JMIR mHealth and uHealth

Journal Impact Factorâ ¢ (JIF) (2023): 5.8
Volume 12 (2024) ISSN 1438-8871 Editor in Chief: Gunther Eysenbach, MD, MPH

Contents

Viewpoints

Reviews

Original Papers

Corrigenda and Addendas

Letter to the Editor

Commentary

Advances and Opportunities of Mobile Health in the Postpandemic Era: Smartphonization of Wearable Devices and Wearable Deviceization of Smartphones

Wonki Hong, PhD

Department of Digital Healthcare, Daejeon University, Daejeon, Republic of Korea

Corresponding Author:
Wonki Hong, PhD
Department of Digital Healthcare, Daejeon University, , Daejeon, , Republic of Korea

Abstract

Mobile health (mHealth) with continuous real-time monitoring is leading the era of digital medical convergence. Wearable devices and smartphones optimized as personalized health management platforms enable disease prediction, prevention, diagnosis, and even treatment. Ubiquitous and accessible medical services offered through mHealth strengthen universal health coverage to facilitate service use without discrimination. This viewpoint investigates the latest trends in mHealth technology, which are comprehensive in terms of form factors and detection targets according to body attachment location and type. Insights and breakthroughs from the perspective of mHealth sensing through a new form factor and sensor-integrated display overcome the problems of existing mHealth by proposing a solution of smartphonization of wearable devices and the wearable deviceization of smartphones. This approach maximizes the infinite potential of stagnant mHealth technology and will present a new milestone leading to the popularization of mHealth. In the postpandemic era, innovative mHealth solutions through the smartphonization of wearable devices and the wearable deviceization of smartphones could become the standard for a new paradigm in the field of digital medicine.

(JMIR Mhealth Uhealth 2024;12:e48803) doi:10.2196/48803

KEYWORDS

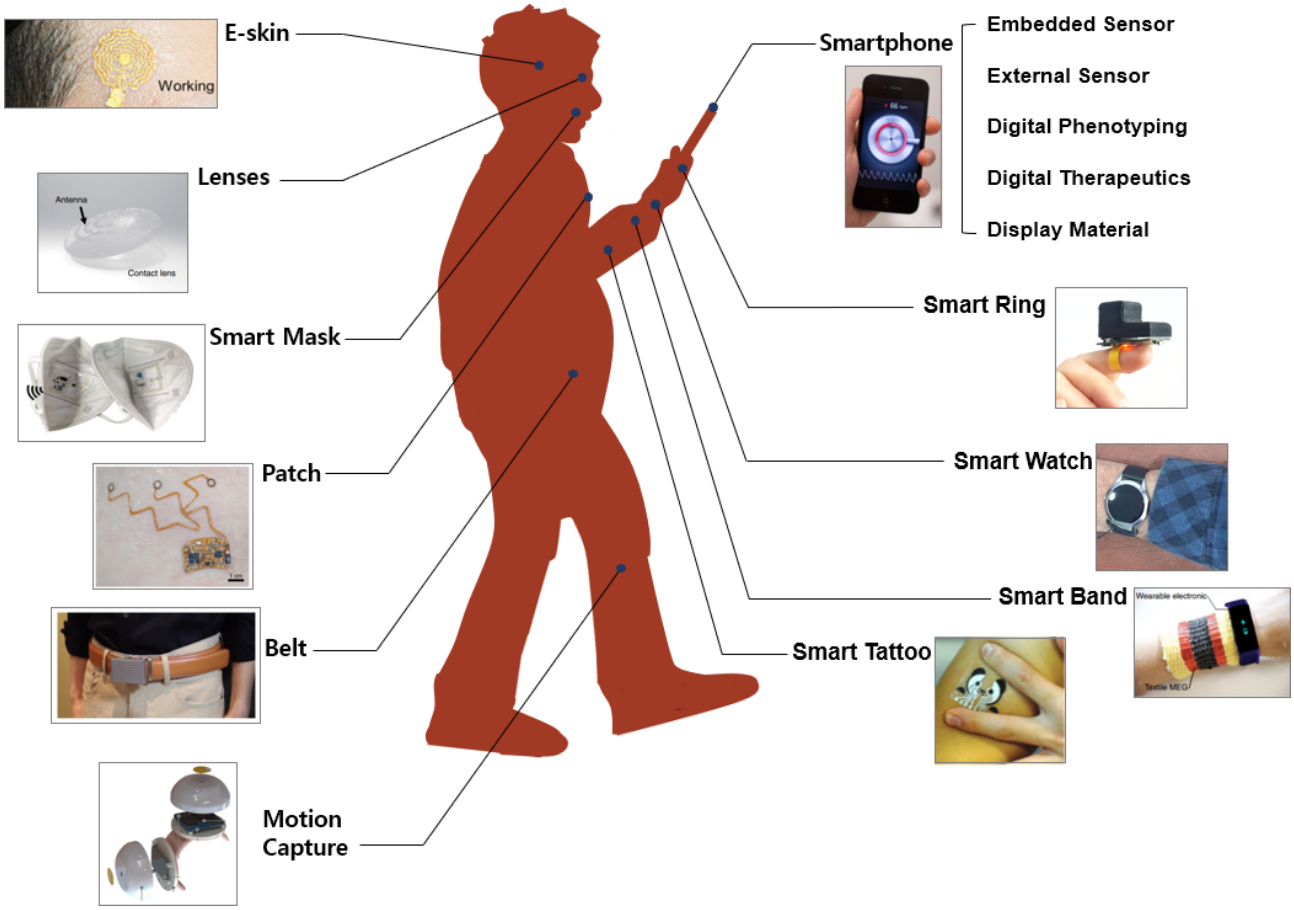
mobile health; mHealth; smartphonization; wearable deviceization; new form factor; sensor-integrated display

Background

In the postpandemic era, the significance of mobile health (mHealth) has been highlighted, and explosive growth in this area is expected to continue [1,2]. Cutting-edge technologies are converging with health care, and mHealth, based on hyperconnected intelligence, is leading the paradigm shift in medical care [3,4]. Many countries have already entered a superaged society, and the proportion of gross domestic product expenditures for medical care is increasing due to an upsurge in the number of people with chronic diseases. In addition, the excessive demand compared to the available supply, the lack of health care infrastructure, and the unbalanced distribution of medical staff are also problems. Therefore, prediction, prevention, and management through artificial intelligence (AI)–based medical big data analysis are required, and for this purpose, ubiquitous and accessible medical services using personalized devices must be provided [5]. mHealth is a strong candidate to make this possible, and the ultimate goal is to dramatically improve the standard and satisfaction of living by providing quality services at affordable prices [6,7].

Wearable electronics and smartphones are representative types of mobile systems optimized for personalized health care sensing. As shown in Figure 1 [8-19], wearable devices that cover the human body and smartphones, a necessity for modern people, enable comprehensive health management in real time.

Figure 1. Application and placement schematic illustration of wearable devices by body part and smartphone for mHealth management. The images were reprinted from Shin et al [8,19], Kim et al [9,19], Escobedo et al [10,19], Hwang et al [11], Nakamura et al [12,19], Hua et al [13,19], Moon et al [14,19], Zhao et al [15,19], Kim et al [16,19], Liu et al [17,19], and Chan et al [18,19]. e-Skin: electronic skin; mHealth: mobile health.



However, the pace of the development and popularization of mHealth technology is progressing more slowly than expected. From the perspective of a paradigm shift from the smartphonization of wearable devices and the wearable deviceization of smartphones, this viewpoint aimed to propose ways to unleash the potentiality of mHealth in the postpandemic era. The smartphonization of wearable devices and the wearable deviceization of smartphones do not simply mean that current smartphones become wearable devices and that current wearable devices maintain the functions of current smartphones. The smartphonization of wearable devices is to completely replace the smartphone function with a wearable device, while upgrading health care performance by embedding the current smartphone's computational power and sensor-integrated display, including large-area panels and user interaction, in the wearable system. In addition, the wearable deviceization of smartphones refers to a change in the form factor so that health care sensing can be performed by switching from the current rigid form to a form that can be attached to a curved skin surface. The new form factor, which features both wearable computer and smartphone functions, will improve detection performance through large-area sensing and increase the penetration rate.

This viewpoint investigated recent trends in health care sensing methods using wearable devices and smartphones, which are the central axis of mHealth. In the case of wearable devices, the form factor for each detailed location on the body and the corresponding detection target technology was described. In the case of smartphones, it covered the detection target and principles of health care according to the application of internal and external sensors, materials, and software. This viewpoint also analyzed the prospects of and current challenges in existing mHealth systems and considered new health care solutions using flexible displays for the convergence form factor of smartphones and wearables. The differentiating point was to consider the direction of mHealth from the perspective of a sensor-integrated and new form factor display. Ultimately, from a display perspective, solutions for the smartphonization of wearable devices and the wearable deviceization of smartphones will provide insight into the health care paradigm shift.

Recent Progress in the Development of Wearable Electronics for Health Care

The primary classification of wearable electronics based on the attachment position can be divided into the face, upper body, limbs, and whole body. Wearable clothes all over the body can also be classified separately.

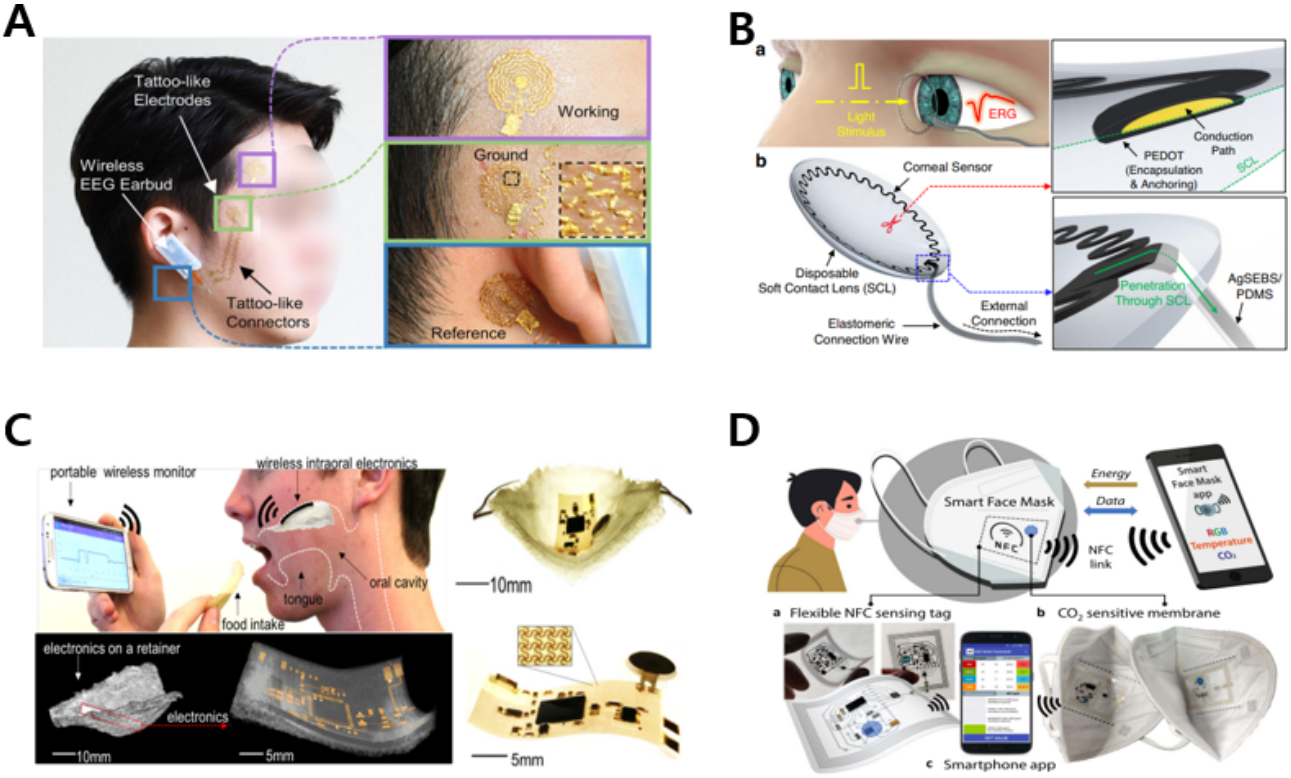
Face

Head

The face, which is closest to the brain, is significant from a sensory point of view because it is where the 5 senses are concentrated. Face-wearable devices with various form factors,

such as bands, caps, headsets, lenses, glasses, tattoos, mouthguards, and masks, may be distributed at each part of the head, eyes, nose, mouth, and ears to sense critical biosignals. In the case of the head, a wearable system that can analyze brain waves and psychological states can be applied [8,20,21]. Figure 2A [8] shows a wireless wearable electroencephalogram (EEG) measurement device based on a tattoo. AI can enhance decision-making by deep learning classification of received EEG data. Namely, it advances the decision performance of AI by feedback through brain waves. Additionally, it would be possible to grasp the degree of brain activation and mental condition of the frontal lobe and temporal lobe through the measurement of biosignals, such as brain waves.

Figure 2. Wearable devices attached to the face for mHealth. (A) Wearable EEG analysis platform with tattoo electrodes for EEG measurement and earbuds for wireless interaction. The images were reprinted from Shin et al [8,19]. (B) Stretchable corneal lenses for ocular electrodiagnosis. The images were reprinted from Kim et al [19,22]. (C) Intraoral electronics for sodium intake analysis through wireless remote control. The images were reprinted from Lee et al [23]. (D) Sensing platform for gaseous CO2 real-time determination inside filtering face piece 2 (FFP2) facemasks. The images were reprinted from Escobedo et al [10,19]. EEG: electroencephalogram; mHealth: mobile health.



Eyes and Nose

System form factors worn on the eye may be divided mainly into lenses and glasses. In the case of lenses, eye health factors, such as glucose, intraocular pressure, and electroretinographic measurements, can be determined using noninvasive methods [9,22,24-26]. For example, a corneal sensor embedded in a disposable soft contact lens can be deployed for electroretinography based on electrochemical anchoring, as shown in Figure 2B [22]. These corneal lenses are functional sensors tailored for ophthalmic electroretinographic testing in human eyes via a user-friendly interface and a design that can be deployed noninvasively. Glasses for health care are prescribed by doctors as an auxiliary tool for surgery and can also analyze the electrolyte and metabolite content of sweat flowing from the head [27-29]. In addition, a wearable system placed on the nose in the form of a nose pad on the glasses can sense the pulse wave, respiratory rate, and electrooculographic measurements [30,31].

Mouth and Ears

Wearable electronics related to the mouth take the form of mouth guards, tooth sensors, and masts and can analyze saliva and nutrients and monitor air quality [10,23,32-35]. For example, a small stretchable circuit and sensor that can be inserted into the human oral cavity may be integrated into a breathable, flexible microporous membrane for a tissue-friendly design, as shown in Figure 2C [23]. Such a device may be used in research to study the prevention of hypertension by facilitating continuous quantification analysis of sodium intake. Figure 2D [10] shows a sensing platform for detecting gaseous CO_2 inside a face mask via stable inorganic phosphors whose luminescence is controlled by a pH indicator. A mask combining a battery-free printed near-field communication (NFC) tag and a photochemical sensor for noninvasive CO_2 measurement was used to achieve detection performance with a resolution of 103 ppm. Practicality in physical activity has been increased through the compensation of the temperature noise and characterized analytical specifications of measurement systems. Moreover, health care wearable systems attached to the ears use earbuds to perform heart rate and sleep monitoring functions [36,37].

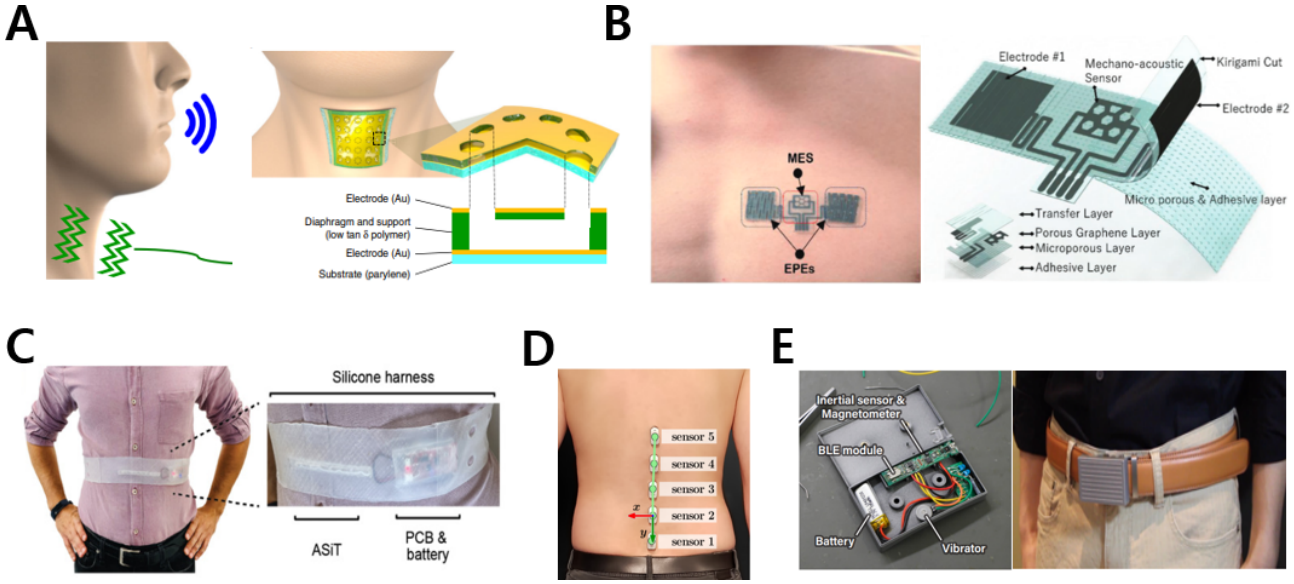
Upper Body

In addition to the face, wearable systems can be applied to the neck, chest, abdomen, internal organs, back, and waist to extract significant health values.

Neck

In the case of the neck, wearable devices with a necklace and patch form factor can record an electrocardiogram (ECG) and voice pressure and monitor the diet through an electroglottogram (EGG) using a neckband [38-40]. For example, a neck-attached wearable device incorporating a cross-linked polymer film and hole-patterned diaphragm structure detects and quantifies voice with an excellent sensitivity of 5.5 V Pa^{-1} over the voice frequency range, as shown in Figure 3A [39]. This device can be used for voice health management and security authentication by eliminating vibration distortions on the curved skin surface through excellent skin compatibility via using ultrathin profiles of ≥5 μm.

Figure 3. Wearable electronics mounted onto the upper body. (A) Vibration-responsive patch for sensing voice pressure. The images were reprinted from Lee et al [19,39]. (B) Epidermal cardiopulmonary patch based on laser fabrication. The images were reprinted from Rachim et al [41]. (C) Air-silicon composite transducer (ASiT) for breathing pattern monitoring. The images were reprinted from Cotur et al [19,42]. (D) Spine tracker sensor system. The image was reprinted from Stollenwerk et al [19,43]. (E) A belt for waistline measurement. The images were reprinted from Nakamura et al [12,19]. EPE: electrophysiological electrode; MES: mechano-acoustic sensor; PCB: printed circuit board.



Thorax

Thorax-related wearable electronics, such as patches, chest belts, and brassieres, enable ECG recording, temperature measurement, sleep monitoring, posture analysis, and galvanic skin response (GSR) assessment [11,41,44-48]. Figure 3B [41] shows a sensor designed for continuous monitoring of the cardiopulmonary biosignal via a CO_2 laser–based manufacturing process. The epidermal patch consisting of a mechanoacoustic sensor and electrophysiological electrodes provides advanced functionality through a gas-permeable and biocompatible layer.

Abdomen

Abdomen-attached mHealth systems can sense glucose and breathing patterns through patches and straps [42,49]. For instance, an air-silicon composite transducer monitors respiratory activity by continuously measuring the force applied to the air channel embedded in the silicon-based elastomer, as shown in Figure 3C [42]. The system, which uses a pressure sensor and mixed-signal radio electronics, follows the principle of sensing the air pressure change inside the channel when breathing force is applied to the transducer surface. In particular, tactile sensing, including pressure sensing, is critical in health care. This is because tactile sensors attached to the skin detect physical stimuli, such as breathing patterns, heart rate, pulse, muscle activity, and body temperature, linked to biological signals. Skin, the most widely distributed organ among the five sense organs in the human body, is a tactile sensor with receptors that detect pressure, delicate movements, and temperature and is also an actuating organ that emits the same physical stimulation. Flexibility is a crucial element for the tactile sensor to be conformally attached to the skin to detect minute physical changes in detail and increase user convenience [50-52].

Furthermore, digestible pills check medication compliance. Management of medication adherence can prevent patients with severe mental illness from experiencing relapses and hospitalizations [53]. In addition, capsule endoscopy can monitor the colon health or bladder pressure state [54,55].

Back

A wearable system attached to the back can be used to analyze changes in the spine's shape during training. A spine tracker device shown in Figure 3D consists of 5 sensors, with each sensor attached to the lumbar spine, and can correct posture by providing real-time feedback [43].

Waist

In addition, a waist belt can be useful for obesity management [12,56]. The belt automatically measures waist circumference with high accuracy, with an F_1-score of 0.95, and monitors the daily lifestyle using a magnetometer, an accelerometer, and a gyroscope, as shown in Figure 3E [12].

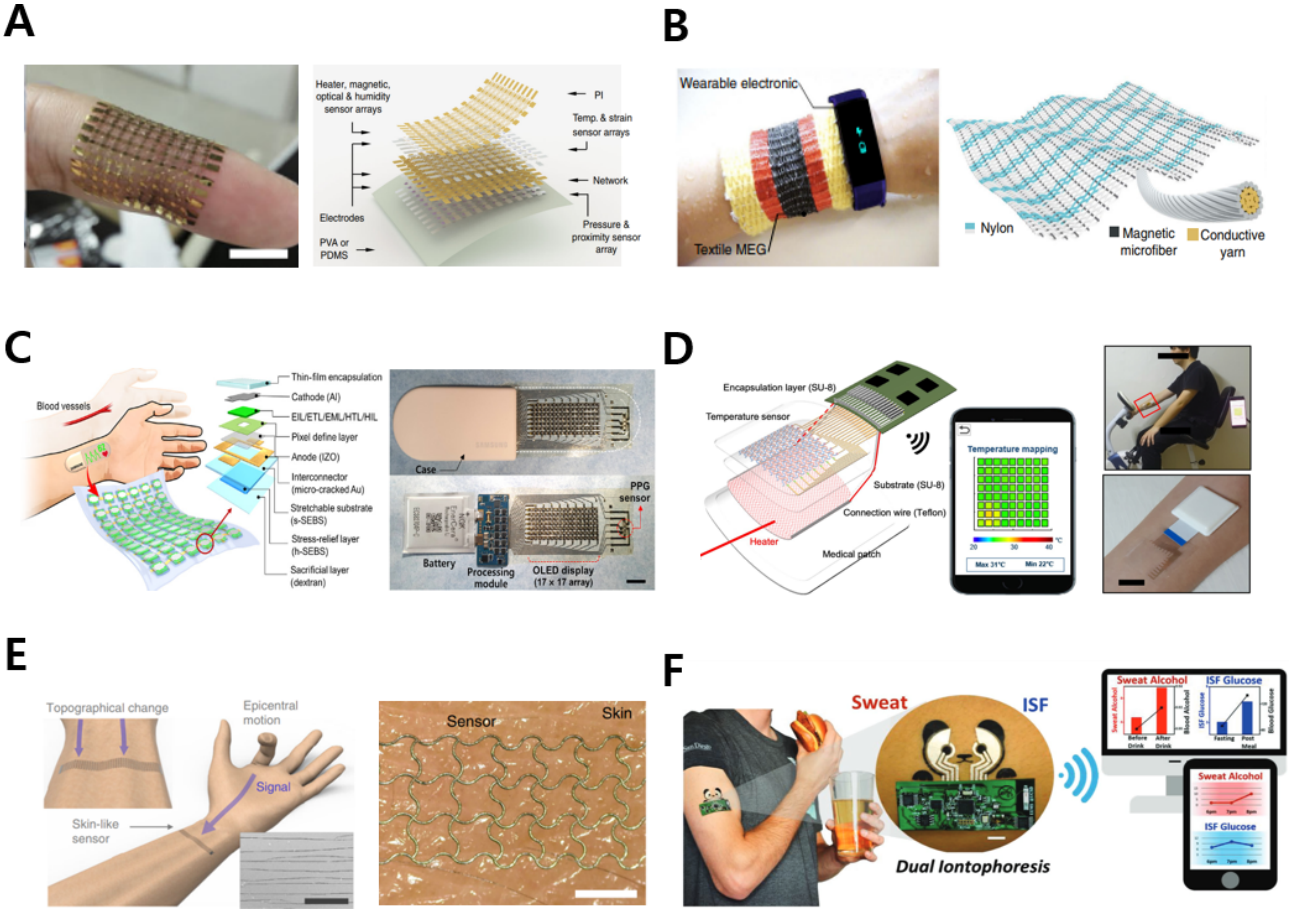
Limbs

In the case of the limbs, the main categories include the hands, arms, legs, and feet by attachment location.

Hands

The measurable health factors in a hand-related wearable device, such as a patch, ring, or glove, include rehabilitation evaluation analysis, ECG characteristics, oxygen saturation, dietary monitoring, pulse wave, and temperature [13,57-61]. For instance, a multisensory electronic skin integrated into a polyimide network simultaneously detects physical properties, such as temperature, strain, humidity, light, magnetic field, pressure, and proximity, in real time, as shown in Figure 4A [13]. It can also be used for rehabilitation evaluation using personalized intelligent prostheses.

Figure 4. Wearable devices attached to the hands and arms. (A) Stretchable and conformable electronic skin for multifunctional sensing. The images were reprinted from Hua et al [13,19]. (B) Power generation textile for wearable health care. The images were reprinted from Zhao et al [15,19]. (C) Stand-alone patch for health monitoring based on a stretchable organic optoelectronic system. The images were reprinted from Lee et al [62]. (D) Thermal patch for self-care treatment through temperature distribution sensing and thermotherapy based on wireless graphene. The images were reprinted from Kang et al [63]. (E) Sensor conformably attached to skin decoding epicentral human motions. The images were reprinted from Kim et al [19,64]. (F) A single wearable biosensor platform that simultaneously monitors sweat and interstitial fluid (ISF). The images were reprinted from Kim et al [16,19]. MEG: magnetoelastic generator; OLED: organic light-emitting diode; PDMS: polydimethylsiloxane; PI: polyimide; PPG: photoplethysmogram; PVA: polyvinyl alcohol; Temp.: temperature.



Arms

mHealth systems of various form factors related to the arm can also be useful for health management. Among them, wristwatches, bands, and bracelet devices can detect health factors, such as the heart rate, oxygen saturation, number of steps, blood pressure, ECG characteristics, glucose, blood sugar, and sweat metabolites [14,15,65-75]. Figure 4B [15] shows a magnetoelastic generator that provides the power to drive a wearable biosensor system. This generator can help measure cardiovascular parameters underwater without encapsulation for telemedicine and has excellent water vapor transmission characteristics.

A patch sensor attached to the arm can measure the pH, sweat rate, lactate, heart rate, temperature, electromyogram (EMG) and ECG characteristics, blood pressure, and water content and can also be applied for wound treatment and rehabilitation evaluation [62-64,76-85]. For instance, a stand-alone organic skin patch for health care with an organic light-emitting display with sufficient pixels reports the heart rate via a stretchable photoplethysmogram (PPG) sensor, as shown in Figure 4C [62]. An ultrathin patch of 15 µm is configured on a soft elastomer substrate and can operate stably at 30% strain using a combination of a stress relief layer and deformable microcracks. Figure 4D [63] shows a wireless graphene patch that simultaneously provides thermal sensing and thermotherapy capabilities. This thermal patch consists of a graphene-based capacitive sensor, a graphene thermal pad, and a flexible

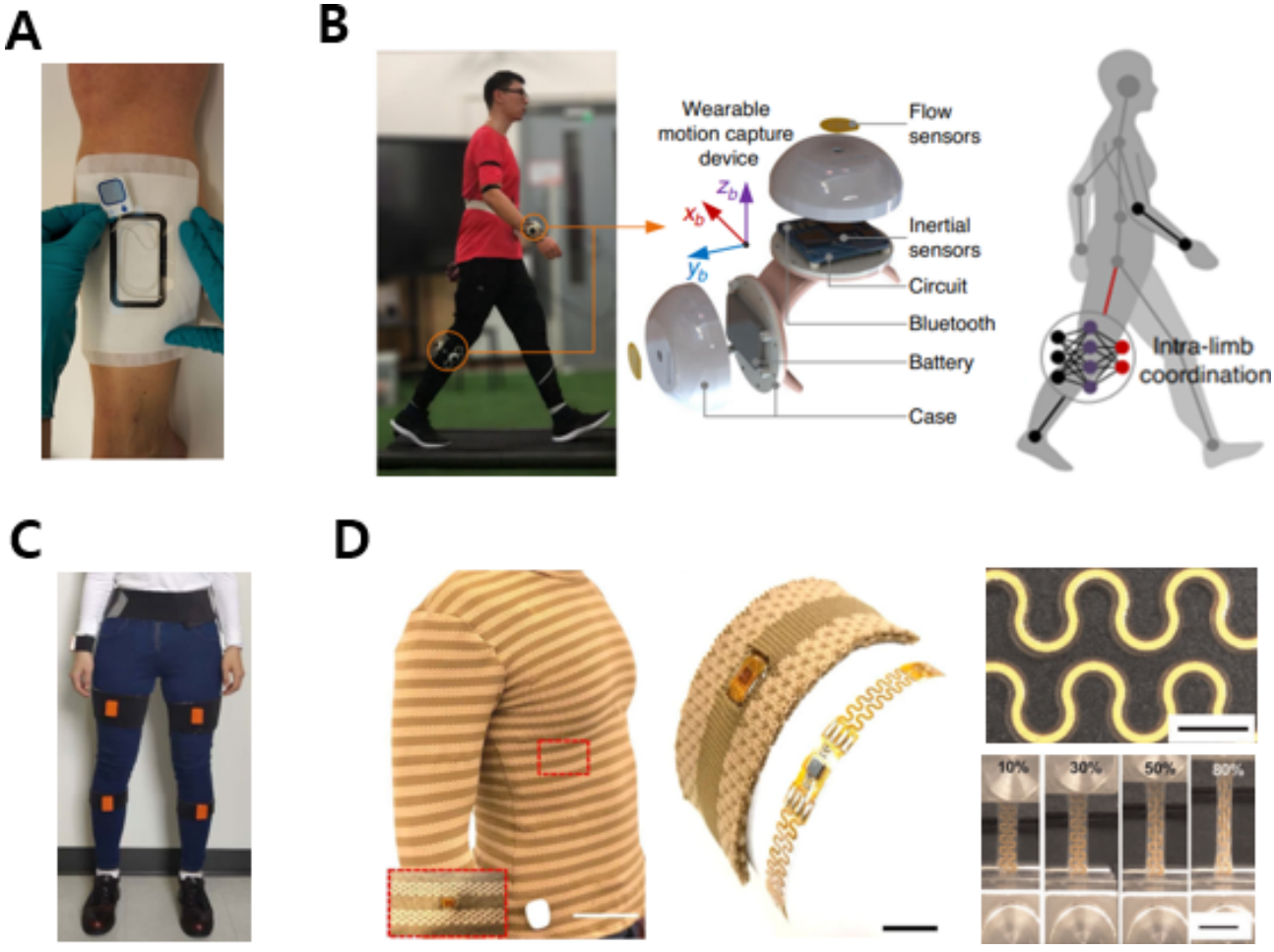
wireless communication module to continuously monitor temperature changes with high resolution and sensitivity and perform thermal treatment through a graphene-based heater. Beyond the existing complex multisensor structure, skin patches alone may decode movements of 5-finger gestures by detecting microdeformation using the laser-induced crack structure, as shown in Figure 4E [64]. Based on the same principle, it can be attached to various body parts to track physical movements.

Furthermore, ECG, EMG, temperature, sweat, and interstitial fluid analyses can be performed following health care monitoring through arm tattoos [16,86]. For instance, a noninvasive epidermal biosensing system includes physically separated electrochemical biosensors for the extraction of interstitial fluid at the cathode and sweat stimulus extraction at the anode, as shown in Figure 4F [16]. Namely, this biomarker monitoring system is a single wearable epidermal platform that simultaneously samples and analyses different biofluids.

Legs and Feet

Figure 5 describes a wearable health care device that may be applied to the legs, feet, or whole body. The mobile form factors applicable to the legs include patches, wearable robots, and straps, which perform moisture analysis at the wound area, gait analysis, ECG measurement, and rehabilitation evaluation [17,87-92]. For instance, appropriate dressing changes for exudative wounds are essential. Using a moisture sensor mounted on the bandage, as shown in Figure 5A [89], the change in the amount of dressing on the wound can be detected and the replacement time determined, increasing patient convenience. A motion capture device can accurately measure the movement of limbs during daily activities, strenuous exercise, and long-term exercise, as shown in Figure 5B [17]. Existing drift and instability problems are solved by integrating microtriaxis inertial and microtriaxis flow sensors. Additionally, it is possible to evaluate gait performance on irregular and uneven surfaces using a wearable sensor in the form of a strap with 6 inertial measurement units (IMUs) and an analysis algorithm, as shown in Figure 5C [92]. It is possible to implement edema measurement, gait analysis, and ulcer detection via plantar pressure analysis using wearable sensors attached to the shoes, socks, or soles of the feet [93-97].

Figure 5. mHealth apps for the legs and the whole body. (A) Moisture sensor for exudative wounds. The image was reprinted from Henricson et al [19,89]. (B) A motion capture device capable of detecting limb movements with high accuracy. The images were reprinted from Liu et al [17,19]. (C) Wearable strap sensor for gait analysis. The image was reprinted from Luo et al [19,92]. (D) An electronic textile conformable suit for distributed sensing wirelessly. The images were reprinted from Wicaksono et al [19,98]. mHealth: mobile health.



Whole Body

Furthermore, clothes worn on the whole body are also a type of wearable device. Figure 5D [98] shows a personalized and conformable suit of an electronics-based textile for multimodal health care sensing. The platform's elasticity ensures intimate contact between the electronic device and the skin, and it can detect the skin temperature, heart rate, and respiration with high accuracy and precision. The suit with electronic textiles can measure the body temperature, respiratory rate, heart rate, oxygen saturation, and EMG and ECG characteristics and can also perform phototherapy [98-103]. As described before, form factor and detection targets by body part on wearable devices are summarized in Tables 1-4.

Table . Summary of form factor and detection targets on wearable devices for the face.

Body position and form factor	Target(s) of detection	Reference(s)
Head		
Tattoo	EEG[a]	[8]
Band, cap, headset	Mental stress through EEG	[20]
Band, cap, headset	EEG	[21]
Eyes		
Lenses	Glucose	[24,25]
Lenses	Intraocular pressure	[9,26]
Lenses	Electroretinogram	[22]
Glasses	Auxiliary surgical tool	[27,28]
Glasses	Sweat electrolytes, metabolites	[29]
Nose		
Nose pad	Pulse wave, respiration rate	[30]
Nose pad	Electrooculogram	[31]
Mouth		
Mouthguard	Saliva monitoring	[32-34]
Mouthguard	Nutrition analysis	[23]
Tooth sensor	Nutrition analysis	[35]
Mask	Air quality monitoring	[10]
Ears		
Earbuds	Heart rate	[36]
Earbuds	Sleep monitoring using EEG	[37]

[a]EEG: electroencephalogram.

Table . Summary of form factor and detection targets on wearable devices for the upper body.

Body position and form factor	Target(s) of detection	Reference(s)
Neck		
Necklace	ECG[a]	[38]
Patch	Voice pressure	[39]
Band	EGG[b]	[40]
Thorax		
Patch	ECG	[11,41,44]
Patch	ECG, temperature	[45]
Patch	Sleep monitoring	[46]
Chest belt	Trunk posture	[47]
Brassiere	Galvanic skin response	[48]
Abdomen		
Patch	Glucose	[49]
Strap	Respiratory patterns	[42,50-52]
Internal organs		
Ingestible pill/capsule	Medication compliance	[53]
Ingestible pill/capsule	Intravesical pressure and colon monitoring	[54,55]
Back		
Strap	Spine monitoring	[43]
Waist		
Belt	Obesity management	[12,56]

[a]ECG: electrocardiogram.

[b]EGG: electroglottogram.

Table . Summary of form factor and detection targets on wearable devices for the limbs.

Body position and form factor	Target(s) of detection	Reference(s)
Hands		
Patch	Rehabilitation	[13]
Ring	ECG[a]	[57]
Ring	SpO_2[b]	[58]
Ring	Dietary management	[59]
Ring	Pulse wave, temperature	[60]
Glove	Rehabilitation	[61]
Wrist		
Watch/band/bracelet	Heart rate, step number	[65]
Watch/band/bracelet	SpO_2	[66]
Watch/band/bracelet	SpO_2, heart rate, energy expenditure	[67,68]
Watch/band/bracelet	Blood pressure	[14,69]
Watch/band/bracelet	Pulse management	[15]
Watch/band/bracelet	ECG	[70,71]
Watch/band/bracelet	Diagnosis of Parkinson disease	[72]
Watch/band/bracelet	Glucose	[73,74]
Watch/band/bracelet	Sweat metabolites (glucose, lactate)	[75]
Patch	Sweat rate, pH, lactate, glucose, chloride	[76]
Patch	Heart rate	[62]
Patch	Wound management	[77,78]
Patch	Temperature, thermotherapy	[63]
Patch	ECG, EMG[c]	[79-81]
Patch	EMG	[82,83]
Patch	Blood pressure, skin hydration, temperature	[84]
Patch	Biometrics	[85]
Patch	Rehabilitation	[64]
Tattoo	ECG, EMG, temperature	[86]
Tattoo	Sweat and Interstitial fluid analysis	[16]
Legs		
Patch	ECG	[87]
Patch	Moisture analysis at the wound area	[88,89]
Wearable robot	Rehabilitation	[17,90,91]
Strap	Gait analysis	[92]
Feet		
Patch	Edema	[93]
Shoes	Gait analysis	[94-96]
Socks	Foot pressure ulcer	[97]

[a]ECG: electrocardiogram.

[b]SpO_2:oxygen saturation.

[c]EMG: electromyogram.

Table . Summary of form factor and detection targets on wearable devices for the whole body (clothes using electronic textiles).

Target(s) of detection	Reference(s)
Temperature, respiration, heart rate	[98]
SpO_2[a], heart rate, temperature	[99]
Phototherapy, temperature, heart rate	[100,101]
EMG[b]	[102]
ECG[c]	[103]

[a]SpO_2:oxygen saturation.

[b]EMG: electromyogram.

[c]ECG: electrocardiogram.

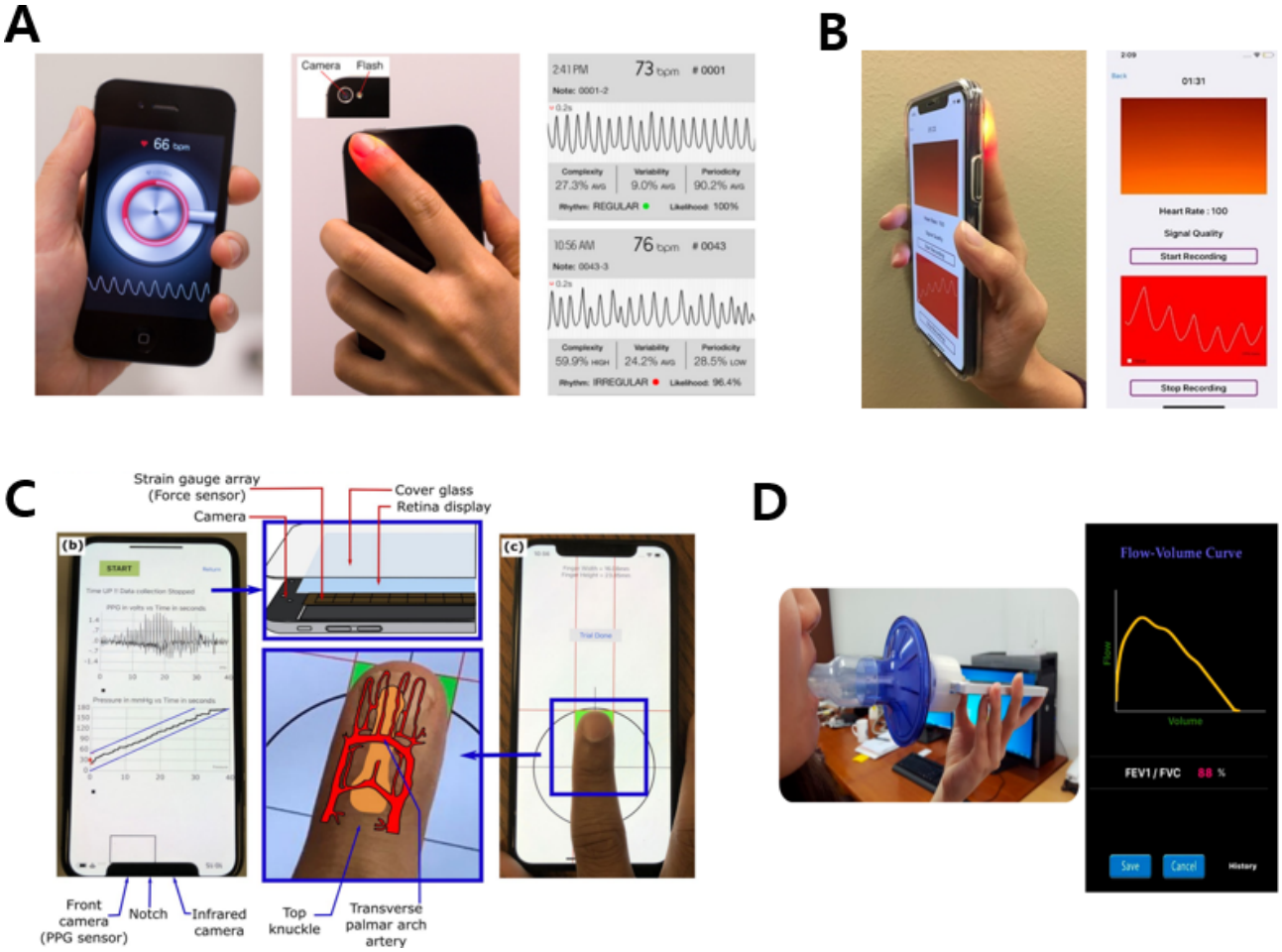
Recent Progress in the Development of Smartphone-Based Health Care Apps

In addition to wearable devices, health care delivery is also possible using smartphones through built-in sensors, smartphone-interlocked gadgets, display-related materials, and apps.

CMOS Only

Smartphones have built-in 20-30 sensors; in particular, complementary metal-oxide-semiconductor (CMOS) image sensors may be used to monitor heart, eye, and skin-related diseases [18,104-109]. As shown in Figure 6A [18], the atrial fibrillation screening ability using PPG pulse analysis based on a smartphone camera and a commercialized app showed a similar performance level to that of patches used for single-lead ECG monitoring. It has been proven that prodromal stroke symptoms can be detected using only a smartphone in a primary care setting. In addition, the fingertip motion signal and color intensity signal, both heterogeneous signals, are acquired and analyzed using a camera to remove finger movement and optical noise, as shown in Figure 6B [104]. In this way, a clean heart rhythm signal with high accuracy can be extracted via smartphone monitoring, while minimizing noise artifacts.

Figure 6. Health care apps using built-in smartphone sensors. (A) Smartphone built-in camera and app-based atrial fibrillation diagnosis. The images were reprinted from Chan et al [18,19]. (B) Heart rhythm analysis using CMOS image sensor. The images were reprinted from Tabei et al [19,104]. (C) Smartphone-based blood pressure measurement through the oscillometric finger-pressing method. The images were reprinted from Chandrasekhar et al [19,110]. (D) Set and acquisition graph of smartphone and 3D-printed mouthpiece adapter for spirometry. The images were reprinted from Thap et al [19,111]. CMOS: complementary metal-oxide-semiconductor; PPG: photoplethysmogram.



Hybrid Including CMOS

New functions, such as blood pressure measurement and temperature and dietary monitoring, can be established by combining pressure sensors, temperature sensors, and the phone microphone instead of CMOS alone [110,112-115]. For instance, as shown in Figure 6C [110], absolute blood pressure is measured via a blood flow oscillometric signal through finger pressure using a strain gauge on the front of the smartphone, in addition to CMOS. A light-emitting display may also be added to this, so it is possible to measure blood pressure ultimately with pure smartphone components.

IMU/Microphone/Ultrasonic Sensor

In addition, sleep position monitoring and treatment can be performed by detecting body movements through an IMU of the smartphone, and the gait of patients with Parkinson disease can also be analyzed [116-118]. The smartphone's built-in microphone sensor can also assess lung capacity and breathing sounds and monitor sleep [111,119-121]. Figure 6D [111] reports lung capacity and function parameter measurements following smartphone microphone–based, high-resolution time-frequency spectral analysis. A moisture-resistant ultrasonic sensor using polyvinylidene fluoride can be used for biometric authentication through fingerprinting [122].

Touch Sensor/Digitizer

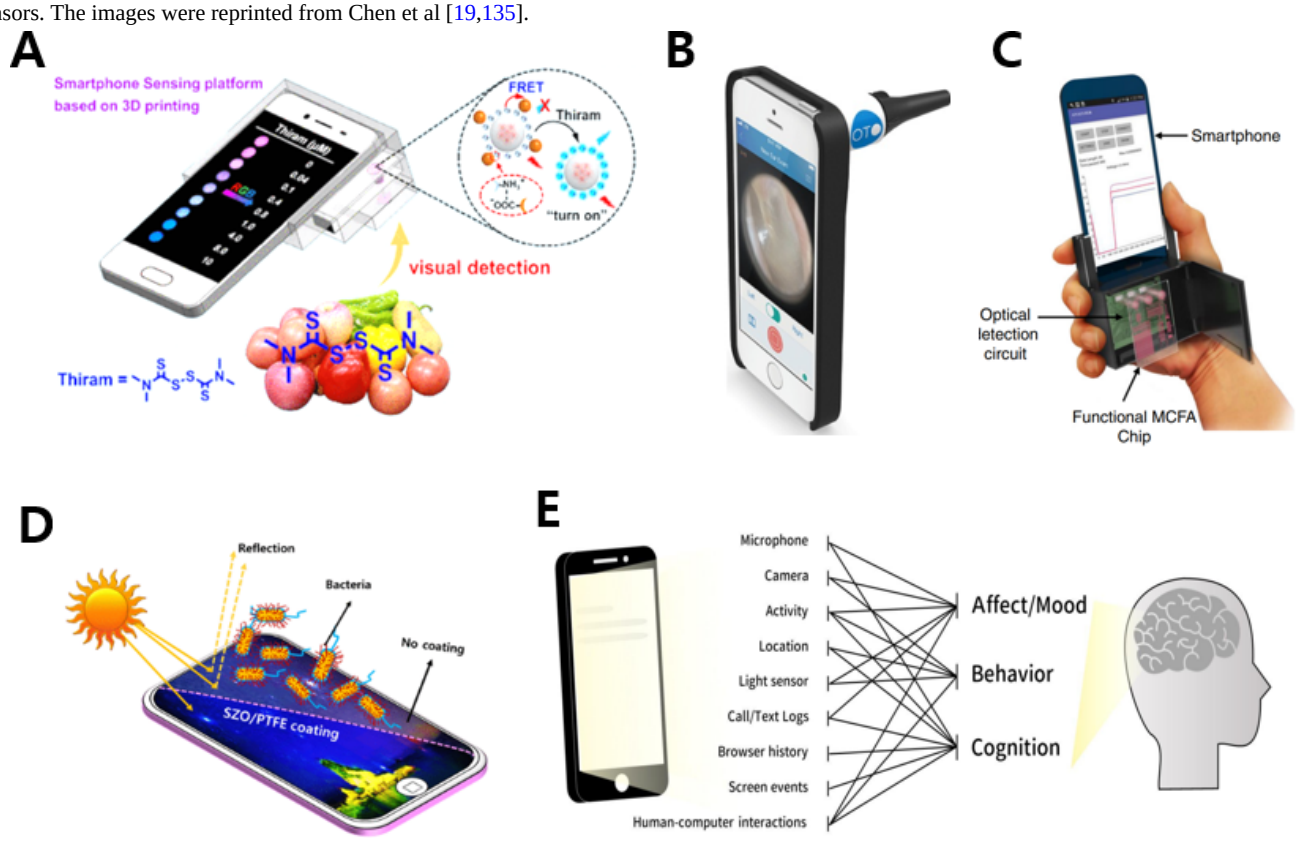
Moreover, general user interfaces, such as a touch sensor and digitizer, can also be used for health care purposes. For example, the heart rate can be checked by assessing capacitance changes according to the heartbeat with a capacitive touch sensor. The touch sensor is also helpful in diagnosing Parkinson disease through touch accuracy analysis [123,124]. In addition, a digitizer for writing can be applied to biometric authentication through handwritten signature recognition [125,126].

Interlocked Gadgets

There is a case of combining various mHealth sensing techniques, such as pesticide analysis, otitis media diagnosis, malaria infection detection, and ECG measurement, by adding a separate gadget rather than using just the smartphone itself [127-133]. The platform shown in Figure 7A [128] performs a visual, quantitative analysis of pesticides using an optical system that combines a dark cavity and an ultraviolet lamp with a smartphone. In other words, integrating a smartphone and a gadget-based paper strip enables real-time and on-site food evaluation. Additionally, it was confirmed that the diagnosis of acute otitis media is possible with the same level of accuracy as that attained with existing otoscopes through the combination of a commercialized optical system and a camera in a

smartphone, as shown in Figure 7B [130]. Figure 7C [132] shows a smartphone-based immunodiagnostic platform that performs a chemiluminescence-based enzyme-linked immunosorbent assay using a lyophilized chemiluminescence reagent. This hand-held point-of-care-testing analyzer can detect active malaria infections with a sensitivity of 8 ng/mL.

Figure 7. Health care apps using gadgets mounted on smartphones. (A) Smartphone platform for pesticide evaluation of food, integrated with an ultraviolet lamp and a dark cavity by 3D printing. The images were reprinted from Chu et al [128]. (B) Smartphone otoscope for diagnosis of acute otitis media. The images were reprinted from Mousseau et al [130]. (C) Smartphone-based immunodiagnosis using microfluidic assays. The images were reprinted from Ghosh et al [19,132]. (D) Antibacterial touchscreen for preventing contamination. The images were reprinted from Ippili et al [134]. (E) Digital biomarkers that reflect users' moods, behaviors, and cognitions using text logs, browser history, human-computer interactions, and various sensors. The images were reprinted from Chen et al [19,135].



Display Materials

Health care delivery can also be achieved through materials used in manufacturing smartphones, such as window coatings for antireflection and display processes. An ecofriendly antibacterial coating with Zn-doped silicon oxide thin films can prevent infectious diseases caused by microbial contamination of touch events, as shown in Figure 7D [134]. In addition, it is possible to reduce the deformation of retinal cells by decreasing the blue light of the display through the material development of organic light-emitting or color filters [136].

Apps

Furthermore, health care sensing is possible through apps incorporating digital phenotypes and digital therapeutics [135,137-140]. A digital phenotype refers to a disease or health condition that is unintentionally reflected in patterns of use of digital devices. Mobile apps can collect human-smartphone interaction data to monitor smartphone usage and construct long-term patterns and trend changes. As shown in Figure 7E [135], analyzing a digital biomarker that reflects human effects, moods, behaviors, and cognition can predict psychiatric conditions, such as depression and smartphone addiction. In addition, digital therapeutics delivered through games, education, coaching, and counselling are based on cognitive behavioral therapy and can treat insomnia, alcohol addiction, drug addiction, panic disorder, and attention deficit hyperactivity disorder. Additionally, it effectively improves physical diseases, such as obesity and high blood glucose. Table 5 summarizes the sensing methods and targets using smartphones.

Table . Summary of sensing methods and targets using smartphones.

Type and sensing methods	Target(s) of detection	Reference(s)
Built-in sensors		
CMOS[a]	Atrial fibrillation	[18]
CMOS	Heart rate	[104,105]
CMOS	Diabetic retinopathy	[106]
CMOS	Skin cancer	[107-109]
CMOS + microphone	Heart rate, SpO_2[b], blood pressure	[112]
CMOS + microphone + speaker	Diet management	[113]
CMOS + strain gauge + display	Blood pressure	[110,114]
CMOS + temperature sensor	Temperature, heart rate	[115]
IMU[c]	Sleep monitoring	[116,117]
IMU	Gait analysis	[118]
Microphone	Spirometry	[111,119]
Microphone	Breathing sound analysis	[120]
Microphone	Sleep monitoring	[121]
Ultrasonic sensor	Biometric using fingerprint	[122]
Touch sensor	Heart rate	[123]
Touch sensor	Parkinson disease	[124]
Digitizer	Biometrics using signature	[125,126]
Gadgets interlocked with smartphones		
Optical platform	Pesticide evaluation in food	[127-129]
Smartphone CMOS + lens	Otoscopy	[130,131]
Microfluidic platform	Malaria infection	[132]
Patch electrode	ECG[d]	[133]
Materials		
Window coating	Antibacterial	[134]
Light emitting	Blocking of blue light	[136]
Apps		
Digital phenotyping	Addiction, attention deficit hyperactivity disorder	[135,137,138]
Digital therapeutics	Mental health	[139,140]

[a]CMOS: complementary metal-oxide-semiconductor.

[b]SpO_2:oxygen saturation.

[c]IMU: inertial measurement unit.

[d]ECG: electrocardiogram.

Prospects for mHealth

The industry of mHealth is expected to grow explosively in the future. In particular, the third generation of medicine and therapies that rely on novel solutions are emerging beyond the existing state of mHealth. Among them, bioelectronic medicine is a nonpharmacological treatment category that stimulates nerve functions with energy, such as electricity, light, and ultrasonic waves. This approach uses an electronic device that controls metabolic function to maintain homeostasis by regulating hormones [141]. To date, electroceuticals have been used for obesity, asthma, sleep apnea, brain tumors, epilepsy, and Parkinson disease and have shown substantial and significant therapeutic effects [142-144]. It is also one of the most innovative fields in medicine because it has significant advantages when considering the development time and cost of existing drugs.

Using digital therapeutics, also referred to as “software as a medical device,” it is possible to manage and treat not only physical diseases but also psychiatric conditions, such as posttraumatic stress disorder and schizophrenia [145,146]. It is

of great significance in terms of patient convenience that personal and sensitive mental health conditions can be diagnosed in real life, not in hospitals, through digital phenotypic analysis, such as smartphone usage patterns and uploaded social networking service (SNS) content.

From the point of view of the wearable form factor, since much of health care sensing is possible on the wrist, the smartwatch is currently playing a pivotal role in health care. The finger (as well as the wrist) is a body part to focus on as it can be used to assess health factors, such as the heart rate, oxygen saturation, ECG characteristics, blood pressure, blood sugar, biometric authentication, body temperature, and dietary monitoring. Therefore, it is expected that in the future, the ring type of device for health care will pair with the smartwatch as the 2 main pillars.

Challenges for mHealth

This viewpoint investigated comprehensive health care sensing technology using wearable electronics and smartphones. However, mHealth is less widely used than expected, unfortunately. Wearable devices are relatively more optimized for continuous and real-time health care sensing compared to smartphones [147,148]. However, the penetration rate compared to smartphones worldwide is sluggish [149-151]. A smartwatch, a representative wearable device, needs to be connected to a smartphone to operate, so users do not recognize the wearable device as an independent entity. Independent use is required to be fully positioned as a separate device. These devices lack effectiveness due to reduced user convenience because of their small screens, poor battery performance, low usage rate, clunky design, and high price. Wearable devices are recognized as a kind of subdevice rather than an essential and leading product because they do not have as much impact as smartphones. Therefore, in the case of wearable devices, innovative solutions are required to make them universal necessities for human beings, such as smartphones.

However, in the case of smartphones, the penetration rate is high worldwide, including low- and middle-income countries [152]. In the case of current smartphones, the fundamental value in terms of user experience as well as utility is high. However, it is not such a great solution from the perspective of health care. It is challenging to conduct biosignal sensing using a smartphone while being in close contact with human skin all day long, so it is challenging to implement continuous real-time big data–based predictive and preventive medical care using smartphones from the health care perspective. Smartphones desperately require a breakthrough that can allow them to monitor health in real time continuously, 24 hours a day, through a form more closely adherent to the skin, while maintaining the current phone function.

Breakthroughs for mHealth

The display is a crucial component of a health care system. In other words, smartphones and wearable devices, as central axes of the mHealth system, are inseparable from their displays. In addition, displays and sensors in mobile devices are closely related. To improve the convenience of user interaction, the proportion of the active area of mobile displays is increasing. However, the increase in the active area has a limitation that reduces the sensing performance, including sensitivity. To overcome this, the upper part of the sensor covers the display by lowering the resolution of the display to prevent the deterioration of the sensing transmittance. A typical example is under-panel camera (UPC) technology that covers the camera with the display by reducing the display resolution on the top of the CMOS image sensor to increase light transmittance.

Sensor-Integrated Display Solution

However, the ultimate and ideal method is a sensor-integrated display solution. A sensor-integrated display has many advantages from a health care sensing point of view. This is because (1) many mHealth sensors use an optical approach, (2) it is relatively easy to manufacture large-area sensors, and (3) the application of a new form factor display can lead to an increase in the body contact area.

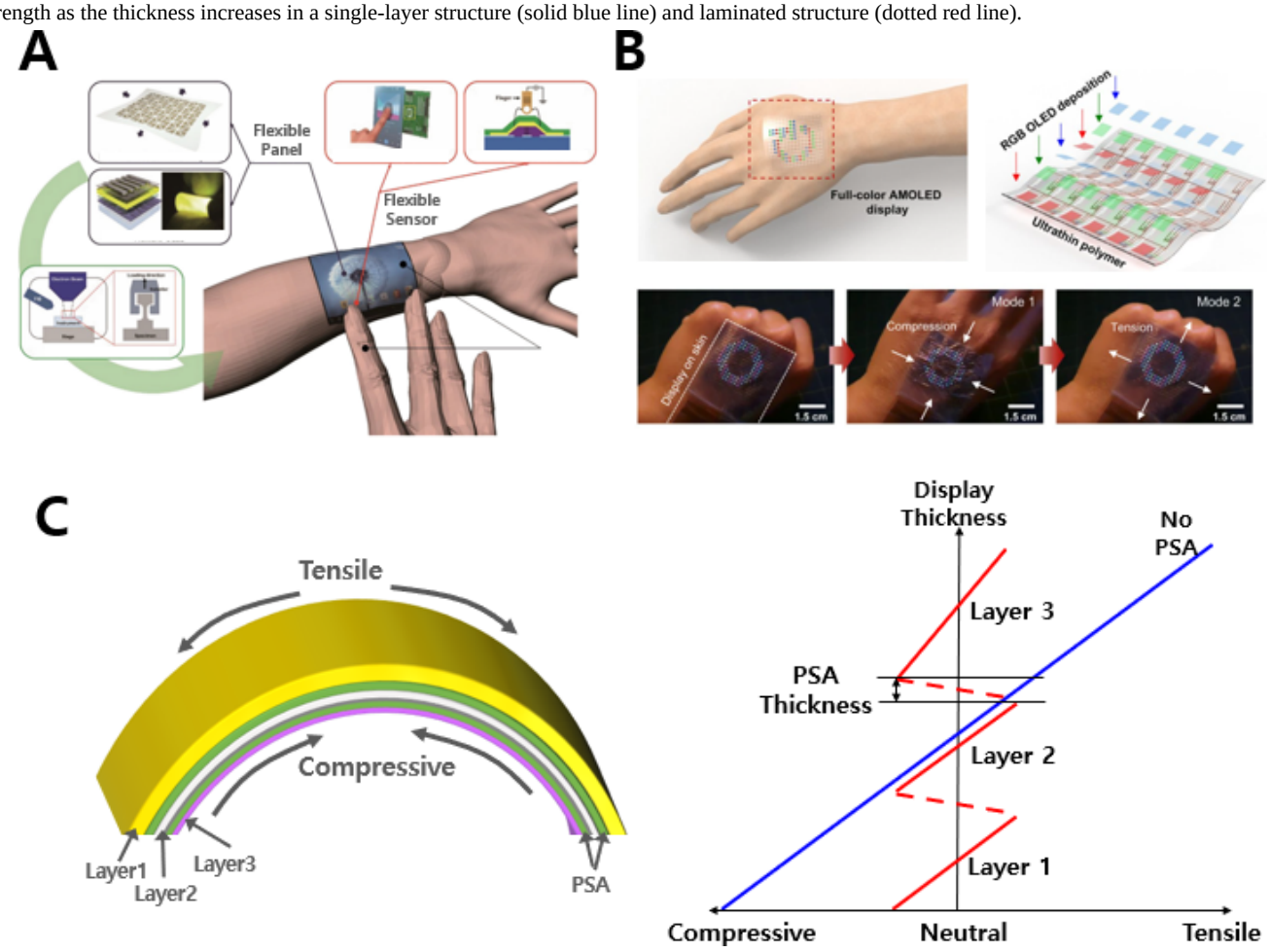
First, the majority of mHealth sensing approaches are optical methods. Various health care parameters, such as the heart rate, oxygen saturation, blood pressure, blood sugar, body temperature, environmental monitoring, and ECH characteristics, can be measured optically. A display is an optical system that already has the means to transmit light. Therefore, a sensor-integrated display could be an optimized health care solution. To implement health care devices using optical systems, in addition to optical transmitters, receiver systems must also be equipped. For advanced performance, the light-emitting wavelength band needs to be expanded and supplemented, including infrared as well as visible light, through the development of materials for the light-emitting layer.

Second, since the sensing area and detection performance are proportional, health care ability can be improved through a sensor embedded in a wide display area. It enables health care sensing in a large area over the entire display area when the built-in optical system is applied, considering design rules. In addition, it is more advantageous for wearability because of a reduction in volume due to the implementation of microlevel thickness because of the sensor-integrated display. Additionally, compared to the number of photomasks needed to manufacture a conventional display, the number of additional photomasks required to implement a display health care system with built-in sensors is far less. It can contribute to popularization due to the low manufacturing price according to the integral type. Ultrathin, low-cost health care devices with relatively simple processes have significant benefits over conventional, bulky, and expensive wearable computers.

Finally, the new form factor device, such as a stretchable sensor-integrated display, increases the area of contact with the body and improves detection capability through health care sensing in close contact with the skin. Flexible panels with user convenience could be applied to the human skin, considering ergonomic factors [153-159]. The flexibility of not only the active matrix backplane and core of the panel but also the touch sensor, fingerprint sensor, and pressure sensor must be ensured, as shown in Figure 8A [153]. In a complete sensor-integrated

display, the flexibility of the backplane allows the sensor part to gain flexibility naturally.

Figure 8. New form factor display and principle. (A) Wearable display with flexible and ultrathin active matrix backplane, touch screen panel, and fingerprint sensor components. The image was reprinted from Park et al [19,153]. (B) A flexible active matrix organic light-emitting diode (AMOLED) with large-area MoS2-based backplane for human skin display. The images were reprinted from Choi et al [154]. (C) A graph of compressive and tensile strength as the thickness increases in a single-layer structure (solid blue line) and laminated structure (dotted red line).



Furthermore, Figure 8B [154] shows a wearable full-color organic light-emitting diode (OLED) display using a 2D material–based backplane transistor suitable for complex skin shapes. The 18×18 thin-film transistor array was fabricated on ultrathin MoS_2 film and then transferred to Al_2O_3 (30 nm)/polyethene terephthalate (6 µm), providing mechanical flexibility beyond conventional OLED technology.

New Form Factor Display

The left picture of Figure 8C simulates a multilayered display, and when this display is bent, tensile strength is applied at the top and compressive strength is applied at the bottom. Assuming that it is formed with only a single layer of the same thickness rather than a laminated structure, extreme tensile and compressive forces occur on the upper and lower surfaces, resulting in cracks in the display, as shown by the solid blue line in the right graph. However, in the stacked structure, a pressure-sensitive adhesive (PSA) between the display layers continues to create new neutral planes, as shown by the dotted red line. In response, the magnitude of the tension and compression force at the top and bottom surfaces does not increase, even if the thickness of the display increases. In other words, using the PSA, it is possible to implement a flexible display without cracks.

No part of the human body is flat. When the health care system and the skin conformally adhere, sensing performance improves. Display technology based on PSA with the harmony of creep and recovery characteristics induces form changes in wearable devices and smartphones. A new form factor with flexibility based on PSA technology that creates a new neutral plane will facilitate a critical conversion of the mHealth system.

Standard of the Medical Paradigm in the Postpandemic Era

A new form factor display for health care with flexibility and display convergence technology using an optical method attaches a large-area health care system to the human skin conformally and continuously detects health care factors in real time, thereby providing a framework for collecting big data. As a result, the existing smartphone becomes a wearable device attached to the body, and the existing wearable device is equipped with smartphone functions suitable for user convenience. Namely, convergence health care technology with the sensor-integrated and new form factor display is an indispensable element that enables the smartphonization of a wearable device and the wearable deviceization of the

smartphone. Of course, health care systems with new form factors and sensor-integrated displays do not solve all mHealth problems. In other words, advances in big data AI software analysis and medical security should go hand in hand with the smartphonization of wearable devices and the wearable deviceization of smartphones. Furthermore, it will be necessary to supplement the medical system policy so that these benefits do not become the exclusive property of the upper class of the economy and so that people from lower social classes can also benefit. Advanced and popularized mHealth system technology could ensure universal health coverage so that everyone can use essential, high-quality medical services without discrimination. In other words, the authentic democratization of health care could become a reality, and a standard for a future health care paradigm in the post-pandemic era could arise.

Conclusion

Personalized platforms, such as wearable devices and smartphones, can be applied to AI-based disease prediction, prevention, and treatment. This viewpoint researched the latest technology trends in mHealth regarding form factors and detection targets according to body attachment location and type. In particular, the sensor convergence technology of the new form factor display provides a framework to analyze health factors in real time by conformally adhering a large-area system to the skin. Innovation in form factors in sensor-integrated displays and convergence health care solutions enable the smartphonization of wearable devices and the wearable deviceization of smartphones. In addition, the strategy for the smartphonization of wearable devices and the wearable deviceization of smartphones can accelerate the development of mHealth, realizing the democratization of medical care so that anyone can use essential services of high quality. Furthermore, it is expected to create a new milestone for the medical paradigm shift in the postpandemic era.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT RS-2022-00165631). This research was also supported by Daejeon University Research Grants (2022).

Authors' Contributions

WH conceived the idea and concept of the viewpoint, created the figure sets, summarized the tables, and wrote the manuscript.

Conflicts of Interest

None declared.

References

1. Weinstein RS, Lopez AM, Joseph BA, et al. Telemedicine, telehealth, and mobile health applications that work: opportunities and barriers. Am J Med 2014 Mar;127(3):183-187. [doi: 10.1016/j.amjmed.2013.09.032] [Medline: 24384059]
2. Figueroa CA, Luo T, Aguilera A, Lyles CR. The need for feminist intersectionality in digital health. Lancet Digit Health 2021 Aug;3(8):e526-e533. [doi: 10.1016/S2589-7500(21)00118-7] [Medline: 34325855]
3. Gunasekeran DV, Tham YC, Ting DSW, Tan GSW, Wong TY. Digital health during COVID-19: lessons from operationalising new models of care in ophthalmology. Lancet Digit Health 2021 Feb;3(2):e124-e134. [doi: 10.1016/S2589-7500(20)30287-9] [Medline: 33509383]
4. Kimberly J, Cronk I. Making value a priority: how this paradigm shift is changing the landscape in health care. Ann N Y Acad Sci 2016 Oct;1381(1):162-167. [doi: 10.1111/nyas.13209] [Medline: 27599092]
5. Bartold PM, Ivanovski S. P4 medicine as a model for precision periodontal care. Clin Oral Investig 2022 Sep;26(9):5517-5533. [doi: 10.1007/s00784-022-04469-y] [Medline: 35344104]
6. Muzny M, Henriksen A, Giordanengo A, et al. Wearable sensors with possibilities for data exchange: analyzing status and needs of different actors in mobile health monitoring systems. Int J Med Inform 2020 Jan;133:104017. [doi: 10.1016/j.ijmedinf.2019.104017] [Medline: 31778885]
7. Mitratza M, Goodale BM, Shagadatova A, et al. The performance of wearable sensors in the detection of SARS-Cov-2 infection: a systematic review. Lancet Digit Health 2022 May;4(5):e370-e383. [doi: 10.1016/S2589-7500(22)00019-X] [Medline: 35461692]
8. Shin JH, Kwon J, Kim JU, et al. Wearable EEG electronics for a brain–AI closed-loop system to enhance autonomous machine decision-making. npj Flex Electron 2022 May 30;6(1):32. [doi: 10.1038/s41528-022-00164-w]
9. Kim J, Kim M, Lee MS, et al. Wearable smart sensor systems integrated on soft contact lenses for wireless ocular diagnostics. Nat Commun 2017 Apr 27;8(1):14997. [doi: 10.1038/ncomms14997]
10. Escobedo P, Fernández-Ramos MD, López-Ruiz N, et al. Smart facemask for wireless CO monitoring. Nat Commun 2022 Jan 10;13(1):72. [doi: 10.1038/s41467-021-27733-3]
11. Hwang W, Kim J, Park S, et al. A Breathable and stretchable metastructure for a versatile hybrid electronic skin patch with long‐term skin comfort. Adv Mater Technol 2022 Jul 28;8(1):2200477. [doi: 10.1002/admt.202200477]

12. Nakamura Y, Matsuda Y, Arakawa Y, Yasumoto K. Waistonbelt X: a belt-type Wearable device with sensing and intervention toward health behavior change. Sensors (Basel) 2019 Oct 22;19(20):4600. [doi: 10.3390/s19204600] [Medline: 31652647]
13. Hua Q, Sun J, Liu H, et al. Skin-inspired highly stretchable and conformable matrix networks for multifunctional sensing. Nat Commun 2018 Jan 16;9(1):244. [doi: 10.1038/s41467-017-02685-9] [Medline: 29339793]
14. Moon JH, Kang MK, Choi CE, Min J, Lee HY, Lim S. Validation of a wearable cuff-less wristwatch-type blood pressure monitoring device. Sci Rep 2020 Nov 4;10(1):19015. [doi: 10.1038/s41598-020-75892-y]
15. Zhao X, Zhou Y, Xu J, et al. Soft fibers with magnetoelasticity for wearable electronics. Nat Commun 2021 Nov 19;12(1):6755. [doi: 10.1038/s41467-021-27066-1]
16. Kim J, Sempionatto JR, Imani S, et al. Simultaneous monitoring of sweat and interstitial fluid using a single wearable biosensor platform. Advanced Science 2018 Oct;5(10):1800880 [FREE Full text] [doi: 10.1002/advs.201800880]
17. Liu S, Zhang J, Zhang Y, Zhu R. A wearable motion capture device able to detect dynamic motion of human limbs. Nat Commun 2020 Nov 5;11(1):5615. [doi: 10.1038/s41467-020-19424-2]
18. Chan PH, Wong CK, Poh YC, et al. Diagnostic performance of a smartphone‐based photoplethysmographic application for atrial fibrillation screening in a primary care setting. J Am Heart Assoc 2016 Jul 21;5(7):e003428. [doi: 10.1161/JAHA.116.003428] [Medline: 27444506]
19. Attribution 4.0 International (CC BY 4.0). Creative Commons. URL: https://creativecommons.org/licenses/by/4.0/ [accessed 2024-01-11]
20. Kamińska D, Smółka K, Zwoliński G. Detection of mental stress through EEG signal in virtual reality environment. Electronics 2021 Nov 18;10(22):2840. [doi: 10.3390/electronics10222840]
21. McDowell K, et al. Biosensor technologies for augmented brain-computer interfaces in the next decades. Proc IEEE 2012 May;100(Special Centennial Issue):1553-1566. [doi: 10.1109/JPROC.2012.2184829]
22. Kim K, Kim HJ, Zhang H, et al. All-printed stretchable corneal sensor on soft contact lenses for noninvasive and painless ocular electrodiagnosis. Nat Commun 2021 Mar 9;12(1):1544. [doi: 10.1038/s41467-021-21916-8]
23. Lee Y, Howe C, Mishra S, et al. Wireless, intraoral hybrid electronics for real-time quantification of sodium intake toward hypertension management. Proc Natl Acad Sci USA 2018 May 22;115(21):5377-5382. [doi: 10.1073/pnas.1719573115]
24. Falk M, Andoralov V, Silow M, Toscano MD, Shleev S. Miniature biofuel cell as a potential power source for glucose-sensing contact lenses. Anal Chem 2013 Jul 2;85(13):6342-6348. [doi: 10.1021/ac4006793] [Medline: 23735164]
25. Yao H, Liao Y, Lingley AR, et al. A contact lens with integrated telecommunication circuit and sensors for wireless and continuous tear glucose monitoring. J Micromech Microeng 2012 Jul 1;22(7):075007. [doi: 10.1088/0960-1317/22/7/075007]
26. Kim J, Park J, Park YG, et al. A soft and transparent contact lens for the wireless quantitative monitoring of intraocular pressure. Nat Biomed Eng 2021 Jul 3;5(7):772-782. [doi: 10.1038/s41551-021-00719-8]
27. Kulak O, Drobysheva A, Wick N, et al. Smart glasses as a surgical pathology grossing tool. Arch Pathol Lab Med 2021 Apr 1;145(4):457-460. [doi: 10.5858/arpa.2020-0090-OA] [Medline: 32823276]
28. Hiranaka T, Nakanishi Y, Fujishiro T, et al. The use of smart glasses for surgical video streaming. Surg Innov 2017 Apr;24(2):151-154. [doi: 10.1177/1553350616685431] [Medline: 28068887]
29. Sempionatto JR, Nakagawa T, Pavinatto A, et al. Eyeglasses based wireless electrolyte and metabolite sensor platform. Lab Chip 2017 May 16;17(10):1834-1842. [doi: 10.1039/c7lc00192d] [Medline: 28470263]
30. Nguyen TV, Ichiki M. MEMS-based sensor for simultaneous measurement of pulse wave and respiration rate. Sensors (Basel) 2019 Nov 13;19(22):4942. [doi: 10.3390/s19224942] [Medline: 31766200]
31. Rostaminia S, Lamson A, Maji S, Rahman T, Ganesan D. W!NCE: unobtrusive sensing of upper facial action units with EOG-based eyewear. Proc ACM Interact Mob Wearable Ubiquitous Technol 2019 Mar 29;3(1):1-26. [doi: 10.1145/3314410]
32. Kim J, Imani S, de Araujo WR, et al. Wearable salivary uric acid mouthguard biosensor with integrated wireless electronics. Biosens Bioelectron 2015 Dec 15;74:1061-1068. [doi: 10.1016/j.bios.2015.07.039] [Medline: 26276541]
33. Arakawa T, Kuroki Y, Nitta H, et al. Mouthguard biosensor with telemetry system for monitoring of saliva glucose: a novel Cavitas sensor. Biosens Bioelectron 2016 Oct 15;84:106-111. [doi: 10.1016/j.bios.2015.12.014] [Medline: 26725934]
34. Moonla C, Lee DH, Rokaya D, et al. Review—lab-in-a-mouth and advanced point-of-care sensing systems: detecting bioinformation from the oral cavity and saliva. ECS Sens Plus 2022 Jun 1;1(2):021603. [doi: 10.1149/2754-2726/ac7533]
35. Tseng P, Napier B, Garbarini L, Kaplan DL, Omenetto FG. Functional, RF-Trilayer sensors for tooth-mounted, wireless monitoring of the oral cavity and food consumption. Adv Mater 2018 May;30(18):e1703257. [doi: 10.1002/adma.201703257] [Medline: 29572979]
36. Boukhayma A, Barison A, Haddad S, Caizzone A. Earbud-embedded micro-power mm-sized optical sensor for accurate heart beat monitoring. IEEE Sensors J 2021 Sep 15;21(18):19967-19977. [doi: 10.1109/JSEN.2021.3098861]
37. Nakamura T, Goverdovsky V, Morrell MJ, Mandic DP. Automatic sleep monitoring using ear-EEG. IEEE J Transl Eng Health Med 2017;5:2800108. [doi: 10.1109/JTEHM.2017.2702558] [Medline: 29018638]
38. Santala OE, Lipponen JA, Jäntti H, et al. Necklace-embedded electrocardiogram for the detection and diagnosis of atrial fibrillation. Clin Cardiol 2021 May;44(5):620-626. [doi: 10.1002/clc.23580] [Medline: 33629410]
39. Lee S, Kim J, Yun I, et al. An Ultrathin conformable vibration-responsive electronic skin for quantitative vocal recognition. Nat Commun 2019 Jun 18;10(1):2468. [doi: 10.1038/s41467-019-10465-w]

40. Farooq M, Fontana JM, Sazonov E. A novel approach for food intake detection using electroglottography. Physiol Meas 2014 May;35(5):739-751. [doi: 10.1088/0967-3334/35/5/739] [Medline: 24671094]
41. Rachim VP, Lee J, Kim Y, Oh J, Jeong U, Park S. A scalable laser‐centric fabrication of an epidermal cardiopulmonary patch. Adv Materials Technologies 2022 Nov;7(11):2200242 [FREE Full text] [doi: 10.1002/admt.202200242]
42. Cotur Y, Olenik S, Asfour T, et al. Bioinspired stretchable transducer for wearable continuous monitoring of respiratory patterns in humans and animals. Adv Mater 2022 Aug;34(33):e2203310. [doi: 10.1002/adma.202203310] [Medline: 35730340]
43. Stollenwerk K, Müller J, Hinkenjann A, Krüger B. Analyzing spinal shape changes during posture training using a wearable device. Sensors (Basel) 2019 Aug 20;19(16):3625. [doi: 10.3390/s19163625] [Medline: 31434320]
44. Vuorinen T, Noponen K, Vehkaoja A, et al. Validation of printed, skin‐mounted multilead electrode for ECG measurements. Adv Materials Technologies 2019 Sep;4(9):1900246 [FREE Full text] [doi: 10.1002/admt.201900246]
45. Yamamoto Y, Yamamoto D, Takada M, et al. Efficient skin temperature sensor and stable gel‐less sticky ECG sensor for a wearable flexible healthcare patch. Adv Healthc Mater 2017 Sep;6(17):1700495. [doi: 10.1002/adhm.201700495] [Medline: 28661047]
46. Hsu YS, Chen TY, Wu D, Lin CM, Juang JN, Liu WT. Screening of obstructive sleep apnea in patients who snore using a patch-type device with electrocardiogram and 3-axis accelerometer. J Clin Sleep Med 2020 Jul 15;16(7):1149-1160. [doi: 10.5664/jcsm.8462] [Medline: 32267228]
47. Lee W, Seto E, Lin KY, Migliaccio GC. An evaluation of wearable sensors and their placements for analyzing construction worker's trunk posture in laboratory conditions. Applied Ergonomics 2017 Nov;65:424-436. [doi: 10.1016/j.apergo.2017.03.016]
48. Kim J, Kwon S, Seo S, Park K. Highly wearable galvanic skin response sensor using flexible and conductive polymer foam. Presented at: 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; Aug 26-30, 2014; Chicago, IL. [doi: 10.1109/embc.2014.6945148] [Medline: 25571516]
49. Castorino K, Polsky S, O'Malley G, et al. Performance of the dexcom G6 continuous glucose monitoring system in pregnant women with diabetes. Diabetes Technol Ther 2020 Dec;22(12):943-947. [doi: 10.1089/dia.2020.0085] [Medline: 32324061]
50. Hu J, Dun G, Geng X, Chen J, Wu X, Ren TL. Recent progress in flexible micro-pressure sensors for wearable health monitoring. Nanoscale Adv 2023 Jun 13;5(12):3131-3145. [doi: 10.1039/D2NA00866A]
51. Seesaard T, Wongchoosuk C. Flexible and stretchable pressure sensors: from basic principles to state-of-the-art applications. Micromachines (Basel) 2023 Aug 20;14(8):1638. [doi: 10.3390/mi14081638] [Medline: 37630177]
52. Mishra S, Mohanty S, Ramadoss A. Functionality of flexible pressure sensors in cardiovascular health monitoring: a review. ACS Sens 2022 Sep 23;7(9):2495-2520. [doi: 10.1021/acssensors.2c00942] [Medline: 36036627]
53. Alipour A, Gabrielson S, Patel PB. Ingestible sensors and medication adherence: focus on use in serious mental illness. Pharmacy (Basel) 2020 Jun 16;8(2):103. [doi: 10.3390/pharmacy8020103] [Medline: 32560178]
54. Kalantar-Zadeh K, Ha N, Ou JZ, Berean KJ. Ingestible sensors. ACS Sens 2017 Apr 28;2(4):468-483. [doi: 10.1021/acssensors.7b00045] [Medline: 28723186]
55. Liao CH, Cheng CT, Chen CC, et al. An Ingestible electronics for continuous and real-time intraabdominal pressure monitoring. J Pers Med 2020 Dec 24;11(1):12. [doi: 10.3390/jpm11010012] [Medline: 33374271]
56. Lee M, Shin J. Change in waist circumference with continuous use of a smart belt: an observational study. JMIR Mhealth Uhealth 2019 May 2;7(5):e10737. [doi: 10.2196/10737] [Medline: 31045500]
57. Kwon S, Hong J, Choi EK, et al. Detection of atrial fibrillation using a ring-type wearable device (cardiotracker) and deep learning analysis of photoplethysmography signals: prospective observational proof-of-concept study. J Med Internet Res 2020 May 21;22(5):e16443. [doi: 10.2196/16443] [Medline: 32348254]
58. Lochner CM, Khan Y, Pierre A, Arias AC. All-organic optoelectronic sensor for pulse oximetry. Nat Commun 2014 Dec 10;5(1):5745. [doi: 10.1038/ncomms6745] [Medline: 25494220]
59. Hong W, Lee J, Lee WG. A finger-perimetric tactile sensor for analyzing the gripping force by chopsticks towards personalized dietary monitoring. Sensors and Actuators A: Physical 2022 Jan;333:113253. [doi: 10.1016/j.sna.2021.113253]
60. Wu Y, Chen PF, Hu ZH, Chang CH, Lee GC, Yu WC, editors. A mobile health monitoring system using RFID ring-type pulse sensor. Presented at: Eighth IEEE International Conference on Dependable, Autonomic and Secure Computing; Dec 12-14, 2009; Chengdu, China. [doi: 10.1109/dasc.2009.136]
61. Yang SH, Koh CL, Hsu CH, et al. An instrumented glove-controlled portable hand-exoskeleton for bilateral hand rehabilitation. Biosensors (Basel) 2021 Dec 3;11(12):495. [doi: 10.3390/bios11120495] [Medline: 34940252]
62. Lee Y, Chung JW, Lee GH, et al. Standalone real-time health monitoring patch based on a stretchable organic optoelectronic system. Sci Adv 2021 Jun 4;7(23):eabg9180. [doi: 10.1126/sciadv.abg9180]
63. Kang M, Jeong H, Park SW, et al. Wireless graphene-based thermal patch for obtaining temperature distribution and performing thermography. Sci Adv 2022 Apr 15;8(15):eabm6693. [doi: 10.1126/sciadv.abm6693]
64. Kim KK, Ha I, Kim M, et al. A deep-learned skin sensor decoding the epicentral human motions. Nat Commun 2020 May 1;11(1):2149. [doi: 10.1038/s41467-020-16040-y] [Medline: 32358525]
65. Bai Y, Hibbing P, Mantis C, Welk GJ. Comparative evaluation of heart rate-based monitors: Apple Watch vs Fitbit Charge HR. J Sports Sci 2018 Aug;36(15):1734-1741. [doi: 10.1080/02640414.2017.1412235] [Medline: 29210326]

66. Phillips C, Liaqat D, Gabel M, de Lara E. Wristo2: reliable peripheral oxygen saturation readings from wrist-worn pulse oximeters. Presented at: 2021 IEEE International Conference on Pervasive Computing and Communications Workshops and Other Affiliated Events (PerCom Workshops); Mar 22-26, 2021; Kassel, Germany. [doi: 10.1109/PerComWorkshops51409.2021.9430986]
67. Spaccarotella C, Polimeni A, Mancuso C, Pelaia G, Esposito G, Indolfi C. Assessment of non-invasive measurements of oxygen saturation and heart rate with an apple Smartwatch: comparison with a standard pulse oximeter. J Clin Med 2022 Mar 8;11(6):1467. [doi: 10.3390/jcm11061467] [Medline: 35329793]
68. Yamagami K, Nomura A, Kometani M, et al. Early detection of symptom exacerbation in patients with SARS-Cov-2 infection using the Fitbit Charge 3 (DEXTERITY): pilot evaluation. JMIR Form Res 2021 Sep 16;5(9):e30819. [doi: 10.2196/30819] [Medline: 34516390]
69. Golbus JR, Dempsey W, Jackson EA, Nallamothu BK, Klasnja P. Microrandomized trial design for evaluating just-in-time adaptive interventions through mobile health technologies for cardiovascular disease. Circ Cardiovasc Qual Outcomes 2021 Feb;14(2):e006760. [doi: 10.1161/CIRCOUTCOMES.120.006760] [Medline: 33430608]
70. Strik M, Caillol T, Ramirez FD, et al. Validating QT-interval measurement using the apple watch ECG to enable remote monitoring during the COVID-19 pandemic. Circulation 2020 Jul 28;142(4):416-418. [doi: 10.1161/CIRCULATIONAHA.120.048253] [Medline: 32478565]
71. Kobel M, Kalden P, Michaelis A, et al. Accuracy of the apple watch iECG in children with and without congenital heart disease. Pediatr Cardiol 2022 Jan;43(1):191-196. [doi: 10.1007/s00246-021-02715-w] [Medline: 34468775]
72. Wile DJ, Ranawaya R, Kiss ZHT. Smart watch accelerometry for analysis and diagnosis of tremor. J Neurosci Methods 2014 Jun 15;230:1-4. [doi: 10.1016/j.jneumeth.2014.04.021] [Medline: 24769376]
73. Hadar E, Chen R, Toledano Y, Tenenbaum-Gavish K, Atzmon Y, Hod M. Noninvasive, continuous, real-time glucose measurements compared to reference laboratory venous plasma glucose values. J Matern Fetal Neonatal Med 2019 Oct;32(20):3393-3400. [doi: 10.1080/14767058.2018.1463987] [Medline: 29635953]
74. Alsunaidi B, Althobaiti M, Tamal M, Albaker W, Al-Naib I. A review of non-invasive optical systems for continuous blood glucose monitoring. Sensors (Basel) 2021 Oct 14;21(20):6820. [doi: 10.3390/s21206820] [Medline: 34696033]
75. Gao W, Emaminejad S, Nyein HYY, et al. Fully integrated wearable sensor arrays for multiplexed in situ perspiration analysis. Nature 2016 Jan 28;529(7587):509-514. [doi: 10.1038/nature16521] [Medline: 26819044]
76. Bandodkar AJ, Gutruf P, Choi J, et al. Battery-free, skin-interfaced microfluidic/electronic systems for simultaneous electrochemical, colorimetric, and volumetric analysis of sweat. Sci Adv 2019 Jan;5(1):eaav3294. [doi: 10.1126/sciadv.aav3294] [Medline: 30746477]
77. Wang C, Sani ES, Gao W. Wearable bioelectronics for chronic wound management. Adv Funct Mater 2022 Apr 25;32(17):2111022. [doi: 10.1002/adfm.202111022] [Medline: 36186921]
78. Jeon Y, Choi H, Lim M, et al. A wearable photobiomodulation patch using a flexible red - wavelength OLED and its in vitro differential cell proliferation effects. Adv Materials Technologies 2018 May;3(5):1700391 [FREE Full text] [doi: 10.1002/admt.201700391]
79. Shahandashti PF, Pourkheyrollah H, Jahanshahi A, Ghafoorifard H. Highly conformable stretchable dry electrodes based on inexpensive flex substrate for long-term Biopotential (EMG/ECG) monitoring. Sensors and Actuators A: Physical 2019 Aug;295:678-686. [doi: 10.1016/j.sna.2019.06.041]
80. Myers AC, Huang H, Zhu Y. Wearable silver nanowire dry electrodes for electrophysiological sensing. RSC Adv 2015;5(15):11627-11632. [doi: 10.1039/C4RA15101A]
81. Kisannagar RR, Jha P, Navalkar A, Maji SK, Gupta D. Fabrication of silver nanowire/polydimethylsiloxane dry electrodes by a vacuum filtration method for electrophysiological signal monitoring. ACS Omega 2020 May 12;5(18):10260-10265. [doi: 10.1021/acsomega.9b03678] [Medline: 32426582]
82. Roland T, Wimberger K, Amsuess S, Russold MF, Baumgartner W. An insulated flexible sensor for stable electromyography detection: application to prosthesis control. Sensors (Basel) 2019 Feb 24;19(4):961. [doi: 10.3390/s19040961] [Medline: 30813504]
83. Ng CL, Reaz MBI, Chowdhury MEH. A low noise capacitive electromyography monitoring system for remote healthcare applications. IEEE Sens J 2020 Mar 15;20(6):3333-3342. [doi: 10.1109/jsen.2019.2957068]
84. Nassar JM, Mishra K, Lau K, Aguirre - Pablo AA, Hussain MM. Recyclable nonfunctionalized paper - based ultralow - cost wearable health monitoring system. Adv Materials Technologies 2017 Apr;2(4):1600228 [FREE Full text] [doi: 10.1002/admt.201600228]
85. Yokota T, Nakamura T, Kato H, et al. A conformable Imager for biometric authentication and vital sign measurement. Nat Electron 2020 Jan 20;3(2):113-121. [doi: 10.1038/s41928-019-0354-7]
86. Kim DH, Lu N, Ma R, et al. Epidermal electronics. Science 2011 Aug 12;333(6044):838-843. [doi: 10.1126/science.1206157] [Medline: 21836009]
87. Garabelli P, Stavrakis S, Albert M, et al. Comparison of QT interval readings in normal sinus rhythm between a smartphone heart monitor and a 12 - Lead ECG for healthy volunteers and inpatients receiving sotalol or dofetilide. J Cardiovasc Electrophysiol 2016 Jul;27(7):827-832. [doi: 10.1111/jce.12976] [Medline: 27027653]

88. Mehmood N, Hariz A, Templeton S, Voelcker NH. A flexible and low power telemetric sensing and monitoring system for chronic wound diagnostics. Biomed Eng Online 2015 Mar 1;14(1):17. [doi: 10.1186/s12938-015-0011-y] [Medline: 25884377]
89. Henricson J, Sandh J, Iredahl F. Moisture sensor for exudative wounds - a pilot study. Skin Res Technol 2021 Sep;27(5):918-924. [doi: 10.1111/srt.13042] [Medline: 33788321]
90. Huo W, Mohammed S, Moreno JC, Amirat Y. Lower limb wearable robots for assistance and rehabilitation: a state of the art. IEEE Systems Journal 2016 Sep;10(3):1068-1081. [doi: 10.1109/JSYST.2014.2351491]
91. Birch N, Graham J, Priestley T, et al. Results of the first interim analysis of the RAPPER II trial in patients with spinal cord injury: ambulation and functional exercise programs in the REX powered walking aid. J Neuroeng Rehabil 2017 Jun 19;14(1):60. [doi: 10.1186/s12984-017-0274-6] [Medline: 28629390]
92. Luo Y, Coppola SM, Dixon PC, Li S, Dennerlein JT, Hu B. A database of human gait performance on irregular and uneven surfaces collected by wearable sensors. Sci Data 2020 Jul 8;7(1):219. [doi: 10.1038/s41597-020-0563-y] [Medline: 32641740]
93. Hong W, Lee J, Lee WG. A size-cuttable, skin-interactive wearable sensor for digital deciphering of epidermis wavy deformation. Biosensors (Basel) 2022 Jul 29;12(8):580. [doi: 10.3390/bios12080580] [Medline: 36004976]
94. Lee SI, Park E, Huang A, et al. Objectively Quantifying walking ability in degenerative spinal disorder patients using sensor equipped smart shoes. Med Eng Phys 2016 May;38(5):442-449. [doi: 10.1016/j.medengphy.2016.02.004] [Medline: 26970892]
95. Lin F, Wang A, Zhuang Y, Tomita MR, Xu W. Smart Insole: a wearable sensor device for unobtrusive gait monitoring in daily life. IEEE Trans Ind Inf 2016 Dec;12(6):2281-2291. [doi: 10.1109/TII.2016.2585643]
96. Lin Z, Wu Z, Zhang B, et al. A triboelectric nanogenerator‐based smart insole for multifunctional gait monitoring. Adv Mater Technol 2018 Oct 31;4(2):1800360. [doi: 10.1002/admt.201800360]
97. Waaijman R, de Haart M, Arts MLJ, et al. Risk factors for plantar foot ulcer recurrence in neuropathic diabetic patients. Diabetes Care 2014 Jun;37(6):1697-1705. [doi: 10.2337/dc13-2470] [Medline: 24705610]
98. Wicaksono I, Tucker CI, Sun T, et al. A tailored, electronic textile conformable suit for large-scale spatiotemporal physiological sensing in vivo. npj Flex Electron 2020 Apr 23;4(1):5. [doi: 10.1038/s41528-020-0068-y]
99. Islam MR, Afroj S, Beach C, et al. Fully printed and multifunctional graphene-based wearable E-textiles for personalized healthcare applications. iScience 2022 Mar 18;25(3):103945. [doi: 10.1016/j.isci.2022.103945] [Medline: 35281734]
100. Libanori A, Chen G, Zhao X, Zhou Y, Chen J. Smart textiles for personalized healthcare. Nat Electron 2022 Mar 28;5(3):142-156. [doi: 10.1038/s41928-022-00723-z]
101. Lyu S, He Y, Yao Y, Zhang M, Wang Y. Photothermal clothing for thermally preserving pipeline transportation of crude oil. Adv Funct Materials 2019 Jul;29(27):1900703 [FREE Full text] [doi: 10.1002/adfm.201900703]
102. Shafti A, Ribas Manero RB, Borg AM, Althoefer K, Howard MJ. Embroidered electromyography: a systematic design guide. IEEE Trans Neural Syst Rehabil Eng 2017 Sep;25(9):1472-1480. [doi: 10.1109/TNSRE.2016.2633506]
103. Cho G, Jeong K, Paik MJ, Kwun Y, Sung M. Performance evaluation of textile-based electrodes and motion sensors for smart clothing. IEEE Sensors J 2011 Dec;11(12):3183-3193. [doi: 10.1109/JSEN.2011.2167508]
104. Tabei F, Zaman R, Foysal KH, Kumar R, Kim Y, Chong JW. A novel diversity method for smartphone camera-based heart rhythm signals in the presence of motion and noise artifacts. PLoS One 2019;14(6):e0218248. [doi: 10.1371/journal.pone.0218248] [Medline: 31216314]
105. De Ridder B, Van Rompaey B, Kampen JK, Haine S, Dilles T. Smartphone apps using photoplethysmography for heart rate monitoring: meta-analysis. JMIR Cardio 2018 Feb 27;2(1):e4. [doi: 10.2196/cardio.8802] [Medline: 31758768]
106. Sengupta S, Sindal MD, Baskaran P, Pan U, Venkatesh R. Sensitivity and specificity of smartphone-based retinal imaging for diabetic retinopathy: a comparative study. Ophthalmol Retina 2019 Feb;3(2):146-153. [doi: 10.1016/j.oret.2018.09.016] [Medline: 31014763]
107. Freeman K, Dinnes J, Chuchu N, et al. Algorithm based smartphone apps to assess risk of skin cancer in adults: systematic review of diagnostic accuracy studies. BMJ 2020 Feb 10;368:m127. [doi: 10.1136/bmj.m127] [Medline: 32041693]
108. Rat C, Hild S, Rault Sérandour J, et al. Use of smartphones for early detection of melanoma: systematic review. J Med Internet Res 2018 Apr 13;20(4):e135. [doi: 10.2196/jmir.9392] [Medline: 29653918]
109. de Carvalho TM, Noels E, Wakkee M, Udrea A, Nijsten T. Development of smartphone apps for skin cancer risk assessment: progress and promise. JMIR Dermatol 2019 Jul 11;2(1):e13376. [doi: 10.2196/13376]
110. Chandrasekhar A, Natarajan K, Yavarimanesh M, Mukkamala R. An iPhone application for blood pressure monitoring via the oscillometric finger pressing method. Sci Rep 2018 Sep 3;8(1):13136. [doi: 10.1038/s41598-018-31632-x] [Medline: 30177793]
111. Thap T, Chung H, Jeong C, et al. High-resolution time-frequency spectrum-based lung function test from a smartphone microphone. Sensors (Basel) 2016 Aug 17;16(8):1305. [doi: 10.3390/s16081305] [Medline: 27548164]
112. Nemcova A, Jordanova I, Varecka M, et al. Monitoring of heart rate, blood oxygen saturation, and blood pressure using a smartphone. Biomedical Signal Processing and Control 2020 May;59:101928. [doi: 10.1016/j.bspc.2020.101928]
113. Gao J, Tan W, Ma L, Wang Y, Tang W. MUSEFood: multi-sensor-based food volume estimation on smartphones. arXiv. Preprint posted online in 2019 . [doi: 10.48550/arXiv.1903.07437]

114. Chandrasekhar A, Kim CS, Naji M, Natarajan K, Hahn JO, Mukkamala R. Smartphone-based blood pressure monitoring via the oscillometric finger-pressing method. Sci Transl Med 2018 Mar 7;10(431):eaap8674. [doi: 10.1126/scitranslmed.aap8674] [Medline: 29515001]
115. Rethnakumar R, Johar M, Alkawaz M, Helmi R, Tahir N. Smartphone based application for body temperature and heart rate measurements. Presented at: ICSGRC 2021: 12th IEEE Control and System Graduate Research Colloquium; Aug 7, 2021; Shah Alam, Malaysia. [doi: 10.1109/icsgrc53186.2021.9515284]
116. Ferrer-Lluis I, Castillo-Escario Y, Montserrat JM, Jané R. Sleeppos app: an automated smartphone application for angle based high resolution sleep position monitoring and treatment. Sensors (Basel) 2021 Jul 1;21(13):4531. [doi: 10.3390/s21134531] [Medline: 34282793]
117. Fino E, Mazzetti M. Monitoring healthy and disturbed sleep through smartphone applications: a review of experimental evidence. Sleep Breath 2019 Mar;23(1):13-24. [doi: 10.1007/s11325-018-1661-3] [Medline: 29687190]
118. Pierce A, Ignasiak NK, Eiteman-Pang WK, Rakovski C, Berardi V. Mobile phone sensors can discern medication-related gait quality changes in Parkinson's patients in the home environment. Comput Methods Programs Biomed Update 2021;1:100028. [doi: 10.1016/j.cmpbup.2021.100028]
119. Chung H, Jeong C, Luhach AK, Nam Y, Lee J. Remote pulmonary function test monitoring in cloud platform via smartphone built-in microphone. Evol Bioinform Online 2019;15:1176934319888904. [doi: 10.1177/1176934319888904] [Medline: 31798298]
120. Faezipour M, Abuzneid A. Smartphone-based self-testing of COVID-19 using breathing sounds. Telemed J E Health 2020 Oct;26(10):1202-1205. [doi: 10.1089/tmj.2020.0114] [Medline: 32487005]
121. Nakano H, Hirayama K, Sadamitsu Y, et al. Monitoring sound to quantify snoring and sleep apnea severity using a smartphone: proof of concept. J Clin Sleep Med 2014 Jan 15;10(1):73-78. [doi: 10.5664/jcsm.3364] [Medline: 24426823]
122. Peng C, Chen M, Wang H, Shen J, Jiang X. P(VDF-Trfe) thin-film-based transducer for under-display ultrasonic fingerprint sensing applications. IEEE Sensors J 2020 Oct 1;20(19):11221-11228. [doi: 10.1109/JSEN.2020.2997375]
123. Kim J, Song W, Jung S, et al. Capacitive heart-rate sensing on touch screen panel with laterally interspaced electrodes. Sensors 2020 Jul 17;20(14):3986. [doi: 10.3390/s20143986]
124. Aghanavesi S, Nyholm D, Senek M, Bergquist F, Memedi M. A smartphone-based system to quantify dexterity in Parkinson's disease patients. Inform Med Unlocked 2017;9:11-17. [doi: 10.1016/j.imu.2017.05.005]
125. Abazid M, Houmani N, Garcia-Salicetti S. Enhancing security on touch-screen sensors with augmented handwritten signatures. Sensors (Basel) 2020 Feb 10;20(3):933. [doi: 10.3390/s20030933] [Medline: 32050606]
126. Blanco-Gonzalo R, Miguel-Hurtado O, Sanchez-Reillo R, Gonzalez-Ramirez A. Usability analysis of a handwritten signature recognition system applied to mobile scenarios. Presented at: ICCST 2013: 47th International Carnahan Conference on Security Technology; Oct 8-11, 2013; Medellín, Colombia. [doi: 10.1109/ccst.2013.6922053]
127. Coskun AF, Wong J, Khodadadi D, Nagi R, Tey A, Ozcan A. A personalized food allergen testing platform on a cellphone. Lab Chip 2013 Feb 21;13(4):636-640. [doi: 10.1039/c2lc41152k] [Medline: 23254910]
128. Chu S, Wang H, Ling X, Yu S, Yang L, Jiang C. A portable smartphone platform using a ratiometric fluorescent paper strip for visual quantitative sensing. ACS Appl Mater Interfaces 2020 Mar 18;12(11):12962-12971. [doi: 10.1021/acsami.9b20458]
129. Liang PS, Park TS, Yoon JY. Rapid and reagentless detection of microbial contamination within meat utilizing a smartphone-based biosensor. Sci Rep 2014 Aug 5;4(1):5953. [doi: 10.1038/srep05953] [Medline: 25092261]
130. Mousseau S, Lapointe A, Gravel J. Diagnosing acute otitis media using a smartphone otoscope; a randomized controlled trial. Am J Emerg Med 2018 Oct;36(10):1796-1801. [doi: 10.1016/j.ajem.2018.01.093] [Medline: 29544905]
131. Demant MN, Jensen RG, Bhutta MF, Laier GH, Lous J, Homøe P. Smartphone otoscopy by non-specialist health workers in rural Greenland: a cross-sectional study. Int J Pediatr Otorhinolaryngol 2019 Nov;126:109628. [doi: 10.1016/j.ijporl.2019.109628] [Medline: 31445480]
132. Ghosh S, Aggarwal K, U VT, Nguyen T, Han J, Ahn CH. A new microchannel capillary flow assay (MCFA) platform with lyophilized chemiluminescence reagents for a smartphone-based POCT detecting malaria. Microsyst Nanoeng 2020;6(1):5. [doi: 10.1038/s41378-019-0108-8] [Medline: 34567620]
133. Peritz DC, Howard A, Ciocca M, Chung EH. Smartphone ECG AIDS real time diagnosis of palpitations in the competitive college athlete. J Electrocardiol 2015;48(5):896-899. [doi: 10.1016/j.jelectrocard.2015.07.010] [Medline: 26271889]
134. Ippili S, Kim B, Jella V, Jung JS, Vuong VH, Yoon SG. Antireflective, transparent, water-resistant, and antibacterial Zn-doped silicon oxide thin films for touchscreen-based display applications. ACS Sustainable Chem Eng 2022 Feb 14;10(6):2136-2147. [doi: 10.1021/acssuschemeng.1c07404]
135. Chen IM, Chen YY, Liao SC, Lin YH. Development of digital biomarkers of mental illness via mobile apps for personalized treatment and diagnosis. JPM 2022 Jun 6;12(6):936. [doi: 10.3390/jpm12060936]
136. Moon J, Yun J, Yoon YD, et al. Blue light effect on retinal pigment epithelial cells by display devices. Integr Biol (Camb) 2017 May 22;9(5):436-443. [doi: 10.1039/c7ib00032d] [Medline: 28386617]
137. Kumar A, Sharma K, Sharma A. Hierarchical deep neural network for mental stress state detection using iot based biomarkers. Pattern Recogn Lett 2021 May;145:81-87. [doi: 10.1016/j.patrec.2021.01.030]

138. Jacobson NC, Summers B, Wilhelm S. Digital biomarkers of social anxiety severity: digital phenotyping using passive smartphone sensors. J Med Internet Res 2020 May 29;22(5):e16875. [doi: 10.2196/16875] [Medline: 32348284]
139. Sverdlov O, van Dam J, Hannesdottir K, Thornton-Wells T. Digital therapeutics: an integral component of digital innovation in drug development. Clin Pharmacol Ther 2018 Jul;104(1):72-80. [doi: 10.1002/cpt.1036] [Medline: 29377057]
140. Makin S. The emerging world of digital therapeutics. Nature 2019 Sep;573(7775):S106-S109. [doi: 10.1038/d41586-019-02873-1] [Medline: 31554990]
141. Alliance for Advancing Bioelectronic Medicine. Building a bioelectronic medicine movement 2019: insights from leaders in industry, academia, and research. Bioelectron Med 2020 Jan 31;6(1):1. [doi: 10.1186/s42234-020-0037-8]
142. Pavlov VA, Tracey KJ. Bioelectronic medicine: preclinical insights and clinical advances. Neuron 2022 Nov 2;110(21):3627-3644. [doi: 10.1016/j.neuron.2022.09.003] [Medline: 36174571]
143. Sevcencu C. Single-interface bioelectronic medicines-concept, clinical applications and preclinical data. J Neural Eng 2022 Jun 6;19(3):031001. [doi: 10.1088/1741-2552/ac6e08] [Medline: 35533654]
144. Peeples L. Core concept: the rise of bioelectric medicine sparks interest among researchers, patients, and industry. Proc Natl Acad Sci U S A 2019 Dec 3;116(49):24379-24382. [doi: 10.1073/pnas.1919040116] [Medline: 31796581]
145. Henson P, Wisniewski H, Hollis C, Keshavan M, Torous J. Digital mental health apps and the therapeutic alliance: initial review. BJPsych Open 2019 Jan;5(1):e15. [doi: 10.1192/bjo.2018.86] [Medline: 30762511]
146. Dang A, Arora D, Rane P. Role of digital therapeutics and the changing future of healthcare. J Family Med Prim Care 2020 May;9(5):2207-2213. [doi: 10.4103/jfmpc.jfmpc_105_20] [Medline: 32754475]
147. Seshadri DR, Bittel B, Browsky D, et al. Accuracy of the apple watch 4 to measure heart rate in patients with atrial fibrillation. IEEE J Transl Eng Health Med 2020;8:2700204. [doi: 10.1109/JTEHM.2019.2950397] [Medline: 32128290]
148. Seshadri DR, Bittel B, Browsky D, et al. Accuracy of apple watch for detection of atrial fibrillation. Circulation 2020 Feb 25;141(8):702-703. [doi: 10.1161/CIRCULATIONAHA.119.044126] [Medline: 32091929]
149. Namwongsa S, Puntumetakul R, Neubert MS, Chaiklieng S, Boucaut R. Ergonomic risk assessment of smartphone users using the rapid upper limb assessment (RULA) tool. PLoS One 2018;13(8):e0203394. [doi: 10.1371/journal.pone.0203394] [Medline: 30161266]
150. Mallinson K. Smartphone revolution: technology patenting and licensing fosters innovation, market entry, and exceptional growth. IEEE Consumer Electron Mag 2015 Apr;4(2):60-66. [doi: 10.1109/mce.2015.2392954]
151. Gupta S, Mahmoud A, Massoomi MR. A clinician's guide to smartwatch "interrogation". Curr Cardiol Rep 2022 Aug;24(8):995-1009. [doi: 10.1007/s11886-022-01718-0] [Medline: 35635678]
152. James J. The smart feature phone revolution in developing countries: bringing the Internet to the bottom of the pyramid. Inf Soc 2020 May 19;36(4):226-235. [doi: 10.1080/01972243.2020.1761497]
153. Park J, Heo S, Park K, et al. Research on flexible display at Ulsan National Institute of Science and Technology. npj Flex Electron 2017 Nov 13;1(1):9. [doi: 10.1038/s41528-017-0006-9]
154. Choi M, Bae SR, Hu L, Hoang AT, Kim SY, Ahn JH. Full-color active-matrix organic light-emitting diode display on human skin based on a large-area MoS backplane. Sci Adv 2020 Jul;6(28):eabb5898. [doi: 10.1126/sciadv.abb5898] [Medline: 32923597]
155. Yokota T, Fukuda K, Someya T. Recent progress of flexible image sensors for biomedical applications. Adv Mater 2021 May;33(19):e2004416. [doi: 10.1002/adma.202004416] [Medline: 33527511]
156. Sekitani T, Someya T. Stretchable, large-area organic electronics. Adv Mater 2010 May 25;22(20):2228-2246. [doi: 10.1002/adma.200904054] [Medline: 20229571]
157. Li S, Peele BN, Larson CM, Zhao H, Shepherd RF. A stretchable multicolor display and touch interface using photopatterning and transfer printing. Adv Mater 2016 Nov;28(44):9770-9775 [FREE Full text] [doi: 10.1002/adma.201603408]
158. Kim EH, Han H, Yu S, et al. Interactive skin display with epidermal stimuli electrode. Adv Sci 2019 Jul 3;6(13):1802351. [doi: 10.1002/advs.201970076]
159. Tajima R, Miwa T, Oguni T, et al. Truly wearable display comprised of a flexible battery, flexible display panel, and flexible printed circuit. J Soc Info Display 2014 Oct 21;22(5):237-244. [doi: 10.1002/jsid.242]

Abbreviations

AI: artificial intelligence
CMOS: complementary metal-oxide-semiconductor
ECG: electrocardiogram
EEG: electroencephalogram
EGG: electroglottogram
EMG: electromyogram
IMU: inertial measurement unit mHealth: mobile health
OLED: organic light-emitting diode
PPG: photoplethysmogram
PSA: pressure-sensitive adhesive

Edited by L Buis; submitted 07.05.23; peer-reviewed by C Wongchoosuk, D Parry; revised version received 08.11.23; accepted 20.12.23; published 22.01.24.

Please cite as:
Hong W
Advances and Opportunities of Mobile Health in the Postpandemic Era: Smartphonization of Wearable Devices and Wearable Deviceization of Smartphones
JMIR Mhealth Uhealth 2024;12:e48803
URL: https://mhealth.jmir.org/2024/1/e48803
doi:10.2196/48803

An Introduction to Smart Home Ward–Based Hospital-at-Home Care in China

Weibin Cheng[1], MD; Xiaowen Cao[1], MSc; Wanmin Lian[2], MSc; Junzhang Tian[1], MD

[1]Institute for Healthcare Artificial Intelligence Application, Guangdong Second Provincial General Hospital, Guangzhou, China

[2]Information Department, Guangdong Second Provincial General Hospital, Guangzhou, China

Corresponding Author:
Junzhang Tian, MD
Institute for Healthcare Artificial Intelligence Application, Guangdong Second Provincial General Hospital, , Guangzhou, , China

Abstract

Hospital-at-home has been gaining increased attention as a potential remedy for the current shortcomings of our health care system, allowing for essential health services to be provided to patients in the comfort of their own homes. The advent of digital technology has revolutionized the way we provide medical and health care, leading to the emergence of a "hospital without walls." The rapid adoption of novel digital health care technologies is revolutionizing remote health care provision, effectively dismantling the conventional boundary separating hospitals from the comfort of patients' homes. The Guangdong Second Provincial General Hospital has developed a 5G-powered Smart Home Ward (SHW) that extends medical care services to the home setting and is tailored to meet the needs and settings of each patient's household. The SHW was initially tested for its suitability for treating 4 specialized diseases, including cardiovascular disease, stroke, Parkinson disease, and Alzheimer disease. Understanding and addressing the potential challenges and risks associated with SHWs is essential for the successful implementation and maintenance of safe and effective home hospitalization.

(JMIR Mhealth Uhealth 2024;12:e44422) doi:10.2196/44422

KEYWORDS

smart home ward; telemonitoring; telemedicine; home care; hospital at home; healthcare delivery; implementation; smart ward; medical monitoring; medical care; rehabilitation; health care

The Future of Older Adult Care in China: Innovations in Health Care Delivery

China has emerged as a rapidly aging society. Aging contributes significantly to the health care burden in China due to the increased prevalence of chronic diseases and disabilities among older adults. The health care burden for the population of older individuals in China is further compounded by inadequate health care infrastructure and limited access to health care services in rural areas. The Chinese government has implemented various policies and initiatives to address these challenges, including expanding health care coverage, promoting preventative care, and increasing investment in health care infrastructure. Currently, the top-level design of China's older adult care model is "9073," which means that 90% of older people receive older adult care services at home, 7% of them receive short-term care in the community, and 3% of them receive institutional care. This implies that the age-friendly and livable environment at home affects 90% of the population of older adults.

Hospital-at-home, which seeks to provide essential health services to patients in the comfort of their own homes, has been gaining increasing attention as a feasible solution for at-home older adult care and medical services [1-3]. The COVID-19 pandemic has underscored the risks of overreliance on physical medical institutions, emphasizing the continued need to develop a decentralized medical service ecosystem that revolves around patients' families and communities. Hospital-at-home programs have been implemented in many high-income countries for years [4-7]; however, successful programs are limited [8]. Experience translating from high-income countries to low- and middle-income countries can be challenging, where medical resources are limited and public health literacy is low [2,3,9,10].

The Growth of Emerging Technologies for Home Hospitalization

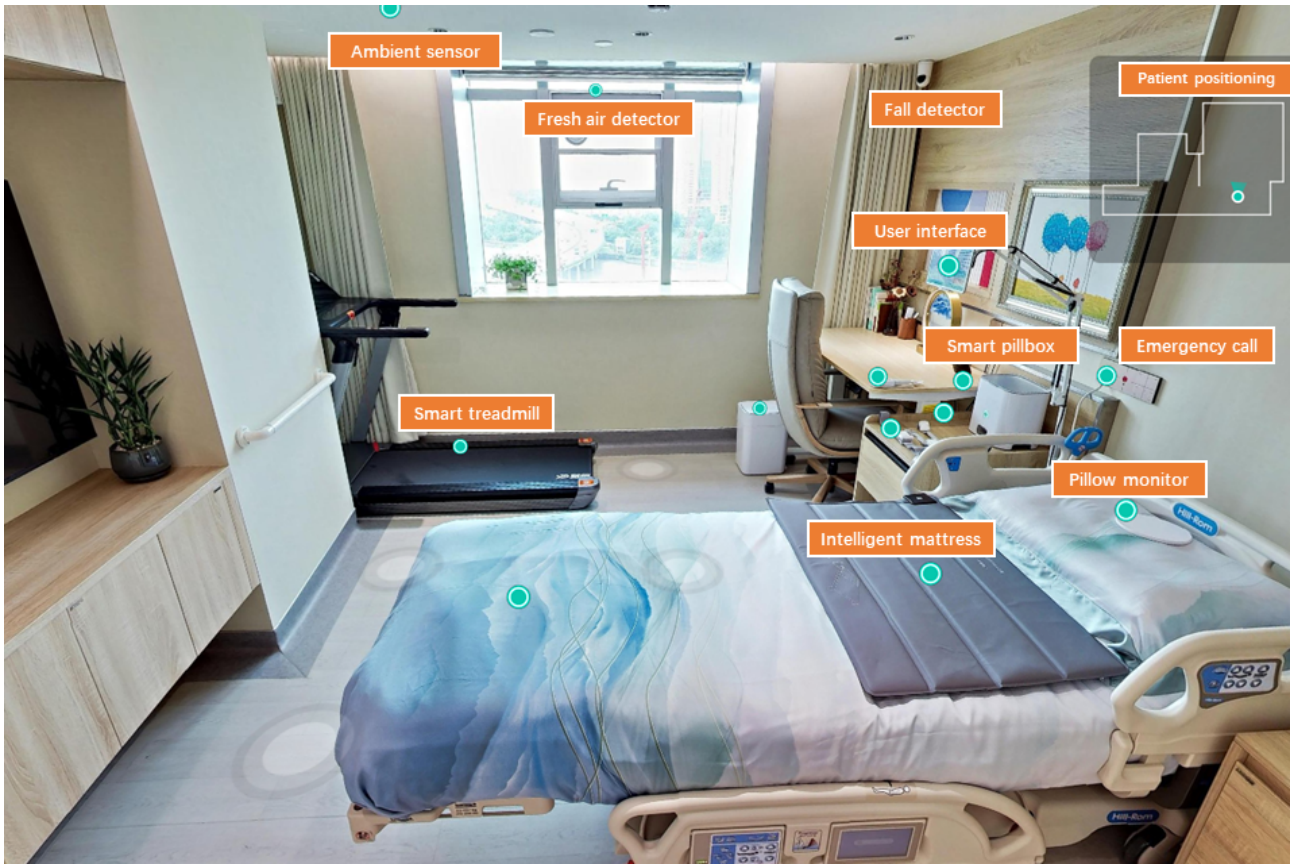
New digital health care technologies are being rapidly adopted for remote health care provision, which is breaking down the traditional barrier between the hospital and home. The emergence of 5G technology posed the potential to further enhance home hospitalization by enabling remote patient monitoring and real-time communication between health care providers and patients. The introduction of sensor-based wearables and devices has changed the way that clinical data are collected and stored, leading to groundbreaking advancements in how care is provided [11,12]. Digital biomarkers—a set of objectives, quantifiable measures of physiological and behavioral characteristics that are acquired via wearables, implants, and other devices—are becoming

increasingly essential in this process [13,14]. Home hospitalization can greatly benefit from the use of digital biomarkers, providing care teams with a more comprehensive understanding of a patient's health. These biomarkers enable the tracking of vital signs such as heart rate and respiration rate, as well as changes in sleep patterns, activity levels, and dietary habits [12,15,16].

Smart Home Ward for Hospital-at-Home Care

In 2021, Guangdong Second Provincial General Hospital developed the "Hospital Intelligent Twins," a 5G-powered smart hospital that integrates the Internet of Things (IoT), artificial intelligence (AI), cloud computing, and 5G applications to create all-scenario intelligence for health care and hospital management [17,18]. Leveraging this 5G smart hospital infrastructure, Guangdong Second Provincial General Hospital further explored the Smart Home Ward (SHW), which seeks to break through the "wall" of centralized hospital-based health care services by extending care services to the home setting (Figure 1). The SHW is a seamless hospital unit managed by health professionals in the hospital and is designed to provide patients with equivalent hospital-level services at home, including medical monitoring, ward rounds, consultation, and medical care, and so on. SHW can support 2 hospital-at-home models, early-supported discharge and admission avoidance, offering integrated health monitoring, medical care, and rehabilitation services all within the comfort of their homes [19,20].

Figure 1. Demonstration of the Smart Home Ward in the hospital.



The management of patients in the SHW is handled by an interdisciplinary care team overseen by responsible specialists from relevant clinical departments. At present, the Departments of Cardiology and Neurology are pilot testing SHWs for patients with cardiovascular conditions and those with cerebrovascular conditions, respectively. Each specialty establishes a dedicated home ward team comprising physicians from that department, along with nurses, pharmacists, physical therapists, occupational therapists, and social workers. For example, the post–coronary heart disease treatment team includes cardiologists, cardiac nurses, and rehabilitation therapists. Patients undergo an initial clinical assessment by their specialist physician to determine their suitability for remote home care. The inclusion and exclusion criteria for specific conditions such as coronary heart disease and stroke are provided in Multimedia Appendix 1. If deemed feasible based on the assessment and if the patient provides consent, the interprofessional SHW staff conduct a home environment evaluation. Necessary modifications are made and monitoring equipment or devices installed to safely support care at home with remote specialist oversight.

In the SHW, series of digital biomarkers are monitored, including blood pressure, heart rate, respiratory rate, blood oxygen saturation, body temperature, and electrocardiographic signals, with patients' daily activities, treatment compliance, and risk occurrences monitored via contact or contactless devices. These data can be leveraged to tailor treatments or notify care providers of any deviations from expected parameters.

Data generated from the SHW are transmitted securely to the hospital's electronic medical record (EMR) system in real time. The integration follows privacy and security guidelines set forth by our hospital's Health Insurance Portability and Accountability Act (HIPAA)–compliant policies. Only deidentified data points that are relevant for clinical care, such as vital signs, activity levels, and medication adherence data, are integrated into the EMR. Data transmission from the SHW to the EMR is one-way to ensure the security of sensitive hospital information. Authorized care team members, including attending physicians, nurses, pharmacists, physical therapists, and case managers, have access to consolidated patient data reports within the EMR system. This allows them to monitor trends, recognize any deviations from normal ranges, and act accordingly without needing to use separate systems.

The primary care team, led by the patient's attending physician, is responsible for reviewing the daily reports and contacting patients if follow-up care is needed, based on the remote patient monitoring data. Family members and caregivers participate in care conferences to stay updated on the patient's progress but do not have access to the EMR. We aim to expand access to aggregated reports to allow for greater caregiver involvement while maintaining privacy and security.

A pretraining ward has been established in the hospital to help patients transit from hospital to home living. Patients are trained on the use of smart home devices and receive guidance on remote rehabilitation training. In addition, the pretraining ward includes medical assessments to ensure that the patient is eligible for admission to SHW care and will benefit from the treatments provided at home. Specialized staff consisting of nurses, physical therapists, and occupational therapists are dedicated to supporting the SHW program. These clinicians provide personalized patient education, demonstrations, and skills training. They ensure that patients and caregivers are comfortable with the remote monitoring system and therapy program prior to discharge. The goal is to maximize treatment adherence and outcomes through empowering patients and families with knowledge. A dedicated tech support team is also available to assist families in successfully setting up the in-home system and addressing any technical issues that may arise post discharge. Readily available support from clinical and technical experts further enables safe and independent living at home with remote care and monitoring.

Characteristics of the SHW

The SHW constructed in this project not only integrates advanced medical IoT technology and products, but also involves the renovation of the environment to make it suitable for older adults and patients, extending homogeneous medical services provided in the hospital wards to patients' homes. Its characteristics are mainly reflected in the following 4 aspects.

First, the SHW has a multitude of functions including ambient sensing, medical monitoring, rehabilitation training, exercise and diet guidance, psychological counseling, and sleep management. To achieve these capabilities, a variety of cutting-edge technologies such as 5G or Wi-Fi 6, Internet of Medical Things (IoMT), smart wearable devices, smart home appliances, and health monitoring equipment have been deployed in the home setting. For instance, an intelligent closetool can detect urine levels and monitor heart rate, body fat, and the length of time spent on the toilet [21]. Additionally, an intelligent mattress and pillow monitor heart rate, respiration rate, and body movement frequency during sleep [22,23]. Applications such as those for fall detection in the bathroom and the amount of stillness in daily activities can trigger an automatic alarm for an emergency, which connects to the hospital [24].

Second, to facilitate hospital-at-home management, a platform was developed with the integration of cross-system and cross-ecological IoMT devices (Figure 2; see the screenshot of the SHW Medical Management Information System). Thus, IoMT devices and ambient sensors are interconnected, and data are generated, gathered, managed, and processed by various terminals, including data from a hospital and a SHW (including data for the home environment, daily activities, treatment, rehabilitation, etc). This enables interoperability for home ward management, allowing patients, health professionals, and caregivers to use mobile terminals such as tablets to view the patient's physiological data in real time and manage the home ward setting with ease. As medical teams have the capability to monitor a patient's health from a distance, fewer in-person visits are needed, thus decreasing the cost of care [25].

Figure 2. Screenshot of the Smart Home Ward Medical Management Information System.

Third, deployment of an SHW is tailored to cater the needs and settings of a patient's household. To promote patient autonomy in the home setting, the SHW undergoes various levels of renovations suitable for aging individuals, creating a barrier-free, safe, and secure living space. Furthermore, ambient sensors enable automatic configuration to ensure the comfort of the environment [26]. These sensors continuously detect the air, temperature, and humidity in the ward. Whenever the indoor concentrations of particulate matter of diameter ≤ 2.5 μm, formaldehyde, and carbon dioxide exceed standard levels, the SHW automatically adjusts its equipment to improve indoor air quality.

Fourth, the extension of hospital care services to the home setting is achieved through the integration of various technologies and resources. This challenge is addressed by the SHW through the deployment of advanced IoMT technology and products, which allow for the passive monitoring of patients' health and the implementation of timely medical interventions. Additionally, a comprehensive EMR system has been implemented, enabling seamless sharing of patients' medical records between the hospital and home health care systems. Furthermore, a robust 5G-based telemedicine infrastructure has been established, enabling real-time communication between patients and doctors, including digital consultations, remote monitoring, and telemedicine-enabled home visits, as well as providing on-demand home care services [27].

SHW Care Plans for Specialized Diseases

The SHW's feasibility of treating 4 specialized diseases, namely Parkinson disease, cardiovascular disease, stroke, and Alzheimer disease, is being tested. For each disease, corresponding care plans have been developed, and necessary monitoring and treatment facilities have been suggested.

In the case of Parkinson disease, exercise rehabilitation treatment is mainly used as the primary treatment modality in the home ward. Patients are guided by doctors remotely to implement various forms of internet-based rehabilitation exercise training, such as relaxation training, joint range of motion training, muscle training, breathing training, gait training, balance training, and cognitive training. During exercise, the patient's heart rate is monitored using a smart watch to prevent excessive exercise intensity. A smart lunch box is provided for daily monitoring of the frequency and amplitude of the patient's hand tremors while eating to evaluate the effectiveness of rehabilitation treatment [28].

For patients with cardiovascular disease, the home ward caters to those with ST-segment elevation myocardial infarction, non–ST-segment elevation acute coronary syndrome, stable angina pectoris, ischemic cardiomyopathy, chronic systolic heart failure, and sudden cardiac death syndrome. Cardiac rehabilitation exercise therapy is guided by health professionals from the hospital, which includes aerobic exercise, resistance training, and neuromuscular training. A treadmill with heart and lung function monitoring is set up as a rehabilitation tool at home, which records and monitors the patient's exercise status and heart rate, blood oxygen saturation, and calorie consumption. This allows doctors to adjust the program promptly [29].

For stroke, given that individuals who have experienced a stroke often face difficulties with mobility, the adaptation of living spaces to suit their needs is crucial to promote independent living and facilitate rehabilitation. To this end, we have implemented appropriate accessibility retrofits within the home

setting to establish a barrier-free environment for patients. Such efforts can contribute to not only increased mobility but also heightened patient confidence. To evaluate the efficacy of functional rehabilitation, regular assessments are conducted through both home visits and hospital evaluations.

In the case of Alzheimer disease, a comprehensive treatment approach is adopted, including cognitive training, task training, and music therapy. These treatments are supported by the augmented reality and virtual reality applications that allow doctors to simulate complex scenarios to stimulate patients' brain activities in a safe and controlled environment. Several physiological signals, such as heart rate variability, eye movements, and sleep patterns, are recorded and are used to train models for treatment evaluation. Furthermore, an AI-powered camera is installed in the home environment to monitor the patient's activity status, such as sitting and lying time. A smart watch with GPS positioning and communication functions is also used to prevent patients from getting lost [30].

Challenges for the Scale-Up of the SHW Care Program in China

The widespread adoption of SHWs has been faced with several challenges. First, authoritative guidelines for the implementation of medical service standards for home ward care and those for setting up technology-enabled digital wards in China are lacking. Second, a sustainable fee-based model for home ward care has yet to be established. Third, the effective operation of home ward care requires close collaboration among hospitals, community health service organizations, and family members. However, there are currently no unified regulations to delineate the responsibilities of various stakeholders involved in providing home ward care. Fourth, the popularity and reliability of the technology needs improvement. SHWs require specialized equipment and technologies, a secure and trust-based environment, and staffing competencies to ensure patient safety and privacy. A unique standard needs to be established to ensure the interconnectedness and interoperability of various devices. Meanwhile, potential problems associated with the cost of equipment, access to necessary data, and data privacy also exist.

Implications of Hospital-at-Home Care Using SHWs for Practitioners, Researchers, and Policy Makers

The implementation of SHWs has the potential to extend the quantity and quality of hospital care services in response to the increasing demand for medical care in China's aging society.

For health care practitioners, the integration of smart home digital health technologies can facilitate remote patient monitoring and management, enabling timely and effective care delivery in patients' homes. This approach can reduce the burden on hospitals and clinics, while improving patient outcomes and satisfaction. To enable the successful implementation of SHWs, new team structures with corresponding workflows must be created in clinical settings to optimize health care systems and patients' usage of this technology. Clinical physicians must work closely with technicians in developing workflows and integrated AI tools and in the process of care provision, such as remote monitoring.

For researchers, the implementation of SHWs presents a unique opportunity to study the impact of digital health technology on health care delivery, patient outcomes, and cost-effectiveness. However, ethical considerations must be navigated when accessing the vast amounts of data gathered at SHWs to ensure patient privacy and security [31]. All stakeholders must understand and adapt to these implications to fully harness their potential benefits while mitigating associated risks.

Policy makers can benefit from the use of SHW technologies, as they show promising potential in reducing health care costs and improving access to care, particularly for patients in rural or underserved areas. However, policy makers must establish regulations to safeguard patient privacy and security, ensure equitable access to technology, and enhance digital infrastructure for widespread adoption. With careful consideration of these factors, the implementation of SHW can revolutionize hospital-at-home care, improving the quality of care, reducing health care costs, and ultimately benefiting patients and the health care system as a whole.

In Summary

The advent of digital technology has revolutionized the way we provide medical and health care, leading to the emergence of "hospitals without walls." This SHW is an innovative care model that promises to bring a host of improvements to health care by providing convenient access to "boundless ambulatory care," as well as "boundless inpatient care." Despite the potential benefits, these digital projects are still faced with some challenges, such as public acceptance and adoption of the technology, willingness to pay for services, and encouraging medical insurance uptake. Therefore, it is essential for governments to promote public awareness of the advantages of digital technology and introduce incentives that motivate people to take up medical insurance in order for these services to be widely available.

Acknowledgments

This study was funded by the National Key R&D (research and development) Program of China (number 2021YFC2009400).

Authors' Contributions

WC, XC, and WL conceived the study and drafted the manuscript. JT provided professional support and made critical revisions to the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Inclusion/exclusion criteria for specific conditions.
[DOCX File, 18 KB - mhealth_v12i1e44422_app1.docx]

References

1. Levi B, Borow M, Wapner L, Feldman Z. Home hospitalization worldwide and in Israel. Isr Med Assoc J 2019 Aug;21(8):565-567. [Medline: 31474024]
2. Ritchie C, Leff B. Home-based care reimagined: a full-fledged health care delivery ecosystem without walls. Health Aff (Millwood) 2022 May;41(5):689-695. [doi: 10.1377/hlthaff.2021.01011] [Medline: 35500180]
3. Knight T, Lasserson D. Hospital at home for acute medical illness: the 21st century acute medical unit for a changing population. J Intern Med 2022 Apr;291(4):438-457. [doi: 10.1111/joim.13394] [Medline: 34816527]
4. Summerfelt WT, Sulo S, Robinson A, Chess D, Catanzano K. Scalable hospital at home with virtual physician visits: pilot study. Am J Manag Care 2015 Oct;21(10):675-684. [Medline: 26633092]
5. Hospital at Home. East and North Hertfordshire NHS Trust. URL: www.enherts-tr.nhs.uk/services/hospital-at-home [accessed 2023-02-15]
6. Hospital in the Home. Department of Health, Victoria. URL: www.health.vic.gov.au/patient-care/hospital-in-the-home [accessed 2023-02-15]
7. Montalto M. The 500-bed hospital that isn't there: the Victorian Department of Health review of the Hospital in the Home program. Med J Aust 2010 Nov 15;193(10):598-601. [doi: 10.5694/j.1326-5377.2010.tb04070.x] [Medline: 21077817]
8. Klein S. "Hospital at Home" programs improve outcomes, lower costs but face resistance from providers and payers. The Commonwealth Fund. URL: www.commonwealthfund.org/publications/newsletter-article/hospital-home-programs-improve-outcomes-lower-costs-face-resistance [accessed 2022-09-22]
9. de Sousa Vale J, Franco AI, Oliveira CV, Araújo I, Sousa D. Hospital at home: an overview of literature. Home Health Care Manag Pract 2020 May;32(2):118-123. [doi: 10.1177/1084822319880930]
10. Foubister V. Hospital at home program in New Mexico improves care quality and patient satisfaction while reducing costs. The Commonwealth Fund. URL: www.commonwealthfund.org/publications/newsletter-article/hospital-home-program-new-mexico-improves-care-quality-and-patient [accessed 2024-01-19]
11. Huang Y, Hsieh M, Chao H, Hung S, Park J. Pervasive, secure access to a hierarchical sensor-based healthcare monitoring architecture in wireless heterogeneous networks. IEEE J Select Areas Commun 2009 May;27(4):400-411. [doi: 10.1109/JSAC.2009.090505]
12. Bijalwan V, Semwal VB, Mandal TK. Fusion of multi-sensor-based biomechanical gait analysis using vision and wearable sensor. IEEE Sensors J 2021 Jul;21(13):14213-14220. [doi: 10.1109/JSEN.2021.3066473]
13. Babrak LM, Menetski J, Rebhan M, et al. Traditional and digital biomarkers: two worlds apart? Digit Biomark 2019 Aug;3(2):92-102. [doi: 10.1159/000502000] [Medline: 32095769]
14. Motahari-Nezhad H, Al-Abdulkarim H, Fgaier M, et al. Digital biomarker-based interventions: systematic review of systematic reviews. J Med Internet Res 2022 Dec 21;24(12):e41042. [doi: 10.2196/41042] [Medline: 36542427]
15. Qiu S, Wang Z, Zhao H, et al. Body sensor network-based robust gait analysis: toward clinical and at home use. IEEE Sensors J 2019 Oct;19(19):8393-8401. [doi: 10.1109/JSEN.2018.2860938]
16. Coravos A, Khozin S, Mandl KD. Developing and adopting safe and effective digital biomarkers to improve patient outcomes. NPJ Digit Med 2019;2(1):14. [doi: 10.1038/s41746-019-0090-4] [Medline: 30868107]
17. Cheng W, Lian W, Tian J. Building the hospital intelligent twins for all-scenario intelligence health care. Digit Health 2022 Jun 12;8:20552076221107894. [doi: 10.1177/20552076221107894] [Medline: 35720617]
18. Li G, Lian W, Qu H, Li Z, Zhou Q, Tian J. Improving patient care through the development of a 5G-powered smart hospital. Nat Med 2021 Jun;27(6):936-937. [doi: 10.1038/s41591-021-01376-9] [Medline: 34017134]
19. Leff B, Burton L, Mader SL, et al. Hospital at home: feasibility and outcomes of a program to provide hospital-level care at home for acutely ill older patients. Ann Intern Med 2005 Dec 6;143(11):798-808. [doi: 10.7326/0003-4819-143-11-200512060-00008] [Medline: 16330791]
20. Leong MQ, Lim CW, Lai YF. Comparison of hospital-at-home models: a systematic review of reviews. BMJ Open 2021 Jan 29;11(1):e043285. [doi: 10.1136/bmjopen-2020-043285] [Medline: 33514582]
21. Zhu SY, Chen Y, Wang HH, Xie FJ. Design of control system for an intelligent closestool. In: Huang B, Yao Y, editors. Proceedings of the 5th International Conference on Electrical Engineering and Automatic Control. Lecture Notes in Electrical Engineering: Springer; 2016, Vol. 367:743-749. [doi: 10.1007/978-3-662-48768-6]
22. Peng M, Ding Z, Wang L, Cheng X. Detection of sleep biosignals using an intelligent mattress based on piezoelectric ceramic sensors. Sensors (Basel) 2019 Sep 5;19(18):3843. [doi: 10.3390/s19183843] [Medline: 31492027]

23. Moshayedi AJ, Hosseinzadeh M, Joshi BP, Andani ME. Recognition system for ergonomic mattress and pillow: design and fabrication. IETE J Res 2023 Jan 10:1-19. [doi: 10.1080/03772063.2022.2163927]
24. Huang Y, Chen W, Chen H, Wang L, Wu KS. G-Fall: device-free and training-free fall detection with geophones. Presented at: 2019 16th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON); Jun 10-13, 2019; Boston, MA p. 1-9. [doi: 10.1109/SAHCN.2019.8824827]
25. Puleo V, Gentili A, Failla G, et al. Digital health technologies: a systematic review of their cost-effectiveness. Eur J Public Health 2021 Oct 20;31(Supplement_3). [doi: 10.1093/eurpub/ckab164.273]
26. Marques G, Pires IM, Miranda N, Pitarma R. Air quality monitoring using assistive robots for ambient assisted living and enhanced living environments through Internet of Things. Electronics 2019 Dec 8;8(12):1375. [doi: 10.3390/electronics8121375]
27. Tu J, Wang C, Wu S. The internet hospital: an emerging innovation in China. Lancet Glob Health 2015 Aug;3(8):e445-e446. [doi: 10.1016/S2214-109X(15)00042-X] [Medline: 26187488]
28. Chandrabhatla AS, Pomeraniec IJ, Ksendzovsky A. Co-evolution of machine learning and digital technologies to improve monitoring of Parkinson's disease motor symptoms. NPJ Digit Med 2022 Mar 18;5(1):32. [doi: 10.1038/s41746-022-00568-y] [Medline: 35304579]
29. Liu S, Wang X, Wu Z, He Y. PhysioTreadmill: an auto-controlled treadmill featuring physiological-data-driven visual/audio feedback. In: 2020 International Conference on Cyberworlds (CW) 2020:219-226. [doi: 10.1109/CW49994.2020.00042]
30. Kim E, Baskys A, Law AV, Roosan MR, Li Y, Roosan D. Scoping review: the empowerment of Alzheimer's disease caregivers with mHealth applications. NPJ Digit Med 2021 Sep 7;4(1):131. [doi: 10.1038/s41746-021-00506-4] [Medline: 34493819]
31. Talal M, Zaidan AA, Zaidan BB, et al. Smart home-based IoT for real-time and secure remote health monitoring of triage and priority system using body sensors: multi-driven systematic review. J Med Syst 2019 Jan 15;43(3):42. [doi: 10.1007/s10916-019-1158-z] [Medline: 30648217]

Abbreviations

AI: artificial intelligence
EMR: electronic medical record
HIPAA: Health Insurance Portability and Accountability Act
IoMT: Internet of Medical Things
IoT: Internet of Things
SHW: Smart Home Ward

Edited by L Buis; submitted 18.11.22; peer-reviewed by A Al-Ani, H Veldandi, H Motahari-Nezhad, W Tang; revised version received 19.10.23; accepted 05.12.23; published 30.01.24.

Please cite as:
Cheng W, Cao X, Lian W, Tian J
An Introduction to Smart Home Ward–Based Hospital-at-Home Care in China
JMIR Mhealth Uhealth 2024;12:e44422
URL: https://mhealth.jmir.org/2024/1/e44422
doi:10.2196/44422

Data Collection and Management of mHealth, Wearables, and Internet of Things in Digital Behavioral Health Interventions With the Awesome Data Acquisition Method (ADAM): Development of a Novel Informatics Architecture

I Wayan Pulantara[1], PhD; Yuhan Wang[1], MSIS; Lora E Burke[2,3], MPH, PhD; Susan M Sereika[2,3], PhD; Zhadyra Bizhanova[3], MBChB; Jacob K Kariuki[4], PhD; Jessica Cheng[3], PhD; Britney Beatrice[2], MS, RD; India Loar[2], MPH; Maribel Cedillo[5], MD, MPH; Molly B Conroy[5], PhD; Bambang Parmanto[1], PhD

[1]School of Health and Rehabilitation Science, University of Pittsburgh, Pittsburgh, PA, United States

[2]School of Nursing, University of Pittsburgh, Pittsburgh, PA, United States

[3]School of Public Health, University of Pittsburgh, Pittsburgh, PA, United States

[4]Nell Hodgson Woodruff School of Nursing, Emory University, Atlanta, GA, United States

[5]School of Medicine, University of Utah, Salt Lake City, UT, United States

Corresponding Author:
Bambang Parmanto, PhD
School of Health and Rehabilitation Science, University of Pittsburgh, , Pittsburgh, PA, , United States

Abstract

The integration of health and activity data from various wearable devices into research studies presents technical and operational challenges. The Awesome Data Acquisition Method (ADAM) is a versatile, web-based system that was designed for integrating data from various sources and managing a large-scale multiphase research study. As a data collecting system, ADAM allows real-time data collection from wearable devices through the device's application programmable interface and the mobile app's adaptive real-time questionnaires. As a clinical trial management system, ADAM integrates clinical trial management processes and efficiently supports recruitment, screening, randomization, data tracking, data reporting, and data analysis during the entire research study process. We used a behavioral weight-loss intervention study (SMARTER trial) as a test case to evaluate the ADAM system. SMARTER was a randomized controlled trial that screened 1741 participants and enrolled 502 adults. As a result, the ADAM system was efficiently and successfully deployed to organize and manage the SMARTER trial. Moreover, with its versatile integration capability, the ADAM system made the necessary switch to fully remote assessments and tracking that are performed seamlessly and promptly when the COVID-19 pandemic ceased in-person contact. The remote-native features afforded by the ADAM system minimized the effects of the COVID-19 lockdown on the SMARTER trial. The success of SMARTER proved the comprehensiveness and efficiency of the ADAM system. Moreover, ADAM was designed to be generalizable and scalable to fit other studies with minimal editing, redevelopment, and customization. The ADAM system can benefit various behavioral interventions and different populations.

(JMIR Mhealth Uhealth 2024;12:e50043) doi:10.2196/50043

KEYWORDS

integrated system; IoT integration; wearable; mHealth Fitbit; Nokia; clinical trial management; research study management; study tracking; remote assessment; tracking; Fitbit; wearable devices; device; management; data analysis; behavioral; data collection; Internet of Things; IoT; mHealth; mobile health

Introduction

In recent years, the health care industry has focused a considerable amount of attention on wearable devices because of their capabilities in monitoring and collecting clinically relevant information such as heart rate (HR), blood pressure, respiratory rate, sleep, and physical activity [1]. One of the main drivers of this increased attention is the high level of adoption of wearable devices in the United States. Among all US adults, around 30% have experience with wearable devices and nearly half of those who use their wearable devices regularly are willing to share their data for research purposes [2]. Furthermore, there is an increase in the accuracy offered by these off-the-shelf wearables in providing health-related measurements. For example, several studies have been conducted on consumer wearable devices such as the Apple Watch, Samsung Gear, Fitbit, Huawei Band, and Xiaomi Mi Band, which reported high accuracy in measuring steps,

distance, and HR [3-6]. These trends and the maturity of consumer wearable technologies have led more researchers to use consumer wearable devices in their research studies.

However, this increasing use of wearable devices in research studies also presents challenges. Managing and analyzing the large amounts of data generated by wearable devices in research studies in real time is not a simple task, especially for research teams without dedicated specialized technical support in wearable data management and data engineering. For example, the Fitbit fitness band, one of the most widely used consumer wearable devices in research, has been used to track steps [4,5], physical activity [4-14], HR [15], and sleep patterns [16]. Among the publications we reviewed (n=15), many (n=6, 40%) of the studies using the Fitbit fitness band used a third-party, proprietary, web-based Fitbit data collection and visualization platform—Fitabase [6-9,11,17]; 3 (20%) studies reported that they manually checked and downloaded participants' data from participants' individual Fitbit account at the end of the study [5,12,17]; 1 (7%) study used the MySantéMobile system [13]; and the rest (n=5, 33%) did not report on the system they used for collecting and displaying data from the wearable devices [4,10,14-16]. At the time of writing, Fitabase is the most popular data collection system, having been used in more than 450 institutions and more than 1100 studies [18]. The advantage of using Fitabase is that it can automatically collect data for all participants in real time and provides a basic dashboard function for individual users. However, Fitabase can only access data from Fitbit-related products, and the dashboard cannot be adapted to study-specific needs. For example, although Nokia is one of the widely used personal weight scales in weight-related research studies [19-24], it is not supported by Fitabase. The lack of a data collection and management system that is useful across wearable technologies is a major barrier for researchers who would like to use multiple wearable devices in a study.

Mobile health (mHealth) methods use advanced interactive technologies for the assessment of behaviors and delivery of digital health interventions, an area of research that is growing rapidly. Typical mHealth systems [25-31] consist of a smartphone app for data collection and intervention delivery, sensors integration, and occasionally clinical or administrative portals. However, although mHealth approaches may be readily adopted in behavioral research, it does not encompass end-to-end research management functions such as participant recruitment, screening, randomization, and retention.

A comprehensive and integrated implementation of a clinical trial management system in addition to wearable and mHealth integrations would allow the tracking of the study based on its specified recruitment needs, guarantee that the study starts and ends on time, make sure the necessary action has been taken at participants' specialized milestones, and minimize study costs [32]. However, disjointed implementations and noninteroperable third-party systems have built inseparable barriers for researchers to take advantage of valuable data and track their study process efficiently. The three main barriers are as follows: (1) integrating data from various sources into 1 database poses difficulties [4-6,11,17]; (2) data are seldom ready to be processed for study-specific needs in real time, with previous attempts seeing varying degrees of success [12,14-17,19-24]; and (3) no end-to-end study management systems that address (1) and (2) are readily available on the market. Therefore, we designed and implemented the Awesome Data Acquisition Method (ADAM) system to address these barriers by providing a digital platform that enabled the management and collection of long-term health data from a variety of the Internet of Things (IoT) and wearable devices in an accurately and timely manner, thus allowing an accurate display of tailored and aggregated data in real time. The ADAM system significantly reduced data integration burdens, that is, (1) by having the most common third-party application programming interface (API) integration templates readily available and (2) saving costs in research management areas, such as recruitment, screening, enrollment, assessment scheduling, participant tracking, and recruitment reporting, by streamlining and digitalizing those processes, all of which support an efficient and rigorous study management experience.

General Design

The Iterative and Incremental Development model [33] was used in the development of ADAM. In implementing the Iterative and Incremental Development model, the development team devised a road map outlining core functionalities essential for supporting capabilities that will be mentioned later. Subsequent iterations were planned based on participants' and study coordinators' (end-users') feedback, as well as the latest technological advancements, with each iteration focusing on delivering a functional and deployable software increment. Criteria for transitioning from one increment to the next were determined by predefined milestones, including successful feature implementation, user experience enhancements, and adherence to performance metrics. End-user feedback played a pivotal role throughout the process, guiding feature enhancements and usability improvements, while rigorous testing procedures ensured the stability and reliability of each increment before release. This approach facilitated the creation of a responsive and user-centric software solution that continually evolved to meet user needs throughout the project.

The ADAM system was designed to support the entire management of research studies that include the use of commercial wearable devices and smartphone apps, including designing questionnaires and collecting the response data remotely, recruitment, screening, enrollment, data tracking, data monitoring, and data visualization. Those components were developed iteratively in each development cycle until the desired study flow was reached. The ADAM system consists of 3 main components: a mobile app, a portal, and a connection with the APIs of wearable devices.

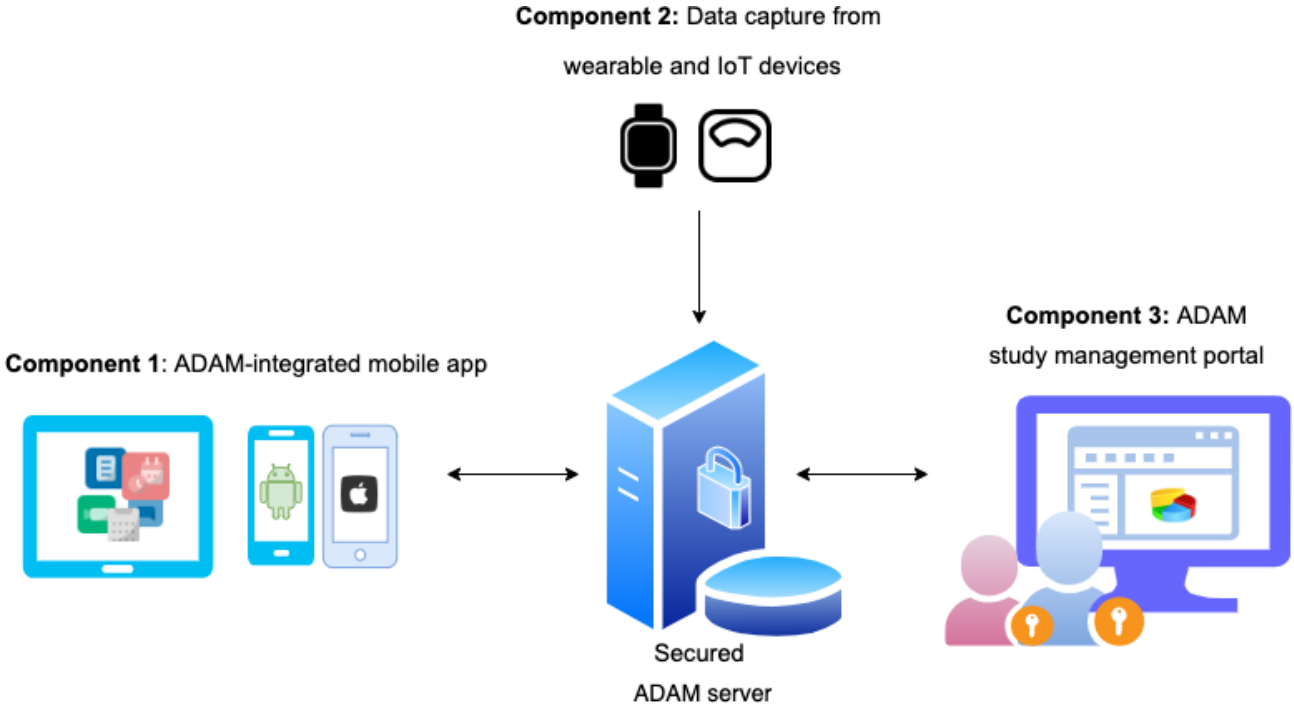
Architecture

The architecture for the ADAM system is illustrated in Figure 1. The 3 main components of ADAM include the ADAM-integrated mobile app, called the SMARTER app in the SMARTER trial, which is a cross-platform app that can run on Android and iOS devices (smartphones or tablets); a data capture component, which is used to capture and track the

study-related data from the wearable and IoT devices for all participants; and the study management portal, which allows study coordinators to collect data and manage the entire study process. These components are connected to a secured ADAM server. We designed the ADAM server to have 1-way data communication with wearable device APIs and 2-way data communication between the ADAM-related mobile app and the clinical portal.

Figure 1 illustrates the architecture of ADAM (the overall system will be referred to as simply ADAM or the ADAM system, to differentiate from its components), which consists of components for data capture from commercial digital wearables and IoT devices through APIs. Currently, ADAM has implemented integration with open APIs from Fitbit and Nokia. Participants' data were automatically collected from the wearable and IoT vendor's servers regularly, with a frequency that was customized to the study. The data transfer between wearable devices, servers, and ADAM was secured through a unique token. Furthermore, all data at transit were secured via Transfer Layer Security (TLS/HTTPS). A similar method of pulling data from other off-the-shelf wearable devices could be used beyond the 2 vendors, Fitbit and Nokia.

Figure 1. The architecture of the Awesome Data Acquisition Method (ADAM) system. IoT: Internet of Things.



We implemented ADAM as a system to support the SMARTER trial, which was conducted to evaluate the efficacy of 2 different approaches to a technology-supported (mHealth, wearable, and IoTs) behavioral intervention for weight loss [34-36]. The study used a randomized controlled trial design involving 502 adult participants who were randomized with equal allocation to 1 of 2 groups: the first group (n=250) received self-monitoring (SM) alone, while the second group (n=250) received remotely delivered feedback messages tailored to the recorded SM data. The main component of the SMARTER study intervention was personalized feedback messages via the ADAM-integrated SMARTER app, an investigator-developed mHealth app. The SMARTER app was connected to the ADAM portal through 2-way secure communications. Overall, ADAM, as a platform, was designed for the delivery of the study's feedback intervention that included secure real-time messaging, adaptive questionnaires that could be delivered promptly, and reminders. As a study tool, we designed the ADAM system to track participants' activities, such as the time when participants received or sent messages related to the study and the time when participants opened the app. All these data have been collected for future analysis.

Another component of the ADAM system was the ADAM portal, which includes a dashboard for study coordinators to review data and engage study participants. Study coordinators were able to get a bird's-eye view of all participants in the study, including a map of their geographical locations. The coordinator was also able to view the detailed data of individual participants, including questionnaire data, screening data, randomization status, and real-time wearable device data. The ADAM portal consists of 8 modules: a wearable device module for pulling data from the Fitbit and Nokia servers to the ADAM server; a questionnaire module for designing and collecting questionnaires; a screening and enrollment module for tracking participants' status in the study; a study phase–tracking module for providing a snapshot of the whole study process; a randomization module for study-customized randomization needs to help the interventionist performed the randomization at the baseline visit [34]; an assessment-scheduling module for remote or in-person assessment alerts and scheduling; a data-tracking and visualization module for reviewing participants' data; and a retention management module for follow-up assessment data collection and remuneration management. All features worked closely with each other throughout the study data flow.

Capabilities

Customized Questionnaire Design and Remote Assessment Collection

Questionnaire-based assessment is one of the most common methods of data collection in research settings. Therefore, the ADAM portal was designed around this need, from the process of questionnaire and assessment design to the remote delivery of questionnaires. Designing a questionnaire based on the study needs was made simpler for researchers with the inclusion of an easy-to-use, drag-and-drop interface and a library of frequently used question templates. The adaptive questionnaire feature in ADAM was designed to be used by study coordinators and investigators with no need to do programming, with a similar ease of use to that of proprietary solutions such as Qualtrics, but unlike most of these proprietary solutions, ADAM allowed integrations with an mHealth app. The ADAM portal also supported more advanced questionnaire features such as input-type validation, error messages, required field validation, and question branching. After a questionnaire was published in the ADAM portal, depending on the purpose, a public link or a participant-specific private link was generated for study use. Once the participant had answered and submitted this questionnaire via the given link, the ADAM portal automatically formatted the answers, placed time stamps, calculated scores, and triggered any necessary flags and status updates that could be immediately reviewed by the study coordinators. As the ADAM portal was designed to support multiphase study needs, questionnaires could be generated for each study phase while considering different restrictions for each phase, for example, adding certain filtering restrictions in the screening phase for certain criteria in the screening phase's questionnaire. This questionnaire module can also be adapted to fit the need for remote participant consent delivery, as the study consent form usually shares the same structure and data requirements as an assessment questionnaire.

Study Management and Multiphase Study Tracking

The ADAM portal also supported the management of other steps in the research study, which commonly included multiple phases or processes such as recruitment, enrollment, engagement, and retention. For the recruitment phase, the ADAM portal supported a multistep screening process. Each sequential screening step could be triggered by the answers or the calculated scores of the current step's screening or assessment questionnaire. It could also support manual evaluation by study coordinators. Once screening decisions had been made, the system automatically moved the eligible participants to the next phase in the ADAM system. This real-time process reduced the study coordinators' manual data review and data calculation workload. It also helped reduce the study cost by improving the overall recruitment process's efficiency.

For the enrollment and engagement phases, the ADAM portal has a built-in account management system that can assign a unique study-specific ID for each participant for efficient progress tracking while protecting their identifiable information. Furthermore, these IDs can also be used as common IDs linking data coming in from outside sources such as the Fitbit and Nokia APIs. With the benefit of the study-specific ID, the ADAM portal can easily record participants' digital footprints within the system and track the milestones and progress of all participants within the study. The portal also provides the study coordinator with access to aggregated data, such as that of important milestones or sorted by group, which is represented in a CONSORT (Consolidated Standards of Reporting Trials) flow diagram. Such a flow diagram is valuable for study management in 3 aspects: it allows the study coordinator to switch the recruitment effort to a specific group of people during recruitment in cases where a group has a less-than-optimal recruitment rate; it facilitates the adjustment of the speed of recruitment to help make sure the study can start and end on time; and it can monitor the screening process and rate of attrition, which may provide some hints for adjustments that may be needed in screening methods.

For the retention process, the ADAM portal provides a dashboard for showing study-related alerts or flags and participants' performance according to the data gathered from wearable devices, the digital scale, and the study app. Using this dashboard, the study coordinators have a near real-time report on whether participants are following the study protocol strictly, in other words, adhering to the study; as a result, they can intervene or take alleviating actions at a point closer to when participants began to show a trend toward lower adherence. This feature also helps to estimate the milestones for participants' assessments. Other features are study compensation and remuneration management, which were considered indispensable parts of the ADAM system. Study coordinators used the remuneration feature of the ADAM portal to record all payments made or planned at different time points in the study.

Data Synchronization With Wearable and IoT Devices

Data synchronization with wearable and IoT devices is one of the key features of the ADAM system. As more people rely on commercially available wearable and IoT devices to get their health data and track their fitness, a growing number of investigators have used this type of off-the-shelf wearable device as part of their studies. This trend indicates the need for a study management system that can integrate IoT and wearable data from multiple providers and vendors. Currently, the ADAM portal provides integration with Fitbit (wearable) and Nokia (IoT), since these companies have the largest market share, a secure authentication process, and stable open APIs. The steps to integrate data from wearables in ADAM from the users' (study coordinators') perspective are made simple. Study coordinators first need to create participant accounts in the ADAM portal and get participants' permission and authorization to collect their data from outside APIs. Once ADAM has been granted authorization, data synchronization for the given participant will be scheduled daily. The ADAM portal also supports subscription-based data integration, where any new data that become available in the wearable will trigger synchronization or a data import into the ADAM database in a real-time fashion, although we did not implement this feature in the SMARTER study.

Randomization

The ADAM system provided randomization by minimization, a more advanced randomization method, which was used to ensure the balance of prognostic factors between treatment groups and set equal treatment allocations [37].

Real-Time Data Analysis for Study-Related Apps

ADAM adopts Firebase Analytics libraries to anonymously monitor app use and statistics. This part of the data mainly describe the use of the app within a period. The study team can obtain a real-time visualization on the ADAM main page. This information can provide an overview of the app use among all study participants and additional background information for the efficiency of message delivery.

Evaluation of the ADAM Platform in the SMARTER Trial

Overview

An evaluation of the ADAM platform was conducted as part of a weight-loss clinical trial called the SMARTER trial [34]. This study was registered on ClinicalTrials.gov (NCT033677936) and approved by the Institutional Review Board of the University of Pittsburgh. The SMARTER trial was a randomized controlled trial that followed a single-site, 2-group design, with a 1-year study duration for both comparison and intervention groups [34]. In the SMARTER trial, participants were allocated equally to the 2 treatment conditions: SM only or SM with real-time tailored feedback messages (SM+feedback) [34]. The study provided all participants with a Fitbit Charge 2 and a Nokia Scale for physical activity and weight SM; participants used the Fitbit app for dietary SM. The SMARTER app generated 3 tailored messages per day responding to the SM data provided, for example, diet, physical activity, and weight. More details for the SMARTER trial design and screening criteria can be found in the study design paper [34].

Recruitment and Screening

The SMARTER trial's recruitment began in August 2018, and the trial itself was completed in March 2021. This study was designed to include a 3-phase screening process. The first 2 screenings were based on web-based questionnaires, where the ADAM portal provided autoscreening services based on the answers submitted by the participants. As shown in Figure 2, there were 3 steps for generating a questionnaire: Figure 2A lists all the questionnaires and consent forms from the SMARTER trial that were available on the ADAM portal; Figure 2B shows the straightforward drag-and-drop interface to design or edit a questionnaire; and Figure 2C shows that the ADAM portal could autocreate a specific questionnaire link for each participant on the user list page in the ADAM portal. In the SMARTER trial, phase-1 and phase-2 questionnaires were the study's basic screening surveys, and the ADAM portal provided automatic screening of the candidates based on preset criteria such as aged ≥18 years, BMI between 27-43 kg/m^2, smartphone user, general health history, lifestyle, and health and medical history. In phase 3, the ADAM portal still handled the display, submission, and scoring processes of the questionnaire but also gave the screening decisions back to the study coordinators. In this phase, participants were asked to complete a 5-day SM food diary using the Fitbit app. As illustrated in Figure 3, in our ADAM portal, the food diary data were automatically pulled from the Fitbit server for the study coordinators to review. Collecting data in one place relieved the pressure on study coordinators from switching back and forward between several different data management systems and tools.

Figure 2. Customized questionnaire module of the ADAM portal: (A) questionnaire list, (B) edit or create a questionnaire, and (C) personal link for the questionnaire. ADAM: Awesome Data Acquisition Method.

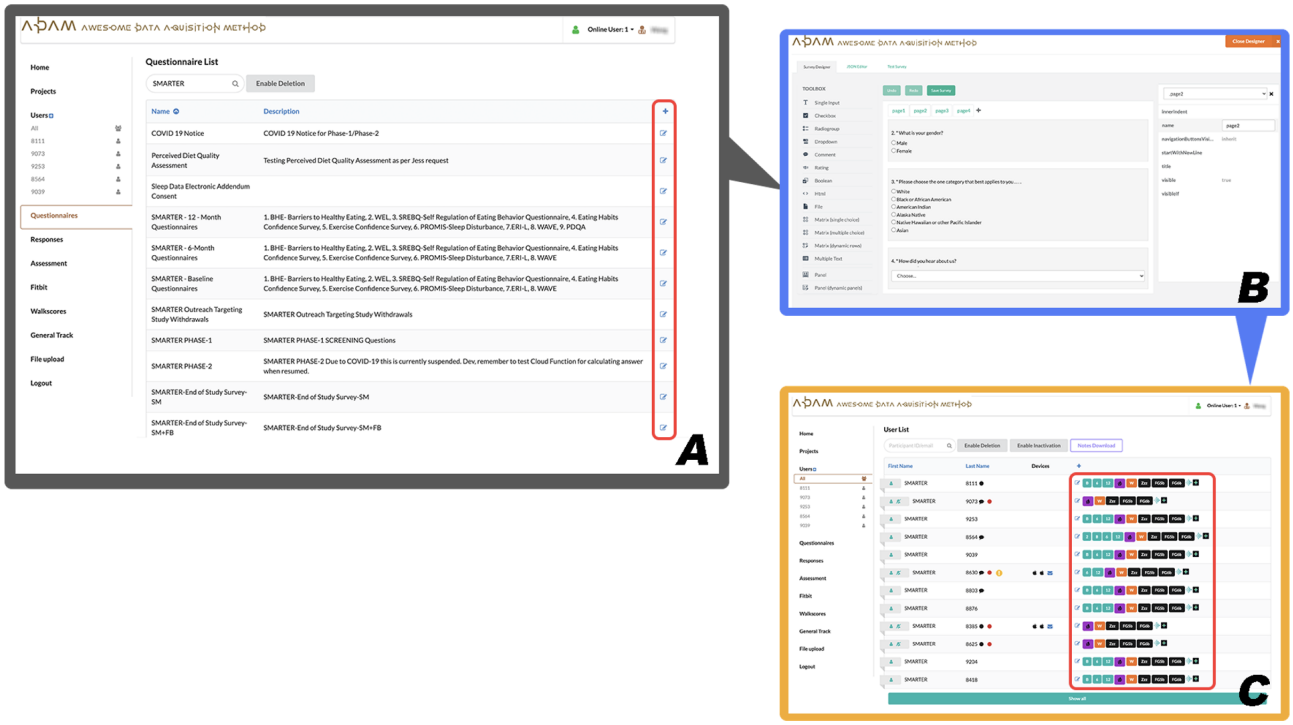
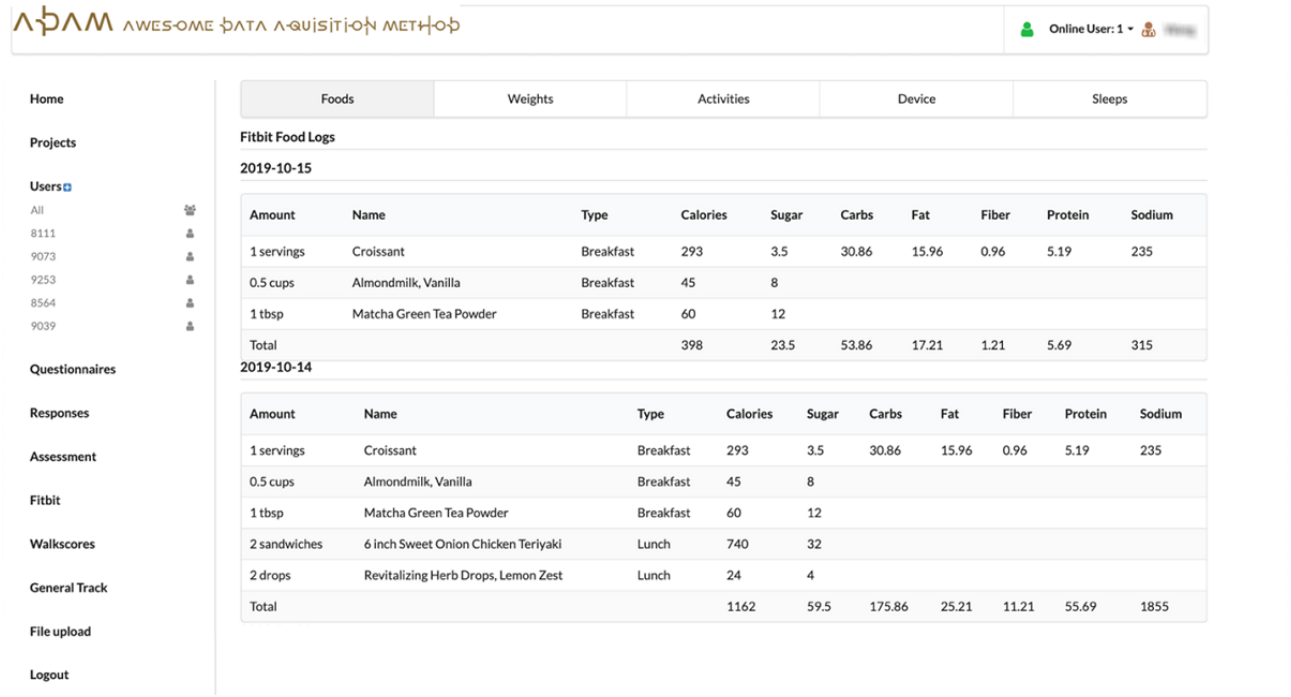


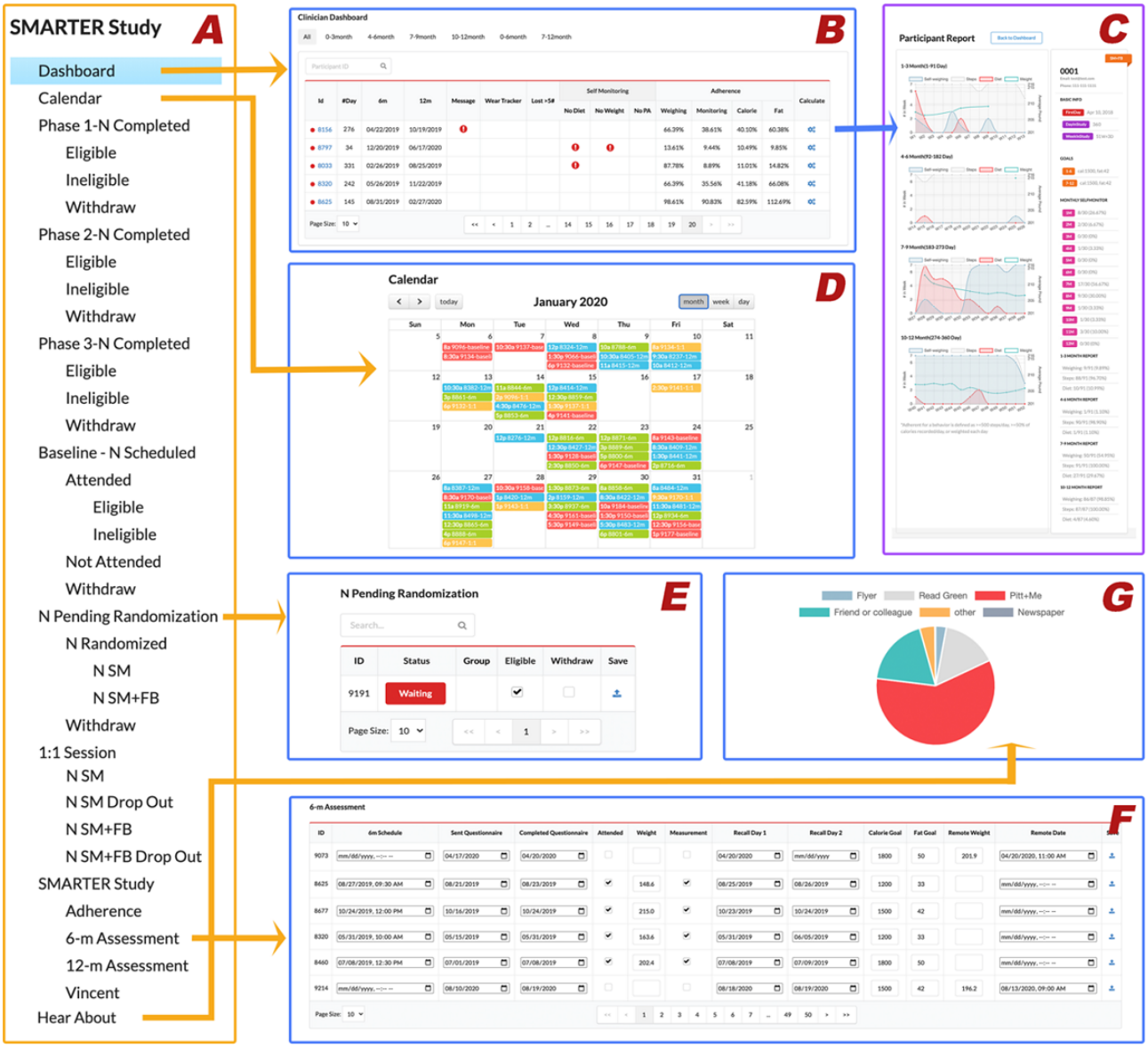
Figure 3. Food diary data collected from Fitbit. ADAM: Awesome Data Acquisition Method.



Randomization and Enrollment

As displayed in Figure 4A, all the SMARTER trial recruitment steps were visualized and cataloged in the ADAM system. By clicking on each recruitment step, study coordinators could review the current remaining number of participants in that phase and make decisions on moving them to the next phase of the study. Once participants were ready for the randomization step, they would be displayed on the randomization page, as shown in Figure 4E. The ADAM system provided a minimization randomization algorithm to support this randomization process. After randomization, participants' Fitbit and Nokia accounts were linked to our ADAM system by study coordinators; all the Fitbit-related and Nokia-related data including physical activity, food diary, and self-weighing data could then be collected by the ADAM system daily for the following 1-year intervention period.

Figure 4. The general data-tracking module for SMARTER trial management in the ADAM portal: (A) recruitment steps, (B) clinician dashboard, (C) participant report, (D) calendar, (E) randomization page, (F) 6-month assessment page, and (G) referral breakdown. ADAM: Awesome Data Acquisition Method.



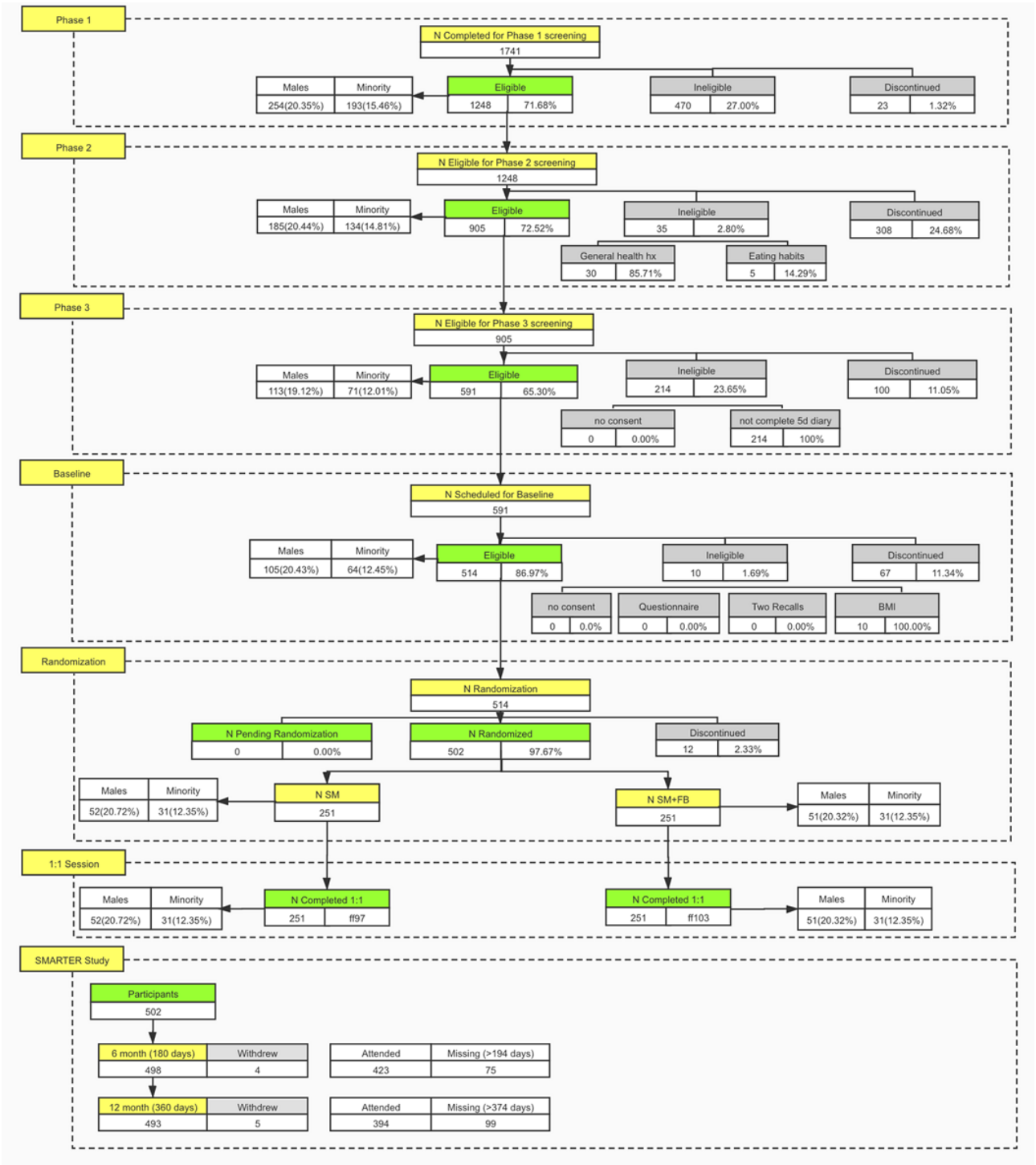
Engagement and Retention

After randomization, participants were asked to do daily SM during the 1-year study and complete 3 assessments at baseline, 6 months, and 12 months. During the 1-year study, participants were asked to adhere to the daily SM protocol by wearing a Fitbit tracker to record their daily physical activities, report food intake using the Fitbit app, and weigh themselves using a smart scale. All data were synchronized to the ADAM system and were displayed in a raw format, as shown in Figure 3, as well as a study-tailored precalculated report format, as shown in Figure 4B and C. The *clinical dashboard* in Figure 4B acts as a valuable resource for study coordinators to track participants' adherence to all SM protocols and to receive alerts if participants lose weight too rapidly (a precaution to prevent unwanted clinical effects) or have poor study adherence. Study coordinators used this page to mark withdrawals at any time during their study period. The ADAM portal then automatically recorded and marked the withdrawal date appropriately. The trial included 3 assessments that heavily used the ADAM portal as a method for appointment scheduling and web-based survey completion. Figure 4F represents an example of a 6-month assessment page, and Figure 4D is an all-in-one assessment calendar. Due to the COVID-19 pandemic, in-person assessments were no longer permitted as of March 2020. The study team made the decision to change all the assessments to be conducted remotely and to obtain the primary outcome of weight directly from participants' smart scales in their homes instead of the usual scale at the research center. Due to the remote-ready adaptive design of the ADAM system, this COVID-19–triggered transition was able to be performed quickly and with almost no interruption to the study flow.

The ADAM portal was also used for participants' compensation and reimbursement management. For example, once the payment was made, study coordinators recorded this on the Adam portal's remuneration page. Another example of a useful feature provided by the ADAM portal for the trial is a real-time flowchart (Figure 5). The flowchart provided an overview of

the entire study at any given time, which allowed for an adaptive recruitment strategy and effective adjustments to study resource allocation as needed. Furthermore, using the ADAM portal, the research staff coordinated activities with each other via a single platform remotely, since everything related to the study was accessible in real time and on the web.

Figure 5. Real-time study flowchart. FB: feedback; SM: self-monitoring.



SMARTER App

The SMARTER app is an ADAM-integrated app that was designed for the SMARTER trial. Only the intervention group had access to the app. Figure 6 shows the SMARTER app use data based on reports provided by Google's Firebase Analytics, which included aggregated data for the following aspects: number of daily active users, popular app versions, most common mobile devices among participants, and active users by city. This report worked as supporting information for the study team to track app use. As for the app's content, Figure 7 shows sample messages tailored and delivered by the SMARTER app. The ADAM system tracked and logged these messages' time stamps and status on whether the messages were delivered, opened, or missed—information for future data analysis that may provide auxiliary insight into the overall outcome of the trial.

Figure 6. SMARTER app use. ADAM: Awesome Data Acquisition Method.

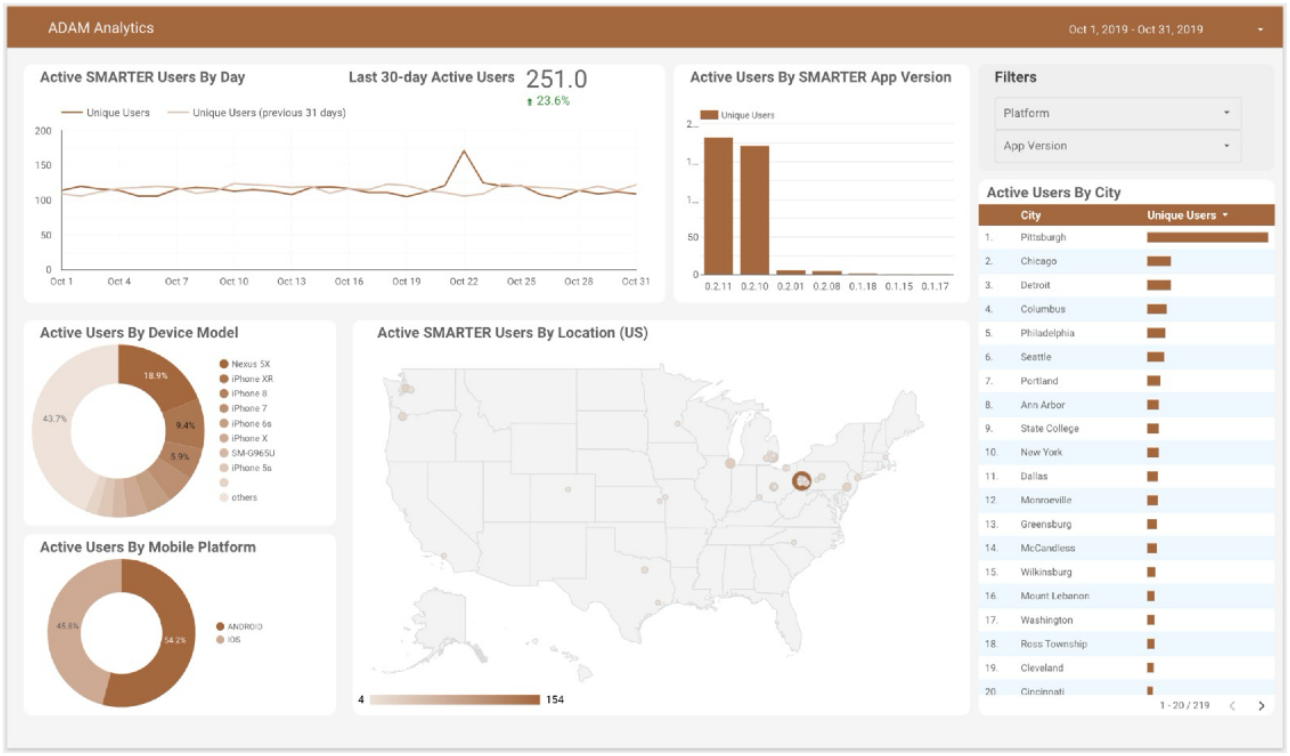
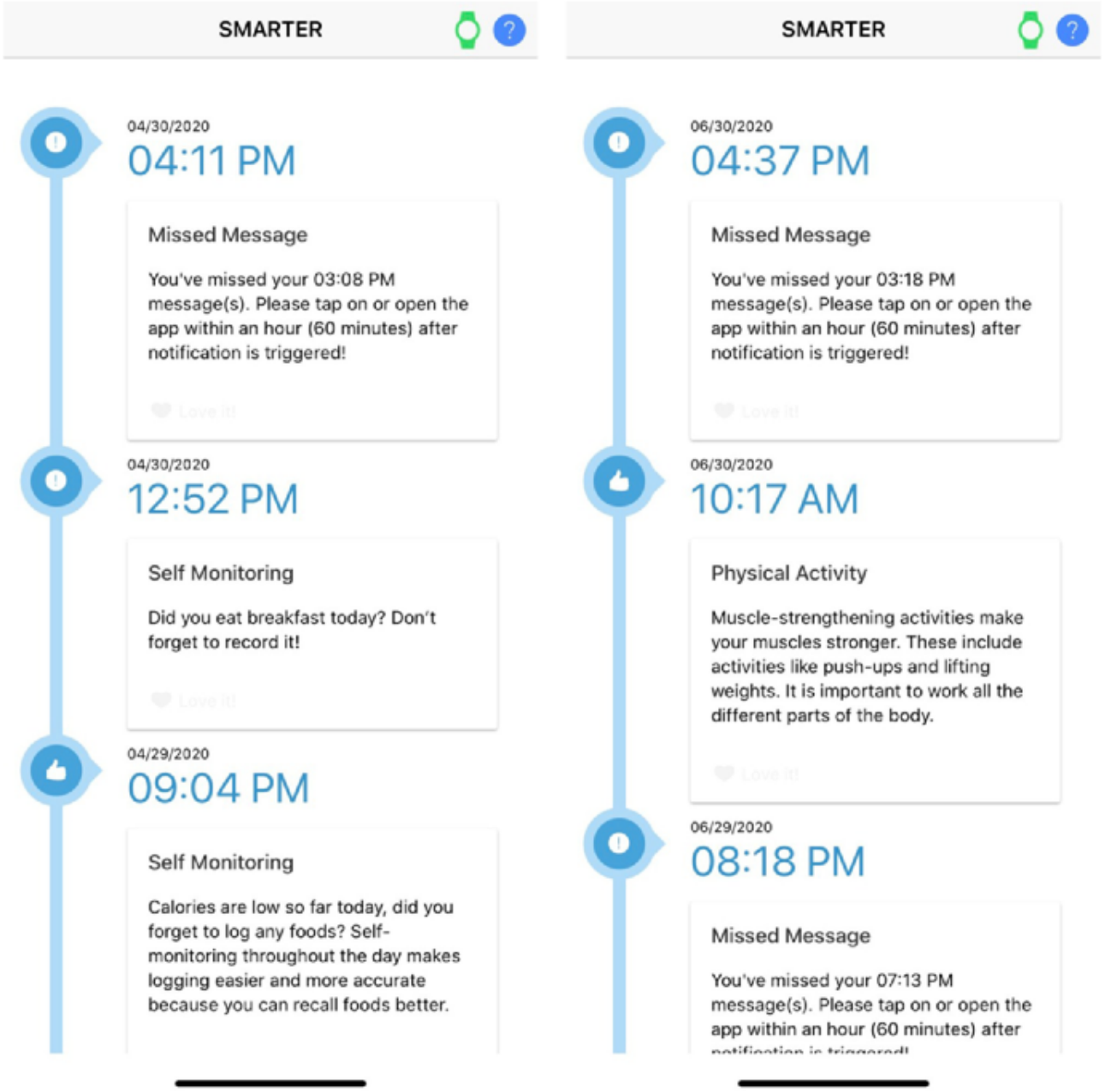


Figure 7. Screenshots of the SMARTER app.



Data Validation and Integrity

The ADAM system incorporated 4 monitoring mechanisms to ensure the quality of collected data. First, as illustrated in Figure 4B on the clinical dashboard, ADAM can generate alerts. To guarantee data quality from various sources, the system sends automatic alerts for the following reasons: (1) if there are no food data recorded for 7 consecutive days; (2) if no weight data are recorded for 7 days or if there is a recorded weight change exceeding 5 lbs within a 7-day period; and (3) if there are no Fitbit tracker data received for 7 consecutive days. Second, the system performed automatic checks on API connections and promptly reported any issues to the study team. Third, to further ensure data integrity, study coordinators manually checked and reviewed selected records during the in-person assessments. Fourth, before data analysis, the research team also ran an algorithm to check for outliers in the collected data.

The SMARTER Trial Evaluation Results

In the SMARTER trial, 1741 unique individuals submitted their phase-1 survey through the ADAM portal; 1248 participants were eligible for the phase-2 screening; 905 participants were eligible for the phase-3 screening; and 591 participants were eligible for the baseline in-person assessment. After the baseline assessment, 502 eligible participants were randomized to 1 of the 2 treatment conditions. During the 12-month study period, 4 participants officially withdrew before their 6-month assessment, and 5 withdrew before the final 12-month assessment, all due to health issues. Even including the COVID-19 pandemic, the overall retention rate for the 6-month assessment was 84.93% (423/498) and for the 12-month assessment was 79.92% (394/493). The ADAM system efficiently coordinated study information and supported the study and study team, especially in a situation of limited human

resources and having to work fully remotely. It helped to recruit, screen, maintain, and track all individuals wishing to participate in the study; scheduled more than 1500 in-person assessments; and received 4131 questionnaire answers from the screening process to the study completion. Using the ADAM system allowed us to achieve every milestone in the study and minimize the disruptions by the COVID-19 pandemic.

Discussion

Principal Findings

We introduced the architecture and implementation of a data collecting and clinical trial management system called ADAM as a solution to current barriers in clinical or behavioral studies involving wearables, IoT, and mHealth. The ADAM system's adoption of the newest available interfaces in integration with multiple wearable devices makes it an attractive choice for behavioral and population-based health intervention studies. Using the ADAM system, the study team in the SMARTER trial no longer needed to switch between different systems or documents when working with study data from diverse sources and avoided most data missingness and data synchronization issues. ADAM's study management functions supported recruitment, screening, and randomization and helped maintain retention with a dashboard and flagging. In addition, it provided an efficient and manageable environment for study tracking and adjustment as necessary. ADAM's study flowchart was extensively used in the team's weekly meetings and contributed widely to decision-making during the entire study period.

Limitations

The system evaluations throughout the SMARTER trial indicated that the ADAM system was able to meet most of the data collection and study management demands. However, the ADAM system was not without limitations. The first limitation was not due to the ADAM system per se, but due to the limited availability of open APIs for third-party data collection sources also used in the study; therefore, the ADAM system had issues including all datasets for the study. For example, in the SMARTER trial, study participants needed to complete two 24-hour dietary recalls using the ASA24 system, a dietary assessment tool, at 3 assessment points. We could not directly synchronize the dietary recalls from the ASA24 system with the ADAM system due to the API limitations of the ASA24 platform. This limitation prevented study coordinators from receiving real-time data visualization and data analysis in the ADAM system for ASA24's data. Therefore, all the data from ASA24 needed to be manually downloaded and uploaded to the ADAM system. Although the ADAM system can integrate more datasets and be an all-in-one data solution for the study, we still need to consider approaches that may have API restrictions or the unavailability of such APIs by different companies and organizations that provide additional data required by the study.

The second limitation to consider is the data missingness and data synchronization issues with Fitbit or Nokia devices. Fitbit's open web API limits the number of API calls per hour for each participant or Fitbit account to 100. This stringent quota restricted the number of days that we could pull the data from the Fitbit server, as we needed to have at least a single call for each day of data. Furthermore, specifically for the SMARTER trial, we were restricted to pulling data only for the last 7 days. Since we were pulling data from several different areas of Fitbit's API, such as physical activity data, diary data, and sleep data, it quickly consumed the quota for the API calls. At the same time, from time to time, some Fitbit and Nokia devices experienced issues sending data to the Fitbit server when there were internet or Bluetooth connection problems or participants failed to open the Fitbit or HealthMate app for an extended amount of time. Those situations caused a data transfer delay and ultimately led to data missingness and failed data synchronization in the ADAM system.

Furthermore, the version of ADAM used in the SMARTER trial did not have the capabilities to send emails, SMS text messages, or push notifications. Therefore, the interventionists and study coordinators needed to use different modalities of communication and tools to send emails or SMS text messages to the participants. The next iteration of ADAM will include these capabilities to streamline communications and monitoring within one system, ADAM. This would increase the efficiency and the general ease of using the system in reaching participants as needed.

Finally, the ADAM system is more suitable for digital intervention studies using consumer technologies. If the study is highly dependent on in-clinic assessments using offline tools and closed or proprietary systems, the ADAM system provides a limited benefit since the data may be hard or even impossible to transmit to the ADAM system. In addition, studies that target recruitment groups who are unfamiliar with mobile technology or who have limited access to the internet could experience barriers to the adoption of the ADAM system technology. According to the questionnaires, these groups represent a small portion of potential study participants, but this is an important limitation to consider in future studies that may target these groups. This limitation was not an issue in the SMARTER trial, as the study was designed to exclude participants who were not smartphone users.

To address these limitations, first, we are considering supporting additional, widely used wearable devices. Second, we plan to generalize the study management components and make them easier to set up and adapt to distinct types of study design. However, overall, the ADAM system provides a solution to digitalizing a research study by providing and integrating multiple methods for collecting study data and ensuring an efficient and compliant study process.

Acknowledgments

The trial has been clinically registered at ClinicalTrials.gov with the identifier NCT03367936. This research was supported by grants from the National Institutes of Health (NIH): R01HL131583 (LEB), R01HL1311583S (LEB, JKK), F31HL156278 (JC), and T32HL098048 (JC). It was also supported by a grant from the University of Pittsburgh Clinical Translational Science Institute to Dr. S. Reis (UL1TR001857).

Conflicts of Interest

None declared.

References

1. Guk K, Han G, Lim J, et al. Evolution of wearable devices with real-time disease monitoring for personalized healthcare. Nanomaterials (Basel) 2019 May 29;9(6):813. [doi: 10.3390/nano9060813] [Medline: 31146479]
2. Chandrasekaran R, Katthula V, Moustakas E. Patterns of use and key predictors for the use of wearable health care devices by US adults: insights from a national survey. J Med Internet Res 2020 Oct 16;22(10):e22443. [doi: 10.2196/22443] [Medline: 33064083]
3. Xie J, Wen D, Liang L, Jia Y, Gao L, Lei J. Evaluating the validity of current mainstream wearable devices in fitness tracking under various physical activities: comparative study. JMIR Mhealth Uhealth 2018 Apr 12;6(4):e94. [doi: 10.2196/mhealth.9754] [Medline: 29650506]
4. Sushames A, Edwards A, Thompson F, McDermott R, Gebel K. Validity and reliability of Fitbit Flex for step count, moderate to vigorous physical activity and activity energy expenditure. PLoS One 2016 Sep 2;11(9):e0161224. [doi: 10.1371/journal.pone.0161224] [Medline: 27589592]
5. Alharbi M, Bauman A, Neubeck L, Gallagher R. Validation of Fitbit-Flex as a measure of free-living physical activity in a community-based phase III cardiac rehabilitation population. Eur J Prev Cardiol 2016 Sep;23(14):1476-1485. [doi: 10.1177/2047487316634883] [Medline: 26907794]
6. Semanik P, Lee J, Pellegrini CA, Song J, Dunlop DD, Chang RW. Comparison of physical activity measures derived from the Fitbit Flex and the ActiGraph GT3X+ in an employee population with chronic knee symptoms. ACR Open Rheumatol 2020 Jan;2(1):48-52. [doi: 10.1002/acr2.11099] [Medline: 31943970]
7. Ward S, Orme M, Zatloukal J, Singh S. Adherence to walking exercise prescription during pulmonary rehabilitation in COPD with a commercial activity monitor: a feasibility trial. BMC Pulm Med 2021 Jan 18;21(1):30. [doi: 10.1186/s12890-021-01406-9] [Medline: 33461515]
8. Harrell C. Knee replacement patient activity [Presentation]. Presented at: South Carolina Junior Academy of Science (SCJAS) 2020; Mar 28, 2020; New York, NY p. 164 URL: https://scholarexchange.furman.edu/scjas/2020/all/164 [accessed 2024-07-22]
9. Franzen-Castle L, Dunker T, Chai W, Krehbiel M. Fitbit and Fitabase technology: tracking and evaluating youth physical activity. J Ext 2017;55(2):5. [doi: 10.34068/joe.55.02.05]
10. Cadmus-Bertram LA, Marcus BH, Patterson RE, Parker BA, Morey BL. Randomized trial of a Fitbit-based physical activity intervention for women. Am J Prev Med 2015 Sep;49(3):414-418. [doi: 10.1016/j.amepre.2015.01.020] [Medline: 26071863]
11. Hartman SJ, Nelson SH, Weiner LS. Patterns of Fitbit use and activity levels throughout a physical activity intervention: exploratory analysis from a randomized controlled trial. JMIR Mhealth Uhealth 2018 Feb 5;6(2):e29. [doi: 10.2196/mhealth.8503] [Medline: 29402761]
12. Naslund JA, Aschbrenner KA, Bartels SJ. Wearable devices and smartphones for activity tracking among people with serious mental illness. Ment Health Phys Act 2016 Mar;10:10-17. [doi: 10.1016/j.mhpa.2016.02.001] [Medline: 27134654]
13. Hermsen S, Moons J, Kerkhof P, Wiekens C, de Groot M. Determinants for sustained use of an activity tracker: observational study. JMIR Mhealth Uhealth 2017 Oct 30;5(10):e164. [doi: 10.2196/mhealth.7311] [Medline: 29084709]
14. Chung AE, Skinner AC, Hasty SE, Perrin EM. Tweeting to health: a novel mHealth intervention using Fitbits and Twitter to foster healthy lifestyles. Clin Pediatr (Phila) 2017 Jan;56(1):26-32. [doi: 10.1177/0009922816653385] [Medline: 27317609]
15. Benedetto S, Caldato C, Bazzan E, Greenwood DC, Pensabene V, Actis P. Assessment of the Fitbit Charge 2 for monitoring heart rate. PLoS One 2018 Feb 28;13(2):e0192691. [doi: 10.1371/journal.pone.0192691] [Medline: 29489850]
16. de Zambotti M, Goldstone A, Claudatos S, Colrain IM, Baker FC. A validation study of Fitbit Charge 2™ compared with polysomnography in adults. Chronobiol Int 2018 Apr;35(4):465-476. [doi: 10.1080/07420528.2017.1413578] [Medline: 29235907]
17. Cadmus-Bertram L, Marcus BH, Patterson RE, Parker BA, Morey BL. Use of the Fitbit to measure adherence to a physical activity intervention among overweight or obese, postmenopausal women: self-monitoring trajectory during 16 weeks. JMIR Mhealth Uhealth 2015 Nov 19;3(4):e96. [doi: 10.2196/mhealth.4229] [Medline: 26586418]
18. Fitabase customers. Fitabase. URL: https://www.fitabase.com/how-it-works/customers [accessed 2024-07-22]

19. Alencar M, Johnson K, Gray V, Mullur R, Gutierrez E, Dionico P. Telehealth-based health coaching increases m-Health device adherence and rate of weight loss in obese participants. Telemed J E Health 2020 Mar;26(3):365-368. [doi: 10.1089/tmj.2019.0017] [Medline: 30994410]
20. Campo D, Khettab H, Yu R, et al. Measurement of aortic pulse wave velocity with a connected bathroom scale. Am J Hypertens 2017 Sep 1;30(9):876-883. [doi: 10.1093/ajh/hpx059] [Medline: 28520843]
21. Benson L, Zhang F, Espel-Huynh H, Wilkinson L, Lowe MR. Weight variability during self-monitored weight loss predicts future weight loss outcome. Int J Obes (Lond) 2020 Jun;44(6):1360-1367. [doi: 10.1038/s41366-020-0534-6] [Medline: 31949298]
22. Sperrin M, Rushton H, Dixon WG, et al. Who self-weighs and what do they gain from it? a retrospective comparison between smart scale users and the general population in England. J Med Internet Res 2016 Jan 21;18(1):e17. [doi: 10.2196/jmir.4767] [Medline: 26794900]
23. Vuorinen AL, Helander E, Pietilä J, Korhonen I. Frequency of self-weighing and weight change: cohort study with 10,000 smart scale users. J Med Internet Res 2021 Jun 28;23(6):e25529. [doi: 10.2196/25529] [Medline: 34075879]
24. Dutton GR, Kinsey AW, Howell CR, et al. The daily Self-Weighing for Obesity Management in Primary Care study: rationale, design and methodology. Contemp Clin Trials 2021 Aug;107:106463. [doi: 10.1016/j.cct.2021.106463] [Medline: 34082075]
25. Parmanto B, Pramana G, Yu DX, Fairman AD, Dicianno BE, McCue MP. iMHere: a novel mHealth system for supporting self-care in management of complex and chronic conditions. JMIR Mhealth Uhealth 2013 Jul 11;1(2):e10. [doi: 10.2196/mhealth.2391] [Medline: 25100682]
26. Pulantara IW, Parmanto B, Germain A. Development of a just-in-time adaptive mHealth intervention for insomnia: usability study. JMIR Hum Factors 2018 May 17;5(2):e21. [doi: 10.2196/humanfactors.8905] [Medline: 29773529]
27. Alnosayan N, Chatterjee S, Alluhaidan A, Lee E, Houston Feenstra L. Design and usability of a heart failure mHealth system: a pilot study. JMIR Hum Factors 2017 Mar 24;4(1):e9. [doi: 10.2196/humanfactors.6481] [Medline: 28341615]
28. Yu DX, Parmanto B, Dicianno BE, Pramana G. Accessibility of iMHere smartphone apps for self-care. Presented at: 2014 IEEE International Conference on Healthcare Informatics (ICHI); Sep 15-17, 2014; Verona, Italy p. 367. [doi: 10.1109/ICHI.2014.62]
29. Pulantara IW, Parmanto B, Germain A. Clinical feasibility of a just-in-time adaptive intervention app (iREST) as a behavioral sleep treatment in a military population: feasibility comparative effectiveness study. J Med Internet Res 2018 Dec 7;20(12):e10124. [doi: 10.2196/10124] [Medline: 30530452]
30. Shuwandy ML, Zaidan BB, Zaidan AA, Albahri AS. Sensor-based mHealth authentication for real-time remote healthcare monitoring system: a multilayer systematic review. J Med Syst 2019 Jan 6;43(2):33. [doi: 10.1007/s10916-018-1149-5] [Medline: 30612191]
31. Georgsson M, Staggers N. Quantifying usability: an evaluation of a diabetes mHealth system on effectiveness, efficiency, and satisfaction metrics with associated user characteristics. J Am Med Inform Assoc 2016 Jan;23(1):5-11. [doi: 10.1093/jamia/ocv099] [Medline: 26377990]
32. Gupta A, Calfas KJ, Marshall SJ, et al. Clinical trial management of participant recruitment, enrollment, engagement, and retention in the SMART study using a marketing and information technology (MARKIT) model. Contemp Clin Trials 2015 May;42:185-195. [doi: 10.1016/j.cct.2015.04.002] [Medline: 25866383]
33. Larman C, Basili VR. Iterative and incremental developments. a brief history. Computer 2003 Jun;36(6):47-56. [doi: 10.1109/MC.2003.1204375]
34. Burke LE, Sereika SM, Parmanto B, et al. The SMARTER trial: design of a trial testing tailored mHealth feedback to impact self-monitoring of diet, physical activity, and weight. Contemp Clin Trials 2020 Apr;91:105958. [doi: 10.1016/j.cct.2020.105958] [Medline: 32087342]
35. Burke LE, Sereika SM, Bizhanova Z, et al. The effect of tailored, daily, smartphone feedback to lifestyle self-monitoring on weight loss at 12 months: the SMARTER randomized clinical trial. J Med Internet Res 2022 Jul 5;24(7):e38243. [doi: 10.2196/38243] [Medline: 35787516]
36. Burke LE, Sereika SM, Parmanto B, et al. Effect of tailored, daily feedback with lifestyle self-monitoring on weight loss: the SMARTER randomized clinical trial. Obesity (Silver Spring) 2022 Jan;30(1):75-84. [doi: 10.1002/oby.23321] [Medline: 34898011]
37. Han B, Enas NH, McEntegart D. Randomization by minimization for unbalanced treatment allocation. Stat Med 2009 Nov 30;28(27):3329-3346. [doi: 10.1002/sim.3710] [Medline: 19739238]

Abbreviations

ADAM: Awesome Data Acquisition Method
API: application programming interface
CONSORT: Consolidated Standards of Reporting Trials
HR: heart rate
IoT: Internet of Things

mHealth: mobile health
SM: self-monitoring

Edited by L Buis; submitted 17.06.23; peer-reviewed by J Bowen, J Xie; revised version received 30.04.24; accepted 05.06.24; published 07.08.24.

Please cite as:
Pulantara IW, Wang Y, Burke LE, Sereika SM, Bizhanova Z, Kariuki JK, Cheng J, Beatrice B, Loar I, Cedillo M, Conroy MB, Parmanto B
Data Collection and Management of mHealth, Wearables, and Internet of Things in Digital Behavioral Health Interventions With the Awesome Data Acquisition Method (ADAM): Development of a Novel Informatics Architecture
JMIR Mhealth Uhealth 2024;12:e50043
URL: https://mhealth.jmir.org/2024/1/e50043
doi:10.2196/50043

Viewpoint

SOMAScience: A Novel Platform for Multidimensional, Longitudinal Pain Assessment

Chloe Zimmerman Gunsilius[1,2,3*], BA; Joseph Heffner[4*], PhD; Sienna Bruinsma[1,5], BA; Madison Corinha[1], BSc; Maria Cortinez[3], BSc; Hadley Dalton[6], BA; Ellen Duong[6], MSc; Joshua Lu[6], BSc; Aisulu Omar[6], MSc; Lucy Long Whittington Owen[1], PhD; Bradford Nazario Roarr[6], BA; Kevin Tang[7], BA; Frederike H Petzschner[1,8,9], PhD

[1]Robert J. and Nancy D. Carney Institute for Brain Science, Brown University, Providence, RI, United States

[2]Neuroscience Graduate Program, Department of Neuroscience, Brown University, Providence, RI, United States

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

[4]Department of Cognitive, Linguistic, and Psychological Sciences, Brown University, Providence, RI, United States

[5]Department of Neuroscience, Brown University, Providence, RI, United States

[6]Center for Computation and Visualization, Brown University, Providence, RI, United States

[7]Industrial Design, Rhode Island School of Design, Providence, RI, United States

[8]Department of Psychiatry and Human Behavior, Brown University, Providence, RI, United States

[9]Center for Digital Health, Brown University, Lifespan, Providence, RI, United States

[*]these authors contributed equally

Corresponding Author:
Chloe Zimmerman Gunsilius, BA
Robert J. and Nancy D. Carney Institute for Brain Science
Brown University
Carney Institute for Brain Science, 4th floor
164 Angell Street
Providence, RI, 02912
United States
Phone: 1 401 863 6272
Email: chloe_zimmerman@brown.edu

Abstract

Chronic pain is one of the most significant health issues in the United States, affecting more than 20% of the population. Despite its contribution to the increasing health crisis, reliable predictors of disease development, progression, or treatment outcomes are lacking. Self-report remains the most effective way to assess pain, but measures are often acquired in sparse settings over short time windows, limiting their predictive ability. In this paper, we present a new mobile health platform called SOMA*Science*. SOMA*Science* serves as an easy-to-use research tool for scientists and clinicians, enabling the collection of large-scale pain datasets in single- and multicenter studies by facilitating the acquisition, transfer, and analysis of longitudinal, multidimensional, self-report pain data. Data acquisition for SOMA*Science* is done through a user-friendly smartphone app, SOMA, that uses experience sampling methodology to capture momentary and daily assessments of pain intensity, unpleasantness, interference, location, mood, activities, and predictions about the next day that provide personal insights into daily pain dynamics. The visualization of data and its trends over time is meant to empower individual users' self-management of their pain. This paper outlines the scientific, clinical, technological, and user considerations involved in the development of SOMA*Science* and how it can be used in clinical studies or for pain self-management purposes. Our goal is for SOMA*Science* to provide a much-needed platform for individual users to gain insight into the multidimensional features of their pain while lowering the barrier for researchers and clinicians to obtain the type of pain data that will ultimately lead to improved prevention, diagnosis, and treatment of chronic pain.

(JMIR Mhealth Uhealth 2024;12:e47177) doi:10.2196/47177

KEYWORDS

acute pain; acute-chronic pain transition; chronic pain; clinical outcome measurement; digital health; ecological momentary assessment; EMA; ESM; experience sampling methodology; mHealth; mobile health; pain management; pain self-management; patient reported outcomes; smartphone app

Introduction

More of us are in chronic pain than you might think—20% of adults in the United States reported pain on most or several days in 2019 [1]. This presents a substantial burden on society, costing up to US $635 billion annually [2]. In 2016, chronic back and neck pain alone accounted for the highest amount of US health care spending across 154 conditions, including diabetes and heart disease [3]. Moreover, pain is the leading cause of health care use across all illnesses [4-6]. Against this backdrop, it has never been more important to develop accurate pain symptom assessment and prediction methods to help patients, caregivers, and other stakeholders make informed decisions about treatment and care.

Accurately measuring pain is crucial for predicting an individual's pain trajectory [7]. Methods to identify objective biomarkers of pain intensity [8-10] are still in their infancy and have yet to be proven effective in predicting future self-reported pain [11]. To date, the most common way to assess if someone is in pain is to simply ask them [12]. Typically, this is done using an 11-point pain intensity scale, where individuals are asked to rate their pain from 0 (no pain) to 10 (worst pain imaginable) [12-14]. Since the 1980s, this simple pain intensity scale has played a significant role in the clinical assessment of pain by enabling defined targets for pain management and the dosing of pain-relieving medications [15,16]. If a person consistently reports pain intensity as more than 3 out of 10 for more than 3 months, the scale becomes part of the diagnostic criteria for chronic pain [17]. For patients, self-reporting their pain on the scale can validate and quantify their pain experience, leading to improved shared decision-making and enhanced communication with health care providers [18].

Despite its ease of use, the pain intensity scale has not led to significant advancements in pain management or patient satisfaction [18-23]. When measured in medical settings, people tend to over- or underreport their pain intensity depending on difficulties with recall, expected treatments, care standards, or other subjective factors such as mood [24,25]. Additionally, regulatory approval for pharmaceutical companies to promote "titration to effect" practices, whereby physicians were encouraged to increase opioid doses to achieve continued reductions on the pain intensity scale, contributed to opioid overprescribing that fueled the opioid epidemic [23,25]. These limitations highlight the need for more nuanced self-report measures of pain.

In recent years, clinical approaches to pain have sought to better assess the multidimensional experience of pain from a biopsychosocial perspective [12,26]. Multidisciplinary and individualized assessment and treatment of pain with both pharmacologic and nonpharmacologic interventions is now considered the ideal way to treat both acute and chronic pain [27]. To overcome the limitations of the unidimensional pain intensity scale, expert panelists at the Food and Drug Administration (FDA) and National Institutes of Health (NIH) have established "core outcome sets" of multidimensional questionnaires for use in research studies [13,28,29]. Dimensions assessed include pain, unpleasantness, interference, and impacts on mood and activity [25,30]. However, completing multiple questionnaires is time-consuming, and therefore not ideal for daily longitudinal studies. As a result, a large barrier remains in the ability of researchers to collect comprehensive, multidisciplinary pain data sets [31]. It is therefore imperative in pain research to implement a reliable method for multidimensional pain measurements that accurately captures the most important dimensions of pain symptoms and treatments as they evolve over time in the context of people's daily lives.

A comprehensive approach to pain assessment necessitates the acquisition of both deep and wide pain data. Deep data involve in-depth evaluations of the multidimensional aspects of pain within individuals over extended periods, while wide data refer to data sets that encompass a large number of individuals across different demographic factors like age, geographic location, race, ethnicity, and socioeconomic status, as well as across different pain diagnoses (eg, arthritis, fibromyalgia, and postsurgical pain). The acquisition of deep data enables a deeper understanding of the mechanisms that trigger and sustain pain in individuals, while wide data provide the foundation for generalizing findings and developing biomarkers for pain persistence or recovery. Smartphone apps can provide large-scale platforms for data collection while also helping users track their daily symptom experience [32]. Such digital tools provide a promising solution for acquiring deep and wide data sets that enable new behavioral and scientific insights into the dynamics and evolution of pain.

Here, we introduce a novel mobile health (mHealth) platform for longitudinal pain assessment, called SOMA*Science*. This platform has been conceptualized to comprehensively capture multiple facets of pain through its smartphone app, SOMA. Unlike conventional pain assessment tools that focus primarily on daily pain intensity, SOMA offers an enriched multidimensional pain assessment. This includes factors like daily pain intensity, unpleasantness, interference, mood, expectations, and activities, based on experience sampling methodologies (ESMs) [33]. Our choice of measures is in accordance with the current Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials (IMMPACT) recommendations for pain assessment [28,34]. In addition, the design and validation of the SOMA app have been executed following the guidelines set forth by the American Psychiatric Association (APA) [35,36].

SOMA was designed with users in mind, offering a free and user-friendly tracking feature that facilitates real-time tracking of pain, medications, and treatment regimens. A "Trends" section distills the multidimensional data to visualize pain trends.

This feature aids users in self-managing their pain, recognizing patterns, and discerning between effective pain management and areas needing improvement. With these insights, users can communicate more efficiently with their health care providers, positioning SOMA as a valuable "companion" tool to be used alongside standard medical advice.

SOMA*Science* has been engineered to provide researchers with the ability to carry out independent single- and multicenter studies. Its pain assessments can be used as end points in research studies and within clinical practice. To support expansive research initiatives, the SOMA*Science* platform ensures seamless transfer of app-sourced data to researchers and affiliated institutions and offers open-source code to streamline data preprocessing.

Contemporary pain research standards underscore the significance of multidimensional pain assessments through established batteries of pain questionnaires [13,29]. However, we identified a scarcity of available tools that effectively capture such assessments in a format meaningful for researchers aiming to develop better measures of pain outcomes or symptom burden over time. While there are smartphone apps, like Manage My Pain and Pain Scale-Pain Tracker App, they primarily cater to users as self-management tools [37,38]. The foundational architecture of such apps and the nature of the data they procure do not typically align with the specifications researchers need for conducting or analyzing comprehensive, large-scale studies that meaningfully contribute to pain research. A further limitation is the proprietary nature of many of these apps, which can impede data quality, accessibility, and transparency.

To our knowledge, no other tools currently exist that are specifically designed to capture multidimensional and longitudinal pain metrics in compliance with the recommended standards and are simultaneously tailored for extensive open-source academic research. This distinctiveness sets SOMA*Science* apart as an unparalleled mHealth platform.

In this paper, we discuss the choices and technological considerations for the development of SOMA*Science* as well as the scientific rationale behind the selection of measures. Our aim is to outline how SOMA*Science* can be used by scientists and institutions to acquire large-scale, longitudinal, multidimensional pain data in single and multicenter studies in order to gain new insights into pain that will benefit patients.

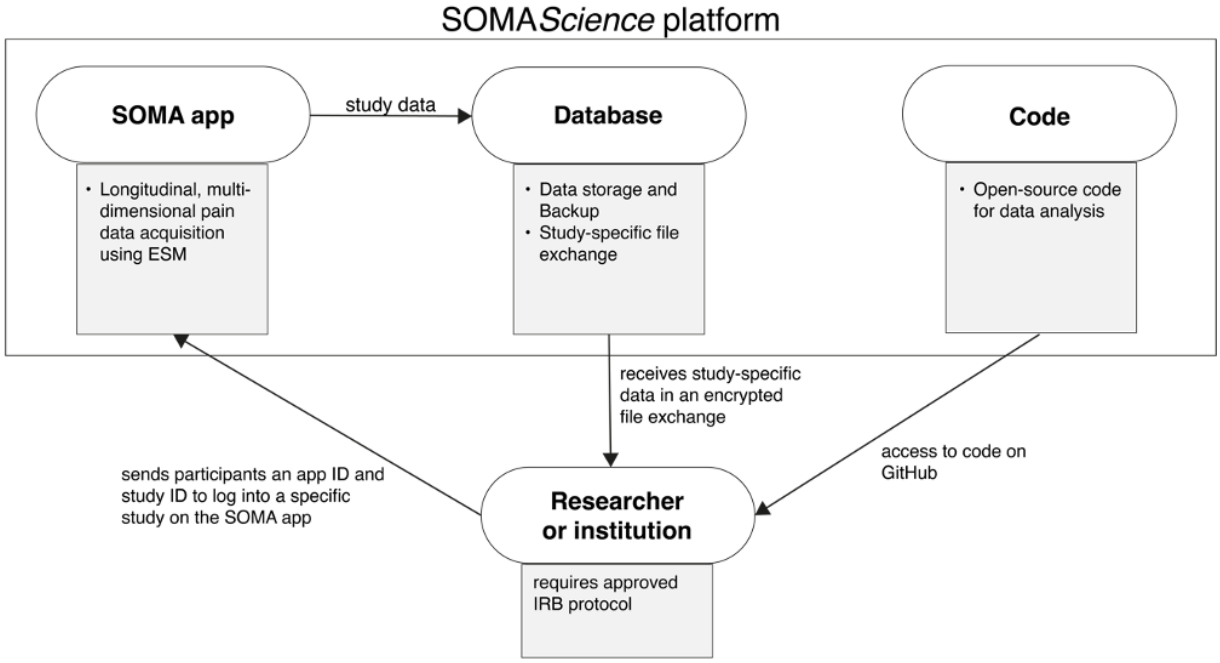
Section 1: SOMA*Science* Platform

Overview

SOMA*Science* aims to address the current shortage of platforms for acquiring both deep (multidimensional and longitudinal) and wide (cross-spectrum and large-scale) pain data to create novel insights into the dynamics and evolution of acute and chronic pain. The term SOMA stems from the Greek word σόμα (meaning body or entire person), signifying that it takes a holistic approach to pain.

SOMA*Science* represents the combination of 3 branches: the SOMA smartphone app used for data acquisition, an application programming interface for data transfer, and open-source analysis code distributed through GitHub [39] (Figure 1).

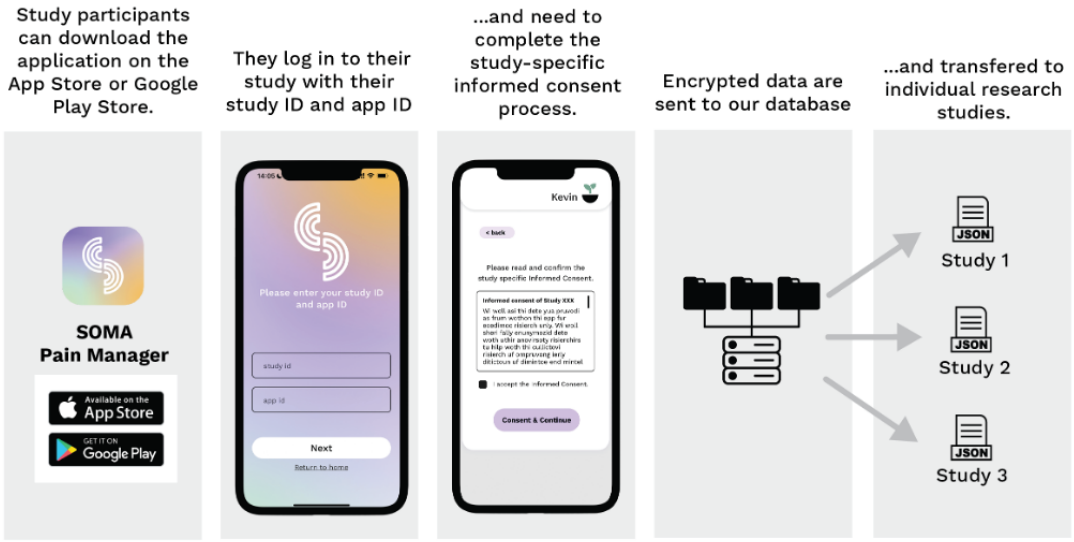
Figure 1. Schematic overview of the SOMA*Science* platform. Data will be acquired using the SOMA app, a user-friendly smartphone app available on Google Play and the Apple App Store. Encrypted data are sent to our application programming interface (currently located at Brown University) and shared in a study-specific manner with individual researchers and institutions. To facilitate data analysis, we created a GitHub repository where researchers can download, modify, or even create new versions of our template scripts for data preprocessing and certain analysis techniques through GitHub. ESM: experience sampling methodology; IRB: institutional review board.



To request to run a study through the platform, researchers need to submit a research inquiry detailing the study purpose on the SOMA website [40]. Data for SOMA*Science* are acquired through the SOMA app, which is freely available on Google

Play [41] and the Apple App Store [42] and can be found by searching for "SOMA Pain Manager." Anyone is able to download and use the app, regardless of whether they are participating in a research study. For associating people's app data with a specific study, the researcher will be assigned a unique study ID (1 per study) and a list of individual app IDs (1 per expected participant), which need to be sent to the study participants. Participants can then install and register on the SOMA app and enroll in a specific study using the study ID and unique app ID (instruction videos on [43]; Figure 2). This use of study and app IDs allows the assignment of individual participants' data to single and multicenter studies.

Figure 2. Schematic overview of study enrollment on the SOMA app. Study participants can download and register on the SOMA app directly using the links to Google Play or the Apple App Store or by searching for "SOMA Pain Manager." Inside the main menu of the app, they can sign into a specific research study using a pre-sent study ID and App ID. They are shown a copy of their institutional review board–approved study-specific informed consent form for their records. Encrypted research data on SOMA will be sent to our database and transferred to the researchers of each individual study.



Upon enrollment, participants will be sent regular reminders through the app to fill out short pain surveys (details about the ESM and data content are in the following sections). At the time of publication, assessments are restricted to the features listed below. Future releases may offer the option to request additional features and questions. After each assessment, encrypted data are transferred to our application programming interface, stored on an actively managed secure database (see "Maximizing Privacy and Security" in Multimedia Appendix 1), and then shared with researchers from individual studies (Figures 1 and 2).

Implementation of APA Guidelines in SOMA

The APA's app evaluation model stands as a notable benchmark for evaluating the suitability of health-related smartphone apps intended for patient populations [35]. It delineates 5 pivotal criteria to assess apps: accessibility, privacy and security, clinical foundation, engagement style, and therapeutic goal [44]. To keep pace with the rapidly advancing field of health apps, the system is regularly updated and refined [45]. Stemming from this APA model is a comprehensive database [46], which facilitates app evaluations across the 5 core domains [47]. Such initiatives are crucial in establishing public-facing, user-friendly standards for health apps and ensuring the development of safe and efficacious apps that benefit users [48-50].

In the development of SOMA*Science*, we have deeply integrated the cardinal principles of the APA model. Recognizing the emphasis that this model (and other akin evaluation frameworks) places on robust privacy, security, usability, and clinical foundations, we meticulously factored in specific technological elements during the app's creation [50]. For readers interested in the technological nuances and our dedicated approaches to privacy and security aligned with the APA guidelines, we direct you to Multimedia Appendix 1 [51-60].

Development of SOMA*Science*

A multidisciplinary team at the Psychiatry, Embodiment, and Computation Lab at Brown University, comprising academic neuroscientists, psychologists, pain physicians, software engineers, and designers, collaboratively conceptualized and initialized the early design and features of the SOMA app. The SOMA*Science* platform is supported by the Brainstorm program at the Carney Institute for Brain Science at Brown University, a new program to accelerate the translation of computational brain science into real-world applications that benefit patients, the scientific community, and society as a whole.

Incorporation of Patient and User Feedback

Any app meant for long-term use must provide a simple, user-friendly interface tailored to its target audience. With this in mind, we actively sought initial feedback from individuals experiencing chronic pain to shape our app's delivery structure. Through comprehensive one-on-one Zoom (Zoom Video Communications, Inc) discussions with a select patient group (n=4), representing a diverse pain spectrum including migraines, postsurgical musculoskeletal pain, cancer pain, and inflammatory bowel disease, we gleaned insights into our early app prototypes. These discussions involved a detailed

walkthrough of each interface, where patients aired their thoughts in a guided interview.

Drawing from this feedback, we refined the "Check-In" interface's design and flow. A common sentiment among initial users was the importance of daily tracking of pain and mood fluctuations in tandem with daily activities. To accommodate this, we introduced the "Trends" screen, a visual tool designed not only to foster self-management and a deeper understanding of pain dynamics but also to facilitate effective communication with health care providers. To further enhance the app's utility, we incorporated screens to monitor various treatment modalities and transitioned the mood and emotion tracking interfaces to use intuitive visual analog scales (VASs) in lieu of a 2D rating system.

Responding to the patient's desire for a more personalized experience, we introduced an interactive chatbot during the onboarding phase. This chatbot briefly engages users, gathering foundational demographic details and an introductory snapshot of their pain experiences. Existing studies vouch for the efficacy of chatbots in extending support to people with chronic pain [61], making this an evolving component with forthcoming features in SOMA that focus on pain interventions.

Our iterative refinement strategy incorporated a beta-testing phase. Initially, 30 internal testers actively engaged with the app daily across a gamut of devices and operating platforms, enabling us to identify and correct technical glitches and enhance the user experience, especially regarding the "Trends" data visualization. To expand our feedback, SOMA was then shared with a larger patient interest group (over 250 people with chronic pain), leading to critical refinements and the inclusion of user-suggested enhancements. This ongoing feedback mechanism ensures the continuous improvement and evolution of SOMA.

Section 2: Pain Data Acquisition Through SOMA*Science*

Deep Data Acquisition Using ESMs

Overview

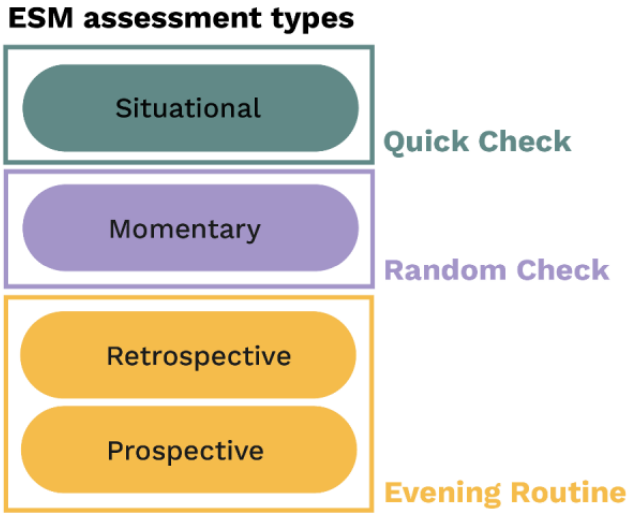
The SOMA app uses ESMs to gather multidimensional and longitudinal pain data for SOMA*Science*. ESMs, also known as ecological momentary assessments, provide real-time, self-report data about individuals' thoughts, feelings, and experiences (eg, "How do you feel right now?") in the context of people's daily lives [33]. Previous ESM studies on pain have shown high completion rates (>85%) and demonstrated the feasibility of using these daily self-reports for pain [62], in line with findings about the high completion rates of mobile-delivered ESM studies in general [63].

ESMs have several benefits over traditional self-report measures. First, they offer real-time data that are less prone to recall bias, allowing for the capture of critical experiences that might be missed by retrospective long-term self-report measures [64-67]. Second, ESMs can capture contextual information about an individual's thoughts and experiences, such as knowing what activities a person engaged in when they experienced pain [68]. Thus, they provide insight into the longitudinal dynamics of multidimensional aspects of pain in people's natural ecological environments across time and context [69,70]. This enables the generation of rich data sets that could be used to identify candidate behavioral "biomarkers" or "assays" that predict transitions in disease states based on self-report alone [68,71]. For example, there is preliminary evidence that longitudinal measures of pain can predict acute pain state transitions after surgery [72] and identify treatment response time courses in patients with chronic pain [73].

*Momentary, Situational, Retrospective, and Prospective Assessments in SOMA*Science*

One limitation of existing ESM studies is that they typically solicit several short, momentary reports throughout the day [74]. While this approach reduces bias in pain reports resulting from memory recall or pain beliefs, it may still miss important short-term pain dynamics, such as flare-ups, and fail to assess the role of expectations in the development and treatment of pain [75]. To address this limitation, SOMA*Science* uses a multifaceted approach, which includes 4 daily assessment types on the SOMA app: momentary assessments (called random check-ins), voluntary self-initiated entries (called quick check-ins), and both retrospective and prospective assessments or coverage assessments [75] (which are both part of an evening routine at the end of the day; Figure 3).

Figure 3. ESM assessment types in SOMA*Science*. Users complete 3 main types of daily check-ins per day (Quick Check, Random Check, and Evening Routine) that assess different domains of their daily experiences. ESM: experience sampling methodology.



Random and quick check-ins capture various aspects of mood, activities, pain, and pain location and can be completed in less than 30 seconds. Quick check-ins can be performed at any time, for example, during or shortly after a flare-up. Random check-ins reflect classical ESM assessments and only occur during randomly selected moments within a specific time window (eg, 3 checks per day between 8 AM and 6 PM). Users receive notifications on their phone when the random check-in is available and have the option to snooze the notification for a predefined time window (eg, a maximum of 60 minutes).

The evening routine assesses recall of pain, mood, activities, and any pain-related treatments over the past day (retrospective), as well as predictions of pain, mood, and activities for the next day (prospective). This routine is available during a prolonged, preset, but fixed time window at night (eg, 6 PM-11 PM) to promote habit formation that increases the likelihood of long-term app use. It takes less than 3 minutes to complete.

Longitudinal Assessments with SOMAScience

Pain is inherently dynamic, fluctuating not just daily but hourly and even on a minute-to-minute scale, even without changes in physiological markers [18,47,62]. Traditional methods, which measure pain intensity sporadically during clinical trials or medical visits, might not capture a patient's holistic pain experience due to their limited assessment windows.

While some studies aim for more granular pain assessments, they often focus on brief periods. A systematic review found the median duration for ESM studies in pain to be 14 days and a mere 7 days for general mHealth ESM studies [40,61]. Such short durations can overlook pivotal phases, like the transition from acute to chronic pain over 3-6 months. Consequently, the role of self-reported pain dynamics in acute pain recovery or its evolution to chronic pain remains ambiguous.

The SOMA app is designed to bridge this gap. Its check-ins are concise, using straightforward design principles for ease of use. The chosen metrics cater to diverse pain experiences, facilitating collaboration and data sharing among researchers. Moreover, the app's "Trends" feature empowers users to track their pain, treatments, activities, and mood over diverse timeframes (ie, weekly, monthly, and annually). This aids in providing users with a deeper understanding of their pain journey, ultimately supporting more effective self-management.

Wide Data Acquisition Using Smartphones

While large data sets on repeated multidimensional pain ratings beyond intensity alone are still few and far between, smartphones offer a unique opportunity to expand data acquisition beyond classical experimental settings [74]. Smartphone access has increased tremendously in the past decade (84% of US households reported owning at least 1 smartphone) [76]. Data acquired remotely through smartphone apps facilitate large-scale, real-world studies without the constraints of traditional laboratory studies. The results of such pragmatic studies are more generalizable than highly selective traditional randomized controlled trials [77-79]. SOMA*Science* was built to allow remote monitoring of longitudinal symptoms and treatments to maximize high-quality data in large-scale pragmatic studies. To further facilitate a much wider array of user input, the SOMA app is compatible with both Android and iOS devices, meaning anyone in the United States with a smartphone can use it.

Smartphone-based pain assessments offer a solution to the limited geographic, economic, and cultural diversity in traditional pain studies. Smartphones are pervasive, even in low-resource [80] or rural areas [72,81], where almost half of the world's smartphone owners live [82]. They are also widely used by older adults [83], who are often left out of laboratory-based pain studies. Additionally, there is a need to consider how pain and its treatment vary across racial, ethnic, and cultural backgrounds for comprehensive care. Even within a specific culture, there are important differences in how pain is experienced and treated across different racial and ethnic backgrounds that need to be accounted for to deliver the best pain care [84-86]. Upcoming translations of the app into languages like Spanish and German, in collaboration with experts familiar with the culture, will further diversify data and

insights. Translation into other languages will follow, and collaboration to translate the app is welcomed.

Section 3: Data Content

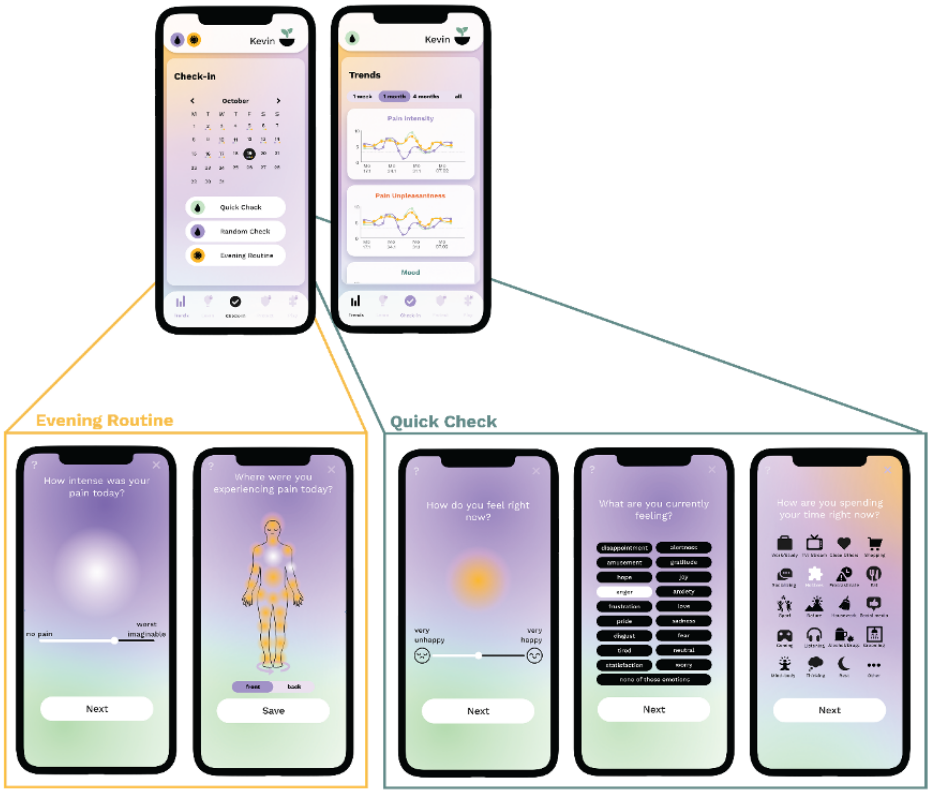
Overview

With the rising number of health-focused smartphone apps, there is also a growing need for transparency in the selection of measures for the app. Here we briefly detail the process of selection for each measure in SOMA*Science* as assessed through the SOMA app, its basis in the scientific and clinical literature, and what gaps it was chosen to address. The goal is to provide transparency in the design and selection process to facilitate the development of research studies using SOMA*Science*. Specific measures may be refined over time with user feedback and as scientific studies using SOMA*Science* identify areas for improvement.

Measuring Pain Intensity, Unpleasantness, and Interference

Multidimensional pain assessment is a crucial aspect of clinical practice as it helps determine the effectiveness of treatment and recovery. The SOMA app assesses the pain intensity scale in addition to pain interference and pain unpleasantness to provide a more holistic measure of daily pain symptom load [87]. Participants are asked to rate each pain question on a sliding VAS, providing a simple and comparable assessment of daily pain experience (Figure 4). This operationalizes a practice advocated in the International Statistical Classification of Diseases and Related Health Problems for providers to use a composite rating of pain intensity, distress, and interference of pain to determine overall burden [88]. The composite pain score provides a multidimensional solution that balances the limitations of the unidimensional pain scale while still being easily quantifiable. The ability to analyze the 3 measures individually and as a composite score helps identify meaningful individual variability, enabling specific predictions between individuals. For example, pain unpleasantness could be more predictive of future pain in individual A, while pain interference may be more important for individual B's long-term outcomes.

Figure 4. Upper panel: example screens on the SOMA app for check-ins and trends. Lower panel: example screens on the SOMA app from Quick Check, Random Check, and Evening Routine. From left to right: assessing pain on a visual analog scale (VAS), indicating pain location, assessing mood on a VAS, emotion selection, and activity selection.



Importantly, SOMA's 3 pain questions were chosen because they are directly comparable to results from established pain questionnaires, such as the Brief Pain Inventory [89] or the McGill Pain Inventory [90]. They also satisfy the standards set by major scientific and regulatory bodies, such as the IMMPACT recommendations, the NIH Helping to End Addiction Long-term initiative, and the FDA guidelines for assessing multidimensional components of pain [13,28,29,87]. In this way, SOMA's multidimensional pain assessment of intensity, unpleasantness, and interference can provide important supplemental measures that are directly comparable to established clinical benchmarks and standards of care. This is

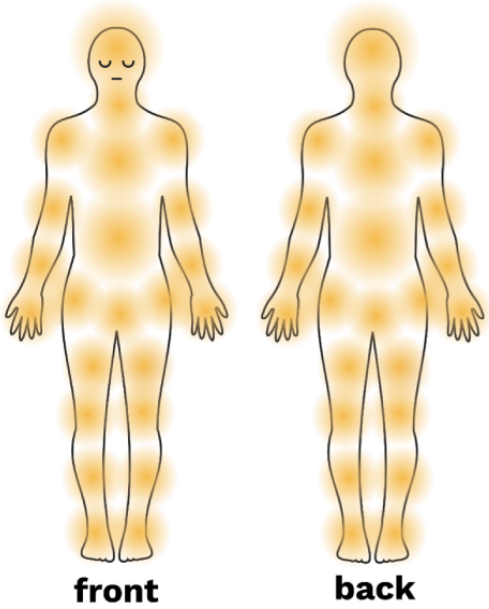
critical for researchers looking to establish and validate novel pain biomarkers or end points.

Measuring Pain Locations

Pain localization is an important aspect of pain assessment. Conventional methods of measuring pain location in medical appointments and research studies involve having individuals indicate it on a body map, such as the Brief Pain Inventory [89], the McGill Pain Questionnaire [91], or the Michigan Body Map [92]. This approach can pinpoint differences in peripheral and central pain pathology based on the localization and stability of pain representation over time. For instance, nociceptive or inflammatory pain is usually precisely localized somatically and does not change much over time, while neuropathic or chronic primary pain is often experienced in multiple bodily locations, radiates, or changes over time [93].

More recent methods of digital quantification, like the ones used on the SOMA app, have established the reliability and validity of body maps for pain assessments [94,95]. Interactive body maps delivered through digital or tablet apps are more effective than traditional paper or laptop assessments [92,96]. Yet a review of smartphone apps that use the body map for tracking pain found that few actually quantified the location ratings or provided any summary feedback [97]. The SOMA app's interactive body map offers 46 different discrete location options on the front and back of the body that participants indicate in every daily check-in (Figure 5 and Multimedia Appendix 1). The use of discrete points ensures uniformity across devices and accounts for differences in participants' finger size or dexterity. The "Trends" section of the SOMA app displays the body map with the percentage of times a location has been selected, enabling users to visualize the frequency of pain at a given location. For participants who experience nonspecific, difficult-to-localize, or widespread pain, such as fibromyalgia, there is an additional option to indicate "My pain is everywhere" on the body map.

Figure 5. Pain map included on the SOMA app covering 46 discrete pain locations.



Measuring Interventions

The treatment of pain has been incredibly difficult to get right. The newest clinical guidelines advocate the use of multimodal, multidisciplinary approaches [27,98,99]. Such approaches emphasize a combination of pain treatments that include medications, restorative therapies (eg, physical therapy), interventional procedures (eg, epidural injections), behavioral interventions (eg, cognitive behavioral interventions), and complementary and integrative medicine (eg, acupuncture). Combinations of these therapies have been associated with the best long-term pain outcomes [100,101] and satisfy a biopsychosocial approach to pain [102].

It can be challenging for individuals and providers to determine which treatments are most effective for them, as the effects of many treatments for pain may not become apparent for weeks or even months (eg, cognitive and physical interventions, certain medications, and surgery) [103,104]. In determining how to measure treatments through the SOMA app, we followed the recommendations of the 2019 Department of Health and Human Services Pain Strategy [27].

In the case of medications, many different pathophysiologic mechanisms are targeted with different classes of medications. The use of different medications often changes over time, so we designed the medication screen on the SOMA app to be able to capture such changes. We, therefore, included 20 options across the main classes of pain medications for both acute and chronic pain on the SOMA app, detailed further in Multimedia Appendix 1.

A second treatment screen includes the recommended nonpharmacologic approaches to pain. Combinations of these treatments are often used by a single person over time to target different pain mechanisms [105]. SOMA*Science* currently provides the ability for people to track up to 20 different nonpharmacologic therapies across these 4 major classes, detailed in Multimedia Appendix 1. SOMA*Science*'s broad treatment tracking capabilities therefore facilitate the type of

wide data needed to understand differences in treatment use across users.

Measuring Emotions

Emotion and pain are fundamentally related. Definitions of pain acknowledge that pain is partially an emotional experience [106,107], yet few researchers would reduce pain down to just another emotional state such as sadness or happiness. The complexity of the pain-emotion relationship is highlighted by reviews of the neural circuits of each construct, showing both shared and functionally dissociable brain regions [108,109]. Unfortunately, the theoretical and empirical understanding of how emotion and pain are connected is limited, as much previous research only focuses on cross-sectional correlations between the 2 constructs [110,111].

Despite the renaissance of emotion research since the 1960s [112], emotion researchers continue to disagree on what constitutes an appropriate emotion measure [113]. The discrete emotion perspectives suggest that specific emotions such as anger, fear, happiness, sadness, disgust, and surprise are special kinds of biologically distinct responses associated with unique behavioral, physiological, and experiential correlates [114,115]. The dimensional perspectives consider emotions to be organized along a set of common dimensions such as valence (unpleasantness) and arousal (intensity) that are combined with cognitive and cultural knowledge to form emotional states.

When designing an emotion list for the SOMA app, we wanted a set of emotions that are commonly experienced, had a balance of positive and negative emotions, and were relevant to pain experiences (eg, worry). We started by selecting emotions experienced more than 5% of the time from a large-scale ecological momentary assessment study [116], removing rare emotions such as contempt or awe. Next, we balanced the remaining emotion set by using positive and negative emotions from the clinically relevant scale (Positive and Negative Affect Schedule) [117]. Lastly, we allowed participants to indicate no emotion (neutral) or emotion not listed (other).

A complementary method from a dimensional perspective is to assess general mood through a single-dimensional VAS of valence ranging from very unpleasant to very pleasant. Although there are other dimensional scales that could also be assessed, such as arousal [118,119] or goal congruence [120], valence is known to capture the majority of variance between emotion states [121], as self-reports of emotions tend to be highly correlated within a positive or negative valence [122]. Accordingly, on the SOMA app, we ask users to evaluate their current, past, or future mood on this VAS, ranging from unpleasant (0) to very pleasant (100). One major benefit of this measure is that we can quantify daily emotional experiences even if the participant does not self-report any discrete emotions from our finite list, reducing the need to interpolate or remove missing data from our analysis of the dynamics of emotion and pain (Figure 4).

Measuring Activities

People's experiences of both pain and emotion are intrinsically tied to the activities they are engaging in each day. There have been many large-scale studies investigating the relationship between emotions and behavior, showing that physical exercise is meaningfully associated with reduced mental health difficulties [123] and that people's choice of activities is motivated by minimizing negative affect and maximizing positive affect [124,125]. In particular, people seem to engage in mood-elevating activities (eg, socializing) when they are feeling down and mood-depressing activities (eg, work and chores) when they are feeling up. At the same time, being in pain affects both mood and daily activities.

Most people intuitively reduce their activities when they experience new-onset pain. Acute pain generally functions as an alarm bell in the brain to signal tissue damage, with the urge to rest considered a protective mechanism to prevent further injury and promote healing [126]. However, outside of the initial acute phase, a lack of activity can hinder long-term recovery and may signal underlying changes in affective and motivational brain circuits that have been causally linked to the transition from subacute to chronic low back pain [126,127]. The synergistic impact of activity engagement on mood, pain, and physical function is why activity engagement is promoted by pain self-management and rehabilitation programs alike [128]. While many people assume that patients with chronic pain move less than those not in pain, objective actigraphy data do not differ between patients with chronic pain and no-pain controls [129,130]. What has been less studied is the types and range of daily activities and how they change between acute and chronic pain stages. While people may recognize changes in their daily activities as a result of pain (eg, no longer walking as much), they may not have insight into the relationship between pain, mood, emotion, and activities (ie, which activities increase or decrease pain in the short vs long term). It remains unclear whether certain patterns of activity engagement at different stages of pain experience are important for long-term outcomes.

For this reason, we added an activities-tracking feature alongside mood and pain tracking. The activities screen on the SOMA app contains 20 activities that are known to have a dynamic relationship with mood based on large-scale, longitudinal data sets [125] or to be common among patients with pain (eg, medical visits). Specific activities selected are detailed in Multimedia Appendix 1. In addition to having the person report either momentary or daily activities, we also have participants reflect on how each activity they completed made them feel that day and how much it affected their pain. Taken together, these measurements provide a full picture of participants' daily activities and help understand the potential bidirectional relationships between emotion, pain, and behavior. Users are also able to visualize their ratings of how much a given activity affected their pain and mood over time in the dedicated trends screen, which may help people develop insight into how certain activities help or hinder their recovery.

Measuring Predictions

Expectations play an undeniably large role in pain perception. In artificial laboratory settings where healthy participants receive painful stimulation, a wealth of findings show that expectations about pain can increase the pain experience (nocebo effect) or conversely decrease it (placebo effect) [131,132]. This has important implications for the experience of pathological pain

[133,134], where expectations about pain treatment are reliable predictors of treatment response [135]. These studies suggest that an ongoing cognitive modulation of pain is an important determinant of ongoing pain perception [136]. Expectations for pain relief most likely shift over time, the longer someone has experienced pain the more difficult the pain is to control [137,138]. Subtle longitudinal changes in expectations are believed to occur as pain becomes chronic, but capturing these changes in research studies is challenging. For this reason, we included an assessment of daily predictions about pain, mood, and activities on the SOMA app to capture how predictions vary over time within the same person or between different types of users (eg, patients with acute vs chronic pain).

We decided to assess expectations on the SOMA app by asking users to predict their expected levels of pain intensity, unpleasantness, interference, mood, and activities for the next day using the same scales used to capture their actual rating for that day. This allows the assessment of the bilateral influence of pain and mood expectations on actual experiences of pain, mood, and activities that are entered the following day.

Discussion

There is a great need for easy-to-use tools that help those in pain, their medical providers, and the larger health care system identify risk factors and predict the onset of chronic pain. Pain management is a rapidly evolving field that increasingly relies on assessments and treatments that are multimodal and multidisciplinary. Traditional, unidimensional assessments of self-reported pain fail to capture the nuances of pain experience and multimodal pain management. Therefore, there is an urgent need for research tools that have been specifically designed to capture this complexity.

To address this gap, we built the SOMA*Science* platform. Briefly, the platform uses the smartphone app SOMA to collect longitudinal, multidimensional, ESM-based pain data that capture daily pain intensity, unpleasantness, inference, mood, activities, and predictions. The SOMA app provides a simple and pleasing user interface that can promote pain self-management through visualization of pain trends over time, helping encourage individual insight into factors that exacerbate or alleviate pain. The visualizations provided can also be used to improve communication of multidimensional pain burdens to health care providers. At the same time, the larger SOMA*Science* platform enables user data to be included in registered single and multicenter studies.

In this paper, we have detailed the clinical and technological considerations taken into account in developing SOMA*Science* and the scientific rationale behind its measurements. We believe this platform is capable of meeting the requirement for tools to acquire deep and wide-ranging pain data over time, which has been largely absent from existing pain data sets. As such, SOMA*Science* can be used to answer a broad range of research questions, such as the correlation between initial pain dynamics and the eventual development of chronic pain (ie, predicting the transition from acute to chronic pain), evaluating both short-term and long-term effects of various treatments on pain experiences, or identifying distinct symptom clusters (ie, pain phenotypes). Moreover, the data available are sufficient to calculate more detailed multidimensional and longitudinal clinical trial or observational study end points.

The primary focus of this paper was to introduce SOMA*Science* as a platform for scientific studies. In the long term, we also plan to build platforms with a more clinical focus that facilitate integration with medical care (SOMA*Clinic*) and the support of treatments (SOMA*Therapeutics*). This will involve connecting the SOMA app to electronic medical records and including interfaces to health trackers (eg, daily actigraphy, heart rate, or sleep data from health kits or wearables). The intention is to have a significant positive impact, both in terms of advancing research on pain and improving the lives of people with pain.

Acknowledgments

This work was supported by the Brainstorm program at the Robert J & Nancy D Carney Institute for Brain Science (FHP), the COBRE Center for Nervous System Function (grant NIGMS 5P20GM103645; FHP), and conducted with the help of research staff at the Center for Computation and Visualization at Brown University and volunteers from the Rhode Island School of Design. CZG has been awarded an individual research grant that covered her effort on this paper from the National Center for Complementary and Integrative Health (grant F30AT012306). We want to thank everyone who has supported the development of SOMA*Science* and the SOMA app over the past years, including but not limited to Benjamin Andrew, Rashi Dhar, Louis Rakovich, and Isabel Restrepo.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The technological and design details regarding the building of the SOMA app, as well as specific measures selected for SOMA app screens, that may be of interest to some readers.
[DOCX File , 22 KB - mhealth_v12i1e47177_app1.docx]

References

1. Yong RJ, Mullins PM, Bhattacharyya N. Prevalence of chronic pain among adults in the United States. Pain 2022;163(2):e328-e332. [doi: 10.1097/j.pain.0000000000002291] [Medline: 33990113]
2. Gaskin DJ, Richard P. The economic costs of pain in the United States. J Pain 2012;13(8):715-724 [FREE Full text] [doi: 10.1016/j.jpain.2012.03.009] [Medline: 22607834]
3. Dieleman JL, Cao J, Chapin A, Chen C, Li Z, Liu A, et al. US health care spending by payer and health condition, 1996-2016. JAMA 2020;323(9):863-884 [FREE Full text] [doi: 10.1001/jama.2020.0734] [Medline: 32125402]
4. Mäntyselkä P, Kumpusalo E, Ahonen R, Kumpusalo A, Kauhanen J, Viinamäki H, et al. Pain as a reason to visit the doctor: a study in Finnish primary health care. Pain 2001;89(2-3):175-180. [doi: 10.1016/s0304-3959(00)00361-4] [Medline: 11166473]
5. Cordell WH, Keene KK, Giles BK, Jones JB, Jones JH, Brizendine EJ. The high prevalence of pain in emergency medical care. Am J Emerg Med 2002;20(3):165-169. [doi: 10.1053/ajem.2002.32643] [Medline: 11992334]
6. St Sauver JL, Warner DO, Yawn BP, Jacobson DJ, McGree ME, Pankratz JJ, et al. Why patients visit their doctors: assessing the most prevalent conditions in a defined American population. Mayo Clin Proc 2013;88(1):56-67 [FREE Full text] [doi: 10.1016/j.mayocp.2012.08.020] [Medline: 23274019]
7. Chapman CR, Donaldson GW, Davis JJ, Bradshaw DH. Improving individual measurement of postoperative pain: the pain trajectory. J Pain 2011;12(2):257-262 [FREE Full text] [doi: 10.1016/j.jpain.2010.08.005] [Medline: 21237721]
8. Davis KD, Flor H, Greely HT, Iannetti GD, Mackey S, Ploner M, et al. Brain imaging tests for chronic pain: medical, legal and ethical issues and recommendations. Nat Rev Neurol 2017;13(10):624-638 [FREE Full text] [doi: 10.1038/nrneurol.2017.122] [Medline: 28884750]
9. Davis KD, Aghaeepour N, Ahn AH, Angst MS, Borsook D, Brenton A, et al. Discovery and validation of biomarkers to aid the development of safe and effective pain therapeutics: challenges and opportunities. Nat Rev Neurol 2020;16(7):381-400 [FREE Full text] [doi: 10.1038/s41582-020-0362-2] [Medline: 32541893]
10. Tracey I. Neuroimaging enters the pain biomarker arena. Sci Transl Med 2021;13(619):eabj7358. [doi: 10.1126/scitranslmed.abj7358] [Medline: 34757804]
11. Mouraux A, Iannetti GD. The search for pain biomarkers in the human brain. Brain 2018;141(12):3290-3307 [FREE Full text] [doi: 10.1093/brain/awy281] [Medline: 30462175]
12. Fillingim RB, Loeser JD, Baron R, Edwards RR. Assessment of chronic pain: domains, methods, and mechanisms. J Pain 2016;17(Suppl 9):T10-T20 [FREE Full text] [doi: 10.1016/j.jpain.2015.08.010] [Medline: 27586827]
13. Cook KF, Dunn W, Griffith JW, Morrison MT, Tanquary J, Sabata D, et al. Pain assessment using the NIH toolbox. Neurology 2013;80(11 Suppl 3):S49-S53 [FREE Full text] [doi: 10.1212/WNL.0b013e3182872e80] [Medline: 23479545]
14. Basch E, Barbera L, Kerrigan CL, Velikova G. Implementation of patient-reported outcomes in routine medical care. Am Soc Clin Oncol Educ Book 2018;38:122-134 [FREE Full text] [doi: 10.1200/EDBK_200383] [Medline: 30231381]
15. Morone NE, Weiner DK. Pain as the fifth vital sign: exposing the vital need for pain education. Clin Ther 2013;35(11):1728-1732 [FREE Full text] [doi: 10.1016/j.clinthera.2013.10.001] [Medline: 24145043]
16. Haefeli M, Elfering A. Pain assessment. Eur Spine J 2006;15(Suppl 1):S17-S24 [FREE Full text] [doi: 10.1007/s00586-005-1044-x] [Medline: 16320034]
17. Treede RD, Rief W, Barke A, Aziz Q, Bennett MI, Benoliel R, et al. A classification of chronic pain for ICD-11. Pain 2015;156(6):1003-1007 [FREE Full text] [doi: 10.1097/j.pain.0000000000000160] [Medline: 25844555]
18. Holmes MM, Lewith G, Newell D, Field J, Bishop FL. The impact of patient-reported outcome measures in clinical practice for pain: a systematic review. Qual Life Res 2017;26(2):245-257 [FREE Full text] [doi: 10.1007/s11136-016-1449-5] [Medline: 27815820]
19. Phillips S, Gift M, Gelot S, Duong M, Tapp H. Assessing the relationship between the level of pain control and patient satisfaction. J Pain Res 2013;6:683-689 [FREE Full text] [doi: 10.2147/JPR.S42262] [Medline: 24049457]
20. Rothman ML, Beltran P, Cappelleri JC, Lipscomb J, Teschendorf B, Mayo/FDA Patient-Reported Outcomes Consensus Meeting Group. Patient-reported outcomes: conceptual issues. Value Health 2007;10(Suppl 2):S66-S75 [FREE Full text] [doi: 10.1111/j.1524-4733.2007.00269.x] [Medline: 17995476]
21. Robinson ME, Staud R, Price DD. Pain measurement and brain activity: will neuroimages replace pain ratings? J Pain 2013;14(4):323-327 [FREE Full text] [doi: 10.1016/j.jpain.2012.05.007] [Medline: 23548484]
22. Mularski RA, White-Chu F, Overbay D, Miller L, Asch SM, Ganzini L. Measuring pain as the 5th vital sign does not improve quality of pain management. J Gen Intern Med 2006;21(6):607-612 [FREE Full text] [doi: 10.1111/j.1525-1497.2006.00415.x] [Medline: 16808744]
23. Scher C, Meador L, Van Cleave JH, Reid MC. Moving beyond pain as the fifth vital sign and patient satisfaction scores to improve pain care in the 21st century. Pain Manag Nurs 2018;19(2):125-129 [FREE Full text] [doi: 10.1016/j.pmn.2017.10.010] [Medline: 29249620]
24. Broderick JE, Stone AA, Calvanese P, Schwartz JE, Turk DC. Recalled pain ratings: a complex and poorly defined task. J Pain 2006;7(2):142-149 [FREE Full text] [doi: 10.1016/j.jpain.2005.09.012] [Medline: 16459280]
25. Sullivan MD, Ballantyne JC. Must we reduce pain intensity to treat chronic pain? Pain 2016;157(1):65-69. [doi: 10.1097/j.pain.0000000000000336] [Medline: 26307855]

26. Nicholas MK. The biopsychosocial model of pain 40 years on: time for a reappraisal? Pain 2022;163(Suppl 1):S3-S14. [doi: 10.1097/j.pain.0000000000002654] [Medline: 36252231]
27. Pain management best practices Inter-Agency Task Force report: updates, gaps, inconsistencies, and recommendations. US Department of Health and Human Services. 2019. URL: https://www.hhs.gov/sites/default/files/pain-mgmt-best-practices-draft-final-report-05062019.pdf [accessed 2023-12-08]
28. Gewandter JS, Dworkin RH, Turk DC, Devine EG, Hewitt D, Jensen MP, et al. Improving study conduct and data quality in clinical trials of chronic pain treatments: IMMPACT recommendations. J Pain 2020;21(9-10):931-942 [FREE Full text] [doi: 10.1016/j.jpain.2019.12.003] [Medline: 31843583]
29. Zigler C, O'Brien E, Weinfurt KP. NIH HEAL, FDA, and other core outcome sets. Rethinking Clinical Trials. URL: https://rethinkingclinicaltrials.org/chapters/conduct/real-world-evidence-patient-reported-outcomes-pros/nih-heal-fda-and-other-core-outcome-sets/ [accessed 2022-11-14]
30. Lazaridou A, Edwards RR. Getting personal: the role of individual patient preferences and characteristics in shaping pain treatment outcomes. Pain 2016;157(1):1-2. [doi: 10.1097/j.pain.0000000000000376] [Medline: 26447708]
31. Cohen SP, Vase L, Hooten WM. Chronic pain: an update on burden, best practices, and new advances. Lancet 2021;397(10289):2082-2097. [doi: 10.1016/S0140-6736(21)00393-7] [Medline: 34062143]
32. Moreno-Ligero M, Moral-Munoz JA, Salazar A, Failde I. mHealth intervention for improving pain, quality of life, and functional disability in patients with chronic pain: systematic review. JMIR Mhealth Uhealth 2023;11:e40844 [FREE Full text] [doi: 10.2196/40844] [Medline: 36729570]
33. Larson R, Csikszentmihalyi M. The experience sampling method. In: Csikszentmihalyi M, editor. Flow and the Foundations of Positive Psychology: The Collected Works of Mihaly Csikszentmihalyi. Dordrecht: Springer Netherlands; 2014:21-34.
34. Edwards RR, Schreiber KL, Dworkin RH, Turk DC, Baron R, Freeman R, et al. Optimizing and accelerating the development of precision pain treatments for chronic pain: IMMPACT review and recommendations. J Pain 2023;24(2):204-225 [FREE Full text] [doi: 10.1016/j.jpain.2022.08.010] [Medline: 36198371]
35. Torous JB, Chan SR, Gipson SYMT, Kim JW, Nguyen TQ, Luo J, et al. A hierarchical framework for evaluation and informed decision making regarding smartphone apps for clinical care. Psychiatr Serv 2018;69(5):498-500 [FREE Full text] [doi: 10.1176/appi.ps.201700423] [Medline: 29446337]
36. Torous J, Andersson G, Bertagnoli A, Christensen H, Cuijpers P, Firth J, et al. Towards a consensus around standards for smartphone apps and digital mental health. World Psychiatry 2019;18(1):97-98 [FREE Full text] [doi: 10.1002/wps.20592] [Medline: 30600619]
37. Bhatia A, Kara J, Janmohamed T, Prabhu A, Lebovic G, Katz J, et al. User engagement and clinical impact of the manage my pain app in patients with chronic pain: a real-world, multi-site trial. JMIR Mhealth Uhealth 2021;9(3):e26528 [FREE Full text] [doi: 10.2196/26528] [Medline: 33661130]
38. Devan H, Farmery D, Peebles L, Grainger R. Evaluation of self-management support functions in apps for people with persistent pain: systematic review. JMIR Mhealth Uhealth 2019;7(2):e13080 [FREE Full text] [doi: 10.2196/13080] [Medline: 30747715]
39. SOMA science: analysis repository for SOMA science. GitHub. URL: https://github.com/SOMAanalysis/ [accessed 2023-12-08]
40. Research inquiry. SOMA. URL: https://somatheapp.com/researchinquiry/ [accessed 2023-12-08]
41. SOMA pain manager. Google Play Store. URL: https://play.google.com/store/apps/details?id=edu.brown.ccv.soma&pli=1 [accessed 2023-12-08]
42. Soma pain manager. Apple Store. URL: https://apps.apple.com/us/app/soma-pain-manager/id6444110898 [accessed 2023-12-08]
43. SOMA study registration tutorial. YouTube. URL: https://www.youtube.com/watch?v=z1hpomiyXFM [accessed 2023-12-08]
44. The app evaluation model. American Psychiatric Association. URL: https://www.psychiatry.org/psychiatrists/practice/mental-health-apps/the-app-evaluation-model [accessed 2023-12-08]
45. Lagan S, Emerson MR, King D, Matwin S, Chan SR, Proctor S, et al. Mental health app evaluation: updating the American Psychiatric Association's framework through a stakeholder-engaged workshop. Psychiatr Serv 2021;72(9):1095-1098 [FREE Full text] [doi: 10.1176/appi.ps.202000663] [Medline: 33882716]
46. M-Health Index and Navigation Database. MindApps. URL: https://mindapps.org [accessed 2023-12-08]
47. Lagan S, Aquino P, Emerson MR, Fortuna K, Walker R, Torous J. Actionable health app evaluation: translating expert frameworks into objective metrics. NPJ Digit Med 2020;3:100 [FREE Full text] [doi: 10.1038/s41746-020-00312-4] [Medline: 32821855]
48. Gordon WJ, Landman A, Zhang H, Bates DW. Beyond validation: getting health apps into clinical practice. NPJ Digit Med 2020;3:14 [FREE Full text] [doi: 10.1038/s41746-019-0212-z] [Medline: 32047860]
49. Mathews SC, McShea MJ, Hanley CL, Ravitz A, Labrique AB, Cohen AB. Digital health: a path to validation. NPJ Digit Med 2019;2:38 [FREE Full text] [doi: 10.1038/s41746-019-0111-3] [Medline: 31304384]
50. Lagan S, Sandler L, Torous J. Evaluating evaluation frameworks: a scoping review of frameworks for assessing health apps. BMJ Open 2021;11(3):e047001 [FREE Full text] [doi: 10.1136/bmjopen-2020-047001] [Medline: 33741674]
51. Hoober S, Berkman E. Designing Mobile Interfaces, 1st Edition. Sebastopol, CA: O'Reilly; 2012.

52. mindLAMP digital app. Division of Digital Psychiatry at BIDMC. URL: https://www.digitalpsych.org/lamp.html [accessed 2023-12-12]
53. Lamp Platform. URL: https://docs.lamp.digital/ [accessed 2023-12-12]
54. MongDB. URL: https://www.mongodb.com/ [accessed 2023-12-12]
55. Kato T. Functional value vs emotional value: a comparative study of the values that contribute to a preference for a corporate brand. Int J Inf Manag Data Insights 2021;1(2):100024. [doi: 10.1016/j.jjimei.2021.100024]
56. Lazard AJ, Brennen JSB, Belina SP. App designs and interactive features to increase mHealth adoption: user expectation survey and experiment. JMIR Mhealth Uhealth 2021;9(11):e29815 [FREE Full text] [doi: 10.2196/29815] [Medline: 34734829]
57. Iten GH, Troendle A, Opwis K. Aesthetics in context—the role of aesthetics and usage mode for a website's success. Interact Comput 2018;30(2):133-149. [doi: 10.1093/iwc/iwy002]
58. React Native. URL: https://reactnative.dev/ [accessed 2023-12-12]
59. Akbar S, Coiera E, Magrabi F. Safety concerns with consumer-facing mobile health applications and their consequences: a scoping review. J Am Med Inform Assoc 2020;27(2):330-340 [FREE Full text] [doi: 10.1093/jamia/ocz175] [Medline: 31599936]
60. SOMA app privacy policy. SOMA App. URL: https://somatheapp.com/privacy-policy-2/ [accessed 2023-12-12]
61. Hauser-Ulrich S, Künzli H, Meier-Peterhans D, Kowatsch T. A smartphone-based health care Chatbot to promote self-management of chronic pain (SELMA): pilot randomized controlled trial. JMIR Mhealth Uhealth 2020;8(4):e15806 [FREE Full text] [doi: 10.2196/15806] [Medline: 32242820]
62. Ono M, Schneider S, Junghaenel DU, Stone AA. What affects the completion of ecological momentary assessments in chronic pain research? An individual patient data meta-analysis. J Med Internet Res 2019;21(2):e11398 [FREE Full text] [doi: 10.2196/11398] [Medline: 30720437]
63. Williams MT, Lewthwaite H, Fraysse F, Gajewska A, Ignatavicius J, Ferrar K. Compliance with mobile ecological momentary assessment of self-reported health-related behaviors and psychological constructs in adults: systematic review and meta-analysis. J Med Internet Res 2021;23(3):e17023 [FREE Full text] [doi: 10.2196/17023] [Medline: 33656451]
64. Robinson MD, Clore GL. Belief and feeling: evidence for an accessibility model of emotional self-report. Psychol Bull 2002;128(6):934-960. [doi: 10.1037/0033-2909.128.6.934] [Medline: 12405138]
65. Verhagen SJW, Hasmi L, Drukker M, van Os J, Delespaul PAEG. Use of the experience sampling method in the context of clinical trials. Evid Based Ment Health 2016;19(3):86-89 [FREE Full text] [doi: 10.1136/ebmental-2016-102418] [Medline: 27443678]
66. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
67. Mikati N, Suffoletto B, Pacella-LaBarbara M, Callaway C. Comparison of musculoskeletal pain severity and pain interference using ecological momentary assessment: a longitudinal analysis. J Pain 2021;22(5):594 [FREE Full text] [doi: 10.1016/j.jpain.2021.03.066]
68. Hektner JM, Schmidt JA, Csikszentmihalyi M. Experience Sampling Method: Measuring the Quality of Everyday Life. Thousand Oaks, CA: Sage Publications; 2007.
69. Winger JG, Vilardaga JCP, Keefe FJ. Indices of pain variability: a paradigm shift. Pain 2019;160(11):2411-2412 [FREE Full text] [doi: 10.1097/j.pain.0000000000001627] [Medline: 31145216]
70. Mun CJ, Suk HW, Davis MC, Karoly P, Finan P, Tennen H, et al. Investigating intraindividual pain variability: methods, applications, issues, and directions. Pain 2019;160(11):2415-2429. [doi: 10.1097/j.pain.0000000000001626] [Medline: 31145212]
71. van der Krieke L, Emerencia AC, Bos EH, Rosmalen JG, Riese H, Aiello M, et al. Ecological momentary assessments and automated time series analysis to promote tailored health care: a proof-of-principle study. JMIR Res Protoc 2015;4(3):e100 [FREE Full text] [doi: 10.2196/resprot.4000] [Medline: 26254160]
72. Tighe PJ, Bzdega M, Fillingim RB, Rashidi P, Aytug H. Markov chain evaluation of acute postoperative pain transition states. Pain 2016;157(3):717-728 [FREE Full text] [doi: 10.1097/j.pain.0000000000000429] [Medline: 26588689]
73. Wang Y, Jacques JJ, Li Z, Sibille KT, Cook RL. Health outcomes among adults initiating medical cannabis for chronic pain: a 3-month prospective study incorporating Ecological Momentary Assessment (EMA). Cannabis 2021;4(2):69-83 [FREE Full text] [doi: 10.26828/cannabis/2021.02.006] [Medline: 34671723]
74. May M, Junghaenel DU, Ono M, Stone AA, Schneider S. Ecological momentary assessment methodology in chronic pain research: a systematic review. J Pain 2018;19(7):699-716 [FREE Full text] [doi: 10.1016/j.jpain.2018.01.006] [Medline: 29371113]
75. Stone AA, Obbarius A, Junghaenel DU, Wen CKF, Schneider S. High-resolution, field approaches for assessing pain: ecological momentary assessment. Pain 2021;162(1):4-9 [FREE Full text] [doi: 10.1097/j.pain.0000000000002049] [Medline: 32833794]
76. Computer and internet use in the United States: 2018. United States Census Bureau. 2021. URL: https://www.census.gov/newsroom/press-releases/2021/computer-internet-use.html [accessed 2023-12-08]

77. Rothwell PM. External validity of randomised controlled trials: "to whom do the results of this trial apply?". Lancet 2005;365(9453):82-93. [doi: 10.1016/S0140-6736(04)17670-8] [Medline: 15639683]
78. Fiore LD, Lavori PW. Integrating randomized comparative effectiveness research with patient care. N Engl J Med 2016;374(22):2152-2158 [FREE Full text] [doi: 10.1056/NEJMra1510057] [Medline: 27248620]
79. Hohenschurz-Schmidt D, Kleykamp BA, Draper-Rodi J, Vollert J, Chan J, Ferguson M, et al. Pragmatic trials of pain therapies: a systematic review of methods. Pain 2022;163(1):21-46 [FREE Full text] [doi: 10.1097/j.pain.0000000000002317] [Medline: 34490854]
80. Kopf A, Patel NB. Guide to pain management in low-resource settings. International Association for the Study of Pain. 2010. URL: https://iaspfiles.s3.amazonaws.com/production/public/2021/IASP-Guide_to_Pain_Management_in_Low-Resource_Settings-English.pdf [accessed 2023-12-08]
81. Vogels EA. Some digital divides persist between rural, urban and suburban America. Pew Research Center. 2021. URL: https://www.pewresearch.org/fact-tank/2021/08/19/some-digital-divides-persist-between-rural-urban-and-suburban-america/ [accessed 2023-12-08]
82. Kemp S. Digital 2021: global overview report. DataReportal. 2021. URL: https://datareportal.com/reports/digital-2021-global-overview-report [accessed 2023-12-08]
83. Busch PA, Hausvik GI, Ropstad OK, Pettersen D. Smartphone usage among older adults. Comput Hum Behav 2021;121:106783 [FREE Full text] [doi: 10.1016/j.chb.2021.106783]
84. Ali SM, Lee RR, McBeth J, James B, McAlister S, Chiarotto A, et al. Exploring the cross-cultural acceptability of digital tools for pain self-reporting: qualitative study. JMIR Hum Factors 2023;10:e42177 [FREE Full text] [doi: 10.2196/42177] [Medline: 36753324]
85. Meints SM, Cortes A, Morais CA, Edwards RR. Racial and ethnic differences in the experience and treatment of noncancer pain. Pain Manag 2019;9(3):317-334 [FREE Full text] [doi: 10.2217/pmt-2018-0030] [Medline: 31140916]
86. Palermo TM, Davis KD, Bouhassira D, Hurley RW, Katz JD, Keefe FJ, et al. Promoting inclusion, diversity, and equity in pain science. Pain Med 2023;24(2):105-109 [FREE Full text] [doi: 10.1093/pm/pnac204] [Medline: 36622210]
87. Morasco B, Lovejoy T, Hyde S, Shull S, Dobscha S. Limitations of pain numeric rating scale scores collected during usual care: need for enhanced assessment. J Pain 2018;19(3):S57-S58. [doi: 10.1016/j.jpain.2017.12.145]
88. Treede RD, Rief W, Barke A, Aziz Q, Bennett MI, Benoliel R, et al. Chronic pain as a symptom or a disease: the IASP classification of chronic pain for the International Classification of Diseases (ICD-11). Pain 2019;160(1):19-27. [doi: 10.1097/j.pain.0000000000001384] [Medline: 30586067]
89. Keller S, Bann CM, Dodd SL, Schein J, Mendoza TR, Cleeland CS. Validity of the brief pain inventory for use in documenting the outcomes of patients with noncancer pain. Clin J Pain 2004;20(5):309-318. [doi: 10.1097/00002508-200409000-00005] [Medline: 15322437]
90. Dworkin RH, Turk DC, Trudeau JJ, Benson C, Biondi DM, Katz NP, et al. Validation of the Short-Form McGill Pain Questionnaire-2 (SF-MPQ-2) in acute low back pain. J Pain 2015;16(4):357-366 [FREE Full text] [doi: 10.1016/j.jpain.2015.01.012] [Medline: 25640290]
91. Melzack R. The McGill pain questionnaire: major properties and scoring methods. Pain 1975;1(3):277-299. [doi: 10.1016/0304-3959(75)90044-5] [Medline: 1235985]
92. Brummett CM, Bakshi RR, Goesling J, Leung D, Moser SE, Zollars JW, et al. Preliminary validation of the Michigan body map. Pain 2016;157(6):1205-1212 [FREE Full text] [doi: 10.1097/j.pain.0000000000000506] [Medline: 26835782]
93. McMahon SB, Tracey I, Koltzenberg M, Turk DC. Wall and Melzack's Textbook of Pain, 6th Edition. Philadelphia, PA: Elsevier/Saunders; 2013.
94. Caseiro M, Woznowski-Vu A, De Oliveira AS, Reis FJJ, Wideman TH. From paper to digitalized body map: a reliability study of the pain area. Pain Pract 2019;19(6):602-608. [doi: 10.1111/papr.12780] [Medline: 30884135]
95. Foxen-Craft E, Scott EL, Kullgren KA, Philliben R, Hyman C, Dorta M, et al. Pain location and widespread pain in youth with orthopaedic conditions: exploration of the reliability and validity of a body map. Eur J Pain 2019;23(1):57-65 [FREE Full text] [doi: 10.1002/ejp.1282] [Medline: 29978523]
96. Jaatun EAA, Haugen DF, Dahl Y, Kofod-Petersen A. Proceed with caution: transition from paper to computerized pain body maps. Procedia Comput Sci 2013;21:398-406 [FREE Full text] [doi: 10.1016/j.procs.2013.09.052]
97. Ali SM, Lau WJ, McBeth J, Dixon WG, van der Veer SN. Digital manikins to self-report pain on a smartphone: a systematic review of mobile apps. Eur J Pain 2021;25(2):327-338 [FREE Full text] [doi: 10.1002/ejp.1688] [Medline: 33113241]
98. Nahin RL. Use of multimodal multidisciplinary pain management in the US. JAMA Netw Open 2022;5(11):e2240620 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.40620] [Medline: 36342720]
99. Institute of Medicine (U.S.), Committee on Advancing Pain Research, Care, and Education. Relieving Pain in America: A Blueprint for Transforming Prevention, Care, Education, and Research. Washington, DC: National Academies Press; 2011.
100. Elbers S, Wittink H, Konings S, Kaiser U, Kleijnen J, Pool J, et al. Longitudinal outcome evaluations of interdisciplinary multimodal pain treatment programmes for patients with chronic primary musculoskeletal pain: a systematic review and meta-analysis. Eur J Pain 2022;26(2):310-335 [FREE Full text] [doi: 10.1002/ejp.1875] [Medline: 34624159]

101. Bujak BK, Regan E, Beattie PF, Harrington S. The effectiveness of interdisciplinary intensive outpatient programs in a population with diverse chronic pain conditions: a systematic review and meta-analysis. Pain Manag 2019;9(4):417-429. [doi: 10.2217/pmt-2018-0087] [Medline: 31237177]
102. Eucker SA, Knisely MR, Simon C. Nonopioid treatments for chronic pain-integrating multimodal biopsychosocial approaches to pain management. JAMA Netw Open 2022;5(6):e2216482 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.16482] [Medline: 35687341]
103. Gewandter JS, McDermott MP, He H, Gao S, Cai X, Farrar JT, et al. Demonstrating heterogeneity of treatment effects among patients: an overlooked but important step toward precision medicine. Clin Pharmacol Ther 2019;106(1):204-210 [FREE Full text] [doi: 10.1002/cpt.1372] [Medline: 30661240]
104. Edwards RR, Dworkin RH, Turk DC, Angst MS, Dionne R, Freeman R, et al. Patient phenotyping in clinical trials of chronic pain treatments: IMMPACT recommendations. Pain Rep 2021;6(1):e896 [FREE Full text] [doi: 10.1097/PR9.0000000000000896]
105. Vardeh D, Mannion RJ, Woolf CJ. Toward a mechanism-based approach to pain diagnosis. J Pain 2016;17(Suppl 9):T50-T69 [FREE Full text] [doi: 10.1016/j.jpain.2016.03.001] [Medline: 27586831]
106. Raja SN, Carr DB, Cohen M, Finnerup NB, Flor H, Gibson S, et al. The revised International Association for the Study of Pain definition of pain: concepts, challenges, and compromises. Pain 2020;161(9):1976-1982 [FREE Full text] [doi: 10.1097/j.pain.0000000000001939] [Medline: 32694387]
107. Melzack R, Casey KL. Sensory, motivational, and central control determinants of pain: a new conceptual model. Skin Senses 1968;1:423-443 [FREE Full text]
108. Gilam G, Gross JJ, Wager TD, Keefe FJ, Mackey SC. What is the relationship between pain and emotion? Bridging constructs and communities. Neuron 2020;107(1):17-21 [FREE Full text] [doi: 10.1016/j.neuron.2020.05.024] [Medline: 32562660]
109. Sheng J, Liu S, Wang Y, Cui R, Zhang X. The link between depression and chronic pain: neural mechanisms in the brain. Neural Plast 2017;2017:9724371 [FREE Full text] [doi: 10.1155/2017/9724371] [Medline: 28706741]
110. Lumley MA, Cohen JL, Borszcz GS, Cano A, Radcliffe AM, Porter LS, et al. Pain and emotion: a biopsychosocial review of recent research. J Clin Psychol 2011;67(9):942-968 [FREE Full text] [doi: 10.1002/jclp.20816] [Medline: 21647882]
111. Linton SJ. A transdiagnostic approach to pain and emotion. J Appl Biobehav Res 2013;18(2):82-103 [FREE Full text] [doi: 10.1111/jabr.12007] [Medline: 24143062]
112. Gendron M, Barrett LF. Reconstructing the past: a century of ideas about emotion in psychology. Emot Rev 2009;1(4):316-339 [FREE Full text] [doi: 10.1177/1754073909338877] [Medline: 20221412]
113. Mauss IB, Robinson MD. Measures of emotion: a review. Cogn Emot 2009;23(2):209-237 [FREE Full text] [doi: 10.1080/02699930802204677] [Medline: 19809584]
114. Ekman P, Friesen WV. Constants across cultures in the face and emotion. J Pers Soc Psychol 1971;17(2):124-129. [doi: 10.1037/h0030377] [Medline: 5542557]
115. Tomkins SS. Affect, Imagery, Consciousness, Vol. 1: The Positive Affects. New York: Springer Publishing Co; 1962.
116. Trampe D, Quoidbach J, Taquet M. Emotions in everyday life. PLoS One 2015;10(12):e0145450 [FREE Full text] [doi: 10.1371/journal.pone.0145450] [Medline: 26698124]
117. Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: the PANAS scales. J Pers Soc Psychol 1988;54(6):1063-1070. [doi: 10.1037//0022-3514.54.6.1063] [Medline: 3397865]
118. Russell JA, Weiss A, Mendelsohn GA. Affect grid: a single-item scale of pleasure and arousal. J Pers Soc Psychol 1989;57(3):493-502. [doi: 10.1037/0022-3514.57.3.493]
119. Heffner J, FeldmanHall O. A probabilistic map of emotional experiences during competitive social interactions. Nat Commun 2022;13(1):1718 [FREE Full text] [doi: 10.1038/s41467-022-29372-8] [Medline: 35361768]
120. Ellsworth PC, Scherer KR. Appraisal processes in emotion. In: Davidson RJ, Scherer KR, Goldsmith HH, editors. Handbook of Affective Sciences. New York, NY: Oxford University Press; 2003:572-595.
121. Bliss-Moreau E, Williams LA, Santistevan AC. The immutability of valence and arousal in the foundation of emotion. Emotion 2020;20(6):993-1004. [doi: 10.1037/emo0000606] [Medline: 31192659]
122. Watson D, Clark LA. On traits and temperament: general and specific factors of emotional experience and their relation to the five-factor model. J Pers 1992;60(2):441-476. [doi: 10.1111/j.1467-6494.1992.tb00980.x] [Medline: 1635050]
123. Chekroud SR, Gueorguieva R, Zheutlin AB, Paulus M, Krumholz HM, Krystal JH, et al. Association between physical exercise and mental health in 1·2 million individuals in the USA between 2011 and 2015: a cross-sectional study. Lancet Psychiatry 2018;5(9):739-746. [doi: 10.1016/S2215-0366(18)30227-X] [Medline: 30099000]
124. Taquet M, Quoidbach J, de Montjoye YA, Desseilles M, Gross JJ. Hedonism and the choice of everyday activities. Proc Natl Acad Sci U S A 2016;113(35):9769-9773 [FREE Full text] [doi: 10.1073/pnas.1519998113] [Medline: 27528666]
125. Taquet M, Quoidbach J, Gross JJ, Saunders KEA, Goodwin GM. Mood homeostasis, low mood, and history of depression in 2 large population samples. JAMA Psychiatry 2020;77(9):944-951 [FREE Full text] [doi: 10.1001/jamapsychiatry.2020.0588] [Medline: 32320005]
126. Baliki MN, Apkarian AV. Nociception, pain, negative moods, and behavior selection. Neuron 2015;87(3):474-491 [FREE Full text] [doi: 10.1016/j.neuron.2015.06.005] [Medline: 26247858]

127. Hashmi JA, Baliki MN, Huang L, Baria AT, Torbey S, Hermann KM, et al. Shape shifting pain: chronification of back pain shifts brain representation from nociceptive to emotional circuits. Brain 2013;136(Pt 9):2751-2768 [FREE Full text] [doi: 10.1093/brain/awt211] [Medline: 23983029]
128. Booth J, Moseley GL, Schiltenwolf M, Cashin A, Davies M, Hübscher M. Exercise for chronic musculoskeletal pain: a biopsychosocial approach. Musculoskeletal Care 2017;15(4):413-421. [doi: 10.1002/msc.1191] [Medline: 28371175]
129. Aebischer O, Suter MR, Vollenweider P, Marques-Vidal P. Association between chronic pain and physical activity in a Swiss population-based cohort: a cross-sectional study. BMJ Open 2022;12(7):e057288 [FREE Full text] [doi: 10.1136/bmjopen-2021-057288] [Medline: 35906050]
130. van den Berg-Emons RJ, Schasfoort FC, de Vos LA, Bussmann JB, Stam HJ. Impact of chronic pain on everyday physical activity. Eur J Pain 2007;11(5):587-593. [doi: 10.1016/j.ejpain.2006.09.003] [Medline: 17107818]
131. Colloca L, Klinger R, Flor H, Bingel U. Placebo analgesia: psychological and neurobiological mechanisms. Pain 2013;154(4):511-514 [FREE Full text] [doi: 10.1016/j.pain.2013.02.002] [Medline: 23473783]
132. Hird EJ, Charalambous C, El-Deredy W, Jones AKP, Talmi D. Boundary effects of expectation in human pain perception. Sci Rep 2019;9(1):9443 [FREE Full text] [doi: 10.1038/s41598-019-45811-x] [Medline: 31263144]
133. Klinger R, Colloca L, Bingel U, Flor H. Placebo analgesia: clinical applications. Pain 2014;155(6):1055-1058 [FREE Full text] [doi: 10.1016/j.pain.2013.12.007] [Medline: 24333780]
134. Bingel U, Placebo Competence Team. Avoiding nocebo effects to optimize treatment outcome. JAMA 2014;312(7):693-694 [FREE Full text] [doi: 10.1001/jama.2014.8342] [Medline: 25003609]
135. Müller-Schrader M, Heinzle J, Müller A, Lanz C, Häussler O, Sutter M, et al. Individual treatment expectations predict clinical outcome after lumbar injections against low back pain. Pain 2023;164(1):132-141 [FREE Full text] [doi: 10.1097/j.pain.0000000000002674] [Medline: 35543638]
136. Wiech K. Deconstructing the sensation of pain: the influence of cognitive processes on pain perception. Science 2016;354(6312):584-587. [doi: 10.1126/science.aaf8934] [Medline: 27811269]
137. Main CJ, Foster N, Buchbinder R. How important are back pain beliefs and expectations for satisfactory recovery from back pain? Best Pract Res Clin Rheumatol 2010;24(2):205-217. [doi: 10.1016/j.berh.2009.12.012] [Medline: 20227642]
138. Kamper SJ, Kongsted A, Haanstra TM, Hestbaek L. Do recovery expectations change over time? Eur Spine J 2015;24(2):218-226. [doi: 10.1007/s00586-014-3380-1] [Medline: 24913213]

Abbreviations

APA: American Psychiatric Association
ESM: experience sampling methodology
FDA: Food and Drug Administration
IMMPACT: Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials
mHealth: mobile health
NIH: National Institutes of Health
VAS: visual analog scale

Edited by L Buis; submitted 13.03.23; peer-reviewed by S Ali, S Rahmatizadeh; comments to author 08.08.23; revised version received 03.10.23; accepted 30.11.23; published 12.01.24.

Please cite as:
Gunsilius CZ, Heffner J, Bruinsma S, Corinha M, Cortinez M, Dalton H, Duong E, Lu J, Omar A, Owen LLW, Roarr BN, Tang K, Petzschner FH
SOMAScience: A Novel Platform for Multidimensional, Longitudinal Pain Assessment
JMIR Mhealth Uhealth 2024;12:e47177
URL: https://mhealth.jmir.org/2024/1/e47177
doi:10.2196/47177
PMID:38214952

Viewpoint

Raw Photoplethysmography as an Enhancement for Research-Grade Wearable Activity Monitors

Paul R Hibbing[1], PhD; Maryam Misal Khan[1,2], MS

[1]Department of Kinesiology and Nutrition, University of Illinois Chicago, Chicago, IL, United States

[2]Department of Kinesiology and Health Sciences, University of Waterloo, Waterloo, ON, Canada

Corresponding Author:
Paul R Hibbing, PhD
Department of Kinesiology and Nutrition
University of Illinois Chicago
1919 W Taylor St
Rm 650, Mail Code 517
Chicago, IL, 60612
United States
Phone: 1 312 355 1088
Fax: 1 312 413 0319
Email: phibbing@uic.edu

Abstract

Wearable monitors continue to play a critical role in scientific assessments of physical activity. Recently, research-grade monitors have begun providing raw data from photoplethysmography (PPG) alongside standard raw data from inertial sensors (accelerometers and gyroscopes). Raw PPG enables granular and transparent estimation of cardiovascular parameters such as heart rate, thus presenting a valuable alternative to standard PPG methodologies (most of which rely on consumer-grade monitors that provide only coarse output from proprietary algorithms). The implications for physical activity assessment are tremendous, since it is now feasible to monitor granular and concurrent trends in both movement and cardiovascular physiology using a single noninvasive device. However, new users must also be aware of challenges and limitations that accompany the use of raw PPG data. This viewpoint paper therefore orients new users to the opportunities and challenges of raw PPG data by presenting its mechanics, pitfalls, and availability, as well as its parallels and synergies with inertial sensors. This includes discussion of specific applications to the prediction of energy expenditure, activity type, and 24-hour movement behaviors, with an emphasis on areas in which raw PPG data may help resolve known issues with inertial sensing (eg, measurement during cycling activities). We also discuss how the impact of raw PPG data can be maximized through the use of open-source tools when developing and disseminating new methods, similar to current standards for raw accelerometer and gyroscope data. Collectively, our comments show the strong potential of raw PPG data to enhance the use of research-grade wearable activity monitors in science over the coming years.

(JMIR Mhealth Uhealth 2024;12:e57158) doi:10.2196/57158

KEYWORDS

measurement; optical sensors; sensor fusion; wearable electronic devices; accelerometry; photoplethysmography; digital health; exercise; sedentary behavior

Introduction

Wearable monitors are increasingly used to measure physical activity in research, and new tools and techniques are continually emerging [1]. Recent innovations have improved the cost, size, and technical capability of various monitors [2], but accuracy has not increased at a commensurate pace [3-6]. Thus, there is a need for further innovation. Successful innovation will likely entail novel measurement paradigms, rather than incremental improvements on current techniques [6]. One of the most promising and underexplored paradigms is to integrate data from multiple types of sensors, rather than the traditional use of only accelerometer sensors [7,8].

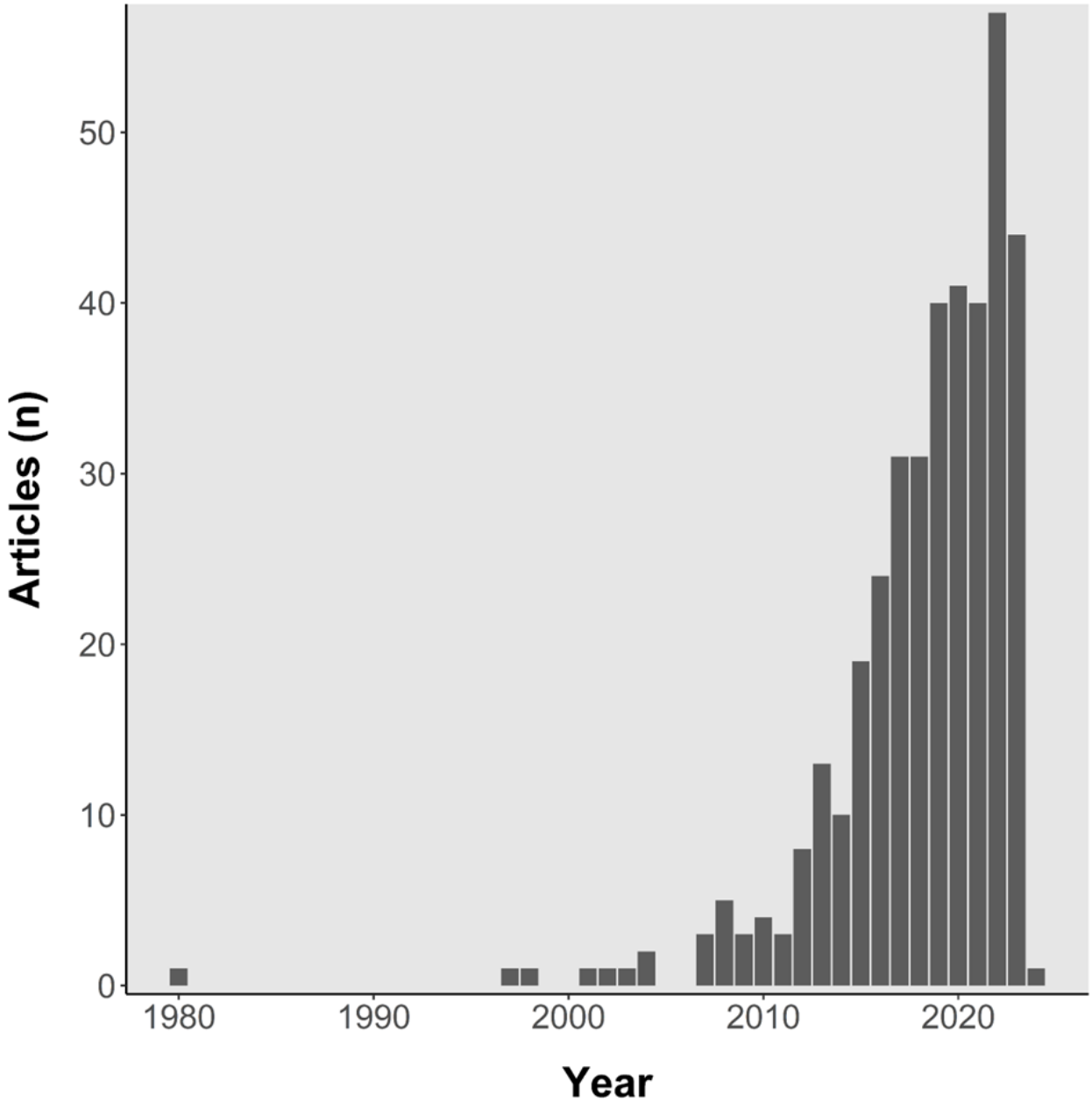
Photoplethysmography (PPG) is an optical technology that may have potential to enhance physical activity measurement when combined with established inertial sensors (accelerometers and gyroscopes) [9]. Although PPG was first described nearly 90 years ago, it has only recently gained a high level of visibility for physical activity assessment [10-12]. This growth is reflected in Figure 1, which shows the results of a Scopus search for

documents addressing physical activity and PPG. Roughly half of the identified studies (183/385, 48%) were published in 2020 or later, and roughly three-quarters (285/385, 74%) were published in 2017 or later.

To date, most applications of PPG for physical activity assessment have involved consumer-grade smartwatches [13-16]. A wealth of developmental research has also been reported in the engineering literature [17-20], but commercial products for research have rarely incorporated PPG sensors and even more rarely given access to raw PPG data (ie, the data recorded by the sensor itself, without any preprocessing applied) [21]. This is beginning to change, and as it does, there is a need to raise awareness of PPG and its potential contribution to monitor-based physical activity assessment. In particular, awareness is needed for *raw* PPG data since it provides an avenue for device-agnostic measurement and iterative, open-source refinement, similar to the standard for inertial sensing [22].

In this viewpoint paper, we present raw PPG as a new frontier in monitor-based methodology. To do this, we first provide an overview of the fundamentals of PPG for physical activity assessment, after which we describe the importance and availability of raw PPG data, as well as specific applications where it holds the most potential. Throughout, we highlight ways that raw PPG data can synergize with raw data from inertial sensors to overcome long-standing challenges (eg, measurement during cycling).

Figure 1. Annual publication counts over time, drawn from a Scopus.com search for "TITLE-ABS-KEY (physical AND activity AND photoplethysmography OR ppg)," conducted on December 6, 2023.



Fundamentals of PPG for Physical Activity Assessment

Technology, Techniques, and Theory

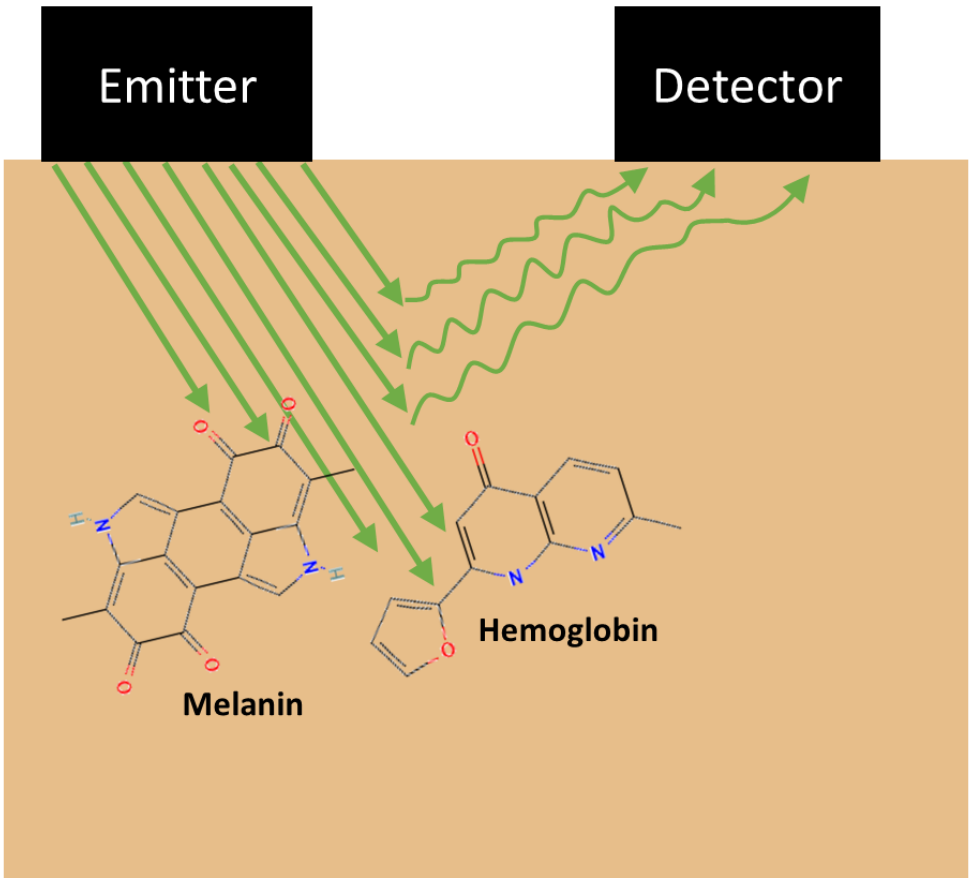
There are 2 types of PPG, namely, transmission and reflectance [23]. Transmission PPG is common in clinical settings where it is used for pulse oximetry [12]. It typically involves red and near-infrared lights, which are shone into one side of a tissue (commonly a finger or an earlobe) and measured upon exiting the other side [10,24,25]. In physical activity assessments, transmission PPG has limited use compared with reflectance PPG. Therefore, we do not provide further comments on transmission PPG.

Reflectance PPG has been investigated using both "wearable" and "remote" instruments, the latter referring to cameras that do not touch the skin. Similar to transmission PPG, remote applications of reflectance PPG have minimal relevance for physical activity assessment, and thus we forgo additional comments on them. Instead, we focus our comments on wearable applications of reflectance PPG, particularly those embedded in wrist-worn devices. Hereafter, we use the term "PPG" to refer exclusively to such applications.

As noted by Mannheimer [26], the term "reflectance" is a misnomer, since there are no mirrors in the skin. Instead, light is scattered by various components of the tissue, and portions of the scattered light return to the surface where they can be measured by a photodetector. Thus, the defining characteristic of this PPG technique is that emission and measurement of light occur on the same side of the tissue [27].

There is some debate around what exactly PPG captures, but the prevailing theory is that it detects pulsatile changes in blood volume [28-30]. Figure 2 depicts the mechanics of this proposed process, with light being shone into the skin while cyclical fluctuations in scatter are monitored. These fluctuations occur because blood concentration is increased when a pulse wave passes under the light, resulting in more light absorption in accordance with the Beer-Lambert law [26,31]. Consequently, a waveform emerges in the PPG signal, which can be analyzed to detect pulse waves and calculate related parameters such as heart rate [9,10,23]. Green light is typically used because it offers shallower penetration and greater robustness against motion artifacts and other noise [17,19,32-35].

Figure 2. Basic representation of PPG technology. An emitter shines light into the skin. The light is absorbed by some components—mainly hemoglobin and melanin—and scattered by others toward a photodetector. Pulse waves cause increases in local blood concentration, leading the balance of absorption and scatter to shift in favor of more absorption. The photodetector signal thus diminishes as the pulse wave passes, creating a waveform that can be analyzed to predict cardiovascular parameters such as heart rate and blood pressure. Public domain icons from PubChem are shown for melanin (CID 6325610) and (deoxy)hemoglobin (CID 135310457).



Common Difficulties and Sources of Error

It is important to understand not only the theoretical workings of PPG but also practical issues that affect its operation. Fine et al [36] have grouped such issues into 3 categories (individual differences, physiology, and external factors), while Bent et al [37] have highlighted the unique importance of one overarching individual factor (skin type) and 2 external factors (motion artifact and signal crossover). In this section, we provide a brief overview of these latter 3 factors and their potential implications for physical activity research, with additional comments on broader sources of error for general applications of PPG.

Skin type encompasses adiposity, pigmentation, and other factors that influence tissue composition (see the study by Fine et al [36] for a detailed listing). Differences in tissue composition can affect optical scattering and absorption in ways that are difficult to predict. Accordingly, prior work has shown the accuracy of PPG-based estimates to vary depending on age, sex, obesity status, and skin tone [13,20,38-40]. The latter is an especially important variable to consider because melanin is one of the skin's main absorbers of light at various wavelengths [41,42]. When using PPG for the measurement of physical activity, there is thus a clear need to ensure that new methods have consistent accuracy across diverse skin types. This is especially important given the implications for equity in health research.

Motion artifact is movement-induced noise in the PPG signal, which can occur due to both mechanical and physiological aspects of the movement [43]. Efforts to address motion artifact often rely on frequency-domain analyses, since it is expected that the rhythmicity of the pulse will create a sharper contrast between signal and noise in that domain [44]. Increasingly, these analyses involve cross-referencing PPG against data from concurrently worn accelerometer and gyroscope sensors to aid in differentiating between inertial and cardiovascular signal [45]. Thus, when using PPG for physical activity assessments, future studies may benefit from using devices that provide access to raw data from both PPG and inertial sensors.

Signal crossover is closely related to motion artifact and refers to confusion between rhythmic motions of the monitor itself (eg, during locomotion) and the inherent rhythmicity of the cardiovascular signal from PPG [37]. Specifically, cardiovascular signal is expected to have dominant frequencies between roughly 1.0 Hz and 3.5 Hz (corresponding to heart rates between 60 and 210 beats per minute), and human movements can generate considerable amplitude in the same range [46-48]. Thus, it is likely that some motions will result in overlap of inertial and pulsatile signal components, making it hard to tell which is which. This is one reason that PPG-based devices have frequently been shown to have lower accuracy during physical activity than during other behaviors [37,49,50]. Signal crossover is uniquely important to highlight because it suggests that device accuracy may vary based on not only the amount of movement but also the type of movement. This could have major implications for physical activity assessments, making it crucial to address in future work.

Apart from skin type, motion artifact, and signal crossover, Fine et al [36] have highlighted difficulties posed by physiological factors (respiration, venous pulsations, attachment site of the device, and body temperature) and additional external factors (ambient light and pressure of the sensor on the skin). These difficulties are important to acknowledge and address, but their implications may not be substantively different for physical activity research than what has been described for other disciplines.

Many of the difficulties with raw PPG resemble what is already faced when dealing with inertial data from accelerometers and gyroscopes [51]. The latter sensors have enhanced the measurement of physical activity despite their limitations [52], and thus PPG may have similar potential. Furthermore, the impact of device limitations may diminish over time through ongoing innovation in technology and analytics. Thus, the difficult aspects of PPG can be viewed as opportunities for refinement rather than insurmountable barriers.

Importance and Availability of Raw PPG Data

To understand the revolutionary potential of raw PPG data for physical activity assessment, it is helpful to consider a similar revolution that has already taken place with accelerometer data [53-55]. Historically, accelerometer-based devices provided only proprietary "activity counts" as their output, which led to intermonitor differences and a lack of flexibility to innovate with new data processing techniques [56-58]. Over time, raw acceleration data became commonplace, opening doors for standardization and innovation in physical activity research [2]. An especially noticeable result was that many researchers began to focus on techniques that combined research-grade products with open-source tools for data processing and analysis, thereby promoting streamlined and coordinated progress in the field [59-64].

The potential parallels for PPG data are striking. To date, most research with PPG has relied on proprietary outputs from consumer-grade devices, which have been used to track heart rate, atrial fibrillation, blood pressure, and more [65-76]. Intermonitor differences and lack of flexibility are thus limitations of current standards for PPG, in much the same way as they once were for accelerometry. Furthermore, concerns have frequently been raised about unannounced algorithm and firmware updates that can make consumer-grade technology undesirable in certain research contexts [77-81]. The advent of raw PPG data therefore offers many of the same benefits that have already been derived from raw accelerometer data, especially when pairing research-grade devices with open-source resources. But the full potential of raw PPG for physical activity research cannot be realized unless the market provides devices that are scalable for use in large studies.

Existing research involving raw PPG data has generally involved small-scale devices (sometimes custom-made) [19,21,27,30,82], specialized tools for hospital use [83-86], or smartphone technology [87-89]. While these studies have shown strong proof-of-concept, they have only sometimes been oriented toward physical activity research, and the availability of suitable devices for large assessments remains an issue. The best-known

research-grade devices are likely the E4 and EmbracePlus from Empatica Inc, the Shimmer3+ GSR from Shimmer Sensing, and the LEAP from ActiGraph LLC. Each device has strengths and limitations, with major points of comparison being cost, comfort, and access to raw data. The Shimmer3+ GSR is the most affordable option and provides access to fully raw data from both PPG and inertial sensors. However, a potential limitation is its reliance on physical components (eg, wired probes that wrap around the fingers) that may be unappealing or uncomfortable for some participants. The EmbracePlus is the most expensive device and is a replacement for the E4. It is designed like a standard smartwatch and is therefore very comfortable, but it does not provide truly raw PPG data (nor did the E4 [21]). Specifically, the EmbracePlus preprocesses PPG data using a proprietary algorithm that produces a blood volume pulse waveform, which resembles but does not replace raw data as it would appear in a direct recording from the photodetector. The LEAP device falls between the other 2 in terms of cost and comfort but does provide access to fully raw data from PPG and inertial sensors.

As new and upgraded devices continue to emerge and provide access to raw PPG data, a key objective will be to apply, extend, and standardize the techniques from earlier proof-of-concept studies for use in large-scale physical activity assessments for research. The following section outlines several specific areas in which there may be greatest warrant for these efforts.

Potential Applications of Raw PPG in Assessments of Physical Activity

The most obvious application of raw PPG for physical activity assessment is heart rate monitoring, where a notable contrast exists between the wrist-based optical approach and standard electrode-based approaches involving chest-worn monitors (eg, heart rate straps and Holter monitors). The latter tend to have greater accuracy than the former [90,91] and yet can also be uncomfortable to wear, especially over long periods [92]. Conversely, wrist-based PPG devices can be comfortably worn over long periods and yet have lower accuracy than chest-worn monitors. One implication is that raw PPG may encourage participant compliance in long assessment protocols (eg, lasting a week or more, which is common in physical activity research). This could be especially valuable for interventions that assess change over time, since responsiveness is generally a greater concern than accuracy in those contexts. Moreover, the diminished accuracy compared with chest-worn monitors may be less of an issue in cases where the key outcome is categorical intensity rather than continuous heart rate (eg, if assessing time in heart rate zones, where measurement error would be a concern only at the boundaries between zones, rather than across the spectrum of continuous heart rates). Nevertheless, there is a definite trade-off between accuracy and comfort, and neither chest-worn nor wrist-worn monitors are the optimal choice for every research question. This makes the accuracy-comfort trade-off a critical consideration when selecting a monitor for research. With continued innovation and refinement, the accuracy gap may narrow between chest- and wrist-worn devices, and trade-off–related considerations may change accordingly. But it is unlikely that the issue will ever disappear completely.

While heart rate monitoring is an obvious application for raw PPG, it may not be the most impactful one. Rather, there may be greater promise when combining raw data from PPG and inertial sensors to predict other physical activity–related outcomes such as energy expenditure. This multimodal approach not only allows for robust correction of motion artifact in the PPG signal (as described previously) but also enables concurrent analysis of movement and cardiovascular data. While this is not an entirely new concept, the ability to carry it out using purely raw data from a single wrist-worn device is quite recent. Thus far, the primary multimodal methods have relied on either separately worn movement and cardiovascular monitors [93-95] or the chest-worn Actiheart device (CamNtech Ltd) [96-99] when predicting energy expenditure. These approaches have shown clear synergy between movement and cardiovascular data but have ultimately had limited uptake compared with the widespread use of wrist-worn monitors in field-based research. Furthermore, heart rate has been the only cardiovascular parameter emphasized with the earlier methods, whereas raw PPG can potentially lead to enhanced predictions through the capture of additional aspects of cardiovascular response to activity (eg, pulse wave parameters and variability). This highlights the warrant for translating and extending earlier concepts of multimodal assessment to the use of raw PPG and inertial data from wrist-worn devices.

The combination of raw PPG and inertial data may also help overcome known limitations of movement-only techniques in the prediction of energy expenditure. For example, wrist-worn monitors are generally unable to register any motion during cycling despite the level of lower-limb exertion, resulting in poor measurement validity [100]. In contrast, PPG may still detect exertion during cycling because it relies on optical and physiological signal rather than inertial signal. This advantage reflects the known benefit of using not only multiple sensors but multiple types of sensors [51,56]. Similar benefits may also arise for other activities where the body's inertial profile is altered, such as when carrying an external load or pushing a stroller [101,102]. The potential to overcome these limitations with virtually no change in participant burden highlights the strong potential of raw PPG to improve physical activity research, especially in the area of energy expenditure prediction.

Combining raw PPG and inertial data may also be beneficial for activity recognition, promoting greater understanding about certain elements of activity context [103]. Activity recognition is also important as a precursor to energy expenditure prediction, since it is much easier to predict energy expenditure if the type of activity is first known [104-106]. This is the basis for several prior models of energy expenditure, including well-known 2-regression models [107-114]. The utility of raw PPG for activity recognition was recently highlighted by Hnoohom et al [115], who calibrated models using data from 3 open-access datasets [116-118]. The parent studies used devices from Shimmer, Empatica, and Maxim Integrated (Analog Devices Inc), and the models were calibrated using deep learning and different combinations of accelerometer, PPG, and electrocardiographic data. When combining accelerometer and

PPG data, the dataset-specific models each achieved near-perfect accuracy during 10-fold cross-validation. However, external validations are needed to confirm the effectiveness of the models and apply them with scalable devices, as described previously.

Outside of energy expenditure and activity recognition, raw PPG may have use for measuring a range of other physical activity–related variables as well, from specific hemodynamic parameters related to exertion (eg, pulse transit time [119]) to consequences of physical activity such as fatigue and recovery [120]. Steps warrant mention as well, given their status as a well-known output of most wearable activity monitors. Raw PPG could potentially enhance or refine the measurement of steps and related variables (eg, cadence), including during periods where the inertial movement profile is altered, as described previously for energy expenditure. The combination of raw data from PPG and inertial sensors could also enable automated measurement of highly specialized outcomes in free-living settings, such as cardiac-locomotor coupling (ie, synchrony of footfalls with systole or diastole) [121].

Finally, although our focus has been on physical activity–related applications of raw PPG, it is important to acknowledge potential contributions in the broader context of 24-hour measurement as well. This refers to the growing emphasis on interrelationships between physical activity, sedentary behavior, and sleep as part of a daily composite [122,123]. The importance of the concept is reflected in the release of 24-hour movement guidelines from numerous governments and the World Health Organization over the last few years [124-129], and there are at least 2 key contributions raw PPG can make to 24-hour assessments. One is to differentiate between nonwear, sedentary behavior, and sleep, all of which produce minimal accelerometer and gyroscope signal and are therefore hard to tell apart using only inertial data. Raw PPG may exhibit richer variation across the categories, thereby assisting with disambiguation. Some proof-of-concept already exists in this area as well, given the amount of prior work using PPG for sleep measurement [130]. The other benefit may be to assist with classifying posture (seated or lying vs upright) [131], which is essential for differentiating sedentary behavior from light-intensity physical activity [132]. These possibilities highlight the strong potential and exceptional flexibility of raw PPG, which will be an asset for a broad range of movement-oriented research in the coming years.

Discussion and Conclusion

In this viewpoint paper, we have introduced raw PPG and highlighted its potential benefits for physical activity assessments. Our specific focus on raw PPG data (as opposed to preprocessed or aggregated data) was critical and timely, given the recent emergence of mainstream devices that provide access to them. A key strength of raw PPG is that its optical basis complements the inertial basis of familiar accelerometer and gyroscope sensors. Furthermore, raw PPG can be used to assess not only heart rate but also broader aspects of cardiovascular physiology. These are the driving forces behind the potential we laid out in the prior sections.

Going forward, it will be critical to obtain raw data from both PPG and inertial sensors, not only to facilitate merging them but also to make new algorithms both transparent and device agnostic (ie, applicable to data from any PPG-inclusive device). These characteristics help combat the “black box” phenomenon of closed-source devices [133]. Device agnosticism also plays an important role in “future proofing” new methods by reducing dependence on individual monitors that can leave the market at any time (eg, as seen with the SenseWear Armband, Phillips Actiwatch, Empatica E4, and ActiGraph GT9X). The use of raw PPG can be further advanced by using open-source channels when developing and disseminating of new resources, consistent with growing standards for existing wearable devices in physical activity assessment [59-64].

This viewpoint paper was among the first to suggest the value of integrating raw PPG data into large-scale assessments of physical activity, where the importance of raw accelerometer and gyroscope data has already been established. In doing so, the viewpoint paper serves to orient new users to the wealth of prior work on PPG from other research areas, where critical reference points have been provided that can spur a paradigm shift in physical activity research. A noteworthy limitation of the viewpoint paper was that it was not a systematic review. As such, it did not fully summarize the available literature, whether in general or focused specifically on physical activity assessment. Nevertheless, our overall conclusion is that there is warrant for vigorous exploration of raw PPG going forward.

Acknowledgments

This study was not supported by a specific award. PRH led the literature review and drafted the manuscript. MK provided support for the review and edited the manuscript.

Conflicts of Interest

PRH has received funding from ActiGraph LLC, who recently released a photoplethysmography-inclusive monitor to the market. ActiGraph had no role in the preparation of this paper.

References

1. Arvidsson D, Fridolfsson J, Börjesson M. Measurement of physical activity in clinical practice using accelerometers. J Intern Med 2019;286(2):137-153 [FREE Full text] [doi: 10.1111/joim.12908] [Medline: 30993807]

2. Troiano RP, McClain JJ, Brychta RJ, Chen KY. Evolution of accelerometer methods for physical activity research. Br J Sports Med 2014;48(13):1019-1023 [FREE Full text] [doi: 10.1136/bjsports-2014-093546] [Medline: 24782483]
3. Crouter SE, DellaValle DM, Haas JD, Frongillo EA, Bassett DR. Validity of ActiGraph 2-regression model, Matthews cut-points, and NHANES cut-points for assessing free-living physical activity. J Phys Act Health 2013;10(4):504-514 [FREE Full text] [doi: 10.1123/jpah.10.4.504] [Medline: 22975460]
4. Bersch SD, Azzi D, Khusainov R, Achumba IE, Ries J. Sensor data acquisition and processing parameters for human activity classification. Sensors (Basel) 2014;14(3):4239-4270 [FREE Full text] [doi: 10.3390/s140304239] [Medline: 24599189]
5. Toth LP, Park S, Springer CM, Feyerabend MD, Steeves JA, Bassett DR. Video-recorded validation of wearable step counters under free-living conditions. Med Sci Sports Exerc 2018;50(6):1315-1322. [doi: 10.1249/MSS.0000000000001569] [Medline: 29381649]
6. O'Driscoll R, Turicchi J, Beaulieu K, Scott S, Matu J, Deighton K, et al. How well do activity monitors estimate energy expenditure? A systematic review and meta-analysis of the validity of current technologies. Br J Sports Med 2020;54(6):332-340 [FREE Full text] [doi: 10.1136/bjsports-2018-099643] [Medline: 30194221]
7. Liu S, Gao RX, John D, Staudenmayer JW, Freedson PS. Multisensor data fusion for physical activity assessment. IEEE Trans Biomed Eng 2012;59(3):687-696. [doi: 10.1109/TBME.2011.2178070] [Medline: 22156943]
8. Intille SS, Lester J, Sallis JF, Duncan G. New horizons in sensor development. Med Sci Sports Exerc 2012;44(1 Suppl 1):S24-S31 [FREE Full text] [doi: 10.1249/MSS.0b013e3182399c7d] [Medline: 22157771]
9. Castaneda D, Esparza A, Ghamari M, Soltanpur C, Nazeran H. A review on wearable photoplethysmography sensors and their potential future applications in health care. Int J Biosens Bioelectron 2018;4(4):195-202 [FREE Full text] [doi: 10.15406/ijbsbe.2018.04.00125] [Medline: 30906922]
10. Allen J. Photoplethysmography and its application in clinical physiological measurement. Physiol Meas 2007;28(3):R1-R39. [doi: 10.1088/0967-3334/28/3/R01] [Medline: 17322588]
11. Hertzman A, Spielman C. Observations on the finger volume pulse recorded photo-electrically. Am J Physiol 1937;119:334-335.
12. Alian AA, Shelley KH. Photoplethysmography. Best Pract Res Clin Anaesthesiol 2014;28(4):395-406. [doi: 10.1016/j.bpa.2014.08.006] [Medline: 25480769]
13. Shcherbina A, Mattsson CM, Waggott D, Salisbury H, Christle JW, Hastie T, et al. Accuracy in wrist-worn, sensor-based measurements of heart rate and energy expenditure in a diverse cohort. J Pers Med 2017;7(2):3 [FREE Full text] [doi: 10.3390/jpm7020003] [Medline: 28538708]
14. Bai Y, Hibbing P, Mantis C, Welk GJ. Comparative evaluation of heart rate-based monitors: apple watch vs fitbit charge HR. J Sports Sci 2018;36(15):1734-1741. [doi: 10.1080/02640414.2017.1412235] [Medline: 29210326]
15. Boudreaux BD, Hebert EP, Hollander DB, Williams BM, Cormier CL, Naquin MR, et al. Validity of wearable activity monitors during cycling and resistance exercise. Med Sci Sports Exerc 2018;50(3):624-633. [doi: 10.1249/MSS.0000000000001471] [Medline: 29189666]
16. Benedetto S, Caldato C, Bazzan E, Greenwood DC, Pensabene V, Actis P. Assessment of the fitbit charge 2 for monitoring heart rate. PLoS One 2018;13(2):e0192691 [FREE Full text] [doi: 10.1371/journal.pone.0192691] [Medline: 29489850]
17. Cui WJ, Ostrander LE, Lee BY. In vivo reflectance of blood and tissue as a function of light wavelength. IEEE Trans Biomed Eng 1990;37(6):632-639. [doi: 10.1109/10.55667] [Medline: 2354845]
18. Reddy KA, George B, Kumar JV. Use of fourier series analysis for motion artifact reduction and data compression of photoplethysmographic signals. IEEE Trans Instrum Meas 2009;58(5):1706-1711. [doi: 10.1109/TIM.2008.2009136]
19. Lee J, Matsumura K, Yamakoshi KI, Rolfe P, Tanaka S, Yamakoshi T. Comparison between red, green and blue light reflection photoplethysmography for heart rate monitoring during motion. Annu Int Conf IEEE Eng Med Biol Soc 2013;2013:1724-1727. [doi: 10.1109/EMBC.2013.6609852] [Medline: 24110039]
20. Preejith SP, Alex A, Joseph J, Sivaprakasam M. Design, development and clinical validation of a wrist-based optical heart rate monitor. 2016 Presented at: IEEE International Symposium on Medical Measurements and Applications (MeMeA); 2016 May 15-18; Benevento, Italy.
21. Wolling F, Van Laerhoven K. The quest for raw signals: a quality review of publicly available photoplethysmography datasets. 2020 Presented at: Proceedings of the Third Workshop on Data: Acquisition To Analysis; 2020 November 16-19; Virtual Event, Japan p. 14-19.
22. Welk GJ, Bai Y, Lee JM, Godino J, Saint-Maurice PF, Carr L. Standardizing analytic methods and reporting in activity monitor validation studies. Med Sci Sports Exerc 2019;51(8):1767-1780 [FREE Full text] [doi: 10.1249/MSS.0000000000001966] [Medline: 30913159]
23. Kamal AA, Harness JB, Irving G, Mearns AJ. Skin photoplethysmography—a review. Comput Methods Programs Biomed 1989;28(4):257-269. [doi: 10.1016/0169-2607(89)90159-4] [Medline: 2649304]
24. Webster JG. Design of Pulse Oximeters. Boca Raton, FL: CRC Press; 1997.
25. Shelley KH. Photoplethysmography: beyond the calculation of arterial oxygen saturation and heart rate. Anesth Analg 2007;105(6 Suppl):S31-S36. [doi: 10.1213/01.ane.0000269512.82836.c9] [Medline: 18048895]

26. Mannheimer PD. The light-tissue interaction of pulse oximetry. Anesth Analg 2007;105(6 Suppl):S10-S17. [doi: 10.1213/01.ane.0000269522.84942.54] [Medline: 18048891]
27. Yan J, Ye Z, Shi F, Dai Y, Yang L, Wu J, et al. Reflection-type photoplethysmography pulse sensor based on an integrated optoelectronic chip with a ring structure. Biomed Opt Express 2021;12(10):6277-6283 [FREE Full text] [doi: 10.1364/BOE.437805] [Medline: 34745736]
28. de Trafford J, Lafferty K. What does photoplethysmography measure? Med Biol Eng Comput 1984;22(5):479-480. [doi: 10.1007/BF02447713] [Medline: 6482540]
29. Lindberg LG, Oberg PA. Optical properties of blood in motion. Opt Eng 1993;32(2):253. [doi: 10.1117/12.60688]
30. Shvartsman LD, Fine I. Optical transmission of blood: effect of erythrocyte aggregation. IEEE Trans Biomed Eng 2003;50(8):1026-1033. [doi: 10.1109/TBME.2003.814532] [Medline: 12892330]
31. Lambert J. Photometria Sive de Mensura et Gradibus Luminis, Colorum et Umbrae. Klett, Norway: Nabu Press; 1760:586.
32. Bergstrand S, Lindberg LG, Ek AC, Lindén M, Lindgren M. Blood flow measurements at different depths using photoplethysmography and laser Doppler techniques. Skin Res Technol 2009;15(2):139-147. [doi: 10.1111/j.1600-0846.2008.00337.x] [Medline: 19622122]
33. Spigulis J, Gailite L, Lihachev A, Erts R. Simultaneous recording of skin blood pulsations at different vascular depths by multiwavelength photoplethysmography. Appl Opt 2007;46(10):1754-1759. [doi: 10.1364/ao.46.001754] [Medline: 17356618]
34. Maeda Y, Sekine M, Tamura T. Relationship between measurement site and motion artifacts in wearable reflected photoplethysmography. J Med Syst 2011;35(5):969-976. [doi: 10.1007/s10916-010-9505-0] [Medline: 20703691]
35. Matsumura K, Rolfe P, Lee J, Yamakoshi T. iPhone 4s photoplethysmography: which light color yields the most accurate heart rate and normalized pulse volume using the iPhysioMeter application in the presence of motion artifact? PLoS One 2014;9(3):e91205 [FREE Full text] [doi: 10.1371/journal.pone.0091205] [Medline: 24618594]
36. Fine J, Branan KL, Rodriguez AJ, Boonya-Ananta T, Ajmal, Ramella-Roman JC, et al. Sources of inaccuracy in photoplethysmography for continuous cardiovascular monitoring. Biosensors (Basel) 2021;11(4):126 [FREE Full text] [doi: 10.3390/bios11040126] [Medline: 33923469]
37. Bent B, Goldstein BA, Kibbe WA, Dunn JP. Investigating sources of inaccuracy in wearable optical heart rate sensors. NPJ Digit Med 2020;3:18 [FREE Full text] [doi: 10.1038/s41746-020-0226-6] [Medline: 32047863]
38. Colvonen PJ. Response to: investigating sources of inaccuracy in wearable optical heart rate sensors. NPJ Digit Med 2021;4(1):38 [FREE Full text] [doi: 10.1038/s41746-021-00408-5] [Medline: 33637822]
39. Nowara EM, McDuff D, Veeraraghavan A. A meta-analysis of the impact of skin type and gender on non-contact photoplethysmography measurements. 2020 Presented at: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW); 2020 June 14-19; Seattle, WA, USA.
40. Puranen A, Halkola T, Kirkeby O, Vehkaoja A. Effect of skin tone and activity on the performance of wrist-worn optical beat-to-beat heart rate monitoring. 2020 Presented at: 2020 IEEE SENSORS; 2020 October 25-28; Rotterdam, Netherlands.
41. Lister T, Wright PA, Chappell PH. Optical properties of human skin. J Biomed Opt 2012;17(9):90901 [FREE Full text] [doi: 10.1117/1.JBO.17.9.090901] [Medline: 23085902]
42. Anderson RR, Parrish JA. The optics of human skin. J Invest Dermatol 1981;77(1):13-19 [FREE Full text] [doi: 10.1111/1523-1747.ep12479191] [Medline: 7252245]
43. Hayes MJ, Smith PR. Artifact reduction in photoplethysmography. Appl Opt 1998;37(31):7437-7446. [doi: 10.1364/ao.37.007437] [Medline: 18301578]
44. Han H, Kim J. Artifacts in wearable photoplethysmographs during daily life motions and their reduction with least mean square based active noise cancellation method. Comput Biol Med 2012;42(4):387-393. [doi: 10.1016/j.compbiomed.2011.12.005] [Medline: 22206810]
45. Ismail S, Akram U, Siddiqi I. Heart rate tracking in photoplethysmography signals affected by motion artifacts: a review. EURASIP J Adv Signal Process 2021;2021(1):5. [doi: 10.1186/s13634-020-00714-2]
46. Bhattacharya A, McCutcheon EP, Shvartz E, Greenleaf JE. Body acceleration distribution and O2 uptake in humans during running and jumping. J Appl Physiol Respir Environ Exerc Physiol 1980;49(5):881-887. [doi: 10.1152/jappl.1980.49.5.881] [Medline: 7429911]
47. Antonsson EK, Mann RW. The frequency content of gait. J Biomech 1985;18(1):39-47. [doi: 10.1016/0021-9290(85)90043-0] [Medline: 3980487]
48. Sun M, Hill JO. A method for measuring mechanical work and work efficiency during human activities. J Biomech 1993;26(3):229-241. [doi: 10.1016/0021-9290(93)90361-h] [Medline: 8468336]
49. Parak J, Korhonen I. Evaluation of wearable consumer heart rate monitors based on photopletysmography. Annu Int Conf IEEE Eng Med Biol Soc 2014;2014:3670-3673. [doi: 10.1109/EMBC.2014.6944419] [Medline: 25570787]
50. Jo E, Lewis K, Directo D, Kim MJ, Dolezal BA. Validation of biofeedback wearables for photoplethysmographic heart rate tracking. J Sports Sci Med 2016;15(3):540-547 [FREE Full text] [Medline: 27803634]
51. Chen KY, Janz KF, Zhu W, Brychta RJ. Redefining the roles of sensors in objective physical activity monitoring. Med Sci Sports Exerc 2012;44(1 Suppl 1):S13-S23 [FREE Full text] [doi: 10.1249/MSS.0b013e3182399bc8] [Medline: 22157770]

52. Esliger DW, Copeland JL, Barnes JD, Tremblay MS. Standardizing and optimizing the use of accelerometer data for free-living physical activity monitoring. J Phys Act Health 2005;3(3):366-383.
53. Rowlands AV, Stiles VH. Accelerometer counts and raw acceleration output in relation to mechanical loading. J Biomech 2012;45(3):448-454 [FREE Full text] [doi: 10.1016/j.jbiomech.2011.12.006] [Medline: 22218284]
54. van Hees VT, Gorzelniak L, Dean León EC, Eder M, Pias M, Taherian S, et al. Separating movement and gravity components in an acceleration signal and implications for the assessment of human daily physical activity. PLoS One 2013;8(4):e61691 [FREE Full text] [doi: 10.1371/journal.pone.0061691] [Medline: 23626718]
55. Hildebrand M, VAN Hees VT, Hansen BH, Ekelund U. Age group comparability of raw accelerometer output from wrist- and hip-worn monitors. Med Sci Sports Exerc 2014;46(9):1816-1824. [doi: 10.1249/MSS.0000000000000289] [Medline: 24887173]
56. Chen KY, Bassett DR. The technology of accelerometry-based activity monitors: current and future. Med Sci Sports Exerc 2005;37(11 Suppl):S490-S500. [doi: 10.1249/01.mss.0000185571.49104.82] [Medline: 16294112]
57. Trost SG, McIver KL, Pate RR. Conducting accelerometer-based activity assessments in field-based research. Med Sci Sports Exerc 2005;37(11 Suppl):S531-S543. [doi: 10.1249/01.mss.0000185657.86065.98] [Medline: 16294116]
58. Welk GJ. Principles of design and analyses for the calibration of accelerometry-based activity monitors. Med Sci Sports Exerc 2005;37(11 Suppl):S501-S511. [doi: 10.1249/01.mss.0000185660.38335.de] [Medline: 16294113]
59. Procter DS, Page AS, Cooper AR, Nightingale CM, Ram B, Rudnicka AR, et al. An open-source tool to identify active travel from hip-worn accelerometer, GPS and GIS data. Int J Behav Nutr Phys Act 2018;15(1):91 [FREE Full text] [doi: 10.1186/s12966-018-0724-y] [Medline: 30241483]
60. John D, Tang Q, Albinali F, Intille S. An open-source monitor-independent movement summary for accelerometer data processing. J Meas Phys Behav 2019;2(4):268-281 [FREE Full text] [doi: 10.1123/jmpb.2018-0068] [Medline: 34308270]
61. Carlson JA, Ridgers ND, Nakandala S, Zablocki R, Tuz-Zahra F, Bellettiere J, et al. CHAP-child: an open source method for estimating sit-to-stand transitions and sedentary bout patterns from hip accelerometers among children. Int J Behav Nutr Phys Act 2022;19(1):109 [FREE Full text] [doi: 10.1186/s12966-022-01349-2] [Medline: 36028890]
62. Migueles JH, Rowlands A, Huber F, Sabia S, van Hees VT. GGIR: a research community? Driven open source R package for generating physical activity and sleep outcomes from multi-day raw accelerometer data. J Meas Phys Behav 2019;2(3):188-196.
63. de Looff P, Duursma R, Noordzij M, Taylor S, Jaques N, Scheepers F, et al. Wearables: an R package with accompanying shiny application for signal analysis of a wearable device targeted at clinicians and researchers. Front Behav Neurosci 2022;16:856544 [FREE Full text] [doi: 10.3389/fnbeh.2022.856544] [Medline: 35813597]
64. Helsel BC, Hibbing PR, Montgomery RN, Vidoni ED, Ptomey LT, Clutton J, et al. agcounts: an R package to calculate ActiGraph activity counts from portable accelerometers. J Meas Phys Behav 2024;7(1):1-7.
65. Zhang Y, Weaver RG, Armstrong B, Burkart S, Zhang S, Beets MW. Validity of wrist-worn photoplethysmography devices to measure heart rate: a systematic review and meta-analysis. J Sports Sci 2020;38(17):2021-2034. [doi: 10.1080/02640414.2020.1767348] [Medline: 32552580]
66. Nachman D, Gepner Y, Goldstein N, Kabakov E, Ishay AB, Littman R, et al. Comparing blood pressure measurements between a photoplethysmography-based and a standard cuff-based manometry device. Sci Rep 2020;10(1):16116 [FREE Full text] [doi: 10.1038/s41598-020-73172-3] [Medline: 32999400]
67. Nachman D, Gilan A, Goldstein N, Constantini K, Littman R, Eisenkraft A, et al. Twenty-four-hour ambulatory blood pressure measurement using a novel noninvasive, cuffless, wireless device. Am J Hypertens 2021;34(11):1171-1180. [doi: 10.1093/ajh/hpab095] [Medline: 34143867]
68. Nachman D, Eisenkraft A, Goldstein N, Ben-Ishay A, Fons M, Merin R, et al. Influence of sex, BMI, and skin color on the accuracy of non-invasive cuffless photoplethysmography-based blood pressure measurements. Front Physiol 2022;13:911544 [FREE Full text] [doi: 10.3389/fphys.2022.911544] [Medline: 35846008]
69. Solà J, Vybornova A, Fallet S, Grossenbacher O, De Marco B, Olivero E, et al. Aktiia bracelet: monitoring of blood pressure using off-the-shelf optical sensors. 2019 Presented at: 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC; 2019 July 24; Berlin, Germany URL: https://embc.embs.org/2019/wp-content/uploads/sites/42/2019/07/EMBC-2019-Program.pdf
70. Vybornova A, Polychronopoulou E, Wurzner-Ghajarzadeh A, Fallet S, Sola J, Wuerzner G. Blood pressure from the optical Aktiia bracelet: a 1-month validation study using an extended ISO81060-2 protocol adapted for a cuffless wrist device. Blood Press Monit 2021;26(4):305-311 [FREE Full text] [doi: 10.1097/MBP.0000000000000531] [Medline: 33675592]
71. Sola J, Vybornova A, Fallet S, Polychronopoulou E, Wurzner-Ghajarzadeh A, Wuerzner G. Validation of the optical Aktiia bracelet in different body positions for the persistent monitoring of blood pressure. Sci Rep 2021;11(1):20644 [FREE Full text] [doi: 10.1038/s41598-021-99294-w] [Medline: 34667230]
72. Seshadri DR, Bittel B, Browsky D, Houghtaling P, Drummond CK, Desai MY, et al. Accuracy of Apple Watch for detection of atrial fibrillation. Circulation 2020;141(8):702-703. [doi: 10.1161/CIRCULATIONAHA.119.044126] [Medline: 32091929]
73. Walch O, Huang Y, Forger D, Goldstein C. Sleep stage prediction with raw acceleration and photoplethysmography heart rate data derived from a consumer wearable device. Sleep 2019;42(12):zsz180 [FREE Full text] [doi: 10.1093/sleep/zsz180] [Medline: 31579900]

74. Gayathri Priyadarshini R, Kalimuthu M, Nikesh S, Bhuvaneshwari M. Review of PPG signal using machine learning algorithms for blood pressure and glucose estimation. IOP Conf Ser Mater Sci Eng 2021;1084(1):012031.
75. Chu J, Yang WT, Lu WR, Chang YT, Hsieh TH, Yang FL. 90% Accuracy for photoplethysmography-based non-invasive blood glucose prediction by deep learning with cohort arrangement and quarterly measured HbA1c. Sensors (Basel) 2021;21(23):7815 [FREE Full text] [doi: 10.3390/s21237815] [Medline: 34883817]
76. Althobaiti MM, Al-Naib I. Optimization of dual-channel near-infrared non-invasive glucose level measurement sensors based on monte-carlo simulations. IEEE Photonics J 2021;13(3):1. [doi: 10.1109/jphot.2021.3079408]
77. Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. Int J Behav Nutr Phys Act 2015;12:159 [FREE Full text] [doi: 10.1186/s12966-015-0314-1] [Medline: 26684758]
78. Wright SP, Hall Brown TS, Collier SR, Sandberg K. How consumer physical activity monitors could transform human physiology research. Am J Physiol Regul Integr Comp Physiol 2017;312(3):R358-R367 [FREE Full text] [doi: 10.1152/ajpregu.00349.2016] [Medline: 28052867]
79. Collins T, Woolley SI, Oniani S, Pires IM, Garcia NM, Ledger SJ, et al. Version reporting and assessment approaches for new and updated activity and heart rate monitors. Sensors (Basel) 2019;19(7):1705 [FREE Full text] [doi: 10.3390/s19071705] [Medline: 30974755]
80. Leung W, Case L, Jung J, Yun J. Factors associated with validity of consumer-oriented wearable physical activity trackers: a meta-analysis. J Med Eng Technol 2021;45(3):223-236. [doi: 10.1080/03091902.2021.1893395] [Medline: 33750250]
81. LaMunion SR, Blythe AL, Hibbing PR, Kaplan AS, Clendenin BJ, Crouter SE. Use of consumer monitors for estimating energy expenditure in youth. Appl Physiol Nutr Metab 2020;45(2):161-168 [FREE Full text] [doi: 10.1139/apnm-2019-0129] [Medline: 31269409]
82. Liang Y, Chen Z, Liu G, Elgendi M. A new, short-recorded photoplethysmogram dataset for blood pressure monitoring in China. Sci Data 2018;5:180020 [FREE Full text] [doi: 10.1038/sdata.2018.20] [Medline: 29485624]
83. Liang Y, Chen Z, Ward R, Elgendi M. Photoplethysmography and deep learning: enhancing hypertension risk stratification. Biosensors (Basel) 2018;8(4):101 [FREE Full text] [doi: 10.3390/bios8040101] [Medline: 30373211]
84. Saeed M, Villarroel M, Reisner AT, Clifford G, Lehman LW, Moody G, et al. Multiparameter intelligent monitoring in intensive care II: a public-access intensive care unit database. Crit Care Med 2011;39(5):952-960 [FREE Full text] [doi: 10.1097/CCM.0b013e31820a92c6] [Medline: 21283005]
85. Johnson AEW, Pollard TJ, Shen L, Lehman LH, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. Sci Data 2016;3:160035 [FREE Full text] [doi: 10.1038/sdata.2016.35] [Medline: 27219127]
86. Sun X, Zhou L, Chang S, Liu Z. Using CNN and HHT to predict blood pressure level based on photoplethysmography and its derivatives. Biosensors (Basel) 2021;11(4):120 [FREE Full text] [doi: 10.3390/bios11040120] [Medline: 33924324]
87. Plews DJ, Scott B, Altini M, Wood M, Kilding AE, Laursen PB. Comparison of heart-rate-variability recording with smartphone photoplethysmography, Polar H7 chest strap, and electrocardiography. Int J Sports Physiol Perform 2017;12(10):1324-1328. [doi: 10.1123/ijspp.2016-0668] [Medline: 28290720]
88. De Ridder B, Van Rompaey B, Kampen JK, Haine S, Dilles T. Smartphone apps using photoplethysmography for heart rate monitoring: meta-analysis. JMIR Cardio 2018;2(1):e4 [FREE Full text] [doi: 10.2196/cardio.8802] [Medline: 31758768]
89. Dey J, Gaurav A, Tiwari V. InstaBP: cuff-less blood pressure monitoring on smartphone using single PPG sensor. Annu Int Conf IEEE Eng Med Biol Soc 2018;2018:5002-5005. [doi: 10.1109/EMBC.2018.8513189] [Medline: 30441464]
90. Gillinov S, Etiwy M, Wang R, Blackburn G, Phelan D, Gillinov A, et al. Variable accuracy of wearable heart rate monitors during aerobic exercise. Med Sci Sports Exerc 2017;49(8):1697-1703. [doi: 10.1249/MSS.0000000000001284] [Medline: 28709155]
91. Laukkanen RM, Virtanen PK. Heart rate monitors: state of the art. J Sports Sci 1998;16 Suppl:S3-S7. [doi: 10.1080/026404198366920] [Medline: 22587712]
92. Pasadyn SR, Soudan M, Gillinov M, Houghtaling P, Phelan D, Gillinov N, et al. Accuracy of commercially available heart rate monitors in athletes: a prospective study. Cardiovasc Diagn Ther 2019;9(4):379-385 [FREE Full text] [doi: 10.21037/cdt.2019.06.05] [Medline: 31555543]
93. Strath SJ, Bassett DR, Swartz AM, Thompson DL. Simultaneous heart rate-motion sensor technique to estimate energy expenditure. Med Sci Sports Exerc 2001;33(12):2118-2123. [doi: 10.1097/00005768-200112000-00022] [Medline: 11740308]
94. Strath SJ, Bassett DR, Thompson DL, Swartz AM. Validity of the simultaneous heart rate-motion sensor technique for measuring energy expenditure. Med Sci Sports Exerc 2002;34(5):888-894. [doi: 10.1097/00005768-200205000-00025] [Medline: 11984311]
95. Strath SJ, Brage S, Ekelund U. Integration of physiological and accelerometer data to improve physical activity assessment. Med Sci Sports Exerc 2005;37(11 Suppl):S563-S571. [doi: 10.1249/01.mss.0000185650.68232.3f] [Medline: 16294119]
96. Brage S, Brage N, Franks PW, Ekelund U, Wareham NJ. Reliability and validity of the combined heart rate and movement sensor actiheart. Eur J Clin Nutr 2005;59(4):561-570. [doi: 10.1038/sj.ejcn.1602118] [Medline: 15714212]
97. Crouter SE, Churilla JR, Bassett DR. Accuracy of the actiheart for the assessment of energy expenditure in adults. Eur J Clin Nutr 2008;62(6):704-711. [doi: 10.1038/sj.ejcn.1602766] [Medline: 17440515]

98. Villars C, Bergouignan A, Dugas J, Antoun E, Schoeller DA, Roth H, et al. Validity of combining heart rate and uniaxial acceleration to measure free-living physical activity energy expenditure in young men. J Appl Physiol (1985) 2012;113(11):1763-1771 [FREE Full text] [doi: 10.1152/japplphysiol.01413.2011] [Medline: 23019315]
99. McMinn D, Acharya R, Rowe DA, Gray SR, Allan JL. Measuring activity energy expenditure: accuracy of the GT3X+ and Actiheart monitors. Int J Exerc Sci 2013;6(3):29.
100. Swartz AM, Strath SJ, Bassett DR, O'Brien WL, King GA, Ainsworth BE. Estimation of energy expenditure using CSA accelerometers at hip and wrist sites. Med Sci Sports Exerc 2000;32(9 Suppl):S450-SS46. [doi: 10.1097/00005768-200009001-00003] [Medline: 10993414]
101. Chen M, Kuo C, Pellegrini C, Hsu M. Accuracy of wristband activity monitors during ambulation and activities. Med Sci Sports Exerc 2016;48(10):1942-1949. [doi: 10.1249/MSS.0000000000000984] [Medline: 27183123]
102. Donisi L, Cesarelli G, Pisani N, Ponsiglione AM, Ricciardi C, Capodaglio E. Wearable sensors and artificial intelligence for physical ergonomics: a systematic review of literature. Diagnostics (Basel) 2022;12(12):3048 [FREE Full text] [doi: 10.3390/diagnostics12123048] [Medline: 36553054]
103. Kern N, Schiele B, Schmidt A. Multi-sensor activity context detection for wearable computing. In: Aarts E, Collier RW, van Loenen E, de Ruyter B, editors. Ambient Intelligence. EUSAI 2003. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer; 2003.
104. Bonomi AG, Plasqui G, Goris AHC, Westerterp KR. Improving assessment of daily energy expenditure by identifying types of physical activity with a single accelerometer. J Appl Physiol (1985) 2009;107(3):655-661 [FREE Full text] [doi: 10.1152/japplphysiol.00150.2009] [Medline: 19556460]
105. Bonomi AG, Plasqui G. "Divide and conquer": assessing energy expenditure following physical activity type classification. J Appl Physiol (1985) 2012;112(5):932-933 [FREE Full text] [doi: 10.1152/japplphysiol.01403.2011] [Medline: 22383498]
106. Plasqui G. Smart approaches for assessing free-living energy expenditure following identification of types of physical activity. Obes Rev 2017;18 Suppl 1:50-55. [doi: 10.1111/obr.12506] [Medline: 28164455]
107. Heil DP. Predicting activity energy expenditure using the actical activity monitor. Res Q Exerc Sport 2006;77(1):64-80. [doi: 10.1080/02701367.2006.10599333] [Medline: 16646354]
108. Crouter SE, Clowers KG, Bassett DR. A novel method for using accelerometer data to predict energy expenditure. J Appl Physiol (1985) 2006;100(4):1324-1331 [FREE Full text] [doi: 10.1152/japplphysiol.00818.2005] [Medline: 16322367]
109. Crouter SE, Bassett DR. A new 2-regression model for the actical accelerometer. Br J Sports Med 2008;42(3):217-224. [doi: 10.1136/bjsm.2006.033399] [Medline: 17761786]
110. Crouter SE, Kuffel E, Haas JD, Frongillo EA, Bassett DR. Refined two-regression model for the ActiGraph accelerometer. Med Sci Sports Exerc 2010;42(5):1029-1037 [FREE Full text] [doi: 10.1249/MSS.0b013e3181c37458] [Medline: 20400882]
111. Rothney MP, Brychta RJ, Meade NN, Chen KY, Buchowski MS. Validation of the ActiGraph two-regression model for predicting energy expenditure. Med Sci Sports Exerc 2010;42(9):1785-1792 [FREE Full text] [doi: 10.1249/MSS.0b013e3181d5a984] [Medline: 20142778]
112. Crouter SE, Horton M, Bassett DR. Use of a two-regression model for estimating energy expenditure in children. Med Sci Sports Exerc 2012;44(6):1177-1185 [FREE Full text] [doi: 10.1249/MSS.0b013e3182447825] [Medline: 22143114]
113. Hibbing PR, Lamunion SR, Kaplan AS, Crouter SE. Estimating energy expenditure with ActiGraph GT9X inertial measurement unit. Med Sci Sports Exerc 2018;50(5):1093-1102. [doi: 10.1249/MSS.0000000000001532] [Medline: 29271847]
114. Crouter SE, Oody JF, Bassett DR. Estimating physical activity in youth using an ankle accelerometer. J Sports Sci 2018;36(19):2265-2271 [FREE Full text] [doi: 10.1080/02640414.2018.1449091] [Medline: 29517959]
115. Hnoohom N, Mekruksavanich S, Jitpattanakul A. Physical activity recognition based on deep learning using photoplethysmography and wearable inertial sensors. Electronics 2023;12(3):693. [doi: 10.3390/electronics12030693]
116. Jarchi D, Casson A. Description of a database containing wrist PPG signals recorded during physical exercise with both accelerometer and gyroscope measures of motion. Data 2016;2(1):1 [FREE Full text] [doi: 10.3390/data2010001]
117. Reiss A, Indlekofer I, Schmidt P, Van Laerhoven K. Deep PPG: large-scale heart rate estimation with convolutional neural networks. Sensors (Basel) 2019;19(14):3079 [FREE Full text] [doi: 10.3390/s19143079] [Medline: 31336894]
118. Biagetti G, Crippa P, Falaschetti L, Saraceni L, Tiranti A, Turchetti C. Dataset from PPG wireless sensor for activity monitoring. Data Brief 2020;29:105044 [FREE Full text] [doi: 10.1016/j.dib.2019.105044] [Medline: 31989005]
119. Jang DG, Park SH, Hahn M. A Gaussian model-based probabilistic approach for pulse transit time estimation. IEEE J Biomed Health Inform 2016;20(1):128-134. [doi: 10.1109/JBHI.2014.2372047] [Medline: 25420274]
120. Adão Martins NR, Annaheim S, Spengler CM, Rossi RM. Fatigue monitoring through wearables: a state-of-the-art review. Front Physiol 2021;12:790292 [FREE Full text] [doi: 10.3389/fphys.2021.790292] [Medline: 34975541]
121. Constantini K, Stickford ASL, Bleich JL, Mannheimer PD, Levine BD, Chapman RF. Synchronizing gait with cardiac cycle phase alters heart rate response during running. Med Sci Sports Exerc 2018;50(5):1046-1053 [FREE Full text] [doi: 10.1249/MSS.0000000000001515] [Medline: 29240004]
122. Rosenberger ME, Fulton JE, Buman MP, Troiano RP, Grandner MA, Buchner DM, et al. The 24-hour activity cycle: a new paradigm for physical activity. Med Sci Sports Exerc 2019;51(3):454-464 [FREE Full text] [doi: 10.1249/MSS.0000000000001811] [Medline: 30339658]

123. Vandelanotte C. A journal dedicated to studying the combined effects of activity, sedentary and sleep behaviours. J Act Sedentary Sleep Behav 2022;1(1):1. [doi: 10.1186/s44167-022-00008-y]
124. Tremblay MS, Carson V, Chaput JP, Connor Gorber S, Dinh T, Duggan M, et al. Canadian 24-hour movement guidelines for children and youth: an integration of physical activity, sedentary behaviour, and sleep. Appl Physiol Nutr Metab 2016;41(6 Suppl 3):S311-S327 [FREE Full text] [doi: 10.1139/apnm-2016-0151] [Medline: 27306437]
125. Okely AD, Ghersi D, Hesketh KD, Santos R, Loughran SP, Cliff DP, et al. A collaborative approach to adopting/adapting guidelines - the Australian 24-hour movement guidelines for the early years (birth to 5 years): an integration of physical activity, sedentary behavior, and sleep. BMC Public Health 2017;17(Suppl 5):869 [FREE Full text] [doi: 10.1186/s12889-017-4867-6] [Medline: 29219094]
126. World Health Organization. Guidelines on physical activity, sedentary behaviour and sleep for children under 5 years of age. Geneva, Switzerland: World Health Organization; 2019.
127. Tremblay MS, Rollo S, Saunders TJ. Sedentary behavior research network members support new Canadian 24-hour movement guideline recommendations. J Sport Health Sci 2020;9(6):479-481 [FREE Full text] [doi: 10.1016/j.jshs.2020.09.012] [Medline: 33071162]
128. Ross R, Chaput JP, Giangregorio LM, Janssen I, Saunders TJ, Kho ME, et al. Canadian 24-Hour movement guidelines for adults aged 18-64 years and adults aged 65 years or older: an integration of physical activity, sedentary behaviour, and sleep. Appl Physiol Nutr Metab 2020;45(10 (Suppl 2):S57-S102 [FREE Full text] [doi: 10.1139/apnm-2020-0467] [Medline: 33054332]
129. Draper CE, Tomaz SA, Biersteker L, Cook CJ, Couper J, de Milander M, et al. The South African 24-hour movement guidelines for birth to 5 years: an integration of physical activity, sitting behavior, screen time, and sleep. J Phys Act Health 2020;17(1):109-119 [FREE Full text] [doi: 10.1123/jpah.2019-0187] [Medline: 31877557]
130. Ryals S, Chiang A, Schutte-Rodin S, Chandrakantan A, Verma N, Holfinger S, et al. Photoplethysmography-new applications for an old technology: a sleep technology review. J Clin Sleep Med 2023;19(1):189-195 [FREE Full text] [doi: 10.5664/jcsm.10300] [Medline: 36123954]
131. Linder SP, Wendelken SM, Wei E, McGrath SP. Using the morphology of photoplethysmogram peaks to detect changes in posture. J Clin Monit Comput 2006;20(3):151-158. [doi: 10.1007/s10877-006-9015-2] [Medline: 16688391]
132. Tremblay MS, Aubert S, Barnes JD, Saunders TJ, Carson V, Latimer-Cheung AE, SBRN Terminology Consensus Project Participants. Sedentary Behavior Research Network (SBRN)—terminology consensus project process and outcome. Int J Behav Nutr Phys Act 2017;14(1):75 [FREE Full text] [doi: 10.1186/s12966-017-0525-8] [Medline: 28599680]
133. Albrecht BM, Flaßkamp FT, Koster A, Eskofier BM, Bammann K. Cross-sectional survey on researchers' experience in using accelerometers in health-related studies. BMJ Open Sport Exerc Med 2022;8(2):e001286 [FREE Full text] [doi: 10.1136/bmjsem-2021-001286] [Medline: 35601138]

Abbreviations

PPG: photoplethysmography

Edited by L Buis; submitted 06.02.24; peer-reviewed by MS Arefin, Q Dong, ABMM Rahman; comments to author 24.05.24; revised version received 09.07.24; accepted 26.08.24; published 27.09.24.

Please cite as:
Hibbing PR, Khan MM
Raw Photoplethysmography as an Enhancement for Research-Grade Wearable Activity Monitors
JMIR Mhealth Uhealth 2024;12:e57158
URL: https://mhealth.jmir.org/2024/1/e57158
doi:10.2196/57158
PMID:39331461

Viewpoint

Digital Health Readiness: Making Digital Health Care More Inclusive

Timothy Bober[1,2,3,4], MS, MD; Bruce L Rollman[1,2,3], MPH, MD; Steven Handler[5,6,7], MD, PhD; Andrew Watson[8,9,10], MLitt, MD; Lyndsay A Nelson[11], PhD; Julie Faieta[6], MOT, OTL/L, PhD; Ann-Marie Rosland[1,2,4], MS, MD

[1]Division of General Internal Medicine, University of Pittsburgh School of Medicine, Pittsburgh, PA, United States

[2]Center for Research on Health Care, University of Pittsburgh, Pittsburgh, PA, United States

[3]Center for Behavioral Health, Media, and Technology, University of Pittsburgh Department of Medicine, Pittsburgh, PA, United States

[4]Center for Health Equity Research and Promotion, VA Pittsburgh Healthcare System, Pittsburgh, PA, United States

[5]Division of Geriatric Medicine, University of Pittsburgh, Pittsburgh, PA, United States

[6]Department of Rehab Science and Technology, University of Pittsburgh, Pittsburgh, PA, United States

[7]Technology Enhancing Cognition and Health – Geriatrics Research Education and Clinical Center, VA Pittsburgh Healthcare System, Pittsburgh, PA, United States

[8]Department of Surgery, University of Pittsburgh, Pittsburgh, PA, United States

[9]UPMC Health Plan, Pittsburgh, PA, United States

[10]UPMC Enterprises, Pittsburgh, PA, United States

[11]Division of General Internal Medicine and Public Health, Department of Medicine, Vanderbilt University Medical Center, Nashvillie, TN, United States

Corresponding Author:
Timothy Bober, MS, MD
Division of General Internal Medicine
University of Pittsburgh School of Medicine
UPMC Montefiore Hospital
Suite W933
Pittsburgh, PA, 15213
United States
Phone: 1 412 692 4821
Email: bobertm@upmc.edu

Abstract

This paper proposes an approach to assess digital health readiness in clinical settings to understand how prepared, experienced, and equipped individual people are to participate in digital health activities. Existing digital health literacy and telehealth prediction tools exist but do not assess technological aptitude for particular tasks or incorporate available electronic health record data to improve efficiency and efficacy. As such, we propose a multidomain digital health readiness assessment that incorporates a person's stated goals and motivations for use of digital health, a focused digital health literacy assessment, passively collected data from the electronic health record, and a focused aptitude assessment for critical skills needed to achieve a person's goals. This combination of elements should allow for easy integration into clinical workflows and make the assessment as actionable as possible for health care providers and in-clinic digital health navigators. Digital health readiness profiles could be used to match individuals with support interventions to promote the use of digital tools like telehealth, mobile apps, and remote monitoring, especially for those who are motivated but do not have adequate experience. Moreover, while effective and holistic digital health readiness assessments could contribute to increased use and greater equity in digital health engagement, they must also be designed with inclusivity in mind to avoid worsening known disparities in digital health care.

(JMIR Mhealth Uhealth 2024;12:e58035) doi:10.2196/58035

KEYWORDS

digital health; digital health literacy; informatics; digital disparities; digital health readiness; inclusivity; digital health tool; literacy; patient support; health system

Introduction

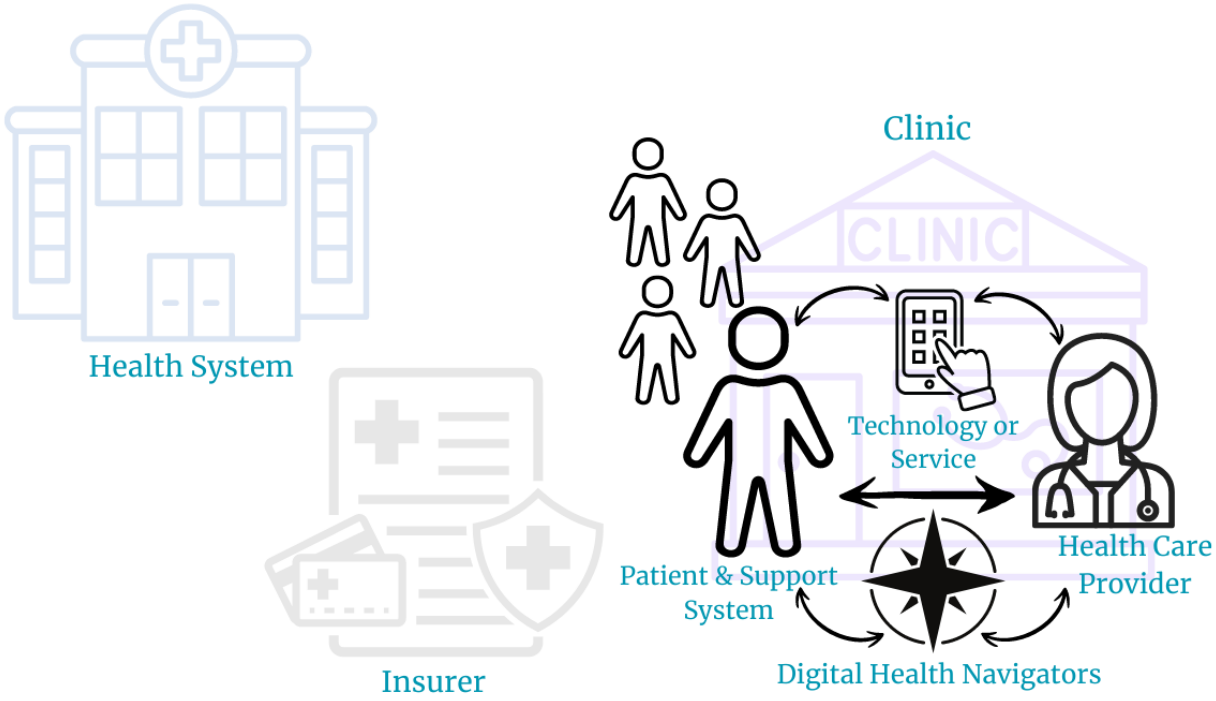
The use of digital tools for health care—including video visits, patient portals, mobile apps, and remote monitors—has risen exponentially over the last decade and become more essential for care access during and after the COVID-19 pandemic [1,2]. Patients using digital health tools have been shown to have better outcomes in managing many outpatient health conditions, including diabetes [3,4], anxiety and mood disorders [5], hypertension [6], and chronic pain [7]. Still, despite their growing incorporation into health care and potential to improve health outcomes, many who could benefit from these tools are not using them [1,2,8,9]. If health systems can develop approaches to close this gap with innovative and tailored pathways to digital health care, they could improve access, inclusivity, and outcomes.

Prior approaches to increase digital health engagement focused on several domains, including such logistical factors as broadband internet access [10], access to smartphones, and the ability of individuals to use technology to participate in health care and understand their health (ie, digital health literacy) [11]. Initial assessments of digital health literacy in the mid-2000s focused on the ability to use the internet, but they have since expanded to encompass smartphones, mobile apps, and social media [12,13]. As a construct, digital health literacy has also grown to reflect multiple domains of health technology use, including personal aspects like prior experiences, digital self-efficacy, motivation to use digital health, and access to technology [14]. The evolution of these assessments reflects changes in the technological environment but also demonstrates the multifaceted nature of digital health literacy overall. Future approaches to facilitating further equitable growth of digital health could consider the ecosystem of factors that drive engagement with these tools. General health literacy is increasingly understood as a relational concept in which patients and health care providers (HCPs) balance their skills and abilities against the demands of health care systems [15,16]. Digital health readiness for individual patients exists within similar contexts and is impacted by the technological tools themselves (particularly the demands that they place on patients), the HCPs prescribing and monitoring their use, the clinics and digital health navigation services where technological instruction occurs, the health systems and their approach toward digital health implementation, and the insurers that control coverage of these services and tools (Figure 1). In this paper, we review current digital health literacy measures to assess and predict a person's ability to engage with digital health, discuss their relative strengths and weaknesses, and describe our holistic vision for health care systems to assess digital health readiness efficiently with health record data.

Multidomain digital health readiness assessments could create a phenotype for each patient representing how prepared, experienced, and equipped they are to use a particular digital health tool at a certain point in time [17]. Prior studies have established approaches to understand readiness within health systems (ie, how prepared and experienced a system is for digital care implementation) [18], within individual health care facilities [19], and among health professionals themselves [20]. Approaches for comprehensively defining and assessing individual-level digital health readiness could become central to health system and payor operations, as signaled by the Center for Medicare and Medicaid Services (CMS) mandate that Medicare Advantage organizations offer "digital health education" for telehealth to their members [21].

Creating effective and holistic digital health readiness assessments could contribute to increased use of and access to these tools among patients and their families. In this paper, we focus only on assessing individual, patient-level digital health readiness, but we acknowledge that this construct can be applied to any node within the digital health readiness ecosystem, as noted above and in Figure 1.

Figure 1. Health care system components of a proposed digital health readiness ecosystem. These are the possible health care nodes of a digital health readiness ecosystem—all of which impact digital health use—including the technology or service, patients and their support system, health care providers, digital health navigators, insurers, and clinics and health systems.



Strengths and Weaknesses of Current Digital Health Readiness Measures

Current methods to assess digital health readiness have several strengths and weaknesses.

One strength of these measures is that they assess relevant aspects of digital health participation and are often short enough to be incorporated into clinical practice; however, these measures assess personal attitudes alone without considering technological aptitude. For example, the eHealth Literacy Scale (eHEALS) is the most cited digital health literacy measure and focuses on assessing a person's attitudes, confidence, and subjective skill level in using internet search engines and evaluating online information, yet it does not assess the experience needed for smartphones and wearable monitors or address such structural factors as device access (either through personal ownership or sharing) [13]. Newer measures such as the Digital Health Care Literacy Scale do capture skills for using and troubleshooting mobile apps and videoconferencing apps in a brief manner that is primed for clinical settings, but they also do not assess technical aptitude or device access [22]. For digital health readiness assessments to be useful in the clinical operations of health systems, they should have an aptitude assessment to stratify individuals into levels with matched support interventions. Additionally, research will be needed on what demonstrated skills are most important for a particular care modality (like a video visit versus wearing a remote monitor).

More thorough digital health readiness assessments cover many relevant aspects of the digital health care experience; however, they may be logistically challenging to administer in clinical settings. For instance, the recent Digital Health Readiness Questionnaire (from 2023) gathers a more detailed assessment of a person's experiences with digital health by asking about their skills, digital literacy, digital health literacy, device use, and learnability, but its 20 items might be cumbersome to administer in a busy primary care setting, do not assess actual aptitude, and do not include questions about device or internet access [23]. Even more robust assessments, including the eHealth Assessment Toolkit [17] and eHealth Literacy Questionnaire [24], are validated and available, though their comprehensiveness also likely makes them unwieldy for application in clinical settings. For example, the eHealth Assessment Toolkit [17] has 44 questions encompassing 7 different tools for digital health care.

One strength of contemporary digital health readiness measures is that they are grounded in updated theoretical constructs of digital health equity that aim to improve engagement with populations facing health disparities and reflect our current technological environment.

The framework for digital health equity augmented the National Institute on Minority Health and Health Disparities research framework by adding individual, interpersonal, community, and societal aspects of the digital environment and patient experience [14]. Previously elaborated digital health readiness research strategies like those from Lyles et al [25] and Jaworski et al [26] were built on components such as "access, motivation and trust, and digital health literacy" that are also fundamental for boosting digital health engagement. Despite being published relatively recently, these frameworks are widely cited and are being incorporated into wide-ranging fields, including behavioral health research, addiction medicine, and cardiovascular medicine—among others [27,28]. While these updated

constructs reflect the current experiences of being a digital health care user, they will also likely need to be updated over time to match the dynamic nature of digital health innovation and remain relevant in the frantic pace of clinical care. Moreover, as seen in the following scenario, approaches to digital health readiness will need to be agile and adaptable to meet the unique needs of each individual.

Scenario 1: Digital Health Readiness and Wearable Health Monitors

This hypothetical patient scenario (Textbox 1) reflects the challenges of applying individual digital health readiness assessments and how clinical teams could be responsive to each person's unique needs.

Ms T's case demonstrates the importance of aptitude testing (eg, prompting a user to show an instructor how they might use a phone app) and how a care team might adjust a digital health care modality to best meet the needs of a patient.

Another weakness of current digital health literacy and readiness measures is that they do not integrate passively collected data from the electronic health record to improve efficiency and efficacy. Using available metrics—such as a visualized breakdown of previous in-person care, completed video visits, completed phone visits, and patient portal use— can increase the efficiency of digital health readiness assessments and portray a person's actual care use compared with their stated goals. Examples include the Telemedicine ImPACT Score [29] and EpicCare Video Visit Technical Risk Score [30], which use data on the number of prior completed video visits and portal messages sent to forecast future digital engagement without the need to administer a questionnaire. These data seamlessly contribute information about an individual's digital determinants of health—that is, the larger social, personal, and structural barriers that impact digital health engagement [14,31]—and could focus on particular factors that are most predictive of certain tasks (like completing a video visit) [32-34]. Looking at a person's health record data in a digital health readiness profile, in-clinic technology navigators may find that a person has no broadband access or internet experience and recommend in-person care over virtual care until these factors are addressed. Passive health record data could refine in-person and digital care delivery so that patients are accessing resources in a way that matches their personal situations.

The essential elements needed for comprehensive and practical digital health readiness assessments will include aptitude testing, in addition to evaluating attitudes toward technology, customizing skill assessment to address emerging technologies, and incorporating passively collected health system data. Existing digital health literacy screening metrics and digital health prediction tools each have strengths that could create a more comprehensive profile of a person's prior technological experience and could be adapted to the use of new technologies over time.

Textbox 1. Hypothetical patient scenario 1.

Ms T is a woman aged 63 years with a laptop computer and a smartphone who regularly searches for health information on the internet. Ms T qualifies for a continuous glucose monitor (CGM) to track her blood sugars; however, the device typically downloads data to a smartphone for users to view their trends. She has nerve damage from diabetes that limits her ability to navigate smartphone screens, but she is able to use computer keyboards without issue. Once the CGM is ordered, the diabetes education team asks her to bring whichever devices she most commonly uses to her CGM training session. During her visit, the diabetes nurse educator evaluates her for digital health literacy using the 3-item Digital Health Care Literacy Scale and feels that she is prepared to use the CGM interface. After the educator downloads the CGM app on her smartphone, Ms T is prompted to sign in and create an account. Immediately, the staff notices that she has issues navigating the smartphone interface. Pivoting to make the technology more usable for her, they set up the CGM application on her laptop so that she can view her blood sugar trends more easily.

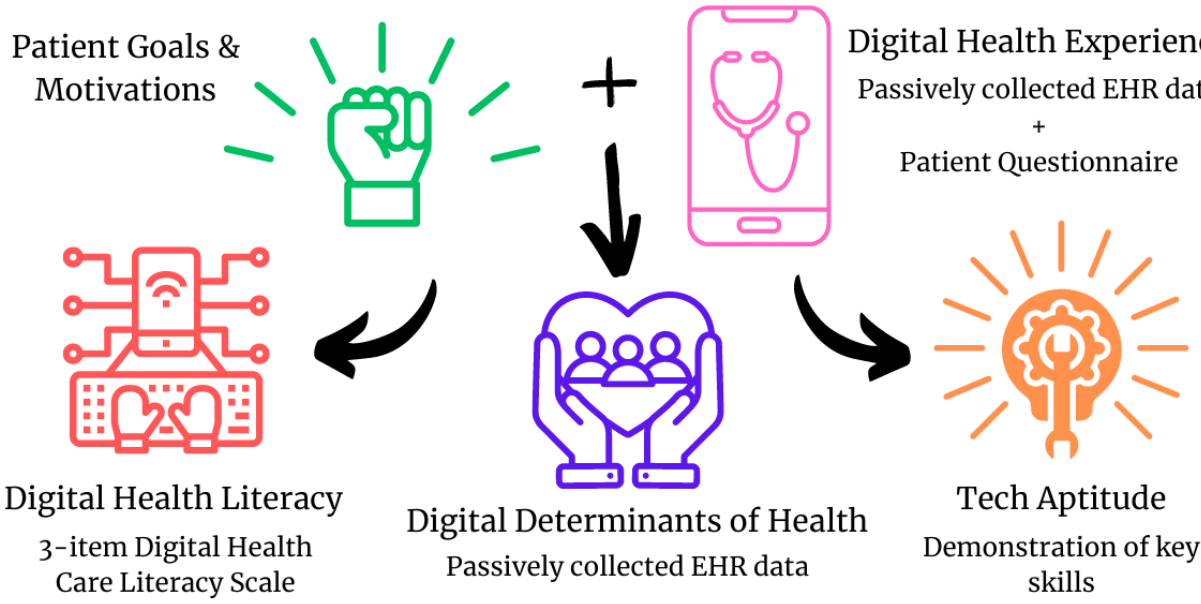
Making Digital Health Readiness Assessments Practical and Efficient for Clinical Settings

We envision a holistic digital heath readiness assessment that will enable health systems to deliver targeted support to those who need it most and close gaps in use. Similar to the Conversational Health Literacy Assessment Tool (CHAT), which is designed to assess multiple dimensions of a person's health literacy in health care settings, digital health readiness assessments could be designed to provide a more comprehensive and pragmatic picture of a patient's digital health strengths and obstacles [35]. In particular, the Health Promotion Barriers and Support, Health Information Access and Comprehension, and Current Health Behaviors domains from the CHAT could be adapted to a digital health context. Digital health readiness assessments could begin with questions about personal goals for health technology use and prior digital health experience, followed by focused aptitude testing for a particular digital health tool or goal, a brief digital health literacy assessment, and visualization of that person's health systems data to probe into their digital determinants of health. Figure 2 reflects the proposed elements of an individual digital health readiness profile that would allow HCPs and care navigators to understand a person's digital phenotype and act to meet their unique needs. The components of Figure 2 [14] represent our thoughts on ways to address the strengths and weaknesses outlined above and were informed by the framework for digital health equity [14]. This multi-domain approach would incorporate patient-reported data with passive data from health systems and payors to make responses more relevant and able to be added to busy clinical workflows. The key difference from existing digital health literacy assessments is the incorporation of a focused aptitude test assessment (such as having a patient show how they use a mobile app for 1-2 minutes) and the integration of passively collected clinical data. These aspects would make digital health readiness phenotyping more efficient, systematic, and, hopefully, effective for clinical settings.

As technology evolves and alters the required skills to participate in modern health care, digital health readiness assessments will

need to grow in kind to reflect these skills. Ideally, the collection of inputs will differ for specific tasks. For example, completing a video visit may involve downloading a mobile app, registering an account, checking in online, and signing in to the appointment. In contrast, registering for a patient portal may involve only some of these steps. Domain-specific digital health readiness assessments could make the assessment most relevant to patients and their goals. The following fictional vignette shows how digital health readiness assessments could be tailored to help patients complete a specific task—such as how to log on to and complete a video visit.

Figure 2. The proposed components of a holistic digital health readiness assessment based on the thoughts of the authors and the digital health equity framework of Richardson et al [14].



Scenario 2: Digital Health Readiness and Telehealth

The hypothetical patient scenario shown in Textbox 2 reflects how passively collected data could link patients with digital health navigation services to improve digital health care outcomes.

Looking at Mr P's case, he is a person who has ostensibly high digital health readiness through demonstrated skills, access to a network, and use of a health system app; however, he has also consistently had issues logging in for video visits, which adversely impacted his digital health care use and increased his risk of hospital readmissions. In this scenario, an automated alert based on previous patterns of digital health care use from electronic health record data triggered help with navigating video visits from a digital health navigator, which many health systems offer [36-38]. That alert could have triggered office staff to arrange an in-person appointment or home visit to assess his ability to use telehealth and provide help if he could not.

Having systematic processes in place to assess who is most appropriate for in-person versus remote or asynchronous care could guide efficient service delivery and use of resources. With their abundance of claims data and the opportunity to trial different variations of digital support pathways, integrated delivery and finance systems represent a unique setting where digital health readiness measures could be deployed, tested, and refined.

Digital health readiness assessments could be a key step toward making digital health implementation more systematic for all people, leading to greater equity and effectiveness. In many clinics, the process of selecting in-person care versus telemedicine could be tied to the nature of the medical issue, the judgment of scheduling and treating team members, and personal preferences (ie, a subset of patients who always want in-person care). Adding more specificity to digital health implementation through the creation of care delivery phenotypes—that is, providing navigation support for patients who are motivated to use digital health but are inexperienced—would optimize this care. It is likely that many opportunities for digital engagement and adoption of new tools are missed simply because health systems do not have robust ways to screen for who is best equipped and motivated for digital health but has not used it. Rather than limiting digital health to those patients who are already confident with technology, streamlined and methodical digital onboarding guided by a digital health readiness assessment could expand the reach of these tools to more patients. In turn, this could provide greater efficiency and, in some cases, reduced costs [39] for patients in scenarios where similar treatment outcomes have been achieved with video versus in-person visits [40]. Differentiating those who can complete a telemedicine appointment on their own from those who might need additional support would further expand digital health as a standard of care and improve the service experience for all patients.

To fully assess digital health readiness, we should also consider how a person's situation may change over time as well as how personal and community resources could help them succeed.

With an aging and increasingly medically complex population, digital health readiness phenotypes will likely be dynamic and may need to be repeated in certain circumstances, such as a major health event, functional decline, cognitive impairment, financial insecurity, or loss of family support [17]. In the event that a person can no longer use a particular tool, a support person may be best suited to provide digital health support in a convenient environment like a health center–affiliated or community-embedded internet clinic [41]. Furthermore, studies have shown that patients with limited technology experience are often able to complete a telehealth visit with the help of a family member, friend, or caregiver—thereby providing an opportunity to engage those with lower digital health readiness from the onset [42,43]. Partnering with patients, families, and communities could help to personalize digital care delivery pathways even further and improve engagement.

Textbox 2. Hypothetical patient scenario 2.

Mr P is a man aged 75 years who has been hospitalized 5 times in the past year for decompensated heart failure. He has a smartphone that enables him to message his primary care provider and heart failure specialist via his health system's patient portal. As he transitioned between hospitals, skilled nursing facilities, and home, he missed multiple follow-ups. His primary care office proactively contacts him at home and sets up a video visit to reestablish care. When the time for the appointment arrives, his primary care provider begins the visit but Mr P cannot log in. After he spends 10 minutes of the 30-minute appointment trying to use the videoconferencing platform, his doctor switches to a phone visit. At the end of the visit, his doctor receives an automated alert from the electronic health record noting that prior scheduled video visits have been converted to phone visits. Looking deeper into the situation, the doctor notices that recurrent telehealth platform issues have taken time away from health care providers to discuss all aspects of his health issues in prior visits—especially dietary counseling (a key reason for his hospitalizations). After the visit, Mr P is referred for an in-person digital health navigation session where he is instructed on ways to troubleshoot the telehealth platform and demonstrate that he can use the videoconferencing service independently.

Challenges of Implementing Digital Health Readiness Assessments

While digital health readiness assessments apply to individual patients, health systems will also need to build infrastructure to respond to the results of these assessments in a meaningful way to realize their full potential. There are established standards to promote organizational health literacy within health systems that could be applied to digital health implementation—including fostering a culture among employees that promotes communication and engagement with patients and families using technology [44,45].

Moreover, HCPs and team members also have varying levels of digital health readiness that affect the implementation of digital health readiness assessments. Similar to medication prescribing, HCPs often serve as gatekeepers for recommending and promoting digital health tools. HCPs' awareness and perceptions of the benefits of digital tools have been identified as determinants of mobile app uptake for chronic disease management [46]. While one might assume that HCPs would have more than adequate digital health readiness and literacy, some studies of hospitals in resource-limited settings worldwide (including one from Ethiopia during the COVID-19 pandemic) have found that less than half of HCPs had high digital health literacy [47]. Health care systems must consider the levels of technological awareness, comfort, and competence among their HCPs when considering more equitable digital health implementation.

There are also potential risks and ethical concerns involved in digital health implementation. With studies showing that digital health engagement is lower among older people, those who require an interpreter, and those who live in more deprived areas [48], efforts to shift more and more health care to digital platforms could exacerbate gaps in care. Furthermore, while the aforementioned evaluation frameworks for digital health tools do consider inclusivity and equity for diverse populations, studies have suggested that only 58% of mobile app evaluation frameworks do so, meaning that vital perspectives on technological tools may still be left behind [49]. Tying back to digital health literacy and health literacy, patients could experience delays in care if they were to choose telehealth or a patient portal message for a condition that warrants in-person evaluation. Personal health data collection and security are also important considerations for making sure that participating in digital health care is safe for all users.

A challenge of aptitude- and analytics-based digital health readiness assessment approaches is that they could amplify societal inequities if not designed carefully and evaluated among minoritized populations. Assessments based solely on aptitude may be biased against other-abled individuals with visual or hearing impairments or people whose primary language is not English. Moreover, given the complex array of factors that impact digital health engagement, digital health readiness assessments cannot be perfectly comprehensive. Digital health literacy is a single digital determinant of health that incorporates a person's underlying literacy, numeracy, and general health literacy—each of which could not be measured or acted upon in a single clinic visit. Using passively collected data carries the risk of perpetuating systemic biases through algorithmic determinism (eg, the perpetuation of systemic bias through algorithms trained on biased data) [50,51] and underrepresentation of marginalized groups in data overall [52], which could further contribute to the digital health divide [21]. It will be important to test and validate digital health readiness assessments among diverse patients. If the evidence for these assessments has not yet been established among certain groups, this should be noted in the electronic health record and factored into how they are deployed and understood.

Conclusion

Assessing and supporting individual patient-level digital health readiness is a crucial step toward maximizing benefits from digital health care and could provide a path toward greater digital health equity. More systematic approaches to support patients with low digital health readiness could ensure that

assessments are actionable for clinicians, payors, and health systems. If we can work to increase the reach of health technology to keep up with the evolution of the consumer electronics market, more patients could be empowered to enter the digital health care age and benefit from these new tools.

Acknowledgments

We acknowledge Ethan Lennox, MA, Division of General Internal Medicine, University of Pittsburgh School of Medicine, for his assistance in editing this manuscript.

Conflicts of Interest

None declared.

References

1. Samson L, Tarazi W, Turrini G. Medicare beneficiaries' use of telehealth in 2020: trends by beneficiary characteristics and location. Office of the Assistant Secretary for Planning and Evaluation. 2021. URL: https://aspe.hhs.gov/sites/default/files/documents/a1d5d810fe3433e18b192be42dbf2351/medicare-telehealth-report.pdf [accessed 2024-07-02]
2. Lee E, Grigorescu V, Enogieru I. Updated national survey trends in telehealth utilization and modality (2021-2022). Office of the Assistant Secretary for Planning and Evaluation. 2023. URL: https://aspe.hhs.gov/reports/updated-hps-telehealth-analysis-2021-2022 [accessed 2014-07-02]
3. Lee EY, Cha S, Yun J, Lim S, Lee J, Ahn Y, et al. Efficacy of personalized diabetes self-care using an electronic medical record-integrated mobile app in patients with type 2 diabetes: 6-month randomized controlled trial. J Med Internet Res 2022 Jul 28;24(7):e37430 [FREE Full text] [doi: 10.2196/37430] [Medline: 35900817]
4. Sabharwal M, Misra A, Ghosh A, Chopra G. Efficacy of digitally supported and real-time self-monitoring of blood glucose-driven counseling in patients with type 2 diabetes mellitus: a real-world, retrospective study in north India. Diabetes Metab Syndr Obes 2022;15:23-33 [FREE Full text] [doi: 10.2147/DMSO.S345785] [Medline: 35023937]
5. Rollman BL, Herbeck Belnap B, Abebe KZ, Spring MB, Rotondi AJ, Rothenberger SD, et al. Effectiveness of online collaborative care for treating mood and anxiety disorders in primary care: a randomized clinical trial. JAMA Psychiatry 2018 Jan 01;75(1):56-64 [FREE Full text] [doi: 10.1001/jamapsychiatry.2017.3379] [Medline: 29117275]
6. Li R, Liang N, Bu F, Hesketh T. The effectiveness of self-management of hypertension in adults using mobile health: systematic review and meta-analysis. JMIR Mhealth Uhealth 2020 Mar 27;8(3):e17776 [FREE Full text] [doi: 10.2196/17776] [Medline: 32217503]
7. Fanning J, Brooks AK, Hsieh KL, Kershner K, Furlipa J, Nicklas BJ, et al. The effects of a pain management-focused mobile health behavior intervention on older adults' self-efficacy, satisfaction with functioning, and quality of life: a randomized pilot trial. Int J Behav Med 2022 Apr;29(2):240-246 [FREE Full text] [doi: 10.1007/s12529-021-10003-3] [Medline: 34018138]
8. Lucas JW, Villarroel MA. Telemedicine use among adults: United States, 2021. NCHS Data Brief 2022 Oct(445):1-8 [FREE Full text] [Medline: 36255940]
9. Bober T, Rothenberger S, Lin J, Ng JM, Zupa M. Factors associated with receipt of diabetes self-management education and support for type 2 diabetes: potential for a population health management approach. J Diabetes Sci Technol 2023 Sep;17(5):1198-1205 [FREE Full text] [doi: 10.1177/19322968231176303] [Medline: 37264614]
10. Reddick CG, Enriquez R, Harris RJ, Sharma B. Determinants of broadband access and affordability: An analysis of a community survey on the digital divide. Cities 2020 Nov;106:102904 [FREE Full text] [doi: 10.1016/j.cities.2020.102904] [Medline: 32921864]
11. Norman CD, Skinner HA. eHealth literacy: essential skills for consumer health in a networked world. J Med Internet Res 2006 Jun 16;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
12. van der Vaart R, Drossaert C. Development of the Digital Health Literacy Instrument: measuring a broad spectrum of health 1.0 and health 2.0 skills. J Med Internet Res 2017 Jan 24;19(1):e27 [FREE Full text] [doi: 10.2196/jmir.6709] [Medline: 28119275]
13. Norman CD, Skinner HA. eHEALS: the eHealth Literacy Scale. J Med Internet Res 2006 Nov 14;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
14. Richardson S, Lawrence K, Schoenthaler AM, Mann D. A framework for digital health equity. NPJ Digit Med 2022 Aug 18;5(1):119 [FREE Full text] [doi: 10.1038/s41746-022-00663-0] [Medline: 35982146]
15. Baker DW. The meaning and the measure of health literacy. J Gen Intern Med 2006 Aug;21(8):878-883 [FREE Full text] [doi: 10.1111/j.1525-1497.2006.00540.x] [Medline: 16881951]
16. Pelikan JM. The evolving concept of the health literate health care organization. Tervise Arengu Instituut. 2014. URL: https://www.tai.ee/sites/default/files/images/eventlist/events/19-11-14_TEH15_Pelikan.pdf [accessed 2024-09-25]

17. Karnoe A, Furstrand D, Christensen KB, Norgaard O, Kayser L. Assessing competencies needed to engage with digital health services: development of the eHealth Literacy Assessment Toolkit. J Med Internet Res 2018 May 10;20(5):e178 [FREE Full text] [doi: 10.2196/jmir.8347] [Medline: 29748163]
18. Lennon MR, Bouamrane M, Devlin AM, O'Connor S, O'Donnell C, Chetty U, et al. Readiness for delivering digital health at scale: lessons from a longitudinal qualitative evaluation of a national digital health innovation program in the United Kingdom. J Med Internet Res 2017 Feb 16;19(2):e42 [FREE Full text] [doi: 10.2196/jmir.6900] [Medline: 28209558]
19. Getachew E, Woldeamanuel Y, Manyazewal T. Capacity and readiness assessment of healthcare facilities for digital health interventions against tuberculosis and HIV in Addis Ababa, Ethiopia. Front Digit Health 2022;4:821390 [FREE Full text] [doi: 10.3389/fdgth.2022.821390] [Medline: 35295619]
20. Bingham JM, Rossi MA, Truong H. Addressing the need for a telehealth readiness assessment tool as a digital health strategy. J Am Pharm Assoc (2003) 2022;62(5):1524-1527 [FREE Full text] [doi: 10.1016/j.japh.2022.04.016] [Medline: 35752603]
21. 2024 Medicare Advantage and Part D Final Rule (CMS-4201-F). Centers for Medicare and Medicaid Services. URL: https://www.cms.gov/newsroom/fact-sheets/2024-medicare-advantage-and-part-d-final-rule-cms-4201-f [accessed 2024-09-25]
22. Nelson LA, Pennings JS, Sommer EC, Popescu F, Barkin SL. A 3-item measure of digital health care literacy: development and validation study. JMIR Form Res 2022 Apr 29;6(4):e36043 [FREE Full text] [doi: 10.2196/36043] [Medline: 35486413]
23. Scherrenberg M, Falter M, Kaihara T, Xu L, van Leunen M, Kemps H, et al. Development and validation of the Digital Health Readiness Questionnaire: prospective single-center survey study. J Med Internet Res 2023 Mar 10;25:e41615 [FREE Full text] [doi: 10.2196/41615] [Medline: 36897627]
24. Kayser L, Karnoe A, Furstrand D, Batterham R, Christensen KB, Elsworth G, et al. A multidimensional tool based on the eHealth Literacy Framework: development and initial validity testing of the eHealth Literacy Questionnaire (eHLQ). J Med Internet Res 2018 Feb 12;20(2):e36 [FREE Full text] [doi: 10.2196/jmir.8371] [Medline: 29434011]
25. Lyles C, Aguilera A, Nguyen O. Bridging the digital health divide: how providers and plans can help communities better adopt digital health tools. California Health Care Foundation. 2022. URL: https://www.chcf.org/wp-content/uploads/2022/02/BridgingDigitalDivideProvidersPlans.pdf [accessed 2024-09-25]
26. Jaworski BK, Webb Hooper M, Aklin WM, Jean-Francois B, Elwood WN, Belis D, et al. Advancing digital health equity: Directions for behavioral and social science research. Transl Behav Med 2023 Apr 03;13(3):132-139 [FREE Full text] [doi: 10.1093/tbm/ibac088] [Medline: 36318232]
27. Miller-Rosales C, Morden NE, Brunette MF, Busch SH, Torous JB, Meara ER. Provision of digital health technologies for opioid use disorder treatment by US health care organizations. JAMA Netw Open 2023 Jul 03;6(7):e2323741 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.23741] [Medline: 37459098]
28. Vijay A, Yancy CW. Health equity in heart failure. Prog Cardiovasc Dis 2024;82:55-60. [doi: 10.1016/j.pcad.2024.01.005] [Medline: 38215916]
29. Crump CA, Wernz C, Schlachta-Fairchild L, Steidle E, Duncan A, Cathers L. Closing the digital health evidence gap: development of a predictive score to maximize patient outcomes. Telemed J E Health 2021 Sep;27(9):1029-1038. [doi: 10.1089/tmj.2020.0334] [Medline: 33170109]
30. Hughes H, Canino R, Sisson S. A simple way to identify patients who need tech support for telemedicine. Harvard Business Review. 2021 Aug 10. URL: https://hbr.org/2021/08/a-simple-way-to-identify-patients-who-need-tech-support-for-telemedicine [accessed 2024-07-02]
31. NTIA Data Explorer. National Telecommunications and Information Administration. URL: https://ntia.gov/other-publication/2022/digital-nation-data-explorer#sel=wiredHighSpeedAtHome&demo=&pc=prop&disp=chart [accessed 2024-07-02]
32. Yan X, Stults CD, Deng S, Liang S, Dillon E, Mudiganti S, et al. Do patients continue to use video visits? Factors related to continued video visit use. Popul Health Manag 2022 Aug;25(4):462-471. [doi: 10.1089/pop.2021.0353] [Medline: 35353619]
33. Dang S, Muralidhar K, Li S, Tang F, Mintzer M, Ruiz J, et al. Gap in willingness and access to video visit use among older high-risk veterans: cross-sectional study. J Med Internet Res 2022 Apr 08;24(4):e32570 [FREE Full text] [doi: 10.2196/32570] [Medline: 35394440]
34. Chen J, Li KY, Andino J, Hill CE, Ng S, Steppe E, et al. Predictors of audio-only versus video telehealth visits during the covid-19 pandemic. J Gen Intern Med 2022 Apr;37(5):1138-1144 [FREE Full text] [doi: 10.1007/s11606-021-07172-y] [Medline: 34791589]
35. O'Hara J, Hawkins M, Batterham R, Dodson S, Osborne RH, Beauchamp A. Conceptualisation and development of the Conversational Health Literacy Assessment Tool (CHAT). BMC Health Serv Res 2018 Mar 22;18(1):199 [FREE Full text] [doi: 10.1186/s12913-018-3037-6] [Medline: 29566755]
36. Mechanic OJ, Lee EM, Sheehan HM, Dechen T, O'Donoghue AL, Anderson TS, et al. Evaluation of telehealth visit attendance after implementation of a patient navigator program. JAMA Netw Open 2022 Dec 01;5(12):e2245615 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.45615] [Medline: 36480202]
37. Rodriguez JA, Charles J, Bates DW, Lyles C, Southworth B, Samal L. Digital healthcare equity in primary care: implementing an integrated digital health navigator. J Am Med Inform Assoc 2023 Apr 19;30(5):965-970 [FREE Full text] [doi: 10.1093/jamia/ocad015] [Medline: 36795062]

38. Perret S, Alon N, Carpenter-Song E, Myrick K, Thompson K, Li S, et al. Standardising the role of a digital navigator in behavioural health: a systematic review. Lancet Digit Health 2023 Dec;5(12):e925-e932 [FREE Full text] [doi: 10.1016/S2589-7500(23)00152-8] [Medline: 38000876]
39. Patel KB, Turner K, Alishahi Tabriz A, Gonzalez BD, Oswald LB, Nguyen OT, et al. Estimated indirect cost savings of using telehealth among nonelderly patients with cancer. JAMA Netw Open 2023 Jan 03;6(1):e2250211 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.50211] [Medline: 36626174]
40. Sammour Y, Main ML, Austin BA, Magalski A, Sperry BW. Outpatient management of guideline-directed medical therapy for heart failure using telehealth: a comparison of in-office, video, and telephone visits. J Card Fail 2022 Jul;28(7):1222-1226. [doi: 10.1016/j.cardfail.2022.02.016] [Medline: 35318127]
41. Joseph N, Hider A, Contreras D. Bridging the digital divide through on-site, health center–based internet clinics. NEJM Catal Innov Care Deliv 2023 Jul 6;4(4):4. [doi: 10.1056/CAT.23.0099]
42. Chung GS, Ellimoottil CS, McCullough JS. The role of social support in telehealth utilization among older adults in the United States during the COVID-19 pandemic. Telemed Rep 2021;2(1):273-276 [FREE Full text] [doi: 10.1089/tmr.2021.0025] [Medline: 35720763]
43. Semere W, Makaroun LK, Beach S, Schillinger D, Rosland A. Family caregivers navigating the health care system: Evolving roles during the COVID-19 pandemic. Fam Syst Health 2022 Jun;40(2):268-273 [FREE Full text] [doi: 10.1037/fsh0000688] [Medline: 35549487]
44. Brach C, Keller D, Hernandez L. Ten attributes of health literate health care organizations. NAM Perspect 2012:1-26. [doi: 10.32388/543501]
45. Brega AG, Hamer MK, Albright K, Brach C, Saliba D, Abbey D, et al. Organizational health literacy: quality improvement measures with expert consensus. Health Lit Res Pract 2019 Apr;3(2):e127-e146 [FREE Full text] [doi: 10.3928/24748307-20190503-01] [Medline: 31294314]
46. Alaslawi H, Berrou I, Al Hamid A, Alhuwail D, Aslanpour Z. Diabetes self-management apps: systematic review of adoption determinants and future research agenda. JMIR Diabetes 2022 Jul 28;7(3):e28153 [FREE Full text] [doi: 10.2196/28153] [Medline: 35900826]
47. Ahmed MH, Guadie HA, Ngusie HS, Teferi GH, Gullslett MK, Hailegebreal S, et al. Digital health literacy during the COVID-19 pandemic among health care providers in resource-limited settings: cross-sectional study. JMIR Nurs 2022 Nov 14;5(1):e39866 [FREE Full text] [doi: 10.2196/39866] [Medline: 36301671]
48. Chapman R, Haroon S, Simms-Williams N, Bhala N, Miah F, Nirantharakumar K, et al. Socioeconomic deprivation, age and language are barriers to accessing personal health records: a cross-sectional study of a large hospital-based personal health record system. BMJ Open 2022 Jan 20;12(1):e054655 [FREE Full text] [doi: 10.1136/bmjopen-2021-054655] [Medline: 35058264]
49. Ramos G, Ponting C, Labao JP, Sobowale K. Considerations of diversity, equity, and inclusion in mental health apps: A scoping review of evaluation frameworks. Behav Res Ther 2021 Dec;147:103990 [FREE Full text] [doi: 10.1016/j.brat.2021.103990] [Medline: 34715396]
50. Milossi M, Alexandropoulou-Egyptiadou E, Psannis KE. AI ethics: algorithmic determinism or self-determination? The GPDR approach. IEEE Access 2021;9:58455-58466. [doi: 10.1109/access.2021.3072782]
51. Cary MP, Zink A, Wei S, Olson A, Yan M, Senior R, et al. Mitigating racial and ethnic bias and advancing health equity in clinical algorithms: a scoping review. Health Aff (Millwood) 2023 Oct;42(10):1359-1368 [FREE Full text] [doi: 10.1377/hlthaff.2023.00553] [Medline: 37782868]
52. Morey BN, Chang RC, Thomas KB, Tulua ', Penaia C, Tran VD, et al. No equity without data equity: data reporting gaps for Native Hawaiians and Pacific Islanders as structural racism. J Health Polit Policy Law 2022 Apr 01;47(2):159-200 [FREE Full text] [doi: 10.1215/03616878-9517177] [Medline: 34522960]

Abbreviations

CGM: continuous glucose monitor
CHAT: Conversational Health Literacy Assessment Tool
CMS: Center for Medicare and Medicaid Services
eHEALS: eHealth Literacy Scale
HCP: health care provider

Edited by L Buis; submitted 03.03.24; peer-reviewed by E Hernandez-Encuentra, K Lubi; comments to author 09.05.24; revised version received 03.07.24; accepted 12.09.24; published 09.10.24.

Please cite as:
Bober T, Rollman BL, Handler S, Watson A, Nelson LA, Faieta J, Rosland AM
Digital Health Readiness: Making Digital Health Care More Inclusive
JMIR Mhealth Uhealth 2024;12:e58035
URL: https://mhealth.jmir.org/2024/1/e58035
doi:10.2196/58035
PMID:

The Role of Smartwatch Technology in the Provision of Care for Type 1 or 2 Diabetes Mellitus or Gestational Diabetes: Systematic Review

Sergio Diez Alvarez[1], MMBBCh, MHA; Antoni Fellas[2,3], BHSc (Hons), BPod, PhD; Katie Wynne[1,3,4], MBBS, PhD; Derek Santos[5,6], PhD; Dean Sculley[7], PhD; Shamasunder Acharya[1,4], MBBS; Pooshan Navathe[8], MBBS; Xavier Gironès[9], PhD; Andrea Coda[2,3], BHSc (Hons), BPod, PhD

[1]School of Medicine and Public Health, College of Health Medicine and Wellbeing, University of Newcastle, University Drive, Callaghan, Newcastle, Australia

[2]School of Health Sciences, College of Health, Medicine and Wellbeing, University of Newcastle, Newcastle, Australia

[3]Equity in Health and Wellbeing Research Program, Hunter Medical Research Institute, Newcastle, Australia

[4]Endocrinology, John Hunter Hospital, Newcastle, Australia

[5]Queen Margaret University, School of Health Sciences, Edinburgh, United Kingdom

[6]University of Gibraltar, Gibraltar, Gibraltar

[7]School of Biomedical Sciences and Pharmacy, College of Health Medicine and Wellbeing, University of Newcastle, Newcastle, Australia

[8]Central Queensland Health, Rockhampton, Australia

[9]Department of Research and Universities, Government of Catalonia–Generalitat de Catalunya, Barcelona, Spain

Corresponding Author:
Sergio Diez Alvarez, MMBBCh, MHA
School of Medicine and Public Health, College of Health Medicine and Wellbeing, University of Newcastle, , University Drive, Callaghan, Newcastle, , Australia

Abstract

Background: The use of smart technology in the management of all forms of diabetes mellitus has grown significantly in the past 10 years. Technologies such as the smartwatch have been proposed as a method of assisting in the monitoring of blood glucose levels as well as other alert prompts such as medication adherence and daily physical activity targets. These important outcomes reach across all forms of diabetes and have the potential to increase compliance of self-monitoring with the aim of improving long-term outcomes such as hemoglobin A_{1c} (HbA_{1c}).

Objective: This systematic review aims to explore the literature for evidence of smartwatch technology in type 1, 2, and gestational diabetes.

Methods: A systematic review was undertaken by searching Ovid MEDLINE and CINAHL databases. A second search using all identified keywords and index terms was performed on Ovid MEDLINE (January 1966 to August 2023), Embase (January 1980 to August 2023), Cochrane Central Register of Controlled Trials (CENTRAL, the Cochrane Library, latest issue), CINAHL (from 1982), IEEE Xplore, ACM Digital Libraries, and Web of Science databases. Type 1, type 2, and gestational diabetes were eligible for inclusion. Quantitative studies such as prospective cohort or randomized clinical trials that explored the feasibility, usability, or effect of smartwatch technology in people with diabetes were eligible. Outcomes of interest were changes in blood glucose or HbA_{1c}, physical activity levels, medication adherence, and feasibility or usability scores.

Results: Of the 8558 titles and abstracts screened, 5 studies were included for qualitative synthesis in this review. A total of 322 participants with either type 1 or type 2 diabetes mellitus were included in the review. A total of 4 studies focused on the feasibility and usability of smartwatch technology in diabetes management. One study conducted a proof-of-concept randomized clinical trial including smartwatch technology for exercise time prescriptions for participants with type 2 diabetes mellitus. Adherence of participants to smartwatch technology varied between included studies, with one reporting input submissions of 58% and another reporting that participants logged 50% more entries than they were required to. One study reported significantly improved glycemic control with integrated smartwatch technology, with increased exercise prescriptions; however, this study was not powered and required a longer observational period.

Conclusions: This systematic review has highlighted the lack of robust randomized clinical trials that explore the efficacy of smartwatch technology in the management of patients with type 1, type 2, and gestational diabetes. Further research is required to establish the role of integrated smartwatch technology in important outcomes such as glycemic control, exercise participation, drug adherence, and diet monitoring in people with all forms of diabetes mellitus.

Trial Registration: PROSPERO International Prospective Register of Systematic Reviews CRD42019136825; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=136825

(JMIR Mhealth Uhealth 2024;12:e54826) doi:10.2196/54826

KEYWORDS

diabetes mellitus; flash glucose monitoring; digital health; smartwatch; smartphones; mHealth; mobile health; glucose monitoring; diabetes; gestational diabetes; systematic review; smartwatch technology; blood glucose; medication adherence; self-monitoring; usability; feasibility; mobile phone

Introduction

Background

Mobile health (mHealth) solutions, which include mobile apps, have rapidly gained popularity as part of the overall management of chronic diseases and have further created opportunities and potential to enhance the ability for self-management in patients with diabetes mellitus (DM) and reduce long-term and irreversible health complications [1-3]. More recently, the introduction of smartwatch technology has opened up the opportunity for the increased use of these wearable devices for health and wellness [4], as well as for the management of chronic medical conditions, such as monitoring chronic obstructive pulmonary disease [5], and health outcomes, such as physical activity monitoring in the older adult population [6]. Smartwatch technology has also seen increasing adoption for use in common chronic conditions such as atrial fibrillation [7] and other cardiac arrhythmias. Most of the early smartwatch intervention research has focused on the accelerometer smartwatch data as a surrogate measure of exercise; however, as innovation in smartwatch technology evolves, more sources of data input become available [8].

Companies manufacturing smartwatches have focused on developing applications that address a broader range of general health and lifestyle interventions such as weight, diets, and exercise. Examples of popular applications include Sugar Sense, Grab Manager, and Diabetes Tracker; however, none have been evaluated for their effects on clinical endpoints. Traditionally, technological companies developing glucose monitoring devices have subsequently launched proprietary applications such as Freestyle Libre Link App and Dexcom Clarity for either personal computers or mobile devices. Henriksen et al [9] showed that research into the area of wrist-worn fitness wearable devices has been accelerating, and technology giants include companies such as Fitbit, Garmin, Misfit, Apple, and Polar, with possible applications in patient diagnostics and treatment. The lack of glucose monitoring technology in smartwatches has served as a real barrier to their broader usefulness in patients with diabetes. This has somewhat been addressed by mobile phones that can integrate data from stand-alone glucose monitoring devices, most recently continuous glucose monitoring (CGM), and other minimally invasive devices such as flash glucose monitoring (FGM). However, early attempts to develop technology such as the GlucoWatch G2 biographer, which was designed to detect trends and track patterns in glucose levels, had setbacks in commercial development [10].

The ability of smartwatch technology to display goals visually and offer motivational graphics upon completion of these goals should not be underestimated. These devices deliver behavioral influences that encourage patients to set specific daily, weekly, or monthly targets and achieve them. Wearable devices can also be wirelessly connected in real time to smartphones that record important outcome data, such as blood glucose levels, which are then automatically synchronized with a freely available and password-protected cloud network. This network is accessible live at any time and anywhere by the patient, family members, and clinicians, allowing potential multidisciplinary clinical decisions to be made in a much more integrated, informed, and prompt manner [11].

The aim of this systematic review is to explore the evidence for the effect and provision of smartwatch technology in patients with type 2 DM (T2DM). This review will also include studies exploring the use of this technology in type 1 DM (T1DM) and gestational DM (GDM).

Research Questions

The following were our research questions:

1. What role can wearable smartwatch-based technology play in the provision of care and improvement of behavioral and clinical outcomes for patients with T1DM, T2DM, or GDM?
2. What is the acceptability and usability of smartwatch-based technology in the care of patients with DM?

Methods

This review was registered with the PROSPERO (International Prospective Register of Systematic Reviews) database for systematic reviews (CRD42019136825).

Search Strategy

The following databases were searched: PubMed, Ovid MEDLINE, Ovid Embase, Cochrane Library, Scopus, Web of Science (Social Sciences Citation Index and Science Citation Index Expanded), IEEEXplore, and CINAHL in August 2023. The search strategy was developed using keywords from systematic reviews that focus on smartwatch technology, mHealth, or mobile app–based interventions in patients with either T1DM, T2DM, or GDM. The OVID search strategy has been provided in Multimedia Appendix 1.

No restrictions were placed on the dates of articles or type of articles during the initial search. Specific search terms were adopted to reflect the requirements of each database. Thesaurus or MeSH (Medical Subject Headings) terms and truncation appropriate to each database were used. Furthermore, ProQuest Dissertations and Theses Global (plus full-text 1997-present)

were searched for relevant dissertations. Theses and conference proceedings were searched via Scopus to capture any additional pertinent research in this emerging field. Reference lists of included studies, as well as the reference lists of related systematic reviews and meta-analyses, were searched to identify any relevant studies. If potential eligible articles were not in English, they were to be translated; however, this was not required.

Types of Studies to Be Included

To capture the full extent of the research literature, any relevant quantitative research addressing the use of smartwatch technology in the treatment of DM was considered. Eligible study designs included randomized controlled trials (including randomized crossover studies and cluster randomized trials), quasi-experimental studies (including interrupted time series studies), controlled before and after studies and observational studies (cohort, case-control, and observational studies (cohort, case-control, and cross-sectional studies). Partially published work (eg, conference abstracts) was eligible only if the full-text reports could be identified.

Participants/population

Participants of any age with a current or previous diagnosis of T1DM, T2DM, and GDM were included.

Intervention(s) and Exposure(s)

Studies using smartwatch technology as an intervention for patients with T1DM, T2DM, or GDM were included. Both wrist-worn and clip-on smartwatches were included; however, they must be "smart devices" (Bluetooth or Wi-Fi enabled to allow synchronizing with a mobile phone app or website) and must be commercially available.

Comparator(s)/control

Potential studies were not specifically excluded if they did not include a control group. Should studies include a comparator group, they may include alternative mHealth interventions, non–smartwatch app-based interventions, or standard-of-care or placebo interventions.

Exclusion Criteria

The rationale for including nonrandomized studies and extension to studies including T1DM and GDM, beyond T2DM, was pragmatic, as a limited number of randomized clinical trials (RCTs) evaluating the use of smartwatch technology in the management of DM were identified. Studies evaluating patients only with prediabetes were excluded.

Outcomes of Interest

Primary Outcomes

The primary outcomes were as follows:

1. Glycemic control: including measures of hemoglobin A_{1c} (HbA_{1c}), CGM parameters using subcutaneous CGM or interstitial FGM, or measures of diabetic complications (eg, fetal or maternal outcomes in GDM).
2. Lifestyle and medication: at least one measure of either objective or subjective record of physical activity, dietary records, or medication adherence.

Secondary Outcomes

The secondary outcomes were acceptance, usability, and acceptability of the smartwatch as an intervention in patients with T1DM, T2DM, or GDM. This included study retention rates, how often patients wore the smartwatch, and any comments/qualitative data on the patient's perception of the watch.

Data Extraction

The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses; Checklist 1) guidelines were used in the identification of eligible studies. Of the studies retrieved from the initial search, titles and abstracts were screened by the authors (AF and DS) using Covidence software. Full texts of potential eligible studies were then screened by two independent authors (AF and DS) to identify eligible studies, with discrepancies resolved by a third reviewer (SDA) if required. The review assessed and discussed the effect of missing data and the degree of its effect on the overall synthesized results.

Data were extracted using a standardized form including the following parameters: publication details (authors, year, and country of study), participant characteristics (number of participants, baseline characteristics, inclusion and exclusion criteria, and type of diabetes), methods (study design, baseline measure, time points [when data were collected: at baseline and the end of the study], and study setting [location, year, and environment]), intervention duration and description including manufacturer/brand of smartwatch used, outcome measures used to identify the effects of the smartwatch intervention, smartwatch-related process evaluation outcomes (usability, acceptability, adherence, and interaction), measures of glycemic clinical outcomes and diabetic complications; and limitations of the study.

Risk of Bias Assessment

The Downs and Black (1998) methodological quality assessment checklist was used as different types of study were anticipated [12]. Two independent authors (DS and AF) assessed the risk of bias, with discrepancies in checklist items resolved by the team; if resolution was not achieved, an arbitrator (AC) was nominated. This was never required.

After analysis, study outcomes were reported using summarized descriptive analysis, focusing on the types of study participants, types of interventions, types of smartwatches, length of follow-up, clinical and behavioral outcomes, and smartwatch-related results.

The overall quality of evidence for an outcome was assessed using a Grading of Recommendations, Assessment, Development, and Evaluations (GRADE) approach, which uses the summary of the risk of bias of the outcome across studies to assess the robustness of the evidence [13]. The GRADE approach uses assessments across 5 domains—study limitations, consistency of results, directness of the evidence, precision of the results, and publication/reporting bias—to categorize the levels of quality as high, moderate, low, and very low [13].

Ethical Considerations

This study did not include human subjects research (no human subjects experimentation or intervention was conducted) and so did not require institutional review board approval.

Results

Overview

A total of 11,470 articles were retrieved from the database searches, of which 2912 were immediately removed as duplicates. Overall, 8558 articles were screened by the two independent reviewers; ultimately 30 studies were considered as potentially eligible, and full-text articles were retrieved. In general, 12 studies were excluded, as they did not incorporate smartwatch technology [14-25]. In total, 5 studies were systematic or literature reviews [26-30]. The reference lists of these reviews were screened for potentially eligible papers, but none were identified. There were 4 studies that were excluded, as they were conference abstracts only and a full-text paper was not available for adequate data extraction [31-34]. Among these studies, 3 studies were the wrong study designs, meaning they did not investigate the use of smartwatches or digital technology [35-37]. Finally, 1 study was excluded, as the intervention was not eligible for inclusion [38]. Multimedia Appendix 2 provides a list of excluded studies and the reasons for their exclusion. A total of 5 studies met the inclusion and exclusion criteria listed and were included for analysis in this review [11,39-42]. Figure 1 depicts the full PRISMA flow diagram and Table 1 provides the table of characteristics of included studies.

To date, 5 studies [11,39-42] have evaluated the integration of smartwatch technology in patients with a diagnosis of DM. Each included study's results have been separately provided below due to the heterogeneous nature of methods and reporting.

Figure 1. Depicts a Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram that was generated in Covidence software during the screening phases of this systematic review.

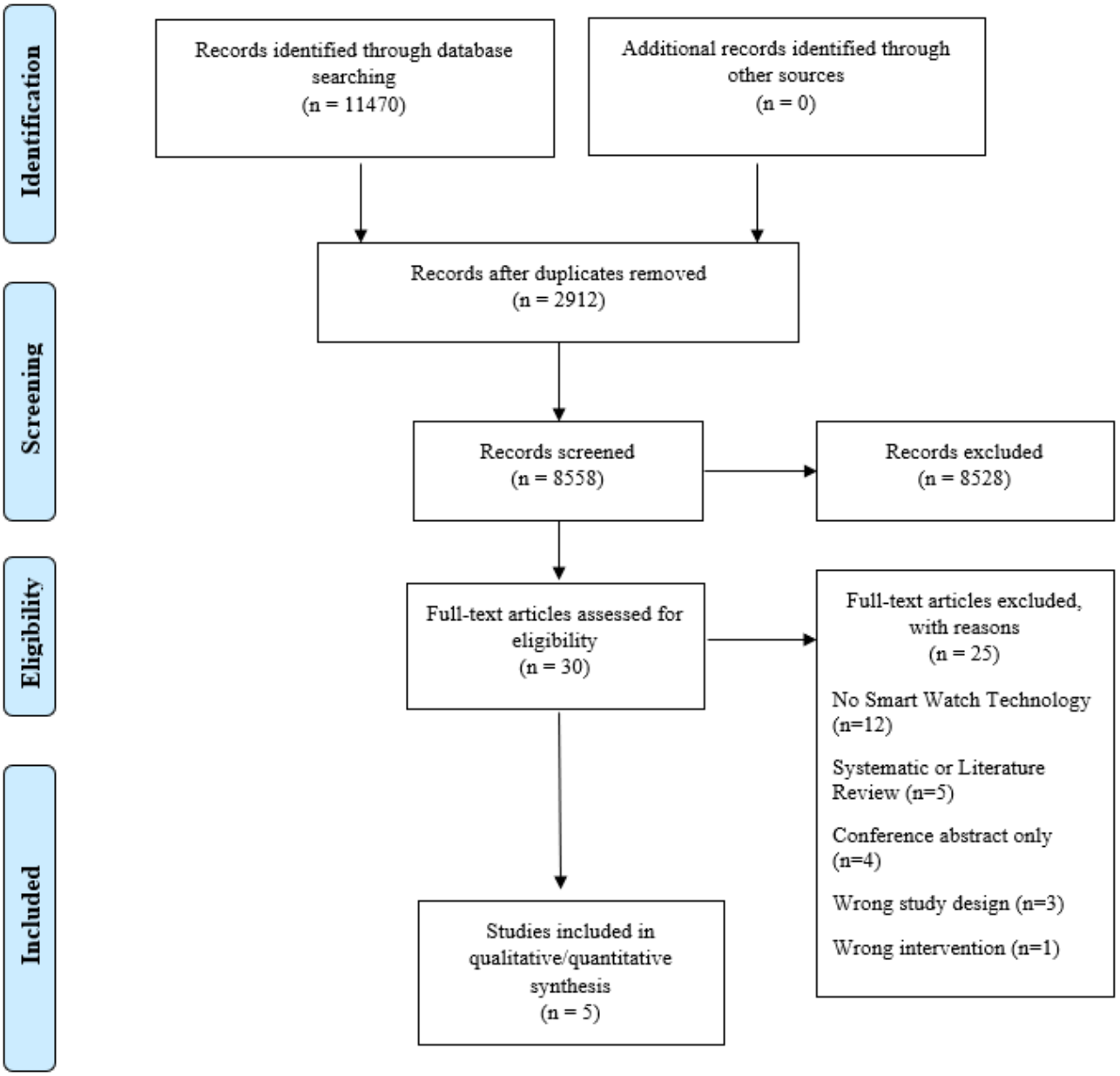


Table . Characteristics of included studies.

Study	Population/sample/context	Aim/objectives	Technologies	Study design	Outcome measures	Results	Limitations
Årsand et al [11]	4 Participants with type 1 diabetes mellitus. Motol University Hospital, Prague.	Feasibility study to test the usability and acceptability of smartwatch technology in patients with type 1 diabetes mellitus.	Pebbles smartwatch on android+ Diabetes Diary App	A novel 6-participant cohort study designed to give the research group feedback on smartwatch technology.	Outcome measures focused on participant feedback on the technology as well as usability of the smartwatch and Diabetes Diary App	Authors concluded that overall users reported positive usability. Participants provided good feedback such as the user interface and being able to input measurements quickly and accessing them. Participants did suggest areas of improvement such as battery usage not able to delete inputs if added accidentally.	Limited sample size with no control group as study was novel and exploratory only. Short observation period.
Kim et al [39]	29 adults with type 2 diabetes mellitus. Seoul National University Hospital, South Korea.	Test the feasibility of HbA_{1c}[a] reduction using a patient-centered, smartphone-based, diabetes care system.	(1) Android-based application with four modules: glucose; diet; physical activity; and social network system (2) Bluetooth glucometer (3) Bluetooth activity tracker	12-week feasibility pilot study. One-arm group.	HbA_{1c}, fasting plasma glucose, body weight, blood pressure, various cholesterol measures. Summary of diabetes self-care activities (SDSCA) was used to evaluate the overall self-management activities for diabetes	After 12 weeks, participants had significantly decreased HbA_{1c} and FPG[b]. Reduction in HbA_{1c} was correlated with the number of daily glucometer inputs. Inputs were generally higher in older patients. Body weight and cholesterol measures were not statistically significant after 12 weeks.	Pilot study. No control group. Short observation time. Small sample size.
Shaw et al [40]	60 adults with type 2 diabetes mellitus. South Eastern United States.	To determine feasibility and acceptability of using multiple mHealth technologies in patients with type 2 diabetes mellitus.	(1) Glucometer – “iHealth” (2) Fitbit (3) Self-report mobile SMS text messaging (4) Cellular enabled Scale by Body Trace	6-month cohort prospective study.	(1) Blood glucose (2) Physical activity – daily steps, distance travelled, and activity intensity (3) Medication adherence (4) Weight	mHealth interventions not used as to improve outcomes listed. Most used technology was the Fitbit. Participants who were younger, had higher HbA_{1c}, and identified as Black were less likely to be engaged with their mHealth devices.	Only observational study. Did not use control group for interventional impact. Small sample size.

Study	Population/sample/context	Aim/objectives	Technologies	Study design	Outcome measures	Results	Limitations
Zahedani et al [41]	665 participants: healthy (448); prediabetic (25); type 2 diabetics (192)	Investigate combined use of continuous glucose monitoring and mobile app (Sugar AI) on glucose tracing, heart rate, and physical activity.	(1) Abbott Freestyle Libre (2) MiBand 3 or Garmin watch (3) Sugar Artificial intelligence app	10-day observational study.	(1) Blood glucose, measured as TIR[c]: 54 - 140 mg/dL for healthy and prediabetes, and 54 - 180 mg/dL for type 2 diabetes mellitus	Authors concluded that a subgroup of those showing poor TIR (combined type 2 diabetes mellitus and prediabetic participants) demonstrated an average of 22.7% improvement in TIR. 62.9% of diabetic participants who showed improved TIR had greater improvement in their daily variation.	Only observation study. Not randomized clinical trial. Short follow-up. Limited results provided on use of Garmin watch or MiBand 3 to improve outcome measures such as blood glucose levels or heart rate, etc.
Chang et al [42]	35 adults with type 2 diabetes mellitus. New South Wales, Australia.	To test the feasibility of prescribing an individualized daily exercise time.	(1) Abbott Freestyle Libre 2 (2) Accelerometer smartwatch (ActiGraph Bluetooth Smart wGT3X-BT	Two-week observational period.	(1) Exercise adherence–feasibility. Proportion of participants completing more than 15 minutes of physical activity (moderate to high intensity) (2) Continuous glucose monitoring (3) Dietary intake	Authors found that participants increased their daily physical activity by an average of 10 minutes per day. However, authors did report adherence issues with a participants not conducting physical activity in the prescribed time.	Proof-of-concept randomized clinical trial. Small sample size. Short observational period.

[a]HbA_{1c}: hemoglobin A_{1c}.

[b]FPG: fasting plasma glucose.

[c]TIR: time in range.

Årsand et al 2015

The “Pebble” smartwatch was tested on 4 participants with T1DM ranging from ages 20 to 46 years. Usability scores for this study averaged 4.4/5, with 5 being the best score. The highest scoring question was that all values were clearly shown on the smartwatch (4.8/5) and the lowest was the ability to make a new registration with the “Pebble” app (4.3/5). Qualitative feedback for appreciated features included the ability to see the last values and user interface while nonappreciated feedback included battery usage, not being able to delete entries from the watch, and the possibility of entering wrong values. A total of 2 of the 4 participants preferred to use their phone to input data as they found it more convenient [11].

Kim et al 2016

An “LG LifeGram” activity tracker combined with a mobile-centered mobile app was tested on 29 participants with T2DM over 12 weeks. The intervention focused on delivering information on glucose control, diet, exercise, and finally a social support network for motivation. After 12 weeks, participants experienced a statistically significant reduction in HbA_{1c} (P=<.0001 with a decrease of 0.6%) and fasting plasma glucose (P=.0088 with a decrease of 20.8 mg/dL) levels. No changes were observed in body weight and cholesterol levels. Authors reported high input scores that correlated with the amount of change in HbA_{1c} levels (P=.0013) [39].

Shaw et al 2020

A total of 60 adults with T2DM were observed for 6 months while using the “Fitbit Alta” (Google Fitbit), which is a reliable and valid accelerometer used for tracking physical activity. This smartwatch was paired with the participant’s smartphone app. A total of 87.45% of daily physical activity submissions for the “Fitbit Alta” were completed by participants, which was the highest achieving outcome for input completions. Other outcomes such as glucose monitoring (60.99%), drug adherence (71.19%), and weight (53.83%) did not score as high as physical activity and, interestingly, were not tracked using smartwatch technology. Technical support was also monitored during this study in which authors reported that 35/60 (58%) participants

required assistance due to devices not syncing or logging-in problems [40].

Zahedani et al 2021

The "Sugar AI" app was used in conjunction with a "MiBand 3" or "Garmin" in 192 participants with T2DM. It is noted that this study also involved healthy individuals and participants with self-diagnosed "prediabetes." The primary outcome in this study was time in range that was defined as individuals with T2DM to be between 3 and 10 mmol/L. The authors reported that 58.3% (112/192) of participants with T2DM and a poor time in range showed an average improvement of 22.7%. Authors did not report adherence, feasibility, or adherence to the intervention [41].

Chang et Al 2023

In this double-blinded proof-of-concept study, 35 participants with physician-diagnosed T2DM were issued a Freestyle Libre 2 (Abbott) and an "ActiGraph Bluetooth Smart wGT3X-BT." These technologies were used and worn for a 14-day baseline period followed by a 14-day interventional period with the purpose of testing the feasibility of prescribing an exercise time to target peak hyperglycemia. The participants were further randomized into the "ExPeak" group, which aimed at conducting physical activity for 30 minutes before peak hyperglycemia, or the "NonPeak" group, which conducted physical activity for 90 minutes after peak hyperglycemia. Adherence to prescribed exercise times showed that only 29% (10/35) of participants adhered to their prescribed times. Overall, 66% (23/35) of participants completed their exercises but not during the times recommended to target peak hyperglycemia. the authors concluded that participation in moderate physical activity increased by 26% during the intervention period versus the baseline period but was not statistically significant between groups (P=.26). There was no statistically significant difference in the intensity of physical activity between groups (P=.77). Authors did not provide specific data on the usability of the smartwatch technology used in this study.

Data extraction revealed the heterogeneous nature of the studies included in this review. This extended to the risk of bias assessment, with results presented in Table 2. Only one study attempted to blind participants to their intervention and randomized this process with robust methods [42].

Table . Risk of bias of included studies.

Items	Årsand et al [11]	Kim et al [39]	Shaw et al [40]	Zahedani et al [41]	Chang et al [42]
1. Is the hypothesis/aim/objective of the study clearly described?	1[a]	1	1	1	1
2. Are the main outcomes to be measured clearly described in the *Introduction* or *Methods* section?	0[b]	1	1	1	1
3. Are the characteristics of the patients included in the study clearly described?	0	1	1	1	1
4. Are the interventions of interest clearly described?	1	1	1	1	1
5. Are the distributions of principal confounders in each group of subjects to be compared clearly described?	0	1	1	1	0 (UTD)[c]
6. Are the main findings of the study clearly described?	0	1	1	1	1
7. Does the study provide estimates of the random variability in the data for the main outcomes?	0	1	1	1	1
8. Have all important adverse events that may be a consequence of the intervention been reported?	0	1	0	0	1
9. Have the characteristics of patients lost to follow-up been described?	1	1	1	1	1
10. Have actual probability values been reported (eg, 0.035 rather than <0.05) for the main outcomes except where the probability value is less than 0.001?	0	1	1	1	1
11. Were the subjects asked to participate in the study representative of the entire population from which they were recruited?	0	0 (UTD)	0 (UTD)	1	0 (UTD)
12. Were those subjects who were prepared to participate representative of the entire population from which they were recruited?	0	0 (UTD)	0 (UTD)	1	0 (UTD)

Items	Årsand et al [11]	Kim et al [39]	Shaw et al [40]	Zahedani et al [41]	Chang et al [42]
13. Were the staff, places, and facilities where the patients were treated, representative of the treatment the majority of patients receive?	0 (UTD)	0	0	1	1
14. Was an attempt made to blind study subjects to the intervention they have received?	0	0	0	0 (UTD)	1
15. Was an attempt made to blind those measuring the main outcomes of the intervention?	0	0 (UTD)	0 (UTD)	0 (UTD)	1
16. If any of the results of the study were based on “data dredging,” was this made clear?	0 (UTD)	0	0 (UTD)	0 (UTD)	0 (UTD)
17. In trials and cohort studies, do the analyses adjust for different lengths of follow-up of patients, or in case-control studies, is the time period between the intervention and outcome the same for cases and controls?	0 (UTD)	1	1	1	1
18. Were the statistical tests used to assess the main outcomes appropriate?	1	1	1	1	1
19. Was compliance with the intervention(s) reliable?	1	1	1	0	0
20. Were the main outcome measures used accurate (valid and reliable)?	0 (UTD)	1	1	1	1
21. Were the patients in different intervention groups (trials and cohort studies) or were the cases and controls (case-control studies) recruited from the same population?	0 (UTD)	1	1	1	0 (UTD)
22. Were study subjects in different intervention groups (trials and cohort studies) or were the cases and controls (case-control studies) recruited over the same period of time?	0 (UTD)	1	1	0 (UTD)	1

Items	Årsand et al [11]	Kim et al [39]	Shaw et al [40]	Zahedani et al [41]	Chang et al [42]
23. Were study subjects randomized to intervention groups?	0	0	0	0	1
24. Was the randomized intervention assignment concealed from both patients and health care staff until recruitment was complete and irrevocable?	*0*	*0*	0	0	1
25. Was there adequate adjustment for confounding in the analyses from which the main findings were drawn?	0 (UTD)	0 (UTD)	0 (UTD)	0 (UTD)	0 (UTD)
26. Were losses of patients to follow-up taken into account?	1	1	1	1	1
27. Did the study have sufficient power to detect a clinically important effect where the probability value for a difference being due to chance is less than 5%?	0	0 (UTD)	0	0 (UTD)	0 (UTD)
Total score	6	17	16	17	19

[a]1: yes.

[b]0: no.

[c]0 (UTD): unable to determine.

Discussion

Principal Findings

This systematic review found that among the current literature, only 5 studies have explored the use of smartwatch technology in people with DM (T1DM, T2DM, and GDM). A total of 4 of the 5 included studies recruited participants with T2DM, while no studies included participants with GDM. This review also provides evidence of the paucity of robust RCTs that evaluate the effect of smartwatch technology in the management of patients diagnosed with DM. Recent research developments have evaluated transcutaneous noninvasive sampling technology such as noninvasive optical glucose monitoring, based on optical glucose monitoring, and noninvasive fluid sampling, based on fluid sample glucose estimation, but these are in their infancy [43]. These technologies aim to reduce the need for patients with diabetes to use the more traditional finger-prick method of measuring their blood glucose levels, which can be painful, more time-consuming, and reduce compliance [44]. It may be an additional impetus for technology development if smartwatches are able to harness these technologies to monitor patients' glucose effectively and noninvasively. T1DM is the area that is most suitable for the development of smartwatch technology for the management of diabetes. Patients with T1DM tend to be younger and thus often have more agile technological literacy and are often early adopters of technological developments. T1DM is also a condition prone to hypoglycemia and thus an important area for technologies that may reduce the risk of developing serious complications that may arise due to hypoglycemia unawareness, which may be where smartwatch technology can most appropriately intercept in patients that exhibit a lack of warning symptoms [45]. The same could also be said for patients with hyperglycemia and life-threatening diabetic ketoacidosis. The carers of patients with T1DM also value the ability to use technology for distant monitoring of children and adolescents with T1DM. Emerging evidence on smartwatch technology may be available soon, with Sehgal et al [46] recently publishing a protocol that will assess the safety and efficacy of the addition of smartwatch technology to usual CGM in adults with T1DM. The authors have stated the primary endpoint will be glucose time in range expressed as a percentage time interstitial glucose between 3.9 and 10 mmol/L. Other outcomes will include quality of life, distress, and sleep quality [46]. Finally, Corbett et al [47] published data on the use of smartwatch gesture–based meal reminders using a proprietary app that picks up "eating motions" by using the smartwatch and uses this to remind patients with T1DM to inject insulin if appropriate.

DM is a heterogenous condition and there is provision for future research in the use of smartwatch technology in all types of diabetes. There is an opportunity for this technology to combine with others like the FreeStyle Libre 2 to prompt regular blood glucose level monitoring, improve time in range and variability, and reduce the risk of serious complications such as

hypoglycemia. Moreover, as more evidence emerges on other subtypes of DM such as diabetes of the exocrine pancreas (type 3c), the provision for evidence-based technologies to monitor and improve outcomes increases [48]. Detailed usability and acceptability trials followed by robust long-term clinical trials should be conducted to test these emerging smart technologies prior to their introduction into the management of patients with diabetes.

Limitations of the Study

This review does not include quantitative analyses or meta-analyses due to a lack of homogenous RCTs. Most papers included were feasibility, usability, and/or acceptability studies and did not include a comparator group. The quality of the papers included was sound. Improvements in blinding of participants and team members to intervention groups need to be considered in future research as well as the sampling of patients selected for studies to promote generalizability of results. This is an area that is rapidly developing and even though an attempt has been made to encompass all contemporaneous literature available, there may be new evidence that will inform the topic. Technology is rapidly developing within the commercial and proprietary arena, making it unavailable for systematic review.

Conclusion

This systematic review highlights the current paucity of evidence supporting the use of smartwatch technology in the management of patients with all types of DM. As smartwatch usage increases with greater affordability and comfort with the technology grows, so does the potential for their significant role in the management of DM. This would especially be the case if wearable technology seamlessly interfaces with glucose monitoring devices or can serve to monitor glucose directly through innovative developments in noninvasive glucose monitoring. Smartwatches would allow patients the ability to check their glucose level frequently and discretely if they are linked to a CGM device. Nutrition and physical activity are cornerstones of diabetes management. Mobile technologies, such as smartphones and wearables, have the potential to educate, motivate, and prompt individuals to optimize lifestyle interventions and thereby favorably affect health.

Authors' Contributions

SDA, AF, and AC conceptualized the research question and methodology. AF and D Sculley screened the titles and abstracts. AF and D Santos assessed the risk of bias. All authors were responsible for drafting the manuscript and ultimately approving the final paper for publication submission.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOCX File, 20 KB - mhealth_v12i1e54826_app1.docx]

Multimedia Appendix 2
Excluded studies.
[DOCX File, 15 KB - mhealth_v12i1e54826_app2.docx]

Checklist 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[PDF File, 157 KB - mhealth_v12i1e54826_app3.pdf]

References

1. Brzan PP, Rotman E, Pajnkihar M, Klanjsek P. Mobile applications for control and self management of diabetes: a systematic review. J Med Syst 2016 Sep;40(9):1-10. [doi: 10.1007/s10916-016-0564-8] [Medline: 27520615]
2. Aikens JE, Zivin K, Trivedi R, Piette JD. Diabetes self-management support using mHealth and enhanced informal caregiving. J Diabetes Complicat 2014 Mar;28(2):171-176. [doi: 10.1016/j.jdiacomp.2013.11.008]
3. Whitehead L, Seaton P. The effectiveness of self-management mobile phone and tablet apps in long-term condition management: a systematic review. J Med Internet Res 2016 May 16;18(5):e97. [doi: 10.2196/jmir.4883] [Medline: 27185295]
4. Reeder B, David A. Health at hand: a systematic review of smart watch uses for health and wellness. J Biomed Inform 2016 Oct;63:269-276. [doi: 10.1016/j.jbi.2016.09.001] [Medline: 27612974]
5. Wu R, Liaqat D, de Lara E, et al. Feasibility of using a smartwatch to intensively monitor patients with chronic obstructive pulmonary disease: prospective cohort study. JMIR Mhealth Uhealth 2018 Jun 14;6(6):e10046. [doi: 10.2196/10046] [Medline: 29903700]
6. Teixeira E, Fonseca H, Diniz-Sousa F, et al. Wearable devices for physical activity and healthcare monitoring in elderly people: a critical review. Geriatrics (Basel) 2021 Apr 7;6(2):38. [doi: 10.3390/geriatrics6020038] [Medline: 33917104]

7. Chon KH, McManus DD. Detection of atrial fibrillation using a smartwatch. Nat Rev Cardiol 2018 Nov;15(11):657-658. [doi: 10.1038/s41569-018-0057-1] [Medline: 29985453]
8. Rafl J, Bachman TE, Rafl-Huttova V, Walzel S, Rozanek M. Commercial smartwatch with pulse oximeter detects short-time hypoxemia as well as standard medical-grade device: validation study. Digit Health 2022;8:20552076221132127. [doi: 10.1177/20552076221132127] [Medline: 36249475]
9. Henriksen A, Haugen Mikalsen M, Woldaregay AZ, et al. Using fitness trackers and smartwatches to measure physical activity in research: analysis of consumer wrist-worn wearables. J Med Internet Res 2018 Mar 22;20(3):e110. [doi: 10.2196/jmir.9157] [Medline: 29567635]
10. Tierney MJ, Tamada JA, Potts RO, et al. The GlucoWatch biographer: a frequent automatic and noninvasive glucose monitor. Ann Med 2000 Dec;32(9):632-641. [doi: 10.3109/07853890009002034] [Medline: 11209971]
11. Årsand E, Muzny M, Bradway M, Muzik J, Hartvigsen G. Performance of the first combined smartwatch and smartphone diabetes diary application study. J Diabetes Sci Technol 2015 May;9(3):556-563. [doi: 10.1177/1932296814567708] [Medline: 25591859]
12. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. J Epidemiol Community Health 1998 Jun;52(6):377-384. [doi: 10.1136/jech.52.6.377] [Medline: 9764259]
13. Brozek JL, Akl EA, Alonso-Coello P, et al. Grading quality of evidence and strength of recommendations in clinical practice guidelines. Part 1 of 3. An overview of the GRADE approach and grading quality of evidence about interventions. Allergy 2009 May;64(5):669-677. [doi: 10.1111/j.1398-9995.2009.01973.x] [Medline: 19210357]
14. Garg SK, Shah VN, Akturk HK, Beatson C, Snell-Bergeon JK. Role of mobile technology to improve diabetes care in adults with type 1 diabetes: the Remote-T1D Study iBGStar® in type 1 diabetes management. Diabetes Ther 2017 Aug;8(4):811-819. [doi: 10.1007/s13300-017-0272-5] [Medline: 28555339]
15. Wang HC, Chang FY, Tsai TM, Chen CH, Chen YY. Development and clinical trial of a smartphone-based colorimetric detection system for self-monitoring of blood glucose. Biomed Opt Express 2020 Apr 1;11(4):2166-2177. [doi: 10.1364/BOE.389638] [Medline: 32341874]
16. Hirst JE, Mackillop L, Loerup L, Gibson OJ, Farmer AJ, Bartlett K, et al. GDm-Health: a pilot study examining acceptability of mobile phone assisted remote blood glucose monitoring for women with gestational diabetes mellitus. Reprod Sci 2014;21(3):252A-253A.
17. Keith-Hynes P, Mize B, Robert A, Place J. The Diabetes Assistant: a smartphone-based system for real-time control of blood glucose. Electronics (Basel) 2014;3(4):609-623. [doi: 10.3390/electronics3040609]
18. Asante E, Bam V, Diji AKA, et al. Pilot mobile phone intervention in promoting type 2 diabetes management in an urban area in Ghana: a randomized controlled trial. Diabetes Educ 2020 Oct;46(5):455-464. [doi: 10.1177/0145721720954070] [Medline: 32998649]
19. Zhai Y, Yu W. A mobile app for diabetes management: impact on self-efficacy among patients with type 2 diabetes at a community hospital. Med Sci Monit 2020 Nov 16;26:e926719. [doi: 10.12659/MSM.926719] [Medline: 33196634]
20. Poulter SE, Meloncelli N, Mack M. Use of a smartphone-based, interactive blood glucose management system in women with gestational diabetes mellitus: a pilot study. Diabetes Res Clin Pract 2022 Mar;185:109224. [doi: 10.1016/j.diabres.2022.109224] [Medline: 35122904]
21. Pamungkas RA, Usman AM, Chamroonsawasdi K, Abdurrasyid. A smartphone application of diabetes coaching intervention to prevent the onset of complications and to improve diabetes self-management: a randomized control trial. Diabetes Metab Syndr Clin Res Rev 2022 Jul;16(7):102537. [doi: 10.1016/j.dsx.2022.102537]
22. Yang Y, Lee EY, Kim HS, Lee SH, Yoon KH, Cho JH. Effect of a mobile phone-based glucose-monitoring and feedback system for type 2 diabetes management in multiple primary care clinic settings: cluster randomized controlled trial. JMIR Mhealth Uhealth 2020 Feb 26;8(2):e16266. [doi: 10.2196/16266] [Medline: 32130172]
23. Yu Y, Yan Q, Li H, et al. Effects of mobile phone application combined with or without self-monitoring of blood glucose on glycemic control in patients with diabetes: A randomized controlled trial. J Diabetes Investig 2019 Sep;10(5):1365-1371. [doi: 10.1111/jdi.13031] [Medline: 30815973]
24. Gunawardena KC, Jackson R, Robinett I, et al. The influence of the smart glucose manager mobile application on diabetes management. J Diabetes Sci Technol 2019 Jan;13(1):75-81. [doi: 10.1177/1932296818804522] [Medline: 30264583]
25. Garnweidner-Holme L, Henriksen L, Torheim LE, Lukasse M. Effect of the Pregnant+ smartphone app on the dietary behavior of women with gestational diabetes mellitus: secondary analysis of a randomized controlled trial. JMIR Mhealth Uhealth 2020 Nov 4;8(11):e18614. [doi: 10.2196/18614] [Medline: 33146620]
26. Martos-Cabrera MB, Velando-Soriano A, Pradas-Hernández L, et al. Smartphones and apps to control glycosylated hemoglobin (HbA1c) level in diabetes: a systematic review and meta-analysis. J Clin Med 2020 Mar 4;9(3):693. [doi: 10.3390/jcm9030693] [Medline: 32143452]
27. Kerr D, King F, Klonoff DC. Digital health interventions for diabetes: everything to gain and nothing to lose. Diabetes Spectr 2019 Aug;32(3):226-230. [doi: 10.2337/ds18-0085] [Medline: 31462878]
28. Hartz J, Yingling L, Powell-Wiley TM. Use of mobile health technology in the prevention and management of diabetes mellitus. Curr Cardiol Rep 2016 Dec;18(12):130. [doi: 10.1007/s11886-016-0796-8] [Medline: 27826901]

29. Kebede MM, Zeeb H, Peters M, Heise TL, Pischke CR. Effectiveness of digital interventions for improving glycemic control in persons with poorly controlled type 2 diabetes: a systematic review, meta-analysis, and meta-regression analysis. Diabetes Technol Ther 2018 Nov;20(11):767-782. [doi: 10.1089/dia.2018.0216] [Medline: 30257102]
30. Constantin J, Mesana L, Pruce D, Syed IA. PDB14 - mobile apps for medication adherence in diabetes. Val Health 2018 Oct;21:S121. [doi: 10.1016/j.jval.2018.09.720]
31. Al-Hajri M, Al-Kawari S, Al-Kharusi A, et al. Why mobile apps for type II diabetes fail to actualize behavioral change among adolescents? Diabetes Technol Ther 2020;22:A235.
32. Yang Y, Lee J, Kang B, et al. Effect of smartphone-based glucose monitoring and feedback system on diabetes management in multiple private clinics setting. Diabetes 2016;65:A318 [FREE Full text]
33. Li C. The effect of case management model based on mobile health on self-management and blood glucose control of type 2 diabetes patients. J Diabetes Investig 2018 Dec;9(S1):47-161 [FREE Full text] [doi: 10.1111/jdi.12939]
34. Aravamudhan A, Jain N, Lewis I, Samuels T, Hogan V, Lipova O. Impact of continuous and flash glucose monitoring on HbA1c, quality of life indicators and hypoglycaemia confidence in children and young people with type 1 diabetes. Pediatr Diabetes 2018 Oct;19:41-150 [FREE Full text] [doi: 10.1111/pedi.12746]
35. Fleming GA, Petrie JR, Bergenstal RM, Holl RW, Peters AL, Heinemann L. Diabetes digital app technology: benefits, challenges, and recommendations. A consensus report by the European Association for the Study of Diabetes (EASD) and the American Diabetes Association (ADA) Diabetes Technology Working Group. Diabetes Care 2020 Jan;43(1):250-260. [doi: 10.2337/dci19-0062] [Medline: 31806649]
36. Kebede MM, Pischke CR. Popular diabetes apps and the impact of diabetes app use on self-care behaviour: a survey among the digital community of persons with diabetes on social media. Front Endocrinol (Lausanne) 2019;10:135. [doi: 10.3389/fendo.2019.00135] [Medline: 30881349]
37. Cappon G, Cossu L, Boscari F, Bruttomesso D, Sparacino G, Facchinetti A. An integrated mobile platform for automated data collection and real-time patient monitoring in diabetes clinical trials. J Diabetes Sci Technol 2022 Nov;16(6):1555-1559. [doi: 10.1177/19322968211024620] [Medline: 34218721]
38. Diabetes Research in Children Network (DIRECNET) Study Group. The accuracy of the GlucoWatch G2 biographer in children with type 1 diabetes: results of the diabetes research in children network (DirecNet) accuracy study. Diabetes Technol Ther 2003;5(5):791-800. [doi: 10.1089/152091503322526996] [Medline: 14633344]
39. Kim EK, Kwak SH, Baek S, et al. Feasibility of a patient-centered, smartphone-based, diabetes care system: a pilot study. Diabetes Metab J 2016 Jun;40(3):192-201. [doi: 10.4093/dmj.2016.40.3.192] [Medline: 27098508]
40. Shaw RJ, Yang Q, Barnes A, et al. Self-monitoring diabetes with multiple mobile health devices. J Am Med Inform Assoc 2020 May 1;27(5):667-676. [doi: 10.1093/jamia/ocaa007] [Medline: 32134447]
41. Dehghani Zahedani A, Shariat Torbaghan S, Rahili S, et al. Improvement in glucose regulation using a digital tracker and continuous glucose monitoring in healthy adults and those with type 2 diabetes. Diabetes Ther 2021 Jul;12(7):1871-1886. [doi: 10.1007/s13300-021-01081-3] [Medline: 34047962]
42. Chang CR, Russell BM, Cyriac T, Francois ME. Using continuous glucose monitoring to prescribe a time to exercise for individuals with type 2 diabetes. J Clin Med 2023 Apr 30;12(9):3237. [doi: 10.3390/jcm12093237] [Medline: 37176677]
43. Valero M, Pola P, Falaiye O, et al. Development of a noninvasive blood glucose monitoring system prototype: pilot study. JMIR Form Res 2022 Aug 26;6(8):e38664. [doi: 10.2196/38664] [Medline: 36018623]
44. Jernelv IL, Milenko K, Fuglerud SS, Hjelme DR, Ellingsen R, Aksnes A. A review of optical methods for continuous glucose monitoring. Appl Spectrosc Rev 2019 Aug 9;54(7):543-572. [doi: 10.1080/05704928.2018.1486324]
45. Martín-Timón I, Del Cañizo-Gómez FJ. Mechanisms of hypoglycemia unawareness and implications in diabetic patients. World J Diabetes 2015 Jul 10;6(7):912-926. [doi: 10.4239/wjd.v6.i7.912] [Medline: 26185599]
46. Sehgal S, De Bock M, Williman J, et al. Study protocol: safety and efficacy of smart watch integrated do-it-yourself continuous glucose monitoring in adults with type 1 diabetes, a randomised controlled trial. J Diabetes Metab Disord 2021 Dec;20(2):2103-2113. [doi: 10.1007/s40200-021-00923-y]
47. Corbett JP, Hsu L, Brown SA, et al. Smartwatch gesture-based meal reminders improve glycaemic control. Diabetes Obes Metab 2022 Aug;24(8):1667-1670. [doi: 10.1111/dom.14737] [Medline: 35491517]
48. Wynne K, Devereaux B, Dornhorst A. Diabetes of the exocrine pancreas. J Gastroenterol Hepatol 2019 Feb;34(2):346-354. [doi: 10.1111/jgh.14451] [Medline: 30151918]

Abbreviations

CENTRAL: Cochrane Central Register of Controlled Trials
CGM: continuous glucose monitoring
DM: diabetes mellitus
GDM: gestational diabetes mellitus
GRADE: Grading of Recommendations, Assessment, Development, and Evaluations
HbA_{1c}: hemoglobin A_{1c}
MeSH: Medical Subject Headings

mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROSPERO: International Prospective Register of Systematic Reviews
RCT: randomized clinical trial
T1DM: type 1 diabetes mellitus
T2DM: type 2 diabetes mellitus

Edited by L Buis; submitted 23.11.23; peer-reviewed by J Birks, K Tufts; revised version received 09.07.24; accepted 26.08.24; published 03.12.24.

Please cite as:
Diez Alvarez S, Fellas A, Wynne K, Santos D, Sculley D, Acharya S, Navathe P, Gironès X, Coda A
The Role of Smartwatch Technology in the Provision of Care for Type 1 or 2 Diabetes Mellitus or Gestational Diabetes: Systematic Review
JMIR Mhealth Uhealth 2024;12:e54826
URL: https://mhealth.jmir.org/2024/1/e54826
doi:10.2196/54826

Cross-Cutting mHealth Behavior Change Techniques to Support Treatment Adherence and Self-Management of Complex Medical Conditions: Systematic Review

Cyd K Eaton[1], PhD; Emma McWilliams[2], BA; Dana Yablon[3], BS; Irem Kesim[3], BS; Renee Ge[3], BS; Karissa Mirus[3], MHSA; Takeera Sconiers[3], BS; Alfred Donkoh[3], BS; Melanie Lawrence[4], BS; Cynthia George[5], MSN; Mary Leigh Morrison[4], MA; Emily Muther[6], PhD; Gabriela R Oates[7], PhD; Meghana Sathe[8], MD; Gregory S Sawicki[2], MD, MPH; Carolyn Snell[2], PhD; Kristin Riekert[3], PhD

[1]Division of General Pediatrics, Department of Pediatrics, Johns Hopkins School of Medicine, Baltimore, MD, United States

[2]Boston Children's Hospital, Harvard Medical School, Boston, MA, United States

[3]Division of Pulmonary and Critical Care Medicine, Department of Medicine, Johns Hopkins School of Medicine, Baltimore, MD, United States

[4]Success with Therapies Research Consortium CF Community Member Advisory Board, Bethesda, MD, United States

[5]Cystic Fibrosis Foundation, Bethesda, MD, United States

[6]Children's Hospital Colorado, University of Colorado School of Medicine, Aurora, CO, United States

[7]Division of Pediatric Pulmonary & Sleep Medicine, Preventive Medicine, University of Alabama at Birmingham, Birmingham, AL, United States

[8]Children's Health Dallas, University of Texas Southwestern Medical Center, Dallas, TX, United States

Corresponding Author:
Cyd K Eaton, PhD
Division of General Pediatrics, Department of Pediatrics, Johns Hopkins School of Medicine, , Baltimore, MD, , United States

Abstract

Background: Mobile health (mHealth) interventions have immense potential to support disease self-management for people with complex medical conditions following treatment regimens that involve taking medicine and other self-management activities. However, there is no consensus on what discrete behavior change techniques (BCTs) should be used in an effective adherence and self-management–promoting mHealth solution for any chronic illness. Reviewing the extant literature to identify effective, cross-cutting BCTs in mHealth interventions for adherence and self-management promotion could help accelerate the development, evaluation, and dissemination of behavior change interventions with potential generalizability across complex medical conditions.

Objective: This study aimed to identify cross-cutting, mHealth-based BCTs to incorporate into effective mHealth adherence and self-management interventions for people with complex medical conditions, by systematically reviewing the literature across chronic medical conditions with similar adherence and self-management demands.

Methods: A registered systematic review was conducted to identify published evaluations of mHealth adherence and self-management interventions for chronic medical conditions with complex adherence and self-management demands. The methodological characteristics and BCTs in each study were extracted using a standard data collection form.

Results: A total of 122 studies were reviewed; the majority involved people with type 2 diabetes (28/122, 23%), asthma (27/122, 22%), and type 1 diabetes (19/122, 16%). mHealth interventions rated as having a positive outcome on adherence and self-management used more BCTs (mean 4.95, SD 2.56) than interventions with no impact on outcomes (mean 3.57, SD 1.95) or those that used >1 outcome measure or analytic approach (mean 3.90, SD 1.93; *P*=.02). The following BCTs were associated with positive outcomes: self-monitoring outcomes of behavior (39/59, 66%), feedback on outcomes of behavior (34/59, 58%), self-monitoring of behavior (34/59, 58%), feedback on behavior (29/59, 49%), credible source (24/59, 41%), and goal setting (behavior; 14/59, 24%). In adult-only samples, prompts and cues were associated with positive outcomes (34/45, 76%). In adolescent and young adult samples, information about health consequences (1/4, 25%), problem-solving (1/4, 25%), and material reward (behavior; 2/4, 50%) were associated with positive outcomes. In interventions explicitly targeting medicine taking, prompts and cues (25/33, 76%) and credible source (13/33, 39%) were associated with positive outcomes. In interventions focused on self-management and other adherence targets, instruction on how to perform the behavior (8/26, 31%), goal setting (behavior; 8/26, 31%), and action planning (5/26, 19%) were associated with positive outcomes.

Conclusions: To support adherence and self-management in people with complex medical conditions, mHealth tools should purposefully incorporate effective and developmentally appropriate BCTs. A cross-cutting approach to BCT selection could accelerate the development of much-needed mHealth interventions for target populations, although mHealth intervention developers should continue to consider the unique needs of the target population when designing these tools.

Trial Registration: PROSPERO CRD42021224407; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=224407

(JMIR Mhealth Uhealth 2024;12:e49024) doi:10.2196/49024

KEYWORDS

cystic fibrosis; mobile health; technology; self-management; patient adherence; behavior intervention; mHealth intervention; systematic review; evaluation of mHealth; treatment adherence; mHealth

Introduction

Ever-advancing mobile health (mHealth) technologies hold immense potential to deliver behavior change techniques (BCTs) to diverse audiences, including people with complex medical conditions that involve treatment adherence and other self-management activities. mHealth refers to "medical and public health practice supported by mobile devices, such as mobile phones, patient monitoring devices, personal digital assistants, and other wireless devices" [1]. Common examples include sending smartphone notifications as medication reminders or recording in an app when treatments are completed. Prior mHealth reviews have broadly summarized mHealth interventions as "reminders, education, or behavioral" [2], which included a wide range of study outcomes beyond adherence or self-management [3] or limited the outcome to medication taking [4,5]. Therefore, existing reviews have a limited impact on exactly how mHealth can most effectively support adherence and disease self-management or can be adapted and tailored for chronic illnesses with complex regimens beyond simply taking medicine.

The BCT Taxonomy [6] was created to define discrete, cross-cutting techniques (or approaches) to changing behavior to facilitate the design and evaluation of behavior change interventions, as well as the comparison of BCTs across interventions to identify which BCTs are the most efficacious. The BCT Taxonomy is disease agnostic such that BCTs found to effectively improve treatment adherence and self-management in one complex medical condition should, in theory, generalize to other complex medical conditions with similar adherence and self-management demands. Reviewing mHealth interventions of diseases with complex adherence and self-management demands using BCT Taxonomy could accelerate the design of mHealth solutions by identifying "essential elements" of effective mHealth interventions. Unfortunately, there is no consensus on what essential features should be included in an adherence or self-management mHealth solution for any chronic medical condition.

Our group's interest in cross-cutting BCTs for adherence and self-management stems from our work with the cystic fibrosis (CF) community. CF is a rare, multisystemic medical condition affecting an estimated 162,428 people worldwide [7]. CF self-management is complex and typically involves a combination of daily oral medications, inhaled treatment, high calorie diet, chest physiotherapy, airway clearance, and exercise [8]. Not surprisingly, people with CF have demonstrated high rates of nonadherence across various aspects of the multicomponent treatment regimen, including low medication adherence (48%-68%) [9,10], nonadherence to caloric goals (24%-40%) [11], and low adherence to airway clearance therapy (28%) [12]. Effective behavioral interventions are needed to promote CF self-management and, in turn, support health outcomes and quality of life. However, rare diseases with complex regimens are rarely the target population for technology developers, and for almost a decade, people with CF have expressed interest in an app but noted that existing apps do not provide the necessary functionality to address their CF management needs [13-15]. A recent search of the Google Play Store (Android) and Apple App Store (iOS) for health-related apps found that only 29 (1.3%) out of 2272 apps address a rare disease population [16], including CF, with none having empirical evidence of their efficacy.

Recognizing that there is a dearth of empirical research on mHealth solutions for treatment adherence and self-management of CF and other rare diseases, we aimed to learn from the BCTs used in effective mHealth interventions for other chronic medical conditions with complex treatment adherence and self-management demands. We, therefore, purposefully designed our systematic review to include people with complex diseases and regimens with overlapping characteristics to CF. Our research questions were (1) Which BCTs have been used in mHealth interventions? and (2) Which BCTs have a positive impact on adherence and self-management behaviors? Differences in BCTs in adult-only studies compared to adolescent and young adult studies were examined, as well as interventions explicitly targeting medicine taking compared to studies targeting broader self-management and other areas of treatment adherence. A systematic review was used because heterogeneity in measuring adherence and self-management outcomes across studies precludes a meta-analysis [17-19] (in contrast to a systematic review, a meta-analysis involves statistically summarizing results across reviewed studies using effect sizes [20]). Our overarching goal was to identify the essential, cross-cutting BCTs delivered via mHealth to effectively facilitate long-term adherence and self-management for people with complex medical regimens, thereby accelerating intervention development, evaluation, and dissemination.

Methods

Overview

Standardized search strategies, eligibility evaluations, and data extraction procedures were used (detailed below and in Multimedia Appendix 1). This review was registered with the International Prospective Register of Systematic Reviews (PROSPERO CRD42021224407), in accordance with PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) guidelines (Checklist 1).

Ethical Considerations

As this was a systematic review, institutional review board approval was not required.

Search Strategy

A literature search in the PubMed, Scopus, Embase, CENTRAL, Web of Science, and PsycINFO databases identified potentially relevant articles published from 2015 through 2020, to enhance relevance to current technology. Given our group's focus on CF, 2 categories of search terms were used: "CF-specific" and "other chronic conditions," which included conditions identified by the study authors as having similar adherence and self-management characteristics to CF (eg, conditions with complex daily medical regimens and diseases often diagnosed in childhood, thus involving caregivers in self-management tasks).

Eligibility Criteria

Peer-reviewed, English language articles published between 2015 and 2020 reporting original empirical findings of mHealth interventions for selected medical conditions and targeting adherence and self-management were included. The mHealth interventions must be accessed on a mobile device (smartphones, cell phones, or tablets, including internet browser programs) and used by a person managing a medical condition or their caregiver.

Post Hoc Exclusions

After executing the search strategy, 3 post hoc exclusion criteria were added. People with chronic obstructive pulmonary disease or engaging in pulmonary rehabilitation were excluded, as it was decided that the former population was too different from people with CF and the latter included medical conditions. Reminder-only text messaging and exclusively synchronous telephone or web videoconferencing interventions were excluded, as our interest was in automated BCTs beyond simple reminders and interventions requiring real-time human interaction. Investigations conducted in low- to middle-income countries were excluded due to potential technology access limitations (unreliable internet or cellular service) that would likely affect the types of interventions tested.

Selection Process

Study records were compiled in a database; duplicates were removed based on DOI number or title. Reviewers (CKE, E McWilliams, DY, TS, and Brandi Blackshear) evaluated each study record (title and citation; blinded double review) for eligibility criteria. The reviewers screened studies for final inclusion and data abstraction using a REDCap (Research Electronic Data Capture; Vanderbilt University) [21,22] form developed for this study.

Data Collection

Overview

A reviewer independently abstracted the study data. A second reviewer read the article, reviewed the initial data abstraction, and identified items of disagreement. Discrepancies were discussed and resolved with all team members. Study characteristics were abstracted for each study in the final review (publication year, study location, study design, sample size, medical condition, age group, and theoretically derived intervention). Missing study details were noted.

Key Definitions

The adherence and self-management measurement method was abstracted. Reviewers categorized adherence and self-management measurement as (1) objective behavior (eg, electronic medication monitoring), (2) subjective behavior (eg, patient-reported medication adherence), (3) psychosocial outcome (eg, disease knowledge and adherence self-efficacy), (4) objective health outcome (eg, hemoglobin A_{1c} and viral load), or (5) subjective health outcome (eg, patient-reported asthma control level). Health outcomes were included if the authors conceptualized them as adherence and self-management indicators.

mHealth tools (eg, app and text messaging) and targets of intervention (eg, taking a specific medicine, airway clearance therapy, diabetes self-management activities, dietary recommendations, exercise and physical activity, managing disease activity and symptoms, etc) were abstracted. mHealth intervention results were categorized based on authors' conclusion of the results as follows:

- Positive: intervention was associated with improved adherence and self-management.
- Negative: intervention was associated with worse adherence and self-management.
- No impact: intervention had no effect on adherence and self-management.
- Mixed: intervention had different effects (positive effect, negative effect, or no impact) on adherence and self-management due to multiple outcome measures and analytic approaches.

The abstracter used information in the manuscript and the BCT Taxonomy to assign discrete BCTs to each intervention component.

Risk of Bias

The Revised Cochrane Risk-of-Bias tool (RoB 2) [23] for randomized controlled trials (RCTs) and the Risk of Bias in Non-randomized Studies-of Interventions (ROBINS-I) [24] tool for nonrandomized studies (excluding qualitative studies) were used to assess risk of bias, certainty, and quality of evidence among the studies reviewed. Blinded double assessments were conducted by 2 independent reviewers (RG, IK, E McWilliams, DY, and AD). The RoB-2 assessed risk of bias due to the randomization process, deviations from intended interventions, missing outcome data, measurement of outcome, and selection of the reported result. The ROBINS-I assessed risk of bias due to confounding, deviations from intended interventions, missing data, bias in measurement of outcomes, and bias in selection of the reported result. Discrepancies in ratings were identified and resolved. If multiple outcomes were assessed, an average risk score was calculated to derive a single rating.

Synthesis

Statistical analyses were conducted in Stata 15 software (StataCorp LLC). Abstracted data were summarized using

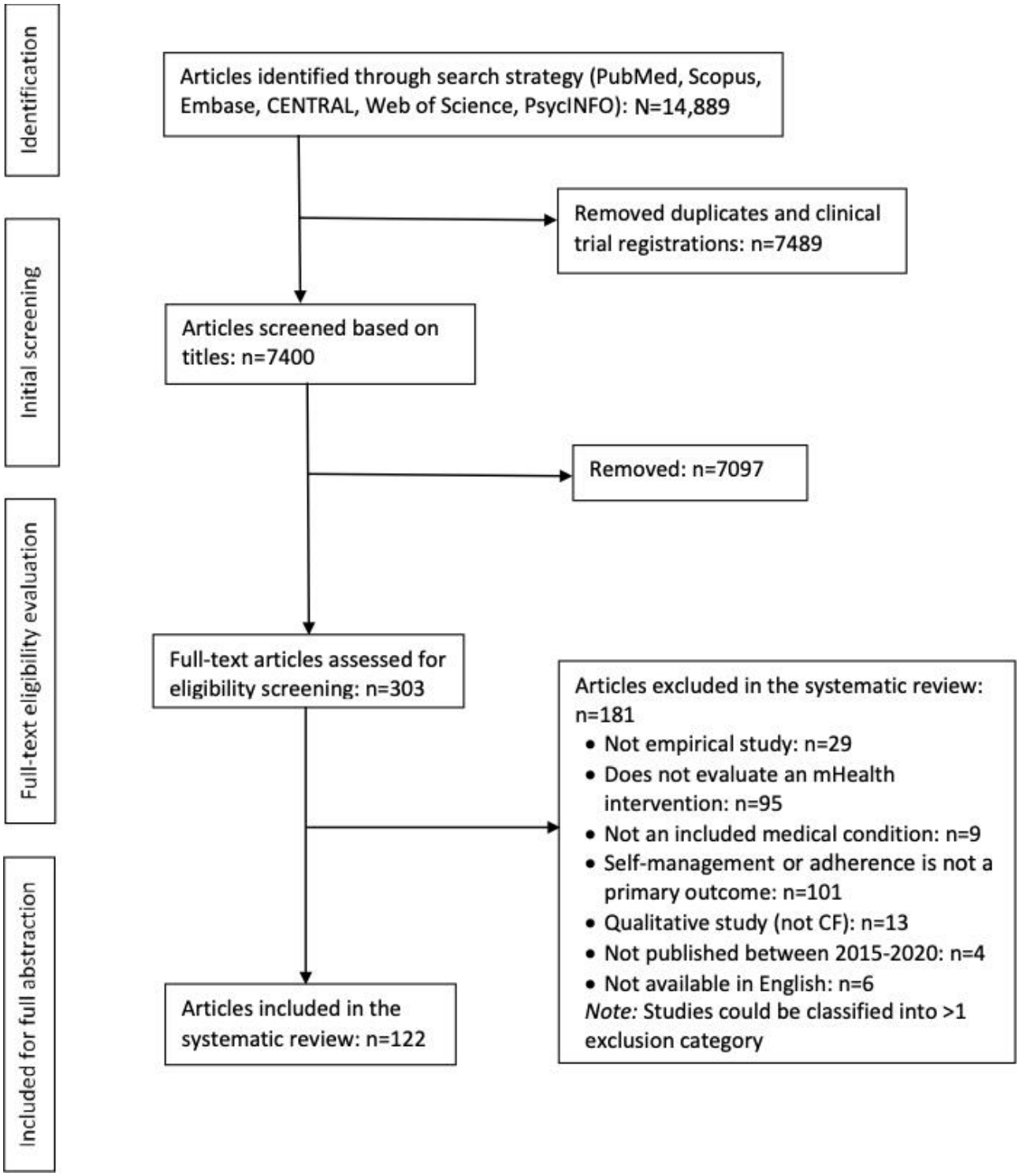
frequencies and percentages. Subgroup analyses examined differences in study characteristics, including BCTs used, age group (adult only [≥18 years and older] vs adolescent and young adult [11-25 years or sample characterized by authors as "adolescents and young adults"]), study design (RCT vs non-RCT), and whether the intervention was theoretically derived. We also conducted an exploratory subgroup analysis to examine which BCTs appeared the most often in interventions explicitly focusing on medicine taking compared to interventions focusing on self-management and other adherence targets. The results highlight BCTs (1) appearing in ≥5% of studies and (2) with a difference of >10% between positive effects versus no impact on adherence and self-management outcomes. This does not mean that rarely used BCTs are ineffective or that 10% is a verified benchmark of clinically meaningful difference. This pragmatic decision supported the interpretation of a large number of BCTs and comparisons. Statistically significant (P<.05) differences in the number of BCTs based on the direction of results were tested using 1-way ANOVA. No effect measures, missing summary statistics or data conversions, or meta-regression were used for this systematic review.

Results

Screening Process

Figure 1 presents this review's PRISMA diagram. The initial search returned 14,889 articles. After removing duplicates and clinical trial registrations, 7400 titles were screened for initial eligibility, 303 articles were potentially eligible, and 122 manuscripts met the criteria for data extraction (see Table S1 in Multimedia Appendix 1 for all included studies and characteristics).

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) diagram. CF: cystic fibrosis; mHealth: mobile health.



Study Characteristics

The most represented medical conditions were type 1 or 2 diabetes and asthma (Table 1). Only 6 (4.9%) out of 122 studies involved people with CF. Most studies were published in 2020 (32/122, 26.2%), were conducted outside of the United States (64/122, 52.5%), and used an RCT design (75/122, 61.5%). Nonrandomized study designs were primarily observational pre-post (23/122, 18.9%), observational without pre-post measurement (6/122, 4.9%), or mixed methods (6/122, 4.9%) studies. Study sample sizes ranged from 10 to 14,085 (median 92, IQR 44-179) participants. Most studies involved adult-only (81/122, 66.4%) or adolescent and young adult–only samples (22/122, 18%), followed by child, adolescent, young adult (11/122, 9%); child, adolescent, young adult, and adult (6/122, 4.9%); and child-only (2/122, 1.6%) samples.

Table . Study characteristics in all studies included for final abstraction (overall) and by direction of primary study results[a].

Study characteristic		Overall (N=122), n (%)	Positive (n=59), n (%)	No impact (n=23), n (%)	Mixed (n=40), n (%)
Study design					
	Randomized controlled trial	75 (61)	26 (44)	21 (91)	28 (70)
	Nonrandomized study	47 (39)	33 (56)	2 (9)	12 (30)
Use of theory in intervention design					
	Yes, theory guided intervention design	22 (18)	12 (20)	5 (22)	5 (13)
	Theory mentioned in manuscript, unclear if theory guided intervention design	15 (12)	4 (7)	7 (30)	4 (10)
	No underpinning theory mentioned in manuscript	85 (70)	44 (73)	11 (48)	30 (77)
Medical condition					
	Type 2 diabetes	28 (23)	18 (31)	5 (22)	5 (13)
	Asthma	27 (22)	13 (22)	4 (17)	10 (25)
	Type 1 diabetes	19 (16)	7 (12)	4 (17)	8 (20)
	HIV	11 (9)	3 (5)	4 (17)	4 (10)
	Multiple medical conditions	8 (7)	6 (10)	2 (9)	0 (0)
	Inflammatory bowel disease, Crohn disease, or ulcerative colitis	7 (6)	2 (3)	0 (0)	5 (13)
	Cystic fibrosis	6 (5)	3 (5)	1 (4)	2 (5)
	Solid organ transplant	5 (4)	2 (3)	2 (9)	1 (3)
	Diabetes (type not specified)	3 (2)	1 (2)	1 (4)	1 (3)
	Epilepsy	3 (2)	2 (3)	0 (0)	1 (3)
	Kidney disease	2 (2)	0 (0)	0 (0)	2 (5)
	Sickle cell disease	2 (2)	1 (2)	0 (0)	1 (3)
	Rheumatoid arthritis	1 (1)	1 (2)	0 (0)	0 (0)

[a]Medical conditions are listed from most frequently to least frequently observed in the overall study sample.

Adherence and self-management outcomes were typically evaluated with objective health outcomes or subjective behavior measures (61/122, 50% for each; Table S2 in Multimedia Appendix 1). mHealth interventions were delivered via app (75/122, 61.5%), SMS text messaging (34/122, 27.9%), or website (30/122, 24.6%). Nearly all studies presented mHealth tools used by patients (118/122, 96.7%), but many included health care providers (48/122, 39.3%) or caregivers (21/122, 17.2%). Only 22 (18%) interventions were clearly informed by scientific theory. mHealth interventions most often targeted taking medication (68/122, 55.7%), diabetes self-management activities (46/122, 37.7%), dietary recommendations (32/122, 26.2%), exercise and physical activity (27/122, 22.1%), asthma self-management activities (13/122, 10.7%), managing disease activity and symptoms (11/122, 9%), and general "self-care" behaviors (5/122, 4.1%). One (0.8%) study targeted airway clearance therapy.

Study results were characterized as having a positive effect on the outcome or outcomes (59/122, 48.4%), followed by mixed results (40/122, 32.8%) or no impact (23/122, 18.9%). No studies were characterized as having negative effects (Table 1). Comparing studies reporting positive effects to no impact, 34% (20/59) of the positive studies used objective behavior adherence measures compared to 13% (3/23) of no-impact studies (Table S2 in Multimedia Appendix 1).

BCTs Used

Across all reviewed studies, 32 different BCTs were used (mean 4.30, SD 2.32). Table S3 in Multimedia Appendix 1 provides

the frequencies, definitions, and examples of BCTs appearing in ≥5% of reviewed studies.

BCTs by Intervention Effect on Outcomes

Overview

Interventions with positive effects contained significantly more BCTs (mean 4.95, SD 2.56) than interventions with mixed effects (mean 3.90, SD 1.93) or no impact (mean 3.57, SD 1.95; *P*=.02). BCTs used in >10% of studies with positive results versus no impact (Multimedia Appendix 2) were self-monitoring of behavior, self-monitoring of outcomes of behavior, feedback on outcomes of behavior, feedback on behavior, credible source, and goal setting (behavior).

Subgroup Analysis: Age

Table S3 in Multimedia Appendix 1 includes the 15 most common BCTs used in adult-only and adolescent and young adult–only studies. Among adult-only studies (n=81), interventions with positive effects contained significantly more BCTs (mean 5.02, SD 2.13) than studies with mixed effects (mean 4.28, SD 2.27) or no impact (mean 3.28, SD 1.96; *P*=.02). BCTs used in >10% of studies with positive results compared to no impact included prompts and cues, self-monitoring of outcomes of behavior, feedback on outcomes of behavior, self-monitoring of behavior, feedback on behavior, credible source, and goal setting (behavior; Multimedia Appendix 3).

In adolescent and young adult–only studies (n=22), interventions with positive effects contained significantly more BCTs (mean 7.75, SD 5.91) than studies with mixed effects (mean 3.50, SD 1.67) or no impact (mean 5.00, SD 1.41; *P*=.04). BCTs used in >10% of studies with positive results compared to no impact included self-monitoring of behavior, feedback on behavior, goal setting (behavior), information about health consequences, problem-solving, and material reward (behavior; Multimedia Appendix 4).

Subgroup Analysis: Study Design and Theory

Non-RCT studies tended to report positive results (33/59, 56%), whereas RCT designs more commonly reported no impact (21/23, 91%) or mixed results (28/40, 70%). Theory rarely guided intervention design; a small proportion (12/59, 20%) of theory-informed interventions were shown to have a positive effect (Table 1).

Subgroup Analysis: Intervention Target

Table S4 in Multimedia Appendix 1 includes the 16 most common BCTs used in studies targeting medicine taking versus self-management and other adherence targets (appeared in >5% of studies). A total of 62% (42/68) of studies targeting medicine taking included people with diabetes (21/68, 31%) or asthma (21/68, 31%). A total of 80% (43/54) of studies targeting self-management and other adherence targets included people with diabetes (37/54, 69%) or asthma (6/54, 11%). There were no significant differences in the number of BCTs used in interventions targeting medicine taking (mean 4.07, SD 1.94) compared to interventions targeting self-management and other adherence targets (mean 4.69, SD 2.70; *P*=.16).

Within interventions explicitly targeting medicine taking (n=68), BCTs used in >10% of studies with positive results compared to no impact included prompts and cues, self-monitoring outcomes of behavior, self-monitoring of behavior, feedback on behavior, feedback on outcomes of behavior, and credible source (Multimedia Appendix 5). There were no significant differences in the number of BCTs used based on the direction of results (*P*=.06).

Within interventions focused on self-management and other adherence targets (n=54), BCTs used in >10% of studies with positive results compared to no impact included self-monitoring outcomes of behavior, feedback on outcomes of behavior, self-monitoring of behavior, feedback on behavior, instruction on how to perform the behavior, goal setting (behavior), and action planning (Multimedia Appendix 6). There were no significant differences in the number of BCTs used based on the direction of results (*P*=.21).

Risk of Bias

Table S1 in Multimedia Appendix 1 reports each study's risk of bias rating. No study was excluded due to bias rating. For RCTs, 57% (43/75) received an overall risk of bias rating of "Some concerns," 39% (29/75) had "High" concerns, and only 4% (3/75) had "Low" concerns. "High" concern ratings were generally due to deviations from the intended interventions (18/29, 62%), the randomization process (11/29, 38%), or missing outcome data (11/29, 38%). For nonrandomized studies, 82% (36/44) received an overall risk of bias rating of "Serious" concerns, 9% (4/44) had "Critical" concerns, and 2% (1/44) had "Moderate" concerns. No nonrandomized study had "Low" risk of bias. "Serious" or "Critical" ratings were generally due to confounding (38/40, 95%) or deviations from intended interventions (13/40, 33%).

Discussion

Principal Findings

Our literature review of mHealth adherence and self-management interventions returned 122 studies, from which we identified discrete behavioral strategies using the BCT Taxonomy [6] with promise to promote adherence and self-management for people living with medical conditions requiring complex, daily self-management activities. The BCT Taxonomy provides, to date, the most rigorously tested, standardized method to identify cross-cutting BCTs with potential applicability across chronic medical conditions with overlapping adherence and self-management demands. The BCT Taxonomy also helps compare mHealth interventions and provides a shared language about BCTs for clinicians, researchers, mHealth innovators, and other key stakeholders such as patients and caregivers. As technological advances can quickly outdate mHealth, focusing on BCT principles, rather than the technology to deliver them, enhances the research's relevance and potential generalizability to a range of complex medical conditions, including rare diseases (an area of focus for our group), which often have significant need for such tools in contrast to the finite resources available to conduct large-scale, multistep mHealth design and evaluation studies.

Consistent with prior research [25,26], using more BCTs was associated with improved adherence and self-management. However, 6 BCTs appear particularly promising: self-monitoring of behavior, self-monitoring of outcomes of behavior, feedback on behavior, feedback on outcomes of behavior, credible source, and goal setting. *Self-monitoring of behavior* and *outcomes of behavior* involve tracking health behavior engagement (eg, logging in an app when medicine is taken) or outcomes of behavior (eg, using a Bluetooth-enabled glucometer to monitor blood glucose levels), whereas *feedback on behavior* and *outcomes of behavior* involve providing users with a summarized interpretation of the tracked data (eg, providing in-app graphical representations of one's daily step count over the past month). Consistent with our results, a prior meta-analysis showed that monitoring medication adherence and providing feedback improve medication adherence [27]. These strategies may build awareness for when the mHealth user engages in a health behavior, provide opportunity to reflect on successes and challenges, and ultimately help the user make informed behavior changes. *Credible source* involves providing expert-generated information about managing the user's medical condition (eg, the app contains information about etiology, symptoms, and treatment), which presents users with knowledge to understand the condition and its management. *Goal setting (behavior*) involves setting measurable and attainable goals for a target health behavior (eg, set a goal for number of days to exercise in a month), which can help the mHealth user focus on key health behavior and build self-efficacy as goals are met.

Developmental differences emerged between adult samples and adolescents and young adult samples. In adult-only studies, *prompts and cues* (reminders) were associated with positive outcomes, consistent with reviews showing that reminders are associated with a 2- to 3-fold increase in adherence [28,29], but they were less effective in adolescent and young adult studies. Indeed, a pre-post study of children and adolescents with CF found that adherence did not change after delivering reminders only (therefore excluded from this review) for 6 months [30]. Adolescents and young adults may benefit from improving knowledge (*information about health consequences*), improving skills (*problem-solving*), and building motivation (*material reward* [*behavior*]). Given the small number of adolescent and young adult studies, these results and interpretations should be seen as hypothesis generating.

Differences emerged between interventions explicitly targeting medicine taking versus those focused on disease self-management and other adherence targets. In interventions targeting medicine taking, *prompts and cues* and *credible source* were associated with positive outcomes. Reminders and expert information may be the most effective when focused on discrete, clearly defined behaviors rather than complex, multicomponent self-management activities. In interventions focused on self-management and other adherence targets, *instruction on how to perform the behavior, goal setting (behavior)*, and *action planning* were associated with positive outcomes. Over three-quarters (43/54, 80%) of studies focused on self-management and other adherence targets involved people with diabetes or asthma, which are relatively common yet complex medical conditions involving self-management behaviors that extend beyond simply taking medicine. Skills training, behavioral goals, and assistance with creating a detailed plan for managing a complex medical condition may be the most effective for multicomponent self-management activities that may involve monitoring and intervening upon changes in disease activity (eg, managing fluctuations in blood glucose levels for people with diabetes or managing asthma exacerbations) and self-managing lifestyle and environmental considerations (eg, diet in diabetes and environmental triggers in asthma). Careful consideration of the intervention target will likely help to further guide appropriate BCT selection from the BCTs found to be associated with improved adherence and self-management in our review.

This review has limitations. A meta-analysis was not conducted due to heterogeneous outcomes [17-19], thus we could not conclude which BCTs were statistically the most effective. Our risk-of-bias assessment highlighted methodological concerns across the studies reviewed. No-impact studies were more likely to be RCTs, and positive studies were more likely to be nonrandomized, raising concerns about publication bias toward positive results irrespective of study quality. We excluded reminder-only interventions; thus, most studies incorporated more than 1 BCT. Our reported average number of BCTs is likely higher than that of all adherence-promoting mHealth interventions. Although we identified some BCTs that may be effective, others may be as or more effective in supporting disease self-management but were rarely used in the reviewed studies. Moreover, no BCT was found to do harm. Thus, mHealth innovators should continue to integrate and evaluate how a wide variety of technology-delivered BCTs may support people living with chronic diseases, including rare diseases such as CF. An inherent limitation of conducting literature reviews is that a cutoff date must be selected, yet scientific literature is constantly being published; there may be utility in conducting an updated systematic review of this topic in the future. We only included studies that were published in peer-reviewed journals to focus on interventions with clear evidence of scientific evaluation; however, expanding our review to "gray literature" may have provided more insight into the most current interventions and reduced publication bias. Our review characterized mHealth BCTs generally. Other metrics including digital literacy and socioeconomic barriers to mHealth were not evaluated. Future researchers should evaluate these factors to support sustained mHealth use among diverse audiences. Additionally, BCTs were analyzed across the included chronic medical conditions given the disproportionate number of studies in diabetes and asthma compared to other medical conditions. Although the BCTs are disease agnostic, intervention developers and researchers should carefully consider the applicability of the BCT to the target patient population.

Future Directions

Our review identified discrete BCTs that may have broad cross-cutting applicability across chronic diseases with complex medical regimens, including people with CF, the community with which our group primarily works with. We consider our systematic review approach to be a model for gathering key findings from the extant scientific literature to inform the development of multicomponent behavioral mHealth

interventions tailored for a patient population that may be smaller and with less existing research, yet has significant self-management needs warranting further research, such as CF [9,10,31,32]. Research involving people with chronic medical conditions following complex treatment regimens should prioritize the design and evaluation of mHealth interventions incorporating cross-cutting, evidence-based, and age-appropriate BCTs to promote adherence and self-management. Such an approach could help accelerate mHealth intervention design and evaluation to create effective products that may be efficiently disseminated to communities with significant need for such tools.

Accelerating mHealth design and evaluation by taking a cross-cutting approach to BCT selection would also help answer remaining "unknowns" about mHealth BCTs and strengthen mHealth intervention quality. For example, although interventions including more BCTs appear to have greater benefit, the optimal number, type, and combination of BCTs to include in mHealth interventions have not been determined. For BCTs demonstrating potential to promote adherence or self-management, the ideal delivery method must be determined (eg, should the BCT *self-monitoring of behavior* be delivered via manual data entry of treatment completion or using an electronic monitoring device to automatically track data?). An overrepresentation of certain BCTs (eg, *prompts and cues* and *self-monitoring of outcomes of behavior*) and underuse of other, potentially more effective techniques (eg, *feedback on behavior* and *goal setting* [*behavior*]) highlight mHealth's focus on simpler technologies at the expense of innovation and efficacy. Collaborations between behavioral scientists, care teams, patients, caregivers, and industry could answer these questions and produce mHealth solutions that are transformative and effective.

When incorporating BCTs that are expected to effectively and appropriately generalize to a range of complex medical conditions and associated regimens, mHealth intervention developers must still consider the unique needs of the target population. In CF, for example, highly effect CF transmembrane conductance regulator modulator therapies have the potential to simplify the regimen and reduce treatment burden [33,34]. The implementation of key BCTs may need to be adapted as new therapies roll out, although the core theory behind the BCT itself is not expected to change. It is critical to build the scientific evidence base for effective adherence and self-management mHealth interventions that maintain pace with rapidly advancing medical management across complex medical conditions.

Acknowledgments

We are thankful for the contributions of Brandi Blackshear, Angela Green, Kirsten Kulik, and Anne Bowen to this important investigation.

Conflicts of Interest

CKE, KR, CS, E McWilliams, MS, KM, E Muther, GRO, TS, and DY receive or received salary and/or grant support from the Cystic Fibrosis Foundation's Success with Therapies Research Consortium. GSS and KR receive honoraria and/or speaker fees from Vertex Pharmaceuticals. CG is an employee of the Cystic Fibrosis Foundation, which funded this research through the Success with Therapies Research Consortium. GRO receives grants from the Kael Pediatric Research Institute, National Institutes of Health, Health Resources and Services Administration's Children's National Research Institute, and the Alabama Department of Public Health, in addition to grant support from the Cystic Fibrosis Foundation. She also receives consulting fees from International Biophysics Corporation. CS serves as an unpaid advisor and consultant to MMNTS, Inc in addition to receiving salary support from the Cystic Fibrosis Foundation's Success with Therapies Research Consortium. MS receives grant support from Cystic Fibrosis Foundation and Anagram Therapeutics, Inc. AD, RG, IK, MLM, and ML declare no conflicts of interest.

Multimedia Appendix 1
Additional details on search and screening strategies, study characteristics, adherence measure types, behavior change technique definitions, and behavior change techniques by intervention target.
[DOCX File, 130 KB - mhealth_v12i1e49024_app1.docx]

Multimedia Appendix 2
The most common BCTs (>5% of all abstracted studies) by the direction of results in the overall sample. BCTs with >10% difference in how often they appear in studies with positive results compared to no-impact results are highlighted in red boxes. BCT: behavior change technique.
[PNG File, 105 KB - mhealth_v12i1e49024_app2.png]

Multimedia Appendix 3
The most common BCTs (>5% of all abstracted studies) by the direction of results in adult-only samples. BCTs with >10% difference in how often they appear in studies with positive results compared to no-impact results are highlighted in red boxes. BCT: behavior change technique.
[PNG File, 120 KB - mhealth_v12i1e49024_app3.png]

Multimedia Appendix 4
The most common BCTs (>5% of all abstracted studies) by the direction of results in adolescent and young adult samples. BCTs with >10% difference in how often they appear in studies with positive results compared to no-impact results are highlighted in red boxes. BCT: behavior change technique.
[PNG File, 125 KB - mhealth_v12i1e49024_app4.png]

Multimedia Appendix 5
The most common BCTs (>5% of all abstracted studies) by the direction of results in interventions explicitly targeting medicine taking. BCTs with >10% difference in how often they appear in studies with positive results compared to no-impact results are highlighted in red boxes. BCT: behavior change technique.
[PNG File, 100 KB - mhealth_v12i1e49024_app5.png]

Multimedia Appendix 6
The most common BCTs (>5% of all abstracted studies) by the direction of results in interventions targeting self-management or other adherence targets. BCTs with >10% difference in how often they appear in studies with positive results compared to no-impact results are highlighted in red boxes. BCT: behavior change technique.
[PNG File, 204 KB - mhealth_v12i1e49024_app6.png]

Checklist 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) 2020 checklist.
[PDF File, 63 KB - mhealth_v12i1e49024_app7.pdf]

References

1. WHO Global Observatory for eHealth. mHealth: new horizons for health through mobile technologies: second global survey on eHealth. World Health Organization. 2011. URL: https://iris.who.int/handle/10665/44607 [accessed 2023-03-23]
2. Ahmed I, Ahmad NS, Ali S, et al. Medication adherence apps: review and content analysis. JMIR Mhealth Uhealth 2018 Mar 16;6(3):e62. [doi: 10.2196/mhealth.6432] [Medline: 29549075]
3. Aguiar M, Trujillo M, Chaves D, Álvarez R, Epelde G. mHealth apps using behavior change techniques to self-report data: systematic review. JMIR Mhealth Uhealth 2022 Sep 9;10(9):e33247. [doi: 10.2196/33247] [Medline: 36083606]
4. Carmody JK, Denson LA, Hommel KA. Content and usability evaluation of medication adherence mobile applications for use in pediatrics. J Pediatr Psychol 2019 Apr 1;44(3):333-342. [doi: 10.1093/jpepsy/jsy086] [Medline: 30358863]
5. Morrissey EC, Corbett TK, Walsh JC, Molloy GJ. Behavior change techniques in apps for medication adherence: a content analysis. Am J Prev Med 2016 May;50(5):e143-e146. [doi: 10.1016/j.amepre.2015.09.034] [Medline: 26597504]
6. Michie S, Richardson M, Johnston M, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
7. Guo J, Garratt A, Hill A. Worldwide rates of diagnosis and effective treatment for cystic fibrosis. J Cyst Fibros 2022 May;21(3):456-462. [doi: 10.1016/j.jcf.2022.01.009] [Medline: 35125294]
8. Bishay LC, Sawicki GS. Strategies to optimize treatment adherence in adolescent patients with cystic fibrosis. Adolesc Health Med Ther 2016 Oct 21;7:117-124. [doi: 10.2147/AHMT.S95637] [Medline: 27799838]
9. Eakin MN, Bilderback A, Boyle MP, Mogayzel PJ, Riekert KA. Longitudinal association between medication adherence and lung health in people with cystic fibrosis. J Cyst Fibros 2011 Jul;10(4):258-264. [doi: 10.1016/j.jcf.2011.03.005] [Medline: 21458391]
10. Quittner AL, Zhang J, Marynchenko M, et al. Pulmonary medication adherence and health-care use in cystic fibrosis. Chest 2014 Jul;146(1):142-151. [doi: 10.1378/chest.13-1926] [Medline: 24480974]
11. Simon SL, Duncan CL, Horky SC, Nick TG, Castro MM, Riekert KA. Body satisfaction, nutritional adherence, and quality of life in youth with cystic fibrosis. Pediatr Pulmonol 2011 Nov;46(11):1085-1092. [doi: 10.1002/ppul.21477] [Medline: 21626713]
12. Oates GR, Stepanikova I, Gamble S, Gutierrez HH, Harris WT. Adherence to airway clearance therapy in pediatric cystic fibrosis: socioeconomic factors and respiratory outcomes. Pediatr Pulmonol 2015 Dec;50(12):1244-1252. [doi: 10.1002/ppul.23317] [Medline: 26436321]
13. Floch J, Zettl A, Fricke L, et al. User needs in the development of a health app ecosystem for self-management of cystic fibrosis: user-centered development approach. JMIR Mhealth Uhealth 2018 May 8;6(5):e113. [doi: 10.2196/mhealth.8236] [Medline: 29739742]
14. Rutland SB, Bergquist RP, Hager A, et al. A mobile health platform for self-management of pediatric cystic fibrosis: qualitative study of adaptation to stakeholder needs and integration in clinical settings. JMIR Form Res 2021 Jan 26;5(1):e19413. [doi: 10.2196/19413] [Medline: 33496667]

15. Hilliard ME, Hahn A, Ridge AK, Eakin MN, Riekert KA. User preferences and design recommendations for an mHealth app to promote cystic fibrosis self-management. JMIR Mhealth Uhealth 2014 Oct 24;2(4):e44. [doi: 10.2196/mhealth.3599] [Medline: 25344616]
16. Hatem S, Long JC, Best S, Fehlberg Z, Nic Giolla Easpaig B, Braithwaite J. Mobile apps for people with rare diseases: review and quality assessment using Mobile App Rating Scale. J Med Internet Res 2022 Jul 26;24(7):e36691. [doi: 10.2196/36691] [Medline: 35881435]
17. Nieuwlaat R, Wilczynski N, Navarro T, et al. Interventions for enhancing medication adherence. Cochrane Database Syst Rev 2014 Nov 20;2014(11):CD000011. [doi: 10.1002/14651858.CD000011.pub4] [Medline: 25412402]
18. Mellon L, Doyle F, Hickey A, et al. Interventions for increasing immunosuppressant medication adherence in solid organ transplant recipients. Cochrane Database Syst Rev 2022 Sep 12;9(9):CD012854. [doi: 10.1002/14651858.CD012854.pub2] [Medline: 36094829]
19. Al-Aqeel S, Gershuni O, Al-Sabhan J, Hiligsmann M. Strategies for improving adherence to antiepileptic drug treatment in people with epilepsy. Cochrane Database Syst Rev 2020 Oct 22;10(10):CD008312. [doi: 10.1002/14651858.CD008312.pub4] [Medline: 33089492]
20. Hedges LV, Tipton E. Meta-analysis. In: Steptoe A, editor. Handbook of Behavioral Medicine: Springer; 2010:909-921. [doi: 10.1007/978-0-387-09488-5_58]
21. Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208. [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
22. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381. [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
23. Sterne JAC, Savović J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019 Aug 28:l4898. [doi: 10.1136/bmj.l4898] [Medline: 31462531]
24. Sterne JA, Hernán MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ 2016 Oct 12:i4919. [doi: 10.1136/bmj.i4919] [Medline: 27733354]
25. van Rhoon L, Byrne M, Morrissey E, Murphy J, McSharry J. A systematic review of the behaviour change techniques and digital features in technology-driven type 2 diabetes prevention interventions. Digit Health 2020 Mar 24;6:2055207620914427. [doi: 10.1177/2055207620914427] [Medline: 32269830]
26. Bond Z, Scanlon T, Judah G. Systematic review of RCTs assessing the effectiveness of mHealth interventions to improve statin medication adherence: using the behaviour-change technique taxonomy to identify the techniques that improve adherence. Healthcare (Basel) 2021 Sep 28;9(10):1282. [doi: 10.3390/healthcare9101282] [Medline: 34682962]
27. Demonceau J, Ruppar T, Kristanto P, et al. Identification and assessment of adherence-enhancing interventions in studies assessing medication adherence through electronically compiled drug dosing histories: a systematic literature review and meta-analysis. Drugs 2013 May;73(6):545-562. [doi: 10.1007/s40265-013-0041-3] [Medline: 23588595]
28. Zhao YY, Dang FP, Zhai TT, Li HJ, Wang RJ, Ren JJ. The effect of text message reminders on medication adherence among patients with coronary heart disease: a systematic review and meta-analysis. Medicine (Baltimore) 2019 Dec;98(52):e18353. [doi: 10.1097/MD.0000000000018353] [Medline: 31876709]
29. Salisbury KR, Ranpariya VK, Feldman SR. Accountability in reminder-based adherence interventions: a review. Patient Educ Couns 2022 Aug;105(8):2645-2652. [doi: 10.1016/j.pec.2021.12.009] [Medline: 34953618]
30. Morton RW, Elphick HE, Edwards E, Daw WJ, West NS. Investigating the feasibility of text message reminders to improve adherence to nebulized medication in children and adolescents with cystic fibrosis. Patient Prefer Adherence 2017 May 8;11:861-869. [doi: 10.2147/PPA.S123723] [Medline: 28507430]
31. Sawicki GS, Sellers DE, Robinson WM. High treatment burden in adults with cystic fibrosis: challenges to disease self-management. J Cyst Fibros 2009 Mar;8(2):91-96. [doi: 10.1016/j.jcf.2008.09.007] [Medline: 18952504]
32. Davies G, Rowbotham NJ, Smith S, et al. Characterising burden of treatment in cystic fibrosis to identify priority areas for clinical trials. J Cyst Fibros 2020 May;19(3):499-502. [doi: 10.1016/j.jcf.2019.10.025] [Medline: 31735561]
33. Schechter MS, Sabater-Anaya N, Oster G, et al. Impact of elexacaftor/tezacaftor/ivacaftor on healthcare resource utilization and associated costs among people with cystic fibrosis in the US: a retrospective claims analysis. Pulm Ther 2023 Dec;9(4):479-498. [doi: 10.1007/s41030-023-00241-z] [Medline: 37874528]
34. Wajda KE, Roesch EA, Gifford AH. Chronic daily respiratory care needs in people with cystic fibrosis treated with highly effective cystic fibrosis transmembrane conductance regulator modulators. Curr Opin Pulm Med 2023 Nov 1;29(6):580-586. [doi: 10.1097/MCP.0000000000001006] [Medline: 37611027]

Abbreviations

BCT: behavior change technique
CF: cystic fibrosis
mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analysis

PROSPERO: International Prospective Register of Systematic Reviews
RCT: randomized controlled trial
REDCap: Research Electronic Data Capture
RoB 2: Revised Cochrane Risk-of-Bias tool
ROBINS-I: Risk of Bias in Non-randomized Studies-of Interventions

Edited by L Buis; submitted 26.05.23; peer-reviewed by C Bossley, UK Chalwadi; revised version received 26.01.24; accepted 26.02.24; published 01.05.24.

Please cite as:
Eaton CK, McWilliams E, Yablon D, Kesim I, Ge R, Mirus K, Sconiers T, Donkoh A, Lawrence M, George C, Morrison ML, Muther E, Oates GR, Sathe M, Sawicki GS, Snell C, Riekert K
Cross-Cutting mHealth Behavior Change Techniques to Support Treatment Adherence and Self-Management of Complex Medical Conditions: Systematic Review
JMIR Mhealth Uhealth 2024;12:e49024
URL: https://mhealth.jmir.org/2024/1/e49024
doi:10.2196/49024

mHealth-Based Gamification Interventions Among Men Who Have Sex With Men in the HIV Prevention and Care Continuum: Systematic Review and Meta-Analysis

Qianqian Luo[1], PhD; Yue Zhang[1], BA; Wei Wang[2], MA; Tianyu Cui[1], BA; Tianying Li[1], BA

[1]School of Nursing, Binzhou Medical University, Yantai, China

[2]Department of Nursing, The People's Hopstial of Laoling City, Dezhou, China

Corresponding Author:
Qianqian Luo, PhD
School of Nursing, Binzhou Medical University, , Yantai, , China

Abstract

Background: In the past few years, a burgeoning interest has emerged in applying gamification to promote desired health behaviors. However, little is known about the effectiveness of such applications in the HIV prevention and care continuum among men who have sex with men (MSM).

Objective: This study aims to summarize and evaluate research on the effectiveness of gamification on the HIV prevention and care continuum, including HIV-testing promotion; condomless anal sex (CAS) reduction; and uptake of and adherence to pre-exposure prophylaxis (PrEP), postexposure prophylaxis (PEP), and antiretroviral therapy (ART).

Methods: We comprehensively searched PubMed, Embase, the Cochrane Library, Web of Science, Scopus, and the *Journal of Medical Internet Research* and its sister journals for studies published in English and Chinese from inception to January 2024. Eligible studies were included when they used gamified interventions with an active or inactive control group and assessed at least one of the following outcomes: HIV testing; CAS; and uptake of and adherence to PrEP, PEP, and ART. During the meta-analysis, a random-effects model was applied. Two reviewers independently assessed the quality and risk of bias of each included study.

Results: The systematic review identified 26 studies, including 10 randomized controlled trials (RCTs). The results indicated that gamified digital interventions had been applied to various HIV outcomes, such as HIV testing, CAS, PrEP uptake and adherence, PEP uptake, and ART adherence. Most of the studies were conducted in the United States (n=19, 73%). The most frequently used game component was gaining points, followed by challenges. The meta-analysis showed gamification interventions could reduce the number of CAS acts at the 3-month follow-up (n=2 RCTs; incidence rate ratio 0.62, 95% CI 0.44-0.88). The meta-analysis also suggested an effective but nonstatistically significant effect of PrEP adherence at the 3-month follow-up (n=3 RCTs; risk ratio 1.16, 95% CI 0.96-1.38) and 6-month follow-up (n=4 RCTs; risk ratio 1.28, 95% CI 0.89-1.84). Only 1 pilot RCT was designed to evaluate the effectiveness of a gamified app in promoting HIV testing and PrEP uptake. No RCT was conducted to evaluate the effect of the gamified digital intervention on PEP uptake and adherence, and ART initiation among MSM.

Conclusions: Our findings suggest the short-term effect of gamified digital interventions on lowering the number of CAS acts in MSM. Further well-powered studies are still needed to evaluate the effect of the gamified digital intervention on HIV testing, PrEP uptake, PEP initiation and adherence, and ART initiation in MSM.

Trial Registration: PROSPERO CRD42023392193; https://www.crd.york.ac.uk/PROSPERO/display_record.php?RecordID=392193

(JMIR Mhealth Uhealth 2024;12:e49509) doi:10.2196/49509

KEYWORDS

mHealth; gamification; HIV; men who have sex with men; meta-analysis; PRISMA; mobile health; Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Introduction

Background

The HIV epidemic among men who have sex with men (MSM) has become a global concern [1]. Current evidence suggests that MSM accounted for two-thirds of all new HIV infections in the United States in 2019 [2], and systematic reviews showed that the HIV prevalence in MSM in China increased from 1.4% in 2001 to 8% in 2015 [3,4]. Given the burden of HIV, substantial efforts have been made to address the unmet needs of MSM. For instance, there is overwhelming scientific evidence of the efficacy and safety of pre-exposure prophylaxis (PrEP) and postexposure prophylaxis (PEP) to prevent HIV infection in MSM. However, studies indicate that uptake of PrEP and PEP has been suboptimal in these risk groups, especially among racial or ethnic minorities [5]. In a recent systematic review and meta-analysis, the PEP uptake was only between 4% and 6% in MSM; major obstacles included insufficient knowledge, underestimated risk of exposure to HIV, and social stigma [6]. Interventions that improve effective HIV prevention adoption among MSM are urgently needed to maximize HIV prevention benefits.

Nearly universal mobile phone ownership provides MSM an opportunity to move away from traditional ways of meeting partners to seeking sexual partners through geo–social networking applications [7]. The widespread use of smartphones also creates a unique opportunity to design innovative internet-based digital interventions for diverse MSM in a scalable manner. Internet-based digital interventions, defined as using internet-based information and communication technology (eg, mobile apps, websites, and social media) to support health [8], benefit from being implemented at a large scale with low costs and can deliver health services at the time and place chosen by users [9]. However, previous research showed that user engagement in digital interventions is suboptimal (due to, for example, high participant attrition) [10], resulting in a significant gap in intervention efficacy.

Gamification, which first emerged in 2008 and gained popularity since the 2010s, refers to applying game components such as leaderboards, points, and badges into nongame contexts [11]. Unlike serious games, which refer to full-fledged video games for health purposes, gamification is relatively open to varying situational engagement models [12]. Previous studies have been widely conducted on the use of gamification as a means to increase the initiation of desired health behaviors [13,14]. Given its increasing use in public health, gamification might also be useful for interventions to promote HIV prevention and control services. To our knowledge, two previous reviews explored how gamification was used during the HIV prevention and care continuum; the first review was conducted in 2017 and only summarized studies published between January 2016 and March 2017 [15], and the second review searched studies published in the *Journal of Medical Internet Research* and its sister journals [16], both of which would result in an incomplete study search. Moreover, to our knowledge, no meta-analysis has hitherto examined the effectiveness of gamification applied to HIV prevention and control. Therefore, it is essential to conduct an in-depth review and provide a meta-analysis combining evidence on the effectiveness of gamified HIV digital prevention interventions. Findings from this review may have important implications for HIV digital health prevention development and future research.

Objectives

This study was divided into two parts. The first part was a systematic review of HIV prevention and control gamification with the following aims: to describe the characteristics of included studies that applied gamification, describe the various gamification elements that are commonly applied, and evaluate the methodological quality of included studies. The second part is a meta-analysis of the effectiveness of gamification applied to HIV prevention and control to assess the impact of gamification on HIV testing; condomless anal sex (CAS) reduction; and uptake of and adherence to PrEP, nonoccupational PEP (nPEP), and antiretroviral therapy (ART). It was hypothesized that MSM participating in the gamified intervention would display higher levels of HIV prevention and care than those in the control groups.

Methods

We conducted this systematic review and meta-analysis according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 guideline [17]. The PRISMA checklist is listed in Checklist 1, and the study was registered in PROSPERO on January 27, 2023 (CRD42023392193).

Search Strategy and Identification of Studies

The electronic databases PubMed/MEDLINE, Embase, the Cochrane Library, Web of Science, Scopus, and the *Journal of Medical Internet Research* and its sister journals were searched for scientific articles published from their inception until January 15, 2024, using a combination of Medical Subject Headings (MeSH) terms and text words: *Men who have sex with men* AND *HIV/AIDS* AND *gamification/game-based learning* AND *telemedicine*. The search strategy used for each database is listed in Multimedia Appendix 1 .

Study Selection

The full review screening was conducted by two independent reviewers who independently checked the titles and abstracts for inclusion in the review. Full-text articles were obtained for closer inspection when an article met the inclusion criteria. In addition, reference lists of retrieved articles, existing relevant systematic reviews, and all articles citing the included studies on Google Scholar were manually searched to allocate studies not identified in the database searches. Any disagreement on study selection was resolved by discussion with a third reviewer.

To be eligible for inclusion, studies were required to describe or evaluate gamified interventions for HIV prevention and control in MSM. For the systematic review, the inclusion criteria were peer-reviewed original articles that explicitly addressed the use of game elements or gamification; the intervention described involved a task specifically designed for HIV prevention and control, and at least one game element was

involved in the task; the task was delivered via a digital device (eg, smartphones, tablets, or laptops); the study population should be exclusively MSM or more than 50% of the participants should be MSM; and the primary outcome should contain at least one of the following outcomes related to HIV prevention and control, namely, HIV testing, CAS, and uptake of and adherence to PrEP, nPEP, and ART for MSM living with HIV. Given the limited number of published randomized controlled trials (RCTs), we included a heterogeneous range of study designs, including controlled pre-post studies, quasi-randomized studies, cohort studies, and case-control studies. If a study included both a protocol and efficacy study, we only included the efficacy study; otherwise, we included the study protocol. The studies included in the meta-analysis represented a subset of studies in the systematic review. An additional inclusion criterion for the meta-analysis was that the study design should be an RCT. The definition of gamification by Deterding et al [18] was adopted for this review, which considers gamification as an umbrella term for using game elements (rather than full-fledged games) to improve user experience and user engagement in nongame services and applications. Studies were excluded if they were not peer-reviewed articles including letters, commentary, conference abstracts, etc; were not written in English or Chinese; were systematic reviews or meta-analyses; the full text was not available in published form; did not include a well-described intervention; exclusively used text-based technology; or involved serious or full-fledged games.

Study Quality Assessment

Two reviewers independently assessed the quality and risk of bias of each included RCT using the revised Cochrane Risk of Bias Tool (RoB2) [19,20], which contains 5 domains, namely, bias arising from the randomization process, bias due to deviations from the intended intervention, bias due to missing outcome data, bias in outcome measurement, and bias in the selection of the reported result. The risk of bias in each domain was judged as *low risk*, *some concerns*, or *high risk*. We used the Risk of Bias in Nonrandomized Studies of Interventions (ROBINS-I) tool to assess the risk of bias in nonrandomized studies of interventions (cohort studies, case-control studies, controlled pre-post studies, and quasi-randomized studies), which contains 7 domains, and the risk of bias in each domain was judged as *low risk*, *moderate risk*, *serious risk*, *critical risk*, or *no information* [21]. Any disputes between the two reviewers were resolved through discussion with a third author. We summarized the quality evaluation results with a risk-of-bias plot by *robvis* [22]. The average score between the two authors was calculated for each study, and the weighted Cohen κ coefficient was measured to test the interrater reliability.

Data Extraction

A structured data extraction form was developed and revised by the authors. The extracted data comprised study characteristics, participants, intervention characteristics, comparison activity, and outcomes. The study characteristics included first author, publication year, country, study design, and study setting. Participant characteristics included participant recruitment method, recruitment period, sample size, follow-up times, mean and SD of participants' age, and HIV status. For intervention characteristics, we extracted intervention developers, intervention names if available, the number of modules in each intervention, intervention delivery modality, game components used in each intervention, and theories used to apply gamification. The primary outcomes were CAS reduction; HIV testing; and uptake of and adherence to PrEP, nPEP, and ART. Authors of the initial publications were contacted when additional information was needed. Because of multiple classifications proposed for the term gamification, the components in each digital gamified intervention were analyzed through the lens of the Octalysis gamification framework [23], which comprises 8 core drives into which various gamification elements can be placed, namely, epic meaning and calling, development and accomplishment, empowerment creativity and feedback, ownership and possession, social influence and relatedness, scarcity and impatience, unpredictability and curiosity, and loss and avoidance (Table S3-2 in Multimedia Appendix 2 [13,24-48]). Laine et al [49] mapped their interview results from schoolchildren to the Octalysis framework to better establish a digital gamification-based intervention for promoting active school transport.

Data Synthesis and Analysis

First, data were synthesized and summarized in a narrative form assessing the characters of included studies and game components in each intervention. This qualitative review consolidated all studies that met the eligibility criteria, including those for which we could not extract data. Second, a meta-analysis was conducted to evaluate the effectiveness of gamification applied to HIV prevention and care, and variables that mediated these relationships. For each trial, we estimated the risk ratio (RR) of HIV testing; CAS; and uptake of and adherence to PrEP, PEP, and ART, comparing the intervention group and the control group together with the SE of the log RR. Random-effects summary effect estimates were obtained from DerSimonian and Laird's [50] random-effects meta-analyses of (log) RRs and 95% CIs from each study. When studies included more than one intervention group with gamification features, they were first combined into one group following the recommendation by the Cochrane Handbook. Studies with multiple control groups were integrated into different subgroup analyses if they compared their gamified interventions to inactive and active control groups. For studies reporting outcomes at more than 1 time point, we abstracted data in each time point after randomization. Higgins I^2 and 95% CI were estimated to measure between-study heterogeneity (low heterogeneity: <25%; moderate heterogeneity: 25%-75%; large heterogeneity: >75%) [51]. When heterogeneity was detected by I^2 (P<.05), a subgroup analysis was performed to investigate the possible source of heterogeneity: study designs, type of control group, selected samples (eg, participants with different risk levels) versus unselected samples, intervention duration, and outcome diagnosis (clinical diagnosis vs self-reporting). A meta-regression analysis was conducted if the number of RCTs for each outcome of interest was >10 [52]. Potential publication bias was explored by the Egger test and contour-enhanced funnel plots. If publication bias was present, nonparametric trim-and-fill analysis was performed to explore its impact on the

meta-analysis results. All *P* values were 2-sided, and *P*<.05 was considered statistically significant. All analyses were conducted using Stata 15 (Stata Corp) using “metan” commands.

Role of Funding Source

The study funder had no role in the study design, data collection, data analysis, and result interpretation. The corresponding author had full access to all study data and held the final responsibility for the decision to submit the manuscript for publication.

Ethics Approval

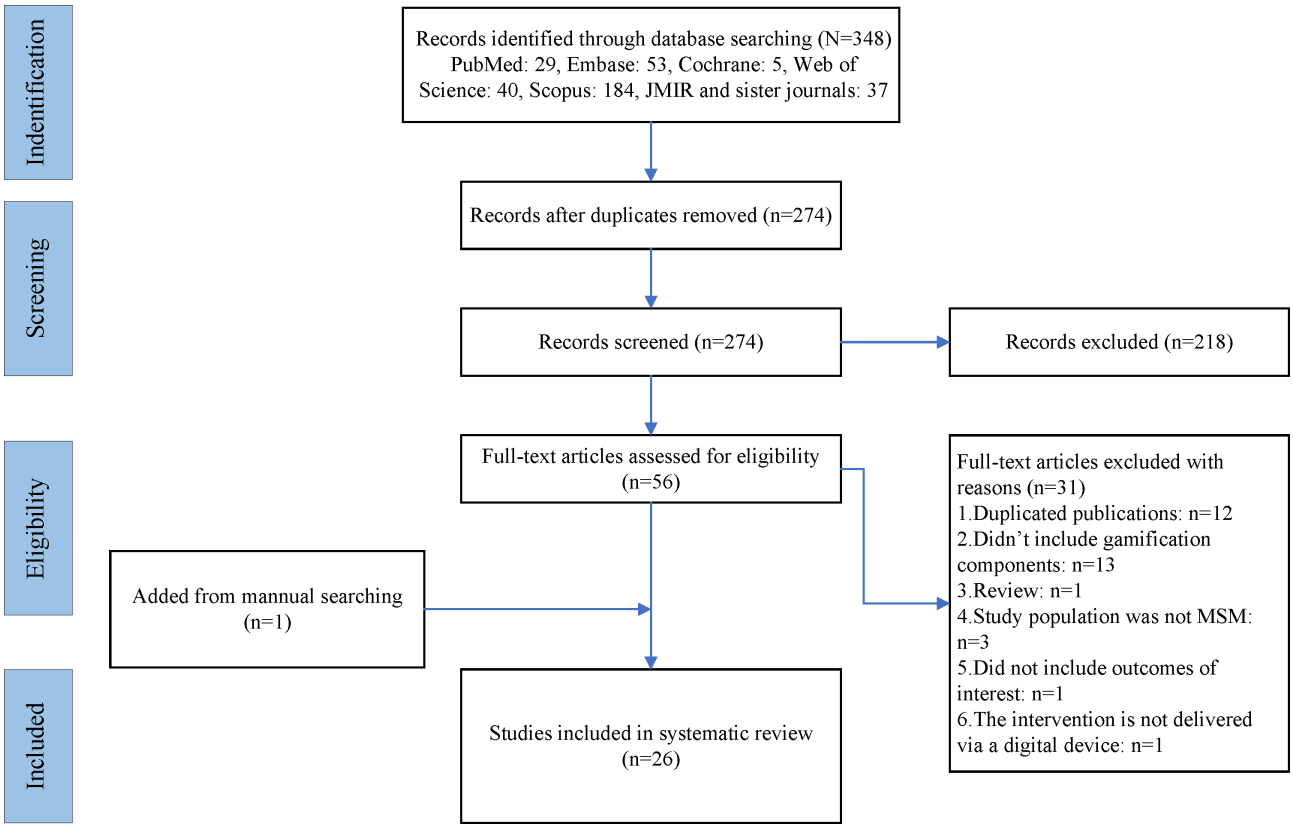
This study was approved by the medical ethics board of Binzhou Medical University (#2021-007).

Results

Search Results

Figure 1 illustrates the literature search and the screening process. A total of 348 records were initially retrieved by the database search. After removing 74 duplicates, the remaining 274 records underwent screening based on their title and abstract. Among these, 218 records were determined to be irrelevant and excluded. Next, the eligibility of 56 full-text records was assessed, yielding 25 records meeting the inclusion criteria and being included in the study. In addition, a secondary reference search was performed, yielding 1 more relevant record. Consequently, a total of 26 articles were included in the qualitative analysis [13,24-48]. Furthermore, of the 26 articles, 6 RCTs were chosen for the meta-analysis [24,29,30,32,33,39].

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart of the literature-screening process. *JMIR*: *Journal of Medical Internet Research*; MSM: men who have sex with men.



Study Characteristics

The features of each included paper regarding study information, interventions, and outcomes are listed in Multimedia Appendix 2. As shown in Table 1 and Table S3-1 in Multimedia Appendix 2, the included studies were published between 2013 and 2023, with the majority published after 2018, which indicated that research on adopting gamification during the HIV prevention and care continuum among MSM is an emerging field. A total of 73% (n=19) of studies were conducted in the United States. As for study designs, most adopted an RCT design (n=10, 38%). In terms of sample sizes, the studies had a range of participants from 5 to 901. When considering all the studies included in the review, a total of 4436 participants were involved. The mean age of participants reported in the studies ranged from 16.2 to 42.7 years. Concerning primary outcomes, most studies explored the effects of gamified intervention on PrEP adherence (n=9), followed by HIV-testing promotion (n=7), changes in CAS (n=6), ART adherence (n=5), PrEP uptake (n=4), and PEP uptake (n=2).

Table . General characteristics of included studies (N=26).

Characteristics		Studies, n (%)
Publication year		
	2013	1 (4)
	2018-2019	10 (38)
	2020-2021	9 (35)
	2022-2023	6 (23)
Country		
	United States	19 (73)
	Malaysia	1 (4)
	Tanzania	1 (4)
	Indonesian	1 (4)
	Spain	1 (4)
	Mexico	1 (4)
	Thailand	1 (4)
	China	1 (4)
Study design		
	Randomized controlled trial	10 (38)
	Pre-post	5 (19)
	Historical control design	1 (4)
	Feasibility test	5 (19)
	Study protocol	5 (19)
Main eHealth modes		
	App	18 (69)
	Internet	5 (19)
	Both	2 (8)
	Not reported	1 (4)
Primary outcomes		
	Condomless anal sex reduction	6 (18)
	HIV testing	7 (21)
	PrEP[a] uptake	4 (12)
	PrEP adherence	9 (27)
	Postexposure prophylaxis	2 (6)
	Antiretroviral therapy adherence	5 (15)
Behavioral theories		
	Motivational interviewing	1 (4)
	Information-motivation-behavioral skills	11 (42)
	Social cognitive theory	5 (19)
	Integrated behavioral model	1 (4)
	Information System Research Framework	1 (4)
	Social learning theory	1 (4)
	Principle of self-learning	1 (4)

Characteristics		Studies, n (%)
	Dyadic HIV care engagement	1 (4)
	Levesque framework	1 (4)
	Not reported	3 (12)

[a]PrEP: pre-exposure prophylaxis.

Gamified Digital Intervention Characteristics

As for intervention characteristics (Table S3-3 in Multimedia Appendix 2), the number of intervention modules ranged from 3 to 13. Of the 26 studies, 22 (85%) included members of the target population (MSM) during the developmental phase of the intervention. Gamification was delivered mostly by mobile apps (n=18, 69%), followed by the internet (n=5, 19%). The duration of study follow-ups ranged from 2 weeks to 15 months. Of the 26 studies, 88% (n=23) used theories for gamified intervention development. Information-motivation-behavioral skills (IMB) were used in 42% (n=11) of studies, social cognitive theory in 19% (n=5), motivational interviewing in 4% (n=1), integrated behavioral model in 4% (n=1), Information System Research Framework in 4% (n=1), social learning theory in 4% (n=1), principle of self-learning in 4% (n=1), dyadic HIV care engagement in 4% (n=1), and Levesque framework in 4% (n=1). Studies included in the systematic review used a range of game core drives and game components (Table 2 and Table S3-3 in Multimedia Appendix 2). The most frequently used game core drives were ownership and possession, followed by social influence and relatedness, development and accomplishment, empowerment creativity and feedback, and unpredictability and curiosity. The most frequently used game components were points, followed by challenge, discussion forums, and mentorship or character narrative.

Table . Type of game core drives and game components used in the included studies (N=26).

Game core drives		Studies, n (%)
Development and accomplishment		11 (42)
	Challenge	9 (35)
	Bottom line	1 (4)
	Leaderboard	1 (4)
Empowerment creativity and feedback		6 (23)
	Tailored message	2 (8)
	Online timely feedback	2 (8)
	Progress bar	2 (8)
Ownership and possession		17 (65)
	Points	10 (38)
	Money	1 (4)
	In-game currency	2 (8)
	Badges	5 (19)
Social influence and relatedness		13 (50)
	Mentorship/character narrative	6 (23)
	Discussion forum	6 (23)
	Allegiance	1 (4)
Unpredictability and curiosity		3 (12)
	Unpredictable incentives	2 (8)
	Gumball/bonus draws	2 (8)

Quality Assessments of Included Studies

An overview of the different risks of bias in each study is presented in Tables S4-1 and S4-2 in Multimedia Appendix 3 [13,24,25,28-30,32-36,39,42,43,45,48]. The weighted Cohen κ coefficient was 0.88, suggesting good agreement between raters [53]. Overall, 10 RCTs were assessed using the RoB2, with 6 rated as having a high risk of bias. A high risk of bias occurred in the domains outcome measurement [24,28,30,42] and missing outcome data [29,43] (Figure S4-1 in Multimedia Appendix 3). Of the 6 nonrandomized studies assessed using the ROBINS-I tool, 2 presented a serious risk of bias [34,35] and 4 a critical risk of bias [25,36,45,48] (Figure S4-2 in Multimedia Appendix 3). These nonrandomized studies were

not further included in the meta-analysis. We did not evaluate the study quality of the other 10 studies [26,27,31,37,38,40,41,44,46,47], as they were designed as study protocols or feasibility studies that did not report results for the outcomes of interest in our review.

Effects of Gamification on CAS

The 10 RCTs under each outcome of interest in this review are listed in Table S5-1 in Multimedia Appendix 4 [13,24-39,41-48]. The meta-analysis on the effect of gamification on CAS reduction contained 3 RCTs (total N=2138). One RCT reported the effect of gamification on engaging in CAS [43], while the other 2 evaluated the effect of gamification on the number of CAS acts and were included in the meta-analysis [32,33]. The gamified digital intervention conferred significant protection against self-reported numbers of CAS acts at the 3-month follow-up (Figure 2; incidence rate ratio [IRR] 0.62, 95% CI 0.44-0.88). However, this effect was not statistically significant 6 months post intervention (Figure 3; IRR 0.71, 95% CI 0.38-1.36). We did not find evidence of publication bias (Figure S6-1 in Multimedia Appendix 5).

Figure 2. Meta-analysis of randomized controlled trials assessing the effect of gamified digital interventions on the number of condomless anal sex acts among men who have sex with men 3 months post intervention [32,33]. IRR: incidence rate ratio.

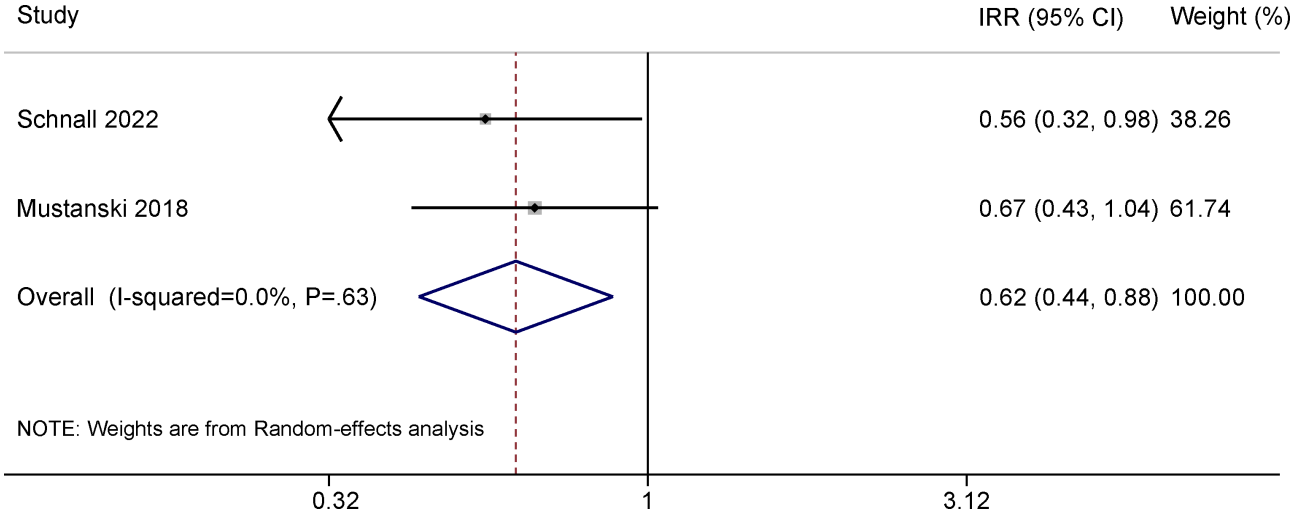
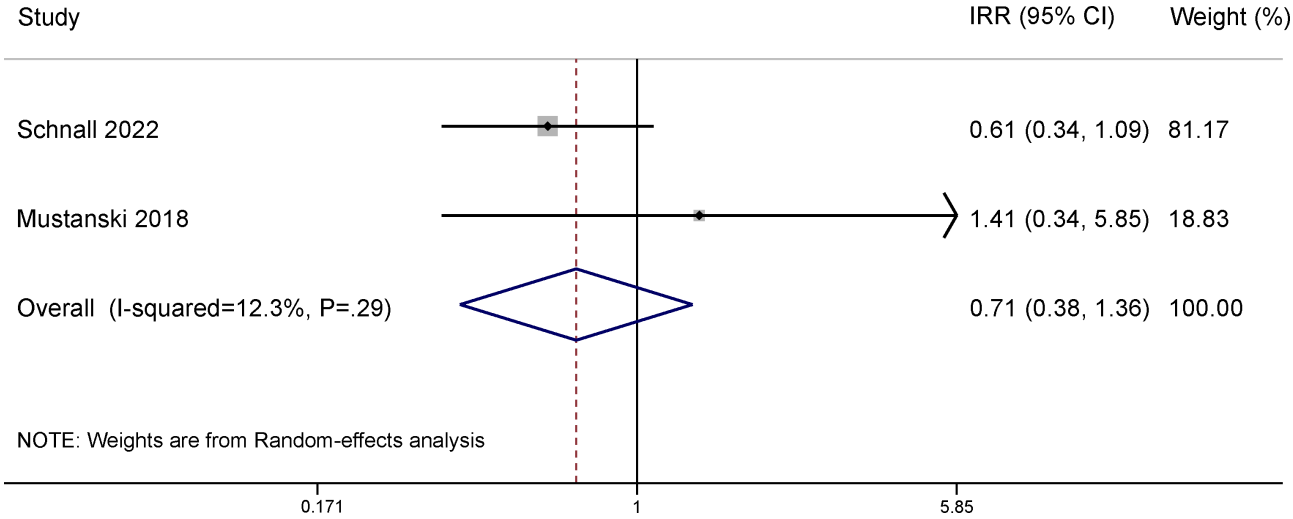


Figure 3. Meta-analysis of randomized controlled trials assessing the effect of gamified digital interventions on the number of condomless anal sex acts among men who have sex with men 6 months post intervention [32,33]. IRR: incidence rate ratio.



Effects of Gamification on PrEP Adherence

A total of 4 RCTs compared PrEP adherence in MSM receiving a gamified digital intervention to non–PrEP-related mobile games [29], youth-friendly services [30], the standard of care from a health educator and a study clinician [39], or educational videos on sleep hygiene and diet [24]. The meta-analysis suggested an effective but nonstatistically significant effect of PrEP adherence at the 3-month follow-up (Figure 4; RR 1.16, 95% CI 0.96-1.38) and 6-month follow-up (Figure 5; RR 1.28, 95% CI 0.89-1.84). We did not find evidence of publication bias (Figure S6-1 in Multimedia Appendix 5).

Only 1 RCT was designed to evaluate a gamified mobile app (MyChoices) to increase HIV testing and PrEP uptake, and the results showed that the MyChoices arm had a 22% higher prevalence of HIV testing over the 6-month follow-up compared to those in the standard of care arm, while there was no difference in PrEP uptake between the different intervention groups [28]. Two RCTs were implemented to evaluate the effects on ART adherence among MSM living with HIV [13,42].

They used different questions to measure ART adherence, and we did not conduct a further meta-analysis for this outcome. For example, the study conducted by Hightow-Weidman et al [13] showed that the proportion of individuals self-reporting ≥90% adherence in the past 7 days rose markedly at 13 weeks post intervention with no significant difference between study arms. Horvath et al [42] evaluated an online social support intervention (Thrive With Me [TWM]) with an RCT design, and the result did not show a significant group difference for the overall ART adherence (≥90% ART adherence in the past 30 days 1 month post intervention). We found that no RCT was conducted to evaluate the effect of the gamified digital intervention on PEP uptake, PEP adherence, and ART initiation. Further studies are still needed to evaluate the effect of the gamified digital intervention on PEP initiation, PEP adherence, and ART uptake in MSM. A meta-regression analysis was not conducted since the number of RCTs in each outcome of interest was fewer than 10.

Figure 4. Meta-analysis of randomized controlled trials assessing the effect of gamified digital interventions on pre-exposure prophylaxis adherence among men who have sex with men at the 3-month follow-up [29,30,39]. RR: risk ratio.

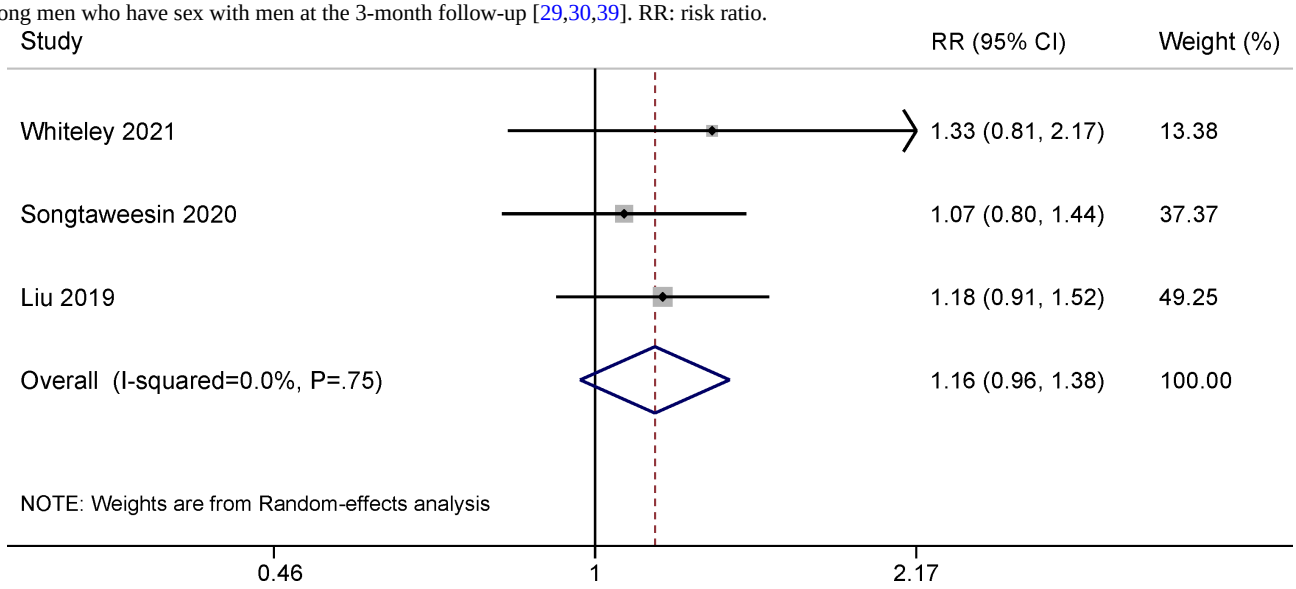
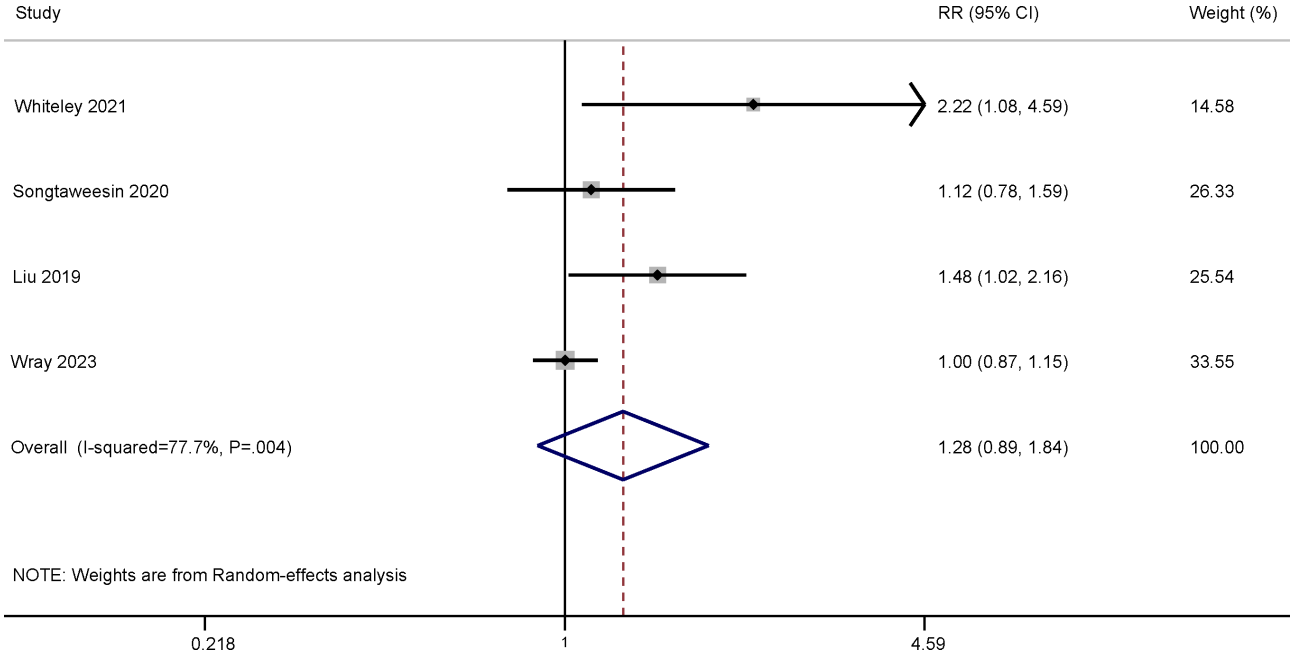


Figure 5. Meta-analysis of randomized controlled trials assessing the effect of gamified digital interventions on pre-exposure prophylaxis adherence among men who have sex with men at the 6-month follow-up [24,29,30,39]. RR: risk ratio.



Discussion

Principal Findings

This study aimed to conduct a systematic review and meta-analysis to evaluate the effectiveness of gamification on the HIV prevention and care continuum among MSM. A total of 26 studies involving 10 RCTs were included in the systematic review. Our findings revealed that gamification of HIV prevention and control addressed HIV-specific outcomes, such as CAS, HIV testing, PrEP uptake, PrEP adherence, nPEP uptake, and ART adherence in MSM living with HIV, with most RCTs focusing on PrEP adherence (4/10, 40%), followed

by CAS (3/10, 30%). The majority of studies were published after 2018 (25/26, 96%) and conducted in the United States (19/26, 73%), and only 6 studies were conducted in low- and middle-income countries (Malaysia, Tanzania, Indonesia, Mexico, Thailand, and China), indicating that the gamification of the HIV prevention and care continuum is an area under development. Our review showed that the most frequently used game core drive was ownership and possession, and points emerged as a frequently used game component. Besides, the most used theory in gamified digital intervention development was IMB. Overall, results from the meta-analysis of RCTs showed significant positive effects of gamification on reducing the number of CAS acts and improving PrEP adherence in MSM. However, the robustness of these findings was limited by the small number of studies included in the meta-analysis.

Gamification Being Applied in the HIV Prevention and Care Continuum

Gamification is a technique that uses game components to motivate participants to engage positively in healthy behaviors. Cugelman [12] identified 7 game components that are linked to proven behavior change strategies, including goal setting, offering a challenge, feedback, reinforcement, compare progress, social connectivity, and fun and playfulness. In this review, we found the most frequently used game element was reinforcement (gaining points), followed by the capacity to overcome a challenge, which is consistent with reviews in other health areas, such as physical activity [54] and cognitive training [55]. Reinforcement represents the motivation driven by feelings of owning something and the desire to get more of it. Hence, healthy behaviors are mostly reinforced by the desire for possession. Challenge is a game component that falls under the core drive development and accomplishment, which motivates individuals through personal growth and the drive to achieve specific targeted goals [23]. Although the above game elements have been widely used in HIV prevention and care interventions, other game elements are less reported in research studies, such as unpredictability and curiosity, loss and avoidance, and scarcity and impatience. Hence, researchers must collaborate with companies focusing on gamification technology to include different game components for the future development of digital interventions.

Effects of Gamification on CAS

The meta-analysis of 2 RCTs demonstrated a 38% reduction in the number of CAS acts in the intervention group compared with the control group at the 3-month follow-up. In addition, a 29% reduction in the number of CAS acts was reported at the 6-month follow-up, with no statistical significance (IRR 0.71, 95% CI 0.38-1.36), which suggests a short-term intervention effect. These findings have important implications, as all of these RCTs enrolled young MSM (ages 13-29 years) disproportionately affected by HIV [56], and few HIV prevention interventions were targeted for young MSM [57]. Moreover, given that mobile technology provides greater access to the target population who may not be able or willing to come in person to receive HIV health service, it provides an ideal platform to reach these young gay, bisexual, and transgender individuals whose ownership of smartphones and use of mobile phone apps are increasing.

Effects of Gamification on HIV PrEP Adherence

Our pooled results indicated that gamified digital interventions effectively improved PrEP adherence in MSM. At the 3-month follow-up, those in the intervention group were 1.16 times more likely to engage in optimal PrEP dosing (≥700 fmol/punch measured by dried blood spots) than the control. The proportion of participants who engaged in optimal PrEP dosing was 28% higher in the intervention versus control groups at the 6-month follow-up. Similar results on behavior change have also been found in asthma and vaccination uptake [58,59]. Notably, in the development stage of their interventions, 3 of the 4 RCTs adopted the IMB theory, which is a behavioral change theory that may explain the effectiveness of these interventions [29,30,39]. Researchers have also suggested that interventions guided by theories are more efficient than those not driven by theories [60]. The IMB model asserts that informed and motivated participants with adequate skills for enacting adherence-related behaviors would optimally adhere to their PrEP regimen over time [61]. Interventions that integrate gamification components can educate users to interface in a more dynamic, immersive, and engaging way to make the experience more informative and motivating for individuals to promote positive health behavior change.

Some concerns should be mentioned after this review. First, more than half of studies on gamification were conducted in the United States, limiting generalizability to other countries. Second, the intervention module and duration were different in different interventions. For example, MSM had access to TWM for 5 months [62], while participants only received access to MyPEEPS over 3 months [32]. Even in an assigned intervention duration, participants' use patterns for different intervention components varied [39]. Reasons may include a lack of ongoing new features embedded in apps, and the effectiveness may be not sustained long-term if the gamification is not continuously improved.

Study Limitations

There were some limitations in this review. First, the number of RCTs available for each outcome of interest in the meta-analysis was limited, suggesting the need for further validation of our findings through additional studies. Second, as previous studies established that intervention engagement metrics (eg, frequency, duration, or amount of mobile health [mHealth] use) were positively associated with interested outcomes [13,63], we were unable to conduct an in-depth analysis of participants' electronic paradata and how these data could affect our result because of the small size of studies. Indeed, further trials are needed to better measure the effective mediators in gamification for healthy behavior. Additionally, the effects of gamification on CAS and PrEP adherence were only pooled 3 and 6 months post intervention. Accordingly, little is currently known about the longer-term impact of the gamified digital intervention.

Conclusions

This study indicated that mHealth-based gamification interventions in the HIV prevention and care continuum among MSM are in a phase of rapid evolution and continued development. Our study substantiated the short-term effect of gamification on CAS in MSM, although the long-term impact of gamified digital interventions remains to be determined.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOCX File, 22 KB - mhealth_v12i1e49509_app1.docx]

Multimedia Appendix 2
Summary of studies included in the review.
[DOCX File, 60 KB - mhealth_v12i1e49509_app2.docx]

Multimedia Appendix 3
Study quality assessment.
[DOCX File, 626 KB - mhealth_v12i1e49509_app3.docx]

Multimedia Appendix 4
Study effects.
[DOCX File, 19 KB - mhealth_v12i1e49509_app4.docx]

Multimedia Appendix 5
Publication bias.
[DOCX File, 206 KB - mhealth_v12i1e49509_app5.docx]

Checklist 1
The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOCX File, 32 KB - mhealth_v12i1e49509_app6.docx]

References

1. Mayer KH, Bekker LG, Stall R, Grulich AE, Colfax G, Lama JR. Comprehensive clinical care for men who have sex with men: an integrated approach. Lancet 2012 Jul 28;380(9839):378-387. [doi: 10.1016/S0140-6736(12)60835-6] [Medline: 22819653]
2. Pitasi MA, Beer L, Cha S, et al. Vital signs: HIV infection, diagnosis, treatment, and prevention among gay, bisexual, and other men who have sex with men - United States, 2010-2019. MMWR Morb Mortal Wkly Rep 2021 Dec 3;70(48):1669-1675. [doi: 10.15585/mmwr.mm7048e1] [Medline: 34855721]
3. Chow EPF, Wilson DP, Zhang J, Jing J, Zhang L. Human immunodeficiency virus prevalence is increasing among men who have sex with men in China: findings from a review and meta-analysis. Sex Transm Dis 2011 Sep;38(9):845-857. [doi: 10.1097/OLQ.0b013e31821a4f43] [Medline: 21844741]
4. Tang S, Tang W, Meyers K, Chan P, Chen Z, Tucker JD. HIV epidemiology and responses among men who have sex with men and transgender individuals in China: a scoping review. BMC Infect Dis 2016 Oct 20;16(1):588. [doi: 10.1186/s12879-016-1904-5] [Medline: 27765021]
5. Kanny D, Jeffries WL, Chapin-Bardales J, et al. Racial/ethnic disparities in HIV preexposure prophylaxis among men who have sex with men - 23 urban areas, 2017. MMWR Morb Mortal Wkly Rep 2019 Sep 20;68(37):801-806. [doi: 10.15585/mmwr.mm6837a2] [Medline: 31536484]
6. Wang Z, Yuan T, Fan S, et al. HIV nonoccupational postexposure prophylaxis among men who have sex with men: a systematic review and meta-analysis of global data. AIDS Patient Care STDS 2020 May;34(5):193-204. [doi: 10.1089/apc.2019.0313] [Medline: 32396477]
7. Phillips G 2nd, Magnus M, Kuo I, et al. Use of geosocial networking (GSN) mobile phone applications to find men for sex by men who have sex with men (MSM) in Washington, DC. AIDS Behav 2014 Sep;18(9):1630-1637. [doi: 10.1007/s10461-014-0760-9] [Medline: 24682866]

8. Burke LE, Ma J, Azar KMJ, et al. Current science on consumer use of mobile health for cardiovascular disease prevention: a scientific statement from the American Heart Association. Circulation 2015 Sep 22;132(12):1157-1213. [doi: 10.1161/CIR.0000000000000232] [Medline: 26271892]
9. Muessig KE, Nekkanti M, Bauermeister J, Bull S, Hightow-Weidman LB. A systematic review of recent smartphone, internet and web 2.0 interventions to address the HIV continuum of care. Curr HIV/AIDS Rep 2015 Mar;12(1):173-190. [doi: 10.1007/s11904-014-0239-3] [Medline: 25626718]
10. McCoy SI, Packel L. Lessons from early stage pilot studies to maximize the impact of digital health interventions for sexual and reproductive health. Mhealth 2020 Jul 5;6:22. [doi: 10.21037/mhealth.2020.02.03] [Medline: 32632360]
11. Zainuddin Z, Chu SKW, Shujahat M, Perera CJ. The impact of gamifcation on learning and instruction: a systematic review of empirical evidence. Educa Res Rev 2020 Jun;30:100326. [doi: 10.1016/j.edurev.2020.100326]
12. Cugelman B. Gamification: what it is and why it matters to digital health behavior change developers. JMIR Serious Games 2013 Dec 12;1(1):e3. [doi: 10.2196/games.3139] [Medline: 25658754]
13. Hightow-Weidman L, Muessig KE, Egger JR, Vecchio A, Platt A. Epic allies: a gamified mobile app to improve engagement in HIV care and antiretroviral adherence among young men who have sex with men. AIDS Behav 2021 Aug;25(8):2599-2617. [doi: 10.1007/s10461-021-03222-y] [Medline: 33740213]
14. Lister C, West JH, Cannon B, Sax T, Brodegard D. Just a fad? Gamification in health and fitness apps. JMIR Serious Games 2014 Aug 4;2(2):e9. [doi: 10.2196/games.3413] [Medline: 25654660]
15. Hightow-Weidman LB, Muessig KE, Bauermeister JA, LeGrand S, Fiellin LE. The future of digital games for HIV prevention and care. Curr Opin HIV AIDS 2017 Sep;12(5):501-507. [doi: 10.1097/COH.0000000000000399] [Medline: 28692490]
16. Warsinsky S, Schmidt-Kraepelin M, Rank S, Thiebes S, Sunyaev A. Conceptual ambiguity surrounding gamification and serious games in health care: literature review and development of game-based intervention reporting guidelines (GAMING). J Med Internet Res 2021 Sep 10;23(9):e30390. [doi: 10.2196/30390] [Medline: 34505840]
17. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71. [doi: 10.1136/bmj.n71] [Medline: 33782057]
18. Deterding S, Sicart M, Nacke L, O'Hara K, Dixon D. Gamification. Using game-design elements in non-gaming contexts. Presented at: CHI Conference on Human Factors in Computing Systems; May 7-12, 2011; Vancouver, BC. [doi: 10.1145/1979742.1979575]
19. Sterne JAC, Savović J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019 Aug 28;366:l4898. [doi: 10.1136/bmj.l4898] [Medline: 31462531]
20. Higgins JPT, Altman DG, Gøtzsche PC, et al. The Cochrane collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011 Oct 18;343:d5928. [doi: 10.1136/bmj.d5928] [Medline: 22008217]
21. Sterne JA, Hernán MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ 2016 Oct 12;355:i4919. [doi: 10.1136/bmj.i4919] [Medline: 27733354]
22. McGuinness LA, Higgins JPT. Risk-of-bias visualization (Robvis): an R package and shiny web app for visualizing risk-of-bias assessments. Res Synth Methods 2021 Jan;12(1):55-61. [doi: 10.1002/jrsm.1411] [Medline: 32336025]
23. Chou YK. Actionable Gamification: Beyond Points, Badges, and Leaderboards: Packt Publishing Ltd; 2014.
24. Wray TB, Chan PA, Kahler CW, Ocean EMS, Nittas V. Pilot randomized controlled trial of game plan for PrEP: a brief, web and text message intervention to help sexual minority men adhere to PrEP and reduce their alcohol use. AIDS Behav 2024 Apr;28(4):1356-1369. [doi: 10.1007/s10461-023-04223-9] [Medline: 37971613]
25. Weitzman PF, Zhou Y, Kogelman L, Rodarte S, Vicente SR, Levkoff SE. mHealth for pre-exposure prophylaxis adherence by young adult men who have sex with men. Mhealth 2021 Jul 20;7:44. [doi: 10.21037/mhealth-20-51] [Medline: 34345621]
26. Kim HC, Pollack LM, Saberi P, et al. Study protocol: a pilot randomised waitlist-controlled trial of a dyadic mobile health intervention for black sexual-minority male couples with HIV in the USA. BMJ Open 2021 Sep 2;11(9):e055448. [doi: 10.1136/bmjopen-2021-055448] [Medline: 34475191]
27. Bauermeister J, Sullivan PS, Gravens L, et al. Reducing HIV vulnerability through a multilevel life skills intervention for adolescent men (the iREACH project): protocol for a randomized controlled trial. JMIR Res Protoc 2018 Jul 10;7(7):e10174. [doi: 10.2196/10174] [Medline: 29991470]
28. Biello KB, Daddario SR, Hill-Rorie J, et al. Uptake and acceptability of Mychoices: results of a pilot RCT of a mobile app designed to increase HIV testing and prep uptake among young American MSM. AIDS Behav 2022 Dec;26(12):3981-3990. [doi: 10.1007/s10461-022-03724-3] [Medline: 35829971]
29. Whiteley L, Craker L, Haubrick KK, et al. The impact of a mobile gaming intervention to increase adherence to pre-exposure prophylaxis. AIDS Behav 2021 Jun;25(6):1884-1889. [doi: 10.1007/s10461-020-03118-3] [Medline: 33483897]
30. Songtaweesin WN, Kawichai S, Phanuphak N, et al. Youth-friendly services and a mobile phone application to promote adherence to pre-exposure prophylaxis among adolescent men who have sex with men and transgender women at-risk for HIV in Thailand: a randomized control trial. J Int AIDS Soc 2020 Sep;23(Suppl 5):e25564. [doi: 10.1002/jia2.25564] [Medline: 32869511]
31. Shrestha R, Altice FL, Khati A, et al. Clinic-integrated smartphone app (JomPrEP) to improve uptake of HIV testing and pre-exposure prophylaxis among men who have sex with men in Malaysia: mixed methods evaluation of usability and acceptability. JMIR Mhealth Uhealth 2023 Feb 16;11:e44468. [doi: 10.2196/44468] [Medline: 36795465]

32. Schnall R, Kuhns LM, Pearson C, et al. Efficacy of Mypeeps mobile, an HIV prevention intervention using mobile technology, on reducing sexual risk among same-sex attracted adolescent males: a randomized clinical trial. JAMA Netw Open 2022 Sep 1;5(9):e2231853. [doi: 10.1001/jamanetworkopen.2022.31853] [Medline: 36129712]
33. Mustanski B, Parsons JT, Sullivan PS, Madkins K, Rosenberg E, Swann G. Biomedical and behavioral outcomes of Keep It Up!: an eHealth HIV prevention program RCT. Am J Prev Med 2018 Aug;55(2):151-158. [doi: 10.1016/j.amepre.2018.04.026] [Medline: 29937115]
34. Mitchell JT, LeGrand S, Hightow-Weidman LB, et al. Smartphone-based contingency management intervention to improve pre-exposure prophylaxis adherence: pilot trial. JMIR Mhealth Uhealth 2018 Sep 10;6(9):e10456. [doi: 10.2196/10456] [Medline: 30201601]
35. Mitchell JT, Burns CM, Atkinson B, et al. Feasibility, acceptability, and preliminary efficacy of a gamified mobile health contingency management intervention for PrEP adherence among Black MSM. AIDS Behav 2022 Oct;26(10):3311-3324. [doi: 10.1007/s10461-022-03675-9] [Medline: 35416595]
36. McCoy SI, Buzdugan R, Grimball R, et al. Stick To It: pilot study results of an intervention using gamification to increase HIV screening among young men who have sex with men in California. Mhealth 2018 Sep 25;4:40. [doi: 10.21037/mhealth.2018.09.04] [Medline: 30363751]
37. Mauka W, Mbotwa C, Moen K, et al. Development of a mobile health application for HIV prevention among at-risk populations in urban settings in East Africa: a participatory design approach. JMIR Form Res 2021 Oct 7;5(10):e23204. [doi: 10.2196/23204] [Medline: 34617904]
38. Luo Q, Luo Y, Li T, Cui T. An integrated online-to-offline model for HIV post-exposure prophylaxis (O2O-PEP) scale-up among men who have sex with men (MSM): protocol for developing a pilot randomized controlled trial. Front Public Health 2022 Nov 16;10:1026137. [doi: 10.3389/fpubh.2022.1026137] [Medline: 36466536]
39. Liu AY, Vittinghoff E, von Felten P, et al. Randomized controlled trial of a mobile health intervention to promote retention and adherence to preexposure prophylaxis among young people at risk for human immunodeficiency virus: the EPIC study. Clin Infect Dis 2019 May 30;68(12):2010-2017. [doi: 10.1093/cid/ciy810] [Medline: 30239620]
40. Liu A, Coleman K, Bojan K, et al. Developing a mobile app (LYNX) to support linkage to HIV/sexually transmitted infection testing and pre-exposure prophylaxis for young men who have sex with men: protocol for a randomized controlled trial. JMIR Res Protoc 2019 Jan 25;8(1):e10659. [doi: 10.2196/10659] [Medline: 30681964]
41. LeGrand S, Knudtson K, Benkeser D, et al. Testing the efficacy of a social networking gamification app to improve pre-exposure prophylaxis adherence (P3: Prepared, Protected, emPowered): protocol for a randomized controlled trial. JMIR Res Protoc 2018 Dec 18;7(12):e10448. [doi: 10.2196/10448]
42. Horvath KJ, Oakes JM, Rosser BRS, et al. Feasibility, acceptability and preliminary efficacy of an online peer-to-peer social support ART adherence intervention. AIDS Behav 2013 Jul;17(6):2031-2044. [doi: 10.1007/s10461-013-0469-1] [Medline: 23553347]
43. Hightow-Weidman LB, LeGrand S, Muessig KE, et al. A randomized trial of an online risk reduction intervention for young black MSM. AIDS Behav 2019 May;23(5):1166-1177. [doi: 10.1007/s10461-018-2289-9] [Medline: 30269231]
44. Hightow-Weidman L, Muessig K, Knudtson K, et al. A gamified smartphone app to support engagement in care and medication adherence for HIV-positive young men who have sex with men (AllyQuest): development and pilot study. JMIR Public Health Surveill 2018 Apr 30;4(2):e34. [doi: 10.2196/publichealth.8923] [Medline: 29712626]
45. Dworkin MS, Lee S, Chakraborty A, et al. Feasibility, and preliminary efficacy of a theory-based relational embodied conversational agent mobile phone intervention to promote HIV medication adherence in young HIV-positive African American MSM. AIDS Educ Prev 2019 Feb;31(1):17-37. [doi: 10.1521/aeap.2019.31.1.17] [Medline: 30742481]
46. Besoain F, Perez-Navarro A, Jacques Aviñó C, Caylà JA, Barriga NA, Garcia de Olalla P. Prevention of HIV and other sexually transmitted infections by geofencing and contextualized messages with a gamified app, UBESAFE: design and creation study. JMIR Mhealth Uhealth 2020 Mar 17;8(3):e14568. [doi: 10.2196/14568] [Medline: 32181752]
47. Andrade-Romo Z, Chavira-Razo L, Buzdugan R, Bertozzi E, Bautista-Arredondo S. Hot, horny and healthy-online intervention to incentivize HIV and sexually transmitted infections (STI) testing among young Mexican MSM: a feasibility study. Mhealth 2020 Jul 5;6:28. [doi: 10.21037/mhealth.2020.03.01] [Medline: 32632366]
48. Garg PR, Uppal L, Mehra S, Mehra D. Mobile health app for self-learning on HIV prevention knowledge and services among a young Indonesian key population: cohort study. JMIR Mhealth Uhealth 2020 Sep 8;8(9):e17646. [doi: 10.2196/17646] [Medline: 32896831]
49. Laine TH, Duong N, Lindvall H, Oyelere SS, Rutberg S, Lindqvist AK. A reusable multiplayer game for promoting active school transport: development study. JMIR Serious Games 2022 Mar 14;10(1):e31638. [doi: 10.2196/31638] [Medline: 35285815]
50. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials 1986 Sep;7(3):177-188. [doi: 10.1016/0197-2456(86)90046-2] [Medline: 3802833]
51. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003 Sep 6;327(7414):557-560. [doi: 10.1136/bmj.327.7414.557] [Medline: 12958120]
52. Higgins JPT, Thomas J, Chandler J, et al. Cochrane Handbook for systematic reviews of interventions. Cochrane Training. 2023. URL: https://training.cochrane.org/handbook [accessed 2024-01-23]

53. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159. [doi: 10.2307/2529310] [Medline: 843571]
54. Xu L, Shi H, Shen M, et al. The effects of mHealth-based gamification interventions on participation in physical activity: systematic review. JMIR Mhealth Uhealth 2022 Feb 3;10(2):e27794. [doi: 10.2196/27794] [Medline: 35113034]
55. Vermeir JF, White MJ, Johnson D, Crombez G, Van Ryckeghem DML. The effects of gamification on computerized cognitive training: systematic review and meta-analysis. JMIR Serious Games 2020 Aug 10;8(3):e18644. [doi: 10.2196/18644] [Medline: 32773374]
56. HIV surveillance reports. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/hiv/library/reports/hiv-surveillance.html [accessed 2024-03-20]
57. Mustanski B, Fisher CB. HIV rates are increasing in gay/bisexual teens: IRB barriers to research must be resolved to bend the curve. Am J Prev Med 2016 Aug;51(2):249-252. [doi: 10.1016/j.amepre.2016.02.026] [Medline: 27061893]
58. Drummond D, Monnier D, Tesnière A, Hadchouel A. A systematic review of serious games in asthma education. Pediatr Allergy Immunol 2017 May;28(3):257-265. [doi: 10.1111/pai.12690] [Medline: 27992659]
59. Montagni I, Mabchour I, Tzourio C. Digital gamification to enhance vaccine knowledge and uptake: scoping review. JMIR Serious Games 2020 May 18;8(2):e16983. [doi: 10.2196/16983] [Medline: 32348271]
60. Simoni JM, Frick PA, Pantalone DW, Turner BJ. Antiretroviral adherence interventions: a review of current literature and ongoing studies. Top HIV Med Nov-Dec 2003;11(6):185-198. [Medline: 14724327]
61. Fisher JD, Amico KR, Fisher WA, Harman JJ. The information-motivation-behavioral skills model of antiretroviral adherence and its applications. Curr HIV/AIDS Rep 2008 Nov;5(4):193-203. [doi: 10.1007/s11904-008-0028-y] [Medline: 18838059]
62. Horvath KJ, Amico KR, Erickson D, et al. Thrive With Me: protocol for a randomized controlled trial to test a peer support intervention to improve antiretroviral therapy adherence among men who have sex with men. JMIR Res Protoc 2018 May 31;7(5):e10182. [doi: 10.2196/10182] [Medline: 29853437]
63. Hightow-Weidman LB, Bauermeister JA. Engagement in mHealth behavioral interventions for HIV prevention and care: making sense of the metrics. Mhealth 2020 Jan 5;6:7. [doi: 10.21037/mhealth.2019.10.01] [Medline: 32190618]

Abbreviations

ART: antiretroviral therapy
CAS: condomless anal sex
IMB: information-motivation-behavioral skills
IRR: incidence rate ratio
MeSH: Medical Subject Headings
mHealth: mobile health
MSM: men who have sex with men
nPEP: nonoccupational postexposure prophylaxis
PEP: postexposure prophylaxis
PrEP: pre-exposure prophylaxis
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
RoB2: Risk of Bias Tool
ROBINS-I: Risk of Bias in Nonrandomized Studies of Interventions
RR: risk ratio
TWM: Thrive With Me

Edited by L Buis; submitted 01.06.23; peer-reviewed by S Chen, S John; revised version received 26.01.24; accepted 27.02.24; published 15.04.24.

Please cite as:
Luo Q, Zhang Y, Wang W, Cui T, Li T
mHealth-Based Gamification Interventions Among Men Who Have Sex With Men in the HIV Prevention and Care Continuum: Systematic Review and Meta-Analysis
JMIR Mhealth Uhealth 2024;12:e49509
URL: https://mhealth.jmir.org/2024/1/e49509
doi:10.2196/49509

Effects of Telemedicine on Informal Caregivers of Patients in Palliative Care: Systematic Review and Meta-Analysis

Xiaoyu Yang[1], BSN; Xueting Li[2], BSN; Shanshan Jiang[2], BSN; Xinying Yu[3], MSN

[1]Department of Oncology, Shengjing Hospital of China Medical University, Shenyang, China

[2]College of Nursing, China Medical University, Shenyang, China

[3]Department of Pediatrics, Shengjing Hospital of China Medical University, Shenyang, China

Corresponding Author:
Xinying Yu, MSN
Department of Pediatrics, Shengjing Hospital of China Medical University, , Shenyang, , China

Abstract

Background: Telemedicine technology is a rapidly developing field that shows immense potential for improving medical services. In palliative care, informal caregivers assume the primary responsibility in patient care and often face challenges such as increased physical and mental stress and declining health. In such cases, telemedicine interventions can provide support and improve their health outcomes. However, research findings regarding the use of telemedicine among informal caregivers are controversial, and the efficacy of telemedicine remains unclear.

Objective: This study aimed to evaluate the impacts of telemedicine on the burden, anxiety, depression, and quality of life of informal caregivers of patients in palliative care.

Methods: A systematic literature search was conducted using the PubMed, Embase, Web of Science, CENTRAL, PsycINFO, CINAHL Plus with Full Text, CBM, CNKI, WanFang, and VIP databases to identify relevant randomized controlled trials published from inception to March 2023. Two authors independently screened the studies and extracted the relevant information. The methodological quality of the included studies was assessed using the Cochrane risk-of-bias tool. Intervention effects were estimated and sensitivity analysis was conducted using Review Manager 5.4, whereas 95% prediction intervals (PIs) were calculated using R (version 4.3.2) and RStudio.

Results: A total of 9 randomized controlled trials were included in this study. The meta-analysis indicated that telemedicine has reduced the caregiving burden (standardized mean differences [SMD] −0.49, 95% CI −0.72 to −0.27; *P*<.001; 95% PI −0.86 to −0.13) and anxiety (SMD −0.23, 95% CI −0.40 to −0.06; *P*=.009; 95% PI −0.98 to 0.39) of informal caregivers; however, it did not affect depression (SMD −0.21, 95% CI −0.47 to 0.05; *P*=.11; 95% PI −0.94 to 0.51) or quality of life (SMD 0.35, 95% CI −0.20 to 0.89; *P*=.21; 95% PI −2.15 to 2.85).

Conclusions: Although telemedicine can alleviate the caregiving burden and anxiety of informal caregivers, it does not significantly reduce depression or improve their quality of life. Further high-quality, large-sample studies are needed to validate the effects of telemedicine. Furthermore, personalized intervention programs based on theoretical foundations are required to support caregivers.

Trial Registration: PROSPERO CRD42023415688; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=415688

(JMIR Mhealth Uhealth 2024;12:e54244) doi:10.2196/54244

KEYWORDS

telemedicine; palliative care; informal caregivers; caregiver burden; anxiety; depression; quality of life; systematic review; meta-analysis; PRISMA; Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Introduction

Background

With the continued increase in the number of individuals with multiple and severe diseases, the global demand for palliative care services is also growing [1]. Given that most patients who require palliative care prefer to spend time at home and receive the necessary care [2-4], informal caregivers play a crucial role in caring for patients. However, the cumbersome and complex care tasks may have negative impacts on their physical, psychological, and social well-being [5-7]. In recent years, telemedicine, as an emerging technology, has been increasingly used in home care [2], benefiting informal caregivers [8,9]. It may serve as a pathway to support informal caregivers of patients in palliative care, improve their health outcomes, and thus enhance the quality of palliative care [10].

The World Health Organization estimates that 56.8 million people require palliative care yearly [11]. However, there is a

prevailing shortage of professional palliative care personnel, and the majority of patients prefer to receive such care at home [2-4]. Hence, informal caregivers, usually family members or friends, assume the primary responsibility for patient care. In doing so, they adapt to changes in their role, family, and social life to provide long-term, unpaid care for patients [12,13]. Informal caregivers frequently lack professional training [4]. Thus, they face unmet supportive care needs, such as symptom management, psychological counseling, and social support [4,13], and experience anxiety, depression, physical overload, and a decline in the quality of life (QOL) [5-7]. Studies reveal that the state of informal caregivers and the condition of the patients mutually affect each other. The quality of care provided by caregivers in poor condition can be diminished, exacerbating the patient's condition. In turn, the patient's worsening condition can negatively affect informal caregivers [7,14,15]. Therefore, the demand to assist informal caregivers and address their physiological, psychological, and social health needs is urgent.

With the development of the information age, telemedicine has demonstrated tremendous potential in providing health care. Telemedicine refers to the use of information and communication technologies to facilitate communication between patients and health care workers for the assessment, diagnosis, treatment, and prevention of diseases, thereby improving patient health [16]. As a personalized medical approach, telemedicine overcomes the conventional care constraints of time and space; facilitates remote treatment, supervision, education, and care services; and promotes the rational distribution and refinement of medical resources [17]. Telemedicine has been widely applied in medical fields such as diabetes, chronic wounds, and cardiovascular diseases [18,19], benefiting patients and improving the health outcomes of informal caregivers [8,9].

In recent years, telemedicine has also provided novel ideas to guide palliative care [20]. In the field of palliative care, an increasing number of informal caregivers are opting to provide home care for patients who require palliative care [21]. Telemedicine facilitates real-time communication between professionals and family caregivers. This promotes information sharing; assists in the patient's symptom management; and helps in providing health education, psychological counseling, and social support [22]. However, the outcomes of using telemedicine with informal caregivers are controversial. For example, a randomized controlled trial (RCT) by Chen et al [23] indicated that telemedicine could alleviate caregiving burden and enhance the QOL of informal caregivers. However, Dionne-Odom et al [24] found no significant difference between the telemedicine and control groups in terms of improvements in informal caregivers' QOL, burden, or emotional state. Of the few available systematic reviews, most provide a descriptive summary of results without performing a meta-analysis to quantify the outcomes of the studies [25-27]. Thus, the intervention effects of telemedicine remain unclear.

Objectives

Given the limitations of previous reviews, we conducted a systematic review and meta-analysis. We summarized articles on the intervention effects of telemedicine among the informal caregivers of patients in palliative care, focusing on 4 health outcomes: caregiver burden, anxiety, depression, and QOL. This provides a reference for the clinical practice of telemedicine. This is the first systematic review and meta-analysis to verify the effects of telemedicine on the outcomes for informal caregivers of patients in palliative care.

Methods

Overview

This systematic review adhered to the guidelines in the 2020 PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist [28] (Checklist 1). In addition, it was registered on PROSPERO (CRD42023415688).

Ethical Considerations

As all data used were obtained from previously published articles, this research did not require ethical approval from an institutional review board or informed consent from participants.

Search Strategy

A literature search was conducted in 10 electronic databases (PubMed, Embase, Web of Science, CENTRAL, PsycINFO, CINAHL Plus with Full Text, CBM, CNKI, WanFang, and VIP) for publications dating from the establishment of each database until March 31, 2023. Following the Population, Intervention, Comparison, Outcome, and Study design principles, the searches in this systematic review were performed using Medical Subject Headings, the title or abstract, and keywords, as well as Boolean logical operations. Multimedia Appendix 1 describes the search strategy for all databases. In addition, relevant systematic reviews and references were manually screened to identify additional eligible studies.

Study Eligibility Criteria

The inclusion criteria were as follows. (1) The study population was informal adult caregivers (aged ≥18 y) caring for patients receiving palliative care for severe diseases (eg, advanced stage, incurable "stage 4" diseases). (2) Intervention measures were being provided through the internet, applications, telephone, video, or other telemedicine technologies. (3) The control group received usual care or enhanced usual care or was on a waiting list. (4) The study reported outcomes for informal caregivers focusing on 1 or more of the following aspects: caregiver burden, anxiety, depression, or QOL. (5) The study was designed as an RCT. (6) The article was published in English or Chinese.

The exclusion criteria were as follows. (1) The publications were qualitative research, conference abstracts, letters, comments, reviews, or protocols. (2) Patients were underage (aged <18 y), or palliative care indications were unrelated to life-limiting diseases (eg, chronic diseases or nonmalignant pain). (3) Interventions were not being targeted at informal caregivers. (4) The full-text article or relevant data were not accessible.

Study Selection and Data Extraction

The titles and abstracts of the retrieved literature were first downloaded and imported into Endnote X9 (Clarivate; a reference management program) to remove duplicates. Two

evaluators then independently screened the studies based on the inclusion and exclusion criteria. Any disagreements were resolved through consultation or discussion with a third researcher. Data were extracted through a predesigned table, including the name of the first author, year of publication, country, age of the caregiver, sample size, type of disease diagnosed in the patient, type and content of intervention measures, study duration, and time of the outcome assessment.

Quality Assessment

Two evaluators independently assessed the methodological quality of the included studies using the Cochrane Collaboration's tool for assessing the risk of bias [29]. Seven aspects were evaluated: (1) random sequence generation, (2) allocation concealment, (3) blinding of participants and personnel, (4) blinding of outcome assessment, (5) incomplete outcome data, (6) selective reporting, and (7) other biases. Each study was categorized as "low risk," "uncertain risk," or "high risk," with disagreements resolved through consultation or discussion with a third researcher.

Data Analysis

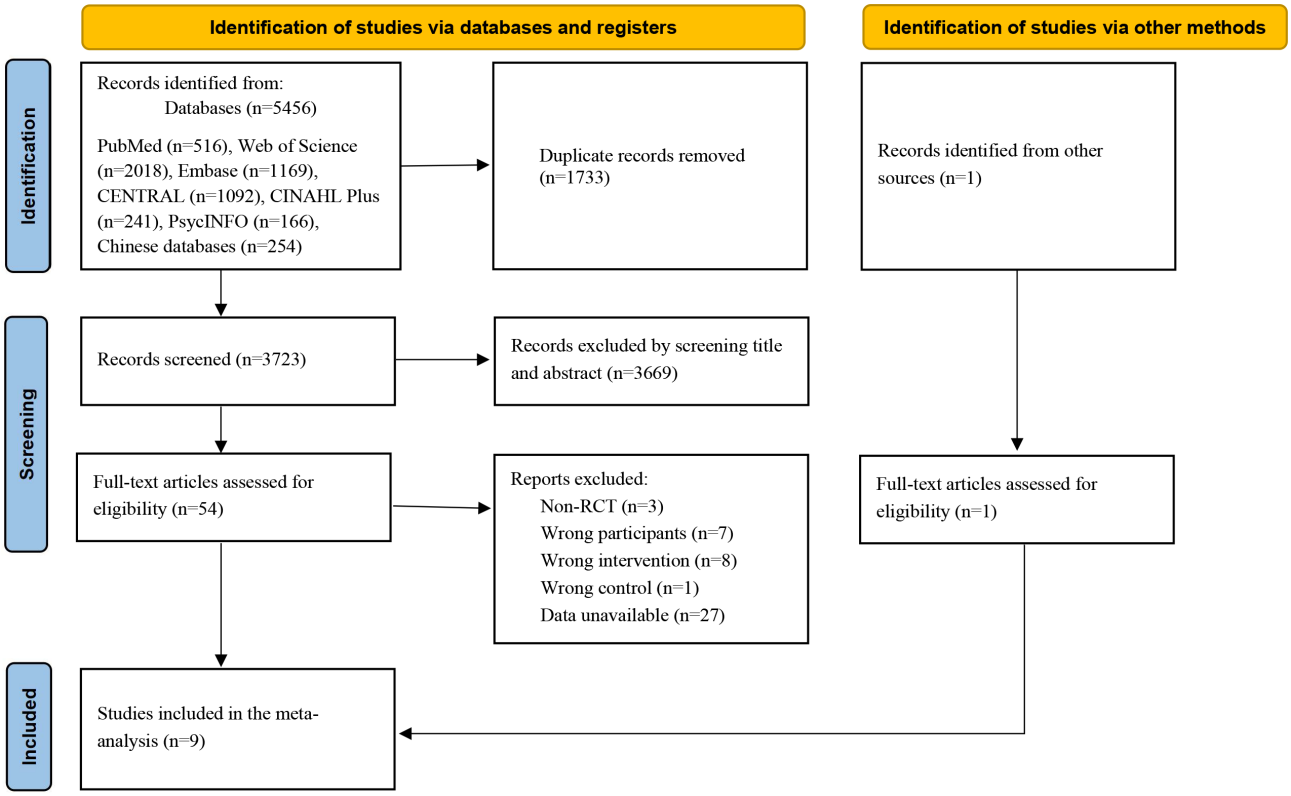
For studies with multiple measurements, only data from the last measurement were extracted for analysis. SDs were calculated according to the *Cochrane Handbook for Systematic Reviews of Interventions* if not reported [29]. If required data were not reported, we contacted the first authors of the relevant publication. Heterogeneity testing and the meta-analysis were conducted using Review Manager 5.4 (The Cochrane Collaboration). Intervention effects were estimated through standardized mean differences (SMDs) and 95% CIs, and forest plots were generated. A 2-sided P value <.05 was considered statistically significant. Heterogeneity was evaluated using the χ^2 test (with P<.10 indicating heterogeneity) and I^2 test (with I^2>50% indicating moderate heterogeneity and I^2>75% indicating high heterogeneity). If I^2≤50% and P>.10, a fixed-effect model was adopted for data merging and analysis; otherwise, a random-effects model was used. A sensitivity analysis was conducted using a one-study-out method to evaluate the robustness of the combined results. In addition, 95% prediction intervals (PIs) were calculated using R (version 4.3.2; R Foundation for Statistical Computing) and RStudio (Posit) to explain the heterogeneity across studies and estimate the true effects in similar future studies [30].

Results

Search Results and Selection

A preliminary search of the electronic databases yielded 5456 articles: 254 in Chinese and 5202 in English. After removing 1733 duplicated articles, an additional 3669 unrelated articles were excluded after evaluating their titles and abstracts, leaving 54 articles for the full-text review. From these, 8 articles were included, and with the addition of 1 more article, 9 studies were ultimately included in the meta-analysis. The screening process is detailed in Figure 1 [28].

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart. RCT: randomized controlled trial.



Characteristics of the Included Studies

Study Characteristics

Multimedia Appendix 2 [23,24,31-37] summarizes the main characteristics of the included studies. These studies were all RCTs published in 3 countries between 2015 and 2023: a total of 6 from the United States, 1 from the Netherlands, and 2 from China. Four studies mentioned the theoretical or conceptual framework of the intervention, including Erikson's psychosocial development theory and Bowen's family system theory [23], self-determination theory [31], shared decision-making [33], and cognitive behavioral stress management [35].

Characteristics of Informal Caregivers

The studies involved 1215 informal caregivers, with the number of participants in each study ranging from 35 to 334. The average age of the informal caregivers ranged from 45.71 (SD 11.85) to 60.1 (SD 12.5) years, and they were predominantly patients' parents, spouses or partners, and children. The types of diseases of the patient included advanced cancer, advanced heart failure, and advanced dementia.

Characteristics of Telemedicine Interventions

Telemedicine was practiced via websites, web conferences, applications, or the telephone, but primarily through websites and the telephone. A total of 4 studies provided interventions through a website. Oliver et al [33] performed a 3-arm clinical trial, where 1 group received an intervention via Facebook, which offered education and social support to informal caregivers, whereas a separate group received the ACCESS intervention. Here, in addition to the Facebook-based intervention, web conferences were incorporated to facilitate the engagement of informal caregivers in joint decision-making in palliative care. The project aimed to alleviate informal family caregivers' anxiety and depression. Pensak et al [35] implemented a 12-week intervention named Pep-Pal, which provided stress management support to informal caregivers of patients with advanced cancer via a website. The intervention program of Parker Oliver et al [34], ACTIVE, used web conferences or telephone calls to link informal caregivers to end-of-life care teams to improve caregivers' perceptions of pain management. Similarly, Fu et al [37] established real-time communication between medical staff and family caregivers of patients with advanced cancer via an internet platform to provide relevant health guidance. Furthermore, 2 studies provided intervention through an application. Schuit et al [36] developed a program called Oncokompas to provide personalized information, suggestions, and supportive care solutions tailored to the caregiver's situation. Chen et al [23] created a dyadic life review program for patients with advanced cancer and their caregivers using WeChat software to promote their QOL. In addition, 3 studies offered interventions via the telephone. Dionne-Odom et al [24] implemented a telephone intervention program named ENABLE CHF-PC, which offered psychological and problem-solving support for patients with heart failure in palliative care and their caregivers to improve their emotions, burden, and QOL. Two years later, Dionne-Odom et al [32] performed a similar intervention for patients with advanced cancer and their caregivers. Finally, Badr et al [31] provided a telephonic psychosocial intervention to enhance the QOL of patients with advanced cancer and their caregivers.

Characteristics of Controls

The control group in 1 study received enhanced usual care [33], whereas those in the remaining studies received usual care. Furthermore, in the study by Schuit et al [36], informal caregivers in the control group were allowed to use telemedicine equipment after the research ended.

Risk of Bias

A quality assessment of the included studies was conducted using the Cochrane risk-of-bias tool. Although most studies (6/9, 67%) reported using randomization, some did not detail allocation concealment, potentially leading to selection bias. Only 2 studies were determined as having a low risk of implementation bias owing to the challenge of blinding researchers and participants in telemedicine intervention trials [24,32]. Approximately half (4/9, 45%) the studies blinded the outcome assessors, and thus, their risk of measurement bias was classified as low. Three studies were determined to have a high risk of attrition bias due to elevated loss to follow-up rates or a lack of appropriate data processing methods [32,35,37]. However, no selective reporting bias was detected in the included studies. Four studies were categorized as having a high risk of other biases due to baseline differences [24,34] and small sample sizes [31,32]. The results are shown in Figure 2.

Figure 2. Risk of bias in each study [23,24,31-37]. Red, green, and yellow colors indicate high, low, and unclear risk of bias, respectively.

Meta-Analysis

Caregiver Burden

A total of 5 studies that evaluated caregiver burden were included in the meta-analysis [23,31,35-37]. Since no significant heterogeneity was observed among the included studies (I^2=0%; *P*=.64), a fixed-effect model was used for merging the data. The results revealed that telemedicine intervention could mitigate the burden on informal caregivers (SMD −0.49, 95% CI −0.72 to −0.27; *P*<.001; 95% PI −0.86 to −0.13), as shown in Figure 3A. The sensitivity analysis showed that the results were stable, as shown in Figure 4A. The results remained unchanged when studies were merged using a random-effects model.

Figure 3. Forest plot of telemedicine versus control group: (A) caregiver burden, (B) anxiety, (C) depression, and (D) quality of life [23,24,31-37]. IV: inverse variance.

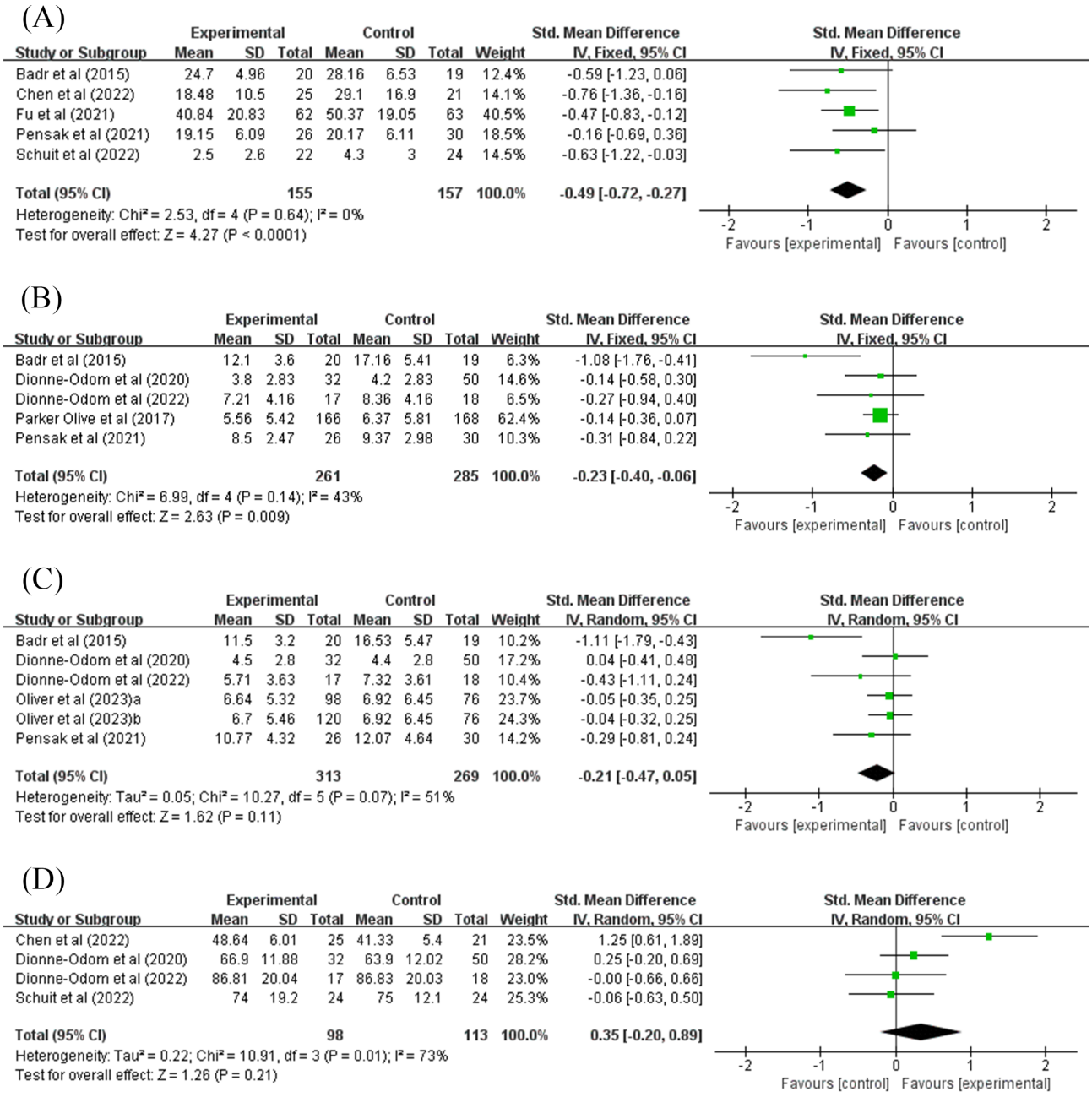
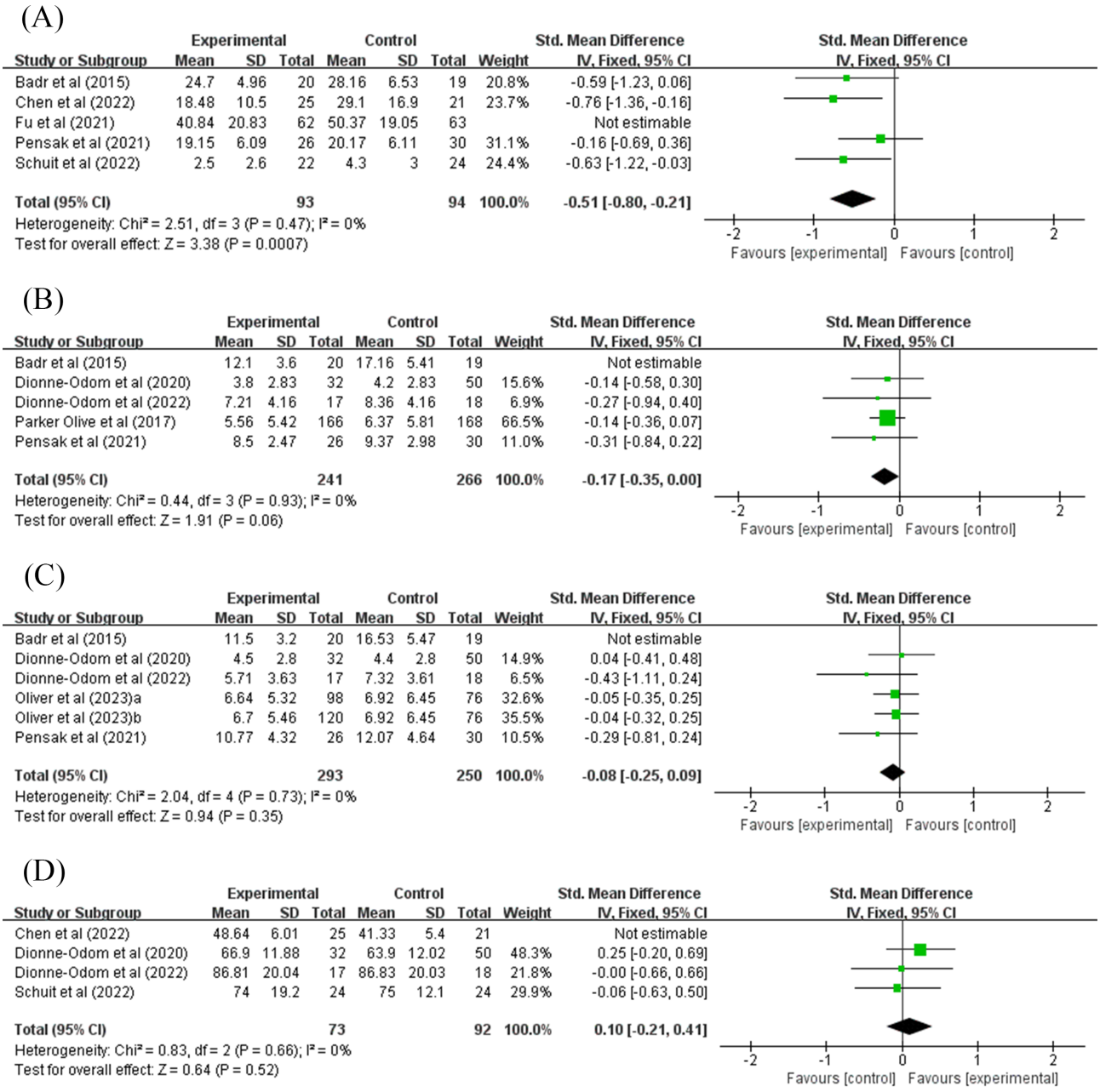


Figure 4. Sensitivity analyses: (A) caregiver burden, (B) anxiety, (C) depression, and (D) quality of life [23,24,31-37]. IV: inverse variance.



Anxiety

A total of 5 studies assessed the anxiety level of informal caregivers [24,31,32,34,35]. Due to mild heterogeneity among the included studies (I^2=43%; *P*=.14), a fixed-effect model was adopted to pool the data for analysis. The results demonstrated that telemedicine intervention could reduce informal caregivers' level of anxiety (SMD −0.23, 95% CI −0.40 to −0.06; *P*=.009; 95% PI −0.98 to 0.39), as shown in Figure 3B. The sensitivity analysis showed that after excluding Badr et al [31], the pooled results were the opposite, with I^2 decreasing to 0%, as shown in Figure 4B. This could be attributed to a higher risk of bias in this study. However, when merging studies using a random-effects model, the results still retained statistical significance (SMD −0.30, 95% CI −0.57 to −0.03; *P*=.03).

Depression

A total of 5 studies assessed the depression level of informal caregivers [24,31-33,35]. In the study by Oliver et al [33], "a" represents the ACCESS intervention and "b" represents the Facebook intervention. Due to moderate heterogeneity among the included studies (I^2=51%; *P*=.07), a random-effects model was used for merging the data. The analysis indicated that the telemedicine intervention did not result in a statistically significant difference in reducing depression among informal caregivers (SMD −0.21, 95% CI −0.47 to 0.05; *P*=.11; 95% PI −0.94 to 0.51), as shown in Figure 3C. Furthermore, the sensitivity analysis showed that no individual trial could change the results. However, after excluding Badr et al [31], the I^2 decreased to 0%, as shown in Figure 4C.

QOL of Caregivers

A total of 4 studies that assessed QOL were included in the meta-analysis [23,24,32,36]. Due to moderate heterogeneity among the studies (I^2=73%; P=.01), a random-effects model was used. The results indicated that the telemedicine intervention did not result in a statistically significant difference in improving the overall QOL of informal caregivers (SMD 0.35, 95% CI −0.20 to 0.89; P=.21; 95% PI −2.15 to 2.85), as shown in Figure 3D. Furthermore, the sensitivity analysis indicated that no individual trial could change the results. However, after excluding Chen et al [23], the I^2 decreased to 0%, as shown in Figure 4D.

Discussion

Principal Findings

The results of this review indicate that compared to conventional care, telemedicine interventions can alleviate the caregiving burden and anxiety of informal caregivers; however, they do not significantly alleviate depression or improve their QOL. The 95% PIs indicate considerable heterogeneity among the studies, and the effects of future telemedicine interventions on these outcomes remain uncertain, except for reducing caregiver burden.

Caregiver Burden

The results of the study demonstrated that telemedicine interventions could relieve the caregiving burden of informal caregivers, which is consistent with previous research [9,12,31]. The systematic review by Hu et al [9] demonstrates that internet-based interventions can effectively alleviate the stress of informal caregivers of patients with chronic diseases and improve their well-being. Chih et al [38] developed the Comprehensive Health Enhancement Support System for informal caregivers of patients with advanced cancer. The tool reduced the negative emotions of family caregivers and subsequently decreased their caregiving burden. Caregiver burden consists of both subjective and objective levels. Subjective burden includes the perceived physical, emotional, social, and economic difficulties caused by caring for individuals with serious diseases, whereas objective burden refers to the time and number of tasks devoted to patient care [39]. Telemedicine facilitates health education, assists in decision-making, helps develop problem-solving skills, and provides social support. It also improves and conserves the resources and time of informal caregivers. Thus, it is conducive to alleviating the caregiving burden at both the subjective and objective levels.

Anxiety

The results revealed that telemedicine interventions can alleviate anxiety in informal caregivers, which is consistent with the findings of previous research [8,27]. Research indicates that the likelihood of anxiety occurring in caregivers of patients with advanced cancer is 3 times that of the general population [40]. Here, factors such as overwhelming nursing pressure, inadequate self-care, and the lack of supportive care can lead to anxiety [41,42]. Currently, the proposed interventions to reduce the anxiety of informal caregivers focus on psychological education, skill training, and treatment counseling [43]. In this case, telemedicine enables monitoring, assessing, and managing patient symptoms, which can enhance informal caregivers' symptom management skills [1,37]. Moreover, it allows them to join discussions on the disease and participate in clinical decisions [20,33,44]. This can help satisfy informal caregivers' information needs and enhance their caregiving confidence and ability. In addition, telemedicine can provide psychological interventions, improve interpersonal relationships, and offer training in stress management skills [31,32], thereby alleviating symptoms of anxiety. However, the sensitivity analysis indicated that the result was unstable. To ensure greater stability, it will be necessary to gather additional data for further investigation.

Depression

The results of our research indicate that telemedicine does not have a statistically significant effect on alleviating depression among caregivers, which is consistent with the findings of previous research [45]. In addition to influencing factors such as the high nursing stress and insufficient social support observed for anxiety, an increased economic burden may also contribute to the onset of depression [42,46]. Despite its potential advantages, telemedicine requires a stable internet connection and available electronic devices. The initial investment cost of such equipment may negatively affect informal caregivers [47]. Furthermore, researchers only offered counseling on disease knowledge and mental and emotional well-being, but not economic and welfare support. Subsequently, factors including a lower baseline depression level in the study population [24], small sample size, and significant differences in intervention measures in various studies may negatively impact the combined results. These findings differ from those of Northouse et al [48], possibly due to variations in the target population. The study by Northouse et al [48] focused on informal caregivers of patients with cancer. In contrast, our research noted higher loss to follow-up rates in the population with advanced diseases, potentially impeding the discovery of beneficial outcomes. Moreover, Northouse et al [48] conducted a self-controlled study, whereas we included RCTs in which conventional palliative care can alleviate depression in informal caregivers [46]. Consequently, the extent to which telemedicine can improve depression is limited.

Quality of Life

This study found that telemedicine does not significantly improve the QOL of informal caregivers of patients in palliative care, which is consistent with the findings of earlier research [25,49]. Most informal caregivers consistently place the needs of patients above their own [50], leading to various unmet supportive care needs, such as physical, psychological, and social needs [6,51], and a subsequent decline in QOL. As QOL is a multidimensional construct, a multidisciplinary intervention is often more effective than single-faceted approaches. However, the majority of studies (3/4, 75%) in this review targeted interventions at the social-psychological level, and the results might not be ideal. Furthermore, the small sample sizes in the included studies and variations in intervention measures may have limited the possibility of revealing meaningful results.

Finally, the effectiveness of intervention measures may further be moderated by other factors such as the characteristics of informal caregivers, preexisting mental health issues, and the caregiver-patient relationship. Therefore, future efforts should aim to devise personalized interventions for specific informal caregiver populations to ensure the best possible support.

Strengths and Limitations

This study meticulously adhered to the systematic review writing process, developed a comprehensive search strategy, and selected appropriate methods for meta-analysis. To enhance the reliability of the results, only RCT studies were included. However, this review also had limitations. First, we only included available data for the analysis. Missing data may impact the combined results of the meta-analysis. Second, some included studies, especially pilot studies, have small sample sizes, which requires a cautious approach to generalizing the results. Third, the inclusion of only English- and Chinese-language articles may lead to publication bias. Last, the included studies all measured the results immediately after intervention, without evaluating the persistence of the intervention effect. Therefore, in the future, large samples and high-quality research are required to further validate the intervention effects of telemedicine and explore the most suitable intervention duration for informal caregivers.

Implications for Practice and Future Research

The findings indicated that telemedicine interventions have beneficial effects on the informal caregivers of patients in palliative care. We recommend that professional palliative care personnel consider the needs of informal caregivers, incorporating telemedicine into care plans to optimize and complement existing health care measures. When implementing such interventions, several considerations arise. First, the needs of informal caregivers are diverse and require multidisciplinary team collaboration. Second, personalized interventions should be tailored based on the demographic characteristics of informal caregivers. Finally, cost-effectiveness should be considered. In this regard, we suggest that relevant organizations establish regulations to minimize health care costs as much as possible.

Moreover, a theoretical or conceptual framework can provide the foundation of interventions, drive their development, and facilitate the prediction and explanation of their mechanism to achieve the desired effect [22]. For example, in the included studies, Badr et al [31] conducted a study based on self-determination theory. The authors hypothesized that telephone-based psychosocial interventions could improve the mental state and burden of patients and informal caregivers. Those results were statistically significant. The study by Pensak et al [35] was based on cognitive behavioral stress management theory and provided stress management training to alleviate informal caregivers' burden. In contrast, studies lacking theoretical support failed to improve patient and informal caregiver outcomes. Therefore, it is recommended that researchers have a relevant theoretical foundation when devising intervention measures to enhance the intervention effect, which will be more likely to benefit informal caregivers.

Conclusion

In summary, telemedicine can alleviate caregiving burden and anxiety in informal caregivers but does not significantly impact their depression and QOL. Despite certain outcomes lacking statistical significance, they retain clinical relevance for those engaged in family palliative care. We believe that support provided through telemedicine represents a viable means to ensure the continuity of care, address the needs of informal caregivers, and foster favorable outcomes. Future studies that involve large samples and high-quality research are still required to further validate the effects of telemedicine. Furthermore, intervention measures should be designed with a solid theoretical basis to the fullest extent.

Acknowledgments

This study was supported by the scientific research project of the Chinese Nursing Association (ZHKY202101). The funder had no role in the study design, data extraction process, data analysis, results interpretation, or manuscript preparation.

Authors' Contributions

X Yang and X Yu were responsible for the topic and design of this review. X Yu obtained funding, and X Yang drafted the manuscript. X Yang, XL, and SJ performed the study selection and data extraction. X Yu supervised the project. All authors were responsible for the analysis and interpretation. All authors reviewed and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOCX File, 22 KB - mhealth_v12i1e54244_app1.docx]

Multimedia Appendix 2
Characteristics of the included studies.
[DOCX File, 19 KB - mhealth_v12i1e54244_app2.docx]

Checklist 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[PDF File, 148 KB - mhealth_v12i1e54244_app3.pdf]

References

1. Bauer EH, Schultz A, Brandt F, Smith AC, Bollig G, Dieperink KB. Protocol for an integrative review: patient and families' perspectives on telehealth in palliative care. BMJ Open 2022 Sep 7;12(9):e062723. [doi: 10.1136/bmjopen-2022-062723] [Medline: 36688736]
2. Steindal SA, Nes AAG, Godskesen TE, et al. Patients' experiences of telehealth in palliative home care: scoping review. J Med Internet Res 2020 May 5;22(5):e16218. [doi: 10.2196/16218] [Medline: 32369037]
3. Gomes B, Calanzani N, Curiale V, McCrone P, Higginson IJ. Effectiveness and cost-effectiveness of home palliative care services for adults with advanced illness and their caregivers. Cochrane Database Syst Rev 2013 Jun 6;2013(6):CD007760. [doi: 10.1002/14651858.CD007760.pub2] [Medline: 23744578]
4. Michaels J, Chen C, Ann Meeker M. Navigating the caregiving abyss: a metasynthesis of how family caregivers manage end-of-life care for older adults at home. Palliat Med 2022 Jan;36(1):81-94. [doi: 10.1177/02692163211042999] [Medline: 34477025]
5. Candy B, Jones L, Drake R, Leurent B, King M. Interventions for supporting informal Caregivers of patients in the terminal phase of a disease. Cochrane Database Syst Rev 2011 Jun 15(6):CD007617. [doi: 10.1002/14651858.CD007617.pub2] [Medline: 21678368]
6. Williams AM, Wang L, Kitchen P. Impacts of care-giving and sources of support: a comparison of end-of-life and non-end-of-life caregivers in Canada. Health Soc Care Community 2016 Mar;24(2):214-224. [doi: 10.1111/hsc.12205] [Medline: 25727849]
7. Valero-Cantero I, Casals C, Carrión-Velasco Y, Barón-López FJ, Martínez-Valero FJ, Vázquez-Sánchez M. The influence of symptom severity of palliative care patients on their family Caregivers. BMC Palliat Care 2022 Feb 28;21(1):27. [doi: 10.1186/s12904-022-00918-3] [Medline: 35227246]
8. Sherifali D, Ali MU, Ploeg J, et al. Impact of internet-based interventions on caregiver mental health: systematic review and meta-analysis. J Med Internet Res 2018 Jul 3;20(7):e10668. [doi: 10.2196/10668] [Medline: 29970358]
9. Hu C, Kung S, Rummans TA, Clark MM, Lapid MI. Reducing caregiver stress with Internet-based interventions: a systematic review of open-label and randomized controlled trials. J Am Med Inform Assoc 2015 Apr;22(e1):e194-e209. [doi: 10.1136/amiajnl-2014-002817] [Medline: 25125686]
10. Payne S, Tanner M, Hughes S. Digitisation and the patient-professional relationship in palliative care. Palliat Med 2020 Apr;34(4):441-443. [doi: 10.1177/0269216320911501] [Medline: 32253996]
11. Palliative care. World Health Organization. 2020 Aug 5. URL: https://www.who.int/news-room/fact-sheets/detail/palliative-care [accessed 2024-02-07]
12. Pop RS, Puia A, Mosoiu D. Factors influencing the quality of life of the primary caregiver of a palliative patient: narrative review. J Palliat Med 2022 May;25(5):813-829. [doi: 10.1089/jpm.2021.0322] [Medline: 35007441]
13. Ullrich A, Ascherfeld L, Marx G, Bokemeyer C, Bergelt C, Oechsle K. Quality of life, psychological burden, needs, and satisfaction during specialized inpatient palliative care in family caregivers of advanced cancer patients. BMC Palliat Care 2017 May 10;16(1):31. [doi: 10.1186/s12904-017-0206-z] [Medline: 28486962]
14. Kershaw T, Ellis KR, Yoon H, Schafenacker A, Katapodi M, Northouse L. The interdependence of advanced cancer patients' and their family caregivers' mental health, physical health, and self-efficacy over time. Ann Behav Med 2015 Dec;49(6):901-911. [doi: 10.1007/s12160-015-9743-y] [Medline: 26489843]
15. Dionne-Odom JN, Hull JG, Martin MY, et al. Associations between advanced cancer patients' survival and family caregiver presence and burden. Cancer Med 2016 May;5(5):853-862. [doi: 10.1002/cam4.653] [Medline: 26860217]
16. Huang Z, Wu S, Yu T, Hu A. Efficacy of telemedicine for patients with chronic wounds: a meta-analysis of randomized controlled trials. Adv Wound Care (New Rochelle) 2021 Feb;10(2):103-112. [doi: 10.1089/wound.2020.1169] [Medline: 32320355]
17. Pengput A, Schwartz DG. Telemedicine in Southeast Asia: a systematic review. Telemed J E Health 2022 Dec;28(12):1711-1733. [doi: 10.1089/tmj.2021.0516] [Medline: 35417250]
18. Chen L, Cheng L, Gao W, Chen D, Wang C, Ran X. Telemedicine in chronic wound management: systematic review and meta-analysis. JMIR Mhealth Uhealth 2020 Jun 25;8(6):e15574. [doi: 10.2196/15574] [Medline: 32584259]
19. Timpel P, Oswald S, Schwarz PEH, Harst L. Mapping the evidence on the effectiveness of telemedicine interventions in diabetes, dyslipidemia, and hypertension: an umbrella review of systematic reviews and meta-analyses. J Med Internet Res 2020 Mar 18;22(3):e16791. [doi: 10.2196/16791] [Medline: 32186516]
20. Finucane AM, O'Donnell H, Lugton J, Gibson-Watt T, Swenson C, Pagliari C. Digital health interventions in palliative care: a systematic meta-review. NPJ Digit Med 2021 Apr 6;4(1):64. [doi: 10.1038/s41746-021-00430-7] [Medline: 33824407]
21. Starr LT, Washington KT, McPhillips MV, Pitzer K, Demiris G, Oliver DP. "It was terrible, I didn't sleep for two years": a mixed methods exploration of sleep and its effects among family caregivers of in-home hospice patients at end-of-life. Palliat Med 2022 Dec;36(10):1504-1521. [doi: 10.1177/02692163221122956] [Medline: 36151698]

22. Zhai S, Chu F, Tan M, Chi NC, Ward T, Yuwen W. Digital health interventions to support family caregivers: an updated systematic review. Digit Health 2023;9:20552076231171967. [doi: 10.1177/20552076231171967] [Medline: 37223775]
23. Chen Y, Sun L, Xiao H, Zheng J, Lin X. Evaluation of a WeChat-based dyadic life review program for people with advanced cancer and family caregivers: a mixed-method feasibility study. Palliat Med 2022 Mar;36(3):498-509. [doi: 10.1177/02692163211066736] [Medline: 35184629]
24. Dionne-Odom JN, Ejem DB, Wells R, et al. Effects of a telehealth early palliative care intervention for family caregivers of persons with advanced heart failure: the ENABLE CHF-PC randomized clinical trial. JAMA Netw Open 2020 Apr 1;3(4):e202583. [doi: 10.1001/jamanetworkopen.2020.2583] [Medline: 32282044]
25. Zheng Y, Head BA, Schapmire TJ. A systematic review of telehealth in palliative care: caregiver outcomes. Telemed J E Health 2016 Apr;22(4):288-294. [doi: 10.1089/tmj.2015.0090] [Medline: 26360181]
26. Cameron P, Munyan K. Systematic review of telehospice telemedicine and e-health. Telemed J E Health 2021 Nov;27(11):1203-1214. [doi: 10.1089/tmj.2020.0451] [Medline: 33512303]
27. Chen Y, Qiao C, Zhang X, Li W, Yang H. The effect of TELE-palliative care on patient and caregiver outcomes: a systematic review. Am J Hosp Palliat Care 2023 Aug;40(8):907-925. [doi: 10.1177/10499091221123566] [Medline: 36113129]
28. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:71. [doi: 10.1136/bmj.n71] [Medline: 33782057]
29. Higgins JPT, Thomas J, Chandler J, et al. Cochrane Handbook for Systematic Reviews of Interventions (Version 64): The Cochrane Collaboration; 2023. URL: https://training.cochrane.org/handbook/current [accessed 2023-08-27]
30. IntHout J, Ioannidis JPA, Rovers MM, Goeman JJ. Plea for routinely presenting prediction intervals in meta-analysis. BMJ Open 2016 Jul 12;6(7):e010247. [doi: 10.1136/bmjopen-2015-010247] [Medline: 27406637]
31. Badr H, Smith CB, Goldstein NE, Gomez JE, Redd WH. Dyadic psychosocial intervention for advanced lung cancer patients and their family caregivers: results of a randomized pilot trial. Cancer 2015 Jan 1;121(1):150-158. [doi: 10.1002/cncr.29009] [Medline: 25209975]
32. Dionne-Odom JN, Azuero A, Taylor RA, et al. A lay navigator-led, early palliative care intervention for African American and rural family caregivers of individuals with advanced cancer (Project Cornerstone): results of a pilot randomized trial. Cancer 2022 Mar 15;128(6):1321-1330. [doi: 10.1002/cncr.34044] [Medline: 34874061]
33. Oliver DP, Washington K, Benson J, et al. Access for cancer caregivers to education and support for shared decision making (ACCESS) intervention: a cluster cross-over randomised clinical trial. BMJ Support Palliat Care 2023 Mar 2:spcare-2022-004100. [doi: 10.1136/spcare-2022-004100] [Medline: 36863862]
34. Parker Oliver D, Demiris G, Washington K, Kruse RL, Petroski G. Hospice family caregiver involvement in care plan meetings: a mixed-methods randomized controlled trial. Am J Hosp Palliat Care 2017 Nov;34(9):849-859. [doi: 10.1177/1049909116661816] [Medline: 27465403]
35. Pensak NA, Carr AL, Jones J, et al. A pilot study of mobilized intervention to help caregivers of oncology patients manage distress. Psychooncology 2021 Apr;30(4):520-528. [doi: 10.1002/pon.5597] [Medline: 33217070]
36. Schuit AS, Rienks MM, Hooghiemstra N, et al. Reach and efficacy of the eHealth application Oncokompas, facilitating partners of incurably ill cancer patients to self-manage their caregiver needs: a randomized controlled trial. Support Care Cancer 2022 Dec;30(12):10191-10201. [doi: 10.1007/s00520-022-07441-4] [Medline: 36367563]
37. Fu YZ, Wang JX, Huang L. Application of “Internet+” extension model in family caregivers of advanced cancer [Article in Chinese]. Bachu Medical Journal 2021;4(4):44-49. [doi: 10.3969/j.issn.2096-6113.2021.04.011]
38. Chih MY, DuBenske LL, Hawkins RP, et al. Communicating advanced cancer patients' symptoms via the internet: a pooled analysis of two randomized trials examining caregiver preparedness, physical burden, and negative mood. Palliat Med 2013 Jun;27(6):533-543. [doi: 10.1177/0269216312457213] [Medline: 22988042]
39. Soto-Guerrero S, Palacios J, Langer P, et al. Objective burden, caregiver psychological distress, and patient religion and quality of life are associated with high-intensity burden of care among caregivers of advanced cancer patients in a Latino population. Palliat Support Care 2023 Jun 26:1-9. [doi: 10.1017/S1478951523000834] [Medline: 37357922]
40. Trevino KM, Prigerson HG, Maciejewski PK. Advanced cancer caregiving as a risk for major depressive episodes and generalized anxiety disorder. Psychooncology 2018 Jan;27(1):243-249. [doi: 10.1002/pon.4441] [Medline: 28426918]
41. Moss KO, Kurzawa C, Daly B, Prince-Paul M. Identifying and addressing family caregiver anxiety. J Hosp Palliat Nurs 2019 Feb;21(1):14-20. [doi: 10.1097/NJH.0000000000000489] [Medline: 30608354]
42. Tay DL, Iacob E, Reblin M, et al. What contextual factors account for anxiety and depressed mood in hospice family caregivers? Psychooncology 2022 Feb;31(2):316-325. [doi: 10.1002/pon.5816] [Medline: 34510631]
43. Northouse LL, Katapodi MC, Song L, Zhang L, Mood DW. Interventions with family caregivers of cancer patients: meta-analysis of randomized trials. CA Cancer J Clin 2010;60(5):317-339. [doi: 10.3322/caac.20081] [Medline: 20709946]
44. Rogante M, Giacomozzi C, Grigioni M, Kairy D. Telemedicine in palliative care: a review of systematic reviews. Ann Ist Super Sanita 2016;52(3):434-442. [doi: 10.4415/ANN_16_03_16] [Medline: 27698303]
45. Bauernschmidt D, Hirt J, Langer G, et al. Technology-based counselling for people with dementia and their informal carers: a systematic review and meta-analysis. J Alzheimers Dis 2023;93(3):891-906. [doi: 10.3233/JAD-221194] [Medline: 37125549]

46. Alam S, Hannon B, Zimmermann C. Palliative care for family caregivers. J Clin Oncol 2020 Mar 20;38(9):926-936. [doi: 10.1200/JCO.19.00018] [Medline: 32023152]
47. Walton L, Courtright K, Demiris G, Gorman EF, Jackson A, Carpenter JG. Telehealth palliative care in nursing homes: a scoping review. J Am Med Dir Assoc 2023 Mar;24(3):356-367. [doi: 10.1016/j.jamda.2023.01.004] [Medline: 36758619]
48. Northouse L, Schafenacker A, Barr KLC, et al. A tailored web-based psychoeducational intervention for cancer patients and their family caregivers. Cancer Nurs 2014;37(5):321-330. [doi: 10.1097/NCC.0000000000000159] [Medline: 24945270]
49. Söylemez BA, Özgül E, Küçükgüçlü Ö, Yener G. Telehealth applications used for self-efficacy levels of family caregivers for individuals with dementia: a systematic review and meta-analysis. Geriatr Nurs 2023;49:178-192. [doi: 10.1016/j.gerinurse.2022.12.001] [Medline: 36565592]
50. Norinder M, Årestedt K, Lind S, et al. Higher levels of unmet support needs in spouses are associated with poorer quality of life - a descriptive cross-sectional study in the context of palliative home care. BMC Palliat Care 2021 Aug 28;20(1):132. [doi: 10.1186/s12904-021-00829-9] [Medline: 34454454]
51. Cheng HL, Leung DYP, Ko PS, et al. Mediating role of self-efficacy between unmet needs and quality of life in palliative cancer caregivers. Psychooncology 2023 Mar;32(3):457-464. [doi: 10.1002/pon.6099] [Medline: 36650108]

Abbreviations

PI: prediction interval
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QOL: quality of life
RCT: randomized controlled trial
SMD: standardized mean differences

Edited by L Buis; submitted 03.11.23; peer-reviewed by C Laranjeira, S Zhai; revised version received 28.02.24; accepted 28.02.24; published 08.04.24.

Please cite as:
Yang X, Li X, Jiang S, Yu X
Effects of Telemedicine on Informal Caregivers of Patients in Palliative Care: Systematic Review and Meta-Analysis
JMIR Mhealth Uhealth 2024;12:e54244
URL: https://mhealth.jmir.org/2024/1/e54244
doi:10.2196/54244

Accuracy of Fitbit Charge 4, Garmin Vivosmart 4, and WHOOP Versus Polysomnography: Systematic Review

An-Marie Schyvens[1,2], MSc; Nina Catharina Van Oost[3], MSc; Jean-Marie Aerts[3], MSc, PhD; Federica Masci[3], MSc, PhD; Brent Peters[4], MSc; An Neven[4], MSc, PhD; Hélène Dirix[4], MSc, PhD; Geert Wets[4], MSc, PhD; Veerle Ross[4,5], MSc, PhD; Johan Verbraecken[1,2], MD, PhD

[1]Multidisciplinary Sleep Disorders Centre, Antwerp University Hospital, Edegem, Belgium

[2]Laboratory of Experimental Medicine and Pediatrics, University of Antwerp, Wilrijk, Belgium

[3]Department of Biosystems, KU Leuven, Leuven, Belgium

[4]Transportation Research Institute (IMOB), School of Transportation Sciences, UHasselt, Hasselt, Belgium

[5]Faresa, Evidence-Based Psychological Centre, Hasselt, Belgium

Corresponding Author:
An-Marie Schyvens, MSc
Multidisciplinary Sleep Disorders Centre, Antwerp University Hospital, , Edegem, , Belgium

Abstract

Background: Despite being the gold-standard method for objectively assessing sleep, polysomnography (PSG) faces several limitations as it is expensive, time-consuming, and labor-intensive; requires various equipment and technical expertise; and is impractical for long-term or in-home use. Consumer wrist-worn wearables are able to monitor sleep parameters and thus could be used as an alternative for PSG. Consequently, wearables gained immense popularity over the past few years, but their accuracy has been a major concern.

Objective: A systematic review of the literature was conducted to appraise the performance of 3 recent-generation wearable devices (Fitbit Charge 4, Garmin Vivosmart 4, and WHOOP) in determining sleep parameters and sleep stages.

Methods: Per the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement, a comprehensive search was conducted using the PubMed, Web of Science, Google Scholar, Scopus, and Embase databases. Eligible publications were those that (1) involved the validity of sleep data of any marketed model of the candidate wearables and (2) used PSG or an ambulatory electroencephalogram monitor as a reference sleep monitoring device. Exclusion criteria were as follows: (1) incorporated a sleep diary or survey method as a reference, (2) review paper, (3) children as participants, and (4) duplicate publication of the same data and findings.

Results: The search yielded 504 candidate articles. After eliminating duplicates and applying the eligibility criteria, 8 articles were included. WHOOP showed the least disagreement relative to PSG and Sleep Profiler for total sleep time (−1.4 min), light sleep (−9.6 min), and deep sleep (−9.3 min) but showed the largest disagreement for rapid eye movement (REM) sleep (21.0 min). Fitbit Charge 4 and Garmin Vivosmart 4 both showed moderate accuracy in assessing sleep stages and total sleep time compared to PSG. Fitbit Charge 4 showed the least disagreement for REM sleep (4.0 min) relative to PSG. Additionally, Fitbit Charge 4 showed higher sensitivities to deep sleep (75%) and REM sleep (86.5%) compared to Garmin Vivosmart 4 and WHOOP.

Conclusions: The findings of this systematic literature review indicate that the devices with higher relative agreement and sensitivities to multistate sleep (ie, Fitbit Charge 4 and WHOOP) seem appropriate for deriving suitable estimates of sleep parameters. However, analyses regarding the multistate categorization of sleep indicate that all devices can benefit from further improvement in the assessment of specific sleep stages. Although providers are continuously developing new versions and variants of wearables, the scientific research on these wearables remains considerably limited. This scarcity in literature not only reduces our ability to draw definitive conclusions but also highlights the need for more targeted research in this domain. Additionally, future research endeavors should strive for standardized protocols including larger sample sizes to enhance the comparability and power of the results across studies.

(JMIR Mhealth Uhealth 2024;12:e52192) doi:10.2196/52192

KEYWORDS

sleep; wearable device; validation; polysomnography; assessing sleep; PRISMA; Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Introduction

Sleep problems have emerged as a widespread concern with implications on health and quality of life for many people worldwide [1]. It has been suggested that 67% of adults worldwide have sleep problems [2,3]. The amount and quality of sleep that someone enjoys have a lasting impact during wakefulness. It affects mental health; physical well-being; and even the risk of developing lifestyle diseases such as cardiovascular diseases, obesity, depression, and type 2 diabetes [4-7]. Considering that sleep is vital to our health and quality of life, it is reasonable to wonder how long someone actually sleeps each night and if someone is getting enough restful and restorative sleep to keep the body and mind in optimal condition. Hence, this is why sleep tracking has gained immense popularity over the past few years. The majority of sleep trackers provide data on sleep architecture and hypnograms through their associated apps, offering insights into sleep stages and patterns [8]. In addition, these wearables can notify you about specific factors that might be affecting your sleep patterns such as drinking water, exercise, meditation, and regular bedtimes. As such, they can be a useful tool to obtain more insights into sleeping habits and patterns and to help optimize your sleep hygiene and quality. However, to improve sleep, an accurate, objective measurement is mandatory.

Polysomnography (PSG) is the gold-standard method for objectively assessing sleep. PSG records signals of brain activity, eye movements, and muscle tone, as well as audio and video, enabling it to classify sleep stages [9]. However, PSG may not be ideal for monitoring sleep in particular settings, as it is expensive, labor-intensive, and time-consuming; requires various equipment and technical expertise; and is impractical for long-term use or in-home environment settings [10,11]. PSG involves real-time monitoring of various parameters. In addition, applying and removing the sensors, organizing the patient administration, and thoroughly analyzing the data that PSG add is quite labor-intensive for sleep technicians. In addition to applying and removing the sensors and possibly completing questionnaires and other administrative tasks, PSG requires an overnight stay at a sleep clinic or laboratory, which makes PSG time-consuming for both the sleep technicians and patients. Due to the inherent limitations of PSG, several alternatives have been proposed. First, the use of sleep diaries is inexpensive and straightforward for consumers, but the subjective self-ratings they require result in frequent inaccuracies and incompleteness [12]. Additionally, they fail to measure sleep architecture and stages. Second, electroencephalogram (EEG) wearables can provide a home evaluation of sleep architecture and stages; however, they come with a high cost and can be technologically complex [12]. Therefore, wearables using an accelerometer and photoplethysmography (PPG) are being explored as a feasible alternative, largely due to their lower cost, convenience, and ability to measure sleep in clinical and personal settings [13].

Accelerometers and PPG sensors monitor different physiological and movement patterns throughout the night. Accelerometers are small, electromechanical devices that measure acceleration along multiple axes (usually 3: x, y, and z) to detect position changes, turning over, or significant body movements during the night [14]. Due to the body movement variations specific to each sleep stage, accelerometers can provide information about wakefulness and general sleep stages. However, they may tend to overestimate sleep due to poorly distinguishing between sleep and sedentary supine wake periods (eg, lying down while reading or watching television), or they could underestimate sleep due to potential body movements during sleep being categorized as awakenings [12,15-18]. In addition, they may not be as accurate in distinguishing between different non–rapid eye movement (REM) stages and detecting subtle changes in sleep architecture [14]. By combining an accelerometer with a PPG sensor in wearables, a more comprehensive and accurate assessment of sleep could be provided. PPG sensors are a noninvasive technology that uses a light source and a photodetector at the surface of the skin to measure the volumetric variations of blood circulation and thus can be used to monitor heart rate, heart rate variability, blood flow, and blood oxygen levels [19-21]. Due to the specific cardiovascular features of each sleep stage, PPG can provide more information about the sleep stages in addition to the accelerometer [21,22]. The benefits of these sensors used in wearables are their low-cost, noninvasive nature and their ability to provide continuous monitoring and real-time data. However, the readings of PPG can be affected by motion artifacts, skin pigmentation, or tissue thickness. In addition, they could be susceptible to environmental factors such as ambient light and temperature [23-25].

Although many have doubts about their accuracy in monitoring sleep, wearable sleep-tracking devices are widely used and becoming more technologically advanced, creating strong interest from researchers and clinicians for their possible use as alternatives to PSG. This was demonstrated by Nguyen et al [26], who used wearables that provide inactivity alerts and personal feedback to increase physical activity and improve sleep for survivors of breast cancer [26].

Given the disadvantages of PSG and the corresponding growing popularity of wearable devices for sleep tracking among consumers and medical organizations, the objective of this paper was to appraise the performance of recent-generation wearable devices in determining sleep parameters and sleep stages through a review of relevant publications. To limit the overwhelming amount of wearables and their corresponding research papers, we performed a search to select a limited number of recent, frequently used wearables, using the following criteria: recent generation; good ease of use (affordable, unobtrusive, and sufficient battery life); and assessment of variables that could also be used for monitoring sleep, stress, fatigue, and sleepiness—namely, heart rate, heart rate variability, stress indicator, and activity. The candidate wearables selected up-front were Fitbit Charge 4, Garmin Vivosmart 4, and WHOOP.

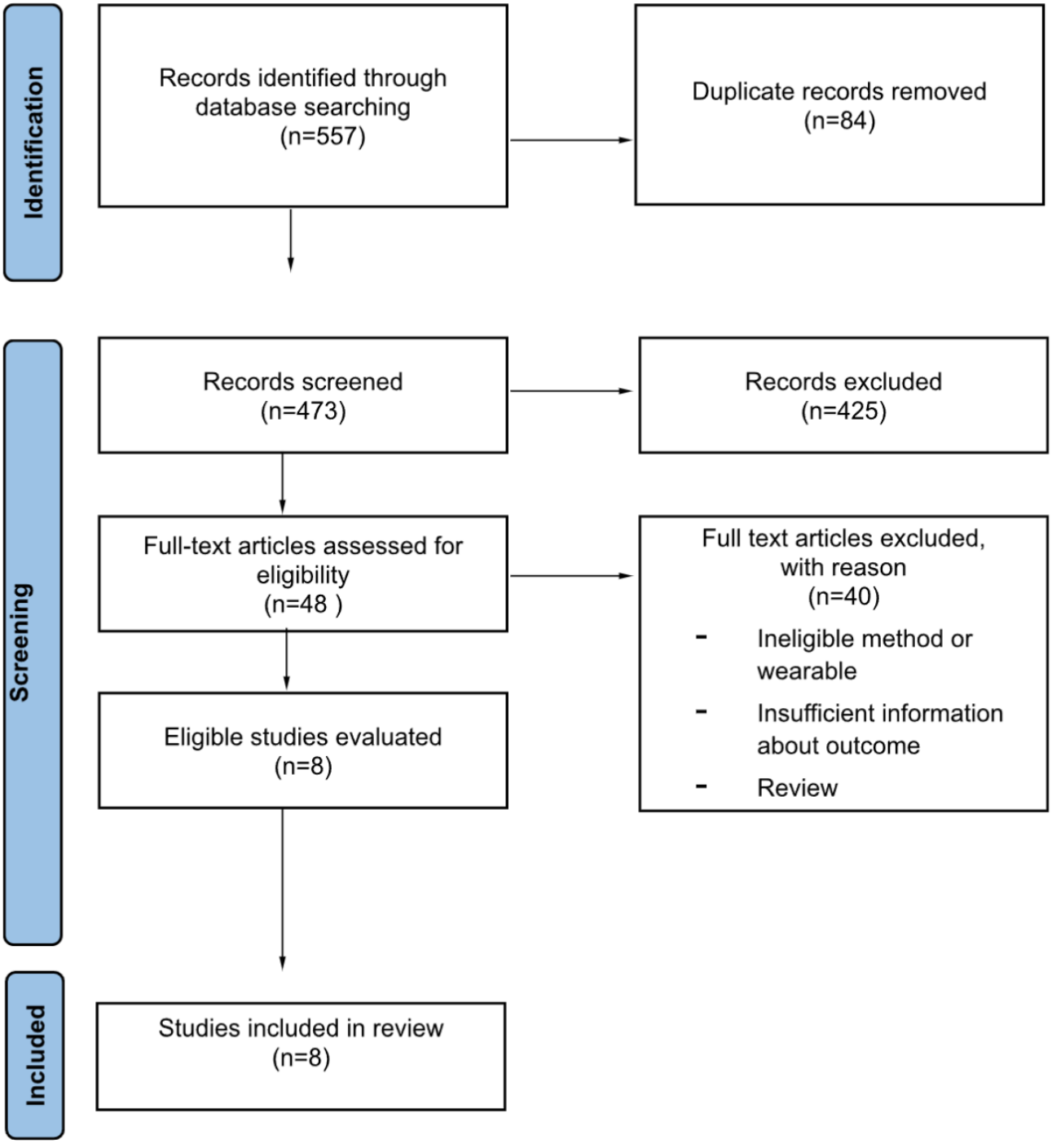
Methods

Search Strategies

In adherence with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement (Checklist 1), a comprehensive search using the PubMed, Web of Science, Google Scholar, Scopus, and Embase databases was conducted

[27]. Relevant keywords such as "validity," "accuracy," "assessment," "performance," "wearable," "sleep tracker," "sleep-tracking," "polysomnography," "wristband," "Whoop," "Fitbit Charge 4," and "Garmin Vivosmart 4" were used (see Multimedia Appendix 1). This search was initially completed by May 16, 2023, and was repeated by November 23, 2023 (see Figure 1). This systematic literature review was not registered and a protocol was not predefined.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram describing the search strategy of databases to retrieve and qualify publications of relevance for this systematic literature review.



Eligibility Criteria

Retrieved publications qualified for the review if they (1) involved the validity of sleep data of any marketed model of the candidate wearables and (2) incorporated PSG or an ambulatory EEG monitor as a reference sleep monitoring device. Exclusion criteria were as follows: (1) incorporated a sleep diary or survey method as a reference, (2) review paper, (3) children as participants, and (4) duplicate publication of the same data and findings.

Data Extraction

The following items were extracted: type of sleep tracker; number, gender, and age of participants; number of nights of sleep assessment; reference sleep monitoring device; and study outcomes relative to the denoted reference standard—the

precision of measuring the parameters of total sleep time (TST), light sleep (LS), deep sleep (DS), and REM sleep, as well as the sensitivity to sleep (the proportion of correctly classified sleep epochs by the wearable); specificity for sleep (the proportion of correctly classified wake epochs by the wearable); agreement; and if applicable, Cohen κ for multistate categorization of sleep periods. These Cohen κ values are a measure of interdevice reliability, often used in the context of validation studies where 2 or more methods of devices are used to assess a particular characteristic or condition [28]. In the case of wearable validation studies for sleep, κ values are often used to assess the agreement between the wearable device's sleep detection algorithm and a reference standard, such as PSG. In addition to the observed agreement between the devices, κ values also take into account the possibility that the agreement comes by chance. The interpretation of κ values is often categorized as follows: values ≤0 indicate no agreement, 0.01-0.20 indicate none to slight agreement, 0.21-0.40 indicate fair agreement, 0.41-0.60 indicate moderate agreement, 0.61-0.80 indicate substantial agreement, and 0.81-1.00 indicate almost perfect agreement [28].

Results

Overview of Included Studies

Figure 1 presents a visual summary of the selection and qualification of articles for this review. A total of 8 publications were retrieved through a search of databases performed by May 2023 and again by November 2023. Table 1 presents an overview of all included studies and the extracted details of each qualifying study involving the different wearable models. Few papers that met the eligibility criteria for the candidate wearables were found; we found 3 papers that satisfy the criteria for Fitbit Charge 4 [28-30], 2 for Garmin Vivosmart 4 [27,28], and 4 for WHOOP [22,30-32].

Table . Summary table of included papers in this systematic literature review.

Author (year)	Wearable	Reference	Participants, n	Sex, n	Age (y), mean (SD)	Duration
Doheny et al [33] (2021)	Fitbit Charge 4	PSG[a]	2	• N/A[b]	N/A	1 night, laboratory based; followed by 7 nights at home
Renerts et al [34] (2022)	Fitbit Charge 4	PSG	8	• N/A	N/A	1 night, laboratory based
Dong et al [18] (2022)	Fitbit Charge 4	PSG	37	• Female: 20 • Male: 17	48.8 (2.1)	1 night, laboratory based
Mouritzen et al [29] (2020)	Garmin Vivosmart 4	PSG	18	• Female: 13 • Male: 5	56.1 (12.0)	1 night, laboratory based
Stone et al [30] (2020)	Garmin Vivosmart 4 and WHOOP	Sleep Profiler (ambulatory EEG[c] monitor)	5	• Female: 3 • Male: 2	27.8 (7.6)	Home environment, 98 nights for all study devices
Miller et al [31] (2020)	WHOOP	PSG	12	• Female: 6 • Male: 6	22.9 (3.4)	10 days, laboratory based
Miller et al [32] (2021)	WHOOP	PSG	6	• Female: 3 • Male: 3	23.0 (2.2)	9 nights, laboratory based
Miller et al [22] (2022)	WHOOP	PSG	53	• Male: 27 • Female: 26	25.4 (5.9)	1 night, laboratory based

[a]PSG: polysomnography.

[b]N/A: not applicable.

[c]EEG: electroencephalogram.

Participants were diverse: healthy adults as well as participants diagnosed with Huntington disease [33,34] and chronic insomnia [18]. Sample size varied substantially between investigations, from 2 to 53 participants. The average age of the participants was 34 years old. Out of the 8 studies, 6 (75%) were fully conducted in a sleep laboratory [18,22,29,31,32,34], 1 (12%) was conducted in the home environment [30], and 1 (12%) conducted an overnight laboratory-based PSG followed by 7 nights in the home environment [33]. In all, 3 (38%) laboratory-based studies had a duration of more than 1 night [30-32].

Comparison of Sleep Parameters Assessed by Wearables Versus PSG

Fitbit Charge 4

As shown in Table 2, of the 3 Fitbit Charge 4 versus PSG comparisons, 2 (67%) reported an overestimation of TST (5 and 23 minutes), whereas 1 (33%) reported a nonsignificant underestimation of 11 minutes. For LS, 2 (67%) of the 3 comparisons reported a similar significant overestimation (37.5 and 37.7 minutes). For DS, all 3 papers reported an underestimation (4.1, 12.5, and 41.4 minutes). REM sleep was

overestimated in 2 (67%) out of 3 papers by 5.2 and 11.5 minutes and underestimated in 1 (33%) paper by 4.7 minutes. Table 3 shows the results for the sensitivity to and specificity for sleep. The sensitivity to sleep was quite high for all 3 papers (89.9%-93.6%). In contrast, the results for specificity for sleep were quite divergent (48.8%, 62.2%, and 73%). The sensitivities to DS and REM sleep could only be extracted from 2 papers: 54% and 96% for DS and 76% and 97% for REM sleep.

Table . The total sleep time (TST), light sleep (LS), deep sleep (DS), and rapid eye movement (REM) sleep of Fitbit Charge 4, Garmin Vivosmart 4, and WHOOP versus polysomnography.

Wearable and paper		TST (min)		LS (min)		DS (min)		REM sleep (min)	
		Mean (SD)	*P* value	Mean (SD)	*P* value	Mean (SD)	*P* value	Mean (SD)	*P* value
Fitbit Charge 4									
	Doheny et al [33]	23 (N/A[a])	N/A	37.5 (N/A)	N/A	−12.5 (N/A)	N/A	11.5 (N/A)	N/A
	Renerts et al [34]	5 (26.8)	N/A	N/A	N/A	−4.1 (21.8)	N/A	5.21 (22.35)	N/A
	Dong et al [18]	−11.0 (N/A)	.16	37.69 (N/A)	.001	−41.38 (N/A)	<.0001	−4.7 (N/A)	.44
	Mean	5.67 (N/A)	N/A	37.6 (N/A)	N/A	−19.33 (N/A)	N/A	4 (N/A)	N/A
Garmin Vivosmart 4									
	Mouritzen et al [29]	27.8 (29.5)	.001	36.5 (71.7)	.045	13.4 (98.1)	.57	−22.1 (54.7)	.11
	Stone et al [30]	66 (N/A)	.06	19.8 (N/A)	1	33.6 (N/A)	1	−3 (N/A)	1
	Mean	46.9 (N/A)	N/A	28 (N/A)	N/A	23.5 (N/A)	N/A	−12.55 (N/A)	N/A
WHOOP									
	Stone et al [30]	16.2	.61	−10.2	1	1.8	1	6	1
	Miller et al [31]	8.2 (32.9)	.54	−3.7 (44.4)	.62	−3.7 (26.4)	.33	15.6 (39.7)	.01
	Miller et al [32]	−17.8 (61.1)	N/A	−8.9 (55.9)	<.05	−15.5 (30.1)	<.001	6.5 (39.5)	N/A
	Miller et al [22]	−12.2 (36.3)	N/A	−15.6 (50.7)	N/A	−19.6 (34.3)	N/A	22.9 (45.4)	N/A
	Mean	−1.4 (N/A)	N/A	−9.6 (N/A)	N/A	−9.25 (N/A)	N/A	12.75 (N/A)	N/A

[a]N/A: not applicable.

Table . The sensitivity to sleep, specificity for sleep, agreement, and Cohen κ coefficient for multistate categorization of Fitbit Charge 4, Garmin Vivosmart 4, and WHOOP versus polysomnography.

Wearable and paper		Sensitivity to sleep (%), mean (SD)	Specificity for sleep (%), mean (SD)	Sensitivity to light sleep (%), mean (SD)	Sensitivity to deep sleep (%), mean (SD)	Sensitivity to REM[a] sleep (%), mean (SD)	Agreement (%), mean (SD)	Cohen κ for multistate categorization, mean (SD)
Fitbit Charge 4								
	Doheny et al [33]	93.6 (2.6)	48.8 (17.7)	N/A[b]	96 (N/A)	97 (N/A)	N/A	N/A
	Renerts et al [34]	90 (N/A)	73 (N/A)	N/A	54 (N/A)	76 (N/A)	N/A	N/A
	Dong et al [18]	89.9 (4.0)	62.2 (26.2)	N/A	N/A	N/A	N/A	N/A
	Mean	91.2 (N/A)	61.3 (N/A)	N/A	75 (N/A)	86.5 (N/A)	NA	N/A
Garmin Vivosmart 4								
	Mouritzen et al [29]	98 (3)	30 (17)	60 (17)	45 (26)	34 (26)	48 (10)	0.20 (0.11)
	Stone et al [30]	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	Mean	98 (3)	30 (17)	60 (17)	45 (26)	34 (26)	48 (10)	N/A
WHOOP								
	Stone et al [30]	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	Miller et al [31]	95 (N/A)	51 (N/A)	62 (N/A)	68 (N/A)	70 (N/A)	64 (N/A)	0.47 (N/A)
	Miller et al [32]	90 (N/A)	60 (N/A)	61 (N/A)	64 (N/A)	66 (N/A)	63 (N/A)	0.47 (N/A)
	Miller et al [22]	90 (N/A)	56 (N/A)	58 (N/A)	62 (N/A)	66 (N/A)	60 (N/A)	0.44 (N/A)
	Mean	91.7 (N/A)	56 (N/A)	60 (N/A)	65 (N/A)	67 (N/A)	62 (N/A)	0.46 (N/A)

[a]REM: rapid eye movement.

[b]N/A: not applicable.

Garmin Vivosmart 4

Both papers comparing Garmin Vivosmart 4 to PSG or Sleep Profiler reported an overestimation of TST (27.8 and 66 minutes), as shown in Table 2. For LS, both papers again reported an overestimation (19.8 and 36.5 minutes). For DS, both papers reported an overestimation (13.4 and 33.6 minutes). On the contrary, REM sleep was underestimated by both papers by 3 and 22.1 minutes. The study from Mouritzen et al [29] reported sensitivities to LS, DS, and REM sleep of 60%, 45%, and 34%, respectively; a sensitivity to sleep of 98%; a specificity for sleep of 30%; an agreement of 48%; and a Cohen κ for multistate categorization of sleep periods of 0.20 (see Table 3).

WHOOP

As shown in Table 2, of the 4 WHOOP versus PSG comparisons, 2 (50%) reported an overestimation of TST (8.2 and 16.2 minutes), whereas the other 2 (50%) reported an underestimation (12.2 and 17.8 minutes). All 4 comparisons reported an underestimation of LS, from 3.7 to 15.6 minutes. For DS, 3 (75%) papers reported an underestimation (3.7, 15.5, and 19.6 minutes). Stone et al [30] reported a small overestimation of DS of 1.8 minutes. REM sleep was overestimated in all 4 studies, from 6 to 22.9 minutes. Table 3 shows that the studies from Miller et al [22,31,32] reported sensitivities to LS, DS, and REM sleep ranging from 58% to 62%, from 62% to 68%, and from 66% to 70%, respectively. The studies also reported a sensitivity to sleep ranging from 90% to 95%, a specificity for sleep ranging from 51% to 60%, an agreement ranging from 60% to 64%, and a Cohen κ for multistate categorization of sleep periods ranging from 0.44 to 0.47.

Comparison of the Mean Values of the Different Wearables

From Tables 2 and 3, the means of the different values extracted from the papers (ie, TST, LS, DS, REM sleep, sensitivity to sleep, specificity for sleep, sensitivity to LS, sensitivity to DS, sensitivity to REM sleep, agreement, and Cohen κ coefficient) were calculated, which are summarized in Table 4. Altogether, WHOOP deviated the least compared to the gold-standard PSG for TST, LS, and DP but showed the highest difference from PSG for REM sleep (ie, a mean overestimation of 21 min).

Table . Mean differences of the sleep parameters in minutes and means of sensitivities to sleep, specificity for sleep, and agreement assessed by the wearables compared to polysomnography.

Variable	Fitbit Charge 4	Garmin Vivosmart 4	WHOOP
TST[a] (min)	5.7	46.9	−1.4
LS[b] (min)	37.6	27.9	−9.6
DS[c] (min)	−19.2	23.5	−9.3
REM[d] sleep (min)	4.0	−12.5	21.0
Sensitivity to sleep (%)	91.2	98.0	91.7
Sensitivity to LS (%)	N/A[e]	60.0	60.0
Sensitivity to DS (%)	75.0	45.0	65.0
Sensitivity to REM sleep (%)	86.5	34.0	67.0
Specificity for sleep (%)	61.3	30.0	55.7
Agreement (%)	N/A	48	62
Cohen κ	N/A	0.20	0.46
Papers, n	3	2	4

[a]TST: total sleep time.

[b]LS: light sleep.

[c]DS: deep sleep.

[d]REM: rapid eye movement.

[e]N/A: not applicable.

Garmin Vivosmart 4 showed the largest deviations out of the 3 different wearables compared to PSG (ie, a mean overestimation of 46.9 minutes for TST, a mean overestimation of 27.9 minutes for LS, a mean overestimation of 23.5 minutes for DS, and a mean underestimation of 12.5 minutes for REM sleep). Additionally, the sensitivities to LS, DS, and REM sleep were lower compared to those of Fitbit Charge 4 and WHOOP.

The mean values of Fitbit Charge 4 deviated the least from PSG for REM sleep, with a mean overestimation of only 4 minutes. On the contrary, it showed the largest deviation to PSG for LS, with a mean overestimation of 37.6 minutes. For TST and DS, Fitbit Charge 4 showed on average better results than Garmin Vivosmart 4 but worse results than WHOOP, namely, a mean overestimation of 5.7 minutes and a mean underestimation of 19.2 minutes, respectively. The sensitivities to LS, DS, and REM sleep were higher compared to those of Garmin Vivosmart 4 and WHOOP.

Discussion

PSG is still the gold-standard method to objectively assess sleep. However, PSG is not ideal for monitoring sleep in particular settings and for long-term follow-up. To overcome these limitations, consumer sleep-tracking devices are becoming more widely used and technologically advanced, creating strong interest from researchers and clinicians for their possible application as alternatives to PSG. Since limited research has been performed to validate the different consumer sleep-tracking wearables, we aimed to review the available literature on the selected wearables to determine the most accurate, commercially available wrist-worn device that can be used in a clinical setting for long-term sleep monitoring.

Thus far, at most, only 7 relevant studies investigated the performance of Fitbit Charge 4, Garmin Vivosmart 4, and WHOOP against PSG. The study of Stone et al [30] investigated the performance of both Garmin Vivosmart 4 and WHOOP relative to Sleep Profiler. After reviewing these studies, the results suggest that WHOOP presented the least amount of disagreement relative to PSG and Sleep Profiler for TST, LS, and DS but showed the largest amount of disagreement for REM sleep. Fitbit Charge 4 and Garmin Vivosmart 4 both showed moderate accuracy in assessing sleep stages and TST compared to PSG. Fitbit Charge 4 showed the least amount of disagreement for REM sleep relative to PSG. In addition, Fitbit Charge 4 showed higher sensitivities for LS, DS, and REM sleep compared to Garmin Vivosmart 4 and WHOOP. Garmin Vivosmart 4 showed the lowest sensitivities to LS, DS, and REM sleep compared to Fitbit Charge 4 and WHOOP.

Some of the studies performed evaluations of the accuracy of wearables in detecting sleep by using an epoch-by-epoch analysis. It involves breaking down the continuous stream of sleep data of PSG into discrete time intervals called “epochs” [35]. Afterward, each epoch is compared individually to the corresponding epoch generated by the wearables. The results of the epoch-by-epoch analysis in this review showed high sensitivity to sleep, ranging from 91.2% to 98.0%, but lower specificity for sleep, ranging from 30% to 61%. The low specificity for sleep, or variability in specificity for sleep, is a commonly observed phenomenon in the validation of devices that primarily rely on actigraphy to estimate sleep [15,36,37].

The challenge of accurately separating wake episodes during sleep stems from the similarities in movement between restful wakefulness and sleep. Hence, it can be inferred that devices that have improved their ability to detect wake epochs during sleep have refined their proprietary algorithms to include metrics other than movement, such as heart rate and heart rate variability, in the detection of wakefulness [22].

Furthermore, it is crucial to contextualize the comparison of the wearables' agreement with that of PSG (see Multimedia Appendix 2), taking into consideration that the scoring of PSG is subject to variability among technicians [38]. As reported by Danker-Hopfe et al [39], the interrater reliability ranges from 86.5% to 97.5% depending on the sleep stage, with an overall accuracy of 81% and a Cohen κ coefficient of 0.7505 (see Table 5). Given this benchmark, Fitbit Charge 4 seems to provide reasonable estimations of multistate sleep. However, it is imperative to acknowledge that achieving the same level of accuracy as PSG may pose a significant challenge for wearables, particularly considering the observed low Cohen κ coefficients of the wearables. The finding of a low κ value indicates poor sleep stage differentiation despite including PPG signals. However, the beneficial role of PPG signals in accelerometer-based sleep tracking remains unilluminated, since we do not know how these signals are processed and applied [29].

Table . Sleep stage–specific degree of agreement according to American Academy of Sleep Medicine standards [38].

Stage	Agreement (%)	Cohen κ
Overall	81	0.7505
Wake	95.6	0.4608
REM[a]	97.5	0.9054
N3[b]	93.8	0.7285
N2[c]	86.5	0.7188
N1[d]	90.1	0.4608

[a]REM: rapid eye movement.

[b]N3: stage 3 non-REM sleep.

[c]N2: stage 2 non-REM sleep.

[d]N1: stage 1 non-REM sleep.

On a similar note, companies usually do not share the methodology they use to score the sleep data from the wearables, nor do they publish the kind of rigorous research sleep experts need to establish the credibility of the sleep reports they produce.

The findings of this systematic literature review about Fitbit Charge 4, Garmin Vivosmart 4, and WHOOP are based on our recent comprehensive search of databases for relevant published articles. The included research studies have certain limitations. For example, several investigations evaluating the sleep-tracking capabilities of these wearables involved a relatively limited number of participants, potentially impacting the generalizability of the results. Efforts were made to get additional information about the sample sizes and participants directly from the authors of some included studies. Unfortunately, despite our attempts to contact the authors, we did not receive a response. It is crucial to address these issues transparently, as they may impact the generalizability of our findings. Small sample sizes can introduce variability and limit the statistical power of the findings, underscoring the need for larger and more diverse cohorts to validate the devices' accuracy across different demographics and sleep conditions. Additionally, it is essential to consider that most of the studies have been conducted in controlled environments, namely sleep laboratories. Although this controlled setting allows for precise data collection and monitoring, it may not fully reflect the real-world sleep experiences of individuals in their natural environments [40]. Sleep laboratory conditions may differ substantially from home environments, where factors such as ambient light, noise, and personal sleep habits can vary widely [41]. Therefore, the generalizability of the findings from laboratory studies to everyday scenarios should be approached with caution. In addition to acknowledging the potential limitations associated with conducting studies in controlled sleep laboratory environments, it is important to recognize the presence of the "first-night effect" in both PSG and wearable sleep-tracking technologies. The first-night effect refers to the phenomenon where an individual's sleep patterns and quality may be altered due to the unfamiliarity with the sleep-monitoring setup, regardless of whether it occurs in a sleep laboratory or at home [42,43]. This phenomenon is not exclusive to sleep laboratories as it extends to home environments where individuals may experience similar disruptions during the initial adaptation to sleep-tracking devices [44,45]. Although some wearable users may initially find it uncomfortable or unfamiliar to wear a device on their wrist while sleeping, which could potentially impact their sleep quality on the first night of use, there is not a widely recognized first-night effect associated with the use of wrist-worn wearables [46]. In addition, the data of PSG and wearables were collected each time under the same circumstances and environmental factors either in a sleep laboratory or in the home environment, making the first-night effect less relevant. Another notable aspect of the studies is the variations in the duration of the validation of the wearables. The diverse durations across studies are likely due to practical constraints or differences in research protocols. Our decision to include studies with varying protocols was motivated by the aim to include as many relevant papers as possible in our

comprehensive review, considering the scarcity of literature on this topic. Future research endeavors should strive for standardized protocols including larger sample sizes to enhance the comparability and power of the results across studies.

Despite these limitations, the findings of this review indicate that the devices with higher relative agreement and sensitivities for multistate sleep (ie, Fitbit Charge 4 and WHOOP) seem appropriate for deriving suitable estimates of sleep parameters and could be used to monitor sustained, meaningful changes in sleep architecture (ie, time spent in different stages of sleep). However, analyses regarding the multistate categorization of sleep (as a specific sleep stage or wake) indicate that all devices can benefit from further improvement. Providers are continuously developing new versions and variants of wearables, which present difficulties for those undertaking independent validation studies. Nevertheless, it can be reasonably assumed that newer models from the same provider will perform at least as well, if not better, than older models when compared against the relevant gold standards [22]. However, although the wearable technology market keeps developing wearable devices, the scientific research on these wearables against PSG remains considerably limited. This scarcity in literature not only reduces our ability to draw definitive conclusions but also highlights the need for more targeted research in this domain. Therefore, the data presented here should not be considered obsolete when the models analyzed are superseded by newer models. Instead, these data should serve as the best approximation of the expected performance of any subsequent models that may be released.

Acknowledgments

This work was supported by Flanders Innovation & Entrepreneurship - VLAIO (Vlaams Agentschap Innoveren & Ondernemen; HBC.2021.0387).

Authors' Contributions

AMS and JV conceptualized and designed the structure of this systematic literature review, outlining the key themes and objectives. AMS conducted an extensive literature review, critically analyzed relevant studies, and synthesized the information to provide a comprehensive overview. AMS, NCVO, JMA, FM, BP, AN, HD, VR, GW, and JV actively participated in the review's writing and revision processes. The collaborative effort of the consortium resulted in a comprehensive and well-balanced review. Each author has approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search string. In adherence with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement, a comprehensive search of the PubMed, Web of Science, Google Scholar, Scopus, and Embase databases was conducted as shown in the search string.
[PDF File, 49 KB - mhealth_v12i1e52192_app1.pdf]

Multimedia Appendix 2
Normative values of sleep parameters. These normative values in healthy male and female individuals of different age groups (mean and SD), derived from the widely recognized Rechtschaffen and Kales scoring system of polysomnography, can provide an additional benchmark when delving into the comparative analysis of sleep parameters measured by wearables [47].
[PDF File, 21 KB - mhealth_v12i1e52192_app2.pdf]

Checklist 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[PDF File, 64 KB - mhealth_v12i1e52192_app3.pdf]

References

1. Lemoine P, Nir T, Laudon M, Zisapel N. Prolonged-release melatonin improves sleep quality and morning alertness in insomnia patients aged 55 years and older and has no withdrawal effects. J Sleep Res 2007 Dec;16(4):372-380. [doi: 10.1111/j.1365-2869.2007.00613.x] [Medline: 18036082]
2. Baranowski M, Jabkowski P. Gender and socioeconomic patterning of self-reported sleep problems across European countries. Eur J Public Health 2023 Apr 1;33(2):242-248. [doi: 10.1093/eurpub/ckad012] [Medline: 36805658]
3. Bhaskar S, Hemavathy D, Prasad S. Prevalence of chronic insomnia in adult patients and its correlation with medical comorbidities. J Family Med Prim Care 2016;5(4):780-784. [doi: 10.4103/2249-4863.201153] [Medline: 28348990]
4. Bernert RA, Kim JS, Iwata NG, Perlis ML. Sleep disturbances as an evidence-based suicide risk factor. Curr Psychiatry Rep 2015 Mar;17(3):554. [doi: 10.1007/s11920-015-0554-4] [Medline: 25698339]

5. Depner CM, Stothard ER, Wright KP Jr. Metabolic consequences of sleep and circadian disorders. Curr Diab Rep 2014 Jul;14(7):507. [doi: 10.1007/s11892-014-0507-z] [Medline: 24816752]
6. Hungin APS, Close H. Sleep disturbances and health problems: sleep matters. Br J Gen Pract 2010 May 1;60(574):319-320. [doi: 10.3399/bjgp10X484147]
7. St-Onge MP, Grandner MA, Brown D, et al. Sleep duration and quality: impact on lifestyle behaviors and cardiometabolic health: a scientific statement from the American Heart Association. Circulation 2016 Nov 1;134(18):e367-e386. [doi: 10.1161/CIR.0000000000000444] [Medline: 27647451]
8. Wulterkens BM, Fonseca P, Hermans LWA, et al. It is all in the wrist: wearable sleep staging in a clinical population versus reference polysomnography. Nat Sci Sleep 2021 Jun 28;13(885-97):885-897. [doi: 10.2147/NSS.S306808] [Medline: 34234595]
9. Kushida CA, Littner MR, Morgenthaler T, et al. Practice parameters for the indications for polysomnography and related procedures: an update for 2005. Sleep 2005 Apr;28(4):499-521. [doi: 10.1093/sleep/28.4.499] [Medline: 16171294]
10. Yi H, Shin K, Shin C. Development of the Sleep Quality Scale. J Sleep Res 2006 Sep;15(3):309-316. [doi: 10.1111/j.1365-2869.2006.00544.x] [Medline: 16911033]
11. de Zambotti M, Baker FC, Colrain IM. Validation of sleep-tracking technology compared with polysomnography in adolescents. Sleep 2015 Sep 1;38(9):1461-1468. [doi: 10.5665/sleep.4990] [Medline: 26158896]
12. Haghayegh S, Khoshnevis S, Smolensky MH, Diller KR, Castriotta RJ. Accuracy of wristband Fitbit models in assessing sleep: systematic review and meta-analysis. J Med Internet Res 2019 Nov 28;21(11):e16273. [doi: 10.2196/16273] [Medline: 31778122]
13. Kosmadopoulos A, Sargent C, Darwent D, Zhou X, Roach GD. Alternatives to polysomnography (PSG): a validation of wrist actigraphy and a partial-PSG system. Behav Res Methods 2014 Dec;46(4):1032-1041. [doi: 10.3758/s13428-013-0438-7] [Medline: 24442593]
14. Nam Y, Kim Y, Lee J. Sleep monitoring based on a tri-axial accelerometer and a pressure sensor. Sensors (Basel) 2016 May 23;16(5):750. [doi: 10.3390/s16050750] [Medline: 27223290]
15. Chase JD, Busa MA, Staudenmayer JW, Sirard JR. Sleep measurement using wrist-worn accelerometer data compared with polysomnography. Sensors (Basel) 2022 Jul 4;22(13):5041. [doi: 10.3390/s22135041] [Medline: 35808535]
16. van de Water ATM, Holmes A, Hurley DA. Objective measurements of sleep for non-laboratory settings as alternatives to polysomnography--a systematic review. J Sleep Res 2011 Mar;20(1 Pt 2):183-200. [doi: 10.1111/j.1365-2869.2009.00814.x] [Medline: 20374444]
17. Hamill K, Jumabhoy R, Kahawage P, de Zambotti M, Walters EM, Drummond SPA. Validity, potential clinical utility and comparison of a consumer activity tracker and a research-grade activity tracker in insomnia disorder II: outside the laboratory. J Sleep Res 2020 Feb;29(1):e12944. [doi: 10.1111/jsr.12944] [Medline: 31680327]
18. Dong X, Yang S, Guo Y, Lv P, Wang M, Li Y. Validation of Fitbit Charge 4 for assessing sleep in Chinese patients with chronic insomnia: a comparison against polysomnography and actigraphy. PLoS One 2022 Oct 18;17(10):e0275287. [doi: 10.1371/journal.pone.0275287] [Medline: 36256631]
19. Chinoy ED, Cuellar JA, Huwa KE, et al. Performance of seven consumer sleep-tracking devices compared with polysomnography. Sleep 2021 May 14;44(5):zsaa291. [doi: 10.1093/sleep/zsaa291] [Medline: 33378539]
20. Soon S, Svavarsdottir H, Downey C, Jayne DG. Wearable devices for remote vital signs monitoring in the outpatient setting: an overview of the field. BMJ Innov 2020 Apr;6(2):55-71. [doi: 10.1136/bmjinnov-2019-000354]
21. Walch O, Huang Y, Forger D, Goldstein C. Sleep stage prediction with raw acceleration and photoplethysmography heart rate data derived from a consumer wearable device. Sleep 2019 Dec 24;42(12):zsz180. [doi: 10.1093/sleep/zsz180] [Medline: 31579900]
22. Miller DJ, Sargent C, Roach GD. A validation of six wearable devices for estimating sleep, heart rate and heart rate variability in healthy adults. Sensors 2022 Aug 22;22(16):6317. [doi: 10.3390/s22166317]
23. Castaneda D, Esparza A, Ghamari M, Soltanpur C, Nazeran H. A review on wearable photoplethysmography sensors and their potential future applications in health care. Int J Biosens Bioelectron 2018 Aug 6;4(4):195-202. [doi: 10.15406/ijbsbe.2018.04.00125] [Medline: 30906922]
24. Conley S, Knies A, Batten J, et al. Agreement between actigraphic and polysomnographic measures of sleep in adults with and without chronic conditions: a systematic review and meta-analysis. Sleep Med Rev 2019 Aug;46:151-160. [doi: 10.1016/j.smrv.2019.05.001] [Medline: 31154154]
25. Danzig R, Wang M, Shah A, Trotti LM. The wrist is not the brain: estimation of sleep by clinical and consumer wearable actigraphy devices is impacted by multiple patient- and device-specific factors. J Sleep Res 2020 Feb;29(1):e12926. [doi: 10.1111/jsr.12926] [Medline: 31621129]
26. Nguyen NH, Vallance JK, Buman MP, et al. Effects of a wearable technology-based physical activity intervention on sleep quality in breast cancer survivors: the ACTIVATE trial. J Cancer Surviv 2021 Apr;15(2):273-280. [doi: 10.1007/s11764-020-00930-7] [Medline: 32875536]
27. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71. [doi: 10.1136/bmj.n71] [Medline: 33782057]
28. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb) 2012;22(3):276-282. [Medline: 23092060]

29. Mouritzen NJ, Larsen LH, Lauritzen MH, Kjær TW. Assessing the performance of a commercial multisensory sleep tracker. PLoS One 2020 Dec 11;15(12):e0243214. [doi: 10.1371/journal.pone.0243214] [Medline: 33306678]
30. Stone JD, Rentz LE, Forsey J, et al. Evaluations of commercial sleep technologies for objective monitoring during routine sleeping conditions. Nat Sci Sleep 2020 Oct 27;12:821-842. [doi: 10.2147/NSS.S270705] [Medline: 33149712]
31. Miller DJ, Lastella M, Scanlan AT, et al. A validation study of the WHOOP strap against polysomnography to assess sleep. J Sports Sci 2020 Nov;38(22):2631-2636. [doi: 10.1080/02640414.2020.1797448] [Medline: 32713257]
32. Miller DJ, Roach GD, Lastella M, et al. A validation study of a commercial wearable device to automatically detect and estimate sleep. Biosensors (Basel) 2021 Jun 8;11(6):185. [doi: 10.3390/bios11060185] [Medline: 34201016]
33. Doheny EP, Renerts K, Baumann C, et al. F35 sleep monitoring in Huntington's disease using Fitbit compared to polysomnography. Journal of Neurology, Neurosurgery & Psychiatry 2021 Sep;92:A33-A34. [doi: 10.1136/jnnp-2021-EHDN.78]
34. Renerts K, Doheny EP, Baumann C, et al. F63 validation of Fitbit Charge 4 for sleep monitoring in participants with Huntington's disease. Journal of Neurology, Neurosurgery & Psychiatry 2022 Sep;93:A59. [doi: 10.1136/jnnp-2022-ehdn.154]
35. Depner CM, Cheng PC, Devine JK, et al. Wearable technologies for developing sleep and circadian biomarkers: a summary of workshop discussions. Sleep 2020 Feb 13;43(2):zsz254. [doi: 10.1093/sleep/zsz254] [Medline: 31641776]
36. Sargent C, Lastella M, Halson SL, Roach GD. The validity of activity monitors for measuring sleep in elite athletes. J Sci Med Sport 2016 Oct;19(10):848-853. [doi: 10.1016/j.jsams.2015.12.007] [Medline: 26794719]
37. Sargent C, Lastella M, Romyn G, Versey N, Miller DJ, Roach GD. How well does a commercially available wearable device measure sleep in young athletes? Chronobiol Int 2018 Jun;35(6):754-758. [doi: 10.1080/07420528.2018.1466800] [Medline: 29750580]
38. Collop NA. Scoring variability between polysomnography technologists in different sleep laboratories. Sleep Med 2002 Jan;3(1):43-47. [doi: 10.1016/s1389-9457(01)00115-0] [Medline: 14592253]
39. Danker-Hopfe H, Anderer P, Zeitlhofer J, et al. Interrater reliability for sleep scoring according to the Rechtschaffen & Kales and the new AASM standard. J Sleep Res 2009 Mar;18(1):74-84. [doi: 10.1111/j.1365-2869.2008.00700.x] [Medline: 19250176]
40. Zhang N, Cao B, Zhu Y. Indoor environment and sleep quality: a research based on online survey and field study. Build Environ 2018 Jun;137:198-207. [doi: 10.1016/j.buildenv.2018.04.007]
41. Bruyneel M, Sanida C, Art G, et al. Sleep efficiency during sleep studies: results of a prospective study comparing home-based and in-hospital polysomnography. J Sleep Res 2011 Mar;20(1 Pt 2):201-206. [doi: 10.1111/j.1365-2869.2010.00859.x] [Medline: 20561176]
42. Mayeli A, Janssen SA, Sharma K, Ferrarelli F. Examining first night effect on sleep parameters with HD-EEG in healthy individuals. Brain Sci 2022 Feb 8;12(2):233. [doi: 10.3390/brainsci12020233] [Medline: 35203996]
43. Byun JH, Kim KT, Moon HJ, Motamedi GK, Cho YW. The first night effect during polysomnography, and patients' estimates of sleep quality. Psychiatry Res 2019 Apr;274:27-29. [doi: 10.1016/j.psychres.2019.02.011] [Medline: 30776709]
44. Hof Zum Berge A, Kellmann M, Jakowski S. The first-night effect in elite sports: an initial glance on polysomnography in home-based settings. Front Psychol 2021 Mar 25;12:641451. [doi: 10.3389/fpsyg.2021.641451] [Medline: 33841275]
45. Tamaki M, Nittono H, Hayashi M, Hori T. Examination of the first-night effect during the sleep-onset period. Sleep 2005 Feb;28(2):195-202. [doi: 10.1093/sleep/28.2.195] [Medline: 16171243]
46. Driller MW, Dunican IC. No familiarization or 'first-night effect' evident when monitoring sleep using wrist actigraphy. J Sleep Res 2021 Aug;30(4):e13246. [doi: 10.1111/jsr.13246] [Medline: 33289237]
47. Saletu B, Anderer P, Saletu-Zyhlarz GM. Recent advances in sleep research. Psychiatr Danub 2013 Dec;25(4):426-434. [Medline: 24247058]

Abbreviations

DS: deep sleep
EEG: electroencephalogram
LS: light sleep
PPG: photoplethysmography
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PSG: polysomnography
REM: rapid eye movement
TST: total sleep time

Edited by L Buis; submitted 25.08.23; peer-reviewed by A Baharav, K Martiny; revised version received 01.02.24; accepted 01.02.24; published 27.03.24.

Please cite as:
Schyvens AM, Van Oost NC, Aerts JM, Masci F, Peters B, Neven A, Dirix H, Wets G, Ross V, Verbraecken J
Accuracy of Fitbit Charge 4, Garmin Vivosmart 4, and WHOOP Versus Polysomnography: Systematic Review
JMIR Mhealth Uhealth 2024;12:e52192
URL: https://mhealth.jmir.org/2024/1/e52192
doi:10.2196/52192

Effects of a Virtual Reality Cycling Platform on Lower Limb Rehabilitation in Patients With Ataxia and Hemiparesis: Pilot Randomized Controlled Trial

Ana Rojo[1,2*], PhD; Arantxa Castrillo Calvillo[3*], BPT; Cristina López[3*], BPT; Rafael Raya[1*], PhD; Juan C Moreno[2*], PhD

[1]Departamento de Tecnologías de la Información, Escuela Politécnica Superior, Universidad San Pablo-CEU, CEU Universities, Madrid, Spain

[2]Neural Rehabilitation Group, Spanish National Research Council, Cajal Institute, Madrid, Spain

[3]Centro Lescer, Madrid, Spain

[*]all authors contributed equally

Corresponding Author:
Ana Rojo, PhD
Departamento de Tecnologías de la Información, Escuela Politécnica Superior, Universidad San Pablo-CEU, CEU Universities, , Madrid, , Spain

Abstract

Background: New interventions based on motor learning principles and neural plasticity have been tested among patients with ataxia and hemiparesis. Therapies of pedaling exercises have also shown their potential to induce improvements in muscle activity, strength, and balance. Virtual reality (VR) has been demonstrated as an effective tool for improving the adherence to physical therapy, but it is still undetermined if it promotes greater improvements than conventional therapy.

Objective: Our objective was to compare the effect on lower limb range of motion (ROM) when using VR technology for cycling exercise versus not using VR technology.

Methods: A randomized controlled trial with 20 patients with ataxia and hemiparesis was carried out. The participants were divided into 2 groups: the experimental group (n=10, 50%) performed pedaling exercises using the VR system and the control group (n=10, 50%) performed pedaling exercises without using VR. Measurements of the active and passive ROM of the hip and knee joint were taken before and after a cycling intervention, which consisted of 3 sessions of the same duration but with progressively increasing speeds (4, 5, and 6 km/h). Repeated measures ANOVAs were conducted to compare the preintervention (T_i) and postintervention (T_e) assessments within each group. Additionally, the improvement effect of using the VR system was analyzed by comparing the variation coefficient ($\Delta = 1 - [T_e / T_i]$) between the preintervention and postintervention assessments for each group. Group comparisons were made using independent 1-tailed *t* tests.

Results: Significant improvements were shown in active left hip flexion (*P*=.03) over time, but there was no group-time interaction effect (*P*=.67). Passive left hip flexion (*P*=.93) did not show significant improvements, and similar results were observed for active and passive right hip flexion (*P*=.39 and *P*=.83, respectively). Neither assessments of knee flexion (active left: *P*=.06; passive left: *P*=.76; active right: *P*=.34; passive right: *P*=.06) nor knee extension showed significant changes (active left: *P*=.66; passive left: *P*=.92; active right: *P*=.12; passive right: *P*=.38). However, passive right knee extension (*P*=.04) showed a significant improvement over time. Overall, although active and passive ROM of the knee and hip joints showed a general improvement, no statistically significant differences were found between the groups.

Conclusions: In this study, participants who underwent the cycling intervention using the VR system showed similar improvement in lower limb ROM to the participants who underwent conventional training. Ultimately, the VR system can be used to engage participants in physical activity.

Trial Registration: ClinicalTrials.gov NCT05162040; https://www.clinicaltrials.gov/study/NCT05162040

(JMIR Serious Games 2024;12:e39286) doi:10.2196/39286

KEYWORDS

ataxia; cycling; hemiparesis; lower limb; neuropathology; rehabilitation; virtual reality; limb; intervention; neural; neural plasticity; therapy; muscle; strength; balance; tool; exercise; physical activity; neuroplasticity

Introduction

Background

Ataxia is an umbrella term for describing deficits in limb movement coordination such as dysmetria, dyspraxia, and dyssynergia [1]. The persistence of these deficits affects an individual's functional ability and poses a health challenge for both patients and clinicians.

Current scientific evidence indicates that the most effective treatment for ataxia should combine balance and coordination retraining and constraint-induced functional movement therapy [2]. However, the scientific literature still lacks a consensus on the details of these interventions and the timing of their implementation to enhance the recovery of the functionality of motor deficits in an individual [3].

On the other hand, in the field of neurophysiology, it is well known that to induce changes in neuroplasticity to achieve the functional recovery of motor deficits, the application of therapies based on the repetition of movements is required [4]. Some studies point out that the principles of motor learning are directly related to the regeneration of structures and the reorganization of neuronal function [5,6]. Moreover, the amount of practice is a key factor in motor learning, as well as the feedback provided during practice [7]. In fact, physical therapists must consider both the error feedback and activity guidance as 2 fundamental components of patient interaction during therapy to promote neuromotor learning [8]. Thus, interventions that promote normal function rather than the compensation of deficits are more recommended and should be applied to generate a physical activity plan based on the principles of motor learning and neural plasticity for patients with ataxic hemiparesis.

Prior Work

The scientific literature in the field of neurorehabilitation shows that pedaling exercises have the potential to induce improvements in muscle activity, strength, and balance [9]. This is mainly due to the fact that pedaling exercises based on the use of a cycloergometer provide a high number of flexion and extension repetitions [10] in the lower extremities for considerable periods of time. Because pedaling and walking are cyclical locomotor tasks that require the lower limb to alternate between flexion and extension [11,12], both share similar locomotor patterns of alternating muscle activation of antagonists [10,13]. Thus, cycling exercises are found to be useful for strengthening the lower limb muscles while acting as a pseudowalking task-oriented exercise. Some studies eluded that those biomechanical functions may be altered by the muscle groups involved in the pedaling tasks [14-16]. In fact, it was found that the degradation of pedaling performance in adults with hemiparesis was related to abnormalities in the execution of specific biomechanical functions [15]. Subsequently, it has been proven that human walking and cycling shared similar muscle synergies [16]. This evidence is the basis for rehabilitation treatments based on pedaling movements with potential positive outcomes for walking [16].

The ergometer is an equipment designed to perform cardiovascular work based on the alternative circular movement of the lower limb. Its use is advantageous for a muscle coordination study because balance is not an applicable factor in this kinematically constrained task [13]. In fact, applying an ergometer-based cycling routine could be useful because it requires no balance. Moreover, the exercise intensity of the ergometer-based cycling can be adapted to the user by adjusting the resistance of the pedal or the target speed. The ability to personalize the intensity of the exercise is a relevant factor for the patient's rehabilitation process. For these reasons, regular ergometer-based cycling is found to be a safer unsupervised exercise that is recommended for lower limb rehabilitation. Nevertheless, cycling exercise is also a static and repetitive form of exercise that leads to boredom and listlessness in patients. To deal with this discouragement factor, emerging technologies have been applied to elicit intrinsic motivation for rehabilitation patients [17]. Several studies pointed out the usefulness of gaming elements and virtual environments as assistive technology [18,19] and their potential effectiveness in physical therapies as opposed to conventional therapies [20].

Quite a few studies have focused on the analysis of functional metrics in virtual pedaling. A recent study evaluated the functionality of a virtual reality (VR) cycling training program that was applied to 10 patients with stroke [21]. It assessed the improvement of the bilateral asymmetry between the experimental group and the control group after the VR cycling intervention program. To evaluate this index, they equipped the ergometer pedals with force plates to determine the effect of the VR cycling training on each limb. The improvement of bilateral strength and standing balance was significantly different between VR cycling training and traditional physical training. Similarly, a previous study compared the effects of a cycling training program with extrinsic biofeedback and a nonimmersive interface versus traditional physical training on lower limb functional recovery in patients with stroke [22]. The results showed that improvements in walking endurance, walking speed, and muscle spasticity of the group using VR were significantly better than the group who underwent traditional physical training.

Objectives

The main objective of this study was to evaluate 2 different interventions: pedaling with VR and pedaling without VR. This study focused on comparing the improvements in lower limb range of motion (ROM) in pedaling activity between the group using VR and the group not using VR. To this end, a randomized controlled trial was carried out with patients with ataxia and hemiparesis. Hip and knee ROMs were measured before and after the cycling intervention. The overall aim of these analyses was to determine the effects of the 2 different interventions on short-term improvement of lower limb function and ROM.

Methods

VR System

The VR system implements extrinsic feedback strategies, gamification by levels, and personalization of the sessions with the aim of achieving greater adherence to pedaling exercise sessions. Its immersive nature means an increase in the sense of "presence," promoting the active involvement of the user.

The VR system is based on the transmission of the cycling kinematic data captured by the inertial sensors to the Oculus Quest 2 (Meta) head-mounted display (HMD) via Bluetooth. Therefore, the virtual application estimates the pedaling cycles, cadence, and distance during the exercise activity. The VR scenarios generated for this therapy consist of mapping the cycling cadence to the vehicle speed. Thus, the patient is placed inside a vehicle and visualizes the session data on the control panel while moving at the speed of the pedaling motion.

The design of the VR experience has been technically validated computationally to ensure low latency in motion analysis and visual representation of motion [23], thus preserving the embodiment effect and the sense of presence. Subsequently, the platform has also been validated from the point of view of satisfaction and ease of use of the system [24]. Additionally, considering that it is a stationary experience with an HMD that simulates a displacement, we evaluated to which extent the VR experience generates the type of motion sickness that causes fatigue, nausea, disorientation, postural instability, or visual fatigue [25]. Indeed, we verified that the platform does not generate adverse effects due to cybersickness [24].

Recruitment

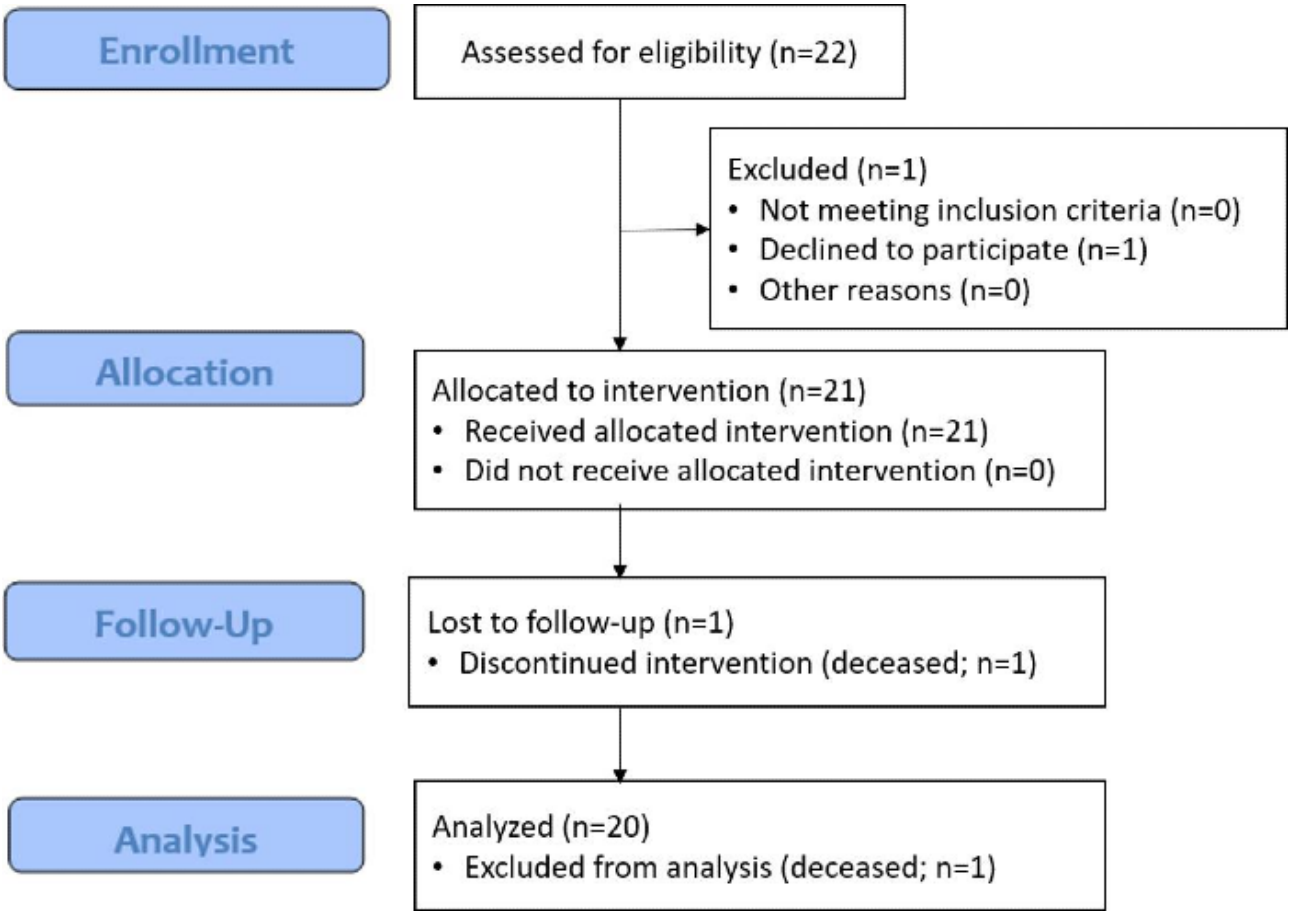
The participants were patients of both sexes between 18 and 90 years of age, recruited at the Lescer Clinic applying the inclusion and exclusion criteria. Inclusion criteria were as follows: individuals were eligible if they (1) had been prescribed pedaling exercise as treatment for lower limb rehabilitation and (2) were able to perform a pedaling session with VR technology. Exclusion criteria were as follows: (1) an insufficient cognitive state, (2) an unbound bone fracture, (3) severe disorders of vision or audition (inability to perceive visual or auditory information coming from VR), and (4) any incompatibility with the use of a VR system according to the clinical record. A sample of 22 participants (n=13, 59% male and n=7, 32% female; mean age 59.90, SD 13.56 y) volunteered to participate in this pilot randomized controlled trial (Table 1). Of this 22-person cohort, 1 participant dropped out of the study and 1 participant did not complete the study (Figure 1). The cohort was randomly divided into the experimental group (EG; 9/10, 90% male and 1/10, 10% female; mean age 60.80, SD 12.26 y) with VR cycling exercises or the control group (CG; 4/10, 40% male and 6/10, 60% female; mean age 59.00, SD 14.69 y) with traditional cycling exercises.

Table . Clinical and epidemiological features of the experimental group (EG) and control group (CG) participants.

Group and participant number	Sex	Age (y)	Etiology	Condition
EG				
1	Male	57	Ischemic stroke	Hemiparesis
2	Male	71	Hemorrhagic stroke	Ataxia
3	Male	53	Hemorrhagic stroke	Ataxia
4	Male	72	MCA[a] stroke	Hemiparesis
5	Male	53	MCA stroke	Hemiparesis
6	Male	62	Ischemic stroke	Hemiparesis
7	Male	59	Hemorrhagic stroke	Ataxia
8	Male	56	Progressive multifocal leukoencephalopathy	Ataxia
9	Female	86	Hemorrhagic stroke	Hemiparesis
10	Male	39	Ischemic stroke	Hemiparesis
CG				
1	Male	45	MCA stroke	Hemiparesis
2	Female	64	Hemorrhagic stroke	Ataxia
3	Male	58	Guillain-Barré syndrome	Hemiparesis
4	Female	41	Hemorrhagic stroke	Ataxia
5	Female	49	Ischemic stroke	Ataxia
6	Male	83	Ischemic stroke	Ataxia
7	Female	80	Hemorrhagic stroke	Hemiparesis
8	Female	72	Traumatic brain injury	Hemiparesis
9	Male	57	Ischemic stroke	Hemiparesis
10	Female	41	Guillain-Barré syndrome	Ataxia

[a]MCA: middle cerebral artery.

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) diagram.



Ethical Considerations

Ethical approval was obtained from the Research Ethics Committee of the San Pablo CEU University (550/21/51). This study has been registered at ClinicalTrials.gov (NCT05162040). All the participants were given written information in accordance with the Research Ethics Committee. The informed consent and the ability for participants to opt out was provided. Additionally, participants were informed that the data collected in this study can only be used for this study, not for secondary studies. The approval of the Research Ethics Committee of San Pablo CEU University only covers this study and does not cover a secondary analysis without additional consent. However, no additional analysis had been carried out.

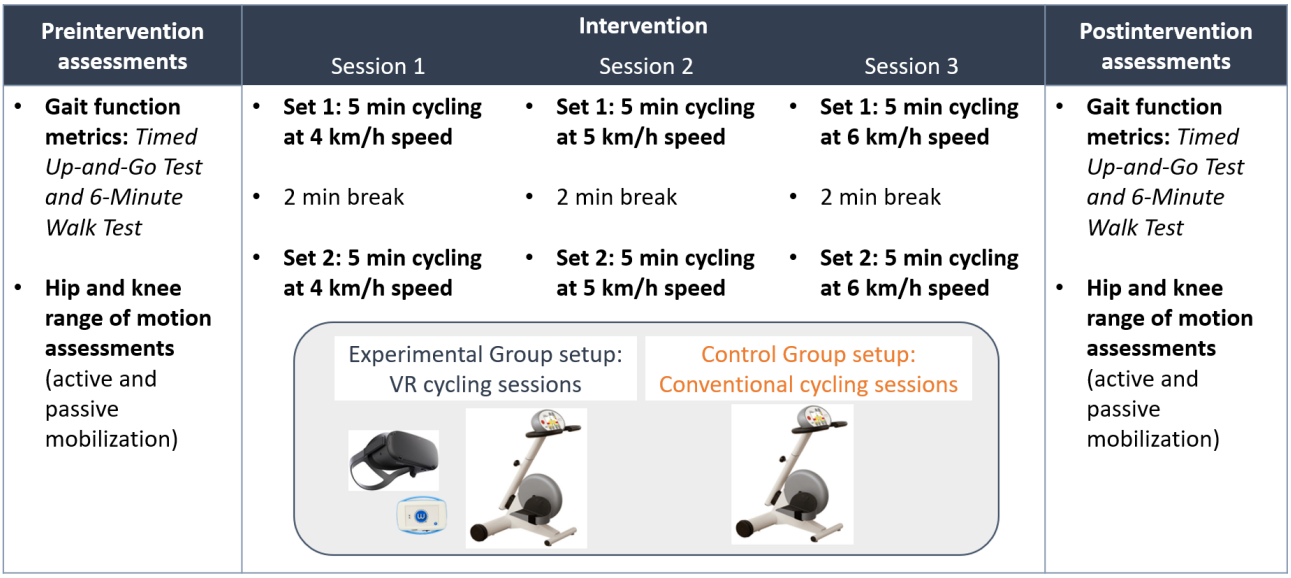
To ensure privacy and confidentiality, data are collected by employees of the agencies participating in the study. Each participant is assigned a unique code along with personal sociodemographic data and informed consent. These files remain in the custody of the principal investigator in charge of the project, while the assigned number is the one that identifies the anonymized data that was later analyzed. Finally, the participation in this study is completely voluntary; no compensation of any nature is offered to the human participants.

Intervention

This study was designed as a randomized controlled trial with 20 participants divided into 2 groups, following a block randomization method. The participants of the EG (n=10) performed pedaling exercises while using the VR system, whereas the participants of the CG (n=10) performed pedaling exercises without using the VR system. Before and after completing the exercise program, measurements of gait function metrics and joint ranges were performed to assess the effect of using VR stimulus during the cycling exercises.

The participants completed the cycling intervention simultaneously with their rehabilitation sessions. Afterward, for each participant, 3 cycling sessions were scheduled over 1 week with a maximum of 48 hours between sessions. Each session consisted of 2 sets of a 5-minute pedaling exercise spaced with a 2-minute break (to rest). Similar studies [19,26] have tested robotic unicycles in pedaling sessions at a cadence of 60 revolutions per minute. In our case, the pedaling speed of 1 cycle per second is equivalent to a target speed of 6 km/h. For this reason, it was decided to set this speed as the maximum speed and to start the first session with a slightly more comfortable speed (4 km/h) and increase it progressively (Figure 2). The participants of both groups performed the exercise following a set pedaling speed so that they received visual feedback according to the set target speed of 4-6 km/h for each session. The EG participants received visual feedback through the immersive VR application, whereas the CG participants received visual feedback on the ergometer display. All participants were instructed to maintain a constant pedaling speed throughout the session at the target cadence.

Figure 2. Summary of the intervention program for experimental and control group participants. VR: virtual reality.



Physical Assessment

For the assessment of active and passive ROM of the hip and knee joint, a specifical ROM assessment tool was used. Measurements were extracted from biomechanical analysis using an inertial motion capture system (Werium; Werium Solutions) consisting of 2 inertial sensors: 1 placed in the distal part of the extremity (moving sensor) and the other in the proximal part (fixed sensor). Both sensors send their measurements via Bluetooth to a PC that runs the data acquisition software, Pro Motion Capture (Werium Solutions). This software computes the relative angle from both angle measurements (avoiding compensations) with an accuracy of 1 degree.

Protocol

The cycling sessions for both groups consisted of the use of a leg ergometer that allows training of the lower limb. Additionally, the EG used an inertial sensor placed on the right thigh and the Oculus Quest 2 HMD (Figure 3).

Figure 3. Cycling session of a participant in the experimental group using the virtual cycling platform.

The EG underwent the following procedure each session:

- The clinician connected the inertial sensor to the Oculus Quest 2 HMD.
- The patient was seated in a nonmovable chair (with no armrests) in front of the pedaling station during the entire session. The inertial sensor was placed on the right thigh of the patient by adjusting an elastic band, and the sensor was turned on.
- The clinician fitted the Oculus Quest 2 HMD comfortably on the patient and guided him or her through the selection of the game scene. Once the game environment was entered, the clinician indicated the number of minutes of exercise and the target speed of the session so that the patient could configure these parameters on the interactive settings panel.
- Finally, the user performed 2 sets of a 5-minute cycling exercise with a 2-minute break between the sets.

Similarly, the CG underwent the following procedure each session:

- The patient was seated in a nonmovable chair (with no armrests) in front of the pedaling station during the entire session.
- The clinician turned on the ergometer's display and entered the number of minutes of exercise and the target speed of the session.
- Finally, the user performed 2 sets of a 5-minute cycling exercise with a 2-minute break between the sets.

Statistical Analysis

The data analysis model is the repeated measures model between 2 groups and the analysis of the longitudinal effect in increments of the measurements. Multifactor ANOVA analysis (with $P<.05$) were computed with SPSS Statistics (version 27.0; IBM Corp). The sample size was calculated using the software tool G*Power (version 3.1.9.7; Heinrich Heine Universität Düsseldorf). Ideally, assuming an effect size of 0.7, a minimum sample of 20 participants was required for the study to provide consistent statistical results. Since the effect size shows the strength of the relationships, it represents a minimum clinically meaningful difference. Of the many different types of effect sizes, the

G*Power software uses Cohen *d* to characterize effect size by relating the mean difference to variability. Therefore, his study standardized the effect size to 0.7 for sample size calculation and power analysis.

Results

To identify the underlying differences between the preintervention (T_i) and postintervention (T_e) assessments in each group, repeated measures ANOVAs were conducted with time ($T_i - T_e$) as the dependent variable and group as the main within-subjects factor. When the ANOVA was significant, the Bonferroni post hoc test was used. To ensure that the error variance of the dependent variables is equal across groups, the Levene test was applied beforehand for all the metrics.

In addition, to identify the improvement effect due to the use or nonuse of the VR system, the variation coefficient between the preintervention and postintervention assessments was analyzed for each group as follows: $\Delta = 1 - (T_e / T_i)$. The variation coefficient outcomes were compared between groups by the independent 1-tailed *t* test. The mean and SD of the ROM outcomes for the hip and knee of each group are shown in Table 2. The mean increase Δ for each measurement is shown in Figures 4 and 5.

Table . Hip and knee range-of-motion outcomes.

Outcome	Experimental group, mean (SD)			Control group, mean (SD)		
	Preintervention (°)	Postintervention (°)	Variation coefficient (%)	Preintervention (°)	Postintervention (°)	Variation coefficient (%)
ALHF[a]	81.25 (36.09)	94.23 (32.26)	26.30 (33.52)	92.84 (21.40)	94.37 (25.83)	1.21 (14.20)
PLHF[b]	106.07 (21.16)	107.94 (17.63)	2.61 (5.81)	112.92 (17.76)	110.70 (16.83)	–1.43 (9.40)
ARHF[c]	97.55 (20.94)	97.13 (21.26)	0.28 (10.94)	97.11 (28.05)	101.79 (27.35)	5.60 (10.50)
PRHF[d]	106.63 (17.06)	109.82 (14.99)	3.69 (8.72)	119.74 (14.73)	117.71 (13.42)	–1.13 (8.72)
ALKF[e]	46.07 (14.62)	45.97 (11.47)	4.27 (26.31)	37.47 (12.03)	35.65 (8.47)	1.63 (30.86)
PLKF[f]	58.82 (9.84)	55.96 (9.79)	–3.48 (17.40)	57.14 (13.92)	54.58 (12.15)	–1.66 (19.96)
ARKF[g]	39.13 (16.54)	37.81 (10.68)	8.98 (35.88)	43.03 (10.00)	44.58 (13.32)	5.36 (30.29)
PRKF[h]	50.57 (10.02)	49.81 (10.31)	–0.65 (15.15)	63.35 (12.28)	57.28 (13.95)	–9.35 (15.17)
ALKE[i]	61.72 (14.86)	62.92 (13.11)	3.28 (9.74)	55.57 (17.13)	63.41 (11.77)	26.70 (51.68)
PLKE[j]	66.46 (11.74)	69.95 (15.09)	4.94 (15.92)	64.75 (11.94)	72.30 (12.46)	14.91 (24.40)
ARKE[k]	64.00 (10.11)	68.02 (10.14)	8.33 (20.67)	57.49 (14.91)	57.19 (14.76)	2.22 (23.52)
PRKE[l]	66.67 (11.53)	67.18 (10.93)	1.58 (14.10)	57.65 (11.21)	68.78 (6.67)	25.29 (34.70)

[a]ALHF: active left hip flexion.
[b]PLHF: passive left hip flexion.
[c]ARHF: active right hip flexion.
[d]PRHF: passive right hip flexion.
[e]ALKF: active left knee flexion.
[f]PLKF: passive left knee flexion.
[g]ARKF: active right knee flexion.
[h]PRKF: passive right knee flexion.
[i]ALKE: active left knee extension.
[j]PLKE: passive left knee extension.
[k]ARKE: active right knee extension.
[l]PRKE: passive right knee extension.

Figure 4. Summary of increments in active and passive hip ROM parameters with SD bars. The vertical axis represents the percentage of postintervention increase or decrease of each hip ROM parameter. ALHF: active left hip flexion; ARHF: active right hip flexion; PLHF: passive left hip flexion; PRHF: passive right hip flexion; ROM: range of motion.

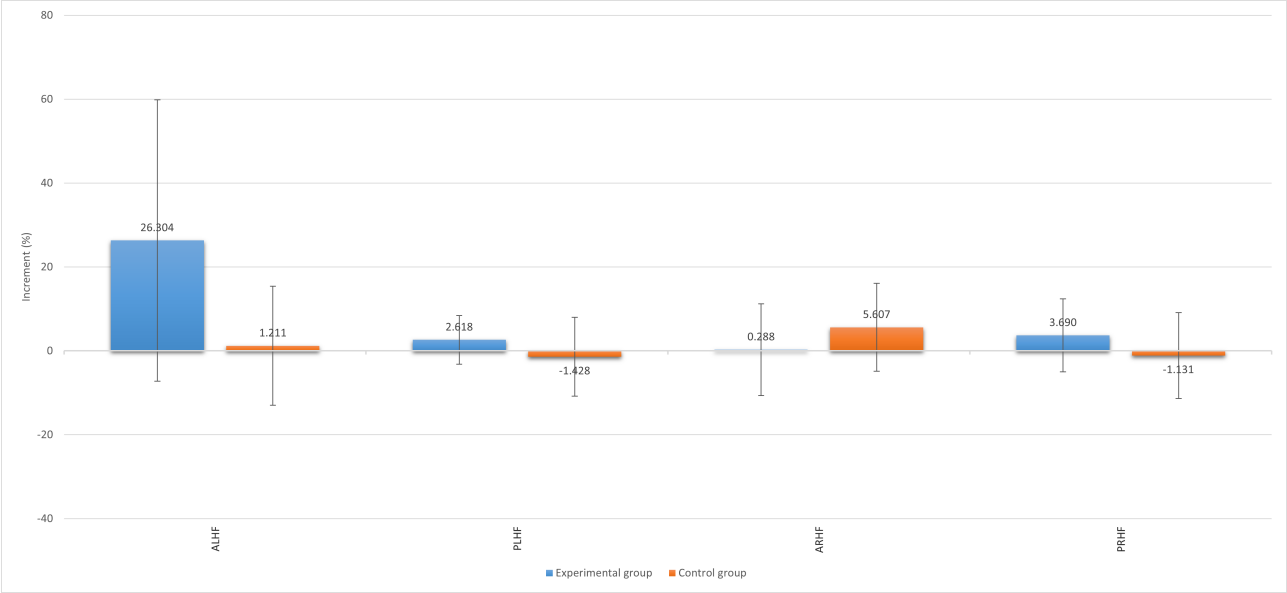
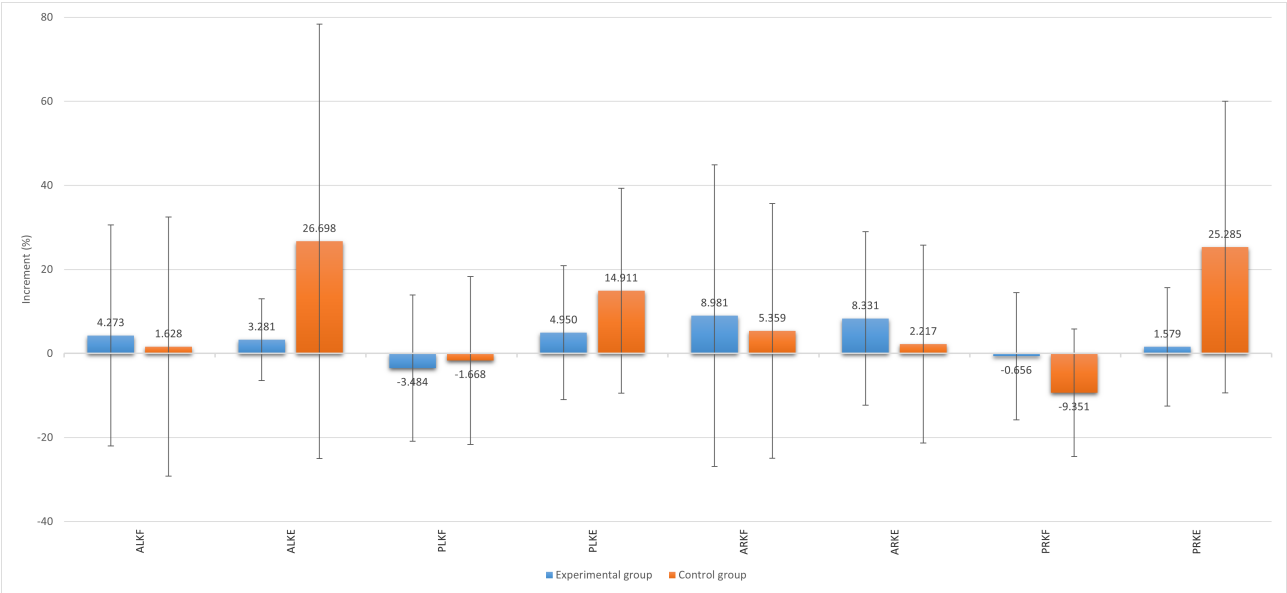


Figure 5. Summary of increments in active and passive knee ROM parameters with SD bars. The vertical axis represents the percentage of postintervention increase or decrease of each knee ROM parameter. ALKE: active left knee extension; ALKF: active left knee flexion; ARKE: active right knee extension; ARKF: active right knee flexion; PLKE: passive left knee extension; PLKF: passive left knee flexion; PRKE: passive right knee extension; PRKF: passive right knee flexion; ROM: range of motion.



With regard to the hip flexion outcomes, the active left hip flexion results were significant by ANOVA (P=.03), with no significance observed for the between-subjects effects test (P=.67). However, the within-subjects effects test was significant for the time factor (P=.03), but no significant group-time interaction effect was found (P=.08). Despite the opposing results showing passive left hip flexion improvements for each group, there was no significance difference by ANOVA (P=.93) and no statistically significant result was obtained by the between-subjects effects test. Passive left hip flexion was statistically significant in the within-subjects effects test for the time factor (P=.008). The active and passive right hip flexion results were not significant by ANOVA (P=.39 and P=.83, respectively). In both cases, no significant results were obtained for the between- and within-subjects effects tests.

For the knee ROM measurements, when analyzing the left knee assessments, the active and passive left knee flexion outcomes were not significant by ANOVA (P=.06 and P=.76, respectively). No statistically significant results were obtained by the between- and within-subjects effects tests in both cases. Similar results were obtained for the active left knee extension outcomes. Although reasonable differences in the active and passive left knee extension increases between groups can be observed in Figure 5, neither active nor passive left knee extension were significant by ANOVA (P=.66 and P=.92,

respectively). No statistically significant results were obtained by the between- and within-subjects effects tests in both cases.

Regarding the right knee assessments outcomes, all outcomes were not significant by ANOVA (active flexion: *P*=.34; passive flexion: *P*=.06; active extension: *P*=.12; passive extension: *P*=.38). No statistically significant results were obtained by the between- and within-subjects effects tests for all cases, except for passive right knee extension, which was statistically significant for the time factor (*P*=.04) by the within-subjects effects test.

Discussion

Principal Findings

The aim of this study was to test the short-term effects of 2 different interventions on short-term improvement of lower limb function and ROM. For this purpose, a randomized controlled trial was carried out with participants with ataxia and hemiparesis.

In this study, the improvement outcomes of active and passive knee and hip joint ROMs due to the use of VR technology were inconclusive. Likewise, no statistically significant differences in the results between groups can be indicated. Even so, all the active ROMs measured—that is, performed by the patients—showed an increase with respect to the initial values. A greater disparity was observed in the passive measurements, although this may be attributed to the different passive mobilizations performed at each time by different physiotherapists. In this case, the active measurement is of special relevance in clinical terms because it indicates a ROM that the patient is able to achieve autonomously. On the other hand, large SDs in outcome variables clearly indicate that the improvements in the functional gait outcomes are not entirely consistent or represent a group effect. We observe that no significant effect can be attributed to VR intervention based on the statistical analysis of the immediate effects on gait function and joint ROM.

However, considering this similarity between groups, it can be pointed out that the use of VR has similar positive effects as the use of the conventional pedaling treatment. Thus, this immediate observation of effects leads us to conclude that the use of VR during pedaling exercise has similar effects to non-VR exercise training. Therefore, given that the use of VR technology does not worsen the improvement of lower limb ROM, and in line with the scientific literature [17-20], it may be advantageous to use it to maintain the patient's motivation.

Strengths and Limitations

A limitation of this study is the short-term nature of the intervention program. It is arguable that a longer intervention program would have shown more notable effects on functional improvement. However, assuming that it is precisely the treatment time that is one of the main causes of progress in physical improvement, the motivational impact of VR technology over time would need to be assessed. Therefore, further studies on the motivational impact of VR cycling versus conventional cycling on long-term physical activity remain to be addressed. Regarding these future studies, we suggest that cohort studies should be conducted among a population with more homogeneous neurological conditions. This recommendation is based on the limitations encountered in this study, where the difficulty of drawing conclusions about group changes or improvements with such wide SDs is presumably a reflection of the heterogeneity of the group.

Another factor to consider is that different physiotherapists were involved in taking the ROM measurements of the participants, although the measurement system was the same. This fact could be considered in future studies to evaluate interrater effects.

Future Directions

We consider it relevant to analyze, in future studies, whether these improvements in active and passive ROM are accompanied by greater muscle activation, in particular, the hamstrings, rectus femoris, gastrocnemius, and tibialis anterior muscles, as suggested by scientific literature [27].

Conclusions

The results of this trial demonstrate that pedaling exercises coordinated with VR technology works as successfully as conventional training for patients with lower limb disorders such as ataxia and hemiparesis. In this study, it was found that participants who performed the pedaling exercise program using the VR system showed similar results to the participants who performed the exercise activity without using VR technology. Overall, VR technologies can be a useful tool to help patients with ataxia and hemiparesis engage in lower limb exercise therapies.

Acknowledgments

The authors would like to thank all the participants who collaborated in this study, as well as the therapists and health care professionals from Centro Lescer for their participation in this study. The financial support for the industrial doctorate project "Desarrollo y estudio de una plataforma interactiva y un sistema electrónico de pedaleo para rehabilitación funcional de personas mayores" of the Autonomous Community of Madrid (IND2019/TIC17090) toward this research is hereby acknowledged. Grant PID2021-127096OB-I00 funded by MCIN/AEI/ 10.13039/501100011033 and by "ERDF A way of making Europe."

The funding sponsors have no role in the design of the study; the collection, analyses, or interpretation of data; the writing of the manuscript; and the decision to publish the result

Data Availability

The data sets generated or analyzed during this study are available on the GitHub repository [28].

Authors' Contributions

AR contributed to software, data curation, formal analysis, and writing–original draft. ACC contributed to data curation and methodology. CL contributed to methodology, resources, and supervision. RR contributed to funding acquisition, supervision, and writing–review and editing. JCM contributed to funding acquisition, supervision, and writing–review and editing.

Conflicts of Interest

RR is the chief executive officer of Werium Solutions, and AR is a software developer at Werium Solutions. The other authors declare no conflicts of interest.

Checklist 1
CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) checklist (V 1.6.1).
[PDF File, 1274 KB - games_v12i1e39286_app1.pdf]

References

1. Blumenfeld H. Neuroanatomy Through Clinical Cases: Oxford University Press; 2010.
2. Crum EO, Baltz MJ, Krause DA. The use of motor learning and neural plasticity in rehabilitation for ataxic hemiparesis: a case report. Physiother Theory Pract 2020 Nov;36(11):1256-1265. [doi: 10.1080/09593985.2019.1566941] [Medline: 30686101]
3. Januário F, Campos I, Amaral C. Rehabilitation of postural stability in ataxic/hemiplegic patients after stroke. Disabil Rehabil 2010;32(21):1775-1779. [doi: 10.3109/09638281003734433] [Medline: 20350209]
4. Lohse KR, Hilderman CGE, Cheung KL, Tatla S, van der Loos HFM. Virtual reality therapy for adults post-stroke: a systematic review and meta-analysis exploring virtual environments and commercial games in therapy. PLoS One 2014 Mar 28;9(3):e93318. [doi: 10.1371/journal.pone.0093318] [Medline: 24681826]
5. Katz-Leurer M, Sender I, Keren O, Dvir Z. The influence of early cycling training on balance in stroke patients at the subacute stage. Results of a preliminary trial. Clin Rehabil 2006 May;20(5):398-405. [doi: 10.1191/0269215505cr960oa] [Medline: 16774090]
6. Lucareli PRG, Greve JMD. Alteration of the load-response mechanism of the knee joint during hemiparetic gait following stroke analyzed by 3-dimensional kinematic. Clinics (Sao Paulo) 2006 Aug;61(4):295-300. [doi: 10.1590/s1807-59322006000400004] [Medline: 16924319]
7. Krakauer JW. Motor learning: its relevance to stroke recovery and neurorehabilitation. Curr Opin Neurol 2006 Feb;19(1):84-90. [doi: 10.1097/01.wco.0000200544.29915.cc] [Medline: 16415682]
8. Winstein C, Lewthwaite R, Blanton SR, Wolf LB, Wishart L. Infusing motor learning research into neurorehabilitation practice: a historical perspective with case exemplar from the accelerated skill acquisition program. J Neurol Phys Ther 2014 Jul;38(3):190-200. [doi: 10.1097/NPT.0000000000000046] [Medline: 24828523]
9. Peng CW, Chen SC, Lai CH, et al. Review: clinical benefits of functional electrical stimulation cycling exercise for subjects with central neurological impairments. J Med Biol Eng 2011;31(1):1-11 [FREE Full text] [doi: 10.5405/jmbe.718]
10. Ting LH, Kautz SA, Brown DA, Zajac FE. Phase reversal of biomechanical functions and muscle activity in backward pedaling. J Neurophysiol 1999 Feb;81(2):544-551. [doi: 10.1152/jn.1999.81.2.544] [Medline: 10036258]
11. Coast JR, Welch HG. Linear increase in optimal pedal rate with increased power output in cycle ergometry. Eur J Appl Physiol Occup Physiol 1985;53(4):339-342. [doi: 10.1007/BF00422850] [Medline: 4039261]
12. Winter DA. Biomechanical motor patterns in normal walking. J Mot Behav 1983 Dec;15(4):302-330. [doi: 10.1080/00222895.1983.10735302] [Medline: 15151864]
13. Raasch CC, Zajac FE. Locomotor strategy for pedaling: muscle groups and biomechanical functions. J Neurophysiol 1999 Aug;82(2):515-525. [doi: 10.1152/jn.1999.82.2.515] [Medline: 10444651]
14. Neptune RR, Kautz SA, Hull ML. The effect of pedaling rate on coordination in cycling. J Biomech 1997 Oct;30(10):1051-1058. [doi: 10.1016/s0021-9290(97)00071-7] [Medline: 9391872]
15. Brown DA, Kautz SA. Increased workload enhances force output during pedaling exercise in persons with poststroke hemiplegia. Stroke 1998 Mar;29(3):598-606. [doi: 10.1161/01.str.29.3.598] [Medline: 9506599]
16. Barroso FO, Torricelli D, Moreno JC, et al. Shared muscle synergies in human walking and cycling. J Neurophysiol 2014 Oct 15;112(8):1984-1998. [doi: 10.1152/jn.00220.2014] [Medline: 25057144]
17. Deci EL, Ryan RM. Intrinsic Motivation and Self-Determination in Human Behavior: Springer; 2013. [doi: 10.1007/978-1-4899-2271-7]

18. de Bruin ED, Schoene D, Pichierri G, Smith ST. Use of virtual reality technique for the training of motor control in the elderly. Z Gerontol Geriatr 2010 Aug;43(4):229-234. [doi: 10.1007/s00391-010-0124-7] [Medline: 20814798]
19. Bruun-Pedersen JR, Pedersen KS, Serafin S, Kofoed LB. Augmented exercise biking with virtual environments for elderly users: a preliminary study for retirement home physical therapy. Presented at: 2014 2nd Workshop on Virtual and Augmented Assistive Technology (VAAT); Mar 30, 2014; Minneapolis, MN p. 23-27. [doi: 10.1109/VAAT.2014.6799464]
20. Laver K, Ratcliffe J, George S, Burgess L, Crotty M. Is the Nintendo Wii Fit really acceptable to older people? A discrete choice experiment. BMC Geriatr 2011 Oct 20;11:64. [doi: 10.1186/1471-2318-11-64] [Medline: 22011360]
21. Yin C, Hsueh YH, Yeh CY, Lo HC, Lan YT. A virtual reality-cycling training system for lower limb balance improvement. Biomed Res Int 2016;2016:9276508. [doi: 10.1155/2016/9276508] [Medline: 27034953]
22. Yang HC, Lee CL, Lin R, et al. Effect of biofeedback cycling training on functional recovery and walking ability of lower extremity in patients with stroke. Kaohsiung J Med Sci 2014 Jan;30(1):35-42. [doi: 10.1016/j.kjms.2013.07.006] [Medline: 24388057]
23. Rojo A, Raya R, Moreno JC. Virtual reality application for real-time pedalling cadence estimation based on hip ROM tracking with inertial sensors: a pilot study. Virtual Real 2023 Mar;27:3-17. [doi: 10.1007/s10055-022-00668-w]
24. Rojo A, Castrillo A, López C, et al. PedaleoVR: usability study of a virtual reality application for cycling exercise in patients with lower limb disorders and elderly people. PLoS One 2023 Feb 22;18(2):e0280743. [doi: 10.1371/journal.pone.0280743] [Medline: 36812248]
25. LaViola JJ. A discussion of cybersickness in virtual environments. ACM SIGCHI Bulletin 2000 Jan;32(1):47-56. [doi: 10.1145/333329.333344]
26. Costa-Felix R, Pomer-Escher A, Longo BB, et al. Neurorehabilitation platform based on EEG, sEMG and virtual reality using Robotic Monocycle. In: Costa-Felix R, Machado JC, Alvarenga AV, editors. XXVI Brazilian Congress on Biomedical Engineering: Springer; 2019:315-321. [doi: 10.1007/978-981-13-2119-1_48]
27. Kuo AD, Zajac FE. A biomechanical analysis of muscle strength as a limiting factor in standing posture. J Biomech 1993;26 Suppl 1:137-150. [doi: 10.1016/0021-9290(93)90085-s] [Medline: 8505348]
28. DataAvailability_JMIR. GitHub. URL: https://github.com/anarojoagusti/DataAvailability_JMIR [accessed 2023-12-20]

Abbreviations

CG: control group
EG: experimental group
HMD: head-mounted display
ROM: range of motion
VR: virtual reality

Edited by T Leung; submitted 05.05.22; peer-reviewed by I Cikajlo, P Postigo-Martin; revised version received 01.11.23; accepted 07.11.23; published 04.01.24.

Please cite as:
Rojo A, Castrillo Calvillo A, López C, Raya R, Moreno JC
Effects of a Virtual Reality Cycling Platform on Lower Limb Rehabilitation in Patients With Ataxia and Hemiparesis: Pilot Randomized Controlled Trial
JMIR Serious Games 2024;12:e39286
URL: https://games.jmir.org/2023/1/e39286
doi:10.2196/39286

Original Paper

Application of eHealth Tools in Anticoagulation Management After Cardiac Valve Replacement: Scoping Review Coupled With Bibliometric Analysis

Ying Wu[1,2*], MSc; Xiaohui Wang[2*], MSc; Mengyao Zhou[2], BSc; Zhuoer Huang[2], MSc; Lijuan Liu[3], MSc; Li Cong[2], PhD

[1]Center for Moral Culture, Hunan Normal University, Changsha, China

[2]School of Medicine, Hunan Normal University, Changsha, China

[3]Teaching and Research Section of Clinical Nursing, Xiangya Hospital of Central South University, Changsha, China

[*]these authors contributed equally

Corresponding Author:
Li Cong, PhD
School of Medicine
Hunan Normal University
371 Tongzipo Road
Changsha, 410013
China
Phone: 86 0731 889124
Email: congli@hunnu.edu.cn

Abstract

Background: Anticoagulation management can effectively prevent complications in patients undergoing cardiac valve replacement (CVR). The emergence of eHealth tools provides new prospects for the management of long-term anticoagulants. However, there is no comprehensive summary of the application of eHealth tools in anticoagulation management after CVR.

Objective: Our objective is to clarify the current state, trends, benefits, and challenges of using eHealth tools in the anticoagulation management of patients after CVR and provide future directions and recommendations for development in this field.

Methods: This scoping review follows the 5-step framework developed by Arksey and O'Malley. We searched 5 databases such as PubMed, MEDLINE, Web of Science, CINAHL, and Embase using keywords such as "eHealth," "anticoagulation," and "valve replacement." We included papers on the practical application of eHealth tools and excluded papers describing the underlying mechanisms for developing eHealth tools. The search time ranged from the database inception to March 1, 2023. The study findings were reported according to the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews). Additionally, VOSviewer (version 1.6.18) was used to construct visualization maps of countries, institutions, authors, and keywords to investigate the internal relations of included literature and to explore research hotspots and frontiers.

Results: This study included 25 studies that fulfilled the criteria. There were 27,050 participants in total, with the sample size of the included studies ranging from 49 to 13,219. The eHealth tools mainly include computer-based support systems, electronic health records, telemedicine platforms, and mobile apps. Compared to traditional anticoagulation management, eHealth tools can improve time in therapeutic range and life satisfaction. However, there is no significant impact observed in terms of economic benefits and anticoagulation-related complications. Bibliometric analysis suggests the potential for increased collaboration and opportunities among countries and academic institutions. Italy had the widest cooperative relationships. Machine learning and artificial intelligence are the popular research directions in anticoagulation management.

Conclusions: eHealth tools exhibit promise for clinical applications in anticoagulation management after CVR, with the potential to enhance postoperative rehabilitation. Further high-quality research is needed to explore the economic benefits of eHealth tools in long-term anticoagulant therapy and the potential to reduce the occurrence of adverse events.

(JMIR Mhealth Uhealth 2024;12:e48716) doi:10.2196/48716

KEYWORDS

eHealth tool; cardiac valve replacement; anticoagulation management; scoping review; bibliometrics analysis; rehabilitation

Introduction

Valvular heart disease involves damage to the cardiac valves caused by various factors such as valve degeneration and rheumatic heart disease [1]. Currently, an estimated 209 million people suffer from valvular heart disease worldwide. With the aging population, it is expected that the prevalence of valvular heart disease among those aged 65 years and older will increase by 50% globally by 2030 [2,3]. Furthermore, valvular heart disease caused approximately 300,000 deaths in 2019, accounting for nearly 2% of cardiovascular disease deaths worldwide. Every year, 8.7 million life years and 10.7 million disability-adjusted life years are lost to valvular heart disease [2]. The fundamental treatment for valvular heart disease is the repair or replacement of damaged heart valves through surgery or interventional therapy. Patients need long-term oral anticoagulants after cardiac valve replacement (CVR) surgery; however, improper use of anticoagulants can lead to bleeding, embolism, eventual valve failure, repeat surgeries, and even death [4,5]. Efficient anticoagulation management (AM) plays a critical role in promoting safe medication therapy after CVR, reducing adverse events, and improving the long-term prognosis of patients.

The traditional AM model has many obstacles and challenges. On the one hand, patients need to travel between their home and hospital regularly for follow-up and laboratory examinations and long-term anticoagulant medication, which costs them a lot of time, energy, and economy after CVR and also reduces patients' confidence in rehabilitation and treatment compliance [6]. On the other hand, it is difficult for doctors to monitor and manage patients and adjust treatment plans promptly, which may affect the quality and effect of anticoagulant therapy, and increase the risk of adverse events [7]. Therefore, the exploration of personalized, intelligent, and efficient AM models is crucial in promoting cardiac recovery among patients undergoing CVR. Several guidelines highlight the potential of eHealth tools to facilitate AM and improve medical outcomes [8] and recommend the use of eHealth tools to assist physicians with anticoagulation decision-making support [9].

eHealth tools generally refer to digital devices based on information and communication technology, which can be used for disease prevention, diagnosis, treatment, monitoring, and management [10]. eHealth tools can provide patients with convenient, efficient, and accurate medical services through digital communication and remote disease management. Some studies have found that eHealth tools may become a potentially cost-effective and effective alternative to traditional "face-to-face" anticoagulation therapy. A meta-analysis of 12 randomized controlled trials (RCTs) showed that telemedicine combined with portable coagulometers significantly improved the time in the therapeutic range (TTR) and reduced the incidence of thrombotic events in patients with cardiovascular disease [11]. Inpatients undergoing anticoagulation therapy receiving the recommended dosage of anticoagulation software can effectively reduce readmissions and 30-day mortality as well as hospitalization costs [12]. A comprehensive understanding of the application status and effectiveness of eHealth tools is essential to establish a scientific, traceable, and integrated AM model.

Although several narrative reviews and meta-analyses have summarized the usage of mobile health technology in patients with cardiovascular diseases [13-15], uncertainties remain regarding the effectiveness and current status of using eHealth tools in AM of patients after CVR. A comprehensive summary of the development, application, and interrelationships among different research papers or groups is lacking. Scoping reviews entail extensive searches and a rigorous selection of research literature on a specific topic, which can include a comprehensive overview of the current research status and trends of the topic [16]. Scientometric analysis can quantitatively evaluate cooperative relationships by statistically analyzing publications and graphically presenting the social and intellectual connections of relevant literature [17]. Through the joint analysis of scoping review and scientometric analysis, the research status and trends of a certain field can be comprehensively evaluated from different perspectives, and the blind spots and unsolved problems can be determined to provide guidance and enlightenment for further research.

By conducting a scoping review and scientometric analysis, this study aimed to (1) summarize the application status of eHealth tools in AM of patients after CVR, (2) identify the hotspots and provide guidance for future research and practice, and (3) to provide a reference for promoting the wider application and sustainable development of eHealth tools in AM of patients after CVR.

Methods

Design

We used Arksey and O'Malley's [18] 5-step framework for the scoping review. This review also followed the recommended items in the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist (Multimedia Appendix 1). Five commonly used databases were searched, and the papers were screened based on the title, abstract, and full text. We exported the complete records of filtered papers as plain text files and imported them into VOSviewer software (version 1.6.18; Centre for Science and Technology Studies) to build visualization maps. We chose coauthorship analysis to construct network visualization maps of countries, institutions, and authors to understand the status of research collaboration in the field of eHealth tools. By cluster analysis and keyword overlay visualization, keywords are divided into different clusters and stacked over time, which can identify different themes and current research hotspots.

Scoping Review

In contrast to systematic and narrative reviews, scoping reviews focus on an initial appraisal of the current extent, scope, and

nature of the research literature and take the dissemination process further by summarizing the relevant existing research activities. It is also an appropriate way to map the key concepts and identify knowledge gaps [19]. We aimed to provide an overview of the use of eHealth tools to assist anticoagulation therapy in patients after CVR and highlight the current status, trends, and challenges in this field. Therefore, the scoping review was appropriate for this study.

Identifying the Research Question for Scoping Review

The first question that guided our scoping review was what are the range and effectiveness of eHealth tools services in AM of patients after CVR? The second one was what are the benefits and barriers of applying eHealth tools in AM?

Identifying Databases and Studies

PubMed, MEDLINE, Web of Science, CINAHL, and Embase were searched from inception to March 1, 2023. Searches were not limited to a specific geographic region, and any literature published in non-English languages was excluded. Searches included combinations of free text words and index terms using Boolean operators. Moreover, a manual retrospective search of the references was conducted as a supplement. Detailed search strategies for each database are described in Multimedia Appendix 2.

Study Selection

The Population, Concept, and Context framework is recommended by the Joanna Briggs Institute to identify the main concepts in primary review questions, guide the search strategy, and ensure application of the inclusion and exclusion criteria [20]. Therefore, we used the Population, Concept, and Context framework to regulate the scoping review process. Population was defined as patients who had undergone CVR surgery and received anticoagulant therapy postoperatively. Concept referred to the practical application and effectiveness of eHealth tools in the AM of patients after CVR. Context had no special restrictions, as eHealth tools can be applied in the patient's home, hospital, anticoagulant therapy clinic, primary care center, and so forth. The inclusion criteria were as follows: (1) publication types including cross-sectional studies, longitudinal studies, cohort studies, case-control studies, pilot studies, and RCTs; (2) published in full text; and (3) published from the inception of each database to March 1, 2023. The exclusion criteria were as follows: (1) non-English publications; (2) publication types including empirical research, reviews, editorials, reports, case reports, letters, and conference proceedings or papers or abstracts; (3) qualitative studies reporting user experiences about eHealth tools in AM of patients after CVR; and (4) studies describing only the potential mechanisms or development process of eHealth tools, rather than their practical applications. The literature records retrieved were imported into Note Express software to screen for duplicate papers. Two researchers (YW and XW) independently conducted the initial screening of the titles and abstracts based on the inclusion and exclusion criteria. Full texts were then examined for secondary screening of potentially eligible papers. Any disagreement during the screening process was resolved through discussion with the third researcher (LC), and the final selection of papers was determined based on the established criteria.

Data Extraction and Analysis

Two researchers (YW and XW) independently extracted data from the included studies using standardized tables and cross-checking their findings. Any discrepancies were resolved through discussion with the third researcher (LC). The information extracted included authors, year, country, study design, study objective, participants' characteristics, study location, content elements, outcome measures, and study conclusions. We summarized the literature on the services and effects of eHealth tools in the AM of patients after CVR and presented the main concepts and findings of the literature using data charts and tables.

Bibliometric Analysis

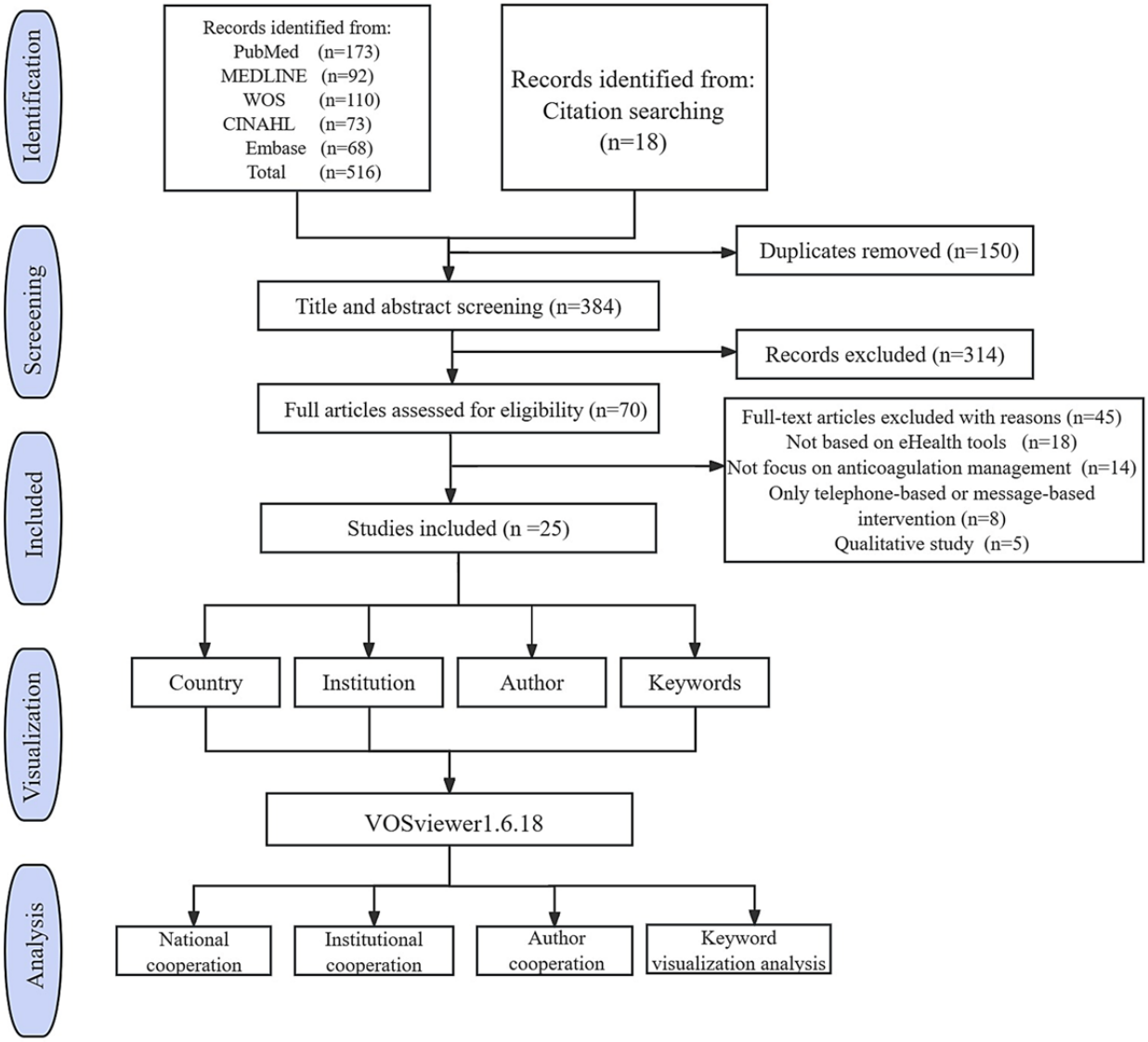
VOSviewer is a widely used tool by researchers for bibliometric analysis, providing effective visualization and revealing connections between research materials [21]. The process of scientometric analysis in this study included 2 parts. First, coauthorship analysis was performed to explore the relationship among researchers, research institutions, and countries, which contributed to understanding the trends of scientific cooperation. We chose "coauthorship" as the analysis type, selected "countries," "organizations," and "authors" as the units of analysis and "full counting" as the counting method. To achieve clearer effects, the minimum number for each project was set as 1. Second, keywords network visualization and overlay visualization were analyzed to explore the hotspots and frontiers. We chose the "author keywords" in the "co-occurrence" type for analysis, the counting method was "full counting," and the minimum cluster size to 6. We merged keywords with the same meaning and deleted redundant keywords. The minimum number was set to 1, and we obtained 40 keywords in total.

Results

Characteristics of the Included Studies

We conducted a systematic search of 5 databases, which yielded 534 studies that may be relevant to the topic. After eliminating duplicates, screening 384 titles and abstracts, and reviewing 70 full texts, we found that 25 papers met the inclusion and exclusion criteria. Figure 1 shows the selection process flowchart based on PRISMA-ScR [22]. The literature included in this study comprised 12 RCTs [23-34], 8 cohort studies [35-42], 2 pilot studies [43,44], 1 longitudinal study [45], 1 cross-sectional study [46], and 1 cross-over study [47]. The 25 papers comprised 27,050 participants in total, with the sample size for each study ranging from 49 to 13,219, the duration of intervention was 1-24 months, and the follow-up period was 1-514 months. Most of the participants were older than 40 years of age and had undergone mechanical valve replacement surgery. A table in Multimedia Appendix 3 provides a summary of the study characteristics and participant demographics.

Figure 1. Flow diagram of the article selection process based on PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines.



Types of Anticoagulation Services Offered by eHealth Tools

Through the comprehensive understanding and interpretation of the included literature, we found that eHealth tools can be broadly classified into 4: computer-based support systems, electronic health records (EHRs), telemedicine platforms, and mobile apps. The application scenarios of eHealth tools are gradually expanding from large general hospitals to primary health care centers and finally to patients' homes. Most studies have reported the beneficial effects of eHealth tools on the rehabilitation outcome and anticoagulation quality in patients undergoing CVR, reflecting the potential and application prospect of eHealth tools in this field. Table 1 shows the specific types of eHealth tools and their application scenarios. Multimedia Appendix 3 shows the specific services provided by eHealth tools and their impacts on anticoagulation outcomes.

Table 1. Types and application scenarios of eHealth tools (N=25).

Author, Year, Country	eHealth tools	Places
Fitzmaurice et al (1996) [23], UK	Decision support systems (Anticoagulation Management Support System, Warwick)	Primary care clinic
Fitzmaurice et al (2000) [25], UK	Softop Information (Warwick, UK)	Primary care practices
Testa et al. (2006) [44], Italy	Electronic patient records (TaoNet, EDP-Progetti, Bolzano, Italy)	Peripheral health units
Poller et al (2008) [27], UK	PARMA5[a] (Instrumentation Laboratory, Milan, Italy) and DAWN AC (program 4S Dawn Clinical Software, Milnthorpe, UK)	Hospital
Ryan et al (2009) [28], Ireland	CoagCare (ZyCare Inc., Chapel Hill, NC, United States)	Home
Cafolla et al (2011) [39], Italy	Automated computer systems coagulometer (HemoSense, San Jose, CA, United States)	Hospital
Bussey et al (2013) [43], United States	ClotFree system (ClotFree; Genesis Advanced Technologies, Lakehills, TX, United States)	Clinical research center
Ferrando et al (2015) [40], Spain	SintromacWebInternet-based system (Grifols, Barcelona, Spain)	Home
Cao et al (2018) [37], China	Digital anticoagulation clinic	Home and hospital
Zhu et al (2021) [32], China	A mobile user interface medical network follow-up platform	Home
Cao et al (2021) [42], China	Alfalfa	Home
Alanazi et al (2022) [47], Saudi Arabia	WhatsApp	Home
Amruthlal et al (2022) [46], India	Smartphone app	Home
Ageno and Turpie (1998) [24], Canada	DAWN AC (Business Technology, Milnthorpe, Cumbria, the UK)	Hospital
Manotti et al (2001) [26], Italy	PARMA system (release 3.2, Parma, Italy)	Manotti et al (2001), Italy
O'Shea et al (2008) [35], United States	Management program (CoagCare, ZyCare Inc, Chapel Hill, NC, United States)	Home
Soliman Hamad et al (2009) [29], the Netherlands	Anticoagulant aid website	Home
Christensen et al (2011) [30], Denmark	Computer system (CSO/AC; IntraMed A/S, Værløse, Denmark)	Home
Thompson et al (2013) [33], United States	Automated computer systems coagulometer (HemoSense, San Jose, CA, United States)	Hospital
Koertke et al (2015) [34], Germany	The Institute for Applied Telemedicine	Heart centers
Talboom et al (2017) [38], the Netherlands	Portavita eHealth platform	Home
Brasen et al (2019) [31], Denmark	Computer system (CSO/AC; IntraMed A/S, Værløse, Denmark)	Home
Jiang et al (2021) [41], China	Yixing App	Home
Jiang et al (2022) [36], China	Alfalfa	Home
Erba et al (2022) [45], Italy	PARMA GTS (Werfen, Milan, Italy) and WhatsApp	Anticoagulant clinic and home

[a]PARMA: Program for Archive, Refertation, and Monitoring of Anticoagulated patients.

Of the 25 included studies, 6 used computer-based support systems [23-27,39], 4 used EHR [33,38,43,44], 9 applied telemedicine platforms [28-32,34,35,37,40], and 6 used mobile apps [36,41,45-47]. Computer-based support systems are computer applications designed to aid clinicians in making diagnostic and therapeutic decisions in patient care [48]. Such support systems are typically installed in hospitals, large clinics for anticoagulant therapy, and primary health care centers. They can help doctors predict the appropriate dose of anticoagulants and the timing of the next international normalized ratio (INR) test based on the patient's latest INR value [23-27,39]. EHRs are generated by 1 or more interactions between medical institutions and patients [49]. Doctors use these tools to record the patients' medication information, laboratory measurement data, clinical history, or symptoms of discomfort for reference in the next visit. They use this information to make the best decision possible and summarize their clinical experience, which provides an important basis for improving the overall medical

level. EHR can be shared between primary health care clinics and hospitals through a central database to support CVR surgical follow-up [33].

A telemedicine platform is a means of receiving medical services remotely through various communication technologies [50]. It enables remote communication and data sharing between doctors and patients, thus providing digital medical services for patients after CVR, such as web-based consultation, real-time communication, and disease management [34,37,40,43]. Patients can use portable devices to measure INR at home and transmit the results to the telemedicine platforms. Doctors can then adjust the dosage of anticoagulants and the treatment plan remotely and determine the changes in anticoagulants and conditions of patients after CVR on time [28-30,35]. Moreover, telemedicine platforms can help to supervise the patient's self-management regime at home for a long time, helping patients correctly understand and implement treatment plans and improving treatment compliance and outcomes [28,29,31,37].

Mobile apps are services that run on smartphones, tablets, or other mobile devices. Some have been developed specifically for AM [36,41,42,46], while others are smartphone-based messaging apps [45,47]. In the research we included, mobile apps mainly mentioned the following four services: (1) Medication assistant: the system automatically generated an oral warfarin regimen based on the patient's latest INR value and previous warfarin dose, which were reviewed by specialists and sent to patients [42,46]. At the same time, the apps set intelligent reminders every day to urge the patient to take medicines regularly [36,41]. (2) Digital consultation: patients could communicate with doctors on the internet about their condition and anticoagulation treatment [47]. (3) Push health information: apps could send videos or pictures to help patients understand disease-related information [41,42]. (4) Patients' interactive community: the patients could communicate with other postoperative patients, share their own rehabilitation experiences, and gain support and encouragement [41].

Application Effects of eHealth Tools in AM

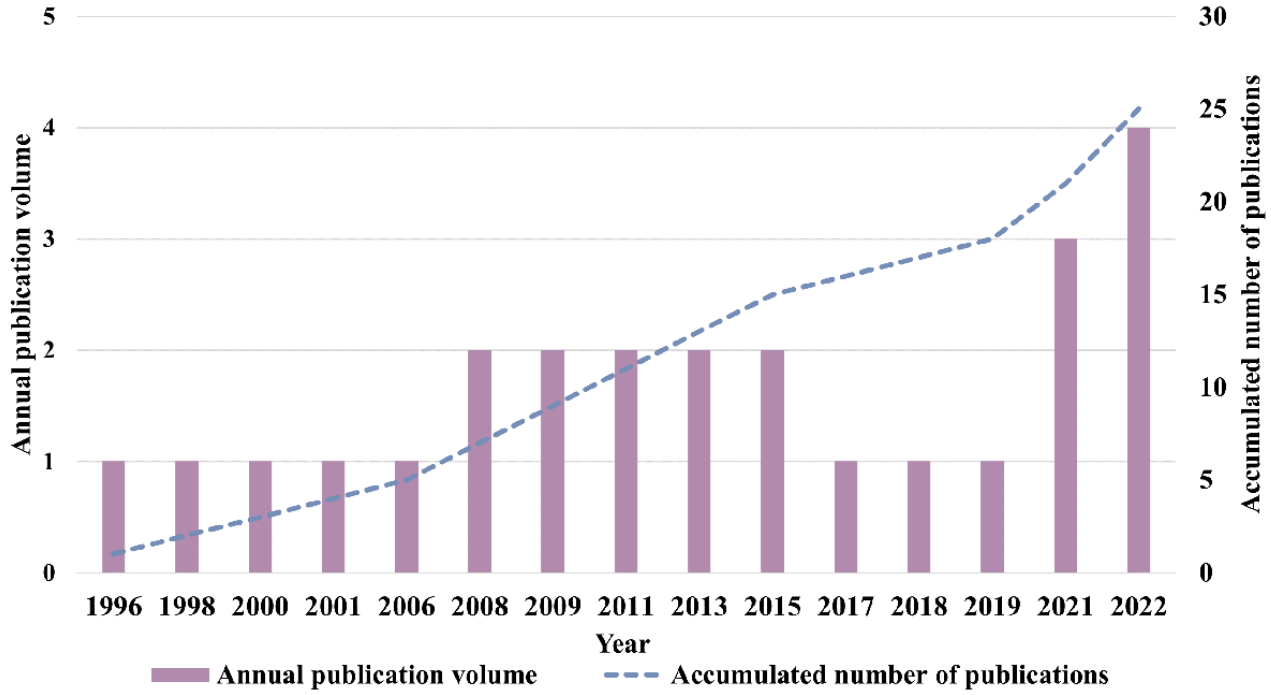
Comprehensive considering the evaluation indexes of anticoagulation effects on patients after CVR can provide better scientific guidance for clinicians and patients. In the literature included, the evaluation of eHealth tools in AM mainly involved three aspects. (1) Clinical outcome: This indicator was reported in 24 papers, including TTR (n=19) [23-28,30,32,33,35-40,42-44,47], rate of achieving target INR (n=5) [27,29,31,42,43], bleeding or thrombotic events (n=18) [23,25,27,28,30,32-42,44,45], and mortality (n=1) [34]. Seven RCTs have demonstrated that using eHealth tools for AM in patients after CVR is a secure and efficacious approach that significantly enhances TTR (P<.05) [23,25,26,28,30,32,33]. Computer-based support systems dosing was found to be more effective in improving TTR than medical staff–monitored dosage at the majority of centers (P<.001) [24,27,42]. Meanwhile, eHealth tools had a significantly higher number of INR within the target range compared to the conventional administration group (P<.05) [27,29]. However, several studies showed no statistical difference (P>.05) in the occurrence of bleeding or thrombotic events between the eHealth tools group and the conventional group [23,25,27,30,33]. (2) Economic benefits: Five papers focused on the health economic benefits of eHealth tools, including the frequency of INR tests (n=3) [24,33,43] and cost-effectiveness (n=2) [23,25]. Research showed that the cost of using computerized decision support software was offset by not going to the hospital [23]. However, when considering the costs associated with establishing a nurse-led clinic, the total cost was higher than traditional treatment [25]. In addition, patient self-testing at home could accelerate INR results, but it is not statistically significant in reducing INR test frequency and generating economic benefits [33,34]. (3) Patient satisfaction: The conclusion of 2 studies revealed that most patients were satisfied with the use of eHealth tools (P<.001) [23,44], which also improved their quality of life [29,44]. These findings demonstrate the potential of eHealth tools in enhancing AM. However, further studies are warranted to explore the economic benefits of long-term anticoagulant therapy and determine whether it can reduce the incidence of clinical adverse events in patients after CVR.

Scientometric Analysis

Publication Trends

The earliest literature on using eHealth tools for AM in patients after CVR was published in 1996. Figure 2 shows the trend of publications, with the number of published papers gradually increasing after 2019. The purple bar above the year label represents the publication volume, and the blue line shows cumulative publication trends. China (n=5) and Italy (n=4) stand out as the 2 countries with the most published papers, the United States and the United Kingdom have each published 3 papers, and other nations have published only 1 or 2 papers. Developed countries had research published around 2000, while transitional countries, such as China and India, only paid attention to this topic around 2019, which may be influenced by cultural differences and medical and economic levels. The majority of the 25 papers in this review were published in thrombus-related journals. However, the application of eHealth tools in the AM of patients after CVR involves not only cardiology and antithrombotic therapy but also the intersection and interdisciplinary cooperation of medical informatics, electronic technology, and health sciences. Some of the papers included in this review were also published in internet medicine journals. Therefore, when seeking the latest progress, readers should not focus only on traditional thrombosis-related journals.

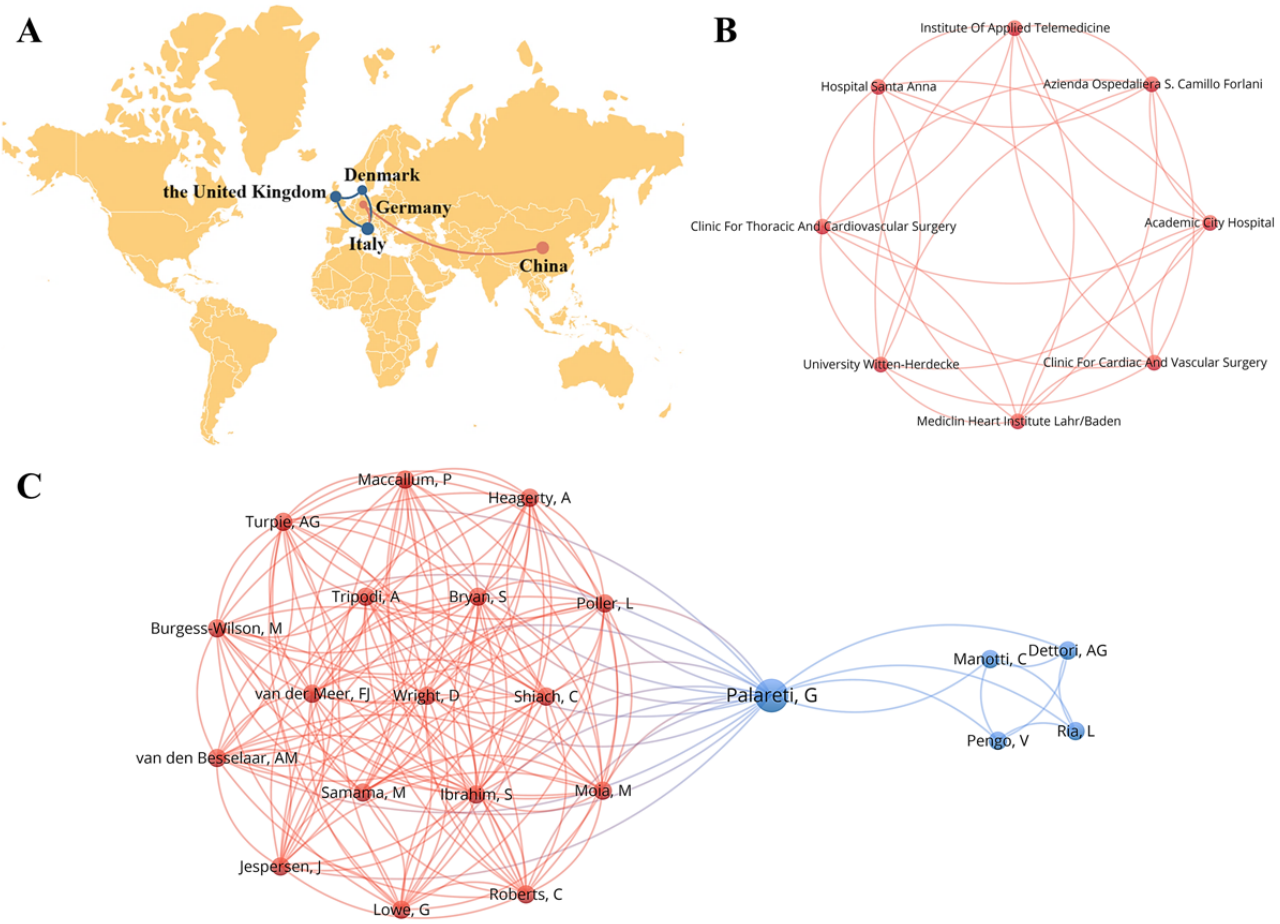
Figure 2. Publication trends of research papers of eHealth tools in anticoagulation management (AM) of patients undergoing cardiac valve replacement (CVR).



Collaborative Analysis of Countries, Research Institutions, and Authors

The 25 papers selected for this review originated from 12 countries and were associated with 46 institutions. Figure 3A depicts the academic collaboration among 5 of these countries, with China being the only transitional country represented. Each node represents a country, and the connection between nodes represents the cooperative relationship. Figure 3B shows that only 8 of 46 participating institutions show close cooperative relationships, and most of them are from Italy. Each node represents an institution, and the connection between nodes represents the cooperative relationship. It should be noted that there are deficiencies in the exchanges and cooperation between transitional and high-income countries, which poses challenges in enhancing the sharing and complementarity among international academic resources. Furthermore, 89 authors participated in the publication of 25 papers. Each node represents an author, the node size depends on the number of authorial publications, and the different colors of the line represent different cooperation networks (Figure 3C). Two collaborative networks have been formed centered on Palareti G, who is affiliated with the Orsola-Malpighi Hospital in Italy and plays a crucial role in the development of eHealth tools in the AM of patients after CVR.

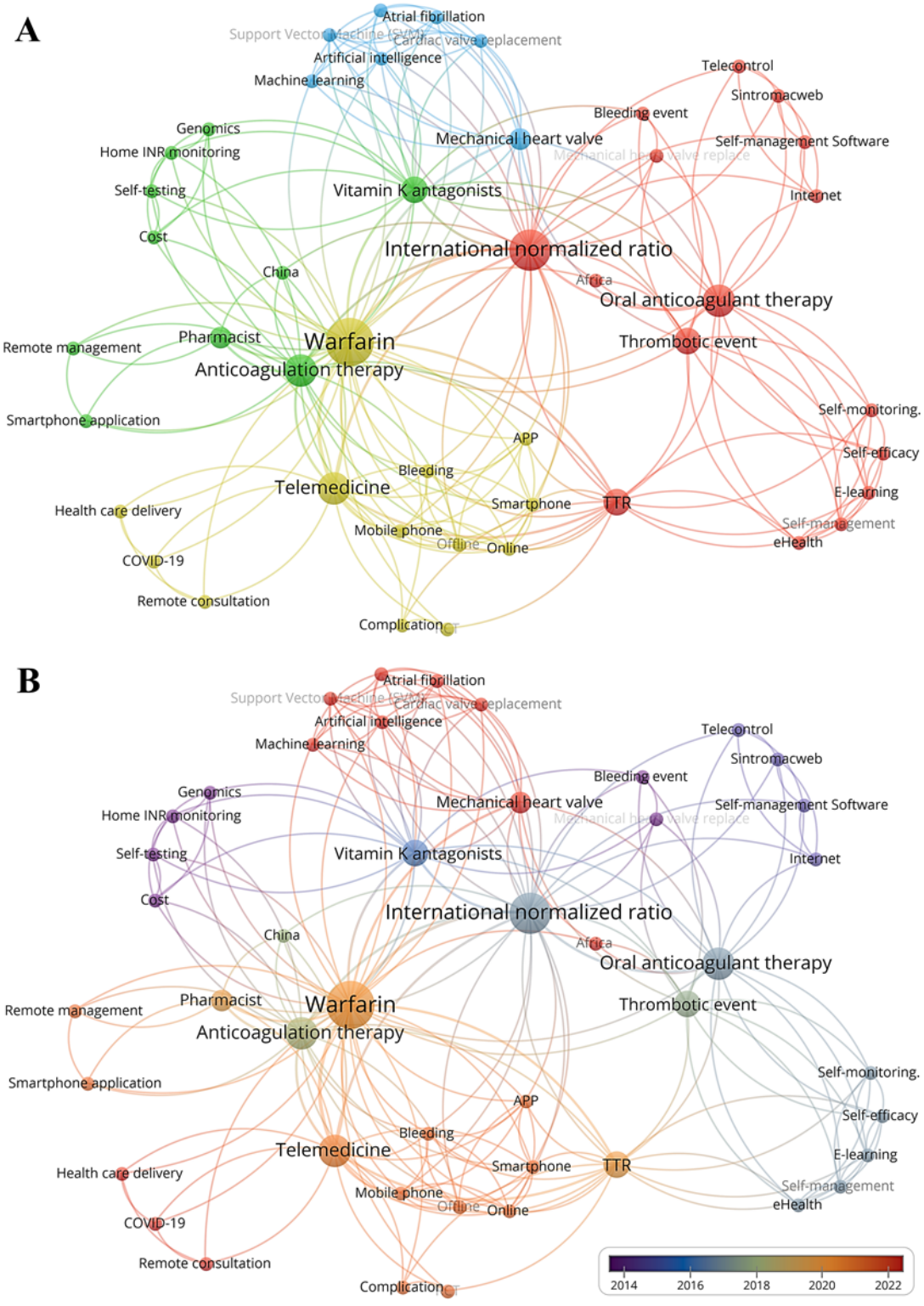
Figure 3. Collaborative network among countries, research institutions, and authors in publications related to the use of eHealth tools for anticoagulation management (AM) in patients after cardiac valve replacement (CVR). (A) Cooperation map among countries. (B) Collaboration network among research institutions. (C) Cooperation network of authors.



Keyword Co-Occurrence Analysis and Cluster Analysis

To present the keyword map more clearly, a cluster analysis was conducted for keywords. Figure 4A indicates that current research on the application of eHealth tools in AM after CVR is mainly focused on 4 distinct areas. The node size indicates the frequency of keyword occurrence, and different colors represent different clusters. We found that 16 keywords in the red cluster were closely related to thrombus formation and bleeding, 10 keywords in the green cluster related to the innovation of INR test techniques (home testing, self-testing, etc), keywords in the blue cluster referred to artificial intelligence (AI) and machine learning, and yellow cluster included telemedicine, health care delivery, and remote consultation.

Figure 4. Visualization of keywords of eHealth tools in anticoagulation management (AM) of patients undergoing cardiac valce replacement (CVR). (A) Keywords clustering visualization. (B) Keywords overlay time visualization.



To further explore the hotspots and future directions, we superimposed time on keyword clustering to create keyword overlay visualization (Figure 4B). The larger the node size, the higher the frequency of keywords and the color of nodes corresponded to the average appearance time of keywords. According to high-frequency keyword analysis, anticoagulant drugs for patients undergoing CVR mainly included vitamin K antagonists led by warfarin. Additionally, the research hotspots of eHealth tools in AM have focused on AI, machine learning, and smartphone apps in recent years. Amruthlal et al [46] constructed warfarin dosage prediction models by linear regression, support vector regression, logistic regression, and multilayer perceptron and found that the prediction model based on support vector regression showed the best predictive performance. Support vector regression was installed in a simple user-friendly Android mobile app. The introduction of advanced

machine learning algorithms provides more choice and support for AM, which may help to improve the efficiency and accuracy of anticoagulant treatment.

Discussion

Summary

This review provides a full evaluation and scientometric analysis of the use of eHealth tools in the AM of patients undergoing CVR surgery. We identified and included 25 studies investigating eHealth tools, such as computer-based support systems, EHRs, telemedicine platforms, and mobile apps. The development of eHealth tools is becoming increasingly important in promoting AM, and the emergence of machine learning and AI has introduced new opportunities for innovation and promotion. Over the past decade, there has been a lack of large-sample and long-term follow-up RCT studies on AM based on eHealth tools, highlighting the need for further research to verify their roles and effects. There remain many areas for improvement to promote the application of eHealth tools in the AM of patients after CVR.

Key Developments and Benefits Provided by eHealth Tools for AM

The AM of patients after CVR involves a large workload, complex patient information, high technical content, and a long follow-up period. The establishment and application of eHealth tools can improve the efficiency of clinical and management work [36,41,48,49]. The computer-based support system facilitates the standardization of anticoagulant recommendations and avoids differences in the dosage prescribed by medical staff, especially in primary health care institutions lacking anticoagulation therapy experience [51]. However, previously, due to technological limitations, these systems could not closely approximate doctors' judgment and decision-making skills, and manual verification and intervention were needed to ensure accuracy [24]. With the development of machine learning, the use of deep learning, reinforcement learning, and ensemble learning are increasing gradually in warfarin management after CVR, which can further improve the accuracy of prediction [52,53]. Additionally, doctors can adjust anticoagulant doses of patients more accurately according to their EHR, thus avoiding excessive or insufficient therapy and reducing the occurrence of complications. Although these 2 eHealth tools are highly targeted, easy to operate, and improve work efficiency, their functions are relatively simple and not enough for more comprehensive management of patients undergoing anticoagulant therapy after CVR.

Telemedicine platforms and mobile apps offer a promising solution for expanding access to high-quality medical resources by providing a digital communication channel between clinicians and patients who have undergone CVR [32-36,38,39,41,44]. Through these 2 ways, doctors can provide electronic prescriptions, digital consultations, health education, and self-management support to not only help patients better understand postoperative rehabilitation but also develop their knowledge and skills and improve the compliance, safety, and effectiveness of AM. However, telemedicine platforms depend on high-quality communication devices to be compatible with other medical devices and systems to perform remote processing of medical data and real-time monitoring of patients. In contrast, mobile apps are more portable and can be used anytime and anywhere, with a more intuitive interface and personalized services for anticoagulant therapy [39,44-47]. However, the older population, who comprise the majority of patients undergoing CVR surgery, may have difficulty learning and adapting to mobile apps and data on the internet. This may affect the accuracy of communication and information transfer between doctors and patients and limit postoperative follow-up. Additionally, practical limitations, such as network instability, data security, low degree of automation, and the need for medical institution support, could affect the effectiveness of telemedicine platforms and mobile apps.

Research on Hotspots of eHealth Tools Based on Scientometric Analysis

Research on the use of eHealth tools for AM after CVR surgery is lacking on a global scale. Developed nations have conducted more studies in this area, which may be attributed to their advanced economic status, superior medical quality, technical proficiency, and greater patient acceptance [54,55]. To promote the development of eHealth tools, it is suggested that countries provide a platform for exchanging and sharing the latest research results and invite scholars from different fields to participate in medical exchange programs, visiting scholar programs, medical conferences, and seminars. Moreover, universities can encourage students to effectively establish a cooperative culture, organize interdisciplinary research groups, conduct interdisciplinary training and education, and cultivate medical talents with an interdisciplinary background and international vision.

Based on the results of keyword clustering and overlay visualization, AI and machine learning have been adopted as research methods for developing eHealth tools and become a hotspot in recent years. These methods have provided valuable insights for future research. Genetic algorithms, backpropagation neural networks, and adapted neural-fuzzy inference system models have the potential to establish more accurate and stable prediction models of warfarin individual maintenance dose for patients after CVR [52,56]. Medical professionals can use biometric technology to visually identify the patient, the medication, and the confirmed ingestion, thereby reducing the risk of noncompliance in patients' anticoagulant therapy [57]. To promote the wide application and sustainable development of eHealth tools, it is necessary to research intelligent auxiliary tools, combining sophisticated machine learning techniques and AI models to enhance the precision and dependability of prediction and innovate anticoagulant therapy tools.

Future Directions for eHealth Tools in AM of Patients After CVR

In 25 studies included, although the main outcome indicators of eHealth tools in AM covered most of the clinical indicators, they could not fully reflect the overall situation of patients from the perspectives of economy, society, psychology, and user experience. Therefore, it is suggested that improvements can be made in the following aspects: (1) Patient interaction indicators: Researchers can monitor the frequency of patients

logging in or using eHealth tools in the background, learn about their experience and needs, and then adjust the tool design and functions accordingly. For example, a user interface with voice support, large font, and novice guidance can help patients overcome the potential obstacles in using eHealth tools [58]. (2) Indicators of acceptability: Researchers should evaluate the acceptance of patients and health care providers using patient-physician satisfaction, use rate, and patient-medication compliance. (3) Anticoagulant dosage index: This can be used to evaluate the accuracy and consistency of manual administration and eHealth tool–assisted administration to ensure the correctness of AM decisions. (4) Sustainability indicators: Patients can evaluate the user experience of eHealth tools from engagement, function, esthetics, and information quality to ensure long-term use and promotion [59].

In addition to the aforementioned optimization metrics, several other important factors should be considered to facilitate eHealth tools. First, it is critical to ensure that the design of eHealth tools is compatible with other medical devices and systems that the patients may use (eg, blood clotting machines and electrocardiographs). To minimize errors and improve data accuracy, these tools should connect seamlessly, allowing data to be shared and integrated across different platforms. Lubitz et al [60] used compatible, wearable devices and Android or iOS smartphones–assisted electrocardiogram patch monitoring, which effectively identified undiagnosed atrial fibrillation at an early stage. Second, providing effective training and support to older or less educated patients is critical for increasing the acceptance and use of eHealth tools. Previous studies have shown that tailored coaching and education programs can help patients operate these tools effectively [61]. Third, the use of advanced machine learning algorithms and AI models helps further improve the accuracy and reliability of eHealth tools. Zeng et al [62] used various machine learning to construct a dynamic anticoagulant treatment scheme for hospitalized patients after CVR surgery. The results showed that the performance of reinforcement learning was significantly better than the other algorithms, and the quality of anticoagulation was significantly optimized [62]. Finally, implementing effective evaluation mechanisms for patients lost to follow-up is critical to ensure they receive appropriate anticoagulant treatment over the long term. Porter et al [63] conducted a 2-year follow-up of patients receiving oral anticoagulants and found that 12-week INR test intervals were feasible, which saved patient's time and improved their compliance [63]. By addressing these factors and optimizing the application of eHealth tools in the AM of patients after CVR, it is possible to significantly improve the overall therapeutic effect and reduce the risk of complications.

Limitations

This review has some limitations. First, we did not assess the quality of the included studies because the study was to provide a comprehensive overview of the existing research landscape related to the application of eHealth tools in AM for patients after CVR. However, we believe that evaluating the quality of the literature will increase the significance of the study. Second, we did not include studies published in the form of other sources, such as qualitative studies, because we were unable to identify eligible studies. Third, concluding trends based on 25 papers are limited, and the conclusions of scientometric analysis should be interpreted with caution. The scientometric analysis in our research is influenced by several limitations, such as the omission of qualitative data, language barriers, and the potential for sample bias. To address the constraints associated with scientometric analysis, researchers can broaden their data sources, integrate qualitative data, and encompass papers published in various languages.

Conclusions

The application of AM based on eHealth tools is expected to truly reflect the social and economic benefits of digital intelligence, thus benefiting patients undergoing CVR. This study provides an overview of the scope, benefits, and future development of eHealth tools in AM for researchers, health care professionals, and post-CVR patients. However, the economic benefits and long-term impact of adverse events need further explored. It is suggested that future eHealth tools in AM should concentrate on enhancing patient contact, acceptance, cost-effectiveness, and sustainability while combining sophisticated algorithms to enhance the precision and dependability of eHealth tools.

Acknowledgments

The authors specially thank the National Natural Science Foundation of China, Hunan Provincial Natural Science Foundation, and Hunan Normal University. This work was supported by the National Natural Science Foundation of China (82100490 to LC), Hunan Provincial Natural Science Foundation (2023JJ30425 to LC), Hunan Normal University Students' Innovation and Entrepreneurship Training Program (2022263 to MZ), and Hunan Normal University School of Medicine Open Foundation Project (KF2022043 to XW).

Data Availability

The data sets used and analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews.
[PDF File (Adobe PDF File), 685 KB - mhealth_v12i1e48716_app1.pdf]

Multimedia Appendix 2
Search strategy per database.
[DOCX File , 20 KB - mhealth_v12i1e48716_app2.docx]

Multimedia Appendix 3
Summary of the study characteristics.
[DOCX File , 40 KB - mhealth_v12i1e48716_app3.docx]

References

1. Coisne A, Lancellotti P, Habib G, Garbi M, Dahl JS, Barbanti M, et al. ACC/AHA and ESC/EACTS guidelines for the management of valvular heart diseases: JACC guideline comparison. J Am Coll Cardiol 2023;82(8):721-734. [doi: 10.1016/j.jacc.2023.05.061] [Medline: 37587584]
2. Santangelo G, Bursi F, Faggiano A, Moscardelli S, Simeoli PS, Guazzi M, et al. The global burden of valvular heart disease: from clinical epidemiology to management. J Clin Med 2023;12(6):2178 [FREE Full text] [doi: 10.3390/jcm12062178] [Medline: 36983180]
3. Coffey S, Roberts-Thomson R, Brown A, Carapetis J, Chen M, Enriquez-Sarano M, et al. Global epidemiology of valvular heart disease. Nat Rev Cardiol 2021;18(12):853-864. [doi: 10.1038/s41569-021-00570-z] [Medline: 34172950]
4. Oz A, Tsoumas I, Lampropoulos K, Xanthos T, Karpettas N, Papadopoulos D. Cardiac rehabilitation after TAVI—a systematic review and meta-analysis. Curr Probl Cardiol 2023;48(3):101531. [doi: 10.1016/j.cpcardiol.2022.101531] [Medline: 36493915]
5. Tsoumas I, Oz A, Lampropoulos K. Optimal anticoagulation on TAVI patients based on thrombotic and bleeding risk and the challenge beyond: a systematic review and meta-analysis. Curr Probl Cardiol 2023;48(6):101632. [doi: 10.1016/j.cpcardiol.2023.101632] [Medline: 36724817]
6. Cromheecke ME, Levi M, Colly LP, de Mol BJ, Prins MH, Hutten BA, et al. Oral anticoagulation self-management and management by a specialist anticoagulation clinic: a randomised cross-over comparison. Lancet 2000;356(9224):97-102. [doi: 10.1016/S0140-6736(00)02470-3] [Medline: 10963245]
7. Menéndez-Jándula B, Souto JC, Oliver A, Montserrat I, Quintana M, Gich I, et al. Comparing self-management of oral anticoagulant therapy with clinic management: a randomized trial. Ann Intern Med 2005;142(1):1-10. [doi: 10.7326/0003-4819-142-1-200501040-00006] [Medline: 15630104]
8. Jain R, Stone JA, Agarwal G, Andrade JG, Bacon SL, Bajaj HS, et al. Canadian Cardiovascular Harmonized National Guideline Endeavour (C-CHANGE) guideline for the prevention and management of cardiovascular disease in primary care: 2022 update. CMAJ 2022;194(43):E1460-E1480 [FREE Full text] [doi: 10.1503/cmaj.220138] [Medline: 36343954]
9. Holbrook A, Schulman S, Witt DM, Vandvik PO, Fish J, Kovacs MJ, et al. Evidence-based management of anticoagulant therapy: antithrombotic therapy and prevention of thrombosis, 9th ed: American College of Chest Physicians evidence-based clinical practice guidelines. Chest 2012;141(2 Suppl):e152S-e184S [FREE Full text] [doi: 10.1378/chest.11-2295] [Medline: 22315259]
10. Scheerens C, Gilissen J, Volow AM, Powell JL, Ferguson CM, Farrell D, et al. Developing eHealth tools for diverse older adults: lessons learned from the prepare for your care program. J Am Geriatr Soc 2021;69(10):2939-2949 [FREE Full text] [doi: 10.1111/jgs.17284] [Medline: 34081773]
11. Huang Y, Xie Y, Huang L, Han Z. The value of anticoagulation management combining telemedicine and self-testing in cardiovascular diseases: a meta-analysis of randomized controlled trials. Ther Clin Risk Manag 2023;19:279-290 [FREE Full text] [doi: 10.2147/TCRM.S395578] [Medline: 36941980]
12. Wang JJ, Koulas I, Myrka A, Spyropoulos AC. Implementation of the management of anticoagulation in the periprocedural period app into an electronic health record: a cost-effectiveness analysis. Clin Appl Thromb Hemost 2023;29:10760296231154553 [FREE Full text] [doi: 10.1177/10760296231154553] [Medline: 36872909]
13. Huang Y, Huang L, Han Z. Combining portable coagulometers with the Internet: a new model of warfarin anticoagulation in patients following mechanical heart valve replacement. Front Surg 2022;9:1016278 [FREE Full text] [doi: 10.3389/fsurg.2022.1016278] [Medline: 36311931]
14. Jang I. A systematic review on mobile health applications' education program for patients taking oral anticoagulants. Int J Environ Res Public Health 2021;18(17):8902 [FREE Full text] [doi: 10.3390/ijerph18178902] [Medline: 34501492]
15. Ni R, Liu M, Huang S, Yang J. Effects of eHealth interventions on quality of life and psychological outcomes in cardiac surgery patients: systematic review and meta-analysis. J Med Internet Res 2022;24(8):e40090 [FREE Full text] [doi: 10.2196/40090] [Medline: 35972792]
16. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci 2010;5:69 [FREE Full text] [doi: 10.1186/1748-5908-5-69] [Medline: 20854677]

17. Hasan M, Abedin MZ, Amin MB, Nekmahmud M, Oláh J. Sustainable biofuel economy: a mapping through bibliometric research. J Environ Manage 2023;336:117644 [FREE Full text] [doi: 10.1016/j.jenvman.2023.117644] [Medline: 36893543]
18. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005;8(1):19-32. [doi: 10.1080/1364557032000119616]
19. Munn Z, Peters MDJ, Stern C, Tufanaru C, McArthur A, Aromataris E. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. BMC Med Res Methodol 2018;18(1):143 [FREE Full text] [doi: 10.1186/s12874-018-0611-x] [Medline: 30453902]
20. Pollock D, Peters MDJ, Khalil H, McInerney P, Alexander L, Tricco AC, et al. Recommendations for the extraction, analysis, and presentation of results in scoping reviews. JBI Evid Synth 2023;21(3):520-532 [FREE Full text] [doi: 10.11124/JBIES-22-00123] [Medline: 36081365]
21. van Eck NJ, Waltman L. Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics 2010;84(2):523-538 [FREE Full text] [doi: 10.1007/s11192-009-0146-3] [Medline: 20585380]
22. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
23. Fitzmaurice DA, Hobbs FD, Murray ET, Bradley CP, Holder R. Evaluation of computerized decision support for oral anticoagulation management based in primary care. Br J Gen Pract 1996;46(410):533-535 [FREE Full text] [Medline: 8917873]
24. Ageno W, Turpie AG. A randomized comparison of a computer-based dosing program with a manual system to monitor oral anticoagulant therapy. Thromb Res 1998;91(5):237-240. [doi: 10.1016/s0049-3848(98)00092-9] [Medline: 9755836]
25. Fitzmaurice DA, Hobbs FD, Murray ET, Holder RL, Allan TF, Rose PE. Oral anticoagulation management in primary care with the use of computerized decision support and near-patient testing: a randomized, controlled trial. Arch Intern Med 2000;160(15):2343-2348 [FREE Full text] [doi: 10.1001/archinte.160.15.2343] [Medline: 10927732]
26. Manotti C, Moia M, Palareti G, Pengo V, Ria L, Dettori AG. Effect of computer-aided management on the quality of treatment in anticoagulated patients: a prospective, randomized, multicenter trial of APROAT (Automated Program for Oral Anticoagulant Treatment). Haematologica 2001;86(10):1060-1670. [Medline: 11602412]
27. Poller L, Keown M, Ibrahim S, Lowe G, Moia M, Turpie AG, et al. An international multicenter randomized study of computer-assisted oral anticoagulant dosage vs. medical staff dosage. J Thromb Haemost 2008;6(6):935-943 [FREE Full text] [doi: 10.1111/j.1538-7836.2008.02959.x] [Medline: 18489430]
28. Ryan F, Byrne S, O'Shea S. Randomized controlled trial of supervised patient self-testing of warfarin therapy using an internet-based expert system. J Thromb Haemost 2009;7(8):1284-1290 [FREE Full text] [doi: 10.1111/j.1538-7836.2009.03497.x] [Medline: 19496921]
29. Hamad MAS, van Eekelen E, van Agt T, van Straten AHM. Self-management program improves anticoagulation control and quality of life: a prospective randomized study. Eur J Cardiothorac Surg 2009;35(2):265-269 [FREE Full text] [doi: 10.1016/j.ejcts.2008.10.020] [Medline: 19041254]
30. Christensen H, Lauterlein JJ, Sørensen PD, Petersen ERB, Madsen JS, Brandslund I. Home management of oral anticoagulation via telemedicine versus conventional hospital-based treatment. Telemed J E Health 2011;17(3):169-176 [FREE Full text] [doi: 10.1089/tmj.2010.0128] [Medline: 21254841]
31. Brasen CL, Madsen JS, Parkner T, Brandslund I. Home management of warfarin treatment through a real-time supervised telemedicine solution: a randomized controlled trial. Telemed J E Health 2019;25(2):109-115. [doi: 10.1089/tmj.2017.0260] [Medline: 29877778]
32. Zhu Z, Li C, Shen J, Wu K, Li Y, Liu K, et al. New internet-based warfarin anticoagulation management approach after mechanical heart valve replacement: prospective, multicenter, randomized controlled trial. J Med Internet Res 2021;23(8):e29529 [FREE Full text] [doi: 10.2196/29529] [Medline: 34397393]
33. Thompson JL, Burkhart HM, Daly RC, Dearani JA, Joyce LD, Suri RM, et al. Anticoagulation early after mechanical valve replacement: improved management with patient self-testing. J Thorac Cardiovasc Surg 2013;146(3):599-604 [FREE Full text] [doi: 10.1016/j.jtcvs.2012.03.088] [Medline: 22921821]
34. Koertke H, Zittermann A, Wagner O, Secer S, Sciangula A, Saggau W, et al. Telemedicine-guided, very low-dose international normalized ratio self-control in patients with mechanical heart valve implants. Eur Heart J 2015;36(21):1297-1305 [FREE Full text] [doi: 10.1093/eurheartj/ehu330] [Medline: 25205534]
35. O'Shea SI, Arcasoy MO, Samsa G, Cummings SE, Thames EH, Surwit RS, et al. Direct-to-patient expert system and home INR monitoring improves control of oral anticoagulation. J Thromb Thrombolysis 2008;26(1):14-21. [doi: 10.1007/s11239-007-0068-y] [Medline: 17616845]
36. Jiang S, Lv M, Zeng Z, Fang Z, Chen M, Qian J, et al. Efficacy and safety of app-based remote warfarin management during COVID-19-related lockdown: a retrospective cohort study. J Thromb Thrombolysis 2022;54(1):20-28 [FREE Full text] [doi: 10.1007/s11239-021-02630-0] [Medline: 35092566]
37. Cao H, Wu J, Zhang J. Outcomes of warfarin therapy managed by pharmacists via hospital anticoagulation clinic versus online anticoagulation clinic. Int J Clin Pharm 2018;40(5):1072-1077. [doi: 10.1007/s11096-018-0674-0] [Medline: 29956134]

38. Talboom-Kamp EPWA, Verdijk NA, Kasteleyn MJ, Harmans LM, Talboom IJSH, Numans ME, et al. Effect of a combined education and eHealth programme on the control of oral anticoagulation patients (PORTALS study): a parallel cohort design in Dutch primary care. BMJ Open 2017;7(9):e017909 [FREE Full text] [doi: 10.1136/bmjopen-2017-017909] [Medline: 28963313]
39. Cafolla A, Melizzi R, Baldacci E, Pignoloni P, Dragoni F, Campanelli M, et al. "Zeus" a new oral anticoagulant therapy dosing algorithm: a cohort study. Thromb Res 2011;128(4):325-330. [doi: 10.1016/j.thromres.2011.04.012] [Medline: 21600633]
40. Ferrando F, Mira Y. Effective and safe management of oral anticoagulation therapy in patients who use the internet-accessed telecontrol tool SintromacWeb. Interact J Med Res 2015;4(2):e10 [FREE Full text] [doi: 10.2196/ijmr.3610] [Medline: 25900775]
41. Jiang S, He Q, Yan J, Zhao L, Zheng Y, Chen P, et al. Evaluation of a pharmacist-led remote warfarin management model using a smartphone application (Yixing) in improving patients' knowledge and outcomes of anticoagulation therapy. Front Pharmacol 2021;12:677943 [FREE Full text] [doi: 10.3389/fphar.2021.677943] [Medline: 34276368]
42. Cao H, Jiang S, Lv M, Wu T, Chen W, Zhang J. Effectiveness of the Alfalfa App in warfarin therapy management for patients undergoing venous thrombosis prevention and treatment: cohort study. JMIR Mhealth Uhealth 2021;9(3):e23332 [FREE Full text] [doi: 10.2196/23332] [Medline: 33650976]
43. Bussey HI, Bussey M, Bussey-Smith KL, Frei CR. Evaluation of warfarin management with international normalized ratio self-testing and online remote monitoring and management plus low-dose vitamin K with genomic considerations: a pilot study. Pharmacotherapy 2013;33(11):1136-1146. [doi: 10.1002/phar.1343] [Medline: 24038412]
44. Testa S, Alatri A, Paoletti O, Morstabilini G, Medagliani MA, Denti N, et al. Reorganisation of an anticoagulation clinic using a telemedicine system: description of the model and preliminary results. Intern Emerg Med 2006;1(1):24-29. [doi: 10.1007/BF02934716] [Medline: 16941809]
45. Erba N, Tosetto A, Langer M, Abdallah SA, Giovanella E, Lentini S, et al. Oral anticoagulant management of patients with mechanical heart valves at the Salam Centre of Khartoum: observations on quality of anticoagulation and thrombotic risk. Thromb Res 2022;219:155-161 [FREE Full text] [doi: 10.1016/j.thromres.2022.09.018] [Medline: 36191535]
46. Amruthlal M, Devika S, Krishnan V, Suhail PAA, Menon AK, Thomas A, et al. Development and validation of a mobile application based on a machine learning model to aid in predicting dosage of Vitamin K antagonists among Indian patients post mechanical heart valve replacement. Indian Heart J 2022;74(6):469-473 [FREE Full text] [doi: 10.1016/j.ihj.2022.10.002] [Medline: 36243102]
47. Alanazi Z, Almutairi N, AlDukkan L, Arafat AA, Albabtain MA. Time in therapeutic range for virtual anticoagulation clinic versus in-person clinic during the COVID-19 pandemic: a crossover study. Ann Saudi Med 2022;42(5):305-308 [FREE Full text] [doi: 10.5144/0256-4947.2022.305] [Medline: 36252144]
48. Blum D, Raj SX, Oberholzer R, Riphagen II, Strasser F, Kaasa S, EURO IMPACT, European Intersectorial Multidisciplinary Palliative Care Research Training. Computer-based clinical decision support systems and patient-reported outcomes: a systematic review. Patient 2015;8(5):397-409. [doi: 10.1007/s40271-014-0100-1] [Medline: 25432150]
49. Xie F, Yuan H, Ning Y, Ong MEH, Feng M, Hsu W, et al. Deep learning for temporal data representation in electronic health records: a systematic review of challenges and methodologies. J Biomed Inform 2022;126:103980 [FREE Full text] [doi: 10.1016/j.jbi.2021.103980] [Medline: 34974189]
50. Cui F, He X, Zhai Y, Lyu M, Shi J, Sun D, et al. Application of telemedicine services based on a regional telemedicine platform in China from 2014 to 2020: longitudinal trend analysis. J Med Internet Res 2021;23(7):e28009 [FREE Full text] [doi: 10.2196/28009] [Medline: 34255686]
51. Vo AT, Gustafson DL. Telepharmacy in oncology care: a scoping review. J Telemed Telecare 2023;29(3):165-176. [doi: 10.1177/1357633X20975257] [Medline: 33377820]
52. Jahmunah V, Chen S, Oh SL, Acharya UR, Chowbay B. Automated warfarin dose prediction for Asian, American, and Caucasian populations using a deep neural network. Comput Biol Med 2023;153:106548. [doi: 10.1016/j.compbiomed.2023.106548] [Medline: 36652867]
53. Liu Y, Chen J, You Y, Xu A, Li P, Wang Y, et al. An ensemble learning based framework to estimate warfarin maintenance dose with cross-over variables exploration on incomplete data set. Comput Biol Med 2021;131:104242. [doi: 10.1016/j.compbiomed.2021.104242] [Medline: 33578070]
54. de Batlle J, Benítez ID, Moncusí-Moix A, Androutsos O, Barbastro RA, Antonini A, et al. GATEKEEPER's strategy for the multinational large-scale piloting of an eHealth platform: tutorial on how to identify relevant settings and use cases. J Med Internet Res 2023;25:e42187 [FREE Full text] [doi: 10.2196/42187] [Medline: 37379060]
55. Cascini F, Gentili A, Causio FA, Altamura G, Melnyk A, Beccia F, et al. Strengthening and promoting digital health practice: results from a global digital health partnership's survey. Front Public Health 2023;11:1147210 [FREE Full text] [doi: 10.3389/fpubh.2023.1147210] [Medline: 37404277]
56. Deniz-Garcia A, Fabelo H, Rodriguez-Almeida AJ, Zamora-Zamorano G, Castro-Fernandez M, Ruano MDPA, et al. Quality, usability, and effectiveness of mHealth apps and the role of artificial intelligence: current scenario and challenges. J Med Internet Res 2023;25:e44030 [FREE Full text] [doi: 10.2196/44030] [Medline: 37140973]

57. Labovitz DL, Shafner L, Gil MR, Virmani D, Hanina A. Using artificial intelligence to reduce the risk of nonadherence in patients on anticoagulation therapy. Stroke 2017;48(5):1416-1419 [FREE Full text] [doi: 10.1161/STROKEAHA.116.016281] [Medline: 28386037]
58. Portz JD, Bayliss EA, Bull S, Boxer RS, Bekelman DB, Gleason K, et al. Using the technology acceptance model to explore user experience, intent to use, and use behavior of a patient portal among older adults with multiple chronic conditions: descriptive qualitative study. J Med Internet Res 2019;21(4):e11604 [FREE Full text] [doi: 10.2196/11604] [Medline: 30958272]
59. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
60. Lubitz SA, Faranesh AZ, Selvaggi C, Atlas SJ, McManus DD, Singer DE, et al. Detection of atrial fibrillation in a large population using wearable devices: the fitbit heart study. Circulation 2022;146(19):1415-1424 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.122.060291] [Medline: 36148649]
61. Kraaijkamp JJM, van Dam van Isselt EF, Persoon A, Versluis A, Chavannes NH, Achterberg WP. eHealth in geriatric rehabilitation: systematic review of effectiveness, feasibility, and usability. J Med Internet Res 2021;23(8):e24015 [FREE Full text] [doi: 10.2196/24015] [Medline: 34420918]
62. Zeng J, Shao J, Lin S, Zhang H, Su X, Lian X, et al. Optimizing the dynamic treatment regime of in-hospital warfarin anticoagulation in patients after surgical valve replacement using reinforcement learning. J Am Med Inform Assoc 2022;29(10):1722-1732. [doi: 10.1093/jamia/ocac088] [Medline: 35864720]
63. Porter AL, Margolis AR, Staresinic CE, Nagy MW, Schoen RR, Ray CA, et al. Feasibility and safety of a 12-week INR follow-up protocol over 2 years in an anticoagulation clinic: a single-arm prospective cohort study. J Thromb Thrombolysis 2019;47(2):200-208 [FREE Full text] [doi: 10.1007/s11239-018-1760-9] [Medline: 30368762]

Abbreviations

AI: artificial intelligence
AM: anticoagulation management
CVR: cardiac valve replacement
EHR: electronic health record
INR: international normalized ratio
RCT: randomized controlled trial
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Review
TTR: time in therapy range

Edited by T de Azevedo Cardoso, L Buis, M Dorsch; submitted 05.05.23; peer-reviewed by Z Zhu, C Li, Z Zheng; comments to author 26.06.23; revised version received 20.07.23; accepted 07.12.23; published 05.01.24.

Please cite as:
Wu Y, Wang X, Zhou M, Huang Z, Liu L, Cong L
Application of eHealth Tools in Anticoagulation Management After Cardiac Valve Replacement: Scoping Review Coupled With Bibliometric Analysis
JMIR Mhealth Uhealth 2024;12:e48716
URL: https://mhealth.jmir.org/2024/1/e48716
doi:10.2196/48716
PMID:38180783

Review

Dissemination Strategies for mHealth Apps: Systematic Review

Henri Claude Moungui[1], MPH, MSc; Hugues Clotaire Nana-Djeunga[2], MSc, PhD; Che Frankline Anyiang[3], MSc; Mireia Cano[4], MSc; Jose Antonio Ruiz Postigo[5], MD, PhD; Carme Carrion[4], MSc, PhD

[1]Universitat Oberta de Catalunya, Barcelona, Spain

[2]Higher Institute for Scientific and Medical Research, Yaounde, Cameroon

[3]Texila American University, Georgetown, Guyana

[4]eHealth Lab Research Group, eHealth Center & School of Health Sciences, Universitat Oberta de Catalunya, Barcelona, Spain

[5]Prevention, Treatment and Care Unit, Department of Control of Neglected Tropical Diseases, World Health Organization, Geneva, Switzerland

Corresponding Author:
Henri Claude Moungui, MPH, MSc
Universitat Oberta de Catalunya
Rambla del Poblenou, 156
Barcelona, 08018
Spain
Phone: 34 672192283
Email: henrimoungui@yahoo.fr

Abstract

Background: Among the millions of mobile apps in existence, thousands fall under the category of mobile health (mHealth). Although the utility of mHealth apps has been demonstrated for disease diagnosis, treatment data management, and health promotion strategies, to be effective they must reach and be used by their target audience. An appropriate marketing strategy can ensure that apps reach potential users and potentially convert them to actual users. Such a strategy requires definitions of target end users, communication channels, and advertising content, as well as a timeline for effectively reaching and motivating end users to adopt and maintain engagement with the mHealth app.

Objective: The aim of this study was to identify strategies and elements that ensure that end users adopt and remain engaged with mHealth apps.

Methods: A systematic search of the PubMed, PsycINFO, Scopus, and CINAHL databases was conducted for suitable studies published between January 1, 2018, and September 30, 2022. Two researchers independently screened studies for inclusion, extracted data, and assessed the risk of bias. The main outcome was dissemination strategies for mHealth apps.

Results: Of the 648 papers retrieved from the selected databases, only 10 (1.5%) met the inclusion criteria. The marketing strategies used in these studies to inform potential users of the existence of mHealth apps and motivate download included both paid and unpaid strategies and used various channels, including social media, emails, printed posters, and face-to-face communication. Most of the studies reported a combination of marketing concepts used to advertise their mHealth apps. Advertising messages included instructions on where and how to download and install the apps. In most of the studies (6/10, 60%), instructions were oriented toward how to use the apps and maintain engagement with a health intervention. The most frequently used paid marketing platform was Facebook Ads Manager (2/10, 20%). Advertising performance was influenced by many factors, including but not limited to advertising content. In 1 (10%) of the 10 studies, animated graphics generated the greatest number of clicks compared with other image types. The metrics used to assess marketing strategy effectiveness were number of downloads; nonuse rate; dropout rate; adherence rate; duration of app use; and app usability over days, weeks, or months. Additional indicators such as cost per click, cost per install, and clickthrough rate were mainly used to assess the cost-effectiveness of paid marketing campaigns.

Conclusions: mHealth apps can be disseminated via paid and unpaid marketing strategies using various communication channels. The effects of these strategies are reflected in download numbers and user engagement with mHealth apps. Further research could provide guidance on a framework for disseminating mHealth apps and encouraging their routine use.

(JMIR Mhealth Uhealth 2024;12:e50293) doi:10.2196/50293

KEYWORDS

mobile health; mHealth; mobile health apps; mHealth apps; dissemination; marketing strategies; digital marketing; engagement; onboarding; systematic review; systematic; market; marketing; app; apps; adoption; consumer; mobile phone

Introduction

Mobile Health Apps

Among the vast array of mobile apps currently available, health care apps serve various purposes, including disease diagnosis [1], health promotion, and disease prevention [2,3]. Such uses of mobile technology to provide patients with health care support or health service providers with technical support in a direct, low-cost, and engaging manner fall under the category of mobile health (mHealth) [4].

With approximately 200 new mHealth apps released every day, the number available now exceeds 300,000 [5]. One factor in this proliferation is the increasing use of mHealth technology by health service providers who not only seek advice from apps but also prescribe them to their patients [6]. In October 2020, Germany became the first country to cover the prescription costs of certain mHealth apps through statutory health insurance [7]. Moreover, because mHealth apps have the potential to replace a number of health provider tasks, it has been suggested that expertise in the use of mHealth-related technologies should be recognized as an essential competency for providers [1].

Dissemination of mHealth Apps to Users

The amount of academic research on mHealth apps has also increased, particularly in the areas of usability, effectiveness, adoption, and assessment. However, the highly important aspects of dissemination and marketing are as yet underexamined. App marketing refers to measures aimed at making a mobile app better known and acquiring users (ie, generating app downloads) and, moreover, contacting users and encouraging them to reach a specified goal [8].

An mHealth app is not a stand-alone product that can work effectively without human interaction, which cannot take place if users are unaware that a certain app exists and is accessible. Such “human touches,” although extraneous to the app itself, can be crucial for promoting use [9].

As users are key actors in mHealth adoption, it is critical to understand how they navigate the various stages from app discovery to frequent use. Google has created such a model [10]. It consists of four key stages: (1) discover (users come across an app and download it to their device), (2) onboard (the process of first use and registration), (3) engage (users start using the app regularly), and (4) embed (the desired outcome as users view the app as “vital” to their lives). Only a small proportion of users currently reach the embedment stage with any app [10]; for instance, the literature is sparse regarding the long-term integration and penetration of mobile interventions within mental health and other support service settings [11].

Increasing the chances of an app achieving embedment requires understanding users and placing them at the core of mHealth services. This process would start with persona definitions: fictional archetypes of actual product users. A persona enables program designers to create high-quality programs that effectively meet user needs [2]. In the marketing world, this also means segmenting users and locating them on the marketing funnel, which is a visual representation of the different phases in a customer’s journey toward conversion and their relationship with a product. By segmenting customers based on where they are located in the funnel, marketers target these groups much more effectively [12].

Positive customer experiences and journeys rely on ensuring that the consumer sees value in an app as a channel for accessing products and services and as a 2-way platform for seamless interactions. Although marketing strategies play a crucial role during the early stages of the customer journey, they have been subjected to very little analysis [13].

Marketing of Health Apps

A successful marketing strategy can ensure that an mHealth app reaches potential users and ease the adoption process. Such a strategy would clearly define target end users; determine the appropriate communication channels, content, and timelines to effectively reach users; and market the app as an attractive product, encouraging people to download it and become regular users. The strategy would include a mix of activities, depending upon the type of app and upon the stage of the launch period (from before the launch to after the launch), including email marketing, targeted advertising, and social media promotion [14].

Marketing services have evolved alongside information and communication technologies. In turn, digital marketing has provided a series of customized platforms for communicating with specific stakeholders using computers, smartphones, and tablet computers [15]. These channels enable information to be gathered and include websites as well as various social media platforms such as Facebook, YouTube, X (the platform formerly known as Twitter), Pinterest, TikTok, and LinkedIn. Traditional marketing also remains an option, with products being promoted on radio and television channels, as well as via printed posters in public spaces, flyers, and face-to-face conversations [16].

The cost of promoting an app will depend upon where money is spent, and those promotional activities that do not cost money will demand time. Although potential customers may be offered incentives such as money or supplemental products, the marketer or marketing firm involved in digital marketing can be offered a fee per click, download, or install. A small pilot trial of activities that cost money is recommended to assess results [14].

Specialized services are available for driving digital campaigns; for example, Facebook Ads Manager is a paid service that oversees paid digital marketing campaigns across the Facebook platform. Google Universal App Campaigns (UAC) is another paid service that promotes mobile apps by distributing marketing messages across Google formats and networks, such as the first page of applicable Google search results and small banner advertisements on relevant YouTube channels [17]. As Google

shares information among platforms, including Google Display Network, YouTube, and Google Play Store, Google UAC can capture the number of Google-driven impressions, clicks, and installs on Android devices.

The effectiveness of a marketing strategy can be observed not only through the number of app downloads but also by the effects of users interacting with the app. Referred to as mobile app engagement, this is defined by a variety of operational metrics, such as the number of log-ins, the number of days of use, the number of pages visited, and the number of tasks or modules completed [18]. Another gauge of marketing effectiveness is user onboarding. In the context of mobile apps, user onboarding is the process of providing instructions and highlighting key benefits and features via a set of example screens when the user first launches the app [19].

Objectives

The aim of this study was to review existing evidence on strategies and elements relevant to how the dissemination of mHealth apps is currently carried out and how these elements contribute to encourage user onboarding and engagement with mHealth apps.

Methods

Overview

This study was carried out following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [20] (Multimedia Appendix 1). The protocol of this review was registered with PROSPERO (CRD42022352369) [21].

Eligibility Criteria

Eligible sources were academic papers. All cohort studies, cross-sectional studies, and randomized controlled trials reporting on the dissemination and marketing of an mHealth app were included. The outcome expected from included studies was strategies and elements relevant to how the dissemination of mHealth apps is currently carried out and how these elements contribute to encourage user onboarding and engagement with mHealth apps.

Editorials, letters to the editor, scoping reviews, systematic reviews, meta-analyses, conference abstracts, commentaries, essays, book chapters, and study protocols were excluded, as were studies with any other study design (eg, bibliometric analysis, modeling study, systematic or web search or review of apps, landscape analysis, and scorecard analysis). We also excluded studies with participants aged <18 years and those not reporting the expected outcome. Search languages were limited to French and English.

Information Sources and Search Strategy

Searches were conducted on PubMed, PsycINFO, Scopus, and CINAHL databases over the 5-year period from January 1, 2018, to September 30, 2022 (refer to Multimedia Appendix 2 for the search strategy). The following search terms were used individually or combined according to Medical Subject Headings terms: “apps,” “mHealth,” “marketing,” “promotion,” and “dissemination.” Moreover, we conducted searches on JMIR and mHealth journals and cross-checked the reference lists of the selected studies to locate additional studies that met the inclusion criteria. The main outcome was dissemination strategies for mHealth apps.

Study Selection and Data Collection Process

All retrieved studies were imported into Rayyan (Rayyan Systems Inc) [22] and duplicate records eliminated. Screening consisted of blind peer review by 2 independent investigators. Any conflict was resolved through discussion or the adjudication of a third investigator.

Results

Selection of Studies

We identified 638 records through database searches (PubMed: n=215, 33.7%; PsycINFO: n=60, 9.4%; Scopus: n=283, 44.4%; and CINAHL: n=80, 12.5%) and 10 records through JMIR and mHealth journal searches. Of the total 648 records, 127 (19.6%) duplicates were removed. Next, of the remaining 521 articles, 502 (96.4%) were removed after title and abstract screening. The reasons for exclusion were unrelated outcome or outcome other than the subject of our review (309/502, 61.6%), study protocol (136/502, 27.1%), published review (48/502, 9.6%), study participants aged <18 years (6/502, 1.2%), and other study design (3/502, 0.6%). We then assessed the remaining 19 full texts for eligibility and excluded 9 (47%) for unrelated outcomes; thus, 10 (53%) papers were included in this review. A PRISMA-compliant flow diagram [20] of the paper selection process is shown in Figure 1. The characteristics of the studies selected are summarized in Table 1.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram of the paper selection procedure.

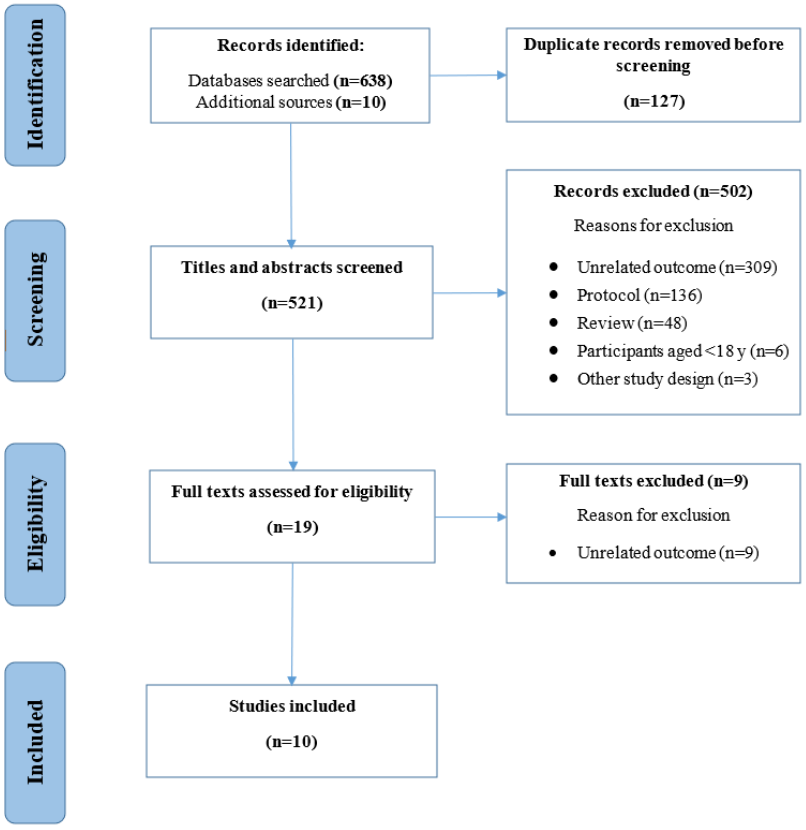


Table 1. Characteristics of the selected studies.

Authors, year; country	Population	Intervention	Outcomes	Design	Quality	Limitations
Kvedarienė et al [23], 2019; Lithuania	• N=149 • Sex: 55% female • Age: 18-60 (mean 37.2, SD 10.4) years • Other details: Lithuanians with allergic rhinitis and asthma	Monitoring of allergic rhinitis and asthma in real life in Lithuanian MASK - air app users	• High app engagement • User retention rate was 107 days of use/user • Patients are satisfied with the app overall	Longitudinal study	Low	• Selection bias: patients recruited from an allergy clinic
Buss et al [24], 2022; Australia	• N=46 • Sex: 50% female • Age: ≥45 (median 62, IQR 56-67) years • Other details: residing in Australia	Test the feasibility of an app-based intervention for cardiovascular and diabetes risk awareness and prevention	• The app scored the highest for the information section and the lowest for the engagement section of the scale • Nonuse and dropouts were too high, and adherence was too low to consider the intervention in its current form feasible • Asking people aged ≥45 years to download the app and expect them to use it over 3 months without additional interaction was not feasible	Nonrandomized controlled trial (cohort study)	Low	• Small sample size • Nonuse and dropouts too high, adherence too low
Arshanapally et al [17], 2022; United States	• N=NI[a] • Sex: NI • Age: 18-45 years • Other details: parents with children aged <5 years, speaking English or Spanish	Investigate the outcomes of a paid digital marketing campaign to promote an mHealth[b] app about parent-engaged developmental monitoring	• Paid digital marketing can be an effective strategy to promote mHealth apps targeting parents of young children • Google-driven marketing messages in English had a higher clickthrough rate than those in Spanish	Implementation study	Low	• No relevant limitations
Resnick et al [25], 2021; United States	• N=41 • Sex: 76% female • Age: ≥18 (mean 51, SD 12) years • Other details: non-Hispanic Black patients, speaking English	Assess the usability, acceptability, and user engagement of the Healthier Together mobile app	• The app strongly engaged participants, with promising results on participants' knowledge of cancer prevention behaviors and success in achieving their cancer prevention behavioral goals	Mixed methods intervention	Low	• Small sample size • Nonprobabilistic purposive sample of non-Hispanic Black patients at 2 internal medicine primary care clinics • Participants were incentivized US $40 to complete the baseline, in-person, 40-month enrollment process and interview and US $60 for completing the 45-month exit telephone interview

Authors, year; country	Population	Intervention	Outcomes	Design	Quality	Limitations
Zlotorzynska et al [26], 2021; United States	• N=NI • Sex: 0% female • Age: 18-24 years • Other details: YMSM[c]	Paid web-based recruitment campaign to recruit HIV-negative or unknown status YMSM for 4 randomized controlled trials of mHealth HIV prevention interventions	• Instagram advertisements yielded the highest proportions of eligible contacts who were racial or ethnic minority individuals and aged <18 years	Randomized controlled trials	Moderate	• Participants offered incentive to enroll in study
Rajani et al [27], 2021; United Kingdom	• N=154 • Sex: 38.8% female • Age: 18-65 years • Other details: smokers	Examine the association among perceived usefulness, perceived ease of use, and frequency of use of gamification features embedded in smoking cessation apps on self-efficacy and motivation to quit smoking	• The use of the apps was associated with increased self-efficacy and motivation-to-quit levels 4 weeks after app use compared with baseline • Perceived frequency of use of gamification features was associated with an increase in self-efficacy and motivation to quit • Higher baseline self-efficacy and motivation to quit were both associated with smaller increases in self-efficacy and motivation-to-quit levels 4 weeks after using the mobile apps compared with preapp use	Observational study	Low	• Participants incentivized
Roberts et al [28], 2019; United Kingdom	• N=32 • Sex: 68.8% female • Age: ≥18 (mean 60, SD 11; range 37-78) years • Other details: diagnosed with breast, prostate, or colorectal cancer	To seek opinions of survivors of breast, prostate, and colorectal cancer regarding using apps to promote PA[d]	• Multiple factors affect engagement with PA apps, and this is highly personalized • Apps that promote walking are the most appealing for survivors of cancer • PA apps should be integrated into cancer care	Cross-sectional study	Low	• Small sample • Participants were offered a £10 (US $12.7) voucher as an incentive for completion of study, and costs were reimbursed if asked to install an app that was not free to download
Bidargaddi et al [29], 2018; United States and Australia	• N=1255 • Sex: NI • Age: NI • Other details: NI	To study the effect of time-varying push notifications on engagement in self-monitoring activity	• Pushing a notification with a tailored health message affects near-time proximal engagement with the self-monitoring activity in the app	Microrandomized trial design	High	• No relevant limitations
Hui et al [30], 2018; United States	• N=101 • Sex: 87.1% female • Age: ≥16 years • Other details: patients with active asthma	The impact of different recruitment strategies and app features on adoption and continued use	• Adherence was dependent upon motivation derived from a sense that the health care professional and researcher were interested in the results and that using an app to support their self-management could improve their asthma control	Cohort study	Low	• Selection bias: participants recruited at clinics
		Market tests of the CycleBeads app in 7 countries		In-app microsurveys	Low	• No relevant limitations

Authors, year; country	Population	Intervention	Outcomes	Design	Quality	Limitations
Haile et al [31], 2018; Egypt, Ghana, India, and Jordan	• N=NI • Sex: 100% female • Age: ≥18 years • Other details: NI		• Social media campaigns proved to be an easy low-cost approach to advertising the CycleBeads app			

[a]NI: no information.

[b]mHealth: mobile health.

[c]YMSM: young men who have sex with men.

[d]PA: physical activity.

Of the 10 selected studies, 6 (60%) were observational studies (longitudinal, cohort, implementation study, mixed methods, and cross-sectional), 1 (10%) reported on 4 randomized controlled trials [26], 1 (10%) reported on in-app microsurveys [31], and 1 (10%) reported on a microrandomized trial [29].

According to the Scottish Intercollegiate Guidelines Network (SIGN) [32] criteria, 8 (80%) of the 10 studies were of low quality, 1 (10%) of moderate quality, and 1 (10%) of high quality. A low-quality rating resulted from a small sample size, the study design (mostly observational), or possible selection bias. A summary of the design, quality, and limitations of the included studies can be found in Table 1.

The selected 10 studies covered 8 countries—Lithuania, the United Kingdom, Australia, the United States, Egypt, Ghana, India, and Jordan—with 50% (5/10) of the studies conducted in the United States and 20% (2/10) in the United Kingdom.

The number of participants in the selected studies ranged from 32 to 1255. Of the 10 studies, 3 (30%) that recruited participants through social media and used impressions and clicks as a proxy measure of their number did not state the number of participants. In 6 (60%) of the 10 studies, 50% to 100% of the participants were female; sex information was not given in 2 (20%) of the 10 papers, 1 (10%) study targeted adolescent and young male individuals, and 1 (10%) targeted female individuals only.

Objectives of the Marketing Strategy

In each of the studies reviewed, we looked for the objective of the marketing strategy implemented through the lens of the Google model of mobile app user journey (discover, onboard, engage, and embed) [10] (Table 2). We found that 60% (6/10) of the studies aimed to drive people to discover their mHealth apps (come across the apps and download them), get onboard (start using the apps), and stay engaged by using the apps for a couple of weeks or months. Marketing strategy in 30% (3/10) of the studies targeted the discover and onboard stages, whereas 1 (10%) of the 10 studies aimed at the engage stage only.

Table 2. Summarized results of reviewed studies.

Authors, year	Channels	Procedures of each dissemination strategy: How was it done?	Effects and results of each dissemination strategy	Objective with regard to the Google model for users' journey with apps
Kvedarienė et al [23], 2019	Face-to-face	• Health personnel trained patients how to use the app at a clinic	• The duration of app use in patients ranged from 1 to 680 (median 54, IQR 23-151) days • There was an increased frequency when the reported days were >200 (18%) • Adherence to mobile apps was higher when the app was promoted by physicians and when the users were taught how to use it	Discover, onboard, and engage
Buss et al [24], 2022	Email	• Participants received an email and a user guide that included instructions to download an app from the app store on their mobile phones and then use it for 3 months • Researchers encouraged regular use • If they had questions or technical issues, participants could get in touch with researchers via email	• Of the 46 participants, 20 (43%) never used the app, and 15 (33%) dropped out • Adherence to app use (using the app at least once/week over 3 months) was 17% (8/46) • The median time between the first and last app use was 54 days • The research team did not actively engage with participants early in the study to verify that all participants could install the app • The intervention did not involve direct contact with health care professionals • The app did not contain enough interactive features	Discover, onboard, and engage
Arshanapally et al [17], 2022	Social media	• Google- or Facebook-driven marketing message: when a user clicked on a marketing message, they were directed to either the Google Play Store or the Apple App Store, depending upon device type • Google Universal App Campaigns distributed marketing messages across several Google formats and networks • Facebook Ads Manager distributed marketing messages across the Facebook platform	• The Google-driven marketing messages garnered a total of 4,879,722 impressions and 73,956 clicks (clickthrough rate: 1.52%); from these clicks, there were 13,707 installs of the app on Google Play Store (18.53% install rate); the overall cost/install was US $0.93 • The Facebook-driven marketing messages garnered a total of 2,434,320 impressions and 44,698 clicks (clickthrough rate: 1.84%); the average cost/install for Facebook could not be calculated because of the limitation of collecting Facebook-driven installs data • Animated graphics generated the greatest number of clicks among both English and Spanish audiences on Facebook when compared with other types of images	Discover, onboard, and engage
Resnick et al [25], 2021	Face-to-face	• In-person information about the app, followed by installation of the mobile app on the participant's mobile phone and instructions on how to select a goal, choose share settings, and invite other social ties • Weekly SMS text messages reminding participants to check in, share goal progress, and invite relatives	• Of the 171 participants, 41 (24%) downloaded the app • Engagement with the app (mean number of check-ins/participant out of 8 possible weekly check-ins during the study period) was 5.7 • Of the 41 participants, 31 (76%) checked in during at least 4 of the 8 weeks • The mean System Usability Scale score was 87 (SD 12; median 90, IQR 78-95)	Discover, onboard, and engage

Authors, year	Channels	Procedures of each dissemination strategy: How was it done?	Effects and results of each dissemination strategy	Objective with regard to the Google model for users' journey with apps
Zlotorzynska et al [26], 2021	Advertisements on Grindr, Snapchat, Instagram, and Facebook	• Advertising was purchased on Facebook, Instagram, Snapchat, Twitter, and Grindr; users who clicked on banner advertisements were taken directly to a study-specific eligibility screener and, if eligible, were asked to provide contact information for follow-up by respective study site staff • The Facebook Ads Manager proprietary algorithm allocated the distribution of advertisement placements (news feed, right-hand column, and Instagram feed and stories) that would provide the best performance • Advertising copy was written to give a very brief description of the study, identify the organization conducting study recruitment, mention the study incentive, and provide a call to action for potential participants seeking to learn more • Snapchat advertisements were created using Snap Publisher and were in the form of short videos up to 10 seconds long; these videos rotated through various stock photo images and superimposed text • Instagram placements were used to test advertisement performance • Images used in advertisements included stock photos, which were either purchased from web-based stock photo vendors (eg, Shutterstock) or accessed through Facebook's free stock photo catalog integrated within the Facebook Ads Manager • Recruitment through various in-person and community outreach efforts included posting flyers, tabling at community events, reaching out to past study participants, and recruiting through clinics serving youth	• Grindr and Snapchat advertisements produced the highest clickthrough rate compared with Facebook advertisements; these advertisements had the lowest proportions of users who initiated eligibility screeners • Facebook advertisements yielded the lowest cost/eligible contact, whereas advertisements on Twitter had the highest • Facebook or Instagram advertisements had much higher rates of screening and ultimately yielded much lower costs/eligible participant • The cost/eligible participant was markedly higher for the Instagram advertisements campaigns	Discover and onboard
Rajani et al [27], 2021	Social media and paper posters	• Participants were recruited via social media, and posters were displayed in public places in London, United Kingdom • Participants were provided instructions on the internet on how to download and start using the app	• Of the 138 participants who installed the app, 116 (84.1%) completed all 4 weeks of the study • A 1-point increase in the average perceived frequency of gamification features was statistically significantly associated with a 3.35-point increase in self-efficacy from baseline to study end (β=3.35, 95% CI 0.31-6.40)	Discover and onboard
Roberts et al [28], 2019	Face-to-face, paper posters, email, and social media			Discover, onboard, and engage

Authors, year	Channels	Procedures of each dissemination strategy: How was it done?	Effects and results of each dissemination strategy	Objective with regard to the Google model for users' journey with apps
		• Participants recruited via advertisements within community-based cancer support groups (either by verbal descriptions from group leaders at meetings or via posters, flyers, and email mailing lists), Facebook cancer support groups, and charitable organizations • Participant randomly assigned to download 2 of 4 apps (Human, The Walk, The Johnson & Johnson Official 7 Minute Workout, and Gorilla Workout) • Guidance in downloading and installing each app was provided, if required; participants were asked to spend approximately 2 consecutive weeks using the apps	• Of the 40 participants, 32 (80%) completed the study (dropout rate: 20%) • Reasons for dropping out were lack of time, family circumstances (eg, bereavement), and not wanting to update their smartphone's operating system or register credit card details with Google Play • Factors affecting engagement included participants' perceptions of the advantages and disadvantages of using apps to support physical activity, the relevance of the app to the user, the quality of the app, and the behavior change techniques used to promote physical activity	
Bidargaddi et al [29], 2018	Push notifications	• Push notifications were sent at 1 of 6 chosen time points throughout the day, and a user could either receive or not receive a push notification at a chosen time point • At each considered time point, users were randomized to either receive or not receive a push notification containing a tailored health message with a 50% probability; once a time point was considered, the user was then considered unavailable for the remainder of the day • To mitigate the risk of users either turning off notifications or deleting the app owing to receiving too many push notifications, users were classified as either "available" or "unavailable" at each time point, and only those time points when users were "available" were considered for the push notification decision; several rules were applied to determine availability	• Sending a push notification containing a tailored health message was associated with greater engagement in a mobile health app • The effect of the pushed notifications is sustained over time; push notifications containing tailored health messages can attenuate the rate at which users disengage • Users who used the app less frequently were "unavailable" to receive push notifications on a greater number of days than those who used the app more often	Engage
Hui et al [30], 2018	Social media and face-to-face	• Researchers sent letters inviting people to use the app for up to 3 months • Practice nurses approached adults with active asthma to try out a prototype app • The app was also advertised via social media (Facebook and Twitter [now known as X]) • Nurses were not expected to teach patients to download and use the app • Technological support was provided by the research team and the app developer • The messages on social media included brief information and a link to information for patients on how to download the app	• A total of 300 patients received information on the app and study from Facebook, Twitter, or organic searches • Only 135 patients downloaded the app, with 111 (82.2%) registering an account on the app • Social media attracted 87 users, but only 15 (17%) used the app for 30 days • A total of 24 patients were recruited, and 13 (54%) continued for 30 days • Successful adoption was dependent upon the ease of downloading and sufficient motivation • Some patients needed technological assistance with downloading the app and starting to use the features	Discover, onboard, and engage
Haile et al [31], 2018	Social media and face-to-face	Series of culturally appropriate Facebook campaigns for each country of interest		Discover and onboard

Authors, year	Channels	Procedures of each dissemination strategy: How was it done?	Effects and results of each dissemination strategy	Objective with regard to the Google model for users' journey with apps
			• Within 10 months after advertising, there were 356,520 app downloads; the cost/download (paid by the advertiser) was lowest in Nigeria at US $0.17, followed by Egypt (US $0.26), Ghana (US $0.27), India (US $0.30), Jordan (US $0.36), Kenya (US $0.41), and Rwanda (US $0.69) • Most of the app users were aged 20-29 years and married or in exclusive relationships	

Channels and Procedures of Each mHealth App Dissemination Strategy

As shown in Table 2, mHealth app promotion was conducted through various channels, including paper posters, face-to-face communications, email, and social media. The most reported channels were face-to-face [23,25,28,30,31] and social media [17,27,28,30,31], both of which appeared in 50% (5/10) of the studies. Messages were intended to inform people about the existence of the app, what it does and the benefits of using it, where to find the downloading link, downloading instructions, and information on how to use the app. Messages through these channels used text, voice-overs, and short videos. Messages on social media and email included a shortened URL to download the app. Face-to-face interactions [23,25,28,30,31] provided the opportunity for training support, including demonstrating app features, installing and using the app, and solving technical issues.

Google Play Store and Apple App Store were the main web repositories to which potential users were redirected for download. In addition, the social media pages of various organizations, such as the Facebook pages of cancer support groups [28], were also used to promote apps and provide download links.

Effects of Dissemination Strategies

Advertising performance was influenced by many factors, including but not limited to advertising content and creativity, competing advertisements within platforms, and emerging platforms that attracted users to new digital spaces [26]. The reviewed papers revealed that social media advertisements attracted substantial downloads over a short period with a large number of impressions [13] and prompted patients to respond to invitations to use the app. However, without the motivation provided by a trusted professional, users quickly ceased their use of the app; a dual promotion strategy was then used to increase app adoption: using social media for its reach and ease of response as well as promoting personal invitations from trusted health care professionals [25].

In 1 (10%) of the 10 studies, embedded push notifications were sent to users to keep them engaged with the app [24]. This effect was sustained over time, and push notifications containing tailored health messages reduced the rate at which users disengaged. The timing of notifications was a key factor, with the results suggesting that users were more likely to engage with an app within 24 hours when push notifications were sent at midday on a weekend. Animated graphics generated a greater number of clicks than other image types [17].

The success of mHealth app dissemination strategies in these studies was measured in terms of adherence, app use, the duration of app use, impressions, clicks, app downloads, and installs. Marketing messages with simple direct calls to action, such as "Track your child's development," generated high clickthrough rates (CTRs) [17].

Marketing Messages and Content of Communications

In some cases, promotional messages included questions prompting people to download and test the app, such as "Would you like to test an app that helps you manage your asthma better? [shortened URL] Need a helping hand to manage your asthma? Try our new app [shortened URL]" [30]. One study included the inquiring headline "Ever think of testing at home?" and explanatory body text: "Fenway Health is looking for young men to help test out a mobile app to support sexual health" [26].

In a study promoting an mHealth app about parent-engaged developmental monitoring [17], marketing messages with animated graphic interchange format (GIF) images as well as images of younger children performed higher than messages without GIF images and images of older children.

Paid or Nonpaid Marketing Campaigns

The paid marketing campaign platforms that were used included Facebook Ads Manager and Google UAC; in addition, advertisements were placed on Grindr, Snapchat, and Instagram [17,26,31]. In 50% (5/10) of the studies, participants were offered incentives to participate [24-28]. Of these 5 studies, 4 (80%) were conducted in the United States and 1 (20%) in Australia. No study reported that paid marketing resulted in higher app adoption than unpaid campaigns.

Key Indicators of mHealth App Marketing Strategies

Various metrics were used in the reviewed studies to assess mHealth app marketing strategies. Key indicators included nonuse rate, dropout rate, and adherence rate, as well as median time between first and last app use, which was generally measured in days or weeks. Nonuse rate was defined as the proportion of participants who never used the app [24], and the duration of use was assessed by determining the reported number

of days of use [23]. The longest duration of app use was 680 days [23].

Dropout rate was defined as the proportion of participants who completely stopped using the app at least 14 days before they received an invitation to complete the end-of-study survey. Adherence rate was defined as the proportion of participants who used the app at least once every week over 3 months of the study [24].

Specific indicators for paid marketing strategies were cost per click (CPC), defined as the amount of money spent per click secured in each advertisement campaign or advertisement set; CTR, the number of clicks divided by impressions; and cost per impression [26]. Impression is the number of times an advertisement is viewed by a user on an advertising platform. For advertisements seen multiple times by the same user, each view is counted as an impression. Additional indicators were the number and percentage of clicks that took people through to the eligibility screener and of those who were eligible for the study [26].

People Implementing the Marketing Processes

In the studies reviewed, we found that promotional messages were mainly disseminated by researchers, trained nurses, and other health practitioners. Technological support was only provided by the research team and app developers. In some cases, recently enrolled patients contributed to promoting the app by sending download invitations to their family and friends [25].

Marketing Concepts Reported in the Reviewed Studies

Inspired by a book on how to successfully distribute apps [12], we screened the included studies to identify the marketing concepts that were used to promote mHealth apps in these studies (Table 3).

Table 3. Marketing concepts [12] reported in the reviewed studies.

Authors, year	Incentivization[a]	Personalization[b]	Mobile app attribution[c]	Loyalty marketing[d]	Remarketing or retargeting[e]	A/B testing[f]	Programmatic marketing[g]	Predictive marketing[h]	Thought-leadership marketing[i]	Content marketing[j]	Behavioral marketing[k]
Kvedarienė et al [23], 2019[l]		✓									✓
Buss et al [24], 2022[l]	✓	✓									
Arshanapally et al [17], 2022[l]			✓			✓					✓
Resnick et al [25], 2021[l]	✓	✓	✓								✓
Zlotorzynska et al [26], 2021[l]	✓	✓	✓			✓				✓	
Rajani et al [27], 2021[m]	✓		✓								
Roberts et al [28], 2019[m]	✓									✓	
Bidargaddi et al [29], 2018[l]		✓	✓	✓				✓			✓
Hui et al [30], 2018[m]			✓						✓	✓	
Haile et al [31], 2018[m]			✓							✓	

[a]The incentivized model is the strategy of making a product, program, or other offering more attractive to customers by offering an incentive in exchange for buying or participating. In the app business, incentivization is normally used to quickly amass app installs. Of the 10 studies, half of the studies (n=5, 50%) used incentivization.

[b]Personalization involves customizing the timing and content of marketing messages to the target user based on their preferences, habits, and behavior patterns. Personalized messages refer to every user by name and entice the user to become more engaged with an app with the right kind of incentive based on their characteristics, such as age, gender, location, profession, and financial segment. Half of the studies (5/10, 50%) used personalization.

[c]Mobile app attribution is the process of recording and measuring the actions of app users, such as installs, level completions, in-app purchases, and other milestones. Mobile app attribution is essential to app marketing because it helps produce the data that are gathered and analyzed to measure how well marketing campaigns are working. The majority of the studies (7/10, 70%) used mobile app attribution.

[d]Loyalty marketing is a marketing strategy that focuses on nurturing existing customers rather than acquiring new ones. Only 1 (10%) of the 10 studies used loyalty marketing.

[e]Remarketing or retargeting targets every individual who has come into contact with the product but has not converted or who converted but later abandoned the app. It allows marketers to reconnect with these categories of users and "bring them back" or increase the time they spend engaging with the app. None of the studies used remarketing or retargeting.

[f]A/B testing involves the use of several versions of the same advertisement distributed to different groups with different designs, color coding, calls to action, and message content to determine which version produces the highest conversion rate. Of the 10 studies, only 2 (20%) used A/B testing.

[g]Programmatic marketing is the automated algorithm-based real-time buying and selling of advertising space through a bidding system, with the aim of reaching the right customers at the right time. None of the studies used programmatic marketing.

[h]Predictive marketing involves using data science based on customer behavior and habits to make smarter marketing decisions. By gathering and analyzing data about user behavior and identifying patterns, marketers can make forecasts about user behavior and make informed decisions about the likelihood of the success of their marketing content and offerings. Of the 10 studies, only 1 (10%) used predictive marketing.

[i]Thought-leadership marketing is the process of positioning a company as a leader in a specific domain by supplying customers with top-quality information. Only 1 (10%) of the 10 studies used thought-leadership marketing.

[j]Content marketing is a marketing strategy that involves producing content that potential customers find useful, valuable, and relevant. Content marketing is highly effective at building a loyal user base and converting leads into customers. More than one-third of the studies (4/10, 40%) used content marketing.

[k]Behavioral marketing involves segmenting the app's user base based on user behavior with the aim of refining the marketing strategy and more effectively targeting users. More than one-third of the studies (4/10, 40%) used behavioral marketing.

[l]These studies (n=6, 60%) used single-channel marketing, which involves reaching users through a single channel, eg, Facebook advertisements.

[m]These studies (n=4, 40%) used multichannel marketing, which involves >1 channel, as opposed to an omni-channel marketing campaign, which attempts to reach users through all available channels.

In summary, most of the studies (7/10, 70%) reported using a combination of marketing concepts to advertise their mHealth apps. The most used concept was mobile app attribution (7/10, 70%). Mobile app attribution is essential to app marketing because it helps produce the data that are gathered and analyzed to measure how well marketing campaigns are working [12]. Furthermore, 40% (4/10) of the studies reported using at least 2 channels to market their mHealth apps.

Discussion

Principal Findings

We found that the marketing strategies used in almost all included studies (9/10, 90%) were aiming to drive at least app discovery and onboarding, with more than half (6/10, 60%) also targeting user engagement to mHealth apps. Social media, emails, television or radio, posters or flyers, and face-to-face communications were all used in the reviewed studies to inform people about the existence of mHealth apps, invite them to download, encourage them to use the apps, and maintain engagement. This is consistent with an integrative review of methods used to promote mobile apps, which also cited app store optimization via keywords and the inclusion of screenshots and videos for greater conversion rate, the use of push notifications, the promotion of apps via influencers, and the leveraging of user review and ratings [13].

Social media attracted many downloads over a short period, whereas emails were most often used for sharing instructions on how to download apps and interact with research teams for technical support.

The strategies used to promote mHealth apps included paid and unpaid marketing, and metrics such as CPC and CTR were used to measure effectiveness. Offering incentives to people to download and use the app did increase app downloads and use. However, it has been shown that people are less likely to keep using an app after incentivization, although the monetary value of the incentive could have a significant effect on the adherence [3]. Furthermore, a focused strategy is required to maintain a low app churn rate [12].

Factors Influencing Engagement, Onboarding, or Adherence to mHealth Apps

Reasons for User Engagement, Onboarding, or Adherence to the App

Onboarding was often reliant on the ease of downloading and sufficient motivation. In the study by Hui et al [30], adherence stemmed from awareness that a health care professional and a researcher were interested in the results and that using an app to support self-management could improve a participant's control of their health condition. Additional factors in engagement included the perceptions of the advantages and disadvantages of using apps to support self-management specific to an individual's health needs, the relevance of the app to the user, the quality of the app, and the behavior change techniques used to promote health [28]. All these factors are among those reported by a literature review that identified retention factors related to apps, such as feedback, appropriate reminders, and in-app support from peers or coaches [33]. Our findings also align with those of another systematic review that listed individualized reminders, user friendliness and technical stability of the apps, and personal support from health care professionals as intervention-related factors influencing adherence [3].

Push notifications and weekly SMS text messages inviting check-in were also used to influence user engagement [26], herein defined by number of check-ins. This is also consistent with the findings from a systematic review [34] that assessed 15 commercial apps for diabetes prevention and found that the app that included the notification features for activity tips, goal progress tips, goals adjustment, and completed goals had the highest engagement mean score (4.5 points out of 5). However, we argue that push notifications should not be a one-size-fits-all solution because marketing research suggests that among app users, just 50% accept push notifications from their favorite app, and 30% disable all push notifications [35].

In our review, user engagement was also influenced by the health goals that participants selected. This finding also aligns with that of a previous systematic review that further suggested that users could disengage at any time and re-engage at a later stage when needed. Thus, this feature might be particularly useful for addiction research targeting relapse prevention strategies [36].

Barriers to User Engagement, Onboarding, or Adherence to the App

From the studies included in this review, reasons for nonadoption included problems in app installation [24,30]; the use of other health apps that better suited participants' needs and preferences; and other concerns, such as prioritizing COVID-19 over the condition addressed by the app [28]. Indeed, marketing can be influenced by competing health information targeting the same users. When people were concerned about contracting COVID-19 and seeking a pandemic-related app, they were less motivated to use an app being promoted to tackle other health issues [17,24]. This could lead to competition among mHealth apps for potential users or health care providers.

Other potential barriers to adoption included a lack of early active engagement with participants to verify that they could install the app, no direct contact with health care professionals,

and not enough interactive features. Indeed, as reported in another systematic review, being less informative and less interactive can lead to a very low engagement mean score [34].

The study by Roberts et al [28] reported that technical issues and concerns about data security reduced engagement. Similar concerns about data security with mHealth apps had been previously pointed out in the literature [37,38].

Return on Investment in Marketing

Although dependent upon the budget available, a decision threshold for the cost of marketing campaigns for mHealth apps to reach a certain level of engagement would be beneficial. According to a study on industry-specific Google benchmarks, for the health care sector overall, these costs amount to a CTR of 3.27% and a CPC of US $2.62 for Google Search and a CTR of 0.59% and a CPC of US $0.63 for the Google Display Network [39]. A similar study on industry-specific Facebook benchmarks produced a CTR of 0.83% and a CPC of US $1.32 for health care [40]. However, it is important to note that comparing CTR and CPC with industry-specific benchmarks for the entire health and medical field should not be the sole method of evaluating effectiveness because these benchmarks may be too broad [17].

People Involved in Marketing mHealth Apps

The most productive marketing team is a multitude of satisfied users championing the app on social media, encouraging their friends and colleagues to download it, and giving it 5-star reviews. Creating a strong user support system and feedback loop, regularly updating the app based on user feedback, and doing whatever it takes to keep users happy are the most important marketing tactics that can be deployed [12].

We found that in studies that investigated age and sex differences in engagement with apps, the age of research participants did not predict app engagement [24,25]. However, there were statistically significant differences in sex and app use, with more male participants using the apps in question than female participants, but not in the duration of app use [24,30]. This result is contradictory with that of another study that found that female sex positively influenced adherence [3]. Therefore, we could not draw conclusions on the effect of sex and age on app adherence and thus leave it to further research.

Marketers of mHealth apps should always consider the motivations of the app audience; for example, the study by Roberts et al [28] stated that apps promoting walking can be appealing to survivors of cancer. Similar findings were highlighted elsewhere [3]. In addition, consideration must be given to the timing of interventions intended to maintain engagement. Data should be collected to predict the moments that users will be available and receptive to in-app notifications.

Limitations

We have noted confusion surrounding the term *user engagement*. Engagement with digital health interventions and engagement with mHealth apps are not clearly separated. This study focuses on the latter, defined as a set of actions by a user within an mHealth app [18]. This differs from user engagement with digital health interventions, which is conceptualized in terms of both experience and behavior [41]. This confusion may explain why many of the studies identified during the database searches had to be excluded: they addressed outcomes related to changes in health behavior; for example, most of the excluded papers reported levels of user engagement with a health intervention, such as increased physical activity, but not engagement with an mHealth app. However, we recognize that these definitions are interlinked: research has shown that app engagement can motivate behavior change [25]. We have noted the same confusion with the term *user embedment*. In the 10 reviewed studies, only 1 (10%) referred to embedment as integrating a functionality within the app.

Most of the studies we reviewed (9/10, 90%) were conducted in high-income countries, with half being carried out in the United States (5/10, 50%). This may limit generalizability in low- and middle-income countries. Moreover, we only searched for papers written in French and English. We also note that the filters we applied with our search terms combination to avoid noise could have excluded some potentially useful papers. This could explain why the reviewed studies mostly reported research conducted in the United States and Europe (8/10, 80%).

Our findings may also be subject to observer bias [42] because in every reviewed study the research team members were involved in the diffusion processes. In some of the studies (2/10, 20%), participants received in-person physical assistance with app installation. This would be impossible for users in many settings. The generalizability of these findings is also limited by the fact that none of the included studies covered the dissemination of mHealth apps among health care personnel.

Implications and Future Research

Integration of mHealth Apps Into Routine Clinical Practice

None of the studies we reviewed aimed to address the embedment of mHealth apps in routine practice as part of their marketing strategy. To tackle the issue of the embedment of mHealth apps, researchers have proposed a framework for prescribing apps and outlined the key issues that need to be addressed to enable app dissemination in clinical care. This includes education and awareness, the creation of digital formularies, workflow and electronic health record integration, payment models, and patient or provider support [43]. As suggested by this framework, a starting point for the integration of mHealth apps into routine clinical practice would be education and awareness, meaning the promotion of mHealth apps, the aim of which would be to create a base of users downloading the apps because the number of app downloads and interactions over time also provides an indication of sustained uptake over time [11].

At this critical point of creating a user base, inspired by the synthesis of our findings, we offer a set of recommended uses of different channels (Table 4). This would first be applicable during the launching phase of the app and to lead users through their app onboarding stage. The use of these channels could change depending upon the objective of the marketing strategy.

Table 4. Recommended uses of channels to promote mobile health apps.

Channel and recommended use	Targeted people	Supportive marketing concepts
Email		
Inform about the existence of the app, link for download, functionalities, app release notes and app updates, and general information regarding the app (developer, brand owner, and sponsor)	High-level users (influencers and decision-makers) and target base users (health care providers and day-to-day users of the app)	A/B testing
Give instructions on how to download and install and use the app	High-level users (influencers and decision-makers) and target base users (health care providers and day-to-day users of the app)	A/B testing
Provide technical support and answers to users' questions; share user guide and tips	Effective users	A/B testing
Social media		
Inform about the existence of the app, and share the link for download	Potential target base users (health care providers and day-to-day users of the app)	A/B testing, incentivization, loyalty marketing, thought-leadership marketing, and content marketing
Engage in direct interactions to provide technical support and answers to users' questions	Potential target base users (health care providers and day-to-day users of the app)	A/B testing, incentivization, loyalty marketing, thought-leadership marketing, and content marketing
Television		
Short promotional video report on the app and its functionalities, as well as 1- to 3-minute video spots with speech by high-level users (influencers and decision makers) recommending the apps	Potential target base users (health care users and day-to-day users of the app)	Incentivization and thought-leadership marketing
Posters or flyers		
Infographics and key text message to inform about the existence of the app and its main value; include a QR code and text to indicate link to download	Potential target base users (health care providers and day-to-day users of the app)	Predictive marketing and incentivization
Face-to-face interaction: in-person training or meeting		
Inform about the existence of the app, link for download, functionalities, app release notes and app updates, and general information regarding the app (developer, brand owner, and sponsor)	High-level users (influencers and decision makers) and potential target base users (health care providers and day-to-day users of the app)	A/B testing
Give instructions on how to download, install, or use the app; share and explain user guide and discuss tips	High-level users (influencers and decision makers) and potential target base users (health care providers and day-to-day users of the app)	A/B testing
Face-to-face interaction: in-person ad hoc (unplanned) encounter		
Engage in direct interactions to provide technical support and answers to users' questions	Health care providers and day-to-day users of the app	Remarketing or retargeting and personalization
Sell the app (highlight its main value) and manage to install it on users' devices; explain how to use it and discuss tips	Health care providers and day-to-day users of the app	Remarketing or retargeting and personalization

Social media could be a beneficial entry point for motivating people to download an app, and human interaction is key during the engagement phase. Therefore, mHealth app promoters should provide users with training and support to start and continue using the apps. This can be done by maintaining communication through social media, including app-dedicated pages. At this point, content marketing—producing content that potential customers find useful—is valuable. It has been demonstrated that content marketing is highly effective in building a loyal user base and converting leads into customers [12].

It is important to note that although social media marketing also tends to attract people who are not the intended audience, communication through email requires a list of targeted email addresses. This entails contacting people directly and requesting their addresses or interacting with someone who will reveal potential users' email addresses or share an app link with potential users; for instance, a hospital director may share information about an app with hospital staff or share an attendance list containing the email addresses of hospital staff.

Future public health campaigns targeting the parents of young children should consider crafting marketing messages for social media campaigns with animated GIF images as well as images of young children.

As time-varying push notifications have been shown to contribute to mHealth app user engagement, developers should

interact with health care providers to implement this strategy. One approach to this is to apply mobile app attribution: the process of recording and measuring the actions of app users, such as installs, level completions, and in-app purchases [12].

Future Research

Finally, future research could be dedicated to developing a framework on how to disseminate mHealth apps. Such a framework, in addition to various marketing concepts presented in this review, should take into account additional considerations that are specific to mHealth apps, such as data confidentiality and privacy, and segment users on the marketing funnel [12] based on the best available evidence on engaging users with mHealth apps. One issue impeding the dissemination of apps that emerged in our study was the existence of competing apps. Some researchers have suggested that digital formularies or app libraries could help to address this. Digital formularies provide a short list of available apps, and providers could search these formularies and know what is available for a specific diagnosis or purpose [43]. Further research could explore and expand on the effectiveness of digital formularies as a dissemination channel for mHealth apps and the enablers of embedment of mHealth apps into routine practice. Finally, further research could aim to address the gap in identifying specific marketing strategies that would effectively drive the embedment of mHealth apps into routine practice.

Conclusions

The dissemination of mHealth apps takes place via face-to-face interactions, email, and printed posters and social media channels with diverse results. The effects of these strategies are reflected in download numbers and user engagement with mHealth apps. The results of this study will serve to guide future research and guide the marketing of mHealth apps for their routine use within the health sector.

The development of a framework for health care designers to promote their apps within health systems would be immensely beneficial. Such a framework would help systematize the dissemination of mHealth apps and guide the impact assessment of the dissemination strategies.

Acknowledgments

The study received funding from the Universitat Oberta de Catalunya, Barcelona, Spain, for proofreading and article processing fees.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOC File , 74 KB - mhealth_v12i1e50293_app1.doc]

Multimedia Appendix 2
Search strategy.
[DOC File , 57 KB - mhealth_v12i1e50293_app2.doc]

References

1. Wattanapisit A, Teo CH, Wattanapisit S, Teoh E, Woo WJ, Ng CJ. Can mobile health apps replace GPs? A scoping review of comparisons between mobile apps and GP tasks. BMC Med Inform Decis Mak 2020 Jan 06;20(1):5 [FREE Full text] [doi: 10.1186/s12911-019-1016-4] [Medline: 31906985]
2. Be he@lthy, be mobile: personas toolkit. World Health Organization. 2019. URL: https://apps.who.int/iris/handle/10665/329947 [accessed 2023-04-22]
3. Jakob R, Harperink S, Rudolf AM, Fleisch E, Haug S, Mair JL, et al. Factors influencing adherence to mHealth apps for prevention or management of noncommunicable diseases: systematic review. J Med Internet Res 2022 May 25;24(5):e35371 [FREE Full text] [doi: 10.2196/35371] [Medline: 35612886]
4. What is mHealth, what does it do and how does it benefit countries? International Telecommunication Union. URL: https://www.itu.int:443/en/ITU-D/ICT-Applications/eHEALTH/Be_healthy/Pages/guide-01.aspx [accessed 2022-12-26]
5. The growing value of digital health - evidence and impact on human health and the healthcare system. IQVIA Institute. 2017 Nov 07. URL: https://www.iqvia.com/-/media/iqvia/pdfs/institute-reports/the-growing-value-of-digital-health.pdf [accessed 2023-10-11]
6. Berkowitz CM, Zullig LL, Koontz BF, Smith SK. Prescribing an app? Oncology providers' views on mobile health apps for cancer care. JCO Clin Cancer Inform 2017 Nov 13;1(10116):1-7 [FREE Full text] [doi: 10.1200/CCI.17.00107] [Medline: 30657404]
7. Dahlhausen F, Zinner M, Bieske L, Ehlers JP, Boehme P, Fehring L. Physicians' attitudes toward prescribable mHealth apps and implications for adoption in Germany: mixed methods study. JMIR Mhealth Uhealth 2021 Nov 23;9(11):e33012 [FREE Full text] [doi: 10.2196/33012] [Medline: 34817385]

8. Wohllebe A, Hillmers M. Towards a scientific definition of app marketing – a practice-oriented approach using scientific and grey literature. Int J Appl Res Bus Manag 2021;2(1):13-25. [doi: 10.51137/ijarbm.2021.2.1.2]
9. Evans SH, Clarke P, Glovinsky JA. A consumer health app works well. . . but, diffusion is not so easy. J Technol Behav Sci 2019 Dec 26;5:164-170. [doi: 10.1007/s41347-019-00124-8]
10. Retail apps: accelerated adoption. Google. URL: https://www.thinkwithgoogle.com/_qs/documents/11700/Google_MTM_Retail_Apps_Accelerated_Adoption_WP_2021.pdf [accessed 2022-01-25]
11. Bear HA, Ayala Nunes L, DeJesus J, Liverpool S, Moltrecht B, Neelakantan L, et al. Determination of markers of successful implementation of mental health apps for young people: systematic review. J Med Internet Res 2022 Nov 09;24(11):e40347 [FREE Full text] [doi: 10.2196/40347] [Medline: 36350704]
12. Panosian H. Learn iOS Application Distribution: Successfully Distribute Apps. New York, NY: Apress; 2017.
13. Stocchi L, Pourazad N, Michaelidou N, Tanusondjaja A, Harrigan P. Marketing research on mobile apps: past, present and future. J Acad Mark Sci 2022;50(2):195-225 [FREE Full text] [doi: 10.1007/s11747-021-00815-w] [Medline: 34776554]
14. Android apps marketing: secrets to selling your Android app. O'Reilly. URL: https://learning.oreilly.com/library/view/android-apps-marketing/9780132378291/ch21.html [accessed 2023-04-23]
15. Pourkarim M, Nayebzadeh S, Alavian SM, Hataminasab SH. Digital marketing: a unique multidisciplinary approach towards the elimination of viral hepatitis. Pathogens 2022 May 29;11(6):626 [FREE Full text] [doi: 10.3390/pathogens11060626] [Medline: 35745480]
16. Kalmegh GS. Digital marketing vs. traditional marketing- a comparative view. Int J Sci Eng Res 2022;10(5):35-38 [FREE Full text]
17. Arshanapally S, Green K, Slaughter K, Muller R, Wheaton D. Use of a paid digital marketing campaign to promote a mobile health app to encourage parent-engaged developmental monitoring: implementation study. JMIR Pediatr Parent 2022 Apr 05;5(2):e34425 [FREE Full text] [doi: 10.2196/34425] [Medline: 35380545]
18. Sieverink F, Kelders SM, van Gemert-Pijnen JE. Clarifying the concept of adherence to eHealth technology: systematic review on when usage becomes adherence. J Med Internet Res 2017 Dec 06;19(12):e402 [FREE Full text] [doi: 10.2196/jmir.8578] [Medline: 29212630]
19. Fernandes T, Vasconcelos A. Mobile marketing in health: user experience guiding the implementation of a medical booking application. In: Proceedings of the 13th International Joint Conference, BIOSTEC 2020. 2020 Presented at: 13th International Joint Conference, BIOSTEC 2020; February 24-26, 2020; Valletta, Malta. [doi: 10.1007/978-3-030-72379-8_27]
20. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
21. Moungui HC, Ribas CC, Nana-Djeunga HC. Dissemination of mhealth applications: a systematic review. PROSPERO. URL: https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=352369 [accessed 2023-05-29]
22. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016 Dec 05;5(1):210 [FREE Full text] [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
23. Kvedarienė V, Biliute G, Didziokaitė G, Kavaliukaite L, Savonyte A, Rudzikaite-Fergize G, et al. Mobile health app for monitoring allergic rhinitis and asthma in real life in Lithuanian MASK-air users. Clin Transl Allergy 2022 Sep;12(9):e12192 [FREE Full text] [doi: 10.1002/clt2.12192] [Medline: 36178186]
24. Buss VH, Varnfield M, Harris M, Barr M. Remotely conducted app-based intervention for cardiovascular disease and diabetes risk awareness and prevention: single-group feasibility trial. JMIR Hum Factors 2022 Jul 01;9(3):e38469 [FREE Full text] [doi: 10.2196/38469] [Medline: 35776504]
25. Resnick D, Schapira MM, Smith JM, Bautista A, Xu C, Jones L, et al. Promoting collaborative goal setting for cancer prevention among primary care patients through mHealth: mixed methods evaluation of a new app. JMIR Form Res 2021 Jul 14;5(7):e22510 [FREE Full text] [doi: 10.2196/22510] [Medline: 34259162]
26. Zlotorzynska M, Bauermeister JA, Golinkoff JM, Lin W, Sanchez TH, Hightow-Weidman L. Online recruitment of youth for mHealth studies. Mhealth 2021 Apr 20;7:27 [FREE Full text] [doi: 10.21037/mhealth-20-64] [Medline: 33898596]
27. Rajani NB, Mastellos N, Filippidis FT. Impact of gamification on the self-efficacy and motivation to quit of smokers: observational study of two gamified smoking cessation mobile apps. JMIR Serious Games 2021 Apr 27;9(2):e27290 [FREE Full text] [doi: 10.2196/27290] [Medline: 33904824]
28. Roberts AL, Potts HW, Koutoukidis DA, Smith L, Fisher A. Breast, prostate, and colorectal cancer survivors' experiences of using publicly available physical activity mobile apps: qualitative study. JMIR Mhealth Uhealth 2019 Jan 04;7(1):e10918 [FREE Full text] [doi: 10.2196/10918] [Medline: 30609982]
29. Bidargaddi N, Almirall D, Murphy S, Nahum-Shani I, Kovalcik M, Pituch T, et al. To prompt or not to prompt? A microrandomized trial of time-varying push notifications to increase proximal engagement with a mobile health app. JMIR Mhealth Uhealth 2018 Nov 29;6(11):e10123 [FREE Full text] [doi: 10.2196/10123] [Medline: 30497999]
30. Hui CY, McKinstry B, Walton R, Pinnock H. A mixed method observational study of strategies to promote adoption and usage of an application to support asthma self-management. J Innov Health Inform 2019 Jan 09;25(4):243-253 [FREE Full text] [doi: 10.14236/jhi.v25i4.1056] [Medline: 30672405]

31. Haile LT, Fultz HM, Simmons RG, Shelus V. Market-testing a smartphone application for family planning: assessing potential of the CycleBeads app in seven countries through digital monitoring. Mhealth 2018 Jul 23;4:27 [FREE Full text] [doi: 10.21037/mhealth.2018.06.07] [Medline: 30148140]
32. SIGN 50: a guideline developer's handbook. Scottish Intercollegiate Guidelines Network. 2008. URL: https://www.sign.ac.uk/media/2038/sign50_2019.pdf [accessed 2023-05-24]
33. Amagai S, Pila S, Kaat AJ, Nowinski CJ, Gershon RC. Challenges in participant engagement and retention using mobile health apps: literature review. J Med Internet Res 2022 Apr 26;24(4):e35120 [FREE Full text] [doi: 10.2196/35120] [Medline: 35471414]
34. Ranjani H, Nitika S, Hariharan R, Charumeena H, Oliver N, Pradeepa R, et al. Systematic review and scientific rating of commercial apps available in India for diabetes prevention. J Diabetol 2021;12(3):285-292. [doi: 10.4103/JOD.JOD_68_21]
35. 99 need-to-know stats on the mobile customer. Braze. 2016 Apr 11. URL: https://www.braze.com/resources/articles/mobile-customer-99-stats [accessed 2023-10-26]
36. Szinay D, Jones A, Chadborn T, Brown J, Naughton F. Influences on the uptake of and engagement with health and well-being smartphone apps: systematic review. J Med Internet Res 2020 May 29;22(5):e17572 [FREE Full text] [doi: 10.2196/17572] [Medline: 32348255]
37. Tobias G, Spanier AB. Developing a mobile app (iGAM) to promote gingival health by professional monitoring of dental selfies: user-centered design approach. JMIR Mhealth Uhealth 2020 Aug 14;8(8):e19433 [FREE Full text] [doi: 10.2196/19433] [Medline: 32795985]
38. König LM, Attig C, Franke T, Renner B. Barriers to and facilitators for using nutrition apps: systematic review and conceptual framework. JMIR Mhealth Uhealth 2021 Apr 01;9(6):e20037 [FREE Full text] [doi: 10.2196/20037] [Medline: 34254938]
39. Irvine M. Google Ads benchmarks for your industry [updated!]. WordStream. URL: https://www.wordstream.com/blog/ws/2016/02/29/google-adwords-industry-benchmarks [accessed 2023-05-26]
40. Irvine M. Facebook Ad benchmarks for your industry [data]. WordStream. URL: https://www.wordstream.com/blog/ws/2017/02/28/facebook-advertising-benchmarks [accessed 2023-05-26]
41. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
42. Jager KJ, Tripepi G, Chesnaye NC, Dekker FW, Zoccali C, Stel VS. Where to look for the most frequent biases? Nephrology (Carlton) 2020 Jun;25(6):435-441 [FREE Full text] [doi: 10.1111/nep.13706] [Medline: 32133725]
43. Gordon WJ, Landman A, Zhang H, Bates DW. Beyond validation: getting health apps into clinical practice. NPJ Digit Med 2020 Feb 3;3:14 [FREE Full text] [doi: 10.1038/s41746-019-0212-z] [Medline: 32047860]

Abbreviations

CPC: cost per click
CTR: clickthrough rate
GIF: graphic interchange format
mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
SIGN: Scottish Intercollegiate Guidelines Network
UAC: Universal App Campaigns

Edited by L Buis; submitted 26.06.23; peer-reviewed by P Simon, C Gibson, V Mohan; comments to author 29.09.23; revised version received 27.10.23; accepted 03.11.23; published 05.01.24.

Please cite as:
Moungui HC, Nana-Djeunga HC, Anyiang CF, Cano M, Ruiz Postigo JA, Carrion C
Dissemination Strategies for mHealth Apps: Systematic Review
JMIR Mhealth Uhealth 2024;12:e50293
URL: https://mhealth.jmir.org/2024/1/e50293
doi:10.2196/50293
PMID:38180796

Review

Effects of mHealth-Based Lifestyle Interventions on Gestational Diabetes Mellitus in Pregnant Women With Overweight and Obesity: Systematic Review and Meta-Analysis

Yirong He[1,2*], MSc; Chuanya Huang[1,2*], PhD; Qiuyang He[2,3*], MSc; Shujuan Liao[1,2], PhD; Biru Luo[1,2], PhD

[1]Department of Nursing, West China Second University Hospital, Sichuan University, Chengdu, Sichuan, China

[2]Key Laboratory of Birth Defects and Related Diseases of Women and Children, Sichuan University, Ministry of Education, Chengdu, Sichuan, China

[3]Department of Obstetric Nursing, West China Second University Hospital, Sichuan University, Chengdu, Sichuan, China

[*]these authors contributed equally

Corresponding Author:
Biru Luo, PhD
Department of Nursing
West China Second University Hospital
Sichuan University
#No. 20, Section 3, People's South Road
Chengdu, Sichuan, 610041
China
Phone: 86 88570307
Email: luomr@scu.edu.cn

Abstract

Background: The increasing incidence of gestational diabetes mellitus (GDM) is a global health problem that is more likely to occur in pregnant women with overweight or obesity. Adhering to a healthy lifestyle is associated with a reduced risk of GDM. With the development of IT, mobile health (mHealth) interventions have become widely available in health care. However, there are no definitive conclusions on the effectiveness of mHealth-based lifestyle interventions in preventing GDM.

Objective: This study aims to evaluate the impact of mHealth-based lifestyle interventions on GDM and other pregnancy outcomes in pregnant women with overweight or obesity.

Methods: A systematic literature search was conducted in 5 English databases (MEDLINE, Embase, Web of Science, CENTRAL, and CINAHL) and 4 Chinese databases (CBM, CNKI, Vip, and Wanfang) to identify randomized controlled trials (RCTs) on the effectiveness of mHealth-based interventions for GDM from inception to January 10, 2023. In total, 2 authors independently screened the studies and extracted the data. The quality of the included studies was examined using the Cochrane risk-of-bias tool. Data synthesis was conducted using Review Manager (version 5.4; The Cochrane Collaboration).

Results: A total of 16 RCTs with 7351 participants were included in this study. The included studies were published between 2014 and 2021 and were conducted in China, the United States, Australia, New Zealand, the United Kingdom, Ireland, and Norway. The sample sizes of the studies ranged from 75 to 2202, and the duration of the mHealth-based lifestyle interventions ranged from 4 to 28 weeks. Compared with usual care, mHealth-based lifestyle interventions significantly reduced the incidence of GDM (odds ratio [OR] 0.74, 95% CI 0.56-0.96; P=.03; I^2=65%), preterm birth (OR 0.65, 95% CI 0.48-0.87; P=.004; I^2=25%), macrosomia (OR 0.59, 95% CI 0.40-0.87; P=.008; I^2=59%), and gestational weight gain (mean difference=−1.12 kg, 95% CI −1.44 to −0.80; P<.001; I^2=43%). The subgroup analysis showed that interventions delivered via apps (OR 0.55, 95% CI 0.37-0.83; P=.004; I^2=44%), provided by obstetricians (OR 0.69, 95% CI 0.51-0.93; P=.02; I^2=60%), and targeted at Asian populations (OR 0.44, 95% CI 0.34-0.58; P<.001; I^2=0%) and that used the International Association of Diabetes and Pregnancy Study Groups diagnostic criteria (OR 0.58, 95% CI 0.39-0.86; P=.007; I^2=69%) showed a statistically significant reduction in the risk of GDM.

Conclusions: mHealth-based lifestyle interventions had a favorable impact on the prevention of GDM in pregnant women with overweight and obesity. Future studies need to further explore the potential of mHealth-based interventions for GDM through better design and more rigorous large-scale RCTs.

Trial Registration: PROSPERO International Prospective Register of Systematic Reviews CRD42021286995; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=286995

(JMIR Mhealth Uhealth 2024;12:e49373) doi:10.2196/49373

KEYWORDS

mobile health; mHealth; lifestyle intervention; gestational diabetes mellitus; meta-analysis; mobile phone

Introduction

Background

Gestational diabetes mellitus (GDM) is defined as a carbohydrate intolerance of varying severity with onset or first recognition during pregnancy [1]. GDM is one of the most common obstetric complications, with the prevalence varying from 7.5% to 27% in different areas [2]. GDM is associated with substantial adverse pregnancy outcomes such as neonatal hypoglycemia and macrosomia as well as long-term metabolic risk in pregnant women and their offspring [3]. Risk factors for GDM include age, ethnicity, overweight or obesity, and family history of diabetes [4]. As obesity becomes a global epidemic, perinatal overweight and obesity are also a growing concern [5]. In recent decades, the rates of overweight and obesity among women of reproductive age have increased dramatically [6]. In the United States, 60% of women are overweight or obese during pregnancy compared with 30% in Europe and 10% in Asia [7]. In pregnant women with overweight and obesity, it is estimated that the risk of GDM is more than twice that of other pregnant women [8]. Unhealthy lifestyle behaviors are a critical factor that affects overweight and obesity during the perinatal period [9]. Numerous randomized controlled trials (RCTs) have attempted to reduce the incidence of GDM through diet [10], physical activity [11], or combination interventions [12]. However, when lifestyle interventions are provided in a personalized way, they are commonly expensive and may lack scalability from the perspective of public health [13]. Therefore, there is an urgent need for effective real-world solutions that address the demand of pregnant women seeking personalized support, information, and guidance to help reduce the risk of GDM.

As IT advances and more people use electronic devices, mobile health (mHealth) has developed rapidly [14]. mHealth is commonly defined as health care services provided by health care professionals using telecommunications technology [15]. mHealth has been applied in many areas of perinatal health care, with medical staff providing health care services to pregnant women through multimedia methods such as mobile apps, software, SMS text messages, email, web-based diaries, and integrated systems combining various components of digital communications technologies [16]. Studies have shown that mHealth care can reduce gestational weight gain (GWG); improve pregnant women's health behaviors; and reduce the number of medical visits, thereby decreasing financial burden [13]. However, the impact of mHealth interventions on pregnancy outcomes in women with overweight or obesity is uncertain. More and more systematic reviews have found an effect of lifestyle interventions based on mHealth technology on diabetes prevention among adults with overweight and obesity [17,18]; however, little is known about their effectiveness in the perinatal population. Several previous reviews have attempted to synthesize the results of mHealth-based lifestyle programs for pregnant women, but none have evaluated the quantitative effects of these programs [19,20]. There is still no consensus on the impact of mHealth lifestyle interventions on preventing GDM and other pregnancy outcomes in women with overweight or obesity.

Objectives

Therefore, we conducted a systematic review and meta-analysis of RCTs to summarize mHealth interventions delivered in different ways and assess the effectiveness of mHealth-based lifestyle interventions in reducing the risk of GDM.

Methods

This systematic review and meta-analysis follows the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [21] and is presented in Multimedia Appendix 1. It was registered in PROSPERO on November 18, 2021, with registration number CRD42021286995.

Search Strategy

A systematic literature search was conducted in 5 English databases (MEDLINE, Embase, Web of Science, CENTRAL, and CINAHL) and 4 Chinese databases (CBM, CNKI, Vip, and Wanfang) for studies published in English and Chinese. A systematic search was conducted combining Medical Subject Headings (MeSH) and free-text terms, including *overweight* (MeSH) OR *obesity* (MeSH); *pregnan** OR *prenatal* OR *antenatal* OR *maternal* OR *gestational*; *gestational diabetes mellitus* OR *gestational diabetes* OR *GDM* OR *T2DM* OR *impaired fasting glucose* OR *impaired glucose tolerance*; *telemedicine* (MeSH) OR *telerehabilitation* (MeSH) OR *telecommunications* (MeSH) OR *electronic health* OR *eHealth* OR *ehealth* OR *mHealth* OR *mobile health* OR *telecare* OR *eHealthcare* OR *mcare* OR *telemonitor** OR *telerehab** OR *telemanagement* OR *mobile communication* OR *remote consult* OR *mobile technolog** OR *mobile devic** OR *mobile app** OR *internet* (MeSH) OR *web** OR *online* OR *smartphone* (MeSH) OR *telephone* (MeSH) OR *cell phone* (MeSH) OR *cellular phone* (MeSH) OR *mobile phone* OR *messag** OR *SMS*. The searches were unlimited by time up to January 10, 2023, and were limited to RCTs. The full details of the search strategy for each database are provided in Multimedia Appendix 2. We complemented this strategy by manually searching the reference lists of included studies and related reviews.

Inclusion and Exclusion Criteria

In total, 2 researchers (YH and CH) independently screened the titles and abstracts and selected the studies in accordance with

the eligibility criteria. Any disagreement was resolved through discussion with a third researcher (QH). Studies were included if they met the following inclusion criteria: (1) pregnant women with overweight (BMI ≥25 kg/m^2) or obesity (BMI ≥30 kg/m^2; population); (2) mHealth interventions including pregnancy nutrition, physical activity, weight management, and health behavior education delivered via the internet, websites, telephone, app, SMS text message, email, or other types of information and communications technologies (intervention); (3) usual care, routine care, conventional care, or standard care without mHealth (comparison); (4) incidence of GDM, postpartum hemorrhage, preterm birth, cesarean delivery, pregnancy-induced hypertension, macrosomia, neonatal gestational age, and GWG (outcome); (5) RCTs (study design); and (6) English or Chinese (language). There were no restrictions regarding the year of publication. We excluded studies that (1) included women with either type 1 or type 2 diabetes mellitus before pregnancy or with existing GDM; (2) contained incomplete data; (3) lacked data related to GDM; or (4) were study protocols, comments, editorials, and conference abstracts.

Study Selection and Data Extraction

The reference management program EndNote X9 (Clarivate Analytics) was used for data management. The studies were imported into EndNote after an extensive database search. After removing duplicates, 2 authors independently reviewed the titles and abstracts according to the eligibility criteria. Disagreements were resolved through discussion or consultation with a third researcher. Data were extracted by an independent researcher (YH) using the predesigned data collection forms, and the extracted data were verified by a second researcher (CH). Disagreements were resolved through consensus. The extracted data included the authors, year, country, study design, sample size, participant characteristics, intervention, control, GDM criteria, and outcomes.

Quality Assessment

In total, 2 researchers (YH and CH) independently assessed the studies' risk of bias in accordance with the *Cochrane Handbook for Systematic Reviews of Interventions*. This tool consists of 6 items: selection bias, performance bias, detection bias, attrition bias, reporting bias, and other biases. Each item was judged as "low risk," "high risk," or "unclear risk." A third researcher (QH) was available if there was a difference in opinion in assessing the risk of bias.

Data Synthesis and Analysis

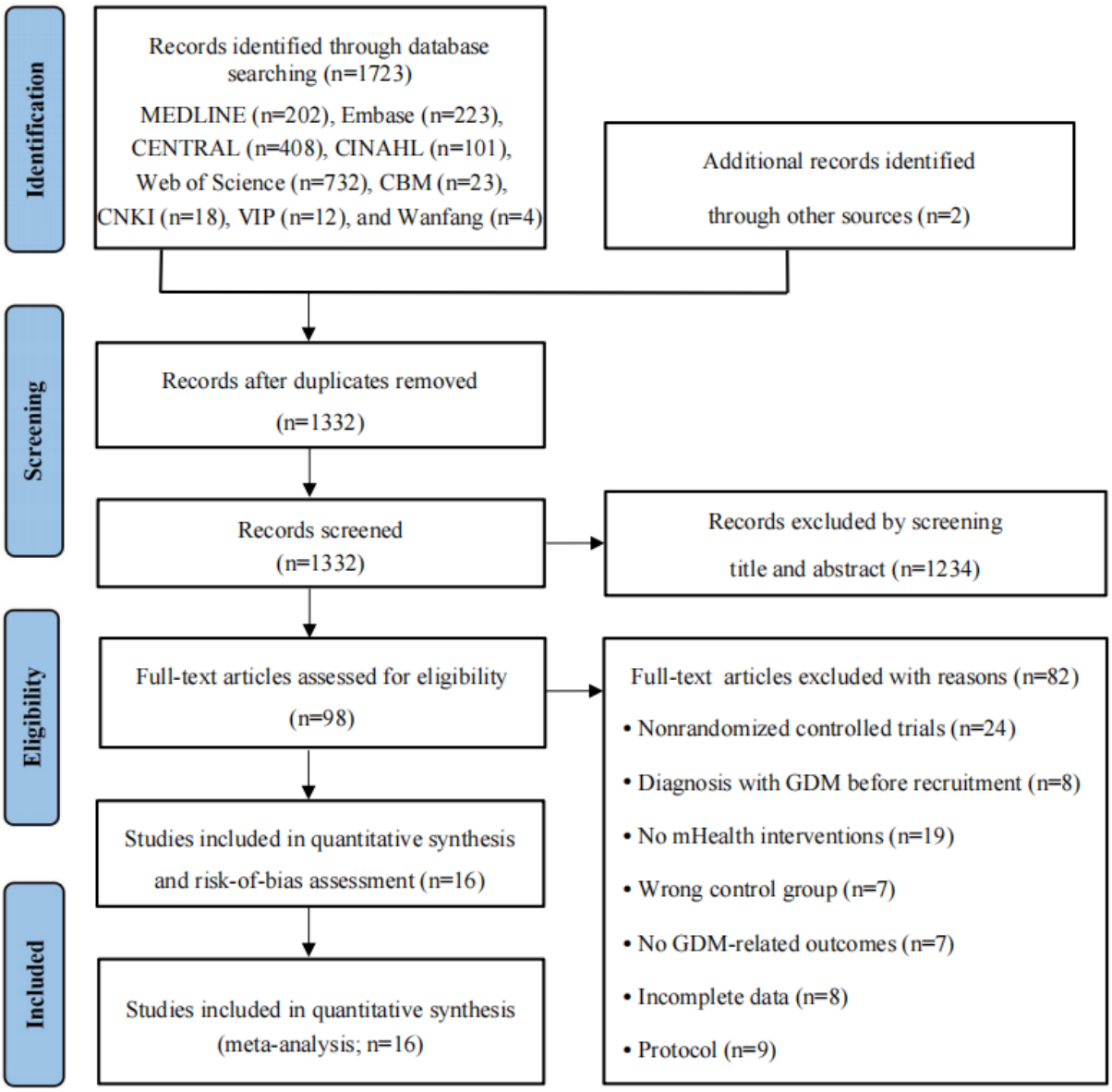
Statistical analyses were conducted using Review Manager (version 5.4; The Cochrane Collaboration). The overall effect difference was considered statistically significant if the 2-tailed *P* value was <.05. Continuous variables were presented using the mean difference, and dichotomous variables were described using the odds ratio (OR) with a 95% CI. Heterogeneity was assessed using the Cochran *Q* test and the I^2 statistic. We considered <25%, 25% to 50%, 50% to 75%, and >75% as low, moderate, high, and severe heterogeneity between the studies, respectively. If I^2≤50% and the *P* value was >.10, a fixed-effects model was considered; otherwise, a random-effects model was used. The sources of heterogeneity were explored using subgroup analysis. A funnel plot was constructed to check for potential publication bias.

Results

Study Selection

Figure 1 shows a PRISMA flowchart of the study selection process. A total of 1725 records were retrieved from 9 electronic databases. After removing duplicates (393/1725, 22.78% of the studies), 1332 studies were included for screening. Of these 1332 studies, we then excluded 1234 (92.64%) based on the relevance of the abstract and title, and the remaining 98 (7.36%) studies were assessed for eligibility. After full-text review, 84% (82/98) of the studies were excluded for the reasons outlined in Figure 1. Finally, 16 studies were included in the review and meta-analysis.

Figure 1. Flowchart of study selection. GDM: gestational diabetes mellitus; mHealth: mobile health.



Characteristics of the Included Studies

Study Characteristics

The characteristics of the included studies are shown in Table 1. The 16 studies included in this review and meta-analysis were RCTs, 4 (25%) of which were multicenter RCTs. The included studies were published between 2014 and 2021 and were conducted in China (8/16, 50%) [22-29], the United States (2/16, 12%) [30,31], Australia (2/16, 12%) [32,33], New Zealand (1/16, 6%) [34], the United Kingdom (1/16, 6%) [35], Ireland (1/16, 6%) [36], and Norway (1/16, 6%) [37]. A total of 7351 participants were included in the studies, and the sample sizes varied from 74 [34] to 2153 [32]. In total, 38% (6/16) of the studies included pregnant women with obesity [23,24,26,28,33,35], and 62% (10/16) included both pregnant women with overweight and pregnant women with obesity [22,25,27,29,30,31,32,34,36, 37]. A total of 81% (13/16) of the articles reported diagnostic criteria for GDM. The incidence of GDM was determined by screening pregnant women using an oral glucose tolerance test. The diagnostic criteria for GDM were inconsistent among the included studies. In total, 56% (9/16) of the RCTs used the International Association of Diabetes and Pregnancy Study Groups (IADPSG) criteria (2010) [22,23,25,27-29,34-36], 12% (2/16) used the World Health Organization 2013 criteria [33,37], 6% (1/16) used the Australasian Diabetes in Pregnancy Society criteria [32], and 6% (1/16) of the RCTs used the Carpenter-Coustan criteria [30]. All the RCTs (16/16, 100%) reported more than 1 outcome.

Table 1. Characteristics of the included studies.

Author, year, and country	Study design	Sample size	Participant characteristics	Intervention	Control	GDM[a] criteria	Outcomes
Dodd et al [32], 2014, Australia	Multicenter RCT[b]	2153; IG[c]: 1080; CG[d]: 1073 (1:1)	Women with a singleton pregnancy between 10 and 20 wk of gestation and a BMI of ≥25 kg/m^2	Phone, in person	Standard care	ADPSC[e]	LGA[f], macrosomia, PIH[g], pre-eclampsia, GDM, PTB[h], CS[i], and PPH[j]
Sagedal et al [37], 2015, Norway	RCT	591; IG: 296; CG: 295 (1:1)	Women with a singleton pregnancy at ≤20 wk of gestation who had a prepregnancy BMI of ≥25 kg/m^2	Phone, website, in person	Standard care	WHO[k]	GWG[l], GDM, LGA, SGA[m], pre-eclampsia, PTB, PPH, and NW[n]
Seneviratne et al [34], 2015, New Zealand	RCT	74; IG: 37; CG: 37 (1:1)	Women aged 18-40 y with a BMI of ≥25 kg/m^2 and a singleton pregnancy of <20 wk of gestation	Software, device, home-based, in person	Usual care	IADPSG[o]	NW, hypoglycemia, GWG, GDM, PIH, PTB, CS, PPH, MVPA[p], and pre-eclampsia
Poston et al [35], 2015, United Kingdom	Multicenter RCT	1280; IG: 629; CG: 651 (1:1)	Women aged ≥16 y with a BMI of ≥30 kg/m² and a singleton pregnancy between 15 and 18 wk plus 6 d of gestation	Phone, email, DVD, pedometer, logbook, in person	Standard care	IADPSG	GDM, FBG[q], pre-eclampsia, CS, PPH, GWG, HOMA-IR[r], GL[s], GI[t], fat, MVPA, LGA, and NW
Simmons et al [33], 2017, Australia	Multicenter RCT	192; IG: 92; CG: 100 (1:1)	Women with a BMI of >29 kg/m^2, ≤19 (–6 to , 6) d of gestation, and a singleton pregnancy	Email, phone, pedometer, device, in person	Usual care	WHO	GDM, FBG, HOMA-IR, insulin, NW, LGA, SGA, MVPA, and sugar intake
Chen [22], 2017, China	RCT	160; IG: 80; CG: 80 (1:1)	Women with a prepregnancy BMI of ≥24 kg/m^2 and a singleton pregnancy between 8 and 12 wk of gestation	Mobile apps, SMS text messages, in person	Usual care	IADPSG	BMI, GDM, CS, and NW
Kennelly et al [36], 2018, Ireland	RCT	498; IG: 241; CG: 257 (1:1)	Singleton pregnant women between 10 and 15 wk of gestation with a BMI between 25.0 and 39.9 kg/m^2	Mobile app, email, in person	Usual care	IADPSG	GDM, GWG, GL, GI, PIH, CS, HOMA-IR, NW, LGA, and SGA
Li [23], 2018, China	RCT	1000; IG: 500; CG: 500 (1:1)	Women with a BMI of ≥28 kg/m^2, aged >18 y, and with a singleton pregnancy between 8 and 12 wk of gestation	Phone, in person	Usual care	IADPSG	GDM, GWG, NW, and macrosomia
Tang et al [24], 2019, China	RCT	136; IG: 68; CG: 68 (1:1)	Pregnant women with a BMI of ≥28 kg/m^2 aged >18 y	Mobile app, software	Standard care	N/A[u]	Macrosomia, CS, hypoglycemia, and GDM
Ferrara et al [30], 2020, United States	Multicenter RCT	389; IG: 195; CG: 194 (1:1)	Women at 8-15 wk of gestation with singletons, with a prepregnancy BMI of 25-40 kg/m², and aged ≥18 y	Phone, device, in person	Usual care	Carpenter and Coustan criteria	GWG, caloric intake, MVPA, FBG, HOMA-IR, GDM, PIH, pre-eclampsia, CS, NW, and macrosomia

Author, year, and country	Study design	Sample size	Participant characteristics	Intervention	Control	GDM[a] criteria	Outcomes
Cao [25], 2020, China	RCT	96; IG: 48; CG: 48 (1:1)	Women aged 22-38 y with a BMI of ≥24 kg/m^2 and a singleton pregnancy between 8 and 12 wk of gestation	Mobile apps, SMS text messages, in person	Usual care	IADPSG	FBG, GDM, BMI, CS, NW, and LGA
Wu and Guang [26], 2020, China	RCT	140; IG: 70; CG: 70 (1:1)	Pregnant women with a BMI of ≥28 kg/m^2 aged >18 y	Mobile app, in person	Usual care	N/A	BMI, NW, GDM, PIH, CS, PTB, macrosomia, and PPH
Liu et al [31], 2021, United States	RCT	217; IG: 112; CG: 105 (1:1)	Pregnant women aged 18-44 y with a gestational age of ≤16 wk and a prepregnancy BMI of ≥25 kg/m^2	Phone, software, SMS text messages, in person	Standard care	N/A	GWG, PTB, LBW[v], macrosomia, SGA, GDM, PIH, CS, and NW
Zhou et al [27], 2021, China	RCT	104; IG: 52; CG: 52 (1:1)	Singleton pregnant women with a BMI of ≥24 kg/m^2 and aged >18 y	Mobile app, in person	Usual care	IADPSG	FBG, 2-hour BG[w], GDM, CS, PTB, and macrosomia
Kang and Sung [28], 2021, China	RCT	106; IG: 53; CG: 53 (1:1)	Women with a BMI of ≥28 kg/m^2, aged >18 y, and with a singleton pregnancy between 12 and 20 wk of gestation	Mobile app, in person	Usual care	IADPSG	GDM, FBG, 2-hour BG, HbA_{1c}[x], CS, PTB, and macrosomia
Ding et al [29], 2021, China	RCT	215; IG: 104; CG: 111 (1:1)	Pregnant women with a BMI of ≥24 kg/m^2 at the onset of pregnancy, aged <35 y, and at <12 wk of gestation	Mobile app, in person	Usual care	IADPSG	Energy intake, GDM, FBG, GWG, CS, PTB, PIH, preeclampsia, PPH, NW, and macrosomia

[a]GDM: gestational diabetes mellitus.

[b]RCT: randomized controlled trial.

[c]IG: intervention group.

[d]CG: control group.

[e]ADPSC: Australasian Diabetes in Pregnancy Society criteria.

[f]LGA: large-for-gestational-age infant.

[g]PIH: pregnancy-induced hypertension.

[h]PTB: preterm birth.

[i]CS: cesarean section.

[j]PPH: postpartum hemorrhage.

[k]WHO: World Health Organization.

[l]GWG: gestational weight gain.

[m]SGA: small-for-gestational-age infant.

[n]NW: neonatal weight.

[o]IADPSG: International Association of Diabetes and Pregnancy Study Groups.

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

[q]FBG: fasting blood glucose.

[r]HOMA-IR: homeostatic model assessment of insulin resistance.

[s]GL: glycemic load.

[t]GI: glycemic index.

[u]N/A: not applicable.

[v]LBW: low birth weight.

[w]BG: blood glucose.

[x]HbA_{1c}: glycated hemoglobin.

Characteristics of the mHealth Interventions

The details of the mHealth interventions are presented in Multimedia Appendix 3 [22-37]. The interventions were divided into 3 groups: exercise only, diet only, and mixed interventions. Regarding the delivery mode, 12% (2/16) of the studies [23,32] provided mHealth interventions through phone counseling, 25% (4/16) of the studies [26-29] only provided the interventions through mobile apps, and 62% (10/16) of the studies [22,24,25,30,31,33-37] adopted a combination of methods. The duration of the mHealth interventions ranged from 4 to 28 weeks. The mHealth interventions were delivered by various types of personnel, including a single provider in 12% (2/16) of the studies [29,32] and multidisciplinary prenatal care providers in 88% (14/16) of the studies [22-28,30,31,33-37].

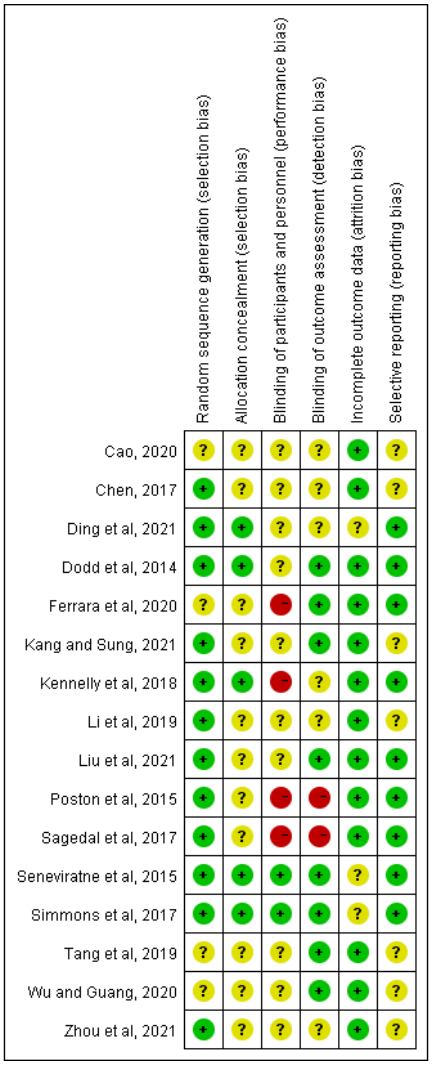
Characteristics of the Comparators

Most studies (14/16, 88%) briefly described the control group, which generally included regular maternity visits and education on diet and exercise during pregnancy. In the control groups, all studies provided standard or usual care for participants, which was based on the different countries' perinatal practices and local hospital guidelines.

Risk of Bias

The Cochrane risk-of-bias tool was used to assess the risk of bias in each study, and the results are shown in Figure 2 [22-37]. A total of 75% (12/16) of the studies described the details of the randomization scheme, and the risk of random sequence generation was low. Only 31% (5/16) of the studies reported allocation concealment, with a low risk of selection bias. Blinding of participants or staff was difficult because of the nature of the mHealth interventions; thus, 25% (4/16) of the studies, which did not blind the participants or staff, were rated as having a high risk of bias, and 62% (10/16) of the studies were rated as unclear regarding the risk of bias. A total of 12% (2/16) of the studies did not report the outcome assessment and were rated as having a high risk of detection bias. In total, 81% (13/16) of the included studies described details of participant dropout and had a low risk of bias. A total of 56% (9/16) of the studies followed preregistered protocols for analysis and outcome reporting, and the remaining 44% (7/16) did not provide information on published protocols or registrations and had an unclear risk of reporting bias. The funnel plot showed no publication bias in any of the included studies (Multimedia Appendix 4).

Figure 2. Results of the risk-of-bias assessment of the included studies [22-37].



Meta-Analysis Results

Primary Outcome: the Incidence of GDM

The pooled analysis of the 16 RCTs with 7351 participants showed that the mHealth interventions significantly decreased the incidence of GDM in women with overweight and obesity (OR 0.74, 95% CI 0.56-0.96; *P*=.03; I^2=65%; Figure 3 [22-37]).

Table 2 provides the results of the subgroup analyses, and the forest plots are presented in Multimedia Appendix 5 [22-37]. The subgroup analysis conducted based on the different interventions showed that diet (OR 0.61, 95% CI 0.29-1.28; *P*=.19; I^2=61%), exercise (OR 1.11, 95% CI 0.57-2.17; *P*=.76; I^2=0%), or a combination of the 2 (OR 0.71, 95% CI 0.51-0.99; *P*=.05; I^2=74%) were not statistically significantly associated with reduced risk of GDM. When compared with the control groups, app-based mHealth interventions were significantly effective in reducing GDM (OR 0.55, 95% CI 0.37-0.83; *P*=.004; I^2=44%). Another subgroup analysis conducted based on different providers indicated that interventions provided by obstetricians and nurses were effective in reducing GDM (OR 0.69, 95% CI 0.51-0.93; *P*=.02; I^2=60%). Interventions targeted at Asian populations (OR 0.44, 95% CI 0.34-0.58; *P*<.001; I^2=0%) and using the IADPSG as a diagnostic criterion for GDM showed a reduction in GDM compared with the control groups (OR 0.58, 95% CI 0.39-0.86; *P*=.007; I^2=69%).

Figure 3. Forest plot of the effect of mobile health interventions on gestational diabetes mellitus [22-37]. IV: inverse variance.

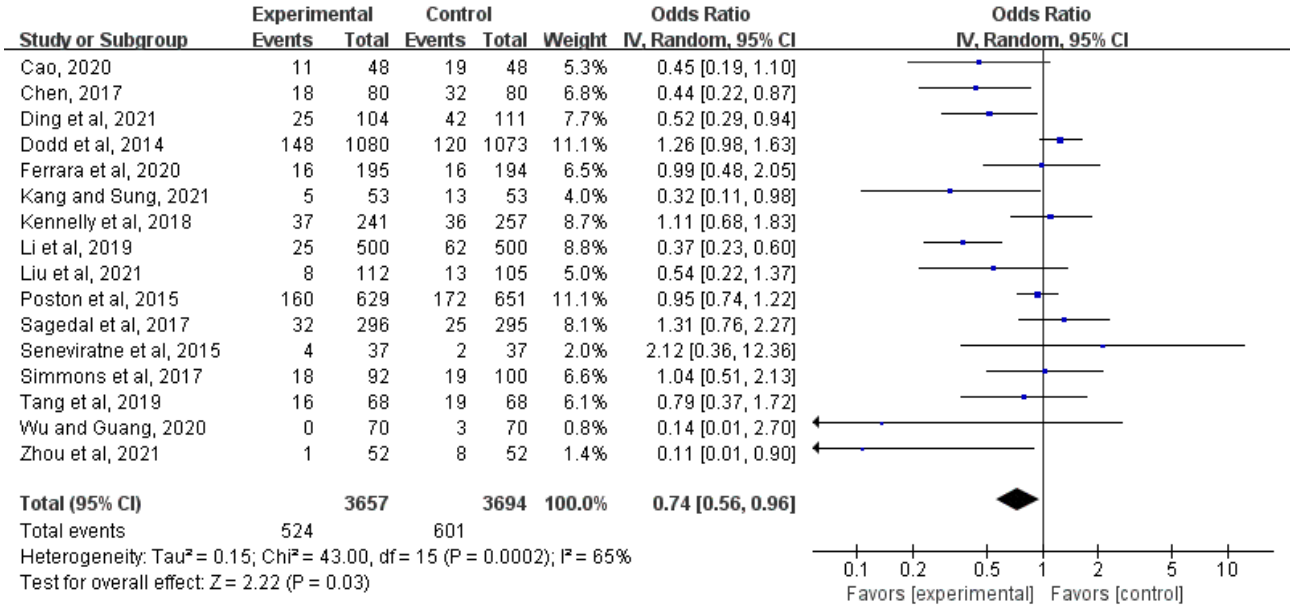


Table 2. Subgroup analyses of the included studies (n=16).

Subgroup analysis	References	Studies, n (%)	Sample size	Effect estimates (95% CI)	*P* value	I^2 (%)
Intervention						
Diet	• Tang et al [24] • Zhou et al [27] • Kang and Sung [28] • Kennelly et al [36]	4 (25)	844	0.61 (0.29-1.28)	.19	61
Exercise	• Ferrara et al [30] • Seneviratne et al [34]	2 (12)	463	1.11 (0.57-2.17)	.76	0
Both	• Chen [22] • Li [23] • Cao [25] • Wu and Guang [26] • Ding et al [29] • Liu et al [31] • Dodd et al [32] • Simmons et al [33] • Poston et al [35] • Sagedal et al [37]	10 (62)	6044	0.71 (0.51-0.99)	.05	74
mHealth[a] technology						
Phone	• Ferrara et al [30] • Simmons et al [33] • Poston et al [35] • Sagedal et al [37] • Liu et al [31] • Dodd et al [32]	6 (38)	4822	1.08 (0.92-1.26)	.37	1
App	• Chen [22] • Tang et al [24] • Cao [25] • Wu and Guang [26] • Zhou et al [27] • Kang and Sung [28] • Ding et al [29] • Kennelly et al [36]	8 (50)	1455	0.55 (0.37-0.83)	.004	44
Computer	• Li [23] • Seneviratne et al [34]	2 (12)	1074	0.72 (0.14-3.74)	.69	71
Provider						
Dietitian	• Ding et al [29] • Dodd et al [32]	2 (12)	2368	0.84 (0.36-2.00)	.70	86
Exercise physiologists	• Seneviratne et al [34]	1 (6)	74	2.12 (0.36-12.36)	.40	N/A[b]
Obstetricians	• Chen [22] • Li [23] • Tang et al [24] • Cao [25] • Wu and Guang [26] • Zhou et al [27] • Kang and Sung [28] • Ferrara et al [30] • Liu et al [31] • Simmons et al [33] • Poston et al [35] • Kennelly et al [36] • Sagedal et al [37]	13 (81)	4909	0.69 (0.51-0.93)	.02	60
Duration						

Subgroup analysis	References	Studies, n (%)	Sample size	Effect estimates (95% CI)	*P* value	I^2 (%)
Short term	• Zhou et al [27] • Liu et al [31] • Poston et al [35]	3 (19)	1601	0.60 (0.26-1.39)	.24	61
Medium to long term	• Chen [22] • Li [23] • Wu and Guang [26] • Kang and Sung [28] • Ding et al [29] • Ferrara et al [30] • Dodd et al [32] • Simmons et al [33] • Seneviratne et al [34] • Kennelly et al [36] • Sagedal et al [37]	11 (69)	5518	0.77 (0.53-1.10)	.15	72
Long term	• Tang et al [24] • Cao [25]	2 (12)	232	0.62 (0.35-1.12)	.11	0
Ethnicity						
Asian	• Chen [22] • Li [23] • Tang et al [24] • Cao [25] • Wu and Guang [26] • Zhou et al [27] • Kang and Sung [28] • Ding et al [29]	8 (50)	1957	0.44 (0.34-0.58)	<.001	0
White	• Ferrara et al [30] • Liu et al [31] • Dodd et al [32] • Simmons et al [33] • Seneviratne et al [34] • Poston et al [35] • Kennelly et al [36] • Sagedal et al [37]	8 (50)	5394	1.09 (0.93-1.26)	.29	0
GDM[c] diagnostic criteria						

Subgroup analysis	References	Studies, n (%)	Sample size	Effect estimates (95% CI)	*P* value	I^2 (%)
IADPSG[d]	• Chen [22] • Li [23] • Cao [25] • Zhou et al [27] • Kang and Sung [28] • Ding et al [29] • Seneviratne et al [34] • Poston et al [35] • Kennelly et al [36]	9 (56)	3533	0.58 (0.39-0.86)	.007	69
WHO[e]	• Simmons et al [33] • Sagedal et al [37]	2 (12)	783	1.20 (0.78-1.86)	.41	0
ADPSC[f]	• Dodd et al [32]	1 (6)	2153	1.26 (0.98-1.63)	.08	N/A
Carpenter-Coustan criteria	• Ferrara et al [30]	1 (6)	389	0.99 (0.48-2.05)	.99	N/A

[a]mHealth: mobile health.

[b]N/A: not applicable.

[c]GDM: gestational diabetes mellitus.

[d]IADPSG: International Association of Diabetes and Pregnancy Study Groups.

[e]WHO: World Health Organization.

[f]ADPSC: Australasian Diabetes in Pregnancy Society criteria.

Secondary Outcomes: Maternal and Neonatal Outcomes

Effect on GWG

A total of 50% (8/16) of the studies examined the effects of mHealth-based interventions on GWG [23,29-31,34-37]. There were statistically significant differences in decreases in GWG between the mHealth intervention groups and the control groups (mean difference=−1.12 kg, 95% CI −1.44 to −0.80; P<.001; I^2=43%; Figure 4 [23, 29-31,34-37]). The effects of the mHealth-based lifestyle interventions on maternal and neonatal outcomes are shown in Table 3.

Figure 4. Forest plot of the effect of mobile health interventions on gestational weight gain [23, 29-31,34-37]. IV: inverse variance.

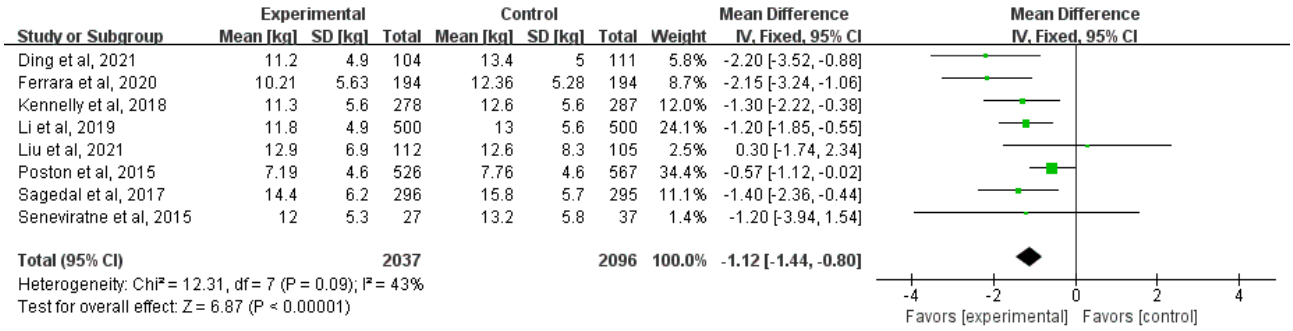


Table 3. Effectiveness of mobile health–based lifestyle interventions on maternal and neonatal outcomes.

Maternal and neonatal outcomes	References	Studies, n (%)	Sample size	Statistical method	Effect estimates (95% CI)	*P* value
Maternal outcomes						
Postpartum hemorrhage (%)	• Wu and Guang [26] • Ding et al [29] • Dodd et all [32] • Seneviratne et al [34] • Poston et al [35] • Sagedal et al [37]	6 (38)	4664	Odds ratio (IV[a]; fixed)	1.06 (0.90 to 1.24)	.49
Cesarean delivery (%)	• Chen [22] • Tang et al [24] • Cao [25] • Wu and Guang [26] • Ferrara et al [30] • Dodd et al [32] • Seneviratne et al [34] • Poston et al [35] • Kennelly et al [36]	9 (56)	5224	Odds ratio (IV; fixed)	0.89 (0.79 to 1.00)	.05
Preeclampsia or PIH[b] (%)	• Wu and Guang [26] • Ding et al [29] • Ferrara et al [30] • Liu et al [31] • Dodd et al [32] • Seneviratne et al [34] • Poston et al [35] • Kennelly et al [36] • Sagedal et al [37]	9 (56)	5829	Odds ratio (IV; fixed)	0.96 (0.80 to 1.15)	.63
GWG[c] (kg)	• Li [23] • Ding et al [29] • Ferrara et al [30] • Liu et al [31] • Seneviratne et al [34] • Poston et al [35] • Kennelly et al [36] • Sagedal et al [37]	8 (50)	4133	Mean difference (IV; fixed)	−1.12 (−1.44 to −0.80)	<.001
Neonatal outcomes						
Preterm birth (%)	• Wu and Guang [26] • Zhou et al [27] • Kang and Sung [28] • Ding et al [29] • Liu et al [31] • Dodd et al [32] • Seneviratne et al [34]	7 (44)	2998	Odds ratio (IV; fixed)	0.65 (0.48 to 0.87)	.004
Macrosomia (%)	• Li [23] • Tang et al [24] • Wu and Guang [26] • Zhou et al [27] • Kang and Sung [28] • Ding et al [29] • Ferrara et al [30] • Liu et al [31] • Dodd et al [32]	9 (56)	4449	Odds ratio (IV; random)	0.59 (0.40 to 0.87)	.008
LGA[d] >90th percentile (%)	• Dodd et al [32] • Simmons et al [33] • Poston et al [35] • Kennelly et al [36] • Sagedal et al [37]	5 (31)	4966	Odds ratio (IV; random)	0.80 (0.60 to 1.06)	.12

Maternal and neonatal outcomes	References	Studies, n (%)	Sample size	Statistical method	Effect estimates (95% CI)	*P* value
SGA[e] <10th percentile (%)	• Liu et al [31] • Simmons et al [33] • Kennelly et al [36] • Sagedal et al [37]	4 (25)	1529	Odds ratio (IV; fixed)	1.10 (0.78 to 1.55)	.60

[a]IV: inverse variance.

[b]PIH: pregnancy-induced hypertension.

[c]GWG: gestational weight gain.

[d]LGA: large for gestational age.

[e]SGA: small for gestational age.

Effect on Cesarean Delivery

A total of 56% (9/16) of the studies explored the effects of mHealth-based interventions on cesarean delivery [22,24-26,30,32,34-36]. The pooled results showed no significant difference between the mHealth intervention and control groups (OR 0.89, 95% CI 0.79-1.00; P=.05; I^2=39%). Forest plots of maternal and neonatal outcomes are shown in Multimedia Appendix 6 [22-37].

Effect on Pregnancy-Induced Hypertension

A total of 56% (9/16) of the studies assessed the effects of mHealth-based interventions on the risk of pregnancy-induced hypertension [26,29-32,34-37]. Compared with the control groups, the groups with mHealth-based interventions did not show statistically significant decreases in pregnancy-induced hypertension (OR 0.96, 95% CI 0.80-1.15; P=.63; I^2=31%).

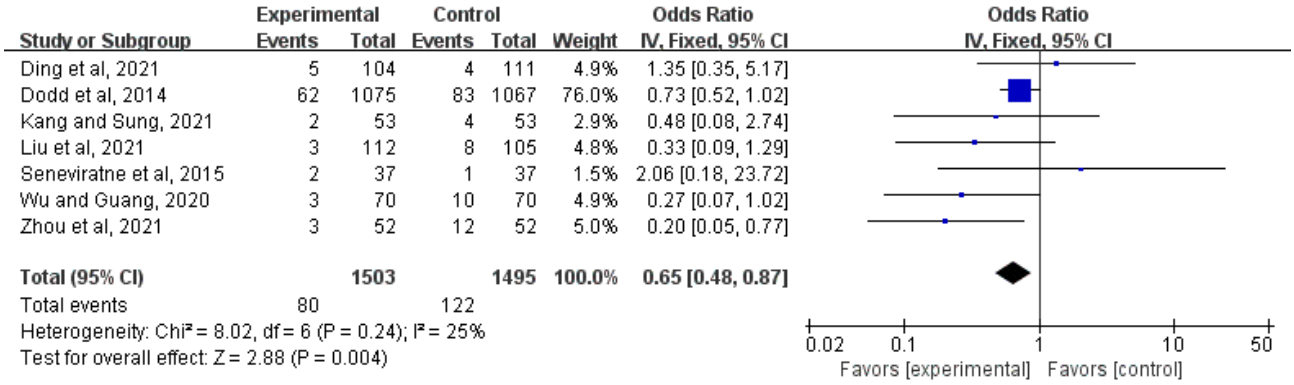
Effect on Postpartum Hemorrhage

A total of 38% (6/16) of the studies assessed the effects of mHealth-based lifestyle interventions on the risk of postpartum hemorrhage [26,29,32,34,35,37]. There were no statistically significant decreases in postpartum hemorrhage between the groups with mHealth-based interventions and the control groups (OR 1.06, 95% CI 0.90-1.24; P=.49; I^2=17%).

Effect on Preterm Birth

A total of 44% (7/16) of the studies explored the effects of mHealth-based interventions on the risk of preterm birth [26-29,31,32,34]. Statistically significant differences in decreases in preterm birth were found between the mHealth intervention groups and the control groups (OR 0.65, 95% CI 0.48-0.87; P=.004; I^2=25%; Figure 5 [26-29,31,32,34]).

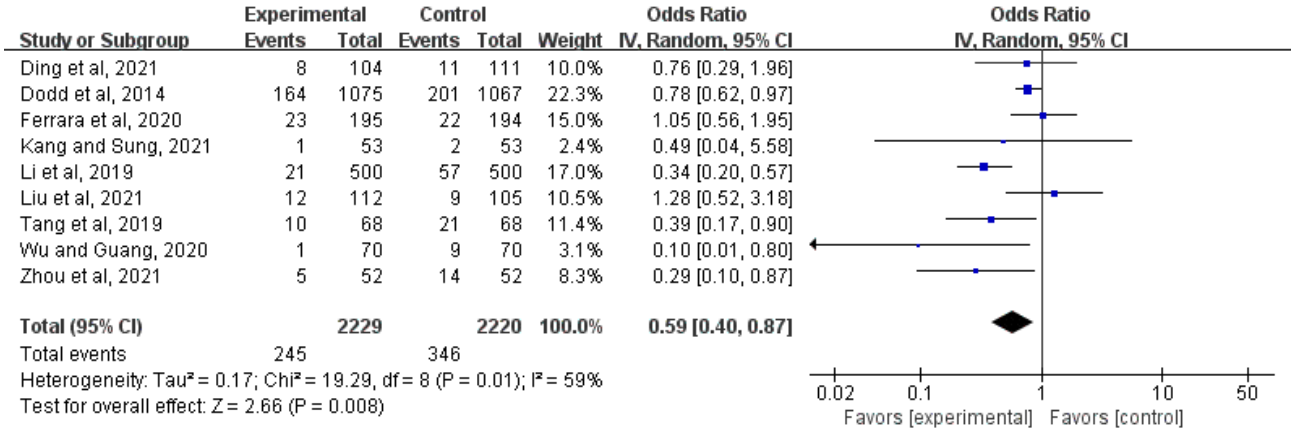
Figure 5. Forest plot of the effect of mobile health interventions on preterm birth [26-29,31,32,34]. IV: inverse variance.



Effect on Macrosomia

A total of 56% (9/16) of the studies evaluated the effects of mHealth-based interventions on the risk of macrosomia [23,24,26-32]. Statistically significant differences in decreases in macrosomia were found between the mHealth intervention groups and the control groups (OR 0.59, 95% CI 0.40-0.87; P=.008; I^2=59%; Figure 6 [23,24,26-32]).

Figure 6. Forest plot of the effect of mobile health interventions on macrosomia [23, 24,26-32]. IV: inverse variance.



Effect on Large-for-Gestational-Age Infants

A total of 31% (5/16) of the studies explored the effects of mHealth-based lifestyle interventions on large-for-gestational-age infants [32,33,35-37]. Compared with the control groups, the mHealth intervention groups did not show statistically significant decreases in large-for-gestational-age infants (OR 0.80, 95% CI 0.60-1.06; P=.12; I^2=51%).

Effect on Small-for-Gestational-Age Infants

In total, 25% (4/16) of the studies assessed the effects of mHealth-based lifestyle interventions on small-for-gestational-age infants [31,33,36,37]. Compared with the control groups, the mHealth intervention groups did not show statistically significant decreases in small-for-gestational-age infants (OR 1.10, 95% CI 0.78-1.55; P=.60; I^2=0%).

Discussion

Methodological Quality of the Included Studies

The methodological quality of the included studies was assessed using the Cochrane risk-of-bias tool. A total of 25% (4/16) of the studies did not report details of randomization methods, and 75% (12/16) of the studies did not report allocation concealment and were at risk of selection bias. In total, 25% (4/16) of the studies did not blind participants and researchers. Therefore, there was a high risk of performance bias in these studies. A total of 12% (2/16) of the studies did not blind the outcome assessors and were rated as having a high risk of detection bias. The included studies reported complete data with low attrition bias. Finally, the funnel plot showed no substantial publication bias among the included studies. Therefore, the overall methodological quality was moderate, and larger samples and well-designed randomized trials are needed in the future.

Summary of Principal Findings

Effectiveness of mHealth Interventions

A total of 16 RCTs were included in this systematic review and meta-analysis. All of these studies offered mHealth lifestyle interventions for perinatal women with overweight or obesity. Pooled results showed that the mHealth interventions reduced the incidence of GDM, which was consistent with the results of a previous study [38]. Various strategies have been proposed to prevent GDM effectively, and the primary intervention strategy is to change the lifestyle. Lifestyle interventions, including dietary guidance, physical exercise, weight management, health education, and blood glucose self-monitoring, are first-line strategies for GDM prevention [39]. The benefits of lifestyle interventions are mediated by mechanisms that improve glycemic variables and outcomes in type 4 and type 2 diabetes by increasing insulin sensitivity and reducing oxidative stress, which has been demonstrated in studies in other populations [40]. Compared with conventional lifestyle interventions, mHealth-based lifestyle interventions can make health education more attractive by providing more intuitive and vivid education and consultation with the help of electronic devices. In addition, pregnant women need to maintain close and continuous contact with the medical team during pregnancy and communicate effectively in the event of pregnancy complications and conscious fetal abnormalities. The advantage of mHealth technology is its ability to provide time-sensitive connectivity and high-quality health care for pregnant women in all regions [41]. mHealth lifestyle interventions enable pregnant women to maintain a healthy diet and engage in appropriate physical activity on a daily basis to encourage a healthier lifestyle during pregnancy, thus reducing the incidence of GDM.

Subgroup Analysis

In this review, we classified mHealth lifestyle interventions into 3 groups based on delivery approach: app-based interventions, phone-based interventions, and mHealth interventions delivered via a computer. Subgroup analysis demonstrated that app-delivered mHealth interventions were highly effective in reducing the risk of GDM in women with overweight or obesity. mHealth telephone-based interventions and interventions delivered via a computer had no significant effect on the prevention of GDM. Perinatal women usually have different problems and specific needs during pregnancy. Personalized applications can focus on individual characteristics and tailor goals and actions to diverse populations [42]. In this way, medical staff can leverage app-based mHealth interventions to provide health care tailored to the specific needs of pregnant women with overweight or obesity. In contrast, the applications help medical staff remotely monitor real-time parameters related

to the health of pregnant women who are overweight or obese during pregnancy, observe whether they are adhering to a healthy and appropriate lifestyle, facilitate communication with them, and help them control their blood sugar and reduce the incidence of GDM [43].

Regarding the different providers of the interventions, the results of the subgroup analysis reported that mHealth interventions delivered by medical staff, including obstetricians and nurses, were effective in preventing GDM. It has been suggested that different models of care provided by various intervention providers may influence outcomes. mHealth interventions delivered by dietitians may focus on the food needs of women during pregnancy [44], whereas exercise physiologists are more concerned with physical activity during pregnancy [45]. Medical staff usually focus on the overall health of pregnant women, providing nutrition, exercise, weight management, and other comprehensive knowledge of pregnancy health care [46].

Subgroup analysis showed that mHealth interventions combining diet and exercise were effective in preventing gestational diabetes in women with overweight or obesity. Combined interventions with diet and exercise appeared to have a greater impact on GDM than interventions with diet or exercise alone, and the effects of diet and exercise on GDM were indistinguishable [47]. Type 2 diabetes has been proven to be preventable through combined diet, exercise, and weight loss interventions [48]. Another subgroup analysis found that mHealth-based lifestyle interventions were effective in preventing GDM in Asian populations. However, the pooled effects of interventions on the risk of GDM in predominantly White populations did not reach statistical significance. The disease burden of GDM varies by race because of multiple factors, including socioeconomic status, lifestyle, and culture, and the prevalence of GDM is significantly higher among Asian and Hispanic populations than among White populations [49]. The results of the subgroup analysis showed that, in studies using the IADPSG diagnostic criteria for GDM, mHealth-based lifestyle interventions had a significantly greater effect on GDM. The prevalence of GDM varies widely worldwide, at least in part because of a lack of consistency in screening and diagnostic criteria. Using lower glucose level thresholds as recommended by the IADPSG, significantly higher numbers of women with GDM were identified compared with using other diagnostic criteria [50].

Maternal and Neonatal Outcomes

Our meta-analyses revealed that the combined effect of mHealth interventions reduced the incidence of preterm birth, macrosomia, and excessive GWG. A review summarized the evidence regarding the influence of maternal diet before and during pregnancy on preterm birth. The results indicated that better maternal diet quality during pregnancy, characterized by a high intake of vegetables, fruits, whole grains, dairy products, and protein, played a significant role in reducing the risk of premature birth [51]. In addition, this study was in line with the findings of Fair and Soltani [52], who conducted a systematic review of the effectiveness of lifestyle interventions on weight gain in women with overweight or obesity during pregnancy and found that lifestyle interventions slightly reduced weight gain during pregnancy by 0.3 to 2.4 kg compared with standard care. There is no robust evidence that mHealth-based interventions are associated with a lower prevalence of postpartum hemorrhage, pregnancy-induced hypertension, cesarean sections, or any alteration in gestational age, consistent with a previous systematic review [53]. One reason for this nonsignificant effect may be the insufficient power of the combined sample size. Another possibility is the short duration of the interventions (median 18 wk), which might have been inadequate to affect some obstetric complications and neonatal outcomes. Further studies are needed to actively explore the associations between mHealth interventions and maternal and neonatal outcomes.

Strengths and Limitations

This systematic review and meta-analysis has several strengths. We combined MeSH terms and keywords covering pregnancy and mHealth to conduct a comprehensive search in 5 primary English electronic databases and 4 main Chinese electronic databases to minimize the possibility of publication bias. We used a robust 3-step search strategy to include databases containing published and unpublished RCTs. To minimize bias, the review methods were preregistered in accordance with the PRISMA statement. In addition, some of the studies included in this meta-analysis were conducted in Western, high-income countries, and some were conducted in Eastern, lower-income countries. Our meta-analyses provided an excellent synthesis of the responses of participants from different cultural backgrounds to mHealth interventions.

However, there were also some limitations to this review. First, we only retrieved studies published in English or Chinese owing to language limitations. Studies published in other languages were not included in the review, which may have led to some publication bias. Second, the methodological quality of the included studies was not optimal, and some studies had a risk of performance bias and detection bias. Third, the included studies varied in sample size, participant characteristics, components of the mHealth interventions, and intervention implementation methods, which may have led to high heterogeneity. Moreover, key variables such as the exact start of the intervention and intensity of physical activity were missing and incomplete in some studies and may have biased the pooled effects. Finally, not all the included studies reported the safety and cost-effectiveness of the mHealth interventions.

Conclusions

This systematic review and meta-analysis demonstrated that mHealth-based lifestyle interventions have a favorable impact on the prevention of GDM in pregnant women with overweight and obesity. mHealth interventions are a convenient and effective way to support pregnant women with overweight and obesity in out-of-hospital self-management in the context of rapid advances in IT and faster transmission speed. However, the potential of mHealth-based interventions for GDM needs to be further explored with better design and more rigorous large-scale RCTs.

Authors' Contributions

All authors made substantial contributions to the study. YH initiated the study. CH and QH conducted the data extraction and analyses. YH wrote the first draft of the manuscript. BL critically reviewed and revised the manuscript. All authors have read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOCX File , 28 KB - mhealth_v12i1e49373_app1.docx]

Multimedia Appendix 2
Search strategies.
[DOCX File , 22 KB - mhealth_v12i1e49373_app2.docx]

Multimedia Appendix 3
The details of the mobile health interventions.
[DOCX File , 23 KB - mhealth_v12i1e49373_app3.docx]

Multimedia Appendix 4
Funnel plot.
[DOCX File , 29 KB - mhealth_v12i1e49373_app4.docx]

Multimedia Appendix 5
Forest plots of subgroup analysis.
[DOCX File , 172 KB - mhealth_v12i1e49373_app5.docx]

Multimedia Appendix 6
Forest plots of secondary outcomes.
[DOCX File , 99 KB - mhealth_v12i1e49373_app6.docx]

References

1. Metzger BE. Summary and recommendations of the third international workshop-conference on gestational diabetes mellitus. Diabetes 1991 Dec;40 Suppl 2:197-201. [doi: 10.2337/diab.40.2.s197] [Medline: 1748259]
2. Wang H, Li N, Chivese T, Werfalli M, Sun H, Yuen L, IDF Diabetes Atlas Committee Hyperglycaemia in Pregnancy Special Interest Group. IDF diabetes atlas: estimation of global and regional gestational diabetes mellitus prevalence for 2021 by international association of diabetes in pregnancy study group's criteria. Diabetes Res Clin Pract 2022 Jan;183:109050. [doi: 10.1016/j.diabres.2021.109050] [Medline: 34883186]
3. Johns EC, Denison FC, Norman JE, Reynolds RM. Gestational diabetes mellitus: mechanisms, treatment, and complications. Trends Endocrinol Metab 2018 Nov;29(11):743-754. [doi: 10.1016/j.tem.2018.09.004] [Medline: 30297319]
4. Natamba BK, Namara AA, Nyirenda MJ. Burden, risk factors and maternal and offspring outcomes of gestational diabetes mellitus (GDM) in sub-Saharan Africa (SSA): a systematic review and meta-analysis. BMC Pregnancy Childbirth 2019 Nov 28;19(1):450 [FREE Full text] [doi: 10.1186/s12884-019-2593-z] [Medline: 31779584]
5. Feghali MN, Catov JM, Zantow E, Mission J, Caritis SN, Scifres CM. Timing of gestational weight gain and adverse perinatal outcomes in overweight and obese women. Obstet Gynecol 2019 May;133(5):962-970. [doi: 10.1097/AOG.0000000000003234] [Medline: 30969214]
6. Mengesha Kassie A, Beletew Abate B, Wudu Kassaw M. Education and prevalence of overweight and obesity among reproductive age group women in Ethiopia: analysis of the 2016 Ethiopian demographic and health survey data. BMC Public Health 2020 Jul 31;20(1):1189 [FREE Full text] [doi: 10.1186/s12889-020-08941-w] [Medline: 32736617]
7. Poston L, Caleyachetty R, Cnattingius S, Corvalán C, Uauy R, Herring S, et al. Preconceptional and maternal obesity: epidemiology and health consequences. Lancet Diabetes Endocrinol 2016 Dec;4(12):1025-1036. [doi: 10.1016/S2213-8587(16)30217-0] [Medline: 27743975]

8. West S, Ollila MM, Franks S, Piltonen T, Jokelainen J, Nevalainen J, et al. Overweight, obesity and hyperandrogenemia are associated with gestational diabetes mellitus: a follow-up cohort study. Acta Obstet Gynecol Scand 2020 Oct;99(10):1311-1319 [FREE Full text] [doi: 10.1111/aogs.13883] [Medline: 32333801]
9. Maas VY, Poels M, Lamain-de Ruiter M, Kwee A, Bekker MN, Franx A, et al. Associations between periconceptional lifestyle behaviours and adverse pregnancy outcomes. BMC Pregnancy Childbirth 2021 Jul 07;21(1):492 [FREE Full text] [doi: 10.1186/s12884-021-03935-x] [Medline: 34233654]
10. Rasmussen L, Poulsen CW, Kampmann U, Smedegaard SB, Ovesen PG, Fuglsang J. Diet and healthy lifestyle in the management of gestational diabetes mellitus. Nutrients 2020 Oct 06;12(10):3050 [FREE Full text] [doi: 10.3390/nu12103050] [Medline: 33036170]
11. Davenport MH, Ruchat SM, Poitras VJ, Jaramillo Garcia A, Gray CE, Barrowman N, et al. Prenatal exercise for the prevention of gestational diabetes mellitus and hypertensive disorders of pregnancy: a systematic review and meta-analysis. Br J Sports Med 2018 Nov;52(21):1367-1375. [doi: 10.1136/bjsports-2018-099355] [Medline: 30337463]
12. Mijatovic-Vukas J, Capling L, Cheng S, Stamatakis E, Louie J, Cheung NW, et al. Associations of diet and physical activity with risk for gestational diabetes mellitus: a systematic review and meta-analysis. Nutrients 2018 May 30;10(6):698 [FREE Full text] [doi: 10.3390/nu10060698] [Medline: 29849003]
13. Sherifali D, Nerenberg KA, Wilson S, Semeniuk K, Ali MU, Redman LM, et al. The effectiveness of eHealth technologies on weight management in pregnant and postpartum women: systematic review and meta-analysis. J Med Internet Res 2017 Oct 13;19(10):e337. [doi: 10.2196/jmir.8006] [Medline: 29030327]
14. Gu D, Li T, Wang X, Yang X, Yu Z. Visualizing the intellectual structure and evolution of electronic health and telemedicine research. Int J Med Inform 2019 Oct;130:103947. [doi: 10.1016/j.ijmedinf.2019.08.007] [Medline: 31450080]
15. mHealth: new horizons for health through mobile technologies. World Health Organization. 2011. URL: https://www.afro.who.int/publications/mhealth-new-horizons-health-through-mobile-technologie [accessed 2023-12-10]
16. Reynolds RM. Telehealth in pregnancy. Lancet Diabetes Endocrinol 2020 Jun;8(6):459-461 [FREE Full text] [doi: 10.1016/S2213-8587(20)30158-3] [Medline: 32445730]
17. Cotie LM, Prince SA, Elliott CG, Ziss MC, McDonnell LA, Mullen KA, et al. The effectiveness of eHealth interventions on physical activity and measures of obesity among working-age women: a systematic review and meta-analysis. Obes Rev 2018 Oct;19(10):1340-1358. [doi: 10.1111/obr.12700] [Medline: 30156044]
18. Hutchesson MJ, Rollo ME, Krukowski R, Ells L, Harvey J, Morgan PJ, et al. eHealth interventions for the prevention and treatment of overweight and obesity in adults: a systematic review with meta-analysis. Obes Rev 2015 May;16(5):376-392. [doi: 10.1111/obr.12268] [Medline: 25753009]
19. Frid G, Bogaert K, Chen KT. Mobile health apps for pregnant women: systematic search, evaluation, and analysis of features. J Med Internet Res 2021 Oct 18;23(10):e25667 [FREE Full text] [doi: 10.2196/25667] [Medline: 34524100]
20. Hayman M, Alfrey KL, Cannon S, Alley S, Rebar AL, Williams S, et al. Quality, features, and presence of behavior change techniques in mobile apps designed to improve physical activity in pregnant women: systematic search and content analysis. JMIR Mhealth Uhealth 2021 Apr 07;9(4):e23649 [FREE Full text] [doi: 10.2196/23649] [Medline: 33825693]
21. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
22. Chen H. Clinical effectiveness analysis of comprehensive nursing measures for preventing gestational diabetes mellitus of women with overweight or obesity before pregnancy. Chin J Fam Plan 2017;25(04):250-253. [doi: 10.3969/j.issn.1004-8189.2017.04.008]
23. Li Y. Evaluation of preventive effect of individualized management on gestational diabetes mellitus in overweight and obese pregnant women: based on behavioral change theory. Chin Rural Health Serv Adm 2018;38(04):509-512.
24. Tang L, Jian C, Yaping W. Application of dietary evaluation WeChat platform in obese pregnant women's nutrition intervention. Mod Med J 2019;47(05):517-521. [doi: 10.3969/j.issn.1671-7562.2019.05.006]
25. Cao L. Analysis of clinical effect of comprehensive nursing measures on prevention of gestational diabetes mellitus in overweight and obese pregnant women before pregnancy. Guide Chin Med 2020;18(10):12-14. [doi: 10.15912/j.cnki.gocm.2020.10.005]
26. Wu X, Guang H. Effect of individualized weight management during pregnancy based on Wechat platform on weight gain and pregnancy outcome of obese pregnant women. Clin Med Res Pract 2020;5(35):65-64. [doi: 10.19347/j.cnki.2096-1413.202035012]
27. Zhou H, Chenru, Wangqin Z, Huimei D, Xiaocong S. Application of pre-pregnancy eugenic guidance combined with pregnancy diet plan in prevention of gestational diabetes mellitus in overweight and obese pregnant women before pregnancy. Matern Child Health Care Chin 2021;36(08):1717-1720. [doi: 10.19829/j.zgfybj.issn.1001-4411.2021.08.004]
28. Kang L, Sung A. Effects of wechat dietary education management on the incidence of gestational diabetes in obese pregnant women. Clin Med 2021;41(07):66-68. [doi: 10.19528/j.issn.1003-3548.2021.07.024]
29. Ding B, Gou B, Guan H, Wang J, Bi Y, Hong Z. WeChat-assisted dietary and exercise intervention for prevention of gestational diabetes mellitus in overweight/obese pregnant women: a two-arm randomized clinical trial. Arch Gynecol Obstet 2021 Sep;304(3):609-618. [doi: 10.1007/s00404-021-05984-1] [Medline: 33570656]

30. Ferrara A, Hedderson MM, Brown SD, Ehrlich SF, Tsai A, Feng J, et al. A telehealth lifestyle intervention to reduce excess gestational weight gain in pregnant women with overweight or obesity (GLOW): a randomised, parallel-group, controlled trial. Lancet Diabetes Endocrinol 2020 Jun;8(6):490-500 [FREE Full text] [doi: 10.1016/S2213-8587(20)30107-8] [Medline: 32445736]
31. Liu J, Wilcox S, Wingard E, Turner-McGrievy G, Hutto B, Burgis J. A behavioral lifestyle intervention to limit gestational weight gain in pregnant women with overweight and obesity. Obesity (Silver Spring) 2021 Apr;29(4):672-680 [FREE Full text] [doi: 10.1002/oby.23119] [Medline: 33619910]
32. Dodd JM, Turnbull D, McPhee AJ, Deussen AR, Grivell RM, Yelland LN, LIMIT Randomised Trial Group. Antenatal lifestyle advice for women who are overweight or obese: LIMIT randomised trial. BMJ 2014 Feb 10;348:g1285 [FREE Full text] [doi: 10.1136/bmj.g1285] [Medline: 24513442]
33. Simmons D, Devlieger R, van Assche A, Jans G, Galjaard S, Corcoy R, et al. Effect of physical activity and/or healthy eating on GDM risk: the DALI lifestyle study. J Clin Endocrinol Metab 2017 Mar 01;102(3):903-913 [FREE Full text] [doi: 10.1210/jc.2016-3455] [Medline: 27935767]
34. Seneviratne S, Jiang Y, Derraik J, McCowan L, Parry G, Biggs J, et al. Effects of antenatal exercise in overweight and obese pregnant women on maternal and perinatal outcomes: a randomised controlled trial. BJOG 2016 Mar 06;123(4):588-597. [doi: 10.1111/1471-0528.13738] [Medline: 26542419]
35. Poston L, Bell R, Croker H, Flynn AC, Godfrey KM, Goff L, UPBEAT Trial Consortium. Effect of a behavioural intervention in obese pregnant women (the UPBEAT study): a multicentre, randomised controlled trial. Lancet Diabetes Endocrinol 2015 Oct;3(10):767-777 [FREE Full text] [doi: 10.1016/S2213-8587(15)00227-2] [Medline: 26165396]
36. Kennelly MA, Ainscough K, Lindsay KL, O'Sullivan E, Gibney ER, McCarthy M, et al. Pregnancy exercise and nutrition with smartphone application support: a randomized controlled trial. Obstet Gynecol 2018 May;131(5):818-826. [doi: 10.1097/AOG.0000000000002582] [Medline: 29630009]
37. Sagedal L, Øverby NC, Bere E, Torstveit M, Lohne-Seiler H, Småstuen M, et al. Lifestyle intervention to limit gestational weight gain: the Norwegian Fit for Delivery randomised controlled trial. BJOG 2017 Jan 14;124(1):97-109. [doi: 10.1111/1471-0528.13862] [Medline: 26768233]
38. Oteng-Ntim E, Varma R, Croker H, Poston L, Doyle P. Lifestyle interventions for overweight and obese pregnant women to improve pregnancy outcome: systematic review and meta-analysis. BMC Med 2012 May 10;10:47 [FREE Full text] [doi: 10.1186/1741-7015-10-47] [Medline: 22574949]
39. Juan J, Yang H. Prevalence, prevention, and lifestyle intervention of gestational diabetes mellitus in China. Int J Environ Res Public Health 2020 Dec 18;17(24):9517 [FREE Full text] [doi: 10.3390/ijerph17249517] [Medline: 33353136]
40. International Weight Management in Pregnancy (i-WIP) Collaborative Group. Effect of diet and physical activity based interventions in pregnancy on gestational weight gain and pregnancy outcomes: meta-analysis of individual participant data from randomised trials. BMJ 2017 Jul 19;358:j3119 [FREE Full text] [doi: 10.1136/bmj.j3119] [Medline: 28724518]
41. Farahani B, Firouzi F, Chakrabarty K. Healthcare IoT. In: Firouzi F, Chakrabarty K, Nassif S, editors. Intelligent Internet of Things: From Device to Fog and Cloud. Cham, Switzerland: Springer; 2020:515-545.
42. Vashist SK, Schneider EM, Luong JH. Commercial smartphone-based devices and smart applications for personalized healthcare monitoring and management. Diagnostics (Basel) 2014 Aug 18;4(3):104-128 [FREE Full text] [doi: 10.3390/diagnostics4030104] [Medline: 26852680]
43. Swayamsiddha S, Mohanty C. Application of cognitive internet of medical things for COVID-19 pandemic. Diabetes Metab Syndr 2020 Sep;14(5):911-915 [FREE Full text] [doi: 10.1016/j.dsx.2020.06.014] [Medline: 32570016]
44. Super S, Beulen YH, Koelen MA, Wagemakers A. Opportunities for dietitians to promote a healthy dietary intake in pregnant women with a low socio-economic status within antenatal care practices in the Netherlands: a qualitative study. J Health Popul Nutr 2021 Jul 31;40(1):35 [FREE Full text] [doi: 10.1186/s41043-021-00260-z] [Medline: 34332642]
45. Lim S, Liang X, Hill B, Teede H, Moran LJ, O'Reilly S. A systematic review and meta-analysis of intervention characteristics in postpartum weight management using the TIDieR framework: a summary of evidence to inform implementation. Obes Rev 2019 Jul 03;20(7):1045-1056. [doi: 10.1111/obr.12846] [Medline: 30942550]
46. Escañuela Sánchez T, Meaney S, O'Connor C, Linehan L, O'Donoghue K, Byrne M, et al. Facilitators and barriers influencing weight management behaviours during pregnancy: a meta-synthesis of qualitative research. BMC Pregnancy Childbirth 2022 Sep 05;22(1):682 [FREE Full text] [doi: 10.1186/s12884-022-04929-z] [Medline: 36064379]
47. Shepherd E, Gomersall JC, Tieu J, Han S, Crowther CA, Middleton P. Combined diet and exercise interventions for preventing gestational diabetes mellitus. Cochrane Database Syst Rev 2017 Nov 13;11(11):CD010443 [FREE Full text] [doi: 10.1002/14651858.CD010443.pub3] [Medline: 29129039]
48. Magkos F, Hjorth MF, Astrup A. Diet and exercise in the prevention and treatment of type 2 diabetes mellitus. Nat Rev Endocrinol 2020 Oct 20;16(10):545-555. [doi: 10.1038/s41574-020-0381-5] [Medline: 32690918]
49. Pu J, Zhao B, Wang EJ, Nimbal V, Osmundson S, Kunz L, et al. Racial/ethnic differences in gestational diabetes prevalence and contribution of common risk factors. Paediatr Perinat Epidemiol 2015 Sep;29(5):436-443 [FREE Full text] [doi: 10.1111/ppe.12209] [Medline: 26201385]

50. Behboudi-Gandevani S, Amiri M, Bidhendi Yarandi R, Ramezani Tehrani F. The impact of diagnostic criteria for gestational diabetes on its prevalence: a systematic review and meta-analysis. Diabetol Metab Syndr 2019;11:11 [FREE Full text] [doi: 10.1186/s13098-019-0406-1] [Medline: 30733833]
51. Gete DG, Waller M, Mishra GD. Effects of maternal diets on preterm birth and low birth weight: a systematic review. Br J Nutr 2020 Feb 28;123(4):446-461. [doi: 10.1017/S0007114519002897] [Medline: 31711550]
52. Fair F, Soltani H. A meta-review of systematic reviews of lifestyle interventions for reducing gestational weight gain in women with overweight or obesity. Obes Rev 2021 May;22(5):e13199 [FREE Full text] [doi: 10.1111/obr.13199] [Medline: 33459493]
53. Lau Y, Klainin-Yobas P, Htun TP, Wong SN, Tan KL, Ho-Lim ST, et al. Electronic-based lifestyle interventions in overweight or obese perinatal women: a systematic review and meta-analysis. Obes Rev 2017 Sep;18(9):1071-1087. [doi: 10.1111/obr.12557] [Medline: 28544551]

Abbreviations

GDM: gestational diabetes mellitus
GWG: gestational weight gain
IADPSG: International Association of Diabetes and Pregnancy Study Groups
MeSH: Medical Subject Headings
mHealth: mobile health
OR: odds ratio
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial

Edited by L Buis; submitted 26.05.23; peer-reviewed by M Jiang, J Birks; comments to author 29.09.23; revised version received 22.11.23; accepted 28.11.23; published 17.01.24.

Please cite as:
He Y, Huang C, He Q, Liao S, Luo B
Effects of mHealth-Based Lifestyle Interventions on Gestational Diabetes Mellitus in Pregnant Women With Overweight and Obesity: Systematic Review and Meta-Analysis
JMIR Mhealth Uhealth 2024;12:e49373
URL: https://mhealth.jmir.org/2024/1/e49373
doi:10.2196/49373
PMID:38231555

Review

Effectiveness of Telecare Interventions on Depression Symptoms Among Older Adults: Systematic Review and Meta-Analysis

Man Wu[1], BSN; Chaoyang Li[1], BSN; Ting Hu[1], BSN; Xueyang Zhao[1], BSN; Guiyuan Qiao[1], BSN; Xiaolian Gao[1], BSN; Xinhong Zhu[1], PhD; Fen Yang[1,2], PhD

[1]School of Nursing, Hubei University of Chinese Medicine, Wuhan, China

[2]Hubei Shizhen Laboratory, Wuhan, China

Corresponding Author:
Fen Yang, PhD
School of Nursing
Hubei University of Chinese Medicine
North District, No 16, Huangjiahu West Road
Hongshan District
Wuhan, 430065
China
Phone: 86 18062003261
Email: fenyang@hbtcm.edu.cn

Abstract

Background: Depression is the most common psychiatric disorder among older adults. Despite the effectiveness of pharmacological and psychological therapies, many patients with late-life depression (LLD) are unable to access timely treatment. Telecare has been shown to be effective in addressing patients' psychosocial issues, while its effectiveness in serving patients with LLD remains unclear.

Objective: This study aimed to evaluate the effectiveness of telecare in reducing depression and anxiety symptoms and improving quality of life (QoL) in patients with LLD.

Methods: Databases including the Cochrane Library, Web of Science, PubMed, Embase, and EBSCO were searched for randomized controlled trials (RCTs) evaluating the effectiveness of telecare for LLD from database establishment to December 28, 2022.

Results: A total of 12 RCTs involving 1663 participants were identified in this study. The meta-analysis showed that (1) telecare significantly reduced depressive symptoms in patients with LLD compared to those in usual care (UC; standardized mean difference [SMD]=–0.46, 95% CI –0.53 to –0.38; *P*<.001), with the best improvement observed within 3 months of intervention (SMD=–0.72, 95% CI –1.16 to –0.28; *P*<.001); (2) other scales appeared more effective than the Patient Health Questionnaire-9 for LLD in telecare interventions (SMD=–0.65, 95% CI –0.96 to –0.35; *P*<.001); (3) telecare was more effective than telephone-based interventions for remote monitoring of LLD (SMD=–1.13, 95% CI –1.51 to –0.76; *P*<.001); (4) the reduction of depressive symptoms was more pronounced in patients with LLD with chronic conditions (SMD=–0.67, 95% CI –0.89 to –0.44; *P*<.001); (5) telecare was more effective for LLD in Europe and the Americas than in other regions (SMD=–0.73, 95% CI –0.99 to –0.47; *P*<.001); (6) telecare significantly reduced anxiety symptoms in patients with LLD (SMD=–0.53, 95% CI –0.73 to –0.33; *P*=.02); and (7) there was no significant improvement in the psychological components of QoL in patients with LLD compared to those receiving UC (SMD=0.30, 95% CI 0.18-0.43; *P*=.80).

Conclusions: Telecare is a promising modality of care for treatment, which can alleviate depression and anxiety symptoms in patients with LLD. Continued in-depth research into the effectiveness of telecare in treating depression could better identify where older patients would benefit from this intervention.

(JMIR Mhealth Uhealth 2024;12:e50787) doi:10.2196/50787

KEYWORDS

telecare; depression; anxiety; quality of life; older adults; meta-analysis

Introduction

Statistics show that the world's population older than 60 years will double between 2015 and 2050, increasing from 12.0% to 22.0% [1]. With the rapid growth of the older population, late-life depression (LLD) has gradually emerged as a hot topic in the field of geriatric medical research. LLD refers to depressive disorders occurring in adults older than 60 years [2,3]. Research findings indicate a global prevalence of LLD of 28.4% [4], which could potentially be higher among individuals with concurrent physical ailments. As a geriatric syndrome with multifactorial etiology, LLD is highly associated with medical problems that pervade later life, including diabetes, hypertension, and dementia [2,5]. LLD is often chronic or recurrent and is associated with functional impairment, diminished health-related quality of life (QoL), and impaired social-psychological functioning [3,6]. A study confirmed that health care costs for patients with LLD were 43.0% to 52.0% higher for outpatient services and 47.0% to 51.0% higher when outpatient and inpatient services were combined, compared to those for individuals without LLD [7].

Despite its high prevalence and severe adverse outcomes, LLD is often overlooked and inadequately treated due to other complications resulting from aging-related issues. Psychopharmacotherapy and psychotherapy have been demonstrated to be effective for people with depression [6]; however, these treatments still have limitations, such as medical side effects and poor treatment adherence [8,9]. Due to mobility issues, geographic isolation, stigma associated with mental illness, and negative beliefs about treatment, older adults have limited access to health care or may be unwilling to seek help from health care institutions [10-12]. Additionally, underuse of professional mental health services, including low detection rates by health care providers and the lack of awareness among older patients regarding the severity of their condition [13,14], is also one of the factors that impede the treatment of LLD. Limited by these factors, only a minority of older adults receive appropriate treatment for depression. Therefore, there is an urgent need to study the clinical effectiveness of alternative therapies for depression, which are more socially acceptable and easily available.

In recent years, there has been increasing attention toward using telecare to support the management and well-being of mental health [15]. Telecare refers to the delivery of health care directly to users, typically in their own homes, supported by information and communication technologies such as telephone, videoconferencing, and applications [16,17]. Health care professionals can remotely provide consultation, assessment, and intervention services to patients [18]. These services include, but are not limited to, lifestyle monitoring, remote monitoring of vital signs for diagnosis, as well as long-distance assessment and education. The benefits of telecare are evident. Evidence suggests that as a promising strategy, telecare services can serve as a medium to overcome certain barriers, thereby enhancing mental health care and increasing opportunities to access evidence-based care under different conditions [19]. Particularly, telecare benefits older adults who are socially isolated or physically frail due to illness, disability, or other familial roles [17,20]. Currently, telecare has been widely used in the management of various chronic conditions among older adults, such as diabetes, hypertension, Parkinson disease, etc, yielding positive outcomes [21-23]. Depression is a commonly observed chronic condition among older adults, closely associated with an approximate 50% increase in chronic disease-related health care costs [24]. Given the significant impact of LLD on patients' QoL and its potential consequences on decreased productivity or suicide, ensuring continuity of care is imperative. Telecare has been proposed as an effective alternative to help bridge this treatment problem. Considering the complexity and severity of LLD, it is necessary to further explore whether telecare is effective in improving health outcomes for patients with LLD.

Previous reviews have assessed the evidence related to the use of telecare for managing mental health issues [11,25]. In the field of psychiatry, telecare has been found to significantly impact mental health outcomes in older adults, including reducing emergency visits and hospitalizations, as well as improving cognitive function [11]. However, the efficacy of telecare for depression is inconsistent. Some studies suggest the effectiveness of telecare in reducing symptoms of depression [11,26], while others indicate that the impact of telecare on improving depressive symptoms is limited, even yielding contradictory results [27,28]. Previous meta-analyses examining the effectiveness of telecare on depression have mostly focused on adult populations [25-27]. However, compared to other age groups, LLD is considered to be different [14]. Differences in study design, intervention methods, and treatment intensity may contribute to varying clinical outcomes in telecare treatments for LLD. Despite recent meta-analyses demonstrating significant efficacy of telemedicine in alleviating depressive symptoms among older adults, the evaluation of its evidence remains limited [29]. Due to inherent heterogeneity in inclusion criteria, interpretation of these results should be approached cautiously. The severe clinical outcomes and interfering factors often pose significant challenges in the treatment of LLD. Determining whether telecare management is effective for LLD is critical. It is unclear how effective telecare is in improving depression, anxiety symptoms, and QoL in patients with LLD. Therefore, this systematic review and meta-analysis explored the efficacy of telecare for LLD.

Methods

This systematic review was conducted in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines (Multimedia Appendix 1) [30].

Search Strategy

We conducted searches in Cochrane Library, Web of Science, PubMed, Embase, and EBSCO for randomized controlled trials (RCTs) published from the inception of the databases up to December 28, 2022, without any language restriction. MeSH (Medical Subject Headings) and free search terms were both used in the literature search. The search terms included "cell Phones," "telemedicine," "smartphone," "mobile applications," "mobile phone*," "telephone*," "telehealth," "tele-healthcare," "electronic health*," "application*," "m-health," "messaging," "depression," "depressive disorder," "depress*,"

"Major depression," "sadness," "late-life depression," "LLD," "aged," "elder*," "geriatric," "senior people," "RCTs," etc. All titles, keywords, and abstracts have been reviewed in accordance with our search criteria. In this study, these research articles were exclusively published in English, focusing on telecare interventions for LLD. The specific search strategy is shown in Multimedia Appendix 2.

Study Selection and Data Exclusion

The inclusion criteria were the following: (1) studies were RCTs reported in full text with their title and abstract; (2) the average age of the study population was at least 60 years; (3) participants were diagnosed with depression in accordance with any established diagnostic criteria or with a score above a cutoff of any established depression rating scale at baseline; (4) the studies compare telecare (mobile phone, telephone, app, video, etc) participants with the control group receiving usual care (UC; routine, offline, or standard care); and (5) any health care professional providing care (ie, psychiatrists, family physicians, nurses, psychologists, etc).

Exclusion criteria were the following: (1) patients with manic or psychotic episodes or symptoms; (2) studies not related to the objective of this review and insufficient data, such as failure to report depression scale scores; and (3) books and studies without full text and studies in the format of abstracts of conference papers.

Data Extraction

Two authors independently reviewed all the databases, with specific search strategies for the relevant articles (MW and CYL). The software EndNote X9 (Clarivate) was used to import all the references and remove duplicates. After removing duplicates, the relevance of the title and abstract of the articles was evaluated. Any disagreements were discussed until a consensus was reached. After screening the title and abstract, the articles were selected for the next step of a full-text review. The 2 authors screened the full-text articles independently (MW and CYL). Finally, eligible articles included in the study were processed based on inclusion and exclusion criteria. Any discrepancies that arose during the assessment were resolved by a third reviewer (FY). Two authors independently extracted data from the included studies and entered them into a predesigned data extraction form. Data extracted for this study included the following: first author, year of publication, country, sample size, mean age, intervention approach, duration, presence or absence of comorbid chronic conditions, depression degree, and outcome measurement tools (Table 1).

Table 1. Basic characteristics of the included studies (N=12; all are randomized controlled trials).

First author (year); country	Sample size, N (TC[a]/UC[b])	Age (years), mean (SD)	Duration	Comorbid chronic diseases	Depression degree	Outcomes
Rollman (2009) [31]; United States	302 (150/152)	TC: 64 (10.8); UC: 64 (11.2)	Baseline, 8 months	Yes	Moderate	HAM-D[c] and SF-36[d]
Aburizik (2013) [32]; United States	52 (29/23)	TC: 66.4 (7.9); UC: 64.1 (10.5)	Baseline, 10 weeks	Yes	Mild	PHQ-9[e] and BDI[f]
Lee (2014) [23]; Korea	25 (12/13)	TC: 66.7 (7.9); UC: 65.4 (8.6)	Baseline, 6 months	Yes	Mild, moderate	CES-D[g]
Villani (2014) [33]; Italy	80 (40/40)	TC: 71 (4); UC: 73 (5)	Baseline, 12 months	Yes	Moderate, severe	PHQ-9 and STAI-6[h]
Pickett (2014) [34]; United States	124 (60/64)	TC: 69.1 (10.9); UC: 68.6 (10.7)	Baseline, 12 weeks	No	Mild	PHQ-9
O'Neil (2014) [35]; Australia	121 (61/60)	TC: 61.0 (10.2); UC: 58.9 (10.7)	Baseline, 6 months	Yes	Mild, moderate	PHQ-9, CDS[i], and SF-12[j]
Gellis (2014) [36]; United States	94 (46/48)	TC: 80.1 (7.8); UC: 78.3 (6.9)	Baseline, 3 months, and 6 months	Yes	Mild, moderate	PHQ-9, HAM-D, and SF-12
Yang (2019) [37]; China	212 (107/105)	TC: 61.25 (8.60); UC: 60.85 (10.80)	Baseline, 12 months	Yes	Mild, moderate	HADS-D[k] and SDS[l]
Naik (2019) [21]; United States	225 (136/89)	61.9 (8.3)	Baseline, 6 months, and 12 months	Yes	Moderate	PHQ-9
Dobkin (2020) [22]; United States	72 (37/35)	TC: 65.62 (9.76); UC: 64.80 (9.62)	Baseline, 3 months, and 6 months	Yes	Moderate	HAM-D, BDI, HAM-A[m], and SF-36
Almeida (2021) [38]; Australia	200 (79/121)	≥65	Baseline, and 52 weeks	No	Mild, moderate	PHQ-9, GAD-7[n], and SF-12
Koehler (2021) [39]; Germany	156 (79/77)	TC: 68.30 (9.13); UC: 64.34 (11.35)	Baseline, 12 months	Yes	Moderate	PHQ-9 and SF-36

[a]TC: telecare.

[b]UC: usual care.

[c]HAM-D: Hamilton Depression Rating Scale.

[d]SF-36: 36-Item Short Form Survey.

[e]PHQ-9: Patient Health Questionnaire-9.

[f]BDI: Beck Depression Inventory.

[g]CES-D: Center for Epidemiological Survey, Depression Scale.

[h]STAI-6: Spielberger's State Trait Anxiety Inventory.

[i]CDS: Cardiac Depression Scale.

[j]SF-12: 12-Item Short Form Survey.

[k]HADS-D: Hospital Anxiety and Depression Scale.

[l]SDS: Zung Self-Rating Depression Scale.

[m]HAM-A: Hamilton Anxiety Rating Scale.

[n]GAD-7: 7-item Generalized Anxiety Disorder Scale.

Quality Assessment

Two authors (MW and CYL) independently assessed the quality of the studies using the Cochrane Risk of Bias tool [40]. The assessment tool included 7 items (random sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessment, incomplete outcome data, selective reporting, and other bias), and authors judged each item individually as "low risk," "high risk," and "unclear risk." The study was considered to be of high quality with a low risk score for at least 4 domains, of which 3 key areas had to be included (random sequence generation, allocation concealment, and incomplete outcome data). Consensus was reached by 2 authors (MW and CYL) through discussion with a third evaluator (FY).

Statistical Analysis

Data were analyzed using Stata (version 16.0; StataCorp) and Review Manager (version 5.3; The Cochrane Collaboration). Intervention effects were estimated by calculating Cohen *d* standardized mean differences (SMDs) and 95% CIs [41]. All studies reported outcomes as continuous data. The Cochran Q

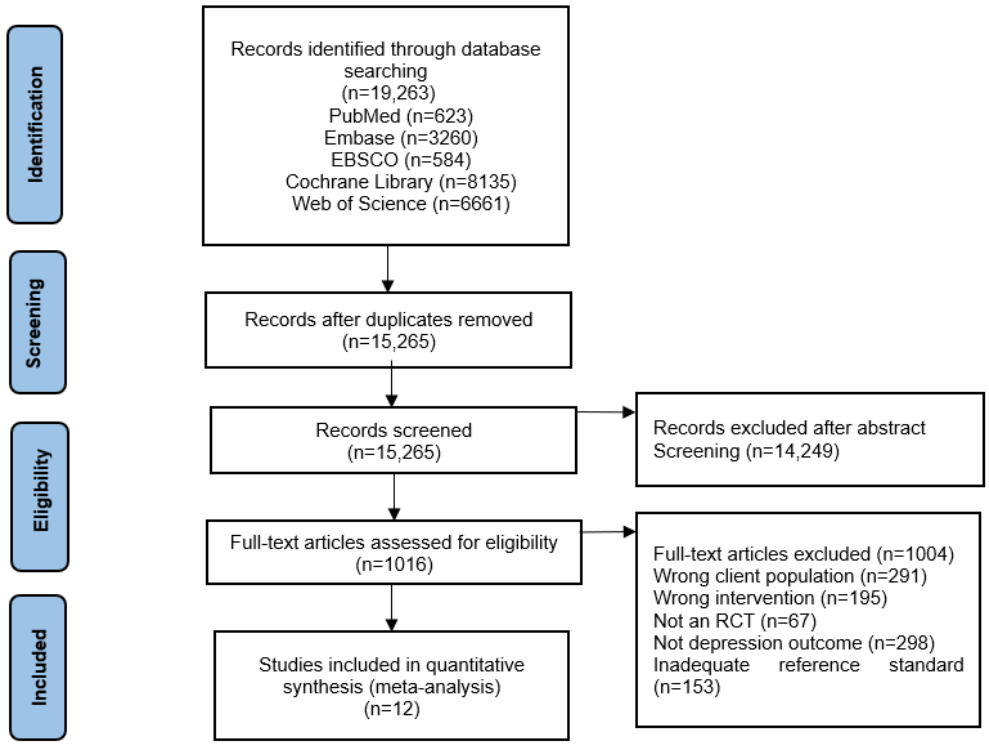
statistic and I^2 statistic were used to assess the statistical heterogeneity between selected studies. Random-effects models were used when study heterogeneity was high (P<.10; I^2>50%); otherwise, a fixed-effects model would be used. When heterogeneity identified across studies was high, we further performed subgroup analyses to explore possible explanations for heterogeneity. Publication bias was measured using a funnel plot and Egger linear regression analysis, and P<.05 on the Egger test indicated statistically significant publication bias [42].

Results

Literature Search

The database search yielded 15,265 articles, of which 14,249 publications were excluded. A total of 1016 full-text articles were assessed for eligibility. Finally, only 12 studies were eligible for inclusion in this meta-analysis [21-23,31-39], all of which were RCTs published between 2009 and 2021. The PRISMA flow diagram is shown in Figure 1.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram. RCT: randomized controlled trial.



Risks of Bias and Quality Assessment

Overall, the quality of the included studies was moderate, of which 5 (41.7%) were of high quality. These studies show that the main bias in the blinding of participants and personnel may be caused by the nature of the intervention measures. All 12 articles reported adequate random sequence generation and, therefore, had a low risk of bias in this regard. In addition, 5 studies reported allocation concealment, which is a low risk of bias. As for detection bias, the assessors were blinded in 7 studies, the presence of blinding was unclear in 3 studies, and 2 studies were not blinded. The risks of study attrition bias and reporting bias were both low. Other risks of bias were also low but were unclear in 1 study. The specific risk of bias and quality assessment results are shown in Figures 2 and 3 [21-23,31-39].

Figure 2. Overall risk of each type of bias.

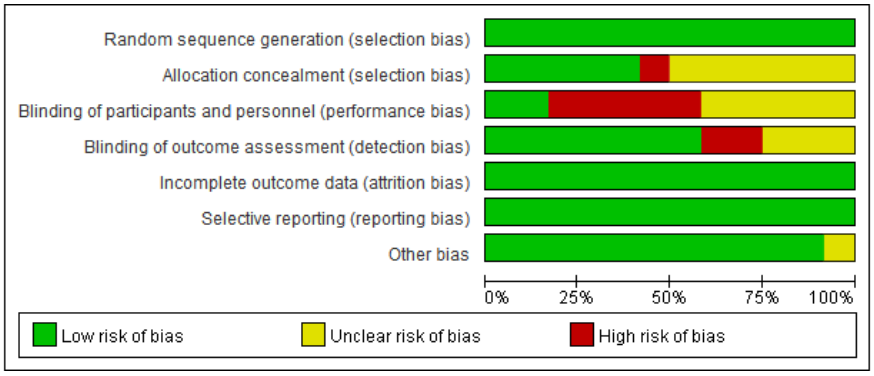
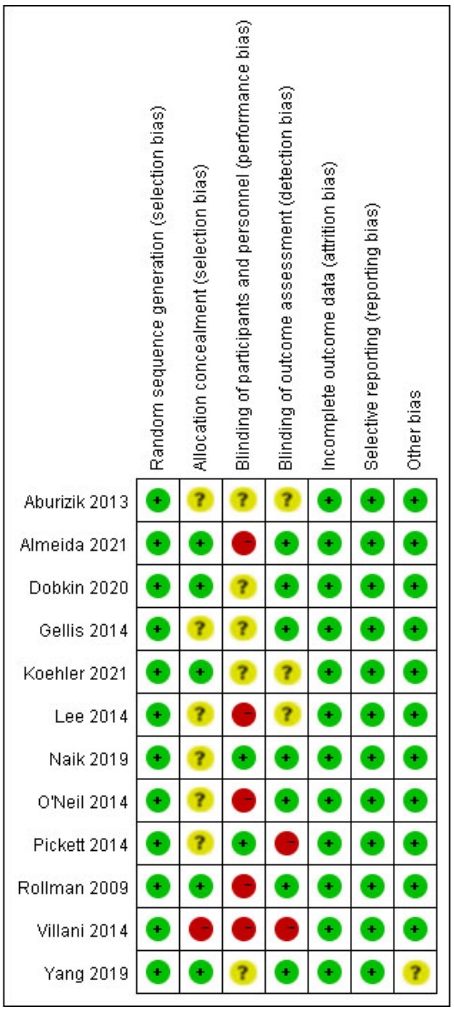


Figure 3. Risk of bias in each study.



Study and Patient Characteristics

The characteristics of the studies included are summarized in Table 1. A total of 1663 patients with LLD were involved, with an average age of over 60 years in each group. The sample size ranged from 25 [23] to 302 [31] participants. Studies were carried out across 6 countries, including the United States (n=6) [21,22,31,32,34,36], Korea (n=1) [23], Italy (n=1) [33], Australia (n=2) [35,38], China (n=1) [37], and Germany (n=1) [39]. Nine of these used telephone-based interventions, while the remaining studies used remote monitoring systems. Durations ranged from 10 weeks to 52 weeks. Depression, anxiety symptoms, and QoL were substantial influencing factors of treatment for older adults. Therefore, our primary outcome of interest was depression, and secondary outcomes were anxiety symptoms and QoL. Depression was evaluated using the Hamilton Depression Rating Scale, Patient Health Questionnaire-9 (PHQ-9), Beck Depression Inventory, Center for Epidemiological Survey, Depression Scale, Cardiac Depression Scale, Hospital Anxiety and Depression Scale, and Zung Self-Rating Depression Scale. Anxiety symptoms were assessed using Spielberger's State Trait Anxiety Inventory, Hamilton Anxiety Rating Scale, and the 7-item Generalized Anxiety Disorder Scale. QoL was assessed using the 12-Item Short Form Survey and the 36-Item Short Form Survey. A higher score on the scales indicated better QoL and

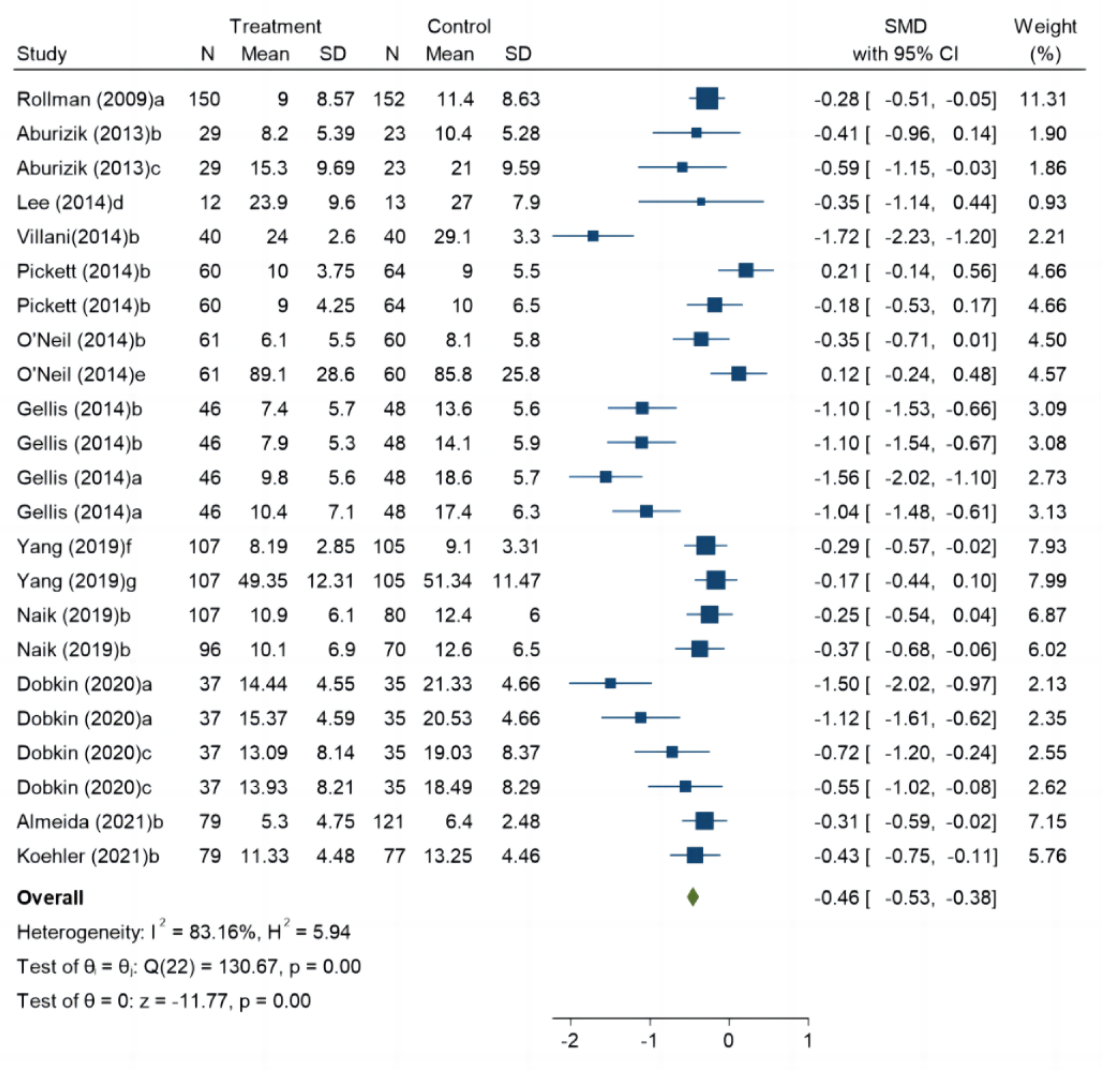
greater severity of depression and anxiety symptoms. The specific interventions are available in Multimedia Appendix 3.

Depression Symptoms

A total of 12 RCTs involving 1663 participants were identified in this meta-analysis to calculate the effectiveness of telecare on depression, anxiety symptoms, and QoL in patients with LLD.

To compare the effects of telecare and UC in improving LLD, we included data from 12 of these studies. Our results show that telecare significantly reduced depressive symptoms in patients with LLD compared to those in UC (SMD=–0.46, 95% CI –0.53 to –0.38; *P*<.001). Fixed-effects model analysis revealed significant heterogeneity among the 12 included studies (I^2=83.16%; *P*<.001; Figure 4) [21-23,31-39].

Figure 4. Forest plot for primary outcomes: depression. a: Hamilton Depression Rating Scale; b: Patient Health Questionnaire-9; c: Beck Depression Inventory; d: Center for Epidemiological Survey, Depression Scale; e: Cardiac Depression Scale; f: Hospital Anxiety and Depression Scale; g: Zung Self-Rating Depression Scale.



To address high heterogeneity, we performed subgroup analyses grouped by the type of scale (PHQ-9 or others), duration time (≤3 months or >3 months), device type (telephone-based or remote monitoring system), comorbid chronic diseases (presence or absence), and region (Europe and the Americas or others).

Random-effects models indicated that telecare significantly reduced depressive symptoms in patients with LLD compared to the UC participants (SMD=–0.59, 95% CI –0.80 to –0.38; *P*<.001). Results of subgroup analysis by duration showed that short-term (≤3 months) interventions (SMD=–0.72, 95% CI –1.16 to –0.28; *P*<.001) were more effective than long-term (>3 months) interventions (SMD=–0.52, 95% CI –0.75 to –0.29; *P*<.001); other scales (SMD=–0.65, 95% CI –0.96 to –0.35; *P*<.001) were more effective than the PHQ-9 (SMD=–0.53, 95% CI –0.83 to –0.22; *P*<.001); the remote monitoring system (SMD=–1.13, 95% CI –1.51 to –0.76; *P*<.001) was more effective than telephone-based interventions (SMD=–0.38, 95% CI –0.56 to –0.20; *P*<.001); the effect on patients with LLD with chronic diseases (SMD=–0.67, 95% CI –0.89 to –0.44; *P*<.001) was better than that on patients with LLD without comorbid chronic diseases (SMD=–0.10, 95% CI –0.41 to 0.20; *P*=.07); and telecare was more effective in Europe and the Americas (SMD=–0.73, 95% CI –0.99 to –0.47; *P*<.001) than in other regions (SMD=–0.22, 95% CI –0.35 to –0.09; *P*=.42; Table 2).

Table 2. Subgroup meta-analysis for patients with late-life depression.

Subgroups	Cohen *d* SMD[a]	95% CI	*P* value	Heterogeneity (I^2; %)
Overall	–0.59	–0.80 to –0.38	<.001	86.42
Duration				
≤3 months	–0.72	–1.16 to –0.28	<.001	86.91
>3 months	–0.52	–0.75 to –0.29	<.001	84.72
Type of scale				
PHQ-9[b]	–0.53	–0.83 to –0.22	<.001	87.13
Others	–0.65	–0.96 to –0.35	<.001	86.21
Device type				
Telephone-based	–0.38	–0.56 to –0.20	<.001	75.30
Remote monitoring system	–1.13	–1.51 to –0.76	<.001	78.32
Comorbid chronic diseases				
Presence	–0.67	–0.89 to –0.44	<.001	85.31
Absence	–0.10	–0.41 to 0.20	.07	61.45
Region				
Europe and the Americas	–0.73	–0.99 to –0.47	<.001	86.38
Others	–0.22	–0.35 to –0.09	.42	0.00

[a]SMD: standardized mean difference.

[b]PHQ-9: Patient Health Questionnaire-9.

Meta-regression analysis showed that heterogeneity may not be related to the year of publication (*P*=.42), total sample size (*P*=.21), study area (*P*=.35), comorbid chronic disease (*P*=.47), duration (*P*=.75), and outcome measurement tools (*P*=.29). However, only the intervention device (*P*=.004) may have contributed to the heterogeneity.

Sensitivity Analysis and Publication Bias

The stability and reliability of the results of this meta-analysis and potential factors contributing to heterogeneity were explored by sensitivity analysis to assess the effect of the data of each study on the combined effect value (ie, SMD). The results of the sensitivity analysis showed that excluding each study individually had no significant effect on the combined effect value, and the study results were stable and reliable (Figure 5) [21-23,31-39]. Publication bias was assessed using funnel plots and Egger test indicators. The funnel plot was symmetrically distributed on both sides (Figure 6), and the Egger test showed no significant publication bias (*P*=.05).

Figure 5. Sensitivity analysis of effect value (standardized mean difference).

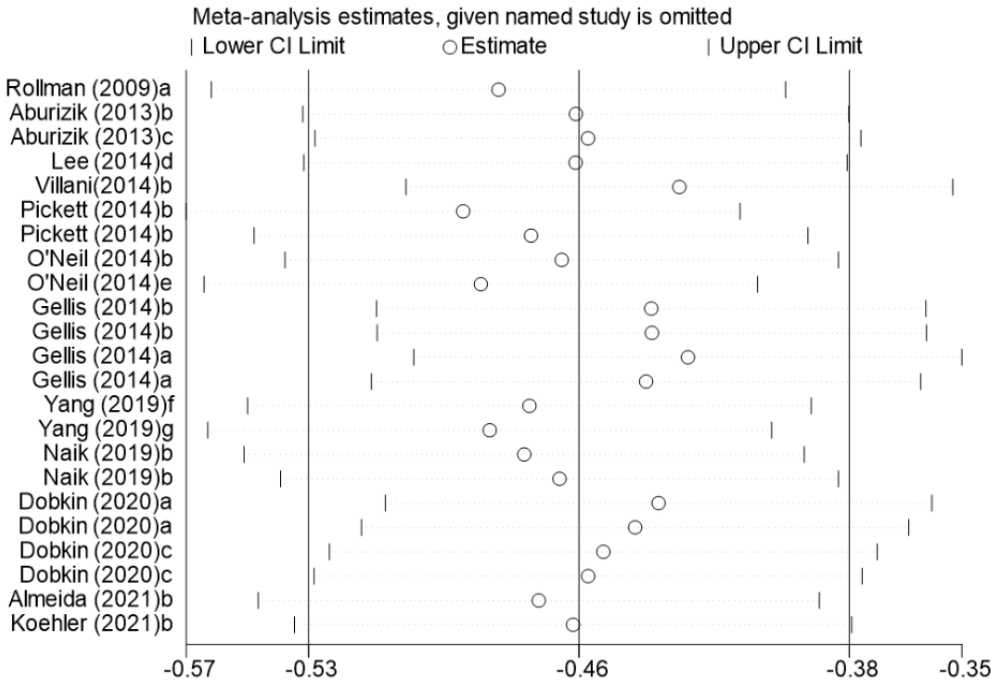
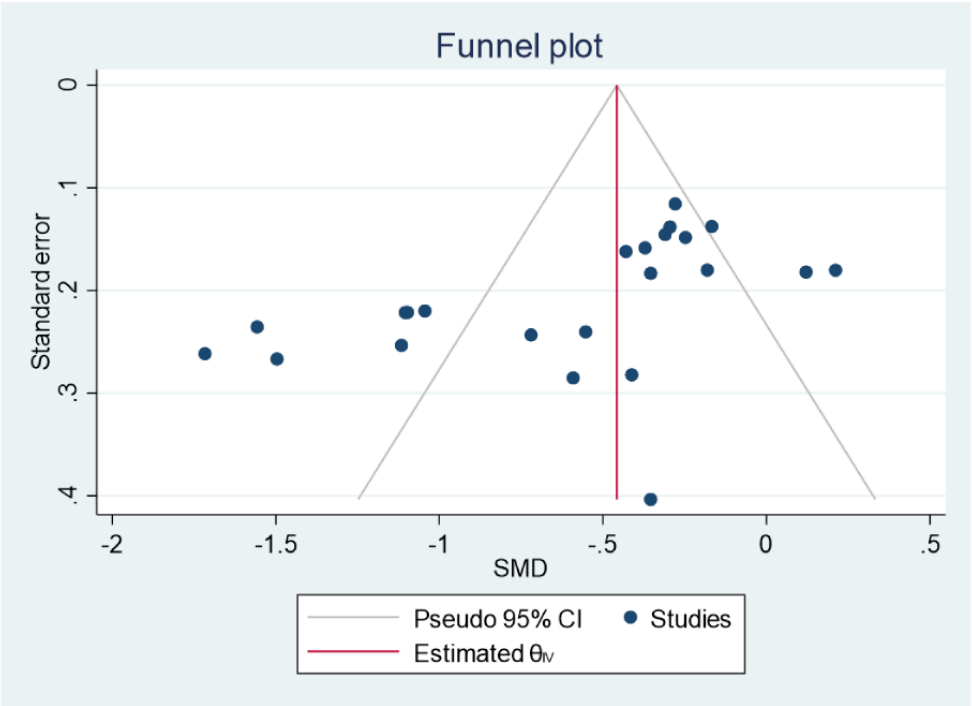


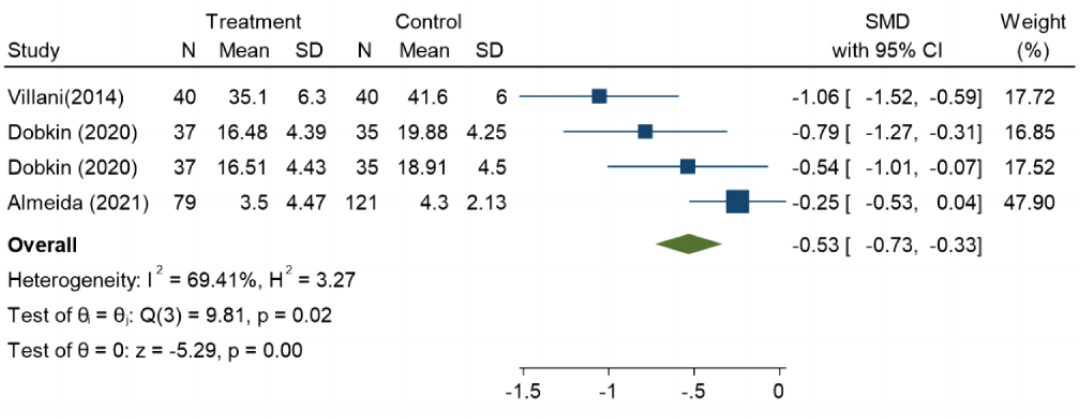
Figure 6. Funnel plot showing publication bias. SMD: standardized mean difference.



Anxiety Symptoms

To examine the efficacy of telecare in reducing anxiety compared with that of UC, we included 3 articles on patients with LLD. The results showed that telecare significantly reduced anxiety symptoms in patients with LLD (SMD=–0.53, 95% CI –0.73 to –0.33; *P*=.02; Figure 7) [22,33,38].

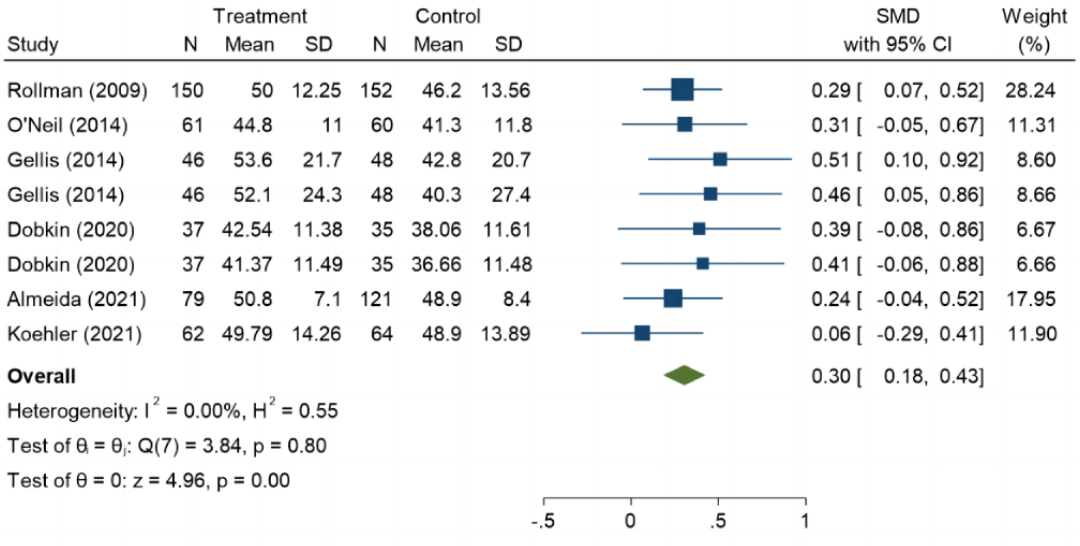
Figure 7. Forest plot of secondary outcome: anxiety.



QoL

Six studies assessed the mental components of QoL by using the Medical Outcomes Study Short Form survey. Our meta-analysis shows that the QoL of patients with LLD improved, but, overall, it was not significant (SMD=0.30, 95% CI 0.18-0.43; *P*=.80; Figure 8) [22,31,35,36,38,39].

Figure 8. Forest plot of secondary outcome: quality of life.



Discussion

Principal Findings

This meta-analysis shows that compared with UC, telecare significantly reduces symptoms of depression and anxiety but has no significant effect on improving QoL in patients with LLD.

Primary Outcome Measures

The pooled results show that telecare has a significant effect on reducing depressive symptoms in patients with LLD, which is consistent with the findings of previous studies [25,26,29]. Apart from dealing with depression itself, the increased severity of LLD is also related to factors such as aging, chronic disease, and socioeconomic stress [5]. Telecare offers unique and innovative opportunities for treating depression symptoms in older adults. Patients with LLD can leverage the advantages of telecare to connect with health care professionals, overcoming geographical distance and physical limitations, thereby reducing the psychological burden of coping with the disease [43]. Furthermore, professional psychological support is crucial for patients with LLD, and it can encourage patients to express their feelings and release stress [44]. However, it is worth noting that despite telecare offering more possibilities for treating LLD, the complexity of the medical population makes it challenging. Telecare can provide greater coverage for health care, yet considerations such as individual needs of older patients or environmental backgrounds need to be factored in [45]. Currently, offering targeted telecare services to a large population of older adults in rural, remote, or underserved areas remains a challenge [10]. In particular, older adults face significant barriers in using telephone and internet connections during the COVID-19 pandemic [46]. As a result, telecare management may not be as effective for this population as for others. The size of the research effect will depend on the nature of the intervention and the quality of the study [47]. High-quality telemedicine will help older adults benefit both physically and mentally. Further investigation and more research are necessary.

Subgroup analysis indicates that the effectiveness of telecare in treating LLD can be influenced by measurement tools,

durations, intervention devices, comorbid chronic conditions, and regions involved. In terms of depression measurement tools, other scales appear to be more effective than PHQ-9 (0.65% vs 0.53%), which may be related to measurement errors caused by differences in specific items and the generalizability of different measurement tools [48]. Results from durations of ≤3 and >3 months showed a reduction in depressive symptoms in patients with LLD, with short-term interventions proving to be more effective (0.72% vs 0.52%). Short-term interventions focus more on addressing specific issues or symptoms, producing immediate effects. For older adults, short-term interventions might be more readily accepted as long-term treatments could induce fatigue or a lack of patience. Our findings differ slightly from those of another study [49], which implemented more targeted interventions based on different treatment responses, confirming the more significant effectiveness of long-term interventions. Therefore, there is insufficient evidence to conclusively establish that telecare is necessarily superior in short-term intervention efficacy for LLD compared to long-term interventions. In fact, for depression management, a combination of short-term and long-term interventions is often required to deliver comprehensive and enduring support and management [50].

Subgroup analysis also found that remote monitoring systems appear to be more effective than telephone-based management (1.13% vs 0.38%). The remote monitoring system ensures timely and accurate transmission of patients' symptom information and data to health care professionals, enabling patients to receive effective treatments [51]. Telecare was more effective in patients with LLD with comorbid chronic conditions compared to those without such comorbidities (0.67% vs 0.10%). Co-occurrence of chronic medical conditions and depression is common. Evidence suggests that older adults with chronic illness are more likely to be affected by depressive symptoms than those without chronic illness [2,5,7]. Older adults with chronic conditions are more likely to seek medical care and adherent to treatment [52]. Therefore, while actively treating chronic conditions, there might be a degree of alleviation in depressive symptoms among older adults. Telecare was more effective in Europe and the Americas in improving depressive symptoms in patients with LLD compared than in other regions (0.73% vs 0.22%). The health care systems in Europe and the Americas are generally more developed, which may lead to more comprehensive support for telecare [53]. In low- and middle-income countries, the resources available for geriatric mental health care are considered severely inadequate [54]. Nevertheless, telecare is beginning to have an important impact on many aspects of health care in transitional countries [55].

Secondary Outcome Measures

Telecare has a positive effect on improving anxiety symptoms of patients with LLD. This result is consistent with findings from other studies [56]. Telecare offers a more convenient access method, allowing patients to receive treatment at home, thereby circumventing the inconvenience and anxiety associated with hospital visits [16,17]. Health care professionals can engage with patients more frequently through telecare, gaining insights into their symptoms and emotional fluctuations. This allows for adjustments in the treatment plan to effectively meet the unique needs of this population [18,57]. Additionally, the symptoms of anxiety and depression are often co-occurring [58], particularly among older adults. Due to the similarity between depression and anxiety symptoms, many treatment approaches are shared between the two. A recent meta-analysis suggests that psychotherapy delivered remotely is as effective as face-to-face therapy for anxiety disorder [59]. This evidence is based on outcomes obtained from different age groups. It may be more challenging to create a trusting relationship remotely than in person [60]. Older adults have negative views about health IT performing accurately and dependably, which will have a significant impact on the acceptance of telecare [61]. In brief, when using telecare for addressing emotional disorders in older adults, closer supervision and guidance might be necessary. Health care professionals need to distinguish the appropriateness of using telecare for communication and, in turn, individually tailor patient care.

We found that the mental component of QoL in patients with LLD improved after using telecare; however, this difference was not significant compared to that with the use of UC. This finding aligns with results from other studies [62,63]. Improving QoL is a comprehensive issue that includes not only improvements in health care but also social, psychological, and emotional factors [64]. Influenced by these factors, it is difficult to compare the results of QoL considering different contexts. Several results from RCTs with older adults using telemonitoring systems showed an improvement in the participants' QoL [65,66]; other telemonitoring RCTs could not achieve congruent results [67]. Improvements in QoL often require deeper interactions and personalized care. In particular, participants with mental disorders may benefit from individual and tailored solutions provided by general practitioners [68]. When using telecare, it is crucial to acknowledge that each subpopulation of marginalized older adults has differing strengths and needs. The studies we included focused more on managing the disease itself, which may weaken overall effectiveness. It is not easy to present telemedicine to the older population. The limitations inherent in older adults may lead to difficulties in receiving telecare, including the lack of technical literacy, equipment access barriers, cognitive function, etc [11]. These reasons could explain why telecare is not significant in improving the QoL of patients with LLD. The potential value of telecare in maintaining the QoL for individuals with LLD warrants further exploration. While this study did not reveal a positive impact of telecare on the QoL for patients with LLD, it has been established that telecare can assist patients with LLD in gaining more information about health services.

Limitations

This study still had some limitations. First, most of the studies included in the review lacked sufficient measure detail, leading to irreversible bias. Our study mainly included 2 interventions based on telephone and remote monitoring to reduce this bias. Second, the measurement tools used in this study lacked standardization and heavily relied on self-reports from participants, introducing a degree of subjectivity and concealment that is not as rigorous as structured interviews. However, we attempted to validate the effectiveness of the results by using authoritative scales. Third, differences in the

specific intervention methods, frequency, and content among the included studies may lead to clinical heterogeneity across different studies.

Conclusions

Our meta-analysis shows that telecare has a positive impact on depressive and anxiety symptoms, despite high heterogeneity in depression symptoms. Therefore, studies with larger sample sizes and homogeneity were required to determine the effects of telecare in patients with LLD. Future research can continue to refine telecare systems and assess the specific needs of older vulnerable populations for more accurate evidence.

Acknowledgments

This project was jointly supported by the Hubei Provincial Natural Science Foundation and the Innovative Development of Chinese Medicine of China (2023AFD160).

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA checklist.
[DOCX File , 28 KB - mhealth_v12i1e50787_app1.docx]

Multimedia Appendix 2
Search strategy.
[DOC File , 17 KB - mhealth_v12i1e50787_app2.doc]

Multimedia Appendix 3
The intervention content of the intervention group.
[DOC File , 26 KB - mhealth_v12i1e50787_app3.doc]

References

1. Ageing and health. World Health Organization. 2022. URL: https://www.who.int/news-room/fact-sheets/detail/ageing-and-health [accessed 2024-01-02]
2. Taylor WD. Depression in the elderly. N Engl J Med 2014 Sep 25;371(13):1228-1236. [doi: 10.1056/nejmcp1402180]
3. Unützer J, Katon W, Callahan CM, Williams JW, Hunkeler E, Harpole L, IMPACT Investigators. Improving Mood-Promoting Access to Collaborative Treatment. Collaborative care management of late-life depression in the primary care setting: a randomized controlled trial. JAMA 2002 Dec 11;288(22):2836-2845. [doi: 10.1001/jama.288.22.2836] [Medline: 12472325]
4. Hu T, Zhao X, Wu M, Li Z, Luo L, Yang C, et al. Prevalence of depression in older adults: A systematic review and meta-analysis. Psychiatry Res 2022 May;311:114511. [doi: 10.1016/j.psychres.2022.114511] [Medline: 35316691]
5. Alexopoulos GS. Depression in the elderly. The Lancet 2005 Jun;365(9475):1961-1970. [doi: 10.1016/s0140-6736(05)66665-2]
6. Gundersen E, Bensadon B. Geriatric Depression. Prim Care 2023 Mar;50(1):143-158. [doi: 10.1016/j.pop.2022.10.010] [Medline: 36822724]
7. Katon WJ, Lin E, Russo J, Unutzer J. Increased medical costs of a population-based sample of depressed elderly patients. Arch Gen Psychiatry 2003 Sep 01;60(9):897-903. [doi: 10.1001/archpsyc.60.9.897] [Medline: 12963671]
8. Li M, Cai J, Zhang P, Fei C, Xu F. Drug brand response and its impact on compliance and efficacy in depression patients. Front Pharmacol 2016 Jan 10;7:540 [FREE Full text] [doi: 10.3389/fphar.2016.00540] [Medline: 28119615]
9. Heidenreich PA. Patient adherence: the next frontier in quality improvement. Am J Med 2004 Jul 15;117(2):130-132. [doi: 10.1016/j.amjmed.2004.03.007] [Medline: 15234651]
10. Caldwell T, Jorm AF, Knox S, Braddock D, Dear KBG, Britt H. General practice encounters for psychological problems in rural, remote and metropolitan areas in Australia. Aust N Z J Psychiatry 2004 Oct 21;38(10):774-780. [doi: 10.1080/j.1440-1614.2004.01461.x] [Medline: 15369535]
11. Harerimana B, Forchuk C, O'Regan T. The use of technology for mental healthcare delivery among older adults with depressive symptoms: a systematic literature review. Int J Ment Health Nurs 2019 Jun;28(3):657-670. [doi: 10.1111/inm.12571] [Medline: 30666762]
12. Barney LJ, Griffiths KM, Jorm AF, Christensen H. Stigma about depression and its impact on help-seeking intentions. Aust N Z J Psychiatry 2006 Jan;40(1):51-54. [doi: 10.1080/j.1440-1614.2006.01741.x] [Medline: 16403038]
13. Bohlmeijer E, Smit F, Cuijpers P. Effects of reminiscence and life review on late‐life depression: a meta‐analysis. Int J Geriat Psychiatry 2003 Dec 09;18(12):1088-1094. [doi: 10.1002/gps.1018]

14. Birren J, Cohen G, Sloane R. Handbook of Mental Health and Aging. San Diego, CA: Academic Press; 2013.
15. Topol E. Digital medicine: empowering both patients and clinicians. The Lancet 2016 Aug;388(10046):740-741. [doi: 10.1016/s0140-6736(16)31355-1]
16. The Audit Commission. Implementing Telecare - The Learning Exchange. London, United Kingdom: The Audit Commission; 2004:2022.
17. Kim EH, Gellis ZD, Bradway CK, Kenaley B. Depression care services and telehealth technology use for homebound elderly in the United States. Aging Ment Health 2019 Sep 24;23(9):1164-1173. [doi: 10.1080/13607863.2018.1481925] [Medline: 30472881]
18. The Knowledge Network. URL: http://www.knowledge.scot.nhs.uk/home.aspx [accessed 2024-01-02]
19. Richardson LK, Frueh BC, Grubaugh AL, Egede L, Elhai JD. Current directions in videoconferencing tele-mental health research. Clin Psychol (New York) 2009 Sep 01;16(3):323-338 [FREE Full text] [doi: 10.1111/j.1468-2850.2009.01170.x] [Medline: 20161010]
20. Gellis ZD, Kenaley B, McGinty J, Bardelli E, Davitt J, Ten Have T. Outcomes of a telehealth intervention for homebound older adults with heart or chronic respiratory failure: a randomized controlled trial. Gerontologist 2012 Aug;52(4):541-552. [doi: 10.1093/geront/gnr134] [Medline: 22241810]
21. Naik AD, Hundt NE, Vaughan EM, Petersen NJ, Zeno D, Kunik ME, et al. Effect of telephone-delivered collaborative goal setting and behavioral activation vs enhanced usual care for depression among adults with uncontrolled diabetes: a randomized clinical trial. JAMA Netw Open 2019 Aug 02;2(8):e198634 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.8634] [Medline: 31390035]
22. Dobkin RD, Menza M, Allen LA, Tiu J, Friedman J, Bienfait KL, et al. Telephone-based cognitive-behavioral therapy for depression in Parkinson disease. J Geriatr Psychiatry Neurol 2011 Dec 05;24(4):206-214 [FREE Full text] [doi: 10.1177/0891988711422529] [Medline: 22228827]
23. Lee H, Yoon J, Lim Y, Jung H, Kim S, Yoo Y, et al. The effect of nurse-led problem-solving therapy on coping, self-efficacy and depressive symptoms for patients with chronic obstructive pulmonary disease: a randomised controlled trial. Age Ageing 2015 May;44(3):397-403. [doi: 10.1093/ageing/afu201] [Medline: 25548124]
24. Katon WJ. Clinical and health services relationships between major depression, depressive symptoms, and general medical illness. Biol Psychiatry 2003 Aug 01;54(3):216-226. [doi: 10.1016/s0006-3223(03)00273-7] [Medline: 12893098]
25. Andersson G, Cuijpers P. Internet-based and other computerized psychological treatments for adult depression: a meta-analysis. Cogn Behav Ther 2009;38(4):196-205. [doi: 10.1080/16506070903318960] [Medline: 20183695]
26. Giovanetti AK, Punt SE, Nelson E, Ilardi SS. Teletherapy versus in-person psychotherapy for depression: a meta-analysis of randomized controlled trials. Telemed J E Health 2022 Aug 01;28(8):1077-1089. [doi: 10.1089/tmj.2021.0294] [Medline: 35007437]
27. García-Lizana F, Muñoz-Mayorga I. Telemedicine for depression: a systematic review. Perspect Psychiatr Care 2010 Apr;46(2):119-126. [doi: 10.1111/j.1744-6163.2010.00247.x] [Medline: 20377799]
28. Kivelitz L, Kriston L, Christalle E, Schulz H, Watzke B, Härter M, et al. Effectiveness of telephone-based aftercare case management for adult patients with unipolar depression compared to usual care: A randomized controlled trial. PLoS One 2017 Oct 27;12(10):e0186967 [FREE Full text] [doi: 10.1371/journal.pone.0186967] [Medline: 29077724]
29. de Oliveira PBF, Dornelles TM, Gosmann NP, Camozzato A. Efficacy of telemedicine interventions for depression and anxiety in older people: A systematic review and meta-analysis. Int J Geriatr Psychiatry 2023 May 19;38(5):e5920. [doi: 10.1002/gps.5920] [Medline: 37204341]
30. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
31. Rollman BL, Belnap BH, LeMenager MS, Mazumdar S, Houck PR, Counihan PJ, et al. Telephone-delivered collaborative care for treating post-CABG depression: a randomized controlled trial. JAMA 2009 Nov 18;302(19):2095-2103 [FREE Full text] [doi: 10.1001/jama.2009.1670] [Medline: 19918088]
32. Aburizik A, Dindo L, Kaboli P, Charlton M, Dawn K, Turvey C. A pilot randomized controlled trial of a depression and disease management program delivered by phone. J Affect Disord 2013 Nov;151(2):769-774. [doi: 10.1016/j.jad.2013.06.028] [Medline: 23871127]
33. Villani A, Malfatto G, Compare A, Rosa FD, Bellardita L, Branzi G, et al. Clinical and psychological telemonitoring and telecare of high risk heart failure patients. J Telemed Telecare 2014 Oct 22;20(8):468-475. [doi: 10.1177/1357633x14555644]
34. Pickett YR, Kennedy GJ, Freeman K, Cummings J, Woolis W. The effect of telephone-facilitated depression care on older, medically ill patients. J Behav Health Serv Res 2014 Jan 10;41(1):90-96 [FREE Full text] [doi: 10.1007/s11414-013-9327-1] [Medline: 23572444]
35. O'Neil A, Taylor B, Sanderson K, Cyril S, Chan B, Hawkes AL, MoodCare Investigator Team. Efficacy and feasibility of a tele-health intervention for acute coronary syndrome patients with depression: results of the "MoodCare" randomized controlled trial. Ann Behav Med 2014 Oct 26;48(2):163-174. [doi: 10.1007/s12160-014-9592-0] [Medline: 24570217]

36. Gellis ZD, Kenaley BL, Ten Have T. Integrated telehealth care for chronic illness and depression in geriatric home care patients: the Integrated Telehealth Education and Activation of Mood (I-TEAM) study. J Am Geriatr Soc 2014 May;62(5):889-895. [doi: 10.1111/jgs.12776] [Medline: 24655228]
37. Yang L, Wang X, Cui X. Patients' intensive telephone-based care program reduces depression in coronary artery disease patients and may contribute to favorable overall survival by decreasing depression. J Cardiovasc Nurs 2019;34(3):236-243. [doi: 10.1097/jcn.0000000000000571]
38. Almeida OP, Patel H, Kelly R, Ford A, Flicker L, Robinson S, et al. Preventing depression among older people living in rural areas: a randomised controlled trial of behavioural activation in collaborative care. Int J Geriatr Psychiatry 2021 Apr 31;36(4):530-539. [doi: 10.1002/gps.5449] [Medline: 33098159]
39. Koehler J, Stengel A, Hofmann T, Wegscheider K, Koehler K, Sehner S, et al. Telemonitoring in patients with chronic heart failure and moderate depressed symptoms: results of the Telemedical Interventional Monitoring in Heart Failure (TIM-HF) study. Eur J Heart Fail 2021 Jan 09;23(1):186-194 [FREE Full text] [doi: 10.1002/ejhf.2025] [Medline: 33063412]
40. Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page M, et al. Cochrane Handbook for Systematic Reviews of Interventions (Version 6.3). Chichester, United Kingdom: John Wiley & Sons; 2022.
41. Cohen J. A power primer. Psychol Bull 1992 Jul;112(1):155-159. [doi: 10.1037//0033-2909.112.1.155] [Medline: 19565683]
42. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997 Sep 13;315(7109):629-634 [FREE Full text] [doi: 10.1136/bmj.315.7109.629] [Medline: 9310563]
43. Bujnowska-Fedak M, Grata-Borkowska U. Use of telemedicine-based care for the aging and elderly: promises and pitfalls. SHTT 2015 May:91. [doi: 10.2147/shtt.s59498]
44. Hamm ME, Karp JF, Lenard E, Dawdani A, Lavretsky H, Lenze EJ, et al. "What else can we do?"-Provider perspectives on treatment-resistant depression in late life. J Am Geriatr Soc 2022 Apr 03;70(4):1190-1197. [doi: 10.1111/jgs.17592] [Medline: 34862593]
45. Watt J, Fahim C, Straus S, Goodarzi Z. Barriers and facilitators to virtual care in a geriatric medicine clinic: a semi-structured interview study of patient, caregiver and healthcare provider perspectives. Age Ageing 2022 Jan 06;51(1):a. [doi: 10.1093/ageing/afab218] [Medline: 34794177]
46. Lam K, Lu AD, Shi Y, Covinsky KE. Assessing telemedicine unreadiness among older adults in the United States during the COVID-19 pandemic. JAMA Intern Med 2020 Oct 01;180(10):1389-1391 [FREE Full text] [doi: 10.1001/jamainternmed.2020.2671] [Medline: 32744593]
47. Lecomte T, Potvin S, Corbière M, Guay S, Samson C, Cloutier B, et al. Mobile apps for mental health issues: meta-review of meta-analyses. JMIR Mhealth Uhealth 2020 May 29;8(5):e17458 [FREE Full text] [doi: 10.2196/17458] [Medline: 32348289]
48. Fried EI. The 52 symptoms of major depression: lack of content overlap among seven common depression scales. J Affect Disord 2017 Jan 15;208:191-197. [doi: 10.1016/j.jad.2016.10.019] [Medline: 27792962]
49. Fortney JC, Pyne JM, Edlund MJ, Williams DK, Robinson DE, Mittal D, et al. A randomized trial of telemedicine-based collaborative care for depression. J Gen Intern Med 2007 Aug;22(8):1086-1093 [FREE Full text] [doi: 10.1007/s11606-007-0201-9] [Medline: 17492326]
50. Shoji M, Maeda H, Watanabe F, Tanuma K, Fujiwara A, Iwanaga Y, et al. A non-randomized, controlled, interventional study to investigate the effects of community pharmacists' cognitive behavioral therapy-based interventions on medication adherence and relevant indicators in patients with depression. BMC Psychiatry 2023 Feb 24;23(1):124 [FREE Full text] [doi: 10.1186/s12888-023-04602-5] [Medline: 36829184]
51. Gokalp H, de Folter J, Verma V, Fursse J, Jones R, Clarke M. Integrated telehealth and telecare for monitoring frail elderly with chronic disease. Telemed J E Health 2018 Dec;24(12):940-957 [FREE Full text] [doi: 10.1089/tmj.2017.0322] [Medline: 30129884]
52. Jiang C, Zhu F, Qin T. Relationships between chronic diseases and depression among middle-aged and elderly people in china: a prospective study from CHARLS. Curr Med Sci 2020 Oct;40(5):858-870. [doi: 10.1007/s11596-020-2270-5] [Medline: 33123901]
53. Sivakumar PT, Mukku SSR, Kar N, Manjunatha N, Phutane VH, Sinha P, et al. Geriatric telepsychiatry: promoting access to geriatric mental health care beyond the physical barriers. Indian J Psychol Med 2020 Oct;42(5 Suppl):41S-46S [FREE Full text] [doi: 10.1177/0253717620958380] [Medline: 33354062]
54. Dlouhy M. Mental health care system and mental health expenditures in the Czech Republic. J Ment Health Policy Econ 2004 Dec;7(4):159-165. [Medline: 15701931]
55. Edworthy SM. Telemedicine in developing countries. BMJ 2001 Sep 08;323(7312):524-525 [FREE Full text] [doi: 10.1136/bmj.323.7312.524] [Medline: 11546681]
56. Ando M, Kao Y, Lee Y, Tai S, Mendez SR, Sasaki K, et al. Remote cognitive behavioral therapy for older adults with anxiety symptoms: a systematic review and meta-analysis. J Telemed Telecare 2023 Feb 16:1357633X2311517. [doi: 10.1177/1357633x231151788]
57. Gustafson DH, Mares M, Johnston DC, Mahoney JE, Brown RT, Landucci G, et al. A web-based eHealth intervention to improve the quality of life of older adults with multiple chronic conditions: protocol for a randomized controlled trial. JMIR Res Protoc 2021 Feb 19;10(2):e25175 [FREE Full text] [doi: 10.2196/25175] [Medline: 33605887]

58. Almeida OP, Draper B, Pirkis J, Snowdon J, Lautenschlager NT, Byrne G, et al. Anxiety, depression, and comorbid anxiety and depression: risk factors and outcome over two years. Int Psychogeriatr 2012 Jun 12;24(10):1622-1632. [doi: 10.1017/s104161021200107x]
59. Krzyzaniak N, Greenwood H, Scott AM, Peiris R, Cardona M, Clark J, et al. The effectiveness of telehealth versus face-to face interventions for anxiety disorders: a systematic review and meta-analysis. J Telemed Telecare 2021 Dec 03:1357633X211053738. [doi: 10.1177/1357633X211053738] [Medline: 34860613]
60. Dorsey ER, Topol EJ. State of telehealth. N Engl J Med 2016 Jul 14;375(2):154-161. [doi: 10.1056/nejmra1601705]
61. Welsh S, Hassiotis A, O'Mahoney G, Deahl M. Big brother is watching you--the ethical implications of electronic surveillance measures in the elderly with dementia and in adults with learning difficulties. Aging Ment Health 2003 Sep;7(5):372-375. [doi: 10.1080/1360786031000150658] [Medline: 12959806]
62. Arian M, Valinejadi A, Soleimani M. Quality of life in heart patients receiving telerehabilitation: an overview with meta-analyses. Iran J Public Health 2022 Nov 19;51(11):2388-2403 [FREE Full text] [doi: 10.18502/ijph.v51i11.11157] [Medline: 36561264]
63. Wong AKC, Bayuo J, Wong FKY, Yuen WS, Lee AYL, Chang PK, et al. Effects of a nurse-led telehealth self-care promotion program on the quality of life of community-dwelling older adults: systematic review and meta-analysis. J Med Internet Res 2022 Mar 21;24(3):e31912 [FREE Full text] [doi: 10.2196/31912] [Medline: 35311680]
64. Hajek A, Brettschneider C, Ernst A, Lange C, Wiese B, Prokein J, et al. Complex coevolution of depression and health-related quality of life in old age. Qual Life Res 2015 Nov 19;24(11):2713-2722. [doi: 10.1007/s11136-015-1005-8] [Medline: 25986907]
65. Cichosz SL, Udsen FW, Hejlesen O. The impact of telehealth care on health-related quality of life of patients with heart failure: results from the Danish TeleCare North heart failure trial. J Telemed Telecare 2020;26(7-8):452-461. [doi: 10.1177/1357633X19832713] [Medline: 30975047]
66. Bohingamu Mudiyanselage S, Stevens J, Watts JJ, Toscano J, Kotowicz MA, Steinfort CL, et al. Personalised telehealth intervention for chronic disease management: a pilot randomised controlled trial. J Telemed Telecare 2018 May 24;25(6):343-352. [doi: 10.1177/1357633x18775850]
67. Pecina JL, Hanson GJ, Van Houten H, Takahashi PY. Impact of telemonitoring on older adults health-related quality of life: the Tele-ERA study. Qual Life Res 2013 Nov 14;22(9):2315-2321. [doi: 10.1007/s11136-013-0361-5] [Medline: 23408299]
68. Sinnott C, Mc Hugh S, Browne J, Bradley C. GPs' perspectives on the management of patients with multimorbidity: systematic review and synthesis of qualitative research. BMJ Open 2013 Sep 13;3(9):e003610 [FREE Full text] [doi: 10.1136/bmjopen-2013-003610] [Medline: 24038011]

Abbreviations

LLD: late-life depression
MeSH: Medical Subject Headings
PHQ-9: Patient Health Questionnaire-9
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QoL: quality of life
RCT: randomized controlled trial
SMD: standardized mean difference
UC: usual care

Edited by L Buis; submitted 12.07.23; peer-reviewed by W He, X Tan, S Peng; comments to author 15.11.23; revised version received 01.12.23; accepted 22.12.23; published 17.01.24.

Please cite as:
Wu M, Li C, Hu T, Zhao X, Qiao G, Gao X, Zhu X, Yang F
Effectiveness of Telecare Interventions on Depression Symptoms Among Older Adults: Systematic Review and Meta-Analysis
JMIR Mhealth Uhealth 2024;12:e50787
URL: https://mhealth.jmir.org/2024/1/e50787
doi:10.2196/50787
PMID:38231546

Review

Effects of Digital Physical Health Exercises on Musculoskeletal Diseases: Systematic Review With Best-Evidence Synthesis

Johanna Nagel[1], MSc; Florian Wegener[1], MSc; Casper Grim[2], PD, Dr med; Matthias Wilhelm Hoppe[1], Prof Dr

[1]Movement and Training Science, Leipzig University, Leipzig, Germany

[2]Center for Musculoskeletal Surgery Osnabrück, Klinikum Osnabrück, Osnabrück, Germany

Corresponding Author:
Johanna Nagel, MSc
Movement and Training Science
Leipzig University
Jahnallee 59
Leipzig, 04109
Germany
Phone: 49 3419731822
Email: jojo_nagel@yahoo.de

Abstract

Background: Musculoskeletal diseases affect 1.71 billion people worldwide, impose a high biopsychosocial burden on patients, and are associated with high economic costs. The use of digital health interventions is a promising cost-saving approach for the treatment of musculoskeletal diseases. As physical exercise is the best clinical practice in the treatment of musculoskeletal diseases, digital health interventions that provide physical exercises could have a highly positive impact on musculoskeletal diseases, but evidence is lacking.

Objective: This systematic review aims to evaluate the impact of digital physical health exercises on patients with musculoskeletal diseases concerning the localization of the musculoskeletal disease, patient-reported outcomes, and medical treatment types.

Methods: We performed systematic literature research using the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. The search was conducted using the PubMed, BISp, Cochrane Library, and Web of Science databases. The Scottish Intercollegiate Guidelines Network checklist was used to assess the quality of the included original studies. To determine the evidence and direction of the impact of digital physical health exercises, a best-evidence synthesis was conducted, whereby only studies with at least acceptable methodological quality were included for validity purposes.

Results: A total of 8988 studies were screened, of which 30 (0.33%) randomized controlled trials met the inclusion criteria. Of these, 16 studies (53%) were of acceptable or high quality; they included 1840 patients (1008/1643, 61.35% female; 3 studies including 197 patients did not report gender distribution) with various musculoskeletal diseases. A total of 3 different intervention types (app-based interventions, internet-based exercises, and telerehabilitation) were used to deliver digital physical health exercises. Strong evidence was found for the positive impact of digital physical health exercises on musculoskeletal diseases located in the back. Moderate evidence was found for diseases located in the shoulder and hip, whereas evidence for the entire body was limited. Conflicting evidence was found for diseases located in the knee and hand. For patient-reported outcomes, strong evidence was found for impairment and quality of life. Conflicting evidence was found for pain and function. Regarding the medical treatment type, conflicting evidence was found for operative and conservative therapies.

Conclusions: Strong to moderate evidence was found for a positive impact on musculoskeletal diseases located in the back, shoulder, and hip and on the patient-reported outcomes of impairment and quality of life. Thus, digital physical health exercises could have a positive effect on a variety of symptoms of musculoskeletal diseases.

(JMIR Mhealth Uhealth 2024;12:e50616) doi:10.2196/50616

KEYWORDS

mobile health; mHealth; electronic health; eHealth; digital health applications; DiGA; musculoskeletal; MSK; home-based; PROM; disorder; mobile phone

Introduction

Background

A total of 1.71 billion people are affected by musculoskeletal diseases worldwide [1]. They are characterized by chronic pain, functional disability, impairment, and reduced quality of life [1,2]. The most commonly affected body regions are the lower back and neck, with a period prevalence over the last 12 months of up to 61.3% and 45.7% [3], respectively, and a common disease is osteoarthritis, with a prevalence of up to 17.9% [4]. In addition to the high biopsychosocial burden [5], the evident increase in the incidence of musculoskeletal diseases over the last decades [6] results in high economic costs because of lost workdays and conservative or operative medical treatments [5]. To overcome such undesirable consequences, evidence-based, effective, and cost-saving health interventions are required. Therefore, the use of digital health interventions is a promising approach.

Digital health interventions aim to manage a wide range of diseases and health issues using digital devices such as smartphones, tablets, computers, or wearables, including mobile apps, telerehabilitation and web-based physician visits, web-based interactive programs, or tracking tools [7]. The use of mobile apps is increasing, with common intervention types categorized as physical exercise and fitness, lifestyle and stress, diet and nutrition, or medication reminders and educational materials [7]. In some countries, such as Germany, so-called digital health applications are also supported by health insurers after being evaluated as medical devices [8]. However, owing to their cost-saving potential and the increasing number of commercially available digital health interventions [7], further research is needed to evaluate the impact of different types of digital health interventions on specific diseases.

Previous systematic reviews have extensively evaluated the impact of digital health interventions on internal diseases. Positive effects have been demonstrated in treating chronic obstructive pulmonary disease [9], cardiovascular disease [10], and diabetes [11]. These effects encompass improvements in clinically relevant outcomes such as quality of life, health-related impairments, amelioration of risk factors and their consequences, as well as the control and management of HbA1c levels. For musculoskeletal diseases, only 2 previous systematic reviews have evaluated the impact of digital health interventions as a primary outcome. One review [12] showed that there are substantial clinical benefits in the management of musculoskeletal diseases for the patient-reported outcomes of pain (9 out of 19 studies) and functional disability (10 out of 16 studies). The results show that digital health interventions as adjuncts and as stand-alone treatments are not inferior but partly superior compared with interventions based on standard therapy, nondigital self-management, noninteractive digital measures, or no intervention. However, in this previous review, no evidence synthesis was performed. In addition, a further review [13] conducted a meta-analysis and showed moderate-quality evidence that digital health interventions are effective in reducing pain and improving function and self-management in patients with musculoskeletal disease. The included studies considered digital health interventions as interventions that are to be used only at home and as adjuncts to standard clinical care, compared with standard care, noninteractive digital interventions, or no intervention. Taken together, the use of digital health interventions as an adjunct to regular therapy could have positive health-related effects for both internal and musculoskeletal diseases, although less evidence is available for the latter.

However, little is known about the relationship between clinically relevant factors, such as the localization of the musculoskeletal diseases, patient-reported outcomes, or the type of applied conservative or operative medical treatments, and the effects of different types of digital health interventions in the treatment of musculoskeletal diseases. In terms of evidence-based medicine, such relationships must first be clarified when using digital health interventions as a regular treatment option for specific musculoskeletal diseases. Because of the increasing number of original studies, more systematic research is needed to review and assess the existing evidence. Previous systematic reviews [12,13] have included all types of digital health interventions, providing a comprehensive overall result across all biopsychosocial domains. As physical exercise is the best clinical practice for the treatment of musculoskeletal diseases [14], digital physical health exercises could have a highly positive impact on musculoskeletal diseases. However, little is known about how the effects of digital physical health exercises are related to the aforementioned clinically relevant factors.

Objective

Therefore, this systematic review aimed to evaluate the impact of digital physical health exercises on patients with musculoskeletal diseases concerning the localization of the musculoskeletal disease, patient-reported outcomes, and medical treatment types. In addition, a best-evidence synthesis was conducted to estimate the direction and strength of the existing evidence.

Methods

Research Design and Eligibility Criteria

The systematic review was conducted according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [15]. Eligibility criteria according to the population, intervention, comparison, outcome, study design (PICOS) scheme [16] were applied. Table 1 presents the inclusion criteria according to the PICOS scheme. Textbox 1 presents the search line.

The corresponding keywords are also presented. Studies were not reviewed if they did not report on a specific musculoskeletal disease, if the digital health intervention included no physical exercises, if no control group was considered, or if none of the included patient-reported outcomes were assessed as a primary outcome. All methodological steps were performed by 1 author and validated by a second author. Uncertainties were discussed until consensus was reached. Because of the literary nature of this study, ethics approval was not required.

Table 1. PICOS[a] scheme for the definition of the inclusion criteria and the presentation of the corresponding keywords.

	Population	Intervention	Comparison	Outcome	Study design
Inclusion criteria[b]	Patients with any musculoskeletal disease according to the definition of the WHO[c]	Any digital health intervention using home-based physical exercises	Any conventional or no therapy	Patient-reported outcomes pain, function, disability, and quality of life assessed by established and validated clinical questionnaires or scales	Randomized controlled trials
Keywords	"Musculoskeletal disease" OR "Musculoskeletal disorder" OR "Musculoskeletal pain" OR "Chronic pain" OR "Acute pain" OR "Overuse pain" OR "Chronic injury" OR "Chronic injuries" OR "Acute injury" OR "Acute injuries" OR "Overuse injury" OR "Overuse injuries" OR "Chronic disease" OR "Acute disease" OR "Overuse disease" OR "Osteoporosis" OR "Osteoarthritis" OR "Rheumatoid arthritis" OR "Tendinopathy" OR "Tendinopathies" OR "Rotator cuff" OR "Lower extremity" OR "Lower extremities" OR "Upper extremity" OR "Upper extremities" OR "Hip" OR "Knee" OR "Foot" OR "Hand" OR "Ankle" OR "Wrist" OR "Elbow" OR "Low back" OR "Neck" OR "Back" OR "Spine" OR "Shoulder" OR "Arm" OR "Leg" OR "Muscle" OR "Tendon" OR "Ligament"	"Digital movement therapy" OR "Digital movement therapies" OR "Mobile health" OR "eTherapy" OR "eTherapies" OR "Web-based intervention" OR "Digital intervention" OR "Computer-based intervention" OR "App-based intervention" OR "Digital health application" OR "Technology-assisted therapy" OR "Technology-assisted therapies" OR "Internet-based intervention" OR "Computer-assisted therapy" OR "Computer-assisted therapies" OR "health app" OR "mobile application" OR "Smartphone" OR "Mobile phone" OR "ehealth" OR "mhealth" OR "telerehabilitation" OR "Telemedicine" OR "online intervention" OR "internet-delivered intervention"	"Osteopathy" OR "movement therapy" OR "movement therapies" OR "physical therapy" OR "physical therapies" OR "therapeutic exercise" OR "medical gymnastic" OR "traditional therapy" OR "traditional therapies" OR "manual therapy" OR "manual therapies" OR "physiotherapy" OR "No therapy" OR "No therapies" OR "conventional therapy" OR "conventional therapies" OR "no treatment" OR "no intervention" OR "watch-and-wait" OR "wait-and-see" OR "watch and wait" OR "wait and see"	N/A[d]	"Randomized controlled trials"

[a]PICOS: Population, intervention, comparison, outcome, and study design.

[b]Others: Studies in English or German language with free full access were included.

[c]WHO: World Health Organization.

[d]N/A: Not applicable.

Textbox 1. Search line.

"(Musculoskeletal disease OR musculoskeletal disorder OR musculoskeletal pain OR chronic pain OR acute pain OR overuse pain OR chronic injury OR chronic injuries OR acute injury OR acute injuries OR overuse injury OR overuse injuries OR chronic disease OR acute disease OR overuse disease OR osteoporosis OR osteoarthritis OR rheumatoid arthritis OR tendinopathy OR tendinopathies OR rotator cuff OR lower extremity OR lower extremities OR upper extremity OR upper extremities OR hip OR knee OR foot OR hand OR ankle OR wrist OR elbow OR low back OR neck OR back OR spine OR shoulder OR arm OR leg OR muscle OR tendon OR ligament) AND (digital movement therapy OR digital movement therapies OR mobile health OR etherapy OR etherapies OR web-based intervention OR digital intervention OR computer-based intervention OR app-based intervention OR digital health application OR technology-assisted therapy OR technology-assisted therapies OR internet-based intervention OR computer-assisted therapy OR computer-assisted therapies OR health app OR mobile application OR smartphone OR mobile phone OR ehealth OR mhealth OR telerehabilitation OR telemedicine OR online intervention OR internet-delivered intervention) AND (osteopathy OR movement therapy OR movement therapies OR physical therapy OR physical therapies OR therapeutic exercise OR medical gymnastic OR traditional therapy OR traditional therapies OR manual therapy OR manual therapies OR physiotherapy OR no therapy OR no therapies OR conventional therapy OR conventional therapies OR no treatment OR no intervention OR watch-and-wait OR wait-and-see OR watch and wait OR wait and see) AND (randomized controlled trials)."

Literature Search, Study Selection, and Risk of Bias

The literature search was performed on July 21, 2022, using the PubMed (MEDLINE), BISp (Federal Institute of Sport Science), Cochrane Library, and Web of Science databases. The search line included terms presented in Table 1. The "outcomes" category was not included in the search strategy but was considered in the subsequent study inclusion and selection process. No filters or other restrictions were used. The retrieved records were exported to a reference manager (EndNote 20, Clarivate). All duplicates were identified using the software and were removed after a manual review. On the basis of the defined eligibility criteria, studies were included or excluded by reviewing the titles, abstracts, and full texts. Full texts were accessed via public or open access and university accounts. If the full texts were not accessible, the authors were contacted. The study quality and the associated risk of bias were assessed using the Scottish Intercollegiate Guidelines Network checklist for randomized controlled trials [17]. The checklist consisted of 10 items related to the internal validity and 2 items related to the overall assessment of the studies. For each included study, all items were answered with "yes," "no," "can't say," or "not applicable." The study quality was then finally rated throughout the "Scottish Intercollegiate Guidelines Network checklist for randomized controlled trials: Notes for completion of checklist" as "not acceptable," "borderline," "acceptable," and "high," as previously done [18]. The definitions of these quality classifications are presented in Textbox 2.

Textbox 2. Definitions for ratings of the overall methodological study quality.

High quality

- Most criteria met. Little or no risk of bias. Results unlikely to be changed by further research.

Acceptable quality

- Most criteria met. Some flaws in the study with an associated risk of bias. Conclusions may change in the light of further studies.

Borderline quality

- Crude effect estimates have been presented or have been calculated (thus no confounders have been considered), but the study is otherwise acceptably sound with respect to other possible biases.

Not acceptable quality

- Either most criteria not met, or significant flaws relating to key aspects of study design. Conclusions likely to change in the light of further studies.

Note: Definitions according to Asker et al [18].

Data Extraction and Synthesis of Results

Data extraction was performed according to the PICOS scheme. A best-evidence synthesis was conducted to clarify the evidence for digital physical health exercises on clinically relevant factors clustered as (1) localization of the musculoskeletal diseases, (2) patient-reported outcomes (according to the eligibility criteria), and (3) medical treatment types (conservative vs operative). Within these clusters, the study results were individually classified as positive, negative, or equal for each clinically relevant factor. In accordance with a previous study [14], the study results were classified as positive or negative if the intervention or control group showed statistically better significant study results than the other group for >50% of the outcome parameters that were used to examine the respective clinically relevant factors. If no statistically significant differences were reported between the intervention and control groups, the study results were classified as equal. With regard to the best-evidence synthesis, the established criteria [18] are summarized in Table 2, and to increase the validity, only studies with at least acceptable study quality were included [19].

Table 2. Criteria of best-evidence synthesis according to Asker et al [18].

Rating	Study quality	Criteria
Strong evidence	≥2 high-quality studies	≥75% consistent findings in these studies
Moderate evidence	1 high-quality study and/or ≥2 acceptable-quality studies	≥75% consistent findings in these studies
Limited evidence	1 acceptable-quality study and/or ≥1 borderline-quality study	N/A[a]
Conflicting evidence	Conflicting results in several studies of any quality	<75% of studies reported concordant results
No evidence	No admissible studies were found	N/A

[a]N/A: Not applicable.

Results

Literature Search, Study Selection, and Risk of Bias

Figure 1 shows the flowchart of the literature search including the study selection process according to the PRISMA guidelines. On the basis of the eligibility criteria, of 10,441 records, 30 (0.29%) studies were finally included in the risk of bias assessment. Although 1453 studies were identified as duplicates, 8958 studies that did not meet the inclusion criteria addressed no musculoskeletal diseases, were not randomized controlled trials, or addressed other outcomes. Table 3 summarizes the results of the risk of bias assessment.

There were 10 studies with high quality [20-29], 6 studies with acceptable quality [30-35], 10 studies with borderline quality [36-45], and 4 studies with not acceptable quality [46-49]. Thus, because of the not acceptable and borderline qualities of 14 studies, 16 studies were further analyzed and finally included in the best-evidence synthesis.

Figure 1. Flowchart of the literature search including the study selection process according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. RCT: randomized controlled trial.

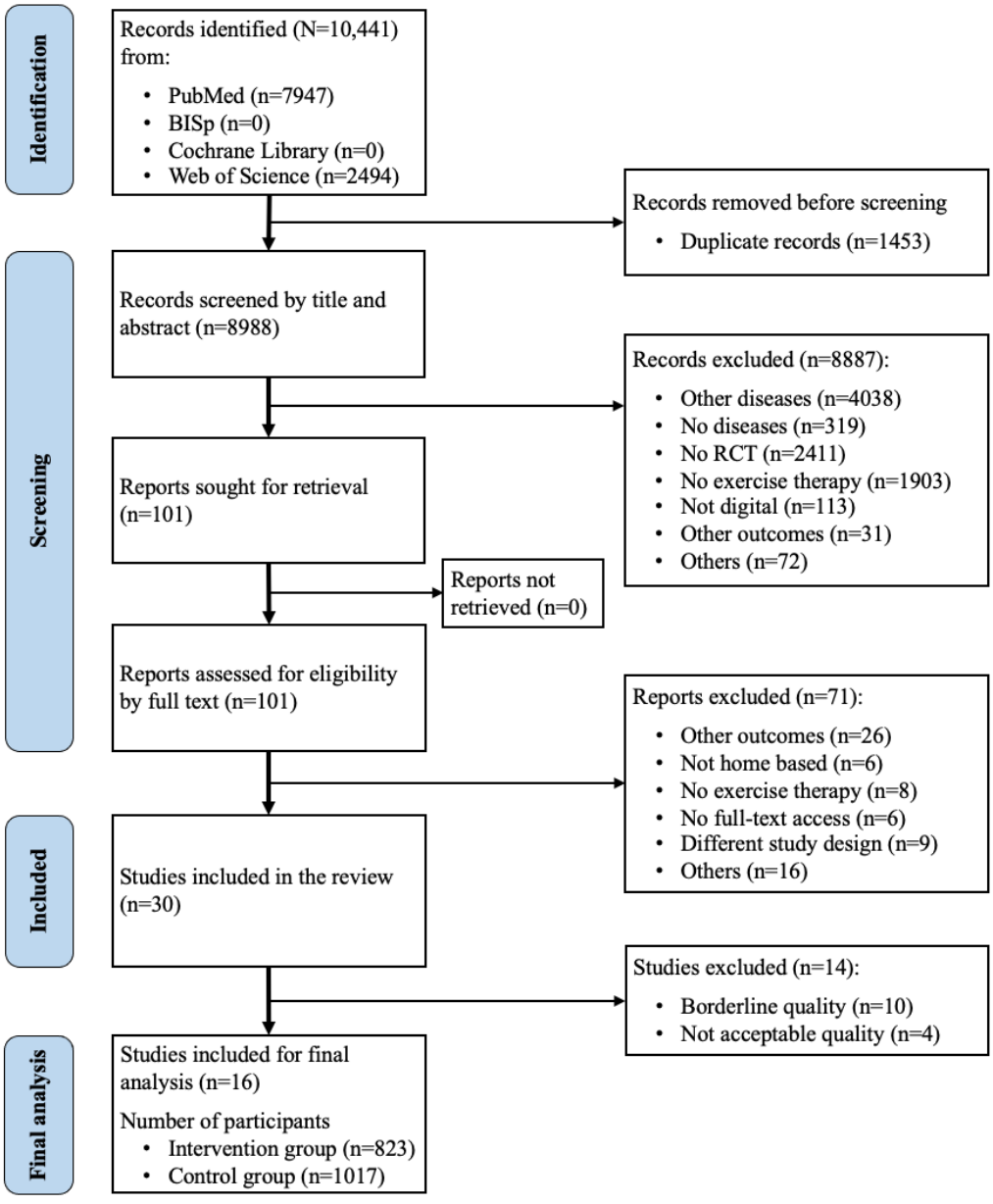


Table 3. Results of the 30 studies checked for the risk of bias assessment using the Scottish Intercollegiate Guidelines Network checklist.

Study	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8 (%)	1.9	1.10	2.1	2.2	2.3	Total				Study quality
														Yes	No	CS[a]	N/A[b]	
Abadiyan et al [20]	Yes	Yes	Yes	No	Yes	Yes	Yes	3	CS	N/A	++[c]	Yes	Yes	8	1	1	1	High
Allen et al [21]	Yes	Yes	Yes	No	Yes	Yes	Yes	13.1	Yes	N/A	++	Yes	Yes	9	1	0	1	High
Blanquero et al [22]	Yes	Yes	Yes	No	Yes	Yes	Yes	0	Yes	N/A	++	Yes	Yes	9	1	0	1	High
Choi et al [23]	Yes	Yes	Yes	CS	Yes	Yes	Yes	0	Yes	N/A	++	Yes	Yes	9	0	1	1	High
Fatoye et al [24]	Yes	Yes	Yes	No	Yes	Yes	Yes	16	CS	N/A	++	Yes	Yes	8	1	1	1	High
Fleischman et al [25]	Yes	Yes	Yes	No	Yes	Yes	Yes	16.6	Yes	N/A	++	Yes	Yes	9	1	0	1	High
Moffet et al [26]	Yes	Yes	Yes	No	Yes	Yes	Yes	6.3	Yes	Yes	++	CS	Yes	9	1	1	0	High
Nelligan et al [27]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	12.6	Yes	Yes	++	Yes	Yes	11	0	0	0	High
Nelson et al [28]	Yes	Yes	Yes	No	CS	Yes	Yes	1	Yes	N/A	++	Yes	Yes	8	1	1	1	High
Özden et al [29]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	7	CS	N/A	++	Yes	Yes	9	0	1	1	High
Bennell et al [30]	Yes	Yes	Yes	No	Yes	Yes	Yes	10.1	Yes	N/A	+[d]	Yes	Yes	9	1	0	1	Acceptable
Chhabra et al [31]	Yes	Yes	Yes	No	No	Yes	Yes	0	Yes	N/A	+	Yes	Yes	8	2	0	1	Acceptable
Hardt et al [32]	Yes	Yes	Yes	No	Yes	CS	Yes	10	Yes	N/A	+	Yes	Yes	8	1	1	1	Acceptable
Hernando-Garijo et al [33]	Yes	Yes	Yes	No	Yes	Yes	Yes	18	Yes	N/A	+	Yes	No	8	2	0	1	Acceptable
Rodríguez Sánchez-Laulhé et al [34]	Yes	Yes	Yes	No	No	Yes	Yes	16	Yes	N/A	+	Yes	Yes	8	2	0	1	Acceptable
Tousignant et al [35]	Yes	Yes	Yes	CS	Yes	Yes	Yes	15	CS	CS	+	Yes	Yes	8	0	3	0	Acceptable
Anan et al [36]	Yes	Yes	Yes	No	Yes	CS	Yes	25.6	Yes	N/A	_[e]	CS	Yes	7	1	2	1	Borderline
Bäcker et al [37]	Yes	Yes	Yes	No	Yes	Yes	Yes	42	No	N/A	−	Yes	Yes	8	2	0	1	Borderline
Bossen et al [38]	Yes	Yes	Yes	No	Yes	Yes	Yes	24.6	Yes	N/A	−	Yes	Yes	9	1	0	1	Borderline
Correia et al [39]	Yes	Yes	Yes	No	Yes	Yes	Yes	36	No	N/A	−	Yes	Yes	8	2	0	1	Borderline
del Pozo-Cruz et al [40]	Yes	CS	CS	No	Yes	Yes	Yes	10	Yes	Yes	−	Yes	CS	7	1	3	0	Borderline

Study	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8 (%)	1.9	1.10	2.1	2.2	2.3	Total				Study quality
														Yes	No	CS[a]	N/A[b]	
Gohir et al [41]	Yes	Yes	Yes	No	Yes	Yes	Yes	28.1	Yes	N/A	−	Yes	Yes	9	1	0	1	Borderline
Kloek et al [42]	Yes	Yes	Yes	No	Yes	Yes	Yes	35.1	Yes	CS	−	Yes	No	8	2	1	0	Borderline
Piqueras et al [43]	Yes	Yes	Yes	No	No	Yes	Yes	26.5	CS	N/A	−	Yes	CS	6	2	2	1	Borderline
Punt et al [44]	Yes	Yes	Yes	No	Yes	Yes	Yes	21	Yes	N/A	−	CS	Yes	8	1	1	1	Borderline
Sandal et al [45]	Yes	Yes	Yes	No	Yes	Yes	Yes	23.6	Yes	CS	−	CS	Yes	8	1	2	0	Borderline
Lara et al [46]	Yes	CS	CS	No	Yes	Yes	Yes	4	Yes	N/A	−−[f]	Yes	No	6	2	2	1	Not acceptable
Lorig et al [47]	Yes	CS	CS	No	Yes	No	Yes	23.9	Yes	CS	−−	No	No	4	4	3	0	Not acceptable
Shebib et al [48]	Yes	Yes	Yes	No	No	No	Yes	40.7	Yes	CS	−−	CS	Yes	6	3	2	0	Not acceptable
Toelle et al [49]	Yes	No	No	No	Yes	Yes	Yes	14.9	No	N/A	−−	CS	CS	4	4	2	1	Not acceptable

[a]CS: Cannot say.

[b]N/A: Not applicable.

[c]Low or no risk of bias.

[d]Associated risk of bias.

[e]Crucial risk of bias.

[f]High risk of bias.

Study Characteristics

Table 4 presents the study characteristics of the 16 included studies according to the PICOS scheme.

The publication period ranged from 2011 [35] to 2022 [29,34], whereby 3 studies were published each in 2018 [21,31,32], 2019 [22,23,25], and 2021 [20,27,33]. The most common publication country was Australia, with 3 studies [27,28,30], followed by 2 studies each published by Spain [22,34], Canada [26,35], and the United States [21,25]. Across the 16 studies, 1840 patients were investigated, and the sample sizes ranged from 34 [33] to 350 patients [21]. The reported dropout rate was up to 18% (6/34) [33]. The average age of the patients varied from 38.5 [20] to 66 years [26,35], and the average female proportion across all studies reported was 61.35% (1008/1643) and varied from 51% (148/290) [25] to 100% (34/34) [33]. Regarding the localization of the musculoskeletal diseases, 7 studies were on knee-specific diseases such as total knee arthroplasty [25,26,32,35], knee osteoarthritis [21,27], and chronic knee pain [30]. This was followed by 4 studies on back-specific diseases such as low back pain [24,29,31] and chronic neck pain [20]. A total of 2 studies were on hand-specific diseases [22,34], whereas only 1 study was found for each full body [33], shoulder-specific diseases [23], and hip-specific diseases [28]. Regarding the patient-reported outcomes, 14, 12, 6, and 5 studies investigated pain [20-23,25-27,29-35], function [21,23,25-30,32-35], disability [20,22,24,29,31,34], and quality of life [20,27-30], respectively. In the 16 included studies, 26 different patient-reported outcomes were investigated. With regard to the digital health interventions, 7, 5, and 4 studies used app-based [20,22,23,28,31,32,34], web-based [21,25,27,29,30], and telerehabilitation-based physical exercises [24,26,33,35], respectively, whereby the duration of the digital health interventions ranged from 7 days [32] to 12 months [21]. As control groups, 9 studies used physiotherapy [20,21,23,25,26,28,31,32,35]; 4 studies used paper-based exercises [22,25,29,34]; 2 studies used internet-based information material [27,30]; and 1 study each used global postural re-education [20], waiting list [21], clinic-based McKenzie therapy [24], and no therapy [33].

Table 4. Summary table of all study characteristics according to the population, intervention, comparison, outcome, study design (PICOS) scheme.

Study	Population and setting	Intervention and assessment	Outcomes

Abadiyan et al [20]	• Sample size: n=60
	• Average age: 38.5 y
	• Female: 55%
	• Disease: chronic neck pain
	• Country: Iran
		• I[a] ("Seeb" app+ GPR[b]): n=20
		• C1[c] (GPR alone): n=20
		• C2[d] (conventional PT[e]) n=20
		• Duration: 8 wk
		• Survey dates: baseline, 8 wk
			• Drop out: 3%
			• I: 5%, C1: 5%, and C2: 0%
			• Pain:
			  • app+GPR>GPR
			  • app+GPR>PT
			• Neck disability index:
			  • app+GPR>GPR
			  • app+GPR>PT
			  • GPR>PT
			• Quality of life:
			  • app+GPR>PT
			  • GPR>PT

Allen et al [21]	• Sample size: n=350 • Average age: 65.3 y • Female: 71.7% • Disease: knee osteoarthritis • Country: United States	• I (IBET[f]): n=142 • C1 (conventional PT): n=140 • C2 (WL[g]): n=68 • Duration: 12 mo • Survey dates: baseline, 4, and 12 mo	• Drop out: 13.1% • I: 21.1%, C1: 7.9%, and C2: 7% • Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC[h]) and other functional tests • IBET=PT=WL
Bennell et al [30]	• Sample size: n=148 • Average age: 61.2 y • Female: 56.1% • Disease: chronic knee pain • Country: Australia	• I: internet-based education material supported by videoconferences with physiotherapist for home exercises (n=74) • Control group: internet-based education material only (n=74) • Duration: 9 mo • Survey dates: baseline, 3, and 9 mo	• Drop out: 10.1% • I: 11% and C: 10% • Pain and function: education+PT>education • Quality of life: education+PT>education
Blanquero et al [22]	• Sample size: n=50 • Average age: 50.0 y • Female: 82% • Disease: carpal tunnel release • Country: Spain	• I: ReHand app for physical home training (n=25) • Control group: paper and home-based physical exercise program (n=25) • Duration: 4 wk • Survey dates: baseline, 4 wk	• Drop out: 0% • Hand disability and pain: app based>paper based
Chhabra et al [31]	• Sample size: n=93 • Average age: 41.2 y • Female: not reported • Disease: chronic low back pain • Country: India	• I: Snapcare app for physical home training (n=45) • Control group: conventional therapy (n=48) • Duration: 12 wk • Survey dates: baseline, 12 wk	• Drop out: 0% • Pain: app based=conventional • Disability: app based>conventional • Current Symptom Score: app based>conventional
Choi et al [23]	• Sample size: n=84 • Average age: 54.5 y • Female: 68% • Disease: frozen shoulder • Country: Korea	• I: app (no name given) for physical home training (n=42) • Control group: conventional home-based self-exercises (n=42) • Duration: 3 mo • Survey dates: baseline, 4, 8, and 12 wk	• Drop out: 0% • Pain and range of motion: app based=conventional
Fatoye et al [24]	• Sample size: n=56 • Average age: 48.7 y • Female: not reported • Disease: chronic low back pain • Country: Nigeria	• I: telerehabilitation home-based McKenzie therapy (TBMT[i]; n=24) • Control group: clinic-based McKenzie therapy (CBMT[j]; n=32) • Duration: 8 wk • Survey dates: baseline, 4, and 8 wk	• Drop out: 16% • I: 13% and C: 19% • Disability: TBMT=CBMT

Study	Population and setting	Intervention and assessment	Outcomes
Fleischman et al [25]	• Sample size: n=290 • Average age: 65.0 y • Female: 51% • Disease: total knee arthroplasty • Country: United States	• I: web-based PT at home (n=96) • C1: paper-based PT at home (n=97) • C2: formal outpatient PT (n=97) • Duration: 6 mo • Survey dates: baseline, 4-6 wk, 6 mo	• Drop out: 15.9% • I: 17%, C1: 27%, and C2: 6% • Knee flexion and Knee Injury and Osteoarthritis Outcome Score (KOOS[k]): • Web PT=paper PT=PT
Hardt et al [32]	• Sample size: n=60 • Average age: 65.9 y • Female: 57% • Disease: total knee arthroplasty • Country: Germany	• I: PT+"GenuSport" app (PT+app; n=33) • Control group: PT (n=27) • Duration: 7 d • Survey dates: daily for 7 d	• Drop out: 10% • I: 15% and C: 7% • Active range of motion, pain, function, KOOS, and Knee Society Score: PT+app>PT
Hernando-Garijo et al [33]	• Sample size: n=34 • Average age: 53.4 y • Female: 100% • Disease: fibromyalgia • Country: Mexico	• I: telerehabilitation with home-based aerobic exercises (n=17) • Control group: no additional intervention (n=17) • Duration: 15 wk • Survey dates: baseline, 15 wk	• Drop out: 18% • I: 18% and C: 18% • Pain: telerehabilitation>nothing • Physical function: telerehabilitation=nothing
Moffet et al [26]	• Sample size: n=205 • Average age: 66.0 y • Female: 51.2% • Disease: total knee arthroplasty • Country: Canada	• I: home-based telerehabilitation (n=104) • Control group: home-visiting PT (n=101) • Duration: 2 mo • Survey dates: baseline, 2, and 4 mo	• Drop out: 6.3% • I: 9.6% and C: 2.9% • WOMAC, KOOS, function, and range of motion: telerehabilitation=PT
Nelligan et al [27]	• Sample size: n=206 • Average age: 60.0 y • Female: 61.2% • Disease: knee osteoarthritis • Country: Australia	• I: website (information+active exercises) and text messages (n=103) • Control group: website with information only (n=103) • Duration: 24 wk • Survey dates: baseline, 24 wk	• Drop out: 12.6% • I: 12.6% and C: 12.6% • Pain, WOMAC, KOOS, quality of life: website information+exercise>website information only
Nelson et al [28]	• Sample size: n=70 • Average age: 64.5 y • Female: 63% • Disease: total hip replacement • Country: Australia	• I: telerehabilitation and technology-based home exercise (n=35) • Control group: PT and paper-based home exercise (n=35) • Duration: 6 wk • Survey dates: baseline, 6 wk, 6 mo	• Drop out: 1% • I: 3% and C: 0% • Quality of life and function: telerehabilation+exercise=PT+exercise
Özden et al [29]	• Sample size: n=50 • Average age: 41.3 y • Female: 60% • Disease: low back pain • Country: turkey	• I: telerehabilitation with Fizyoweb software (n=25) • Control group: same exercises with paper-based instructions (n=25) • Duration: 8 wk • Survey dates: baseline, 8 wk	• Drop out: 7% • I: 7% and C: 7% • Pain, function, disability, and quality of life: telerehabilation>paper based
Rodríguez Sánchez-Laulhé et al [34]	• Sample size: n=36 • Average age: 59.8 y • Female: 61% • Disease: rheumatoid arthritis • Country: Spain	• I: CareHand app for exercises and self-management and monitoring tools (n=14) • Control group: paper-based home exercises (n=22) • Duration: 3 mo • Survey dates: baseline, 1, 3, and 6 mo	• Drop out: 16% • I: 7% and C: 22% • Function: app based>paper based • Pain and disability for upper extremity: app based=paper based
Tousignant et al [35]	• Sample size: n=48 • Average age: 66.0 y • Female: not reported • Disease: total knee arthroplasty • Country: Canada	• I: telerehabilitation by videoconference with a physiotherapist (n=24) • Control group: conventional PT (n=24) • Duration: 2 mo • Survey dates: baseline, 2, and 6 mo	• Drop out: 15% • I: 12% and C: 17% • Disability: telerehabilation=conventional PT • Function: telerehabilation>conventional PT • Functional activity, physical functioning, and physical pain: conventional PT>telerehabilation

[a]I: Intervention group.

[b]GPR: Global postural re-education.

[c]C1: Control group 1.
[d]C2: Control group 2.
[e]PT: Physiotherapy.
[f]IBET: Internet-based exercise training.
[g]WL: Waitlist.
[h]WOMAC: Western Ontario and McMaster Universities Osteoarthritis Index.
[i]TBMT: Telerehabilitation home-based McKenzie therapy.
[j]CBMT: Clinic-based McKenzie therapy.
[k]KOOS: Knee Injury and Osteoarthritis Outcome Score.

Synthesis of Results by Best‐Evidence Synthesis

Tables 5, 6, and 7 show the results of the best-evidence synthesis with regard to the cluster of the localization of the musculoskeletal diseases, patient-reported outcomes, and medical treatment types, respectively.

Regarding the localization of the musculoskeletal diseases, there was strong evidence that digital physical health exercises had a positive impact on the musculoskeletal diseases located in the back. Although moderate evidence was obtained for diseases located in the shoulder and hip, evidence for fibromyalgia (the entire body) is limited. Conflicting evidence was found for diseases located in the knee and hand. For the patient-reported outcomes, there was strong evidence that digital physical health exercises had a positive impact on disability and quality of life. Conflicting evidence was found for pain and function. Regarding the medical treatment types, operative and conservative therapies both achieved conflicting evidence. Figure 2 shows the evidence found across the 3 defined clusters for studies included in the best-evidence synthesis.

Table 5. Best-evidence synthesis for the localization of the musculoskeletal diseases.

Localization	Study	Musculoskeletal disease	Results	Study quality	Evidence
Back	Abadiyan et al [20]	Chronic neck pain	+[a]	High	Strong[b]
Back	Chhabra et al [31]	Chronic low back pain	+	Acceptable	Strong[b]
Back	Fatoye et al [24]	Chronic low back pain	=[c]	High	Strong[b]
Back	Özden et al [29]	Chronic low back pain	+	High	Strong[b]
Shoulder	Choi et al [23]	Frozen shoulder	=	High	Moderate
Hip	Nelson et al [28]	Total hip arthroplasty	=	High	Moderate
Full body	Hernando-Garijo et al [33]	Fibromyalgia	=	Acceptable	Limited
Knee	Allen et al [21]	Knee osteoarthritis	=	High	Conflicting[b]
Knee	Bennell et al [30]	Chronic knee pain	+	Acceptable	Conflicting[b]
Knee	Fleischman et al [25]	Total knee arthroplasty	=	High	Conflicting[b]
Knee	Hardt et al [32]	Total knee arthroplasty	+	Acceptable	Conflicting[b]
Knee	Moffet et al [26]	Total knee arthroplasty	=	High	Conflicting[b]
Knee	Nelligan et al [27]	Knee osteoarthritis	+	High	Conflicting[b]
Knee	Tousignant et al [35]	Total knee arthroplasty	+	Acceptable	Conflicting[b]
Hand	Blanquero et al [22]	Carpal tunnel release	+	High	Conflicting[b]
Hand	Rodríguez Sánchez−Laulhéet al [34]	Rheumatoid arthritis	=	Acceptable	Conflicting[b]

[a]>50% of the outcomes were significantly better in the intervention group than in the control group.
[b]The level of evidence was determined from all studies in the same localization.
[c]No statistically significant difference between the intervention and control groups.

Table 6. Best-evidence synthesis for the patient-reported outcomes of the musculoskeletal diseases.

Outcomes	Study	Assessment tools	Results	Study quality	Evidence
Disability	Abadiyan et al [20]	Neck Disability Index	+[a]	High	Strong[b]
Disability	Blanquero et al [22]	Disabilities of Arm, Shoulder and Hand Questionnaire	+	High	Strong[b]
Disability	Chhabra et al [31]	Modified Oswestry Disability Index	+	Acceptable	Strong[b]
Disability	Fatoye et al [24]	Oswestry Disability Index	=[c]	High	Strong[b]
Disability	Özden et al [29]	Oswestry Disability Index	+	High	Strong[b]
Disability	Rodríguez Sánchez-Laulhé et al [34]	Disabilities of Arm, Shoulder and Hand Questionnaire	=	Acceptable	Strong[b]
Quality of life	Abadiyan et al [20]	Short Form Health 36 Questionnaire	+	High	Strong[b]
Quality of life	Bennell et al [30]	Assessment of Quality of Life−2	+	Acceptable	Strong[b]
Quality of life	Nelligan et al [27]	Assessment of Quality of Life-6D	+	High	Strong[b]
Quality of life	Nelson et al [28]	Short Form Health 12 Questionnaire/European Quality of Life 5 Dimensions 5 Level Version	=	High	Strong[b]
Quality of life	Özden et al [29]	Short Form Health 36 Questionnaire	+	High	Strong[b]
Pain	Abadiyan et al [20]	Visual analog scale	+	High	Conflicting[b]
Pain	Allen et al [21]	WOMAC[d]	=	High	Conflicting[b]
Pain	Bennell et al [30]	Numeric rating scale	+	Acceptable	Conflicting[b]
Pain	Blanquero et al [22]	Visual analog scale	+	High	Conflicting[b]
Pain	Chhabra et al [31]	Numeric rating scale, Current Symptom Score	=	Acceptable	Conflicting[b]
Pain	Choi et al [23]	Visual analog scale	=	High	Conflicting[b]
Pain	Fleischman et al [25]	KOOS[e]	=	High	Conflicting[b]
Pain	Hardt et al [32]	Numeric rating scale	+	Acceptable	Conflicting[b]
Pain	Hernando-Garijo et al [33]	Visual analog scale	+	Acceptable	Conflicting[b]
Pain	Moffet et al [26]	WOMAC	=	High	Conflicting[b]
Pain	Nelligan et al [27]	Numeric rating scale	+	High	Conflicting[b]
Pain	Özden et al [29]	Visual analog scale	+	High	Conflicting[b]
Pain	Rodríguez Sánchez-Laulhé et al [34]	Visual analog scale	=	Acceptable	Conflicting[b]
Pain	Tousignant et al [35]	WOMAC	−[f]	Acceptable	Conflicting[b]
Function	Allen et al [21]	WOMAC/30-s chair stand test/Timed up and go test/2-min step test, single-leg stand	=	High	Conflicting[b]
Function	Bennell et al [30]	WOMAC	+	Acceptable	Conflicting[b]
Function	Choi et al [23]	Range of motion	=	High	Conflicting[b]
Function	Fleischman et al [25]	KOOS	=	High	Conflicting[b]
Function	Hardt et al [32]	Range of motion/Timed up and go test/10-m walk test/30-s chair stand test/Knee Society Score	+	Acceptable	Conflicting[b]
Function	Hernando-Garijo et al [33]	Arm curl test, 6-min walk test	=	Acceptable	Conflicting[b]
Function	Moffet et al [26]	KOOS/Stair test/6-min walk test	=	High	Conflicting[b]

Outcomes	Study	Assessment tools	Results	Study quality	Evidence
Function	Nelligan et al [27]	WOMAC, KOOS	+	High	Conflicting[b]
Function	Nelson et al [28]	Timed up and go test	=	High	Conflicting[b]
Function	Özden et al [29]	Timed up and go test	+	High	Conflicting[b]
Function	Rodríguez Sánchez-Laulhé et al [34]	Michigan Hand Outcome Questionnaire	+	Acceptable	Conflicting[b]
Function	Tousignant et al [35]	WOMAC/Timed up and go test/Functional Autonomy Measurement System	−	Acceptable	Conflicting[b]

[a]>50% of the outcomes were significantly better in the intervention group than in the control group.

[b]The level of evidence was determined from all studies in the same outcomes.

[c]No statistically significant differences between the intervention and control groups.

[d]WOMAC: Western Ontario and McMaster Universities Osteoarthritis Index.

[e]KOOS: Knee Injury and Osteoarthritis Outcome Score.

[f]>50% of the outcomes were significantly better in the control group than in the intervention group.

Table 7. Best-evidence synthesis for the medical treatment types.

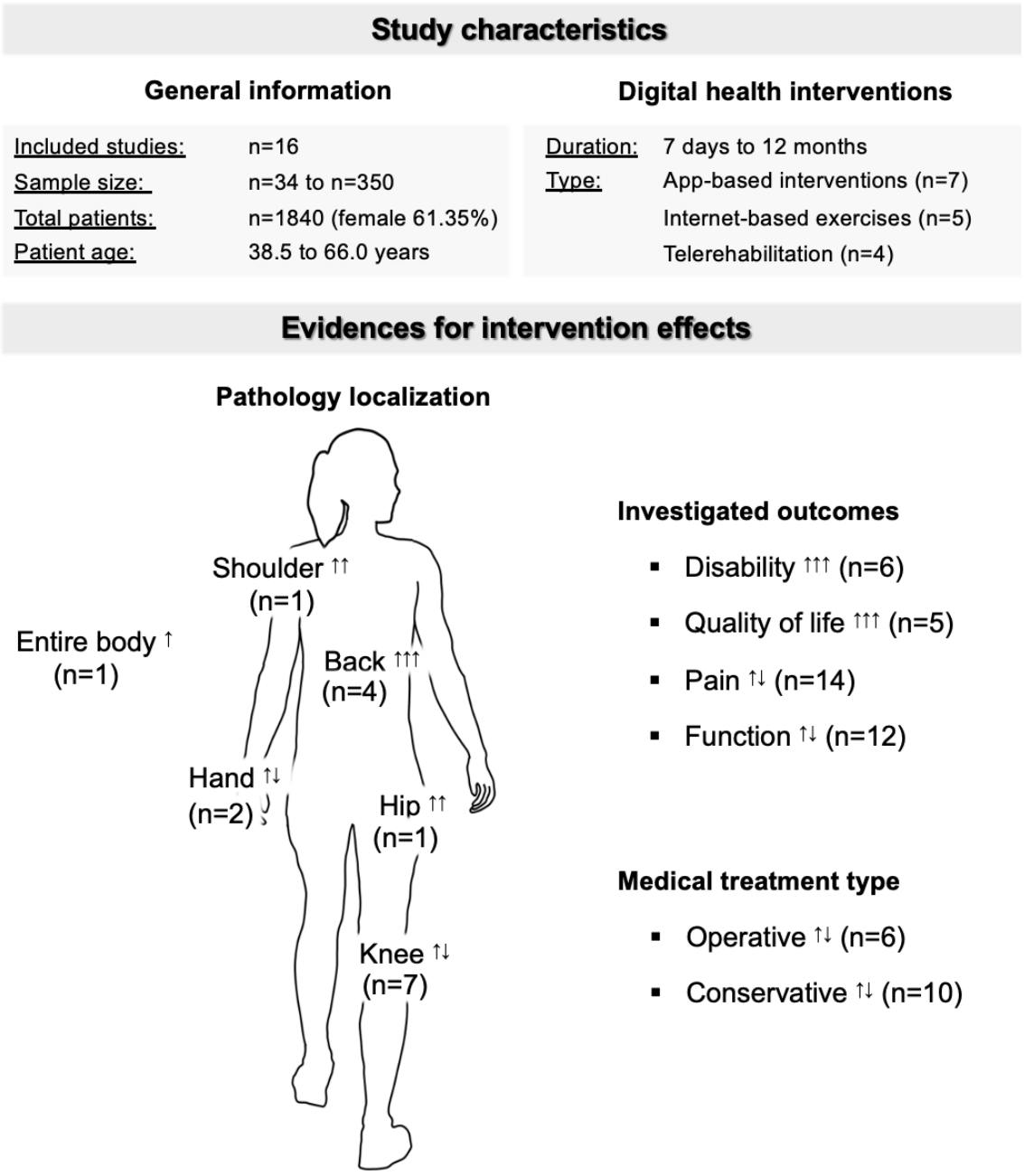
Therapy	Study	Musculoskeletal disease	Results	Study quality	Evidence
Operative	Blanquero et al [22]	Carpal tunnel release	+[a]	High	Conflicting[b]
Operative	Fleischman et al [25]	Total knee arthroplasty	=[c]	High	Conflicting[b]
Operative	Hardt et al [32]	Total knee arthroplasty	+	Acceptable	Conflicting[b]
Operative	Moffet et al [26]	Total knee arthroplasty	=	High	Conflicting[b]
Operative	Nelson et al [28]	Total hip arthroplasty	=	High	Conflicting[b]
Operative	Tousignant et al [35]	Total knee arthroplasty	+	Acceptable	Conflicting[b]
Conservative	Abadiyan et al [20]	Chronic neck pain	+	High	Conflicting[b]
Conservative	Allen et al [21]	Knee osteoarthritis	=	High	Conflicting[b]
Conservative	Bennell et al [30]	Chronic knee pain	+	Acceptable	Conflicting[b]
Conservative	Chhabra et al [31]	Chronic low back pain	+	Acceptable	Conflicting[b]
Conservative	Choi et al [23]	Frozen shoulder	=	High	Conflicting[b]
Conservative	Fatoye et al [24]	Chronic low back pain	=	High	Conflicting[b]
Conservative	Hernando-Garijo et al [33]	Fibromyalgia	=	Acceptable	Conflicting[b]
Conservative	Nelligan et al [27]	Knee osteoarthritis	+	High	Conflicting[b]
Conservative	Özden et al [29]	Chronic low back pain	+	High	Conflicting[b]
Conservative	Rodríguez Sánchez-Laulhé et al [34]	Rheumatoid Arthritis	=	Acceptable	Conflicting[b]

[a]>50% of the outcomes were significantly better in the intervention group than in the control group.

[b]The level of evidence was determined from all studies in the same therapy.

[c]No statistically significant difference between the intervention and control groups.

Figure 2. Visualization of the evidence found across the 3 defined clusters for studies included in the best-evidence synthesis. ↑↑↑: strong evidence, ↑↑: moderate evidence, ↑: limited evidence, and ↑↓: conflicting evidence.



Discussion

Principal Findings

This systematic review aimed to evaluate the impact of digital physical health exercises on patients with musculoskeletal diseases concerning the localization of the musculoskeletal disease, patient-reported outcomes, and medical treatment types. In addition, a best-evidence synthesis was conducted to estimate the direction and strength of the existing evidence. The main findings were that (1) strong evidence was found for a positive impact on musculoskeletal diseases located in the back and on the patient-reported outcomes of disability and quality of life and (2) moderate evidence was found for a positive impact on musculoskeletal diseases located in the shoulder and hip.

The first main finding was that strong evidence was found for a positive impact on musculoskeletal diseases located in the back and on the patient-reported outcomes of disability and quality of life (Figure 2). Our findings are partly supported by a previous systematic review with a meta-analysis [13] showing moderate-quality evidence for the positive impact on the patient-reported outcome of disability. In contrast to the previous review [13] and to another systematic review [12], conflicting evidence for the patient-reported outcomes of pain and function was found. It should be noted that 1 study [12] found some clinical benefits for pain and function but did not conduct an evidence synthesis or a meta-analysis. In addition, the outcomes of pain and function represent health-related outcomes, and the outcomes of disability and quality of life are the resulting consequences. Therefore, pain acts as a protective mechanism and can lead to disability [50]. With appropriate exercises,

patients learn to compensate for their disabilities [20,30], whereas exercise alone can provoke pain [51]. As disabilities are part of the concept of health-related quality of life [52], these outcomes are mutually dependent, and identical strong evidence is plausible.

In addition, it should be mentioned that both previous systematic reviews included all types of digital health interventions, and we explicitly focused our systematic review on the impact of digital physical health exercises. Regarding this, our findings add that this type of intervention shows strong evidence to have an overall positive impact on the musculoskeletal diseases located in the back, independent of the investigated outcomes [20,24,29,31]. Back-related musculoskeletal diseases usually arise because of muscular causes and are often caused by a lack of physical activity [53]. Participants recruited in back pain–related studies are often middle-aged and have an office occupation [20]. The use of digital physical health exercises in such patients can be considered highly effective because of the increased physical activity targeting muscle strengthening and the teaching of exercise techniques [20,24,29,31]. Overall, the application of digital physical health exercise in patients with musculoskeletal diseases shows versatile positive effects, especially for musculoskeletal diseases located in the back and for the improvement of disabilities and quality of life. However, the type of digital health interventions seems to influence the effects on the specific patient-reported outcome, and more studies to investigate this relationship are needed.

The second main finding was that moderate evidence was found for a beneficial effect on musculoskeletal diseases of the shoulder and hip (Figure 2). As this systematic review is the first to evaluate the association between digital physical health exercises and different localizations of musculoskeletal diseases, no evidence levels from previous research is available for clarification. There is only 1 other systematic review on the effectiveness of digital health interventions for total hip arthroplasty [54]. The review found no significant improvements in the studied patient-reported outcomes. For the shoulder, another systematic review examined the effectiveness of telerehabilitation for musculoskeletal diseases compared with normal in-person physiotherapy [55] and found very low to low evidence. In this context, our findings suggest that digital physical health exercises may also be effective in treating musculoskeletal diseases of the shoulder and hip. However, it should be noted that only 1 study each was found for shoulder- and hip-specific musculoskeletal diseases, whereas several studies were found for back- or knee-specific musculoskeletal diseases (Table 5). Therefore, our results must be interpreted with caution, as a small number of high-quality studies may result in stronger evidence, according to the definitions of the best-evidence synthesis [18], than the presence of many lower-quality studies. Overall, the results demonstrated that digital physical health exercises could have a positive effect on a variety of health-related outcomes, regardless of the localization of the musculoskeletal diseases. However, the number of studies investigating the relationship between the effectiveness of digital health interventions and the localization of musculoskeletal diseases is small, and more studies are needed, especially for localizations other than the knee and back.

An additional interesting finding is the conflicting evidence in the medical treatment types concerning operative and conservative approaches (Figure 2). Although the underlying reasons remain unknown, it can be stated that the operative treatment (ie, carpal tunnel release and total knee arthroplasty) has no impact on the overall stimulus-response mechanism of the digital intervention, requiring further studies for clarification.

Limitations

Although this systematic review increases knowledge of the positive impacts of digital physical health exercises on musculoskeletal diseases, there are a few limitations. Because of the heterogeneity of the included studies (eg, different numbers of patients, interventions, body regions, and control groups), a meta-analysis could not be performed. Instead, and as an established alternative approach, a best-evidence synthesis [18] was used. A strength of this approach is that it is possible to estimate an evidence level for various categories despite the large study heterogeneity. However, a limitation is that no quantitative analysis (eg, in terms of statistical significance) can be conducted [56]. An additional limitation of our review is that we did not register the study plan in PROSPERO a priori. The reason is that according to the PRISMA guidelines, registration is currently recommended but not mandatory [15]. Furthermore, all studies that included digital health interventions beyond active exercises were not included. Therefore, some studies could be lost, but the aspect of physical exercise as an established clinical treatment for musculoskeletal diseases could be focused on for the first time.

Conclusions

There is strong to moderate evidence for the beneficial impact of digital physical health exercises for musculoskeletal diseases located in the back, shoulder, and hip. There is limited or conflicting evidence for other localizations. In addition, strong evidence was found for the patient-reported outcomes of disability and quality of life, whereas conflicting evidence exists for other commonly investigated patient-reported outcomes such as pain and function. Thus, digital physical health exercises could have a positive effect on a variety of health-related outcomes of musculoskeletal diseases. To implement digital physical health exercises in evidence-based medicine for musculoskeletal diseases, more high-quality randomized controlled trials are needed to clarify the relationship between the impact of digital physical health exercises and clinically relevant factors such as localization, patient-reported outcomes, and medical treatment types.

Acknowledgments

The study was funded by the Open Access Publishing Fund of Leipzig University supported by the German Research Foundation within the Open Access Publication Funding program. The authors acknowledge the support from Leipzig University for Open Access Publishing.

Authors' Contributions

JN, FW, and MWH conceived the study and were involved in the conceptual process. JN performed the literature search and data analysis, and MWH validated these steps. JN wrote a raw version of the manuscript, and FW and MWH revised the raw version and wrote the final manuscript together with JN. CG critically revised the final manuscript and made additional changes. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA 2020 Checklist.
[DOCX File , 31 KB - mhealth_v12i1e50616_app1.docx]

References

1. Musculoskeletal health. World Health Organization. URL: https://www.who.int/news-room/fact-sheets/detail/musculoskeletal-conditions [accessed 2023-07-03]
2. Humphreys JH, Verstappen SM. The burden of musculoskeletal disease. Medizin 2022 Feb;50(2):82-84. [doi: 10.1016/j.mpmed.2021.11.002]
3. von der Lippe E, Krause L, Prost M, Wengler A, Leddin J, Müller A, BURDEN 2020 Studiengruppe. Prävalenz von Rücken- und Nackenschmerzen in Deutschland. Ergebnisse der Krankheitslast-Studie BURDEN 2020. J Gesundheit Monit 2019 Dec 22;6(S3):1-14 [FREE Full text] [doi: 10.1007/s00103-019-03054-1] [Medline: 31758220]
4. Fuchs J, Kuhnert R, Scheidt-Nave C. 12-Monats-Prävalenz von Arthrose in Deutschland. J Gesundheit Monit 2017;2(3):55-60. [doi: 10.17886/rki-gbe-2017-054]
5. Musculoskeletal health status in Europe - summery report. v5.0. eumusc.net. URL: http://www.eumusc.net/workpackages_wp4.cfm [accessed 2023-07-03]
6. Sebbag E, Felten R, Sagez F, Sibilia J, Devilliers H, Arnaud L. The world-wide burden of musculoskeletal diseases: a systematic analysis of the World Health Organization Burden of Diseases Database. Ann Rheum Dis 2019 Jun;78(6):844-848. [doi: 10.1136/annrheumdis-2019-215142] [Medline: 30987966]
7. Aitken M, Clancy B, Nass D. The growing value of digital health, evidence and impact on human health and the healthcare system. IQVIA Institute for Human Data Science. 2017. URL: https://www.iqvia.com/insights/the-iqvia-institute/reports/the-growing-value-of-digital-health [accessed 2023-07-03]
8. Das Fast-Track-Verfahren für digitale Gesundheitsanwendungen (DiGA) nach § 139e SGB V. BfAM. 2023 Apr 13. URL: https://www.bfarm.de/SharedDocs/Downloads/DE/Medizinprodukte/diga_leitfaden.html [accessed 2023-07-03]
9. Janjua S, Banchoff E, Threapleton CJ, Prigmore S, Fletcher J, Disler RT. Digital interventions for the management of chronic obstructive pulmonary disease. Cochrane-Datenbank-System Rev 2021 Apr 19;4(4):CD013246 [FREE Full text] [doi: 10.1002/14651858. CD013246.pub2] [Medline: 33871065]
10. Widmer RJ, Collins MM, Collins C, Westen CP, Lerman LO, Lerman A. Digital health interventions for the prevention of cardiovascular disease: a systematic review and meta-analysis. Mayo Clin Proc 2015 Apr;90(4):469-480 [FREE Full text] [doi: 10.1016/j.mayocp.2014.12.026] [Medline: 25841251]
11. Eberle C, Stichling S. Clinical improvements by telemedicine interventions managing type 1 and type 2 diabetes: systematic meta-review. J Med Internet Res 2021 Feb 19;23(2):e23244 [FREE Full text] [doi: 10.2196/23244] [Medline: 33605889]
12. Hewitt S, Sephton R, Yeowell G. The effectiveness of digital health interventions in the management of musculoskeletal conditions: systematic literature review. J Med Internet Res 2020 Jun 05;22(6):e15617 [FREE Full text] [doi: 10.2196/15617] [Medline: 32501277]
13. Valentijn PP, Tymchenko L, Jacobson T, Kromann J, Biermann CW, AlMoslemany M, et al. Digital health interventions for musculoskeletal pain conditions: systematic review and meta-analysis of randomized controlled trials. J Med Internet Res 2022 Sep 06;24(9):e37869 [FREE Full text] [doi: 10.2196/37869] [Medline: 36066943]
14. Lin IC, Wiles L, Waller R, Goucke R, Nagree Y, Gibberd M, et al. What does best practice care for musculoskeletal pain look like? Eleven consistent recommendations from high-quality clinical practice guidelines: systematic review. Br J Sport Med 2020 Jan;54(2):79-86. [doi: 10.1136/bjsports-2018-099878] [Medline: 30826805]

15. Seite MJ, McKenzie JE, Bossuyt P, Boutron IC, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
16. Kang T, Zou S, Weng C. Pretraining to recognize PICO rlements from randomized controlled trial literature. Stud Health Technol Informieren 2019 Aug 21;264:188-192 [FREE Full text] [doi: 10.3233/SHTI190209] [Medline: 31437911]
17. Scottish Scottish Intercollegiate Guidelines Network (SIGN). URL: http://www.sign.ac.uk [accessed 2023-07-03]
18. Asker M, Brooke HL, Waldén M, Tranaeus U, Johansson F, Skillgate E, et al. Risk factors for, and prevention of, shoulder injuries in overhead sports: a systematic review with best-evidence synthesis. Br J Sport Med 2018 Oct;52(20):1312-1319. [doi: 10.1136/bjsports-2017-098254] [Medline: 29581141]
19. Hoppe MW, Brochhagen J, Tischler T, Beitzel K, Seil R, Grimm C. Risk factors and prevention strategies for shoulder injuries in overhead sports: an updated systematic review. J Exp Orthop 2022 Aug 16;9(1):78 [FREE Full text] [doi: 10.1186/s40634-022-00493-9] [Medline: 35971013]
20. Abadiyan F, Hadadnezhad M, Khosrokiani Z, Letafatkar A, Akhshik H. Adding a smartphone app to global postural re-education to improve neck pain, posture, quality of life, and endurance in people with nonspecific neck pain: a randomized controlled trial. Versuche 2021 Apr 12;22(1):274 [FREE Full text] [doi: 10.1186/s13063-021-05214-8] [Medline: 33845880]
21. Allen KD, Arbeeva L, Callahan LF, Golightly YM, Goode AP, Heiderscheit BC, et al. Physical therapy vs internet-based exercise training for patients with knee osteoarthritis: results of a randomized controlled trial. Arthrose- Knorpel 2018 Mar;26(3):383-396 [FREE Full text] [doi: 10.1016/j.joca.2017.12.008] [Medline: 29307722]
22. Blanquero J, Cortés-Vega M, García-Frasquet M, Sánchez-Laulhé PR, Nieto Díaz de Los Bernardos E, Suero-Pineda A. Exercises using a touchscreen tablet application improved functional ability more than an exercise program prescribed on paper in people after surgical carpal tunnel release: a randomised trial. J Physiother 2019 Apr;65(2):81-87 [FREE Full text] [doi: 10.1016/j.jphys.2019.02.008] [Medline: 30926400]
23. Choi Y, Nam J, Yang D, Jung W, Windschatten HR, Kim SH. Effect of smartphone application-supported self-rehabilitation for frozen shoulder: a prospective randomized control study. Clin Rehabil 2019 Apr;33(4):653-660. [doi: 10.1177/0269215518818866] [Medline: 30526016]
24. Fatoye F, Gebrye T, Fatoye C, Mbada CE, Olaoye E, Odole W, et al. The clinical and cost-effectiveness of telerehabilitation for people with nonspecific chronic low back pain: randomized controlled trial. JMIR Mhealth Uhealth 2020 Jun 24;8(6):e15375 [FREE Full text] [doi: 10.2196/15375] [Medline: 32357128]
25. Fleischman EI, Crizer M, Tarabichi M, Schmied S, Rothman RH, Lonner JH, et al. Insall Award: Recovery of knee flexion with unsupervised home exercise is not inferior to outpatient physical therapy after TKA: A randomized trial. Clin Orthop Relat Res 2019 Jan;477(1):60-69 [FREE Full text] [doi: 10.1097/CORR.0000000000000561] [Medline: 30794229]
26. Moffet H, Tousignant M, Nadeau S, Mérette C, Boissy P, Corriveau H, et al. In-home telerehabilitation compared with face-to-face rehabilitation after total knee arthroplasty: a noninferiority randomized controlled trial. J Knochengelenk Surg Am 2015 Jul 15;97(14):1129-1141. [doi: 10.2106/JBJS.N.01066] [Medline: 26178888]
27. Nelligan RK, Hinman RS, Kasza J, Crofts SJ, Bennell KL. Effects of a self-directed web-based strengthening exercise and physical activity program supported by automated text messages for people with knee osteoarthritis: a randomized clinical trial. JAMA Praktikant Med 2021 Jun 01;181(6):776-785 [FREE Full text] [doi: 10.1001/jamainternmed.2021.0991] [Medline: 33843948]
28. Nelson M, Bourke M, Crossley K, Russell T. elerehabilitation is non-inferior to usual care following total hip replacement - a randomized controlled non-inferiority trial. Physiotherapie 2020 Jun;107:19-27. [doi: 10.1016/j.physio.2019.06.006] [Medline: 32026820]
29. Özden F, Sarı Z, Karaman Ö, Aydoğmuş H. The effect of video exercise-based telerehabilitation on clinical outcomes, expectation, satisfaction, and motivation in patients with chronic low back pain. Ir J Med Sci 2022 Jun;191(3):1229-1239. [doi: 10.1007/s11845-021-02727-8] [Medline: 34357527]
30. Bennell KL, Nelligan R, Dobson F, Rini C, Keefe F, Kasza J, et al. Effectiveness of an internet-delivered exercise and pain-coping skills training intervention for persons with chronic knee pain: a randomized trial. Ann Praktikantin Med 2017 Apr 04;166(7):453-462. [doi: 10.7326/M16-1714] [Medline: 28241215]
31. Chhabra PW, Sharma S, Verma S. Smartphone app in self-management of chronic low back pain: a randomized controlled trial. Eur Wirbelsäule J 2018 Nov;27(11):2862-2874. [doi: 10.1007/s00586-018-5788-5] [Medline: 30324496]
32. Hardt S, Schulz HE, Pfitzner T, Wassilew G, Horstmann H, Liodakis E, et al. Improved early outcome after TKA through an app-based active muscle training programme-a randomized-controlled trial. Knie-Surg-Sport Traumatol Arthrosc 2018 Nov;26(11):3429-3437. [doi: 10.1007/s00167-018-4918-2] [Medline: 29589050]
33. Hernando-Garijo IC, Ceballos-Laita L, Mingo-Gómez MT, Medrano-de-la-Fuente R, Estébanez-de-Miguel E, Martínez-Pérez M, et al. Immediate effects of a telerehabilitation program based on aerobic exercise in women with fibromyalgia. Int J Environ Res Öffentliche Gesundheit 2021 Feb 20;18(4):2075 [FREE Full text] [doi: 10.3390/ijerph18042075] [Medline: 33672691]
34. Rodríguez Sánchez-Laulhé P, Luque-Romero LG, Barrero-García FJ, Biscarri-Carbonero Á, Blanquero J, Suero-Pineda A, et al. An exercise and educational and self-management program delivered with a smartphone app (CareHand) in adults

with rheumatoid arthritis of the hands: randomized controlled trial. JMIR Mhealth Uhealth 2022 Apr 07;10(4):e35462 [FREE Full text] [doi: 10.2196/35462] [Medline: 35389367]
35. Tousignant M, Moffet H, Boissy P, Corriveau H, Cabana F, Marquis F. A randomized controlled trial of home telerehabilitation for post-knee arthroplasty. J Telemed Telecare 2011;17(4):195-198. [doi: 10.1258/jtt.2010.100602] [Medline: 21398389]
36. Anan T, Kajiki S, Oka H, Fujii T, Kawamata K, Mori K, et al. Effects of an artificial intelligence-assisted health program on workers with neck/shoulder pain/stiffness and low back pain: randomized controlled trial. JMIR Mhealth Uhealth 2021 Sep 24;9(9):e27535 [FREE Full text] [doi: 10.2196/27535] [Medline: 34559054]
37. Bäcker HC, Wu CH, Schulz HE, Weber-Spickschen TS, Perka C, Hardt S. App-based rehabilitation program after total knee arthroplasty: a randomized controlled trial. Arch Orthop Trauma Surg 2021 Sep;141(9):1575-1582 [FREE Full text] [doi: 10.1007/s00402-021-03789-0] [Medline: 33547927]
38. Bossen D, Veenhof C, Van Beek KE, Spreeuwenberg P, Dekker J, De Bakker DH. Effectiveness of a web-based physical activity intervention in patients with knee and/or hip osteoarthritis: randomized controlled trial. J Med Internet Res 2013 Nov 22;15(11):e257 [FREE Full text] [doi: 10.2196/jmir.2662] [Medline: 24269911]
39. Correia FD, Molinos M, Luís S, Carvalho D, Carvalho C, Costa P, et al. Digitally assisted versus conventional home-based rehabilitation after arthroscopic rotator cuff repair: a randomized controlled trial. Am J Phys Med Rehabil 2022 Mar 01;101(3):237-249 [FREE Full text] [doi: 10.1097/PHM.0000000000001780] [Medline: 33935152]
40. del Pozo-Cruz B, Parraca JA, del Pozo-Cruz J, Adsuar JC, Hill J, Gusi N. An occupational, internet-based intervention to prevent chronicity in subacute lower back pain: a randomised controlled trial. J Rehabil Med 2012 Jun;44(7):581-587 [FREE Full text] [doi: 10.2340/16501977-0988] [Medline: 22674240]
41. Gohir SA, Eek F, Kelly A, Abhishek A, Valdes AM. Effectiveness of internet-based exercises aimed at treating knee osteoarthritis: the iBEAT-OA randomized clinical trial. JAMA Netw Geöffnet 2021 Feb 01;4(2):e210012 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.0012] [Medline: 33620447]
42. Kloek CJ, Bossen D, Spreeuwenberg P, Dekker J, de Bakker DH, Veenhof C. Effectiveness of a blended physical therapist intervention in people with hip osteoarthritis, knee osteoarthritis, or both: a cluster-randomized controlled trial. Phys Ther 2018 Jul 01;98(7):560-570 [FREE Full text] [doi: 10.1093/ptj/pzy045] [Medline: 29788253]
43. Piqueras M, Marco E, Coll M, Escalada F, Ballester A, Cinca C, et al. Effectiveness of an interactive virtual telerehabilitation system in patients after total knee arthoplasty: a randomized controlled trial. J Rehabil Med 2013 Apr;45(4):392-396 [FREE Full text] [doi: 10.2340/16501977-1119] [Medline: 23474735]
44. Punt IM, Ziltener JL, Monnin D, Allet L. Wii Fit™ exercise therapy for the rehabilitation of ankle sprains: its effect compared with physical therapy or no functional exercises at all. Scand J Med Sci Sports 2016 Jul;26(7):816-823. [doi: 10.1111/sms.12509] [Medline: 26076737]
45. Sandal LF, Bach K, Øverås CK, Svendsen MJ, Dalager T, Stejnicher Drongstrup Jensen J, et al. Effectiveness of app-delivered, tailored self-management support for adults with lower back pain-related disability: a selfBACK randomized clinical trial. JAMA Intern Med 2021 Oct 01;181(10):1288-1296 [FREE Full text] [doi: 10.1001/jamainternmed.2021.4097] [Medline: 34338710]
46. Lara TR, Kagan RP, Hiratzka SL, Thompson AR, Nazir VO, Mirarchi AJ. Traditional versus digital media-based hand therapy after distal radius fracture. J Hand Surg Am 2022 Mar;47(3):291.e1-291.e8. [doi: 10.1016/j.jhsa.2021.06.018] [Medline: 34366180]
47. Lorig KR, Ritter PL, Laurent DD, Plant K. The internet-based arthritis self-management program: a one-year randomized trial for patients with arthritis or fibromyalgia. Arthritis Rheum 2008 Jul 15;59(7):1009-1017 [FREE Full text] [doi: 10.1002/art.23817] [Medline: 18576310]
48. Shebib R, Bailey JF, Smittenaar P, Perez DA, Mecklenburg G, Hunter S. Randomized controlled trial of a 12-week digital care program in improving low back pain. NPJ Digit Med 2019;2:1 [FREE Full text] [doi: 10.1038/s41746-018-0076-7] [Medline: 31304351]
49. Toelle TR, Utpadel-Fischler DA, Haas K, Priebe JA. App-based multidisciplinary back pain treatment versus combined physiotherapy plus online education: a randomized controlled trial. NPJ Digit Med 2019;2:34 [FREE Full text] [doi: 10.1038/s41746-019-0109-x] [Medline: 31304380]
50. Swieboda P, Filip R, Prystupa A, Drozd M. Assessment of pain: types, mechanism and treatment. Ann Agric Environ Med 2013;Spec no. 1:2-7. [Medline: 25000833]
51. Fullwood D, Means S, Merriwether EN, Chimenti RL, Ahluwalia S, Booker SQ. Toward understanding movement-evoked pain (MEP) and its measurement: a scoping review. Clin J Pain 2021 Jan;37(1):61-78 [FREE Full text] [doi: 10.1097/AJP.0000000000000891] [Medline: 33093342]
52. Bullinger M, Morfeld M. Der SF-36 Health Survey. In: Schöffski O, Schulenburg JM, editors. Gesundheitsökonomische Evaluationen. Berlin, Deutschland: Springer; 2007:387-402.
53. Hildebrandt J. [Musculature as a source of back pain]. Schmerz 2003 Dec;17(6):412-418. [doi: 10.1007/s00482-003-0251-9] [Medline: 14648313]

54. Wang X, Hunter DJ, Vesentini G, Pozzobon D, Ferreira ML. Technology-assisted rehabilitation following total knee or hip replacement for people with osteoarthritis: a systematic review and meta-analysis. BMC Musculoskelet Disord 2019 Nov 03;20(1):506 [FREE Full text] [doi: 10.1186/s12891-019-2900-x] [Medline: 31679511]
55. Gava V, Ribeiro LP, Barreto RP, Camargo PR. Effectiveness of physical therapy given by telerehabilitation on pain and disability of individuals with shoulder pain: A systematic review. Clin Rehabil 2022 Jun;36(6):715-725. [doi: 10.1177/02692155221083496] [Medline: 35230167]
56. Slavin D. Best evidence synthesis: an intelligent alternative to meta-analysis. J Clin Epidemiol 1995 Jan;48(1):9-18 [FREE Full text] [doi: 10.1016/0895-4356(94)00097-a] [Medline: 7853053]

Abbreviations

PICOS: Population Intervention Comparison Outcome Study design
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by L Buis; submitted 06.07.23; peer-reviewed by F Eek, L Tymchenko; comments to author 01.09.23; revised version received 21.10.23; accepted 02.11.23; published 23.01.24.

Please cite as:
Nagel J, Wegener F, Grim C, Hoppe MW
Effects of Digital Physical Health Exercises on Musculoskeletal Diseases: Systematic Review With Best-Evidence Synthesis
JMIR Mhealth Uhealth 2024;12:e50616
URL: https://mhealth.jmir.org/2024/1/e50616
doi:10.2196/50616
PMID:38261356

Review

Mobile and Computer-Based Applications for Rehabilitation Monitoring and Self-Management After Knee Arthroplasty: Scoping Review

Sabhya Pritwani[1], BDS, MPH; Purnima Shrivastava[1], MPT; Shruti Pandey[1], BPT, MPH; Ajit Kumar[2], MPT; Rajesh Malhotra[2], MS; Ralph Maddison[3], PhD; Niveditha Devasenapathy[1], MBBS, PhD

[1]Department of Research & Development, The George Institute for Global Health India, Delhi, India

[2]Department of Orthopaedics, All India Institute of Medical Sciences, Delhi, India

[3]Department of School of Exercise & Nutrition, Institute for Physical Activity and Nutrition, Deakin University, Geelong, Australia

Corresponding Author:
Niveditha Devasenapathy, MBBS, PhD
Department of Research & Development
The George Institute for Global Health India
308, Third floor, Elegance Tower
Plot no 8, Jasola District Centre
Delhi, 110025
India
Phone: 91 11 4158 8091 93
Email: NDevasenapathy@georgeinstitute.org.in

Abstract

Background: Successful post-knee replacement rehabilitation requires adequate access to health information, social support, and periodic monitoring by a health professional. Mobile health (mHealth) and computer-based technologies are used for rehabilitation and remote monitoring. The extent of technology use and its function in post-knee replacement rehabilitation care in low and middle-income settings are unknown.

Objective: To inform future mHealth intervention development, we conducted a scoping review to map the features and functionality of existing technologies and determine users' perspectives on telerehabilitation and technology for self-management.

Methods: We followed the Joanna Briggs Institute methodology for scoping reviews. We searched the Embase, Medline, PsycINFO via OVID, and Cochrane Central Register of Controlled Trials databases for manuscripts published from 2001 onward. We included original research articles reporting the use of mobile or computer-based technologies by patients, health care providers, researchers, or family members. Studies were divided into the following 3 categories based on the purpose: validation studies, clinical evaluation, and end user feedback. We extracted general information on study design, technology features, proposed function, and perspectives of health care providers and patients. The protocol for this review is accessible in the Open Science Framework.

Results: Of the 5960 articles, 158 that reported from high-income settings contributed to the qualitative summary (64 studies on mHealth or telerehabilitation programs, 28 validation studies, 38 studies describing users' perceptions). The highest numbers of studies were from Europe or the United Kingdom and North America regarding the use of a mobile app with or without wearables and reported mainly in the last decade. No studies were from low and middle-income settings. The primary functions of technology for remote rehabilitation were education to aid recovery and enable regular, appropriate exercises; monitoring progress of pain (n=19), activity (n=20), and exercise adherence (n=30); 1 or 2-way communication with health care professionals to facilitate the continuum of care (n=51); and goal setting (n=23). Assessment of range of motion (n=16) and gait analysis (n=10) were the commonly validated technologies developed to incorporate into a future rehabilitation program. Few studies (n=14) reported end user involvement during the development stage. We summarized the reasons for satisfaction and dissatisfaction among users across various technologies.

Conclusions: Several existing mobile and computer-based technologies facilitate post-knee replacement rehabilitation care for patients and health care providers. However, they are limited to high-income settings and may not be extrapolated to low-income settings. A systematic needs assessment of patients undergoing knee replacement and health care providers involved in rehabilitation,

involving end users at all stages of development and evaluation, with clear reporting of the development and clinical evaluation can make post-knee replacement rehabilitation care in resource-poor settings accessible and cost-effective.

(JMIR Mhealth Uhealth 2024;12:e47843) doi:10.2196/47843

KEYWORDS

knee arthroplasty; telerehabilitation; mHealth; rehabilitation; monitoring; self-management; knee; arthroplasty; social support; mHealth intervention; development; scoping review; knee replacement

Introduction

Knee arthroplasty is the gold standard treatment for end-stage osteoarthritis when conservative treatments fail to relieve symptoms [1]. Wound care and postarthroplasty physiotherapy are essential components of this treatment. Poor adherence to physiotherapy could delay the recovery and lead to suboptimal functional outcomes [2]. Beyond in-hospital clinical care and initiation of physical therapy before discharge, continued and reliable access to information, support from health care providers, awareness of the recovery pathway, easy access to rehabilitation centers, and periodic monitoring are influential factors for optimal recovery [3-6]. In addition to an uneventful surgery, postarthroplasty outcomes are associated with several patient-related factors such as their preoperative physical and mental state, comorbidities, social support, and socioeconomic status, emphasizing the need for personalized approaches [7]. Hence, monitoring of the rehabilitation phase is essential, whether at clinics, in rehabilitation units, or at home [8-10].

Technology-assisted remote monitoring methods are increasingly being advocated in high-income countries. There is low to moderate-quality evidence on the superiority of telerehabilitation compared with unsupported home-based rehabilitation and noninferiority compared with clinic-based monitoring with respect to range of motion (ROM), pain, function, quality of life, and cost-effectiveness at 3 months between clinic-based and home-based rehabilitation strategies using technology [11-17]. Hence, current evidence supports the adaptation of technology-based rehabilitation as feasible, as safe, and as good as clinic-based monitoring with an additional benefit of saving out-of-pocket expenditure. Technology-based approaches are diverse, varying from telehealth [17] to virtual reality techniques [13] aimed at improving adherence to physical therapy and facilitating remote monitoring [12] of patient progress during the post-acute rehabilitation phase [18].

Therefore, the aim of this scoping review was to summarize the extent, range, and nature of technology used for provision of rehabilitation or to monitor progress following knee arthroplasty. This scoping review aimed to address the following objectives:

1. To map the characteristic features and functionality of the technologies, guiding or theoretical framework for designing the technology, and evaluation methodologies of mobile technology–based apps for rehabilitation monitoring and self-management following knee arthroplasty
2. To understand the patient and physical therapist perspectives regarding the use of mobile technology–based apps for rehabilitation monitoring and self-management following knee arthroplasty

To our knowledge, there are no existing scoping reviews that address our aims [19]. The information from this review will help us and other researchers make an informed decision on future mobile health (mHealth) interventions for monitoring post-knee arthroplasty rehabilitation care by physiotherapists and orthopedic surgeons and to promote self-management by individuals. This review will also help highlight existing gaps in the context of low and middle-income countries (LMICs).

Methods

We conducted this scoping review following the Joanna Briggs Institute (JBI) methodology for JBI Scoping Reviews [20] and consulted the PRISMA-ScR (Preferred Reporting Items for Systematic Review and Meta-Analyses Extension for Scoping Reviews) checklist for reporting [21]. The protocol was registered at the Open Science Framework [22].

Data Sources and Searches

To identify relevant studies, an electronic database literature search was conducted in the Embase, Medline, PsycINFO via OVID, and Cochrane Central Register of Controlled Trials (CENTRAL) databases using the following key terms: “Knee arthroplasty OR Knee replacement,” “mobile,” “web,” “remote sensor,” “computer,” “telerehabilitation,” and “m-health” (Tables S1 and S2 in Multimedia Appendix 1). The search was executed in October 2021 and updated in August 2023. The search was restricted to 2001 onward. There were no language restrictions during the search. We searched the reference list of included articles to identify potentially eligible studies.

Study Selection

Predefined inclusion criteria were articles reporting the use of mobile or computer apps or any other technologies such as sensor-based devices for delivering or monitoring rehabilitation either scheduled or following knee joint replacement. We also included proof- of-concept papers that described the development process of a mobile or technology-based solution for rehabilitation. The purpose of technology could be for a health care provider to monitor rehabilitation adherence, to aid patient-health care provider communication, to promote self-management, to act as reminders, or to act as a source of education or any other function that is aimed at rehabilitation care following knee replacement. The app or technology could be used by patients, health care providers, researchers, or a family member. Included studies could have been conducted in the community or home for any clinical setting in any geographic region. The studies were required to be original research articles, and we included experimental and observational studies using quantitative or qualitative research methods. Reviews (narrative or systematic reviews), non-English

articles, and articles without abstracts or full texts were excluded.

Data Extraction

Screening of manuscript titles and abstracts was conducted by 2 independent reviewers using the web app Rayyan [23]. Prior to screening, reviewers discussed inclusion and exclusion criteria to ensure consistency between individuals. Two reviewers assessed the eligibility of the full text, and disagreements were resolved by discussion. Systematic reviews were not included in the review but were used to obtain potentially relevant references. Multiple publications originating from a single technology were grouped and presented as 1 study.

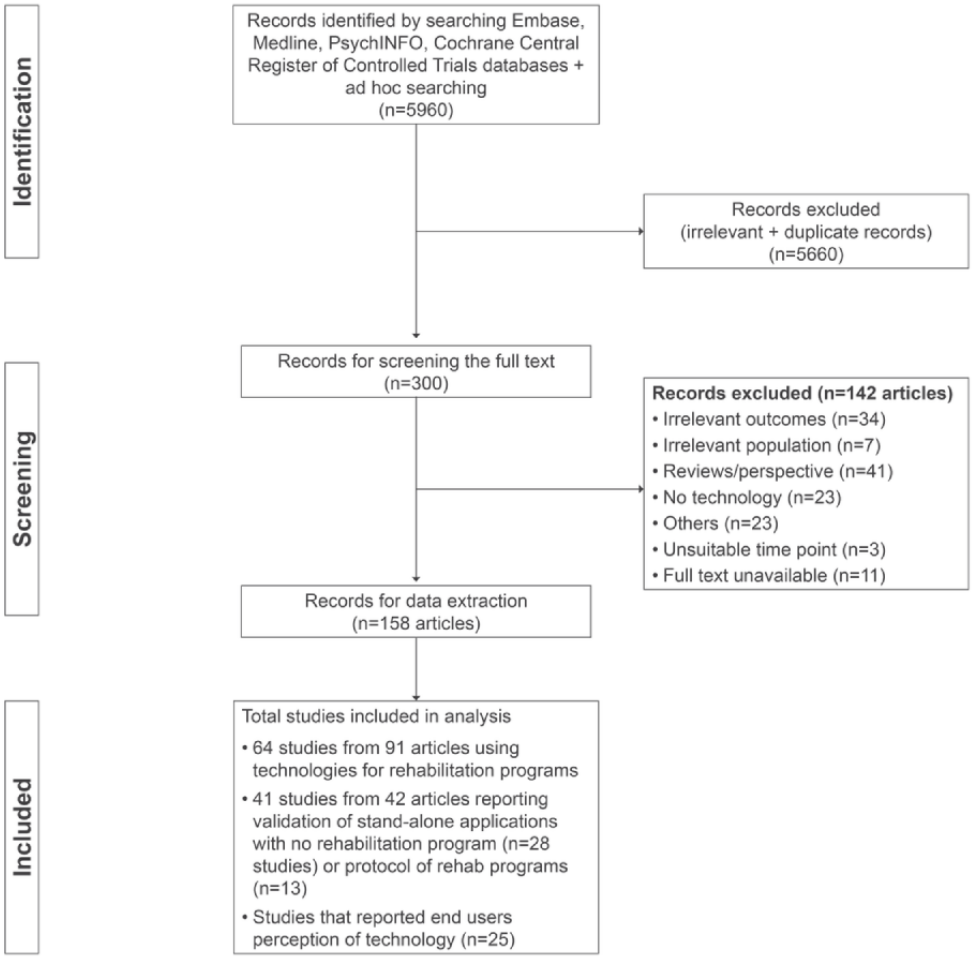
For data charting purposes, the studies were divided into the following 3 categories: (1) studies that had no rehabilitation program but included an app or a technology to assess ROM or gait and were validation studies, (2) studies reporting the use of a mobile or computer app or a telehealth delivery platform for a rehabilitation program with or without sensor-based devices and wearable sensors, (3) studies that reported end users' perceptions of the technology used for rehabilitation monitoring. Data on the general information for the studies, features of the technology, the proposed function, and perspectives of health care providers and patients were extracted and entered in Microsoft Excel. If only the protocol of a planned study was available, there was no information on clinical evaluation, or the study included <6 individuals, we did not extract data beyond the general information.

Results

Search Results

The database search, including the ad hoc search, yielded 5960 articles. Of these articles, 300 articles were considered potentially relevant. Of these, 158 articles were included for data extraction, 131 articles were excluded, and 11 articles were not available (Table S3 in Multimedia Appendix 1). Of the 158 articles, 91 articles (64 studies) reported the clinical evaluation of a technology-based rehabilitation program, 29 articles (28 studies) reported the validation or a proof of concept of technology intended to be used for rehabilitation, and 13 articles were protocols of evaluation studies. In addition, 25 articles reported end users' perceptions on technology (Figure 1) as stand-alone articles or part of clinical evaluation studies (n=13), totaling 38 studies. The 13 studies that reported the perceptions of technology that were also included in rehabilitation program studies were removed from the final list of included full-text articles to avoid double counting.

Figure 1. Process of identifying and including studies according to PRISMA-ScR (Preferred Reporting Items for Systematic Review and Meta-Analyses Extension for Scoping Reviews).

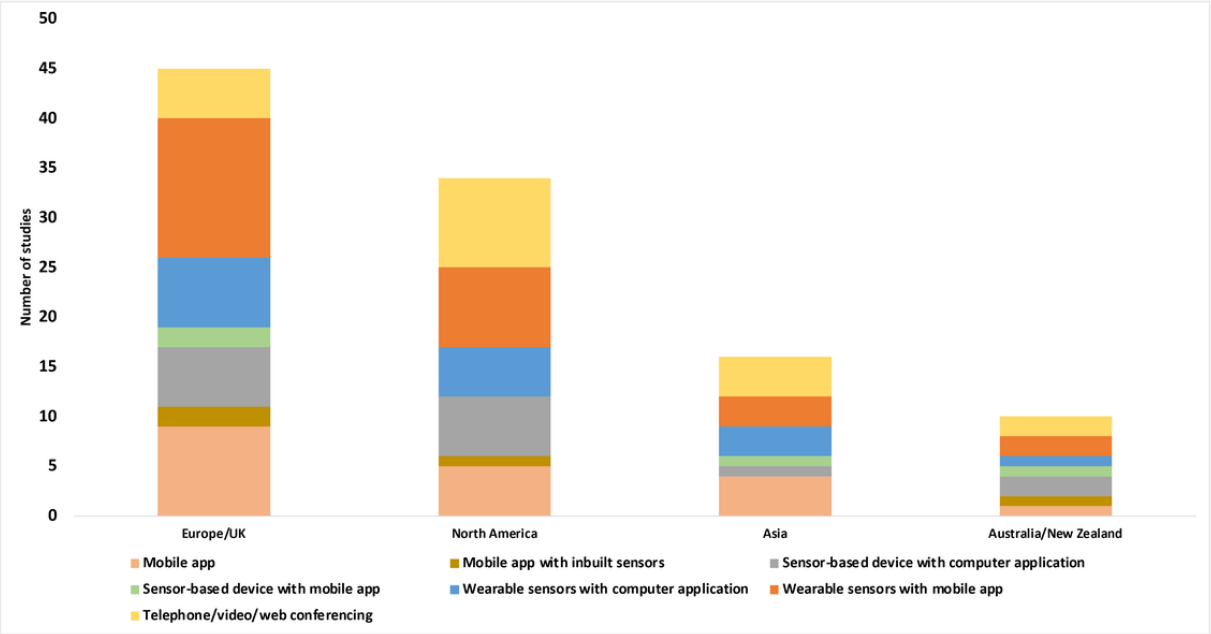


Technology for Rehabilitation

Characteristics of the 105 Studies

Studies were reported from Europe and the United Kingdom (n=45) [11, 24-60, 62-66, 169, 170], North America (n=34) [6, 67-99], Australia and New Zealand (n=10) [100-109], and Asia (n=16) [110-125]. None of the studies were from LMICs. Reports of mobile-based technologies represented the highest number (54/105, 51.4%) [6, 25-27, 31, 32, 36, 37, 41-50, 52, 55, 56, 58-60, 62, 64, 67-69, 74, 75, 78, 80, 82, 83, 90-92, 95, 101, 103, 106-108, 112, 113, 117, 121-126, 169], followed by computer applications (31/105, 29.5%) [24, 29, 30, 33-35, 39, 53, 54, 57, 65, 66, 70, 73, 76, 79, 84-86, 89, 93, 97, 98, 100, 102, 104, 111, 114, 116, 120, 127], and tele/video/web conferencing (20/105, 19%) for rehabilitation monitoring [28, 38, 51, 63, 71, 72, 77, 81, 87, 88, 94, 96, 99, 105, 109, 110, 115, 118, 119, 170]. The highest use of mobile apps associated with or without a wearable was in Europe and the United Kingdom, followed by North America. Tele/video/web conferencing was used across regions, with the highest number in North America (Figure 2).

Figure 2. Number of studies published by region based on different technologies (n=105).



Validation Studies

There were 28 validation studies. Studies that validated stand-alone technologies included those to assess ROM (n=16) [24, 29, 43, 45, 48, 49, 52, 68, 73, 84, 100, 101, 111, 112, 116, 125] or gait or posture (n=10) [29, 30, 33, 53, 57, 74, 89, 97, 102, 124], and 2 studies involved technologies to monitor exercises [98, 114]. The technologies involved were either wearables (n=20) [24, 29, 30, 33, 45, 48, 49, 57, 68, 73, 84, 89, 97, 98, 100, 111, 112, 114, 116, 125], sensor-based devices (nonwearables; n=4) [53, 66, 102, 124], or inbuilt sensors available within a smartphone (n=4) [43, 52, 74, 101] (Table S4 in Multimedia Appendix 1).

In terms of study design, 9 were cross-sectional studies [33, 48, 52, 57, 84, 89, 97, 101, 116], 7 were cohort or longitudinal studies [45, 53, 68, 74, 100, 111, 125], 5 were pre-post studies [29, 30, 43, 73, 102], 1 was an uncontrolled trial [112], 1 was a randomized controlled trial (RCT) [66], and 5 were articles that described the proof of concept or development plan for the technologies [24, 49, 98, 114, 124]. The participant sample size ranged from 1 to 60. Most studies reported reliability between a standard or universal goniometer and smartphone app goniometry and the clinical evaluation of sensors to measure gait parameters (Table S4 in Multimedia Appendix 1). In 7 studies, gait was measured using sensors provided by a health care provider in a hospital setting [29, 33, 57, 74, 89, 97, 102], and 3 studies did not describe the measurement setting [30, 53, 124].

Clinical Evaluation Studies

There were 64 clinical evaluation studies. The technology consisted of a mobile or computer app with a wearable device (n=18) [6, 26, 31, 32, 39, 44, 46, 50, 54, 64, 67, 69, 90, 92, 95, 106, 108, 169], a mobile or computer app with a sensor-based device (n=13) [25, 34, 35, 40, 42, 65, 70, 76, 79, 85, 86, 93, 120], only a mobile app (n=14) [36, 37, 55, 56, 62, 75, 78, 80, 83, 107, 113, 117, 123, 128], or only telephone or videoconferencing (n=19) for remote monitoring [28, 38, 51, 63, 71, 72, 77, 81, 88, 94, 96, 99, 105, 109, 110, 115, 118, 119, 170]. Of the studies that used a mobile app, 9 studies were developed only for iOS [55, 67, 69, 71, 77, 92, 106, 107, 109], 1 was an Android app [42], 7 were for both Android and iOS devices [28, 36, 56, 88, 108, 115, 117], and 21 studies did not specify the platform (Multimedia Appendix 2). A web-based clinician portal for synchronous or asynchronous remote monitoring of patients was reported by 36 studies (Table 1). The number of published studies and the intervention arm sample size (ranging from 7 to 2292), especially for those that included wearable sensors and mobile apps, steadily increased over the last 2 decades (Figure 3).

Table 1. Summary of application functionality (N=64).

First author, year	Web portal	Devices		Peer	App name
	Monitoring	Wearables	Sensor-based devices		
Alexander, 2023 [67]	✓	Apple Watch	__[a]	—	mymobility
An, 2021 [110]	—	—	—	—	—
Argent, 2019 [169]	—	IMU[b]	Avatar	—	—
Bäcker, 2021 [25]	—	GenuSport	—	—	GenuSport
Bade, 2020 [166]	—	In-shoe sensors	—	—	—
Bell, 2020 [90]	✓	InterACTION IMU	—	—	—
Bini, 2017 [71]	✓	—	—	—	Capture proof
Blasco, 2022 [28]	—	—	—	—	WeChat app
Campbell, 2019 [72]	✓	—	—	—	StreaMD
Chughtai, 2018 [76]	✓	—	VERA[c]	—	VERA
Chughtai, 2019 [75]	—	—	—	✓	PReHab
Colomina, 2021 [31]	✓	Fitbit Flex 2	—	—	—
Correia, 2019 [32]	✓	IMU	—	—	—
De Berardinis, 2022 [26]	✓	Magnetic sensors with Velcro bands	—	—	kari
Doiron-Cadrin, 2020 [77]	—	—	—	—	Reacts Lite
Duong, 2023 [106]	✓	Fitbit, ActivPal, Goniometer Pro	—	—	—
Eichler, 2019 [34]	✓	Kinect sensor	—	—	MainReha app
Eisermann, 2004 [39]	✓	Accelerometers, wrist band, chest sensors	Web cams	—	—
Farr-Wharton, 2020 [108]	✓	Garmin Vivosmart heart rate activity tracker	—	—	—
Fung, 2012 [79]	—	—	Wii sensor balance	—	—
Gianola, 2020 [35]	—	—	Avatar	—	—
Gohir, 2021 [36]	✓	—	—	—	i-Beat app
Gray, 2022 [37]	✓	—	—	—	Digital Joint School using GoWell health program
Gunduz, 2021 [38]	—	—	—	—	—
Hadamus, 2022 [40]	—	—	Kinetic camera	—	—
Hardwick-Morris, 2022 [107]	✓	—	—	—	Physitrack
Hong, 2022 [80]	—	—	—	—	Digital Musculoskeletal Surgical Care Program app
Huang, 2017 [113]	—	—	—	—	Yishu
Janhunen, 2023 [42]	—	—	Kinect sensor with TV and tablet	—	—
Juhl, 2016 [44]	✓	IMU	—	—	ICURA app
Klement, 2019 [81]	✓	—	—	—	—
Knapp, 2021 [83]	✓	—	—	—	—
Kramer, 2003 [99]	—	—	—	—	—

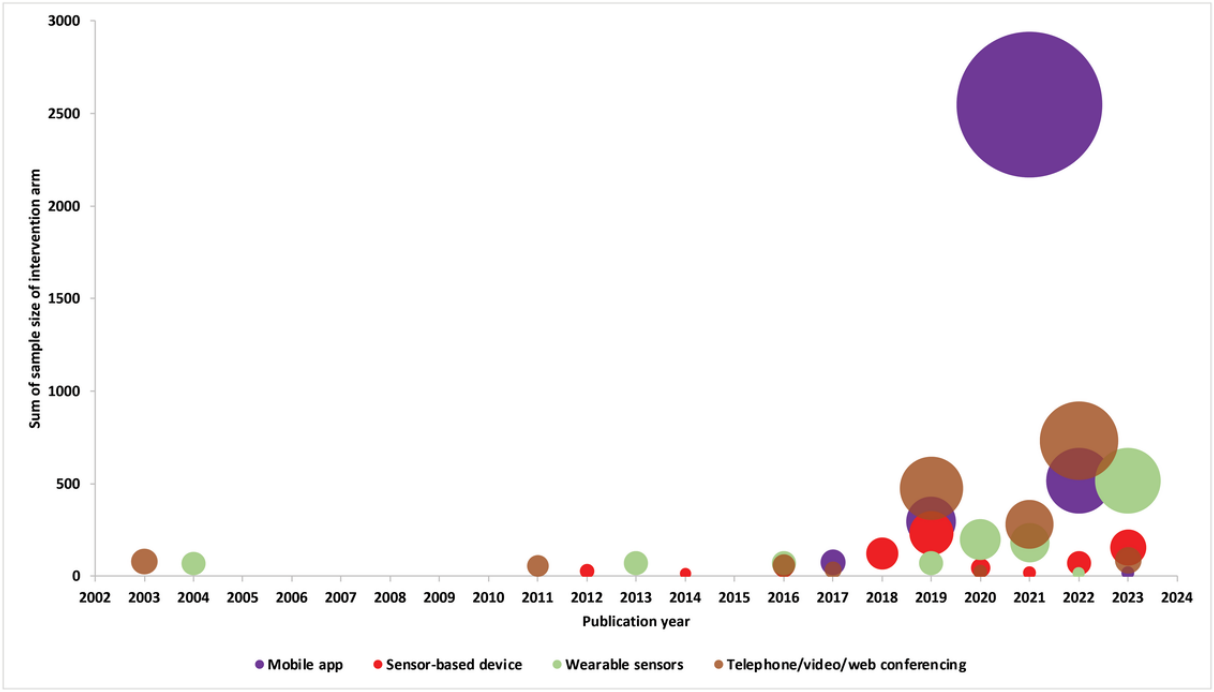
First author, year	Web portal	Devices		Peer	App name
	Monitoring	Wearables	Sensor-based devices		
Kuether, 2019 [85]	✓	—	VERA	—	—
Lam, 2016 [86]	✓	IMU	—	—	ReHab system
Lebleu, 2023 [46]	✓	Activity tracker Garmin vívofit 4	—	—	moveUP Therapy
LeBrun, 2022 [78]	✓	—	—	—	MyChart app
Li, 2023 [115]	—	—	—	✓	—
Lu, 2021 [117]	—	—	—	✓	—
McDonall, 2022 [147]	✓	—	—	—	—
Mehta, 2020 [6]	✓	Activity tracker	—	✓	—
Milliren, 2022 [88]	—	—	—	—	Ubicare Smart X
Nuevo, 2023 [50]	✓	Accelerometer, gyroscope, magnetometer (DyCare)	—	—	ReHub
Osterloh, 2023 [51]	✓	—	—	✓	YOLii
Park, 2017 [118]	—	—	—	—	—
Park, 2023 [119]	—	—	—	—	—
Piqueras, 2013 [54]	✓	(WAGYRO)	Avatar	—	—
Pournajaf, 2022 [65]	—	—	—	—	—
Pronk, 2020 [55]	—	—	—	—	Pain coach app
Prvu Bettger, 2019 [70]	✓	—	VERA	—	—
Ramkumar, 2019 [92]	✓	Motion sensors	—	—	Focus ventures RPM
Russell, 2011 [105]	—	—	—	—	—
Scheper, 2019 [56]	✓	—	—	—	Woundcare app
Su, 2015 [120]	—	Kinect sensor	—	—	—
Summers, 2023 [93]	✓	—	Electro-mechanical device	—	—
Szöts, 2016 [170]	—	—	—	—	—
Timmers, 2019 [62]	—	—	—	—	The Patient Journey app
Torpil, 2022 [63]	—	—	—	—	—
Tousignant, 2011 [94]	✓	—	—	—	—
Tripuraneni, 2021 [95]	✓	Smart watch	—	—	—
van Dijk-Huisman, 2020 [64]	✓	MOX activity monitor	—	—	—
Visperas, 2021 [96]	✓	—	—	—	—
Wang, 2023 [121]	—	—	—	✓	WeChat app
Zhang, 2021 [123]	—	—	—	✓	WeChat app

[a]Not applicable.

[b]IMU: inertial motion unit.

[c]VERA: Virtual Exercise Rehabilitation Assistant.

Figure 3. Technologies developed over the years by sample size (n=64), with the size of the bubble indicating the sample size of the intervention arm of all the studies published that particular year per technology category. Mobile app = mobile app + mobile app with inbuilt sensors; sensor-based device = sensor-based device with a mobile app + sensor-based device with a computer application; wearable sensors = wearable sensors with a mobile app + wearable sensors with a computer application + wearable sensors.



Although most studies described the features and functionality of the technology to deliver the intervention, they lacked details about the technological aspects that could benefit future researchers. For example, 2 studies [31, 86] explicitly reported information on the software, programming language and tools used, or calibration procedures either along with the main study or cited the article that described the development phase. Information on conceptualization of the technology-based intervention was described in only 3 studies [37, 38, 120]. End users' involvement was typically late during the development phase (ie, prototype stage) and involved refining the functionalities and features of technology [32, 51, 64, 65, 75-77, 80, 86, 88, 115, 118, 169] prior to deployment. Patient feedback on their needs at an early development phase was reported only by Blasco et al [28].

Clinical effectiveness was tested using an RCT design in 57.8% (37/64) of the studies [6, 25, 28, 34-36, 39, 42, 44, 50, 51, 54, 55, 62, 63, 65, 67, 70-72, 77, 79, 90, 94-96, 99, 105, 106, 108-110, 117-119, 128, 170], and the rest of the studies were either retrospective comparative cohort studies (n=3) [26, 37, 78], uncontrolled cohort studies (n=9) [46, 56, 69, 76, 81, 83, 85, 92, 169], cross-sectional studies (n=1) [86], or non-RCTs (n=14) [31, 32, 38, 40, 64, 75, 80, 88, 93, 107, 113, 115, 120, 123].

We found 13 study protocols, of which 12 were RCTs published between 2013 and 2023 [11, 41, 47, 58-60, 82, 87, 91, 103, 104, 122], for which we could not find a published report and hence were not included in this summary. User experience was measured in trials using quantitative (n=9) [32, 34, 38, 50, 55, 94, 96, 105, 123], qualitative (n=2) [61, 128], and mixed methods (n=3) [39, 90, 109] approaches.

Application Functionality for Rehabilitation Programs

The key functionalities of the telerehabilitation technologies extracted from 64 studies are summarized under 4 themes, namely education and enablement, monitoring progress, communication, and goal setting (Table 2).

Table 2. Themes of the key functionalities of the telerehabilitation technologies.

First author, year	Exercise						Monitoring progress						Functions		Communication	
	Repository	Diary	Tracker or reminder	Biofeedback	VR[a]	Feedback to patient	Pain	ROM[b]	Knee function	Physical activity	Sedentary time	Sleep	Triggers	Goal setting	Direction	Mode
Alexander, 2023 [67]	✓	—[c]	✓	—	—	SP[d]	—	—	✓	✓	—	—	—	—	2-way	Text, F2F[e]
An, 2021 [110]	✓	—	—	—	—	SP	—	—	—	—	—	—	—	—	2-way	Video
Argent, 2019 [169]	✓	—	✓	✓	✓	SA[f]	✓	✓	—	—	—	—	—	Exercise	2-way	F2F
Bäcker, 2021 [25]	✓	—	✓	✓	—	SA	—	—	—	—	—	—	—	Exercise	—	—
Bade, 2020 [166]	✓	—	✓	✓	—	SA	—	—	—	—	—	—	—	—	2-way	F2F
Bell, 2020 [90]	✓	—	✓	✓	—	SA, AP[g]	—	✓	—	—	—	—	—	—	2-way	Video
Bini, 2017 [71]	✓	—	—	—	—	AP	—	—	—	—	—	—	—	—	2-way	Text, video, F2F
Blasco, 2022 [28]	—	—	—	—	—	—	—	—	—	—	—	—	SC[h]	—	1-way, 2-way	Text, audio, F2F
Campbell, 2019 [72]	✓	—	✓	—	—	AP	—	—	—	—	—	—	—	—	1-way (SMS text messaging bot)	Video, text
Chughtai, 2018 [76]	✓	—	✓	✓	✓	SA, SP	✓	✓	✓	✓	—	—	—	—	2-way	Video
Chughtai, 2019 [75]	✓	✓	—	—	—	—	✓	—	—	—	—	—	—	—	—	—
Colomina, 2021 [31]	—	—	✓	—	—	SA, AP	✓	—	—	✓	✓	✓	SC	Exercise	2-way	Text
Correia, 2019 [32]	✓	—	✓	✓	—	SA, AP	—	✓	—	—	—	—	—	—	2-way	Audio, F2F
De Berardinis, 2022 [26]	✓	—	✓	✓	—	SA	—	—	—	—	—	—	SC	Exercise	2-way	F2F
Doiron-Cadrin, 2019 [77]	✓	—	✓	—	—	SP	—	—	—	—	—	—	—	—	2-way	Video
Duong, 2023 [106]	✓	—	✓	—	—	AA[i], AP	✓	✓	✓	✓	✓	✓	SC	Activity	1-way, 2-way	Text, video

First author, year	Exercise						Monitoring progress						Functions		Communication	
	Repository	Diary	Tracker or reminder	Biofeedback	VR[a]	Feedback to patient	Pain	ROM[b]	Knee function	Physical activity	Sedentary time	Sleep	Triggers	Goal setting	Direction	Mode
Eichler, 2019 [34]	✓	—	✓	✓	✓	SA, AP	—	✓	—	—	—	—	—	Exercise	1-way, 2-way	Audio, video, text, F2F
Eisermann, 2004 [39]	✓	—	—	✓	—	SA, AP	✓	—	—	✓	—	—	—	—	2-way	Text
Farr-Wharton, 2020 [108]	✓	—	✓	—	—	AA, AP	✓	✓	—	✓	—	✓	DS[j]	Function	1-way	Text, audio
Fung, 2012 [79]	✓	—	✓	✓	✓	SA	—	—	—	—	—	—	—	Lower extremity function	2-way	F2F
Gianola, 2020 [35]	✓	—	✓	✓	✓	SA	—	—	—	—	—	—	—	Exercise	—	—
Gohir, 2021 [36]	✓	—	✓	—	—	AA, AP	—	—	—	—	—	—	—	Exercise	1-way, 2-way	Text, audio (tele)
Gray, 2022 [37]	✓	—	—	—	—	SP	—	—	—	—	—	—	—	—	1-way, 2-way	Text
Gunduz, 2021 [38]	✓	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Hadamus, 2022 [40]	✓	—	—	✓	✓	SA, SP	—	—	—	—	—	—	—	Exercise	2-way	F2F
Hardwick-Morris, 2022 [107]	✓	✓	—	—	—	SP	✓	—	—	—	—	—	SC	—	2-way	Video, text
Hong, 2022 [80]	✓	—	—	—	—	SP	—	—	—	—	—	—	—	Recovery goals	2-way	Video
Huang, 2017 [113]	✓	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Janhunen, 2023 [42]	✓	✓	—	✓	✓	SA	—	—	—	—	—	—	—	Exercise	—	—
Juhl, 2016 [44]	✓	—	—	✓	—	SP	—	—	—	—	—	—	SC	—	2-way	Unclear
Klement, 2019 [81]	✓	—	✓	—	—	—	—	—	—	—	—	—	—	—	1-way, 2-way	Text, videos, F2F
Knapp, 2021 [83]	✓	—	✓	—	—	—	—	—	—	—	—	—	NU[k]	—	—	—

First author, year	Exercise						Monitoring progress						Functions		Communication	
	Repository	Diary	Tracker or reminder	Biofeedback	VR[a]	Feedback to patient	Pain	ROM[b]	Knee function	Physical activity	Sedentary time	Sleep	Triggers	Goal setting	Direction	Mode
Kramer, 2003 [99]	—	—	✓	—	—	—	—	—	—	—	—	—	SC	—	2-way	Audio
Kuether, 2019 [85]	✓	—	✓	✓	✓	SA, SP	—	—	✓	✓	—	—	—	—	2-way	F2F, video
Lam, 2016 [86]	✓	—	✓	✓	✓	SA, SP	—	✓	—	—	—	—	—	ROM, strength	—	—
Lebleu, 2023 [46]	✓	✓	—	✓	—	SA, AP	✓	✓	✓	✓	—	—	DS	—	2-way	Text
LeBrun, 2022 [78]	✓	—	—	—	—	SP	—	—	—	—	—	—	SC	—	2-way	Audio, video
Li, 2023 [115]	✓	—	✓	—	—	SP	—	—	—	—	—	—	—	—	2-way	Video, text
Lu, 2021 [117]	✓	—	—	—	—	SP	—	—	—	—	—	—	SC	—	2-way	Video
McDonall, 2022 [147]	✓	—	—	—	—	—	—	—	—	—	—	—	—	Pain management, knee function, avoiding complications	—	—
Mehta, 2020 [6]	—	—	✓	—	—	AA	✓	—	—	✓	—	—	DS, NU	Activity	1-way, 2-way	Text, F2F
Milliren, 2022 [88]	—	—	—	—	—	—	—	—	—	—	—	—	—	Discharge goal	1-way	Text (automatic)
Nuevo, 2023 [50]	✓	—	✓	✓	—	SA	✓	✓	—	—	—	—	DS, NU	—	2-way	Video, text
Osterloh, 2023 [51]	✓	—	—	—	—	SP	—	—	—	—	—	—	SC	—	2-way	Video
Park, 2017 [118]	—	—	✓	—	—	—	—	—	—	—	—	—	SC	—	1-way, 2-way	Text, audio (tele),
Park, 2023 [119]	—	—	—	—	—	SP	—	—	—	—	—	—	SC	—	2-way	Audio calls
Piqueras, 2013 [54]	✓	—	✓	✓	✓	SA, AP	—	✓	—	—	—	—	—	—	2-way	Audio (tele)

First author, year	Exercise						Monitoring progress						Functions		Communication	
	Repository	Diary	Tracker or reminder	Biofeedback	VR[a]	Feedback to patient	Pain	ROM[b]	Knee function	Physical activity	Sedentary time	Sleep	Triggers	Goal setting	Direction	Mode
Pournajaf, 2022 [65]	✓	—	—	✓	✓	SA	—	—	—	—	—	—	VR-based balance board	Exercise	✓	Exercise
Pronk, 2020 [55]	✓	—	—	—	—	Unclear	✓	—	—	—	—	—	—	—	—	—
Prvu Bettger, 2019 [70]	✓	—	—	✓	✓	SA, SP, AP	—	—	✓	✓	✓	—	—	Exercise	2-way	Video, F2F
Ramkumar, 2019 [92]	✓	—	✓	✓	✓	SA	✓	✓	—	✓	—	—	DS	Exercise	1-way	Text
Russell, 2011 [105]	✓	✓	—	—	—	SP	—	✓	—	✓	—	—	SC	Unclear	2-way	Video
Scheper, 2019 [56]	—	—	—	—	—	—	✓	—	—	—	—	—	DS	—	1-way	Text
Su, 2015 [120]	✓	—	—	✓	✓	SA	—	—	—	—	—	—	—	Exercise	—	—
Summers, 2023 [93]	✓	—	✓	✓	—	SA, SP	✓	✓	✓	✓	—	—	DS	—	2-way	Video
Szöts, 2016 [170]	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2-way	Audio (tele)
Timmers, 2019 [62]	✓	—	—	—	—	—	✓	—	—	—	—	—	—	—	1-way	Audio, video, text
Torpil, 2022 [63]	—	—	—	—	—	—	—	—	—	—	—	—	SC	Occupation related	2-way	Video
Tousignant, 2011 [94]	—	—	—	—	—	SP	—	—	—	—	—	—	SC	—	2-way	Video
Tripuraneni, 2021 [95]	✓	—	✓	—	—	AA	—	—	—	✓	✓	—	—	—	1-way	Text
van Dijk-Huisman, 2020 [64]	✓	—	✓	✓	—	SA, AP	—	—	—	✓	—	—	SC	—	2-way	Video
Visperas, 2021 [96]	✓	—	—	—	—	AP	✓	—	✓	—	—	—	DS, SC	—	1-way, 2-way	Text, audio (telephone)
Wang, 2023 [121]	✓	✓	✓	—	—	AP	—	—	—	—	—	—	—	Task	2-way	Text

First author, year	Exercise						Monitoring progress						Functions		Communication	
	Repository	Diary	Tracker or reminder	Biofeedback	VR[a]	Feedback to patient	Pain	ROM[b]	Knee function	Physical activity	Sedentary time	Sleep	Triggers	Goal setting	Direction	Mode
Zhang, 2021 [123]	✓	—	—	—	—	—	—	—	—	—	—	—	—	—	2-way	Audio, text, video

[a]VR: virtual reality.

[b]ROM: range of motion.

[c]Not applicable.

[d]SP: synchronous from physiotherapist.

[e]F2F: face to face.

[f]SA: synchronous from app.

[g]AP: asynchronous from physiotherapist.

[h]SC: scheduled call.

[i]AA: asynchronous from app.

[j]DS: danger signs.

[k]NU: non-use.

Education and Enablement

An exercise repository in the form of videos, text, or infographics was one of the main features in the studies (n=53), of which only 20 studies described the list of exercises (Table S5 in Multimedia Appendix 1). Education for patients was part of the rehabilitation program in 17 studies. Table S6 in Multimedia Appendix 1 lists the topic areas covered in the education materials. Regarding exercise, 6 studies reported using an e-diary for maintaining an exercise log, 11 studies reported using reminders to perform exercises, and 13 studies reported using a tracker for exercise adherence (Multimedia Appendix 2). Feedback on the appropriateness of exercise performance was synchronous (biofeedback or virtual reality) from the app (n=19), directly from the health care provider via a video call with the patients (patient performing exercise live, measurement of ROM during video call, transmission of virtual avatar data to health care provider; n=14), or provided via both (n=6; Table 2). Feedback to the patient, which was either in the form of push notifications or a progress summary, was asynchronous from the app using automated programs in 2 studies. Asynchronous feedback from a health care provider in the form of instructions, messages, or an exercise regimen was reported in 13 studies. Feedback via both the app and a health care provider was provided in 3 studies (Table 2). Only 7 studies [6, 51, 75, 115, 117, 123, 128] had an option for peer support for patients.

Measuring Progress

Measurement of patient-reported outcomes such as pain (n=19) was an inbuilt feature in the app. Changes in knee function and activity were monitored directly via wearables or captured using patient-reported outcome measures. These included ROM in 15 studies, knee function in 8 studies, physical activity in 20 studies, sedentary behavior in 5 studies, and sleep in 4 studies. Automatic alerts were provided to the health care provider for any danger signs such as knee pain, wound health, opioid consumption, function, ROM, number of steps, exercise adherence, and any negative response to questions after entering the postoperative follow-up in 9 studies; for non-use of the technology by patients in 4 studies; and for scheduled consultations in 18 studies (Table 2).

Communication

Mobile app–enabled 1-way communication included push messages, notifications, reminders, patients' replies to inbuilt questions in the app, information sent to the patient by the health care team, and an SMS text messaging bot (n=10). Two-way communication, either via an app or in face-to-face visits, was reported in 41 studies. In addition, 11 studies reported a combination of both 1 and 2-way communication, and 1 study did not provide sufficient information about communication. Electronic communication was delivered in the form of text, audio or video messages, and direct communication (Table 2).

Goal Setting

Goal setting for exercises, activity, pain management, knee function, ROM, muscle strength, rehabilitation, and discharge as part of the rehabilitation program was reported in 23 studies. The goals were set by either the health care provider or the patient (Table 2).

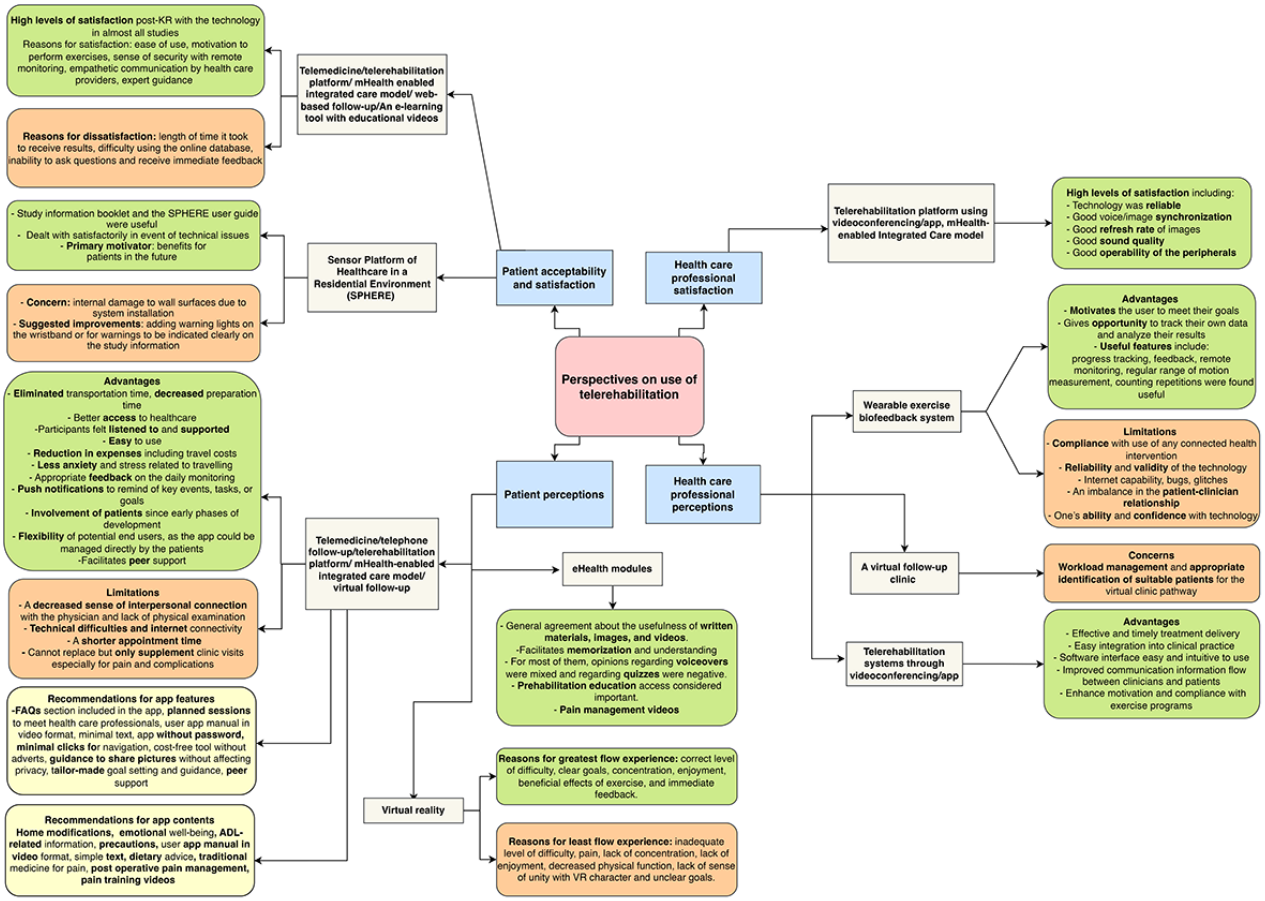
End Users' Perceptions

Of the 38 studies that reported user perspectives, 2 focused on the perspectives of health care providers, 27 focused on the perspectives of patients and caregivers, and 9 focused on the perspectives of both groups (health care providers and patients and caregivers). The approach for data collection was quantitative (n=23), qualitative (n=9), or mixed methods (n=6). The sample size ranged from 2 to 200 health care providers and from 5 to 2292 patients (Tables S7 and S8 in Multimedia Appendix 1).

Commonly used quantitative questionnaires to assess satisfaction were the System Usability Scale [129] and the net promoter score [130]. To ratify the experience with telerehabilitation, the Telemedicine Perception Questionnaire was used [131]. Acceptability and usability were assessed using the acceptance of information technology questionnaire [132] and the Telemedicine Usability Questionnaire [133]. Some studies used bespoke questionnaires to report user experience and satisfaction [32, 39, 61, 90, 94, 105, 109, 134-146].

Overall, health care providers perceived telerehabilitation and the use of technology such as biosensors as a way of improving efficiency in providing care [146], patient adherence to exercises [39, 136, 146], patient-physician communication [136], and case management [137, 146]. The main factors associated with user satisfaction with e-consultations were reliable technology, good voice or image synchronization, the refresh rate of images, sound quality, and operability of the peripherals [94, 96, 138, 139]. The key factors they perceived would influence use and uptake of technology were decreased workload (rather than increased) [140], reliability of measurements aided by technology [146], ability to measure functional outcomes objectively [141], clearer criteria when choosing appropriate patients to be enrolled in the program [140], self-efficacy in the use of technology [94, 138, 146], and ease of reporting and tracking of patient data [90]. Patients and health care providers felt e-learning modules, push notifications, and appropriate feedback from sensors and virtual reality improved self-management [138, 142-144] (Figure 4).

Figure 4. Perceptions of patients and health care providers about the technology used. ADL: activities of daily living; FAQs: frequently asked questions; KR: knee replacement; mHealth: mobile health; VR: virtual reality.



Patient satisfaction levels were reported when teleconsultation was provided via a computer, smartphone, or tablet [34, 39, 55, 56, 80, 92, 105, 121, 123, 134, 135, 145, 147-149]; telephone [61], videoconferencing [38, 77, 94, 105, 139, 141, 150-152], a web-based system [32, 50, 90, 96, 140, 153], and an mHealth-enabled integrated care model [46, 88, 138]. Patients were satisfied with telemonitoring due to improved access to services, continued support after discharge from hospital, ability for self-management, reduced need for clinic visits, reduction in cost and travel time, ability of health care providers to provide personalized care [32, 61, 94, 121, 136, 138, 140, 141, 145, 153-155], ease of use [34, 50, 55, 56, 92, 105, 135, 138, 147, 148], motivation to perform exercises [134, 135], sense of security with remote monitoring [134, 155], and empathetic communication by a health care provider [121, 135, 136, 145, 152, 155]. The reasons for dissatisfaction were lack of an in-person examination, shorter appointment times, delay in receiving reports (eg, x-ray), and an inability to transfer pictures from one technology to another [140, 145, 149, 153]. Patients provided suggestions for the app functionalities to improve the ease of use such as minimal clicks, an instructional video for app navigation, and restriction of commercial advertisements [149]. Home modifications [149], emotional well-being, information related to activities of daily living in simple text,

dietary advice, frequently asked questions, and use of traditional medicine for postoperative pain management were a few of the suggestions for app content [121].

Patients were generally satisfied with the telerehabilitation program and were ready to recommend it to others [39, 80, 85, 96, 121, 135, 151]. The use of technology for rehabilitation was influenced by computer literacy [141, 150]. However, interruption of virtual physiotherapy sessions due to poor internet issues [139] was not commonly reported (Figure 4).

Discussion

Principal Findings

This scoping review summarized the extent, user perceptions, range, and nature of technologies used to support rehabilitation following knee arthroplasty. All studies reported in this review were from upper and middle-to-upper–income countries, with a steep increase in studies in the last decade. The technologies focused on enabling patients to remember prescribed exercises as well as be able to perform them appropriately by providing synchronous and asynchronous feedback via biosensors or virtual reality. Motivation and support during recovery via technology-enabled 1-way or 2-way communication gave patients access to health care providers. Self-management and monitoring of progress were dependent on active input using e-diaries by patients or passive input through wearables. In the context where these technologies were evaluated, end users were satisfied and found remote monitoring to be acceptable for routine use.

The last decade has seen an exponential increase in the number of arthroplasties worldwide [156]; however, a corresponding increase in technological solutions to facilitate remote monitoring is nonexistent in resource-limited settings such as LMICs where the need for monitoring and a continuum of care may be higher due to lower literacy levels and lack of access to rehabilitation clinics. Research on this topic that can inform clinical practice is nonexistent in the LMIC context. Despite a high penetration of the smartphone market [157] in LMICs, a higher initial investment to develop the technology, especially in the health care sector [158], or a lack of publication of such efforts could be reasons. In LMICs, there is an increasing trend of lower limb joint replacement procedures [156]. High out-of-pocket expenditures incurred due to home visits by physiotherapists or clinic visits by patients [159] dictate the need for a cost-effective and feasible technology-based strategy to fit the context while using lessons learned from available research.

There is unequivocal evidence that there is a need for physical and psychological support from professionals during the recovery period for pain management, adherence to exercises, and modifications to therapy planning based on one's progress [3, 160, 161]. The apps were either focused on a single function (such as communication or knowledge transfer) or were multifunctional. They were generally received well by end users; however, the usability and acceptability of these applications or remote monitoring modalities cannot be extrapolated to low health literacy and tech literacy settings. The challenges we expect with using remote monitoring in the LMIC context could be inequitable smartphone access or tech literacy, internet speed, affordability of wearables, the burden to the health system if these needs are provided free of cost, and the need for educational content in multiple languages in countries with a non-native English-speaking, multilingual population such as in India [162].

Implications for Future Research

mHealth interventions have the potential to expand the reach and effectiveness of health support by facilitating behavior change. However, to ensure these "digital behavior health interventions" effectively engage users and are effective, both microengagement (the mHealth interface) and macroengagement (evidence-based behavior change techniques) are essential [163, 164]. However, we found only a handful of studies that reported user involvement during the development stage [28, 32, 51, 58, 64, 65, 75-77, 80, 86, 88, 115, 118, 169]. Studies rarely provided an adequate explanation of the theoretical behavioral framework behind the technology-based interventions [165].

Since the context and technologies are so varied, any new applications that are developed, especially in the LMIC context, should undertake formative research with end users to understand their needs, understand their preferences, and study the local digital regulatory requirements before investing time and effort. Feasibility and pilot testing by a multidisciplinary team should be crucial steps before a full-scale evaluation [69, 166], and embedding end users' involvement and documenting their experiences at every stage are vital to refining future interventions [164]. Further, the rehabilitation protocols should map the application features with the desired function [167, 168], and this should be confirmed by means of a process evaluation embedded within the clinical evaluation to inform the mechanism of the impact in a real-life setting [147].

Limitations

This review needs to be interpreted in light of the following limitations. This scoping review focused only on technology interventions for post-knee replacement rehabilitation and hence cannot be extrapolated to other orthopedic procedures. We did not include articles for which the full text was not available. Further, incomplete reporting on the features and functions of the technology is possible and may have affected our qualitative summary and conclusion.

We did not perform a consultation phase as per the guidelines [20], and the research question was formulated upon discussion between the researchers of the scoping review team, physiotherapists, and clinicians. We limited our search from 2001 onward; however, since knee arthroplasty and mHealth came into practice in the last 2 decades, this restriction in the search may not have an implication for our review findings.

Conclusion

Several technologies have been identified to promote adherence, increase self-efficacy, enhance self-management, and support remote monitoring. However, all the available technologies have been developed and used in developed countries. The need for remote monitoring is compelling in resource-limited

countries where knee arthroplasty is on the rise. However, irrespective of the context, it is important to involve a multidisciplinary team and include users' perspectives during the development stage.

What Was Already Known About the Topic

Computer and mobile technologies to support rehabilitation following knee arthroplasty are in wide use. Telerehabilitation and remote monitoring are as effective and safe as clinic-based rehabilitation programs. They reduce out-of-pocket expenditure or health cost expenditure by reducing the time to discharge following surgery and the number of clinic visits after discharge.

What This Study Adds

This study provides a map of the types of technology and the functionality of mobile and computer-based multifunction applications. We summarized end users' perceptions and reasons for satisfaction or dissatisfaction with available technology. The findings reflect the lack of research and readily available technologies for LMICs.

Acknowledgments

The authors would like to thank distinguished Professor Gordon Guyatt, in the Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada, for providing expert comments. We also thank Dr. Shyamashree Biswas, Research Intern at the George institute, for assisting with the quality check of the extracted data.

This scoping review is a part of a fellowship funded by DBT/Wellcome Trust India Alliance (grant number: IA/CPHI/20/1/505224).

Authors' Contributions

ND conceptualized the protocol and conducted the search. ND, SP, PS, and ShP screened and extracted the data. The first draft was written by SP. RM and AK interpreted the study findings and gave significant feedback to the early drafts. RMad provided expert comments and suggestions and edited the later version of manuscript. ND takes responsibility for the data. All authors read and agreed to the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary information on search strategy, included and excluded studies.
[DOCX File , 546 KB - mhealth_v12i1e47843_app1.docx]

Multimedia Appendix 2
Raw data extraction file for rehabilitation program studies.
[XLSX File (Microsoft Excel File), 39 KB - mhealth_v12i1e47843_app2.xlsx]

Multimedia Appendix 3
PRISMA-ScR checklist.
[PDF File (Adobe PDF File), 105 KB - mhealth_v12i1e47843_app3.pdf]

References

1. Zhang W, Moskowitz RW, Nuki G, Abramson S, Altman RD, Arden N, et al. OARSI recommendations for the management of hip and knee osteoarthritis, Part II: OARSI evidence-based, expert consensus guidelines. Osteoarthritis Cartilage 2008 Feb;16(2):137-162 [FREE Full text] [doi: 10.1016/j.joca.2007.12.013] [Medline: 18279766]
2. Artz N, Elvers KT, Lowe CM, Sackley C, Jepson P, Beswick AD. Effectiveness of physiotherapy exercise following total knee replacement: systematic review and meta-analysis. BMC Musculoskelet Disord 2015 Feb 07;16(1):15 [FREE Full text] [doi: 10.1186/s12891-015-0469-6] [Medline: 25886975]
3. Goldsmith LJ, Suryaprakash N, Randall E, Shum J, MacDonald V, Sawatzky R, et al. The importance of informational, clinical and personal support in patient experience with total knee replacement: a qualitative investigation. BMC Musculoskelet Disord 2017 Mar 24;18(1):127 [FREE Full text] [doi: 10.1186/s12891-017-1474-8] [Medline: 28340610]
4. Groeneveld BS, Dekkers T, Mathijssen NMC, Vehmeijer SBW, Melles M, Goossens RHM. Communication preferences in total joint arthroplasty: exploring the patient experience through generative research. Orthop Nurs 2020;39(5):292-302. [doi: 10.1097/NOR.0000000000000694] [Medline: 32956269]
5. Westby MD, Backman CL. Patient and health professional views on rehabilitation practices and outcomes following total hip and knee arthroplasty for osteoarthritis:a focus group study. BMC Health Serv Res 2010 May 11;10(1):119 [FREE Full text] [doi: 10.1186/1472-6963-10-119] [Medline: 20459834]

6. Mehta SJ, Hume E, Troxel AB, Reitz C, Norton L, Lacko H, et al. Effect of remote monitoring on discharge to home, return to activity, and rehospitalization after hip and knee arthroplasty: a randomized clinical trial. JAMA Netw Open 2020 Dec 01;3(12):e2028328 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.28328] [Medline: 33346847]
7. Schwartz FH, Lange J. Factors that affect outcome following total joint arthroplasty: a review of the recent literature. Curr Rev Musculoskelet Med 2017 Sep 29;10(3):346-355 [FREE Full text] [doi: 10.1007/s12178-017-9421-8] [Medline: 28664450]
8. L Snell D, Hipango J, Sinnott KA, Dunn JA, Rothwell A, Hsieh CJ, et al. Rehabilitation after total joint replacement: a scoping study. Disabil Rehabil 2018 Jul 23;40(14):1718-1731. [doi: 10.1080/09638288.2017.1300947] [Medline: 28330380]
9. Alrawashdeh W, Eschweiler J, Migliorini F, El Mansy Y, Tingart M, Rath B. Effectiveness of total knee arthroplasty rehabilitation programmes: A systematic review and meta-analysis. J Rehabil Med 2021 Jun 02;53(6):jrm00200 [FREE Full text] [doi: 10.2340/16501977-2827] [Medline: 33846757]
10. Buhagiar MA, Naylor JM, Harris IA, Xuan W, Adie S, Lewin A. Assessment of outcomes of inpatient or clinic-based vs home-based rehabilitation after total knee arthroplasty: a systematic review and meta-analysis. JAMA Netw Open 2019 Apr 05;2(4):e192810. [doi: 10.1001/jamanetworkopen.2019.2810] [Medline: 31026026]
11. Jansson MM, Rantala A, Miettunen J, Puhto A, Pikkarainen M. The effects and safety of telerehabilitation in patients with lower-limb joint replacement: A systematic review and narrative synthesis. J Telemed Telecare 2020 Apr 21;28(2):96-114. [doi: 10.1177/1357633x20917868]
12. Bernhardsson S, Larsson A, Bergenheim A, Ho-Henriksson C, Ekhammar A, Lange E, et al. Digital physiotherapy assessment vs conventional face-to-face physiotherapy assessment of patients with musculoskeletal disorders: A systematic review. PLoS One 2023 Mar 21;18(3):e0283013 [FREE Full text] [doi: 10.1371/journal.pone.0283013] [Medline: 36943857]
13. Peng L, Zeng Y, Wu Y, Si H, Shen B. Virtual reality-based rehabilitation in patients following total knee arthroplasty: a systematic review and meta-analysis of randomized controlled trials. Chin Med J (Engl) 2021 Dec 13;135(2):153-163 [FREE Full text] [doi: 10.1097/CM9.0000000000001847] [Medline: 34908004]
14. Constantinescu D, Pavlis W, Rizzo M, Vanden Berge D, Barnhill S, Hernandez VH. The role of commercially available smartphone apps and wearable devices in monitoring patients after total knee arthroplasty: a systematic review. EFORT Open Rev 2022 Jul 05;7(7):481-490 [FREE Full text] [doi: 10.1530/EOR-21-0115] [Medline: 35900191]
15. Özden F, Sarı Z. The effect of mobile application-based rehabilitation in patients with total knee arthroplasty: A systematic review and meta-analysis. Arch Gerontol Geriatr 2023 Oct;113:105058. [doi: 10.1016/j.archger.2023.105058] [Medline: 37172329]
16. Gianzina E, Kalinterakis G, Delis S, Vlastos I, Platon Sachinis N, Yiannakopoulos CK. Evaluation of gait recovery after total knee arthroplasty using wearable inertial sensors: A systematic review. Knee 2023 Mar;41:190-203. [doi: 10.1016/j.knee.2023.01.006] [Medline: 36724578]
17. McKeon JF, Alvarez PM, Vajapey AS, Sarac N, Spitzer AI, Vajapey SP. Expanding role of technology in rehabilitation after lower-extremity joint replacement: a systematic review. JBJS Rev 2021 Sep 13;9(9):A. [doi: 10.2106/JBJS.RVW.21.00016] [Medline: 34516463]
18. Wang X, Hunter DJ, Vesentini G, Pozzobon D, Ferreira ML. Technology-assisted rehabilitation following total knee or hip replacement for people with osteoarthritis: a systematic review and meta-analysis. BMC Musculoskelet Disord 2019 Nov 03;20(1):506 [FREE Full text] [doi: 10.1186/s12891-019-2900-x] [Medline: 31679511]
19. Seron P, Oliveros MJ, Gutierrez-Arias R, Fuentes-Aspe R, Torres-Castro RC, Merino-Osorio C, et al. Effectiveness of telerehabilitation in physical therapy: a rapid overview. Phys Ther 2021 Jun 01;101(6):A [FREE Full text] [doi: 10.1093/ptj/pzab053] [Medline: 33561280]
20. Peters MDJ, Marnie C, Tricco AC, Pollock D, Munn Z, Alexander L, et al. Updated methodological guidance for the conduct of scoping reviews. JBI Evid Synth 2020 Oct;18(10):2119-2126. [doi: 10.11124/JBIES-20-00167] [Medline: 33038124]
21. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, Tunçalp, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473. [doi: 10.7326/m18-0850]
22. Pritwani S, Devasenapathy N. Mobile technology-based applications for rehabilitation monitoring and self-management after knee arthroplasty: A scoping review. Open Science Framework. 2022 Jun 21. URL: https://osf.io/srxkc [accessed 2023-12-11]
23. Rayyan. URL: https://www.rayyan.ai/ [accessed 2023-12-11]
24. Ayoade M, Morton L, Baillie L. Investigating the feasibility of a wireless motion capture system to aid in the rehabilitation of total knee replacement patients. 2011 Presented at: 5th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops; May 23-26, 2011; Dublin, Ireland. [doi: 10.4108/icst.pervasivehealth.2011.246132]
25. Bäcker HC, Wu CH, Schulz MRG, Weber-Spickschen TS, Perka C, Hardt S. App-based rehabilitation program after total knee arthroplasty: a randomized controlled trial. Arch Orthop Trauma Surg 2021 Sep 06;141(9):1575-1582 [FREE Full text] [doi: 10.1007/s00402-021-03789-0] [Medline: 33547927]
26. De Berardinis L, Senarighi M, Ciccullo C, Forte F, Spezia M, Gigante AP. Fast-track surgery and telerehabilitation protocol in unicompartmental knee arthroplasty leads to superior outcomes when compared with the standard protocol: a

propensity-matched pilot study. Knee Surg Relat Res 2022 Dec 12;34(1):44 [FREE Full text] [doi: 10.1186/s43019-022-00173-z] [Medline: 36510279]
27. Bitsaki M, Koutras G, Heep H, Koutras C. Cost-effective mobile-based healthcare system for managing total joint arthroplasty follow-up. Healthc Inform Res 2017 Jan;23(1):67-73 [FREE Full text] [doi: 10.4258/hir.2017.23.1.67] [Medline: 28261533]
28. Blasco J, Roig-Casasús S, Igual-Camacho C, Díaz-Díaz B, Pérez-Maletzki J. Conversational chatbot to promote adherence to rehabilitation after total knee replacement: implementation and feasibility. Archives of Physical Medicine and Rehabilitation 2022 Dec;103(12):e125. [doi: 10.1016/j.apmr.2022.08.764]
29. Bonora S, Pirani R, Amadori E, Fantini C, Chiari L, Merlo A, et al. Use of a haptic biofeedback in the rehabilitation of patients with total knee arthroplasty (TKA): A pilot study. Gait & Posture 2017 Sep;57:16-17. [doi: 10.1016/j.gaitpost.2017.07.069]
30. Calliess T, Bocklage R, Karkosch R, Marschollek M, Windhagen H, Schulze M. Clinical evaluation of a mobile sensor-based gait analysis method for outcome measurement after knee arthroplasty. Sensors (Basel) 2014 Aug 28;14(9):15953-15964 [FREE Full text] [doi: 10.3390/s140915953] [Medline: 25171119]
31. Colomina J, Drudis R, Torra M, Pallisó F, Massip M, Vargiu E, CONNECARE-Lleida Group. Implementing mHealth-enabled integrated care for complex chronic patients with osteoarthritis undergoing primary hip or knee arthroplasty: prospective, two-arm, parallel trial. J Med Internet Res 2021 Sep 02;23(9):e28320 [FREE Full text] [doi: 10.2196/28320] [Medline: 34473068]
32. Correia FD, Nogueira A, Magalhães I, Guimarães J, Moreira M, Barradas I, et al. Medium-term outcomes of digital versus conventional home-based rehabilitation after total knee arthroplasty: prospective, parallel-group feasibility study. JMIR Rehabil Assist Technol 2019 Feb 28;6(1):e13111 [FREE Full text] [doi: 10.2196/13111] [Medline: 30816849]
33. De Vroey H, Staes F, Weygers I, Vereecke E, Vanrenterghem J, Deklerck J, et al. The implementation of inertial sensors for the assessment of temporal parameters of gait in the knee arthroplasty population. Clin Biomech (Bristol, Avon) 2018 May;54:22-27. [doi: 10.1016/j.clinbiomech.2018.03.002] [Medline: 29533844]
34. Eichler S, Salzwedel A, Rabe S, Mueller S, Mayer F, Wochatz M, et al. The effectiveness of telerehabilitation as a supplement to rehabilitation in patients after total knee or hip replacement: randomized controlled trial. JMIR Rehabil Assist Technol 2019 Nov 07;6(2):e14236 [FREE Full text] [doi: 10.2196/14236] [Medline: 31697239]
35. Gianola S, Stucovitz E, Castellini G, Mascali M, Vanni F, Tramacere I, et al. Effects of early virtual reality-based rehabilitation in patients with total knee arthroplasty: A randomized controlled trial. Medicine (Baltimore) 2020 Feb;99(7):e19136 [FREE Full text] [doi: 10.1097/MD.0000000000019136] [Medline: 32049833]
36. Gohir SA, Eek F, Kelly A, Abhishek A, Valdes AM. Effectiveness of internet-based exercises aimed at treating knee osteoarthritis: the iBEAT-OA randomized clinical trial. JAMA Netw Open 2021 Feb 01;4(2):e210012 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.0012] [Medline: 33620447]
37. Gray J, McCarthy S, Carr E, Danjoux G, Hackett R, McCarthy A, et al. The impact of a digital joint school educational programme on post-operative outcomes following lower limb arthroplasty: a retrospective comparative cohort study. BMC Health Serv Res 2022 Apr 29;22(1):580 [FREE Full text] [doi: 10.1186/s12913-022-07989-1] [Medline: 35488258]
38. Gündüz CS, Çalişkan N. The effect of preoperative video based pain training on postoperative pain and analgesic use in patients Undergoing Total Knee Arthroplasty: a non-randomized control group intervention study. Clin Nurs Res 2021 Jul 31;30(6):741-752. [doi: 10.1177/1054773820983361] [Medline: 33383995]
39. Eisermann U, Haase I, Kladny B. Computer-aided multimedia training in orthopedic rehabilitation. Am J Phys Med Rehabil 2004 Sep;83(9):670-680. [doi: 10.1097/01.phm.0000137307.44173.5d] [Medline: 15314531]
40. Hadamus A, Błażkiewicz M, Wydra KT, Kowalska AJ, Łukowicz M, Białoszewski D, et al. Effectiveness of early rehabilitation with exergaming in virtual reality on gait in patients after total knee replacement. J Clin Med 2022 Aug 23;11(17):4950 [FREE Full text] [doi: 10.3390/jcm11174950] [Medline: 36078879]
41. Harmelink KEM, Zeegers AVCM, Tönis TM, Hullegie W, Nijhuis-van der Sanden MWG, Staal JB. The effectiveness of the use of a digital activity coaching system in addition to a two-week home-based exercise program in patients after total knee arthroplasty: study protocol for a randomized controlled trial. BMC Musculoskelet Disord 2017 Jul 05;18(1):290 [FREE Full text] [doi: 10.1186/s12891-017-1647-5] [Medline: 28679400]
42. Janhunen M, Katajapuu N, Paloneva J, Pamilo K, Oksanen A, Keemu H, et al. Effects of a home-based, exergaming intervention on physical function and pain after total knee replacement in older adults: a randomised controlled trial. BMJ Open Sport Exerc Med 2023 Mar 03;9(1):e001416 [FREE Full text] [doi: 10.1136/bmjsem-2022-001416] [Medline: 36896366]
43. Jenny J. Measurement of the knee flexion angle with a smartphone-application is precise and accurate. J Arthroplasty 2013 May;28(5):784-787. [doi: 10.1016/j.arth.2012.11.013] [Medline: 23499407]
44. Juhl C, Roth S, Schierbeck R, Nielsen L, Nordlien A, Hansen N, et al. Effectiveness of technology assisted exercise compared to usual care in total knee arthroplasty. Osteoarthritis and Cartilage 2016 Apr;24:S473. [doi: 10.1016/j.joca.2016.01.864]
45. Kontadakis G, Chasiouras D, Proimaki D, Halkiadakis M, Fyntikaki M, Mania K. Gamified platform for rehabilitation after total knee replacement surgery employing low cost and portable inertial measurement sensor node. Multimed Tools Appl 2018 Sep 12;79(5-6):3161-3188. [doi: 10.1007/s11042-018-6572-6]

46. Lebleu J, Pauwels A, Anract P, Parratte S, Van Overschelde P, Van Onsem S. Digital rehabilitation after knee arthroplasty: a multi-center prospective longitudinal cohort study. J Pers Med 2023 May 13;13(5):824 [FREE Full text] [doi: 10.3390/jpm13050824] [Medline: 37240994]
47. Mark-Christensen T, Thorborg K, Kallemose T, Bandholm T. Physical rehabilitation versus no physical rehabilitation after total hip and knee arthroplasties: Protocol for a pragmatic, randomized, controlled, superiority trial (The DRAW1 trial). F1000Res 2021 Feb 25;10:146. [doi: 10.12688/f1000research.50814.1]
48. Marques CJ, Bauer C, Grimaldo D, Tabeling S, Weber T, Ehlert A, et al. Sensor positioning influences the accuracy of knee ROM data of an e-rehabilitation system: a preliminary study with healthy subjects. Sensors (Basel) 2020 Apr 15;20(8):2237 [FREE Full text] [doi: 10.3390/s20082237] [Medline: 32326616]
49. Msayib Y, Gaydecki P, Callaghan M, Dale N, Ismail S. An intelligent remote monitoring system for total knee arthroplasty patients. J Med Syst 2017 Jun 18;41(6):90 [FREE Full text] [doi: 10.1007/s10916-017-0735-2] [Medline: 28421308]
50. Nuevo M, Rodríguez-Rodríguez D, Jauregui R, Fabrellas N, Zabalegui A, Conti M, et al. Telerehabilitation following fast-track total knee arthroplasty is effective and safe: a randomized controlled trial with the ReHub platform. Disabil Rehabil 2023 Jul 05:1-11. [doi: 10.1080/09638288.2023.2228689] [Medline: 37403684]
51. Osterloh J, Knaack F, Bader R, Behrens M, Peschers J, Nawrath L, et al. The effect of a digital-assisted group rehabilitation on clinical and functional outcomes after total hip and knee arthroplasty-a prospective randomized controlled pilot study. BMC Musculoskelet Disord 2023 Mar 14;24(1):190 [FREE Full text] [doi: 10.1186/s12891-023-06270-8] [Medline: 36915066]
52. Pereira L, Rwakabayiza S, Lécureux E, Jolles B. Reliability of the knee smartphone-application goniometer in the acute orthopedic setting. J Knee Surg 2017 Mar 24;30(3):223-230. [doi: 10.1055/s-0036-1584184] [Medline: 27218479]
53. Pfeufer D, Monteiro P, Gililland J, Anderson MB, Böcker W, Stagg M, et al. Immediate postoperative improvement in gait parameters following primary total knee arthroplasty can be measured with an insole sensor device. J Knee Surg 2022 May 25;35(6):692-697. [doi: 10.1055/s-0040-1716852] [Medline: 33241543]
54. Piqueras M, Marco E, Coll M, Escalada F, Ballester A, Cinca C, et al. Effectiveness of an interactive virtual telerehabilitation system in patients after total knee arthoplasty: a randomized controlled trial. J Rehabil Med 2013 Apr;45(4):392-396 [FREE Full text] [doi: 10.2340/16501977-1119] [Medline: 23474735]
55. Pronk Y, Peters MCWM, Sheombar A, Brinkman J. Effectiveness of a mobile eHealth app in guiding patients in pain control and opiate use after total knee replacement: randomized controlled trial. JMIR Mhealth Uhealth 2020 Mar 13;8(3):e16415 [FREE Full text] [doi: 10.2196/16415] [Medline: 32167483]
56. Scheper H, Derogee R, Mahdad R, van der Wal R, Nelissen R, Visser L, et al. A mobile app for postoperative wound care after arthroplasty: Ease of use and perceived usefulness. Int J Med Inform 2019 Sep;129:75-80. [doi: 10.1016/j.ijmedinf.2019.05.010] [Medline: 31445292]
57. Boekesteijn RJ, Smolders JMH, Busch VJJF, Geurts ACH, Smulders K. Independent and sensitive gait parameters for objective evaluation in knee and hip osteoarthritis using wearable sensors. BMC Musculoskelet Disord 2021 Mar 03;22(1):242 [FREE Full text] [doi: 10.1186/s12891-021-04074-2] [Medline: 33658006]
58. Stauber A, Schüßler N, Palmdorf S, Schürholz N, Bruns D, Osterbrink J, et al. RECOVER-E - a mobile app for patients undergoing total knee or hip replacement: study protocol. BMC Musculoskelet Disord 2020 Feb 04;21(1):71 [FREE Full text] [doi: 10.1186/s12891-020-3090-2] [Medline: 32019529]
59. Straat AC, Maarleveld JM, Smit DJM, Visch L, Hulsegge G, Huirne JAF, et al. (Cost-)effectiveness of a personalized multidisciplinary eHealth intervention for knee arthroplasty patients to enhance return to activities of daily life, work and sports - rationale and protocol of the multicentre ACTIVE randomized controlled trial. BMC Musculoskelet Disord 2023 Mar 04;24(1):162. [doi: 10.1186/s12891-023-06236-w] [Medline: 36869330]
60. Strahl A, Graichen H, Haas H, Hube R, Perka C, Rolvien T, et al. Evaluation of the patient-accompanying app "alley ortho companion" for patients with osteoarthritis of the knee and hip: study protocol for a randomized controlled multi-center trial. Trials 2022 Aug 29;23(1):716 [FREE Full text] [doi: 10.1186/s13063-022-06662-6] [Medline: 36038894]
61. Szöts K, Konradsen H, Solgaard S, Bogø S, Østergaard B. Nurse-led telephone follow-up after total knee arthroplasty--content and the patients' views. J Clin Nurs 2015 Oct 14;24(19-20):2890-2899. [doi: 10.1111/jocn.12905] [Medline: 26178752]
62. Timmers T, Janssen L, van der Weegen W, Das D, Marijnissen W, Hannink G, et al. The effect of an app for day-to-day postoperative care education on patients with total knee replacement: randomized controlled trial. JMIR Mhealth Uhealth 2019 Oct 21;7(10):e15323 [FREE Full text] [doi: 10.2196/15323] [Medline: 31638594]
63. Torpil B, Kaya. The effectiveness of client-centered intervention with telerehabilitation method after total knee arthroplasty. OTJR (Thorofare N J) 2022 Jan 22;42(1):40-49. [doi: 10.1177/15394492211038293] [Medline: 34423693]
64. van Dijk-Huisman HC, Weemaes AT, Boymans TA, Lenssen AF, de Bie RA. Smartphone app with an accelerometer enhances patients' physical activity following elective orthopedic surgery: a pilot study. Sensors (Basel) 2020 Aug 02;20(15):4317 [FREE Full text] [doi: 10.3390/s20154317] [Medline: 32748876]
65. Pournajaf S, Goffredo M, Pellicciari L, Piscitelli D, Criscuolo S, Le Pera D, et al. Effect of balance training using virtual reality-based serious games in individuals with total knee replacement: A randomized controlled trial. Ann Phys Rehabil Med 2022 Nov;65(6):101609 [FREE Full text] [doi: 10.1016/j.rehab.2021.101609] [Medline: 34839056]

66. Ficklscherer A, Stapf J, Meissner KM, Niethammer T, Lahner M, Wagenhäuser M, et al. Testing the feasibility and safety of the Nintendo Wii gaming console in orthopedic rehabilitation: a pilot randomized controlled study. Arch Med Sci 2016 Dec 01;12(6):1273-1278 [FREE Full text] [doi: 10.5114/aoms.2016.59722] [Medline: 27904518]
67. Alexander JS, Redfern RE, Duwelius PJ, Berend KR, Lombardi AV, Crawford DA. Use of a smartphone-based care platform after primary partial and total knee arthroplasty: 1-year follow-up of a prospective randomized controlled trial. J Arthroplasty 2023 Jul;38(7 Suppl 2):S208-S214. [doi: 10.1016/j.arth.2023.02.082] [Medline: 36889524]
68. Antunes R, Jacob P, Meyer A, Conditt MA, Roche MW, Verstraete MA. Accuracy of measuring knee flexion after TKA through wearable IMU sensors. J Funct Morphol Kinesiol 2021 Jul 05;6(3):60 [FREE Full text] [doi: 10.3390/jfmk6030060] [Medline: 34287303]
69. Bade M, Cheuy V, Zeni J, Christiansen C, Stevens-Lapsley J. Effects Of real-time biofeedback using instrumented insoles on recovery after total knee arthroplasty: a pilot study. Osteoarthritis and Cartilage 2023 Mar;31:S124. [doi: 10.1016/j.joca.2023.01.078]
70. Prvu Bettger J, Green CL, Holmes DN, Chokshi A, Mather RC, Hoch BT, et al. Effects of virtual exercise rehabilitation in-home therapy compared with traditional care after total knee arthroplasty. The Journal of Bone and Joint Surgery 2019 Nov 18;102(2):101-109. [doi: 10.2106/jbjs.19.00695]
71. Bini S, Mahajan J. Clinical outcomes of remote asynchronous telerehabilitation are equivalent to traditional therapy following total knee arthroplasty: A randomized control study. J Telemed Telecare 2016 Jul 09;23(2):239-247. [doi: 10.1177/1357633x16634518]
72. Campbell KJ, Louie PK, Bohl DD, Edmiston T, Mikhail C, Li J, et al. A novel, automated text-messaging system is effective in patients undergoing total joint arthroplasty. The Journal of Bone and Joint Surgery 2019 Jan 16;101(2):145-151. [doi: 10.2106/jbjs.17.01505]
73. Chapman RM, Moschetti WE, Van Citters DW. Is clinically measured knee range of motion after total knee arthroplasty 'good enough?': A feasibility study using wearable inertial measurement units to compare knee range of motion captured during physical therapy versus at home. Medicine in Novel Technology and Devices 2021 Sep;11:100085. [doi: 10.1016/j.medntd.2021.100085]
74. Christensen JC, Stanley EC, Oro EG, Carlson HB, Naveh YY, Shalita R, et al. The validity and reliability of the OneStep smartphone application under various gait conditions in healthy adults with feasibility in clinical practice. J Orthop Surg Res 2022 Sep 14;17(1):417 [FREE Full text] [doi: 10.1186/s13018-022-03300-4] [Medline: 36104792]
75. Chughtai M, Shah NV, Sultan AA, Solow M, Tiberi JV, Mehran N, et al. The role of prehabilitation with a telerehabilitation system prior to total knee arthroplasty. Ann Transl Med 2019 Feb;7(4):68-68 [FREE Full text] [doi: 10.21037/atm.2018.11.27] [Medline: 30963063]
76. Chughtai M, Kelly J, Newman J, Sultan A, Khlopas A, Sodhi N, et al. The role of virtual rehabilitation in total and unicompartmental knee arthroplasty. J Knee Surg 2018 Jan 16;32(1):105-110. [doi: 10.1055/s-0038-1637018] [Medline: 29548061]
77. Doiron-Cadrin P, Kairy D, Vendittoli P, Lowry V, Poitras S, Desmeules F. Feasibility and preliminary effects of a tele-prehabilitation program and an in-person prehablitation program compared to usual care for total hip or knee arthroplasty candidates: a pilot randomized controlled trial. Disabil Rehabil 2020 Apr 13;42(7):989-998. [doi: 10.1080/09638288.2018.1515992] [Medline: 30638076]
78. LeBrun DG, Martino B, Biehl E, Fisher CM, Gonzalez Della Valle A, Ast MP. Telerehabilitation has similar clinical and patient-reported outcomes compared to traditional rehabilitation following total knee arthroplasty. Knee Surg Sports Traumatol Arthrosc 2022 Dec 26;30(12):4098-4103. [doi: 10.1007/s00167-022-06931-6] [Medline: 35347376]
79. Fung V, Ho A, Shaffer J, Chung E, Gomez M. Use of Nintendo Wii Fit in the rehabilitation of outpatients following total knee replacement: a preliminary randomised controlled trial. Physiotherapy 2012 Sep;98(3):183-188. [doi: 10.1016/j.physio.2012.04.001] [Medline: 22898573]
80. Hong M, Loeb J, Yang M, Bailey JF. Postoperative outcomes of a digital rehabilitation program after total knee arthroplasty: retrospective, observational feasibility study. JMIR Form Res 2022 Sep 19;6(9):e40703 [FREE Full text] [doi: 10.2196/40703] [Medline: 36121690]
81. Klement MR, Rondon AJ, McEntee RM, Greenky MR, Austin MS. Web-based, self-directed physical therapy after total knee arthroplasty is safe and effective for most, but not all, patients. J Arthroplasty 2019 Jul;34(7S):S178-S182. [doi: 10.1016/j.arth.2018.11.040] [Medline: 30591206]
82. Kline PW, Melanson EL, Sullivan WJ, Blatchford PJ, Miller MJ, Stevens-Lapsley JE, et al. Improving physical activity through adjunct telerehabilitation following total knee arthroplasty: randomized controlled trial protocol. Phys Ther 2019 Jan 01;99(1):37-45 [FREE Full text] [doi: 10.1093/ptj/pzy119] [Medline: 30329126]
83. Knapp PW, Keller RA, Mabee KA, Pillai R, Frisch NB. Quantifying patient engagement in total joint arthroplasty using digital application-based technology. J Arthroplasty 2021 Sep;36(9):3108-3117 [FREE Full text] [doi: 10.1016/j.arth.2021.04.022] [Medline: 33965282]
84. Krebs DE, Huddleston JI, Goldvasser D, Scarborough DM, Harris WH, Malchau H. Biomotion community-wearable human activity monitor: total knee replacement and healthy control subjects. 2006 Presented at: International Workshop on Wearable and Implantable Body Sensor Networks (BSN'06); April 3-5; Cambridge, MA. [doi: 10.1109/BSN.2006.13]

85. Kuether J, Moore A, Kahan J, Martucci J, Messina T, Perreault R, et al. Telerehabilitation for total hip and knee arthroplasty patients: a pilot series with high patient satisfaction. HSS J 2019 Oct 21;15(3):221-225 [FREE Full text] [doi: 10.1007/s11420-019-09715-w] [Medline: 31624476]
86. Lam AWK, Varona-Marin D, Li Y, Fergenbaum M, Kulić D. Automated rehabilitation system: movement measurement and feedback for patients and physiotherapists in the rehabilitation clinic. Human–Computer Interaction 2016 Nov 18;31(3):294-334. [doi: 10.1080/07370024.2015.1093419]
87. Losina E, Collins JE, Daigle ME, Donnell-Fink LA, Prokopetz JJ, Strnad D, et al. The AViKA (Adding Value in Knee Arthroplasty) postoperative care navigation trial: rationale and design features. BMC Musculoskelet Disord 2013 Oct 12;14(1):290 [FREE Full text] [doi: 10.1186/1471-2474-14-290] [Medline: 24119081]
88. Milliren CE, Lindsay B, Biernat L, Smith TA, Weaver B. Can digital engagement improve outcomes for total joint replacements? Digit Health 2022 Apr 24;8:20552076221095322 [FREE Full text] [doi: 10.1177/20552076221095322] [Medline: 35493958]
89. Na A, Buchanan TS. Validating wearable sensors using self-reported instability among patients with knee osteoarthritis. PM R 2021 Feb 12;13(2):119-127. [doi: 10.1002/pmrj.12393] [Medline: 32358908]
90. Bell KM, Onyeukwu C, Smith CN, Oh A, Devito Dabbs A, Piva SR, et al. A portable system for remote rehabilitation following a total knee replacement: a pilot randomized controlled clinical study. Sensors (Basel) 2020 Oct 27;20(21):6118 [FREE Full text] [doi: 10.3390/s20216118] [Medline: 33121204]
91. Pellegrini CA, Lee J, DeVivo KE, Harpine CE, Del Gaizo DJ, Wilcox S. Reducing sedentary time using an innovative mHealth intervention among patients with total knee replacement: Rationale and study protocol. Contemp Clin Trials Commun 2021 Jun;22:100810 [FREE Full text] [doi: 10.1016/j.conctc.2021.100810] [Medline: 34195473]
92. Ramkumar PN, Haeberle HS, Ramanathan D, Cantrell WA, Navarro SM, Mont MA, et al. Remote patient monitoring using mobile health for total knee arthroplasty: validation of a wearable and machine learning-based surveillance platform. J Arthroplasty 2019 Oct;34(10):2253-2259. [doi: 10.1016/j.arth.2019.05.021] [Medline: 31128890]
93. Summers SH, Nunley RM, Slotkin EM. A home-based, remote-clinician-controlled, physical therapy device leads to superior outcomes when compared to standard physical therapy for rehabilitation after total knee arthroplasty. J Arthroplasty 2023 Mar;38(3):497-501. [doi: 10.1016/j.arth.2022.10.009] [Medline: 36252744]
94. Tousignant M, Boissy P, Moffet H, Corriveau H, Cabana F, Marquis F, et al. Patients' satisfaction of healthcare services and perception with in-home telerehabilitation and physiotherapists' satisfaction toward technology for post-knee arthroplasty: an embedded study in a randomized trial. Telemed J E Health 2011 Jun;17(5):376-382. [doi: 10.1089/tmj.2010.0198] [Medline: 21492030]
95. Tripuraneni KR, Foran JR, Munson NR, Racca NE, Carothers JT. A smartwatch paired with a mobile application provides postoperative self-directed rehabilitation without compromising total knee arthroplasty outcomes: a randomized controlled trial. J Arthroplasty 2021 Dec;36(12):3888-3893. [doi: 10.1016/j.arth.2021.08.007] [Medline: 34462184]
96. Visperas AT, Greene KA, Krebs VE, Klika AK, Piuzzi NS, Higuera-Rueda CA. A web-based interactive patient-provider software platform does not increase patient satisfaction or decrease hospital resource utilization in total knee and hip arthroplasty patients in a single large hospital system. J Arthroplasty 2021 Jul;36(7):2290-2296.e1. [doi: 10.1016/j.arth.2021.01.037] [Medline: 33581971]
97. Youn I, Youn J, Zeni J, Knarr B. Biomechanical gait variable estimation using wearable sensors after unilateral total knee arthroplasty. Sensors (Basel) 2018 May 15;18(5):1577 [FREE Full text] [doi: 10.3390/s18051577] [Medline: 29762541]
98. Zhao W, Yang S, Luo X. Towards rehabilitation at home after total knee replacement. Tsinghua Science and Technology 2021;26(6):791-799. [doi: 10.26599/TST.2020.9010034]
99. Kramer JF, Speechley M, Bourne R, Rorabeck C, Vaz M. Comparison of clinic- and home-based rehabilitation programs after total knee arthroplasty. Clin Orthop Relat Res 2003 May(410):225-234. [doi: 10.1097/01.blo.0000063600.67412.11] [Medline: 12771834]
100. Bolam SM, Batinica B, Yeung TC, Weaver S, Cantamessa A, Vanderboor TC, et al. Remote patient monitoring with wearable sensors following knee arthroplasty. Sensors (Basel) 2021 Jul 29;21(15):5143 [FREE Full text] [doi: 10.3390/s21155143] [Medline: 34372377]
101. Castle H, Kozak K, Sidhu A, Khan RJK, Haebich S, Bowden V, et al. Smartphone technology: a reliable and valid measure of knee movement in knee replacement. Int J Rehabil Res 2018 Jun;41(2):152-158. [doi: 10.1097/MRR.0000000000000276] [Medline: 29465474]
102. Levinger P, Zeina D, Teshome AK, Skinner E, Begg R, Abbott JH. A real time biofeedback using Kinect and Wii to improve gait for post-total knee replacement rehabilitation: a case study report. Disabil Rehabil Assist Technol 2016 Sep 04;11(3):251-262. [doi: 10.3109/17483107.2015.1080767] [Medline: 26336875]
103. Liptak MG, Theodoulou A, Kaambwa B, Saunders S, Hinrichs SW, Woodman RJ, et al. The safety, efficacy and cost-effectiveness of the Maxm Skate, a lower limb rehabilitation device for use following total knee arthroplasty: study protocol for a randomised controlled trial. Trials 2019 Jan 10;20(1):36 [FREE Full text] [doi: 10.1186/s13063-018-3102-9] [Medline: 30630494]

104. Negus J, Cawthorne D, Chen J, Scholes C, Parker D, March L. Patient outcomes using Wii-enhanced rehabilitation after total knee replacement - the TKR-POWER study. Contemp Clin Trials 2015 Jan;40:47-53. [doi: 10.1016/j.cct.2014.11.007] [Medline: 25460343]
105. Russell TG, Buttrum P, Wootton R, Jull GA. Internet-based outpatient telerehabilitation for patients following total knee arthroplasty. The Journal of Bone and Joint Surgery-American Volume 2011;93(2):113-120. [doi: 10.2106/jbjs.i.01375]
106. Duong V, Dennis S, Harris A, Robbins S, Venkatesha V, Ferreira M, et al. The effects of a disruptive digital technology intervention following total knee replacement: results from the pathway randomised controlled trial. Osteoarthritis and Cartilage 2023 Mar;31:S30-S31. [doi: 10.1016/j.joca.2023.01.521]
107. Hardwick-Morris M, Carlton S, Twiggs J, Miles B, Liu D. Pre- and postoperative physiotherapy using a digital application decreases length of stay without reducing patient outcomes following total knee arthroplasty. Arthroplasty 2022 Aug 02;4(1):30 [FREE Full text] [doi: 10.1186/s42836-022-00133-8] [Medline: 35915464]
108. Farr-Wharton G, Li J, Hussain MS, Freyne J. Mobile Supported Health Services: Experiences in Orthopaedic Care. Presented at: IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS); July 28-30, 2020; Rochester, MN. [doi: 10.1109/cbms49503.2020.00074]
109. McDonall J, de Steiger R, Reynolds J, Redley B, Livingston PM, Hutchinson AF, et al. Patient activation intervention to facilitate participation in recovery after total knee replacement (MIME): a cluster randomised cross-over trial. BMJ Qual Saf 2019 Oct 11;28(10):782-792 [FREE Full text] [doi: 10.1136/bmjqs-2018-008975] [Medline: 30975730]
110. An J, Ryu H, Lyu S, Yi H, Lee B. Effects of preoperative telerehabilitation on muscle strength, range of motion, and functional outcomes in candidates for total knee arthroplasty: a single-blind randomized controlled trial. Int J Environ Res Public Health 2021 Jun 04;18(11):6071 [FREE Full text] [doi: 10.3390/ijerph18116071] [Medline: 34199913]
111. Chiang C, Chen K, Liu K, Hsu S, Chan C. Data collection and analysis using wearable sensors for monitoring knee range of motion after total knee arthroplasty. Sensors (Basel) 2017 Feb 22;17(2):418 [FREE Full text] [doi: 10.3390/s17020418] [Medline: 28241434]
112. Huang Y, Liu Y, Hsu W, Lai L, Lee MS. Progress on range of motion after total knee replacement by sensor-based system. Sensors (Basel) 2020 Mar 18;20(6):1703 [FREE Full text] [doi: 10.3390/s20061703] [Medline: 32197503]
113. Huang P, He J, Zhang YM. [The mobile application of patient management in education and follow-up for patients following total knee arthroplasty]. Zhonghua Yi Xue Za Zhi 2017 May 30;97(20):1592-1595. [doi: 10.3760/cma.j.issn.0376-2491.2017.20.019] [Medline: 28592069]
114. Hung LP, Chao YH, Tseng YL, Chung YL. Constructing a home-based knee replacement exercise monitoring system with G sensor. Presented at: International Conference on Frontier Computing; July 3-6, 2018; Kuala Lumpur, Malaysia. [doi: 10.1007/978-981-10-3187-8_59]
115. Li Y, Gu Z, Ning R, Yin H. Study on the effect of internet plus continuous nursing on functional recovery and medication compliance of patients with knee joint replacement. J Orthop Surg Res 2023 Jun 11;18(1):424 [FREE Full text] [doi: 10.1186/s13018-023-03907-1] [Medline: 37303038]
116. Lou N, Diao Y, Chen Q, Ning Y, Li G, Liang S, et al. A portable wearable inertial system for rehabilitation monitoring and evaluation of patients with total knee replacement. Front Neurorobot 2022 Mar 23;16:836184 [FREE Full text] [doi: 10.3389/fnbot.2022.836184] [Medline: 35401138]
117. Lu G, Wu T, Tan Q, Wu Z, Shi L, Zhong Y. The effect of a micro-visual intervention on the accelerated recovery of patients with kinesiophobia after total knee replacement during neo-coronary pneumonia. Medicine (Baltimore) 2021 Feb 12;100(6):e24141 [FREE Full text] [doi: 10.1097/MD.0000000000024141] [Medline: 33578519]
118. Park KH, Song MR. The effects of postdischarge telephone counseling and short message service on the knee function, activities of daily living, and life satisfaction of patients undergoing total knee replacement. Orthop Nurs 2017;36(3):229-236 [FREE Full text] [doi: 10.1097/NOR.0000000000000332] [Medline: 28363197]
119. Park SA, Jeong Y. The effect of a multidimensional home rehabilitation program for post-total knee arthroplasty elderly patients. Orthop Nurs 2023;42(1):22-32. [doi: 10.1097/NOR.0000000000000913] [Medline: 36702093]
120. Su C, Cheng C. Developing and evaluating creativity gamification rehabilitation system: the application of PCA-ANFIS based emotions model. Eurasia J Math Sci T 2016;12(5):1. [doi: 10.12973/eurasia.2016.1527a]
121. Wang Q, Hunter S, Lee RL, Chan SW. The effectiveness of a mobile application-based programme for rehabilitation after total hip or knee arthroplasty: A randomised controlled trial. Int J Nurs Stud 2023 Apr;140:104455 [FREE Full text] [doi: 10.1016/j.ijnurstu.2023.104455] [Medline: 36821950]
122. Yang C, Shang L, Yao S, Ma J, Xu C. Cost, time savings and effectiveness of wearable devices for remote monitoring of patient rehabilitation after total knee arthroplasty: study protocol for a randomized controlled trial. J Orthop Surg Res 2023 Jun 27;18(1):461 [FREE Full text] [doi: 10.1186/s13018-023-03898-z] [Medline: 37370130]
123. Zhang X, Chen X, Kourkoumelis N, Gao R, Li G, Zhu C. A social media-promoted educational community of joint replacement patients using the WeChat app: survey study. JMIR Mhealth Uhealth 2021 Mar 18;9(3):e18763 [FREE Full text] [doi: 10.2196/18763] [Medline: 33734094]
124. Zheng Q, Chen H. A monitoring system for walking rehabilitation after THR or TKR surgeries. Annu Int Conf IEEE Eng Med Biol Soc Jul;2017:2373-2376. [doi: 10.1109/EMBC.2017.8037333] [Medline: 29060375]

125. Zhang H, Zhou Y. Concept verification of a remote automatic scoring system for evaluating knee function after total knee arthroplasty. J Knee Surg 2021 Mar 27;34(4):464-470. [doi: 10.1055/s-0040-1710568] [Medline: 32462648]
126. Jansson M, Vuorinen A, Harjumaa M, Similä H, Koivisto J, Puhto A, et al. The digital patient journey solution for patients undergoing elective hip and knee arthroplasty: Protocol for a pragmatic randomized controlled trial. J Adv Nurs 2020 Jun 15;76(6):1436-1448. [doi: 10.1111/jan.14343] [Medline: 32133684]
127. Hadamus A, Białoszewski D, Urbaniak E, Kowalska A, Wydra K, Boratyński R, et al. The impact of training in virtual reality on balance in patients after total knee replacement is relatively slight. Gait & Posture 2020 Sep;81:134-135. [doi: 10.1016/j.gaitpost.2020.07.101]
128. Wang Q, Hunter S, Lee RL, Wang X, Chan SW. Patients' needs regarding rehabilitation services delivered via mobile applications after arthroplasty: A qualitative study. J Clin Nurs 2022 Nov 02;31(21-22):3178-3189. [doi: 10.1111/jocn.16152] [Medline: 34859523]
129. Doak CC, Doak LG, Root JH. Teaching Patients with Low Literacy Skills. Philadelphia, PA: J. B. Lippincott Company; 1996.
130. Net Promoter System. URL: https://www.netpromotersystem.com/about/measuring-your-net-promoter-score/ [accessed 2023-12-12]
131. Demiris G, Speedie S, Finkelstein S. A questionnaire for the assessment of patients' impressions of the risks and benefits of home telecare. J Telemed Telecare 2000 Jun 24;6(5):278-284. [doi: 10.1258/1357633001935914] [Medline: 11070589]
132. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly 1989 Sep;13(3):319-340 [FREE Full text] [doi: 10.2307/249008]
133. Parmanto B, Lewis AN, Graham KM, Bertolet MH. Development of the Telehealth Usability Questionnaire (TUQ). Int J Telerehabil 2016 Jul 01;8(1):3-10 [FREE Full text] [doi: 10.5195/ijt.2016.6196] [Medline: 27563386]
134. Cooper DM, Bhuskute N, Walsh G. Exploring the impact and acceptance of wearable sensor technology for pre- and postoperative rehabilitation in knee replacement patients: a U.K.-based pilot study. JB JS Open Access 2022;7(2):1 [FREE Full text] [doi: 10.2106/JBJS.OA.21.00154] [Medline: 35506018]
135. Booth MW, Riegler V, King JS, Barrack RL, Hannon CP. Patients' perceptions of remote monitoring and app-based rehabilitation programs: a comparison of total hip and knee arthroplasty. J Arthroplasty 2023 Jul;38(7S):S39-S43. [doi: 10.1016/j.arth.2023.04.032] [Medline: 37088222]
136. van Kasteren Y, Freyne J, Hussain MS. Total knee replacement and the effect of technology on cocreation for improved outcomes and delivery: qualitative multi-stakeholder study. J Med Internet Res 2018 Mar 20;20(3):e95 [FREE Full text] [doi: 10.2196/jmir.7541] [Medline: 29559424]
137. Jansson MM, Harjumaa M, Puhto A, Pikkarainen M. Healthcare professionals' proposed eHealth needs in elective primary fast-track hip and knee arthroplasty journey: A qualitative interview study. J Clin Nurs 2019 Dec 18;28(23-24):4434-4446. [doi: 10.1111/jocn.15028] [Medline: 31408555]
138. de Batlle J, Massip M, Vargiu E, Nadal N, Fuentes A, Ortega Bravo M, CONNECARE-Lleida Group. Implementing mobile health-enabled integrated care for complex chronic patients: patients and professionals' acceptability study. JMIR Mhealth Uhealth 2020 Nov 20;8(11):e22136 [FREE Full text] [doi: 10.2196/22136] [Medline: 33216004]
139. Boissy P, Tousignant M, Moffet H, Nadeau S, Brière S, Mérette C, Belzile, et al. Conditions of use, reliability, and quality of audio/video-mediated communications during in-home rehabilitation teletreatment for postknee arthroplasty. Telemed J E Health 2016 Aug;22(8):637-649. [doi: 10.1089/tmj.2015.0157] [Medline: 26958932]
140. Parkes RJ, Palmer J, Wingham J, Williams DH. Is virtual clinic follow-up of hip and knee joint replacement acceptable to patients and clinicians? A sequential mixed methods evaluation. BMJ Open Qual 2019 Mar 01;8(1):e000502 [FREE Full text] [doi: 10.1136/bmjoq-2018-000502] [Medline: 31259271]
141. Russell TG, Buttrum P, Wootton R, Jull GA. Low-bandwidth telerehabilitation for patients who have undergone total knee replacement: preliminary results. J Telemed Telecare 2003 Dec 02;9 Suppl 2:S44-S47. [doi: 10.1258/135763303322596246] [Medline: 14728759]
142. Lee M, Suh D, Son J, Kim J, Eun S, Yoon B. Patient perspectives on virtual reality-based rehabilitation after knee surgery: Importance of level of difficulty. J Rehabil Res Dev 2016;53(2):239-252. [doi: 10.1682/jrrd.2014.07.0164]
143. Reid H, Mohammadi S, Watson W, Robillard JM, Crocker M, Westby MD, et al. Patient and caregiver perspectives on an eHealth tool: a qualitative investigation of preferred formats, features and characteristics of a presurgical eHealth education module. Rehabil Process Outcome 2021 Apr 21;10:11795727211010501 [FREE Full text] [doi: 10.1177/11795727211010501] [Medline: 34497456]
144. Culliton SE, Bryant DM, MacDonald SJ, Hibbert KM, Chesworth BM. Effect of an e-Learning tool on expectations and satisfaction following total knee arthroplasty: a randomized controlled trial. J Arthroplasty 2018 Jul;33(7):2153-2158. [doi: 10.1016/j.arth.2018.02.040] [Medline: 29555496]
145. LeBrun DG, Malfer C, Wilson M, Carroll KM, Wang Ms V, Mayman DJ, et al. Telemedicine in an outpatient arthroplasty setting during the COVID-19 pandemic: early lessons from New York City. HSS J 2021 Feb 21;17(1):25-30 [FREE Full text] [doi: 10.1177/1556331620972659] [Medline: 33967638]

146. Argent R, Slevin P, Bevilacqua A, Neligan M, Daly A, Caulfield B. Clinician perceptions of a prototype wearable exercise biofeedback system for orthopaedic rehabilitation: a qualitative exploration. BMJ Open 2018 Oct 25;8(10):e026326. [doi: 10.1136/bmjopen-2018-026326] [Medline: 30366919]
147. McDonall J, Redley B, Livingston P, Hutchinson A, de Steiger R, Botti M. A nurse-led multimedia intervention to increase patient participation in recovery after knee arthroplasty: hybrid type II implementation study. JMIR Hum Factors 2022 May 19;9(2):e36959 [FREE Full text] [doi: 10.2196/36959] [Medline: 35588363]
148. Giunta NM, Paladugu PS, Bernstein DN, Makhni MC, Chen AF. Telemedicine hip and knee arthroplasty experience during COVID-19. J Arthroplasty 2022 Aug;37(8S):S814-S818.e2 [FREE Full text] [doi: 10.1016/j.arth.2022.02.106] [Medline: 35257819]
149. Joshi R, Joseph A, Mihandoust S, Madathil KC, Cotten SR. A mobile application-based home assessment tool for patients undergoing joint replacement surgery: a qualitative feasibility study. Appl Ergon 2022 Sep;103:103796. [doi: 10.1016/j.apergo.2022.103796] [Medline: 35605308]
150. Glinkowski W, Cabaj D, Kostrubała A, Krawczak K, Górecki A. Pre-surgery and post-surgery telerehabilitation for hip and knee replacement - Treatment options review and patient's attitudes towards telerehabilitation. 2011 Presented at: eChallenges e-2010 Conference; October 27-29, 2010; Warsaw, Poland.
151. Moffet H, Tousignant M, Nadeau S, Mérette C, Boissy P, Corriveau H, Belzile, et al. Patient satisfaction with in-home telerehabilitation after total knee arthroplasty: results from a randomized controlled trial. Telemed J E Health 2017 Feb;23(2):80-87. [doi: 10.1089/tmj.2016.0060] [Medline: 27529575]
152. Williams E, Putnam J, Emkes L, Greenwood J. Patient perspectives and numerical evaluation of a COVID secure hybrid rehabilitation programme following knee replacement surgery. Physiotherapy 2022 Feb;114:e224-e225. [doi: 10.1016/j.physio.2021.12.214]
153. Marsh J, Bryant D, MacDonald SJ, Naudie D, Remtulla A, McCalden R, et al. Are patients satisfied with a web-based followup after total joint arthroplasty? Clin Orthop Relat Res 2014 Jun;472(6):1972-1981 [FREE Full text] [doi: 10.1007/s11999-014-3514-0] [Medline: 24562873]
154. Grant S, Blom AW, Whitehouse MR, Craddock I, Judge A, Tonkin EL, et al. Using home sensing technology to assess outcome and recovery after hip and knee replacement in the UK: the HEmiSPHERE study protocol. BMJ Open 2018 Jul 28;8(7):e021862 [FREE Full text] [doi: 10.1136/bmjopen-2018-021862] [Medline: 30056388]
155. Kairy D, Tousignant M, Leclerc N, Côté AM, Levasseur M, Researchers T. The patient's perspective of in-home telerehabilitation physiotherapy services following total knee arthroplasty. Int J Environ Res Public Health 2013 Aug 30;10(9):3998-4011 [FREE Full text] [doi: 10.3390/ijerph10093998] [Medline: 23999548]
156. Hip and knee replacement. OECD iLibrary. URL: http://tinyurl.com/yv339863 [accessed 2023-12-12]
157. Silver L. Smartphone Ownership Is Growing Rapidly Around the World, But Not Always Equally. Pew Research Center. 2019 Feb 05. URL: http://tinyurl.com/39dkhvvd [accessed 2023-12-12]
158. McCool J, Dobson R, Whittaker R, Paton C. Mobile health (mHealth) in low- and middle-income countries. Annu Rev Public Health 2022 Apr 05;43(1):525-539 [FREE Full text] [doi: 10.1146/annurev-publhealth-052620-093850] [Medline: 34648368]
159. Malik IV, Devasenapathy N, Kumar A, Dogra H, Ray S, Gautam D, et al. Estimation of expenditure and challenges related to rehabilitation after knee arthroplasty: a hospital-based cross-sectional study. Indian J Orthop 2021 Oct 02;55(5):1317-1325 [FREE Full text] [doi: 10.1007/s43465-021-00405-6] [Medline: 34824731]
160. Zhou Z, Hou Y, Lin J, Wang K, Liu Q. Patients' views toward knee osteoarthritis exercise therapy and factors influencing adherence - a survey in China. Phys Sportsmed 2018 May 16;46(2):221-227. [doi: 10.1080/00913847.2018.1425595] [Medline: 29307256]
161. Taylor CE, Murray CM, Stanton TR. Patient perspectives of pain and function after knee replacement: a systematic review and meta-synthesis of qualitative studies. PR9 2022 May 09;7(3):e1006. [doi: 10.1097/pr9.0000000000001006]
162. Reddy H, Joshi S, Joshi A, Wagh V. A critical review of global digital divide and the role of technology in healthcare. Cureus 2022 Sep;14(9):e29739. [doi: 10.7759/cureus.29739] [Medline: 36340549]
163. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun 13;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
164. Voorheis P, Zhao A, Kuluski K, Pham Q, Scott T, Sztur P, et al. Integrating behavioral science and design thinking to develop mobile health interventions: systematic scoping review. JMIR Mhealth Uhealth 2022 Mar 16;10(3):e35799 [FREE Full text] [doi: 10.2196/35799] [Medline: 35293871]
165. Hussain MS, Li J, Brindal E, van Kasteren Y, Varnfield M, Reeson A, et al. Supporting the delivery of total knee replacements care for both patients and their clinicians with a mobile app and web-based tool: randomized controlled trial protocol. JMIR Res Protoc 2017 Mar 01;6(3):e32 [FREE Full text] [doi: 10.2196/resprot.6498] [Medline: 28249832]
166. Bade MJ, Christensen JC, Zeni JA, Christiansen CL, Dayton MR, Forster JE, et al. Movement pattern biofeedback training after total knee arthroplasty: Randomized clinical trial protocol. Contemp Clin Trials 2020 Apr;91:105973 [FREE Full text] [doi: 10.1016/j.cct.2020.105973] [Medline: 32171937]

167. Wang X, Hunter DJ, Robbins S, Capistrano S, Duong V, Melo L, et al. Participatory health through behavioural engagement and disruptive digital technology for postoperative rehabilitation: protocol of the PATHway trial. BMJ Open 2021 Jan 17;11(1):e041328 [FREE Full text] [doi: 10.1136/bmjopen-2020-041328] [Medline: 33455931]
168. McDonall J, de Steiger R, Reynolds J, Redley B, Livingston P, Botti M. Patient participation in postoperative care activities in patients undergoing total knee replacement surgery: Multimedia Intervention for Managing patient Experience (MIME). Study protocol for a cluster randomised crossover trial. BMC Musculoskelet Disord 2016 Jul 18;17(1):294 [FREE Full text] [doi: 10.1186/s12891-016-1133-5] [Medline: 27431239]
169. Argent R, Slevin P, Bevilacqua A, Neligan M, Daly A, Caulfield B. Wearable sensor-based exercise biofeedback for orthopaedic rehabilitation: a mixed methods user evaluation of a prototype system. Sensors (Basel) 2019 Jan 21;19(2):432 [FREE Full text] [doi: 10.3390/s19020432] [Medline: 30669657]
170. Szöts K, Konradsen H, Solgaard S, Østergaard B. Telephone follow-up by nurse after total knee arthroplasty: results of a randomized clinical trial. Orthop Nurs 2016;35(6):411-420. [doi: 10.1097/NOR.0000000000000298] [Medline: 27851679]

Abbreviations

CENTRAL: Cochrane Central Register of Controlled Trials
JBI: Joanna Briggs Institute
LMIC: low and middle-income countries
mHealth: mobile health
PRISMA-ScR: Preferred Reporting Items for Systematic Review and Meta-Analyses Extension for Scoping Reviews
RCT: randomized controlled trial
ROM: range of motion

Edited by L Buis; submitted 04.04.23; peer-reviewed by A Timmermans, T Li; comments to author 03.07.23; revised version received 10.10.23; accepted 01.12.23; published 26.01.24.

Please cite as:
Pritwani S, Shrivastava P, Pandey S, Kumar A, Malhotra R, Maddison R, Devasenapathy N
Mobile and Computer-Based Applications for Rehabilitation Monitoring and Self-Management After Knee Arthroplasty: Scoping Review
JMIR Mhealth Uhealth 2024;12:e47843
URL: https://mhealth.jmir.org/2024/1/e47843
doi:10.2196/47843
PMID:38277195

Review

User Perception of Smart Home Surveillance Among Adults Aged 50 Years and Older: Scoping Review

Jessica Percy Campbell[1,2], PhD; Jacob Buchan[2]; Charlene H Chu[2,3], GNC(c), RN, PhD; Andria Bianchi[2,4], PhD; Jesse Hoey[2,5], PhD; Shehroz S Khan[2,6], PhD

[1]Political Science, University of Victoria, Victoria, BC, Canada

[2]KITE Research Institute, Toronto Rehabilitation Institute, University Health Network, Toronto, ON, Canada

[3]Lawrence S Bloomberg Faculty of Nursing, University of Toronto, Toronto, ON, Canada

[4]Centre for Clinical Ethics, Unity Health Toronto, Toronto, ON, Canada

[5]David R Cheriton School of Computer Science, University of Waterloo, Waterloo, ON, Canada

[6]Institute of Biomedical Engineering, University of Toronto, Toronto, ON, Canada

Corresponding Author:
Shehroz S Khan, PhD
Institute of Biomedical Engineering
University of Toronto
Rosebrugh Bldg, 164 College St
Toronto, ON, M5S 3G9
Canada
Phone: 1 4169787459
Email: shehroz.khan@uhn.ca

Abstract

Background: Smart home technology (SHT) can be useful for aging in place or health-related purposes. However, surveillance studies have highlighted ethical issues with SHTs, including user privacy, security, and autonomy.

Objective: As digital technology is most often designed for younger adults, this review summarizes perceptions of SHTs among users aged 50 years and older to explore their understanding of privacy, the purpose of data collection, risks and benefits, and safety.

Methods: Through an integrative review, we explored community-dwelling adults' (aged 50 years and older) perceptions of SHTs based on research questions under 4 nonmutually exclusive themes: privacy, the purpose of data collection, risk and benefits, and safety. We searched 1860 titles and abstracts from Ovid MEDLINE, Ovid Embase, Cochrane Database of Systematic Reviews, and Cochrane Central Register of Controlled Trials, Scopus, Web of Science Core Collection, and IEEE Xplore or IET Electronic Library, resulting in 15 included studies.

Results: The 15 studies explored user perception of smart speakers, motion sensors, or home monitoring systems. A total of 13 (87%) studies discussed user privacy concerns regarding data collection and access. A total of 4 (27%) studies explored user knowledge of data collection purposes, 7 (47%) studies featured risk-related concerns such as data breaches and third-party misuse alongside benefits such as convenience, and 9 (60%) studies reported user enthusiasm about the potential for home safety.

Conclusions: Due to the growing size of aging populations and advances in technological capabilities, regulators and designers should focus on user concerns by supporting higher levels of agency regarding data collection, use, and disclosure and by bolstering organizational accountability. This way, relevant privacy regulation and SHT design can better support user safety while diminishing potential risks to privacy, security, autonomy, or discriminatory outcomes.

(JMIR Mhealth Uhealth 2024;12:e48526) doi:10.2196/48526

KEYWORDS

smart homes; privacy; surveillance; ambient assisted living; smart speakers; Internet of Things; sensors; sensor; smart home; perception; perceptions; elderly; older adult; older adults; review methods; review methodology; home monitoring; security; safety; ageing; ageing-in-place; integrative review; integrative reviews

Introduction

Smart home technologies (SHTs) typically consist of one or more devices connected through the Internet of Things, which can transmit user data to various stakeholders [1]. Commonly used SHTs include Wi-Fi–enabled cameras, smart speakers with embedded voice assistants, or ambient assisted-living networks of sensors. SHTs are often controllable through smartphones, web platforms, or voice interaction [2]. These networked devices can be useful to the general population for a variety of reasons, but specifically for the aging population, they allow monitoring health status and enable information sharing with health care practitioners, family, or caregivers, potentially alleviating pressure on such networks [3-6]. Until recently, researchers have noted a gap in user perception studies focusing on older adults' (aged 65 years and older) unique needs, preferences, and ethical factors in SHT adoption or decision-making [4]. Others have noted the need for further research that involves users from older age groups outside of the laboratory [2]. Overall, researchers have an active interest in better understanding user perceptions to remove the barriers to SHT adoption for aging populations.

Related studies have also focused on the pressing ethical implications of SHTs in terms of privacy, autonomy, and security [3,5-7]. Insights from surveillance studies and gerontology literature warn that such systems can limit user autonomy by flagging spontaneous behavior as "abnormal or deviant" in ways that could discourage users from deviating from daily routines where movements are continually monitored [3,5]. Others have noted the potential for exploiting vulnerable SHT users through surveillance capitalism, in which user behavioral data are commodified by commercial actors, resulting in exacerbated power imbalances [7,8]. For instance, some commercial-grade smart devices have the potential to make behavioral data available to advertisers, third parties, and insurance companies in ways that can disproportionately and negatively affect vulnerable individuals and groups [1,7,8]. Moreover, security issues with any internet-enabled technology can lead to unauthorized data access by malicious actors, exacerbating the potential for harm [9,10].

With these insights in mind, the aim of this paper is to explore the potential benefits and drawbacks of SHTs from the perspective of users aged 50 years and older. Despite the abovementioned privacy and security risks, it has been well established that SHT users are often limited in their knowledge of the purpose of SHT data collection [11,12]. On the other hand, SHTs are often seen as safety-enhancing [13]. Moreover, as mentioned in our related larger review paper on SHT users of all ages (Percy Campbell et al, unpublished data, January 2024), user perception studies frequently pertain to younger populations and such technology is more often designed for younger groups [14]. Because of the usual emphasis on younger age groups and technology, our goal is to incorporate the views of older demographics regarding the paradoxical benefits and drawbacks of SHTs. To do so, we collected user perception studies related to 4 nonmutually exclusive themes: privacy, the purpose of data collection, risk and benefits, and safety. To our knowledge, we are the first to compile research findings spanning these 4 categories, leading to unique insights that can inform private sector data protection regulation and SHT design, especially for older adults. We constructed four research questions prior to our literature search. (1) Privacy: What are SHT users' privacy attitudes? (2) Purpose: What are SHT users' understandings of the purpose of why and how their data are collected? (3) Risk or benefits: What do users think about the possible benefits and potential risks of harms of SHTs? (4) Safety: What are SHT users' safety perceptions?

Methods

Overview

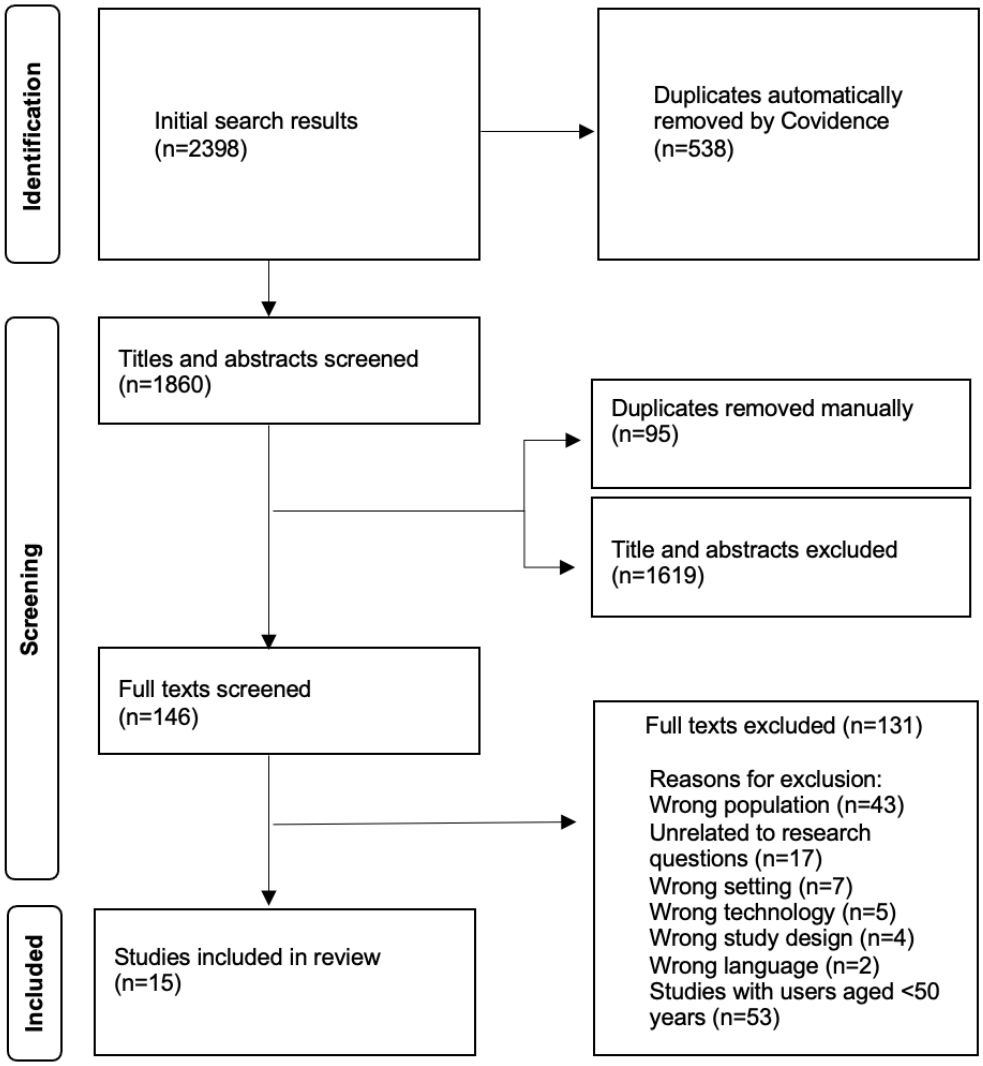
This section outlines our search strategy and the inclusion and exclusion criteria for paper selection. Research questions were crafted to examine the interdisciplinary literature on user perceptions of smart home surveillance. We used an integrative review framework to provide an established, rigorous, and comprehensive review method. An integrative approach is well suited to consolidating an expansive range of articles from varied theoretical backgrounds and empirical methods, allowing for a deeper understanding of a given phenomenon [15].

Search Strategy

The search for peer-reviewed English studies was conducted in October 2021. The research team selected relevant keywords based on 4 research questions listed in the previous section. A health information specialist helped to further identify and refine the search keywords (Multimedia Appendix 1) and selected the following databases to find relevant journal articles using the following databases: Ovid MEDLINE, Ovid Embase, Cochrane Database of Systematic Reviews (Ovid), Cochrane Central Register of Controlled Trials (Ovid), Scopus; Web of Science Core Collection, and IEEE Xplore or IET Electronic Library (IEL). No time frame for the publication date was specified. The results were imported into Covidence reference management software to manage the screening process. The duplicate studies were removed automatically by Covidence and manually by team members.

Following duplicate removal, 4 team members were involved in the review process, which included a title or abstract screening round and a full-text review screening round. Articles were eligible for full-text review if they initially appeared to meet inclusion criteria in the title and abstract phase. Next, in the full-text review phase, each article was read in full and subsequently accepted or rejected based on inclusion and exclusion criteria. To ensure reliability and to mitigate subjective biases, article selection in each research phase required acceptance from 2 team members working independently. The rare instance of disagreement between researchers over whether to accept or reject an article was resolved through the involvement of other team members in weekly team meetings. The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram (Figure 1) outlines the screening processes.

Figure 1. PRISMA flow diagram of extracted studies. PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.



Inclusion and Exclusion Criteria

Eligible studies were those focused predominantly on smart home users' perceptions of privacy, data collection purposes, perceived risks or benefits, and physical safety impacts of SHTs. Selected studies included SHT users aged 50 years and older, aside from 1 study where the participant age range spanned from 48 to 84 years with a mean age of 67.8 years and the aging population was the key focus [16]. The studies were included if participants were community dwelling rather than residing in care home facilities (eg, nursing homes), assisted living facilities, or hospitals. These clinical settings produce additional complexities associated with SHTs (eg, legal requirements, staff, and patient or resident consent), which were beyond the scope of this review. Qualitative, quantitative, and mixed methods empirical peer-reviewed studies that measured user perception of smart home surveillance were included. Common methods from accepted studies included questionnaires, surveys, interviews, and focus groups.

The following enumerates our exclusion criteria: (1) studies focused exclusively on wearables or smartphones due to their use outside of the home and further surveillance potential such as location tracking; (2) smart energy grid studies focused solely on cost or energy sustainability; (3) technical articles (eg, algorithm or system development) or theoretical articles; (4) system feasibility studies that were unrelated to user perception; (5) other review papers; (6) usability or acceptability studies that were unrelated to privacy, data collection purpose, risk, or safety.

Thematic Analysis

Once the full-text screening phase had been completed, key details from each accepted study were entered into a shared Microsoft Excel sheet. The sheet was used to record the article title, author, publication year, country, method, demographic breakdown of participants (age or gender), and a short summary of key findings under the corresponding themes of privacy, the purpose of data collection, risk or benefits, and safety. The articles were classified under 1 or more themes when insights could be linked to our related research questions. These findings were then reported throughout the results section, which has been organized by theme. Summarizing articles by theme allows for patterns and contradictions to emerge from the data, ultimately facilitating analysis in the discussion section.

Results

Study Characteristics

Throughout the first phase, 2398 titles and abstracts were screened by our research team. The team selected 146 articles for full-text review, of which 78 were excluded based on the screening criteria mentioned in the previous section. The 68 remaining articles were selected for a larger user perception review paper on users of all age groups. Of those 68 studies, the 15 studies that focused primarily on adults aged 50 years and older were included here. Publication dates ranged from 2011 to October 2021. The results included 6 qualitative, 5 quantitative, and 4 mixed methods studies. Sample sizes ranged from 8 to 447 participants. Studies were conducted in the United States (n=4), the United Kingdom (n=3), Philippines (n=1), South Korea (n=1), Germany (n=1), Finland (n=1), the Netherlands (n=1), and 3 studies sampled participants from India, Thailand, Indonesia, and Malaysia. Participants' mean ages ranged from 66.4 years to 86.67 years. Further demographic details are available in Table 1.

Table 1. Study characteristics: location, SHT[a] type, user demographics, method, and theme.

Reference	Location	SHT type	Demographic of participants	Method	Key themes
Albina and Hernandez [17]	Philippines	Sensors, cameras	• N=118 • Age range: 60 to ≥81 years • 34.3% female • 65.7% male	Survey	Enhanced safety and privacy concerns
Arthanat et al [18]	United States	Thermostats, voice assistants, home security systems, cameras, and remote controlled lights and appliances	• N=447 • Age range: 65-95 years • 68.8% female • 31.2% male	Survey	Enhanced safety
Choi et al [2]	United States	Smart speakers, cameras, door or window sensor, and multisensor	• N=37 • Age range: N/A[b] • 75% female • 22% male	Semistructured interview	Privacy concerns
Chung et al [19]	United States	Smart speakers (voice assistant)	• N=47 • Age range: N/A • 49% female • 51% male	Survey	Privacy concerns and risk or benefits (lack of affordability)

Ghorayeb et al [13]	United Kingdom	Nonmedical sensors: (SPHERE[c] system: environmental and wearable and video sensors	• N=13
			• Group 1 (users): N=6
			  • Age range: 66-88 years
			  • 67% female
			  • 33% male
			• Group 2 (nonuser): N=7
			  • Age range: 67-89 years
			  • 57% female
			  • 43% male
				Focus groups	Privacy concerns, safety concerns, and purpose of data collection (unaware or forgetful)

Jo et al [20]	South Korea	Environmental sensors, Bluetooth smart bands, and receivers	• N=9 • Age range: 68-87 years • 100% female	Focus groups	Fewer privacy concerns, purpose of data collection (aware), and enhanced safety
Kirchbuchner et al [16]	Germany	Sensors	• N=60 • Age range: 48-84 years • 70% female • 30% male	Survey	Privacy concerns, risk or benefits, and enhanced safety
Leikas and Kulju [21]	Finland	Sensors	• N=8 • Age range: 70-85 years • 88% female • 12% male	Focus groups	Enhanced safety, risk or benefits (improved independent living), privacy concerns, and purpose of data collection (unaware)
O'Brien et al [22]	United States	Smart speakers (voice assistant)	• N=125 • Age range: N/A • Sex: N/A	Thematic analysis of Amazon smart speaker reviews	Enhanced safety
Pal et al [4]	India, Thailand, Indonesia, and Malaysia	Smart homes	• N=239 • Age range: 55 to 75+ years • 34.3% female • 65.7% male	Survey	Privacy concerns and risk or benefits (security concerns)

Reference	Location	SHT type	Demographic of participants	Method	Key themes
Pal et al [23]	India, Thailand, Indonesia, and Malaysia	Smart homes	• N=239 • Age range: 55 to 75+ years • 34.3% female • 65.7% male	Survey	Privacy concerns and risk or benefits (lack of affordability)
Pal et al [24]	India, Thailand, Indonesia, and Malaysia	Smart homes	• N=239 • Age range: 55 to 75+ years • 34.3% female • 65.7% male	Survey	Privacy concerns and risk or benefits (poor design)
Psychoula et al [25]	United Kingdom	Sensors	• N=41 • Age range: N/A • 76% female • 24% male	Survey or semistructured interview	Privacy concerns (limited) and purpose of data collection (aware)
Rogerson et al [26]	United Kingdom	Sensors (light, temperature, and movement)	• N=19 • Age range: N/A • 47% female • 53% male	Semistructured interview	Privacy concerns (limited) and enhanced safety
Van Hoof et al [27]	Netherlands	Mobility monitoring, voice response, fire detection, and wandering prevention	• N=12 • Age range: 63-87 years • 83% female • 17% male	Semistructured qualitative interview	Privacy concerns (limited), enhanced safety, and risk (dependence on internet)

[a]SHT: smart home technology.

[b]N/A: not applicable.

[c]SPHERE: Sensor Platform for Healthcare in a Residential Environment

Thematic Results

Overview

Table 1 shows the same papers classified under our research questions related to privacy, the purpose of data collection, risks and benefits, and safety. A total of 13 studies related to user privacy perception, 4 studies explored users' understandings of the purpose of their data collection, 7 studies related to the benefits and risks of SHTs, while 9 studies pertained to user safety perception.

In summary, our results show that users display a diverse range of perspectives on privacy, the purpose of data collection, risks and benefits, and safety. Although privacy is often seen as an important value in need of protection, nuanced perspectives showed that users were more comfortable with certain types of devices over others, and more comfortable with certain groups of data recipients than others. Participants were sometimes unaware of SHT data collection purposes, although others expressed higher levels of understanding when their SHT functions were adequately explained to them. However, in some cases, the details of SHTs' purpose and function were forgotten over time. Often, users lacked confidence in explaining what data their devices collected or why. Security risks, including hacking and data breaches, were often cited user concerns, while SHT convenience was frequently seen as a major benefit. Overall, despite varying levels of concern in the aforementioned areas, users were generally enthused about safety-enhancing benefits of SHTs. These safety benefits may be especially important to older adults with health concerns in particular, as SHTs can act as emergency contact tools or direct lines of support with health care providers, caregivers, and family. In what follows, the key findings are explained in further detail and separated by theme.

Privacy: What Are SHT Users' Privacy Attitudes?

A total of 13 studies discussed privacy perspectives in general, with some flagging privacy as an important consideration in SHT adoption [2,17]. This section explores a range of user privacy attitudes, where environmental or motion sensors were the most accepted type of SHT [20,26] compared to smart speakers or cameras which were considered invasive [2,16], participants indicated the need to control access to their SHT data [13,21]. Many participants were skeptical about the handling of their personal information by service providers and third parties [4,23,24]. However, over time, some users forgot about the presence of sensors that collected electricity, light, temperature, and movement data [26]. In another study, most participants forgot about the presence of nonmedical sensors that collected environmental and activity data in the home [13].

Certain SHTs were generally viewed as nonthreatening, such as door and window sensors, multisensors [2], fall detection and health monitoring sensors [20], or light, temperature, and movement sensors [26]. Some participants considered sensor systems to be preferable in comparison to alternative arrangements such as institutionalization, which was generally seen as undesirable due to a lack of privacy and restricted visitation rights [27]. In early smart home models, networks of motion sensors, fall detectors, emergency voice response, and fire detection sensors were seen preferably by users, except for 1 participant who removed the technology from her home due

to privacy and autonomy concerns [27]. This participant enjoyed standing in the hallway which would set off alarms, and in turn, alerted staff. However, although most participants did not feel they were being "watched or monitored" the authors also noted that "some are even not fully aware of the presence of the UAS (Unattended Autonomous Surveillance) system at home" [25].

In terms of data sharing, participants expressed mixed attitudes. In a UK study, interviews (n=41) showed older adults were open to having behavioral data collected and shared with family members or health care providers. Here, researchers noted that older adults were more open to sharing data for health care purposes than younger people [25]. One study from South Korea reported that participants (n=9) used environmental sensors for energy management and health-related sensors for fall detection and activity monitoring and reported willingness to share their health-related data with friends, family, and health care practitioners [20]. By contrast, others expressed the need to limit their data sharing to select parties. Those equipped with environmental, wearable, and video sensors in the United Kingdom preferred to share data with health care practitioners instead of family or friends (n=7) [13]. In a survey of 118 older adults (aged 60 years and older) in the Philippines, participants were concerned about assistive technology data access and sharing from environmental sensors and cameras [17].

In Germany, older adults (n=60) perceived cameras to be privacy-invasive compared to other SHTs. Here, privacy was prioritized above other potentially relevant adoption considerations, such as ease of use. Again, study participants preferred to limit data sharing, rejecting commercial service providers as legitimate data recipients [16]. Additionally, in a US study (n=37), the IP web cameras were considered more invasive than other SHTs such as smart speakers, door and window sensors, or multisensors [2]. Some participants expressed concern over smart speaker developers listening in to private conversations, while others were unperturbed [2]. Users were comfortable using smart speakers for certain purposes such as alarms, reminders, and searching for online information. However, many were hesitant to use medical SHTs that shared their health data; 1 user specifically noted their discomfort with the potential for pharmaceutical companies to profile them with targeted advertisements based on health-related data [2]. In another US survey with 47 Amazon smart speaker (Alexa) users, some participants expressed concern over their conversations being monitored, while others were indifferent [19]. The following section discusses the extent to which users understood the purpose of their SHT data collection.

Purpose: What Are SHT Users' Understandings of the Purpose of Why and How Their Data Are Collected?

A total of 4 studies revealed insight into what participants understood about the purposes of their SHT data collection. In using SHTs for health care purposes, participants in Jo et al [20] were generally aware of the purposes of their sensor data collection. In this study, researchers had explained to participants what data were collected, how they were stored, and who had access to their data. However, study participants most often relied on support networks such as friends, family, or neighbors to help with their privacy decisions regarding SHTs. One issue arose, however, when participants in another study were taught about the functions of their SHTs. They eventually forgot the purposes of why those sensors were installed and, by extension, what information was being transmitted [13]. Focus group participants expressed a lack of confidence in their knowledge of whether sensors were measuring water consumption levels, humidity levels, body movement, the number of people in the room, and how alarms are triggered by artificial intelligence (AI) [13]. Similarly, focus group participants in Finland lacked confidence in their knowledge of who had access to their movement sensor data, for what purposes, and whether they had access to it themselves [21]. As noted elsewhere, SHT users often have a sense that privacy issues are present, but are unsure of "what data is collected, or how or why" [19]. Overall, apart from 1 study under this category [20], participants expressed limited understanding of the purposes of the collection of their SHT data [13,21,25]. As will be further discussed, a lack of privacy literacy around the types of surveillance SHTs contribute to is an issue with users of all age groups, younger demographics included. A lack of literacy in this area may result in difficulties in obtaining ongoing consent and informed decision-making regarding SHT use.

Risk and Benefits: What Do Users Think About the Possible Benefits and Potential Risks of Harm of SHTs?

A total of 7 studies explored user perceptions of SHT risks and benefits. Overall, participants identified data security threats as significant risks [16,19,21,23,24]. These perceived risks were mostly divided between disquiet over malicious data breaches, such as through hacking and misuse of personal data by smart home providers. Malicious data breaches were generally characterized by users as the unauthorized access of data by criminal parties; 1 survey (n=60) found that participants, the majority of whom had no prior experience with SHTs, were mainly concerned with criminal access to their data [16]. Similarly, the larger survey sample (n=237) in the study by Pal et al [23] showed that older adults did not trust smart home companies to securely handle their data or prevent data breaches. Specifically, the authors found that SHT users wanted their personal data to be anonymized and did not trust SHT providers to provide adequate or desired protections [23].

Alongside malicious data breaches, the misuse of personal data by SHT providers was consistently described as a risk by study participants. These concerns were mirrored in commercial contexts: focus group discussions (n=14) showed that SHT users were knowledgeable about the collection of their consumer data and were uncomfortable with their lack of agency in the use of the data [21]. Another study by Pal et al [24] (n=239) further affirmed that SHT users are uncomfortable with corporate access to their personal information. From specific medical and commercial contexts to overarching sentiments, users appear dissatisfied with the current levels of data protection offered by SHTs.

Users described additional risks beyond data collection. These included concerns over SHT dependence: semistructured interviews (n=12) demonstrated that power outages or system

failures were flagged as risks by SHT users [27]. Participants also expressed concern over steep learning curves with new SHTs, as well as a potential lack of agency in selecting their own devices and controlling use of the devices [24]. Additionally, affordability was consistently identified as a risk, with concerns that SHTs would not offer benefits worth their price [4]. This finding was reiterated by Chung et al [19] where 24 of 47 (51%) surveyed users reported that affordability surpassed other risks. Overall, malfunction, affordability, and user trust represent additional risks identified by older adults using SHT.

Alongside risks, participants aged 50 years and older noted distinct positive benefits conferred by SHTs. Users were commonly enthusiastic about assistive smart home devices, including mobility monitoring, voice response, fire detection, and wandering prevention technology. Participants believed these SHTs gave them greater independence and reduced the burden on supportive family members and caregivers [21,27]. Similarly, the survey responses (n=239) in the study by Pal et al [23] indicated that users enjoyed home automation, which increased their daily convenience, especially those users who experienced or expected to experience physical or cognitive ailments. Survey participants simultaneously identified the abovementioned risks while reportedly appreciating SHTs' value, thus creating the need to trade their reservations for SHT convenience [24]. Finally, users gained self-confidence with digital technology by mastering newly installed SHTs; however, learning to use the devices was sometimes perceived as a barrier and a deterrent to use [19]. These varied findings comprise the social benefits identified by users; however, enhanced physical safety was among the most noted. The following section explores safety perceptions in greater detail.

Safety: What Are SHT Users' Safety Perceptions?

A total of 9 studies discussed the role of SHTs in safety enhancement, where participants were generally enthused about their devices' safety features and support for aging in place [16,17,20-22,26,27]. Safety has been viewed as an important component of smart home adoption, especially for older adults with health issues [18]. Sensor users have even expressed the need to trade their privacy for increased safety through SHTs, especially for older adults living alone who experience memory problems [23].

All types of SHTs were considered useful for safety purposes, especially in an accident or emergency. In one example, stroke survivors in the United Kingdom felt safer using motion sensors in the home, as these gave them the feeling that they were being looked after [26]. In South Korea, participants found environmental and wearable sensors to be beneficial for aging in place and reported enhanced feelings of safety [20]. The ability to share their behavioral data was seen as a form of safety assurance among users [20]. Participants using ambient intelligent systems also reported enhanced feelings of safety in the home in the Netherlands, especially in the event of a fall or when feeling unwell and unable to access the phone [27]. Fall detection and other health-related safety features and burglary detection were generally well accepted [16]. Likewise, in the Philippines, assistive technology users reported enthusiasm about increased feelings of safety in the home through emergency response features [17]. Sensor Platform for Healthcare in Residential Environment users were subject to environmental, wearable, and video sensors. They were mainly concerned about the limited ability of human operators to react quickly enough in the event of an emergency [13].

Smart speakers embedded with voice assistants were also perceived to improve safety [22]. In a study examining 125 Amazon smart speaker reviews, safety features were commonly mentioned by older adults and caregivers. For example, emergency contact features such as "Ask My Buddy" were popular among reviewers [22]. In the words of one reviewer [22],

> *If I call out "Alexa, tell My Buddy to alert contacts," she sends an alert via cell phone voice and text to my contact list telling them to check on me. This is great in the event of anything from a slip in the shower to any medical or emergency issue or if I feel in danger.*

Throughout our collected studies, this sentiment appears to reflect the preference for the safety-enhancing features that SHTs may provide among many adults aged 50 years and older.

Discussion

Paradoxical Nature of SHTs

Throughout the analysis of 15 studies, study participants identified numerous benefits and drawbacks of SHTs. Overall, our findings indicated that SHT users aged 50 years and older found value in SHTs for several reasons beyond the superficial purposes of convenience or entertainment. Perceived benefits included enhanced independence levels for older adults and increased confidence levels with technology [19,21,27]. They were also widely perceived to support health and well-being through fall prevention or emergency contact features and were seen to enhance physical safety levels at home [16,17,20-22,26,27].

Conversely, participants voiced several concerns pertaining to device affordability [4,19], device reliability, criminal data breaches [16], or a lack of trust in SHT companies in securing user data [23]. If user consent over SHT data collection is to be considered meaningful, it should be ongoing, which poses an issue in cases where users expressed limited understanding of data flows and access [21]. As mentioned by homecare field professionals and related employees, this challenge can be particularly difficult for those who develop memory issues in later years [21]. This is complicated by the fact that privacy concerns sometimes fade over time and participants sometimes forget about the existence of their SHTs altogether [13,26]. Low levels of understanding around SHT data collection purposes, use, and disclosure span all age groups [28] but may be especially detrimental to older people who have SHTs installed by others for health and safety purposes.

It follows that the need for higher levels of user autonomy regarding data access is a consistent finding that requires further attention [4,16,21,23-25]. This is especially important because SHTs are often marketed in ways that promote increased autonomy for older adults. However, if SHT settings are not

carefully configured and managed, they may increase independence in some ways while simultaneously diminishing it in others [28]. Data sharing and intrusive monitoring may create issues related to privacy, autonomy, or attempts at behavioral control [29]. Higher levels of user autonomy would require a strong understanding and access to controls over data monitoring and use. Otherwise, these issues can be mitigated by designers embedding tightly controlled "privacy by default" settings. The SHTs should occasionally prompt users to review and manage privacy settings and restrict data flows where unnecessary for device functionality. Finally, some older adult users may consider creating a set of guidelines and privacy preferences for caregivers to follow should memory or cognitive capacities diminish.

Despite a general unease with data sharing among third parties or service providers, one major limitation of our selected studies is the lack of detailed participant discussion on the potential for SHTs to influence insurance rates, targeted ads, or the increased difficulty in differentiating consumer data from health data. Recall that participants rejected commercial providers as data recipients [2,16]. Commercial SHTs such as smart speakers commodify user data [1,30,31], potentially inferring user health data in the process [32] and sharing such information with third parties with unknown goals or incentives. In cases where commercial-grade SHTs are used as care or safety devices for older people with health issues (eg, [19,22]), should such data still be commodified by commercial actors? SHTs can reveal mental and physical health status, mood, personality traits, and sensitive activity recognition, among other personal details [32-34]. In some cases, SHT or wearable data can also be used to influence personalized insurance rates in ways that may be disadvantage older adults with health issues [7].

Moreover, SHT developers in health care spaces have noted the difficulty in differentiating what is or should be considered medical and health data versus what is not [5]. If users are unaware of what types of data they are sharing (eg, [13,21]), to what extent is autonomous decision-making enabled or respected? When commercial-grade SHTs are used to infer health data, they may be treated as consumer data, facilitating access by public and private sector actors outside of user knowledge or meaningful consent. Many people would likely object to commercial actors gaining access to health care data from hospital settings for the purposes of third-party advertising, yet inferring user behavioral and health patterns through SHTs and wearables is possible. The discriminatory issues with targeted advertising, data brokers, and marketer classifications of different groups of people are well known [35]. Currently, studies linking the ways that SHT data contribute to targeted ads through behavioral patterns or biometric markers such as voice are in their infancy [30,33]. Additional research is needed on how SHT data from older adults are treated by SHT companies; what the subsequent targeted advertising or personalized insurance outcomes may be, either now or in the future; and whether such outcomes are discriminatory in nature.

For these reasons, the ethical implications of inferring health-related data from commercial SHT products should be considered alongside the abovementioned user privacy concerns. In short, the challenges in protecting SHT user privacy and autonomy are ongoing [3] and can be further complicated by the involvement of inferred or self-reported health-related data. As has been recommended elsewhere [2,6,14,29,36] SHT developers should prioritize design choices that better support members of all age groups through user-centric design, considering multiple stakeholders, such as older adults, nurses, and caregivers. Others have advocated for an ethical by-design (EbD) approach to implementing digital technology not only through co-design and product development but also through transdisciplinary research [29,37]. Alongside EbD choices, private sector privacy regulation could further protect users through a data justice approach that privileges human rights over commercial interests.

At a global level, the technological ability to collect and aggregate data for surveillance has outpaced regulatory mechanisms [38]. Using a data justice framework is a logical path forward to the ethical use of technology in ways that benefit both individuals and groups without further disempowering them through surveillance imperatives that do not suit their needs. Taylor's [38] data justice framework includes three pillars: (1) visibility, (2) engagement, and (3) antidiscrimination. The first pillar, visibility, refers to the understanding that representation in certain databases can be beneficial to individuals and groups, such as in health care or welfare services. However, it also recognizes the right to privacy and the need to opt out of databases, such as those aggregated by commercial bodies [38]. As shown throughout our findings, many users indicated preferences in sharing their data with health care providers instead of family or friends [13] and preferred not to share with manufacturers, marketers, or other third parties [16]. The second pillar, digital engagement and disengagement, supports individual autonomy by encouraging personalized decisions regarding a user's preferred level of technological engagement and control over circumstances [38]. For our purposes, the right to digital disengagement would help support older adult SHT users in situations where personalized human care is their preferred option for certain purposes or where only select SHT functions were preferred. The third pillar, the right to challenge data-driven discrimination, allows for the ability to challenge bias in algorithmic decision-making and outcomes [38]. This last pillar may be particularly important as AI capabilities continue to develop alongside rising SHT popularity. Although issues with gender and racial bias with AI platforms are well documented within the literature, digital ageism is currently understudied and is thus in need of further critical analysis [28,36]. The ability to evaluate and challenge ageist bias is an important task as consumer-grade devices become more popular among aging populations. Taken together, regulatory frameworks following Taylor's [38] 3 pillars of data justice can be used to construct meaningful guidelines around how SHT data should be managed by private sector actors. This way, those who choose to engage with such technology in their homes can enjoy the potential health and safety benefits of SHTs while preventing or mitigating challenges to privacy, autonomy, and discrimination that can be detrimental to older age groups.

Strengths and Limitations

To the best of our knowledge, we are the first to research SHT user perception under the 4 themes of privacy, the purpose of

data collection, risk and benefits, and safety. Previous reviews have largely focused on rehabilitation or health care settings exclusively, whereas we have also incorporated user perception of commercial SHT surveillance. Our review engaged with interdisciplinary fields across the social sciences, computer sciences, engineering, legal studies, and nursing. We have also applied insights from the surveillance studies literature to findings from gerontology research. In terms of limitations, as we excluded studies that focused solely on nonusers, we may have missed potential insight into why individuals do not adopt SHTs. We also excluded studies on smartphones or wearable devices, due to their ability to be used outside the home, which may have further limited our findings. We did not include other search methods such as hand searching for references and did not reconduct the search after October 2021 both of which may have resulted in additional relevant studies. We did not conduct a quality appraisal of our included studies, resulting in another potential limitation. As many of these studies were written about users in global North countries, the extent to which these findings are representative of other regions requires further inquiry. Finally, only English language studies were reviewed, so relevant non-English papers may have been omitted.

Conclusions

In conclusion, through our review of 15 studies, we have demonstrated a variety of perceived benefits and drawbacks from research participants over the age of 50 years. Although SHTs are seen as beneficial for safety enhancement such as emergency contact and convenience purposes, many users are also concerned about the privacy and or security risks, such as a lack of knowledge over where their data were going or a lack of control over who had access. These findings add to the growing body of literature highlighting the need for more age-inclusive technology design. This becomes especially important as commercial-grade SHTs are increasingly positioned to be used for care or health-related purposes for aging populations. In tandem with age-inclusive efforts such as EbD approaches [29], we further encourage the use and development of technology that enhances home safety while respecting the need for user privacy and autonomy. To do so, we have recommended data justice [38] as an equitable approach to these issues through regulatory guidelines.

Future directions for research in this area include studies on how privacy regulators can better support adults aged 50 years and older who use SHT or wearable devices for health or safety purposes. Further work is also needed on how privacy settings can be made more easily accessible and flexible to support everyday users in various contexts. As mentioned, robust analysis is needed where there is a current gap in the literature pertaining to the link between older adults, targeted advertisements or personalized insurance pricing, and SHTs or wearables [28], both in the practical application of such commercial relationships and through user perception studies. For further insight on this topic, subsequent user perception research on SHTs in general should actively include participants over the age of 50 years, especially in the oldest age categories, as opposed to targeting younger populations exclusively. Finally, beyond privacy and security, user perception studies on related ethical issues such as AI discrimination and the potential impacts on user autonomy should be further explored.

Acknowledgments

This work was supported by the Social Sciences and Humanities Research Council (435-2021-0879). The authors thank the research assistant Alicia Natividad for the assistance with article screening and selection. They would also like to extend their thanks to Ani Orchanian-Cheff for assistance with database selection and keyword searches.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOCX File , 13 KB - mhealth_v12i1e48526_app1.docx]

Multimedia Appendix 2
PRIMSA Checklist.
[PDF File (Adobe PDF File), 112 KB - mhealth_v12i1e48526_app2.pdf]

References

1. Maalsen S, Sadowski J. The smart home on FIRE: amplifying and accelerating domestic surveillance. Surveill Soc 2019;17(1/2):118-124 [FREE Full text] [doi: 10.24908/ss.v17i1/2.12925]
2. Choi YK, Thompson HJ, Demiris G. Internet-of-things smart home technology to support aging-in-place: older adults' perceptions and attitudes. J Gerontol Nurs 2021;47(4):15-21. [doi: 10.3928/00989134-20210310-03] [Medline: 34038251]
3. Mortenson WB, Sixsmith A, Woolrych R. The power(s) of observation: theoretical perspectives on surveillance technologies and older people. Ageing Soc 2015;35(3):512-530 [FREE Full text] [doi: 10.1017/S0144686X13000846] [Medline: 29307944]

4. Pal D, Funilkul S, Charoenkitkarn N, Kanthamanon P. Internet-of-things and smart homes for elderly healthcare: an end user perspective. IEEE Access 2018;6:10483-10496 [FREE Full text] [doi: 10.1109/access.2018.2808472]
5. Wallace B, Knoefel F. Ethics/law/policy and the supportive smart home. AGE-WELL National Innovation Hub. 2022. URL: https://agewell-nih-appta.ca/wp-content/uploads/2022/06/SAM3-paper_0616-02-compressed.pdf [accessed 2024-01-10]
6. Zhu J, Shi K, Yang C, Niu Y, Zeng Y, Zhang N, et al. Ethical issues of smart home-based elderly care: a scoping review. J Nurs Manag 2022;30(8):3686-3699. [doi: 10.1111/jonm.13521] [Medline: 34806243]
7. Carver LF, Mackinnon D. Health applications of gerontechnology, privacy, and surveillance: a scoping review. Surveill Soc 2020;18(2):216-230 [FREE Full text] [doi: 10.24908/ss.v18i2.13240]
8. Zuboff S. The Age of Surveillance Capitalism. New York, NY: Public Affairs; 2019.
9. Edu JS, Such JM, Suarez-Tangil G. Smart home personal assistants: a security and privacy review. ACM Comput Surv 2020;53(6):1-36. [doi: 10.1145/3412383]
10. Zhang N, Mi X, Feng X, Wang X, Tian Y, Qian F. Understanding and mitigating the security risks of voice-controlled third-party skills on Amazon Alexa and Google Home. ArXiv. Preprint posted online on May 3 2018 2018 [FREE Full text]
11. Lau J, Zimmerman B, Schaub F. Alexa, are you listening?: Privacy perceptions, concerns and privacy-seeking behaviors with smart speakers. Proc ACM Hum-Comput Interact 2018;2(CSCW):1-31. [doi: 10.1145/3274371]
12. Malkin N, Deatrick J, Tong A, Wijesekera P, Egelman S, Wagner D. Privacy attitudes of smart speaker users. Proc Priv Enh Technol 2019;2019(4):250-271. [doi: 10.2478/popets-2019-0068]
13. Ghorayeb A, Comber R, Gooberman-Hill R. Older adults' perspectives of smart home technology: are we developing the technology that older people want? Int J Hum Comput 2021;147:102571. [doi: 10.1016/j.ijhcs.2020.102571]
14. Helbostad JL, Vereijken B, Becker C, Todd C, Taraldsen K, Pijnappels M, et al. Mobile health applications to promote active and healthy ageing. Sensors (Basel) 2017;17(3):622 [FREE Full text] [doi: 10.3390/s17030622] [Medline: 28335475]
15. Whittemore R, Knafl K. The integrative review: updated methodology. J Adv Nurs 2005;52(5):546-553. [doi: 10.1111/j.1365-2648.2005.03621.x] [Medline: 16268861]
16. Kirchbuchner F, Grosse-Puppendahl T, Hastall MR, Distler M, Kuijper A. Ambient intelligence from senior citizens' perspectives: understanding privacy concerns, technology acceptance, and expectations. In: De Ruyter B, Kameas A, Chatzimisios P, Mavrommati I, editors. Ambient Intelligence. AmI 2015. Lecture Notes in Computer Science, vol 9425. Cham, Switzerland: Springer; 2015:15.
17. Albina EM, Hernandez AA. Assessment of the elderly on perceived needs, benefits and barriers: inputs for the design of intelligent assistive technology. 2018 Presented at: 16th International Conference on ICT and Knowledge Engineering (ICT&KE); November 21-23, 2018; Bangkok, Thailand p. 1-10. [doi: 10.1109/ictke.2018.8612447]
18. Arthanat S, Chang H, Wilcox J. Determinants of information communication and smart home automation technology adoption for aging-in-place. J Enabling Technol 2020;14(2):73-86 [FREE Full text] [doi: 10.1108/jet-11-2019-0050] [Medline: 36196218]
19. Chung J, Bleich M, Wheeler DC, Winship JM, McDowell B, Baker D, et al. Attitudes and perceptions toward voice-operated smart speakers among low-income senior housing residents: comparison of pre- and post-installation surveys. Gerontol Geriatr Med 2021;7:23337214211005869 [FREE Full text] [doi: 10.1177/23337214211005869] [Medline: 35047655]
20. Jo TH, Ma JH, Cha SH. Elderly perception on the internet of things-based integrated smart-home system. Sensors (Basel) 2021;21(4):1284 [FREE Full text] [doi: 10.3390/s21041284] [Medline: 33670237]
21. Leikas J, Kulju M. Ethical consideration of home monitoring technology: a qualitative focus group study. Gerontechnology 2018;17(1):38-47 [FREE Full text] [doi: 10.4017/gt.2018.17.1.004.00]
22. O'Brien K, Liggett A, Ramirez-Zohfeld V, Sunkara P, Lindquist LA. Voice-controlled intelligent personal assistants to support aging in place. J Am Geriatr Soc 2020;68(1):176-179 [FREE Full text] [doi: 10.1111/jgs.16217] [Medline: 31617581]
23. Pal D, Funilkul S, Vanijja V, Papasratorn B. Analyzing the elderly users' adoption of smart-home services. IEEE Access 2018;6:51238-51252 [FREE Full text] [doi: 10.1109/access.2018.2869599]
24. Pal D, Papasratorn B, Chutimaskul W, Funilkul S. Embracing the smart-home revolution in Asia by the elderly: an end-user negative perception modeling. IEEE Access 2019;7:38535-38549 [FREE Full text] [doi: 10.1109/access.2019.2906346]
25. Psychoula I, Singh D, Chen L, Chen F, Holzinger A, Ning H. Users' privacy concerns in IoT based applications. In: IEEE Smartworld. 2018 Presented at: 2018 IEEE SmartWorld, Ubiquitous Intelligence and Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI); October 08-12, 2018; Guangzhou, China p. 1887-1894. [doi: 10.1109/smartworld.2018.00317]
26. Rogerson L, Burr J, Tyson S. The feasibility and acceptability of smart home technology using the Howz system for people with stroke. Disabil Rehabil Assist Technol 2020;15(2):148-152. [doi: 10.1080/17483107.2018.1541103] [Medline: 30663921]
27. van Hoof J, Kort HSM, Rutten PGS, Duijnstee MSH. Ageing-in-place with the use of ambient intelligence technology: perspectives of older users. Int J Med Inform 2011;80(5):310-331 [FREE Full text] [doi: 10.1016/j.ijmedinf.2011.02.010] [Medline: 21439898]

28. Percy Campbell J. Aging in place with Google and Amazon smart speakers: privacy and surveillance implications for older adults. University of Victoria. 2023. URL: http://hdl.handle.net/1828/15095 [accessed 2024-01-10]
29. Boger J. Culture change, human-centered design, and ethical by design as transactional cornerstone concepts in the development of technology for supporting aging. 2022 Presented at: Proceedings of the 15th International Conference on PErvasive Technologies Related to Assistive Environments; June 29-July 1, 2022; Corfu, Greece p. 556-561. [doi: 10.1145/3529190.3535692]
30. Iqbal U, Bahrami PN, Trimananda R, Cui H, Gamero-Garrido A, Dubois D, et al. Tracking, profiling, and ad targeting in the Alexa echo smart speaker ecosystem. 2023 Presented at: Proceedings of the 2023 ACM on Internet Measurement Conference; October 24-26, 2023; Montreal, QC p. 569-583. [doi: 10.1145/3618257.3624803]
31. Pridmore J, Zimmer M, Vitak J, Mols A, Trottier D, Kumar PC, et al. Intelligent personal assistants and the intercultural negotiations of dataveillance in platformed households. Surveill Soc 2019;17(1/2):125-131 [FREE Full text] [doi: 10.24908/ss.v17i1/2.12936]
32. Kröger J, Gellrich L, Pape S, Brause SR, Ullrich S. Personal information inference from voice recordings: user awareness and privacy concerns. Proc Priv Enh Technol 2022;2022(1):6-27 [FREE Full text] [doi: 10.2478/popets-2022-0002]
33. Turow J. The Voice Catchers: How Marketers Listen in to Exploit Your Feelings, Your Privacy, and Your Wallet. New Haven, NJ: Yale University Press; 2021.
34. Kröger J. Unexpected inferences from sensor data: a hidden privacy threat in the internet of things. In: Strous L, Cerf VG, editors. Internet of Things: Information Processing in an Increasingly Connected World. Cham, Switzerland: Springer International Publishing; 2019:147-159.
35. Kuempel A. The invisible middlemen: a critique and call for reform of the data broker industry. Northwest J Int Law Bus 2016;36(1):207 [FREE Full text]
36. Chu CH, Nyrup R, Leslie K, Shi J, Bianchi A, Lyn A, et al. Digital ageism: challenges and opportunities in artificial intelligence for older adults. Gerontologist 2022;62(7):947-955 [FREE Full text] [doi: 10.1093/geront/gnab167] [Medline: 35048111]
37. Sixsmith A. Ethical challenges in aging and technology. 2022 Presented at: Proceedings of the 15th International Conference on PErvasive Technologies Related to Assistive Environments; June 29-July 1, 2022; Corfu, Greece p. 552-555. [doi: 10.1145/3529190.3534756]
38. Taylor L. What is data justice? The case for connecting digital rights and freedoms globally. Big Data Soc 2017;4(2):1-14 [FREE Full text] [doi: 10.1177/2053951717736335]

Abbreviations

AI: artificial intelligence
EbD: ethical by-design
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
SHT: smart home technology

Edited by L Buis; submitted 26.04.23; peer-reviewed by Z Zandesh, S Nedjat, V Genaro Motti; comments to author 10.08.23; revised version received 02.10.23; accepted 15.12.23; published 09.02.24.

Please cite as:
Percy Campbell J, Buchan J, Chu CH, Bianchi A, Hoey J, Khan SS
User Perception of Smart Home Surveillance Among Adults Aged 50 Years and Older: Scoping Review
JMIR Mhealth Uhealth 2024;12:e48526
URL: https://mhealth.jmir.org/2024/1/e48526
doi:10.2196/48526
PMID:38335026

Review

Mobile Apps for COVID-19 Detection and Diagnosis for Future Pandemic Control: Multidimensional Systematic Review

Mehdi Gheisari[1,2], PhD; Mustafa Ghaderzadeh[3*], PhD; Huxiong Li[1*], PhD; Tania Taami[4], PhD; Christian Fernández-Campusano[5], PhD; Hamidreza Sadeghsalehi[6], PhD; Aaqif Afzaal Abbasi[7], PhD

[1]Institute of Artificial Intelligence, Shaoxing University, Shaoxing, China

[2]Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, India

[3]School of Nursing and Health Sciences of Boukan, Urmia University of Medical Sciences, Urmia, Iran

[4]Florida State University, Tallahassee, FL, United States

[5]Departamento de Ingeniería Eléctrica, Facultad de Ingeniería, Universidad de Santiago de Chile, Santiago, Chile

[6]Department of Neuroscience, Faculty of Advanced Technologies in Medicine, Tehran, Iran

[7]Department of Earth and Marine Sciences, University of Palermo, Palermo, Italy

[*]these authors contributed equally

Corresponding Author:
Mustafa Ghaderzadeh, PhD
School of Nursing and Health Sciences of Boukan
Urmia University of Medical Sciences
Kurdistan Blv Boukan
Urmia, 5951715161
Iran
Phone: 98 9129378390
Email: Mustaf.ghaderzadeh@sbmu.ac.ir

Related Article:

This is a corrected version. See correction statement: https://mhealth.jmir.org/2024/1/e58810

Abstract

Background: In the modern world, mobile apps are essential for human advancement, and pandemic control is no exception. The use of mobile apps and technology for the detection and diagnosis of COVID-19 has been the subject of numerous investigations, although no thorough analysis of COVID-19 pandemic prevention has been conducted using mobile apps, creating a gap.

Objective: With the intention of helping software companies and clinical researchers, this study provides comprehensive information regarding the different fields in which mobile apps were used to diagnose COVID-19 during the pandemic.

Methods: In this systematic review, 535 studies were found after searching 5 major research databases (ScienceDirect, Scopus, PubMed, Web of Science, and IEEE). Of these, only 42 (7.9%) studies concerned with diagnosing and detecting COVID-19 were chosen after applying inclusion and exclusion criteria using the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) protocol.

Results: Mobile apps were categorized into 6 areas based on the content of these 42 studies: contact tracing, data gathering, data visualization, artificial intelligence (AI)–based diagnosis, rule- and guideline-based diagnosis, and data transformation. Patients with COVID-19 were identified via mobile apps using a variety of clinical, geographic, demographic, radiological, serological, and laboratory data. Most studies concentrated on using AI methods to identify people who might have COVID-19. Additionally, symptoms, cough sounds, and radiological images were used more frequently compared to other data types. Deep learning techniques, such as convolutional neural networks, performed comparatively better in the processing of health care data than other types of AI techniques, which improved the diagnosis of COVID-19.

Conclusions: Mobile apps could soon play a significant role as a powerful tool for data collection, epidemic health data analysis, and the early identification of suspected cases. These technologies can work with the internet of things, cloud storage, 5th-generation

technology, and cloud computing. Processing pipelines can be moved to mobile device processing cores using new deep learning methods, such as lightweight neural networks. In the event of future pandemics, mobile apps will play a critical role in rapid diagnosis using various image data and clinical symptoms. Consequently, the rapid diagnosis of these diseases can improve the management of their effects and obtain excellent results in treating patients.

(JMIR Mhealth Uhealth 2024;12:e44406) doi:10.2196/44406

KEYWORDS

COVID-19; detection; diagnosis; internet of things; cloud computing; mobile applications; mobile app; mobile apps; artificial intelligence: AI; mobile phone; smartphone

Introduction

Following the widespread and rapid outbreak of COVID-19, the disease crossed geographical borders and had a devastating impact on the health, economy, and well-being of the worldwide population. According to the World Health Organization (WHO), high-severity COVID-19 was reported in 16%-21% of patients and almost 3% died. In the case of other variants, however, local statistics in many countries indicated a high mortality rate, with some studies estimating a mortality rate of 4% or higher. Due to the novelty of the disease, ways to deal with it were not known early on; still, researchers considered the screening and rapid diagnosis of patients and their separation from healthy people to be significant steps in fighting the disease [1,2].

Early and low-cost diagnosis of infections in any pandemic is essential for pandemic control. Therefore, if it is possible to diagnose and quarantine infected cases in the earliest phases of an outbreak, the outbreak can be managed in the epidemic phase and will not become a pandemic. In the COVID-19 pandemic, various diagnostic methods have been used, with the polymerase chain reaction (PCR) test being the primary diagnostic tool. Nevertheless, PCR is a time-consuming and costly method; until being diagnosed with this diagnostic test, patients might freely transmit COVID-19 and accelerate its conversion into a pandemic by increasing the transmission rate [3,4].

In the COVID-19 pandemic, a wide range of technologies came to aid in the faster diagnosis and screening of people with infection, many of which proved successful. Meanwhile, mobile phones, as highly abundant tools and an information gateway, helped people update their information and make more accurate decisions. In addition to being a platform for installing essential and valuable apps to detect people with infection, smartphones help track people and examine the keywords used by them for making a diagnosis [5-7].

Numerous technical and review studies have addressed smartphone apps for COVID-19 management. Many of these studies have dealt with various mobile apps used to estimate the prevalence, lessons, opportunities, and challenges of these devices and disease management. Still, none of them has adequately focused on detection and diagnosis. This study is the first to systematically review all the studies that have used smartphone technology to detect and diagnose COVID-19. Previous reviews have examined a limited number of studies. Herein, by covering the maximum number of databases, an effort was made to explore all published papers that used smartphones to diagnose COVID-19. In the area of mobile apps for the diagnosis of COVID-19, a thorough and complete study has not been conducted. Alnazi [8] examined apps related to COVID-19 released on Google Play. The 12 studies reviewed included mobile-based software for contact tracing, awareness building, appointment booking, and online consultation. The author only reviewed free apps, however, and although this study was conducted in 2021, it did not include many scientific or commercial apps [8].

Asadzadeh et al [9] determined the scope of mobile solutions in the COVID-19 pandemic and reviewed 16 mobile app studies on COVID-19–related data processing. The reviewed studies were classified into 4 categories: prevention, diagnosis, treatment, and protection. Despite noting a different range of mobile apps, this study did not mention the details and algorithms of these studies [9].

Aslani et al [10] studied mobile health apps in pandemics and epidemics. They examined 17 studies and explored common respiratory diseases and lung infections. Although this study was published during the COVID-19 pandemic, it did not mention mobile apps related to COVID-19 [10].

Kondylakis et al [11] examined 12 studies on mobile apps for COVID-19 data analysis in a more comprehensive investigation. These 12 studies covered the following domains: training, information sharing, risk assessment, self-management of symptoms, contact tracing, home monitoring, and decision-making. Still, this study did not include studies using machine learning (ML) methods to predict and diagnose COVID-19. It also did not deal with mobile apps for COVID-19 diagnosis and primarily focused on studies on COVID-19 education, care, and management using mobile apps [11].

Almalki et al [12] analyzed and discussed all the apps available on Google Play and the Apple Store and provided a brief explanation. Among its flaws, this study did not review scientific or academic studies and only examined mobile apps available in the market, many of which were developed without scientific or clinical supervision. Therefore, it is difficult to rely on these apps, as they lack scientific support and cannot be introduced to or proposed by the communities [12].

Table 1 lists some studies that have addressed mobile apps for COVID-19 data management, including the first author's name, the country, the main topic, and the number of studies covered.

Table 1. Related studies on mobile apps for detecting and diagnosing COVID-19.

Author and country	Main topic	Studies covered, n
Alanzi [8], Saudi Arabia	Mobile app used during COVID-19	12
Asadzadeh et al [9], Iran	Mobile health solutions	16
Aslani et al [10], Iran	Mobile health apps for epidemic and pandemic outbreaks	17
Kondylakis et al [11], Greece	Mobile app for COVID-19	12
Almalki et al [12], Saudi Arabia	Implemented an app to combat COVID-19	115

This study aimed to fill the gap left by previous reviews by conducting a comprehensive review of studies on smartphone apps for the diagnosis of COVID-19, providing solutions based on technological models, and answering research questions so that researchers and health systems can envision devices and their apps in preventing future pandemics.

Methods

Search Criteria

This systematic review, which was conducted for the first time using this method, aimed to identify relevant studies related to detecting and diagnosing COVID-19 using a variety of smartphone apps. The systematic search strategy was developed based on previous studies and the authors' knowledge. The main objective was to address the following analytical questions (AQs):

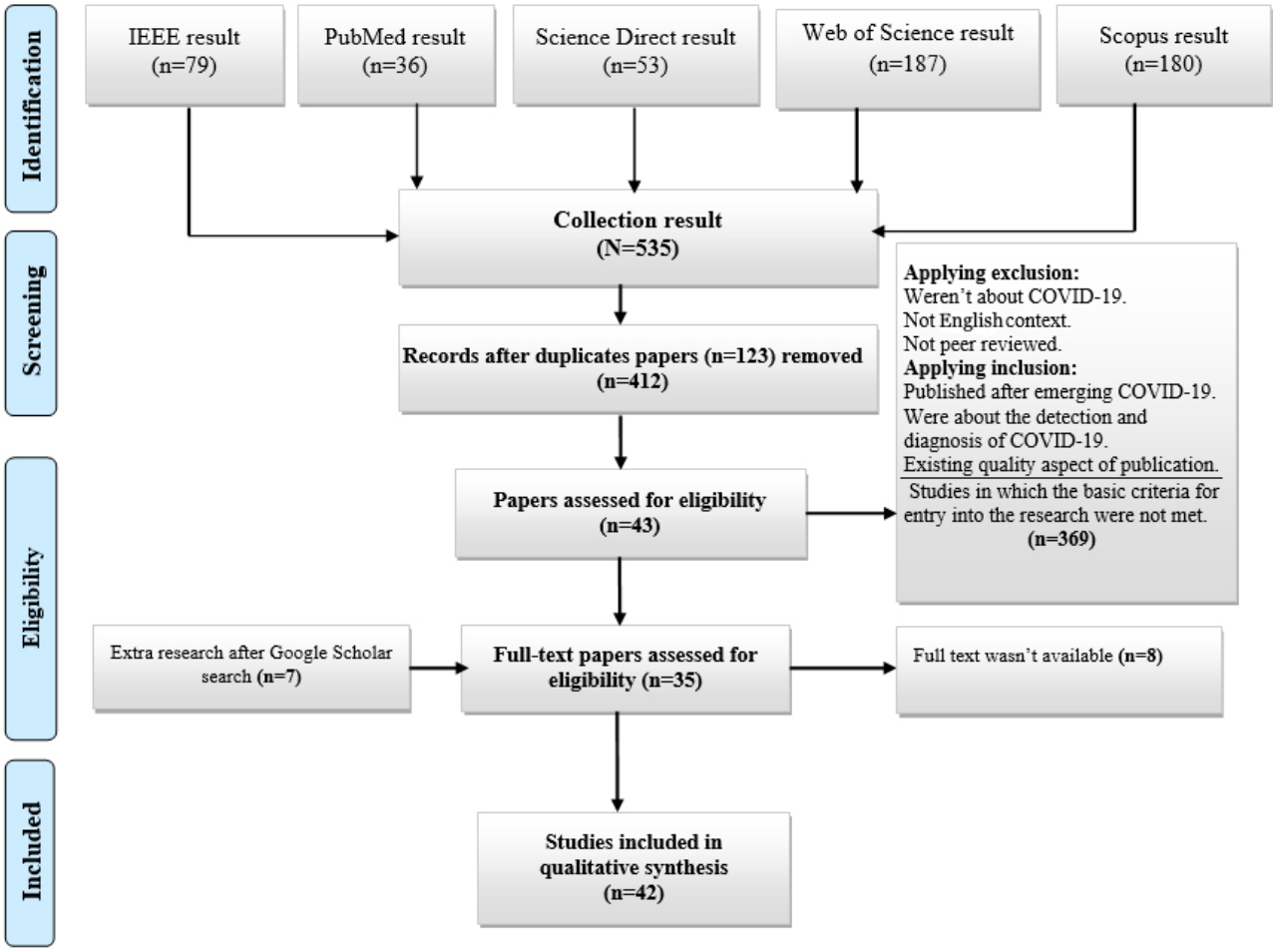
- AQ1: What are the uses of smartphones for COVID-19 detection and diagnosis?
- AQ2: What data do smartphones use to detect and diagnose COVID-19?
- AQ3: Which artificial intelligence (AI) methods and algorithms are used to process smartphone data?
- AQ4: How successful have smartphone apps been in COVID-19 detection and classification?
- AQ5: What suggestions can be made to improve the quality of mobile apps in disease diagnosis and pandemic control?

We reviewed electronic databases publishing papers on medicine and computer science. We concluded that PubMed, Web of Science (WoS), Scopus, IEEE, and ScienceDirect contain the most relevant papers. The search used the following keywords and logical expressions: ((“COVID-19”) AND (Detection OR Diagnosis) AND (Smartphone OR Mobile Application OR Mobile App)). The investigation was conducted from November 1, 2019, to late April 2022, and relevant published papers were extracted. The Embase database was eliminated from the examination due to the proximity of the publications.

Data Extraction

Relevant studies and the main elements of their methodology and results were recorded in data extraction forms in order to identify AI algorithms and techniques. Two researchers (authors AMR and MG) performed data extraction, and discrepancies between the researchers were resolved by discussion with an independent researcher (author AH). The extracted data elements included the first author's name, country of origin, research population, data used, purpose, method, the role of the mobile app, and the evaluation method. The search in reputable databases was performed based on the search strategy, and 535 papers were extracted. After reviewing the papers' abstracts and full texts, applying the inclusion and exclusion criteria, and selecting papers relevant to the title of this study, 42 (7.9%) full-text papers were finally selected. This process was performed based on the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart, as shown in Figure 1.

Figure 1. The review process and how to exclude papers according to the PRISMA flow diagram. PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

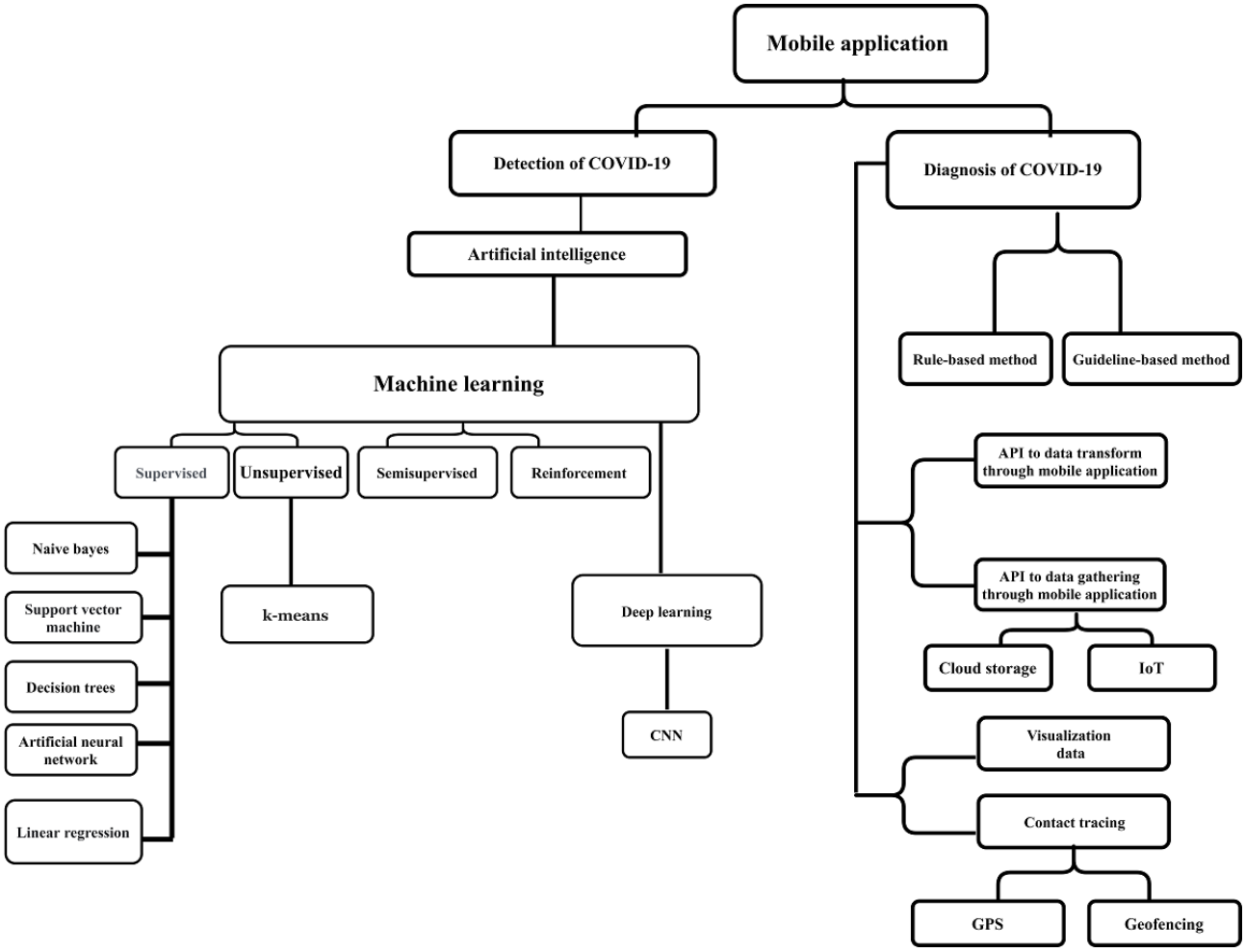


Results

Study Details

Due to the newly emergent status of COVID-19, the titles, abstracts, and keywords of all the papers published between 2020 and 2022 were reviewed, and 42 (7.9%) of 535 papers were ultimately selected as eligible. By comprehensively examining the mobile apps, we found that the role of smartphones was described in 6 areas with different types of COVID-19 data sets, including "Smartphones play the role of a platform for data collection," "visualizing the input data," "installing AI-based processing software," "determining contact tracing," and "COVID-19 data processing based on role-based and guideline-based methods," to detect and diagnose COVID-19. A taxonomy was developed to better organize the content and concepts related to mobile apps for COVID-19 diagnosis and detection, as shown in Figure 2.

Figure 2. Taxonomy of mobile applications for the COVID-19 pandemic. API: application programming interface; CNN: convolutional neural network; IoT: internet of things.



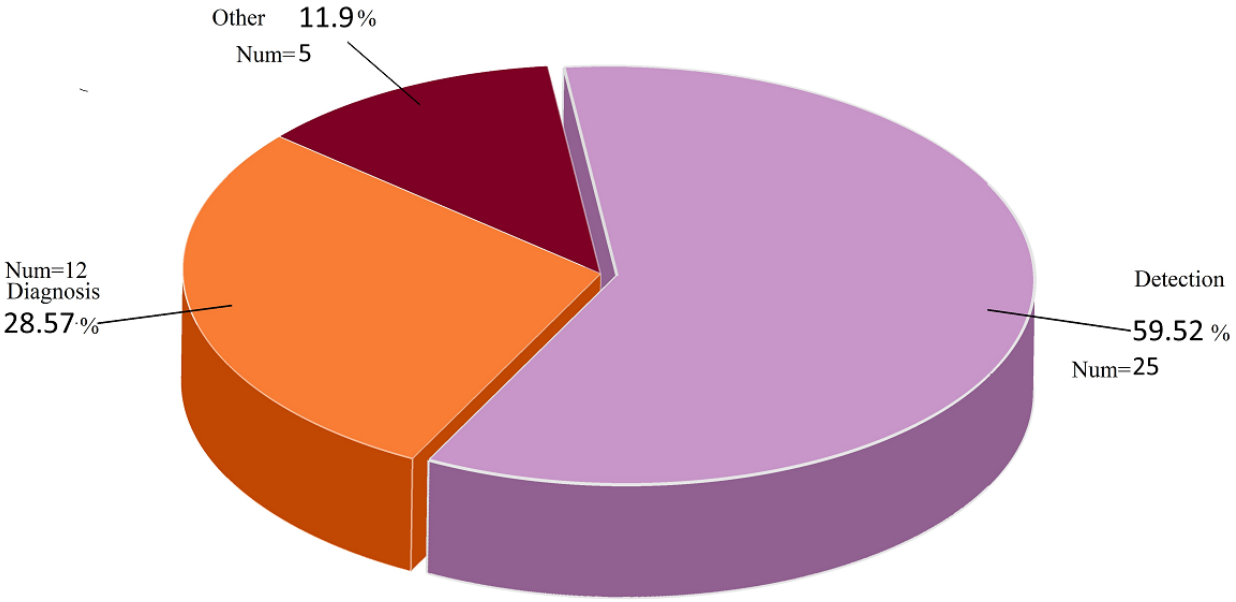
The taxonomy presented in this research classified the studies conducted in line with the investigation into 2 branches: detection and diagnosis. In the field of detection, the AI method was used more, which included traditional ML algorithms, such as supervised, unsupervised, semisupervised, and reinforcement learning; however, deep learning (DL) dealt with images and sounds. In the field of diagnosis, rule- and guideline-based techniques were mainly used to diagnose COVID-19 with mobile apps.

Detection vs Diagnosis of COVID-19 Using Mobile Apps

The terms "detection" and "diagnosis" were used interchangeably in many of the studies, but they differed in meaning and usage. These 2 terms are used differently in different settings. In clinical settings and diagnostic laboratories, the term "diagnosis" is used, while in computer vision and pattern recognition, observing the first definite signs in determining the status of a disease is called "detection." By examining the differences between these 2 terms in the dictionary and by inquiring from clinicians, a 2021 study found that detection identifies diseases in a set of patient and nonpatient cases. In detection, the disease is distinguished from other conditions, such that other cases may or may not have a disease-free status. However, the precise level and type of disease are completely specified during diagnosis. In the diagnosis concept, different cases might have a disease status or belong to other classes of abnormality or nonhealth, which can be determined [4,13-17].

Accordingly, detection was used for studies that distinguished cases of COVID-19 from healthy and normal cases. In contrast, diagnosis was used for studies that distinguished COVID-19 from other infectious pulmonary diseases (eg, different types of pneumonia). Detection makes sense in sets where other conditions (not infected with COVID-19) are specified, and COVID-19 can be distinguished with certainty from types of pneumonia or other coronaviruses [4,13,14]. By reviewing the 42 extracted papers, we found that 25 (59.5%) papers used DL to detect (identify) COVID-19, whereas 12 (28.6%) studies used it to diagnose and distinguish COVID-19 from other diseases. In addition, 3 (7.1%) studies did not precisely mention that diagnosis was their objective, while in 2 (4.8%) studies, the goal was to diagnose and detect. Figure 3 shows the amount of research performed to detect and diagnose COVID-19.

Figure 3. Aim of studies on the analysis of COVID-19 data in mobile applications.



Overview of the Role of Smartphone Apps in the Diagnosis of COVID-19

The apps were classified into the following 6 categories by reviewing all the studies conducted on mobile technologies and apps in line with our research objectives: contact tracing, data gathering, data visualization, AI-based diagnosis, rule- and guideline-based diagnosis, and data transformation.

Mobile Apps for Contact-Tracing Analysis

One of the main uses of apps during the COVID-19 pandemic was for contact tracing to diagnose and classify COVID-19. Thus, smartphones were used to diagnose patients with COVID-19 who had been in contact with people with infection. Many of the studies used mobile apps to trace people via trackers, geofencing, and GPS.

Mobile Apps for Gathering Data From Users

The sole purpose of some mobile apps was to collect data for processing. These apps received data through a standard smartphone from mobile devices as input devices in perception layers and sent them to the edge and cloud layers, where processing took place. In this case, the mobile apps sent data from the output of biosensors to higher-level systems for processing.

Mobile Apps for Data Visualization

Mobile apps were used to visualize the data received from biosensors or other data collection tools in many of the studies. After data were received, they were displayed to users via statistical charts. Smartphones carried out this task as the first step in data processing, which means gathering data and visualizing these data for COVID-19 diagnosis. In the next stage, these data were transferred to higher levels of processing, such as cloud and fog spaces.

Mobile Apps Used to Analyze COVID-19 Data Based on AI Algorithms

One of the most frequently used mobile apps for COVID-19 diagnosis was an AI-based application in which learning occurred. Studies used different clinical data, sounds, and radiology images. Data were collected in 2 ways: by mobile app designers and developers and by clinical app users using mobile apps.

Since most mobile apps processed and predicted COVID-19 infection, they incorporated different AI algorithms that involved 2 types of methods: methods based on traditional ML algorithms and those based on DL algorithms.

Mobile Apps Used to Analyze COVID-19 Data Based on Guidelines and Rules

A notable study on COVID-19 diagnosis using mobile phones involved ontologies, clinical guidelines, and rules. Due to COVID-19 diagnostic guidelines in health care centers, guideline- and rule-based methods for COVID-19 diagnosis are expected to be popular among clinicians and physicians. By incorporating these guidelines and rules into mobile apps, clinicians and stakeholders can see the process as more tangible and acceptable.

Mobile Apps as a Platform to Transform Messages and Data

Another application of mobile phones for COVID-19 diagnosis was processing keywords related to COVID-19 in social media using smartphones or analog messaging methods. In this method of COVID-19 diagnosis, the main terms representing COVID-19 diagnosis were counted and the COVID-19 diagnosis was made based on the number of uses of these words and their relationship with pronouns and sentence components.

Data Used in Mobile Apps for COVID-19 Detection and Diagnosis

A wide range of data were used in mobile apps, and according to specialty domains, different scientific disciplines made the initial diagnosis of COVID-19 differently. Mobile apps used a range of data for collection, classification, analysis, transfer, and visualization. These apps dealt with various data, including nasal swap samples, personal clinical data, signs and symptoms, voice and sounds, radiology images, and words and terms in media.

Several of the studies used a combination of data types. Some others used subjective clinical data and symptoms. Moreover, some apps used a variety of radiology images and patients' voices (cough).

Table 2 lists 11 (26.2%) studies on mobile apps for diagnosing COVID-19. This table includes the study, data, the application method, the mobile role, and the evaluation output.

Table 2. Studies (n=11) evaluating mobile apps used for COVID-19 diagnosis (or a goal similar to diagnosis).

Study	Data	Goal	Method	Mobile role	Evaluation output
Bindra et al [18]	Symptoms, clinical and bibliography data	COVID-19 risk prediction	ML[a]	Platform for the applied model to calculate risk prediction	N/A[b]
Sharma et al [19]	Sound	Diagnosis	ML	Platform for analysis; speech and sound analysis	N/A
Nema et al [20]	Symptoms	Diagnosis	Rule-based reasoning	Gathering symptoms and receiving alerts (SMS)	N/A
Quer et al [21]	Smartwatch, activity tracker data, symptoms, testing results	Differentiating COVID-19–positive status	Single decision threshold	Data collection	Accuracy=83.3%
Elagan et al [22]	Heart rate, blood cell counts, temperature	Diagnosis	Sending patient data to a physician and receiving output from the physician	Estimating the heart rate, receiving data from wireless sensors used to measure white blood cells (WBCs) and red blood cells (RBCs), and estimating air temperature	N/A
Imran et al [23]	Cough sounds	Diagnosis	DL[c]	Receiving cough sounds and analyzing them using a designed app	N/A
Mukhtar et al [24]	Cough, SpO_2[d], temperature	Diagnosis	Rule-based reasoning	Collecting data, sending data, showing the assessment	N/A
Koshti et al [25]	Symptoms	Diagnosis	ML	App platform	Accuracy=99%
Ertuğrul et al [26]	Personal data, observed symptoms (images, sounds)	Prediction	Ontology and rules	App platform	N/A
Maghded et al [27]	CT[e] scan, cough, voice, breath sounds, fatigue	Diagnosis	DL	App platform	N/A
Rangarajan and Ramachandran [28]	CT images	Diagnosis	DL	App platform	N/A

[a]ML: machine learning.

[b]N/A: not applicable.

[c]DL: deep learning.

[d]SpO_2: saturation of peripheral oxygen.

[e]CT: computed tomography.

There are several mobile apps available to detect COVID-19, as shown in Table 3. An overview of 31 (73.8%) studies is provided in the table, which includes the study, data, the application method, the mobile role, and evaluation results.

Table 3. Studies (n=31) that used mobile technologies to detect COVID-19.

Study	Data	Goal	Method	Mobile role	Evaluation output
Gökcen et al [29]	Cough	Detection	DL[a]	Platform for applying a COVID-19 detector via cough sounds	Accuracy=79%; F_1-score=80
Mao et al [30]	Wastewater sample	Detection	Biosensor analysis	Interface to send data	N/A[b]
Stasak et al [31]	Speech voice	Detection	ML[c]	App platform	Accuracy>82%-86%
Alkhodari and Khandoker [32]	Breath, cough, voice	Detection	DL	App platform	Accuracy=94.5% and 92.1%
Al-zubidi et al [33]	Blood index	Detection	ML	App platform	Accuracy=89%
Abdulrazaq Al-shekhly et al [34]	Thermal images and location	Detection	Thermometer and Send location	API[d] to calculate and send data	N/A
Berquedich et al [35]	Contact tracing	Detection and Management	Guideline based	App platform to prescribe drugs and send an alarm	N/A
Karataş et al [36]	Cough, voice	Detection	ML	App platform	Accuracy=96.5%
Awasthi et al [37]	Ultrasound images	Detection	DL	App platform	Accuracy=83%
Tawfik et al [38]	Cough sounds	Detection	ML and DL	App platform	Accuracy=98%
Krisnanik et al [39]	Symptoms	Detection	Rule-based reasoning	App platform	N/A
Ponomarchuk et al [40]	Breath and cough sounds	Detection	DL	App platform	N/A
Shreyas et al [41]	X-ray images	Detection	DL	App platform	Accuracy=98.4%
Mohsin et al [42]	Symptoms	Detection	Rule-based reasoning	App platform	N/A
Sanjeev et al [43]	Cough and clinical data (SpO_2[e] level, body temperature, heart rate, symptoms)	Detection	ML	App platform	Accuracy=85%
Ponomarchuk [40]	Breath and cough sounds	Detection	DL	App platform	N/A
Bushra et al [44]	X-ray images	Detection	DL	Platform to trace and analyze keywords	Accuracy=98.6%
Verde et al [45]	Cough	Detection	ML	Platform to analyze cough sounds	Accuracy=82%
Stanciu et al [46]	Bluetooth data	Virus Detection	Contact management	Contact tracing	N/A
Han et al [47]	Nasal swab sample	Virus Detection	Fluorescent aptasensors	Data visualization	N/A
Fozouni et al [48]	Nasal swab sample	Detection	RNA analysis	Data visualization	Sensitivity in less than 30 minutes
Coppock et al [49]	Audio and sound	Detection	ML	Platform to install an app	N/A
Wong et al [50]	Symptoms	Detection	Medical protocol	Receiving data and analysis using a designed app	N/A
Hijazi et al [51]	Heart rate, feeling features, blood pressure	Detection	ML	Collecting data from users	Mean accuracy= 83.3% (SD 1.6%)
Echeverría et al [52]	Sounds, symptoms	Early detection, management of close contacts	Guidelines	Gathering signs and symptoms	N/A
Verma et al [53]	CT[f] scan	Detection	DL	Process unit and platform for applied model	Accuracy=99.6%; F_1-score=99.6
Chen et al [54]	Spike protein, nucleocapsid protein	Detection	Data transfer	Receiving, gathering, and transmitting data in edge layers	N/A

Study	Data	Goal	Method	Mobile role	Evaluation output
Chen et al [55]	Temperature, heartbeat	Detection	Comparing received data with normal data	Receiving data from a sensor and sending it to the database	N/A
Wang et al [56]	Keywords on social media	Detecting the SARS-CoV-2 outbreak	Analysis of emerged keywords	Platform to install WeChat to trace and analyze keywords	N/A
Udhaya Sankar et al [57]	Speech, voice	Detection	Computational audit techniques	Data collection	N/A
Sun et al [58]	Horse nasal swab samples	Detection	ML	Smartphone-based collection and visualization of data	Results achieved in approx. 30 minutes

[a]DL: deep learning.

[b]N/A: not applicable.

[c]ML: machine learning.

[d]API: application programming interface.

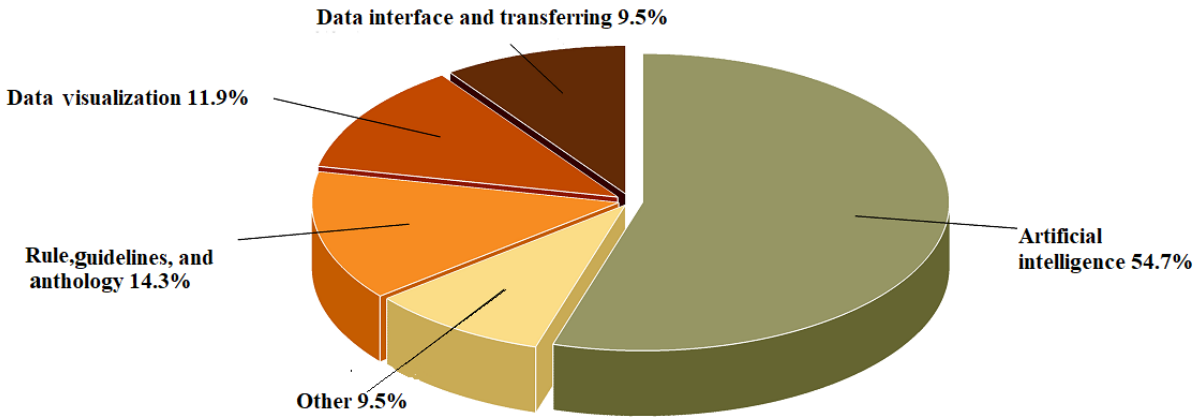
[e]SpO_2: saturation of peripheral oxygen.

[f]CT: computed tomography.

Having thoroughly scrutinized the particulars of the investigations delineated in Tables 2 and 3, the methodologies implemented therein, the scholarly community involved, and the significance of mobile apps in the detection of COVID-19, we were equipped to provide answers to the AQs.

To answer AQ1, although different apps were used, the main apps belonged to 6 functional domains: contact tracing, data gathering, data visualization, AI-based platform to analyze data and signals, rule- and guideline-based methods for decision-making, and data transformation. Contact tracing is a critical application and a strength of this technology for the timely diagnosis of COVID-19 due to its high availability, serving as a module for accessing GPS satellites, determining people's positions, and accessing higher data transaction layers, such as cloud and fog spaces. Berquedich et al [35] adopted a model for contact tracing and designed technology to reduce hospital visits and alert people who were in contact with patients to seek health care. Some domains, such as the processing of patients' voice and audio data for COVID-19 detection, received more attention, whereas biosensor data visualization apps were less frequently designed and presented. Figure 4 depicts the 6 functional dimensions of smartphones for COVID-19 diagnosis.

Figure 4. Areas of application of mobile technology for the diagnosis of COVID-19. AI: artificial intelligence.



As can be seen in Figure 4, AI-based methods formed a large part (about 54.7% of the methodologies at the core of smartphones to diagnose COVID-19). It is believed that the researchers' focus was on mobile phones as on-site processing tools for faster detection of COVID-19. However, the mobile apps that embedded guideline- and rule-based techniques in their processing core to analyze biomarkers are powerful tools for processing all kinds of numerical data and present them to users as a piece of point-of-care equipment. These aspects of mobile apps have been of great importance for researchers due to the ease of faster analysis of input data and the data input gates of mobile phones for more immediate identification of patients. More than 14% of the mobile apps analyzed all kinds of biomarkers using guidelines and rules to access a model in the field of COVID-19 diagnosis by classifying input data. A large number of mobile apps also focused on visualizing (11.9%) and transferring (9.5%) data.

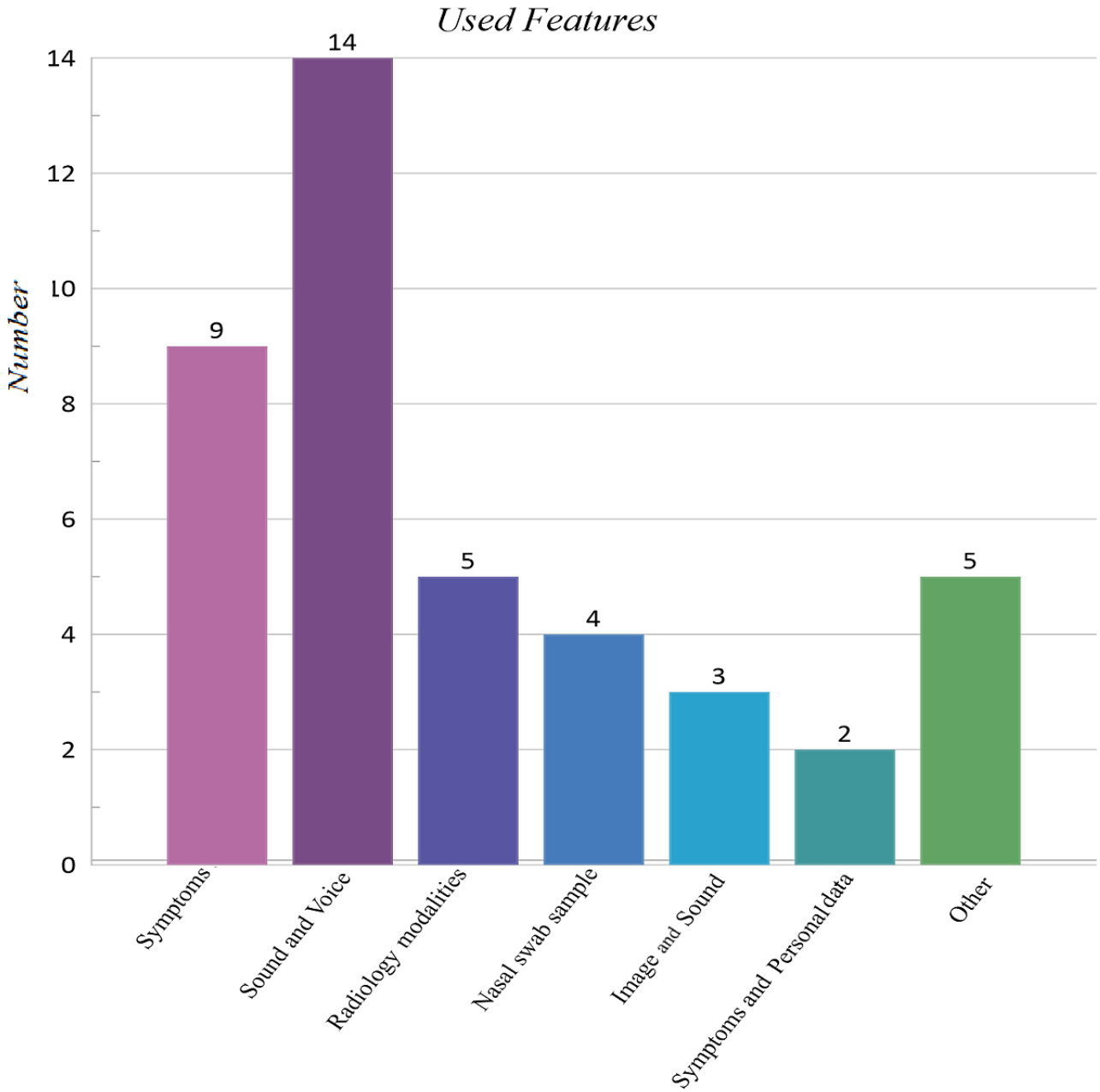
One of the strengths of mobile technology in identifying and diagnosing diseases, especially infectious diseases, is the visualization of the data received from internet of things (IoT) technologies and biosensors. After analyzing the data with

hardware chips and equipment or nucleic acid analysis, the signals are transferred to the mobile app for visualization through various means, such as IoT and Bluetooth. In these studies, blood, serology, and saliva data analysis results, after DNA and amino acid analysis, were sent by biosensors to mobile apps for visualization [27,28,47,52].

In some other studies, one of the most influential features of smartphone technology that had a significant impact on COVID-19 diagnosis and detection was the receiving and collection of data from people suspected of having COVID-19. These data included saliva samples, clinical signs and symptoms, vital signs, demographic information, cough sounds, and medical history. The data were obtained with mobile phones, the equipment connected to and embedded in them (eg, chemical sensors), and the recording of vital and clinical signs was stored in the mobile phones or moved to the cloud, allowing other computing devices to access this information. Some studies dealt with using smartphones for collecting patient data; for computation, they provided the data in a centralized space to specialists or decision support systems [21,22,25,26,32,53].

We observed many features to answer AQ2. The design and development of mobile apps for COVID-19 detection and diagnosis were commensurate with the frequency of COVID-19 diagnosis data. In other words, for frequent data, such as coughs, apps that used this type of data to diagnose COVID-19 were the most frequent. Figure 5 presents the usage of data types in mobile apps.

Figure 5. Data used in the detection and diagnosis of COVID-19.



Several laboratory biomarkers, diagnostic tests, biographical data, histories of diseases, and the voices and cough sounds of patients were used as input features for mobile apps based on methodology. According to Figure 5, it seems that the amount

of research using these features has a linear relationship with the diagnostic methods of COVID-19, as most mobile apps used the features of signs and symptoms as well as the patients' voices and cough sounds. In many apps designed to increase the accuracy of disease diagnosis, a combination of these features was used. Since mobile devices can receive and save sounds, these apps used this feature significantly.

For the analysis of nasal swab samples, apps that used biosensor technology incorporated multidisciplinary knowledge and used additional hardware. Following the sample analysis, the resulting data were transferred to the mobile apps via Wi-Fi or Bluetooth and were then used to automatically diagnose COVID-19. As a notable study, one can mention aptamers as a robust molecular tool for COVID-19 diagnosis [12]. The findings revealed that mobile technologies in pandemic control and prevention were a hot research topic and an exciting and trendy approach.

To answer AQ3, the notable use of smartphone apps in COVID-19 prevention assisted clinicians with the timely diagnosis of COVID-19. AI methods, such as ML and DL algorithms, achieved remarkable results in COVID-19 diagnosis. Many researchers [8,10,13,27,35] who used clinical data, such as serology data, vital signs, and symptoms, adopted ML algorithms. These features were received from users and applied to the ML model to diagnose COVID-19. Different ML algorithms were used for this purpose.

Since the main symptoms of COVID-19 diagnosis are cough sounds, the type of coughs, and respiratory sounds, many researchers adopted ML methods and algorithms, such as support vector machines, to identify patients' sound patterns [10,20,40,48]. Sound feature extraction was carried out based on mathematical algorithms. Using mobile apps to classify these features led to the automatic diagnosis of COVID-19.

Due to the complexity of detecting the pattern of cough sounds, several of the studies required more efficient and voluminous methods, so they used DL algorithms. After training with more significant data and obtaining a more efficient model on systems with more powerful processing units, such as graphics processing units (GPUs), these algorithms were applied to smartphone operating systems with the cough sounds of people suspected of having COVID-19. The methods in this category differentiated infected from noninfected cases by receiving patients' sounds with higher accuracy and precision [9,30]. Tawfik et al [38] used ML and DL methods simultaneously and achieved 98% accuracy.

Radiographic images offered another facility for rapid diagnosis of COVID-19 and, thus, pandemic control. DL methods and the convolutional neural network (CNN) algorithm were also used in studies using these images in mobile technologies. These models were optimally designed by training on radiology images using concepts such as transfer learning. At the point of care (PoC), they helped clinicians diagnose and detect new cases. These mobile apps demonstrated optimal performance and achieved an accuracy of >92%. By automatically extracting the features using manual feature selection, the bias in the results of diagnostic models was eliminated [18,29,45,46].

AI algorithms were used in various techniques in mobile phones, and the developers attempted to enhance their efficiency by adjusting the parameters of these algorithms [59-62]. Some of the studies used no performance evaluation index to determine the success rate of these apps (to answer AQ4). Regarding apps that used AI methods, the studies used the metrics in the confusion matrix. For different AI methods, different levels of accuracy were obtained. Upon evaluating the efficacy of ML and DL approaches in the identification and diagnosis of COVID-19, our analysis revealed that the accuracy of ML methods ranged from 83% to 99.5%, while DL algorithms exhibited accuracy rates within the range of 91% to 99.6%.

Studies that used all kinds of numerical variables as features, such as signs and symptoms, selected ML methods, while studies that used features such as sounds and images as recognition features, selected DL methods due to their high efficiency.

By calculating the approximate average accuracy of AI-based models in the detection and diagnosis of COVID-19, we concluded that ML algorithms used in mining COVID-19–related data achieved good results. In addition, in analyzing sound data, radiology, or their combination, DL methods achieved high accuracy; in many cases, this accuracy was reported to be >95%. As shown by the accuracy rates presented above, the findings demonstrate the acceptable results of AI methods. In addition, studies using apps to visualize RNA and DNA analysis tests detected cases of COVID-19 in a shorter time due to their sensitivity.

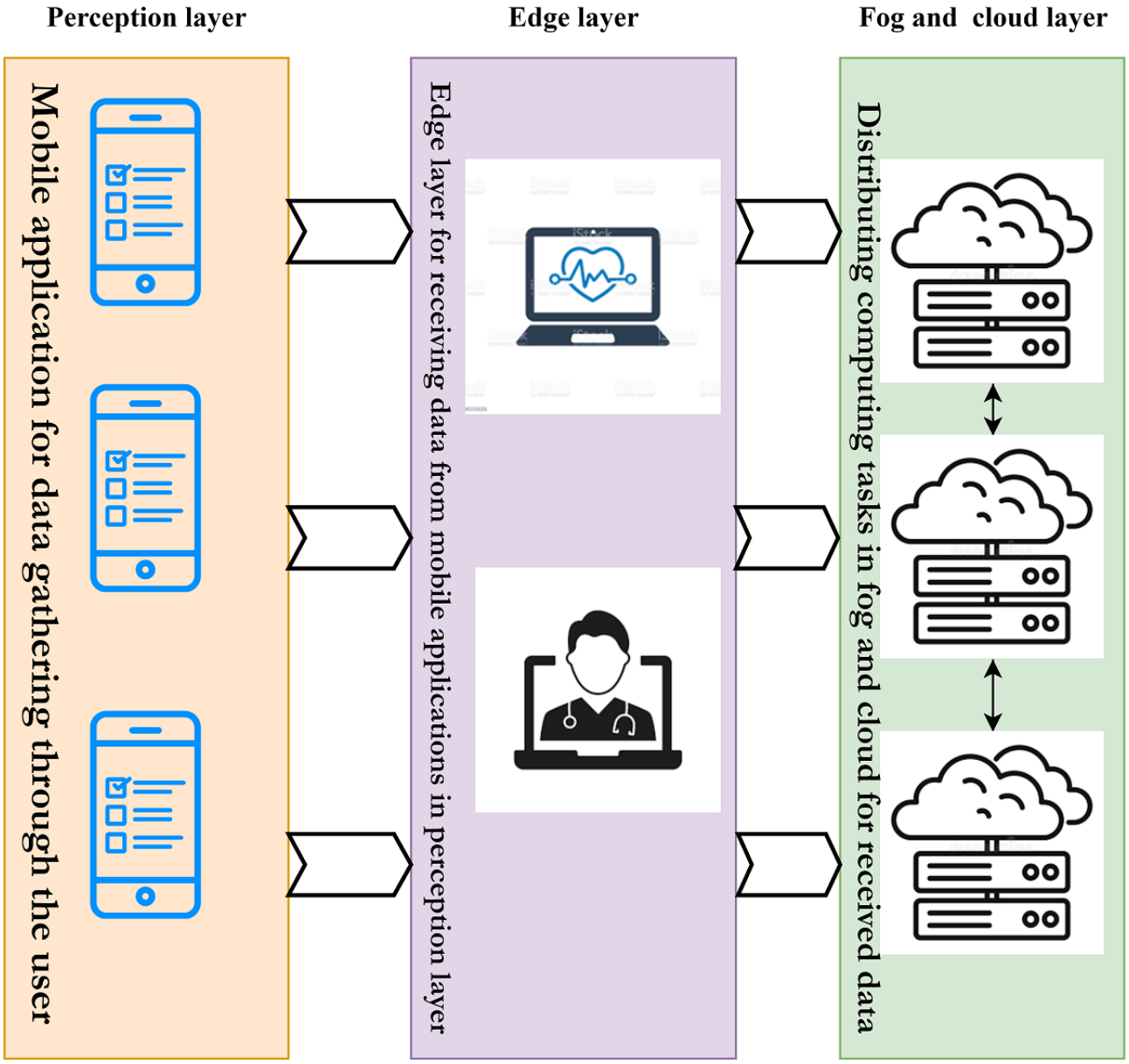
Discussion

Principal Findings

In this study, after reviewing all the research conducted on mobile apps for the diagnosis of COVID-19, several research questions that may have arisen for researchers and health care app development companies were answered. This section provides an answer to the remaining AQ by analyzing and evaluating the data entrance, technological method, and performance metrics of the methodologies outlined in the main tables.

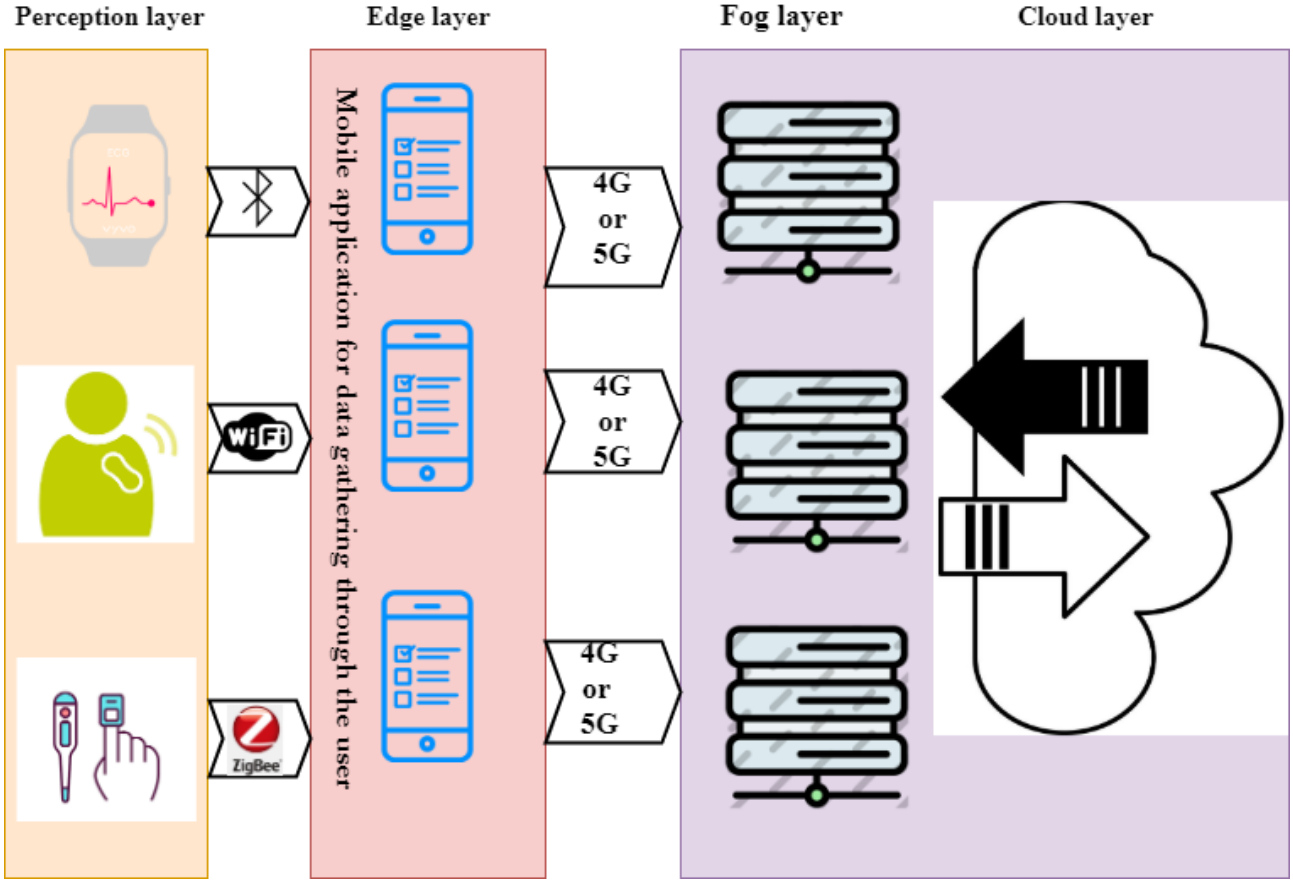
To address AQ5, we recommend using mobile apps for gathering data from PoCs obtained from individuals, monitoring devices, or clinical data generators. In modern techniques, edge layer devices are used in the health care devices layer. Therefore, it is strongly advised to incorporate higher-level technologies, such as fog and cloud spaces, for computing processes or computational units in distributed data clusters. Figure 6 illustrates a model of this mobile app for diagnosis.

Figure 6. Suggested model for mobile applications for gathering data from smartphones.



In the case of mobile apps that are used as data transmitters at the perception level, we suggest using 5th-generation (5G) instead of 4th-generation (4G) technology in centers and geographical areas that have access to this technology to transfer structured and unstructured data to a higher layer. Using this technology, and transferring health care data to the fog and cloud spaces, distributes computing and data processing. The model in Figure 7 is proposed for mobile apps with the purpose of data collection from sensors, biosensors, and monitoring equipment.

Figure 7. Suggested model for mobile applications for transmitting data from health sensors, biosensors, and monitoring devices. 5G: 5th generation.



We compared the 42 studies in terms of the requirements for AI methods. In using AI algorithms, the volume of the data set, overfitting prevention methods, and method lightness (for use in mobile phones) are the parameters for analyzing the quality of studies incorporating AI methodology in mobile phones. Since the included studies did not mention these cases, future studies should adopt methods to prevent overfitting when using ML methods. In ML algorithms, it is advised to use 3 methods: early stopping, dropout, and cross-validation. In addition to the mentioned methods, data augmentation techniques should be used in DL methods so that the resulting model is free of any overfitting.

Studies using DL algorithms in smartphones should carefully consider the following points:

- Data preprocessing plays a vital role in model convergence and speed [54]. It is recommended that future studies use data preprocessing techniques, especially for images.
- Model parameters must be optimized to develop a robust and valid model. Future research should adopt the concept of transfer learning and pretrained networks, and these networks should be customized to the investigation.
- Mobile phones have a limited processing unit. As a result, it is suggested that the models be designed using cutting-edge techniques. It is preferred to use lightweight pretrained CNN networks, such as MobileNet, MobileNetV2, and Efficient, to obtain lightweight models that efficiently run on smartphones.
- In studies using mobile phones for data collection and visualization, we recommend using cloud storage and 5G technologies that significantly contribute to the comprehensiveness of the data and image visualization.

Conclusion

Mobile technology, including various apps, can help with COVID-19 diagnosis and detection and play a vital role in controlling COVID-19 outbreak. Contact tracing can prevent additional contacts during an epidemic or pandemic outbreak of any disease. Thus, on the front lines of outbreak control, healthy people can be separated from people with infection and be alerted through their mobile phones. In the second step, mobile app technology, biosensors (for rapid diagnosis), and AI methods (for diagnosis in the early and acute stages of the disease) can reduce high mortality rates and minimize the consumption of hospital resources. In the third step, mobile technology as a powerful tool can help clinicians form repositories of clinical data and signs and symptoms and collect data from individual smartphones to create such repositories and big data. This can shed more light on COVID-19, its symptoms, the prognosis, and treatment outcomes. In this and future pandemics, smartphones and their apps can be an integral part of controlling the disease and improving patients' survival.

Acknowledgments

We thank the anonymous reviewers.

Authors' Contributions

M Ghaderzadeh was responsible for the search methodology, search strategy, and database search; M Gheisari and CFC, data visualization and data gathering; TT, AAA, and HS, data gathering and data organization; HS and M Ghaderzadeh, main table preparation, writing, and review; and MGZ, writing (review and editing), investigation, and supervision.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA Checklist.
[DOCX File , 36 KB - mhealth_v12i1e44406_app1.docx]

References

1. Sun J, He W, Wang L, Lai A, Ji X, Zhai X, et al. COVID-19: epidemiology, evolution, and cross-disciplinary perspectives. Trends Mol Med 2020 May;26(5):483-495 [FREE Full text] [doi: 10.1016/j.molmed.2020.02.008] [Medline: 32359479]
2. Baud D, Qi X, Nielsen-Saines K, Musso D, Pomar L, Favre G. Real estimates of mortality following COVID-19 infection. Lancet Infect Diseases 2020 Jul;20(7):773. [doi: 10.1016/s1473-3099(20)30195-x]
3. Tahamtan A, Ardebili A. Real-time RT-PCR in COVID-19 detection: issues affecting the results. Expert Rev Mol Diagn 2020 May 22;20(5):453-454 [FREE Full text] [doi: 10.1080/14737159.2020.1757437] [Medline: 32297805]
4. Ghaderzadeh M, Asadi F. Deep learning in the detection and diagnosis of COVID-19 using radiology modalities: a systematic review. J Healthc Eng 2021;2021:6677314 [FREE Full text] [doi: 10.1155/2021/6677314] [Medline: 33747419]
5. Ghaderzadeh M, Aria M, Asadi F. X-ray equipped with artificial intelligence: changing the COVID-19 diagnostic paradigm during the pandemic. Biomed Res Int 2021;2021:9942873 [FREE Full text] [doi: 10.1155/2021/9942873] [Medline: 34458373]
6. Kumar A, Gupta PK, Srivastava A. A review of modern technologies for tackling COVID-19 pandemic. Diabetes Metab Syndr 2020 Jul;14(4):569-573 [FREE Full text] [doi: 10.1016/j.dsx.2020.05.008] [Medline: 32413821]
7. Ghaderzadeh M, Eshraghi M, Asadi F, Hosseini A, Jafari R, Bashash D, et al. Efficient framework for detection of COVID-19 omicron and delta variants based on two intelligent phases of CNN models. Comput Math Methods Med 2022;2022:4838009 [FREE Full text] [doi: 10.1155/2022/4838009] [Medline: 35495884]
8. Alanzi TA. A review of mobile applications available in the app and Google Play stores used during the COVID-19 outbreak. J Multidiscip Healthc 2021 Jan;14:45-57. [doi: 10.2147/jmdh.s285014]
9. Asadzadeh A, Kalankesh L. A scope of mobile health solutions in COVID-19 pandemics. Inform Med Unlocked 2021;23:100558 [FREE Full text] [doi: 10.1016/j.imu.2021.100558] [Medline: 33842688]
10. Aslani N, Lazem M, Mahdavi S, Garavand A. A review of mobile health applications in epidemic and pandemic outbreaks: lessons learned for COVID-19. Arch Clin Infect Dis 2020 Jun 02;15(4):1-7 [FREE Full text] [doi: 10.5812/archcid.103649]
11. Kondylakis H, Katehakis DG, Kouroubali A, Logothetidis F, Triantafyllidis A, Kalamaras I, et al. COVID-19 mobile apps: a systematic review of the literature. J Med Internet Res 2020 Dec 09;22(12):e23170 [FREE Full text] [doi: 10.2196/23170] [Medline: 33197234]
12. Almalki M, Giannicchi A. Health apps for combating COVID-19: descriptive review and taxonomy. JMIR Mhealth Uhealth 2021 Mar 02;9(3):e24322 [FREE Full text] [doi: 10.2196/24322] [Medline: 33626017]
13. Rasekh Eslami Z, Zohoor S. Second language (L2) pragmatics and computer assisted language learning (CALL). Technol Assist Lang Educ 2023;1(2):1-17.
14. Kiaghadi M, Hoseinpour P. University admission process: a prescriptive analytics approach. Artif Intell Rev 2022 Aug 19;56(1):233-256. [doi: 10.1007/s10462-022-10171-y]
15. Borji A, Seifi A, Hejazi TH. An efficient method for detection of Alzheimer's disease using high-dimensional PET scan images. Intell Decis Technol 2023;17:729-749. [doi: 10.3233/IDT-22031]
16. Karimzadeh M, Vakanski A, Xian M, Zhang B. Post-Hoc Explainability of BI-RADS Descriptors in a Multi-Task Framework for Breast Cancer Detection and Segmentation. 2023 Presented at: 2023 IEEE 33rd Int Work Mach Learn Signal Process IEEE p. 1-6 URL: https://ieeexplore.ieee.org/document/10286006 [doi: 10.1109/MLSP55844.2023.10286006]
17. Mousavi A, Arefanjazi H, Sadeghi M, Ghahfarokhi AM, Beheshtinejad F, Masouleh M. Comparison of Feature Extraction with PCA and LTP Methods and Investigating the Effect of Dimensionality Reduction in the Bat Algorithm for Face Recognition. Int J Robot Control Syst 2023;3:501-509. [doi: 10.31763/ijrcs.v3i3.1057]
18. Bindra J, Ahlawat S, Javed M. Modified COVID-19 Indian and international dataset for automatic prediction of risk in an individual using machine learning models using a mobile APP. IJIEI 2021;9(2):142-749. [doi: 10.1504/IJIEI.2021.117057]

19. Sharma A, Baldi A, Kumar Sharma D. Int J Clin Pract 2021 Jun;75(6):e14134 [FREE Full text] [doi: 10.1111/ijcp.14134] [Medline: 33683774]
20. Nema BM, Makki Mohialden Y, Mahmood Hussien N, Ali Hussein N. COVID-19 knowledge-based system for diagnosis in Iraq using IoT environment. IJEECS 2021 Jan 01;21(1):328 [FREE Full text] [doi: 10.11591/ijeecs.v21.i1.pp328-337]
21. Quer G, Radin J, Gadaleta M, Baca-Motes K, Ariniello L, Ramos E, et al. Wearable sensor data and self-reported symptoms for COVID-19 detection. Nat Med 2021 Jan;27(1):73-77 [FREE Full text] [doi: 10.1038/s41591-020-1123-x] [Medline: 33122860]
22. Elagan S, Abdelwahab S, Zanaty E, Alkinani M, Alotaibi H, Zanaty M. Remote diagnostic and detection of coronavirus disease (COVID-19) system based on intelligent healthcare and internet of things. Results Phys 2021 Mar;22:103910 [FREE Full text] [doi: 10.1016/j.rinp.2021.103910] [Medline: 33614401]
23. Imran A, Posokhova I, Qureshi HN, Masood U, Riaz MS, Ali K, et al. AI4COVID-19: AI enabled preliminary diagnosis for COVID-19 from cough samples via an app. Inform Med Unlocked 2020;20:100378 [FREE Full text] [doi: 10.1016/j.imu.2020.100378] [Medline: 32839734]
24. Mukhtar H, Rubaiee S, Krichen M, Alroobaea R. An IoT framework for screening of COVID-19 using real-time data from wearable sensors. Int J Environ Res Public Health 2021 Apr 12;18(8):4022 [FREE Full text] [doi: 10.3390/ijerph18084022] [Medline: 33921223]
25. Koshti D, Kamoji S, Cheruthuruthy K, Shahi S, Mishra M. A detection, tracking and alerting system for COVID-19 using geo-fencing and machine learning. 2021 Presented at: ICICCS 2021: 5th International Conference on Intelligent Computing and Control Systems; May 6-8, 2021; Madurai, India p. 1499-1506 URL: http://tinyurl.com/mrcbep2m [doi: 10.1109/icicccs51141.2021.9432254]
26. Ertuğrul DC, Ulusoy DC. A knowledge-based self-pre-diagnosis system to predict COVID-19 in smartphone users using personal data and observed symptoms. Expert Syst 2022 Mar 21;39(3):e12716 [FREE Full text] [doi: 10.1111/exsy.12716] [Medline: 34177034]
27. Maghded HS, Ghafoor KZ, Sadiq AS, Curran K, Rawat DB, Rabie K. A novel AI-enabled framework to diagnose coronavirus COVID-19 using smartphone embedded sensors: design study. 2020 Presented at: IRI2020: IEEE 21st International Conference on Information Reuse and Integration for Data Science; August 11-13, 2020; Las Vegas, NV p. 180-187. [doi: 10.1109/iri49571.2020.00033]
28. Rangarajan A, Ramachandran H. A preliminary analysis of AI based smartphone application for diagnosis of COVID-19 using chest X-ray images. Expert Syst Appl 2021 Nov 30;183:115401 [FREE Full text] [doi: 10.1016/j.eswa.2021.115401] [Medline: 34149202]
29. Gökcen A, Karadag B, Riva C, Boyaci A. Artificial intelligence-based COVID-19 detection using cough records. Electrica 2021 May 29;21(2):203-208 [FREE Full text] [doi: 10.5152/electrica.2021.21005]
30. Mao K, Zhang H, Yang Z. An integrated biosensor system with mobile health and wastewater-based epidemiology (iBMW) for COVID-19 pandemic. Biosens Bioelectron 2020 Dec 01;169:112617 [FREE Full text] [doi: 10.1016/j.bios.2020.112617] [Medline: 32998066]
31. Stasak B, Huang Z, Razavi S, Joachim D, Epps J. Automatic detection of COVID-19 based on short-duration acoustic smartphone speech analysis. J Healthc Inform Res 2021 Mar 11;5(2):201-217 [FREE Full text] [doi: 10.1007/s41666-020-00090-4] [Medline: 33723525]
32. Alkhodari M, Khandoker A. Detection of COVID-19 in smartphone-based breathing recordings: a pre-screening deep learning tool. PLoS One 2022;17(1):e0262448 [FREE Full text] [doi: 10.1371/journal.pone.0262448] [Medline: 35025945]
33. Al-zubidi A, AL-Bakri NF, Hasoun RK, Hashim SH, Th.Salim Alrikabi H. Mobile application to detect COVID-19 pandemic by using classification techniques: proposed system. Int J Interact Mob Technol 2021 Aug 23;15(16):34. [doi: 10.3991/ijim.v15i16.24195]
34. Abdulrazaq Alshekhly MN, Zuhriyah H, Al-Zubaidi SS, Karim SA. Novel COVID-19 detection and diagnosis system using IoT based smart helmet. Int J Psychosoc Rehabil 2020 Mar;24(7):2296-2303. [doi: 10.37200/IJPR/V24I7/PR270221]
35. Berquedich M, Berquedich A, Kamach O, Masmoudi M, Chebbak A, Deshayes L. Developing a mobile COVID-19 prototype management application integrated with an electronic health record for effective management in hospitals. IEEE Eng Manag Rev 2020 Dec 1;48(4):55-64 [FREE Full text] [doi: 10.1109/emr.2020.3032943]
36. Karataş E, Adali B, Aydin, DalkiliÇ G. Mobile application that detects COVID-19 from cough and image using smartphone recordings and machine learning. 2021 Presented at: ASYU2021: Innovations in Intelligent Systems and Applications Conference; October 6-8, 2021; Turkey p. 1-6. [doi: 10.1109/asyu52992.2021.9598961]
37. Awasthi N, Dayal A, Cenkeramaddi LR, Yalavarthy PK. Mini-COVIDNet: efficient lightweight deep neural network for ultrasound based point-of-care detection of COVID-19. IEEE Trans Ultrason Ferroelect Freq Contr 2021 Jun;68(6):2023-2037. [doi: 10.1109/tuffc.2021.3068190]
38. Tawfik M, Nimbhore S, Al-Zidi N, Ahmed Z, Almadani A. Multi-features extraction for automating COVID-19 detection from cough sound using deep neural networks. 2022 Presented at: ICSSIT2022: 4th International Conference on Smart Systems and Inventive Technology; January 20-22, 2022; Tirunelveli, India p. 944-950. [doi: 10.1109/icssit53264.2022.9716529]

39. Krisnanik E, Imawangi N, Irmanda H. SIMTEKDIN of COVID-19 using forward chaining based on android mobile. 2021 Presented at: ICIMCIS2021: 3rd International Conference on Informatics, Multimedia, Cyber and Information System; October 28-29, 2021; Jakarta, Indonesia p. 207-212. [doi: 10.1109/icimcis53775.2021.9699351]
40. Ponomarchuk A, Burenko I, Malkin E, Nazarov I, Kokh V, Avetisian M, et al. Project Achoo: a practical model and application for COVID-19 detection from recordings of breath, voice, and cough. IEEE J Sel Top Signal Process 2022 Feb;16(2):175-187 [FREE Full text] [doi: 10.1109/jstsp.2022.3142514]
41. Shreyas S, Rao J. Diagnostic decision support for medical imaging and COVID-19 image classification on ARM Mali GPU. 2021 Presented at: 2021 IEEE Globecom Workshops (GC Wkshps); December 7-11, 2021; Madrid, Spain URL: http://tinyurl.com/22suu52e [doi: 10.1109/gcwkshps52748.2021.9682104]
42. Mohsin J, Saleh F, Ali AA. Real-time surveillance system to detect and analyzers the suspects of COVID-19 patients by using IoT under edge computing techniques (RS-SYS). 2020 Presented at: NICST2020: 2nd Al-Noor International Conference for Science and Technology; August 28-30, 2020; Baku, Azerbaijan p. 68 URL: http://tinyurl.com/2zuuxj6v [doi: 10.1109/nicst50904.2020.9280305]
43. Sanjeev S, Balne C, Reddy T, Reddy G. Deep learning-based mixed data approach for COVID-19 detection. 2021 Presented at: INDICON2021: IEEE 18th India Council International Conference; December 19-21, 2021; Guwahati, India URL: http://tinyurl.com/27cks7pt [doi: 10.35940/ijitee.a5063.129219]
44. Bushra K, Ahamed M, Ahmad M. Automated detection of COVID-19 from X-ray images using CNN and Android mobile. Res Biomed Eng 2021 Aug 12;37(3):545-552 [FREE Full text] [doi: 10.1007/s42600-021-00163-2]
45. Verde L, De Pietro G, Sannino G. Artificial intelligence techniques for the non-invasive detection of COVID-19 through the analysis of voice signals. Arab J Sci Eng 2021 Oct 08;48(8):1-11 [FREE Full text] [doi: 10.1007/s13369-021-06041-4] [Medline: 34642613]
46. Stanciu S, Gurzun M, Iancu D, Gheorghita V, Lupascu C, Gugulea G, et al. Innovative solution for SARS-COV2 epidemic management. Rom J Mil Med 2021 Aug 01;124(3):334-338. [doi: 10.55453/rjmm.2021.124.3.10]
47. Han C, Li Q, Ji H, Xing W, Zhang L, Zhang L. Aptamers: the powerful molecular tools for virus detection. Chem Asian J 2021 Jun 01;16(11):1298-1306. [doi: 10.1002/asia.202100242] [Medline: 33851522]
48. Fozouni P, Son S, Díaz de León Derby M, Knott GJ, Gray CN, D'Ambrosio MV, et al. Amplification-free detection of SARS-CoV-2 with CRISPR-Cas13a and mobile phone microscopy. Cell 2021 Jan 21;184(2):323-333.e9 [FREE Full text] [doi: 10.1016/j.cell.2020.12.001] [Medline: 33306959]
49. Coppock H, Jones L, Kiskin I, Schuller B. COVID-19 detection from audio: seven grains of salt. Lancet Digital Health 2021 Sep;3(9):e537-e538 [FREE Full text] [doi: 10.1016/s2589-7500(21)00141-2]
50. Wong C, Ho D, Tam A, Zhou M, Lau Y, Tang M, et al. Artificial intelligence mobile health platform for early detection of COVID-19 in quarantine subjects using a wearable biosensor: protocol for a randomised controlled trial. BMJ Open 2020 Jul 22;10(7):e038555 [FREE Full text] [doi: 10.1136/bmjopen-2020-038555] [Medline: 32699167]
51. Hijazi H, Abu Talib M, Hasasneh A, Bou Nassif A, Ahmed N, Nasir Q. Wearable devices, smartphones, and interpretable artificial intelligence in combating COVID-19. Sensors (Basel) 2021 Dec 17;21(24):8424 [FREE Full text] [doi: 10.3390/s21248424] [Medline: 34960517]
52. Echeverría P, Puig J, Ruiz J, Herms J, Sarquella M, Clotet B, et al. Remote health monitoring in the workplace for early detection of COVID-19 cases during the COVID-19 pandemic using a mobile health application. Int J Environ Res Public Health 2021 Dec 24;19(1):167 [FREE Full text] [doi: 10.3390/ijerph19010167] [Medline: 35010426]
53. Verma A, Amin S, Naeem M, Saha M. Detecting COVID-19 from chest computed tomography scans using AI-driven android application. Comput Biol Med 2022 Apr;143:105298 [FREE Full text] [doi: 10.1016/j.compbiomed.2022.105298] [Medline: 35220076]
54. Guo J, Chen S, Tian S, Liu K, Ni J, Zhao M, et al. 5G-enabled ultra-sensitive fluorescence sensor for proactive prognosis of COVID-19. Biosens Bioelectron 2021 Jun 01;181:113160 [FREE Full text] [doi: 10.1016/j.bios.2021.113160] [Medline: 33740542]
55. Chen Z, Khan S, Abbas M, Nazir S, Ullah K. Enhancing healthcare through detection and prevention of COVID-19 using internet of things and mobile application. Mob Inf Syst 2021 Nov 16;2021:1-11 [FREE Full text] [doi: 10.1155/2021/5291685]
56. Wang W, Wang Y, Zhang X, Jia X, Li Y, Dang S. Using WeChat, a Chinese social media app, for early detection of the COVID-19 outbreak in December 2019: retrospective study. JMIR Mhealth Uhealth 2020 Oct 05;8(10):e19589 [FREE Full text] [doi: 10.2196/19589] [Medline: 32931439]
57. Udhaya Sankar SM, Ganesan R, Katiravan J, Ramakrishnan M, Kouser R. Mobile application based speech and voice analysis for COVID-19 detection using computational audit techniques. IJPCC 2020 Oct 19;18(5):508-517. [doi: 10.1108/ijpcc-09-2020-0150]
58. Sun F, Ganguli A, Nguyen J, Brisbin R, Shanmugam K, Hirschberg D, et al. Smartphone-based multiplex 30-minute nucleic acid test of live virus from nasal swab extract. Lab Chip 2020 May 05;20(9):1621-1627 [FREE Full text] [doi: 10.1039/d0lc00304b] [Medline: 32334422]
59. Ghaderzadeh M, Hosseini A, Asadi F, Abolghasemi H, Bashash D, Roshanpoor A. Automated detection model in classification of B-lymphoblast cells from normal B-lymphoid precursors in blood smear microscopic images based on the majority voting technique. Sci Program 2022 Jan 4;2022:1-8. [doi: 10.1155/2022/4801671]

60. Ghaderzadeh M, Asadi F, Ramezan Ghorbani N, Almasi S, Taami T. Toward artificial intelligence (AI) applications in the determination of COVID-19 infection severity: considering AI as a disease control strategy in future pandemics. Iran J Blood Cancer 2023;15(3):93-111.
61. Fasihfar Z, Rokhsati H, Sadeghsalehi H, Ghaderzadeh M, Gheisari M. AI-driven malaria diagnosis: developing a robust model for accurate detection and classification of malaria parasites. Iran J Blood Cancer 2023;15(3):112-124.
62. Arefanjazi H, Ataei M, Ekramian M, Montazeri A. A robust distributed observer design for Lipschitz nonlinear systems with time-varying switching topology. J Franklin Inst 2023;360(14):10728-10744.

Abbreviations

5G: 5th generation
AI: artificial intelligence
API: application programming interface
AQ: analytical question
CNN: convolutional neural network
CT: computed tomography
IoT: internet of things
DL: deep learning
ML: machine learning
PCR: polymerase chain reaction
PoC: point of care
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
SpO2: saturation of peripheral oxygen

Edited by L Buis; submitted 18.11.22; peer-reviewed by S Paydar, N Jiwani, A Hosseini; comments to author 15.12.22; revised version received 02.01.23; accepted 18.08.23; published 22.02.24.

Please cite as:
Gheisari M, Ghaderzadeh M, Li H, Taami T, Fernández-Campusano C, Sadeghsalehi H, Afzaal Abbasi A
Mobile Apps for COVID-19 Detection and Diagnosis for Future Pandemic Control: Multidimensional Systematic Review
JMIR Mhealth Uhealth 2024;12:e44406
URL: https://mhealth.jmir.org/2024/1/e44406
doi:10.2196/44406
PMID:38231538

Review

Lessons and Untapped Potential of Smartphone-Based Physical Activity Interventions for Mental Health: Narrative Review

Emily E Bernstein[1,2], PhD; Emma C Wolfe[3], BA; Brynn M Huguenel[1,2], PhD; Sabine Wilhelm[1,2], PhD

[1]Department of Psychiatry, Massachusetts General Hospital, Boston, MA, United States

[2]Department of Psychiatry, Harvard Medical School, Boston, MA, United States

[3]Department of Psychology, University of Virginia, Charlottesville, VA, United States

Corresponding Author:
Emma C Wolfe, BA
Department of Psychology
University of Virginia
Gilmer Hall
485 McCormick Road
Charlottesville, VA, 22903
United States
Phone: 1 3014665679
Email: rpu3zk@virginia.edu

Abstract

Background: Physical activity has well-known and broad health benefits, including antidepressive and anxiolytic effects. However, only approximately half of Americans meet even the minimum exercise recommendations. Individuals with anxiety, depression, or related conditions are even less likely to do so. With the advent of mobile sensors and phones, experts have quickly noted the utility of technology for the enhanced measurement of and intervention for physical activity. In addition to being more accessible than in-person approaches, technology-driven interventions may uniquely engage key mechanisms of behavior change such as self-awareness.

Objective: This study aims to provide a narrative overview and specific recommendations for future research on smartphone-based physical activity interventions for psychological disorders or concerns.

Methods: In this paper, we summarized early efforts to adapt and test smartphone-based or smartphone-supported physical activity interventions for mental health. The included articles described or reported smartphone-delivered or smartphone-supported interventions intended to increase physical activity or reduce sedentary behavior and included an emotional disorder, concern, or symptom as an outcome measure. We attempted to extract details regarding the intervention designs, trial designs, study populations, outcome measures, and inclusion of adaptations specifically for mental health. In taking a narrative lens, we drew attention to the type of work that has been done and used these exemplars to discuss key directions to build on.

Results: To date, most studies have examined mental health outcomes as secondary or exploratory variables largely in the context of managing medical concerns (eg, cancer and diabetes). Few trials have recruited psychiatric populations or explicitly aimed to target psychiatric concerns. Consequently, although there are encouraging signals that smartphone-based physical activity interventions could be feasible, acceptable, and efficacious for individuals with mental illnesses, this remains an underexplored area.

Conclusions: Promising avenues for tailoring validated smartphone-based interventions include adding psychoeducation (eg, the relationship between depression, physical activity, and inactivity), offering psychosocial treatment in parallel (eg, cognitive restructuring), and adding personalized coaching. To conclude, we offer specific recommendations for future research, treatment development, and implementation in this area, which remains open and promising for flexible, highly scalable support.

(JMIR Mhealth Uhealth 2024;12:e45860) doi:10.2196/45860

KEYWORDS

smartphone; digital health; exercise; physical activity; mental health; depression; anxiety; mobile phone

Introduction

Background

In the 21st century, anxiety and depression have been among the top 25 causes of global disease burden [1]. The COVID-19 pandemic has only intensified the rising prevalence as well as the personal and societal impacts of these disorders [2,3]. As there are simply not enough mental health professionals to meet this need [4,5], alternative interventions—both preventive and curative—are urgently needed. Targeting physical activity is a clear opportunity. Before the pandemic, more than 1 in 4 adults reported sitting for >8 hours per day [6], and this number rose to >40% during the pandemic [7]. Both of these statistics are likely underestimated [8]. Prolonged sedentary behavior, or extended time spent awake with minimal energy expenditure [9], is associated with more severe anxiety and depression as well as higher odds of developing related disorders [10-14]. In contrast, decades of research have demonstrated that regular physical activity is associated with numerous positive psychological outcomes [15-17]. Cross-sectionally, individuals who engage in regular exercise—a subset of physical activity that involves planned, structured, and repetitive bodily movement intended to improve or maintain fitness [18]—report fewer and less severe symptoms of anxiety and depression [19,20], greater positive affect and well-being [21], less stress [22], and lower rates of anxiety and depressive disorder diagnoses [23-25]. At the individual level, people report feeling better on days when they exercise [26-28].

Exercise as an Intervention for Depression and Anxiety

Prospective data support regular exercise as a potent population-level *prevention* tool, significantly lowering the risk of developing anxiety and depressive disorders [25,29,30]. Encouragingly, even small amounts of physical activity may have an enormous impact on mental health [31,32]. For example, an estimated 12% of new cases of depression could be prevented if the entire population exercised for just 1 hour per week [33]. In addition, among individuals presenting with diagnosable levels of symptomatology, systematically increasing exercise behavior is therapeutic [33-36].

Exercise appears to enhance emotional flexibility, or a person's ability to self-regulate under stress [37-40]. Physiologically, individuals who exercise more regularly show faster heart rate recovery following stressors than their peers who exercise less, and individual bouts of exercise can mitigate the hypothalamic-pituitary-adrenal axis, heart rate, and blood pressure reactivity to acute stress [39,41-43]. Physical activity can also increase the production of brain-derived neurotrophic factors, which are neurobiological changes that are understood to increase resilience [44]. These effects are mirrored in reports of exercise bolstering emotional recovery following stressors, enhancing coping self-efficacy, and mitigating the impact of rumination and other emotion regulation deficits on prolonging distress [34,39,42,45,46]. Furthermore, exercise benefits physical health targets that share bidirectional relationships with mental health, such as better sleep and cardiometabolic health [47-51]. Critically, positive treatment effects have been found for directly alleviating anxiety [52-55] and depressive [56-58] disorders as well as related and frequently comorbid conditions such as posttraumatic stress disorder [59] and obsessive-compulsive disorder [60,61]. Importantly, research has replicated the benefits of physical activity (ie, reducing psychiatric symptoms) in samples of people with severe mental illness, such as schizophrenia [62]. Similarly, physical activity and other health-related behaviors (eg, sleep hygiene) are considered to be an integral component of treatment for bipolar disorder [63]. Exercise has also been successfully used to augment the effects of other validated psychosocial treatments such as cognitive behavioral therapy (CBT) [64-67].

Despite the broad knowledge that regular physical activity is physically, cognitively, and emotionally beneficial, only approximately half of Americans meet even the minimum exercise recommendation of 150 minutes per week of moderate-intensity or equivalent physical activity [68]. Individuals with anxiety, depression, or related conditions are even less likely to do so [69,70]. They are also more likely than peers without mental health disorders to exhibit elevated sedentary behavior [71,72]. Thus, although acceptable and efficacious tools exist to help individuals meaningfully change their behavior and improve psychiatric symptoms, there is a large gap between the evidence and real-world implementation. Few clinicians include physical activity as an explicit treatment target or use it as an intervention tool [73-75]. Moreover, the larger barriers to treatment within our health care system remain, including the inaccessibility of treatment due to the acute shortage of qualified clinicians; stigma; and patients' difficulty with travel, timing, and the cost of regular appointments [76].

Promise of Digital Platforms for Promoting Physical Activity

With the advent of mobile sensors and phones, experts have quickly noted the ability of technology to expand the reach of evidence-based psychiatric care; overcome the aforementioned barriers by providing treatment flexibly; and begin reducing long-standing disparities in treatment access, response, and dropout [77-79]. This could also be an efficient, scalable method of promoting increased physical activity among adults with or at risk of anxiety and depressive disorders.

Digital solutions show strong early benefits for activity measurement and intervention in nonpsychiatric populations. In fact, leveraging technology to measure and increase physical activity was an official recommendation from the National Heart, Lung, and Blood Institute and National Institute on Aging "Influences on sedentary behavior/Interventions to reduce sedentary behavior" joint workshop [80]. First, measuring behavior via mobile sensor is validated and widely used [81-84]. Similarly, people's tendency to carry their phones with them throughout the day allows for more accurate monitoring of physical activity and related progress. The ease of use and, therefore, precision of such technologies (wearable and smartphone-based sensors) is an important boon for research and treatment as self-report measures of activity typically result in underreporting [81-84].

Second, delivering treatment in part or fully via mobile phones is effective in increasing physical activity in nonpsychiatric populations [85-87]. This parallels broader findings that apps

can effectively promote other health behaviors (eg, improved nutrition, smoking cessation, and medication adherence) [88,89]. Currently, 97% of Americans have a mobile phone, and an estimated 85% have a smartphone [90]. Although these numbers are lower in certain populations, such as those with serious mental illness (wherein an estimated 85% own a mobile phone and 60%-70% own a smartphone), the ubiquity of smartphones allows for the promotion of behavior change in real time and with a wider array of individuals [91-94]. Inactivity frequently occurs out of conscious awareness or choice due to people's attention being fixated elsewhere (eg, watching television or taking the elevator at work). As such, personal devices can unobtrusively enhance awareness of behavior, which itself can promote increased activity [95]. Furthermore, technologies can deliver notifications in the moment to interrupt passive episodes while also providing tools to increase activity when individuals are most likely to take action [96]. Mobile app–based physical activity interventions can also gamify exercise to enhance enjoyment, which is a key mechanism for long-term engagement in physical activity [97,98]. In-the-moment enjoyment not only promotes regular exercise but is also in itself beneficial for mental health, contributing to the success of broader interventions such as behavioral activation. Overall, digital interventions are promising as they are low risk (ie, typically focus on reducing sitting and increasing light activity), can be deployed without a clinician, and can be used in the context of a patient's daily life. Ultimately, research conducted thus far in the general population supports mobile technologies as valid, accessible, and effective methods of promoting physical activity and reducing sedentary behavior.

Current Objective

Although it is reasonable to extrapolate that physical activity interventions could be implemented via smartphone in a similarly feasible, acceptable, and effective manner in psychiatric populations or for psychiatric targets, this remains an open question. High-quality trials of in-person exercise programs for mental health often unintentionally include components beyond the activity itself that are potentially therapeutic, such as regular, structured, and supervised sessions [65]. In other words, as part of most exercise programs, participants also receive regular social engagement or support, face-to-face time with a professional, instruction and demonstration of target behaviors, and guidance with behavioral scheduling or activation, all of which may enhance the therapeutic benefits of physical activity. However, the remote and asynchronous nature of technology-based interventions may provide a different experience from that of in-person programs, and thus, the impact may also differ. On the other hand, the aforementioned benefits of digital interventions, such as their ability to increase accessibility, lower logistical barriers to engagement, and enhance self-awareness while also promoting behavior change in real time, may boost response and, thus, lead to comparable—or even stronger—effects than face-to-face trials. As a result, it cannot be assumed that face-to-face physical activity programs or digital programs designed for other populations (eg, medical) will translate when delivered via smartphone or to a new population.

In this study, we explored the potential of physical activity interventions, as delivered (at least in part) via smartphone, to improve mental health in psychiatric populations. As this topic remains relatively new, we also considered available evidence on these tools to address mental health symptoms in nonpsychiatric populations. Specifically, we highlighted in which populations these tools have been tested; what outcomes have been evaluated (eg, acceptability, behavior change, and symptom change); and how strategies and tools have (or have not) been tailored to individuals with depression, anxiety, or related concerns. The goal was to provide a narrative overview and specific recommendations for future research on smartphone-based physical activity interventions for psychological disorders or concerns.

Methods

Literature Search

To provide a narrative overview of this emerging research area, we searched for articles that (1) described or reported an intervention intended to increase physical activity or reduce sedentary behavior; (2) included an emotional disorder, concern, or symptom as an outcome measure; (3) described or reported an intervention delivered entirely or in part via a smartphone app; and (4) were published in English and in peer-reviewed journals. Example search terms include "smartphone," "smartphone application," "mobile application," "mobile app," "digital mental health," "app-based," "app-assisted," "mobile phone," "ehealth," "digital," "mobile," "exercise," "physical activity," "sedentary," "sedentary behavior," "physical inactivity," "depression," "dysthymia," "mood," "anxiety," "phobia," "trauma," "posttraumatic stress," "obsessive compulsive disorder," "post-traumatic stress," "obsessive-compulsive disorder," "stress," "emotional disorder," "emotional problem," "well-being," "wellness," "affective disorder," "OCD," "PTSD," "MDD," "GAD," "mental health," and "mental illness." Web-based database (PubMed, Google Scholar, and Cochrane) searches and additional manual searches (eg, searching the reference sections of articles identified through database searches) were conducted up to March 2022. Records were initially reviewed by one coauthor; in cases of uncertainty about appropriateness for this review, records were reviewed in full by 2 additional coauthors and discussed until a consensus was reached.

Data Review

We attempted to extract the following information, where available, from each paper: sample size, inclusion criteria, demographics of the sample, primary aim, trial design (eg, randomized controlled trial), treatment duration, technology used, other interventions used (ie, in addition to physical activity), outcome measures (eg, primary and secondary measures of physical activity), results, treatment components or behavior change strategies, adaptations for mental health, and inclusion of coaching.

Results

Overview

In taking a narrative lens, we drew attention to the type of work that has been done and used these exemplars to discuss key directions to build on. Table 1 provides a summary of the included articles.

Table 1. Summary of studies investigating the impact of digital physical activity interventions on mental health symptoms.

Study, year	Sample size, N	Population studied	Intervention used	Duration	Physical activity outcome	Primary psychiatric outcome	Primary medical outcome
Aguilera et al [99], 2020[a]	N/A[b]	Adults with diabetes and a score of >5 on the PHQ[c]	Apps: DIAMANTE to track data and deliver adaptive learning algorithm (active only) and HealthySMS to send messages (active and control)	6 months	Steps	Depressive symptoms (PHQ-8[d])	HbA_{1c}[e] levels (blood glucose)
Broers et al [100], 2019	557	Adults diagnosed with hypertension, symptomatic heart failure, or coronary artery disease	Wearable (Fitbit) activity tracker, wearable (Beddit) sleep tracker, app (Moves) GPS tracker, app (Careportal) home monitoring system	6 months	Steps; physical activity level (combined length of active periods and step count)	Anxiety (GAD-7[f]) and depression (PHQ-9[g])	N/A
Damschroder et al [101], 2020	357	Web-based confirmation of veteran status	App (Stay Strong) and wearable (Fitbit Charge 2)	12 months	Active minutes per week; steps	N/A	N/A
Edney et al [102], 2020	444	Adults currently completing <150 minutes of MVPA[h] per week	App (Active Team) and wearables (pedometer and Zencro TW64S)	3 months	Daily minutes of MVPA	Symptoms of anxiety, depression, and stress (DASS-D[i])	N/A
García-Estela et al [103], 2021[a]	N/A	Spanish-speaking adults with MADRS[j] score of >1	App (IDEApp) and wearable (smartband)	8 months	SIMPAQ[k]; functional exercise capacity (6MWT[l] and 1-min sit-to-stand test); short Borg CR-10[m] Scale	Depressive symptoms (PHQ-9); well-being (WHO-5[n])	Global functioning (SF-36v2[o])
Guo et al [104], 2020	300	Adults who are HIV-seropositive with elevated depressive symptoms	App (WeChat) and Run4Love program (adapted CBSM[p] course and physical activity promotion) delivered through the WeChat app	3 months	Chinese version of the GPAQ[q]	Depressive symptoms (CES-D[r])	N/A
Haufe et al [105], 2020	314	Adults with metabolic syndrome	Wearable (activity monitor; Forerunner 35; Garmin)	6 months	Freiburger Questionnaire on Physical Activity; steps	Anxiety severity and depression severity (HADS[s])	Change in metabolic syndrome severity; health-related quality of life (SF-36[t])
Kim et al [106], 2021	21	Adults aged >46 years diagnosed with PD[u] or atypical parkinsonism conditions and regular participation in a PD exercise program at least once a week	App (researcher-created physical activity app)	8 weeks	Total exercise calculated by multiplying the frequency and duration for all exercises; subjective exercise scale (Borg 6-20 scale); IPAQ[v]	Depression (Geriatric Depression Scale–Short Form)	PDQ-39[w]
Lin et al [107], 2020[a]	N/A	Physically inactive adults with a musculoskeletal diagnosis (*ICD-10*[x]) and in rehabilitation following inpatient clinic treatment	App (MoVo)	12 months	BSA[y], sport activity and movement activity subscales	Depression (PHQ-9); anxiety (GAD-7)	Brief Pain Inventory

Study, year	Sample size, N	Population studied	Intervention used	Duration	Physical activity outcome	Primary psychiatric outcome	Primary medical outcome
Ma et al [108], 2015[a]	N/A	Adult participants who were obese and experiencing depression	Wearable (Fitbit) and app or website (MyFitnessPal)	12 months	Minutes of physical activity logged on MyFitnessPal	Depression (SCL-20[z])	Changes in BMI
Nadal et al [109], 2021[a]	N/A	Adult users with mild to moderate depression who were assigned to iCBT[aa] treatment for depression	Wearable (smartwatch; Mood Monitor watch app)	8 weeks	Smartwatch activity data	Depression (PHQ-9), anxiety (GAD-7), and functional impairment (WSAS[ab])	N/A
Park et al [110], 2021	60	Adults with a history of cardiovascular disease who were within 2 weeks of completing cardiac rehabilitation	Wearable (Fitbit Charge 2) and apps (Movn and Fitbit)	2 months	Steps; 6MWT; self-reported physical activity	Quality of life (QLESQ[ac]); depression (PHQ-9)	N/A
Puszkiewicz et al [111], 2016	13	Adults with a diagnosis of breast, prostate, or colorectal cancer who had finished primary curative treatment	App (GAINFitness)	6 weeks	Physical activity (GLTEQ[ad])	Health and quality of life outcomes (EQ-5D); well-being (FACT-G[ae]); anxiety and depression (HADS)	Cancer-related fatigue (FACIT[af]); sleep quality (PSQI[ag])
Puterman et al [112], 2021	334	Adults with a score of 1-3 on the L-CAT[ah] who were cleared to exercise	App (Down Dog suite of apps—HIIT[ai] and yoga)	6 weeks	Sessions of yoga or HIIT completed; minutes of yoga or HIIT completed	Depressive symptoms (CESD)	N/A
Skrepnik et al [113], 2017	172	Adults with osteoarthritis eligible to receive the hylan G-F 20 injection	Wearable (Jawbone UP24) and app (OA GO)	90 days	Steps	VAMS[aj]	Changes in sleep captured by the wearable activity monitor
Stephens et al [114], 2022	15	Youth (aged 11 to ≤18 years) with MS[ak] and a disability rating of <4 on the EDSS[al] and attending a pediatric MS and neuroinflammatory disorder clinic	App (Atomic)	12 weeks	Physical activity measured via accelerometry; time spent in MVPA and sedentary activities; aerobic fitness, musculoskeletal strength, and walking endurance	Depression (CES-DC[am])	N/A
Teychenne et al [115], 2021	62	Mothers 3-9 months post partum, insufficiently active, and experiencing heightened depressive symptoms	App (smartphone app and web forum)	12 weeks	Self-reported physical activity; accelerometer-assessed physical activity and sedentary behavior	Depressive and anxiety symptoms (unstandardized questionnaires)	N/A
Wilczynska et al [116], 2020	N/A	Adults with or at risk of type 2 diabetes	App (eCoFIT)	20 weeks	N/A	Depressive and anxiety symptoms (PHQ-9 and GAD-7)	Social support, self-efficacy, nature relatedness, and perceived sleep quality

Study, year	Sample size, N	Population studied	Intervention used	Duration	Physical activity outcome	Primary psychiatric outcome	Primary medical outcome
Wong et al [117], 2021	79	Adults with moderate depressive symptoms	App (Lifestyle Hubl)	8 weeks	Physical activity level (IPAQ)	Depressive and anxiety symptoms (PHQ-9 and GAD-7)	Insomnia (ISI[an]); health-related quality of life; health-promoting behaviors (HPLP-II[ao]); functional impairment (SDS[ap])

[a]Published protocol.
[b]N/A: not applicable.
[c]PHQ: Patient Health Questionnaire.
[d]PHQ-8: Patient Health Questionnaire-8.
[e]HbA_{1c}: hemoglobin A_{1c}.
[f]GAD-7: Generalized Anxiety Disorder-7.
[g]PHQ-9: Patient Health Questionnaire-9.
[h]MVPA: moderate to vigorous physical activity.
[i]DASS-D: Depression Anxiety Stress Scales: Depression Subscale.
[j]MADRS: Montgomery–Åsberg Depression Rating Scale.
[k]SIMPAQ: Simple Physical Activity Questionnaire.
[l]6MWT: 6-Minute Walk Test.
[m]CR-10: Borg Category-Ratio scale.
[n]WHO-5: World Health Organisation–Five Well-Being Index.
[o]SF-36v2: 36-Item Short Form Health Survey version 2.
[p]CBSM: cognitive behavioral stress management.
[q]GPAQ: Global Physical Activity Questionnaire.
[r]CES-D: Center for Epidemiologic Studies Depression Scale.
[s]HADS: Hospital Anxiety and Depression Scale.
[t]SF-36: 36-Item Short Form Health Survey.
[u]PD: Parkinson disease.
[v]IPAQ: International Physical Activity Questionnaire.
[w]PDQ-39: Parkinson Disease Questionnaire-39.
[x]*CD-10: International Statistical Classification of Diseases, Tenth Revision*I.
[y]BSA: Movement and Sport Activity Questionnaire.
[z]SCL-20: Symptom Checklist Depression Scale.
[aa]iCBT: internet-based cognitive behavioral therapy.
[ab]WSAS: Work and Social Adjustment Scale.
[ac]QLESQ: Quality of Life Enjoyment and Satisfaction Questionnaire.
[ad]GLTEQ: Godin Leisure-Time Exercise Questionnaire.
[ae]FACT-G: Functional Assessment of Cancer Therapy–General.
[af]FACIT: Functional Assessment of Chronic Illness Therapy.
[ag]PSQI: Pittsburgh Sleep Quality Index.
[ah]L-CAT: Stanford Leisure-Time Categorical Activity Item.
[ai]HIIT: high-intensity interval training.
[aj]VAMS: visual analog mood scale.
[ak]MS: multiple sclerosis.
[al]EDSS: Expanded Disability Status Scale.
[am]CES-DC: Center for Epidemiologic Studies Depression Scale for Children.
[an]ISI: Insomnia Severity Index.
[ao]HPLP-II: Health-Promoting Lifestyle Profile-II.
[ap]SDS: Sheehan Disability Scale.

Who Was Included in This Work?

Overall, a review of the literature revealed that little work has been done to test the impact of smartphone-based physical activity interventions on increasing physical activity or reducing mental health symptoms in psychiatric or at-risk populations. Most related trials with clinical populations have been conducted in the area of medicine, with studies investigating the effects of physical activity—encouraged through smartphone- and wearable-based interventions—on physical health conditions (eg, diabetes [99,116], obesity [86,108], cancer [111,118], cardiovascular issues [100,110], and multiple sclerosis [114]). Physical activity is well established as a means of facilitating rehabilitation following serious illness or injury, as well as mitigating the progression of chronic health conditions [119-121]. For example, adults with Parkinson disease who used a minimally supported, customizable home-based exercise app for 8 weeks doubled their amount of weekly exercise (minutes) while also increasing the intensity of such exercise [106]. Similarly, engagement with smartphone-based physical activity interventions led to increased strenuous exercise among adults with cancer [111] as well as increased step count for those with cardiac issues [100,110] and for youths with multiple sclerosis [114]. These changes are noteworthy as medical illness or disease can serve as a barrier to engaging in health-promoting behaviors [122] despite the knowledge that such behaviors can stabilize or even improve such medical conditions [123].

In contrast, the practice of formally integrating exercise into *mental* health care is relatively new and not established in current standards of care. This is mirrored by the disproportion of extant research examining digital tools for increasing physical activity in medical versus psychiatric populations. When mental health targets were examined, they were largely included as secondary or exploratory outcomes and frequently framed in relation to coping with the medical concern of interest [99,100,104-108,110,111,113,114]. We only identified 16% (3/19) of the studies that specifically recruited individuals with psychiatric symptoms, and all (3/3, 100%) were focused on individuals with depressive symptoms [103,109,117]. The most common mental health outcomes were depression, anxiety, general quality of life, and emotional well-being [99,100,102-110,112,114-117]. Specifically, subclinical depressive concerns were the most frequently investigated psychiatric target, followed by subclinical anxiety [100-105,112,115-117].

The impact of smartphone-based physical activity interventions is yet to be investigated explicitly for individuals diagnosed with depressive or anxiety disorders, let alone other mental health conditions, including serious mental illness. Furthermore, although wide age brackets were represented across the studies, with average ages ranging from teenagers to older adults, men and non-White individuals were underrepresented. As research progresses in this space, it will be imperative to include the experiences and perspectives of adults with clinical levels of psychiatric concerns as well as diverse backgrounds and identities.

Do Smartphone-Based Physical Activity Interventions Benefit Mental Health?

Owing to the limited available data; small samples comprising mostly White Western, educated, industrialized, rich, and Democratic women with subclinical depression or anxiety; and heterogeneity of outcomes measured, it is difficult to conclude whether and to what extent existing smartphone-based physical activity interventions benefit mental health. However, with these caveats in mind, we aimed to synthesize the available evidence in the following sections.

Feasibility and Acceptability

There is encouraging evidence that smartphone-based physical activity interventions could be feasible and acceptable for psychiatric populations. However, supporting data were primarily collected in samples of individuals with elevated depressive symptoms or who were at risk of depression rather than in explicitly clinical samples or among individuals with other prominent mental health concerns. One study of postnatal women at risk of depression found *low* engagement with the digital aspects of a 12-week multicomponent physical activity intervention (home exercise equipment and a physical logbook combined with a motivational smartphone app and a web-based social support forum) [115]; however, other studies reported more positive participant response and engagement. For example, retention in smartphone interventions for physical activity tended to be high compared to that in other types of digital health interventions—one systematic review found that completion rates of digital mental health interventions ranged from 1% to 28% [124]. In a study of adults with diabetes, retention in a 20-week digital physical activity intervention (an app that allowed participants to use workout circuits, set goals, monitor progress, and learn cognitive and behavioral strategies) was of >70% [99]. Similarly, compliance with a suite of high-intensity interval training (HIIT) and yoga apps during a 6-week intervention was strong in a community sample with elevated depressive symptoms. More than half of the participants in the yoga and HIIT+yoga group and 40% in the HIIT group continued completing the recommended 4 sessions per week by the end of the trial [112]. Compliance and satisfaction ratings were comparably high when physical activity promotion was combined with cognitive behavioral stress management via a WeChat intervention in a group of adults with HIV and elevated depressive symptoms [104].

These findings are consistent with those of the larger literature showing that digital physical activity interventions tend to be well received by participants [85,125,126]. Public interest is already high, with physical activity and fitness apps dominating the mobile health space. Notably, a 2018 systematic review of the experience of adults who used mobile interventions to promote physical activity highlighted important themes to be considered for future design—self-reported engagement was most enhanced by the availability of social features, prompts, goal setting, personalization or customization, and gamification but was limited by low technological literacy, preference for coached apps, and a desire for social support [127].

Change in Psychiatric Symptoms

The impact of smartphone-based physical activity interventions on psychiatric symptoms was far more mixed. Some studies observed resultant improvements in symptoms. In one study that included patients specifically recruited for having elevated symptoms of depression and anxiety, patients experienced a reduction in depression and anxiety scores following a 6-month exercise intervention (150 minutes of moderate physical activity per week with individual recommendations given via a smartphone app) as compared to a waitlist control. However, it is notable that this intervention did not test a smartphone-based physical activity intervention in isolation but, instead, combined it with nutritional counseling and the option of receiving exercise recommendations through personal meetings or by phone instead of an app [105]. Among adults from the general population with low physical activity scores, using a publicly available suite of exercise apps for 6 weeks significantly improved depressive symptoms compared to a waitlist control [112]. Psychiatric improvements were also observed in patients with medical comorbidities. For example, in a study of adults with obesity using a smartphone-based physical activity app (eCoFIT), depression symptom severity improved after 20 weeks [116]. Finally, in a study of older adults with Parkinson disease, using a mobile app to access and customize a home-based exercise program for 8 weeks led to reduced depression symptoms and improved quality of life [106].

However, other studies reported null effects. For example, among patients with a history of cardiovascular disease, a smartphone-based intervention including motivational prompts and educational messages did not yield significant changes in depressive symptoms from baseline to 2 months [110]. Furthermore, a trial of patients with cancer using a tailored physical activity smartphone app (which included workout videos, spoken instructions, and push notifications) did not observe changes in depression, anxiety, or quality of life after 6 weeks [111]. In a study testing an app-based physical activity program for youth with multiple sclerosis (including personalized coaching and promotion of aerobic fitness, musculoskeletal strength, and walking endurance), there was no change in depression levels over 12 weeks [114]. Furthermore, a meta-analysis of studies examining digital physical activity interventions in cancer survivors found that none of the included studies were successful in improving depression or anxiety [118]. In the general population, one study similarly did not find significant differences in depression, anxiety, stress, or well-being after 3 or 9 months of using an app (Active Team) and wearable pedometer [102].

These results should be interpreted cautiously for 2 reasons. First, there is the confound of potential floor effects—as none of these trials were designed to address questions about mental health, symptom levels at study start were typically already low, thus reducing investigators' abilities to identify possible effects. In addition, studies varied widely in their evidence of behavior change, including outcome measures (eg, minutes of activity, number of sessions completed, exercise intensity, and fitness level; see Table 1 for detailed information on this variance) and the use of objective versus subjective reports, which are known to be discrepant [82-84]. In other words, if an intervention did not produce meaningful physical activity changes, it would be unlikely that downstream emotional changes would occur.

How Have Smartphone-Based Physical Activity Interventions Been Tailored to Individuals With Depression, Anxiety, or Other Psychiatric Concerns?

To date, interventions generally have not been tailored to the specific needs or presentations of individuals with mental health concerns. This is unsurprising as most technology-based physical activity trials have not been designed to target mental health. However, some studies have integrated components that specifically address psychological well-being. One adaptation that is low effort but high return is adding psychoeducation about the relationship between physical activity and mental health. For example, a recent trial for adults with mild to moderate depressive symptoms devoted the first in-person group session to discuss the relationship between depression and exercise to complement the personalized exercise program, smartphone app, and wearable device they received [103].

The second adaptation observed in the literature is offering concurrent psychotherapy-based tools. In some cases, psychotherapeutic content was interspersed with the physical activity intervention; for example, in a trial for adults living with HIV and depression in China, both an exercise promotion intervention and cognitive behavioral stress management course were delivered as multimedia messages through the WeChat app [104]. Relatedly, in a study of the eCoFit app for adults with or at risk of type 2 diabetes, short cognitive behavioral tasks ("FitMind Challenges") were integrated throughout the program [116]. Examples of FitMind Challenges included motivational strategies, relaxation, cognitive restructuring, social support, and problem-solving. In other cases, the approaches were delivered in parallel. In one trial, adults who were overweight and experiencing depression received a 7-step problem-solving therapy via a workbook in addition to live lifestyle coaching, at-home video lessons, the MyFitnessPal app, and a Fitbit for monitoring [108]. In another study, SilverCloud's guided internet-based CBT program for depression was the primary intervention, with smartwatch-based monitoring of sleep, steps, and mood added to promote greater awareness of the relationship between health behaviors and mood, thereby independently encouraging positive lifestyle changes [109].

A third but largely unexplored avenue is the inclusion of personalized or tailored messaging. This is an opportunity for coaches or other support persons to address barriers that may be specific to the experience of someone with mental health concerns (eg, navigating social anxiety to go to the gym and restructuring depressive thoughts). In one open trial of a physical activity app for youth with multiple sclerosis, coaches were trained in social cognitive theory for behavior change as well as motivational interviewing [114].

Although few digital physical activity interventions have been designed or modified to specifically affect mental health, many have been designed using evidence- and theory-based behavior change strategies that are ripe for implementation in psychiatric contexts. Indeed, the most successful interventions are based

on behavioral theory [73,127]—explicitly stated or not—such as the transtheoretical model [128], the theory of planned behavior [129], self-determination theory [130], and social cognitive theory [131]. Social cognitive theory is most often cited given its emphasis on internal, external, and social factors that reinforce learning and contribute to sustained change [131]. Targeting self-efficacy, self-regulation, and social support to engender meaningful, lasting behavior change aligns strongly with principles of psychotherapy as well. Digital physical activity interventions have also experimented with numerous evidence-based behavior change techniques, including goal setting and review, action planning, regular feedback, self-monitoring of behavior, instruction and demonstration of how to perform a new behavior, graded tasks, prompts and cues, and social rewards, to name a few [132-134]. Interventions integrating multiple behavior change strategies are more successful than those that rely on one (eg, self-monitoring or reminders alone [133]). Considering how such strategies could be adapted for individual presentations (eg, those with clinical levels of dysregulation) should be a priority for future iterations of these programs.

Furthermore, technology-driven techniques may uniquely (or at least more strongly than traditional treatments) engage key mechanisms of behavior change. For example, these tools can promote self-awareness. As people tend to keep their devices close to them throughout their daily lives, wearable and mobile platforms can provide objective, continuous monitoring and feedback related to behavioral patterns such as physical activity [102,108]. In addition, these approaches can enhance a person's likelihood of changing their behavior by lowering the cognitive burden involved in initiating physical activity. Strategies include delivering content more flexibly (eg, when it is most convenient for a participant to engage or in doses of their choosing); modeling target behavior via written, image, or video instructions that can be reviewed on demand or infinite times; or tailoring activity suggestions to a person's present context (eg, suggesting at-home activities on rainy days). This may be particularly meaningful for psychiatric audiences as depression and anxiety are associated with attention and memory deficits that can interfere with information processing and learning [135-138]. In addition, in-the-moment rewards and other gamification or reinforcement features could be particularly useful early on [139,140] as individuals with depression and anxiety may not experience initial sessions of exercise as intrinsically gratifying or mood boosting as others do; for example, depression is characterized by deficits in reward processing and motivation [141], and anxiety sensitivity and social anxiety can blunt positive responses or promote avoidance [142,143].

Finally, personal devices may allow for more consistent, flexible social support throughout an intervention. Social support is an established, evidence-based behavior change technique that promotes physical activity [126,144-146]. Smartphone-based physical activity interventions provide a range of avenues for social connection, such as texting with a coach [101,107], access to an web or app-based discussion forum [115], and creation of virtual "teams" [102]. Critically, although social media has been frequently incorporated as a means of facilitating connection, participant reactions have been mixed, and it may not be optimal for psychiatric populations [126]. In general, social support appears to boost engagement when it is perceived to facilitate emotional support, provide tips from peers, enhance motivation, foster social comparison or competition [126,147,148]. How to best leverage social support and social media for psychiatric populations requires nuanced future study.

Discussion

Principal Findings

The primary aim of this narrative synthesis was to examine the status of smartphone-based physical activity interventions for mental health and understand how they have and have not been tailored to or evaluated in psychiatric populations. Ultimately, the literature is limited and difficult to synthesize owing to the high heterogeneity across the studies in terms of sample selection; study design; outcomes of feasibility, acceptability, and efficacy; and degree of tailoring. To date, mental health outcomes have typically been secondary or exploratory within trials focused on medical outcomes (eg, diabetes management) and, when included, have had a narrow focus on measures of depression, anxiety, and general well-being in nonclinical populations. As a result, this review relied significantly on research focusing on medical populations to explore how smartphone-based physical activity interventions could be used to impact mental health outcomes and to infer how they may be used in psychiatric populations. Furthermore, although extant studies have included diversity of age, the samples in the included studies comprised mostly White and female individuals, thus reducing the generalizability of the already limited findings.

The feasibility and acceptability of these interventions for subclinical and at-risk populations are encouraging and suggest that digital physical activity programs may be similarly well received among individuals above diagnostic thresholds. The available data on psychiatric outcomes were mixed. However, it is difficult to draw meaningful conclusions given the limited data; high heterogeneity of intervention approach and target behavior; and lack of standardization in measurement and reporting of use, engagement, and behavior change, as well as the elevated risk of floor effects given the subclinical samples. These inconclusive psychiatric outcomes may also be related to a lack of tailoring of smartphone-based physical activity interventions to the specific needs of those presenting with mental health concerns. The existing tailoring included basic psychoeducation about physical activity as a treatment, adding concurrent psychotherapy-based tools, and including personalized or tailored messages. There was no standardization or evidence base for how this tailoring was applied. The upshot is that many of the papers included in this review presented interventions that were already built around established, evidence-based behavior change strategies, which suggests that psychotherapeutic tailoring could be efficiently integrated into existing smartphone-delivered physical activity interventions. In general, effective physical activity interventions use many of the same fundamental behavior change strategies commonly found in psychotherapy, such as education, goal setting,

self-monitoring, graded tasks, engaging social support, and motivational interviewing [149,150].

Taken together, the primary barrier to advancing the use of smartphone-based physical activity interventions in mental health care is the absence of evidence. The need for research in this area has been highlighted in other reviews as well [151,152]. To construct a more consistent, evidence-based foundation for intervention development, we outline several avenues for future research.

Recommendations for Tailoring Physical Activity Interventions to Psychiatric Populations

More research is needed to better understand *how* existing smartphone interventions can be tailored to fit the needs of psychiatric populations. The following are example adaptations rather than an exhaustive list. One likely critical step is to provide users with psychoeducation early on about the ways in which physical activity can be used to affect psychological health, such as improving mood and reducing anxiety. This should involve making explicit connections between health behaviors (eg, exercise), mental health symptoms, and emotion regulation so that users can better appreciate the bidirectional links between these areas of well-being. Including even a brief text summary of the literature or treatment rationale could likely augment the effects [153,154]. In fact, there is evidence with depression treatment that physical activity interventions lacking such psychoeducation or treatment rationale do not lead to robust clinical changes and can worsen dropout rates [155]. In contrast, attending to the mental and emotional benefits of exercise, particularly the acute or immediate impact on affect or resilience, can further enhance mood and motivation to continue exercising [35,156]. Highlighting these benefits and encouraging users to monitor such positive changes could improve sustained engagement and clinical response. Technology may be particularly helpful for this; apps, for example, can provide in-the-moment reminders through push notifications to attend to one's affect or visual feedback of a user's pretest-posttest change in self-reported mood with exercise. Furthermore, digital tools could provide information about the impact that mental health symptoms may have on program engagement. This can help users recognize that it is normal and expected for symptoms such as fatigue, anxiety sensitivity, or low motivation to serve as barriers to physical exercise and can proactively help users engage in related problem-solving.

Another compelling feature to test is the incorporation of modules or content that specifically address mental health concerns or symptoms. For example, technology-based physical activity interventions aimed at improving anxiety symptoms would benefit from including evidence-based skills such as cognitive restructuring and exposure practices. These approaches can be used to identify and challenge maladaptive beliefs that anxiety symptoms such as a racing heart are dangerous (known as anxiety sensitivity [157]), design a more graded exercise plan, and use activity as an interoceptive exposure by allowing patients to experience and tolerate those feared sensations [158]. Furthermore, CBT skills can be incorporated to address exercise-related social anxiety, such as integrating exposures (eg, walking with a friend or going to a gym first at off-peak hours) and challenging associated negative expectations (eg, "I won't be able to keep up with my friend and they will judge me"). Meanwhile, a digital physical activity intervention aimed at helping people with depression could include skills consistent with behavioral activation, such as tracking the relationship between mood and activities (including physical activity); intentionally adding new behaviors such as gardening, walking, or going to the gym to their weekly schedule; and generating more flexible approaches to regular movement. Furthermore, cognitive skills can be used for participants to identify and evaluate negative thoughts about themselves or the program (eg, "I can never stick to my goal of going for walks, so what's the point") in terms of their accuracy or utility. In this vein, equipping coaches with some knowledge of common mental health symptoms to look out for, destigmatize, and address could enhance outcomes.

In addition, coaches are understood to bolster digital interventions in general by providing further psychoeducation or resources, personalizing content or skill use, and answering questions. Given the known relationships between mental illness and both inactivity and chronic medical conditions, even with basic mental health knowledge, coaches could perform these roles better. It is also important to recognize that digital or physical activity interventions may not be the appropriate or most effective level or type of care for all individuals experiencing mental health concerns. As such, tools should include information for users on the signs or symptoms that may indicate that pursuing psychotherapy could be beneficial as well as resources for doing so.

Finally, physical activity promotion tools could be integrated to augment existing treatments. Currently, there are a number of well-established treatments for psychiatric disorders, such as CBT and mindfulness. CBT has extensive research support as the gold-standard treatment for a range of disorders, including depression and anxiety [159]. This treatment integrates both behavioral and cognitive skills, such as tracking and scheduling activities as well as evaluating and challenging maladaptive thoughts, to reduce the severity and impact of symptoms. Mindfulness—or the purposeful, nonjudgmental awareness of the present moment [160]—is increasingly included in "third wave" interventions to reduce psychiatric symptoms [161] as well as increase well-being, such as positive affect and quality of life [162,163]. Recently, a dominant focus of digital mental health innovation has been translating these gold-standard psychotherapies to digital platforms. There is now a strong foundation of evidence supporting the feasibility, acceptability, and efficacy of delivering these treatments through both face-to-face and digital means [164-166].

Thus far, the development of digital interventions for physical activity and for mental health has largely occurred separately. However, their concurrent delivery provides promising initial evidence [111,118]. This parallels in-person trials demonstrating that increasing physical activity strengthens psychotherapy outcomes [167,168]. The next step in this line of research is to more formally integrate the 2 or even develop technologies in which both sets of skills are delivered within a single coherent platform. For example, in a study conducted by Wilczynska et al [116], adults with diabetes used the eCoFit app, which

integrated guided workouts, goal setting, and cognitive behavioral skills. Some of the commercially available apps targeting mental health have already begun moving in this direction as well. The mindfulness-based app Headspace has recently incorporated a suite of video- and audio-guided exercises that help users engage in activities such as stretching, dancing, and yoga. Within these integrated platforms, it will be important to explicitly link the mental health and physical activity content rather than presenting them side by side as distinct intervention pathways.

Recommendations for Developing and Testing Physical Activity Interventions for Populations With Psychiatric Disorders

Overview

More research is also needed to understand *to what degree* smartphone interventions require tailoring and for whom. Given the wide range of possibilities, an important step in the development process is to have focus groups with the goal of hearing from individuals with lived experience about their wants and needs. Through pilot-testing, intervention design and refinement can be an iterative process wherein individuals of the target audience engage with the program, feedback is elicited, and changes are made in response to that feedback. This user-centered approach fits well within the larger preparation phase of a Multiphase Optimization Strategy. Following such development, digital tools should be scientifically tested and optimized, leading to a randomized controlled trial to examine their efficacy in achieving the outcomes of interest (eg, reduction in depression symptoms). This testing phase is necessary to establish a program as evidence-based, which would allow it to stand out in an otherwise large pool of digital applications that are not backed by research.

Leveraging New Trial Designs

New trial designs, such as sequential multiple-assignment randomized trials, microrandomized trials, and factorial designs, will be useful in intermediate stages to parse issues such as dosing, sequencing, and personalization. For example, research shows that at least 6 weeks are required for new physical activity habits to form [169]; thus, interventions that are, on average, 8 weeks long lead to more lasting changes than shorter ones [169]. Moreover, it remains unclear whether longer treatments, such as ≥24 weeks, have a greater impact on the general population [170]. It is unknown what duration would be sufficient for various clinical populations to observe changes in both the target behavior and in downstream symptoms and for whom extended support would be necessary. Individual components such as the aforementioned tailoring elements or the inclusion of coaching can also be efficiently tested using these new study designs [170].

Understanding the Role of Human Support

Previous work has shown that supervised exercise tends to have a larger impact on anxiety outcomes than unsupervised prescriptions [171]. This parallels guidance from experts in digital mental health that including human support (eg, a lay coach or therapist) alongside internet- or app-based cognitive behavioral and other therapies should enhance retention, engagement, and outcomes. Possible explanations could be greater accountability, the presence of social support, clearer guidelines, opportunities to ask for clarification, in-the-moment personalization, problem-solving, direct affirmation or reinforcement, and a more regular routine. However, guidance does not unilaterally improve outcomes for all digital interventions or all patients [164,172]. There are currently no evidence-based guidelines for implementing human support in digital interventions (ie, when, how often, how much, by whom, and for which users), let alone a nuanced understanding of *how* coaches can effect positive change in adherence or response [173]. Understanding the mechanisms of action would allow developers to maximize automation and most efficiently deploy human support when needed. As human support is the most expensive and scarce resource in digital health solutions, it will be critical to determine how to automate some of these supportive pathways and how to most efficiently identify who needs human support and at what dose.

Considering Individual Factors

Across intervention types, we must also consider the individual factors that may serve as facilitators of or barriers to engagement and success. This is doubly important for smartphone-based physical activity interventions as there are potential barriers inherent in both smartphone use and physical activity uptake.

Barriers to Physical Activity

Research examining barriers to physical activity in those with mental health conditions suggests that individuals with high symptom severity and low self-efficacy may be particularly disinclined to pursue physical activity–based interventions [174,175]. Lack of social support, lack of available time, and fear of injury were also frequently mentioned barriers in a sample of adults with anxiety and depression [174]. In addition, individuals with higher or lower levels of baseline physical activity or fitness may face different barriers and have different needs. A qualitative review suggested that those with lower baseline physical activity wanted an app that had more of a coaching role, whereas those with a higher baseline physical activity preferred an app that helped them intensify or optimize their current physical activity level [126]. In addition, developers should consider the accessibility of exercise suggestions; for example, exercise prescriptions that necessitate equipment, a gym membership, or access to a safe outdoor space may not be generalizable to many otherwise well-suited recipients.

Barriers to Digital Mental Health Use

A recent review by Borghouts et al [176] examining barriers to and facilitators of user engagement found that scoring high on neuroticism and agreeableness was associated with greater interest in using smartphone apps to reduce stress, whereas scoring high on extraversion was a predictor of preferring in-person services to web-based options [177]. The severity of baseline symptoms—both psychiatric and comorbid medical concerns—may also play a role in engagement and adherence. Most smartphone-based digital physical activity interventions have been investigated in those with mild to moderate symptoms, which can hamper engagement with apps [178,179].

Some studies suggest that those with mild depression may actually be at an even greater risk of dropout than those with moderate depression [178,180]. Researchers should also be cautious about potential iatrogenic app components. For example, the tracking components inherent in many smartphone-based physical activity interventions, particularly those related to physical health and tracking activity, run the risk of becoming compulsive or rigid. This could pose an issue for individuals with obsessive-compulsive, anxiety, and related disorders that are often characterized by perfectionism or inflexibility. Furthermore, peer support groups within apps, although often helpful, could also lead to negative social comparisons, thus exacerbating depression.

A possible mitigating approach to this would be to introduce smartphone-based physical activity interventions through a stratified care model in which individuals are allocated to different levels of an intervention depending on their clinical needs. In this model, providers could use patient-level data to decide whether an individual would benefit from the smartphone tool as a stand-alone intervention, as a coached version integrated with another level of care (eg, psychotherapy with a clinician), or delivered after progress with another intervention has been made (eg, medication stabilization).

Barriers Related to Technology

For many, significant barriers may include technology literacy and access. Multiple studies have identified technology literacy as an obstacle to digital physical activity intervention uptake [126,181] and for digital mental health use [178,180]. This, coupled with the lack of technical support provided by many apps, means that individuals who may be motivated to engage with smartphone-based physical activity interventions are stymied by the inability to navigate the app or seek help when issues arise. This necessitates a user interface and navigability features that can be understood or customized by a range of age groups and technological ability levels, and furthermore, it emphasizes the importance of accessible, embedded technical support tools for those who need them. Coached or guided tools may be helpful in mitigating this literacy issue but still require users to have the basic skills needed to contact their coaches or guides for help. In addition, there is still a subset (15%) of the US population that does not own a smartphone, many of whom represent communities that could benefit the most from flexible, low-cost, and accessible support options [182]. An even greater percentage of Americans lack a stable internet connection; this statistic is highly stratified along racial lines—8 in 10 White adults report having a broadband connection at home, whereas only 71% of Black adults and 65% of Hispanic adults report the same [183]. This suggests a need for digital physical activity interventions that can be accessible from communal settings, such as local community centers or publicly available fitness facilities.

It should be acknowledged that the vast majority of digital mental health interventions are designed—intentionally or not—with primarily White, Western, educated, industrialized, rich, and Democratic populations in mind. Research investigating the efficacy of both in-person and digital physical activity interventions also suffers from similarly nonrepresentative samples, calling into question which validated strategies are universally beneficial. For example, there is compelling research that physical activity interventions (both in person [184,185] and virtual [186]) are effective for health behavior change in Black Americans; however, other studies investigating the perspectives of Black Americans suggest that many in this community face unique social and structural barriers to physical activity that may not be considered by extant programs [187,188]. Thus, engagement in smartphone-based physical activity interventions by marginalized populations might be impacted by both the perception and reality that many of these apps are not designed with their community or culture in mind. Future research should engage with a range of underserved populations in qualitative research to understand the community and cultural values surrounding physical activity and technology use. This information should then be used to collaboratively design new programs or features or culturally tailor existing tools to meet the needs of a broader audience.

All these limitations support the importance of qualitative work as a future direction when building smartphone-based physical activity interventions for mental health. The research synthesized by Carter et al [126], for example, presents valuable insights into individual-level concerns and emerging trends in patient preferences for components and design of apps. In particular, their identification of 2 key mechanisms through which mobile health use facilitates physical activity (strengthening motivation and changes in self-awareness and strategizing) is an important step in boosting engagement and exploring the mechanisms by which these apps function. Engaging qualitatively with a broad spectrum of stakeholders would also be foundational in improving the issue of representation, thereby supporting the goal that smartphone-based physical activity interventions for mental health are acceptable and efficacious for all.

Practical Considerations

The unfortunate reality of scientifically validated digital mental health products is that the vast majority do not move beyond their success in the laboratory [173], and those that do make it to market face fierce competition, flagging engagement rates, and a lack of financial means to scale the project. Thus, in developing and testing these promising digital mental health–focused physical activity tools, investigators should integrate elements essential for successful dissemination. One proposed pathway for improving the dissemination and ultimate success of digital mental health tools is to connect consumers through employers or public and private insurance companies, who have indicated a growing interest in expanding services to cover digital mental health. To illustrate, Blue Shield of California is now offering the mindfulness meditation app Headspace to subscribers [189]; Cigna offers the mental health app Ginger as part of its service package [190]; and Kaiser Permanente supports the use of Ginger, Calm, and MyStrength [191]. Physical activity technologies that track and manage exercise and step count are even more prevalently covered by insurance. Blue Cross Blue Shield, United, and others have partnered with Fitbit to offer low-cost wearable devices and use of their apps to promote health behavior change. Aetna and Cigna offer similar programs and occasional incentives to people who use health-tracking apps and devices. Taken together, the

enthusiasm for apps and devices promoting physical health, as well as the recent foray by insurers into the digital mental health space, suggests that smartphone-based physical activity interventions for mental health may be prime for scalable coverage. This also means that academics developing such tools should be mindful when designing research studies to collect outcome data relevant to insurers and other payers, such as outcomes related to health care costs (reduction in insurance claims and physician or therapist visits), disability-adjusted life years (reduction in overall illness burden), adoption and engagement rates, and user data such as acceptability and fidelity (whether people use the tools as intended). In addition, investigators and designers should carefully consider the costs inherent to their interventions, such as relying on "off-the-shelf" versus research-grade devices and other platforms, the extent to which an intervention relies on human support to be administered, and the broader infrastructure required for implementation and sustainment, all of which will alter accessibility and scale. By designing research studies on smartphone-based physical activity interventions with true scalability in mind, researchers will be better poised to expand their intervention beyond academia and better achieve the goal of connecting evidence-based interventions with those who need them.

Conclusions

Physical activity has well-known and broad mental health benefits. However, a minority of at-risk individuals or those with mental disorders meet even the minimum exercise recommendations. Smartphones may bridge this gap given their pervasiveness in daily life, capacity to help concurrently manage multiple dimensions of personal health, and ability to engage key mechanisms of behavior change. Although early data for smartphone-based physical activity interventions reducing psychological symptoms are encouraging, overall, surprisingly little work has been done in this area. Therefore, there is untapped potential for developing and disseminating accessible, beneficial tools that can have a great public health impact.

Acknowledgments

During the preparation of this manuscript, EEB and BMH were supported in part by the Hope Fund, an internal fellowship at Massachusetts General Hospital provided through the generosity of an individual donor to support early-career investigators.

Conflicts of Interest

EEB receives research support from Koa Health, is a presenter for the Massachusetts General Hospital Psychiatry Academy in educational programs supported by independent medical education grants from pharmaceutical companies, and has a consulting agreement with Otsuka Pharmaceutical Development & Commercialization, Inc. ECW has no competing interests to declare. BMH receives research support from Koa Health. SW is a presenter for the Massachusetts General Hospital Psychiatry Academy in educational programs supported by independent medical education grants from pharmaceutical companies and has received royalties from Elsevier Publications, Guilford Publications, New Harbinger Publications, Springer, and Oxford University Press. SW has also received speaking honoraria from various academic institutions and foundations, including the International Obsessive-Compulsive Disorder Foundation, Tourette Association of America, and Centers for Disease Control and Prevention. In addition, she received payment from the Association for Behavioral and Cognitive Therapies for her role as associate editor of the Behavior Therapy journal as well as from John Wiley & Sons, Inc, for her role as associate editor of the Depression & Anxiety journal. SW has also received honoraria from One Mind for her role in the PsyberGuide scientific advisory board. SW is also on the scientific advisory board for Koa Health, Inc, and Noom, Inc. SW has received research and salary support from Koa Health, Inc. In addition, SW has a consulting agreement with Noom, Inc.

References

1. GBD 2019 Diseases Injuries Collaborators. Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet 2020 Oct 17;396(10258):1204-1222 [FREE Full text] [doi: 10.1016/S0140-6736(20)30925-9] [Medline: 33069326]
2. COVID-19 Mental Disorders Collaborators. Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. Lancet 2021 Nov 06;398(10312):1700-1712 [FREE Full text] [doi: 10.1016/S0140-6736(21)02143-7] [Medline: 34634250]
3. Twenge JM, Joiner TE. U.S. Census Bureau-assessed prevalence of anxiety and depressive symptoms in 2019 and during the 2020 COVID-19 pandemic. Depress Anxiety 2020 Oct;37(10):954-956 [FREE Full text] [doi: 10.1002/da.23077] [Medline: 32667081]
4. Thomas KC, Ellis AR, Konrad TR, Holzer CE, Morrissey JP. County-level estimates of mental health professional shortage in the United States. Psychiatr Serv 2009 Oct;60(10):1323-1328. [doi: 10.1176/ps.2009.60.10.1323] [Medline: 19797371]
5. Patients with depression and anxiety surge as psychologists respond to the coronavirus pandemic. American Psychological Association. URL: https://www.apa.org/news/press/releases/2020/11/telehealth-survey-summary.pdf [accessed 2024-01-29]
6. Ussery EN, Fulton JE, Galuska DA, Katzmarzyk PT, Carlson SA. Joint prevalence of sitting time and leisure-time physical activity among US adults, 2015-2016. JAMA 2018 Nov 20;320(19):2036-2038 [FREE Full text] [doi: 10.1001/jama.2018.17797] [Medline: 30458482]

7. Meyer J, Herring M, McDowell C, Lansing J, Brower C, Schuch F, et al. Joint prevalence of physical activity and sitting time during COVID-19 among US adults in April 2020. Prev Med Rep 2020 Dec;20:101256 [FREE Full text] [doi: 10.1016/j.pmedr.2020.101256] [Medline: 33282638]
8. Ku PW, Steptoe A, Liao Y, Hsueh M, Chen LJ. A cut-off of daily sedentary time and all-cause mortality in adults: a meta-regression analysis involving more than 1 million participants. BMC Med 2018 May 25;16(1):74 [FREE Full text] [doi: 10.1186/s12916-018-1062-2] [Medline: 29793552]
9. Tremblay MS, Aubert S, Barnes JD, Saunders TJ, Carson V, Latimer-Cheung AE, SBRN Terminology Consensus Project Participants. Sedentary behavior research network (SBRN) - terminology consensus project process and outcome. Int J Behav Nutr Phys Act 2017 Jun 10;14(1):75 [FREE Full text] [doi: 10.1186/s12966-017-0525-8] [Medline: 28599680]
10. Jiang L, Cao Y, Ni S, Chen X, Shen M, Lv H, et al. Association of sedentary behavior with anxiety, depression, and suicide ideation in college students. Front Psychiatry 2020 Dec 11;11:566098 [FREE Full text] [doi: 10.3389/fpsyt.2020.566098] [Medline: 33424653]
11. Allen MS, Walter EE, Swann C. Sedentary behaviour and risk of anxiety: a systematic review and meta-analysis. J Affect Disord 2019 Jan 01;242:5-13. [doi: 10.1016/j.jad.2018.08.081] [Medline: 30170238]
12. Teychenne M, Costigan SA, Parker K. The association between sedentary behaviour and risk of anxiety: a systematic review. BMC Public Health 2015 Jun 19;15(1):513 [FREE Full text] [doi: 10.1186/s12889-015-1843-x] [Medline: 26088005]
13. Teychenne M, Ball K, Salmon J. Sedentary behavior and depression among adults: a review. Int J Behav Med 2010 Dec;17(4):246-254. [doi: 10.1007/s12529-010-9075-z] [Medline: 20174982]
14. Zhai L, Zhang Y, Zhang D. Sedentary behaviour and the risk of depression: a meta-analysis. Br J Sports Med 2015 Jun 02;49(11):705-709. [doi: 10.1136/bjsports-2014-093613] [Medline: 25183627]
15. Stephens T. Physical activity and mental health in the United States and Canada: evidence from four population surveys. Prev Med 1988 Jan;17(1):35-47. [doi: 10.1016/0091-7435(88)90070-9] [Medline: 3258986]
16. Lee C, Russell A. Effects of physical activity on emotional well-being among older Australian women: cross-sectional and longitudinal analyses. J Psychosom Res 2003 Feb;54(2):155-160. [doi: 10.1016/s0022-3999(02)00414-2] [Medline: 12573737]
17. Penedo FJ, Dahn JR. Exercise and well-being: a review of mental and physical health benefits associated with physical activity. Curr Opin Psychiatry 2005 Mar;18(2):189-193. [doi: 10.1097/00001504-200503000-00013] [Medline: 16639173]
18. Caspersen CJ, Powell KE, Christenson GM. Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. Public Health Rep 1985;100(2):126-131 [FREE Full text] [Medline: 3920711]
19. De Moor MH, Beem AL, Stubbe JH, Boomsma DI, De Geus EJ. Regular exercise, anxiety, depression and personality: a population-based study. Prev Med 2006 Apr;42(4):273-279. [doi: 10.1016/j.ypmed.2005.12.002] [Medline: 16439008]
20. Harris AH, Cronkite R, Moos R. Physical activity, exercise coping, and depression in a 10-year cohort study of depressed patients. J Affect Disord 2006 Jul;93(1-3):79-85. [doi: 10.1016/j.jad.2006.02.013] [Medline: 16545873]
21. Garcia D, Archer T, Moradi S, Andersson-Arntén A. Exercise frequency, high activation positive affect, and psychological well-being: beyond age, gender, and occupation. Psychol 2012;03(04):328-336. [doi: 10.4236/psych.2012.34047]
22. Hassmén P, Koivula N, Uutela A. Physical exercise and psychological well-being: a population study in Finland. Prev Med 2000 Jan;30(1):17-25. [doi: 10.1006/pmed.1999.0597] [Medline: 10642456]
23. Goodwin RD. Association between physical activity and mental disorders among adults in the United States. Prev Med 2003 Jun;36(6):698-703. [doi: 10.1016/S0091-7435(03)00042-2]
24. Paffenbarger Jr RS, Lee IM, Leung R. Physical activity and personal characteristics associated with depression and suicide in American college men. Acta Psychiatr Scand Suppl 1994 Feb;377(s377):16-22. [doi: 10.1111/j.1600-0447.1994.tb05796.x] [Medline: 8053361]
25. Ströhle A, Höfler M, Pfister H, Müller A, Hoyer J, Wittchen H, et al. Physical activity and prevalence and incidence of mental disorders in adolescents and young adults. Psychol Med 2007 Jun 20;37(11):1657-1666. [doi: 10.1017/s003329170700089x]
26. Hyde AL, Conroy DE, Pincus AE, Ram N. Unpacking the feel-good effect of free-time physical activity: between- and within-person associations with pleasant-activated feeling states. J Sport Exerc Psychol 2011 Dec;33(6):884-902 [FREE Full text] [doi: 10.1123/jsep.33.6.884] [Medline: 22262710]
27. Giacobbi PR, Hausenblas HA, Frye N. A naturalistic assessment of the relationship between personality, daily life events, leisure-time exercise, and mood. Psychology of Sport and Exercise 2005 Jan;6(1):67-81. [doi: 10.1016/j.psychsport.2003.10.009]
28. Kanning M, Schlicht W. Be active and become happy: an ecological momentary assessment of physical activity and mood. J Sport Exerc Psychol 2010 Apr;32(2):253-261. [doi: 10.1123/jsep.32.2.253] [Medline: 20479481]
29. Gianfredi V, Blandi L, Cacitti S, Minelli M, Signorelli C, Amerio A, et al. Depression and objectively measured physical activity: a systematic review and meta-analysis. Int J Environ Res Public Health 2020 May 25;17(10):3738 [FREE Full text] [doi: 10.3390/ijerph17103738] [Medline: 32466242]
30. McDowell CP, Dishman RK, Gordon BR, Herring MP. Physical activity and anxiety: a systematic review and meta-analysis of prospective cohort studies. Am J Prev Med 2019 Oct;57(4):545-556. [doi: 10.1016/j.amepre.2019.05.012] [Medline: 31542132]

31. Pearce M, Garcia L, Abbas A, Strain T, Schuch FB, Golubic R, et al. Association between physical activity and risk of depression: a systematic review and meta-analysis. JAMA Psychiatry 2022 Jun 01;79(6):550-559 [FREE Full text] [doi: 10.1001/jamapsychiatry.2022.0609] [Medline: 35416941]
32. Lee IM, Shiroma EJ, Kamada M, Bassett DR, Matthews CE, Buring JE. Association of step volume and intensity with all-cause mortality in older women. JAMA Intern Med 2019 Aug 01;179(8):1105-1112 [FREE Full text] [doi: 10.1001/jamainternmed.2019.0899] [Medline: 31141585]
33. Harvey SB, Øverland S, Hatch SL, Wessely S, Mykletun A, Hotopf M. Exercise and the prevention of depression: results of the HUNT cohort study. Am J Psychiatry 2018 Jan 01;175(1):28-36. [doi: 10.1176/appi.ajp.2017.16111223] [Medline: 28969440]
34. Stathopoulou G, Powers MB, Berry AC, Smits JA, Otto MW. Exercise interventions for mental health: a quantitative and qualitative review. Clin Psychol Sci Pract 2006;13(2):179-193. [doi: 10.1111/j.1468-2850.2006.00021.x]
35. Otto MW, Church TS, Craft LL, Greer TL, Smits JA, Trivedi MH. Exercise for mood and anxiety disorders. Prim Care Companion J Clin Psychiatry 2007 Aug 15;9(4):287-294 [FREE Full text] [doi: 10.4088/pcc.v09n0406] [Medline: 17934553]
36. DiLorenzo TM, Bargman EP, Stucky-Ropp R, Brassington GS, Frensch PA, LaFontaine T. Long-term effects of aerobic exercise on psychological outcomes. Prev Med 1999 Jan;28(1):75-85. [doi: 10.1006/pmed.1998.0385] [Medline: 9973590]
37. Kishida M, Elavsky S. Daily physical activity enhances resilient resources for symptom management in middle-aged women. Health Psychol 2015 Jul;34(7):756-764. [doi: 10.1037/hea0000190] [Medline: 25528176]
38. Flueckiger L, Lieb R, Meyer AH, Witthauer C, Mata J. The importance of physical activity and sleep for affect on stressful days: two intensive longitudinal studies. Emotion 2016 Jun;16(4):488-497. [doi: 10.1037/emo0000143] [Medline: 26709860]
39. Forcier K, Stroud LR, Papandonatos GD, Hitsman B, Reiches M, Krishnamoorthy J, et al. Links between physical fitness and cardiovascular reactivity and recovery to psychological stressors: a meta-analysis. Health Psychol 2006 Nov;25(6):723-739. [doi: 10.1037/0278-6133.25.6.723] [Medline: 17100501]
40. Bernstein EE, Curtiss JE, Wu GW, Barreira PJ, McNally RJ. Exercise and emotion dynamics: an experience sampling study. Emotion 2019 Jun;19(4):637-644. [doi: 10.1037/emo0000462] [Medline: 29999384]
41. Blumenthal JA, Emery CF, Walsh MA, Cox DR, Kuhn CM, Williams RB, et al. Exercise training in healthy type A middle-aged men: effects on behavioral and cardiovascular responses. Psychosom Med 1988;50(4):418-433. [doi: 10.1097/00006842-198807000-00009] [Medline: 3413273]
42. Jackson EM, Dishman RK. Cardiorespiratory fitness and laboratory stress: a meta-regression analysis. Psychophysiology 2006 Jan;43(1):57-72. [doi: 10.1111/j.1469-8986.2006.00373.x] [Medline: 16629686]
43. Rejeski WJ, Thompson A, Brubaker PH, Miller HS. Acute exercise: buffering psychosocial stress responses in women. Health Psychol 1992;11(6):355-362. [doi: 10.1037//0278-6133.11.6.355] [Medline: 1286654]
44. Szuhany KL, Bugatti M, Otto MW. A meta-analytic review of the effects of exercise on brain-derived neurotrophic factor. J Psychiatr Res 2015 Jan;60:56-64 [FREE Full text] [doi: 10.1016/j.jpsychires.2014.10.003] [Medline: 25455510]
45. Bernstein EE, McNally RJ. Acute aerobic exercise helps overcome emotion regulation deficits. Cogn Emot 2017 Jun 04;31(4):834-843. [doi: 10.1080/02699931.2016.1168284] [Medline: 27043051]
46. Bernstein EE, McNally RJ. Acute aerobic exercise hastens emotional recovery from a subsequent stressor. Health Psychol 2017 Jun;36(6):560-567. [doi: 10.1037/hea0000482] [Medline: 28277695]
47. Kline CE. The bidirectional relationship between exercise and sleep: implications for exercise adherence and sleep improvement. Am J Lifestyle Med 2014 Aug 07;8(6):375-379 [FREE Full text] [doi: 10.1177/1559827614544437] [Medline: 25729341]
48. Scott AJ, Webb TL, Martyn-St James M, Rowse G, Weich S. Improving sleep quality leads to better mental health: a meta-analysis of randomised controlled trials. Sleep Med Rev 2021 Dec;60:101556 [FREE Full text] [doi: 10.1016/j.smrv.2021.101556] [Medline: 34607184]
49. Carraça EV, Encantado J, Battista F, Beaulieu K, Blundell JE, Busetto L, et al. Effect of exercise training on psychological outcomes in adults with overweight or obesity: a systematic review and meta-analysis. Obes Rev 2021 Jul 06;22 Suppl 4(Suppl 4):e13261 [FREE Full text] [doi: 10.1111/obr.13261] [Medline: 33960106]
50. Frank E, Wallace ML, Hall M, Hasler B, Levenson JC, Janney CA, et al. An Integrated risk reduction intervention can reduce body mass index in individuals being treated for bipolar I disorder: results from a randomized trial. Bipolar Disord 2015 Jun 12;17(4):424-437 [FREE Full text] [doi: 10.1111/bdi.12283] [Medline: 25495748]
51. van Zanten JJ, Fenton SA, Brady S, Metsios GS, Duda JL, Kitas GD. Mental health and psychological wellbeing in rheumatoid arthritis during COVID-19 - can physical activity help? Mediterr J Rheumatol 2020 Sep;31(Suppl 2):284-287 [FREE Full text] [doi: 10.31138/mjr.31.3.284] [Medline: 33196006]
52. Stubbs B, Vancampfort D, Rosenbaum S, Firth J, Cosco T, Veronese N, et al. An examination of the anxiolytic effects of exercise for people with anxiety and stress-related disorders: a meta-analysis. Psychiatry Res 2017 Mar;249:102-108. [doi: 10.1016/j.psychres.2016.12.020] [Medline: 28088704]
53. Petruzzello SJ, Landers DM, Hatfield BD, Kubitz KA, Salazar W. A meta-analysis on the anxiety-reducing effects of acute and chronic exercise. Outcomes and mechanisms. Sports Med 1991 Mar;11(3):143-182. [doi: 10.2165/00007256-199111030-00002] [Medline: 1828608]

54. Asmundson GJ, Fetzner MG, Deboer LB, Powers MB, Otto MW, Smits JA. Let's get physical: a contemporary review of the anxiolytic effects of exercise for anxiety and its disorders. Depress Anxiety 2013 Apr 08;30(4):362-373. [doi: 10.1002/da.22043] [Medline: 23300122]
55. Wipfli BM, Rethorst CD, Landers DM. The anxiolytic effects of exercise: a meta-analysis of randomized trials and dose-response analysis. J Sport Exerc Psychol 2008 Aug;30(4):392-410. [doi: 10.1123/jsep.30.4.392] [Medline: 18723899]
56. Bailey AP, Hetrick SE, Rosenbaum S, Purcell R, Parker AG. Treating depression with physical activity in adolescents and young adults: a systematic review and meta-analysis of randomised controlled trials. Psychol Med 2017 Oct 10;48(7):1068-1083. [doi: 10.1017/s0033291717002653]
57. Schuch FB, Vancampfort D, Richards J, Rosenbaum S, Ward PB, Stubbs B. Exercise as a treatment for depression: a meta-analysis adjusting for publication bias. J Psychiatr Res 2016 Jun;77:42-51 [FREE Full text] [doi: 10.1016/j.jpsychires.2016.02.023] [Medline: 26978184]
58. Blumenthal JA, Babyak MA, Moore KA, Craighead WE, Herman S, Khatri P, et al. Effects of exercise training on older patients with major depression. Arch Intern Med 1999 Oct 25;159(19):2349-2356. [doi: 10.1001/archinte.159.19.2349] [Medline: 10547175]
59. Fetzner MG, Asmundson GJ. Aerobic exercise reduces symptoms of posttraumatic stress disorder: a randomized controlled trial. Cogn Behav Ther 2015;44(4):301-313. [doi: 10.1080/16506073.2014.916745] [Medline: 24911173]
60. Abrantes AM, Farris SG, Brown RA, Greenberg BD, Strong DR, McLaughlin NC, et al. Acute effects of aerobic exercise on negative affect and obsessions and compulsions in individuals with obsessive-compulsive disorder. J Affect Disord 2019 Feb 15;245:991-997 [FREE Full text] [doi: 10.1016/j.jad.2018.11.074] [Medline: 30699885]
61. Brown RA, Abrantes AM, Strong DR, Mancebo MC, Menard J, Rasmussen SA, et al. A pilot study of moderate-intensity aerobic exercise for obsessive compulsive disorder. J Nerv Ment Dis 2007 Jun;195(6):514-520. [doi: 10.1097/01.nmd.0000253730.31610.6c] [Medline: 17568300]
62. Rosenbaum S, Tiedemann A, Sherrington C, Curtis J, Ward PB. Physical activity interventions for people with mental illness: a systematic review and meta-analysis. J Clin Psychiatry 2014 Sep;75(9):964-974. [doi: 10.4088/JCP.13r08765] [Medline: 24813261]
63. Janney CA, Bauer MS, Kilbourne AM. Self-management and bipolar disorder--a clinician's guide to the literature 2011-2014. Curr Psychiatry Rep 2014 Sep 16;16(9):485. [doi: 10.1007/s11920-014-0485-5] [Medline: 25123130]
64. Rector NA, Richter MA, Lerman B, Regev R. A pilot test of the additive benefits of physical exercise to CBT for OCD. Cogn Behav Ther 2015 Mar 04;44(4):328-340. [doi: 10.1080/16506073.2015.1016448] [Medline: 25738234]
65. Powers MB, Medina JL, Burns S, Kauffman BY, Monfils M, Asmundson GJ, et al. Exercise augmentation of exposure therapy for PTSD: rationale and pilot efficacy data. Cogn Behav Ther 2015;44(4):314-327 [FREE Full text] [doi: 10.1080/16506073.2015.1012740] [Medline: 25706090]
66. Bourbeau K, Moriarty T, Ayanniyi A, Zuhl M. The combined effect of exercise and behavioral therapy for depression and anxiety: systematic review and meta-analysis. Behav Sci (Basel) 2020 Jul 14;10(7):116 [FREE Full text] [doi: 10.3390/bs10070116] [Medline: 32674359]
67. Frederiksen KP, Stavestrand SH, Venemyr SK, Sirevåg K, Hovland A. Physical exercise as an add-on treatment to cognitive behavioural therapy for anxiety: a systematic review. Behav Cogn Psychother 2021 Sep;49(5):626-640. [doi: 10.1017/S1352465821000126] [Medline: 33678210]
68. Whitfield GP, Hyde ET, Carlson SA. Participation in leisure-time aerobic physical activity among adults, national health interview survey, 1998-2018. J Phys Act Health 2021 Aug 01;18(S1):S25-S36 [FREE Full text] [doi: 10.1123/jpah.2021-0014] [Medline: 34465654]
69. McDowell CP, Dishman RK, Vancampfort D, Hallgren M, Stubbs B, MacDonncha C, et al. Physical activity and generalized anxiety disorder: results from The Irish Longitudinal Study on Ageing (TILDA). Int J Epidemiol 2018 Oct 01;47(5):1443-1453. [doi: 10.1093/ije/dyy141] [Medline: 29982489]
70. Janney CA, Brzoznowski KF, Richardson CR, Dopp RR, Segar ML, Ganoczy D, et al. Moving towards wellness: physical activity practices, perspectives, and preferences of users of outpatient mental health service. Gen Hosp Psychiatry 2017 Nov;49:63-66. [doi: 10.1016/j.genhosppsych.2017.07.004] [Medline: 29122150]
71. Vancampfort D, Stubbs B, Herring MP, Hallgren M, Koyanagi A. Sedentary behavior and anxiety: association and influential factors among 42,469 community-dwelling adults in six low- and middle-income countries. General Hospital Psychiatry 2018 Jan;50:26-32. [doi: 10.1016/j.genhosppsych.2017.09.006] [Medline: 28987919]
72. Stubbs B, Vancampfort D, Firth J, Schuch FB, Hallgren M, Smith L, et al. Relationship between sedentary behavior and depression: a mediation analysis of influential factors across the lifespan among 42,469 people in low- and middle-income countries. J Affect Disord 2018 Mar 15;229:231-238 [FREE Full text] [doi: 10.1016/j.jad.2017.12.104] [Medline: 29329054]
73. Seime RJ, Vickers KS. The challenges of treating depression with exercise: from evidence to practice. Clin Psychol Sci Pract 2001 Nov;13:194-197. [doi: 10.1002/jclp.1097] [Medline: 11590615]
74. Stanton R, Franck C, Reaburn P, Happell B. A pilot study of the views of general practitioners regarding exercise for the treatment of depression. Perspect Psychiatr Care 2015 Oct 13;51(4):253-259. [doi: 10.1111/ppc.12088] [Medline: 25307254]
75. Pollock KM. Exercise in treating depression: broadening the psychotherapist's role. J Clin Psychol 2001 Nov;57(11):1289-1300. [doi: 10.1002/jclp.1097] [Medline: 11590615]

76. Andrade LH, Alonso J, Mneimneh Z, Wells JE, Al-Hamzawi A, Borges G, et al. Barriers to mental health treatment: results from the WHO World Mental Health surveys. Psychol Med 2013 Aug 09;44(6):1303-1317. [doi: 10.1017/s0033291713001943]
77. Boisseau CL, Schwartzman CM, Lawton J, Mancebo MC. App-guided exposure and response prevention for obsessive compulsive disorder: an open pilot trial. Cogn Behav Ther 2017 Nov 05;46(6):447-458. [doi: 10.1080/16506073.2017.1321683] [Medline: 28565937]
78. Lee EB, Hoepfl C, Werner C, McIngvale E. A review of tech-based self-help treatment programs for obsessive-compulsive disorder. J Obsessive Compuls Relat Disord 2019 Oct;23:100473. [doi: 10.1016/j.jocrd.2019.100473]
79. Ramos G, Chavira DA. Use of technology to provide mental health care for racial and ethnic minorities: evidence, promise, and challenges. Cogn Behav Pract 2022 Feb;29(1):15-40. [doi: 10.1016/j.cbpra.2019.10.004]
80. Manini TM, Carr LJ, King AC, Marshall S, Robinson T, Rejeski W. Interventions to reduce sedentary behavior. Med Sci Sports Exerc 2015 Jun;47(6):1306-1310 [FREE Full text] [doi: 10.1249/MSS.0000000000000519] [Medline: 25222818]
81. Ward DS, Evenson KR, Vaughn A, Rodgers AB, Troiano RP. Accelerometer use in physical activity: best practices and research recommendations. Med Sci Sports Exerc 2005 Nov;37(11 Suppl):S582-S588. [doi: 10.1249/01.mss.0000185292.71933.91] [Medline: 16294121]
82. Boudet G, Chausse P, Thivel D, Rousset S, Mermillod M, Baker JS, et al. How to measure sedentary behavior at work? Front Public Health 2019 Jul 5;7:167 [FREE Full text] [doi: 10.3389/fpubh.2019.00167] [Medline: 31355172]
83. Germini F, Noronha N, Borg Debono V, Abraham Philip B, Pete D, Navarro T, et al. Accuracy and acceptability of wrist-wearable activity-tracking devices: systematic review of the literature. J Med Internet Res 2022 Jan 21;24(1):e30791 [FREE Full text] [doi: 10.2196/30791] [Medline: 35060915]
84. Ueno DT, Guerra PH, Christofoletti AE, Bonolo A, Nakamura PM, Kokubun E. Mobile health apps to reduce sedentary behavior: a scoping review. Health Promot Int 2022 Apr 29;37(2):daab124. [doi: 10.1093/heapro/daab124] [Medline: 34392354]
85. Stephenson A, McDonough SM, Murphy MH, Nugent CD, Mair JL. Using computer, mobile and wearable technology enhanced interventions to reduce sedentary behaviour: a systematic review and meta-analysis. Int J Behav Nutr Phys Act 2017 Aug 11;14(1):105 [FREE Full text] [doi: 10.1186/s12966-017-0561-4] [Medline: 28800736]
86. Kim HN, Seo K. Smartphone-based health program for improving physical activity and tackling obesity for young adults: a systematic review and meta-analysis. Int J Environ Res Public Health 2019 Dec 18;17(1):15 [FREE Full text] [doi: 10.3390/ijerph17010015] [Medline: 31861359]
87. He Z, Wu H, Yu F, Fu J, Sun S, Huang T, et al. Effects of smartphone-based interventions on physical activity in children and adolescents: systematic review and meta-analysis. JMIR Mhealth Uhealth 2021 Feb 01;9(2):e22601 [FREE Full text] [doi: 10.2196/22601] [Medline: 33522980]
88. Han M, Lee E. Effectiveness of mobile health application use to improve health behavior changes: a systematic review of randomized controlled trials. Healthc Inform Res 2018 Jul;24(3):207-226 [FREE Full text] [doi: 10.4258/hir.2018.24.3.207] [Medline: 30109154]
89. Litman L, Rosen Z, Spierer D, Weinberger-Litman S, Goldschein A, Robinson J. Mobile exercise apps and increased leisure time exercise activity: a moderated mediation analysis of the role of self-efficacy and barriers. J Med Internet Res 2015 Aug 14;17(8):e195 [FREE Full text] [doi: 10.2196/jmir.4142] [Medline: 26276227]
90. Perrin A. Mobile technology and home broadband 2021. Pew Research Center. 2021. URL: https://www.pewresearch.org/internet/2021/06/03/mobile-technology-and-home-broadband-2021/ [accessed 2024-02-20]
91. Compernolle S, DeSmet A, Poppe L, Crombez G, De Bourdeaudhuij I, Cardon G, et al. Effectiveness of interventions using self-monitoring to reduce sedentary behavior in adults: a systematic review and meta-analysis. Int J Behav Nutr Phys Act 2019 Aug 13;16(1):63 [FREE Full text] [doi: 10.1186/s12966-019-0824-3] [Medline: 31409357]
92. Young AS, Cohen AN, Niv N, Nowlin-Finch N, Oberman RS, Olmos-Ochoa TT, et al. Mobile phone and smartphone use by people with serious mental illness. Psychiatr Serv 2020 Mar 01;71(3):280-283 [FREE Full text] [doi: 10.1176/appi.ps.201900203] [Medline: 31744429]
93. Ho'gelen EI, Akdede BB, Alptekin K. M101. prevelance use of technological devices and internet among patients diagnosed with schizophrenia and schizoaffective disorder. Schizophr Bull 2020;46(Suppl 1):S173. [doi: 10.1093/schbul/sbaa030.413]
94. Torous J, Wisniewski H, Liu G, Keshavan M. Mental health mobile phone app usage, concerns, and benefits among psychiatric outpatients: comparative survey study. JMIR Ment Health 2018 Nov 16;5(4):e11715 [FREE Full text] [doi: 10.2196/11715] [Medline: 30446484]
95. Schembre SM, Liao Y, Robertson MC, Dunton GF, Kerr J, Haffey ME, et al. Just-in-time feedback in diet and physical activity interventions: systematic review and practical design framework. J Med Internet Res 2018 Mar 22;20(3):e106 [FREE Full text] [doi: 10.2196/jmir.8701] [Medline: 29567638]
96. Smyth JM, Sliwinski MJ, Zawadzki MJ, Scott SB, Conroy DE, Lanza ST, et al. Everyday stress response targets in the science of behavior change. Behav Res Ther 2018 Feb;101:20-29 [FREE Full text] [doi: 10.1016/j.brat.2017.09.009] [Medline: 29031538]

97. Mazeas A, Duclos M, Pereira B, Chalabaev A. Evaluating the effectiveness of gamification on physical activity: systematic review and meta-analysis of randomized controlled trials. J Med Internet Res 2022 Jan 04;24(1):e26779 [FREE Full text] [doi: 10.2196/26779] [Medline: 34982715]
98. Cotton V, Patel MS. Gamification use and design in popular health and fitness mobile applications. Am J Health Promot 2019 Mar;33(3):448-451 [FREE Full text] [doi: 10.1177/0890117118790394] [Medline: 30049225]
99. Aguilera A, Figueroa CA, Hernandez-Ramos R, Sarkar U, Cemballi A, Gomez-Pathak L, et al. mHealth app using machine learning to increase physical activity in diabetes and depression: clinical trial protocol for the DIAMANTE Study. BMJ Open 2020 Aug 20;10(8):e034723 [FREE Full text] [doi: 10.1136/bmjopen-2019-034723] [Medline: 32819981]
100. Broers ER, Gavidia G, Wetzels M, Ribas V, Ayoola I, Piera-Jimenez J, et al. Usefulness of a lifestyle intervention in patients with cardiovascular disease. Am J Cardiol 2020 Feb 01;125(3):370-375 [FREE Full text] [doi: 10.1016/j.amjcard.2019.10.041] [Medline: 31761149]
101. Damschroder LJ, Buis LR, McCant FA, Kim HM, Evans R, Oddone EZ, et al. Effect of adding telephone-based brief coaching to an mHealth app (stay strong) for promoting physical activity among veterans: randomized controlled trial. J Med Internet Res 2020 Aug 04;22(8):e19216 [FREE Full text] [doi: 10.2196/19216] [Medline: 32687474]
102. Edney SM, Olds TS, Ryan JC, Vandelanotte C, Plotnikoff RC, Curtis RG, et al. A social networking and gamified app to increase physical activity: cluster RCT. Am J Prev Med 2020 Feb;58(2):e51-e62. [doi: 10.1016/j.amepre.2019.09.009] [Medline: 31959326]
103. García-Estela A, Angarita-Osorio N, Alonso S, Polo M, Roldán-Berengué M, Messaggi-Sartor M, et al. Improving depressive symptoms through personalised exercise and activation (IDEA): study protocol for a randomised controlled trial. Int J Environ Res Public Health 2021 Jun 10;18(12):6306 [FREE Full text] [doi: 10.3390/ijerph18126306] [Medline: 34200805]
104. Guo Y, Hong YA, Cai W, Li L, Hao Y, Qiao J, et al. Effect of a WeChat-based intervention (Run4Love) on depressive symptoms among people living with HIV in China: a randomized controlled trial. J Med Internet Res 2020 Feb 09;22(2):e16715 [FREE Full text] [doi: 10.2196/16715] [Medline: 32044751]
105. Haufe S, Kahl KG, Kerling A, Protte G, Bayerle P, Stenner HT, et al. Employers with metabolic syndrome and increased depression/anxiety severity profit most from structured exercise intervention for work ability and quality of life. Front Psychiatry 2020 Jun 18;11:562 [FREE Full text] [doi: 10.3389/fpsyt.2020.00562] [Medline: 32625123]
106. Kim A, Yun SJ, Sung KS, Kim Y, Jo JY, Cho H, et al. Exercise management using a mobile app in patients with parkinsonism: prospective, open-label, single-arm pilot study. JMIR Mhealth Uhealth 2021 Aug 31;9(8):e27662 [FREE Full text] [doi: 10.2196/27662] [Medline: 34463635]
107. Lin J, Wurst R, Paganini S, Hohberg V, Kinkel S, Göhner W, et al. A group- and smartphone-based psychological intervention to increase and maintain physical activity in patients with musculoskeletal conditions: study protocol for a randomized controlled trial ("MoVo-App"). Trials 2020 Jun 08;21(1):502 [FREE Full text] [doi: 10.1186/s13063-020-04438-4] [Medline: 32513291]
108. Ma J, Yank V, Lv N, Goldhaber-Fiebert JD, Lewis MA, Kramer MK, et al. Research aimed at improving both mood and weight (RAINBOW) in primary care: a type 1 hybrid design randomized controlled trial. Contemp Clin Trials 2015 Jul;43:260-278 [FREE Full text] [doi: 10.1016/j.cct.2015.06.010] [Medline: 26096714]
109. Nadal C, Earley C, Enrique A, Vigano N, Sas C, Richards D, et al. Integration of a smartwatch within an internet-delivered intervention for depression: protocol for a feasibility randomized controlled trial on acceptance. Contemp Clin Trials 2021 Apr;103:106323 [FREE Full text] [doi: 10.1016/j.cct.2021.106323] [Medline: 33621632]
110. Park LG, Elnaggar A, Lee SJ, Merek S, Hoffmann TJ, Von Oppenfeld J, et al. Mobile health intervention promoting physical activity in adults post cardiac rehabilitation: pilot randomized controlled trial. JMIR Form Res 2021 Apr 16;5(4):e20468 [FREE Full text] [doi: 10.2196/20468] [Medline: 33861204]
111. Puszkiewicz P, Roberts AL, Smith L, Wardle J, Fisher A. Assessment of cancer survivors' experiences of using a publicly available physical activity mobile application. JMIR Cancer 2016 May 31;2(1):e7 [FREE Full text] [doi: 10.2196/cancer.5380] [Medline: 28410168]
112. Puterman E, Hives B, Mazara N, Grishin N, Webster J, Hutton S, et al. COVID-19 Pandemic and Exercise (COPE) trial: a multigroup pragmatic randomised controlled trial examining effects of app-based at-home exercise programs on depressive symptoms. Br J Sports Med 2022 May;56(10):546-552 [FREE Full text] [doi: 10.1136/bjsports-2021-104379] [Medline: 34580067]
113. Skrepnik N, Spitzer A, Altman R, Hoekstra J, Stewart J, Toselli R. Assessing the impact of a novel smartphone application compared with standard follow-up on mobility of patients with knee osteoarthritis following treatment with Hylan G-F 20: a randomized controlled trial. JMIR Mhealth Uhealth 2017 May 09;5(5):e64 [FREE Full text] [doi: 10.2196/mhealth.7179] [Medline: 28487266]
114. Stephens S, Schneiderman JE, Finlayson M, Berenbaum T, Motl RW, Yeh EA. Feasibility of a theory-informed mobile app for changing physical activity in youth with multiple sclerosis. Mult Scler Relat Disord 2022 Feb;58:103467. [doi: 10.1016/j.msard.2021.103467] [Medline: 34954651]
115. Teychenne M, Abbott G, Stephens LD, Opie RS, Olander EK, Brennan L, et al. Mums on the move: a pilot randomised controlled trial of a home-based physical activity intervention for mothers at risk of postnatal depression. Midwifery 2021 Feb;93:102898. [doi: 10.1016/j.midw.2020.102898] [Medline: 33290891]

116. Wilczynska M, Lubans DR, Plotnikoff RC. The effects of the eCoFit RCT on depression and anxiety symptoms among adults with or at risk of type 2 diabetes. Psychol Health Med 2022 Aug 16;27(7):1421-1430. [doi: 10.1080/13548506.2021.1883684] [Medline: 33726568]
117. Wong VW, Ho FY, Shi NK, Tong JT, Chung KF, Yeung WF, et al. Smartphone-delivered multicomponent lifestyle medicine intervention for depressive symptoms: a randomized controlled trial. J Consult Clin Psychol 2021 Dec;89(12):970-984. [doi: 10.1037/ccp0000695] [Medline: 35025538]
118. Roberts AL, Fisher A, Smith L, Heinrich M, Potts HW. Digital health behaviour change interventions targeting physical activity and diet in cancer survivors: a systematic review and meta-analysis. J Cancer Surviv 2017 Dec;11(6):704-719 [FREE Full text] [doi: 10.1007/s11764-017-0632-1] [Medline: 28779220]
119. Marques A, Santos T, Martins J, Matos MG, Valeiro MG. The association between physical activity and chronic diseases in European adults. Eur J Sport Sci 2018 Feb 14;18(1):140-149 [FREE Full text] [doi: 10.1080/17461391.2017.1400109] [Medline: 29134857]
120. Arietaleanizbeaskoa MS, Sancho A, Olazabal I, Moreno C, Gil E, Garcia-Alvarez A, EfiKroniK group. Effectiveness of physical exercise for people with chronic diseases: the EFIKRONIK study protocol for a hybrid, clinical and implementation randomized trial. BMC Fam Pract 2020 Nov 06;21(1):227 [FREE Full text] [doi: 10.1186/s12875-020-01298-4] [Medline: 33158422]
121. Pedersen BK, Saltin B. Exercise as medicine - evidence for prescribing exercise as therapy in 26 different chronic diseases. Scand J Med Sci Sports 2015 Dec 25;25 Suppl 3:1-72. [doi: 10.1111/sms.12581] [Medline: 26606383]
122. Lascar N, Kennedy A, Hancock B, Jenkins D, Andrews RC, Greenfield S, et al. Attitudes and barriers to exercise in adults with type 1 diabetes (T1DM) and how best to address them: a qualitative study. PLoS One 2014;9(9):e108019 [FREE Full text] [doi: 10.1371/journal.pone.0108019] [Medline: 25237905]
123. Anderson E, Durstine JL. Physical activity, exercise, and chronic diseases: a brief review. Sports Med Health Sci 2019 Dec;1(1):3-10 [FREE Full text] [doi: 10.1016/j.smhs.2019.08.006] [Medline: 35782456]
124. Fleming T, Bavin L, Lucassen M, Stasiak K, Hopkins S, Merry S. Beyond the trial: systematic review of real-world uptake and engagement with digital self-help interventions for depression, low mood, or anxiety. J Med Internet Res 2018 Jun 06;20(6):e199 [FREE Full text] [doi: 10.2196/jmir.9275] [Medline: 29875089]
125. Yang CH, Maher JP, Conroy DE. Implementation of behavior change techniques in mobile applications for physical activity. Am J Prev Med 2015 Apr;48(4):452-455. [doi: 10.1016/j.amepre.2014.10.010] [Medline: 25576494]
126. Carter DD, Robinson K, Forbes J, Hayes S. Experiences of mobile health in promoting physical activity: a qualitative systematic review and meta-ethnography. PLoS One 2018;13(12):e0208759 [FREE Full text] [doi: 10.1371/journal.pone.0208759] [Medline: 30557396]
127. Thraen-Borowski KM, Ellingson LD, Meyer JD, Cadmus-Bertram L. Nonworksite interventions to reduce sedentary behavior among adults: a systematic review. Transl J Am Coll Sports Med 2017 Jun 15;2(12):68-78 [FREE Full text] [doi: 10.1249/TJX.0000000000000036] [Medline: 28993817]
128. Prochaska JO, DiClemente CC. Stages and processes of self-change of smoking: toward an integrative model of change. J Consult Clin Psychol 1983 Jun;51(3):390-395. [doi: 10.1037//0022-006x.51.3.390] [Medline: 6863699]
129. Ajzen I. The theory of planned behavior. Organ Behav Hum Decis Process 1991:179-211 [FREE Full text] [doi: 10.1016/0749-5978(91)90020-t]
130. Deci EL, Ryan EM. Intrinsic Motivation and Self-Determination in Human Behavior. New York, NY: Plenum Press; 1985.
131. Bandura A. Social Foundations of Thought and Action: A Social Cognitive Theory. Englewood Cliffs, NJ: Prentice-Hall, Inc; 1986.
132. Chase JD, Otmanowski J, Rowland S, Cooper PS. A systematic review and meta-analysis of interventions to reduce sedentary behavior among older adults. Transl Behav Med 2020 Oct 12;10(5):1078-1085. [doi: 10.1093/tbm/ibz189] [Medline: 33044538]
133. Ludwig K, Arthur R, Sculthorpe N, Fountain H, Buchan DS. Text messaging interventions for improvement in physical activity and sedentary behavior in youth: systematic review. JMIR Mhealth Uhealth 2018 Sep 17;6(9):e10799 [FREE Full text] [doi: 10.2196/10799] [Medline: 30224335]
134. Stockwell S, Schofield P, Fisher A, Firth J, Jackson SE, Stubbs B, et al. Digital behavior change interventions to promote physical activity and/or reduce sedentary behavior in older adults: a systematic review and meta-analysis. Exp Gerontol 2019 Jun;120:68-87. [doi: 10.1016/j.exger.2019.02.020] [Medline: 30836130]
135. McNally RF, Reese HE. Information-processing approaches to understanding anxiety disorders. In: Antony MM, Stein MB, editors. Oxford Handbook of Anxiety and Related Disorders. Oxford, UK: Oxford University Press; 2009:136-152.
136. Pan Z, Park C, Brietzke E, Zuckerman H, Rong C, Mansur RB, et al. Cognitive impairment in major depressive disorder. CNS Spectr 2019 Feb;24(1):22-29. [doi: 10.1017/S1092852918001207] [Medline: 30468135]
137. Berggren N, Derakshan N. Attentional control deficits in trait anxiety: why you see them and why you don't. Biol Psychol 2013 Mar;92(3):440-446. [doi: 10.1016/j.biopsycho.2012.03.007] [Medline: 22465045]
138. McIntyre RS, Cha DS, Soczynska JK, Woldeyohannes HO, Gallaugher LA, Kudlow P, et al. Cognitive deficits and functional outcomes in major depressive disorder: determinants, substrates, and treatment interventions. Depress Anxiety 2013 Jun 06;30(6):515-527. [doi: 10.1002/da.22063] [Medline: 23468126]

139. Looyestyn J, Kernot J, Boshoff K, Ryan J, Edney S, Maher C. Does gamification increase engagement with online programs? A systematic review. PLoS One 2017;12(3):e0173403 [FREE Full text] [doi: 10.1371/journal.pone.0173403] [Medline: 28362821]
140. Yang Y, Koenigstorfer J. Determinants of fitness app usage and moderating impacts of education-, motivation-, and gamification-related app features on physical activity intentions: cross-sectional survey study. J Med Internet Res 2021 Jul 13;23(7):e26063 [FREE Full text] [doi: 10.2196/26063] [Medline: 34255656]
141. Admon R, Pizzagalli DA. Dysfunctional reward processing in depression. Curr Opin Psychol 2015 Aug 01;4:114-118 [FREE Full text] [doi: 10.1016/j.copsyc.2014.12.011] [Medline: 26258159]
142. Smits JA, Tart CD, Presnell K, Rosenfield D, Otto MW. Identifying potential barriers to physical activity adherence: anxiety sensitivity and body mass as predictors of fear during exercise. Cogn Behav Ther 2010 Jan;39(1):28-36. [doi: 10.1080/16506070902915261] [Medline: 19675961]
143. Mason J, Faller Y, LeBouthillier D, Asmundson G. Exercise anxiety: a qualitative analysis of the barriers, facilitators, and psychological processes underlying exercise participation for people with anxiety-related disorders. Ment Health Phys Act 2019 Mar;16:128-139. [doi: 10.1016/j.mhpa.2018.11.003]
144. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95 [FREE Full text] [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
145. Heath GW, Parra DC, Sarmiento OL, Andersen LB, Owen N, Goenka S, Lancet Physical Activity Series Working Group. Evidence-based intervention in physical activity: lessons from around the world. Lancet 2012 Jul 21;380(9838):272-281 [FREE Full text] [doi: 10.1016/S0140-6736(12)60816-2] [Medline: 22818939]
146. Lewis BA, Napolitano MA, Buman MP, Williams DM, Nigg CR. Future directions in physical activity intervention research: expanding our focus to sedentary behaviors, technology, and dissemination. J Behav Med 2017 Feb;40(1):112-126 [FREE Full text] [doi: 10.1007/s10865-016-9797-8] [Medline: 27722907]
147. Anderson PJ, Bovard RS, Wang Z, Beebe TJ, Murad MH. A survey of social support for exercise and its relationship to health behaviours and health status among endurance Nordic skiers. BMJ Open 2016 Jun 23;6(6):e010259 [FREE Full text] [doi: 10.1136/bmjopen-2015-010259] [Medline: 27338876]
148. Middelweerd A, Te Velde SJ, Abbott G, Timperio A, Brug J, Ball K. Do intrapersonal factors mediate the association of social support with physical activity in young women living in socioeconomically disadvantaged neighbourhoods? A longitudinal mediation analysis. PLoS One 2017 Mar 16;12(3):e0173231 [FREE Full text] [doi: 10.1371/journal.pone.0173231] [Medline: 28301538]
149. Duff OM, Walsh DM, Furlong BA, O'Connor NE, Moran KA, Woods CB. Behavior change techniques in physical activity eHealth interventions for people with cardiovascular disease: systematic review. J Med Internet Res 2017 Aug 02;19(8):e281 [FREE Full text] [doi: 10.2196/jmir.7782] [Medline: 28768610]
150. Samdal GB, Eide GE, Barth T, Williams G, Meland E. Effective behaviour change techniques for physical activity and healthy eating in overweight and obese adults; systematic review and meta-regression analyses. Int J Behav Nutr Phys Act 2017 Mar 28;14(1):42 [FREE Full text] [doi: 10.1186/s12966-017-0494-y] [Medline: 28351367]
151. Bakker D, Kazantzis N, Rickwood D, Rickard N. Mental health smartphone apps: review and evidence-based recommendations for future developments. JMIR Ment Health 2016 Mar 01;3(1):e7 [FREE Full text] [doi: 10.2196/mental.4984] [Medline: 26932350]
152. Guerrero-Jiménez M, Ruiz M, Gutiérrez-Rojas L, Jiménez-Muñoz L, Baca-Garcia E, Porras-Segovia A. Use of new technologies for the promotion of physical activity in patients with mental illness: a systematic review. World J Psychiatry 2023 Apr 19;13(4):182-190 [FREE Full text] [doi: 10.5498/wjp.v13.i4.182] [Medline: 37123096]
153. Horvath P. Treatment expectancy as a function of the amount of information presented in therapeutic rationales. J Clin Psychol 1990 Sep;46(5):636-642. [doi: 10.1002/1097-4679(199009)46:5<636::aid-jclp2270460516>3.0.co;2-u] [Medline: 2246373]
154. Schofield CA, Ponzini GT, Becker SJ. Evaluating approaches to marketing cognitive behavioral therapy: does evidence matter to consumers? Cogn Behav Ther 2020 Jul;49(4):257-269 [FREE Full text] [doi: 10.1080/16506073.2019.1682654] [Medline: 31922456]
155. Nyström CD, Sandin S, Henriksson P, Henriksson H, Trolle-Lagerros Y, Larsson C, et al. Mobile-based intervention intended to stop obesity in preschool-aged children: the MINISTOP randomized controlled trial. Am J Clin Nutr 2017 Jun;105(6):1327-1335 [FREE Full text] [doi: 10.3945/ajcn.116.150995] [Medline: 28446496]
156. Hearon BA, Beard C, Kopeski LM, Smits JA, Otto MW, Björgvinsson T. Attending to timely contingencies: promoting physical activity uptake among adults with serious mental illness with an exercise-for-mood vs. an exercise-for-fitness prescription. Behav Med 2018 Feb 21;44(2):108-115. [doi: 10.1080/08964289.2016.1276428] [Medline: 28027010]
157. McNally RJ. Anxiety sensitivity and panic disorder. Biol Psychiatry 2002 Nov 15;52(10):938-946. [doi: 10.1016/s0006-3223(02)01475-0] [Medline: 12437935]

158. Wallman-Jones A, Perakakis P, Tsakiris M, Schmidt M. Physical activity and interoceptive processing: theoretical considerations for future research. Int J Psychophysiol 2021 Aug;166:38-49 [FREE Full text] [doi: 10.1016/j.ijpsycho.2021.05.002] [Medline: 33965423]
159. van Dis EA, van Veen SC, Hagenaars MA, Batelaan NM, Bockting CL, van den Heuvel RM, et al. Long-term outcomes of cognitive behavioral therapy for anxiety-related disorders: a systematic review and meta-analysis. JAMA Psychiatry 2020 Mar 01;77(3):265-273 [FREE Full text] [doi: 10.1001/jamapsychiatry.2019.3986] [Medline: 31758858]
160. Kabat-Zinn J. Wherever You Go, There You Are: Mindfulness Meditation in Everyday Life. New York, NY: Hyperion; 1994.
161. Schreiner I, Malcolm JP. The benefits of mindfulness meditation: changes in emotional states of depression, anxiety, and stress. Behav change 2012 Feb 22;25(3):156-168 [FREE Full text] [doi: 10.1375/bech.25.3.156]
162. Remmers C, Topolinski S, Koole SL. Why being mindful may have more benefits than you realize: mindfulness improves both explicit and implicit mood regulation. Mindfulness 2016 Apr 5;7(4):829-837 [FREE Full text] [doi: 10.1007/s12671-016-0520-1] [Medline: 28269319]
163. Ribeiro L, Atchley RM, Oken BS. Adherence to practice of mindfulness in novice meditators: practices chosen, amount of time practiced, and long-term effects following a mindfulness-based intervention. Mindfulness (N Y) 2018 Apr;9(2):401-411 [FREE Full text] [doi: 10.1007/s12671-017-0781-3] [Medline: 30881517]
164. Firth J, Torous J, Nicholas J, Carney R, Pratap A, Rosenbaum S, et al. The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. World Psychiatry 2017 Oct;16(3):287-298 [FREE Full text] [doi: 10.1002/wps.20472] [Medline: 28941113]
165. Wilhelm S, Weingarden H, Greenberg JL, McCoy TH, Ladis I, Summers BJ, et al. Development and pilot testing of a cognitive-behavioral therapy digital service for body dysmorphic disorder. Behav Ther 2020 Jan;51(1):15-26 [FREE Full text] [doi: 10.1016/j.beth.2019.03.007] [Medline: 32005333]
166. Kambeitz-Ilankovic L, Rzayeva U, Völkel L, Wenzel J, Weiske J, Jessen F, et al. A systematic review of digital and face-to-face cognitive behavioral therapy for depression. NPJ Digit Med 2022 Sep 15;5(1):144 [FREE Full text] [doi: 10.1038/s41746-022-00677-8] [Medline: 36109583]
167. Merom D, Phongsavan P, Wagner R, Chey T, Marnane C, Steel Z, et al. Promoting walking as an adjunct intervention to group cognitive behavioral therapy for anxiety disorders--a pilot group randomized trial. J Anxiety Disord 2008 Aug;22(6):959-968. [doi: 10.1016/j.janxdis.2007.09.010] [Medline: 17988832]
168. Kaushal N, Rhodes RE. Exercise habit formation in new gym members: a longitudinal study. J Behav Med 2015 Aug;38(4):652-663. [doi: 10.1007/s10865-015-9640-7] [Medline: 25851609]
169. Landais LL, Damman OC, Schoonmade LJ, Timmermans DR, Verhagen EA, Jelsma JG. Choice architecture interventions to change physical activity and sedentary behavior: a systematic review of effects on intention, behavior and health outcomes during and after intervention. Int J Behav Nutr Phys Act 2020 Apr 07;17(1):47 [FREE Full text] [doi: 10.1186/s12966-020-00942-7] [Medline: 32264899]
170. Vandelanotte C, Spathonis KM, Eakin EG, Owen N. Website-delivered physical activity interventions a review of the literature. Am J Prev Med 2007 Jul;33(1):54-64. [doi: 10.1016/j.amepre.2007.02.041] [Medline: 17572313]
171. Conn VS. Anxiety outcomes after physical activity interventions: meta-analysis findings. Nurs Res 2010;59(3):224-231 [FREE Full text] [doi: 10.1097/NNR.0b013e3181dbb2f8] [Medline: 20410849]
172. Benjet C, Zainal NH, Albor Y, Alvis-Barranco L, Carrasco-Tapias N, Contreras-Ibáñez CC, et al. A precision treatment model for internet-delivered cognitive behavioral therapy for anxiety and depression among university students: a secondary analysis of a randomized clinical trial. JAMA Psychiatry 2023 Aug 01;80(8):768-777. [doi: 10.1001/jamapsychiatry.2023.1675] [Medline: 37285133]
173. Bernstein EE, Weingarden H, Wolfe EC, Hall MD, Snorrason I, Wilhelm S. Human support in app-based cognitive behavioral therapies for emotional disorders: scoping review. J Med Internet Res 2022 Apr 08;24(4):e33307 [FREE Full text] [doi: 10.2196/33307] [Medline: 35394434]
174. van Rijen D, Ten Hoor GA. A qualitative analysis of facilitators and barriers to physical activity among patients with moderate mental disorders. Z Gesundh Wiss 2022 Jun 01:1-16 [FREE Full text] [doi: 10.1007/s10389-022-01720-4] [Medline: 35668718]
175. Searle A, Calnan M, Lewis G, Campbell J, Taylor A, Turner K. Patients' views of physical activity as treatment for depression: a qualitative study. Br J Gen Pract 2011 Apr 01;61(585):e149-e156. [doi: 10.3399/bjgp11x567054]
176. Borghouts J, Eikey E, Mark G, De Leon C, Schueller SM, Schneider M, et al. Barriers to and facilitators of user engagement with digital mental health interventions: systematic review. J Med Internet Res 2021 Mar 24;23(3):e24387 [FREE Full text] [doi: 10.2196/24387] [Medline: 33759801]
177. Mikolasek M, Witt CM, Barth J. Adherence to a mindfulness and relaxation self-care app for cancer patients: mixed-methods feasibility study. JMIR Mhealth Uhealth 2018 Dec 06;6(12):e11271 [FREE Full text] [doi: 10.2196/11271] [Medline: 30522990]
178. Arean PA, Hallgren KA, Jordan JT, Gazzaley A, Atkins DC, Heagerty PJ, et al. The use and effectiveness of mobile apps for depression: results from a fully remote clinical trial. J Med Internet Res 2016 Dec 20;18(12):e330 [FREE Full text] [doi: 10.2196/jmir.6482] [Medline: 27998876]

179. Bondaronek P, Alkhaldi G, Slee A, Hamilton FL, Murray E. Quality of publicly available physical activity apps: review and content analysis. JMIR Mhealth Uhealth 2018 Mar 21;6(3):e53 [FREE Full text] [doi: 10.2196/mhealth.9069] [Medline: 29563080]
180. Wallin EE, Mattsson S, Olsson EM. The preference for internet-based psychological interventions by individuals without past or current use of mental health treatment delivered online: a survey study with mixed-methods analysis. JMIR Ment Health 2016 Jun 14;3(2):e25 [FREE Full text] [doi: 10.2196/mental.5324] [Medline: 27302200]
181. Lepore SJ, Rincon MA, Buzaglo JS, Golant M, Lieberman MA, Bauerle Bass S, et al. Digital literacy linked to engagement and psychological benefits among breast cancer survivors in internet-based peer support groups. Eur J Cancer Care (Engl) 2019 Jul 18;28(4):e13134 [FREE Full text] [doi: 10.1111/ecc.13134] [Medline: 31318132]
182. Mobile fact sheet. Pew Research Center. URL: https://www.pewresearch.org [accessed 2022-12-15]
183. Internet/broadband fact sheet. Pew Research Center. 2021. URL: https://www.pewresearch.org [accessed 2022-12-15]
184. Young DR, Stewart KJ. A church-based physical activity intervention for African American women. Fam Community Health 2006;29(2):103-117. [doi: 10.1097/00003727-200604000-00006] [Medline: 16552288]
185. Jenkins F, Jenkins C, Gregoski MJ, Magwood GS. Interventions promoting physical activity in African American women: an integrative review. J Cardiovasc Nurs 2017;32(1):22-29 [FREE Full text] [doi: 10.1097/JCN.0000000000000298] [Medline: 26544170]
186. Pekmezi DW, Williams DM, Dunsiger S, Jennings EG, Lewis BA, Jakicic JM, et al. Feasibility of using computer-tailored and internet-based interventions to promote physical activity in underserved populations. Telemed J E Health 2010 May;16(4):498-503 [FREE Full text] [doi: 10.1089/tmj.2009.0135] [Medline: 20507203]
187. Sayer J, Paniagua D, Ballentine S, Sheehan L, Carson M, Nieweglowski K, Community-Based Participatory Research (CBPR) Team. Perspectives on diet and physical activity among urban African Americans with serious mental illness. Soc Work Health Care 2019 Mar 25;58(5):509-525 [FREE Full text] [doi: 10.1080/00981389.2019.1587662] [Medline: 30907271]
188. Banks-Wallace J, Conn V. Interventions to promote physical activity among African American women. Public Health Nurs 2002 Sep;19(5):321-335. [doi: 10.1046/j.1525-1446.2002.19502.x] [Medline: 12182692]
189. Blue shield of California's Wellvolution program now offers headspace app to help members reduce stress and increase resilience. Blue Shield of California. 2018. URL: https://news.blueshieldca.com/2021/10/26/blue-shield-of-californias-wellvolution-program-now-offers-headspace-app-to-help-members-reduce-stress-and-increase-resilience [accessed 2022-12-16]
190. Ginger for Cigna customers. Ginger. URL: https://www.ginger.com/cigna [accessed 2022-12-16]
191. Self-care apps for your everyday life. Kaiser Foundation Health Plan. URL: https://healthy.kaiserpermanente.org/health-wellness/mental-health/tools-resources/digital [accessed 2022-12-16]

Abbreviations

CBT: cognitive behavioral therapy
HIIT: high-intensity interval training

Edited by L Buis; submitted 19.01.23; peer-reviewed by K Szuhany, C Janney , G Jonathan, J Brinsley; comments to author 21.07.23; revised version received 12.09.23; accepted 30.11.23; published 15.03.24.

Please cite as:
Bernstein EE, Wolfe EC, Huguenel BM, Wilhelm S
Lessons and Untapped Potential of Smartphone-Based Physical Activity Interventions for Mental Health: Narrative Review
JMIR Mhealth Uhealth 2024;12:e45860
URL: https://mhealth.jmir.org/2024/1/e45860
doi:10.2196/45860
PMID:38488834

Review

Attributes, Methods, and Frameworks Used to Evaluate Wearables and Their Companion mHealth Apps: Scoping Review

Preetha Moorthy[1], MSc; Lina Weinert[2,3], MSc; Christina Schüttler[4], PhD; Laura Svensson[2], MSc; Brita Sedlmayr[5], PhD; Julia Müller[2], MSc; Till Nagel[6], PhD

[1]Department of Biomedical Informatics, Center for Preventive Medicine and Digital Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany

[2]Institute of Medical Informatics, Heidelberg University Hospital, Heidelberg, Germany

[3]Section for Oral Health, Heidelberg Institute of Global Health, Heidelberg University Hospital, Heidelberg, Germany

[4]Medical Center for Information and Communication Technology, University Hospital Erlangen, Erlangen, Germany

[5]Institute for Medical Informatics and Biometry, Carl Gustav Carus Faculty of Medicine, Technische Universität Dresden, Dresden, Germany

[6]Human Data Interaction Lab, Mannheim University of Applied Sciences, Mannheim, Germany

Corresponding Author:
Preetha Moorthy, MSc
Department of Biomedical Informatics
Center for Preventive Medicine and Digital Health
Medical Faculty Mannheim, Heidelberg University
Theodor-Kutzer-Ufer 1-3
Mannheim, 68167
Germany
Phone: 49 621 383 8078
Email: preetha.moorthy@medma.uni-heidelberg.de

Abstract

Background: Wearable devices, mobile technologies, and their combination have been accepted into clinical use to better assess the physical fitness and quality of life of patients and as preventive measures. Usability is pivotal for overcoming constraints and gaining users' acceptance of technology such as wearables and their companion mobile health (mHealth) apps. However, owing to limitations in design and evaluation, interactive wearables and mHealth apps have often been restricted from their full potential.

Objective: This study aims to identify studies that have incorporated wearable devices and determine their frequency of use in conjunction with mHealth apps or their combination. Specifically, this study aims to understand the attributes and evaluation techniques used to evaluate usability in the health care domain for these technologies and their combinations.

Methods: We conducted an extensive search across 4 electronic databases, spanning the last 30 years up to December 2021. Studies including the keywords "wearable devices," "mobile apps," "mHealth apps," "physiological data," "usability," "user experience," and "user evaluation" were considered for inclusion. A team of 5 reviewers screened the collected publications and charted the features based on the research questions. Subsequently, we categorized these characteristics following existing usability and wearable taxonomies. We applied a methodological framework for scoping reviews and the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist.

Results: A total of 382 reports were identified from the search strategy, and 68 articles were included. Most of the studies (57/68, 84%) involved the simultaneous use of wearables and connected mobile apps. Wrist-worn commercial consumer devices such as wristbands were the most prevalent, accounting for 66% (45/68) of the wearables identified in our review. Approximately half of the data from the medical domain (32/68, 47%) focused on studies involving participants with chronic illnesses or disorders. Overall, 29 usability attributes were identified, and 5 attributes were frequently used for evaluation: satisfaction (34/68, 50%), ease of use (27/68, 40%), user experience (16/68, 24%), perceived usefulness (18/68, 26%), and effectiveness (15/68, 22%). Only 10% (7/68) of the studies used a user- or human-centered design paradigm for usability evaluation.

Conclusions: Our scoping review identified the types and categories of wearable devices and mHealth apps, their frequency of use in studies, and their implementation in the medical context. In addition, we examined the usability evaluation of these technologies: methods, attributes, and frameworks. Within the array of available wearables and mHealth apps, health care providers encounter the challenge of selecting devices and companion apps that are effective, user-friendly, and compatible with user

interactions. The current gap in usability and user experience in health care research limits our understanding of the strengths and limitations of wearable technologies and their companion apps. Additional research is necessary to overcome these limitations.

(JMIR Mhealth Uhealth 2024;12:e52179) doi:10.2196/52179

KEYWORDS

wearables; mobile health; mHealth; mobile phone; usability methods; usability attributes; evaluation frameworks; health care

Introduction

Background

Wearable technology, also known as *wearable devices*, includes smart electronic devices worn in close proximity to the surface of the human body. These devices can detect, analyze, and transmit information concerning body signals such as vital signs and physiological data, including step count and heartbeat [1-3]. Smart wearable technologies and their high-performance microsensors are of growing importance for patient health monitoring and are being widely accepted into clinical use and trials [4-7]. These technologies have the capability to amplify personal wellness and raise awareness in the spectrum of preventive health care. Consumers continue to rely on smart devices such as mobile phones and smartwatches to engage in healthy behavior [8-10]. They also assist in the self-management of chronic conditions, preventive measures, and aftercare, for example, diabetes monitoring [11], rehabilitation [12-14], fall detection [15,16], wound healing [17], and even monitoring symptoms of long-term illness [18-20]. Wearable technology further enhances the continuum of care within interdisciplinary communication and improves individuals' health and well-being, all in their natural mobile environment [3,21].

Commercial Wearables Versus Medical Wearables

The growing demand for health care technology, particularly wearable devices, has led to the proliferation of various medical and smart health care wearables. However, there is ambiguity in distinguishing between commercial consumer wearable devices and wearable medical devices. The European Union regulations [22] define medical devices as those intended for medical purposes such as disease diagnosis, monitoring, treatment, injury management, and physiological process modification; however, this scope does not include wearable technologies, such as smartwatches, smart bands, and mobile phone–based devices, designed primarily to provide users as tools for health monitoring and management. Fotiadis et al [23] defined wearable medical devices as self-contained, noninvasive devices with specific medical functions. Although a clear definition of wearable medical wearables remains elusive, these devices serve as a convergence point for both conventional medical device manufacturers and consumer-oriented companies aiming to enter the profitable medical market. The traditional distinction between medical and consumer-grade devices relies on the primary intention; however, we found that many commercial devices are being used opportunistically in health care and clinical trials.

Despite the expanding scope of wearable devices, unresolved concerns persist among general consumers regarding the safety, security, and usability of these devices [24-28]. Therefore, ensuring the fit-for-use of these technologies for specific users in clinical settings must be ascertained. The assessment of the usability of these technologies is critical to the success and adoption of wearable and mobile technology or the combination thereof. The identification and consideration of the appropriate attributes and methods for the measurement of usability as early on in the product development process can increase productivity, reduce errors, reduce user training and user support, and improve efficacy, thereby further broadening the acceptance of wearable and mobile technology by users [29-31].

Definition of Usability

In the literature, the definition of *usability* varies, with some studies equating it to assess a device's functionality, whereas others focus on aspects such as feasibility or performance. This ambiguity highlights the need for a comprehensive approach in measuring usability, considering users, devices, environment, and the actions users perform. The International Organization for Standardization (ISO) 9241 clarifies usability as "the extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use" [32]. It is important to recognize that usability extends beyond the immediate outcomes of use. Established standards such as ISO 9241 or other regulatory frameworks primarily view usability as a result of use, emphasizing attributes such as effectiveness, efficiency, and satisfaction. However, a holistic evaluation of usability attributes that goes beyond immediate outcomes contributes to a deeper understanding of user interactions with these technologies and their acceptance in everyday use.

For better adoption of wearables in combination with their companion mobile health (mHealth) apps in clinical settings, usability needs to be considered to safeguard the effectiveness, functionality, and ease of use of these technologies. Concerning the acceptance of wearable technologies, it has been advocated that the devices must be easy to wear, affordable, possess suitable functions, and be appealing to users [33-35]. In such circumstances, designers, developers, and interdisciplinary researchers need to consider the development and use of such devices in a user-centered manner [36-38], thereby affirming that wearable technology is relative to the requirements of the users as it is a vital factor in the adoption of digital health apps and devices because it can be challenging for users owing to their health conditions. Furthermore, usability testing of these technologies allows researchers to understand how the wearable being developed meets users' requirements before being used in health interventions.

Exploring Key Attributes of Usability

Previous scientific literature disclosed the measurements of usability using different entities. These entities are defined as

dimensions, components, scales, or factors of usability. According to Folmer and Bosch [39], these terms are analogous and hold the same meaning. Therefore, as defined by Wixon and Wilson [40], the term usability attribute is the characteristics of a product that can be measured. The most consistently reported usability attributes are effectiveness, efficiency, and satisfaction, which are part of the usability definition of ISO 9241-11:2018 [32,41]. Existing reviews have focused on reporting the usability attributes of mobile apps in health care; however, the shortcoming of applicable attributes for wearables and their companion app poses a challenge in assessing the usability of these technologies.

Exploring Evaluation Methods of Usability

Usability assessment is instrumental in determining how well users learn and use technology to meet their goals. This includes the effectiveness and efficiency of a device and how satisfied the users are with the process. Therefore, different usability evaluation methods should be used to gather this information. Existing literature shows that different methods have been used for the testing of wearables such as field studies and laboratory experiments [42-44]. Although laboratory experiments, field studies, and hands-on measurement are some of the most commonly used methodologies, these are sometimes difficult to apply and have drawbacks. The prevailing usability methods assess different facets of usability, each providing different data. Accordingly, the selection of methods plays a pivotal role in evaluating the desired attributes of usability. Previous reviews have investigated the usability of wearables [45-47] and mHealth apps independently [41,48-50]. These reviews examined the usability assessment of wearables or mobile apps according to specific use cases in the health care domain [51-57]. Moreover, evaluation studies on the combination of the aforementioned devices were not taken into consideration. Research in this area continues to be fragmented, which demonstrates the importance of exploring further the requirements, functionalities, and capabilities of such wearable devices to enhance our comprehensive understanding of their use and acceptance.

Objectives

This study aims to survey the existing literature in the field of medicine and health care that reports on the usability of wearable technology, mHealth apps, or their combination. Our scoping review seeks to analyze the literature in three ways: (1) type (commercial or medical) and category (stand-alone or paired) of wearable devices and their frequency of use in studies; (2) medical use cases; and (3) usability evaluation of these technologies, specifically usability attributes, methods, and frameworks.

Methods

Framework

This study uses the framework developed by Arksey and O'Malley [58] for reporting on scoping reviews, following the recommendations for enhancement of this approach by Levac et al [59]. We followed the five stages of the framework: (1) identifying the research question (RQ), (2) identifying relevant studies, (3) selecting studies, (4) charting the data, and (5) summarizing and reporting the results. In addition, the review followed the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist to report the study selection process of the scoping review (Multimedia Appendix 1) [60,61].

Stage 1: Identifying the RQs

For our scoping review, we used the PICO (Population, Intervention, Comparison, and Outcome) model [62,63] shown in Table 1 to help us regulate our RQs, outline the search strategy, and identify relevant studies within the health care domain. However, for our scoping study, the control or comparison aspect of the PICO methods was eliminated because our focus was not on comparative studies or controlled exposure.

Table 1. The PICO (Population, Intervention, Comparison, and Outcome) method applied to our review.

Aspect	Description	Our review
Patient, population, or problem	Problem to be addressed	• Wearables • Fitness trackers • Physiological data
Intervention, prognostic factor, or exposure	Situation or condition or a characteristic of a patient (technology savvy). Exposure to be considered in treatments and tests	• Mobile devices or smartphones • Mobile apps
Control or comparison	Control or comparison intervention treatment or placebo or standard of care	• Comparison is eliminated as the focus is not on comparative studies or controlled exposure
Outcome to measure or achieve	Outcome of interest—what can be accomplished, ensured, improved, or affected?	• Usability and human factors

This scoping review aims to accomplish its objectives by answering the following RQs:

1. RQ1: What type (commercial or medical) and category (stand-alone or paired) of wearable devices and their companion mHealth apps were implemented and how frequently were they used in the studies?
2. RQ2: What medical use cases and medical data were reported?
3. RQ3: What usability methods, frameworks, and attributes were used for the usability evaluation?

Stage 2: Identifying Relevant Studies

The focus of the second stage of the Arksey and O'Malley framework [58] was to find the relevant studies that match the RQs and the purpose of the scoping review. We began the review with an extensive search using keywords related to the PICO model. However, this raised questions about the sensitivity and specificity of the articles, that is, retrieving and identifying relevant research topic publications. Therefore, redefining the search terms after the initial search added the advantage of prioritizing the sensitivity of the relevant article.

We conducted our search in 4 electronic databases, including ACM Digital Library, IEEE Xplore, PubMed, and Web of Science (Clarivate Analytics), resulting in relevant studies covering the last 30 years up to December 2021. Relevant additional literature was also identified through other resources such as citations and expert recommendations. The search strategy was developed in association with the university librarian at the Medical Faculty Mannheim, Heidelberg University. The search integrated both search terms and Medical Subject Headings associated with the topics of health care, wearables, mHealth apps and terms used under the umbrella term *user experience*. The search strategy for the respective databases can be found in Multimedia Appendix 2.

Stage 3: Study Selection

For our scoping review, all types of articles ranging from journal articles to conference papers were considered, without restrictions on the period of publication. In line with the systematic review methodology, we formulated inclusion and exclusion criteria for this scoping review. The inclusion and exclusion criteria are provided in Textbox 1. This allowed us to reduce the number of papers that were included in the screening of titles and abstracts. Citavi (version 6; Swiss Academic Software GmbH) was used for the collection of the articles.

Textbox 1. Inclusion and exclusion criteria for the scoping review.

Inclusion criteria

- Language: English
- Papers focused on wearables or a prototype of wearables that have used usability testing for their evaluation, where methods such as questionnaires, observations, experimental testing, or surveys were used
- Papers focused on a mobile health (mHealth) app or a prototype of the mHealth app that used usability testing for its evaluation, where methods such as questionnaires, observations, experimental testing, or surveys were used
- Papers that use either a wearable, an mHealth app, or both in a medical use case, for example, chronic diseases
- Papers that use a user-centered design approach for developing wearables or mHealth apps

Exclusion criteria

- Inclusion criteria not fulfilled, for example, papers not written in English or not matching any of the secondary inclusion criteria
- Papers with the theme or topic of augmented reality and virtual reality, which may also include usability studies (eg, Google Glass)
- Papers that have only used audio and visual wearable aids (ie, without additional support from smartphones)
- Papers purely focused only on the technical aspects, technical descriptions, or features of wearables or mobile apps or PDAs in the development and testing processes of materials; self-developed sensors; or wearable sensors such as accelerometers, gyroscopes, or inertial measurement unit
- Papers that are focused mainly on medical professionals rather than patients, for example, describing algorithms or methods used for the optimization of viewing medical data (such as electrocardiogram and electroencephalogram)

We followed the recommendations from Daudt et al [64] for interdisciplinary teamwork in scoping reviews: we incorporated reviewers from different disciplines and backgrounds such as health services research, usability engineering, and medical informatics. The reviewers were divided into 2 groups such that each group had members with diverse backgrounds and expertise. Furthermore, an expert not involved in the screening reviewed mismatched publications from the groups and made discrete decisions for the inclusion and exclusion of articles. Each team member independently reviewed the titles, abstracts, and full text of the publications assigned to them. Studies were considered for the full-text reading if the inclusion criteria were met and cross-verified among team members.

Stage 4: Charting the Data

At this stage, the data from the included studies were extracted. The review team collectively designed a structured data-charting format aligned with the RQs of the scoping review. Each team member individually extracted relevant characteristics from the included studies and adapted them to the data-charting format. Disagreements between the reviewers were resolved through discussions and feedback. The characteristics extracted from the included studies that are associated with the aim of this scoping review are as follows: (1) classification of wearable devices and mHealth apps (only wearables or paired); (2) type of wearable devices; (3) type of mHealth app (stand-alone or interactive); (4) medical use cases (if wearables or mobile apps or combination of both were used in a specific medical use case); (5) physiological data; (6) type of connections between wearables and apps (Wi-Fi or cables); (7) duration of usability studies; and (8) usability evaluation—usability attributes, frameworks, and methods. Excel (Microsoft Corp) was used to facilitate this process.

Stage 5: Summarizing and Reporting the Results

To effectively summarize and organize the extracted data, a comprehensive search of the relevant literature was conducted to identify suitable articles to structure the examination and analysis. Two relevant literature sources were identified to fulfill our objectives. We aimed to find a suitable classification system that could effectively categorize various wearable devices. The classification proposed by Seneviratne et al [65], which provides a comprehensive survey of commercial wearable products grouped into 3 categories—accessories, e-textiles, and e-patches—served as a helpful tool for our analysis. In addition, we used the usability taxonomy hierarchy proposed by Alonso-Ríos et al [66] for our analysis. This taxonomy provided a comprehensive framework for organizing usability attributes in a logical and meaningful order.

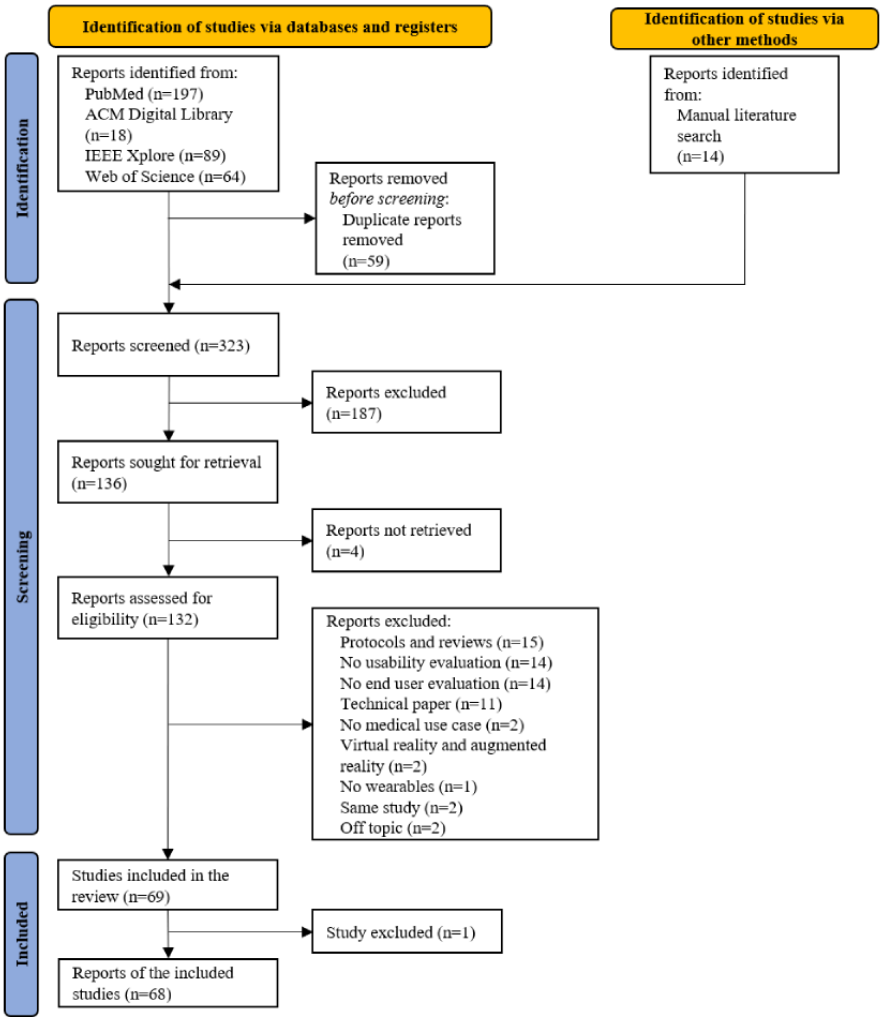
We present our findings by integrating descriptive tables and graphical illustrations of the outcomes. These figures and graphics helped our analysis to directly connect the findings to the objectives of our review and identify the gaps in the literature. In our study, we illustrate the frequencies and percentages of the findings in coherent data visualizations, emphasizing the analysis and reporting of data and giving them a comprehensive meaning.

Results

Eligible Studies

Our search yielded 382 records, including articles about wearables, mHealth apps, or their combination; research about the implementation of these technologies in medical use cases; and evaluations of their usability. Overall, 323 records were evaluated for the initial screening of titles and abstracts after eliminating duplicates. From these, 132 full-text papers were found, of which 62 were excluded, resulting in 69 studies whose data were charted per the study questions. Following the final text reading, a single study had to be excluded from this scoping review because it did not meet the predefined inclusion criteria, despite the presence of relevant keywords in the paper. The process of selection of articles for the scoping review can be seen in the PRISMA-ScR diagram shown in Figure 1.

Figure 1. PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) flow diagram of the selection process for the scoping review.



Wearable Devices and Their Frequency of Use

Most studies (57/68, 84%) used a combination of wearables and mobile apps. Overall, 12% (8/68) of the studies used wearables in conjunction with other technological devices such as smartphones, computers, recording devices, or PDAs; however, these devices were used independently from the wearables. Furthermore, the data extraction indicated that two-thirds of the studies (45/68, 66%) used commercially available wearables for their evaluation studies, of which approximately half of the studies (21/45, 47%) used Fitbit (Fitbit Inc.) devices as their source of data tracking and collection. Our data further showed that only 11% (5/45) of the studies used wearable devices that were certified as medical devices. In addition, only 18% (12/68) of the studies used a self-developed wearable prototype and a mobile app for data tracking and monitoring.

From the included studies, 9% (6/68) of the studies reported using only wearables in their studies. Of these 6 studies, 5 (83%) used commercially available wearable devices such as Fitbit and Samsung Gear S3 (Samsung Electronics Co, Ltd), whereas 1 (1%) study reported using a self-developed wearable prototype. However, 7% (5/68) of outliers were detected where 60% (3/5) of the studies mentioned using only a mobile app; 20% (1/5) of the studies reported using only a smartphone; and in 20% (1/5) of the studies, smartphone was used as a wearable device by attaching a 3D-printed phone holder around the user's neck [67]. Although this is typically not defined as a wearable, the outcome of this study proved imperative in determining the different devices and variables used for evaluation. We used a classification system consisting of 6 distinct groups to categorize the diverse use of wearable technologies and mobile apps (with, without, or a combination thereof). They are (1) only wearables (eg, stand-alone wearables such as Fitbit and Garmin [Garmin Ltd]), (2) wearables+companion apps (eg, Garmin tracker+corresponding Garmin mobile app), (3) smartphone as wearables (eg, smartphone used in close proximity to the skin to track physiological data), (4) wearables+connectivity (not companion) apps (eg, wearables paired with connectivity apps for Bluetooth connection and not for presenting data), (5) wearables+other technologies (eg, Garmin+laptops, recording units, and PDAs), and (6) others (only smartphone or only app). The specified categories and the corresponding studies included in this review are presented in Table 2.

In most studies, the validation, accuracy, and certification of the used wearables were not thoroughly discussed, despite these aspects being considered essential in good research practices. Although some studies briefly touched upon validation or accuracy, they did not necessarily indicate that the wearables had undergone certification, such as Food and Drug Administration approval or Conformité Européenne mark. Authors of the included studies often omitted reporting the inaccuracies and validation limitations of consumer-grade wearables, particularly when usability was of significant importance. Instead, the accuracy of wearables was often assumed based on the authors' validation of wearables' selection through peer-reviewed research and their alignment with traditional instruments for measuring health data. The list of included studies along with the extracted information can be found in Multimedia Appendix 3 [43,67-133].

Most of the wearables (35/68, 51%) covered in the review were wrist worn, such as wristbands or smartwatches. However, only 9% (6/68) of the studies used multiple wearables, such as a smartwatch and chest belt or multiple wrist-worn wearables. Table 3 presents the data on the different categories of the types of wearable devices extracted from the included studies. More than half of the studies (48/68, 71%) deployed stand-alone mobile apps, which implies that users or patients collected health information using wearables and apps without sharing it with their professionals. Overall, 68% (46/68) of the studies determined that Bluetooth connections were the primary means of connectivity for wearables and the mobile apps that accompanied them. Data from the extracted studies revealed no linkage between the technologies; hence, outliers (4/68, 6%) were also recognized.

Table 2. Wearable devices, mobile apps, and their combination along with their frequency of use in studies (n=68).

Category of wearable devices and mobile apps	Frequency of use in studies, n (%)	References
Only wearables	6 (9)	[68-73]
Wearable devices+companion apps	49 (72)	[43,74-121]
Wearable devices+connected with apps (not companion app)	5 (7)	[122-126]
Wearables+other technologies (eg, laptops, recording units, and PDAs)	3 (4)	[127-129]
Smartphone as wearables	1 (1)	[67]
Others	4 (6)	[130] (smartphone only), [131-133] (app only)

Table 3. Categorization of the type of wearable devices according to the classification of wearable devices by Seneviratne et al [65] ordered by their frequency of use in studies (n=68).

Categorization of the type of wearables	Description	Devices	Studies, n (%)
Wrist worn	Wrist-worn devices with fitness tracking capabilities or other functionalities, generally without a touchscreen	Wrist bands	26 (38)
Wrist worn	Wrist-worn devices with a touchscreen display	Smartwatches	9 (13)
>1 wearable device in the study	Study includes >1 wearable device (any type)	Wrist bands, Upper arm bands, e-Patches, Sensor patches	6 (9)
Other accessories	Chest straps, belts, upper arm bands (in contrast to wrist-worn bands), or knee straps equipped with sensors for health tracking or other functionalities	Straps	5 (7)
e-Textiles	Main clothing items that also serve as wearables, such as shirts, pants, and undergarments	Smart garments	5 (7)
Hearables	Fits in or on an ear that contains a wireless link	Hearing devices	4 (6)
Other accessories	Clip-on	Clip-on	3 (4)
e-Patches	Sensor patches that can be adhered to the skin for either fitness tracking or haptic applications	Sensor patches	3 (4)
Outliers	Does not fit the categories	Wrist bands, Upper arm bands, e-Patches, Sensor patches	3 (4)
e-Textiles	Shoes, socks, insoles, or gloves embedded with sensors	Foot or hand-worn	2 (3)
Other accessories	Jewelry designed with features such as health monitoring and handless control	Smart jewelry	1 (1)
e-Patches	Tattoos with flexible and stretchable electronic circuits to realize sensing and wireless data transmission	e-tattoo or e-skin	—[a]
Head-mounted devices	Spectacles or contact lenses with sensing, wireless communication, or other capabilities	Smart eyewear	—
Head-mounted devices	Bluetooth enables headsets or earplugs. Sensor-embedded hats and neck-work devices are also found in research products	Headsets or earbuds	—

[a]Not available.

Medical Use Cases and Reported Data

Our data showed that approximately half of the studies (32/68, 47%) focused on participants with chronic illnesses or disorders, indicating the importance of wearable technologies in managing and monitoring chronic conditions. The remaining 53% (36/68) of the studies encompassed various other medical use cases such as wellness, mental health, rehabilitation, sleeping disorders, otolaryngology, and preventive measures.

Most studies (40/68, 59%) routinely collected physiological data from users or patients. The most commonly collected health data revolved around physical activity, encompassing metrics such as steps taken, stairs climbed, and inertial measurement units. Approximately 35% (24/68) of the health data gathered in the studies focused on cardiac measurements, including electrocardiogram, heart rate variability, heart rate, or blood pressure. In addition to physical activity and cardiac measurements, other data types were also collected, albeit to a lesser extent. Sleep data accounted for 25% (17/68) of the collected information, and brain activity data, such as electroencephalogram recordings, constituted 4% (3/68) of the data. Furthermore, biosignals, including measurements such as skin conductance and respiration rate, were captured in 18% (12/68) of the studies. Other health data and observations such as acoustics, posture, blood glucose, and weight were also monitored in approximately 21% (14/68) of the cases, indicating the broad range of parameters that wearables can track and analyze.

In addition to physiological data, a small proportion of the studies (4/68, 6%) included in our analysis also collected nonphysiological data that were not directly linked to health parameters. These data contained various variables such as the number of cigarettes consumed per day, location data, and dietary intake. Although not directly related to traditional health measurements, the inclusion of such data provides a broader context and enables a more holistic understanding of individuals' behaviors and lifestyle factors.

Usability Attributes

The studies included in this review used various terms such as *usability characteristics* and *attributes*, which we consider to be synonymous. Therefore, we applied the usability attributes from the extracted data to the usability taxonomy [66]. We found that satisfaction (34/68, 50%), ease of use (27/68, 40%), user experience (16/68, 24%), perceived usefulness (18/68, 26%), and effectiveness (15/68, 22%) were the most commonly used attributes for assessing usability. Although user experience is acknowledged as the overarching term encompassing

usability, conceptualizing user experience as a facet of usability captures the comprehensive perception arising from interactions with devices. This extends beyond the mere use of the device, encapsulating the entirety of the experience or the anticipated use of the technologies. Furthermore, we identified 32% (22/68) of the studies that simply reported usability or perceived usability. Moreover, our findings further indicate that out of the 29 identified usability attributes, 6 (21%) can be classified as quality attributes. Table 4 shows the mapping of the attributes identified in the review to the attributes defined in the usability taxonomy. These particular usability characteristics possess qualities that are directly related to the overall quality and performance of the technologies used in the studies. Table 5 presents the quality and product attribute matrix of the attributes ascertained in the review and the defined attributes from the ISO norm 25010 [134]. A detailed explanation of the different attributes in Tables 4 and 5 can be found in Multimedia Appendix 4 [66].

Table 4. Matrix mapping of the attributes identified in the scoping review and the usability taxonomy [66][a].

Usability attributes from scoping review	Usability taxonomy						
	Knowability	Operability	Efficiency	Robustness	Safety	Subjective satisfaction	Number of studies, n (%)
Accuracy		✓					5 (7)
Aesthetics						✓	7 (10)
Attitude						✓	4 (6)
Attractiveness						✓	2 (3)
Clarity	✓						1 (1)
Cognitive load			✓				2 (3)
Controllability		✓					1 (1)
Data quality		✓					1 (1)
Ease of use or perceived ease of use (effort expectancy, easiness, self-descriptiveness, and self-efficacy)	✓						27 (40)
Effectiveness (user errors, ease of executing a task, task completion, and task completeness)		✓					15 (22)
Efficiency (task time)			✓				6 (9)
Engagement						✓	7 (10)
Error tolerance				✓			1 (1)
Functionality		✓					7 (10)
Hedonic motivation						✓	2 (3)
Learnability	✓						4 (6)
Likes and dislikes						✓	1 (1)
Perceived usefulness (usefulness, utility, performance expectancy, and system usefulness)	✓						18 (26)
Satisfaction (subjective app quality, survey and ratings, positive and negative feedback, opinions and reactions, and participants' experience)						✓	34 (50)
Technical difficulties				✓			2 (3)
Trust					✓		1 (1)
User control		✓					1 (1)
User experience (experience, overall subjective quality, and user-friendliness)						✓	16 (24)

[a]The rows list all the usability attributes identified in the scoping review and the columns list the first-level usability attributes from Alonso-Ríos et al [66], with each checkmark symbol indicating a match based on their description and their sublevel attributes. The last column lists the number of studies using the term from that row.

Table 5. Quality and product attributes identified in the scoping review that match the attributes from the International Organization for Standardization (ISO) norm ISO/International Electrotechnical Commission (IEC) 25010 [134][a].

Attributes identified in the scoping review	Quality attribute					Product attribute								
	Effectiveness	Efficiency	Satisfaction	Safety	Usability	Functional suitability	Performance efficiency	Compatibility	Usability	Reliability	Security	Maintainability	Portability	Number of studies (%)
Comfort			✓											8 (12)
Design—app interface and design and interface quality									✓					3 (4)
Effectiveness[b]	✓													15 (22)
Efficiency[b]		✓					✓							6 (9)
Facilitating conditions								✓						1 (1)
Information quality							✓							3 (4)
Interface quality									✓					2 (3)
Reliability										✓				1 (1)
Trust[b]				✓										1 (1)
Satisfaction[b]			✓											34 (50)

[a]The rows list all the attributes identified in the scoping review and the columns list the quality and product attributes from ISO norm 25010 [134], with each checkmark symbol indicating a match based on the description of the attributes. The last column lists the number of studies using the term from that row.

[b]Overlapping attributes also identified as usability attributes.

In addition, during the data extraction process, we obtained insights into the elements and factors that affect the evaluation of usability. Among the 68 included studies, 52 (76%) reported these elements and factors, which played a crucial role in shaping and guiding the measurement of usability. Notably, acceptance (21/68, 31%) emerged as the most commonly used element or factor in assessing usability, indicating its significance in understanding users' acceptance and adoption of wearable technologies.

Usability Evaluation Methods and Frameworks

Only 12% (9/68) of the studies outlined using some sort of framework for usability evaluation. User- or human-centered design (7/68, 10%) was the most commonly used framework. Our findings revealed that more than half of the studies (37/68, 54%) collected data using the mixed methods approach. Only 15% (10/68) of the articles used only qualitative methods. These data collection methods included interviews, focus group discussions, thinking-aloud protocols, cognitive walkthroughs, open-ended discussions, Wizard of Oz, and free-text writing. Approximately two-thirds of the studies (21/68, 31%) used quantitative approaches for data collection during evaluation studies. The System Usability Scale (SUS) outnumbered other usability questionnaires (17/68, 25%) such as the Mobile Application Rating Scale, Net Promoter Score, Single Ease Question, NASA Task Load Index, and Technology Acceptance Model. However, a large percentage (21/68, 31%) of the articles used self-developed surveys or self-reporting questionnaires. Only one-fourth of the articles (16/68, 24%) further implemented statistical analysis including task completion, number of errors, descriptive statistics, or Google Analytics for the assessment of usability. Consequently, only a small proportion (3/68, 4%) of the included studies performed heuristic evaluation as a form of expert evaluation.

Discussion

Principal Findings

Our data suggest that the evaluation of wearables for medical purposes was largely conducted without direct integration with mobile apps. Although some studies used smartphones as a means of connecting with the wearables, users were not assigned companion apps for data viewing. This limits the analysis and data visualization capabilities of the data collected within the studies. Wrist-worn devices were the most common type of wearables identified in the studies, indicating the convenience

of using this type of wearables for measuring physical activities in a research setting. We also found that most studies (21/68, 47%) reported the use of consumer-grade fitness and activity trackers from Fitbit, and only a handful of the studies (5/68, 7%) implemented medical-grade wearable devices. A small number of the studies (10/68, 15%) investigated the data collection and use of the wearable in the aspects of aftercare of patients, wellness, and rehabilitation.

We noted the absence of standards or guidelines to facilitate the analysis of the usability of wearables, mobile apps, or their combination. Although user- and human-centered design frameworks were mentioned in a few studies (7/68, 10%), they are guiding the design and development of systems and devices focusing on users and their needs but not the usability of these devices. Despite the fact that many wearable technologies were included in the study, no usability evaluations of multiple devices, the combination of devices, or multidevice interfaces were reported.

Only a little more than one-third of the included publications (22/68, 32%) in our review explicitly reported the measurement of usability or perceived usability. Some studies (8/68, 12%) primarily focused on assessing the measurement of *usability* or *perceived usability* as user perceptions of the devices, attitudes, and compliance using different qualitative or quantitative methods [71,89,91,95,97,98,102,129]. However, it is worth noting that these studies encountered a challenge in clearly differentiating the evaluation of the usability of the wearable device from that of the accompanying mobile app. Consequently, the intended purpose of the evaluation may have been limited in these studies. Therefore, we address the term *usability*, a broad term that encompasses various factors including technology and user acceptability. Studies might have reported on usability alone either due to missing expertise or due to a high-level summarization of various aspects they investigated. To help mitigate this, we incorporated and synthesized a set of attributes or subattributes to measure the capability and performance of a system based on the extracted data.

Our analysis showed that a subset of usability studies lacked testing with the intended target group and instead relied on healthy adults as participants. Despite the absence of explicit acknowledgment of this limitation in the included studies, it raises concern regarding the extent to which the devices and apps under investigation adequately address the unique needs and requirements of the target users, particularly individuals with chronic health conditions. This observation is in line with the findings reported in previous studies [135-144], emphasizing the concern regarding the devices and apps under investigation that adequately meet the unique needs and requirements of the target users.

Comparison With Prior Work

Wearables

Most of the studies (45/68, 66%) in our review used consumer-grade wearables. This matches the observations of other studies [145-150] that investigated commercially available wearables and reported a wide variety of purposes, ranging from digital diagnostic tools to sports tracking to remote monitoring. Niknejad et al [151] and Ferreira et al [152] reported that consumer-grade wearables have been used to foster self-awareness among users. In contrast, we observed that the studies using consumer-grade wearables in our corpus focused on their use for monitoring chronic conditions such as diabetes, obesity, cardiology, and cancer.

We found that wearables are gradually being used more widely in health care and clinical settings. As stated in the previous paragraph, this is true for consumer-grade wearables such as Fitbit, Jawbone, Apple, and Garmin [153-157]. However, concerns about the safety, reliability, and accuracy of these devices persist. In their work, Piwek et al [146] raised concerns about wearables in terms of user safety, emphasizing the need to better address the reliability and security of the data collected from these devices. Considering the inherent lack of emphasis on user safety in consumer wearables, it is imperative to acknowledge the importance of adhering to standard safety and privacy protocols. This includes ensuring ethical transparency and providing appropriate education to users regarding the privacy and information security risks they may encounter when using such devices [24,158,159]. In addition, the use of consumer wearables in health care settings remains somewhat ambivalent at present [160,161]. Although our review did not specifically address these concerns, we acknowledge that these factors significantly affect the usability of wearable technologies, whether used independently or with companion mHealth apps. Similar to Piwek et al [146], who pointed out acceptance challenges of wearables concerning safety and security, we believe that a structured framework with clear definitions and well-defined methods would allow bringing wearables into more diverse practices in the health care system, encouraging a broader adoption and implementation of use cases with high and tested usability.

Studies by Niknejad et al [151], Dimou et al [162], and Yang et al [163] proposed different categorization approaches for wearables, considering factors such as industry relevance (eg, health care or fashion) or wearable placement on the body. However, owing to the wide variety of wearables available in the market, establishing a standardized classification or hierarchy for these different types of wearables becomes challenging. Thus, to help designers and developers, a standardized classification or hierarchy would be helpful when selecting wearables for specific use cases.

Usability Attributes

Many studies in our corpus did not explicitly state the usability attributes they evaluated. Some mentioned generic terms such as *usability* or *user experience*, but did not define them further for their specific cases. As we have argued, more specific usability attributes can facilitate the development of more appropriate requirements and clearer identification of problems in usability studies. This matches observations in related areas. Meyer et al [164] analyzed usability evaluation practices in wearable robotics and recommended better distinguishing between the different usability dimensions and including qualitative measures for identifying a wider range of usability issues. Chiauzzi et al [139] examined the use of wearable

devices for long-term chronic disease management. Patient concerns regarding technical difficulties and the appeal of the devices were identified, but their investigation did not address usability attributes such as device comfort and usefulness. Furthermore, the authors emphasized the importance of wearables being perceived as usable and generating comprehensible data to facilitate wider adoption among patients.

Among the 29 identified attributes in our review, 6 (21%) attributes were found to be more suitable for capturing the quality or product-related aspects of the technologies investigated in this study (Table 5). However, determining the most appropriate attributes for wearables and their associated mobile apps can be challenging because of the potential overlap between the product and inherent attributes, such as effectiveness and satisfaction (Table 5). This finding is consistent with Bakhshian and Lee [165], who argued that consumers' attitudes and purchase intentions toward wearable technology are influenced by both product attributes and inherent attributes, including functional, expressive, and esthetic characteristics. In contrast, other studies have explored the design attributes and their influence on user interactions and acceptance of different types of wearables for specific use cases, such as electroencephalogram systems [166], autism spectrum disorders [167], sports applications [168], and haptic feedback wearable robots [169]. The importance of assessing usability based on user interactions with wearable devices and their associated app remains crucial, amid the emphasis on the design and quality attributes of wearables.

Our data highlight the importance of considering supplementary attributes such as wearability, perceived usefulness, and connectivity when evaluating individual wearables and companion mobile apps. Consistent with our findings, the existing literature also emphasizes the significance of incorporating auxiliary attributes beyond the conventional usability factors such as effectiveness and satisfaction to enhance the acceptance of wearable devices. The aforementioned supplementary attributes identified in the literature include characteristics such as comfort, user-friendliness, affordability, useful features, and appealing design [33-35,136]. It is imperative to incorporate these attributes into the evaluation process to ensure comprehensive assessments of usability and user acceptance in both wearables and companion mobile apps. Although a few studies provide a general overview of wearable attributes related to design and product quality, there is limited research that specifically focuses on the usability attributes of wearables. Although some reviews have identified specific usability attributes for mHealth apps used in various use cases [41,170-176], these are insufficient when evaluating the combined use of wearables and companion apps because of the complex and multifaceted features and interactions involved.

Our results revealed diverse informal terminology used to describe usability, performance, and quality aspects of the technologies examined in the studies. This variability in terminology posed challenges in accurately distinguishing and classifying the terms based on their usability characteristics. We adopted an existing usability taxonomy from the literature to address this issue and ensure consistency in data interpretation [66].

Usability Evaluation Methods and Frameworks

Our analysis revealed that only 9 studies incorporated the user- and human-centered design framework in the design and development of prototypes. These studies specifically targeted specific user populations and assessed usability attributes such as satisfaction, ease of use, and effectiveness as part of their evaluation process. Although usability frameworks are available individually for the design and development of wearables and mobile apps [32,177-182], a usability evaluation framework for the combination of these technologies or multi-interface devices is unavailable.

In our scoping review, most of the included studies (57/68, 84%) used qualitative methods, with interviews being the primary method. This corresponds with other studies reporting on evaluation methodologies [183-185]. A survey on the evaluation of physical activity apps highlighted that a substantial number of studies specifically used a mixed methods approach, including randomized controlled trials, to assess the acceptability and evaluate the usability of wearable technologies, with or without companion mobile apps [186]. This approach has gained popularity because of its comprehensive nature and ability to capture diverse perspectives. Among the combinations of methods used, questionnaires and interviews have emerged as commonly used techniques [75,187-189].

Among the 68 studies included in our review, only one-third (20/68, 31%) used a standardized usability questionnaire to evaluate perceived usability. These questionnaires comprise a predefined set of questions presented in a specific order and format, with established scoring rules based on respondents' answers [190]. The SUS questionnaire [191] was used most frequently among the studies analyzed. Although SUS was originally designed as a generic tool for usability assessment across a broad range of digital interfaces and software apps, it may not include items tailored to the specific characteristics and challenges posed by wearable technology. Wearables often involve prolonged and continuous interaction with the user, making aspects such as device comfort and user experience crucial; however, these aspects are not comprehensively addressed by the SUS. Researchers have proposed different adaptations of usability questionnaires tailored to assess the usability of wearables or mobile apps [192-194]. Although this helps to better evaluate the particularities of wearables and mobile apps, they lack items assessing novel usability considerations such as ergonomics, comfort, real-time data feedback, and interaction with the wearable independently or combined with their associated mobile app, which has been identified as important factors by studies [33,73,195,196]. The investigation of these aspects, which are crucial for most systems and apps, has been limited and partial. These observations may be attributed to various factors, including inadequate awareness of available usability questionnaires, the perception that these questionnaires do not align with their specific study, or the belief that the items or constructs within the questionnaires do not adequately reflect the purpose of their evaluation.

Concluding Analysis

We have shown that health care professionals and the medical technology industry acknowledge the importance of high or

adequate usability in new medical equipment, including wearables and mobile apps. In addition, our results highlight the shortcomings in the evaluation and reporting of usability for wearable technologies, necessitating further research on human factors and usability. Literature reviews emphasize concerns regarding the lack of standardized study methodology reporting, guidelines for evaluating usability, and the absence of frameworks or theories for designing comprehensive usability assessments [151,197-199]. According to the reviews conducted by Khakurel et al [27] and Keogh et al [144], the current literature lacks a comprehensive usability evaluation method that effectively addresses usability issues throughout the entire life cycle of a wearable device, from the early development stages to product release. Considering the significance of reliability and wearability in wearable devices, it is imperative to establish traceability in the usability evaluation process. Although researchers are actively engaged in assessing usability, further research is required to identify potential usability attributes, develop suitable evaluation methods and frameworks, and successfully integrate these effective assessments into practice.

Limitations

The goal of this scoping review was to investigate wearable devices and their frequency of use in studies as well as their combination with mHealth apps within the medical domain. Our findings may not completely capture the breadth of usability attributes and their effectiveness in wearables across different contexts. Wearables and their companion apps have demonstrated utility in various use cases and recreational activities, including industrial settings, gaming, museums, and entertainment. By limiting our search to health care–related keywords, we may have excluded valuable insights and perspectives from these alternative domains. Future research could broaden the search criteria to include diverse contexts and use cases beyond health care. This will allow for a more holistic exploration of the potential and effectiveness of usability attributes across different industries and settings.

This study did not aim to synthesize evidence on the effectiveness of usability evaluation methods. Instead, it focused on capturing the diversity of the available literature, encompassing various objectives, critical usability measures, and methods. Consequently, this study primarily serves as an exploratory investigation and provides suggestions for future research in the field.

Conclusions

Our scoping review sheds light on the types and categories of wearable devices, frequency of wearables used in the medical context, their use cases, and the evaluation of their usability. With a wide array of wearables and mHealth apps available, health care providers and manufacturers face the challenge of selecting devices and apps that are effective and user-friendly. The evaluation of usability is crucial for ensuring user engagement and the success of these technologies. As our scoping review shows, there is a lack of standardized frameworks for classifying usability attributes and their subattributes as well as structured evaluation guidelines for wearable technologies. This gap in usability and user experience research hinders the understanding of strengths and limitations in the field of wearable technologies. Therefore, further research is needed to address these limitations and enhance the comprehension of researchers in this field.

Acknowledgments

The authors would like to thank Thomas Ganslandt for his early support and initiation of this scoping review. The authors would also like to thank Fabian Siegel for his support in the execution of this review. The library at the Medical Faculty Mannheim of Heidelberg University played a vital role in assisting with the development and implementation of the search strategy for this scoping review. The extensive knowledge and access to resources of the library staff greatly facilitated a thorough and systematic exploration of the relevant literature for our study. The authors extend their gratitude to the librarian Maurizio Grilli for his valuable support, as it made a substantial contribution to the success of this research. In addition, for the publication fee, the authors acknowledge the financial support from Deutsche Forschungsgemeinschaft within the funding program, Open Access Publikationeskosten, and from Heidelberg University.

Authors' Contributions

PM proposed the idea for this review. PM developed the search strategy, and the strategy was approved by LW, LS, CS, BS, and JM. PM performed the search screenings. LW, LS, CS, BS, and JM screened the titles and abstracts and determined the eligibility of the studies based on full-text examination. PM contributed to the screening process and the full-text examination. LW, LS, CS, BS, and JM extracted data from the included studies. PM proposed the framework for the review, and TN approved it. PM wrote the first draft of the manuscript, and LW, LS, CS, BS, JM, and TN contributed to the drafts of the manuscript. All authors approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist.
[DOCX File , 106 KB - mhealth_v12i1e52179_app1.docx]

Multimedia Appendix 2
Search strategies for the 4 databases.
[DOCX File , 29 KB - mhealth_v12i1e52179_app2.docx]

Multimedia Appendix 3
List of the included studies and extracted information.
[XLSX File (Microsoft Excel File), 63 KB - mhealth_v12i1e52179_app3.xlsx]

Multimedia Appendix 4
Short description of the different attributes in Tables 4 and 5.
[DOCX File , 45 KB - mhealth_v12i1e52179_app4.docx]

References

1. Düking P, Achtzehn S, Holmberg HC, Sperlich B. Integrated framework of load monitoring by a combination of smartphone applications, wearables and point-of-care testing provides feedback that allows individual responsive adjustments to activities of daily living. Sensors (Basel) 2018 May 19;18(5):1632 [FREE Full text] [doi: 10.3390/s18051632] [Medline: 29783763]
2. Düking P, Hotho A, Holmberg HC, Fuss FK, Sperlich B. Comparison of non-invasive individual monitoring of the training and health of athletes with commercially available wearable technologies. Front Physiol 2016 Mar 09;7:71 [FREE Full text] [doi: 10.3389/fphys.2016.00071] [Medline: 27014077]
3. Odendaal WA, Anstey Watkins J, Leon N, Goudge J, Griffiths F, Tomlinson M, et al. Health workers' perceptions and experiences of using mHealth technologies to deliver primary healthcare services: a qualitative evidence synthesis. Cochrane Database Syst Rev 2020 Mar 26;3(3):CD011942 [FREE Full text] [doi: 10.1002/14651858.CD011942.pub2] [Medline: 32216074]
4. Izmailova ES, Wagner JA, Perakslis ED. Wearable devices in clinical trials: hype and hypothesis. Clin Pharmacol Ther 2018 Jul 02;104(1):42-52 [FREE Full text] [doi: 10.1002/cpt.966] [Medline: 29205294]
5. Cox SM, Lane A, Volchenboum SL. Use of wearable, mobile, and sensor technology in cancer clinical trials. JCO Clin Cancer Inform 2018 Dec;2:1-11 [FREE Full text] [doi: 10.1200/CCI.17.00147] [Medline: 30652590]
6. Beauchamp UL, Pappot H, Holländer-Mieritz C. The use of wearables in clinical trials during cancer treatment: systematic review. JMIR Mhealth Uhealth 2020 Nov 11;8(11):e22006 [FREE Full text] [doi: 10.2196/22006] [Medline: 33174852]
7. Tsoukalas D, Chatzandroulis S, Goustouridis D. Capacitive microsensors for biomedical applications. In: Webster JG, editor. Encyclopedia of Medical Devices and Instrumentation. Hoboken, NJ: John Wiley & Sons; 2006.
8. Choudhury A, Asan O. Impact of using wearable devices on psychological distress: analysis of the health information national trends survey. Int J Med Inform 2021 Dec;156:104612 [FREE Full text] [doi: 10.1016/j.ijmedinf.2021.104612] [Medline: 34649113]
9. Li H, Wu J, Gao Y, Shi Y. Examining individuals' adoption of healthcare wearable devices: an empirical study from privacy calculus perspective. Int J Med Inform 2016 Apr;88:8-17. [doi: 10.1016/j.ijmedinf.2015.12.010] [Medline: 26878757]
10. Berg M. Making sense with sensors: self-tracking and the temporalities of wellbeing. Digit Health 2017 Mar 28;3:2055207617699767 [FREE Full text] [doi: 10.1177/2055207617699767] [Medline: 29942585]
11. Lee H, Hong YJ, Baik S, Hyeon T, Kim DH. Enzyme-based glucose sensor: from invasive to wearable device. Adv Healthc Mater 2018 Apr 15;7(8):e1701150. [doi: 10.1002/adhm.201701150] [Medline: 29334198]
12. Joo Y, Byun J, Seong N, Ha J, Kim H, Kim S, et al. Silver nanowire-embedded PDMS with a multiscale structure for a highly sensitive and robust flexible pressure sensor. Nanoscale 2015 Apr 14;7(14):6208-6215 [FREE Full text] [doi: 10.1039/c5nr00313j] [Medline: 25779911]
13. Sardini E, Serpelloni M, Pasqui V. Wireless wearable t-shirt for posture monitoring during rehabilitation exercises. IEEE Trans Instrum Meas 2015 Feb;64(2):439-448. [doi: 10.1109/tim.2014.2343411]
14. Park YL, Chen B, Pérez-Arancibia NO, Young D, Stirling L, Wood RJ, et al. Design and control of a bio-inspired soft wearable robotic device for ankle-foot rehabilitation. Bioinspir Biomim 2014 Mar 16;9(1):016007. [doi: 10.1088/1748-3182/9/1/016007] [Medline: 24434598]
15. Shany T, Redmond SJ, Narayanan MR, Lovell NH. Sensors-Based Wearable Systems for Monitoring of Human Movement and Falls. IEEE Sens J 2012 Mar;12(3):658-670. [doi: 10.1109/jsen.2011.2146246]
16. Aziz O, Robinovitch SN. An analysis of the accuracy of wearable sensors for classifying the causes of falls in humans. IEEE Trans Neural Syst Rehabil Eng 2011 Dec;19(6):670-676. [doi: 10.1109/tnsre.2011.2162250]
17. Hattori Y, Falgout L, Lee W, Jung S, Poon E, Lee JW, et al. Multifunctional skin-like electronics for quantitative, clinical monitoring of cutaneous wound healing. Adv Healthc Mater 2014 Oct 26;3(10):1597-1607 [FREE Full text] [doi: 10.1002/adhm.201400073] [Medline: 24668927]
18. Ozanne A, Johansson D, Hällgren Graneheim U, Malmgren K, Bergquist F, Alt Murphy M. Wearables in epilepsy and Parkinson's disease-a focus group study. Acta Neurol Scand 2018 Feb 16;137(2):188-194. [doi: 10.1111/ane.12798] [Medline: 28714112]

19. Teshuva I, Hillel I, Gazit E, Giladi N, Mirelman A, Hausdorff JM. Using wearables to assess bradykinesia and rigidity in patients with Parkinson's disease: a focused, narrative review of the literature. J Neural Transm (Vienna) 2019 Jun 22;126(6):699-710 [FREE Full text] [doi: 10.1007/s00702-019-02017-9] [Medline: 31115669]
20. Del Din S, Elshehabi M, Galna B, Hobert MA, Warmerdam E, Suenkel U, et al. Gait analysis with wearables predicts conversion to Parkinson disease. Ann Neurol 2019 Sep 27;86(3):357-367 [FREE Full text] [doi: 10.1002/ana.25548] [Medline: 31294853]
21. Kumar S, Nilsen W, Pavel M, Srivastava M. Mobile health: revolutionizing healthcare through transdisciplinary research. Computer 2013 Jan;46(1):28-35. [doi: 10.1109/mc.2012.392]
22. Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC (text with EEA relevance). European Commission. 2021. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:02017R0745-20230320 [accessed 2023-10-31]
23. Fotiadis DI, Glaros C, Likas A. Wearable medical devices. In: Akay M, editor. Biomedical Engineering. Hoboken, NJ: John Wiley & Sons; 2006:3816-3827.
24. Loncar-Turukalo T, Zdravevski E, Machado da Silva J, Chouvarda I, Trajkovik V. Literature on wearable technology for connected health: scoping review of research trends, advances, and barriers. J Med Internet Res 2019 Sep 05;21(9):e14017 [FREE Full text] [doi: 10.2196/14017] [Medline: 31489843]
25. Segura Anaya LH, Alsadoon A, Costadopoulos N, Prasad PW. Ethical implications of user perceptions of wearable devices. Sci Eng Ethics 2018 Feb 2;24(1):1-28 [FREE Full text] [doi: 10.1007/s11948-017-9872-8] [Medline: 28155094]
26. Kaewkannate K, Kim S. A comparison of wearable fitness devices. BMC Public Health 2016 May 24;16(1):433 [FREE Full text] [doi: 10.1186/s12889-016-3059-0] [Medline: 27220855]
27. Khakurel J, Porras J, Melkas H, Fu B. A comprehensive framework of usability issues related to the wearable devices. In: Paiva S, Paul S, editors. Convergence of ICT and Smart Devices for Emerging Applications. Cham, Switzerland: Springer; 2020:21-66.
28. Lamb K, Huang HY, Marturano A, Bashir M. Users' privacy perceptions about wearable technology: examining influence of personality, trust, and usability. In: Nicholson D, editor. Advances in Human Factors in Cybersecurity. Cham, Switzerland: Springer; 2016:55-68.
29. Kushniruk AW, Triola MM, Borycki EM, Stein B, Kannry JL. Technology induced error and usability: the relationship between usability problems and prescription errors when using a handheld application. Int J Med Inform 2005 Aug;74(7-8):519-526 [FREE Full text] [doi: 10.1016/j.ijmedinf.2005.01.003] [Medline: 16043081]
30. Horsky J, Zhang J, Patel VL. To err is not entirely human: complex technology and user cognition. J Biomed Inform 2005 Aug;38(4):264-266 [FREE Full text] [doi: 10.1016/j.jbi.2005.05.002] [Medline: 15967732]
31. Peute LW, Jaspers MW. The significance of a usability evaluation of an emerging laboratory order entry system. Int J Med Inform 2007 Feb;76(2-3):157-168 [FREE Full text] [doi: 10.1016/j.ijmedinf.2006.06.003] [Medline: 16854617]
32. ISO 9241-11:2018: ergonomics of human-system interaction - part 11: usability: definitions and concepts. International Organization for Standardization. 2023. URL: https://www.iso.org/standard/63500.html [accessed 2023-06-12]
33. Dvorak JL. Moving Wearables into the Mainstream: Taming the Borg. New York, NY: Springer; 2008.
34. Baig MM, GholamHosseini H, Moqeem AA, Mirza F, Lindén M. A systematic review of wearable patient monitoring systems - current challenges and opportunities for clinical adoption. J Med Syst 2017 Jul 19;41(7):115 [FREE Full text] [doi: 10.1007/s10916-017-0760-1] [Medline: 28631139]
35. Tsertsidis A, Kolkowska E, Hedström K. Factors influencing seniors' acceptance of technology for ageing in place in the post-implementation stage: a literature review. Int J Med Inform 2019 Sep;129:324-333 [FREE Full text] [doi: 10.1016/j.ijmedinf.2019.06.027] [Medline: 31445274]
36. Göttgens I, Oertelt-Prigione S. The application of human-centered design approaches in health research and innovation: a narrative review of current practices. JMIR Mhealth Uhealth 2021 Dec 06;9(12):e28102 [FREE Full text] [doi: 10.2196/28102] [Medline: 34874893]
37. Roberts JP, Fisher TR, Trowbridge MJ, Bent C. A design thinking framework for healthcare management and innovation. Healthc (Amst) 2016 Mar;4(1):11-14 [FREE Full text] [doi: 10.1016/j.hjdsi.2015.12.002] [Medline: 27001093]
38. Zuber CD, Moody L. Creativity and innovation in health care: tapping into organizational enablers through human-centered design. Nurs Adm Q 2018;42(1):62-75. [doi: 10.1097/NAQ.0000000000000267] [Medline: 29194334]
39. Folmer E, Bosch J. Architecting for usability: a survey. J Syst Softw 2004 Feb;70(1-2):61-78 [FREE Full text] [doi: 10.1016/s0164-1212(02)00159-0]
40. Wixon D, Wilson C. The usability engineering framework for product design and evaluation. In: Helander MG, Landauer TK, Prabhu PV, editors. Handbook of Human-Computer Interaction. Cambridge, MA: Elsevier Academic Press; 1997:653-688.
41. Harrison R, Flood D, Duce D. Usability of mobile applications: literature review and rationale for a new usability model. J Interact Sci 2013;1(1):1 [FREE Full text] [doi: 10.1186/2194-0827-1-1]

42. Siegel J, Bauer M. A field usability evaluation of a wearable system. In: Proceedings of the 1st International Symposium on Wearable Computers. 1997 Presented at: ISWC '97; October 13-14, 1997; Cambridge, MA p. 18-22 URL: https://ieeexplore.ieee.org/document/629914 [doi: 10.1109/iswc.1997.629914]
43. Altenhoff B, Vaigneur H, Caine K. One step forward, two steps back: the key to wearables in the field is the app. In: Proceedings of the 9th International Conference on Pervasive Computing Technologies for Healthcare. 2015 Presented at: ICST '15; May 20-23, 2015; Istanbul, Turkey p. 241-244 URL: https://eudl.eu/doi/10.4108/icst.pervasivehealth.2015.259049 [doi: 10.4108/icst.pervasivehealth.2015.259049]
44. Guillodo E, Lemey C, Simonnet M, Walter M, Baca-García E, Masetti V, HUGOPSY Network, et al. Clinical applications of mobile health wearable-based sleep monitoring: systematic review. JMIR Mhealth Uhealth 2020 Apr 01;8(4):e10733 [FREE Full text] [doi: 10.2196/10733] [Medline: 32234707]
45. Wu M, Jake L. Wearable technology applications in healthcare: a literature review. Healthcare Information and Management Systems Society. 2019. URL: https://www.himss.org/resources/wearable-technology-applications-healthcare-literature-review [accessed 2024-03-05]
46. Huhn S, Axt M, Gunga HC, Maggioni MA, Munga S, Obor D, et al. The impact of wearable technologies in health research: scoping review. JMIR Mhealth Uhealth 2022 Jan 25;10(1):e34384 [FREE Full text] [doi: 10.2196/34384] [Medline: 35076409]
47. Gluck S, Chapple LA, Chapman MJ, Iwashyna TJ, Deane AM. A scoping review of use of wearable devices to evaluate outcomes in survivors of critical illness. Crit Care Resusc 2017 Sep;19(3):197-204.e1. [doi: 10.1016/s1441-2772(23)00917-1]
48. Moore K, O'Shea E, Kenny L, Barton J, Tedesco S, Sica M, et al. Older adults' experiences with using wearable devices: qualitative systematic review and meta-synthesis. JMIR Mhealth Uhealth 2021 Jun 03;9(6):e23832 [FREE Full text] [doi: 10.2196/23832] [Medline: 34081020]
49. Jake-Schoffman DE, Silfee VJ, Waring ME, Boudreaux ED, Sadasivam RS, Mullen SP, et al. Methods for evaluating the content, usability, and efficacy of commercial mobile health apps. JMIR Mhealth Uhealth 2017 Dec 18;5(12):e190 [FREE Full text] [doi: 10.2196/mhealth.8758] [Medline: 29254914]
50. Weichbroth P. Usability of mobile applications: a systematic literature study. IEEE Access 2020;8:55563-55577 [FREE Full text] [doi: 10.1109/access.2020.2981892]
51. Brognara L, Palumbo P, Grimm B, Palmerini L. Assessing gait in Parkinson's disease using wearable motion sensors: a systematic review. Diseases 2019 Feb 05;7(1):18 [FREE Full text] [doi: 10.3390/diseases7010018] [Medline: 30764502]
52. Carreiro S, Newcomb M, Leach R, Ostrowski S, Boudreaux ED, Amante D. Current reporting of usability and impact of mHealth interventions for substance use disorder: a systematic review. Drug Alcohol Depend 2020 Oct 01;215:108201 [FREE Full text] [doi: 10.1016/j.drugalcdep.2020.108201] [Medline: 32777691]
53. Imtiaz SA. A systematic review of sensing technologies for wearable sleep staging. Sensors (Basel) 2021 Feb 24;21(5):1562 [FREE Full text] [doi: 10.3390/s21051562] [Medline: 33668118]
54. Knight S, Lipoth J, Namvari M, Gu C, Hedayati M, Syed-Abdul S, et al. The accuracy of wearable photoplethysmography sensors for telehealth monitoring: a scoping review. Telemed J E Health 2023 Jun 01;29(6):813-828. [doi: 10.1089/tmj.2022.0182] [Medline: 36288566]
55. Rodríguez-Fernández A, Lobo-Prat J, Font-Llagunes JM. Systematic review on wearable lower-limb exoskeletons for gait training in neuromuscular impairments. J Neuroeng Rehabil 2021 Feb 01;18(1):22 [FREE Full text] [doi: 10.1186/s12984-021-00815-5] [Medline: 33526065]
56. Prill R, Walter M, Królikowska A, Becker R. A systematic review of diagnostic accuracy and clinical applications of wearable movement sensors for knee joint rehabilitation. Sensors (Basel) 2021 Dec 09;21(24):8221 [FREE Full text] [doi: 10.3390/s21248221] [Medline: 34960315]
57. Rukasha T, I Woolley S, Kyriacou T, Collins T. Evaluation of wearable electronics for epilepsy: a systematic review. Electronics 2020 Jun 10;9(6):968 [FREE Full text] [doi: 10.3390/electronics9060968]
58. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
59. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci 2010 Sep 20;5(1):69 [FREE Full text] [doi: 10.1186/1748-5908-5-69] [Medline: 20854677]
60. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
61. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
62. Luckmann R. Evidence-based medicine: how to practice and teach EBM, 2nd edition: by David L. Sackett, Sharon E. Straus, W. Scott Richardson, William Rosenberg, and R. Brian Haynes, Churchill Livingstone, 2000. J Intensive Care Med 2016 Dec 07;16(3):155-156 [FREE Full text] [doi: 10.1177/088506660101600307]
63. Roever L. PICO: model for clinical questions. Evid Based Med Pract 2018;3(2):1-2 [FREE Full text] [doi: 10.7717/peerj.5061/table-1]

64. Daudt HM, van Mossel C, Scott SJ. Enhancing the scoping study methodology: a large, inter-professional team's experience with Arksey and O'Malley's framework. BMC Med Res Methodol 2013 Mar 23;13(1):48 [FREE Full text] [doi: 10.1186/1471-2288-13-48] [Medline: 23522333]
65. Seneviratne S, Hu Y, Nguyen T, Lan G, Khalifa S, Thilakarathna K, et al. A Survey of Wearable Devices and Challenges. IEEE Commun Surv Tutor 2017;19(4):2573-2620 [FREE Full text] [doi: 10.1109/comst.2017.2731979]
66. Alonso-Ríos D, Vázquez-García A, Mosqueira-Rey E, Moret-Bonillo V. Usability: a critical analysis and a taxonomy. Int J Hum Comput Interact 2010 Jan 06;26(1):53-74. [doi: 10.1080/10447310903025552]
67. Feiz S, Borodin A, Bi X, Ramakrishnan IV. Towards enabling blind people to fill out paper forms with a wearable smartphone assistant. Proc (Graph Interface) 2021 May;2021:156-165 [FREE Full text] [doi: 10.20380/GI2021.18] [Medline: 35185271]
68. Batsis JA, Zagaria A, Kotz DF, Bartels S, Boateng G, Proctor P, et al. Usability evaluation for the amulet wearable device in rural older adults with obesity. Gerontechnology 2018 Sep;17(3):151-159 [FREE Full text] [doi: 10.4017/gt.2018.17.3.003.00] [Medline: 30631251]
69. Hafiz P, Bardram JE. The ubiquitous cognitive assessment tool for smartwatches: design, implementation, and evaluation study. JMIR Mhealth Uhealth 2020 Jun 01;8(6):e17506 [FREE Full text] [doi: 10.2196/17506] [Medline: 32478664]
70. Lunsford-Avery JR, Keller C, Kollins SH, Krystal AD, Jackson L, Engelhard MM. Feasibility and acceptability of wearable sleep electroencephalogram device use in adolescents: observational study. JMIR Mhealth Uhealth 2020 Oct 01;8(10):e20590 [FREE Full text] [doi: 10.2196/20590] [Medline: 33001035]
71. Manini TM, Mendoza T, Battula M, Davoudi A, Kheirkhahan M, Young ME, et al. Perception of older adults toward smartwatch technology for assessing pain and related patient-reported outcomes: pilot study. JMIR Mhealth Uhealth 2019 Mar 26;7(3):e10044 [FREE Full text] [doi: 10.2196/10044] [Medline: 30912756]
72. Simblett SK, Biondi A, Bruno E, Ballard D, Stoneman A, Lees S, RADAR-CNS consortium. Patients' experience of wearing multimodal sensor devices intended to detect epileptic seizures: a qualitative analysis. Epilepsy Behav 2020 Jan;102:106717 [FREE Full text] [doi: 10.1016/j.yebeh.2019.106717] [Medline: 31785481]
73. Bruno E, Biondi A, Böttcher S, Lees S, Schulze-Bonhage A, Richardson MP, RADAR-CNS Consortium. Day and night comfort and stability on the body of four wearable devices for seizure detection: a direct user-experience. Epilepsy Behav 2020 Nov;112:107478 [FREE Full text] [doi: 10.1016/j.yebeh.2020.107478] [Medline: 33181896]
74. Álvarez R, Torres J, Artola G, Epelde G, Arranz S, Marrugat G. OBINTER: a holistic approach to catalyse the self-management of chronic obesity. Sensors (Basel) 2020 Sep 06;20(18):5060 [FREE Full text] [doi: 10.3390/s20185060] [Medline: 32899921]
75. Argent R, Slevin P, Bevilacqua A, Neligan M, Daly A, Caulfield B. Wearable sensor-based exercise biofeedback for orthopaedic rehabilitation: a mixed methods user evaluation of a prototype system. Sensors (Basel) 2019 Jan 21;19(2):432 [FREE Full text] [doi: 10.3390/s19020432] [Medline: 30669657]
76. Beauchemin M, Gradilla M, Baik D, Cho H, Schnall R. A multi-step usability evaluation of a self-management app to support medication adherence in persons living with HIV. Int J Med Inform 2019 Feb;122:37-44 [FREE Full text] [doi: 10.1016/j.ijmedinf.2018.11.012] [Medline: 30623782]
77. Bentley CL, Powell L, Potter S, Parker J, Mountain GA, Bartlett YK, et al. The use of a smartphone app and an activity tracker to promote physical activity in the management of chronic obstructive pulmonary disease: randomized controlled feasibility study. JMIR Mhealth Uhealth 2020 Jun 03;8(6):e16203 [FREE Full text] [doi: 10.2196/16203] [Medline: 32490838]
78. Bootsman R, Markopoulos P, Qi Q, Wang Q, Timmermans AA. Wearable technology for posture monitoring at the workplace. Int J Hum Comput Stud 2019 Dec;132:99-111 [FREE Full text] [doi: 10.1016/j.ijhcs.2019.08.003]
79. Burgdorf A, Güthe I, Jovanović M, Kutafina E, Kohlschein C, Bitsch J, et al. The mobile sleep lab app: an open-source framework for mobile sleep assessment based on consumer-grade wearable devices. Comput Biol Med 2018 Dec 01;103:8-16 [FREE Full text] [doi: 10.1016/j.compbiomed.2018.09.025] [Medline: 30316065]
80. Du J, Wang Q, Baets LD, Markopoulos P. Supporting shoulder pain prevention and treatment with wearable technology. In: Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare. 2017 Presented at: PervasiveHealth '17; May 23-26, 2017; Barcelona, Spain p. 235-243 URL: https://dl.acm.org/doi/10.1145/3154862.3154886 [doi: 10.1145/3154862.3154886]
81. Dufour S, Fedorkow D, Kun J, Deng SX, Fang Q. Exploring the impact of a mobile health solution for postpartum pelvic floor muscle training: pilot randomized controlled feasibility study. JMIR Mhealth Uhealth 2019 Jul 11;7(7):e12587 [FREE Full text] [doi: 10.2196/12587] [Medline: 31298221]
82. Ajdaraga E, Gusev M, Poposka L. Evaluation of user satisfaction with the ECGalert system. In: Proceedings of the 42nd International Convention on Information and Communication Technology. 2019 Presented at: MIPRO '19; November 22-23, 2019; Opatija, Croatia p. 336-341 URL: https://ieeexplore.ieee.org/document/8756722 [doi: 10.23919/mipro.2019.8756722]
83. Ehn M, Eriksson LC, Åkerberg N, Johansson AC. Activity monitors as support for older persons' physical activity in daily life: qualitative study of the users' experiences. JMIR Mhealth Uhealth 2018 Feb 01;6(2):e34 [FREE Full text] [doi: 10.2196/mhealth.8345] [Medline: 29391342]

84. Hardy J, Veinot TC, Yan X, Berrocal VJ, Clarke P, Goodspeed R, et al. User acceptance of location-tracking technologies in health research: implications for study design and data quality. J Biomed Inform 2018 Mar;79:7-19 [FREE Full text] [doi: 10.1016/j.jbi.2018.01.003] [Medline: 29355784]
85. Hauptmann C, Wegener A, Poppe H, Williams M, Popelka G, Tass PA. Validation of a mobile device for acoustic coordinated reset neuromodulation tinnitus therapy. J Am Acad Audiol 2016 Oct 06;27(9):720-731 [FREE Full text] [doi: 10.3766/jaaa.15082] [Medline: 27718349]
86. Jonker LT, Plas M, de Bock GH, Buskens E, van Leeuwen BL, Lahr MM. Remote home monitoring of older surgical cancer patients: perspective on study implementation and feasibility. Ann Surg Oncol 2021 Jan 29;28(1):67-78 [FREE Full text] [doi: 10.1245/s10434-020-08705-1] [Medline: 32602060]
87. Kwon S, Kim J, Kang S, Lee Y, Baek H, Park K. CardioGuard: a brassiere-based reliable ECG monitoring sensor system for supporting daily smartphone healthcare applications. Telemed J E Health 2014 Dec;20(12):1093-1102 [FREE Full text] [doi: 10.1089/tmj.2014.0008] [Medline: 25405527]
88. Maidment DW, Ferguson M. An application of the medical research council's guidelines for evaluating complex interventions: a usability study assessing smartphone-connected listening devices in adults with hearing loss. Am J Audiol 2018 Nov 19;27(3S):474-481. [doi: 10.1044/2018_aja-imia3-18-0019]
89. Mercer K, Giangregorio L, Schneider E, Chilana P, Li M, Grindrod K. Acceptance of commercially available wearable activity trackers among adults aged over 50 and with chronic illness: a mixed-methods evaluation. JMIR Mhealth Uhealth 2016 Jan 27;4(1):e7 [FREE Full text] [doi: 10.2196/mhealth.4225] [Medline: 26818775]
90. Meyer N, Kerz M, Folarin A, Joyce DW, Jackson R, Karr C, et al. Capturing rest-activity profiles in schizophrenia using wearable and mobile technologies: development, implementation, feasibility, and acceptability of a remote monitoring platform. JMIR Mhealth Uhealth 2018 Oct 30;6(10):e188 [FREE Full text] [doi: 10.2196/mhealth.8292] [Medline: 30377146]
91. Selvaraj N. Long-term remote monitoring of vital signs using a wireless patch sensor. In: Proceedings of the 2014 IEEE Healthcare Innovation Conference. 2014 Presented at: HIC '14; October 8-10, 2014; Seattle, WA p. 83-86 URL: https://ieeexplore.ieee.org/document/7038880 [doi: 10.1109/hic.2014.7038880]
92. Naslund JA, Aschbrenner KA, Bartels SJ. Wearable devices and smartphones for activity tracking among people with serious mental illness. Ment Health Phys Act 2016 Mar;10:10-17 [FREE Full text] [doi: 10.1016/j.mhpa.2016.02.001] [Medline: 27134654]
93. Pavic M, Klaas V, Theile G, Kraft J, Tröster G, Guckenberger M. Feasibility and usability aspects of continuous remote monitoring of health status in palliative cancer patients using wearables. Oncology 2020 Jul 23;98(6):386-395 [FREE Full text] [doi: 10.1159/000501433] [Medline: 31336377]
94. Pulantara IW, Parmanto B, Germain A. Development of a just-in-time adaptive mHealth intervention for insomnia: usability study. JMIR Hum Factors 2018 May 17;5(2):e21 [FREE Full text] [doi: 10.2196/humanfactors.8905] [Medline: 29773529]
95. Reyzelman AM, Koelewyn K, Murphy M, Shen X, Yu E, Pillai R, et al. Continuous temperature-monitoring socks for home use in patients with diabetes: observational study. J Med Internet Res 2018 Dec 17;20(12):e12460 [FREE Full text] [doi: 10.2196/12460] [Medline: 30559091]
96. Salvi E, Bosoni P, Tibollo V, Kruijver L, Calcaterra V, Sacchi L, et al. Patient-generated health data integration and advanced analytics for diabetes management: the AID-GM platform. Sensors (Basel) 2019 Dec 24;20(1):128 [FREE Full text] [doi: 10.3390/s20010128] [Medline: 31878195]
97. Shoneye CL, Mullan B, Begley A, Pollard CM, Jancey J, Kerr DA. Design and development of a digital weight management intervention (ToDAy): qualitative study. JMIR Mhealth Uhealth 2020 Sep 09;8(9):e17919 [FREE Full text] [doi: 10.2196/17919] [Medline: 32641284]
98. Silva de Lima AL, Hahn T, Evers LJ, de Vries NM, Cohen E, Afek M, et al. Feasibility of large-scale deployment of multiple wearable sensors in Parkinson's disease. PLoS One 2017 Dec 20;12(12):e0189161 [FREE Full text] [doi: 10.1371/journal.pone.0189161] [Medline: 29261709]
99. Simons D, de Bourdeaudhuij I, Clarys P, de Cocker K, Vandelanotte C, Deforche B. A smartphone app to promote an active lifestyle in lower-educated working young adults: development, usability, acceptability, and feasibility study. JMIR Mhealth Uhealth 2018 Feb 20;6(2):e44 [FREE Full text] [doi: 10.2196/mhealth.8287] [Medline: 29463491]
100. Sonney J, Duffy M, Hoogerheyde LX, Langhauser E, Teska D. Applying human-centered design to the development of an asthma essentials kit for school-aged children and their parents. J Pediatr Health Care 2019 Mar;33(2):169-177 [FREE Full text] [doi: 10.1016/j.pedhc.2018.07.008] [Medline: 30228032]
101. Thorpe JR, Rønn-Andersen KV, Bień P, Özkil AG, Forchhammer BH, Maier AM. Pervasive assistive technology for people with dementia: a UCD case. Healthc Technol Lett 2016 Dec 02;3(4):297-302 [FREE Full text] [doi: 10.1049/htl.2016.0057] [Medline: 28008366]
102. Tong HL, Coiera E, Tong W, Wang Y, Quiroz JC, Martin P, et al. Efficacy of a mobile social networking intervention in promoting physical activity: quasi-experimental study. JMIR Mhealth Uhealth 2019 Mar 28;7(3):e12181 [FREE Full text] [doi: 10.2196/12181] [Medline: 30920379]
103. Uddin AA, Morita PP, Tallevi K, Armour K, Li J, Nolan RP, et al. Development of a wearable cardiac monitoring system for behavioral neurocardiac training: a usability study. JMIR Mhealth Uhealth 2016 Apr 22;4(2):e45 [FREE Full text] [doi: 10.2196/mhealth.5288] [Medline: 27106171]

104. Ummels D, Braun S, Stevens A, Beekman E, Beurskens A. Measure It Super Simple (MISS) activity tracker: (re)design of a user-friendly interface and evaluation of experiences in daily life. Disabil Rehabil Assist Technol 2022 Oct 24;17(7):767-777 [FREE Full text] [doi: 10.1080/17483107.2020.1815089] [Medline: 32970493]
105. Vaughn J, Gollarahalli S, Shaw RJ, Docherty S, Yang Q, Malhotra C, et al. Mobile health technology for pediatric symptom monitoring: a feasibility study. Nurs Res 2020;69(2):142-148 [FREE Full text] [doi: 10.1097/NNR.0000000000000403] [Medline: 31972852]
106. Wang JB, Cataldo JK, Ayala GX, Natarajan L, Cadmus-Bertram LA, White MM, et al. Mobile and wearable device features that matter in promoting physical activity. J Mob Technol Med 2016 Jul;5(2):2-11 [FREE Full text] [doi: 10.7309/jmtm.5.2.2] [Medline: 27493694]
107. Wulfovich S, Fiordelli M, Rivas H, Concepcion W, Wac K. " I must try harder": design implications for mobile apps and wearables contributing to self-efficacy of patients with chronic conditions. Front Psychol 2019 Oct 23;10:2388 [FREE Full text] [doi: 10.3389/fpsyg.2019.02388] [Medline: 31749733]
108. Chen Y, Ji M, Wu Y, Wang Q, Deng Y, Liu Y, et al. An intelligent individualized cardiovascular app for risk elimination (iCARE) for individuals with coronary heart disease: development and usability testing analysis. JMIR Mhealth Uhealth 2021 Dec 13;9(12):e26439 [FREE Full text] [doi: 10.2196/26439] [Medline: 34898449]
109. Ding EY, Pathiravasan CH, Schramm E, Borrelli B, Liu C, Trinquart L, et al. Design, deployment, and usability of a mobile system for cardiovascular health monitoring within the electronic framingham heart study. Cardiovasc Digit Health J 2021 Jun;2(3):171-178 [FREE Full text] [doi: 10.1016/j.cvdhj.2021.04.001] [Medline: 35265906]
110. Elavsky S, Klocek A, Knapova L, Smahelova M, Smahel D, Cimler R, et al. Feasibility of real-time behavior monitoring via mobile technology in Czech adults aged 50 years and above: 12-week study with ecological momentary assessment. JMIR Aging 2021 Nov 10;4(4):e15220 [FREE Full text] [doi: 10.2196/15220] [Medline: 34757317]
111. Elnaggar A, von Oppenfeld J, Whooley MA, Merek S, Park LG. Applying mobile technology to sustain physical activity after completion of cardiac rehabilitation: acceptability study. JMIR Hum Factors 2021 Sep 02;8(3):e25356 [FREE Full text] [doi: 10.2196/25356] [Medline: 34473064]
112. Fanning J, Brooks AK, Ip E, Nicklas BJ, Rejeski WJ, Nesbit B, et al. A mobile health behavior intervention to reduce pain and improve health in older adults with obesity and chronic pain: the MORPH pilot trial. Front Digit Health 2020 Dec 18;2:598456 [FREE Full text] [doi: 10.3389/fdgth.2020.598456] [Medline: 33817686]
113. Ferguson MA, Maidment DW, Gomez R, Coulson N, Wharrad H. The feasibility of an m-health educational programme (m2Hear) to improve outcomes in first-time hearing aid users. Int J Audiol 2021 Apr 01;60(sup1):S30-S41 [FREE Full text] [doi: 10.1080/14992027.2020.1825839] [Medline: 33135503]
114. Jones SL, Hue W, Kelly RM, Barnett R, Henderson V, Sengupta R. Determinants of longitudinal adherence in smartphone-based self-tracking for chronic health conditions. Proc ACM Interact Mob Wearable Ubiquitous Technol 2021 Mar 30;5(1):1-24. [doi: 10.1145/3448093]
115. Marques MM, Matos M, Mattila E, Encantado J, Duarte C, Teixeira PJ, et al. A theory- and evidence-based digital intervention tool for weight loss maintenance (NoHoW toolkit): systematic development and refinement study. J Med Internet Res 2021 Dec 03;23(12):e25305 [FREE Full text] [doi: 10.2196/25305] [Medline: 34870602]
116. Murdin L, Sladen M, Williams H, Bamiou D, Bibas A, Kikidis D, et al. EHealth and its role in supporting audiological rehabilitation: patient perspectives on barriers and facilitators of using a personal hearing support system with mobile application as part of the EVOTION study. Front Public Health 2021 Jan 14;9:669727 [FREE Full text] [doi: 10.3389/fpubh.2021.669727] [Medline: 35118034]
117. Petracca F, Tempre R, Cucciniello M, Ciani O, Pompeo E, Sannino L, et al. An electronic patient-reported outcome mobile app for data collection in type a hemophilia: design and usability study. JMIR Form Res 2021 Dec 01;5(12):e25071 [FREE Full text] [doi: 10.2196/25071] [Medline: 34855619]
118. Piau A, Steinmeyer Z, Charlon Y, Courbet L, Rialle V, Lepage B, et al. A smart shoe insole to monitor frail older adults' walking speed: results of two evaluation phases completed in a living lab and through a 12-week pilot study. JMIR Mhealth Uhealth 2021 Jul 05;9(7):e15641 [FREE Full text] [doi: 10.2196/15641] [Medline: 36260404]
119. Santala OE, Halonen J, Martikainen S, Jäntti H, Rissanen TT, Tarvainen MP, et al. Automatic mobile health arrhythmia monitoring for the detection of atrial fibrillation: prospective feasibility, accuracy, and user experience study. JMIR Mhealth Uhealth 2021 Oct 22;9(10):e29933 [FREE Full text] [doi: 10.2196/29933] [Medline: 34677135]
120. Ter Harmsel A, van der Pol T, Swinkels L, Goudriaan AE, Popma A, Noordzij ML. Development of a wearable biocueing app (Sense-IT) among forensic psychiatric outpatients with aggressive behavior: design and evaluation study. JMIR Form Res 2021 Nov 24;5(11):e29267 [FREE Full text] [doi: 10.2196/29267] [Medline: 34821567]
121. To QG, Green C, Vandelanotte C. Feasibility, usability, and effectiveness of a machine learning-based physical activity chatbot: quasi-experimental study. JMIR Mhealth Uhealth 2021 Nov 26;9(11):e28577 [FREE Full text] [doi: 10.2196/28577] [Medline: 34842552]
122. Bruno E, Biondi A, Thorpe S, Richardson M, RADAR-CNS Consortium. Patients self-mastery of wearable devices for seizure detection: a direct user-experience. Seizure 2020 Oct;81:236-240 [FREE Full text] [doi: 10.1016/j.seizure.2020.08.023] [Medline: 32889512]

123. Bae YS, Kim KH, Choi SW, Ko T, Lim JS, Piao M. Satisfaction and usability of an information and communications technology-based system by clinically healthy patients with COVID-19 and medical professionals: cross-sectional survey and focus group interview study. JMIR Form Res 2021 Aug 26;5(8):e26227 [FREE Full text] [doi: 10.2196/26227] [Medline: 34254946]
124. Chen K, Dandapani H, Guthrie KM, Goldberg E. Can older adult emergency department patients successfully use the apple watch to monitor health? R I Med J (2013) 2021 Aug 02;104(6):49-54 [FREE Full text] [Medline: 34323880]
125. Jimah T, Borg H, Kehoe P, Pimentel P, Turner A, Labbaf S, et al. A technology-based pregnancy health and wellness intervention (two happy hearts): case study. JMIR Form Res 2021 Nov 17;5(11):e30991 [FREE Full text] [doi: 10.2196/30991] [Medline: 34787576]
126. Shelby T, Caruthers T, Kanner OY, Schneider R, Lipnickas D, Grau LE, et al. Pilot evaluations of two bluetooth contact tracing approaches on a university campus: mixed methods study. JMIR Form Res 2021 Oct 28;5(10):e31086 [FREE Full text] [doi: 10.2196/31086] [Medline: 34586078]
127. Villalba E, Salvi D, Peinado I, Ottaviano M, Arredondo MT. Validation results of the user interaction in a heart failure management system. In: Proceedings of the 2009 International Conference on eHealth, Telemedicine, and Social Medicine. 2009 Presented at: eTELEMED '09; February 1-7, 2009; Cancun, Mexico p. 81-86 URL: https://ieeexplore.ieee.org/document/4782637 [doi: 10.1109/etelemed.2009.41]
128. Liang SF, Kuo CE, Lee YC, Lin WC, Liu YC, Chen PY, et al. Development of an EOG-based automatic sleep-monitoring eye mask. IEEE Trans Instrum Meas 2015 Nov;64(11):2977-2985. [doi: 10.1109/tim.2015.2433652]
129. Merilahti J, Pärkkä J, Antila K, Paavilainen P, Mattila E, Malm E, et al. Compliance and technical feasibility of long-term health monitoring with wearable and ambient technologies. J Telemed Telecare 2009 Aug 31;15(6):302-309 [FREE Full text] [doi: 10.1258/jtt.2009.081106] [Medline: 19720768]
130. Lawson S, Jamison-Powell S, Garbett A. Validating a mobile phone application for the everyday, unobtrusive, objective measurement of sleep. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2013 Presented at: CHI '13; April 27-May 2, 2013; Paris, France p. 2497-2506 URL: https://dl.acm.org/doi/10.1145/2470654.2481345 [doi: 10.1145/2470654.2481345]
131. Comello ML, Porter JH. Concept test of a smoking cessation smart case. Telemed J E Health 2018 Dec;24(12):1036-1040 [FREE Full text] [doi: 10.1089/tmj.2017.0196] [Medline: 29621000]
132. Bouça-Machado R, Pona-Ferreira F, Leitão M, Clemente A, Vila-Viçosa D, Kauppila LA, et al. Feasibility of a mobile-based system for unsupervised monitoring in Parkinson's disease. Sensors (Basel) 2021 Jul 21;21(15):4972 [FREE Full text] [doi: 10.3390/s21154972] [Medline: 34372208]
133. Lewinski AA, Vaughn J, Diane A, Barnes A, Crowley MJ, Steinberg D, et al. Perceptions of using multiple mobile health devices to support self-management among adults with type 2 diabetes: a qualitative descriptive study. J Nurs Scholarsh 2021 Sep 29;53(5):643-652 [FREE Full text] [doi: 10.1111/jnu.12667] [Medline: 33928755]
134. ISO/IEC 25010:2011: systems and software engineering: systems and software quality requirements and evaluation (SQuaRE): system and software quality models. International Organization for Standardization. 2011. URL: https://www.iso.org/standard/35733.html [accessed 2023-06-21]
135. Global strategy on digital health 2020-2025. World Health Organization. URL: https://www.who.int/docs/default-source/documents/gs4dhdaa2a9f352b0445bafbc79ca799dce4d.pdf [accessed 2024-03-05]
136. Bryson D. Unwearables. AI Soc 2006 Dec 4;22(1):25-35 [FREE Full text] [doi: 10.1007/s00146-006-0072-3]
137. Rupp MA, Michaelis JR, McConnell DS, Smither JA. The role of individual differences on perceptions of wearable fitness device trust, usability, and motivational impact. Appl Ergon 2018 Jul;70:77-87 [FREE Full text] [doi: 10.1016/j.apergo.2018.02.005] [Medline: 29866329]
138. Sun N, Rau PL. The acceptance of personal health devices among patients with chronic conditions. Int J Med Inform 2015 Apr;84(4):288-297 [FREE Full text] [doi: 10.1016/j.ijmedinf.2015.01.002] [Medline: 25655783]
139. Chiauzzi E, Rodarte C, DasMahapatra P. Patient-centered activity monitoring in the self-management of chronic health conditions. BMC Med 2015 Apr 09;13(1):77 [FREE Full text] [doi: 10.1186/s12916-015-0319-2] [Medline: 25889598]
140. Steinert A, Haesner M, Steinhagen-Thiessen E. Activity-tracking devices for older adults: comparison and preferences. Univ Access Inf Soc 2017 Apr 8;17(2):411-419. [doi: 10.1007/s10209-017-0539-7]
141. Huberty J, Ehlers DK, Kurka J, Ainsworth B, Buman M. Feasibility of three wearable sensors for 24 hour monitoring in middle-aged women. BMC Womens Health 2015 Jul 30;15(1):55 [FREE Full text] [doi: 10.1186/s12905-015-0212-3] [Medline: 26223521]
142. Preusse KC, Mitzner TL, Fausset CB, Rogers WA. Older adults' acceptance of activity trackers. J Appl Gerontol 2017 Feb 07;36(2):127-155 [FREE Full text] [doi: 10.1177/0733464815624151] [Medline: 26753803]
143. Keogh A, Dorn JF, Walsh L, Calvo F, Caulfield B. Comparing the usability and acceptability of wearable sensors among older Irish adults in a real-world context: observational study. JMIR Mhealth Uhealth 2020 Apr 20;8(4):e15704 [FREE Full text] [doi: 10.2196/15704] [Medline: 32310149]
144. Keogh A, Argent R, Anderson A, Caulfield B, Johnston W. Assessing the usability of wearable devices to measure gait and physical activity in chronic conditions: a systematic review. J Neuroeng Rehabil 2021 Sep 15;18(1):138 [FREE Full text] [doi: 10.1186/s12984-021-00931-2] [Medline: 34526053]

145. Chakrabarti S, Biswas N, Jones LD, Kesari S, Ashili S. Smart consumer wearables as digital diagnostic tools: a review. Diagnostics (Basel) 2022 Aug 31;12(9):2110 [FREE Full text] [doi: 10.3390/diagnostics12092110] [Medline: 36140511]
146. Piwek L, Ellis DA, Andrews S, Joinson A. The rise of consumer health wearables: promises and barriers. PLoS Med 2016 Feb 2;13(2):e1001953 [FREE Full text] [doi: 10.1371/journal.pmed.1001953] [Medline: 26836780]
147. Roman DH, Abbott I, Conlee KD, Jones R, Noble A, Rich N. The digital revolution comes to US healthcare. Goldman Sachs. URL: https://www.anderson.ucla.edu/documents/areas/adm/acis/library/DigitalRevolutionGS.pdf [accessed 2024-03-05]
148. Lee SY, Lee K. Factors that influence an individual's intention to adopt a wearable healthcare device: the case of a wearable fitness tracker. Technol Forecast Soc Change 2018 Apr;129:154-163 [FREE Full text] [doi: 10.1016/j.techfore.2018.01.002]
149. Wilson HJ. Wearables in the workplace. Harvard Business Review. 2013. URL: https://hbr.org/2013/09/wearables-in-the-workplace [accessed 2024-03-05]
150. Ye S, Feng S, Huang L, Bian S. Recent progress in wearable biosensors: from healthcare monitoring to sports analytics. Biosensors (Basel) 2020 Dec 15;10(12):205 [FREE Full text] [doi: 10.3390/bios10120205] [Medline: 33333888]
151. Niknejad N, Ismail WB, Mardani A, Liao H, Ghani I. A comprehensive overview of smart wearables: the state of the art literature, recent advances, and future challenges. Eng Appl Artif Intell 2020 Apr;90:103529 [FREE Full text] [doi: 10.1016/j.engappai.2020.103529]
152. Ferreira JJ, Fernandes CI, Rammal HG, Veiga PM. Wearable technology and consumer interaction: a systematic review and research agenda. Comput Human Behav 2021 May;118:106710 [FREE Full text] [doi: 10.1016/j.chb.2021.106710]
153. Henriksen A, Haugen Mikalsen M, Woldaregay AZ, Muzny M, Hartvigsen G, Hopstock LA, et al. Using fitness trackers and smartwatches to measure physical activity in research: analysis of consumer wrist-worn wearables. J Med Internet Res 2018 Mar 22;20(3):e110 [FREE Full text] [doi: 10.2196/jmir.9157] [Medline: 29567635]
154. Edney S, Bogomolova S, Ryan J, Olds T, Sanders I, Maher C. Creating engaging health promotion campaigns on social media: observations and lessons from Fitbit and Garmin. J Med Internet Res 2018 Dec 10;20(12):e10911 [FREE Full text] [doi: 10.2196/10911] [Medline: 30530449]
155. Nguyen NH, Hadgraft NT, Moore MM, Rosenberg DE, Lynch C, Reeves MM, et al. A qualitative evaluation of breast cancer survivors' acceptance of and preferences for consumer wearable technology activity trackers. Support Care Cancer 2017 Nov 24;25(11):3375-3384 [FREE Full text] [doi: 10.1007/s00520-017-3756-y] [Medline: 28540402]
156. Maher C, Ryan J, Ambrosi C, Edney S. Users' experiences of wearable activity trackers: a cross-sectional study. BMC Public Health 2017 Nov 15;17(1):880 [FREE Full text] [doi: 10.1186/s12889-017-4888-1] [Medline: 29141607]
157. Shin G, Jarrahi MH, Fei Y, Karami A, Gafinowitz N, Byun A, et al. Wearable activity trackers, accuracy, adoption, acceptance and health impact: a systematic literature review. J Biomed Inform 2019 May;93:103153 [FREE Full text] [doi: 10.1016/j.jbi.2019.103153] [Medline: 30910623]
158. Sui A, Sui W, Liu S, Rhodes R. Ethical considerations for the use of consumer wearables in health research. Digit Health 2023 Feb 01;9:20552076231153740 [FREE Full text] [doi: 10.1177/20552076231153740] [Medline: 36756643]
159. Tu J, Gao W. Ethical considerations of wearable technologies in human research. Adv Healthc Mater 2021 Sep 18;10(17):e2100127 [FREE Full text] [doi: 10.1002/adhm.202100127] [Medline: 33870653]
160. Lu L, Zhang J, Xie Y, Gao F, Xu S, Wu X, et al. Wearable health devices in health care: narrative systematic review. JMIR Mhealth Uhealth 2020 Nov 09;8(11):e18907 [FREE Full text] [doi: 10.2196/18907] [Medline: 33164904]
161. Cilliers L. Wearable devices in healthcare: privacy and information security issues. Health Inf Manag 2020 May 30;49(2-3):150-156. [doi: 10.1177/1833358319851684] [Medline: 31146589]
162. Dimou E, Manavis A, Papachristou E, Kyratsis P. A conceptual design of intelligent shoes for pregnant women. In: Proceedings of IT4Fashion on Business Models and ICT Technologies for the Fashion Supply Chain. 2016 Presented at: IT4Fashion '16; April 20-22, 2016; Florence, Italy p. 69-77 URL: https://link.springer.com/chapter/10.1007/978-3-319-48511-9_6 [doi: 10.1007/978-3-319-48511-9_6]
163. Yang H, Yu J, Zo H, Choi M. User acceptance of wearable devices: an extended perspective of perceived value. Telemat Inform 2016 May;33(2):256-269 [FREE Full text] [doi: 10.1016/j.tele.2015.08.007]
164. Meyer JT, Gassert R, Lambercy O. An analysis of usability evaluation practices and contexts of use in wearable robotics. J Neuroeng Rehabil 2021 Dec 09;18(1):170 [FREE Full text] [doi: 10.1186/s12984-021-00963-8] [Medline: 34886902]
165. Bakhshian S, Lee YA. Holistic integration of product attributes with consumer behavioral aspects for the use of wearable technology. Int Text Appar Assoc Annu Conf Proc 2018;75(1) [FREE Full text]
166. Lin CT, Ko LW, Chang MH, Duann JR, Chen JY, Su TP, et al. Review of wireless and wearable electroencephalogram systems and brain-computer interfaces--a mini-review. Gerontology 2010 Jul 25;56(1):112-119 [FREE Full text] [doi: 10.1159/000230807] [Medline: 19641287]
167. Koo SH, Gaul K, Rivera S, Pan T, Fong D. Wearable technology design for autism spectrum disorders. Arch Des Res 2018 Feb 28;31(1):37-55. [doi: 10.15187/adr.2018.02.31.1.37]
168. Aroganam G, Manivannan N, Harrison D. Review on wearable technology sensors used in consumer sport applications. Sensors (Basel) 2019 Apr 28;19(9):1983 [FREE Full text] [doi: 10.3390/s19091983] [Medline: 31035333]
169. Al-Sada M, Jiang K, Ranade S, Kalkattawi M, Nakajima T. HapticSnakes: multi-haptic feedback wearable robots for immersive virtual reality. Virtual Reality 2019 Sep 30;24(2):191-209 [FREE Full text] [doi: 10.1007/s10055-019-00404-x]

170. Liew MS, Zhang J, See J, Ong YL. Usability challenges for health and wellness mobile apps: mixed-methods study among mHealth experts and consumers. JMIR Mhealth Uhealth 2019 Jan 30;7(1):e12160 [FREE Full text] [doi: 10.2196/12160] [Medline: 30698528]
171. Gupta K, Roy S, Poonia RC, Nayak SR, Kumar R, Alzahrani KJ, et al. Evaluating the usability of mHealth applications on type 2 diabetes mellitus using various MCDM methods. Healthcare (Basel) 2021 Dec 21;10(1):4 [FREE Full text] [doi: 10.3390/healthcare10010004] [Medline: 35052167]
172. Azad-Khaneghah P, Neubauer N, Miguel Cruz A, Liu L. Mobile health app usability and quality rating scales: a systematic review. Disabil Rehabil Assist Technol 2021 Oct 08;16(7):712-721. [doi: 10.1080/17483107.2019.1701103] [Medline: 31910687]
173. Alzahrani AS, Gay V, Alturki R, AlGhamdi MJ. Towards understanding the usability attributes of AI-enabled eHealth mobile applications. J Healthc Eng 2021 Dec 21;2021:5313027-5313028 [FREE Full text] [doi: 10.1155/2021/5313027] [Medline: 34970424]
174. Zapata BC, Fernández-Alemán JL, Idri A, Toval A. Empirical studies on usability of mHealth apps: a systematic literature review. J Med Syst 2015 Feb 20;39(2):1. [doi: 10.1007/s10916-014-0182-2] [Medline: 25600193]
175. Alturki R, Gay V. Usability attributes for mobile applications: a systematic review. In: Ahmad Jan M, Khan F, Alam M, editors. Recent Trends and Advances in Wireless and IoT-enabled Networks. Cham, Switzerland: Springer; 2019:53-62.
176. Johnson SG, Potrebny T, Larun L, Ciliska D, Olsen NR. Usability methods and attributes reported in usability studies of mobile apps for health care education: scoping review. JMIR Med Educ 2022 Jun 29;8(2):e38259 [FREE Full text] [doi: 10.2196/38259] [Medline: 35767323]
177. Cho H, Flynn G, Saylor M, Gradilla M, Schnall R. Use of the FITT framework to understand patients' experiences using a real-time medication monitoring pill bottle linked to a mobile-based HIV self-management app: a qualitative study. Int J Med Inform 2019 Nov;131:103949 [FREE Full text] [doi: 10.1016/j.ijmedinf.2019.08.009] [Medline: 31561192]
178. Henson P, David G, Albright K, Torous J. Deriving a practical framework for the evaluation of health apps. Lancet Digit Health 2019 Jun;1(2):e52-e54. [doi: 10.1016/s2589-7500(19)30013-5]
179. Connelly K, Molchan H, Bidanta R, Siddh S, Lowens B, Caine K, et al. Evaluation framework for selecting wearable activity monitors for research. Mhealth 2021 Jan;7:6-6 [FREE Full text] [doi: 10.21037/mhealth-19-253] [Medline: 33634189]
180. Wang Y, Yu S, Wang J, Ma N, Liu Z. Introducing an evaluation framework for wearabel devices design: explain reasons of low user adoption. In: Proceedings of the DESIGN 15th International Design Conference. 2018 Presented at: IDC '15; August 22, 2015; Seattle, WA p. 2357-2368 URL: https://www.designsociety.org/publication/40630/INTRODUCING+AN+EVALUATION+FRAMEWORK+FOR+WEARABLE+DEVICES+DESIGN%3A+EXPLAIN+REASONS+OF+LOW+USER+ADOPTION [doi: 10.21278/idc.2018.0179]
181. Caulfield B, Reginatto B, Slevin P. Not all sensors are created equal: a framework for evaluating human performance measurement technologies. NPJ Digit Med 2019 Feb 14;2(1):7 [FREE Full text] [doi: 10.1038/s41746-019-0082-4] [Medline: 31304357]
182. Kim YW, Yoon SH, Hwangbo H, Ji YG. Development of a user experience evaluation framework for wearable devices. In: Proceedings of the 3rd International Conference, ITAP 2017, Held as Part of HCI International conference on Human Aspects of IT for the Aged Population. Applications, Services and Contexts. 2017 Presented at: ITAP '17; July 9-14, 2017; Vancouver, BC p. 53-67 URL: https://link.springer.com/chapter/10.1007/978-3-319-58536-9_5 [doi: 10.1007/978-3-319-58536-9_5]
183. Purnell L, Sierra M, Lisker S, Lim MS, Bailey E, Sarkar U, et al. Acceptability and usability of a wearable device for sleep health among English- and Spanish-speaking patients in a safety net clinic: qualitative analysis. JMIR Form Res 2023 Jun 05;7:e43067 [FREE Full text] [doi: 10.2196/43067] [Medline: 37098152]
184. Voskuil VR, Stroup S, Leyden M. Acceptability and usability of a wearable activity tracker and application among inactive adolescent girls. Phys Act Health 2020;4(1):52-61 [FREE Full text] [doi: 10.5334/paah.51]
185. Thilo FJ, Hahn S, Halfens RJ, Schols JM. Usability of a wearable fall detection prototype from the perspective of older people-a real field testing approach. J Clin Nurs 2019 Jan;28(1-2):310-320. [doi: 10.1111/jocn.14599] [Medline: 29964344]
186. McCallum C, Rooksby J, Gray CM. Evaluating the impact of physical activity apps and wearables: interdisciplinary review. JMIR Mhealth Uhealth 2018 Mar 23;6(3):e58 [FREE Full text] [doi: 10.2196/mhealth.9054] [Medline: 29572200]
187. Ding EY, CastañedaAvila M, Tran KV, Mehawej J, Filippaios A, Paul T, et al. Usability of a smartwatch for atrial fibrillation detection in older adults after stroke. Cardiovasc Digit Health J 2022 Jun;3(3):126-135 [FREE Full text] [doi: 10.1016/j.cvdhj.2022.03.003] [Medline: 35720675]
188. O'Reilly MA, Slevin P, Ward T, Caulfield B. A wearable sensor-based exercise biofeedback system: mixed methods evaluation of formulift. JMIR Mhealth Uhealth 2018 Jan 31;6(1):e33 [FREE Full text] [doi: 10.2196/mhealth.8115] [Medline: 29386171]
189. Toh SF, Gonzalez PC, Fong KN. Usability of a wearable device for home-based upper limb telerehabilitation in persons with stroke: a mixed-methods study. Digit Health 2023 Feb 07;9:20552076231153737 [FREE Full text] [doi: 10.1177/20552076231153737] [Medline: 36776407]
190. Sauro J, Lewis JR. Quantifying the User Experience: Practical Statistics for User Research. 2nd edition. Cambridge, MA: Morgan Kaufmann; 2016.

191. Brooke J. SUS: a retrospective. J Usability Stud 2013;8(2):29-40 [FREE Full text]
192. Liang J, Xian D, Liu X, Fu J, Zhang X, Tang B, et al. Usability study of mainstream wearable fitness devices: feature analysis and system usability scale evaluation. JMIR Mhealth Uhealth 2018 Nov 08;6(11):e11066 [FREE Full text] [doi: 10.2196/11066] [Medline: 30409767]
193. Moumane K, Idri A, Abran A. Usability evaluation of mobile applications using ISO 9241 and ISO 25062 standards. SpringerPlus 2016 Apr 29;5(1):548 [FREE Full text] [doi: 10.1186/s40064-016-2171-z] [Medline: 27190747]
194. Schnall R, Cho H, Liu J. Health information technology usability evaluation scale (Health-ITUES) for usability assessment of mobile health technology: validation study. JMIR Mhealth Uhealth 2018 Jan 05;6(1):e4 [FREE Full text] [doi: 10.2196/mhealth.8851] [Medline: 29305343]
195. Gimhae GN. Six human factors to acceptability of wearable computers. Int J Multim Ubiquit Eng 2013;8(3):130-114 [FREE Full text]
196. Rosenthal J, Edwards N, Villanueva D, Krishna S, McDaniel T, Panchanathan S. Design, implementation, and case study of a pragmatic vibrotactile belt. IEEE Trans Instrum Meas 2011 Jan;60(1):114-125. [doi: 10.1109/tim.2010.2065830]
197. Borsci S, Federici S, Malizia A, De Filippis ML. Shaking the usability tree: why usability is not a dead end, and a constructive way forward. Behav Inf Technol 2018 Nov 02;38(5):519-532. [doi: 10.1080/0144929x.2018.1541255]
198. Francés-Morcillo L, Morer-Camo P, Rodríguez-Ferradas MI, Cazón-Martín A. The role of user-centred design in smart wearable systems design process. In: Proceedings of the DESIGN 15th International Design Conference. 2018 Presented at: IDC '18; September 19-22, 2018; New Orleans, LA p. 2197-2208 URL: https://www.designsociety.org/publication/40616/THE+ROLE+OF+USER-CENTRED+DESIGN+IN+SMART+WEARABLE+SYSTEMS+DESIGN+PROCESS [doi: 10.21278/idc.2018.0405]
199. Jones J, Gouge C, Crilley M. Design principles for health wearables. Commun Des Q Rev 2017 Aug 04;5(2):40-50. [doi: 10.1145/3131201.3131205]

Abbreviations

ISO: International Organization for Standardization
mHealth: mobile health
PICO: Population, Intervention, Comparison, and Outcome
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
RQ: research question
SUS: System Usability Scale

Edited by L Buis; submitted 25.08.23; peer-reviewed by P Hummelsberger, A Keogh, K Wenderott; comments to author 17.10.23; revised version received 15.12.23; accepted 01.02.24; published 05.04.24.

Please cite as:
Moorthy P, Weinert L, Schüttler C, Svensson L, Sedlmayr B, Müller J, Nagel T
Attributes, Methods, and Frameworks Used to Evaluate Wearables and Their Companion mHealth Apps: Scoping Review
JMIR Mhealth Uhealth 2024;12:e52179
URL: https://mhealth.jmir.org/2024/1/e52179
doi:10.2196/52179
PMID:38578671

Review

Users' Acceptability and Perceived Efficacy of mHealth for Opioid Use Disorder: Scoping Review

Lynnette Nathalie Lyzwinski[1], MPhil, PhD; Mohamed Elgendi[2], PhD; Carlo Menon[1,2], PhD

[1]Menrva Research Group, School of Mechatronics Systems Engineering and Engineering Science, Simon Fraser University, Vancouver, BC, Canada
[2]ETH Biomedical and Mobile Health Technology Lab, Zurich, Switzerland

Corresponding Author:
Lynnette Nathalie Lyzwinski, MPhil, PhD
Menrva Research Group, School of Mechatronics Systems Engineering and Engineering Science
Simon Fraser University
250-13450 102 Avenue
Metro Vancouver
Vancouver, BC, V3T 0A3
Canada
Phone: 1 778 302 1151
Email: lnl2@sfu.ca

Abstract

Background: The opioid crisis continues to pose significant challenges to global public health, necessitating the development of novel interventions to support individuals in managing their substance use and preventing overdose-related deaths. Mobile health (mHealth), as a promising platform for addressing opioid use disorder, requires a comprehensive understanding of user perspectives to minimize barriers to care and optimize the benefits of mHealth interventions.

Objective: This study aims to synthesize qualitative insights into opioid users' acceptability and perceived efficacy of mHealth and wearable technologies for opioid use disorder.

Methods: A scoping review of PubMed (MEDLINE) and Google Scholar databases was conducted to identify research on opioid user perspectives concerning mHealth-assisted interventions, including wearable sensors, SMS text messaging, and app-based technology.

Results: Overall, users demonstrate a high willingness to engage with mHealth interventions to prevent overdose-related deaths and manage opioid use. Users perceive mHealth as an opportunity to access care and desire the involvement of trusted health care professionals in these technologies. User comfort with wearing opioid sensors emerged as a significant factor. Personally tailored content, social support, and encouragement are preferred by users. Privacy concerns and limited access to technology pose barriers to care.

Conclusions: To maximize benefits and minimize risks for users, it is crucial to implement robust privacy measures, provide comprehensive user training, integrate behavior change techniques, offer professional and peer support, deliver tailored messages, incorporate behavior change theories, assess readiness for change, design stigma-reducing apps, use visual elements, and conduct user-focused research for effective opioid management in mHealth interventions. mHealth demonstrates considerable potential as a tool for addressing opioid use disorder and preventing overdose-related deaths, given the high acceptability and perceived benefits reported by users.

(JMIR Mhealth Uhealth 2024;12:e49751) doi:10.2196/49751

KEYWORDS

acceptability; addict; addiction; addictions; app; app-based; application; applications; apps; barrier; barriers; challenge; challenges; messaging; mHealth; mobile health; monitoring; opioid; opioids; overdose; overdosing; pharmacology; review methodology; review methods; scoping; sensor; sensors; SMS; substance abuse; substance use; text message; wearable technology; wearable; wearables

Introduction

Overview

Since the COVID-19 pandemic, the incidence of opioid overdose has increased significantly among young adults [1,2]. According to research conducted in Canada, there was a 135% increase in opioid overdose–related deaths per week in Ontario during COVID-19 in comparison to prepandemic times [2]. Excess levels of opioids may lead to respiratory depression and cardiopulmonary failure, resulting in loss of consciousness and death [3,4]. It is treated with naloxone, an antagonist that prevents opioid-related fatalities [5,6]. Thus, it is essential to find ways to manage opioid use disorder (OUD) and prevent overdose. A study reviewing smartphone apps that were commercially available until 2019 found that there was a lack of evidence on their ability to be useful for monitoring OUD [7]. The study's authors also reviewed interventional studies involving mobile health (mHealth) for opioid use; most of the reviewed studies included smartphone apps or personal digital assistants, and only a minority of those devices had a wearable biosensor [7]. The authors concluded that there is a gap in the literature and that more studies are needed to address OUD [7].

Since that study was published, there have been technological developments in this area to meet the pressing need that arose during the COVID-19 pandemic. Specifically, new wearable sensors have been developed that monitor sweat, heart rate, and temperature and that predict overdose in patients [7-13]. Several studies have evaluated the application of wearable opioid sensors for detecting overdose by detecting changes in temperature, movement, respiratory rate, heart rate, and electrodermal activity [8,11-17]. Most of the studies involved a wrist-worn biosensor, usually the Empatica E4 biosensor and Q sensor, which integrate a machine learning algorithm. A few studies found that wearable technology can detect changes in skin temperature, notably a temperature rise during opioid intake and overdose [8,12-14,16,17]. Carreiro et al [14] found that the mean skin temperature increase after opioid intake was 2.62 °C and that differences in body temperature before and after intake were significant (P<.01). In addition to temperature changes, several studies found reduced movement or locomotion by evaluating triaxis acceleration data [8,11,13,14,16,17]. Notably, local extremities, such as the fingers, had reduced motion or there was less "fidgeting" after opioid intake. In 1 study, fidgeting was observed more in heavy opioid users than less heavy users [14]. Emerging research also suggests that these devices may be valid for detecting opioid intake and overdose [8,11,16]. For example, Mahmud et al [11] found that the Q Affectiva sensor had an accuracy of 99% for detecting opioid intake in users.

Although emerging technological research in the wearables and mHealth domain suggests that these devices may be promising for detecting opioid overdose and assisting with OUD, little is presently known about the acceptability of this form of monitoring from the patients' perspectives. In other words, a review of qualitative studies on consumer perspectives has not been undertaken. While a device may work in theory, it may fail to be an effective intervention and be implemented within the community care setting if the technology is not acceptable for opioid users or if they do not adhere to the technological intervention. While the quantitative efficacy of these devices has been established in a past review [7], there is a need to understand how usable these devices are for patients. Often, many patients with substance use disorders are difficult to reach and treat due to marginalization or stigma [18-20], making them resistant to accessing medical help [21]. Thus, there is a compelling research interest to determine how patients and opioid users feel about the emerging technological developments that may assist them in managing their substance use and preventing overdose fatalities.

Aims

This study aimed to undertake a review of opioid users' perspectives on mHealth and wearable technology for OUD and overdose management, with a focus on the latest technology developed over the last 5 years. We aimed to better understand whether users found the technology acceptable to use and helpful. We also aimed to better understand the perceived barriers as well as the benefits of using wearables for OUD and opioid overdose management among opioid-using populations. This study focused on qualitative research in order to better understand consumer experiences with wearable technology for OUD, with the overarching goal of making recommendations for future technological development and intervention design.

Methods

Study Guidelines

A scoping review was performed to summarize the use of wearable devices in managing OUD to better understand the key benefits, preferences, and barriers of use. The guidelines of the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) [22] were followed.

Search Strategy and Study Eligibility

The PubMed (MEDLINE) and Google Scholar databases were searched for all studies published until July 2022 using the search strategy in Textbox 1.

Textbox 1. The search strategy used to search for studies.

"Wearable Electronic Devices"[Mesh] OR wearable*[tiab] OR "Smart Band*"[tiab] OR "Smart Watch*"[tiab] OR mHealth OR "mobile health" OR app OR application OR wristband OR ((sensor*[tiab] OR sensing[tiab] OR biosensor*[tiab]) AND (wear*[tiab] OR worn[tiab]))

(("Analgesics, Opioid"[Mesh] OR "Analgesics, Opioid" [Pharmacological Action] OR Opiate*[TIAB] OR Opioid*[TIAB] OR "fentanyl"[tiab] OR "hydromorphone"[tiab] OR "meperidine"[tiab] OR "morphine"[tiab] OR "oxycodone"[tiab] OR "pentazocine"[tiab] OR "sufentanil"[tiab] OR "tramadol"[tiab] OR "morphine"[MeSH Terms] OR "hydrocodone"[MeSH Terms] OR "hydrocodone"[tiab] OR "buprenorphine"[MeSH Terms] OR "buprenorphine"[tiab] OR "codeine "[MeSH Terms] OR "fentanyl"[MeSH Terms] OR "hydromorphone"[MeSH Terms] OR "meperidine"[MeSH Terms] OR "oxycodone"[MeSH Terms] OR "pentazocine"[MeSH Terms] OR "sufentanil"[MeSH Terms] OR "tramadol"[MeSH Terms] OR OxyContin[tiab] OR Vicodin[tiab] OR "Codeine"[Mesh] OR codeine[tiab] OR morphine[tiab]) AND ("substance use disorder"[tiab] OR "substance abuse"[tiab] OR disorder[tiab] OR abuse[tiab] OR use[tiab])) OR "Opioid-Related Disorders"[Mesh] OR "Opiate use disorder"[tiab] OR "opioid use disorder"[tiab] OR "opiate dependence"[tiab] OR "opioid dependence"[tiab] OR "opiate abuse"[tiab] OR "opioid abuse"[tiab]

Inclusion and Exclusion Criteria

All qualitative studies that evaluated opioid users' perspectives on the latest mHealth wearable devices or portable mHealth devices (including apps and SMS text messaging–based interventions) for monitoring opioid use were included in the review. Quantitative studies were only included if they had a qualitative component or if they had collected more detailed information on acceptability, benefits, and barriers associated with use. The technology must have been emerging and recently developed (within the past 5 years, with a focus on technology during the pandemic). Web-based interventions were excluded if they did not include a wearable or mHealth element.

Data Screening and Extraction

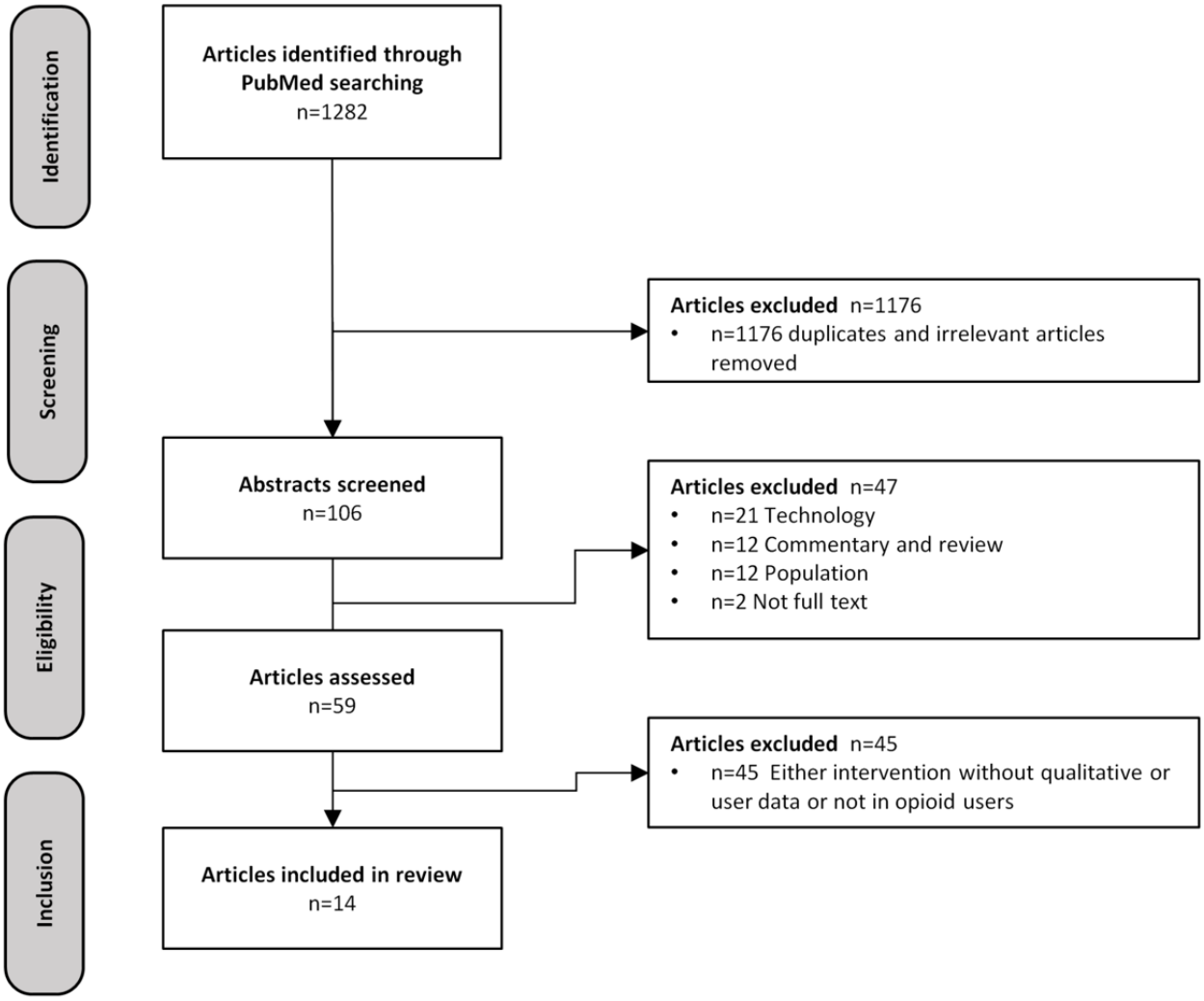
Data were screened according to titles, and abstracts, and followed by full texts of articles meeting inclusion criteria. The 2 reviewers, LNL and ME, screened the articles for inclusion and met when there was disagreement. Data on user perspectives, including benefits, barriers, and preferences, were extracted and summarized in tabular format. General study characteristics, including location, methods, participant age, and gender, as well as the type of technology, were also extracted.

Results

Overview

A total of 14 studies met the criteria and were included in the review [9,23-35]. Details of the search and stages of selection are outlined in Figure 1. Most of the studies were undertaken in the United States. One study was undertaken in Singapore and another in Canada. Most of the studies were relatively evenly distributed by sex, and most participants were middle-aged.

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



Usability, Feasibility, and Acceptability

Overall, the studies that explored the usability and acceptability of mHealth technology for opioid monitoring found relatively high usability and acceptability among the participants. A summary of the key issues and perspectives raised by the participants with OUD is described in Table 1. In Waselewski et al's [34] study, the mean useability score was 86.9 in the participants using opioids; in health care providers, the score was also high at 83.3. Another study found that 71% of the patients had a high level of satisfaction when using a mobile app with telemedicine built-in videos to monitor medication use [23]. In that study, the usability was also high; 93% of the patients did not report having had any technical issues, and 72% submitted their videos using the app [23]. Similarly, a wearable biosensor for overdose monitoring also had a high level of support from participants and a willingness to wear it [25]. Although the study by Hawk et al [24] had a high initial survey response rate (95%), which highlights interest among opioid users, completion of the surveys declined over the course of 1 month, from 97% at baseline to 42% at 4 weeks. Kanter et al [9] found that patients were willing to wear the device to monitor opioid overdose all the time (76%). Most patients were willing to share personal data on their opioid intake [24,31]. For example, in 1 study, 81% of the patients were willing to use the mHealth device to help them taper off opioids [31]. Comfort was also identified as an important factor in determining acceptability and wearability for participants in 2 studies [9,24].

Table 1. Study characteristics.

Study; country; type of study	Number of participants (M%[a] and F%[b]); Age	Technology	Procedure and measures	Results
Waselewski et al (2021) [34]; United States; 6-month pilot with qualitative interview	25 (52% and 48%); mean 33.7 (SD 8.1) years	Hope app for opioid user disorder	Feasibility, usability, and acceptability	• Patient Usability: mean score 86.9 (SD 10.2) • Providers: mean 83.3 (SD 12.8) • Patients liked self-monitoring, enhanced support, communication, and contact with providers
Godersky et al (2020) [23]; United States; N/A[c]	14 (86% and 14%); 18-65 years	mHealth[d] app for opioid and telemedicine (observation of buprenorphine intake) use in the office	Feasibility and acceptability	• 93% had no problems with using the app to share videos • Adherence= 72% • High satisfaction= 71 % of users • Pros: simplicity of use, structure, and personal accountability • Cons: self-portrayal when recording

Hawk et al (2021) [24]; United States; pilot feasibility	101 (43.7% and 57.3%); mean 38.4 (SD 10.25) years	mHealth platform reporting patient outcomes	Feasibility and acceptability	• Registration=95%
				• Declining rate of completion of surveys over time:
				  • Day 1=97%
				  • Day 3= 49%
				  • Month 1=42%
				• Willingness of patients to share data on medication, visits, substance use.
				• Barriers: Wi-Fi connection and access to technology
				• Email retention issues
				• mHealth log-in information retainment
				• patient factors: comfort
				• Privacy is a concern in a minority

Kanter et al (2021) [9]; United States; cross-sectional study with semistructured interviews	97 (57% and 43%); mean 41 (range 37-49) years	Wearable device for opioid detection, overdose, and treatment	Acceptability or likability and willingness to use (usability)	• Theme: privacy or discreetness is important • Theme: comfort when wearing it • Theme: a device that tracks and helps with overdose is needed • Willingness to use an mHealth device for opioid overdose=76% • Wear device continuously=75.5% • Vital sign monitoring=77% • Alerts others of overdose=63% • Watch type of bracelet design preferred=77% • Necklace=51% • Advised giving this to the hardest to reach and treat (eg, opioid intolerant or homeless)
Magee et al (2021) [31]; Australia; N/A	21 (48% and 52%); N/A	mHealth SMS text message supportive intervention	Qualitative acceptability and willingness to use	• Willingness of patients to engage with SMS text messaging to help curtail opioid use (from prescriptions)=81% • App-based support=71% • Desire for education and social support (socioeconomic support) • Barriers: internet; access to technology; visual problems; phone signal strength; poor self-assurance for knowledge around technology such as apps • Adherence could be improved through greater integrated medical or physician care, flexibility with dosing, and regular replies • Pain management advice • Need for support from a person as well • Need for encouragement and motivation
Marcu et al (2020) [32]; United States; N/A	19 (53% and 47%); 18-36 years	Smartphone app for opioid overdose	Acceptability or likability and user perspective	• Positive: mHealth can increase accessibility to care including medical supportive therapy for preventing overdose fatalities • Negative: privacy concerns, geolocation tracking of apps, and theft, and private data in apps

Study; country; type of study	Number of participants (M%[a] and F%[b]); Age	Technology	Procedure and measures	Results
Ahamad et al (2019) [25]; Canada; N/A	1061 (63.1% and 36.9%); median 44.2 years	Wearable opioid biosensor	Factors linked with willingness to wear the device	• Barriers: homelessness • High willingness to wear the opioid sensor • Predictors of openness to try the sensor: Higher in women (aOR[e]=1.41; 95% CI 1.09-1.84) • Those who were on methadone treatment regimes (aOR=1.86; 95% CI 1.45-2.40) and history of overdose (aOR 1.39; 95% CI 1.06-1.83)
Eaves et al (2022) [26]; United States; N/A	27 (0% and 100%); N/A	mHealth app for substance abuse (general opioid and polysubstance abuse, alcohol)	Qualitative 5 focus groups	• Desire to access social services: housing and counseling parenting help • Respectful safe environment • Supportive tone in SMS text messages, connect with social support (peers) and health professionals • Did not want the app to feel like a chore (eg, points earned)
Flickinger et al (2022) [27]; United States; N/A	28 (52% and 48%); mean 33.7 (SD 8.1) years	Hope app and community message board	User engagement and conversation in the app	• Opioid users exchanged supportive messages with one another and physicians through the app • They asked for medical advice=52% • Social support subjects=8% • App related interaction=45%
Glass et al (2021) [28]; United States; N/A	14 (71% and 29%); mean 40.1 (19-65) years	mHealth apps (opioid and cannabis users)	Semistructured interviews	• Apps should be a part of the primary care encounter and communication • Endorsement from medical professionals • Support and interaction with physicians, instructing them • SMS text messaging and professional health support through phone • Convenience • Minimize hassle • Trust essential
Hodges et al (2022) [29]; United States; pilot study	25 (52% and 48%); mean 34 years	Hope app	Self-efficacy and engagement	• Loss to follow-up a problem due to distance • mHealth app usage continued • Reported increases in self-efficacy to abstain
Tofighi et al (2022) [33]; United States; N/A	50 (48% and 52%); mean 44.1 (SD 11.8) years	SMS text message through texting software (ApToto)	Analysis of text messages	• Desire for personally tailored messages • Frequency according to risk profiles • Professional health support • Video-based content and advice • 6% opted out of messages • Reply to SMS text message over month=88% • Buprenorphine advice=2% messages • Cognitive behavioral therapy–based replies=13.8% • Appointments=6.1%
Zhang et al (2019) [35]; Singapore; N/A	30 (66.7% and 33.3%); mean 47.9 (SD 11.8) years	mHealth smartphone intervention for polysubstance (opioid, cannabis)	Thematic analysis and surveys	• App was liked and motivating=54% • Self-efficacy or confidence=33% • Interactive=77% • Easy=100%
Langdon et al (2021) [30]; United States; N/A	24 (50% and 50%); mean 38.9 years	Digital health interventions, computer, and SMS text message–based opioid users	Semistructured interviews on preferences	• Motivational element is essential • Distress tolerance • Assists with learning • Reduces stigma or being judged • Automated and tailored messages • Frequency=2-3 messages per day • Engaging use of media, videos (links), emojis, and GIFs

[a]M%: percentage of male individuals.

[b]F%: percentage of female individuals.

[c]N/A: not available.

[d]mHealth: mobile health.

[e]aOR: adjusted odds ratio.

Perceived Benefits

The synthesis of qualitative research reveals shared advantages among participants regarding mHealth technologies for opioid management. Accessibility stands out as a key benefit, with an emphasis on the pivotal role of behavior change. For example, 1 theme was that the opioid mHealth wearable tracker could provide enhanced access to care, which was perceived to be needed [31,32,34]. This included greater contact with health care services and medical providers [32,34]. These trackers, particularly wearable forms such as bracelets, wristbands, and necklaces, were favored for their noninvasive nature, as mentioned in Kanter et al [9]. Participants in 2 studies recognized a clear need for an opioid monitoring device that would help prevent overdose and fatalities [9,32], echoing the sentiment for an effective opioid monitoring tool. For example, the study by Kanter et al [9] found that 77% of patients reported that wearables were beneficial for monitoring vital signs, while 63% saw their value for alerting bystanders of an overdose. Self-monitoring and accountability were brought up as benefits in 2 studies [23,34]. Convenience and ease of use were also identified as benefits of using wearables for opioid use management [23,28,35]. A total of 2 studies found that a mHealth intervention increased the self-efficacy of opioid users and polysubstance users to overcome substance abuse [29,35], and motivation was brought up as a benefit as well [30,35]. These findings collectively underscore the value of mHealth apps in supporting individuals with OUD through various practical and psychological avenues.

Perceived Barriers

The research across various studies delineates several challenges in the adoption and continuous use of mHealth technologies for opioid management. Privacy stands out prominently as a concern. The apprehension over safeguarding personal health information was paramount, particularly with wearables that track opioid use. Protecting privacy and personal information was one of the main concerns participants had with tracking opioid use through wearable devices [9,32]. Privacy was mainly related to protecting personal and private health information. This included geolocation tracking as well as theft of the device if the patient was unconscious from an overdose [32] as well as concerns with the privacy of electronic health records and log-in information retention [24].

The visibility of the device and the need for it to be inconspicuous during use were significant factors in Kanter et al's [9] study, while Godersky et al [23] identified discomfort with self-viewing during video recordings as a deterrent. These issues underscore the need for privacy in both data handling and physical device usage.

Technological barriers also emerged, with several studies pointing to issues such as unreliable phone reception, limited access to necessary hardware, and inconsistent Wi-Fi connections, as indicated by Hawk et al [24]. These issues are compounded by the digital literacy required to effectively use such apps, with some users lacking the confidence to navigate mHealth technology, as might be inferred from some studies [24,31].

Visual difficulties in interacting with devices were also noted as a hindrance in 1 study [31]. Ahamad et al [25] report that homelessness can be a substantial obstacle to the effective use of wearable biosensors for overdose monitoring. Moreover, Kanter et al [9] suggest that interventions should aim to reach the most vulnerable populations who stand to benefit significantly from mHealth solutions.

The importance of human support in conjunction with technological aids was also mentioned, indicating that personal interaction may bolster adherence to using mHealth apps, a sentiment that could be associated with the insights from Magee et al [31]. These findings highlight the multifaceted nature of the barriers faced and the need for a holistic approach to mHealth implementation for OUD management.

Preferences

The research presents a range of preferences by participants when engaging with mHealth technologies for opioid management. Central to these preferences is the desire for behavior change facilitation through various techniques. The main behavior change techniques (BCTs) that were preferred by participants across the studies included encouragement or motivation [30,31], social and personal support [26,27,31,34], professional support [28,31], education [31,35], and personal tailoring [30,33]. Most notably, participants expressed a desire to have contact with a health care professional and to receive encouragement and educational, motivational, socioeconomic, and social support [26,31,33]. With regard to professional help, some participants wanted to have more medical professionals on board with the app rather than requiring users to navigate the app without professional support [26,28]. Integrated medical care from providers when using the wearables was preferred by participants in 5 studies [26,28,31-33]. For example, the study by Glass et al [28] found that participants preferred to have conversations with their medical providers about mHealth apps as part of their treatment plan. Another study found that 52% of users asked for medical advice from health practitioners while using the app [27].

In addition to this, peer social support was highlighted as being important in 2 studies [26,27], with app-based support being preferred by 71% of participants in another study [31]. Participants in 1 study wanted the mHealth app to integrate social assistance, housing services, counseling, and medical and peer support into the application and treatment regime [26].

Tone was also brought up as a theme in 1 study. Specifically, participants preferred to receive messages with a supportive and respectful tone; moreover, they wanted the technology to feel safe to use [26]. Integration of visual elements, including emojis, video links, and memes, was identified as being important in 1 SMS text messaging study [33]. Trust was an important factor in 2 studies [26,28]. Discreetness was essential, as was the need for mHealth to reduce the stigma associated with substance use [30,32]. In addition to automated messages, personally tailored messages were preferred by the participants

in 2 studies [30,33]. The need for motivation and encouragement was brought up as an important theme in 2 studies [30,31].

Additionally, 2 studies found that participants desired the app to be hassle-free or to not feel like a tiresome exercise to complete [26,28]. The personalization of messages and ease of use are underscored, with participants favoring tailored communication as seen in studies [30,33], and a seamless, nonburdensome interaction with the technology as suggested in 2 studies [26,28]. These preferences highlight the need for mHealth interventions to be user-centered, providing not only the technology but also the necessary human and educational resources to support individuals with OUD.

Discussion

Overview

This study aimed to better understand the perspectives of consumers of wearable technology and mHealth apps for detecting opioid intake and overdose in patients with OUD. Overall, it appears that patients are open to using mHealth technology and wearable devices to help manage opioid-related substance use and detect overdose. The common themes around the ability of technology to save lives and increase access to health care were a benefit that was noted in a few studies. Comfortable wearable sensors also appeared to be important as was technology that protected the user's privacy. Special attention must be paid to protecting the privacy of opioid users, as this was brought up as a concern and a potential disadvantage of the emerging technological interventions. Indeed, patient privacy and data protection are key hallmarks of the ethical principles surrounding modern health technology. Privacy must be protected to prevent unauthorized access to sensitive and personal information.

To manage the technological barriers associated with mHealth apps and wearable devices, users should be trained and reassured of their safety. This will increase their confidence in their ability to use this type of technology to manage their substance use condition. As users in 1 study wanted encouragement and motivation, these BCTs [36] should be incorporated in future interventions [31]. Other BCTs that may be considered include integrating social and professional health support and sending reminders and prompts [37]. Since access to health professionals was identified as being important [31], combining technological interventions with access to physicians and other health care professionals may be desirable. It may also be relevant to integrate theory-informed psychological interventions to increase the users' self-efficacy and motivation to change, such as implementing the Theory of Behavior Change, the Transtheoretical Model of Health Behavior Change, and implementation intentions [38-40]. It would also be relevant to study the stage of change (eg, preparedness vs denial) patients are in and to tailor the interventions accordingly [40]. It would also be interesting to evaluate opioid users' adherence to behavioral changes in a tailored intervention according to their stage of change, their motivation, and their self-efficacy levels.

The technology holds potential given the high usability and willingness of participants to try these wearables and apps. However, future studies should aim to address some adherence issues surrounding completion rates and reaching out to the most vulnerable groups. The decline in survey completion rates over time, as reported by Hawk et al [24], highlights a challenge in maintaining user engagement, and the issue of homelessness, as discussed by Ahamad et al [25], raises concerns about the reach of such interventions among the most vulnerable populations.

Strengths and Limitations

This review is important as it explored user perspectives on mHealth for opioid use management and overdose prevention, providing new insights into consumer acceptability in a newly developing field. We focused on emerging wearables and apps that were evaluated over the past 5 years. A limitation is that we did not explore older devices, but our aim was to focus on the latest technology. It should be noted that the findings of this review are limited by the few studies that have been undertaken on the population with OUD during this period. It is necessary to better understand this population's needs to develop technologically tailored interventions that may best maximize the benefits of assisting with the disorder and preventing fatalities while limiting barriers to using mHealth apps and wearables. More studies are needed to explore future consumer perspectives on mHealth apps for opioid use management. Another limitation is that we did not focus on commercially available apps if they were not evaluated in peer-reviewed medical literature for consumer perspectives. However, our focus was to better understand user experiences; hence, reviewing commercial applications that were not evaluated was not our goal. We also did not focus on efficacy, something that should be further explored alongside mixed methods qualitative research in order to make the best future recommendations.

Here is a list of recommendations for moving this technology forward:

1. Ensure robust privacy measures to address data security concerns, including geolocation tracking and device theft.
2. Provide thorough user training for confident and safe use of mHealth apps and wearables.
3. Integrate more BCTs to potentially increase engagement and positive opioid use management.
4. Provide health care professionals and peer support for enhanced mHealth intervention outcomes.
5. Deliver tailored messages considering participants' needs and preferences for increased intervention effectiveness.
6. Incorporate behavior change theories for enhanced motivation and self-efficacy.
7. Evaluate participants' needs and tailor interventions accordingly.
8. Design stigma-reducing mHealth apps and wearables to promote acceptance and help-seeking.
9. Investigate user perspectives to develop tailored mHealth interventions for effective opioid management.

Conclusion

In conclusion, while mHealth apps and wearables show potential for adoption among patients with OUD, there is a scarcity of studies on consumer perspectives. More research is required to

address the specific needs of this population. Common themes in the reviewed studies highlighted the benefits of using mHealth technology to prevent fatalities and improve accessibility to care. However, participants expressed concerns about data privacy and faced technological challenges. To advance mHealth technology for managing OUD, future development should focus on reducing technical barriers, prioritizing patient privacy, and enhancing access to health care professionals. By implementing our 10 recommendations, we can drive progress in health technology, ensuring user-centered design, increased engagement, and improved outcomes for individuals seeking support in managing their OUD.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA checklist.
[DOCX File , 30 KB - mhealth_v12i1e49751_app1.docx]

References

1. Ghose R, Forati AM, Mantsch JR. Impact of the COVID-19 pandemic on opioid overdose deaths: a spatiotemporal analysis. J Urban Health 2022;99(2):316-327 [FREE Full text] [doi: 10.1007/s11524-022-00610-0] [Medline: 35181834]
2. Gomes T, Kitchen SA, Murray R. Measuring the burden of opioid-related mortality in Ontario, Canada, during the COVID-19 pandemic. JAMA Netw Open 2021;4(5):e2112865 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.12865] [Medline: 34037734]
3. Sakhuja A, Sztajnkrycer M, Vallabhajosyula S, Cheungpasitporn W, Patch R, Jentzer J. National trends and outcomes of cardiac arrest in opioid overdose. Resuscitation 2017;121:84-89. [doi: 10.1016/j.resuscitation.2017.10.010] [Medline: 29037885]
4. Parthvi R, Agrawal A, Khanijo S, Tsegaye A, Talwar A. Acute opiate overdose: an update on management strategies in emergency department and critical care unit. Am J Ther 2019;26(3):e380-e387. [doi: 10.1097/MJT.0000000000000681] [Medline: 28952972]
5. Community management of opioid overdose. World Health Organization. 2014. URL: https://www.who.int/publications/i/item/9789241548816 [accessed 2024-02-25]
6. Wermeling DP. Review of naloxone safety for opioid overdose: practical considerations for new technology and expanded public access. Ther Adv Drug Saf 2015;6(1):20-31 [FREE Full text] [doi: 10.1177/2042098614564776] [Medline: 25642320]
7. Nuamah J, Mehta R, Sasangohar F. Technologies for opioid use disorder management: mobile app search and scoping review. JMIR Mhealth Uhealth 2020;8(6):e15752 [FREE Full text] [doi: 10.2196/15752] [Medline: 32501273]
8. Gullapalli BT, Carreiro S, Chapman BP, Ganesan D, Sjoquist J, Rahman T. OpiTrack: a wearable-based clinical opioid use tracker with temporal convolutional attention networks. Proc ACM Interact Mob Wearable Ubiquitous Technol 2021;5(3):102 [FREE Full text] [doi: 10.1145/3478107] [Medline: 35291374]
9. Kanter K, Gallagher R, Eweje F, Lee A, Gordon D, Landy S, et al. Willingness to use a wearable device capable of detecting and reversing overdose among people who use opioids in Philadelphia. Harm Reduct J 2021;18(1):75 [FREE Full text] [doi: 10.1186/s12954-021-00522-3] [Medline: 34301246]
10. Kulman E, Chapman B, Venkatasubramanian K, Carreiro S. Identifying opioid withdrawal using wearable biosensors. Proc Annu Hawaii Int Conf Syst Sci 2021;54:3583-3592 [FREE Full text] [Medline: 33568965]
11. Mahmud MS, Fang H, Wang H, Carreiro S, Boyer E. Automatic detection of opioid intake using wearable biosensor. Int Conf Comput Netw Commun 2018;2018:784-788 [FREE Full text] [doi: 10.1109/ICCNC.2018.8390334] [Medline: 31853456]
12. Roth AM, Tran NK, Cocchiaro B, Mitchell AK, Schwartz DG, Hensel DJ, et al. Wearable biosensors have the potential to monitor physiological changes associated with opioid overdose among people who use drugs: A proof-of-concept study in a real-world setting. Drug Alcohol Depend 2021;229(Pt A):109138 [FREE Full text] [doi: 10.1016/j.drugalcdep.2021.109138] [Medline: 34781097]
13. Salgado García FI, Indic P, Stapp J, Chintha KK, He Z, Brooks JH, et al. Using wearable technology to detect prescription opioid self-administration. Pain 2022;163(2):e357-e367 [FREE Full text] [doi: 10.1097/j.pain.0000000000002375] [Medline: 34270522]
14. Carreiro S, Wittbold K, Indic P, Fang H, Zhang J, Boyer EW. Wearable biosensors to detect physiologic change during opioid use. J Med Toxicol 2016;12(3):255-262 [FREE Full text] [doi: 10.1007/s13181-016-0557-5] [Medline: 27334894]
15. Nandakumar R, Gollakota S, Sunshine JE. Opioid overdose detection using smartphones. Sci Transl Med 2019;11(474):eaau8914 [FREE Full text] [doi: 10.1126/scitranslmed.aau8914] [Medline: 30626717]
16. Singh R, Lewis B, Chapman B, Carreiro S, Venkatasubramanian K. A machine learning-based approach for collaborative non-adherence detection during opioid abuse surveillance using a wearable biosensor. Biomed Eng Syst Technol Int Jt Conf BIOSTEC Revis Sel Pap 2019;5:310-318 [FREE Full text] [doi: 10.5220/0007382503100318] [Medline: 30993266]

17. Chintha KK, Indic P, Chapman B, Boyer EW, Carreiro S. Wearable biosensors to evaluate recurrent opioid toxicity after naloxone administration: a hilbert transform approach. Proc Annu Hawaii Int Conf Syst Sci 2018;2018:3247-3252 [FREE Full text] [Medline: 29375277]
18. Aglipay M, Wylie JL, Jolly AM. Health research among hard-to-reach people: six degrees of sampling. CMAJ 2015;187(15):1145-1149 [FREE Full text] [doi: 10.1503/cmaj.141076] [Medline: 26124229]
19. Matsuzaki M, Vu QM, Gwadz M, Delaney JAC, Kuo I, Trejo MEP, et al. Perceived access and barriers to care among illicit drug users and hazardous drinkers: findings from the Seek, Test, Treat, and Retain data harmonization initiative (STTR). BMC Public Health 2018;18(1):366 [FREE Full text] [doi: 10.1186/s12889-018-5291-2] [Medline: 29554894]
20. Bennett AS, Freeman R, Des Jarlais DC, Aronson ID. Reasons people who use opioids do not accept or carry no-cost naloxone: qualitative interview study. JMIR Form Res 2020;4(12):e22411 [FREE Full text] [doi: 10.2196/22411] [Medline: 33355094]
21. Arun P, Chavan BS, Kaur H. A study of reasons for not seeking treatment for substance abuse in community. Indian J Psychiatry 2004;46(3):256-260 [FREE Full text] [Medline: 21224907]
22. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev 2015 Jan 01;4(1):1-327 [FREE Full text] [doi: 10.1186/2046-4053-4-1] [Medline: 25554246]
23. Godersky ME, Klein JW, Merrill JO, Blalock KL, Saxon AJ, Samet JH, et al. Acceptability and feasibility of a mobile health application for video directly observed therapy of buprenorphine for opioid use disorders in an office-based setting. J Addict Med 2020;14(4):319-325 [FREE Full text] [doi: 10.1097/ADM.0000000000000608] [Medline: 31972762]
24. Hawk K, Malicki C, Kinsman J, D'Onofrio G, Taylor A, Venkatesh A. Feasibility and acceptability of electronic administration of patient reported outcomes using mHealth platform in emergency department patients with non-medical opioid use. Addict Sci Clin Pract 2021;16(1):66 [FREE Full text] [doi: 10.1186/s13722-021-00276-0] [Medline: 34758881]
25. Ahamad K, Dong H, Johnson C, Hyashi K, DeBeck K, Milloy MJ, et al. Factors associated with willingness to wear an electronic overdose detection device. Addict Sci Clin Pract 2019;14(1):23 [FREE Full text] [doi: 10.1186/s13722-019-0153-5] [Medline: 31269963]
26. Eaves ER, Doerry E, Lanzetta SA, Kruithoff KM, Negron K, Dykman K, et al. Applying user-centered design in the development of a supportive mHealth app for women in substance use recovery. Am J Health Promot 2023;37(1):56-64 [FREE Full text] [doi: 10.1177/08901171221113834] [Medline: 35815770]
27. Flickinger TE, Waselewski M, Tabackman A, Huynh J, Hodges J, Otero K, et al. Communication between patients, peers, and care providers through a mobile health intervention supporting medication-assisted treatment for opioid use disorder. Patient Educ Couns 2022;105(7):2110-2115 [FREE Full text] [doi: 10.1016/j.pec.2022.02.014] [Medline: 35260260]
28. Glass JE, Matson TE, Lim C, Hartzler AL, Kimbel K, Lee AK, et al. Approaches for implementing app-based digital treatments for drug use disorders into primary care: a qualitative, user-centered design study of patient perspectives. J Med Internet Res 2021;23(7):e25866 [FREE Full text] [doi: 10.2196/25866] [Medline: 34255666]
29. Hodges J, Waselewski M, Harrington W, Franklin T, Schorling K, Huynh J, et al. Six-month outcomes of the HOPE smartphone application designed to support treatment with medications for opioid use disorder and piloted during an early statewide COVID-19 lockdown. Addict Sci Clin Pract 2022;17(1):16 [FREE Full text] [doi: 10.1186/s13722-022-00296-4] [Medline: 35255965]
30. Langdon KJ, Scherzer C, Ramsey S, Carey K, Rich J, Ranney ML. Feasibility and acceptability of a digital health intervention to promote engagement in and adherence to medication for opioid use disorder. J Subst Abuse Treat 2021;131:108538 [FREE Full text] [doi: 10.1016/j.jsat.2021.108538] [Medline: 34154869]
31. Magee MR, McNeilage AG, Avery N, Glare P, Ashton-James CE. mHealth interventions to support prescription opioid tapering in patients with chronic pain: qualitative study of patients' perspectives. JMIR Form Res 2021;5(5):e25969 [FREE Full text] [doi: 10.2196/25969] [Medline: 34003133]
32. Marcu G, Aizen R, Roth AM, Lankenau S, Schwartz DG. Acceptability of smartphone applications for facilitating layperson naloxone administration during opioid overdoses. JAMIA Open 2020;3(1):44-52 [FREE Full text] [doi: 10.1093/jamiaopen/ooz068] [Medline: 32607487]
33. Tofighi B, Durr M, Marini C, Lewis CF, Lee JD. A mixed-methods evaluation of the feasibility of a medical management-based text messaging intervention combined with buprenorphine in primary care. Subst Abuse 2022;16:11782218221078253 [FREE Full text] [doi: 10.1177/11782218221078253] [Medline: 35356483]
34. Waselewski ME, Flickinger TE, Canan C, Harrington W, Franklin T, Otero KN, et al. A mobile health app to support patients receiving medication-assisted treatment for opioid use disorder: development and feasibility study. JMIR Form Res 2021;5(2):e24561 [FREE Full text] [doi: 10.2196/24561] [Medline: 33620324]
35. Zhang M, Ying J, Amron SB, Mahreen Z, Song G, Fung DSS, et al. A smartphone attention bias app for individuals with addictive disorders: feasibility and acceptability study. JMIR Mhealth Uhealth 2019;7(9):e15465 [FREE Full text] [doi: 10.2196/15465] [Medline: 31586359]
36. Michie S, Wood CE, Johnston M, Abraham C, Francis JJ, Hardeman W. Behaviour change techniques: the development and evaluation of a taxonomic method for reporting and describing behaviour change interventions (a suite of five studies

involving consensus methods, randomised controlled trials and analysis of qualitative data). Health Technol Assess 2015;19(99):1-188 [FREE Full text] [doi: 10.3310/hta19990] [Medline: 26616119]
37. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013;46(1):81-95 [FREE Full text] [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
38. Tapera R, Mbongwe B, Mhaka-Mutepfa M, Lord A, Phaladze NA, Zetola NM. The theory of planned behavior as a behavior change model for tobacco control strategies among adolescents in Botswana. PLoS One 2020;15(6):e0233462 [FREE Full text] [doi: 10.1371/journal.pone.0233462] [Medline: 32502211]
39. Orbell S, Hodgkins S, Sheeran P. Implementation intentions and the theory of planned behavior. Pers Soc Psychol Bull 1997;23(9):945-954. [doi: 10.1177/0146167297239004] [Medline: 29506445]
40. Prochaska JO, Velicer WF. The transtheoretical model of health behavior change. Am J Health Promot 1997;12(1):38-48. [doi: 10.4278/0890-1171-12.1.38] [Medline: 10170434]

Abbreviations

BCT: behavior change technique
mHealth: mobile health
OUD: opioid use disorder
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews

Edited by L Buis; submitted 07.06.23; peer-reviewed by S Ramsey, M Cozad; comments to author 16.08.23; revised version received 14.11.23; accepted 29.11.23; published 11.04.24.

Please cite as:
Lyzwinski LN, Elgendi M, Menon C
Users' Acceptability and Perceived Efficacy of mHealth for Opioid Use Disorder: Scoping Review
JMIR Mhealth Uhealth 2024;12:e49751
URL: https://mhealth.jmir.org/2024/1/e49751
doi:10.2196/49751
PMID:38602751

Review

Effectiveness of mHealth App–Based Interventions for Increasing Physical Activity and Improving Physical Fitness in Children and Adolescents: Systematic Review and Meta-Analysis

Jun-Wei Wang[1,2], MSc; Zhicheng Zhu[3], MSc; Zhang Shuling[1], PhD; Jia Fan[1], MSc; Yu Jin[1], MSc; Zhan-Le Gao[1], MSc; Wan-Di Chen[4], MSc; Xue Li[1], PhD

[1]School of Sport Medicine and Health, Chengdu Sport University, Chengdu, China

[2]School of Sports Science, Beijing Sport University, Beijing, China

[3]Physical education institute, Xinyu University, Xinyu, China

[4]Academic Administration, Chengdu Sport University, Chengdu, China

Corresponding Author:
Xue Li, PhD
School of Sport Medicine and Health
Chengdu Sport University
No.2, Tiyuan Road, Wuhou District
Chengdu, 610041
China
Phone: 86 13550146822
Email: lixue2078@126.com

Abstract

Background: The COVID-19 pandemic has significantly reduced physical activity (PA) levels and increased sedentary behavior (SB), which can lead to worsening physical fitness (PF). Children and adolescents may benefit from mobile health (mHealth) apps to increase PA and improve PF. However, the effectiveness of mHealth app–based interventions and potential moderators in this population are not yet fully understood.

Objective: This study aims to review and analyze the effectiveness of mHealth app–based interventions in promoting PA and improving PF and identify potential moderators of the efficacy of mHealth app–based interventions in children and adolescents.

Methods: We searched for randomized controlled trials (RCTs) published in the PubMed, Web of Science, EBSCO, and Cochrane Library databases until December 25, 2023, to conduct this meta-analysis. We included articles with intervention groups that investigated the effects of mHealth-based apps on PA and PF among children and adolescents. Due to high heterogeneity, a meta-analysis was conducted using a random effects model. The Cochrane Risk of Bias Assessment Tool was used to evaluate the risk of bias. Subgroup analysis and meta-regression analyses were performed to identify potential influences impacting effect sizes.

Results: We included 28 RCTs with a total of 5643 participants. In general, the risk of bias of included studies was low. Our findings showed that mHealth app–based interventions significantly increased total PA (TPA; standardized mean difference [SMD] 0.29, 95% CI 0.13-0.45; *P*<.001), reduced SB (SMD –0.97, 95% CI –1.67 to –0.28; *P*=.006) and BMI (weighted mean difference –0.31 kg/m^2, 95% CI –0.60 to –0.01 kg/m^2; *P*=.12), and improved muscle strength (SMD 1.97, 95% CI 0.09-3.86; *P*=.04) and agility (SMD –0.35, 95% CI –0.61 to –0.10; *P*=.006). However, mHealth app–based interventions insignificantly affected moderate to vigorous PA (MVPA; SMD 0.11, 95% CI –0.04 to 0.25; *P*<.001), waist circumference (weighted mean difference 0.38 cm, 95% CI –1.28 to 2.04 cm; *P*=.65), muscular power (SMD 0.01, 95% CI –0.08 to 0.10; *P*=.81), cardiorespiratory fitness (SMD –0.20, 95% CI –0.45 to 0.05; *P*=.11), muscular endurance (SMD 0.47, 95% CI –0.08 to 1.02; *P*=.10), and flexibility (SMD 0.09, 95% CI –0.23 to 0.41; *P*=.58). Subgroup analyses and meta-regression showed that intervention duration was associated with TPA and MVPA, and age and types of intervention was associated with BMI.

Conclusions: Our meta-analysis suggests that mHealth app–based interventions may yield small-to-large beneficial effects on TPA, SB, BMI, agility, and muscle strength in children and adolescents. Furthermore, age and intervention duration may correlate with the higher effectiveness of mHealth app–based interventions. However, due to the limited number and quality of included studies, the aforementioned conclusions require validation through additional high-quality research.

Trial Registration: PROSPERO CRD42023426532; https://tinyurl.com/25jm4kmf

(JMIR Mhealth Uhealth 2024;12:e51478) doi:10.2196/51478

KEYWORDS

mobile health; mHealth apps; children and adolescents; physical activity; physical fitness; systematic review; meta-analysis; mobile phone

Introduction

Background

The COVID-19 global pandemic has had adverse effects on the physical fitness (PF) and mental health of children and adolescents [1]. Before the COVID-19 pandemic outbreak, only approximately 30% of children and adolescents worldwide could meet the recommended levels of physical activity (PA) [2,3]. However, the COVID-19 pandemic has exacerbated this issue by decreasing their levels of PA, increasing sedentary behavior (SB), and leading to a decline in their PF [4]. A recent study in the United Kingdom discovered that children exhibited lower performance scores on the seated forward bend and 20-meter shuttle run test and higher BMI values in 2020 than in 2019 [5]. PF is a crucial determinant of children's and adolescents' health status [6], which can be influenced by various factors, including genetic, environmental, and PA-related factors [7]. Epidemiological studies have established a "dose response" relationship between PA and PF, which showed that increased PA levels and reduced SB were positively associated with improved PF in adolescents [8]. Therefore, the way to increase PA and improve PF is still one of the most important social problems to be solved for children and adolescents.

The rapid advancement of intelligent technology has increased the use of smartphones among young generation and the wide use of mobile health (mHealth) technologies [9,10]. At present, mHealth technologies, including wearable devices, smartphones, tablets, mHealth apps, smartwatches, and pedometers, are commonly used in the field of health care [11]. Recently, 2 systematic reviews have investigated the impact of mHealth-based interventions on behavioral changes, including PA and SB, in children and adolescents [12,13]. However, these reviews primarily concentrated on specific mHealth technology interventions, including SMS text messaging, wearable devices, web-based interventions, and others. Moreover, these systematic reviews exclusively focus on one or more behavioral changes, encompassing physical inactivity and SB, and the overall quality of the evidence is deemed low [13]. Current research indicates that app-based interventions on smartphones may represent the most effective strategy [13]. mHealth app–based interventions are among the most commonly used methods within the realm of mHealth technologies. Among these technologies, mHealth apps have been extensively used in the fitness and medical fields due to their affordability, personalization, and diverse features [14]. mHealth apps can provide quantitative visual feedback regarding the health behaviors of users, such as their PA; meanwhile, users upload their personal information to app databases, and apps facilitate personalized, long-distance, and low-contact training to improve the healthy development of users [15]. In recent years, mHealth app–based interventions have shown significant promise in promoting healthy behaviors, including increased PA and reduced SB, among children and adolescents. Nevertheless, there is a lack of systematic reviews comprehensively summarizing the impacts of stand-alone mHealth apps or concerted interventions using apps as one of the multiple components (eg, behavioral counseling combined with app interventions) on various health behaviors, including PF. In addition, studies in this domain have been predominantly centered on adults, with a noticeable dearth of pertinent research within populations such as children and adolescents [16].

Studies have demonstrated that mHealth app–based interventions can lead to effective outcomes in improving the PA behavior of users [17]. However, another study [18] found that such intervention has indicated only small effects on PA and is likely related to potential influencing factors. Furthermore, the efficacy of mHealth app–based interventions on PF is inconsistent. One study linked lower BMI and higher motor competence to the frequency and type of mHealth app use [19], while another study indicated that mHealth app–based interventions were ineffective in improving PF among adolescents, which is possibly due to the characteristics of the intervention [20]. The use of theory-based mHealth app interventions may also be more advantageous in increasing PA and enhancing PF in children and adolescents [21]. Several theoretical paradigms, including self-determination theory (SDT), the transtheoretical model, the health belief model, the theory of planned behavior, and social cognitive theory (SCT), have been used in mHealth app–based interventions [22]. The number and type of behavior change technique (BCT) clusters may also play a significant role in the effectiveness of mHealth app interventions. Michie et al [23] provided a standardized taxonomy of BCT that categorizes them into 16 clusters, such as feedback and monitoring, reward and threat, goals and planning, shaping knowledge, social support, and comparison of outcomes. This taxonomy aids in identifying which BCT clusters are more effectively applied to apps, thereby enhancing PA promotion and PF improvement. In conclusion, various factors, including the type of mHealth app, intervention characteristics, theoretical paradigms, and BCT clusters, are important considerations in the effectiveness of mHealth app interventions. Despite the increasing number of articles summarizing interventions based on mHealth apps, a noteworthy research gap persists. Most of these articles concentrate on interventions using commercially available mHealth apps that lack evidence-based behavior change strategies. Nevertheless, a significant proportion of users and patients rely on commercially available app-based mHealth interventions that lack empirical evaluation and rarely incorporate evidence-based behavior change strategies. Furthermore, current research predominantly highlights the intervention effects of mHealth apps on health-related outcome

measures. However, there is a notable deficiency in evidence-based mHealth apps intervention programs and studies that integrate various target behaviors.

Objectives

The first objective of this systematic review aims to evaluate the effectiveness of mHealth app–based interventions in promoting PA and improving PF among children and adolescents. The second objective is to specifically assess moderating effects (eg, age, types of apps, theoretical paradigm, BCT clusters, and intervention duration) on the effectiveness of mHealth app–based interventions in subgroups within these studies. Unlike previous studies, our review contributes evidence-based, high-quality content for potential mHealth app interventions addressing PA and PF. This contribution results from a meticulous evaluation and meta-analysis of relevant randomized controlled trials (RCTs). In addition, we conducted an extensive analysis of key moderating variables using subgroups and meta-regression, encompassing theoretical paradigms, BCT clusters, intervention duration, and more. This comprehensive approach enhances our understanding of the factors influencing intervention effects and facilitates the precise quantification of the intervention program. Our endeavors significantly expand the research scope beyond previous reviews.

Methods

Registration and Approval

The systematic reviews were registered on the PROSPERO (CRD42023426532). The literature search, reporting guidance, and implementation process of the study followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [24].

Search Strategy

Several databases were searched, including PubMed, Web of Science, EBSCO, and Cochrane Library, to identify relevant RCTs published until December 25, 2023. The search strategy involved a Boolean search using a combination of subject-related words and free words. The following search terms were used: (Child OR Preschool OR Adolescent), ("Mobile health application*" OR "mHealth app*" OR "Portable Software Application*" OR "Mobile Application*" OR App*), and ("Physical Activity" OR PA OR MVPA OR "sedentary behavior" OR SB OR "Physical Fitness"). A detailed search strategy for Web of Science is presented in Multimedia Appendix 1. We opted to update the searches in the same databases used for the initial search to refine the results for this study. In addition, we examined the references cited in previous reviews to identify additional relevant literature. Concurrently, we reached out to authors of potentially eligible studies to obtain complete data. If >2 attempts were made to contact authors without receiving a response, the study was excluded. The literature search was not restricted by language.

Inclusion and Exclusion Criteria

Inclusion Criteria

The following criteria were included for inclusion in the literature review: (1) The study comprised children and adolescents, aged 3 to 18 years, with the majority falling within this range. Participants did not exhibit physical dysfunction, and overweight or obesity, among other factors, were not exclusionary criteria. Children and adolescents were categorized into 3 groups: preschoolers aged 3-6 years; children aged 7-12 years; and adolescents aged 13-18 years. (2) Interventions using smartphone-based and tablet-based mHealth apps may involve either stand-alone apps (ie, solely apps) or concerted intervention (ie, apps combined with another intervention). The control group comprised genuine controls, such as no interventions, waitlist conditions, and usual clinical care. In addition, active controls, including interventions via apps, were considered. Studies using placebo and sham apps also met the inclusion criteria. (3) The study design was an RCT. (4) Primary outcome measures included PA, SB, performance-related PF (eg, coordination and balance), health-related PF (eg, cardiorespiratory endurance, muscle strength, and body composition), and physiological function (eg, body shape and metabolism).

Exclusion Criteria

The exclusion criteria were as follows: (1) articles written in languages other than English and Chinese; (2) repeated published studies, basic studies, observational studies, reviews, and case series articles; (3) studies for which full text was unavailable or data were incomplete; and (4) mHealth apps that only used SMS text messaging interventions or were incompatible with smartphones or tablets.

Study Selection

After the literature search, the initial search results were imported into EndNote 20 (Thomson ResearchSoft) to remove duplicate articles. Predefined inclusion and exclusion criteria were applied to the literature. Two authors (ZZ and ZS) initially screened the titles and abstracts. Articles meeting the criteria were downloaded, and 1 author (ZZ) thoroughly evaluated the full text based on the inclusion and exclusion criteria, while the other author (ZS) conducted a randomized assessment. Another author (JF) was involved in resolving discrepancies between 2 independent reviewers to determine if a study met the inclusion criteria.

Data Extraction

The included literature was independently extracted by 2 researchers (YJ and ZLG). The extracted information included basic details (eg, authors, publication year, country, and study type), characteristics of the study population (eg, age, gender, and sample size), characteristics of mHealth apps (eg, name, type, theoretical paradigm, and number or type of BCT clusters), intervention characteristics, outcome measures, and indicators related to risk of bias assessment. Two authors (YJ and ZLG) assessed the use of BCT in apps using the taxonomy by Michie [23]. Relevant information was primarily extracted from the study descriptions; in cases of incomplete data, the original apps were consulted. Disputes were resolved through third-party consultation (WDC).

Risk of Bias Assessment

Two independent (YJ and ZLG) investigators assessed the risk of bias using the Cochrane Working Group's tool [25]. Any disagreements were resolved by a third independent researcher (WDC) through consultation. Each study underwent evaluation in 7 domains, and the risk of bias was categorized as unclear, low, or high.

Statistical Analysis

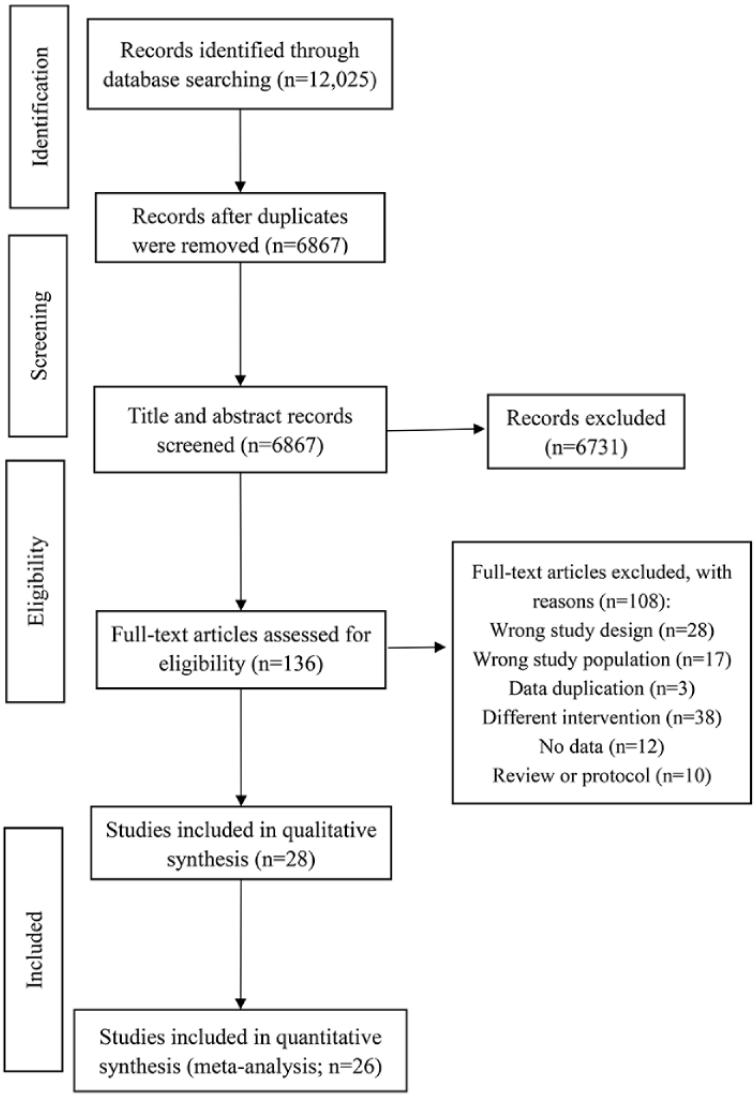
If the number of included papers was <3, then a systematic review was conducted. When a sufficient number of included studies were available for a meta-analysis, we used Revman 5.4 (Nordic Cochrane Center) and Stata 16.0 (StataCorp) to estimate effect sizes, conduct subgroup analysis, and sensitivity analysis. Weighted mean difference (WMD) and 95% CI were used as effect measures when the same measurement method was used. When measurement methods were consistent, the standardized mean difference (SMD) and 95% CI were used. The magnitude of SMD was interpreted as follows: SMD<0.2, negligible; 0.2≤SMD<0.5, small; 0.5≤SMD<0.8, medium; and SMD≥0.8, large [26].

The study indirectly mentioned the mean and SD, and the MD and SMD were calculated as the postintervention mean and SD based on the Cochrane Handbook [27]. The study examined the magnitude of heterogeneity using I^2 and P values. If I^2≤50% and P≥.10, a fixed-effect model was used for data analysis. On the contrary, if I^2>50% and P<.10, a random effect model was used for meta-analysis [27]. The sources of heterogeneity were identified by subgroup analyses and meta-regression analysis based on type of apps, theoretical paradigm, age, and BCT clusters. The robustness of each study was evaluated using sensitivity analysis, and publication bias was assessed with funnel plots and the Egger test. Results

Study Selection

A total of 12,025 relevant articles were retrieved from PubMed (n=69), Web of Science (n=11,033), EBSCO (n=217), and Cochrane Library (n=706). Duplicate references were removed, which resulted in 6867 articles. Following an initial screening by abstracts, 136 articles were identified, which were then assessed by reading the full text. Finally, 28 articles were chosen for inclusion (Figure 1).

Figure 1. Flowchart of the study selection.



Characteristics of the Included Studies

This study included 28 publications, all of which were RCTs published between 2014 and 2023 [20,26,28-49]. The sample size included 5643 subjects, ranging from 12 to 1392 participants per study. As shown in Table 1, the included studies have the following basic characteristics. The age of the participants varied from 3 to 18 years with 5 (18%) of the 28 studies involving preschool children [26,44,46,48,50], 8 (29%) studies involving children [28-31,33,36,37,51], and 15 (53%) studies involving adolescents [20,32,34,35,38-43,45,47,49,52,53]. The objectives of the studies differed, and the selection criteria for the target population varied inconsistently. Most studies (16/28, 57%) focused on healthy children and adolescents; however, of the 28 studies, 9 (32%) included participants with overweight and those with obesity [29-31,35,36,39,43,45,52], 2 (7%) involved children with cancer [34,53], and 1 (4%) enrolled youth with congenital heart disease [40]. Of the 28 studies, 8 (29%) were conducted in Asia [31,36,37,39-41,44,45], while the remaining studies were performed in Oceania (7/28, 25%) [29,30,35,38,46,50,53], North America (5/28, 18%) [28,33,34,43,51], and Europe (8/28, 29%) [20,26,32,42,47-49,52].

This review encompassed 28 studies that used 14 different mHealth apps. These included 9 commercial apps [20,31-33,35,38,41,47,49] and 14 research apps [28-30,39,40,42-46,48,51-53], and 5 mHealth apps did not provide the corresponding information [26,34,36,37,50]. The type of intervention used in 15 studies was stand-alone apps [20,26,28,32,33,40,41,43-47,49-51], and 13 studies used concerted intervention [29-31,34-39,42,48,52,53]. Participant engagement with the intervention was mentioned in 28 studies, 8 studies focused on parent-centered [26,31,36,37,44,46,48,50], the remaining studies were child centered. The mHealth apps were based on various theoretical paradigms, including self-regulation theory (SRT), SDT, and SCT. Different numbers or types of BCT clusters were identified, and they ranged from 1 to 7 clusters. Examples of BCT clusters used included goal setting and planning, feedback and monitoring, and behavioral comparison. Interventions duration lasted 2 to 48 weeks (Table 1). The primary and secondary outcome measures included total PA (TPA), moderate to vigorous PA (MVPA), SB, cardiorespiratory fitness (CRF), BMI, waist circumference (WC), muscle strength, agility, flexibility, muscular power, and endurance.

Table 1. Summary of the intervention characteristics of the included studies.

Study	Participants or population	Age (years)	Sample size（treatment/control）	Interventions	Comparator	Type of mobile health apps	Theoretical paradigm	BCT[a] clusters	Duration（weeks）	Outcomes
Direito et al [20]	Insufficiently active healthy young people	15.7 (+1.2 or –1.2)	32/17	Group 1: immersive app *Zombies Run;* group 2: nonimmersive app *Get Running*	No interventions	Commercial apps	SRT[b]	Feedback and monitoring	8	CRF[c], TPA[d], SB[e], and MVPA[f]
Garde et al [28]	Healthy students	11.3 (+1.2 or –1.2)	26/16	MKMM game app	Waitlist (crossed over after 3 weeks)	Research apps	SDT[g] and theory of motivation	Reward and threat feedback and monitoring	4	TPA
Lubans et al [29]	Adolescent boys “at risk” of obesity	12.7 (+0.5 or –0.5)	181/180	Face-to-face PA[h] sessions+pedometers for PA self-monitoring+purpose-built web-based smartphone apps+other	Regular curriculum	Research apps	SCT[i] and SDT	Goals and planning, shaping knowledge, social support, and feedback and monitoring	32	SB, MVPA, BMI, WC[j], muscle strength, and muscular endurance
Smith et al [30]	Adolescent boys “at risk” of obesity	12.7 (+0.5 or –0.5)	181/180	Face-to-face PA sessions+pedometers for PA self-monitoring+purpose-built web-based smartphone apps+other	Regular curriculum	Research apps	SCT and SDT	Goals and planning, shaping knowledge, social support, feedback and monitoring	20	TPA, SB, MVPA, BMI, WC, and muscle strength
Fernandez-Luque et al [31]	Children with overweight and those with obesity	9-12	108/119	Wearable sensors+mobile and social media (WhatsApp and Instagram)	No intervention	Commercial apps	NR[k]	Goals and planning, social support, feedback and monitoring	12	BMI
Pyky et al [32]	Young adolescent men	17.8 (+0.6 or –0.6)	250/246	*MOPOrtal* app	No intervention	Commercial apps	TTM[l]	Goals and planning, shaping knowledge, feedback and monitoring, and comparison of outcomes	24	SB and BMI
Gaudet et al [33]	Young adolescents	13 (+0.35 or –0.35)	23/23	An individualized goal was set by *Fitbit app*	No intervention	Commercial apps	SRT and self-monitoring theory	Goals and planning, feedback and monitoring, and regulation	7	MVPA
Mendoza et al [34]	Adolescent and young adult childhood cancer survivors	16.6 (+1.5 or –1.5)	29/20	Fitbit Flex wearable wristband and mobile health app+peer-based web-based support group	Usual care	NR	SDT	Goals and planning, feedback and monitoring	0	SB and MVPA
Nyström et al [26]	Healthy Swedish children	4.5 (+0.1 or –0.1)	156/159	MINISTOP app	Information or advice about a healthy diet+PA via a 4-page pamphlet	Research apps	SCT	Shaping knowledge, feedback, and monitoring	24	SB and MVPA

Study	Participants or population	Age (years)	Sample size (treatment/control)	Interventions	Comparator	Type of mobile health apps	Theoretical paradigm	BCT[a] clusters	Duration (weeks)	Outcomes
Chen et al [35]	Adolescents who are overweight or obese	14.9 (+1.67 or –1.67)	23/17	Fitbit Flex app+iStart Smart for Teens web-based educational program+biweekly SMS text messages	Omron HJ-105 pedometer+a blank food and activity diary	Commercial apps	SCT	Goals and planning, feedback and monitoring, shaping knowledge, social support, regulation, natural consequences, and covert learning	12.24	TPA, SB, and BMI
Browne et al 2020 [52]	Children with obesity	9-16	8/12	Usual clinical care+Mandolean training (*myBigO app*)	Usual clinical care	Commercial apps	NR	Goals and planning, feedback and monitoring, comparison of outcomes, shaping knowledge, social support, repetition and substitution, and antecedents	4	BMI
Garde et al [51]	Elementary school students	10.6 (+0.51 or –0.51)	19/18	*MKMM* game app	No intervention	Research apps	SDT and theory of motivation	Goals and planning, feedback and monitoring, comparison of outcomes, and social support	2	TPA
Trost and Brookes [50]	Children	3-6	17/17	*Moovosity* app	No intervention	NR	NR	NR	8	TPA
Devine et al [53]	Adolescent and young adult survivors of childhood cancer	13-25	25/24	In-person group sessions+mobile app+fitness tracker use alone	Waitlist	Research apps	SCT	Goals and planning, feedback and monitoring, shaping knowledge, and social support	12	CRF, SB, MVPA, BMI, WC, muscle strength, and coordination
Kahana et al [36]	Children with overweight and obesity	Median 10	32/47	Structured PA sessions, nutritional and behavioral counseling+"Just Dance Now" and "Motion Sports" app	Structured PA sessions, nutritional, and behavioral counseling	NRNR	NR	—[m]	20	BMI, muscle strength, muscular power, muscular endurance, and agility
Liu et al [37,54]	Primary school children	9.6 (+0.4 or –0.4)	705/687	Health education reinforcement of PA and BMI monitoring and feedback (*Eat Wisely and Move Happily* app)	Health education lessons and physical education sessions	NR	NR	Feedback and monitoring	36	CRF, TPA, MVPA, BMI, WC, muscular power, and muscular endurance
Likhitsawong et al [39]	Children and adolescents with obesity	10-15	35/35	OBEST app+standard care	Standard care	Research apps	Theory of motivation	Goals and planning, feedback and monitoring, and shaping knowledge	24	BMI and WC

Study	Participants or population	Age (years)	Sample size （treatment/control）	Interventions	Comparator	Type of mobile health apps	Theoretical paradigm	BCT[a] clusters	Duration （weeks）	Outcomes
Lin et al [40]	Youth with congenital heart disease	15-24	100/50	Group 1: *COOL Passport* app; group 2: *COOL Passport app* +health promotion cloud	Standard care	Research apps	SRT	Goals and planning, feedback and monitoring, shaping knowledge, and reward and threat	24.48	TPA
Seah and Koh [41]	Adolescent girls	14.9 (+0.3 or –0.3)	13/23	*MapMyFitness* app	No intervention	Commercial apps	SDT	Goals and planning, feedback and monitoring, and social support	2.3	TPA and MVPA
Stasinaki et al [42]	Adolescents with obesity	10-18	18/13	Nutritional education and PA+PathMate2 app	Nutritional education and PA	Research apps	Theory of motivation	Goals and planning, feedback and monitoring, reward and threat, and comparison of behavior	22	CRF, WC, muscular power, muscular endurance, agility, flexibility, and balance
Tugault-Lafleur et al [43]	Children with overweight or obesity	10-17	107/107	*Aim2Be* app	No intervention	Research apps	SCT and SRT	Goals and planning, feedback and monitoring, identity, and social support	12	TPA and SB
Han et al [44]	Preschool children	3-6	66/44	YOUXUE UP app	No intervention	Research apps	Socioecological model	Goals and planning, social support, and reward and threat	8	SB, MVPA, muscle strength, muscular power, agility, flexibility, coordination, and balance
Oh et al [45]	Adolescents with obesity	13.2 (+3.6 or –3.6)	12/12	*SUKIA* app	Nintendo Switch	Research apps	NR	Feedback and monitoring and shaping knowledge	3	CRF and BMI
Staiano et al [46]	Preschoolers	4.0 (+0.8 or –0.8)	32/37	Motor skills app	Free Play app	Research apps	SCT	Goals and planning, social support, shaping knowledge, and feedback and monitoring	12	TPA, SB and MVPA
Mateo-Orcajada et al [47]	Adolescents	13.96 (+1.21 or –1.21)	240/160	*Poksammon Go* app or *Pacer* app or *Strava* app or *MapMyWalk* app	No intervention	Commercial apps	NR	8-10 change techniques per application	10	CRF, TPA, BMI, WC, muscle strength, muscular power, muscular endurance, and flexibility

Study	Participants or population	Age (years)	Sample size（treatment/control）	Interventions	Comparator	Type of mobile health apps	Theoretical paradigm	BCT[a] clusters	Duration（weeks）	Outcomes
Ridgers et al [38]	Inactive adolescents	13.7 (+0.4 or –0.4)	144/131	Wrist-worn Fitbit Flex and accompanying *Fitbit* app and digital behavior change resources	No intervention	Commercial apps	CT and behavioral choice theory	Goals and planning, feedback and monitoring, and self-belief	12	TPA and MVPA
Alexandrou et al [48]	Preschool-aged children	2.5-3	270/271	Standard care+ *MINISTOP 2.0* app	Standard care	Research apps	SCT	Identity, goals and planning, shaping knowledge, and feedback and monitoring	24	SB and MVPA
Mateo-Orcajada et al [49]	Adolescents	13.66 (+1.17 or –1.17)	92/46	Group 1: *Pokémon Go Playing* app continuously; group 2: *Pokémon Go Playing* app intermittently	No intervention	Commercial apps	NR	NR	10	TPA, BM, and WC

[a]BCT: behavior change technique.

[b]SRT: self-regulation theory.

[c]CRF: cardiorespiratory fitness.

[d]TPA: total physical activity.

[e]SB: sedentary behavior.

[f]MVPA: moderate to vigorous physical activity.

[g]SDT: self-determination theory.

[h]PA: physical activity.

[i]SCT: social cognitive theory.

[j]WC: waist circumference.

[k]NR: not reported.

[l]TTM: transtheoretical model.

[m]Not available.

Risk of Bias Assessment

Overall, of the 28 studies, 11 (39%) were classified as having a low risk of bias [20,26,28,29,34,35,37,39,51-53], and a high risk of bias was identified in 17 (61%) studies [30-33,36,38,40-50]. The methods for random sequence generation were adequately reported in 28 studies, and 14 (50%) of the 28 studies described allocation concealment protocols [20,26,29,31,35,37,39,43,46,48,50-53]. Of the 28 studies, blinding of participants and personnel was unclear in 17 studies [20,28-33,38-41,43-45,47,49,51], high risk of bias was identified in 7 (25%) studies [26,34,35,37,50,52,53], and blinding of outcome assessment was unclear in 18 (64%) studies [18,28-34,38,40,41,44,45,47-51,53]. Four studies provided data regarding subjects lost to follow-up [30-32,34], while one study had unclear information on this aspect [51]. None of the 28 studies were found to have selective outcome reporting, and other aspects of bias were evaluated mainly in terms of baseline data and conflicts of interest (Figure 2 [20,26,28-53]).

Figure 2. Risk bias assessment of the included studies. (A) Risk of bias graph and (B) risk of bias summary.

Results of the Meta-Analysis

Effects of mHealth App–Based Interventions on PA

Of the 28 studies, 21 (75%) were examined to assess the impact of mHealth app–based interventions on TPA. The heterogeneity test indicated a substantial level of heterogeneity among the studies (I^2=75%; P<.001), which led to the adoption of a random effects model for the analysis. The meta-analysis results indicated that the intervention group exhibited higher TPA (SMD 0.29, 95% CI 0.13-0.45; P<.001; Figure 3 [20,28,30,35,37,38,40,41,43,46,47,49-51]), but the effect size was small.

Of the 28 studies, 14 (50%) reported the impact of mHealth app–based interventions on SB levels. Heterogeneity tests revealed homogeneity between the studies (I^2=98%; P<.001), which required analysis using a random effect model. Meta-analysis found that mHealth app–based interventions significantly reduced SB (SMD –0.97, 95% CI –1.67 to –0.28; P=.006; Figure 4 [20,26,29,30,32,34,35,43,44,46,48,53]).

Of the 28 studies, 14 (50%) investigated the impact of mHealth app–based interventions on MVPA levels. The heterogeneity test indicated homogeneity among the studies (I^2=67%; P<.001), which allowed for analysis using a random effect model. There were no significant differences between the control and intervention groups (SMD 0.11, 95% CI –0.04 to 0.25; P<.001; Figure 5 [20,26,29,30,33,34,37,38,41,46,48,53]).

Figure 3. Forest plot of the effect of mobile health app–based interventions on increasing total physical activity.

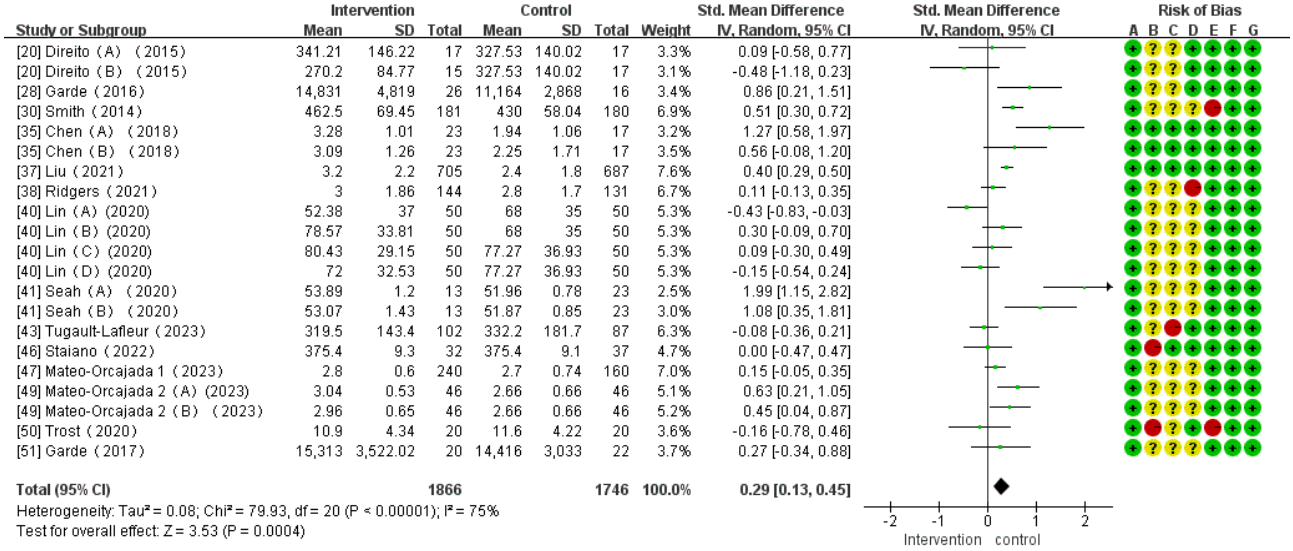


Figure 4. Forest plot of the effect of mobile health app–based interventions on decreasing sedentary behavior.

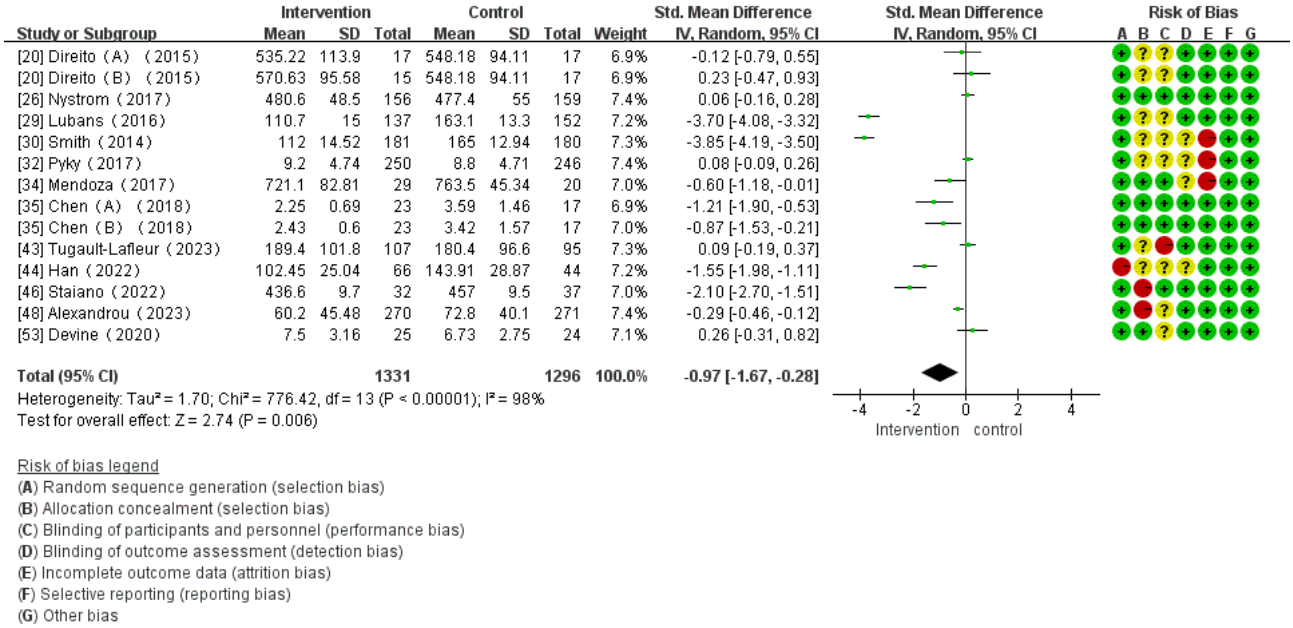
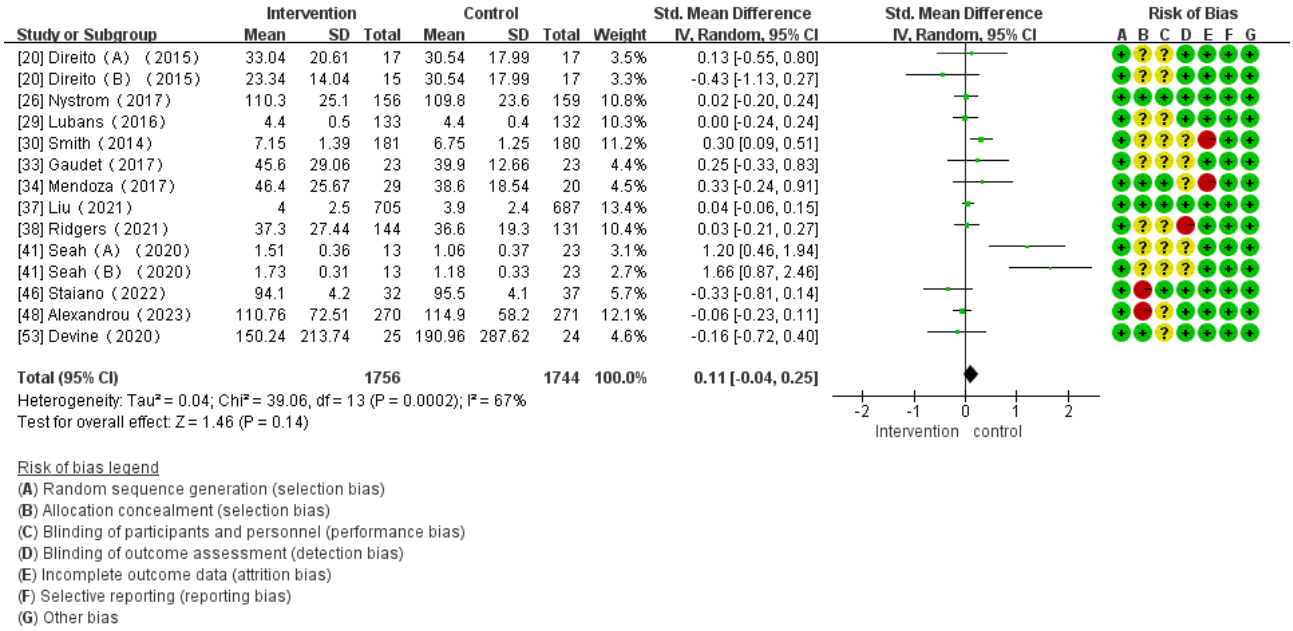


Figure 5. Forest plot of the effect of mobile health app–based interventions on increasing moderate to vigorous physical activity.



Effects of mHealth App–Based Interventions on BMI and WC

Of the 28 studies, 13 (46%) investigated the effects of mHealth app–based interventions on BMI. Of these 13 studies, 2 (15%) did not directly conduct changes in outcome indicators before and after interventions [31,52] and were only included in systematic reviews. The heterogeneity test demonstrated homogeneity among the studies (I^2=32%; P=.12), which enabled analysis using a fixed-effect model. Meta-analysis found that mHealth app–based interventions significantly reduced BMI (WMD –0.31, 95% CI –0.60 to –0.01; P=.12; Figure 6 [29,30,35-37,39,45,47,49,53]). The 2 other studies on BMI reported a significant reduction in BMI among obese children with mHealth app–based interventions [31,52], which is consistent with the meta-analysis results.

Of the 28 studies, 9 (32%) reported the effect of mHealth app–based interventions on WC. The heterogeneity tests showed homogeneity among the studies (I^2=54%; P=.02), which allowed for analysis using a random effect model. There were no significant differences in WC between the intervention and control groups (WMD 0.38 kg/m^2, 95% CI –1.28 to 2.04 kg/m^2; P=.65; Figure 7 [29,37,39,42,47,49,53]).

Figure 6. Forest plot of the effect of mobile health app–based interventions on BMI.

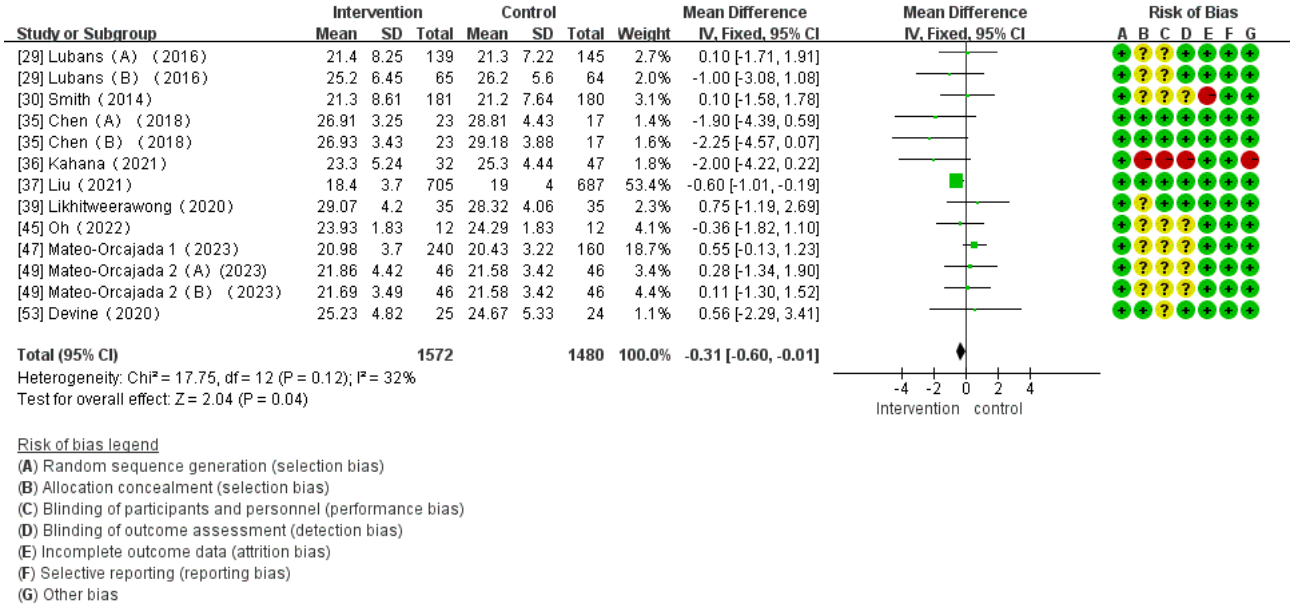
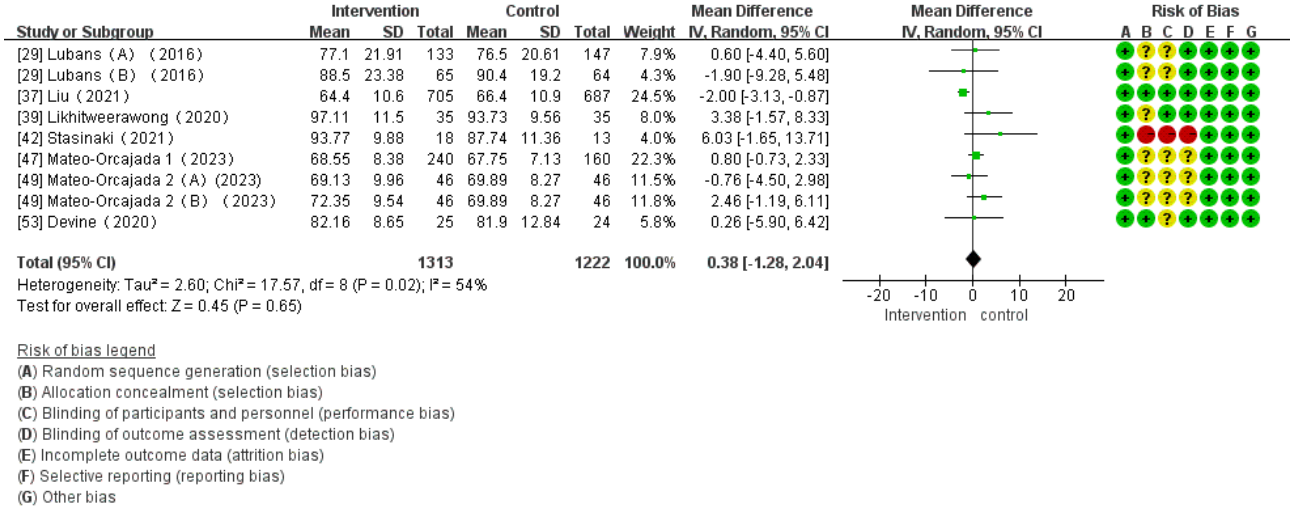


Figure 7. Forest plot of the effect of mobile health app–based interventions on waist circumference.



Effects of mHealth App–Based Interventions on PF

Of the 28 studies, 7 (25%) studies investigated the impact of mHealth app–based interventions on CRF. Heterogeneity tests showed homogeneity among the studies (I^2=66%; P=.007) and were conducted using a random effect model. No significant differences were found between intervention and control groups in CRF (SMD –0.20 cm, 95% CI –0.45 to 0.05 cm; P=.11; Figure 8 [20,37,42,45,47,53]).

Of the 28 studies, 6 (21%) reported the impact of mHealth app–based interventions on muscle strength. Heterogeneity tests revealed homogeneity in the studies (I^2=99%; P<.001) and were conducted using a random effect model. Meta-analysis found that mHealth app–based interventions significantly increased muscle strength (SMD 1.97, 95% CI 0.09-3.86; P=.04; Figure 9 [29,30,36,44,47,53]).

Of the 28 studies, 5 (18%) investigated the impact of mHealth app–based interventions on muscular power. The heterogeneity test indicated homogeneity among the studies (I^2=45%; P=.12), which allowed for analysis using a fixed-effect model. There were no significant differences between the control and intervention groups (SMD 0.01, 95% CI –0.08 to 0.10; P=.81; Figure 10 [36,37,42,44,47]).

Of the 28 studies, 5 (18%) examined the impact of mHealth app–based interventions on muscular endurance. The heterogeneity test indicated a substantial level of heterogeneity among the studies (I^2=98%; P<.001), which led to the adoption of a random effects model for the analysis. The meta-analysis results indicated no significant difference in muscle endurance between the intervention and control groups (SMD 0.47, 95% CI –0.08 to 1.02; P=.10; Figure 11 [29,36,37,42,47]).

Of the 28 studies, 3 (11%) investigated the impact of mHealth app–based interventions on agility. The heterogeneity test indicated homogeneity among the studies (I^2=0%; P=.87), which allowed for analysis using a fixed-effect model. Meta-analysis found that mHealth app–based interventions significantly improved agility (SMD –0.35, 95% CI –0.61 to –0.10; P=.006; Figure 12 [36,42,44]).

Of the 28 studies, 3 (11%) reported the impact of mHealth app–based interventions on flexibility. Heterogeneity tests revealed homogeneity in the studies (I^2=52%; P=.12) and were conducted using a random effect model. There were no significant differences in flexibility between the control and intervention groups (SMD 0.09, 95% CI –0.23 to –0.41; P=.58; Figure 13 [42,44,47]).

Figure 8. Forest plot of the effect of mobile health app–based interventions on cardiorespiratory fitness.

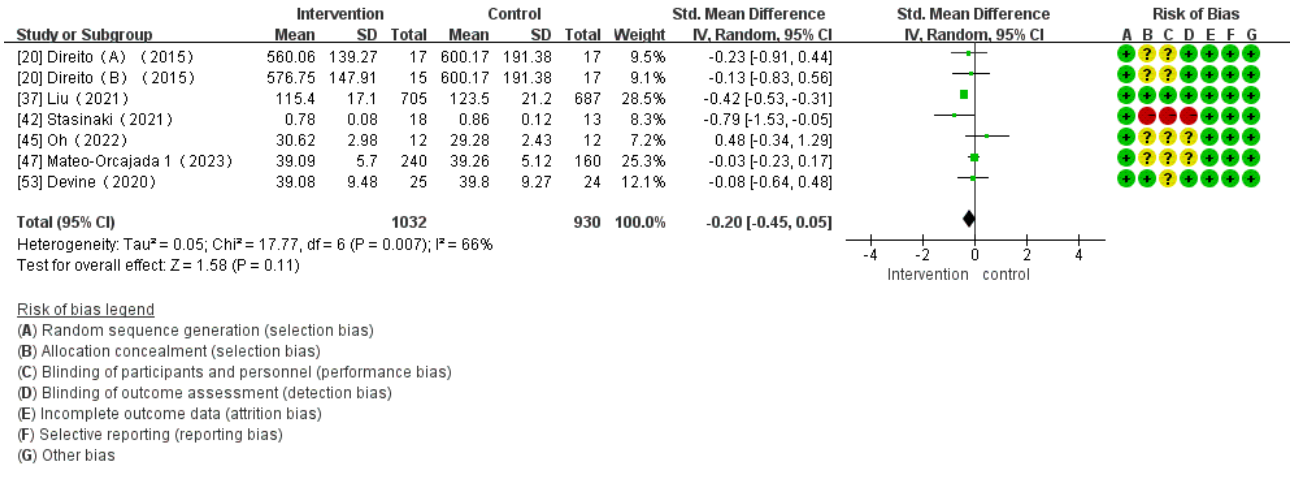


Figure 9. Forest plot of the effect of mobile health app–based interventions on muscular strength.

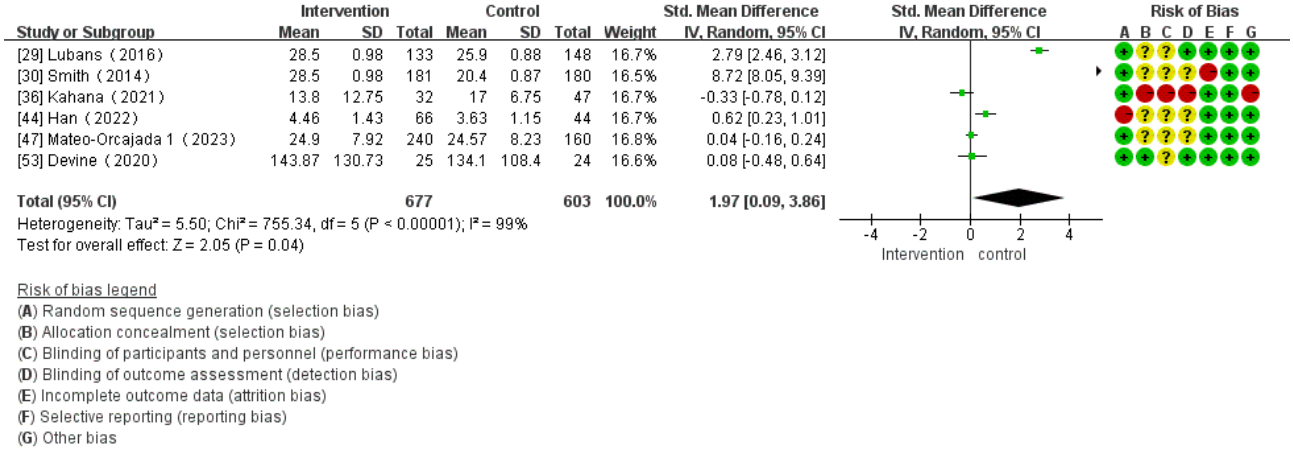


Figure 10. Forest plot of the effect of mobile health app–based interventions on muscular power.

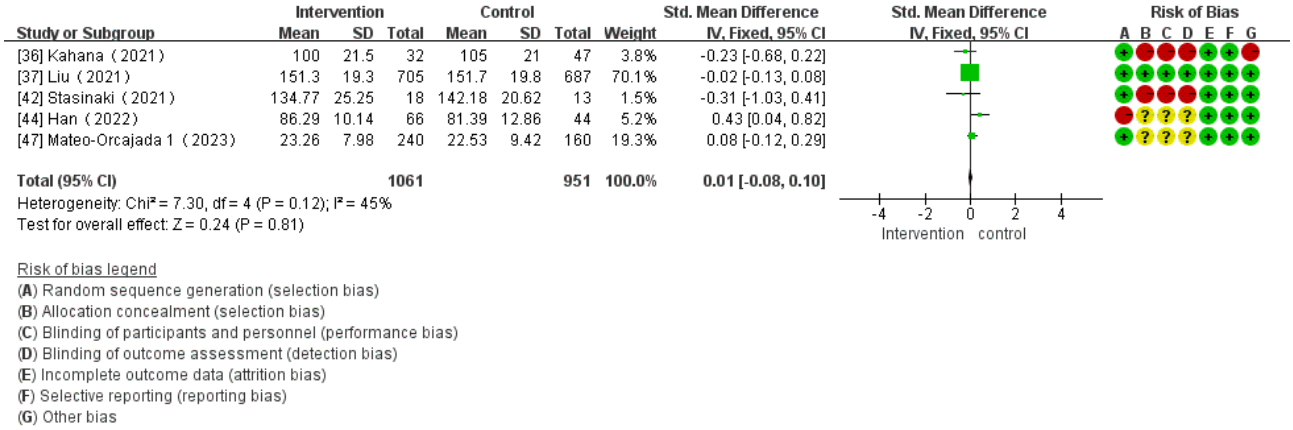


Figure 11. Forest plot of the effect of mobile health app–based interventions on muscular endurance.

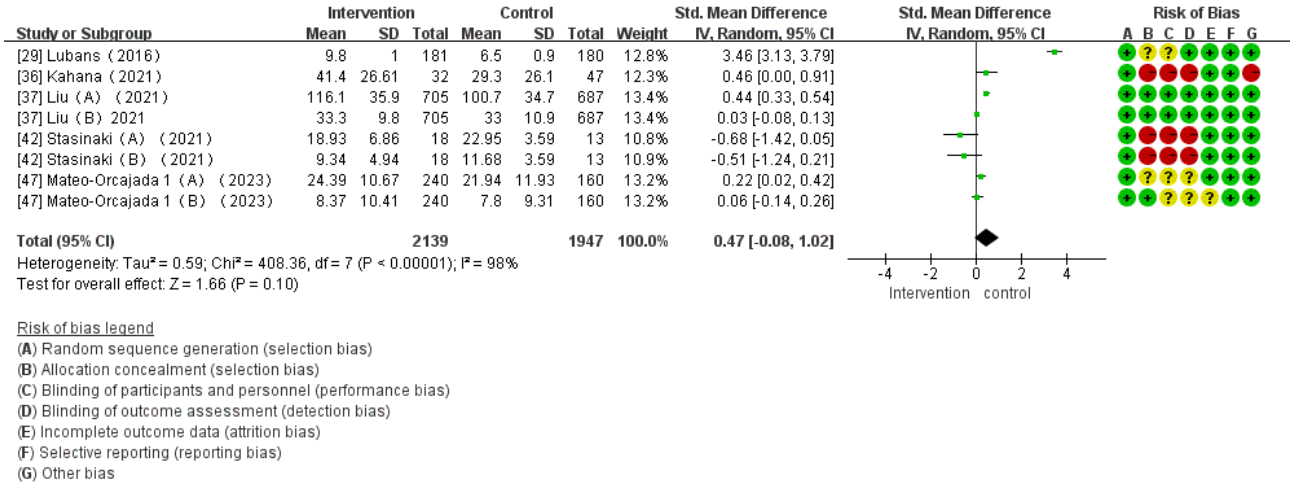


Figure 12. Forest plot of the effect of mobile health app–based interventions on agility.

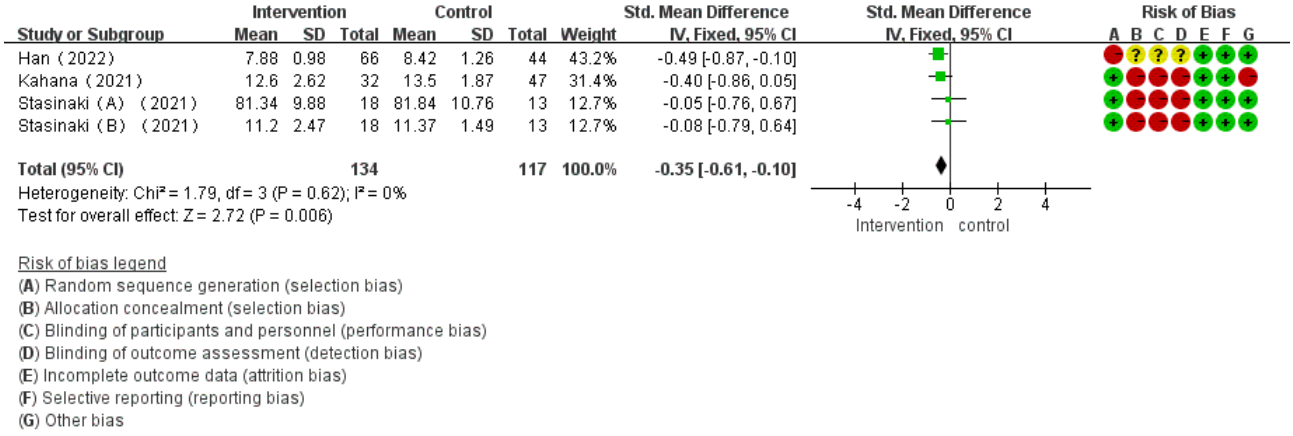
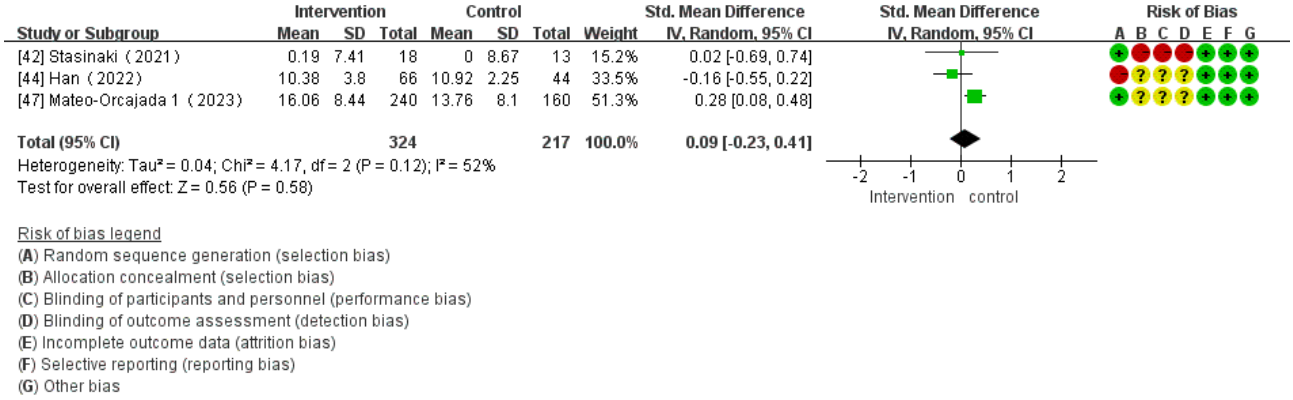


Figure 13. Forest plot of the effect of mobile health app–based interventions on flexibility.



Sensitivity Analysis

In this study, Stata 16.0 was used to conduct sensitivity analyses on TPA, SB, MVPA, and BMI for evaluating the robustness and reliability of the results. The sensitivity analysis results demonstrated that excluding any single study did not impact the effect size of the mHealth app–based intervention in outcomes such as TPA, SB, and MVPA (Multimedia Appendices 2-4), which indicates the robustness and reliability of the study results. In terms of BMI, sensitivity analyses identified 1 study as an outlier [37] (Multimedia Appendix 5). Removing this study altered the overall effect size, indicating that the study results were not sufficiently robust (SMD 0.03 cm, 95% CI –0.41 to –0.46 cm; *P*=.90) and should be interpreted with caution.

Subgroup Analyses

Overview

Subgroup analyses were conducted to investigate the potential sources of heterogeneity and moderating effects, which considered factors such as age, types of apps, theoretical paradigm, number or type of BCT clusters, and intervention duration. The primary outcome indicators included TPA, BMI, SB, and MVPA. Sensitivity analysis, subgroup analysis, and assessment of publication bias were not conducted for the remaining outcome indicators due to the limited number of included studies.

Subgroup Analyses on the Effect of mHealth App–Based Interventions on TPA

The results of subgroup analyses investigating the impact of mHealth app–based interventions on TPA are presented in Multimedia Appendix 6. Subgroup analyses indicated no significant difference in types of intervention, theoretical paradigm, and the number of BCT clusters concerning the improvement in TPA in children and adolescents.

Age-based subgroup analysis revealed significant positive effects of interventions for children in the 7-12 years group (SMD 0.42) and adolescents in the 13-18 years group (SMD 0.29) but negligible effect for the preschool children in the 3-6 years group (SMD –0.06).

Subgroup analysis based on types of apps revealed a significantly greater effect size for the commercial apps (SMD 0.51) compared to the research apps intervention (SMD 0.13). Substantial heterogeneity existed between these 2 groups (I^2=75%; *P*<.001), but the 95% CI for the effect size between the 2 groups overlapped.

Subgroup analysis based on intervention duration revealed a significant increase in TPA for the 2- to 4-week intervention (SMD 1.01) and 8- to 12-week intervention (SMD 0.23). Heterogeneity was observed between the 2 groups (I^2=75%; *P*<.001), but the 95% CI for the effect size between the 2 groups overlapped.

Subgroup Analyses on the Effect of mHealth App–Based Interventions on SB

The results of subgroup analyses on SB are presented in Multimedia Appendix 7. Subgroup analyses showed no significant difference in intervention duration regarding the reduction of SB in children and adolescents.

Age-based subgroup analysis unveiled a notably larger effect size for interventions directed at children aged 7-12 years (SMD –3.78) compared to the group of preschool children aged 3-6 years (SMD –0.92). Heterogeneity was evident among the 2 groups (I^2=99%; P<.001), and the 95% CI for the effect size between the 2 groups did not overlap.

Subgroup analysis based on the types of apps demonstrated a significantly stronger effect size for research app interventions (SMD –1.38) than for commercial app interventions (SMD –0.35). Heterogeneity was observed between the 2 groups (I^2=98%; P<.001), but the 95% CI for the effect size between the 2 groups overlapped.

Subgroup analysis based on types of intervention revealed a significantly greater effect size for the concerted intervention (SMD –1.47) compared to the stand-alone apps intervention (SMD –0.45). Substantial heterogeneity existed between these 2 groups (I^2=98%; P<.001), but the 95% CI for the effect size between the 2 groups overlapped.

Subgroup analyses based on the theoretical paradigm showed a greater effect size for interventions based on the SCT (SMD –0.64) than for interventions based on combination of SCT and other theories (SMD –2.18) and interventions solely based on SRT (SMD 0.05). Substantial heterogeneity existed between these 2 groups (I^2=98%; P<.001), but the 95% CI for the effect size between the 2 groups overlapped.

Subgroup analysis based on the number of BCT clusters demonstrated a significant reduction in SB for interventions based on 7-10 BCT clusters (SMD –1.03) and interventions based on 4 BCT clusters (SMD –1.36), and the 95% CI for the effect size between the 2 groups overlapped.

Subgroup analysis based on intervention duration revealed a significant reduction in SB for the 20- to 48-week intervention (SMD –1.42) but had a negligible effect at the 8- to 12-week intervention (SMD –0.63).

Subgroup Analyses on the Effect of mHealth App–Based Interventions on MVPA

The results of subgroup analyses on MVPA are shown in Multimedia Appendix 8. Subgroup analyses revealed no significant difference in type of apps, types of intervention, and the number of BCT clusters in improving MVPA in children and adolescents, which were not moderators of the effect of mHealth app interventions.

Age-based subgroup analyses unveiled a notably larger effect size for interventions directed at adolescents aged 13-18 years (SMD 0.42) compared to the group of preschool children aged 3-6 years (SMD –0.05) and children aged 7-12 years (SMD 0.11). Heterogeneity was evident among the 2 groups (I^2=68%; P<.001), but the 95% CI for the effect size between the 2 groups overlapped.

Subgroup analyses based on theoretical paradigm demonstrated a significant increase in MVPA for SDT-based intervention (SMD 1.03), while no significant effect was observed for SCT (SMD –0.06), combination of SCT and other theories (SMD 0.12), and SRT (SMD –0.14).

Subgroup analysis based on intervention duration revealed a significant increase in MVPA for the 2- to 4-week intervention (SMD 1.42) but negligible effect for the 8- to 12-week intervention (SMD –0.01) and 20- to 48-week intervention (SMD 0.05).

Subgroup Analyses on the Effect of mHealth App–Based Interventions on BMI

The results of the subgroup analyses based on BMI are shown in Multimedia Appendix 9. Subgroup analyses revealed no significant difference in type of apps, theoretical paradigm, and the number of BCT clusters in decreasing BMI in children and adolescents.

Age-based subgroup analysis revealed a significant decrease in BMI for interventions targeting children aged 7-12 years (WMD –0.59), but no significant difference was observed for adolescents aged 13-18 years (WMD 0.03).

Subgroup analysis based on types of intervention revealed a significantly greater effect size for the concerted intervention (WMD –0.59) compared to the stand-alone apps intervention (WMD 0.34), but the 95% CI for the effect size between the 2 groups overlapped.

Subgroup analysis based on intervention duration revealed a significant reduction in BMI for the 20- to 48-week intervention (WMD –0.57) and the 8- to 12-week intervention (WMD 0.33), and the former showed a clearly superior effect.

Meta-Regression

We conducted meta-regressions for TPA, BMI, SB, and MVPA, focusing on statistically significant moderators identified in the subgroup analyses. The results of the meta-regression for TPA (P=.03, 95% CI –1.127 to –0.082) and MVPA (P=.045, 95% CI –2.052 to –0.033) revealed that intervention duration had a potential moderating effect on high heterogeneity; no statistical differences were found for other variables acting as moderators. In the case of BMI, meta-regression results indicated that age (P=.04, 95% CI 0.041-1.473) and types of intervention (P=.02, 95% CI –1.648 to –0.205) could be potential moderators for high heterogeneity, whereas other factors did not significantly contribute to explaining high heterogeneity.

Reporting Biases

Funnel plots were used to assess publication bias in the effects of mHealth app–based interventions on TPA, SB, BMI, and MVPA in children and adolescents. The funnel plots exhibited mostly symmetrical patterns in the 4 studies (Multimedia Appendices 10-13). In addition, Egger test was performed for TPA (t_{20}=0.01; P=.99), SB (t_{13}=–0.135; P=.20), MVPA (t_{13}=1.22; P=.25), and BMI (t_{12}=–0.07; P=.95), and the results

suggested no significant publication bias (Multimedia Appendices 14-17).

Discussion

Principal Findings

We conducted a systematic review and meta-analysis to assess the effectiveness of mHealth app–based interventions in promoting PA and enhancing PF in children and adolescents. This study also examined the potential moderators that may influence the efficacy of these interventions. The findings of this study suggest that mHealth app–based interventions may yield positive effects on TPA, SB, BMI, agility, and muscle strength in children and adolescents. However, no significant effects were observed for MVPA, WC, CRF, muscular power and endurance, and flexibility. Age, theoretical paradigm, BCT clusters, types of intervention, types of apps, and intervention duration were identified as significant moderating factors associated with the increased effectiveness of mHealth app interventions on PA and PF, but the impact on effect size is not entirely consistent.

Overall Effect

PA Levels

The findings of this study indicated that mHealth app–based interventions increased TPA and reduced SB among children and adolescents but had no significant effect on MVPA. Our research findings represent a valuable expansion of recently published systematic reviews; however, they do not entirely align with the results of previous studies. A prior systematic review reported that mHealth-based interventions increased TPA levels and addressed physical inactivity in children and adolescents but did not lead to reduced SB and improved MVPA [12,13]. The inconsistency in the findings may be attributed to the cointervention effect of technologies such as SMS text messaging, wearable devices, web-based interventions, and smartphone apps [13]. Among these, smartphone app–based interventions might be the most effective strategy. The use of apps may contribute to increased SB time. Nevertheless, results of this study indicate that mHealth app–based interventions can effectively reduce SB in children and adolescents. Distinguishing whether the effect is from stand-alone app interventions or other strategies within concerted interventions is challenging. Subgroup analyses in this study demonstrated the superiority of concerted interventions over stand-alone app interventions. In conclusion, mHealth app–based interventions serve as a valuable adjunct in reducing SB in children and adolescents. Further concerted interventions, such as combining educational policies with mHealth apps interventions, are recommended.

The effectiveness of mHealth app–based interventions was influenced by age and intervention format. Notably, these interventions were more effective in improving PA levels among adolescents compared to children, and greater effects were observed when using mHealth apps than when using only SMS text messaging interventions [35,50]. These findings support the conclusions of this study, but some researchers hold differing views. The study by Trost and Brooks [50] indicated that an 8-week intervention with Moovosity apps (Kinetica Group Pty Ltd) improved proficiency in fundamental movement skills (FMS) but did not increase PA levels. The phenomenon may be associated with the types of design and goal setting of the apps. The apps in the study were primarily used to increase the FMS of children, which may have resulted in the activity of FMS replacing the original PA. As a result, the TPA of children was unchanged. Another study [35] identified that targeted app interventions would be effective in reducing SB in adolescents, which must be based on certain theoretical paradigms and BCT clusters. Gamification-based app interventions are also more favorable for increasing TPA and reducing SB levels in children and adolescents and must be combined with the theoretical paradigm, intervention duration, and features of apps [55]. In conclusion, the reasons for inconsistent intervention results may be related to population characteristics, types of apps, theoretical paradigm, BCT clusters, and intervention duration. Moreover, further research is needed in the future.

PF Levels

Another significant finding of the meta-analysis was that mHealth app–based interventions decreased BMI and improved muscle strength and agility among children and adolescents. However, no significant effects were observed for WC, CRF, muscular power and endurance, and flexibility. These findings are not entirely aligned with the results of previous studies. For example, a previous study reported an increase in PA levels among adolescents and significant reductions in sugary drink consumption and BMI following a 3-month intervention using Fitbit Flex apps (Fitbit, Inc) [35]. Meanwhile, a different study [29] indicated that an 18-month app intervention did not lead to significant reductions in BMI and WC, but it improved exercise capacity and reduced screen time. Several potential factors may contribute to the inconsistent results. First, the choice of outcome measure could be a contributing factor. BMI and WC may indicate distinct aspects of obesity; although BMI primarily assesses body size and shape, WC is a measure of abdominal obesity, leading to a weak correlation between the two [56,57]. Moreover, BMI tends to underestimate overweight prevalence when abdominal obesity is considered [57]. Hence, although mHealth app–based interventions show minor effects on BMI in children and adolescents, they might not induce significant changes in WC. Nonetheless, such interventions can still yield favorable outcomes. Another possible reason for the inconsistency in interventions could be related to the age of the population under study. Some studies have shown that mHealth app–based interventions are effective in reducing body weight and BMI in individuals aged ≥45 years, but they are less effective in children and adolescents [58]. In addition, our meta-regression analysis revealed that age significantly influences the effectiveness of app-based interventions. This result could be attributed to the fact that children and adolescents are in a period of rapid growth and development, and the effects of the interventions may be masked by the significant changes in height and weight during this stage.

In terms of muscular strength, power, and endurance, previous studies have shown that a 6-month intervention using mHealth apps and trackers resulted in increased muscle strength among children and adolescents but did not have a significant effect

on CRF [53]; Stasiak et al [42] discovered that a daily conversational agent intervention, using a mobile app for adolescents with obesity, enhanced subjects' muscular fitness, with no discernible difference between the intervention and control groups. Conversely, other studies have indicated that interventions using AIMFIT apps can positively impact maximal oxygen uptake [20]; Mateo-Orcajada et al [47] found that a 10-week after-school intervention for adolescents using mobile step-tracking apps significantly enhanced subjects' upper limb strength, hamstring and lower back flexibility, explosive power of the lower limbs, as well as abdominal muscular strength and endurance. Various factors can influence the effectiveness of interventions, including subject characteristics, intervention dosage, engagement levels, and willingness to use apps. Some of the studies included in this work involved participants who were patients with cancer [53] or children with overweight and obesity [42], and this factor may have influenced the outcomes of the intervention. Furthermore, low user willingness to engage with apps may reduce PA engagement, which consequently impacts the effectiveness of the intervention. Finally, insufficient dosage of mHealth app–based interventions has been associated with a small effect size on TPA in children and adolescents and has shown no increase in MVPA levels. This insufficiency in dosage may explain the lack of intervention effects observed in children.

Moderating Variables on the Effects of mHealth App–Based Interventions for PA and PF

Age

The study results revealed that age moderates the effects of mHealth app interventions on TPA, MVPA, SB, and BMI. Subgroup analysis indicated that mHealth app–based interventions significantly increased TPA and MVPA and reduced SB in children and adolescents aged 7-18 years; increased TPA and MVPA in adolescents aged 7-18 years; and only reduced SB in children aged 3-6 years. It is noteworthy that previous studies have indicated that mHealth app–based interventions did not effectively decrease SB in children and adolescents. This observation may stem from the fact that the use of apps may result in increased SB time, consequently reallocating the time resources of children and adolescents [12]. However, the findings of this study do not support this conclusion. A separate study reported that the intervention using Fit Survivor apps on SB in adolescents was ineffective, which is possibly due to the characteristics of the population; notably, most participants in these studies had cancer, which may require a longer period of SB to observe significant changes [53]. The findings regarding BMI indicate that parent-focused mHealth app–based interventions have not been successful in reducing BMI among children and adolescents [29], somewhat not entirely consistent with the findings of this study. Our meta-regression analysis revealed that age significantly influences the effectiveness of app-based interventions.

This phenomenon may be closely associated with willingness of users and their behavioral intention to use apps. Individual differences, such as age and gender, play a crucial role in moderating willingness and intention of users. In general, behavioral habits and expected performance serve as predictors of willingness and intention of users [59], and age can increase the frequency of use by moderating behavioral habit and performance [60], which in turn have an influence on SB. In addition, most of the app intervention is parenting-focused intervention among preschoolers [61]. The successful implementation of interventions relies on parental involvement, thus parental attitudes significantly influence the participation of preschoolers aged 3-6 years, potentially impacting the efficacy of mHealth app interventions [13]. In the case of children and adolescents aged 7-18 years, increasing age correlates with improved expected performance [62]. In adolescents, this performance expectation is closely linked to user intention, bolstering their willingness to engage with apps. However, older adolescents, who prioritize academic achievement, may exhibit reduced effectiveness in app interventions [63]. In conclusion, the effects of mHealth app–based interventions may vary across different age groups. However, only 4 studies included subjects aged 3-6 years, potentially affecting the generalizability of the findings. Therefore, further validation through a larger number of high-quality studies is warranted to strengthen the conclusions.

Types of Intervention

Concerning types of interventions, there is a lack of systematic reviews analyzing the impacts of stand-alone apps and concerted interventions on PA and PF, including their distinctions. Subgroup analysis revealed that concerted interventions significantly increased TPA and reduced SB and BMI in children and adolescents, whereas stand-alone app interventions had only a modest positive effect on TPA with a small effect size. The findings diverge somewhat from prior research. Some studies indicate that interventions using COOL Passport apps or a combination of COOL Passport apps and game-based interactive platforms do not enhance PA in youth with congenital heart disease [40]. Conversely, other studies suggest that mobile apps can increase TPA while decreasing SB and BMI in youth [46,47]. This phenomenon may be associated with behavior change interventions. Effective behavior change interventions can enhance health status and reduce health care spending, with theory-based interventions often yielding superior results [64]. Hence, interventions based on the high-scoring theoretical paradigm, or in combination with other specific high-scoring strategies, are more likely to yield positive outcomes [65]. In summary, interventions integrating apps with high-scoring strategies are probably more effective than stand-alone apps, although the optimal number and type of strategies remain undetermined.

Theoretical Paradigm

Effective behavior change interventions can improve health status and reduce health care spending, and theory-based interventions generally yielded superior results [64]. Currently, mHealth app interventions use multiple types of BCT that have been shown to be effective in influencing intervention outcomes [21]. Of the 28 studies, 8 (29%) did not use theoretical frameworks in their mHealth app interventions [31,36,37,45,47,49,50,52], while the remaining 20 (71%) studies were grounded in theoretical paradigms such as SCT, SDT, and SRT.

The most commonly used theoretical paradigms in the research are SCT and SDT, and interventions based on these frameworks have demonstrated positive effects on TPA, MVPA, and SB in children and adolescents. Subgroup analysis revealed that interventions based on SDT significantly increased TPA and MVPA in children and adolescents, while interventions based on SCT significantly reduced SB in this population. Previous studies have consistently shown that interventions based on SCT effectively improve PA among adolescents, which aligns with the findings of this study [54]. SCT is commonly used to elucidate the mechanisms underlying the improvement in PA through behavior change interventions. SCT posits that PA is influenced by personal, social, and environmental factors. Personal factors, including self-efficacy, self-management, and expected performance, play a crucial role in enhancing PA and reducing SB. mHealth apps can effectively promote the increase in PA and the reduction in SB by modifying these factors [66].

SDT is frequently used to account for variations in individual behavior resulting from differences in intrinsic and extrinsic motivation. Intrinsic motivation, which is a crucial predictor of PA levels, is lacking in approximately 40% of Europeans; this situation leads to failure to meet the recommended PA levels [67]. In contrast to extrinsic motivation, intrinsic motivation drives the initiation and sustained engagement in individual behaviors. Thus, interventions grounded in SDT can effectively enhance leisure-time PA and PA in physical education classes and reduce SB among adolescents by fostering intrinsic motivation [68]. Furthermore, interventions using mHealth apps based on the health belief model and theory of planned behavior may yield positive effects on increasing PA in children and adolescents. However, which theory yields optimal intervention effects requires further investigation. In conclusion, interventions based on SCT or SDT within mHealth apps may likely yield more effective results. However, due to the limited number of included studies, further validation through a larger volume of high-quality research is necessary to bolster these conclusions.

BCT Clusters

BCT clusters represent specific implementation strategies in behavior change interventions. Previous studies have demonstrated that interventions incorporating a specific number of BCT clusters are more effective in promoting the PA of users. However, further investigation is needed to determine the optimal combination of the number and type of BCT clusters to achieve optimal intervention effects [69].

The BCT clusters that emerged frequently in this study were feedback and monitoring (n=22), goals and planning (n=19), shaping knowledge (n=12), and social support (n=11). A previous study [70] focusing on app interventions targeting PA and SB in adults also identified feedback and monitoring as well as goals and planning as the most commonly used BCT clusters. Another study [16] also highlighted the positive effects of mHealth app–based interventions on PA and SB, particularly when specific BCT clusters such as goals and planning, feedback and monitoring, and social support were implemented. Consequently, specific BCT clusters, including goals and planning, feedback, and monitoring, play a crucial role in influencing the effectiveness of mHealth app–based interventions, ensuring increased TPA, decreased SB, and reduced BMI by enhancing the willingness and engagement of users [71].

Limited research exists on the impact of varying numbers of BCT clusters implemented on PA and PF in children and adolescents. In this study, we discovered that mHealth app–based interventions incorporating 4 and 7-10 BCT clusters resulted in a significant reduction in SB among child adolescents, and the latter showed a greater effect size than the former. Conroy et al [72] reported a mean of 4.2 for the use of BCT in app. The number of studies included in the analysis varies, and most studies used 4 BCT clusters, which is slightly lower than the number of BCT used in previous studies. This discrepancy may be attributed to the characteristics of the subjects [73]. Moreover, more BCT clusters in mHealth apps do not necessarily lead to better outcomes. Excessive use of BCT can decrease the willingness of users and their frequency of app use [74]. Thus, further research is needed to determine the optimal number of BCT clusters.

This phenomenon appears to be strongly associated with the willingness of users and their engagement in using apps. Feedback and monitoring, which involves design patterns enabling users to track their performance or status, play a crucial role in enhancing the trust, motivation, and engagement with the app of users. Consequently, this situation leads to the effective promotion of PA levels and reduction in BMI [75]. Similarly, goals and planning, which involve planned behaviors and the conversion of the intentions of users into actionable steps through self-regulation and self-efficacy, have been shown to improve PA and reduce BMI [76].

Types of Apps and Intervention Duration

The available evidence suggests that commercial apps did not yield improvements in PA or reductions in SB among children and adolescents [20,31-33,35]. However, Jasmine et al [77] found that integrating commercial apps with web-based social networking platforms was effective in improving PA. However, a certain use frequency was needed, and a significant correlation between frequency of use and PA was observed. Our study discovered that research apps significantly reduced SB, while commercial apps increased TPA in children and adolescents, but the effect size was small. However, the effect size was small for commercial apps, which contrasts with our expectations. This discrepancy may be attributed to the design and characteristics of the apps. Commercial app developers often prioritize user interface simplification and the inclusion of complex features to enhance app engagement. However, they may overlook incorporating theoretical frameworks and BCT clusters that are closely associated with intervention effectiveness [78].

An intriguing finding emerged regarding the duration of the interventions. The 2- to 4-week and 8- to 12-week interventions significantly increased TPA, but the former demonstrated superior effectiveness to the latter; 8- to 12-week interventions significantly increased MVPA. Furthermore, the 20- to 48-week intervention significantly reduced SB and BMI in children and adolescents. This finding aligns with the results of previous

studies that short-term mHealth app interventions effectively increased the PA of children, while longer interventions yielded diminished outcomes [40,51]. The results of this study support the notion that intervention effects are closely linked to the willingness of participants to engage. The reason is that shortly after engaging in an mHealth app–based intervention, children and adolescents will experience a sensitive attraction period [79]; prolonged interventions may lead to diminished interest and compliance, thereby weakening the impact. Another plausible explanation is that behavior change interventions rooted in theoretical frameworks yield better outcomes, but they require sufficient time for users to develop habitual behaviors conducive to lasting positive changes, and interventions targeting SB and BMI take even longer. In addition, our meta-regression analysis revealed that intervention duration significantly influences the effectiveness of app-based interventions.

Limitations

This study has several limitations. The search was restricted to English literature, which may introduce language bias and affect the reliability of the findings. Only a limited number of published articles were included, which potentially omit unpublished manuscripts and other sources (eg, conference papers and dissertations) that could contribute to publication bias. The outcome measures used in this study had inconsistent units, and the interpretation of the results using SMD as effect indicators requires caution. PF encompasses multiple dimensions, and different studies focused on various subdimensions.

Most studies (21/28, 75%) encompassed multiple metrics, and some studies (11/28, 39%) were not primarily designed to enhance PA and PF. These factors could have biased the results, potentially limiting the effect sizes and hindering a full reflection of the true effects in this study. Despite our extensive meta-regression and subgroup analyses, only age and intervention period demonstrated significant associations. Heterogeneity for some outcome metrics remained high, and the 95% CIs for the effect size overlapped between groups. Consequently, caution is advised when interpreting partial conclusions from subgroups. Our findings indicate heterogeneity in some studies, and the source of this heterogeneity remains unclear. The effectiveness of mHealth app–based interventions may vary among children and adolescents based on demographics (eg, ethnic backgrounds, regions, genders, and BMI), economic levels, intervention models (parent centered vs child centered), and the degree of app individualization. These factors could contribute to the observed heterogeneity. However, due to limitations stemming from the number and characteristics of included studies, these factors were not analyzed in this study. In conclusion, the findings of this systematic review and meta-analysis should be interpreted cautiously.

Conclusions

The findings from a systematic review and meta-analysis indicate that mHealth app–based interventions hold significant potential as a therapeutic strategy for increasing TPA levels, reducing BMI and SB, and improving agility and muscle strength among children and adolescents. However, these interventions insignificantly affected MVPA, WC, CRF, muscular power and endurance, and flexibility, which are crucial for promoting PA and enhancing PF. Age, app types, types of intervention, theoretical paradigms, BCT clusters, and intervention duration emerged as important moderating variables that influence the effectiveness of mHealth app–based interventions. These moderating effects should be considered during the design, preparation, and promotion of interventions.

Considering the potential benefits of using mHealth apps, future research should explore the combination of different types and quantities of mHealth app–based interventions to determine the optimal approach for increasing PA levels and improving PF in children and adolescents. Furthermore, future studies could investigate the impact of additional influences on the intervention effectiveness of mHealth app interventions, such as demographics characteristics, economic levels, and intervention models.

Acknowledgments

The authors would like to thank the staff and faculty at the Institute of Sports Medicine and Health at Chengdu Sport University, which offered the funding support and the place to conduct the study. This study was funded by the National Social Science Fund (grant 22BTY063), the Key R&D Projects of the Sichuan Provincial Science and Technology Plan Project (grant 2022YFS0053), the Soft Science Project of Sichuan Provincial Department of Science and Technology (grant 2022JDR0314), the Key Project of Sichuan Provincial Philosophy and Social Science Foundation (grant CJJ23ND77), the Sports Medicine Key Laboratory of Sichuan Province and Sports Medicine Key Laboratory of State Sport General Administration (grant 2024-A002), and the “14th Five Year Plan” Scientific Research and Innovation Team of Chengdu Sport University (grant 23CXTD02).

Authors' Contributions

JWW designed the study protocol and drafted the report; ZZ, ZS, and JF searched the literature, analyzed the data, and interpreted the results; YJ, ZLG, and WDC contributed to the study protocol and reviewed the manuscript; and XL reviewed and revised the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Literature search strategy.
[PDF File (Adobe PDF File), 69 KB - mhealth_v12i1e51478_app1.pdf]

Multimedia Appendix 2
Sensitivity analyses results on total physical activity.
[PNG File , 271 KB - mhealth_v12i1e51478_app2.png]

Multimedia Appendix 3
Sensitivity analyses results on sedentary behavior.
[PNG File , 204 KB - mhealth_v12i1e51478_app3.png]

Multimedia Appendix 4
Sensitivity analyses results on moderate to vigorous physical activity.
[PNG File , 202 KB - mhealth_v12i1e51478_app4.png]

Multimedia Appendix 5
Sensitivity analyses results on BMI.
[PNG File , 198 KB - mhealth_v12i1e51478_app5.png]

Multimedia Appendix 6
Summary of subgroup analysis results of mobile health app–based interventions on total physical activity.
[PDF File (Adobe PDF File), 86 KB - mhealth_v12i1e51478_app6.pdf]

Multimedia Appendix 7
Summary of subgroup analysis results of mobile health app–based interventions on sedentary behavior.
[PDF File (Adobe PDF File), 86 KB - mhealth_v12i1e51478_app7.pdf]

Multimedia Appendix 8
Summary of subgroup analysis results of mobile health app–based interventions on moderate to vigorous physical activity.
[PDF File (Adobe PDF File), 86 KB - mhealth_v12i1e51478_app8.pdf]

Multimedia Appendix 9
Summary of subgroup analysis results of mobile health app–based interventions on BMI.
[PDF File (Adobe PDF File), 85 KB - mhealth_v12i1e51478_app9.pdf]

Multimedia Appendix 10
Funnel plot of total physical activity.
[PDF File (Adobe PDF File), 28 KB - mhealth_v12i1e51478_app10.pdf]

Multimedia Appendix 11
Funnel plot of sedentary behavior.
[PDF File (Adobe PDF File), 27 KB - mhealth_v12i1e51478_app11.pdf]

Multimedia Appendix 12
Funnel plot of BMI.
[PDF File (Adobe PDF File), 19 KB - mhealth_v12i1e51478_app12.pdf]

Multimedia Appendix 13
Funnel plot of moderate to vigorous physical activity.
[PDF File (Adobe PDF File), 27 KB - mhealth_v12i1e51478_app13.pdf]

Multimedia Appendix 14
Egger test results on total physical activity.
[PNG File , 114 KB - mhealth_v12i1e51478_app14.png]

Multimedia Appendix 15
Egger test results sedentary behavior.
[PNG File , 94 KB - mhealth_v12i1e51478_app15.png]

Multimedia Appendix 16
Egger test results moderate to vigorous physical activity.
[PNG File , 96 KB - mhealth_v12i1e51478_app16.png]

Multimedia Appendix 17
Egger test results BMI.
[PNG File , 102 KB - mhealth_v12i1e51478_app17.png]

Multimedia Appendix 18
PRISMA 2020 Checklist.
[DOCX File , 28 KB - mhealth_v12i1e51478_app18.docx]

References

1. Singh S, Roy D, Sinha K, Parveen S, Sharma G, Joshi G. Impact of COVID-19 and lockdown on mental health of children and adolescents: a narrative review with recommendations. Psychiatry Res 2020 Nov;293:113429 [FREE Full text] [doi: 10.1016/j.psychres.2020.113429] [Medline: 32882598]
2. Guthold R, Stevens GA, Riley LM, Bull FC. Global trends in insufficient physical activity among adolescents: a pooled analysis of 298 population-based surveys with 1·6 million participants. Lancet Child Adolesc Health 2020 Jan;4(1):23-35 [FREE Full text] [doi: 10.1016/S2352-4642(19)30323-2] [Medline: 31761562]
3. Neil-Sztramko SE, Caldwell H, Dobbins M. School-based physical activity programs for promoting physical activity and fitness in children and adolescents aged 6 to 18. Cochrane Database Syst Rev 2021 Sep 23;9(9):CD007651 [FREE Full text] [doi: 10.1002/14651858.CD007651.pub3] [Medline: 34555181]
4. Cachón-Zagalaz J, Sánchez-Zafra M, Sanabrias-Moreno D, González-Valero G, Lara-Sánchez AJ, Zagalaz-Sánchez ML. Systematic review of the literature about the effects of the COVID-19 pandemic on the lives of school children. Front Psychol 2020;11:569348 [FREE Full text] [doi: 10.3389/fpsyg.2020.569348] [Medline: 33162910]
5. Basterfield L, Burn NL, Galna B, Batten H, Goffe L, Karoblyte G, et al. Changes in children's physical fitness, BMI and health-related quality of life after the first 2020 COVID-19 lockdown in England: a longitudinal study. J Sports Sci 2022 May;40(10):1088-1096. [doi: 10.1080/02640414.2022.2047504] [Medline: 35262450]
6. Ortega FB, Ruiz JR, Castillo MJ, Sjöström M. Physical fitness in childhood and adolescence: a powerful marker of health. Int J Obes (Lond) 2008 Jan;32(1):1-11. [doi: 10.1038/sj.ijo.0803774] [Medline: 18043605]
7. Lämmle L, Worth A, Bös K. Socio-demographic correlates of physical activity and physical fitness in German children and adolescents. Eur J Public Health 2012 Dec;22(6):880-884. [doi: 10.1093/eurpub/ckr191] [Medline: 22219503]
8. Júdice PB, Silva AM, Berria J, Petroski EL, Ekelund U, Sardinha LB. Sedentary patterns, physical activity and health-related physical fitness in youth: a cross-sectional study. Int J Behav Nutr Phys Act 2017 Mar 04;14(1):25 [FREE Full text] [doi: 10.1186/s12966-017-0481-3] [Medline: 28259140]
9. Krebs P, Duncan DT. Health app use among US mobile phone owners: a national survey. JMIR Mhealth Uhealth 2015 Nov 04;3(4):e101 [FREE Full text] [doi: 10.2196/mhealth.4924] [Medline: 26537656]
10. Brannon EE, Cushing CC. A systematic review: is there an app for that? Translational science of pediatric behavior change for physical activity and dietary interventions. J Pediatr Psychol 2015 May;40(4):373-384. [doi: 10.1093/jpepsy/jsu108] [Medline: 25502745]
11. Rising CJ, Jensen RE, Moser RP, Oh A. Characterizing the US population by patterns of mobile health use for health and behavioral tracking: analysis of the national cancer institute's health information national trends survey data. J Med Internet Res 2020 May 14;22(5):e16299 [FREE Full text] [doi: 10.2196/16299] [Medline: 32406865]
12. Baumann H, Fiedler J, Wunsch K, Woll A, Wollesen B. mHealth interventions to reduce physical inactivity and sedentary behavior in children and adolescents: systematic review and meta-analysis of randomized controlled trials. JMIR Mhealth Uhealth 2022 May 11;10(5):e35920 [FREE Full text] [doi: 10.2196/35920] [Medline: 35544294]
13. He Z, Wu H, Yu F, Fu J, Sun S, Huang T, et al. Effects of smartphone-based interventions on physical activity in children and adolescents: systematic review and meta-analysis. JMIR Mhealth Uhealth 2021 Feb 01;9(2):e22601 [FREE Full text] [doi: 10.2196/22601] [Medline: 33522980]
14. Byambasuren O, Sanders S, Beller E, Glasziou P. Prescribable mHealth apps identified from an overview of systematic reviews. NPJ Digit Med 2018;1:12 [FREE Full text] [doi: 10.1038/s41746-018-0021-9] [Medline: 31304297]
15. Higgins JP. Smartphone applications for patients' health and fitness. Am J Med 2016 Jan;129(1):11-19. [doi: 10.1016/j.amjmed.2015.05.038] [Medline: 26091764]

16. Yerrakalva D, Yerrakalva D, Hajna S, Griffin S. Effects of mobile health app interventions on sedentary time, physical activity, and fitness in older adults: systematic review and meta-analysis. J Med Internet Res 2019 Nov 28;21(11):e14343 [FREE Full text] [doi: 10.2196/14343] [Medline: 31778121]
17. Domin A, Spruijt-Metz D, Theisen D, Ouzzahra Y, Vögele C. Smartphone-based interventions for physical activity promotion: scoping review of the evidence over the last 10 years. JMIR Mhealth Uhealth 2021 Jul 21;9(7):e24308 [FREE Full text] [doi: 10.2196/24308] [Medline: 34287209]
18. Direito A, Carraça E, Rawstorn J, Whittaker R, Maddison R. mHealth technologies to influence physical activity and sedentary behaviors: behavior change techniques, systematic review and meta-analysis of randomized controlled trials. Ann Behav Med 2017 Apr;51(2):226-239. [doi: 10.1007/s12160-016-9846-0] [Medline: 27757789]
19. Litman L, Rosen Z, Spierer D, Weinberger-Litman S, Goldschein A, Robinson J. Mobile exercise apps and increased leisure time exercise activity: a moderated mediation analysis of the role of self-efficacy and barriers. J Med Internet Res 2015 Aug 14;17(8):e195 [FREE Full text] [doi: 10.2196/jmir.4142] [Medline: 26276227]
20. Direito A, Jiang Y, Whittaker R, Maddison R. Apps for IMproving FITness and increasing physical activity among young people: the AIMFIT pragmatic randomized controlled trial. J Med Internet Res 2015 Aug 27;17(8):e210 [FREE Full text] [doi: 10.2196/jmir.4568] [Medline: 26316499]
21. Fiedler J, Eckert T, Wunsch K, Woll A. Key facets to build up eHealth and mHealth interventions to enhance physical activity, sedentary behavior and nutrition in healthy subjects - an umbrella review. BMC Public Health 2020 Oct 23;20(1):1605 [FREE Full text] [doi: 10.1186/s12889-020-09700-7] [Medline: 33097013]
22. Yang Q, Van Stee SK. The comparative effectiveness of mobile phone interventions in improving health outcomes: meta-analytic review. JMIR Mhealth Uhealth 2019 Apr 03;7(4):e11244 [FREE Full text] [doi: 10.2196/11244] [Medline: 30942695]
23. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95 [FREE Full text] [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
24. Jüni P, Egger M. PRISMAtic reporting of systematic reviews and meta-analyses. Lancet 2009 Oct 10;374(9697):1221-1223. [doi: 10.1016/S0140-6736(09)61765-7] [Medline: 19819375]
25. Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Cochrane Bias Methods Group, Cochrane Statistical Methods Group. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011 Oct 18;343:d5928 [FREE Full text] [doi: 10.1136/bmj.d5928] [Medline: 22008217]
26. Nyström CD, Sandin S, Henriksson P, Henriksson H, Trolle-Lagerros Y, Larsson C, et al. Mobile-based intervention intended to stop obesity in preschool-aged children: the MINISTOP randomized controlled trial. Am J Clin Nutr 2017 Jun;105(6):1327-1335 [FREE Full text] [doi: 10.3945/ajcn.116.150995] [Medline: 28446496]
27. Higgins JP, Thomas J, Chandler J, Cumpston MJ, Li T, Page M, et al. Cochrane Handbook for Systematic Reviews of Interventions. Hoboken, NJ: John Wiley & Sons; 2019.
28. Garde A, Umedaly A, Abulnaga SM, Junker A, Chanoine JP, Johnson M, et al. Evaluation of a novel mobile exergame in a school-based environment. Cyberpsychol Behav Soc Netw 2016 Mar;19(3):186-192. [doi: 10.1089/cyber.2015.0281] [Medline: 26882222]
29. Lubans DR, Smith JJ, Plotnikoff RC, Dally KA, Okely AD, Salmon J, et al. Assessing the sustained impact of a school-based obesity prevention program for adolescent boys: the ATLAS cluster randomized controlled trial. Int J Behav Nutr Phys Act 2016 Aug 20;13:92 [FREE Full text] [doi: 10.1186/s12966-016-0420-8] [Medline: 27542825]
30. Smith JJ, Morgan PJ, Plotnikoff RC, Dally KA, Salmon J, Okely AD, et al. Smart-phone obesity prevention trial for adolescent boys in low-income communities: the ATLAS RCT. Pediatrics 2014 Sep;134(3):e723-e731. [doi: 10.1542/peds.2014-1012] [Medline: 25157000]
31. Fernandez-Luque L, Singh M, Ofli F, Mejova YA, Weber I, Aupetit M, et al. Erratum to: implementing 360° quantified self for childhood obesity: feasibility study and experiences from a weight loss camp in Qatar. BMC Med Inform Decis Mak 2017 May 12;17(1):62 [FREE Full text] [doi: 10.1186/s12911-017-0457-x] [Medline: 28499380]
32. Pyky R, Koivumaa-Honkanen H, Leinonen AM, Ahola R, Hirvonen N, Enwald H, et al. Effect of tailored, gamified, mobile physical activity intervention on life satisfaction and self-rated health in young adolescent men: a population-based, randomized controlled trial (MOPO study). Comput Human Behav 2017 Jul;72:13-22. [doi: 10.1016/j.chb.2017.02.032]
33. Gaudet J, Gallant F, Bélanger M. A bit of fit: minimalist intervention in adolescents based on a physical activity tracker. JMIR Mhealth Uhealth 2017 Jul 06;5(7):e92 [FREE Full text] [doi: 10.2196/mhealth.7647] [Medline: 28684384]
34. Mendoza JA, Baker KS, Moreno MA, Whitlock K, Abbey-Lambertz M, Waite A, et al. A Fitbit and Facebook mHealth intervention for promoting physical activity among adolescent and young adult childhood cancer survivors: a pilot study. Pediatr Blood Cancer 2017 Dec;64(12). [doi: 10.1002/pbc.26660] [Medline: 28618158]
35. Chen J, Guedes CM, Lung AE. Smartphone-based healthy weight management intervention for Chinese American adolescents: short-term efficacy and factors associated with decreased weight. J Adolesc Health 2019 Apr;64(4):443-449. [doi: 10.1016/j.jadohealth.2018.08.022] [Medline: 30409751]

36. Kahana R, Kremer S, Dekel Dahari M, Kodesh E. The effect of incorporating an exergame application in a multidisciplinary weight management program on physical activity and fitness indices in children with overweight and obesity. Children (Basel) 2021 Dec 29;9(1):18 [FREE Full text] [doi: 10.3390/children9010018] [Medline: 35053643]
37. Liu Z, Gao P, Gao AY, Lin Y, Feng XX, Zhang F, et al. Effectiveness of a multifaceted intervention for prevention of obesity in primary school children in China: a cluster randomized clinical trial. JAMA Pediatr 2022 Jan 01;176(1):e214375 [FREE Full text] [doi: 10.1001/jamapediatrics.2021.4375] [Medline: 34747972]
38. Ridgers ND, Timperio A, Ball K, Lai SK, Brown H, Macfarlane S, et al. Effect of commercial wearables and digital behaviour change resources on the physical activity of adolescents attending schools in socio-economically disadvantaged areas: the RAW-PA cluster-randomised controlled trial. Int J Behav Nutr Phys Act 2021 Apr 12;18(1):52 [FREE Full text] [doi: 10.1186/s12966-021-01110-1] [Medline: 33845853]
39. Likhitweerawong N, Boonchooduang N, Kittisakmontri K, Chonchaiya W, Louthrenoo O. Effectiveness of mobile application on changing weight, healthy eating habits, and quality of life in children and adolescents with obesity: a randomized controlled trial. BMC Pediatr 2021 Nov 10;21(1):499 [FREE Full text] [doi: 10.1186/s12887-021-02980-x] [Medline: 34758761]
40. Lin PJ, Fanjiang YY, Wang JK, Lu CW, Lin KC, Cheong IM, et al. Long-term effectiveness of an mHealth-tailored physical activity intervention in youth with congenital heart disease: a randomized controlled trial. J Adv Nurs 2021 Aug;77(8):3494-3506. [doi: 10.1111/jan.14924] [Medline: 34151444]
41. Seah ML, Koh KT. The efficacy of using mobile applications in changing adolescent girls' physical activity behaviour during weekends. Eur Phys Educ Rev 2020 Jun 17;27(1):113-131. [doi: 10.1177/1356336x20930741]
42. Stasinaki A, Büchter D, Shih CI, Heldt K, Güsewell S, Brogle B, et al. Effects of a novel mobile health intervention compared to a multi-component behaviour changing program on body mass index, physical capacities and stress parameters in adolescents with obesity: a randomized controlled trial. BMC Pediatr 2021 Jul 09;21(1):308 [FREE Full text] [doi: 10.1186/s12887-021-02781-2] [Medline: 34243738]
43. Tugault-Lafleur CN, De-Jongh González O, Macdonald J, Bradbury J, Warshawski T, Ball GD, et al. Efficacy of the Aim2Be intervention in changing lifestyle behaviors among adolescents with overweight and obesity: randomized controlled trial. J Med Internet Res 2023 Apr 25;25:e38545 [FREE Full text] [doi: 10.2196/38545] [Medline: 37097726]
44. Han X, Tian Z, Zhao M, Zhou Z. An intervention pattern of family parent-child physical activity based on a smartphone app for preschool children during COVID-19. Biomed Res Int 2022;2022:2777079 [FREE Full text] [doi: 10.1155/2022/2777079] [Medline: 36046455]
45. Oh W, An Y, Min S, Park C. Comparative effectiveness of artificial intelligence-based interactive home exercise applications in adolescents with obesity. Sensors (Basel) 2022 Sep 28;22(19):7352 [FREE Full text] [doi: 10.3390/s22197352] [Medline: 36236451]
46. Staiano AE, Newton RL, Beyl RA, Kracht CL, Hendrick CA, Viverito M, et al. mHealth intervention for motor skills: a randomized controlled trial. Pediatrics 2022 May 01;149(5):e2021053362 [FREE Full text] [doi: 10.1542/peds.2021-053362] [Medline: 35415743]
47. Mateo-Orcajada A, Abenza-Cano L, Albaladejo-Saura MD, Vaquero-Cristóbal R. Mandatory after-school use of step tracker apps improves physical activity, body composition and fitness of adolescents. Educ Inf Technol (Dordr) 2023 Jan 24:1-32 [FREE Full text] [doi: 10.1007/s10639-023-11584-0] [Medline: 36714445]
48. Alexandrou C, Henriksson H, Henström M, Henriksson P, Delisle Nyström C, Bendtsen M, et al. Effectiveness of a smartphone app (MINISTOP 2.0) integrated in primary child health care to promote healthy diet and physical activity behaviors and prevent obesity in preschool-aged children: randomized controlled trial. Int J Behav Nutr Phys Act 2023 Feb 21;20(1):22 [FREE Full text] [doi: 10.1186/s12966-023-01405-5] [Medline: 36810069]
49. Mateo-Orcajada A, Vaquero-Cristóbal R, Abenza-Cano L. Influence of Pokémon go playing style on physical activity and its effect on Kinanthropometry variables and body composition in adolescents. J Phys Act Health 2023 Oct 01;20(10):934-947. [doi: 10.1123/jpah.2023-0140] [Medline: 37433524]
50. Trost SG, Brookes DS. Effectiveness of a novel digital application to promote fundamental movement skills in 3- to 6-year-old children: a randomized controlled trial. J Sports Sci 2021 Feb;39(4):453-459. [doi: 10.1080/02640414.2020.1826657] [Medline: 32985373]
51. Garde A, Chowdhury M, Rollinson AU, Johnson M, Prescod P, Chanoine JP, et al. A multi-week assessment of a mobile exergame intervention in an elementary school. Games Health J 2018 Feb;7(1):1-8. [doi: 10.1089/g4h.2017.0023] [Medline: 29394109]
52. Browne S, Kechadi MT, O'Donnell S, Dow M, Tully L, Doyle G, et al. Mobile health apps in pediatric obesity treatment: process outcomes from a feasibility study of a multicomponent intervention. JMIR Mhealth Uhealth 2020 Jul 08;8(7):e16925 [FREE Full text] [doi: 10.2196/16925] [Medline: 32673267]
53. Devine KA, Viola A, Levonyan-Radloff K, Mackowski N, Bozzini B, Chandler A, et al. Feasibility of FitSurvivor: a technology-enhanced group-based fitness intervention for adolescent and young adult survivors of childhood cancer. Pediatr Blood Cancer 2020 Sep 26;67(9):e28530 [FREE Full text] [doi: 10.1002/pbc.28530] [Medline: 32589339]

54. Liu J, Zeng M, Wang D, Zhang Y, Shang B, Ma X. Applying social cognitive theory in predicting physical activity among Chinese adolescents: a cross-sectional study with multigroup structural equation model. Front Psychol 2021;12:695241 [FREE Full text] [doi: 10.3389/fpsyg.2021.695241] [Medline: 35370866]
55. Khamzina M, Parab KV, An R, Bullard T, Grigsby-Toussaint DS. Impact of Pokémon Go on physical activity: a systematic review and meta-analysis. Am J Prev Med 2020 Feb;58(2):270-282. [doi: 10.1016/j.amepre.2019.09.005] [Medline: 31836333]
56. Tully L, Burls A, Sorensen J, El-Moslemany R, O'Malley G. Mobile health for pediatric weight management: systematic scoping review. JMIR Mhealth Uhealth 2020 Jun 03;8(6):e16214 [FREE Full text] [doi: 10.2196/16214] [Medline: 32490849]
57. Booth ML, Hunter C, Gore CJ, Bauman A, Owen N. The relationship between body mass index and waist circumference: implications for estimates of the population prevalence of overweight. Int J Obes Relat Metab Disord 2000 Aug;24(8):1058-1061. [doi: 10.1038/sj.ijo.0801359] [Medline: 10951546]
58. Emberson MA, Lalande A, Wang D, McDonough DJ, Liu W, Gao Z. Effectiveness of smartphone-based physical activity interventions on individuals' health outcomes: a systematic review. Biomed Res Int 2021;2021:6296896 [FREE Full text] [doi: 10.1155/2021/6296896] [Medline: 34409104]
59. Venkatesh V, Thong JY, Xu X. Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. MIS Q 2012;36(1):157. [doi: 10.2307/41410412]
60. Schoeppe S, Alley S, Van Lippevelde W, Bray NA, Williams SL, Duncan MJ, et al. Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: a systematic review. Int J Behav Nutr Phys Act 2016 Dec 07;13(1):127 [FREE Full text] [doi: 10.1186/s12966-016-0454-y] [Medline: 27927218]
61. Webster EK, Kracht CL, Newton Jr RL, Beyl RA, Staiano AE. Intervention to improve preschool children's fundamental motor skills: protocol for a parent-focused, mobile app-based comparative effectiveness trial. JMIR Res Protoc 2020 Oct 20;9(10):e19943 [FREE Full text] [doi: 10.2196/19943] [Medline: 33079066]
62. Mateo-Orcajada A, Vaquero-Cristóbal R, Abenza-Cano L. Gender and academic year as moderators of the efficacy of mobile app interventions to promote physical activity in adolescents: a randomized controlled trial. Humanit Soc Sci Commun 2023 Dec 22;10(1) [FREE Full text] [doi: 10.1057/s41599-023-02502-3]
63. Morales-Vives F, Camps E, Dueñas JM. Predicting academic achievement in adolescents: the role of maturity, intelligence and personality. Psicothema 2020 Feb;32(1):84-91. [doi: 10.7334/psicothema2019.262] [Medline: 31954420]
64. Connell LE, Carey RN, de Bruin M, Rothman AJ, Johnston M, Kelly MP, et al. Links between behavior change techniques and mechanisms of action: an expert consensus study. Ann Behav Med 2019 Jul 17;53(8):708-720 [FREE Full text] [doi: 10.1093/abm/kay082] [Medline: 30452535]
65. Hale K, Capra S, Bauer J. A framework to assist health professionals in recommending high-quality apps for supporting chronic disease self-management: illustrative assessment of type 2 diabetes apps. JMIR Mhealth Uhealth 2015 Sep 14;3(3):e87 [FREE Full text] [doi: 10.2196/mhealth.4532] [Medline: 26369346]
66. Elizabeth Morghen S, Robert WM. Pilot study of social cognitive theory variables as correlates of physical activity among adolescents with pediatric multiple sclerosis. Mult Scler Relat Disord 2020 Jul;42:102026. [doi: 10.1016/j.msard.2020.102026] [Medline: 32272443]
67. Ng JY, Ntoumanis N, Thøgersen-Ntoumani C, Deci EL, Ryan RM, Duda JL, et al. Self-determination theory applied to health contexts: a meta-analysis. Perspect Psychol Sci 2012 Jul;7(4):325-340. [doi: 10.1177/1745691612447309] [Medline: 26168470]
68. Nogg KA, Vaughn AA, Levy SS, Blashill AJ. Motivation for physical activity among U.S. adolescents: a self-determination theory perspective. Ann Behav Med 2021 Mar 16;55(2):133-143. [doi: 10.1093/abm/kaaa037] [Medline: 32756874]
69. Hu L, Ping L. Can sports and fitness apps promote physical exercise? Evidence from a meta-analysis. China Sports Sci Technol 2021;58(06):94-103 [FREE Full text] [doi: 10.16470/j.csst.2020008]
70. Bardus M, van Beurden SB, Smith JR, Abraham C. A review and content analysis of engagement, functionality, aesthetics, information quality, and change techniques in the most popular commercial apps for weight management. Int J Behav Nutr Phys Act 2016 Mar 10;13:35 [FREE Full text] [doi: 10.1186/s12966-016-0359-9] [Medline: 26964880]
71. Muellmann S, Forberger S, Möllers T, Zeeb H, Pischke CR. Effectiveness of eHealth interventions for the promotion of physical activity in older adults: a systematic review protocol. Syst Rev 2016 Mar 16;5:47 [FREE Full text] [doi: 10.1186/s13643-016-0223-7] [Medline: 26984189]
72. Conroy DE, Yang CH, Maher JP. Behavior change techniques in top-ranked mobile apps for physical activity. Am J Prev Med 2014 Jun;46(6):649-652. [doi: 10.1016/j.amepre.2014.01.010] [Medline: 24842742]
73. Schoeppe S, Alley S, Rebar AL, Hayman M, Bray NA, Van Lippevelde W, et al. Apps to improve diet, physical activity and sedentary behaviour in children and adolescents: a review of quality, features and behaviour change techniques. Int J Behav Nutr Phys Act 2017 Jun 24;14(1):83 [FREE Full text] [doi: 10.1186/s12966-017-0538-3] [Medline: 28646889]
74. Bondaronek P, Alkhaldi G, Slee A, Hamilton FL, Murray E. Quality of publicly available physical activity apps: review and content analysis. JMIR Mhealth Uhealth 2018 Mar 21;6(3):e53 [FREE Full text] [doi: 10.2196/mhealth.9069] [Medline: 29563080]

75. Vinnikova A, Lu L, Wei J, Fang G, Yan J. The use of smartphone fitness applications: the role of self-efficacy and self-regulation. Int J Environ Res Public Health 2020 Oct 20;17(20):7639 [FREE Full text] [doi: 10.3390/ijerph17207639] [Medline: 33092090]
76. Wang YS, Lin HH, Luarn P. Predicting consumer intention to use mobile service. Inf Syst J 2006 Mar 02;16(2):157-179 [FREE Full text] [doi: 10.1111/j.1365-2575.2006.00213.x]
77. Petersen JM, Kemps E, Lewis LK, Prichard I. Associations between commercial app use and physical activity: cross-sectional study. J Med Internet Res 2020 Jun 03;22(6):e17152 [FREE Full text] [doi: 10.2196/17152] [Medline: 32490836]
78. Desmet M, Franssen S, Varol T, Fillon A, Thivel D, Roefs A, et al. A smartphone application as a personalized treatment tool for adolescents with overweight: an explorative qualitative study. BMC Public Health 2023 Apr 14;23(1):696 [FREE Full text] [doi: 10.1186/s12889-023-15248-z] [Medline: 37060051]
79. Egilsson E, Bjarnason R, Njardvik U. Usage and weekly attrition in a smartphone-based health behavior intervention for adolescents: pilot randomized controlled trial. JMIR Form Res 2021 Feb 17;5(2):e21432 [FREE Full text] [doi: 10.2196/21432] [Medline: 33481750]

Abbreviations

BCT: behavior change technique
CRF: cardiorespiratory fitness
FMS: fundamental movement skills
mHealth: mobile health
MVPA: moderate to vigorous physical activity
PA: physical activity
PF: physical fitness
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
SB: sedentary behavior
SCT: social cognitive theory
SDT: self-determination theory
SMD: standardized mean difference
SRT: self-regulation theory
TPA: total physical activity
WC: waist circumference
WMD: weighted mean difference

Edited by L Buis; submitted 01.08.23; peer-reviewed by M Hammersley, H Baumann, O Byambasuren, M Zhou; comments to author 13.09.23; revised version received 11.02.24; accepted 14.03.24; published 30.04.24.

Please cite as:
Wang JW, Zhu Z, Shuling Z, Fan J, Jin Y, Gao ZL, Chen WD, Li X
Effectiveness of mHealth App–Based Interventions for Increasing Physical Activity and Improving Physical Fitness in Children and Adolescents: Systematic Review and Meta-Analysis
JMIR Mhealth Uhealth 2024;12:e51478
URL: https://mhealth.jmir.org/2024/1/e51478
doi:10.2196/51478
PMID:38687568

Review

Motion Tracking of Daily Living and Physical Activities in Health Care: Systematic Review From Designers' Perspective

Lai Wei[1], PhD; Stephen Jia Wang[1], Prof Dr

School of Design, The Hong Kong Polytechnic University, Hung Hom, China (Hong Kong)

Corresponding Author:
Stephen Jia Wang, Prof Dr
School of Design
The Hong Kong Polytechnic University
11 Yuk Choi Rd
Hung Hom
China (Hong Kong)
Phone: 852 2766 5478
Email: stephen.j.wang@polyu.edu.hk

Abstract

Background: Motion tracking technologies serve as crucial links between physical activities and health care insights, facilitating data acquisition essential for analyzing and intervening in physical activity. Yet, systematic methodologies for evaluating motion tracking data, especially concerning user activity recognition in health care applications, remain underreported.

Objective: This study aims to systematically review motion tracking in daily living and physical activities, emphasizing the critical interaction among devices, users, and environments from a design perspective, and to analyze the process involved in health care application research. It intends to delineate the design and application intricacies in health care contexts, focusing on enhancing motion tracking data's accuracy and applicability for health monitoring and intervention strategies.

Methods: Using a systematic review, this research scrutinized motion tracking data and their application in health care and wellness, examining studies from Scopus, Web of Science, EBSCO, and PubMed databases. The review used actor network theory and data-enabled design to understand the complex interplay between humans, devices, and environments within these applications.

Results: Out of 1501 initially identified studies, 54 (3.66%) were included for in-depth analysis. These articles predominantly used accelerometer and gyroscope sensors (n=43, 80%) to monitor and analyze motion, demonstrating a strong preference for these technologies in capturing both dynamic and static activities. While incorporating portable devices (n=11, 20%) and multisensor configurations (n=16, 30%), the application of sensors across the body (n=15, 28%) and within physical spaces (n=17, 31%) highlights the diverse applications of motion tracking technologies in health care research. This diversity reflects the application's alignment with activity types ranging from daily movements to specialized scenarios. The results also reveal a diverse participant pool, including the general public, athletes, and specialized groups, with a focus on healthy individuals (n=31, 57%) and athletes (n=14, 26%). Despite this extensive application range, the focus primarily on laboratory-based studies (n=39, 72%) aimed at professional uses, such as precise activity identification and joint functionality assessment, emphasizes a significant challenge in translating findings from controlled environments to the dynamic conditions of everyday physical activities.

Conclusions: This study's comprehensive investigation of motion tracking technology in health care research reveals a significant gap between the methods used for data collection and their practical application in real-world scenarios. It proposes an innovative approach that includes designers in the research process, emphasizing the importance of incorporating data-enabled design framework. This ensures that motion data collection is aligned with the dynamic and varied nature of daily living and physical activities. Such integration is crucial for developing health applications that are accessible, intuitive, and tailored to meet diverse user needs. By leveraging a multidisciplinary approach that combines design, engineering, and health sciences, the research opens new pathways for enhancing the usability and effectiveness of health technologies.

(JMIR Mhealth Uhealth 2024;12:e46282) doi:10.2196/46282

KEYWORDS

motion tracking; daily living; physical activity; health care application; design; public health; systematic review; mobile phone

Introduction

Daily Living and Physical Activity in Health Care

Motion tracking data are pivotal for understanding physical activity in hospitalization and rehabilitation [1,2]. However, the procedures for investigating motion tracking data have not yet been systematically reported, especially in the context of user activity recognition in health care application studies. Several findings indicate that living patterns, such as daily activity trajectories, physical activity, and dietary behaviors, are associated with the initial stages of disease development [3-5]. Inadequate daily physical activity exacerbates symptoms of mental distress and contributes to medical illnesses, including cardiovascular and kidney diseases [6,7]. Conversely, regular physical activity can enhance wellness by activating the immune system and reducing inflammation [8,9]. Monitoring activities of daily living (ADLs) [10] is vital for self-care and long-term health [11,12], making the analysis of daily living and physical activity data crucial for disease prevention and promoting a healthy lifestyle.

A comprehensive analysis of motion tracking data in daily living and physical activities scenarios is essential for delving into human activity patterns and behavior intentions. Research shows that activity patterns are linked to behavior intentions [13,14], and motion data help uncover these patterns, improving diagnostic efficiency [15]. The motion data of ADLs represent activity patterns [16,17], while the formation of these patterns is multifaceted [18], with external influences playing a role [19]. Accumulated movement behaviors, when stimulated by the environment, become habitual actions [20,21]. By analyzing the motion data of ADLs, researchers have found that the formation of activity patterns is attributed to individuals' cognitive representations [22] and environmental factors, such as sedentary behaviors in working and learning spaces [10,23], as well as daily routines at home [24]. Thus, motion tracking data, enriched by environmental context, effectively represent and interpret human behaviors across different settings.

When applied as representations of ADLs, motion data can facilitate behavioral change through interventions. The analysis and visualization of motion data concerning ADLs can potentially enhance mutual understanding between physicians and patients in medical education and remote clinical consultations [25,26]. Patients' trajectories of past daily activity and medical histories are essential indicators for physicians to understand symptoms [27]. Human physical activity patterns can reflect the formation of disease and aid in the prevention and management of illness conditions [28,29]. In addition, many researchers have devised human activity interventions and suggested that early care physical activity interventions are feasible for promoting healthy lifestyles [30,31].

Motion Tracking Technology

The evolution of motion tracking technologies in health care has transitioned from basic motion detection systems aimed at rehabilitation to artificial intelligence (AI)–enhanced wearable sensors for personalized care. Early systems focused on simple motion detection, primarily for rehabilitation [1]. Advancements led to the inclusion of wearable sensors and AI for personalized health care [32]. Recent developments have embraced Internet of Things (IoT) frameworks for more comprehensive health monitoring [33]. In addition, the shift toward patient-centered care has been facilitated by technologies that assess and support daily activities and physical functions [11,34]. The integration of machine learning and sensor data fusion has further enhanced the ability to monitor and analyze patient movement in real time, contributing to preventive health care and improved treatment outcomes [12,35]. Researchers in health care and computer science have discussed the possibility of using motion tracking. Data classification, precision, validity, and diagnostic prediction in activity recognition are popular areas of investigation [36,37]. Several studies have also discussed how collecting voluminous data might contribute to personalized health surveillance systems [32,33]. Furthermore, scholars in computing have claimed that high-quality data would enhance data-driven AI processing [38,39]. IoT, with motion tracking techniques (eg, inertial measurement unit, smartphone, smartwatch, Mocap system, etc) [40,41] and AI processing (eg, human activity recognition, positioning system, etc) [35,42], have been widely used. It is anticipated that AI processing using high-quality motion tracking data would provide accurate and timely health information. For instance, smartphones with built-in motion-tracking sensors can record activity trajectories and routine data to identify activity types and patterns [43]. Studies have also found that motion data can be used to learn about humans' typical behaviors and identify any outlier activities [44,45]. Moreover, it can boost individuals' likelihood of engaging in physical activity over the long term, which is associated with habit formation [46,47].

However, the data interpretation of human activity trajectories, environment, and process has not been discussed as extensively as technology. Conversely, several studies have reported that learning from health information could promote health literacy [48,49]. Yet, the public has a long-standing misunderstanding of technology [50,51]. The technology-based health information needs to be in readable language [52,53]. In the meantime, the foundational criteria (ie, ease of use, usefulness, and enjoyment) of the technology acceptance model (TAM) provide insights to evaluate human-technology interaction [54,55]. TAM exemplifies the 3 needs of humans in interpreting technology-based health information.

The Usability of Motion Tracking Data

Designers play a key role in making data legible, efficiently transmitted, and actively engaging for users [56-58]. To optimize the utility and interpretability of motion tracking for both professionals and users, designers should gather health information needs from the user's cognitive standpoint, thereby refining the functionality and analytical clarity of motion-tracking systems [59,60]. The adoption of data-enabled design (DED) [61-63] in health care emphasizes continuous data refinement and context-specific design. On the basis of these, health care–related applications, as intelligent solutions [62], should integrate data, users, and environments to optimize the transmission of health information and the intelligent ecosystems. Scholars asserted that design studies could illuminate the nature of nonsocial entities, such as data, to create durable systems [64,65]. This approach has been implemented

in designing health interventions [66,67]. It overcomes the barrier between people and technology, offering guidance for enhancing the sustainable hybridization of the social and the technical entities.

Motion tracking serves as a bridge, delivering activity data to health care professionals and users. Understanding data collection and analysis is vital, involving motion data (eg, trajectories and durations) and environmental elements in physical settings. Analysis entails identifying and classifying activities. Yet, interactions among researchers, participants, devices, and environments are seldom explored. Therefore, it is critical to methodically study these interactions within health care applications and their environments.

Integrating Actor Network Theory in Research

In health care application research on ADLs, interactions occur between participants, sensors, cameras, and scenario objects. The actor network theory (ANT) has been instrumental in analyzing these interactions within a network [68,69], especially in human-computer interaction and design studies [70]. ANT outlines three critical aspects for network analysis: (1) actor-led activities, where initiators, such as individuals or devices, lead activities and interact with others; (2) purpose-oriented interactions, which serve as the guiding principle for activities; and (3) the dynamic interplay between devices, environmental factors, and individuals, shaping the network. Integrating ANT into motion tracking data analysis illuminates the movement trajectories and environmental influences, clarifying human-device-environment relationships.

Objectives

Prior research has thoroughly examined motion tracking in health care, mainly focusing on algorithm validation and technology deployment for rehabilitation and hospitalization [42,71]. Meanwhile, several studies have reviewed the benefits of physical activity for medical purposes [10,26]. However, systematic reviews exploring the use of motion tracking techniques in ADLs and their connection with humans for health care and physical wellness research are scarce. Furthermore, studies investigating the process of analyzing ADLs for health care applications are even rarer. This study shifts focus to motion data, participant engagement, and situational contexts as key components in health care technology. It delves into how participants interact with technology in physical environments, technology adoption, and participant behavior. This systematic review aimed to map and collate literature on the motion tracking data of daily living and physical activities for health care and physical wellness application research. It aimed to evaluate the research landscape, identify literature gaps, and endorse a designer-engaged approach to studying ADLs. The overarching research questions are as follows:

1. How was the motion tracking technique used in ADLs for health care–related studies and physical wellness applications?
2. What are the environmental factors, interactions, and processes of ADLs in health care and physical wellness application research?
3. What is the design opportunity in health care and physical wellness application research?

Methods

Overview

Our research focuses on investigating the motion tracking data, interaction, and process of ADLs in health care and physical wellness applications from the designers' perspective. Owing to the exploratory and descriptive nature of our research question, we opted for a systematic review [72,73], with the aim of compiling and comprehensively summarizing the relevant research. To guarantee rigor and coherence with our research objectives, 2 researchers meticulously screened the studies for inclusion. These 2 researchers independently conducted screening work on the same data set and met weekly for discussions. Following 5 detailed discussions about the screening methodology, centered on eligibility criteria and research focus, we conducted 2 pilot searches to refine our search strategy and ensure the accuracy of our results. In addition, we discussed the framing of questions being addressed with reference to participants, interventions, comparisons, and outcomes (PICO) [74], ensuring a comprehensive and structured approach to our review (Table 1). The overarching goal of the systematic review is to identify and map the available evidence investigating the technology use in everyday health care and physical wellness apps and its relationship with humans and environments. We then applied ANT [68,69] to analyze the interaction in activity scenarios and used design narrative through the lens of DED [62] to analyze the process of motion tracking data in health care and physical wellness application research scenarios.

Given the independent nature of technology, user behaviors, and environments in our research, we developed bespoke criteria for bias assessment. This approach enabled 2 independent reviewers to identify potential biases in study design, execution, and reporting accurately. Discrepancies between reviewers were resolved through consensus, ensuring a balanced evaluation. To further validate our review, we engaged relevant scholars to critique the search strategies of both pilot and final searches stages, substantially enhancing the review's validity and comprehensiveness. Our thorough assessments of potential biases, both at the study and outcome levels, were integrated into our data synthesis, paving the way for a credible interpretation of the evidence (detailed in Multimedia Appendix 1 [75-128]).

Table 1. Search term– and strategy-based participants, interventions, comparisons, and outcomes (PICO) framework and research questions.

Component	Details	Search terms	Search strategies
Population	• General public, including special populations such as children, pregnant women, older adults, people with disabilities, and athletes	N/A[a]	N/A
Intervention	• Use of motion tracking techniques in ADLs[b] and physical activity scenarios	Motion tracking	TITLE-ABS-KEY (“motion capture” OR “MoCap” OR “motion analysis” OR “motion tracking” OR “body positioning” OR “human activity recognition” OR “IMU” OR “dynamometry”)
	• Application of AI[c] processing for activity classification and analysis; or using devices (including sensors, motion capture technology, IoT[d] devices, VR[e], AR[f], and MR[g])	IoT, AI, VR, and AR	TITLE-ABS-KEY (“Internet-of-things” OR “IoT” OR “wearable” OR “virtual reality” OR “augmented reality” OR “mixed reality” OR “machine learning” OR “deep learning” OR “decision making” OR “artificial intelligence” OR “AI”)
Comparator	• Variation in technology use, sensor placement, and participant interaction across different studies	N/A	N/A
Outcome	• Identification and classification of daily living and physical activities within health care and physical wellness applications • Understanding of environmental factors, interactions, and processes in health care and physical wellness applications • Exploration of design opportunities in health care and physical wellness applications	Health and sport application	TITLE-ABS-KEY ([“sport” OR “kinematics” OR “sport analytics” OR “wellness” OR “health”] AND [“smart health” OR “healthcare” OR “health monitoring” OR “health of things” OR “digital health” OR “mobile health system” OR “behaviour change” OR “decision making”])

[a]N/A: not applicable.

[b]ADL: activity of daily living.

[c]AI: artificial intelligence.

[d]IoT: Internet of Things.

[e]VR: virtual reality.

[f]AR: augmented reality.

[g]MR: mixed reality.

Selection Criteria

Study Types

Original research articles published in scientific, technical, and medical journals in English from January 2013 to December 2022 were considered. Reviews, conference abstracts, magazines, and newspaper articles were excluded. We concentrated exclusively on studies encompassing ADLs, specifically within the realms of health care and physical wellness application research, to ensure the relevance of motion tracking technology to real-world health and wellness contexts. Eligibility for inclusion was determined for studies using motion-tracking sensors, motion capture technology, IoT devices, multiple sensors capable of motion tracking, virtual reality (VR) or augmented (mixed) reality, or AI, highlighting our focus on advanced technologies that offer innovative approaches to monitoring and enhancing health-related activities. Eligible populations include the general public as well as specific demographics including children, pregnant women, older adults, individuals with disabilities, and athletes. Exclusions were applied to studies that (1) focused on technology or materials development (ie, technical validation), as our interest was in direct applications of technology in health and wellness, rather than preliminary stages of technological development; for instance, one research focused on preliminary technology validation without applying findings to enhance health-related activities, missing our application-focused criteria, although it mentioned motion tracking, smart systems, inertial measurement unit, and ADLs [129]; (2) presented a data set without further analysis, as our aim was to understand the implications of data on ADLs, necessitating detailed data interpretation; (3) investigated activities in clinical scenarios such as injury, impairments, hospitalization, rehabilitation, etc, because our focus was on everyday activities rather than those strictly within clinical settings; (4) were applied to nonhuman subjects to maintain the applicability of findings to human health and wellness; or (5) did not include any user study results or did not clearly explain their findings, as comprehensible and applicable user data are crucial for informing practical health care and wellness interventions (Textbox 1).

Textbox 1. Eligibility criteria for considering the studies in the review.

Inclusion criteria

- Study types
  - Original research articles published in scientific, technical, and medical journals in English from January 2013 to December 2022
  - Studies that encompass activities of daily living (ADLs) or physical activity within the scope of health care and physical wellness application research
  - Studies using motion-tracking sensors, motion capture technology, internet of things (IoT) devices, multiple sensors capable of motion tracking, virtual reality (VR) or augmented (mixed) reality (A[M]R), or artificial intelligence (AI)
- Population
  - General public, including special populations such as children, pregnant women, older adults, people with disabilities, and athletes
- Materials
  - Research investigating the use of motion tracking for ADLs analysis
  - Studies using motion-tracking sensors, motion capture technology, or multiple sensors for motion monitoring, analysis, visualization, or providing feedback
  - Research integrating a combination of AI processing, IoT devices, or VR (A[M]R) technologies
  - Studies integrating motion tracking data and AI technology for the classification of ADLs
- Comparison
  - Variation in technology use, sensor placement, and participant interaction across different studies
- Research outcomes
  - Measures or indexes describing the activity in health care or physical wellness application research (eg, daily human activities, physical activities, or daily activities of special populations)
  - Studies focusing on the use of motion tracking for monitoring, analyzing, visualizing, or providing feedback
  - Research related to health care or physical wellness system design

Exclusion criteria

- Study types
  - Reviews, conference abstracts, magazines, and newspaper articles
  - Studies focused on technology or materials development (technical validation)
  - Research presenting a data set without further analysis
  - Investigations in clinical scenarios such as injury, impairments, hospitalization, rehabilitation, etc
  - Studies applied to nonhuman subjects
  - Studies that did not include any user study results or did not clearly explain their findings
- Population
  - Investigations in clinical scenarios such as injury, impairments, hospitalization, rehabilitation, etc
  - Studies applied to nonhuman subjects
- Materials
  - Studies not using the described technologies or not focused on the specified applications within health care and physical wellness
- Research outcomes
  - Outcomes not related to the activity in health care or physical wellness application research, or those not providing meaningful insights into monitoring, analysis, visualization, or feedback within these contexts

Materials

Health care or physical wellness application research investigated the use of (1) motion tracking for the analysis of ADLs; (2) motion-tracking sensors, motion capture technology, or multiple sensors for motion monitoring, analysis, visualization, or providing feedback in ADL scenarios; (3) a combination of AI processing, IoT devices, or VR or augmented (mixed) reality technologies; or (4) integrating motion tracking data and AI technology for the classification of ADLs.

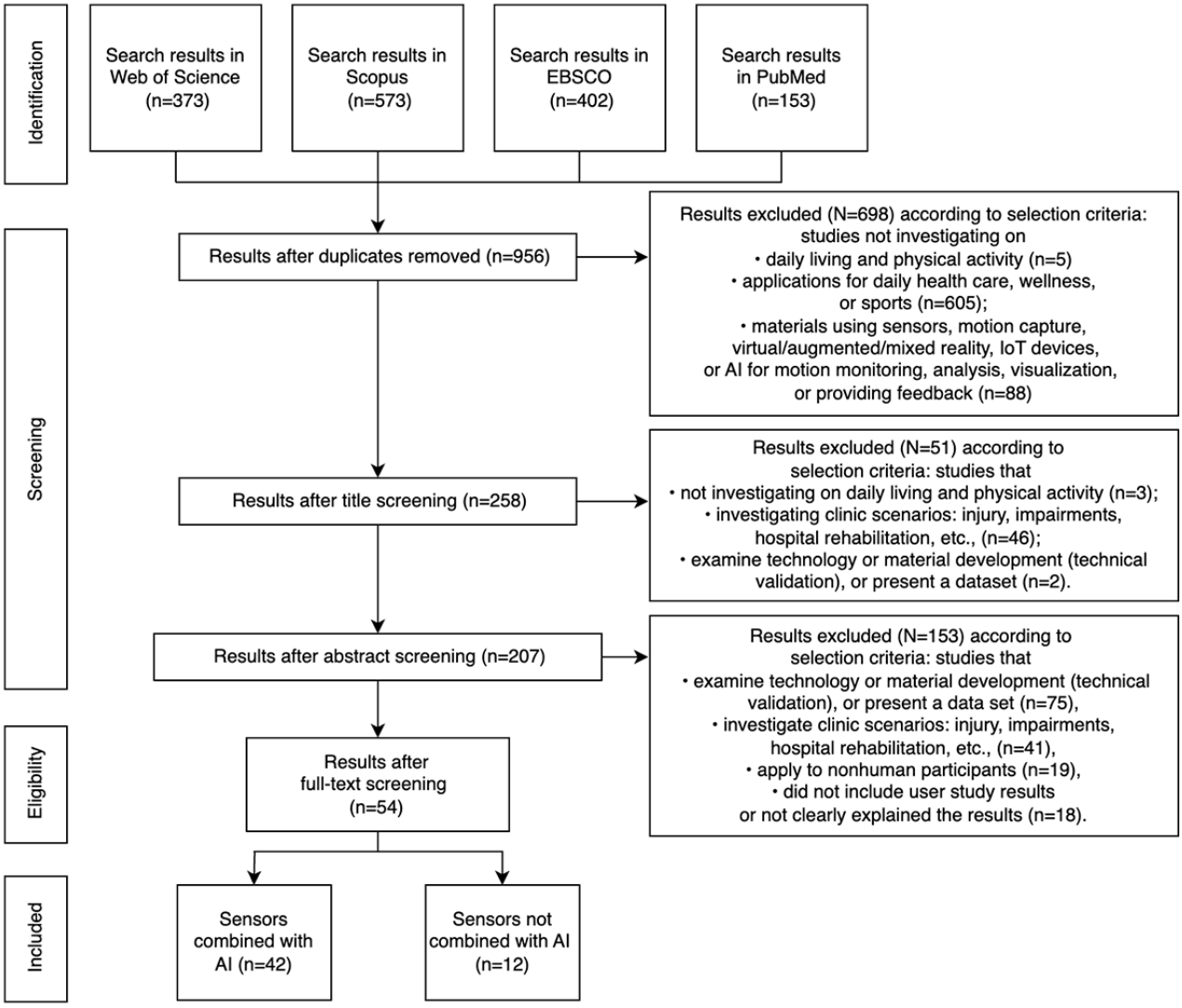
Research Outcomes

Any measure or index that described (1) the activity in health care or physical wellness application research (eg, daily human activities, physical activities, or daily activities of special populations); (2) the use of motion tracking for monitoring, analyzing, visualizing, or providing feedback; and (3) health care– or physical wellness–related system design was included.

Search Strategy

Two pilot searches were conducted using the Scopus and Web of Science electronic databases in August 2022, applying search strategies detailed in Multimedia Appendix 2. The first search yielded an exceedingly limited number of articles, while the second search displayed a broader range. Upon reviewing the titles and abstracts from the pilot search results, we observed that the terms “physical activity” and “daily living activity” might lead to irrelevant research fields, such as heart rate monitoring and step count. In contrast, “motion tracking” and “motion capture” more accurately capture the essence of physical activity within our research context. These terms are commonly used in the field of motion capture technology research. Consequently, we gave these terms precedence in our investigation, which assisted in the strategic identification of pertinent keywords for inclusion. Moreover, from the results of the initial 2 pilot searches, we identified additional keywords related to health care applications, AI, and motion tracking, such as “health of things” and “dynamometry.” Following discussions, we selected these terms for use in our final search queries (Table 1). The search strategy was subsequently refined and executed in September 2022, covering the following 4 databases: Web of Science, Scopus, EBSCO, and PubMed (Figure 1). The titles, abstracts, and index terms were screened to identify the studies that met our stated eligibility criteria as outlined in Textbox 1.

Figure 1. The flowchart of the selection process of articles. AI: artificial intelligence; IoT: Internet of Things.

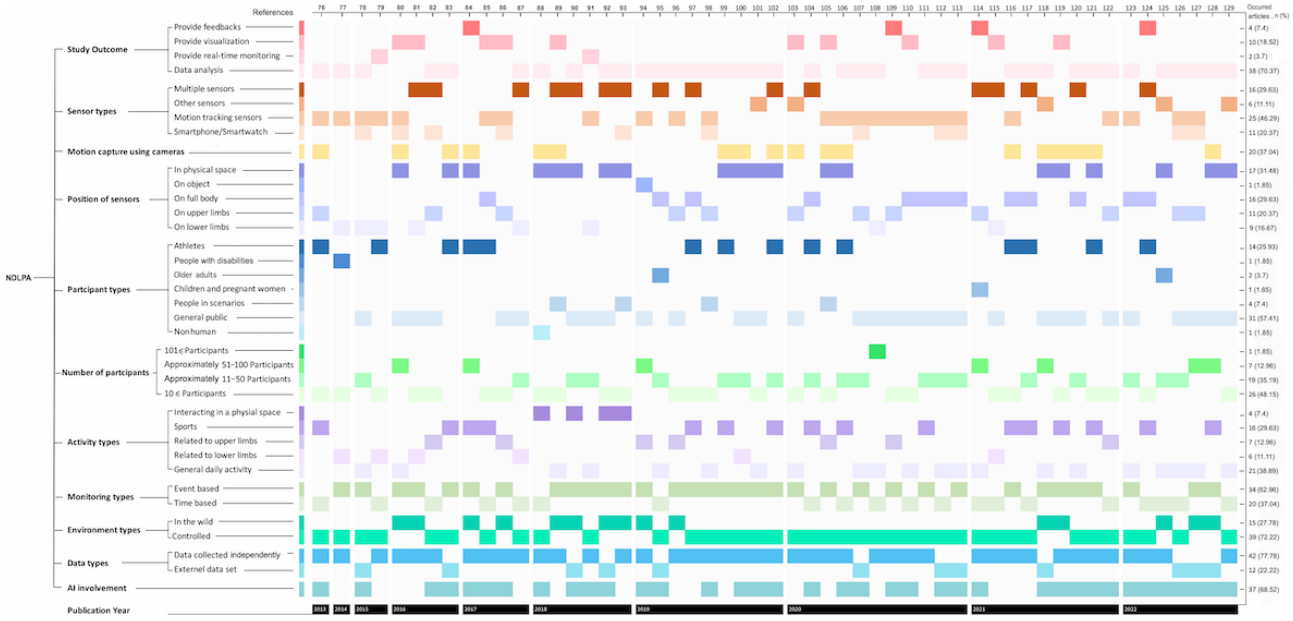


Data Extraction

Drawing from previous research in health care and motion tracking, our study builds upon the insights provided by Straczkiewicz et al [42], who illustrated the relationships among different human activity recognition processing phases through data visualizations. Their approach to data visualization reveals the connections among co-occurring factors identified across the reviewed articles over time. In addition, we also considered the work of van Kollenburg et al [62], who explored design narratives for situated design exploration, thereby enriching our understanding of the data analysis process and its application in health care scenarios [62].

Through the data extraction process from the selected articles, we organized the data into 11 categories: (1) study outcome (subcategories encompassing providing feedback, visualization, real-time monitoring, and data analysis); (2) sensor types (subcategories include multiple sensors, other sensors, motion-tracking sensors, and smartphone or smartwatch); (3) motion capture using cameras; (4) position of sensors (eg, physical space, on object, full body, upper limbs, or lower limbs); (5) participant types (individuals on whom sensors were placed or who were being monitored, including athletes, people with disabilities, older adults, children, pregnant women, people in scenarios, general public, or nonhuman subjects); (6) number of participants (categorized as >101, 51-100, 10-50, or <10); (7) activity types (covering interactions in physical spaces, sports, activities related to upper or lower limbs, and general daily activities); (8) monitoring types (differentiated by event or time-based monitoring); (9) environment types (distinguished as in the wild or controlled); (10) data types (differentiating between data collected independently or from external data sets); and (11) Al involvement. Then, using the timeline and the categories of the included studies, we provided a publication year–based summary that illustrates the Network of Daily Living and Physical Activity (NDLPA) among included studies, as depicted in Figure 2.

Figure 2. The Network of Daily Living and Physical Activity (NDLPA) among included studies. AI: artificial intelligence. For a higher-resolution version of this figure, see Multimedia Appendix 3.



Results

Identification of the Studies

The electronic search yielded 1501 studies. After removing internal duplicates, 956 studies remained. On the basis of the eligibility criteria, 258 articles were chosen through title screening (Figure 1). A total of 54 articles were included for full-text appraisal after the abstract screening process [75-128]. The results from database search indicated that the relevant research from 2013 to 2018 grew steadily, while research articles published from 2019 to 2022, aligning with our interests, experienced a significant uptick and consistently maintained a high volume (Figure 3; Table 2). The findings (Figure 2) revealed that 54 articles used motion tracking data in health care and physical wellness applications, of which 42 (78%) articles used motion-tracking sensors [75-81,84-86,88-97,101,103-116,119,121-123,125,126]. Overall, 70% (38/54) of the articles used AI technology combined with motion tracking data [75,77,79,82-85,87,89-91,93,94,97,99-101,103,105,106,108,109,111-113,117-128]. The role of AI mainly functions for data classification, monitoring, and visualization among the included studies. Furthermore, 80% (43/54) of the studies recruited participants for user study [75,76,78-81,83-88,90,92,93,95-105,107-110,113-116,118-124,127,128]. Twelve articles used public data set [78,82,89,91,94,106,111,112,117,125-127]. Most of the public data set that were used are University of California, Irvine human activity recognition, center of advanced studies in adaptive system, mobile health, University of Milano Bicocca smartphone-based human activity recognition, mobile sensor data anonymization, a large multipurpose human motion and video data set, and wireless sensor data mining data sets.

Figure 3. The distribution of total selection, abstract screening, and included articles based on publication year.

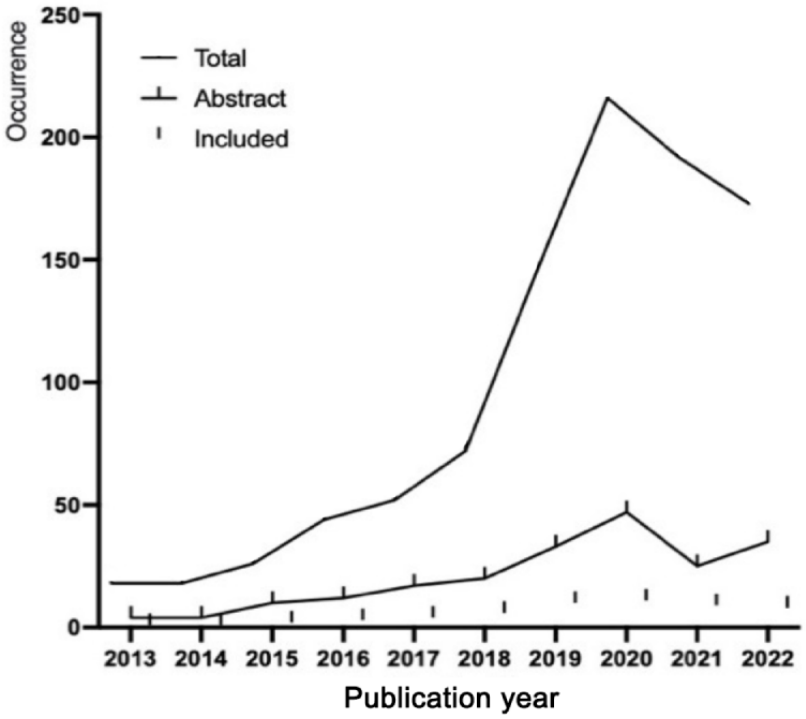


Table 2. The distribution of total selection, abstract screening, and included articles.

Publication year	Total (n=956), n (%)	Abstract screening (n=207), n (%)	Included articles (n=54), n (%)
2013	18 (1.9)	4 (1.9)	1 (1.8)
2014	18 (1.9)	4 (1.9)	1 (1.8)
2015	26 (2.7)	10 (4.8)	2 (3.6)
2016	44 (4.6)	12 (5.8)	4 (7.3)
2017	52 (5.4)	17 (8.2)	4 (7.3)
2018	72 (7.5)	20 (9.7)	6 (10.9)
2019	145 (15.2)	33 (15.9)	9 (16.4)
2020	216 (22.6)	47 (22.7)	11 (20)
2021	192 (20.1)	25 (12.1)	9 (16.4)
2022	173 (18.1)	35 (16.9)	7 (13)

Descriptions of Technology Use, Participants, Activity, and Outcomes

Overview

The contributing studies displayed a broad spectrum of methodologies, participant types, and technology uses. Characteristics of activity varied from studies focusing on lower limb activities to those analyzing full-body movements, using a range of sensors and AI technologies. Most studies used a combination of accelerometers, gyroscopes, and sometimes cameras across diverse settings from controlled environments to in-the-wild scenarios. Participant types ranged from the general public, including specific groups such as athletes and older adults, to nonhuman participants in a few instances.

Technology Use

Most of the included publications (42/54, 78%) used technology based on AI combined with motion tracking techniques or multiple sensors. Among those using sensors, studies incorporating accelerometer and gyroscope sensors formed the core of the investigations (42/54, 78%), including those using a smartphone or smartwatch (11/54, 20%) and those using multiple sensors capable of motion tracking (16/54, 30%). Fewer studies used other types of sensors (eg, temperature sensors, light sensors, etc; 6/54, 11%) or combined them with motion capture using cameras (10/54, 19%). Moreover, 17% (9/54) of the studies exclusively used cameras for motion capture. Details are provided in Figure 2. A combination of accelerometer and gyroscope sensors has been deemed essential for tracking everyday human activity over the past decade. They were used to collect data on typical ADLs, such as standing, walking,

sitting, and jumping, while ambient sensors were used less frequently. Physical activities such as jumping [75,78,98], playing basketball [96,116], soccer [83], baseball [115], running [105], ballet dance [103] skiing [127], karate [84], and Taijiquan [123] were studied. Meanwhile, AI processing in the included articles functions for data extraction and classification to achieve daily living and physical activities recognition.

According to the studies using motion tracking data and classifiers, using a minimal pair of accelerometer and gyroscope sensors to acquire motion tracking data can achieve activity recognition with high precision [75,77,78,81,85,97,107,108,113,121]. Several scholars have suggested that the human activity recognition system has the potential to enhance the efficiency of health care applications using multiple pairs of accelerometer and gyroscope sensors [94,101]. Most studies obtained original motion tracking data of ADLs for evaluating health care systems, whereas fewer studies adopted public data sets for activity recognition applications.

Position of Sensors

The method of sensor placement among the included 54 studies was distributed relatively evenly (Figure 2). The sensor placements in most of the included articles were implemented on the full body (n=16, 30%) and in physical spaces (n=17, 31%), with relatively fewer studies focusing on the upper limb (n=11, 20%) and the lower limb (n=9, 17%). However, 1 (2%) article reported that the sensors were placed on an object. The findings suggest that the placement of sensors is inherently determined by the nature of the activities. ADLs were categorized into lower limb–based activities (eg, walking, standing, sitting, etc); upper limb activities (eg, eating, talking on the phone, washing dishes, cooking, etc); and full-body activities (eg, playing basketball, dancing, skiing, etc). Therefore, sensor placements are specifically tailored to match the inherent types of activities being tracked.

The studies that involved placing sensors on the full body and in environments were widely reported [84,94,96,103,109-112,115,116,119,122,123], and the purposes of using full-body motion tracking are commonly argued for analyzing specific movements or securing the credibility of system evaluations [94,106,110,112,115,116,123]. The investigations involving the placement of sensors in physical spaces mostly use motion capture and image processing techniques [117,119]. Researchers have placed high-speed cameras to capture movements during sports activities [75,82-84,88,98,101,105,115,120,127] and in specialized work situations, such as maritime occupations [104]. One study [93] placed accelerometer and gyroscope sensors on an object, specifically a water container, to determine the frequency and amount of use from the container's motion tracking data. The positioning of sensors has become increasingly precise and intricate over time. Research has primarily targeted the upper limb, full body, or physical spaces.

Participants

The research exploring motion tracking of daily living and physical activities for health care applications has studied a variety of participant types (Figure 2), including the general public (healthy individuals; 31/54, 57%), people in specialized scenarios (4/54, 7%), children and pregnant women (1/54, 2%), older adults (2/54, 4%), people with disabilities (1/54, 2%), and athletes (14/54, 26%) over the past decade. One exception involved using a male cadaver for the analysis of vehicle-cadaver tests. From the results, researchers primarily applied motion capture technology to healthy individuals aged between 18 and 60 years and to athletes. Activities involving healthy individuals can provide representative activity characteristics for training activity recognition models and identifying outliers in activities through activity classification. In the last 3 years, researchers have increasingly prioritized the use of data collected from the public and youths for sports education–related research.

The number of participants recruited varied significantly across the included studies. Approximately half of the selected studies collected data from <10 participants (26/54, 48%), and a smaller number gathered data from 10 to 50 participants (19/54, 35%). Moreover, 13% (7/54) of the studies collected data from 50 to 100 individuals. Half (27/54, 50%) of the studies used 10 to 100 samples from public databases [77,89,94,106,111,112,117,125-127]. Only one study [107] amassed gait data from >1000 healthy individuals. The quantity of the sample size correlates with the types of participants and activities. In sports-related activities, the sample size ranged from 1 to 50 individuals, predominantly involving athletes. For studies on daily living activities, the sample size varied from 10 to 1000s of individuals, primarily focusing on the healthy general public.

Activity

Prior studies have classified activities according to the intensity of exercise [130]. In this study, the activity classification extends the previous categorization and is divided into general daily activity (21/54, 39%; eg, walking, standing, sitting, etc); activities related to the lower limb (6/54, 11%; eg, running, jumping, etc); activities related to the upper limb (7/54, 13%; eg, drinking, writing, cutting, etc); sports (16/54, 30%; eg, playing basketball or baseball, dancing, Taijiquan, etc); and interaction in a physical space (4/54, 7%; eg, living in a laboratory setting house, farming, etc) based on the sensor placements and the purposes of the activities (Figure 2). Most studies (39/54, 72%) specified that they were conducted in laboratory experiment settings and aimed to minimize the intervention of irrelevant environmental factors [75-78,82,84,86,87,90,94,96-116,119-123,125,128]. Each data collection session involved 1 participant, who was pretrained for a few minutes before their data were collected for experimental purposes. Researchers aimed to focus participants' attention on specific movements and perform standardized activities. The results suggest that collecting pretrained ADLs data provides standardized motion data for training ADLs recognition, but this approach might not capture the variety of behaviors necessary to accurately judge specific motion data. Meanwhile, the results also indicate that researchers concentrated on laboratory-based activities for professional use, such as the classification and identification of ADLs and the functionality of joints.

Study Outcome

The included studies (n=54) were all aimed at the identification and classification of activity trajectories (Figure 2). The findings indicate that most included research (n=39, 72%) focused on data analysis. Overall, 2 (4%) studies provided real-time monitoring; 3 (6%) studies offered feedback; and 10 (19%) studies were dedicated to visualization, including 6 studies that visualized original data without AI processing. The purpose of visualization is to provide professionals [88,114] or users [104,118] with the ability to view motion statuses. The visualization of motion data without classification is used in medical research [79,80,88,102,104,114], and the data are intended for use by health care professionals. Excluding the studies using animation and VR techniques, such as the one by Zelck et al [104], who proposed the animation of digital maritime workers' working processes, and another study by Ahmed and Demirel [102], who demonstrated human performance in normal situations and aircraft accidents using VR, and the data from the latter are accessible to the public. Generally, studies focusing on ADLs primarily used data from the healthy public for professional use.

Risk of Bias and Quality Assessment

In conducting a thorough risk of bias and quality assessment of the selected studies, our scrutiny was directed toward several critical dimensions: participant selection transparency, the accuracy and completeness of data acquisition methodologies, the analytical rigor, and the integrity in reporting outcomes. Notably, prevalent biases included selection bias, rooted in participant recruitment strategies; measurement bias, owing to sensor placement inconsistencies or activity categorization; and reporting bias. Studies using public data sets or conducted within laboratory confines typically showed reduced bias risks. In contrast, studies with constrained participant demographics or undertaken in less regulated environments faced heightened bias risks, primarily affecting data representativeness and the influence of external variables on activity assessments.

The included articles variably reported on participants' gender and age specifics—57% (31/54) of the articles provided detailed demographic data [77-79,82,86,89-91,93,94,97-99,101,103,105-107,109-113,115,116,119,122,124-127], while others (22/54, 41%) omitted such specifics [75,76,80,81,83-85,87,88,92,95,96,100,102,104,108,114,117,118,120,121,123,128]. The sensor models used across studies are meticulously documented within the articles and are detailed in the Multimedia Appendix 1. It was observed that the spectrum of daily living activities discussed was somewhat narrow, seldom addressing activities that entail object interaction. The theoretical underpinning, such as that provided by ANT [68,69] and activity classification [130], lends robust support to the motion data analysis, facilitating a nuanced understanding of the data. The selected articles did not address the incorporation of verbal language or cultural customs in the context of data collection or the execution of physical activities. Thus, we consider that the concentration on physical movement to the exclusion of linguistic or cultural contexts likely accounts for the omission of cultural background in the analysis. Data collection in laboratory settings is controlled, thereby eliminating interference from factors such as socioeconomic elements. The physical tasks assigned to participants encompass routine daily movements or activities well-acquainted with athletes, effectively eliminating potential interferences from policy-related factors.

Discussion

Interaction Between Devices, Users, and Environmental Factors

The features of the selected studies were systematically categorized into study outcome, sensor configuration, participant types, activity types, monitoring, environment types, and AI involvement, as illustrated in NDLPA among included studies (Figure 2). This organization was informed by the visualization of Straczkiewicz et al [42], who elucidated the complex interrelations among diverse human activity recognition process phases. Drawing upon the theoretical foundation provided by ANT [64,68,69], our analysis delved into the NDLPA by investigating the dynamics of actor-led activity; purpose-oriented interaction; and interaction among devices, individuals, and environmental factors.

Actor-Led Activity

Overview

Among the 54 articles selected for review, 3 main categories of initiators in health and physical wellness application research were identified: participants (including health information beneficiaries, professionals, caregivers, and interactional factors), researchers in health care, and researchers in computing. According to ANT [68,69], the interaction within the activity network is governed by those who lead the activity. As researchers are the initiators for implementing ADL research, we classified environment types by the extent of the researchers' intervention, namely, controlled and in the wild.

Controlled

In controlled settings, researchers initiated the activity. Participants conducted time- and environment-restricted daily activities following the researchers' activity assignments. Most articles mentioned that participants undergo pretraining before starting data collection to ensure that the data are typical [90,96]. Most athlete participants are in controlled environments. Researchers stated that motion data collection was controlled in specific movements, times, and spaces to guarantee optimal exercise time [81,121]. There is no other human-athlete interaction, except for 1 article using a virtual athlete [96]. This approach ensures the typicality of the data and the precision of the movement trajectories, which are essential for analyzing physical activity.

In the Wild

In uncontrolled or in-the-wild settings, participants are inclined to initiate activities, with interaction between participants and ambiance minimally restricted by researchers. Researchers configure sensors within the range of participants' activities (cameras, smart water meters, pressure mats, etc), allowing participants to produce movement trajectories relating to their intents. Findings include 2 types of in-the-wild activities: one

is video data sets captured in real-world situations, and the other is based chiefly on motion and multiple sensor data (eg, temperature and light use) to infer work and daily routines in living scenarios. The included studies in this review used data from in-the-wild activities, with 11 papers using data from the general population [79,80,85,88,89,91-93,95,118,126] and 1 article using data from older adults [124]. One study mentioned data collected from interactional factors, a water container [93] as the interactional initiator. Placing a motion-tracking device on the container that passively triggers humans' drinking movements may accurately detect the trajectory of upper limb movements.

Generally, most included investigations on everyday living and physical activities have been conducted in controlled settings, where researchers aim to capture accurate and standardized movements for health professionals and activity recognition processing. Although participants receive activity guidance and practice movements before the experiment, inconsistencies in activity objectives among them may arise. Alternatively, research using activity data from in-the-wild settings could provide a more intimate and real-life interaction but generate excess data. Moreover, it can achieve data controllability when the initiators are passive objects.

Purpose-Orientated Network

Overview

The activity initiator guides the purpose. As the research and the content of the activity were led by the researchers, the purpose of the research becomes the purpose of the activity in the network. We consider study outcomes, activity types, and monitoring types as elements forming the purpose of the activity networks.

Study Outcomes

The fields of human activity recognition and motion analysis have consistently been focal points of research, with an increasing volume of publications annually, particularly maintaining a stable high volume over the past 3 years (Figure 3). Studies have primarily focused on data visualization and monitoring, targeting daily activities and local body movements. Studies providing data feedback have addressed movement correction [123], outlier indication [113], and body posture correction [108]. It is noteworthy that outcomes related to feedback provision and data visualization have chiefly used in-the-wild data during the initial 6 years, shifting toward controlled data in the subsequent years, a transition likely influenced by advancements in research methodologies and technological evolution.

Activity Types

Most studies on activity trajectory recognition have focused on whole body movements, with local movements and movements related to environments being rare. Over the past 3 years, researchers have started to shift the target beneficiaries of their sports research to the general public, with physical activity education gradually overtaking professional sports. The dominant type of activities remains predominantly oriented toward daily living, with sports-related activities progressively shifted to ordinary activity.

Monitoring Types

Researchers described the nodes of data acquisition in terms of activity content and duration, and these nodes serve in subsequent data computation and analytical processes. They directed participants to undertake activities conforming to the predetermined criteria associated with these nodes. Moreover, research using in-the-wild data adapt their methodologies to the unpredictable nature of activity content and duration, ensuring a structured approach to capturing motion data despite the inherent variability. This nuanced guidance ensures that even in uncontrolled environments, the collection and analysis of data remain methodically aligned with the research objectives.

Interaction Among Devices, Individuals, and Environmental Factors

Activity trajectory is demonstrated through the interaction between humans, sensors, and environments. Motion-tracking devices, sensor positions, and environments factors are the main features to be evaluated.

Motion-Tracking Devices

Given this study's basis on motion tracking technology and the categorization of sensor types, we divided device types into complex devices (multiple sensors, eg, temperature, lighting, pressure accelerometer, and gyroscope sensors, VR, AR, etc) [91,92,101,113]; portable devices (smartphone or smartwatch) [77,79,81,92]; basic devices (motion-tracking sensors, eg, accelerometer and gyroscope sensors) [78,79,84,90,93]; and environmental devices (other sensors, eg, cameras, radiofrequency ID tags, etc) [100,102,117,124,128]. The configuration of complex devices aims at research related to local motions, complex motions, and environmental interactions. In conditions using complex devices, participants wore portable devices such as smartwatches to record motion data simultaneously with built-in multiple sensors (eg, ambient light sensor, UV light exposure sensor, and skin temperature sensor) [81]. A portable device with built-in accelerometer and gyroscope sensors performed a similar motion tracking task as a basic device. Most of them were used to determine everyday activities' trajectories as they are easy to perform. In conditions using portable and environmental devices, the portable device was placed on the body, and the participant activated sensors placed in the environment. The active and passive motion data form an activity network [92]. Accelerometer and gyroscope sensors were placed on multiple areas of participants' bodies to monitor their full-body motions. However, using cameras to capture athletes' movements and environmental devices (eg, pressure mats, mercury contacts, and float sensors) [91] to record living routine is favored by researchers.

Positions of Sensors

Our study categorizes device placements by areas of the human body and their relative distance from the human body, namely, in physical spaces, on objects, on the full body, on the upper limb, and on the lower limb. The results indicate that sensor placement is related to the study's purpose and the intensity of the activity. Sensors used in physical spaces or on the full body

examined high-intensity activities (eg, professional sports), correlating with accurate data acquisition. High-intensity activities, such as running [105] and throwing darts [110], were primarily studied using devices on local body areas. Devices placed on specific objects received limited daily exercise data but were consistent with the study's objectives. Sensors attached to different areas of the participants' bodies collected data from those areas and integrated with participants' movements. Sensors placed in the environment passively received signals, yet timing and activity patterns could be detected.

Environmental Factors

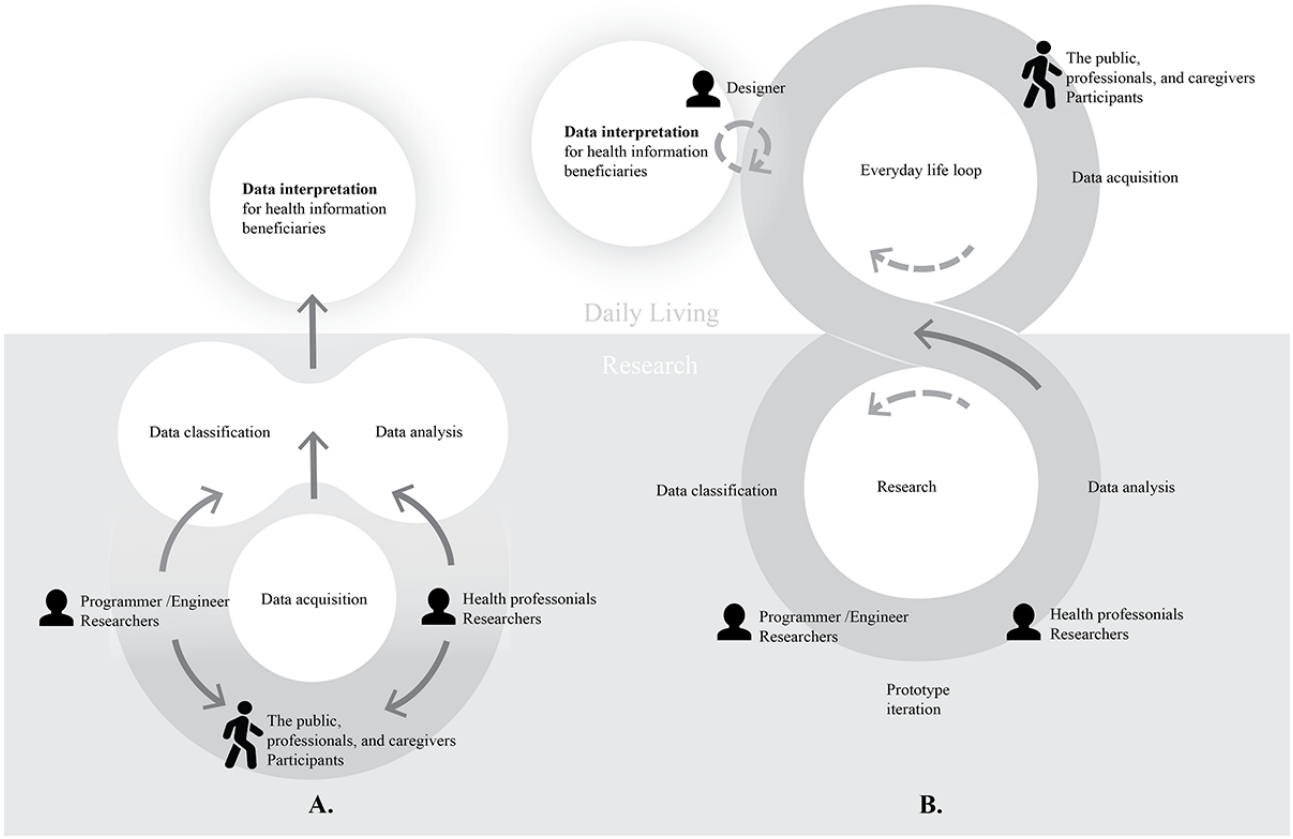
The environments for study implementation were mainly controlled by researchers. The movement trajectory of participants' performance aligned with the research objectives. The content and duration of activities were fixed, and devices were placed according to the research purposes. Participants passively interacted with their surroundings. Performing daily living or physical activities in an experimental setting might influence the participants' intent of the action. In these settings, participants' movements were unrelated to habitual actions. The environment and devices reflected interactions with researchers but did not interact with the participants, leading to a severed connection between the activity network and the participant involved. However, in video recording scenarios, participants performed activities in real-life settings, potentially providing purposeful and personalized movement data for research.

Investigation Process

Activity identification and recognition have been the primary focus of research on ADLs and physical activities in health and wellness applications over the past decade. The fundamental objectives are to enhance the accuracy of motion tracking data recognition and its practical use for health professionals. From the perspective of NDLPA, researchers acted as initiators of activity networks, leading the content and environments of activities in most activity recognition studies. Participants provided standard movement data within a restricted time and environment, lacking personal purpose and awareness of real-life activity. Participants did not fully become the initiators of the activity networks, but real-life activity data could be more complex. The variety of activity scenarios, user-friendly sensor design, data collection at specific time points, and data variances produced by complex activities and various real-life purposes could present research opportunities for studying daily living and physical activities for health and physical wellness applications. Meanwhile, designers have the opportunity to engage in designing user-initiated activities and collect technology-based motion data for health and physical wellness application research.

According to the descriptions in the included articles, the investigation of ADLs and physical activities in health care applications covers 4 phases: data acquisition, data classification, data analysis, and data interpretation, as shown in Figure 4A. In the process of motion tracking data acquisition, researchers recruited participants and conducted data collection. Participants performed assigned activities under full or partial surveillance by researchers. The data were supervised and categorized into standardized movements corresponding to specific activities. At this stage, the activities varied by different engaged areas on participants' bodies but not by the content of daily living activities. Researchers trained various algorithm models for data classification to achieve high activity recognition accuracy and used different data sets to test the system's validity. In the data analysis stage, researchers identified standardized activities and examined the difference values through motion tracking parameters. Using these parameters, researchers analyzed activity norms and behaviors. This stage is also associated with data monitoring. Data classification and analysis can be combined or conducted independently for different research evaluation purposes. Data interpretation, such as data visualization and feedback provision, was created by researchers from the health, engineering, and computing domains. It interacted with health information beneficiaries (eg, professionals, users, and caregivers). The health information was presented in either video or image forms for visualization or in text forms for feedback to users.

Figure 4. (A) The investigation process of motion data on daily living and physical activity for health care application and (B) a designer-involved investigation process of motion data on daily living and physical activity for health care application based on data-enabled design.



On the basis of the investigation process of ADLs and physical activities in health care applications, participants' involvement occurs at an early stage. However, participants are situated in a controlled environment for experimental purposes, with devices attached to their bodies without active interaction. Although the data are clear for classification and analysis, they rarely explain the diversity of ADLs. In the data interpretation stage, there is a weak connection between the data and the general public owing to the dominant purpose of activity identification research aimed at professional use. Designers can aid in bridging communication between information and the public. As described in the literature in previous sections, human motion data are essential components of activity recognition and analysis [1,2], with the public being the beneficiary of physical activity intervention research [11,12]. We propose that designers, coupled with researchers' guidelines, participate in the investigation and application process, which promises to facilitate data collection and interpretation.

Designer-Involved Investigation Process

The research of health care and physical wellness applications is directed toward professional deployment and public health enhancement. However, it has been observed that there exists a gap between the motion data collected from participants and their applicability to real-life situations. In addition, research in health care applications often fails to encompass the breadth of data diversity; there is a noticeable lack of meaningful engagement between devices and participants, and the process of making data understandable for the public is fraught with difficulties. To address these issues, we advocate for a research methodology that integrates DED framework [61-63], a designer-involved research process of motion data on ADLs and physical activity for health care application research, as illustrated in Figure 4B.

In this proposed methodology, designers collaborate with health care researchers to understand the specific information required for participant recruitment in a given project. They assist health care researchers in designing data collection strategies while considering the daily behavior patterns of the target population, thereby refining the data collection plan. Researchers from computing and engineering fields undertake the collection and categorization of data. These data span various demographics, purposes, devices, and scenarios from designers and are tailored for a multitude of applications. Such research process encompasses a cycle of prototype iterations. Researchers from the health field analyze classified data, identify abnormal motion data, and then provide guidelines for designers. Designers, positioned as intermediaries between users and researchers, leverage these guidelines and their nuanced understanding of human daily experiences including participant involvement and activity motivation [131] to reinterpret professional data into user-friendly formats. Furthermore, designers play a pivotal role in collecting motion data that reflect a diverse array of demographics, intentions, devices, and scenarios. They are able to simplify the data and assist with data interpretation for the recipients of health information.

The integration of design principles into the development of health care applications is key to enhancing user engagement, offering more intuitive interfaces that potentially increase adherence to health interventions through user-friendly designs. Our findings indicate that in intricate settings, standardizing data to mitigate external variables might lead to binary misjudgments. Given that daily activities are influenced by individual attributes, environments, and habits, the role of designers becomes critical in tailoring solutions to meet the varied needs of distinct user groups. They fine-tune the implementation methods and locations for motion-tracking devices, such as using cameras or accelerometer and gyroscope sensors and determining specific areas for deployment, such as full-body tracking or environmental interaction. They also specify the nature of the data to be collected, differentiating between event-driven data, which capture specific actions such as walking across varied material surfaces, and time-sequential data, which records information over continuous periods such as walking at night. This process enhances the accuracy and applicability of collected information, facilitating the development of data-driven products that reflect real user experiences. During the experiment, the dialogue between users and designers' aids in identifying actual and hidden needs, thus refining information precision and application. Upon analyzing the data, health care researchers collaborate with designers to reformat health information into accessible, engaging formats on digital interfaces, using techniques such as digital twins, animations, visualizations, and gamification. This method enhances user engagement, elicits valuable feedback, and fosters ongoing improvement in data accuracy, thus promoting a cycle of continuous information enhancement.

In addition, designers can assist engineers in designing user-friendly devices and facilitate actuator developments to provide users with timely feedback. The iterative refinement of technology that monitors ADLs and physical activity motion data in health application research is driven by the diversity of the population, necessitating continuous optimization. Design's contribution is pivotal in promoting effective communication, nurturing relationships, and pursuing sustainable innovation. By involving designers in the process, the development of activities initiated by users and the creation of user-friendly motion-tracking sensors are significantly improved. This, in turn, supports the cyclical development of health and physical wellness applications [64,65], ensuring that they are more aligned with real-world needs and contexts. Thus, in the sphere of health care and physical wellness application research, the involvement of designers is crucial for the effective acquisition and interpretation of data, ensuring that the data are relevant and beneficial in everyday scenarios.

Designers' involvement is essential in shaping design and implementation strategies, ensuring that technology solutions are tailored to meet the inherent human needs for interpreting health-related information. This approach is instrumental in improving the monitoring and management of health conditions, thereby improving patient outcomes. This not only improves usability but also ensures that technologies are perceived as useful and enjoyable, mirroring the foundational criteria of the TAM [54,55]. However, potential challenges include ensuring effective communication and collaboration across multidisciplinary teams, aligning different objectives and methodologies, and the complexity of translating complex health data into accessible and actionable information for users.

Limitations and Future Directions

This review process identified key limitations and future research directions in the use of motion tracking technologies for health care. It pinpointed risks of selection and measurement biases due to participant recruitment strategies and inconsistencies in sensor placement or activity categorization. Highlighting the inadequate exploration of feedback mechanisms, this review emphasized the necessity for future studies to optimize user communication. The use of sensors for motion data collection offers a method to preserve privacy, contrasting with cameras or recorded videos, which might inadvertently compromise it. One included study introduced a Kinect and smartwatch system, focusing on health care professionals' privacy and showcasing a strong commitment to privacy and ethical considerations in system design [79]. The minimal focus on privacy protection in other included studies indicates a potential research gap. Furthermore, this review underscored the unexplored potential of integrating geolocation data into physical activity interventions, proposing future research to delve into scenario-based health care applications. The investigation did not fully address the influences of cultural, socioeconomic, and policy-related factors on physical activity, or the budget constraints that may have curtailed the exploration of these dimensions. Moreover, it acknowledged the potential publication bias and the influence of study quality on the findings, especially given researchers' inclination to report positive outcomes of health care application development. Although the search included major databases, there is a possibility that relevant studies in other databases were overlooked. In addition, database search confined to English language might have excluded pertinent non-English studies. These areas present fertile ground for future research to ensure a more holistic understanding and application of motion tracking technologies in diverse settings.

Implications

This study significantly contributes to future research by demonstrating the importance of a multidisciplinary approach in health care application development, particularly in the integration of design principles with engineering and health sciences. It lays a foundation for further exploration into how designers can enhance the usability and effectiveness of health applications through user-centric design and data interpretation. Future studies can build on this framework to investigate specific design strategies and their impact on user engagement, adherence, and overall health outcomes, potentially leading to more personalized and effective health interventions. For real-world applications, this approach could fundamentally transform the development of health care apps, making them more accessible, intuitive, and tailored to individual needs, ultimately leading to improved health outcomes and patient engagement.

Conclusions

The study systematically reviewed the implementation of motion tracking in health care application research, revealing a discrepancy between the data collection and its effective application in real-world scenarios. Informed by ANT, we explore the dynamics of actor-led activities and purpose-oriented interactions, focusing on participants, sensors, cameras, and environmental factors in health care research concerning ADLs. The guiding roles of activity initiators significantly influence the phases of data acquisition, classification, analysis, and interpretation, underscoring the potential to enhance the accuracy of motion tracking and activity recognition when considering contextual factors. By advocating a designer-involved research process of motion data on ADLs and physical activity for health care application research, this study emphasizes the essential role of incorporating design principles via a DED framework in developing health care applications. This integration is crucial for aligning motion data collection with practical, real-life applications. By adopting a multidisciplinary approach and combining insights from design, engineering, and health sciences, the research demonstrates potential pathways for making health applications more user-friendly and effective. It underscores the necessity of engaging designers in the research process to ensure that health technologies are accessible, intuitive, and tailored to meet the diverse needs of users. Designers play a key role in customizing the deployment and data collection methods of motion-tracking devices, which improves the relevance and accuracy of collected data, leading to better-informed health applications. The iterative dialogue between users and designers during development refines the precision of information and its application, fostering a cycle of continuous improvement. By transforming complex health data into engaging, understandable formats, designers help translate user needs into actionable health solutions, promoting better monitoring and management of health conditions. The study advocates for further research to explore and refine these integrations, offering a direction that promises to transform health care monitoring and interventions, ultimately enhancing patient outcomes and engagement.

Acknowledgments

The authors would like to acknowledge the Hong Kong Polytechnic University for funding this research. This research was supported by the School of Design Collaborative Research Fund (P0035075), Strategic Importance Fund at the Hong Kong Polytechnic University (P0036851), and projects of Research Institute for Intelligent Wearable Systems (P0039255).

Authors' Contributions

LW developed the research framework, identified the research scopes, conducted data collection, screened articles, analyzed the data, and was responsible for writing and revising the manuscript. SJW conceptualized and designed the study, managed project administration, screened articles, provided supervision, contributed to reviewing and editing the draft, supplied resources, and acquired funding. Both authors reviewed and edited the manuscript and approved the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Summary of included studies based on participants, interventions, comparisons, and outcomes methodology.
[XLSX File (Microsoft Excel File), 48 KB - mhealth_v12i1e46282_app1.xlsx]

Multimedia Appendix 2
Search terms and strategies based participants, interventions, comparisons, and outcomes methodology and research questions (including 3 phases).
[XLSX File (Microsoft Excel File), 10 KB - mhealth_v12i1e46282_app2.xlsx]

Multimedia Appendix 3
The Network of Daily Living and Physical Activity (NDLPA) among included studies. AI: artificial intelligence.
[PNG File , 218 KB - mhealth_v12i1e46282_app3.png]

Multimedia Appendix 4
PRISMA checklist.
[PDF File (Adobe PDF File), 58 KB - mhealth_v12i1e46282_app4.pdf]

References

1. Zhou H, Hu H. Human motion tracking for rehabilitation—a survey. Biomed Signal Process Control 2008 Jan;3(1):1-18. [doi: 10.1016/j.bspc.2007.09.001]

2. Desmarais Y, Mottet D, Slangen P, Montesinos P. A review of 3D human pose estimation algorithms for markerless motion capture. Comput Vis Image Underst 2021 Nov;212:103275. [doi: 10.1016/j.cviu.2021.103275]
3. Marteau TM, Hollands GJ, Fletcher PC. Changing human behavior to prevent disease: the importance of targeting automatic processes. Science 2012 Sep 21;337(6101):1492-1495. [doi: 10.1126/science.1226918] [Medline: 22997327]
4. Andrieieva O, Kashuba V, Kovalova N, Khrypko I. Assessment of emotional state and mental activity of 15-16 year-old boys and girls who had a low level of physical activity. J Phys Educ Sport 2019;19(Supplement 3):1022-1029 [FREE Full text]
5. Cena H, Calder PC. Defining a healthy diet: evidence for the role of contemporary dietary patterns in health and disease. Nutrients 2020 Jan 27;12(2):334 [FREE Full text] [doi: 10.3390/nu12020334] [Medline: 32012681]
6. Cunningham C, O' Sullivan R, Caserotti P, Tully MA. Consequences of physical inactivity in older adults: a systematic review of reviews and meta-analyses. Scand J Med Sci Sports 2020 May;30(5):816-827. [doi: 10.1111/sms.13616] [Medline: 32020713]
7. Wilkinson TJ, Clarke AL, Nixon DG, Hull KL, Song Y, Burton JO, et al. Prevalence and correlates of physical activity across kidney disease stages: an observational multicentre study. Nephrol Dial Transplant 2021 Mar 29;36(4):641-649. [doi: 10.1093/ndt/gfz235] [Medline: 31725147]
8. Ford ES. Does exercise reduce inflammation? Physical activity and C-reactive protein among U.S. adults. Epidemiology 2002 Sep;13(5):561-568. [doi: 10.1097/00001648-200209000-00012] [Medline: 12192226]
9. Geffken DF, Cushman M, Burke GL, Polak JF, Sakkinen PA, Tracy RP. Association between physical activity and markers of inflammation in a healthy elderly population. Am J Epidemiol 2001 Feb 01;153(3):242-250. [doi: 10.1093/aje/153.3.242] [Medline: 11157411]
10. Castillo-Retamal M, Hinckson EA. Measuring physical activity and sedentary behaviour at work: a review. Work 2011;40(4):345-357. [doi: 10.3233/wor-2011-1246]
11. Edemekong PF, Bomgaars DL, Sukumaran S, Schoo C. Activities of daily living. StatPearls. Treasure Island, FL: StatPearls Publishing; 2022. URL: https://www.ncbi.nlm.nih.gov/books/NBK470404/ [accessed 2022-11-17]
12. Nweke HF, Teh YW, Mujtaba G, Al-garadi MA. Data fusion and multiple classifier systems for human activity detection and health monitoring: review and open research directions. Inf Fusion 2019 Mar;46:147-170. [doi: 10.1016/j.inffus.2018.06.002]
13. Nilsen P, Roback K, Broström A, Ellström PE. Creatures of habit: accounting for the role of habit in implementation research on clinical behaviour change. Implement Sci 2012 Jun 09;7(1):53 [FREE Full text] [doi: 10.1186/1748-5908-7-53] [Medline: 22682656]
14. Zacks JM, Kumar S, Abrams RA, Mehta R. Using movement and intentions to understand human activity. Cognition 2009 Aug;112(2):201-216. [doi: 10.1016/j.cognition.2009.03.007] [Medline: 19497569]
15. Bo F, Yerebakan M, Dai Y, Wang W, Li J, Hu B, et al. IMU-based monitoring for assistive diagnosis and management of IoHT: a review. Healthcare (Basel) 2022 Jun 28;10(7):1210 [FREE Full text] [doi: 10.3390/healthcare10071210] [Medline: 35885736]
16. Freedson PS, Miller K. Objective monitoring of physical activity using motion sensors and heart rate. Res Q Exerc Sport 2000 Jun 13;71 Suppl 2(sup2):21-29. [doi: 10.1080/02701367.2000.11082782] [Medline: 25680009]
17. Ferretti G, Zipoli Caiani S. Habitual actions, propositional knowledge, motor representations and intentionality. Topoi 2021 Jan 03;40(3):623-635. [doi: 10.1007/s11245-020-09723-0]
18. Smith KS, Graybiel AM. Habit formation. Dialogues Clin Neurosci 2022 Apr 01;18(1):33-43. [doi: 10.31887/dcns.2016.18.1/ksmith]
19. Booth SL, Sallis JF, Ritenbaugh C, Hill JO, Birch LL, Frank LD, et al. Environmental and societal factors affect food choice and physical activity: rationale, influences, and leverage points. Nutr Rev 2001 Mar;59(3 Pt 2):S21-S57. [doi: 10.1111/j.1753-4887.2001.tb06983.x] [Medline: 11330630]
20. Gardner B. A review and analysis of the use of 'habit' in understanding, predicting and influencing health-related behaviour. Health Psychol Rev 2015 Jan 21;9(3):277-295 [FREE Full text] [doi: 10.1080/17437199.2013.876238] [Medline: 25207647]
21. Hagger MS. Habit and physical activity: theoretical advances, practical implications, and agenda for future research. Psychol Sport Exerc 2019 May;42:118-129. [doi: 10.1016/j.psychsport.2018.12.007]
22. Hagger MS, Chatzisarantis NL, Biddle SJ. A meta-analytic review of the theories of reasoned action and planned behavior in physical activity: predictive validity and the contribution of additional variables. Int J Sport Exerc Psychol 2002;24:3-32. [doi: 10.1123/jsep.24.1.3]
23. Cook BG, Li D, Heinrich KM. Obesity, physical activity, and sedentary behavior of youth with learning disabilities and ADHD. J Learn Disabil 2015 Jan 21;48(6):563-576. [doi: 10.1177/0022219413518582] [Medline: 24449262]
24. Suryadevara NK, Mukhopadhyay SC. Determining wellness through an ambient assisted living environment. IEEE Intell Syst 2014 May;29(3):30-37. [doi: 10.1109/mis.2014.16]
25. Jones RB, Hubble S, Taylor L, Gunn H, Logan A, Rowland T, et al. Technologies to support assessment of movement during video consultations: exploratory study. JMIRx Med 2021 Sep 24;2(3):e30233 [FREE Full text] [doi: 10.2196/30233] [Medline: 37725550]

26. Geraedts H, Zijlstra A, Bulstra SK, Stevens M, Zijlstra W. Effects of remote feedback in home-based physical activity interventions for older adults: a systematic review. Patient Educ Couns 2013 Apr;91(1):14-24. [doi: 10.1016/j.pec.2012.10.018] [Medline: 23194823]
27. Larsson US, Säljö R, Aronsson K. Patient-doctor communication on smoking and drinking: lifestyle in medical consultations. Soc Sci Med 1987 Jan;25(10):1129-1137. [doi: 10.1016/0277-9536(87)90354-6] [Medline: 3686078]
28. Ginis KA, Heisz J, Spence JC, Clark IB, Antflick J, Ardern CI, et al. Formulation of evidence-based messages to promote the use of physical activity to prevent and manage Alzheimer's disease. BMC Public Health 2017 Feb 17;17(1):209 [FREE Full text] [doi: 10.1186/s12889-017-4090-5] [Medline: 28212648]
29. Blair SN, Kohl HW, Gordon NF, Paffenbarger Jr RS. How much physical activity is good for health? Annu Rev Public Health 1992 May;13(1):99-126. [doi: 10.1146/annurev.pu.13.050192.000531] [Medline: 1599603]
30. Vaughn AE, Bartlett R, Luecking CT, Hennink-Kaminski H, Ward DS. Using a social marketing approach to develop Healthy Me, Healthy We: a nutrition and physical activity intervention in early care and education. Transl Behav Med 2019 Jul 16;9(4):669-681 [FREE Full text] [doi: 10.1093/tbm/iby082] [Medline: 30107586]
31. Liguori G, American College of Sports Medicine. ACSM's Guidelines for Exercise Testing and Prescription. Philadelphia, PA: Wolters Kluwer Health; 2021.
32. Andreu-Perez J, Leff DR, Ip HM, Yang GZ. From wearable sensors to smart implants--toward pervasive and personalized healthcare. IEEE Trans Biomed Eng 2015 Dec;62(12):2750-2762 [FREE Full text] [doi: 10.1109/TBME.2015.2422751] [Medline: 25879838]
33. Lamiae E, Fatiha E, Mohammed B, Hicham GT. A study on smart home for medical surveillance: contribution to smart healthcare paradigm. In: Proceedings of the 4th International Conference on Smart City Applications. 2019 Presented at: SCA '19; October 2-4, 2019; Casablanca, Morocco p. 1-6 URL: https://dl.acm.org/doi/10.1145/3368756.3368994 [doi: 10.1145/3368756.3368994]
34. Katz S. Assessing self-maintenance: activities of daily living, mobility, and instrumental activities of daily living. J Am Geriatr Soc 1983 Dec 27;31(12):721-727. [doi: 10.1111/j.1532-5415.1983.tb03391.x] [Medline: 6418786]
35. Wichmann J. Indoor positioning systems in hospitals: a scoping review. Digit Health 2022 Mar 14;8:205520762210816. [doi: 10.1177/20552076221081696]
36. Rong G, Mendez A, Bou Assi E, Zhao B, Sawan M. Artificial intelligence in healthcare: review and prediction case studies. Engineering 2020 Mar;6(3):291-301. [doi: 10.1016/j.eng.2019.08.015]
37. Shamshirband S, Fathi M, Dehzangi A, Chronopoulos AT, Alinejad-Rokny H. A review on deep learning approaches in healthcare systems: taxonomies, challenges, and open issues. J Biomed Inform 2021 Jan;113:103627 [FREE Full text] [doi: 10.1016/j.jbi.2020.103627] [Medline: 33259944]
38. Liu X, Song W, Zhang W, Di Mauro M, Liotta A. Efficient subjective video quality assessment based on active learning and clustering. In: Proceedings of the 20th International Conference on Advances in Mobile Computing and Multimedia Intelligence. 2022 Presented at: AMCMI '22; November 28-30, 2022; Virtual Event p. 47-56. [doi: 10.1007/978-3-031-20436-4_5]
39. Wang Y, Song W, Tao W, Liotta A, Yang D, Li X, et al. A systematic review on affective computing: emotion models, databases, and recent advances. Inf Fusion 2022 Jul;83-84:19-52. [doi: 10.1016/j.inffus.2022.03.009]
40. Brookshaw M, Sexton A, McGibbon CA. Reliability and validity of a novel wearable device for measuring elbow strength. Sensors (Basel) 2020 Jun 17;20(12):3412 [FREE Full text] [doi: 10.3390/s20123412] [Medline: 32560409]
41. Reeder B, David A. Health at hand: a systematic review of smart watch uses for health and wellness. J Biomed Inform 2016 Oct;63:269-276 [FREE Full text] [doi: 10.1016/j.jbi.2016.09.001] [Medline: 27612974]
42. Straczkiewicz M, James P, Onnela JP. A systematic review of smartphone-based human activity recognition methods for health research. NPJ Digit Med 2021 Oct 18;4(1):148 [FREE Full text] [doi: 10.1038/s41746-021-00514-4] [Medline: 34663863]
43. Bert F, Giacometti M, Gualano MR, Siliquini R. Smartphones and health promotion: a review of the evidence. J Med Syst 2014 Jan;38(1):9995 [FREE Full text] [doi: 10.1007/s10916-013-9995-7] [Medline: 24346929]
44. Xie G, Shangguan A, Fei R, Ji W, Ma W, Hei X. Motion trajectory prediction based on a CNN-LSTM sequential model. Sci China Inf Sci 2020 Oct 15;63(11):212207. [doi: 10.1007/s11432-019-2761-y]
45. Lin ZH, Chen AY, Hsieh SH. Temporal image analytics for abnormal construction activity identification. Autom Constr 2021 Apr;124:103572. [doi: 10.1016/j.autcon.2021.103572]
46. Kaushal N, Rhodes RE, Spence JC, Meldrum JT. Increasing physical activity through principles of habit formation in new gym members: a randomized controlled trial. Ann Behav Med 2017 Aug 10;51(4):578-586. [doi: 10.1007/s12160-017-9881-5] [Medline: 28188586]
47. Hollands GJ, Marteau TM, Fletcher PC. Non-conscious processes in changing health-related behaviour: a conceptual analysis and framework. Health Psychol Rev 2016 Dec 16;10(4):381-394 [FREE Full text] [doi: 10.1080/17437199.2015.1138093] [Medline: 26745243]
48. Kim H, Chang CF. Effectiveness of using personal health records to improve recommended breast cancer screening and reduce racial and geographic disparities among women. J Cancer Educ 2022 Apr 9;37(2):343-351. [doi: 10.1007/s13187-020-01821-2] [Medline: 32648239]

49. Chen X, Hay JL, Waters EA, Kiviniemi MT, Biddle C, Schofield E, et al. Health literacy and use and trust in health information. J Health Commun 2018;23(8):724-734 [FREE Full text] [doi: 10.1080/10810730.2018.1511658] [Medline: 30160641]
50. Paakkari L, Okan O. COVID-19: health literacy is an underestimated problem. Lancet Public Health 2020 May;5(5):e249-e250. [doi: 10.1016/s2468-2667(20)30086-4]
51. Smith B, Magnani JW. New technologies, new disparities: the intersection of electronic health and digital health literacy. Int J Cardiol 2019 Oct 01;292:280-282 [FREE Full text] [doi: 10.1016/j.ijcard.2019.05.066] [Medline: 31171391]
52. Dash S, Shakyawar SK, Sharma M, Kaushik S. Big data in healthcare: management, analysis and future prospects. J Big Data 2019 Jun 19;6(1):54. [doi: 10.1186/s40537-019-0217-0]
53. Friedman CP, Flynn AJ. Computable knowledge: an imperative for learning health systems. Learn Health Syst 2019 Oct 06;3(4):e10203 [FREE Full text] [doi: 10.1002/lrh2.10203] [Medline: 31641690]
54. Rahimi B, Nadri H, Lotfnezhad Afshar H, Timpka T. A systematic review of the technology acceptance model in health informatics. Appl Clin Inform 2018 Jul 15;9(3):604-634 [FREE Full text] [doi: 10.1055/s-0038-1668091] [Medline: 30112741]
55. Kamal SA, Shafiq M, Kakria P. Investigating acceptance of telemedicine services through an extended technology acceptance model (TAM). Technol Soc 2020 Feb;60:101212. [doi: 10.1016/j.techsoc.2019.101212]
56. Núñez-Pacheco C, Loke L. Getting into someone else's soul: communicating embodied experience. Digital Creativity 2020 Oct 27;31(4):245-258. [doi: 10.1080/14626268.2020.1835987]
57. Crilly N, Good D, Matravers D, Clarkson PJ. Design as communication: exploring the validity and utility of relating intention to interpretation. Des Stud 2008 Sep;29(5):425-457. [doi: 10.1016/j.destud.2008.05.002]
58. Lindley J, Akmal HA, Pilling F, Coulton P. Researching AI legibility through design. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. 2020 Presented at: CHI '20; April 25-30, 2020; Honolulu, HI p. 1-13 URL: https://dl.acm.org/doi/abs/10.1145/3313831.3376792 [doi: 10.1145/3313831.3376792]
59. Scott P, De Keizer N, Georgiou A. Applied Interdisciplinary Theory in Health Informatics. Amsterdam, Netherlands: IOS Press; 2019.
60. Leslie M, Eales J, Fast J, Mortenson WB, Atoyebi O, Khayatzadeh-Mahani A. Towards sustainable family care: using goals to reframe the user-centred design of technologies to support carers. Int J Care Caring 2019;3(3):445-451. [doi: 10.1332/239788219x15597493546652]
61. Bogers S, van Kollenburg J, Deckers E, Frens J, Hummels C. A situated exploration of designing for personal health ecosystems through data-enabled design. In: Proceedings of the 2018 Designing Interactive Systems Conference. 2018 Presented at: DIS '18; June 9-13, 2018; Hong Kong, China p. 109-120 URL: https://dl.acm.org/doi/10.1145/3196709.3196769 [doi: 10.1145/3196709.3196769]
62. van Kollenburg J, Bogers SJ, Hummels, Caroline CM, Frens Joep W, Deckers EJ. Data-Enabled Design : A Situated Design Approach That Uses Data as Creative Material When Designing for Intelligent Ecosystems. Eindhoven, the Netherlands: Technische Universiteit Eindhoven; 2019.
63. Lee B, Cooper R, Hands D, Coulton P. Continuous cycles of data-enabled design: reimagining the IoT development process. Artif Intell Eng Des Anal Manuf 2022 Feb 09;36:1-15 [FREE Full text] [doi: 10.1017/s0890060421000299]
64. Yaneva A. Making the social hold: towards an actor-network theory of design. Des Cult 2015 Nov 05;1(3):273-288. [doi: 10.1080/17547075.2009.11643291]
65. Storni C. Unpacking design practices: the notion of thing in the making of artifacts. Sci Technol Human Values 2010 Dec 28;37(1):88-123. [doi: 10.1177/0162243910392795]
66. Bilodeau A, Potvin L. Unpacking complexity in public health interventions with the actor-network theory. Health Promot Int 2018 Feb 01;33(1):173-181. [doi: 10.1093/heapro/daw062] [Medline: 27492825]
67. Patrício L, de Pinho NF, Teixeira JG, Fisk RP. Service design for value networks: enabling value cocreation interactions in healthcare. Serv Sci 2018 Mar;10(1):76-97. [doi: 10.1287/serv.2017.0201]
68. Latour B. On actor-network theory: a few clarifications. Soziale Welt 1996;47(4):369-381 [FREE Full text]
69. Latour B. Reassembling the Social: An Introduction to Actor-Network-Theory. Oxfordshire, UK: Oxford University Press; 2005.
70. Frauenberger C. Entanglement HCI the next wave? ACM Trans Comput Hum Interact 2019 Nov 19;27(1):1-27. [doi: 10.1145/3364998]
71. Walmsley CP, Williams SA, Grisbrook T, Elliott C, Imms C, Campbell A. Measurement of upper limb range of motion using wearable sensors: a systematic review. Sports Med Open 2018 Nov 29;4(1):53 [FREE Full text] [doi: 10.1186/s40798-018-0167-7] [Medline: 30499058]
72. Booth A, Sutton A, Clowes M, Martyn-St James M. Systematic Approaches to a Successful Literature Review. Thousand Oaks, CA: Sage Publishing; 2021.
73. Denyer D, Tranfield D. Producing a systematic review. In: Buchanan D, Bryman A, editors. The SAGE Handbook of Organizational Research Methods. Thousand Oaks, CA: SagePublications; 2009:671-689.
74. Richardson WS, Wilson MC, Nishikawa J, Hayward RS. The well-built clinical question: a key to evidence-based decisions. ACP J Club 1995 Nov 01;123(3):A12. [doi: 10.7326/acpjc-1995-123-3-a12]

75. Sadi F, Klukas R, Hoskinson R. Precise air time determination of athletic jumps with low-cost MEMS inertial sensors using multiple attribute decision making. Sports Technol 2013 Aug 23;6(2):63-77. [doi: 10.1080/19346182.2013.817058]
76. Makino Y, Tsujiuchi N, Ito A, Koizumi T, Nakamura S, Matsuda Y, et al. Quantitative evaluation of unrestrained human gait on change in walking velocity. Annu Int Conf IEEE Eng Med Biol Soc 2014;2014:2521-2524. [doi: 10.1109/EMBC.2014.6944135] [Medline: 25570503]
77. Ghose S, Mitra J, Karunanithi M, Dowling J. Human activity recognition from smart-phone sensor data using a multi-class ensemble learning in home monitoring. Stud Health Technol Inform 2015;214:62-67. [Medline: 26210419]
78. Jaitner T, Schmidt M, Nolte K, Rheinländer C, Wille S, Wehn N. Vertical jump diagnosis for multiple athletes using a wearable inertial sensor unit. Sports Tech 2016 Jan 11;8(1-2):51-57. [doi: 10.1080/19346182.2015.1117476]
79. Zhao W, Lun R, Gordon C, Fofana AB, Espy DD, Reinthal MA, et al. A privacy-aware Kinect-based system for healthcare professionals. In: Proceedings of the 2016 IEEE International Conference on Electro Information Technology. 2016 Presented at: EIT '16; May 19-21, 2016; Grand Forks, ND p. 0205-0210 URL: https://ieeexplore.ieee.org/document/7535241 [doi: 10.1109/eit.2016.7535241]
80. Lin F, Wang A, Zhuang Y, Tomita MR, Xu W. Smart insole: a wearable sensor device for unobtrusive gait monitoring in daily life. IEEE Trans Ind Inf 2016 Dec;12(6):2281-2291. [doi: 10.1109/tii.2016.2585643]
81. Filippoupolitis A, Takand B, Loukas G. Activity recognition in a home setting using off the shelf smart watch technology. In: Proceedings of the 2016 15th International Conference on Ubiquitous Computing and Communications and 2016 International Symposium on Cyberspace and Security. 2016 Presented at: IUCC-CSS '16; December 14-16, 2016; Granada, Spain p. 39-44 URL: https://ieeexplore.ieee.org/document/7828581 [doi: 10.1109/iucc-css.2016.014]
82. Hachaj T, Ogiela MR. The adaptation of GDL motion recognition system to sport and rehabilitation techniques analysis. J Med Syst 2016 Jun 22;40(6):137 [FREE Full text] [doi: 10.1007/s10916-016-0493-6] [Medline: 27106581]
83. Rojas-Mora J, Chávez-Bustamante F, del Río-Andrade J, Medina-Valdebenito N. A methodology for the analysis of soccer matches based on PageRank centrality. In: Peris-Ortiz M, Álvarez-García J, de la Cruz Del Río-Rama M, editors. Sports Management as an Emerging Economic Activity: Trends and Best Practices. Cham, Switzerland: Springer; 2017:257-272.
84. Hachaj T, Piekarczyk M, Ogiela MR. Human actions analysis: templates generation, matching and visualization applied to motion capture of highly-skilled karate athletes. Sensors (Basel) 2017 Nov 10;17(11):2590 [FREE Full text] [doi: 10.3390/s17112590] [Medline: 29125560]
85. Maradani B, Levkowitz H. The role of visualization in tele-rehabilitation: a case study. In: Proceedings of the 7th International Conference on Cloud Computing, Data Science & Engineering. 2017 Presented at: CCDSE '17; January 12-13, 2017; Noida, India p. 643-648 URL: https://ieeexplore.ieee.org/document/7943231 [doi: 10.1109/confluence.2017.7943231]
86. Martin JP, Li Q. Overground vs. treadmill walking on biomechanical energy harvesting: an energetics and EMG study. Gait Posture 2017 Feb;52:124-128. [doi: 10.1016/j.gaitpost.2016.11.027] [Medline: 27912153]
87. Tierney GJ, Joodaki H, Krosshaug T, Forman JL, Crandall JR, Simms CK. Assessment of model-based image-matching for future reconstruction of unhelmeted sport head impact kinematics. Sports Biomech 2018 Mar 28;17(1):33-47 [FREE Full text] [doi: 10.1080/14763141.2016.1271905] [Medline: 28632058]
88. Debarba HG, d. Oliveira ME, Laedermann A, Chague S, Charbonnier C. Augmented reality visualization of joint movements for rehabilitation and sports medicine. In: Proceedings of the 20th Symposium on Virtual and Augmented Reality. 2018 Presented at: SVR '18; October 28-30, 2018; Foz do Iguacu, Brazil p. 114-121 URL: https://ieeexplore.ieee.org/document/8802515 [doi: 10.1109/svr.2018.00027]
89. Tabatabaee Malazi H, Davari M. Combining emerging patterns with random forest for complex activity recognition in smart homes. Appl Intell 2017 Jul 5;48(2):315-330. [doi: 10.1007/s10489-017-0976-2]
90. Cai XJ, Ignacio JI, Mendoza EF, Rabino DJ, Real RP, Roxas EA. IoT-based gait monitoring system for static and dynamic classification of data. In: Proceedings of the IEEE 10th International Conference on Humanoid, Nanotechnology, Information Technology,Communication and Control, Environment and Management. 2018 Presented at: HNICEM '18; December 2, 2018; Baguio City, Philippines p. 1-4 URL: https://ieeexplore.ieee.org/document/8666277 [doi: 10.1109/hnicem.2018.8666277]
91. Chen G, Wang A, Zhao S, Liu L, Chang CY. Latent feature learning for activity recognition using simple sensors in smart homes. Multimed Tools Appl 2017 Aug 18;77(12):15201-15219. [doi: 10.1007/s11042-017-5100-4]
92. Shimada A. Potential of wearable technology for super-aging societies. In: Proceedings of the 6th International Conference on Distributed, Ambient and Pervasive Interactions: Technologies and Contexts. 2018 Presented at: DAPI '18; July 15-20, 2018; Las Vegas, NV p. 214-226 URL: https://link.springer.com/chapter/10.1007/978-3-319-91131-1_17 [doi: 10.1007/978-3-319-91131-1_17]
93. Griffith H, Shi Y, Biswas S. A container-attachable inertial sensor for real-time hydration tracking. Sensors (Basel) 2019 Sep 17;19(18):4008 [FREE Full text] [doi: 10.3390/s19184008] [Medline: 31533275]
94. Gumaei A, Hassan MM, Alelaiwi A, Alsalman H. A hybrid deep learning model for human activity recognition using multimodal body sensing data. IEEE Access 2019;7:99152-99160. [doi: 10.1109/access.2019.2927134]
95. Zhou H, Lv H, Yi K, Pang Z, Yang H, Yang G. An IoT-enabled telerobotic-assisted healthcare system based on inertial motion capture. In: Proceedings of the IEEE 17th International Conference on Industrial Informatics. 2019 Presented at:

INDIN '19; July 22-25, 2019; Helsinki, Finland p. 1373-1376 URL: https://ieeexplore.ieee.org/document/8972195 [doi: 10.1109/indin41052.2019.8972195]
96. Bieniek A, Szczygio A, Chrzan M, Wodarski P, Morys M, Bacik B, et al. Biomechatronic simulator for fencing training using virtual reality technology. In: Świder J, Kciuk S, Trojnacki M, editors. Mechatronics 2017 - Ideas for Industrial Applications. Cham, Switzerland: Springer; 2019:30-37.
97. Varma AL, Sinha VK, Das D. Health assessment for office workers by tracking the wrist motion centric activity. In: Proceedings of the 2019 IEEE 16th India Council International Conference. 2019 Presented at: INDICON '19; December 13-15, 2019; Rajkot, India p. 1-4 URL: https://ieeexplore.ieee.org/abstract/document/9030309 [doi: 10.1109/indicon47234.2019.9030309]
98. Norouzi S, Esfandiarpour F, Mehdizadeh S, Yousefzadeh NK, Parnianpour M. Lower extremity kinematic analysis in male athletes with unilateral anterior cruciate reconstruction in a jump-landing task and its association with return to sport criteria. BMC Musculoskelet Disord 2019 Oct 27;20(1):492 [FREE Full text] [doi: 10.1186/s12891-019-2893-5] [Medline: 31656192]
99. Cronin NJ, Rantalainen T, Ahtiainen JP, Hynynen E, Waller B. Markerless 2D kinematic analysis of underwater running: a deep learning approach. J Biomech 2019 Apr 18;87:75-82. [doi: 10.1016/j.jbiomech.2019.02.021] [Medline: 30850178]
100. Oguntala GA, Abd-Alhameed RA, Ali NT, Hu Y, Noras JM, Eya NN, et al. SmartWall: novel RFID-enabled ambient human activity recognition using machine learning for unobtrusive health monitoring. IEEE Access 2019;7:68022-68033. [doi: 10.1109/access.2019.2917125]
101. Zago M, Sforza C, Dolci C, Tarabini M, Galli M. Use of machine learning and wearable sensors to predict energetics and kinematics of cutting maneuvers. Sensors (Basel) 2019 Jul 12;19(14):3094 [FREE Full text] [doi: 10.3390/s19143094] [Medline: 31336997]
102. Ahmed S, Demirel O. A framework to assess human performance in normal and emergency situations. ASCE ASME J Risk Uncertain Eng Syst Part B Mech Eng 2020;6(1). [doi: 10.1115/1.4044791]
103. Hendry D, Chai K, Campbell A, Hopper L, O'Sullivan P, Straker L. Development of a human activity recognition system for ballet tasks. Sports Med Open 2020 Feb 07;6(1):10 [FREE Full text] [doi: 10.1186/s40798-020-0237-5] [Medline: 32034560]
104. Zelck S, Verwulgen S, Denteneer L, Scataglini S. Digital human modeling as a risk assessment tool for maritime workers. In: Proceedings of the 6th International Digital Human Modeling Symposium. 2020 Presented at: DHM '20; August 31-September 2, 2020; Skövde, Sweden p. 398-407 URL: https://ebooks.iospress.nl/publication/55325 [doi: 10.1007/978-3-030-79763-8_44]
105. Zrenner M, Küderle A, Roth N, Jensen U, Dümler B, Eskofier BM. Does the position of foot-mounted IMU sensors influence the accuracy of spatio-temporal parameters in endurance running? Sensors (Basel) 2020 Oct 07;20(19):5705 [FREE Full text] [doi: 10.3390/s20195705] [Medline: 33036477]
106. Mukherjee D, Mondal R, Singh PK, Sarkar R, Bhattacharjee D. EnsemConvNet: a deep learning approach for human activity recognition using smartphone sensors for healthcare applications. Multimed Tools Appl 2020 Aug 22;79(41-42):31663-31690. [doi: 10.1007/s11042-020-09537-7]
107. Lee M, Youm C, Noh B, Park H. Gait characteristics based on shoe-type inertial measurement units in healthy young adults during treadmill walking. Sensors (Basel) 2020 Apr 08;20(7):2095 [FREE Full text] [doi: 10.3390/s20072095] [Medline: 32276416]
108. Severin IC. Head posture monitor based on 3 IMU sensors: consideration toward healthcare application. In: Proceedings of the 2020 International Conference on e-Health and Bioengineering. 2020 Presented at: EHB '20; October 29-30, 2020; Iasi, Romania p. 1-4 URL: https://ieeexplore.ieee.org/document/9280106 [doi: 10.1109/ehb50910.2020.9280106]
109. Konak O, Wegner P, Arnrich B. IMU-based movement trajectory heatmaps for human activity recognition. Sensors (Basel) 2020 Dec 15;20(24):7179 [FREE Full text] [doi: 10.3390/s20247179] [Medline: 33333839]
110. Tumialis A, Smirnov A, Fadeev K, Alikovskaia T, Khoroshikh P, Sergievich A, et al. Motor program transformation of throwing dart from the third-person perspective. Brain Sci 2020 Jan 18;10(1):55 [FREE Full text] [doi: 10.3390/brainsci10010055] [Medline: 31963722]
111. Mekruksavanich S, Jitpattanaku A. Smartwatch-based human activity recognition using hybrid LSTM network. In: Proceedings of the 2020 IEEE SENSORS. 2020 Presented at: SENSORS '20; October 25-28, 2020; Rotterdam, Netherlands p. 1-4 URL: https://ieeexplore.ieee.org/document/9278630 [doi: 10.1109/sensors47125.2020.9278630]
112. Jalal A, Batool M, Kim K. Stochastic recognition of physical activity and healthcare using tri-axial inertial wearable sensors. Appl Sci 2020 Oct 13;10(20):7122. [doi: 10.3390/app10207122]
113. Ullah F, Iqbal A, Iqbal S, Kwak D, Anwar H, Khan A, et al. A framework for maternal physical activities and health monitoring using wearable sensors. Sensors (Basel) 2021 Jul 21;21(15):4949 [FREE Full text] [doi: 10.3390/s21154949] [Medline: 34372186]
114. Gao S, He T, Zhang Z, Ao H, Jiang H, Lee C. A motion capturing and energy harvesting hybridized lower-limb system for rehabilitation and sports applications. Adv Sci (Weinh) 2021 Oct;8(20):e2101834 [FREE Full text] [doi: 10.1002/advs.202101834] [Medline: 34414697]
115. Canavan PK, Suderman B, Yang N. A novel approach for baseball pitch analysis using a full body motion analysis suit: A case series study. J Hum Sport Exerc 2020;16(3):689-700. [doi: 10.14198/jhse.2021.163.17]

116. Tsai WL, Su LW, Ko TY, Pan TY, Hu MC. Feasibility study on using AI and VR for decision-making training of basketball players. IEEE Trans Learn Technol 2021 Dec 1;14(6):754-762. [doi: 10.1109/tlt.2022.3145093]
117. Drazan JF, Phillips WT, Seethapathi N, Hullfish TJ, Baxter JR. Moving outside the lab: markerless motion capture accurately quantifies sagittal plane kinematics during the vertical jump. J Biomech 2021 Aug 26;125:110547 [FREE Full text] [doi: 10.1016/j.jbiomech.2021.110547] [Medline: 34175570]
118. Guo H, Zhang L. Online evaluation system of college physical education students based on big data and artificial intelligence technology. In: Chang JW, Yen N, Hung JC, editors. Frontier Computing. Lecture Notes in Electrical Engineering. Cham, Switzerland: Springer; 2021:459-465.
119. Noamani A, Agarwal K, Vette A, Rouhani H. Predicted threshold for seated stability: estimation of margin of stability using wearable inertial sensors. IEEE J Biomed Health Inform 2021 Sep;25(9):3361-3372. [doi: 10.1109/jbhi.2021.3073352]
120. Liu J, Wang L, Zhou H. The application of human-computer interaction technology fused with artificial intelligence in sports moving target detection education for college athlete. Front Psychol 2021 Jul 22;12:677590 [FREE Full text] [doi: 10.3389/fpsyg.2021.677590] [Medline: 34366996]
121. Coffen B, Mahmud MS. TinyDL: edge computing and deep learning based real-time hand gesture recognition using wearable sensor. In: Proceedings of the 2020 IEEE International Conference on E-health Networking, Application & Services. 2020 Presented at: HEALTHCOM '20; March 1-2, 2021; Shenzhen, China p. 1-6 URL: https://ieeexplore.ieee.org/document/9399005 [doi: 10.1109/healthcom49281.2021.9399005]
122. Turetta C, Demrozi F, Pravadelli G. A freely available system for human activity recognition based on a low-cost body area network. In: Proceedings of the 2022 IEEE 46th Annual Computers, Software, and Applications Conference. 2022 Presented at: COMPSAC '22; June 27-July 1, 2022; Los Alamitos, CA p. 395-400 URL: https://ieeexplore.ieee.org/document/9842668 [doi: 10.1109/compsac54236.2022.00062]
123. Gala HJ, Li C, Zheng L. Characteristic analysis of Tai Ji Chuan standard action based on and perceptual IoT network. Secur Commun Netw 2022 Mar 18;2022:1-9. [doi: 10.1155/2022/6713578]
124. Shahid ZK, Saguna S, Åhlund C. Detecting anomalies in daily activity routines of older persons in single resident smart homes: proof-of-concept study. JMIR Aging 2022 Apr 11;5(2):e28260 [FREE Full text] [doi: 10.2196/28260] [Medline: 35404260]
125. Mukherjee A, Bose A, Chaudhuri DP, Kumar A, Chatterjee A, Ray SK, et al. Edge-based human activity recognition system for smart healthcare. J Inst Eng India Ser B 2021 Oct 12;103(3):809-815. [doi: 10.1007/s40031-021-00663-w]
126. Wu Q, Chen X, Zhou Z, Zhang J. FedHome: cloud-edge based personalized federated learning for in-home health monitoring. IEEE Trans Mob Comput 2022 Aug 1;21(8):2818-2832. [doi: 10.1109/tmc.2020.3045266]
127. Song Y, Zhao L. Skill movement trajectory recognition of freestyle skiing u-shaped field based on deep learning and multitarget tracking algorithm. Comput Intell Neurosci 2022 Aug 12;2022:7992045-7992012 [FREE Full text] [doi: 10.1155/2022/7992045] [Medline: 35990161]
128. Yang C, Wang X, Mao S. TARF: technology-agnostic RF sensing for human activity recognition. IEEE J Biomed Health Inform 2023 Feb;27(2):636-647. [doi: 10.1109/jbhi.2022.3175912]
129. Ayman A, Attalah O, Shaban H. Smart system for recognizing daily human activities based on wrist IMU sensors. In: Proceedings of the 2019 International Conference on Advances in the Emerging Computing Technologies. 2019 Presented at: AECT '19; February 10, 2019; Al Madinah Al Munawwarah, Saudi Arabia p. 1-6 URL: https://ieeexplore.ieee.org/document/9194154 [doi: 10.1109/aect47998.2020.9194154]
130. Farooq A, Martin A, Janssen X, Wilson MG, Gibson A, Hughes A, et al. Longitudinal changes in moderate-to-vigorous-intensity physical activity in children and adolescents: a systematic review and meta-analysis. Obes Rev 2020 Jan 23;21(1):e12953 [FREE Full text] [doi: 10.1111/obr.12953] [Medline: 31646739]
131. Molanorouzi K, Khoo S, Morris T. Motives for adult participation in physical activity: type of activity, age, and gender. BMC Public Health 2015 Jan 31;15:66 [FREE Full text] [doi: 10.1186/s12889-015-1429-7] [Medline: 25637384]

Abbreviations

ADL: activity of daily living
AI: artificial intelligence
ANT: actor network theory
DED: data-enabled design
IoT: Internet of Things
NDLPA: Network of Daily Living and Physical Activity
PICO: participants, interventions, comparisons, and outcomes
TAM: technology acceptance model
VR: virtual reality

Edited by L Buis; submitted 07.02.23; peer-reviewed by W Lin, S Mitra; comments to author 20.12.23; revised version received 16.02.24; accepted 14.03.24; published 06.05.24.

Please cite as:
Wei L, Wang SJ
Motion Tracking of Daily Living and Physical Activities in Health Care: Systematic Review From Designers' Perspective
JMIR Mhealth Uhealth 2024;12:e46282
URL: https://mhealth.jmir.org/2024/1/e46282
doi:10.2196/46282
PMID:38709547

Review

Digital Phenotyping for Stress, Anxiety, and Mild Depression: Systematic Literature Review

Adrien Choi[1], BAdvSci(Hons); Aysel Ooi[1], BSocSci, MInfoTech; Danielle Lottridge[1], PhD

School of Computer Science, Faculty of Science, University of Auckland, Auckland, New Zealand

Corresponding Author:
Danielle Lottridge, PhD
School of Computer Science
Faculty of Science
University of Auckland
38 Princes Street
Auckland, 1010
New Zealand
Phone: 64 9 373 7599 ext 82930
Email: d.lottridge@auckland.ac.nz

Abstract

Background: Unaddressed early-stage mental health issues, including stress, anxiety, and mild depression, can become a burden for individuals in the long term. Digital phenotyping involves capturing continuous behavioral data via digital smartphone devices to monitor human behavior and can potentially identify milder symptoms before they become serious.

Objective: This systematic literature review aimed to answer the following questions: (1) what is the evidence of the effectiveness of digital phenotyping using smartphones in identifying behavioral patterns related to stress, anxiety, and mild depression? and (2) in particular, which smartphone sensors are found to be effective, and what are the associated challenges?

Methods: We used the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) process to identify 36 papers (reporting on 40 studies) to assess the key smartphone sensors related to stress, anxiety, and mild depression. We excluded studies conducted with nonadult participants (eg, teenagers and children) and clinical populations, as well as personality measurement and phobia studies. As we focused on the effectiveness of digital phenotyping using smartphones, results related to wearable devices were excluded.

Results: We categorized the studies into 3 major groups based on the recruited participants: studies with students enrolled in universities, studies with adults who were unaffiliated to any particular organization, and studies with employees employed in an organization. The study length varied from 10 days to 3 years. A range of passive sensors were used in the studies, including GPS, Bluetooth, accelerometer, microphone, illuminance, gyroscope, and Wi-Fi. These were used to assess locations visited; mobility; speech patterns; phone use, such as screen checking; time spent in bed; physical activity; sleep; and aspects of social interactions, such as the number of interactions and response time. Of the 40 included studies, 31 (78%) used machine learning models for prediction; most others (n=8, 20%) used descriptive statistics. Students and adults who experienced stress, anxiety, or depression visited fewer locations, were more sedentary, had irregular sleep, and accrued increased phone use. In contrast to students and adults, less mobility was seen as positive for employees because less mobility in workplaces was associated with higher performance. Overall, travel, physical activity, sleep, social interaction, and phone use were related to stress, anxiety, and mild depression.

Conclusions: This study focused on understanding whether smartphone sensors can be effectively used to detect behavioral patterns associated with stress, anxiety, and mild depression in nonclinical participants. The reviewed studies provided evidence that smartphone sensors are effective in identifying behavioral patterns associated with stress, anxiety, and mild depression.

(JMIR Mhealth Uhealth 2024;12:e40689) doi:10.2196/40689

KEYWORDS

digital phenotyping; passive sensing; stress; anxiety; depression; PRISMA; Preferred Reporting Items for Systematic Reviews and Meta-Analyses; mobile phone

Introduction

Background

Digital phenotyping is "the moment-by-moment quantification of the individual level human phenotype in situ using data from personal digital devices" [1]. Digital phenotyping applies the concept of phenotypes, in other words, the observable characteristics resulting from the genotype and environment, to conceptualize observable patterns in individuals' digital data. In the last decade, digital phenotyping studies have been able to compare typical and atypical patterns in daily activities to correlate atypical behavior with negative emotions [2,3]. Behavioral patterns include variations in mobility, frequency of being in various locations, and sleep patterns. In smartphones, user data can be stored, managed, interpreted, and captured in enormous amounts [1,4,5]. This can be done actively or passively. Active data collection requires the user to self-report and complete surveys, whereas passive sensing collects data automatically without user input [5]. Most studies combine active and passive sensing to more accurately detect and predict behavioral abnormalities. Modern smartphone analytics can be used for the discovery of commonalities and abnormalities in user behavior. The ease of using passive sensing makes it an ideal data gathering method for mental health studies [6-8] and an ideal technique for assessing mental health [9].

Digital phenotyping has been successful in the early detection and prediction of behaviors related to neuropharmacology [10]; cardiovascular diseases [11]; diabetes [12]; and major severe injuries, such as spinal cord injury [13], motivating further adoption. Digital phenotyping has also proven useful for the detection of severe mental health issues, such as schizophrenia [14,15], bipolar disorder [16], and suicidal thoughts [17]. Digital phenotyping has been so successful for specialized, clinical populations that it is increasingly considered for mass market use with nonclinical populations. Digital phenotyping applications and software tools have been used to capture employee information, such as their screen time and clicking patterns [18]. However, there are not many digital phenotyping studies that have specifically examined the detection or prediction of stress, anxiety, and mild depression.

Individuals with stress, anxiety, and mild depression can develop chronic mental health symptoms that impact their mobility, satisfaction with life, and social interaction [19,20]. When these symptoms are not detected early, they worsen, and the impact is more significant [21-23], increasing the need for medication and hospitalization. This makes mild mental health symptoms a valid target for digital phenotyping, as its goal is to enable early detection and, subsequently, early treatment. Smartphones are increasingly ubiquitous [24], which makes them an optimal platform for digital phenotyping. We constrained our systematic literature search to the more challenging problem of the detection of mild mental health symptoms using only smartphone sensors and excluded studies that used additional wearable sensors. In general, we believe that additional wearables might increase the effectiveness of digital phenotyping in detecting stress, anxiety, and mild depression. Given the ubiquity of smartphones, we aimed to answer the following question: what is the effectiveness of digital phenotyping using smartphone sensors in detecting stress, anxiety, and mild depression?

Objectives

The objective of this systematic literature review was to better understand the current uses of digital phenotyping and results of using digital phenotyping for the detection and prediction of mild behavioral patterns related to stress, anxiety, and mild depression. The 2 research questions this review sought to answer were as follows:

1. What is the evidence of the effectiveness of digital phenotyping using smartphones in identifying behavioral patterns related to stress, anxiety, and mild depression?
2. In particular, which smartphone sensors are found to be effective, and what are the associated challenges?

For these research questions, we considered statistically significant associations between sensor patterns and behavioral patterns as evidence of effectiveness.

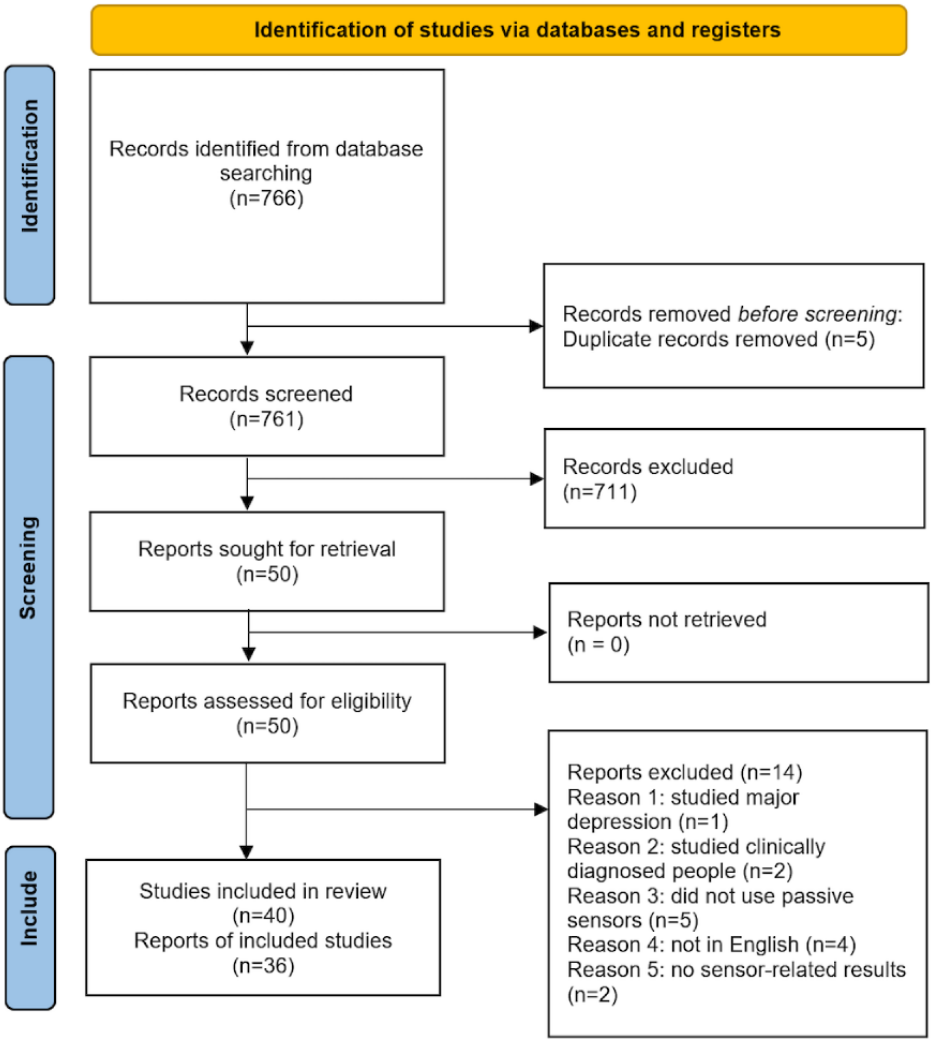
Methods

Type of Studies

This review followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [25] (Multimedia Appendix 1). Figure 1 shows the reviewing process and search results. In the first round of screening studies, 1 author excluded studies that were not relevant to the research questions. Another author reran the queries for confirmation. Studies were included in this review if they were conducted to measure and detect stress, anxiety, or mild depression, even if they included other variables, such as job performance, promotion, or discrimination. We included studies in which data were collected through smartphones with an iOS (Apple Inc) or Android (Google LLC) operating system. Data collected through wearable devices were excluded. We included studies in which the participants were adults aged ≥18 years and were from a nonclinical population. Studies conducted with nonadult participants (eg, teenagers and children) were excluded. Given our research questions, if the studies' participants had or had had any severe mental health disorder, such as schizophrenia, bipolar disorder, or psychosis, they were not included. We also excluded personality and character measurement and phobia studies. The primary research language was English. The studies included were conducted from September 2010 to September 2023. Peer-reviewed conference articles and journal articles were included. The data we wished to extract were the study aim, data collected, operating system in the smartphone used for data collection, behavioral patterns identified, surveys used for verification, and sample size. A total of 3 authors reviewed the studies independently to extract data and confirm the extracted data. After the first round of data extraction, 1 author re-examined the studies to extract the predictive modeling used. These data are presented in the *Results* section. We noticed that participants in the included studies fell into 1 of 3 major groups (ie, students, adults, and employees). We refer to the participants of the studies that recruited adults enrolled in universities as "students," participants of the studies that recruited adults

unaffiliated to any particular organization as "adults," and participants of the studies that recruited adults employed at a particular organization as "employees."

Figure 1. Systematic literature reviewing process and search results with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) diagram.



Search Strategy

A total of 3 databases were queried: Web of Science, ACM, and PubMed. PubMed is a medicine-based database, ACM is a technology-based database, and Web of Science is a cross-domain database. The search query was the same for the 3 platforms: "digital phenotyping" OR "passive sensing" AND (stress OR anxiety OR ((mild OR moderate) AND depression)).

Results

Duration

The study length varied from 10 days [26] to 3 years [27]. One study [28] conducted in-depth interviews with students lasting an average of 4.5 hours per person, and another study was a controlled laboratory study [29]. These 2 studies are not presented in Table 1. In the studies conducted with students, a semester or spring or winter term was a common duration. The studies with general nonclinical adult populations were typically longer than those with students.

Table 1. Duration of the reviewed studies (N=38; 2 studies are excluded, as 1 [28] is interview based and the other [29] is a controlled laboratory study).

Study, year	Length of the study (d)
Adams et al [26], 2014	10
Cai et al [30], 2018	14
Boukhechba et al [31], 2018	14
Di Matteo et al [32], 2021	14
Jacobson et al [33], 2020	16
Wen et al [34], 2021	21
Melcher et al [35], 2023	28
Fukuzawa et al [36], 2019	28
Rashid et al [37], 2020	35
Zakaria et al [38], 2019	35
DaSilva et al [39], 2019	43
Nepal et al [40], 2020	60
Saha et al [41], 2019	68
Morshed et al [42], 2019	70
Acikmese et al [43], 2019	70
Zakaria et al [38], 2019	81
Zakaria et al [38], 2019	81
Boukhechba et al [44], 2017	98
Tseng et al [45], 2016	98
Morshed et al [42], 2019	98
Xu et al [46], 2019	106
Chikersal et al [47], 2021	112
Meyerhoff et al [48], 2021	112
Xu et al [46], 2019	113
Rhim et al [49], 2020	121
Wang et al [50], 2018	121
Currey and Torous [51], 2022	147
Di Matteo et al [52], 2021	153
Sefidgar et al [53], 2019	153
Mendu et al [54], 2020	153
Pratap et al [55], 2017	181
Mirjafari et al [56], 2019	260
Currey et al [57], 2023	336
Huckins et al [58], 2020	458
Mack et al [59], 2021	458
Xu et al [60], 2023	458
Nepal et al [61], 2022	730
Servia-Rodríguez et al [27], 2017	1095

Number of Participants

The number of participants ranged from a minimum of 7 adults [26] to a maximum of 18,000 adults [27]. Apart from the 3-year longitudinal study with 18,000 participants [27], the average number of participants was 129.4 (SD 184.01). We observed a pattern of attrition, where the number of participants who completed the study was lower than the number of the

participants recruited. The number of participants reported in this review is the final sample size. For example, one of the studies [52] recruited 112 participants, of whom 84 (75%) completed the study. In the study by Pratap et al [55], there was a drastic drop in participants, with only 359 (30.42%) of the 1180 enrolled participants completing the study. Another significant drop was seen in the study by Nepal et al [40], where 750 participants were interested in the research, whereas only 141 (18.8%) of them completed the study. Some studies were less affected; for example, 86 participants started the study by Rhim et al [49], and 78 (91%) completed it.

Publication Years of the Studies

Although the query started with the year 2010, the earliest publication was from 2014 [26], extending to articles published as of April 2023 [35]. Over the years, the interest in detecting and predicting stress, anxiety, and mild depression in the nonclinical population has increased (Table 2).

Table 2. Number of reviewed reports (N=36) by year.

Year	Publication, n (%)
2014	1 (3)
2016	1 (3)
2017	3 (8)
2018	4 (11)
2019	10 (28)
2020	6 (17)
2021	6 (17)
2022	2 (6)
2023	3 (8)

Studies With the iOS and Android Operating Systems

The Android operating system was more common than iOS. Among the 40 included studies, only 2 (5%) were compatible with only iOS [29,51]. A total of 27 (68%) studies were available for both iOS and Android [26,28,30,34,35,37-42,45-47,50,53-61]. A total of 11 (28%) studies were for only Android users [27,31-33,36,42-44,48,49,52]. The reasons identified for the use of the Android operating system were that it has more freedom to capture more modalities, such as keyboard typing and use of apps, and that Android devices enable apps to run more easily in the background [49].

Studies With Students

Table 3 presents the data extracted from the studies that were conducted with student populations. The average length of the studies with students was 158.6 (SD 176.4) days. The average number of participants was 137.3 (SD 152.1). There were significantly more studies with students than studies with employees or general adults. The sample sizes of the studies with students were similar to those of the studies with adults but smaller than those of the studies with employees. In the studies with students, various passive sensors were used, and some were found to be effective for detection, prediction, or both.

Of the 28 studies with students, 23 (82%) used machine learning models for prediction. A total of 12 studies (43%) [30,31,33,37,38,44,46,47,54] used decision tree–based methods, and 9 studies (32%) [37,39,42,49-51,57,58] used regression-based methods. A total of 3 (11%) studies conducted in recent years [43,60,61] used deep neural networks because of their enhanced ability to discern underlying patterns in large unstructured data sets. Tree-based models have the best performance when trained with structured data, and the reported studies mostly used tree-based models and structured data. Among the 28 studies, 2 studies [57,60] conducted in 2023 addressed the generalizability of their proposed detection method and verified its applicability across students from various years, classes, and institutions. Two (7%) studies [42,43] in Table 3 used the StudentLife data set [62]. Each study contributed substantial original analyses including different behavioral patterns and was considered a “study” in this systematic review. Entries with “N/A” in the predictive modeling column indicate that the study did not involve any attempts to predict future occurrences. However, these studies may still contain statistical analyses as part of their research approach. Overall, students who experienced depression, anxiety, and stress visited fewer locations [39,44,50,58-60] and were more sedentary [47,50,58-60]. Depression was also associated with shorter or irregular sleep [35,46,47,50,52,59,60] and accrued phone use [46,47,50,51,58-60].

Table 3. Summary of the reviewed studies with student participants.

Study, year	Aim	Data collected	Operating system	Behavioral patterns	Predictive modeling	Verification surveys	Sample size, n
Huckins et al [58], 2020	Understand how students' behavioral health and mental health are affected by the COVID-19 pandemic	GPS, accelerometer, phone lock and unlock, and light sensor data	iOS (Apple Inc) and Android (Google LLC)	At the start of the COVID-19 pandemic, students were more depressed and anxious, used their phones more, visited fewer locations, and spent more time sedentary. Depression and stress were associated with increasing COVID-19–related news coverage.	Linear regressors were used to inspect how behavioral changes were affected by COVID-19 news reports.	PHQ[a]-4	217 students
Melcher et al [35], 2023	Understand how behavioral patterns correlate with mental health for students during the COVID-19 pandemic	GPS, accelerometer, call log, and phone use data	iOS and Android	Individuals with more irregular sleep patterns had worse sleep quality and were experiencing more depression and more stress than those with consistent sleep patterns.	N/A[b]	PHQ-9, DASS[c], SIAS[d], GAD-7[e], PQ[f], PSS[g], PSQI[h], BASIS[i], SF[j]-36, SFS[k], Flourishing Scale, CGI[l], HDRS[m], CAS[n], HAI[o], and UCLA[p]-Loneliness Scale	100 students
Jacobson et al [33], 2020	Predict social anxiety symptom severity and discriminate between depression, negative affect, and positive affect	Accelerometer, call log, and SMS text message data	Android	Measures of SMS text message and call response time discriminated among depression, negative affect, and positive affect. Accelerometer patterns suggested that persons with low social anxiety walked at a steady pace, whereas persons with high social anxiety walked more quickly with more irregularity.	XGBoost[q] with LOOCV[r] was used to predict social anxiety symptom severity.	SIAS, DASS-21, and PANAS[s]	59 students
DaSilva et al [39], 2019	Predict stress	GPS, accelerometer, phone lock and unlock, microphone, and light sensor data	iOS and Android	Students with stress were more likely to spend less time in campus food locations and more time in schoolwork locations. Students with stress traveled less, engaged in fewer conversations, and were in quieter environments during evenings.	Penalized generalized estimating equations were used to prune features and fit a marginal regression model to predict stress.	MPSM[t]	94 students
Acikmese and Alptekin [43], 2019	Predict stress level	Accelerometer, microphone, Bluetooth, light sensor, phone lock and unlock, phone charge, and app use data (GPS and Wi-Fi data were collected but not used)	Android	Students were successfully categorized as stressed or nonstressed using the measured sensors.	LSTM[u], CNN[v], and CNN-LSTM were used to classify stress, with LSTM yielding the best accuracy.	Self-reported stress	48 students

Study, year	Aim	Data collected	Operating system	Behavioral patterns	Predictive modeling	Verification surveys	Sample size, n
Rooksby et al [28], 2019	Understand students' perspectives about digital phenotyping	GPS, phone lock and unlock, phone charge, battery, microphone, Bluetooth, light sensor, SMS text message, email, app use, call log, camera, and keyboard data	iOS and Android	None of the results related sensors to symptoms of depression or anxiety. Students have privacy concerns regarding the use of app use logs, Bluetooth data, call logs, camera data, keyboard data, and microphone data but not regarding the use of battery, or light sensor. Students had privacy concerns with the use of SMS text message content but not with counts of messages.	N/A	PHQ-9, GAD-7, and WEMWBS[w]	15 students
Chikersal et al [47], 2021	Predict postsemester depressive symptoms	GPS, accelerometer, Bluetooth, Wi-Fi, phone use, call log, and microphone data	iOS and Android	Depression was predicted by participants' social context in the afternoons and evenings, phone use throughout the day, long periods without exercise, periods of disturbed sleep at night, and time spent outdoors.	Trained an ensemble classifier with the outputs from models containing features from 1 sensor, with different setting combinations.	BDI[x]-II	138 students
Morshed et al [42], 2019	Predict mood instability	Accelerometer, microphone, Bluetooth, light sensor, Wi-Fi, GPS, phone lock and unlock, and phone charge data	Android	Mood instability was negatively correlated with the duration of sleep, the number of conversations, the amount of activity, and outdoor mobility.	Ridge regression with regularization was used to infer mood instability score.	EMAs[y], PAM[z], and PANAS	48 students
Zakaria et al [38], 2019	Detect depression and stress	Wi-Fi data	iOS and Android	Students with severe stress spent significantly less time on campus and were less involved in work-related activities than students with normal stress. Students with severe stress were more involved in these activities at the start of the semester, but the involvement decreased over time.	The random forest stress model with domain-specific features achieved the best result, with feature sets changed every 6 days.	PSS-4, PHQ-8, and BFI[aa]	62 students
Zakaria et al [38], 2019	Detect depression and stress	Wi-Fi data	iOS and Android	Same patterns as those mentioned earlier.	The random forest model that excluded domain-specific features achieved the best result, with feature sets changed every 6 days.	PSS-4, PHQ-8, and BFI	11 students
Zakaria et al [38], 2019	Detect depression and stress	Wi-Fi data	iOS and Android	Same patterns as those mentioned earlier.	The best model is a random forest model with the neuroticism score added as an additional feature, with sensor data sets calculated with a 6-day interval.	PSS-4, PHQ-8, and BFI	35 students
Wang et al [50], 2018	Predict depression	Light sensor, GPS, accelerometer, microphone, screen on and off, and phone lock and unlock data	iOS and Android	Students who experienced depression had more irregular sleep patterns, used their phones more at study places, spent more time stationary, and visited fewer locations.	LASSO[ab] regression was used to predict presurvey and postsurvey PHQ-9 scores.	PHQ-4 and PHQ-8	83 students
Exposito et al [29], 2018	Detect stress	Keyboard 3D touch data	iOS	Students' typing pressure increased under stress.	N/A	Self-reported stress	11 students

Study, year	Aim	Data collected	Operating system	Behavioral patterns	Predictive modeling	Verification surveys	Sample size, n
Rhim et al [49], 2020	Detect subjective well-being and stress	Accelerometer, GPS, screen on and off, app use, and notification data	Android	Lower subjective well-being was associated with more time spent on campus, more time spent stationary, increased phone use in the evenings, and more expenses.	Hierarchical regression models were used to predict subjective well-being.	COMOSWB[ac], PHQ, SAS[ad], PPC[ae], and BFI	78 students
Sefidgar et al [53], 2019	Detect stress, anxiety, and gender discrimination	Accelerometer, GPS, phone lock and unlock, screen on and off, and call log data	iOS and Android	Students who experienced discrimination became more physically active; their phone use increased in the morning, they had more calls in the evening, and they spent more time in bed on the day of the discrimination.	Linear regression was used to predict long-term changes in mental health states; hierarchical linear modeling was used for short-term prediction.	UCLA Loneliness Scale, SSS[af], MAAS[ag], ERQ[ah], BRS[ai], PSS, CES-D[aj], STAI[ak], and self-reported affect and fairness of treatment	176 students
Cai et al [30], 2018	Detect state affect, stress, anxiety, and depression	Accelerometer, GPS, call log, and SMS text message data	iOS and Android	Negative emotions were related to geographical locations, but this was affected by personal routines and preferences, for example, liking cinema theatres. On Fridays and Saturdays, students reported less negative states.	Compared support vector machine, random forest, and XGboost with LOSOCV[al] and LOOCV to predict negative affect. The best model was support vector machine with LOOCV.	SIAS and self-reported affect (EMAs)	220 students
Boukhechba et al [31], 2018	Predict response rate and latency to EMA	GPS, call log, accelerometer, and SMS text message data	Android	None of the results related sensors to symptoms of depression or anxiety.	Used random forest, support vector machine, and a multilayer perceptron of 1 hidden layer with LOOCV to predict the compliance rate of EMA responses.	Self-reported affect (EMAs)	65 students
Xu et al [46], 2019	Detect depression	Accelerometer, battery or charge, Bluetooth, call log, screen, location, and phone lock and unlock data	iOS and Android	Students who experienced depression had more disturbed sleep patterns and more phone interactions than students who did not experience depression.	AdaBoost[am] with decision tree–based components achieved the best performance when features were hybrid (contextually filtered + unimodal).	BDI-II	138 students
Xu et al [46], 2019	Detect depression	Accelerometer, battery or charge, Bluetooth, call log, screen, location, and phone lock and unlock data	iOS and Android	Same patterns as those mentioned earlier.	AdaBoost with decision tree–based components achieved a similar result to majority-based baseline predictors.	BDI-II	212 students
Boukhechba et al [44], 2017	Predict social anxiety	GPS, call log, and SMS text message data	Android	Students who experienced high social anxiety may be more likely to buy food so they can eat at home; they tended to visit fewer places and had a narrower range of activities.	Decision tree was used to predict SAS.	SIAS	54 students
Rashid et al [37], 2020	Predict social anxiety and evaluate the effectiveness of imputation methods in handling missing data	GPS, pedometer, accelerometer, call log, and SMS text message data	iOS and Android	The level of social anxiety was predicted, but there were no specific patterns relating sensors to symptoms of social anxiety.	Evaluated 7 predictive models: linear regression, decision tree, XBboost, lightGBM[an], random forest, MERF[ao], and CatBoost.	SIAS and self-reported dimensions of social anxiety	80 students

Study, year	Aim	Data collected	Operating system	Behavioral patterns	Predictive modeling	Verification surveys	Sample size, n
Mendu et al [54], 2020	Explore the relationships among private social media messages, personality traits, and symptoms of mental illness	Facebook (Meta Platforms, Inc) private messages	iOS and Android	Students who experienced anxiety received responses later, had more night-time communications, talked less about games and sports, and used more plural pronouns.	Used random forest classifier to select features and support vector machine with LOOCV to predict each psychological measure binarily.	STAI, UCLA Loneliness Scale, and TIPI[ap]	103 students
Tseng et al [45], 2016	Detect stress and its relationship with academic performance	Location, activity, step count (iOS only), audio, accelerometer (iOS only), device use, charging event, battery, light (Android only), SMS text message (Android only) and call (Android only) data and data about currently running apps (Android only)	iOS and Android	Students slept less during examination periods and more during breaks; they felt more stressed during the breaks and examination periods; sensor data were able to capture different routines during weekdays, weekends, and breaks.	N/A	PSQI, ESS[aq], MCTQ[ar], PROMIS[as]-10, BHM[at]-20, CD-RISC[au], Flourishing Scale, Perceived Stress Scale, BFI, PHQ-8, and UCLA Loneliness Scale	22 students
Mack et al [59], 2021	Understand the association between behavioral and mental health and the COVID-19 pandemic	GPS, accelerometer, phone lock and unlock, and light sensor data	iOS and Android	During the COVID-19 pandemic, students experienced more depression and anxiety and increased sedentary time and phone use, whereas sleep and the number of locations visited decreased.	N/A	PHQ-4 and EMAs	217 students
Xu et al [60], 2023	Evaluate the cross–data set generalizability of depression detection	GPS, accelerometer, phone lock and unlock, Bluetooth, Wi-Fi, call log, microphone, gyroscope, and light sensor data	iOS and Android	Individuals who experienced depression had shorter sleep duration, had more interrupted sleep, had more frequent phone locks and unlocks, spent more time at home, were more sedentary, had fewer physical activities, visited fewer uncommon places, and had more consistent mobility patterns.	A multitask learning model with the 1D-CNN[av]–based embedding, fully connected layers for reordering and classification.	Weekly surveys on self-reported depression symptoms and affect, BDI-II, and PHQ-4	534 students
Nepal et al [61], 2022	Explore the association between students' COVID-19 concerns and behavioral and mental health	GPS, accelerometer, phone lock and unlock, light sensor, and phone use data	iOS and Android	Heightened COVID-19 concerns correlated with increased depression, anxiety, and stress. No specific results relating sensors to symptoms of depression, anxiety, or stress were observed.	Evaluated different deep learning models in terms of their classification of COVID-19 concerns: CNN, InceptionTime, MCDCNN[aw], ResNet[ax], multilayer perceptron, TWIESN[ay], LSTM, and FCNN[az]; FCNN performed the best, with an AUROC[ba] score of 0.7.	Self-reported affect and PHQ-4	180 students
Currey and Torous [51], 2022	Predict survey results on mental health from passive sensors	GPS, accelerometer, call, and screen time data	iOS	Individuals at higher risks of psychosis spent less time at home. Individuals who were lonelier had longer sleep duration and fewer calls. Individuals who experienced stress or depression had longer outgoing calls.	Logistic regression was used to predict survey scores.	PHQ-9, GAD-7, PSS, UCLA Loneliness Scale, PQ-16, and PSQI	147 students

Study, year	Aim	Data collected	Operating system	Behavioral patterns	Predictive modeling	Verification surveys	Sample size, n
Currey et al [57], 2023	Explore the cross–data set generalizability of symptom improvement based on the surveys	GPS, accelerometer, and screen time data	iOS and Android	Logistic regression was able to predict changes in mood across 2 data sets of student participants. No results relating sensors to symptoms of depression or anxiety were observed.	Logistic regression was used to predict weekly score improvement from both active and passive features.	PHQ-9, GAD-7, PSS, UCLA Loneliness Scale, PSQI, PQ-16, and DWAI[bb]	698 students

[a]PHQ: Patient Health Questionnaire.
[b]N/A: not applicable.
[c]DASS: Depression Anxiety Stress Scales.
[d]SIAS: Social Interaction Anxiety Scale.
[e]GAD-7: Generalized Anxiety Disorder Scale-7.
[f]PQ: Prodromal Questionnaire.
[g]PSS: Perceived Stress Scale.
[h]PSQI: Pittsburgh Sleep Quality Index.
[i]BASIS: Behavior and Symptom Identification Scale.
[j]SF: Short Form Health Survey.
[k]SFS: Social Functioning Schedule Scale.
[l]CGI: Clinical Global Impressions Scale.
[m]HDRS: Hamilton Depression Rating Scale.
[n]CAS: Coronavirus Anxiety Scale.
[o]HAI: Health Anxiety Inventory.
[p]UCLA: University of California, Los Angeles.
[q]XGBoost: extreme gradient boosting.
[r]LOOCV: leave-one-out cross validation.
[s]PANAS: Positive and Negative Affect Schedule.
[t]MPSM: Mobile Photographic Stress Meter.
[u]LSTM: long short-term memory.
[v]CNN: convolutional neural network.
[w]WEMWBS: Warwick-Edinburgh Mental Well-Being Scale.
[x]BDI: Beck Depression Inventory.
[y]EMA: ecological momentary assessment.
[z]PAM: Patient Activation Measure.
[aa]BFI: Big Five Inventory.
[ab]LASSO: least absolute shrinkage and selection operator.
[ac]COMOSWB: Concise Measure of Subjective Well-Being.
[ad]SAS: Sport Anxiety Scale.
[ae]PPC: Perceived Personal Control.
[af]SSS: Social Support Scale.
[ag]MAAS: Mindful Attention Awareness Scale.
[ah]ERQ: Emotion Regulation Questionnaire.
[ai]BRS: Brief Resilience Scale.
[aj]CES-D: Center for Epidemiological Studies-Depression.
[ak]STAI: State Trait Anxiety Inventory.
[al]LOSOCV: leave-one-subject-out cross validation.
[am]AdaBoost: adaptive boosting.
[an]LightGBM: light gradient boosting machine.
[ao]MERF: mixed-effects random forest.
[ap]TIPI: Ten-Item Personality inventory.
[aq]ESS: Epworth Sleepiness Scale.
[ar]MCTQ: Munich Chronotype Questionnaire.
[as]PROMIS: Patient-Reported Outcomes Measurement Information System.
[at]BHM: Behavioral Health Measure.

[au]CD-RISC: Connor-Davidson Resilience Scale.
[av]1D-CNN: 1-dimensional convolutional neural network.
[aw]MCDCNN: multi-channel deep convolutional neural network.
[ax]ResNet: residual network.
[ay]TWIESN: time warping invariant echo state network.
[az]FCNN: fully convolutional neural network.
[ba]AUROC: area under the receiver operating characteristic curve.
[bb]DWAI: Digital Working Alliance Inventory.

Studies With Adults

Table 4 presents the data extracted from the studies conducted with the general adult population. The average study duration was 201.6 (SD 367) days. Apart from a 3-year longitudinal study with 18,000 participants, the average number of participants was 123.4 (SD 139.8). Of the 8 studies with adults, 2 (25%) [32,52] were conducted with the same set of participants. A total of 3 (38%) studies used predictive modeling, with regression-based models being the most common [34,36,52], and 1 (12%) study identified gender differences in behavioral patterns [27]. Overall, the research with adults showed that GPS, accelerometer, ambient audio, and illuminance data related to individuals' emotional state. Adults with depression were less likely to leave home and were less physically active, whereas adults who were socially anxious were more active and left their home more often but avoided going to places where they needed to socially interact.

Table 4. Summary of the reviewed studies with adult participants.

Study, year	Aim	Data collected	Operating system	Behavioral patterns	Predictive modeling	Verification surveys	Sample size, n
Di Matteo et al [32], 2021	Understand whether ambient speech correlates with social anxiety, generalized anxiety, and depressive symptoms	Microphone data	Android	Generalized anxiety and depression were correlated with reward-related words. Social anxiety was correlated with vision-related words.	N/A[a]	LSAS[b], GAD-7[c], PHQ[d]-8, and SDS[e]	86 Canadian adults
Di Matteo et al [52], 2021	Predict general anxiety disorder, social anxiety disorder, and depression	GPS, microphone, screen on and off, and light sensor data	Android	Depression and social anxiety were associated with increased screen use. Depression was associated with sleep disturbance and death-related word features.	A total of 3 logistic regression models were used to predict social anxiety disorder and generalized anxiety disorder with repeated k-fold cross validation.	LSAS, GAD-7, PHQ-8, and SDS	84 Canadian adults
Wen et al [34], 2021	Detect impulsive behavior, positivity, and stress	Call log, phone lock and unlock, and phone charging data	iOS and Android	Impulsivity was correlated with increased phone use and screen checking.	Used LASSO[f] regularization to first select features and trained a linear regression model to estimate trait impulsivity scores.	BIS[g]-15, UPPS[h], PAM[i], and self-reported feelings	26 adults
Fukazawa et al [36], 2019	Predict anxiety levels and stress	Light sensor, gyroscope, accelerometer, and app use data	Android	Anxiety was higher from Monday to Thursday than on Friday and Saturday. Increased anxiety was associated with decreased mobility. During mild exercise, anxiety was reduced.	Used linear classifier by LASSO and XGBoost[j] to classify the change of anxiety.	STAI[k]	20 adults
Pratap et al [55], 2017	Detect depression	GPS, call log, and SMS text message data	iOS and Android	None of the results related sensors to symptoms of depression.	N/A	PHQ-2 and PHQ-9	359 Hispanic or Latino adults
Adams et al [26], 2014	Detect stress level	Microphone data	iOS and Android	Stress can be recognized from pitch, speaking speed, and vocal energy.	N/A	PANAS[l], PSS[m]-14, MAAS[n], and self-reported affect	7 adults
Meyerhoff et al [48], 2021	Detect anxiety and depression	GPS, call log, app use, and SMS text message data	Android	Changes in the number of locations visited and social activity duration were associated with depression. Time spent at exercise locations was positively correlated with changes in depressive symptoms.	N/A	GAD-7, PHQ-8, and SPIN[o]	282 adults
Servia-Rodríguez et al [27], 2017	Predict mood	GPS, Wi-Fi, cell tower, accelerometer, microphone, SMS text message, and call data	Android	A strong correlation was identified between daily routines and users' personality, well-being perception, and other psychological variables; the participants who were the most emotionally stable tended to be more active, stayed in more noisy places, and texted less than participants who were unstable.	Used stacked RBMs[p] to classify moods.	Big-5 personality test, self-reported mood, and self-reports of locations	18,000 adults mainly

[a]N/A: not applicable.

[b]LSAS: Liebowitz Social Anxiety Scale.

[c]GAD-7: Generalized Anxiety Disorder Assessment-7.

[d]PHQ: Patient Health Questionnaire.
[e]SDS: Sheehan Disability Scale.
[f]LASSO: least absolute shrinkage and selection operator.
[g]BIS: Barratt Impulsiveness Scale.
[h]UPPS: Impulsive Behavior Scale.
[i]PAM: Patient Activation Measure.
[j]XGBoost: extreme gradient boosting.
[k]STAI: State Trait Anxiety Inventory.
[l]PANAS: Positive and Negative Affect Schedule.
[m]PSS: Perceived Stress Scale.
[n]MAAS: Mindful Attention Awareness Scale.
[o]SPIN: Social Phobia Inventory.
[p]RBM: Restricted Boltzmann Machine.

Studies With Employees

Table 5 presents the data extracted from the studies that were conducted with employees. Among the 4 studies with employees, 1 (25%) study recruited its own participants [56], and the other 3 (75%) studies [40-42] used the Tesserae data set [63]. Compared with students and adults, the employee population was the least studied, with the fewest articles. However, the studies with employees had the largest number of participants, with a mean of 427.3 (SD 280.3). All 4 studies used regression-based predictive modeling, and 2 (50%) of them [40,56] evaluated a variety of models, with logistic regression, support vector machine, and random forest being the most common methods. Detecting and predicting employees' stress in workplaces were examined in tandem with employees' work performance. The research goal for these studies was to understand the underlining reasons for lowered work-related productivity. In contrast to the other 2 populations (ie, students and adults), less mobility was seen as positive for employees because less mobility in workplaces was associated with more positivity and higher performance.

Table 5. Summary of the reviewed studies with employee participants.

Study, year	Aim	Data collected	Operating system	Behavioral patterns	Predictive modeling	Verification surveys	Sample size, n
Mirjafari et al [56], 2019	Predict stress and job performance	Accelerometer, GPS, phone lock and unlock, and light sensor data	iOS and Android	Higher performers unlocked their phone fewer times during evenings, had less physical activity, visited fewer locations on weekday evenings, were more mobile, and visited more locations during weekends.	Evaluated logistic regression, support vector machine, random forest, and XGBoost[a] in terms of employee performance classification; XGBoost was the best model with 5-fold cross validation.	ITP[b], IRB[c], OCB[d], and CWB[e]	554 employees
Nepal et al [40], 2020	Detect stress, well-being, and mood	GPS, phone lock and unlock, accelerometer, Bluetooth, and phone use data	iOS and Android	Promoted employees spent more time on their phones during early mornings and late evenings and had more unlocks during the night time than nonpromoted employees. Women's mobility increased after promotion, whereas men's mobility decreased.	Evaluated logistic regression, support vector machine, Gaussian naive Bayes, random forest, and k-nearest neighbor in terms of their classification between promoted and nonpromoted periods; the best model was logistic regression trained on ROCKET[f]-based features.	CWB, OCB, IRB, and ITP	141 employees
Saha et al [41], 2019	Predict stress and workplace performance	Light sensor, GPS, accelerometer, and phone lock and unlock data	iOS and Android	Stress was higher with increased role ambiguity.	Linear regression was used to predict a well-being score.	IRB, ITP, and OCB	257 employees
Morshed et al [42], 2019	Predict mood instability	Light sensor, GPS, accelerometer, and phone lock and unlock data	iOS and Android	Mood instability was negatively correlated with the duration of sleep, the number of conversations, the amount of activity, and outdoor mobility.	Ridge regression with regularization was used to infer a mood instability score.	EMAs[g], PAM[h], and PANAS[i]	757 employees

[a]XGBoost: extreme gradient boosting.

[b]ITP: Psychological Type Indicator.

[c]IRB: in-role behavior.

[d]OCB: organizational citizenship behavior.

[e]CWB: counterproductive work behavior.

[f]ROCKET: random convolutional kernel transform.

[g]EMA: ecological momentary assessment.

[h]PAM: Patient Activation Measure.

[i]PANAS: Positive and Negative Affect Schedule.

Passive Sensors

Overview

Table 6 provides an overview of the range of sensors used to detect patterns related to mild mental health symptoms and summarizes the evidence of the effectiveness of the various sensors. The first column lists the sensor, and the second column presents how the data from that sensor are interpreted; in other words, it presents the behavior-related information that the sensor data are intended to represent. The third column indicates which articles found significant associations between the specific sensor and stress, anxiety, or mild depression. The fourth column indicates which articles found no significant associations between the specific sensor and mental health outcomes (ie, explicitly stated so in the articles). In the subsequent sections, we discuss the types of activities detected by the sensors.

Table 6. Sensor summary of the reviewed studies.

Sensor	Behavior	Evidence for effectiveness	No evidence
GPS	Location and physical activity	[27,30,35,37,39-42,44-53,56-61]	[28,31,55]
Microphone	Voice recognition, ambient sound, and sleep	[26,27,32,39,42,43,45,48,50,52,60]	[28,41,47]
Light sensor	Time spent in darkness and sleep	[36,39,41,43,45,50,52,56,58-61]	[28,42,48]
Accelerometer	Movement and physical activity	[27,30,35-37,39-43,45-47,49-51,53,56-61]	[31,33]
Phone locks and unlocks	Phone use	[34,35,39,40,43,45-47,50,53,56,58-61]	[28,41,42]
Call logs	Social interaction and incoming and outgoing calls	[27,33,34,37,44-46,51,53,60]	[28,30,31,35,47,48,55]
Bluetooth	Social interaction	[40,42,43,46,47,60]	[28,51]
Wi-Fi	Indoor location	[27,38,42,47,60]	None
Keyboard	Typing patterns and muscle activity	[29]	[28]
SMS text messages and emails	Social interaction and incoming and outgoing messages	[27,32,33,37,44,45,52]	[28,30,31,48,55]
App use	Phone use and social media	[28,35,36,40,43,45,48,49,61]	None
Screen on and off	Phone use	[40,45,46,49,50,52,53,55,57]	[51]
Gyroscope	Orientation of the smartphone	[36,60]	None

Social Interaction: Call and Text Logs, Audio, Microphone, and Bluetooth

The social interaction of an individual is reflective of their current mood and mental state [44,64,65,66]. Individuals with depression and stress may be expected to decrease their social interactions. This is measured through the frequency of receiving texts and calls, how fast individuals respond, and the frequency of being around others. Among the 40 included studies, 18 (45%) [27,28,30,31,33-35,37,44-48,51,53,55,60] examined call logs to understand social interaction patterns, mainly through the number of incoming and outgoing calls, the number of missed calls, and the duration of calls. Individuals who experience depression and stress may engage in longer outgoing calls [51]. Evening communications were predictive of depression [47], anxiety, and loneliness [54]. Students who experienced discriminations [53] and anxious participants had more evening communications [54]. Metadata on SMS text messages were examined in 10 (25%) [27,28,30,31,33,37,44,45,48,55] of the 40 studies, including the frequency of receiving SMS text messages and the average time of responses. People who are socially anxious were found to take different amounts of time to respond to SMS text messages and calls [33]. Increases in the number of calls were associated with increased social anxiety [48]. Those who experienced social anxiety were less likely to call or text in public [44]. For students, fewer conversations were associated with more stress [39] and more mood instability [42]. One of the studies found that more emotionally unstable individuals tended to text more than emotionally stable individuals [27].

Location: GPS, Bluetooth, and Wi-Fi

Location data can provide insights into individuals' mental health state in terms of the normal or abnormal variety and frequency of locations visited [67]. As presented in Table 6, GPS has been one of the most commonly used passive sensors for stress, anxiety, and mild depression research. The findings regarding location consistently demonstrate that students and adults who experienced depression, anxiety, or stress tended to visit fewer places [39,44,50,58-60]. One of the studies [48] found that location data are highly inversely correlated with mild depression severity. The main way in which this is measured is through the frequency of exiting the house, the variety of locations visited, and mobility. The frequency of exiting the house is less for individuals who are depressed, and there is less variety in the visited locations for individuals who are socially anxious. Individuals who are feeling depressed often experience being less energetic [68,69]. Overall, negative emotions were associated with time spent at specific locations, but this is also affected by personal routines and preferences [30]. For students, stress and lower subjective well-being were associated with more time spent on campus [39,49] and less time spent at campus food locations [39]. Students who experienced depression spent more time at home [60], whereas individuals at higher risk of psychosis spent less time at home [51]. Time spent at exercise locations was positively correlated with changes in depressive symptoms [48]. Another study [38] distinguished between students experiencing severe stress and those with normal stress levels, revealing that students with severe stress spent significantly less time on campus and were less involved in work-related activities compared with their counterparts with normal stress levels. As for employees, higher performers were found to visit fewer locations on weekday evenings but more locations during weekends [56].

Voice Recognition: Audio

The microphone is used to measure audio data of speech and ambient noises. One of the studies [26] examined how people with stress speak by analyzing their voice, including the speed of speech, how energetic their vocality is, and the pitch. One caveat is that the study by Adams et al [26] used audio captured within laboratory environments and found that stress could be

recognized from the absence of speech. In variable environments, it will be harder to recognize the changing voice patterns. One study found that generalized anxiety and depression related to reward-related words in ambient speech, and social anxiety related to vision-related words [32]. Another study [52] identified that people with depression tend to speak less and use more death-related words.

Sleep: Accelerometer, Audio, and Illuminance

Sleep is highly correlated with individuals' mental state [26,35,36,42,45-47,59,60]. Among the 40 included studies, 5 (13%) [35,46,52,60] found that more disturbed sleep correlated with more depressive symptoms. However, occasional sleep disturbance is not necessarily predictive. For example, for those with social anxiety, sleep disturbance might be positive because it suggests night-time activity and social interactions. Metadata on the time spent in darkness can be indicative of sleep patterns. The study by Fukazawa et al [36] stated that anxiety levels increase when the time spent in darkness increases. The study by Di Matteo et al [52] found that individuals with symptoms related to social anxiety and depression spent less time in darker environments. Another study [39] stated that stress changed students' sleep patterns, where they became less likely to move around between 6 PM and midnight. Of the 40 studies, 6 (15%) found that shorter sleep duration was correlated with more mood instability [42], more depressive symptoms [59,60], and more stress [36,44]. One of the studies [45] also found that the student population, in general, tended to sleep less during examination periods and slept more during breaks, and they felt more stressed during both breaks and examination periods.

Phone Use: On and Off Screen, Lock and Unlock, and App Use

Today, smartphones are used for self-regulated "distractions," such as the use of social media [38]. This type of self-regulated distraction can temporarily reduce stress. The study by Chikersal et al [47] showed that depression can impact concentration levels, so if distraction by phone can be measured, this could be a potential predictive marker. Several studies found that increasing phone use was correlated with more depressive symptoms [46,47,50,52,58-60], anxiety [52,59], impulsivity [34] and lower subjective-wellbeing [49]. The study by Morshed et al [42] outlined that for postsemester depression, phone use at night is not predictive, whereas another study [47] summarized that phone use during the day is predictive of depression. More frequent phone locks or unlocks correlated with higher levels of depressive symptoms [60] and impulsivity [34]. Higher performing employees tended to unlock their phones less frequently in the evenings [56]. Additionally, individuals who were promoted spent more time on their phones during early mornings and late evenings, with more unlocks occurring during nighttime compared with their nonpromoted counterparts [40].

Physical Activity and Mobility: Accelerometer

According to Table 6, along with GPS, accelerometer is one of the most widely used passive sensors in digital phenotyping research to monitor participant's mobility, activity, and sedentary periods. Increased sedentary time was correlated with increased depressive symptoms [47,48,50,58-60], increased mood instability [27,42], increased stress [36] and decreased subjective well-being [49]. Exercise duration was positively correlated with changes in anxiety [36] and depressive symptoms [48]. The study by Mirjafari et al [56] found that the amount of movement and physical activity was related to employee's stress level and highlighted that if the activity is regular, it should reduce stress. Different occupations require different levels of physical activity, social interactions, and mobility. For instance, developers spend most of their time at their desks, and their tasks might require less social interaction and mobility at work, but this does not mean they are more stressed. Project managers have more mobility during the day, and this may be because they need to move around to meet with the stakeholders [56]. Several studies have observed variations in mobility and gait consistency. The study by Boukhechba et al [44] reported that individuals with high social anxiety exhibited a narrower range of activities, whereas the study by Xu et al [60] revealed that students experiencing depression demonstrated more consistent mobility patterns. Additionally, accelerometer data indicated that individuals with low social anxiety maintained a steady walking pace, whereas those with high social anxiety tended to walk more rapidly and with greater irregularity [33].

Muscle Activity: Keyboard

Stress can cause muscle tension [70,71]. One of the studies [29] collected the data of users with stress via a keyboard in a laboratory environment and found that typing pressure significantly increased under stressful conditions.

Challenges

Digital phenotyping for mild mental health symptoms in nonclinical participants can present ethical challenges, limitations to the research, and technical challenges. We review the challenges that were stated in the literature.

Ethical Challenges

Among the 40 included studies, 7 (18%) specifically mentioned privacy-related ethical concerns [28,31,35,36,40,41,43]. A major concern for participants across several studies was whether authorities, such as employers or teachers, will have access to their data. One of the studies [28] conducted in-depth interviews with 15 students to understand their perspectives on digital phenotyping through app prototypes. They found that the students' core concerns were whether the acquainted university staff had access to the data. They also found that students' acceptability of such apps depends on the perceived relevancy of the data collected and the effects on students' devices. The study by Nepal et al [40] with employees reported a similar privacy concern of whether the employees' data would be leaked to their boss; if the boss is aware of a potential mental health issue, it may impact their work performance ratings.

The methods of collecting and storing passive sensing data also present privacy concerns [28,70,72], particularly when the tracked data involve sensitive topics, such as mental health [72]. Sensors that infer individuals' social interactions provide insights into their mental health status [26,36,53]. However, these types of data were less likely to be shared by participants

because of privacy concerns. In the study by Rooksby et al [28], students identified camera, microphone, call log, and keyboard data as highly unacceptable types of data to capture.

Location data were associated with privacy and security concerns. In the study by Wen et al [34], participants felt uncomfortable with location tracking because it might breach their privacy and were hesitant to log their location when they moved from one place to another. Some studies excluded specific sensors to protect the participant's privacy. Location data were not recorded owing to security concerns, even though they could provide valuable insights into the mental state [36,38]. In the study by Adams et al [26], the microphone was disabled to capture calls and conversations while individuals were talking to their family members. Another ethical concern was regarding the misuse of data. The main focus in studies of digital phenotyping using smartphones was on tracking participants' usual behavioral patterns and identifying whether they behaved unusually. There were concerns regarding secondary uses. For example, participants' leaked data can be used for advertising purposes or to create content [34,41].

Limitations to the Research

Coping mechanisms related to stress and anxiety vary among individuals [22]. Individual differences can make it challenging to label individuals as stressed, anxious, or depressed, particularly nonclinical participants. Certain behavioral patterns can be generally expected; however, not all individuals will follow the same pattern. To make generalizable and powerful analyses and understand behavioral patterns associated with mild mental health concerns, it is recommended to study diverse groups for longer than a 2-week period. Of the 40 included studies, 2 (5%) [33,39] focused on a particular demographic subset, namely, undergraduate students. Therefore, the generalizability of the studies is limited. In the studies by Rooksby et al [28], Exposito et al [29], and Wang et al [50], limited variation in representation was seen as a major limitation. The studies by Rhim et al [49], DaSilva et al [39], and Fukazawa et al [36] stressed the importance of selecting a wider age group, as younger people use their smartphones proactively, whereas older people's behavioral patterns might show differences when they are experiencing mild mental health symptoms. The study by Nepal et al [40] suggested that diverse population testing is required for more reliable results, considering interindividual differences. Furthermore, the accuracy and effectiveness of machine learning models are highly affected by data set quality. We noticed that over the last 4 years [38,46,57,60], there has been increased focus on the generalizability of machine learning models, with the goal of assessing generalizability across students from various years, classes, and institutions.

Technical Challenges

Digital phenotyping studies on mild symptoms related to mental health with nonclinical participants presented technical challenges. A main concern was the accuracy of the sensor data collected from smartphones. The study by Fukazawa et al [36] sought to understand the time spent in darkness and its effects on the relationship between stress and anxiety patterns and sleep. However, when individuals carried their smartphone in their pockets or bags, the smartphone could not detect the darkness of the environment. This presented a challenge because illuminance data were captured even when the phone was not used actively. Similar concerns were raised in the study by Di Matteo et al [52]. The time spent in darkness feature did not distinguish whether the device was in a dark room or a dark location (ie, in the pocket). The study by Melcher et al [35] stated that the captured accelerometer data may not accurately represent daily activity, as not all participants constantly carried their phones throughout the day. In the study by Di Matteo et al [32], environmental audio did not produce clear transcripts in louder environments. This study mentioned that transcripts were produced based on dictionaries, so language analysis of complex speech, such as metaphors and sarcasm, was ignored. Therefore, the entire content of the conversation might not be correctly interpreted. In the study by Di Matteo et al [52], similar challenges were identified, as the speech data produced from smartphones were not clear. The recorded voices of the participants were masked by those of the people around them or even sound from other sources such as television or radio. Moreover, it was not possible to identify whether the death-related words came from the participants or from the people they interacted with.

Another technical challenge identified was battery life [47]. As expected, moment-by-moment data collection requires high power use, which might shorten the battery life. Participants had to charge their phones more often, which was inconvenient, and altered their usual behavior because they could not carry their phones as usual when the phones were charging. The study by Chikersal et al [47] mentioned another technical limitation: the transfer rate was affected if the app stopped working randomly. During these times, data were not transferred or collected. With the increase in the use of 5G technology, Wi-Fi data for indoor locations may cease to be relevant. In the study by Zakaria et al [38], some users were on their 5G indoors rather than their Wi-Fi, and this may point to a future trend of the use of 5G. We now turn to the discussion.

Discussion

Principal Findings

This literature review examined digital phenotyping studies that detected and predicted stress, anxiety, and depression in their mild states in nonclinical populations using data collected from smartphones. The primary objective of digital phenotyping in the context of mild mental health was similar among the 3 participant cohorts: students, adults, and employees. However, notable distinctions emerged among these groups. Among university students, the geographical proximity and relevance of the university campus were discerned as influential factors. Moreover, academic pursuits, particularly coursework and study-related activities, assumed significance within this demographic. Conversely, among employees, work aspects held salience, accompanied by the workplace environment. The remaining studies encompassed a general population cohort, delineated by undisclosed characteristics. Overall, we found that identifying behavioral abnormalities related to stress and anxiety was possible but raised certain challenges. Generalized

stress and anxiety symptoms vary largely among individuals, whereas serious diagnoses, such as bipolar disorder or schizophrenia, have well-documented behavioral changes. Sleep was a strong predictor variable, yet some individuals tended to sleep more while they were stressed, whereas others lacked sleep under stress. This may be one of the reasons why there are fewer studies and reviews completed on stress and anxiety compared with studies on serious conditions such as bipolar disorder, severe depression, and schizophrenia. Another reason is that clinical psychologists and psychiatrists who are familiar with clinical populations are leading the digital phenotyping research.

Studies tended to use self-report to categorize nonclinical populations as stressed, depressed, or anxious. It was not always clear whether the identified patterns of the passive sensor data would effectively discriminate among groups. Most studies used prestudy and poststudy surveys to identify participants' mental state. There were concerns raised regarding the accuracy of the categorization of self-report surveys. For instance, the study by Sefidgar et al [53] stated that students with stress may not report themselves as very stressed. Melcher et al [70] conducted a review and found that students were concerned regarding their professors learning about their data [71]. Thus, the accuracy of self-report remains an issue for passive sensing studies that use self-report labels, especially when there are privacy concerns. This may be related to the high dropout rates in the studies.

Many types of data sensors were used in the reviewed studies. Few articles related sensor patterns to specific symptoms validated by relevant psychological evidence. One of the studies [46] extracted interpretable rules (such as intermittent sleep episodes or number of bouts of being asleep or number of outgoing calls during weekends) through association rule mining to distinguish the behavioral patterns between students who were depressed and students who were not. However, although the behavioral patterns were identified, they were not validated to be exclusive to the addressed mental health issue; for example, high mobility and physical activity do not necessarily mean that the person is not stressed. In the study by Tseng et al [45], students were more mobile during the examination week, despite being under high pressure and stress. In the same study, some students were less mobile when studying for their examinations, which we cannot necessarily be interpreted as being under stress. Of the 40 included studies, 4 (10%) [35,58,59,61] explored the effects of the COVID-19 pandemic on behavioral and mental health. Additional recent investigations, which independently gathered their own data sets during the COVID-19 pandemic, have shown that quarantine measures have influenced individual behavioral patterns. For the purpose of making precise predictions in digital phenotyping, it is imperative to consider contextual and environmental factors.

Privacy and secondary data uses were the main concerns identified for digital phenotyping. Individuals using digital phenotyping systems have the right to provide informed consent. This means that they should be made aware of how all their data will be used, who will have access to their data, where their data will be stored, and for how long their data will be stored, and they have the right to decline to participate. We urge researchers and medical practitioners to carefully consider the system design and requirements because data transferred to the cloud and other services may fall under various service agreements. To empower end users and improve the quality of digital phenotyping systems, we recommend that transparent algorithms and explainable artificial intelligence be combined with user-accessible and understandable displays so that adults can engage in the process of identifying and categorizing patterns related to mild mental health symptoms.

The digital phenotyping research focused on in this review may enable the design of tailored intervention programs for nonclinical participants who are showing symptoms of stress, anxiety, and mild depression. Most of the studies included in this review were conducted within a restricted timeline and limited scope of detection and prediction. Only 4 (10%) of the 40 studies mentioned potential intervention programs upon predicting stress, anxiety, and mild depression [31,38,47,53].

Our review has some limitations. We excluded studies conducted with teenagers, children, and adults who were clinically diagnosed. Thus, we missed studies that focused on the detection and prediction of stress, anxiety, and mild depression in these populations. These populations are likely to show different patterns than those in adults who are not clinically diagnosed. Further, we excluded studies conducted using technologies other than smartphones. We chose this more limited subset of technologies to scope findings related to widely available technologies. The availability of technologies is changing rapidly, and wearables such as smartwatches are becoming more common. As wearable technologies become ubiquitous, we recommend including them in future systematic reviews.

This literature review is unique in that it examines studies focused on the behavioral patterns of nonclinical populations, namely students, employees, and adults who are stressed, anxious, or mildly depressed. We examined each type of sensor and indicated when it was significantly associated with mild mental health symptoms. We identified commonalities in the studies in terms of ethical challenges, limitations to the research, and technical challenges.

Conclusions

This systematic literature review found that digital phenotyping can be an effective way of identifying certain behavioral patterns related to stress, anxiety, and mild depression. A range of passive sensors was used in the studies, such as GPS, Bluetooth, ambient audio, light sensors, accelerometers, microphones, illuminance, and Wi-Fi. We found that location, physical activity, and social interaction data were highly related to participants' mental health and well-being. The surveyed literature discussed the ethical and technical challenges that limit the accuracy and generalizability of results. One of the greatest challenges was privacy concerns, and these were primarily related to camera, location, SMS text message, and call log data. Another challenge was the significant variation among individuals and their unique behaviors related to mental health. Finally, technical limitations have not been fully resolved, with issues such as the sensor for illuminance still capturing data while not in use reducing the accuracy of the

collected data. It is hoped that this overview of digital phenotyping and mental health studies conducted in the last decade, including the common privacy and technical concerns, can help move this area of research forward, ultimately improving the quality of passive sensing, and provide benefits in terms of the early detection of relevant mild mental health phenomena.

Acknowledgments

The authors would like to thank the School of Computer Science at the University of Auckland for their financial support.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The PRISMA 2020 checklist.
[DOCX File , 31 KB - mhealth_v12i1e40689_app1.docx]

References

1. Onnela JP, Rauch SL. Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. Neuropsychopharmacology 2016 Jun;41(7):1691-1696 [FREE Full text] [doi: 10.1038/npp.2016.7] [Medline: 26818126]
2. Houts CR, Patrick-Lake B, Clay I, Wirth RJ. The path forward for digital measures: suppressing the desire to compare apples and pineapples. Digit Biomark 2020 Nov 26;4(Suppl 1):3-12 [FREE Full text] [doi: 10.1159/000511586] [Medline: 33442577]
3. Huckvale K, Venkatesh S, Christensen H. Toward clinical digital phenotyping: a timely opportunity to consider purpose, quality, and safety. NPJ Digit Med 2019 Sep 06;2:88 [FREE Full text] [doi: 10.1038/s41746-019-0166-1] [Medline: 31508498]
4. Berry JD, Paganoni S, Carlson K, Burke K, Weber H, Staples P, et al. Design and results of a smartphone-based digital phenotyping study to quantify ALS progression. Ann Clin Transl Neurol 2019 Apr 3;6(5):873-881 [FREE Full text] [doi: 10.1002/acn3.770] [Medline: 31139685]
5. Trifan A, Oliveira M, Oliveira JL. Passive sensing of health outcomes through smartphones: systematic review of current solutions and possible limitations. JMIR Mhealth Uhealth 2019 Aug 23;7(8):e12649 [FREE Full text] [doi: 10.2196/12649] [Medline: 31444874]
6. Torous J, Gershon A, Hays R, Onnela JP, Baker JT. Digital phenotyping for the busy psychiatrist: clinical implications and relevance. Psychiatr Ann 2019;49(5):196-201 [FREE Full text] [doi: 10.3928/00485713-20190417-01]
7. Maharjan SM, Poudyal A, van Heerden A, Byanjankar P, Thapa A, Islam C, et al. Passive sensing on mobile devices to improve mental health services with adolescent and young mothers in low-resource settings: the role of families in feasibility and acceptability. BMC Med Inform Decis Mak 2021 Apr 07;21(1):117 [FREE Full text] [doi: 10.1186/s12911-021-01473-2] [Medline: 33827552]
8. Adler DA, Ben-Zeev D, Tseng VW, Kane JM, Brian RE, Campbell AT, et al. Predicting early warning signs of psychotic relapse from passive sensing data: an approach using encoder-decoder neural networks. JMIR Mhealth Uhealth 2020 Aug 31;8(8):e19962 [FREE Full text] [doi: 10.2196/19962] [Medline: 32865506]
9. Potier R. The digital phenotyping project: a psychoanalytical and network theory perspective. Front Psychol 2020 Jul 15;11:1218 [FREE Full text] [doi: 10.3389/fpsyg.2020.01218] [Medline: 32760307]
10. Kokel D, Rennekamp AJ, Shah AH, Liebel U, Peterson RT. Behavioral barcoding in the cloud: embracing data-intensive digital phenotyping in neuropharmacology. Trends Biotechnol 2012 Aug;30(8):421-425 [FREE Full text] [doi: 10.1016/j.tibtech.2012.05.001] [Medline: 22652049]
11. Teo JX, Davila S, Yang C, Hii AA, Pua CJ, Yap J, et al. Digital phenotyping by consumer wearables identifies sleep-associated markers of cardiovascular disease risk and biological aging. Commun Biol 2019 Oct 4;2:361 [FREE Full text] [doi: 10.1038/s42003-019-0605-1] [Medline: 31602410]
12. Fagherazzi G. Deep digital phenotyping and digital twins for precision health: time to dig deeper. J Med Internet Res 2020 Mar 03;22(3):e16770 [FREE Full text] [doi: 10.2196/16770] [Medline: 32130138]
13. Mercier HW, Hamner JW, Torous J, Onnela JP, Taylor JA. Digital phenotyping to quantify psychosocial well-being trajectories after spinal cord injury. Am J Phys Med Rehabil 2020 Dec;99(12):1138-1144 [FREE Full text] [doi: 10.1097/PHM.0000000000001506] [Medline: 32576743]
14. Rodriguez-Villa E, Rauseo-Ricupero N, Camacho E, Wisniewski H, Keshavan M, Torous J. The digital clinic: implementing technology and augmenting care for mental health. Gen Hosp Psychiatry 2020;66:59-66 [FREE Full text] [doi: 10.1016/j.genhosppsych.2020.06.009] [Medline: 32688094]

15. Barnett I, Torous J, Staples P, Sandoval L, Keshavan M, Onnela JP. Relapse prediction in schizophrenia through digital phenotyping: a pilot study. Neuropsychopharmacology 2018 Jul;43(8):1660-1666 [FREE Full text] [doi: 10.1038/s41386-018-0030-z] [Medline: 29511333]
16. Orsolini L, Fiorani M, Volpe U. Digital phenotyping in bipolar disorder: which integration with clinical endophenotypes and biomarkers? Int J Mol Sci 2020 Oct 16;21(20):7684 [FREE Full text] [doi: 10.3390/ijms21207684] [Medline: 33081393]
17. Kleiman EM, Turner BJ, Fedor S, Beale EE, Picard RW, Huffman JC, et al. Digital phenotyping of suicidal thoughts. Depress Anxiety 2018 Jul;35(7):601-608 [FREE Full text] [doi: 10.1002/da.22730] [Medline: 29637663]
18. Onnela JP. Opportunities and challenges in the collection and analysis of digital phenotyping data. Neuropsychopharmacology 2021 Jan;46(1):45-54 [FREE Full text] [doi: 10.1038/s41386-020-0771-3] [Medline: 32679583]
19. Lardier DT, Lee CY, Rodas JM, Garcia-Reid P, Reid RJ. The effect of perceived college-related stress on depression, life satisfaction, and school satisfaction: the coping strategies of hispanic college students from a hispanic serving institution. Educ Urban Soc 2020 Jan 02;52(8):1204-1222. [doi: 10.1177/0013124519896845]
20. Rezaei A, Mousanezhad Jeddi E. Relationship between wisdom, perceived control of internal states, perceived stress, social intelligence, information processing styles and life satisfaction among college students. Curr Psychol 2018 Feb 17;39:927-933 [FREE Full text] [doi: 10.1007/s12144-018-9804-z]
21. Futo J. Dealing with mental health issues on campus starts with early recognition and intervention. Campus Law Enforc J 2011 Jun;41(3):22 [FREE Full text]
22. Jonason PK, Talbot D, Cunningham ML, Chonody J. Higher-order coping strategies: who uses them and what outcomes are linked to them. Pers Individ Differ 2020 Mar 1;155:109755 [FREE Full text] [doi: 10.1016/j.paid.2019.109755]
23. Jorm AF. Mental health literacy: empowering the community to take action for better mental health. Am Psychol 2012 Apr;67(3):231-243 [FREE Full text] [doi: 10.1037/a0025957] [Medline: 22040221]
24. Number of smartphone mobile network subscriptions worldwide from 2016 to 2022, with forecasts from 2023 to 2028. Statista. URL: https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/ [accessed 2023-11-04]
25. PRISMA flow diagram. PRISMA. URL: http://www.prisma-statement.org/PRISMAStatement/FlowDiagram [accessed 2023-11-04]
26. Adams P, Rabbi M, Rahman T, Matthews M, Voida A, Gay G, et al. Towards personal stress informatics: comparing minimally invasive techniques for measuring daily stress in the wild. In: Proceedings of the 8th International Conference on Pervasive Computing Technologies for Healthcare. 2014 Presented at: 8th International Conference on Pervasive Computing Technologies for Healthcare; May 20-23, 2014; Oldenburg, Germany. [doi: 10.4108/icst.pervasivehealth.2014.254959]
27. Servia-Rodríguez S, Rachuri KK, Mascolo C, Rentfrow PJ, Lathia N, Sandstrom GM. Mobile sensing at the service of mental well-being: a large-scale longitudinal study. In: Proceedings of the 26th International Conference on World Wide Web. 2017 Presented at: WWW '17: 26th International World Wide Web Conference; April 3-7, 2017; Perth, Australia. [doi: 10.1145/3038912.3052618]
28. Rooksby J, Morrison A, Murray-Rust D. Student perspectives on digital phenotyping: the acceptability of using smartphone data to assess mental health. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 2019 Presented at: CHI '19: CHI Conference on Human Factors in Computing Systems; May 4-9, 2019; Glasgow, UK. [doi: 10.1145/3290605.3300655]
29. Exposito M, Hernandez J, Picard RW. Affective keys: towards unobtrusive stress sensing of smartphone users. In: Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct. 2018 Presented at: MobileHCI '18: 20th International Conference on Human-Computer Interaction with Mobile Devices and Services; September 3-6, 2018; Barcelona, Spain. [doi: 10.1145/3236112.3236132]
30. Cai L, Boukhechba M, Wu C, Chow PI, Teachman BA, Barnes LE, et al. State affect recognition using smartphone sensing data. In: Proceedings of the 2018 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies. 2018 Presented at: CHASE '18: ACM/IEEE International Conference on Connected Health: Applications, Systems and Engineering Technologies; September 26-28, 2018; Washington, DC. [doi: 10.1145/3278576.3284386]
31. Boukhechba M, Cai L, Chow PI, Fua K, Gerber MS, Teachman BA, et al. Contextual analysis to understand compliance with smartphone-based ecological momentary assessment. In: Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare. 2018 Presented at: PervasiveHealth '18: 12th EAI International Conference on Pervasive Computing Technologies for Healthcare; May 21-24, 2018; New York, NY. [doi: 10.1145/3240925.3240967]
32. Di Matteo D, Wang W, Fotinos K, Lokuge S, Yu J, Sternat T, et al. Smartphone-detected ambient speech and self-reported measures of anxiety and depression: exploratory observational study. JMIR Form Res 2021 Jan 29;5(1):e22723 [FREE Full text] [doi: 10.2196/22723] [Medline: 33512325]
33. Jacobson NC, Summers B, Wilhelm S. Digital biomarkers of social anxiety severity: digital phenotyping using passive smartphone sensors. J Med Internet Res 2020 May 29;22(5):e16875 [FREE Full text] [doi: 10.2196/16875] [Medline: 32348284]

34. Wen H, Sobolev M, Vitale R, Kizer J, Pollak JP, Muench F, et al. mPulse mobile sensing model for passive detection of impulsive behavior: exploratory prediction study. JMIR Ment Health 2021 Jan 27;8(1):e25019 [FREE Full text] [doi: 10.2196/25019] [Medline: 33502330]
35. Melcher J, Lavoie J, Hays R, D'Mello R, Rauseo-Ricupero N, Camacho E, et al. Digital phenotyping of student mental health during COVID-19: an observational study of 100 college students. J Am Coll Health 2023 Apr;71(3):736-748 [FREE Full text] [doi: 10.1080/07448481.2021.1905650] [Medline: 33769927]
36. Fukazawa Y, Ito T, Okimura T, Yamashita Y, Maeda T, Ota J. Predicting anxiety state using smartphone-based passive sensing. J Biomed Inform 2019 May;93:103151 [FREE Full text] [doi: 10.1016/j.jbi.2019.103151] [Medline: 30880254]
37. Rashid H, Mendu S, Daniel KE, Beltzer ML, Teachman BA, Boukhechba M, et al. Predicting subjective measures of social anxiety from sparsely collected mobile sensor data. Proc ACM Interact Mob Wearable Ubiquitous Technol 2020 Sep 04;4(3):1-24 [FREE Full text] [doi: 10.1145/3411823]
38. Zakaria C, Balan R, Lee Y. StressMon: scalable detection of perceived stress and depression using passive sensing of changes in work routines and group interactions. Proc ACM Hum Comput Interact 2019 Nov 07;3(CSCW):1-29 [FREE Full text] [doi: 10.1145/3359139]
39. DaSilva AW, Huckins JF, Wang R, Wang W, Wagner DD, Campbell AT. Correlates of stress in the college environment uncovered by the application of penalized generalized estimating equations to mobile sensing data. JMIR Mhealth Uhealth 2019 Mar 19;7(3):e12084 [FREE Full text] [doi: 10.2196/12084] [Medline: 30888327]
40. Nepal S, Mirjafari S, Martinez GJ, Audia P, Striegel A, Campbell AT. Detecting job promotion in information workers using mobile sensing. Proc ACM Interact Mob Wearable Ubiquitous Technol 2020 Sep 04;4(3):1-28 [FREE Full text] [doi: 10.1145/3414118]
41. Saha K, Reddy MD, Mattingly S, Moskal E, Sirigiri A, De Choudhury M. LibRA: on LinkedIn based role ambiguity and its relationship with wellbeing and job performance. Proc ACM Hum Comput Interact 2019 Nov 07;3(CSCW):1-30 [FREE Full text] [doi: 10.1145/3359239]
42. Morshed MB, Saha K, Li R, D'Mello SK, De Choudhury M, Abowd GD, et al. Prediction of mood instability with passive sensing. Proc ACM Interact Mob Wearable Ubiquitous Technol 2019 Sep 09;3(3):1-21 [FREE Full text] [doi: 10.1145/3351233]
43. Acikmese Y, Alptekin SE. Prediction of stress levels with LSTM and passive mobile sensors. Procedia Comput Sci 2019;159:658-667 [FREE Full text] [doi: 10.1016/j.procs.2019.09.221]
44. Boukhechba M, Huang Y, Chow P, Fua K, Teachman BA, Barnes LE. Monitoring social anxiety from mobility and communication patterns. In: Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers. 2017 Presented at: UbiComp '17: The 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing; September 11-15, 2017; Maui, Hawaii. [doi: 10.1145/3123024.3125607]
45. Tseng VW, Merrill M, Wittleder F, Abdullah S, Aung MH, Choudhury T. Assessing mental health issues on college campuses: preliminary findings from a pilot study. In: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct. 2016 Presented at: UbiComp '16: The 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing; September 12-16, 2016; Heidelberg, Germany. [doi: 10.1145/2968219.2968308]
46. Xu X, Chikersal P, Doryab A, Villalba DK, Dutcher JM, Tumminia MJ, et al. Leveraging routine behavior and contextually-filtered features for depression detection among college students. Proc ACM Interact Mob Wearable Ubiquitous Technol 2019 Sep 09;3(3):1-33 [FREE Full text] [doi: 10.1145/3351274]
47. Chikersal P, Doryab A, Tumminia M, Villalba DK, Dutcher JM, Liu X, et al. Detecting depression and predicting its onset using longitudinal symptoms captured by passive sensing: a machine learning approach with robust feature selection. ACM Trans Comput Hum Interact 2021 Jan 20;28(1):1-41 [FREE Full text] [doi: 10.1145/3422821]
48. Meyerhoff J, Liu T, Kording KP, Ungar LH, Kaiser SM, Karr CJ, et al. Evaluation of changes in depression, anxiety, and social anxiety using smartphone sensor features: longitudinal cohort study. J Med Internet Res 2021 Sep 03;23(9):e22844 [FREE Full text] [doi: 10.2196/22844] [Medline: 34477562]
49. Rhim S, Lee U, Han K. Tracking and modeling subjective well-being using smartphone-based digital phenotype. In: Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization. 2020 Presented at: UMAP '20; July 14-17, 2020; Genoa, Italy. [doi: 10.1145/3340631.3394855]
50. Wang R, Wang W, daSilva A, Huckins JF, Kelley WM, Heatherton TF, et al. Tracking depression dynamics in college students using mobile phone and wearable sensing. Proc ACM Interact Mob Wearable Ubiquitous Technol 2018 Mar 26;2(1):1-26 [FREE Full text] [doi: 10.1145/3191775]
51. Currey D, Torous J. Digital phenotyping correlations in larger mental health samples: analysis and replication. BJPsych Open 2022 Jun 03;8(4):e106 [FREE Full text] [doi: 10.1192/bjo.2022.507] [Medline: 35657687]
52. Di Matteo D, Fotinos K, Lokuge S, Mason G, Sternat T, Katzman MA, et al. Automated screening for social anxiety, generalized anxiety, and depression from objective smartphone-collected data: cross-sectional study. J Med Internet Res 2021 Aug 13;23(8):e28918 [FREE Full text] [doi: 10.2196/28918] [Medline: 34397386]

53. Sefidgar YS, Seo W, Kuehn KS, Althoff T, Browning A, Riskin E, et al. Passively-sensed behavioral correlates of discrimination events in college students. Proc ACM Hum Comput Interact 2019 Nov;3(CSCW):1-29 [FREE Full text] [doi: 10.1145/3359216] [Medline: 34322658]
54. Mendu S, Baglione A, Baee S, Wu C, Ng B, Shaked A, et al. A framework for understanding the relationship between social media discourse and mental health. Proc ACM Hum Comput Interact 2020 Oct 15;4(CSCW2):1-23 [FREE Full text] [doi: 10.1145/3415215]
55. Pratap A, Anguera JA, Renn BN, Neto EC, Volponi J, Mooney SD, et al. The feasibility of using smartphones to assess and remediate depression in Hispanic/Latino individuals nationally. In: Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers. 2017 Presented at: UbiComp '17: The 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing; September 11-15, 2017; Maui, Hawaii. [doi: 10.1145/3123024.3127877]
56. Mirjafari S, Masaba K, Grover T, Wang W, Audia P, Campbell AT, et al. Differentiating higher and lower job performers in the workplace using mobile sensing. Proc ACM Interact Mob Wearable Ubiquitous Technol 2019 Jun 21;3(2):1-24 [FREE Full text] [doi: 10.1145/3328908]
57. Currey D, Hays R, Torous J. Digital phenotyping models of symptom improvement in college mental health: generalizability across two cohorts. J Technol Behav Sci 2023 Mar 02:1-14 [FREE Full text] [doi: 10.1007/s41347-023-00310-9] [Medline: 37362062]
58. Huckins JF, daSilva AW, Wang W, Hedlund E, Rogers C, Nepal SK, et al. Mental health and behavior of college students during the early phases of the COVID-19 pandemic: longitudinal smartphone and ecological momentary assessment study. J Med Internet Res 2020 Jun 17;22(6):e20185 [FREE Full text] [doi: 10.2196/20185] [Medline: 32519963]
59. Mack DL, DaSilva AW, Rogers C, Hedlund E, Murphy EI, Vojdanovski V, et al. Mental health and behavior of college students during the COVID-19 pandemic: longitudinal mobile smartphone and ecological momentary assessment study, part II. J Med Internet Res 2021 Jun 04;23(6):e28892 [FREE Full text] [doi: 10.2196/28892] [Medline: 33900935]
60. Xu X, Liu X, Zhang H, Wang W, Nepal S, Sefidgar Y, et al. GLOBEM: cross-dataset generalization of longitudinal human behavior modeling. Proc ACM Interact Mob Wearable Ubiquitous Technol 2023 Jan 11;6(4):1-34 [FREE Full text] [doi: 10.1145/3569485]
61. Nepal S, Wang W, Vojdanovski V, Huckins JF, daSilva A, Meyer M, et al. COVID student study: a year in the life of college students during the COVID-19 pandemic through the lens of mobile phone sensing. In: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. 2022 Presented at: CHI '22: CHI Conference on Human Factors in Computing Systems; April 29-May 5, 2022; New Orleans, LA. [doi: 10.1145/3491102.3502043]
62. Wang R, Chen F, Chen Z, Li T, Tignor S, Zhou X, et al. StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones. In: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2014 Presented at: UbiComp '14: The 2014 ACM Conference on Ubiquitous Computing; September 13-17, 2014; Seattle, WA. [doi: 10.1145/2632048.2632054]
63. Mattingly SM, Gregg JM, Audia P, Bayraktaroglu AE, Campbell AT, Chawla NV, et al. The tesserae project: large-scale, longitudinal, in situ, multimodal sensing of information workers. In: Proceedings of the Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems. 2019 Presented at: CHI '19: CHI Conference on Human Factors in Computing Systems; May 4-9, 2019; Glasgow, UK. [doi: 10.1145/3290607.3299041]
64. Teles A, Barros F, Rodrigues I, Barbosa A, Silva F, Coutinho L, et al. Internet of things applied to mental health: concepts, applications, and perspectives. In: Gupta N, Paiva S, editors. IoT and ICT for Healthcare Applications. Cham, Switzerland: Springer; Aug 13, 2020.
65. Jin J, Gao B, Yang S, Zhao B, Luo L, Woo WL. Attention-block deep learning based features fusion in wearable social sensor for mental wellbeing evaluations. IEEE Access 2020 May 13;8:89258-89268 [FREE Full text] [doi: 10.1109/ACCESS.2020.2994124]
66. Nyqvist F, Forsman AK, Giuntoli G, Cattan M. Social capital as a resource for mental well-being in older people: a systematic review. Aging Ment Health 2013;17(4):394-410 [FREE Full text] [doi: 10.1080/13607863.2012.742490] [Medline: 23186534]
67. Diagnostic And Statistical Manual Of Mental Disorders, Fifth Edition, Text Revision (DSM-5-TR). Washington, DC: American Psychiatric Association; 2013.
68. Park SC, Kim D. The centrality of depression and anxiety symptoms in major depressive disorder determined using a network analysis. J Affect Disord 2020 Jun 15;271:19-26 [FREE Full text] [doi: 10.1016/j.jad.2020.03.078] [Medline: 32312693]
69. Peltz JS, Bodenlos JS, Kingery JN, Rogge RD. The role of financial strain in college students' work hours, sleep, and mental health. J Am Coll Health 2021;69(6):577-584 [FREE Full text] [doi: 10.1080/07448481.2019.1705306] [Medline: 31940259]
70. Melcher J, Hays R, Torous J. Digital phenotyping for mental health of college students: a clinical review. Evid Based Ment Health 2020 Nov;23(4):161-166 [FREE Full text] [doi: 10.1136/ebmental-2020-300180] [Medline: 32998937]
71. Cosgrove L, Karter JM, McGinley M, Morrill Z. Digital phenotyping and digital psychotropic drugs: mental health surveillance tools that threaten human rights. Health Hum Rights 2020 Dec;22(2):33-39 [FREE Full text] [Medline: 33390690]

72. Zacher H, Rudolph CW. Individual differences and changes in subjective wellbeing during the early stages of the COVID-19 pandemic. Am Psychol 2021 Jan;76(1):50-62 [FREE Full text] [doi: 10.1037/amp0000702] [Medline: 32700938]

Abbreviations

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by L Buis; submitted 01.07.22; peer-reviewed by J Rooksby, KS Sahu, A Joseph; comments to author 09.08.22; revised version received 03.10.22; accepted 27.09.23; published 23.05.24.

Please cite as:
Choi A, Ooi A, Lottridge D
Digital Phenotyping for Stress, Anxiety, and Mild Depression: Systematic Literature Review
JMIR Mhealth Uhealth 2024;12:e40689
URL: https://mhealth.jmir.org/2024/1/e40689
doi:10.2196/40689
PMID:38780995

Review

Technology-Based Music Interventions to Reduce Anxiety and Pain Among Patients Undergoing Surgery or Procedures: Systematic Review of the Literature

Sunghee Park[1,2], PhD; Sohye Lee[3], PhD; Sheri Howard[3], PhD; Jeeseon Yi[4], MPH, PhD

[1]Nursing Department, Ajou University Hospital, Suwon, Republic of Korea

[2]College of Nursing, Ajou University, Suwon, Republic of Korea

[3]Loewenberg College of Nursing, The University of Memphis, Memphis, TN, United States

[4]College of Nursing & Sustainable Health Research Institute, Gyeongsang National University, Jinju, Republic of Korea

Corresponding Author:
Jeeseon Yi, MPH, PhD
College of Nursing & Sustainable Health Research Institute
Gyeongsang National University
816-15 Jinjudae-ro
Jinju, 52727
Republic of Korea
Phone: 82 55 772 8252
Fax: 82 55 772 8222
Email: jy772@gnu.ac.kr

Abstract

Background: Hospitalized patients undergoing surgery or procedures may experience negative symptoms. Music is a nonpharmacological complementary approach and is used as an intervention to reduce anxiety, stress, and pain in these patients. Recently, music has been used conveniently in clinical situations with technology devices, and the mode of providing music is an important factor in technology-based music interventions. However, many reviews have focused only on the effectiveness of music interventions.

Objective: We aimed to review randomized controlled trials (RCTs) of technology-based music interventions for reducing anxiety and pain among patients undergoing surgery or procedures. We examined the clinical situation, devices used, delivery methods, and effectiveness of technology-based music interventions in primary articles.

Methods: The search was performed in the following 5 electronic databases: PubMed, MEDLINE (OvidSP), CINAHL complete, PSYCINFO, and Embase. This systematic review focused on technology-based music interventions. The following articles were included: (1) RCTs, (2) studies using interactive technology (eg, smartphones, mHealth, tablets, applications, and virtual reality), (3) empirical studies reporting pain and anxiety outcomes, and (4) English articles published from 2018 to 2023 (as of January 18, 2023). The risk of bias was assessed using the Cochrane Risk of Bias tool version 2.

Results: Among 292 studies identified, 21 met the inclusion criteria and were included. Of these studies, 9 reported that anxiety scores decreased after music interventions and 7 reported that pain could be decreased before, during, and after procedures. The methodology of the music intervention was important to the results on anxiety and pain in the clinical trials. More than 50% (13/21, 62%) of the studies included in this review allowed participants to select themes themselves. However, it was difficult to distinguish differences in effects depending on the device or software used for the music interventions.

Conclusions: Technology-based music interventions could help reduce anxiety and pain among patients undergoing surgery or procedures. The findings of this review could help medical teams to choose a practical methodology for music interventions. Future studies should examine the effects of advanced technology-based music interventions using smart devices and software that promote interactions between medical staff and patients.

(JMIR Mhealth Uhealth 2024;12:e48802) doi:10.2196/48802

KEYWORDS

technology; music intervention; anxiety; pain

Introduction

Over 33 million patients were hospitalized in the United States in 2020 [1]. Previous studies have reported that many hospitalized patients had experienced negative symptoms, such as pain, anxiety, depression, sleep disturbance, stress, and fatigue, or a combination of these due to illness responses, medical procedures, and unfamiliar environments in the hospital [2-4]. This consistent psychological distress and the negative symptoms could cause a delay in recovery, disrupting optimal treatment and increasing morbidity and mortality [5,6].

To mitigate psychological distress and negative symptoms for patients, nonpharmacological complementary approaches have been studied, such as muscle relaxation, massage, aromatherapy, acupuncture, and music [7-11]. Among these interventions, music interventions were reported to be safe complementary approaches for patients. Therefore, music intervention studies were conducted, especially for hospitalized patients, to reduce pain, improve sleep quality, and help with successful mechanical ventilator weaning [10,12,13].

With the recent developments in science and technology, music interventions can be delivered using smart devices, such as smartphones, tablets, PCs, laptops, and apps [14]. These smart devices allow researchers to deliver personalized music and video content in interactive ways. In addition, health care providers often interact with patients to provide music interventions, bringing psychological stability to patients in a hospital environment. It is possible to provide this treatment independent of a therapist, which may provide a convenient intervention without the time restraints associated with waiting on a therapist [15-20].

Several of the studies supporting music interventions have explained the process of providing music genres and songs that reflect the taste and choice of subjects rather than the music selected by the provider and have reported that this approach enhances the effect of the intervention by increasing the interaction between the provider and patient [14,21-25]. This indicates that the provider-patient interaction is a key element of an intervention using smart devices, and the consideration of the process and method of providing music should be prioritized. However, previous reviews have only focused on the effects of music interventions according to medical treatment and environment, including surgery, procedures, and respiratory treatment [9,25,26], and there has been no consideration of the process and method of a technology-based music intervention and its impact on psychological issues such as anxiety and pain.

The purpose of this study was to focus on the methods of technology-based music interventions and examine the effectiveness of the interventions for the anxiety and pain of hospitalized patients undergoing procedures. The research questions were as follows: (1) What are the characteristics of technology-based music interventions in primary articles? (2) In what ways were technology-based interventions effective for the anxiety and pain of patients undergoing procedures?

Methods

Information Sources and Search Strategy

A systematic review of the literature was performed using the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [27]. A comprehensive search was completed in January 2023 by 2 authors (SL and JY) using the following 5 electronic databases: PubMed, MEDLINE (OvidSP), CINAHL complete, PSYCINFO, and Embase.

The search keywords were selected from the PICO format (population: adult patients with procedures in the inpatient and outpatient settings; intervention: technology-based music intervention; comparison: standard care or usual care; outcome: pain and anxiety). These included keywords such as (“inpatient*” OR “hospitalization” OR “intensive care unit*” OR “emergency ward*” OR “general ward*” OR “patient*”) AND (“music*” OR “music intervention” OR “music therapy” OR “music medicine” OR “music listen*” OR “music-based” OR “music methods”) AND (“mobile application*” OR “smartphone” OR “telemedicine” OR “tablet*” OR “computer” OR “mhealth*” OR “ehealth* OR “technolog*” OR “cellphone*” OR “internet*” OR “internet-based” OR “mobile-based” OR “technology-based” OR “smartphone-based” OR “mhealth-based” OR “app*” OR “ipad”) AND (“anxiety” OR “pain”). Results were limited to adults (18 years or older), English text, and publication within 5 years (2018-2023).

Inclusion and Exclusion Criteria

This systematic review selected original empirical research studies on technology-based music interventions. The following articles were included: (1) randomized controlled trials (RCTs), (2) studies using interactive technology (eg, smartphones, mHealth, tablets, applications, and virtual reality), (3) empirical studies reporting pain and anxiety outcomes, and (4) English articles published from 2018 to 2023 (as of January 18, 2023). Articles were excluded if they were (1) not full-text articles (eg, conference abstracts and poster abstracts), (2) review articles, (3) study protocols, (4) studies that were not focused on music interventions (eg, therapist-focused), and (5) studies that targeted inpatients and outpatients who were younger than 18 years. We specifically selected the most recent articles published within the last 5 years to ensure the most up-to-date information on technology-based interventions and to improve upon previous systematic reviews [9,25,26,28,29].

Selection Process and Data Items

Database searches were independently carried out by 2 authors (SL and JY) using electronic databases and cross-references in January 2023. Initially, relevant bibliographic details, including article titles, authors, journal names, publication years, keywords, and abstracts, were retrieved from each electronic database. The management of duplications was facilitated through the use of the EndNote program (Clarivate).

Following the deduplication process, the titles and abstracts underwent independent screening by the 2 authors (SL and JY). Any discrepancies encountered during this phase were systematically resolved through consensus-building between

them. Upon the completion of this initial screening stage, the identified primary articles underwent a comprehensive full-text review.

Subsequently, data extraction from the selected studies was conducted with precision to effectively synthesize the study findings. A matrix table employing Excel (Microsoft Corp) spreadsheets was proficiently used throughout the review process to manage and consolidate the extracted data. The extracted information included a comprehensive array of elements, such as authors' names, research objectives, baseline sample characteristics, study designs, intervention modalities related to music, control group specifications, and outcome variables, with a particular emphasis on elucidating findings pertinent to pain and anxiety.

The full-text screening was independently executed by the same 2 authors. In instances where discrepancies arose during this phase, the intervention of a third author (SP) was sought. The role of the third author entailed a meticulous review of the identified articles to ensure the accuracy and consistency of the selection process. Any disparities or ambiguities were meticulously addressed and resolved under the scrutiny of the third author.

Throughout the review process, adherence to academic standards and methodology was paramount. Any disagreements or discrepancies encountered at any stage were effectively addressed through consensus-building, thereby enhancing the reliability of the synthesized evidence.

Study Risk of Bias Assessment

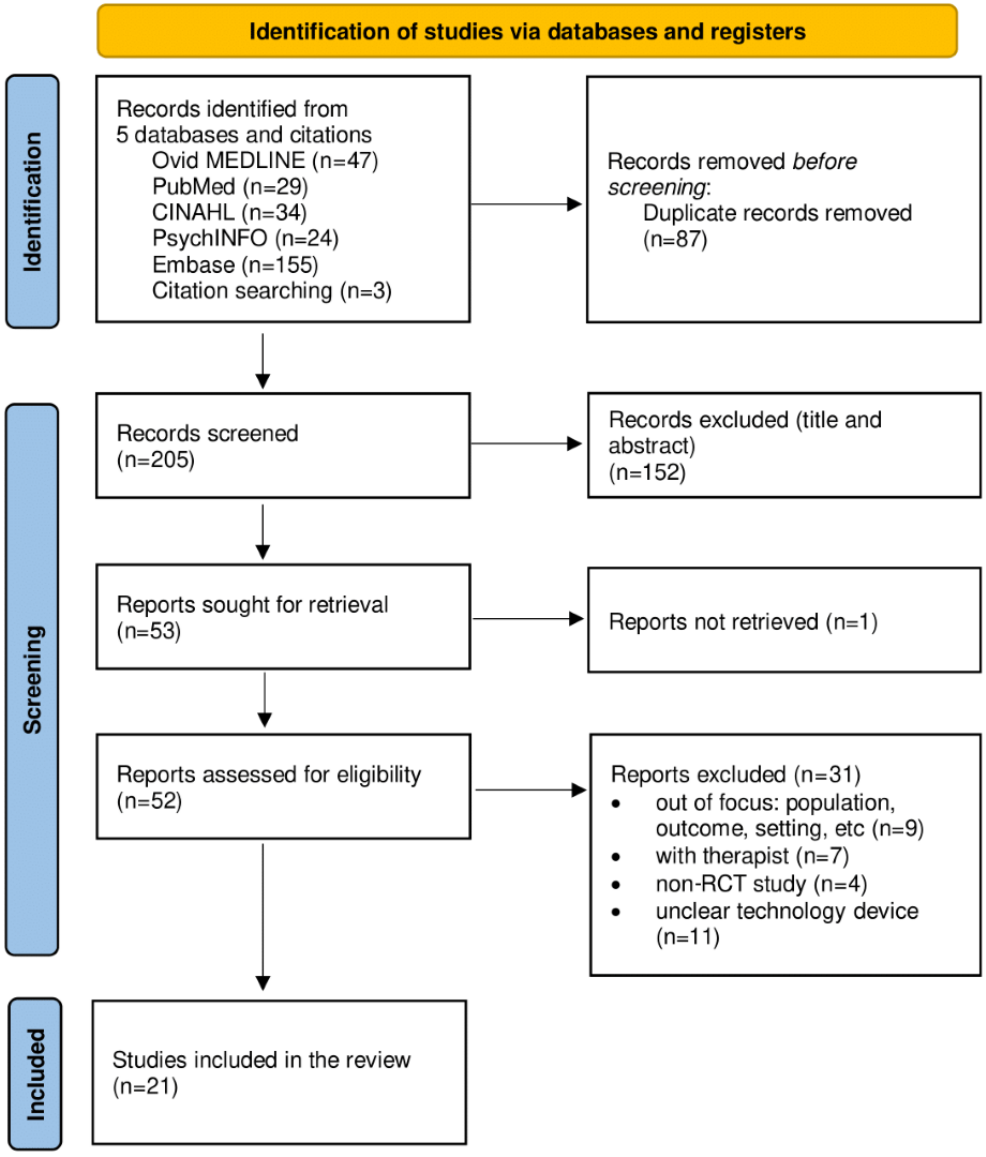
The risk of bias was assessed using the Cochrane Risk of Bias version 2 tool [30]. This tool is used to evaluate the risk of bias for individual RCTs. There are five domains: (1) randomization process, (2) deviations from intended interventions, (3) missing outcome data, (4) measurement of the outcomes, and (5) selection of the reported results. The 3 researchers (SP, SL, and JY) assessed primary articles independently using the Risk of Bias version 2 tool. All disagreements and discrepancies were discussed and resolved through meetings until a consensus was achieved.

Results

Study Selection

Figure 1 shows the flow diagram of the literature search and selection process. A total of 292 articles were identified through 5 electronic database searches. Initially, 87 duplicate articles were eliminated. Subsequently, 205 articles were screened based on titles and abstracts. Among these, 153 articles were excluded due to either irrelevance (n=152) or unavailability of full text (n=1). Following thorough full-text reviews, 31 articles were further excluded for various reasons, including being out of focus (n=9), being therapist-focused (n=7), having a non-RCT design (n=4), and using an unclear technology device (n=11). Consequently, 21 articles met the inclusion criteria for this study.

Figure 1. Study flow chart.



Study Characteristics

The characteristics of the 21 articles included in this study are shown in Table 1. The purpose of all the studies was to examine the effects of technology-based music interventions on patient outcomes, including anxiety and pain. Two studies aimed to examine particularly interactive technology-based interventions [31,32]. Drzymalski et al [31] focused on the impact of self-selected or preselected music, and Anglin et al [32] focused on the effects of a patient's choice. Most studies included male and female participants; however, 6 studies targeted only women [15,20,21,24,31,33]. The sample size ranged from 18 to 330. Most studies had 2 groups (ie, experimental group and control group), except 2 studies [31,33]. The interventions were performed for patients who underwent surgery or invasive procedures. Surgery included gynecologic surgery [20], cardiothoracic surgery [34], cataract surgery [14], orthopedic surgery [35,36], nasal bone fracture reduction [37], and cesarean delivery [21,31]. The procedures included radiation therapy [15], catheter insertion [38], colonoscopy [18], biopsy [24], bronchoscopy [39], steroid injection [40], transvaginal ultrasound-guided oocyte retrieval [33], wound care [16], eye procedures [41], pleural procedures [42], and urologic procedures [32]. The studies were conducted in many countries, including the United States [15,24,31,32,40], France [14,20,38], Hong Kong [18,33], Australia [42], Germany [21], Taiwan [34], India [35,41], Turkey [43], Spain [44], Chile [37], Malaysia [36], Iran [39], and Brazil [16].

Table 1. Characteristics of the included studies.

Article (author, year)	Purpose	Sample characteristics at baseline				Study design	Country and setting
		Sample size, n	Sex, n	Age (years)	Procedure/surgery		
Abdul Hamid et al [36], 2022	"To assess the effects of music on patients undergoing total knee replacement surgery under subarachnoid anesthesia."	90; EG[a]: 45, CG[b]: 45	Male: 57, female: 33	EG: 21-40 (n=15), 41-60 (n=16), >60 (n=14); CG: <20 (n=2), 21-40 (n=14), 41-60 (n=10), >60 (n=19)	Total knee replacement under subarachnoid anesthesia	RCT[c]	Malaysia
Chen et al [34], 2022	"To investigate the effectiveness of intermittent positive pressure breathing with and without music intervention."	18; EG: 9, CG: 9	Male: 14, female: 4	EG: 58.44 (SD 10.06)[d]; CG: 63.11 (SD 11.80)[d]	Cardiothoracic surgery	RCT	Taiwan, maybe the surgery department
Jacquier et al [38], 2022	"To evaluate the effect of a musical intervention on patient anxiety during a central venous catheter or dialysis catheter implantation in an intensive care unit."	72; EG: 37, CG: 35	Male: 39, female: 33	EG: 60 (47-70)[e]; CG: 61 (48-70)[e]	Insertion of a central venous catheter or a dialysis catheter	Prospective single-center controlled open-label 2-arm RCT	France, medical intensive care unit
Kaur et al [35], 2022	"To evaluate the role of music on perioperative anxiety, hemodynamic parameters, and patient satisfaction in patients undergoing orthopedic surgeries under spinal anesthesia."	70; EG: 35, CG: 35	Male: 54, female: 16	EG: 37.66 (SD 11.67)[d]; CG: 36.97 (SD 12.06)[d]	Elective orthopedic surgeries under spinal anesthesia	RCT	India, tertiary care hospital
Anglin et al [32], 2021	"To determine if listening to music of a patient's choice would decrease pain during various outpatient clinic urological procedures."	91; EG: 53, CG: 38	Male: 32, female: 59	Not stated	Urological procedure	Unblinded, single-center RCT	United States, outpatient clinic
Ferraz et al [16], 2021	"To evaluate the analgesic effect of music on acute procedural pain during the care of surgical tibial fracture wounds."	70; EG: 35, CG: 35	Male: 64, female: 6	EG: 20-29 (n=15), 30-39 (n=7), 40-49 (n=9), 50-59 (n=4); CG: 20-29 (n=9), 30-39 (n=12), 40-49 (n=11), 50-59 (n=3)	Care of surgical tibial fracture wounds managed under spinal anesthesia	Randomized, controlled, blinded clinical trial	Brazil, hospital
Guerrier et al [14], 2021	"To describe the effects of a web app–based music intervention on the incidence of hypertension in patients during cataract surgery performed under local anesthesia."	309; EG: 154, CG: 155	Male: 133, female: 176	68.9 (SD 10.8)[d]; EG: 68.5 (SD 11.2)[d]; CG: 69.2 (SD 10.8)[d]	First (unilateral) eye cataract surgery under local anesthesia	Single-masked RCT	France, hospital (operating room and recovery room)
Muddana et al [41], 2021	"To determine whether preoperative and perioperative music exposure reduced patient self-rated anxiety and physiologic indicators of stress during first-time phacoemulsification cataract surgery."	330; EG: 165, CG: 165	Male: 158, female: 172	EG: 57.8 (SD 7.72)[d]; CG: 58.79 (SD 7.57)[d]	First eye procedure (phacoemulsification with topical anesthesia)	Prospective open-label single-masked RCT	India, eye hospital
O'steen et al [15], 2021	"To evaluate the influence of genre-based music chosen by the study participant on anxiety during the first radiation therapy session."	102; EG: 51, CG: 51	Female: 102	62 (32-92)[f]; EG: 63 (38-85)[f]; CG: 62 (32-92)[f]	First radiation therapy treatment session	Prospective RCT	United States, oncology unit
Reynaud et al [20], 2021	"To determine whether listening to self-selected music decreases preoperative anxiety in women scheduled to undergo gynecologic surgery."	171; EG: 84, CG: 87	Female: 171	41.5 (SD 10.0)[d]; EG: 42.1 (SD 10.1)[d]; CG: 41.1 (SD 10.0)[d]	Gynecological surgery under general or spinal anesthesia	Single-blind, monocentric, parallel, superiority RCT	France, surgery department

Article (author, year)	Purpose	Sample characteristics at baseline				Study design	Country and setting
		Sample size, n	Sex, n	Age (years)	Procedure/surgery		
Bennett et al [24], 2020	"To determine whether listening to self-selected music during image-guided breast biopsy lowers anxiety."	129; EG: 75, CG: 54	Female: 129	49.7 (18-75)[g]; EG: 50.7 (30-72)[g]; CG: 48.4 (18-75)[g]	Stereotactic or ultrasound-guided core biopsy	Open-label RCT	United States, breast imaging center
Drzymalski et al [31], 2020	"To determine the effects of patient-selected or preselected music on anxiety in a parturient undergoing scheduled cesarean delivery."	149; EG1: 49, EG2: 50, CG: 50	Female: 149	EG1: 35 (SD 4)[d]; EG2: 35 (SD 5)[d]; CG: 33 (SD 5)[d]	Cesarean delivery	Prospective RCT	United States
Li et al [40], 2020	"To investigate the potential role of music on patients undergoing routine image-guided musculoskeletal corticosteroid injections for pain."	126; EG: 70, CG: 56	Male: 52, female: 74	EG: 56.8; CG: 58.9	Image-guided joint or spine corticosteroid injections	Prospective, single-blind RCT	United States
Lopez-Yufera et al [44], 2020	"To evaluate the effect of a music intervention on anxiety, blood pressure, and heart rate in adult patients with potentially malignant oral disorders."	80; EG: 40, CG: 40	Male: 24, female: 56	68.3 (SD 2.8)[d]; EG: 68.1 (SD 1.3)[d]; CG: 67.3 (SD 1.1)[d]	Consultation in a unit of oral medicine due to potentially malignant oral disorders	RCT	Spain, unit of oral medicine
Ko et al [18], 2019	"To examine the effects of an easy-listening music intervention on satisfaction, anxiety, pain, sedative and analgesic medication requirements, and physiological parameters in adult patients undergoing colonoscopy."	80; EG: 40, CG: 40	Male: 41, female: 39	EG: 57.68 (SD 11.07)[d]; CG: 57.68 (SD 11.92)[d]	Colonoscopy	Prospective, parallel RCT	Hong Kong, electromedical diagnostic unit
Ortega et al [37], 2019	"To evaluate whether the use of a fixed list of rhythmically slow music delivered by over-the-ear binaural headphones during a nasal fracture reduction with local anesthesia decreases the perception of pain and anxiety associated with the procedure."	36; EG: 17, CG: 19	Male: 22, female: 14	30.5 (18-60)[f]; EG: 35 (13)[e]; CG: 30 (10)[e]	Nasal bone fracture reduction	RCT	Chile, otorhinolaryngology department
Cheung et al [33], 2018	"To investigate the effect of music therapy on the perception of pain in patients undergoing a transvaginal ultrasound–guided oocyte retrieval (TUGOR) procedure."	196; EG1: 66, CG1: 65, CG2: 65	Female: 196	EG1: 35 (SD 3.2)[d]; CG1: 35.7 (SD 3.6)[d]; CG2: 34.7 (SD 3.0)[d]	TUGOR with conscious sedation	Open-label RCT	Hong Kong, assisted reproductive technology unit
Ergin et al [43], 2018	"To determine the effect of music on the severity of dyspnea, anxiety, blood pressure, breathing rate, pulse rate, and blood oxygen levels in patients with dyspnea."	60; EG: 30, CG: 30	Male: 36, female: 24	61.21 (SD 1.13)[d]; EG: 60.00 (SD 12.12)[d]; CG: 62.43 (SD 10.65)[d]	Being followed up with a complaint of dyspnea (no procedure)	RCT	Turkey, chest diseases service of a public hospital
Hepp et al [21], 2018	"To examine the anxiolytic and stress-reducing effect of a music intervention during cesarean delivery."	304; EG: 154, CG: 150	Female: 304	33.6 (18-47)[g]; EG: 33.5 (SD 5.4)[d]; CG: 33.7 (SD 5.4)[d]	Primary cesarean delivery under regional anesthesia	Single-center controlled RCT	Germany, department of gynecology and obstetrics
Mackintosh et al [42], 2018	"To evaluate the benefits of music therapy during pleural procedures on patient anxiety, perceived pain, and satisfaction with the procedure."	60; EG: 30, CG: 30	Male: 30, female: 30	67 (SD 14)[d]; EG: 65 (SD 15)[d]; CG: 68 (SD 13)[d]	Therapeutic pleural procedure	Prospective RCT	Australia, respiratory ward

Article (author, year)	Purpose	Sample characteristics at baseline				Study design	Country and setting
		Sample size, n	Sex, n	Age (years)	Procedure/surgery		
Navidian et al [39], 2018	"To investigate the effect of audiovisual distraction on the tolerability of flexible bronchoscopy."	60; EG: 30, CG: 30	Male: 34, female: 26	EG: 54.53 (SD 7.33)[d]; CG: 46.37 (SD 14.06)[d]	Flexible bronchoscopy	Single-center, prospective RCT	Iran

[a]EG: experimental group.

[b]CG: control group or comparison group.

[c]RCT: randomized controlled trial.

[d]Mean (SD).

[e]Median (IQR).

[f]Median (range).

[g]Mean (range).

Music Intervention Characteristics

Music Playing and Listening Devices

The experimental groups conducted the music intervention using web-based music applications (including QR code access) [14,15,20,24,31,37,38,42], smartphones [20], tablets [14,20], computers [20,42], CD players [20,21,34,43], MP3 players [16,18,33,36,41,44], projectors [39], iPods [31], or personal/cellular devices [32,40] as the music play devices or software. Over 90% (19/21, 91%) of studies used headphones [14,16,18,33-39,41,42,44], earphones [24,43], or speakers [15,21,31,40] for delivery of the intervention.

Music Selection Strategy

In over 50% (13/21, 62%) of studies, participants could select the music theme themselves [14-16,20,24,31,32,36,38-42], and most of the identified music content or genres did not limit the participants' choices [14,15,20,24,31,36,38,40,42]. However, some studies required participants to choose music from playlists available on the music app [38] or select pop-rock, romantic, or religious music [36]. If researchers decided on the music content or genre, there were variations in the selections, such as nature sounds [34], popular songs [18], slow-tempo songs [21,37], relaxation music [35,41,44], folk music [39], piano sonata [31], and regional music [16]. In the experimental groups, music was played before [14,18,20,37,39,41,42,44], during [15,18,21,24,31-33,35,37,38,40,42], or after [16,31,36,37,39,42] the procedure or surgery. The duration was from 10 to 60 minutes [14,16,18,20,36-39,42-44]. Some of the studies did not clearly report the played time and duration. The effect of the intervention was assessed by comparing the experimental group with a no-music intervention group [14,15,31-34,36,37,40-44], standard care group [16,18,21,24,35,38,39], or listening predetermined music group [20].

Effect of the Music Intervention

Anxiety

Seventeen studies employed measures, such as the State-Trait Anxiety Inventory, visual analog scale (VAS), Hospital Anxiety and Depression validated Scale, and Corah dental anxiety scale, to assess anxiety levels. Of these, 13 studies reported a decrease in anxiety scores following music intervention. Out of these 13 studies, 9 (69%) reported a statistically significant decrease in anxiety associated with the use of the music intervention compared to controls [14,21,24,35-37,41-43]. Four studies reported that there were no significant differences between the experimental and control groups, even though there was a reduction in anxiety scores in the experimental group [15,20,31,34]. Studies that reported decreased anxiety scores in the experimental group compared with the control group tended to use music selected by the participants [14,24,36,41,42] and use participants' choices, including classical music [41], traditional music of the nation [43], slow music [21,37], and relaxation music [35,41]. Ten studies assessed music playing during the procedure or surgery [15,18,21,24,31,33,35,37,38,42], and the music playing device or software did not show distinct characteristics (Table 2).

Table 2. Intervention and outcomes of the studies.

Article (author, year)	Music intervention of the experimental group				Comparison or control group	Outcomes (tool)	Key findings
	Music playing device or software	Music selection by participants	Music content or genre	Played time/duration			
Abdul Hamid et al [36], 2022	MP3 with headphones	Yes	Participants' choice: pop-rock, romantic, or religious	After regional anesthesia/30 min	No music	Anxiety (VAS[a], STAI-S[b])	Changes in anxiety from pre- to postoperation were significantly different between the groups (VAS; *P*=.002). Anxiety was higher in the CG[c] than in the EG[d] (STAI-S).
Chen et al [34], 2022	CD player with noise-cancelling headphones	No (nature sounds)	Nature sounds	Not stated	No music	Anxiety (STAI-S, STAI-T[e]); pain (VAS)	Anxiety was not significantly different between the groups (*P*>.05). Reduced postoperative pain and anxiety in cardiothoracic surgery patients, but no significance in the interaction between music intervention and time (*P*=.16).
Jacquier et al [38], 2022	Music Care app with headphones	Yes	Participants' choice: one of the playlists available on the Music Care app	During the procedure/20-60 min	Standard care without music	Anxiety (VAS); pain (VAS)	The music intervention did not reduce patients' anxiety as compared with usual care (anxiety, *P*=.24; pain, *P*=.40).
Kaur et al [35], 2022	Music player with noise-cancelling headphones	Probably no, initial volume setting by participants	Relaxation music	During surgery/not stated	Standard care with a standard operation theater tape sound without music	Anxiety (VAS-A[f])	Anxiety scores were comparable in both groups preoperatively and before anesthesia induction and were lower in the EG intra- and postoperatively (*P*<.001).
Anglin et al [32], 2021	Cellular device	Yes	Not stated	During the procedure/not stated	No music	Pain (VAS-P[g])	Among men, pain scores worsened in both groups (*P*=.38). Among women, changed pain scores significantly differed between the music group and nonmusic group (*P*=.005).
Ferraz et al [16], 2021	MP3 with headphones	Yes	Regional, others	Change of dressing, immediate postoperative period/30 min	Standard care	Pain (NRS[h])	Pain was lower in the EG than in the CG (*P*<.001).
Guerrier et al [14], 2021	Web application–based music (Music Care) with a tablet interface via headphones	Yes	Participants' choice	Before cataract surgery/20 min	Headphones without music	Anxiety (VAS)	Anxiety was lower in the EG than in the CG at the second end point (*P*=.005).
Muddana et al [41], 2021	MP3 with headphones	Yes	Relaxing classical, instrumental, or devotional music	Before surgery/not stated	Headphones without music	Anxiety (State anxiety)	Reduction in self-reported anxiety preoperatively in the EG. A higher proportion in the EG reported feeling not at all or a little anxious compared to the CG peri- and postoperatively (*P*<.05).
O'steen et al [15], 2021	A web-based music application with speakers	Yes	Participants' choice	During radiation therapy/not stated	No music	Anxiety (STAI[i])	Reduction in anxiety scores in the music group relative to the no music group, without statistical significance (*P*=.22).
Reynaud et al [20], 2021	Smartphone, tablet, computer, CD player, or Music Care app (participants' choice)	Yes	Participants' choice, self-selected playlist	One hour before surgery/20 min	Predetermined music using the Music Care app	Anxiety (STAI); pain (NRS)	No significant difference in the reduction of anxiety and pain scores between the groups (anxiety, *P*=.80; pain, *P*=.48).

Article (author, year)	Music intervention of the experimental group				Comparison or control group	Outcomes (tool)	Key findings
	Music playing device or software	Music selection by participants	Music content or genre	Played time/duration			
Bennett et al [24], 2020	A personalized internet radio station (Pandora) via earphones	Yes	Participants' choice	During biopsy/not stated	Standard care without music	Anxiety (STAI)	Anxiety reduction was significantly greater in the EG than in the CG (*P*=.03).
Drzymalski et al [31], 2020	Pandora station broadcast on the iPod with speakers	Yes; EG1=participants, EG2=preselected	EG1=participants' choice; EG2=Mozart piano sonata	Preoperative, during the procedure, and 1 hour after the procedure/not stated	No music	Anxiety (NRS); pain (NRS)	Postoperative anxiety: not different between the EG1 and CG (*P*=.43) and between the EG2 and CG (*P*=.15). Postoperative pain: not different between the EG1 and CG (*P*=.10), but significantly different between the EG2 and CG (*P*=.03).
Li et al [40], 2020	Personal devices with speakers or a radio station (participants' choice)	Yes	Participants' choice	During the procedure	No music	Pain (subjective questionnaire)	The EG had significantly lower postprocedural pain and a decrease in pain compared with the CG (*P*=.03).
Lopez-Yufera et al [44], 2020	MP3 with headphones	Not stated, access volume control	Relaxing music	Before medical intervention/10 min	Headphones without music	Anxiety (HADS[j], Corah dental anxiety score)	No significant differences in anxiety (*P*=.08).
Ko et al [18], 2019	MP3 with headphones	No; adjusted volume by participants	15 easy-listening Chinese popular songs	Before and during the procedure/20 min	Standard care	Anxiety (STAI-C[k]); pain (VAS)	No significant differences between the 2 groups in terms of anxiety (*P*>.05) and pain (*P*=.83).
Ortega et al [37], 2019	Spotify (QR code access) and Bluetooth headphones	No; set music intensity by participants	Rhythmically slow songs	Prior to the intervention, during the procedure, and postoperatively/each 10 min	No music	Anxiety (STAI); pain (VAS)	The EG had significantly lower levels of anxiety (*P*<.001) and pain (*P*<.001) compared with the CG.
Cheung et al [33], 2018	MP3 with headphones	Not stated	EG1=recommended music, EG2=mute music	During TUGOR[l]/not stated	No music	Pain (VAS-P); anxiety (STAI)	Pain in the EG1 was significantly lower than that in the EG2 and CG (*P*=.005). Anxiety was not significantly different between the groups.
Ergin et al [43], 2018	CD player with earphones	No; adjusted volume by participants	Husseini maqam	Not stated/30 min	No music	Anxiety (STAI)	The difference in anxiety in the EG before and after the intervention was statistically significant (*P*<.05).
Hepp et al [21], 2018	CD player using speakers	No	Slow tempo songs	During cesarean delivery/not stated	Standard care without music	Anxiety (STAI, VAS-A)	Significantly lower anxiety levels in the EG by time and group (STAI, *P*=.004; VAS, *P*=.02).
Mackintosh et al [42], 2018	A popular video-sharing website through a laptop computer with ear-bud headphones	Yes	Participants' choice	During the pleural procedure, and 10 min before and after the procedure	No music	Anxiety (STAI)	The EG had significantly improved state anxiety scores between pre- and postprocedure (*P*<.001). However, the pre- and postprocedure trait anxiety scores were not significantly different in both groups (*P*=.80).

Article (author, year)	Music intervention of the experimental group				Comparison or control group	Outcomes (tool)	Key findings
	Music playing device or software	Music selection by participants	Music content or genre	Played time/duration			
Navidian et al [39], 2018	Projector with headphones	Yes	Iranian folk music	From 10 min before to 10 min after the procedure	Standard care	Pain (VAS)	Pain was significantly less severe in the EG compared with the CG (*P*=.01).

[a]VAS: visual analog scale.

[b]STAI-S: State-Trait Anxiety Inventory for state-anxiety.

[c]CG: control group.

[d]EG: experimental group.

[e]STAI-T: State-Trait Anxiety Inventory for trait-anxiety.

[f]VAS-A: visual analog scale for anxiety.

[g]VAS-P: visual analog scale for pain.

[h]NRS: Numeric Rating Scale.

[i]STAI: State-Trait Anxiety Inventory.

[j]HADS: Hospital Anxiety and Depression validated Scale.

[k]STAI-C: State-Trait Anxiety Inventory, Chinese version.

[l]TUGOR: transvaginal ultrasound–guided oocyte retrieval.

Pain

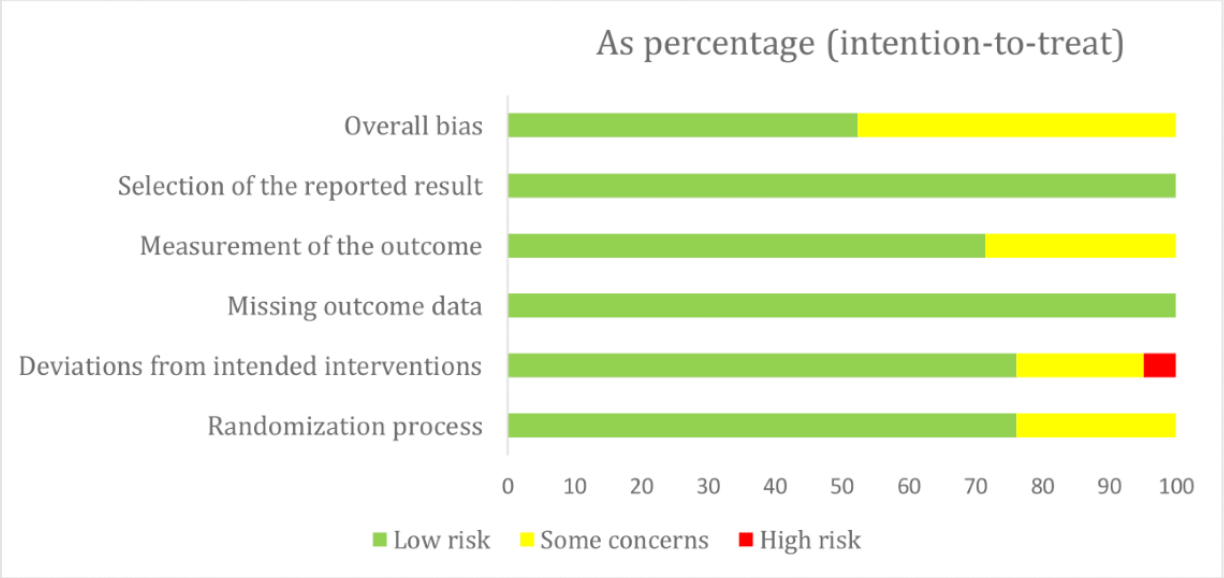
Eleven studies measured pain using the Numeric Rating Scale, VAS, or a subjective questionnaire. All studies reported that pain scores decreased after the music intervention. Among these 11 studies, 7 (64%) reported a statistically significant reduction in pain with the use of the music intervention [16,31-33,37,39,40]. Among them, Drzymalski et al [31] identified significant differences in the effectiveness of the music intervention in only 1 experimental group compared with the control group. Anglin et al [32] reported that male participants showed an overall increase in pain scores, whereas female participants in the intervention group exhibited improvements in pain scores compared with the worsening of scores in the control group. The other 4 articles reported no significant differences between the experimental and control groups, even though there was a reduction in pain scores in the experimental group [18,20,34,38]. Studies that reported decreases in pain in the experimental group tended to allow music selection by participants [16,31,32,39,40]. The played music content or genre varied (eg, slow songs, folk music, participant's choice, and regional music), and the music medium involved a QR code, projector, iPod, MP3 player, or personal or cellular device. Moreover, the timing of the music intervention was before or after the procedure or surgery (Table 2).

Risk of Bias in Studies

The quality assessment of the selected studies was conducted using the Risk of Bias 2 criteria [30]. These criteria provide a more nuanced approach compared to the previous version of the tool, allowing for a more detailed evaluation of bias across different domains of study conduct. This structured assessment helps researchers and readers of systematic reviews to better understand and interpret the quality of evidence presented in the studies. "Low risk" indicates that there are sufficient measures in place within the study design and conduct to minimize bias, thus providing confidence in the validity of the study results. Studies categorized as having low risk of bias are considered to have minimal risk of distorting the intervention effect estimates. "Some concerns" indicates that there are certain aspects of the study design or conduct that raise concerns about the potential for bias. "High risk" indicates that there are significant issues in the design, conduct, or reporting that substantially increase the risk of bias. Studies categorized as high risk of bias are deemed to have limitations that seriously compromise the validity of the findings.

Of the 21 studies, 11 (52%) were classified as having low risk of bias and 10 (48%) were identified as having some concerns. More specifically, in terms of the randomization process, 76% (16/21) of studies were deemed to have low risk and 24% (5/21) were flagged as having some concerns. Regarding deviation from the intended intervention, 76% (16/21) of studies were categorized as having low risk, 19% (4/21) were identified as having some concerns, and 5% (1/21) were categorized as having high risk. In terms of the measurement of outcomes, 71% (15/21) of studies were classified as having low risk and 29% (6/21) were identified as having some concerns. All 21 studies (100%) were rated as low risk for missing outcome data and the selection of the reported results (Figure 2). The evaluation was conducted independently by 3 authors, and any disparities were resolved through discussion.

Figure 2. Risk of bias.



Discussion

Research Trends and Strategies

Many patients undergoing surgical procedures to manage illnesses often endure physical discomfort and psychological strain, which can significantly impact medical expenses and clinical outcomes [45]. Considering these aspects, this systematic review investigated the efficacy of music interventions for alleviating anxiety and pain, drawing insights from recent clinical trials. Specifically, it delves into various methodological approaches, such as employing music-dispensing devices and considering individual preferences regarding music selection and genre. The advent of technological advancements has facilitated the integration of diverse devices for music interventions within clinical settings. These innovations have not only enhanced the accessibility of music interventions in such environments but also empowered patients to personalize their music experiences according to their preferences. This signifies the potential of music interventions in not only mitigating anxiety and pain but also fostering meaningful patient-clinician interactions during the intervention process. Thus, our findings underscore the significance of methodological nuances in music intervention studies, highlighting the pivotal role in yielding favorable outcomes.

A total of 21 studies met the inclusion criteria for this review. Among these 21 studies, 17 addressed anxiety, 11 addressed pain, and 6 addressed both. These studies aimed to identify the effectiveness of music interventions for pain and anxiety among hospital patients. Among the 17 studies on anxiety, 13 reported a significant reduction in anxiety in the experimental group compared with the control group, while 4 studies showed inconsistent results. Similarly, among the 11 studies on pain, 7 reported a significant decrease in pain in the experimental group compared with the control group, while 4 studies showed inconsistent results. These findings support earlier literature that music interventions can be used to manage physical and psychological symptoms [46,47] and can significantly benefit improvements in pain [46-49] and anxiety [46,47,49,50], but they indicated the need for repeated studies.

Previous reviews have underscored the value of music interventions across various patient populations and treatments [46,51], and they have been recommended to decrease anxiety, stress, and postoperative pain [46,47,51]. In this review, music interventions using technology were implemented across a range of medical settings, with experimental groups using diverse music playback devices or software, including web-based music applications (eg, QR code access), smartphones, tablets, computers, CD players, MP3 players, projectors, iPods, and personal or portable devices. However, the findings regarding which device delivered the music intervention better were unclear, and the difference in the effect of the medium on anxiety and pain was not directly discussed. Therefore, these findings indicate that selecting an appropriate device is an important factor in music interventions [52], and the effectiveness of the medium of the music intervention should be further assessed in future studies.

More than 50% (13/21, 62%) of studies included in this review allowed participants to select themes themselves, and of these, more than 70% (10/13, 77%) reported positive effects. Previous studies reported that anxiety and pain reductions were more effective when self-selected music was played [28,29,52,53]. The researchers interpreted that participants' familiarity with self-selected music evoked a sense of pleasure and control that could act to reduce pain [29]. On this basis, this review suggests that the method of the selection of music is also an important factor in a music intervention for reducing anxiety and pain. Some studies required participants to select music from playlists available in music applications, which included pop-rock music, romantic music, religious music, natural sounds, popular songs, slow-tempo music, relaxing music, folk music, piano sonatas, and local folk music. Music content or genre and tempo are important for stimulation and relaxation [51]. Pop, film soundtrack, jazz, classical, folk, and instrumental music are widely used music genres [54]. Moreover, slow-tempo music

(60-80 beats/min) supports relaxation, whereas fast-tempo music can cause discomfort [55]. Hatwar and Gawande [56] reported that consulting an expert while selecting music is beneficial. Yangöz and Özer [47] reported that a music genre should be chosen after consulting with an expert, and the patient or researcher should then select a music playlist. However, in our review, it was unclear whether the music was chosen after expert consultation. Therefore, consultation with experts can be considered in the process of selecting music playlists, and more studies are required to examine the difference between the effects of music selected by oneself and those of music selected by researchers [51]. Although music interventions can be used with individualized interaction, group interaction, or individual listening [47], most of the reviewed literature involved individual listening. Thus, identifying the effect of the music intervention in the context of group interaction in future studies will help implement appropriate music interventions.

Consistently in previous literature [57], the interventions were performed before, during, and after the surgery or procedure, and the duration of playing music varied from 10 to 60 minutes. Music duration is a factor in the intervention's effect on pain and anxiety [52]. Yangöz and Özer [47] reported that music duration ranged from 15 to 180 minutes, and they recommended the duration be restricted to less than 20 minutes because listeners could get bored if the duration is too long. Gillen et al [58] reported that the most common duration for listening to music was 15 to 30 minutes, and Pittman and Kridli [51] found that a duration of 15 to 20 minutes was effective in music interventions. The appropriate playing duration for a music intervention could vary according to the treatment environment (eg, the required time for a surgery or procedure differs). However, studies that determine the appropriate duration for music interventions are limited [51,52].

The risk of bias assessment is an established approach to evaluate the credibility of results at the systematic review level [59]. Nevertheless, the results of this review need to be interpreted with caution. One of the studies included in this review had a high risk of bias in deviations from intended interventions since we could not clarify the interventionist. Given the nature of a music intervention as a therapy, a professional may be required for intervention delivery. However, if researchers perform the intervention, they could be aware of the potential influence on the results. Thus, researchers should design the approach by clearly considering the importance of study quality.

Future studies should examine the effect of advanced technology-based music interventions in the context of interaction using various smart devices and software. Advanced technologies, such as artificial intelligence, robotics, and smartphone apps, are continually evolving for better patient outcomes. More interactive and patient-tailored music interventions using these advanced technologies can be developed and tested to reduce anxiety and pain by allowing to control the type of music and music delivery method. It is also necessary to review the effectiveness of in-patient mobile communication interventions between medical staff and patients, and to promote the development of different field application methods using this technology.

Limitations and Strengths

While our systematic review offers valuable insights, it is important to acknowledge its limitations. Despite our diligent efforts to include comprehensive studies meeting the eligibility criteria, there remains the possibility of incomplete retrieval owing to the search keywords used. Considering the significant evolution in music delivery technology, we decided to include articles published within the past 5 years. Moreover, our search was confined to 5 databases and included only studies written in English, potentially leading to an incomplete list of included studies.

Despite its limitations, our review has significant value as it has conducted an in-depth analysis of the current literature and has provided a lucid summary of intervention protocols and outcomes concerning anxiety and pain. Emphasizing technology use and its outcomes, our findings underscore the critical importance of selecting suitable devices in music interventions, presenting a notable challenge for future intervention development. Notably, our review distinguishes itself with its comprehensive summary of primary articles, intervention methodologies, and outcomes related to anxiety and pain. By highlighting technology use and its outcomes, we offer a global perspective on intervention protocols and the efficacy of technology-based music interventions for alleviating pain and anxiety among hospitalized patients.

While acknowledging the potential impact of a meta-analysis, our review prioritizes elucidating intervention methodologies specific to technology-based music interventions. The diverse array of technologies employed in the included studies can pose challenges in clearly discerning the effectiveness of the used devices. By focusing on methodologies related to technology use, our review underscores significant challenges associated with selecting appropriate devices in intervention protocols aimed at mitigating pain and anxiety in hospitalized patients. This approach can help to lay the foundation for validating the effectiveness of technology-based device use in future studies.

Conclusions

Music interventions are valuable for various patients and treatments and have been recommended to decrease anxiety, stress, and postoperative pain. A music intervention is a nonpharmacological complementary approach and shows a positive effect on patients, and up to 70% of studies reported a positive effect when patients were allowed to select the music themselves. This systematic review provided an in-depth review of the current literature on technology-based music interventions for patients undergoing various procedures. Future studies are needed to examine the effectiveness of interactive technology-based music interventions for reducing anxiety and pain among hospitalized patients undergoing procedures. This review contributes to the research on technology-based music interventions and can help to select a practical methodology for the interventions. Moreover, further meta-analyses should be conducted to enhance statistical power through the combination of studies.

Acknowledgments

This work was supported by the fund of the research promotion program of Gyeongsang National University (2021) and the University of Memphis.

Data Availability

This paper includes all data generated during this review.

Authors' Contributions

SP and JY conceived the idea. SP, SL, and JY reviewed the literature. SP and JY led the work. All authors wrote this manuscript. SL and JY revised it critically for important intellectual content. All authors have read and agreed to the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[PDF File (Adobe PDF File), 883 KB - mhealth_v12i1e48802_app1.pdf]

References

1. Fast Facts on U.S. Hospitals. American Hospital Association. URL: https://www.aha.org/statistics/fast-facts-us-hospitals [accessed 2023-04-02]
2. Davis MP, Rybicki LA, Samala RV, Patel C, Parala-Metz A, Lagman R. Pain or fatigue: which correlates more with suffering in hospitalized cancer patients? Support Care Cancer 2021 Aug 21;29(8):4535-4542. [doi: 10.1007/s00520-021-05996-2] [Medline: 33479793]
3. Wesselius HM, van den Ende ES, Alsma J, Ter Maaten J, Schuit SCE, Stassen PM, "Onderzoeks Consortium Acute Geneeskunde" Acute Medicine Research Consortium. Quality and quantity of sleep and factors associated with sleep disturbance in hospitalized patients. JAMA Intern Med 2018 Sep 01;178(9):1201-1208 [FREE Full text] [doi: 10.1001/jamainternmed.2018.2669] [Medline: 30014139]
4. Weiss R, Vittinghoff E, Fang MC, Cimino JEW, Chasteen KA, Arnold RM, et al. Associations of physician empathy with patient anxiety and ratings of communication in hospital admission encounters. J Hosp Med 2017 Oct 06;12(10):805-810. [doi: 10.12788/jhm.2828] [Medline: 28991945]
5. Fattouh N, Hallit S, Salameh P, Choueiry G, Kazour F, Hallit R. Prevalence and factors affecting the level of depression, anxiety, and stress in hospitalized patients with a chronic disease. Perspect Psychiatr Care 2019 Oct 02;55(4):592-599. [doi: 10.1111/ppc.12369] [Medline: 30825393]
6. IsHak WW, Collison K, Danovitch I, Shek L, Kharazi P, Kim T, et al. Screening for depression in hospitalized medical patients. J Hosp Med 2017 Feb;12(2):118-125. [doi: 10.12788/jhm.2693] [Medline: 28182810]
7. Brinkhaus B, Roll S, Jena S, Icke K, Adam D, Binting S, et al. Acupuncture in patients with allergic asthma: a randomized pragmatic trial. J Altern Complement Med 2017 Apr;23(4):268-277. [doi: 10.1089/acm.2016.0357] [Medline: 28287818]
8. De Paolis G, Naccarato A, Cibelli F, D'Alete A, Mastroianni C, Surdo L, et al. The effectiveness of progressive muscle relaxation and interactive guided imagery as a pain-reducing intervention in advanced cancer patients: A multicentre randomised controlled non-pharmacological trial. Complement Ther Clin Pract 2019 Feb;34:280-287. [doi: 10.1016/j.ctcp.2018.12.014] [Medline: 30712739]
9. Hole J, Hirsch M, Ball E, Meads C. Music as an aid for postoperative recovery in adults: a systematic review and meta-analysis. The Lancet 2015 Oct;386(10004):1659-1671. [doi: 10.1016/s0140-6736(15)60169-6]
10. Kakar E, Venema E, Jeekel J, Klimek M, van der Jagt M. Music intervention for sleep quality in critically ill and surgical patients: a meta-analysis. BMJ Open 2021 May 10;11(5):e042510 [FREE Full text] [doi: 10.1136/bmjopen-2020-042510] [Medline: 33972331]
11. Ozkaraman A, Dügüm Ö, Özen Yılmaz H, Usta Yesilbalkan Ö. Aromatherapy: the effect of lavender on anxiety and sleep quality in patients treated with chemotherapy. Clin J Oncol Nurs 2018 Apr 01;22(2):203-210. [doi: 10.1188/18.CJON.203-210] [Medline: 29547610]
12. Hetland B, Lindquist R, Weinert C, Peden-McAlpine C, Savik K, Chlan L. Predictive associations of music, anxiety, and sedative exposure on mechanical ventilation weaning trials. Am J Crit Care 2017 May;26(3):210-220 [FREE Full text] [doi: 10.4037/ajcc2017468] [Medline: 28461543]

13. Lee C, Lee C, Hsu M, Lai C, Sung Y, Lin C, et al. Effects of music intervention on state anxiety and physiological indices in patients undergoing mechanical ventilation in the intensive care unit. Biol Res Nurs 2017 Mar;19(2):137-144. [doi: 10.1177/1099800416669601] [Medline: 27655993]
14. Guerrier G, Abdoul H, Jilet L, Rothschild P, Baillard C. Efficacy of a web app-based music intervention during cataract surgery: a randomized clinical trial. JAMA Ophthalmol 2021 Sep 01;139(9):1007-1013 [FREE Full text] [doi: 10.1001/jamaophthalmol.2021.2767] [Medline: 34323929]
15. O'steen L, Lockney NA, Morris CG, Johnson-Mallard V, Pereira D, Amdur RJ. A prospective randomized trial of the influence of music on anxiety in patients starting radiation therapy for cancer. Int J Radiat Oncol Biol Phys 2021 Mar 01;109(3):670-674. [doi: 10.1016/j.ijrobp.2020.09.048] [Medline: 33121861]
16. Ferraz MCL, Santana-Santos E, Pinto JS, Nunes Ribeiro CJ, Santana JFNB, Alves JAB, et al. Analgesic effect of music during wound care among patients with diaphyseal tibial fractures: Randomized controlled trial. Eur J Pain 2021 Mar 26;25(3):541-549. [doi: 10.1002/ejp.1692] [Medline: 33135319]
17. Khan SH, Wang S, Harrawood A, Martinez S, Heiderscheit A, Chlan L, et al. Decreasing Delirium through Music (DDM) in critically ill, mechanically ventilated patients in the intensive care unit: study protocol for a pilot randomized controlled trial. Trials 2017 Nov 29;18(1):574 [FREE Full text] [doi: 10.1186/s13063-017-2324-6] [Medline: 29187230]
18. Ko SY, Leung DY, Wong EM. Effects of easy listening music intervention on satisfaction, anxiety, and pain in patients undergoing colonoscopy: a pilot randomized controlled trial. Clin Interv Aging 2019;14:977-986 [FREE Full text] [doi: 10.2147/CIA.S207191] [Medline: 31213784]
19. Padam A, Sharma N, Sastri OSKS, Mahajan S, Sharma R, Sharma D. Effect of listening to Vedic chants and Indian classical instrumental music on patients undergoing upper gastrointestinal endoscopy: A randomized control trial. Indian J Psychiatry 2017;59(2):214. [doi: 10.4103/psychiatry.indianjpsychiatry_314_16]
20. Reynaud D, Bouscaren N, Lenclume V, Boukerrou M. Comparing the effects of self-selected MUsic versus predetermined music on patient ANXiety prior to gynaecological surgery: the MUANX randomized controlled trial. Trials 2021 Aug 13;22(1):535 [FREE Full text] [doi: 10.1186/s13063-021-05511-2] [Medline: 34389022]
21. Hepp P, Hagenbeck C, Gilles J, Wolf OT, Goertz W, Janni W, et al. Effects of music intervention during caesarean delivery on anxiety and stress of the mother a controlled, randomised study. BMC Pregnancy Childbirth 2018 Nov 03;18(1):435 [FREE Full text] [doi: 10.1186/s12884-018-2069-6] [Medline: 30390639]
22. Walter S, Gruss S, Neidlinger J, Stross I, Hann A, Wagner M, et al. Evaluation of an objective measurement tool for stress level reduction by individually chosen music during colonoscopy-results from the study "ColoRelaxTone". Front Med (Lausanne) 2020 Sep 15;7:525 [FREE Full text] [doi: 10.3389/fmed.2020.00525] [Medline: 33043027]
23. Lee W, Sung H, Liu S, Chang S. Meditative music listening to reduce state anxiety in patients during the uptake phase before positron emission tomography (PET) scans. Br J Radiol 2017 Feb;90(1070):20160466 [FREE Full text] [doi: 10.1259/bjr.20160466] [Medline: 27897034]
24. Bennett DL, Swan JS, Gazelle GS, Saksena M. Music during image-guided breast biopsy reduces patient anxiety levels. Clin Imaging 2020 Sep;65:18-23. [doi: 10.1016/j.clinimag.2020.03.018] [Medline: 32353714]
25. Sibanda A, Carnes D, Visentin D, Cleary M. A systematic review of the use of music interventions to improve outcomes for patients undergoing hip or knee surgery. J Adv Nurs 2019 Mar 19;75(3):502-516. [doi: 10.1111/jan.13860] [Medline: 30230564]
26. Cheng J, Zhang H, Bao H, Hong H. Music-based interventions for pain relief in patients undergoing hemodialysis: A PRISMA-compliant systematic review and meta-analysis. Medicine (Baltimore) 2021 Jan 15;100(2):e24102 [FREE Full text] [doi: 10.1097/MD.0000000000024102] [Medline: 33466179]
27. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
28. Garza-Villarreal EA, Pando V, Vuust P, Parsons C. Music-induced analgesia in chronic pain conditions: a systematic review and meta-analysis. Pain Physician 2017 Nov;20(7):597-610 [FREE Full text] [Medline: 29149141]
29. Martin-Saavedra JS, Vergara-Mendez LD, Pradilla I, Vélez-van-Meerbeke A, Talero-Gutiérrez C. Standardizing music characteristics for the management of pain: A systematic review and meta-analysis of clinical trials. Complement Ther Med 2018 Dec;41:81-89. [doi: 10.1016/j.ctim.2018.07.008] [Medline: 30477868]
30. Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page M, et al. Cochrane Handbook for Systematic Reviews of Interventions Version 6.3, 2022. Cochrane. URL: https://training.cochrane.org/handbook/archive/v6.3 [accessed 2024-06-24]
31. Drzymalski DM, Lumbreras-Marquez MI, Tsen LC, Camann WR, Farber MK. The effect of patient-selected or preselected music on anxiety during cesarean delivery: a randomized controlled trial. J Matern Fetal Neonatal Med 2020 Dec 27;33(24):4062-4068. [doi: 10.1080/14767058.2019.1594766] [Medline: 30880522]
32. Anglin C, Knoll P, Mudd B, Ziegler C, Choi K. Music's effect on pain relief during outpatient urological procedures: a single center, randomized control trial focusing on gender differences. Transl Androl Urol 2021 Jun;10(6):2332-2339 [FREE Full text] [doi: 10.21037/tau-20-1311] [Medline: 34295720]

33. Cheung CWC, Yee AWW, Chan PS, Saravelos SH, Chung JPW, Cheung LP, et al. The impact of music therapy on pain and stress reduction during oocyte retrieval - a randomized controlled trial. Reprod Biomed Online 2018 Aug;37(2):145-152. [doi: 10.1016/j.rbmo.2018.04.049] [Medline: 30078420]
34. Chen G, Guo L, Chuang I, Kuo H, Tsai Y, Liu S. Effect of music intervention on lung expansion exercises after cardiothoracic surgery. J Clin Med 2022 Mar 13;11(6):1589 [FREE Full text] [doi: 10.3390/jcm11061589] [Medline: 35329916]
35. Kaur H, Saini N, Singh G, Singh A, Dahuja A, Kaur R. Music as an aid to allay anxiety in patients undergoing orthopedic surgeries under spinal anesthesia. Noise Health 2022;24(112):7-12 [FREE Full text] [doi: 10.4103/nah.nah_58_21] [Medline: 35645134]
36. Abdul Hamid MR, Mansor MB, Zainal Abidin MF. Music therapy for reducing anxiety in patients undergoing total knee replacement surgery under subarachnoid anesthesia. J Orthop Surg (Hong Kong) 2022 Aug 27;30(2):10225536221122347 [FREE Full text] [doi: 10.1177/10225536221122347] [Medline: 36031855]
37. Ortega A, Gauna F, Munoz D, Oberreuter G, Breinbauer HA, Carrasco L. Music therapy for pain and anxiety management in nasal bone fracture reduction: randomized controlled clinical trial. Otolaryngol Head Neck Surg 2019 Oct 11;161(4):613-619. [doi: 10.1177/0194599819856604] [Medline: 31184266]
38. Jacquier S, Nay M, Muller G, Muller L, Mathonnet A, Lefèvre-Benzekri D, et al. Effect of a musical intervention during the implantation of a central venous catheter or a dialysis catheter in the intensive care unit: a prospective randomized pilot study. Anesth Analg 2022 Apr 01;134(4):781-790. [doi: 10.1213/ANE.0000000000005696] [Medline: 35299213]
39. Navidian A, Moulaei N, Ebrahimi Tabas E, Solaymani S. The effect of audiovisual distraction on the tolerability of flexible bronchoscopy: a randomized trial. Clin Respir J 2018 Jan 10;12(1):76-83. [doi: 10.1111/crj.12486] [Medline: 27116084]
40. Li W, Bryan RG, Kheterpal A, Simeone FJ, Chang CY, Torriani M, et al. The effect of music on pain and subjective experience in image-guided musculoskeletal corticosteroid injections: a randomized controlled trial. Skeletal Radiol 2020 Mar 21;49(3):435-441. [doi: 10.1007/s00256-019-03298-7] [Medline: 31435716]
41. Muddana SK, Hess OM, Sundar S, Venkatesh R. Preoperative and perioperative music to reduce anxiety during first-time phacoemulsification cataract surgery in the high-volume setting: randomized controlled trial. J Cataract Refract Surg 2021 Apr 01;47(4):471-475. [doi: 10.1097/j.jcrs.0000000000000485] [Medline: 33196574]
42. Mackintosh J, Cone G, Harland K, Sriram KB. Music reduces state anxiety scores in patients undergoing pleural procedures: a randomised controlled trial. Intern Med J 2018 Sep 04;48(9):1041-1048. [doi: 10.1111/imj.13738] [Medline: 29345398]
43. Ergin E, Sagkal Midilli T, Baysal E. The effect of music on dyspnea severity, anxiety, and hemodynamic parameters in patients with dyspnea. J Hosp Palliat Nurs 2018;20(1):81-87. [doi: 10.1097/njh.0000000000000403]
44. Lopez-Yufera E, López-Jornet P, Toralla O, Pons-Fuster López E. Non-pharmacological interventions for reducing anxiety in patients with potentially malignant oral disorders. J Clin Med 2020 Feb 26;9(3):622 [FREE Full text] [doi: 10.3390/jcm9030622] [Medline: 32110871]
45. van der Meij E, Huirne JA, Bouwsma EV, van Dongen JM, Terwee CB, van de Ven PM, et al. Substitution of usual perioperative care by eHealth to enhance postoperative recovery in patients undergoing general surgical or gynecological procedures: study protocol of a randomized controlled trial. JMIR Res Protoc 2016 Dec 21;5(4):e245 [FREE Full text] [doi: 10.2196/resprot.6580] [Medline: 28003177]
46. Kühlmann A, de Rooij A, Kroese L, van Dijk M, Hunink M, Jeekel J. Meta-analysis evaluating music interventions for anxiety and pain in surgery. Br J Surg 2018 Jun;105(7):773-783 [FREE Full text] [doi: 10.1002/bjs.10853] [Medline: 29665028]
47. Yangöz Ş, Özer Z. Effects of music intervention on physical and psychological problems in adults receiving haemodialysis treatment: A systematic review and meta-analysis. J Clin Nurs 2022 Dec 03;31(23-24):3305-3326. [doi: 10.1111/jocn.16199] [Medline: 35118718]
48. Lee JH. The effects of music on pain: a meta-analysis. J Music Ther 2016;53(4):430-477. [doi: 10.1093/jmt/thw012] [Medline: 27760797]
49. Wu X, Ji B, Yao S, Wang L, Jiang Z. Effect of music intervention during hemodialysis: a comprehensive meta-analysis. Eur Rev Med Pharmacol Sci 2021 May;25(10):3822-3834 [FREE Full text] [doi: 10.26355/eurrev_202105_25950] [Medline: 34109591]
50. Corbijn van Willenswaard K, Lynn F, McNeill J, McQueen K, Dennis C, Lobel M, et al. Music interventions to reduce stress and anxiety in pregnancy: a systematic review and meta-analysis. BMC Psychiatry 2017 Jul 27;17(1):271 [FREE Full text] [doi: 10.1186/s12888-017-1432-x] [Medline: 28750631]
51. Pittman S, Kridli S. Music intervention and preoperative anxiety: an integrative review. Int Nurs Rev 2011 Jun 15;58(2):157-163. [doi: 10.1111/j.1466-7657.2011.00888.x] [Medline: 21554287]
52. See C, Ng M, Ignacio J. Effectiveness of music interventions in reducing pain and anxiety of patients in pediatric and adult emergency departments: A systematic review and meta-analysis. Int Emerg Nurs 2023 Jan;66:101231. [doi: 10.1016/j.ienj.2022.101231] [Medline: 36528945]
53. Jafari H, Emami Zeydi A, Khani S, Esmaeili R, Soleimani A. The effects of listening to preferred music on pain intensity after open heart surgery. Iran J Nurs Midwifery Res 2012 Jan;17(1):1-6 [FREE Full text] [Medline: 23493927]

54. Zhang Y, Cai J, An L, Hui F, Ren T, Ma H, et al. Does music therapy enhance behavioral and cognitive function in elderly dementia patients? A systematic review and meta-analysis. Ageing Res Rev 2017 May;35:1-11. [doi: 10.1016/j.arr.2016.12.003] [Medline: 28025173]
55. de Witte M, Pinho ADS, Stams G, Moonen X, Bos AE, van Hooren S. Music therapy for stress reduction: a systematic review and meta-analysis. Health Psychol Rev 2022 Mar 27;16(1):134-159. [doi: 10.1080/17437199.2020.1846580] [Medline: 33176590]
56. Hatwar NR, Gawande UH. Can Music Therapy Reduce Human Psychological Stress: A Review. In: Zhang YD, Mandal J, So-In C, Thakur N, editors. Smart Trends in Computing and Communications. Smart Innovation, Systems and Technologies, vol 165. Singapore: Springer; 2020:405-411.
57. Joanna Briggs Institute. The Joanna Briggs Institute Best Practice Information Sheet: music as an intervention in hospitals. Nurs Health Sci 2011 Mar 23;13(1):99-102. [doi: 10.1111/j.1442-2018.2011.00583.x] [Medline: 21426462]
58. Gillen E, Biley F, Allen D. Effects of music listening on adult patients pre-procedural state anxiety in hospital. JBI Library of Systematic Reviews 2008;6(17):686-732. [doi: 10.11124/jbisrir-2008-180]
59. Schwingshackl L, Schünemann H, Meerpohl JJ. Improving the trustworthiness of findings from nutrition evidence syntheses: assessing risk of bias and rating the certainty of evidence. Eur J Nutr 2021 Sep 30;60(6):2893-2903 [FREE Full text] [doi: 10.1007/s00394-020-02464-1] [Medline: 33377996]

Abbreviations

RCT: randomized controlled trial
VAS: visual analog scale

Edited by L Buis; submitted 07.05.23; peer-reviewed by P Watson, T Parker, SY Kim; comments to author 23.10.23; revised version received 19.12.23; accepted 03.06.24; published 08.07.24.

Please cite as:
Park S, Lee S, Howard S, Yi J
Technology-Based Music Interventions to Reduce Anxiety and Pain Among Patients Undergoing Surgery or Procedures: Systematic Review of the Literature
JMIR Mhealth Uhealth 2024;12:e48802
URL: https://mhealth.jmir.org/2024/1/e48802
doi:10.2196/48802
PMID:

Review

Acceptability, Effectiveness, and Roles of mHealth Applications in Supporting Cancer Pain Self-Management: Integrative Review

Weizi Wu[1], RN, PhD; Teresa Graziano[1], RN, PhD; Andrew Salner[2], MD; Ming-Hui Chen[3], PhD; Michelle P Judge[1], RN, PhD; Xiaomei Cong[4], RN, PhD; Wanli Xu[1], RN, PhD

[1]School of Nursing, University of Connecticut, Storrs, CT, United States

[2]Hartford HealthCare Cancer Institute, Hartford, CT, United States

[3]Department of Statistics, University of Connecticut, Storrs, CT, United States

[4]Yale School of Nursing, Orange, CT, United States

Corresponding Author:
Wanli Xu, RN, PhD
School of Nursing
University of Connecticut
231 Glenbrook Rd
Storrs, CT, 06269
United States
Phone: 1 8604860580
Email: wanli.xu@uconn.edu

Abstract

Background: Cancer pain remains highly prevalent and persistent throughout survivorship, and it is crucial to investigate the potential of leveraging the advanced features of mobile health (mHealth) apps to empower individuals to self-manage their pain.

Objective: This review aims to comprehensively understand the acceptability, users' experiences, and effectiveness of mHealth apps in supporting cancer pain self-management.

Methods: We conducted an integrative review following Souza and Whittemore and Knafl's 6 review processes. Literature was searched in PubMed, Scopus, CINAHL Plus with Full Text, PsycINFO, and Embase, from 2013 to 2023. Keywords including "cancer patients," "pain," "self-management," "mHealth applications," and relevant synonyms were used in the search. The Johns Hopkins research evidence appraisal tool was used to evaluate the quality of eligible studies. A narrative synthesis was conducted to analyze the extracted data.

Results: A total of 20 studies were included, with the overall quality rated as high (n=15) to good (n=5). Using mHealth apps to monitor and manage pain was acceptable for most patients with cancer. The internal consistency of the mHealth in measuring pain was 0.96. The reported daily assessment or engagement rate ranged from 61.9% to 76.8%. All mHealth apps were designed for multimodal interventions. Participants generally had positive experiences using pain apps, rating them as enjoyable and user-friendly. In addition, 6 studies reported significant improvements in health outcomes, including enhancement in pain remission (severity and intensity), medication adherence, and a reduced frequency of breakthrough pain. The most frequently highlighted roles of mHealth apps included pain monitoring, tracking, reminders, education facilitation, and support coordination.

Conclusions: mHealth apps are effective and acceptable in supporting pain self-management. They offer a promising multi-model approach for patients to monitor, track, and manage their pain. These findings provide evidence-based insights for leveraging mHealth apps to support cancer pain self-management. More high-quality studies are needed to examine the effectiveness of digital technology–based interventions for cancer pain self-management and to identify the facilitators and barriers to their implementation in real-world practice.

(JMIR Mhealth Uhealth 2024;12:e53652) doi:10.2196/53652

KEYWORDS

cancer pain; self-management; mHealth applications; integrative review; cancer survivors

Introduction

Cancer remains a significant health concern in the United States [1]. The cost of cancer health care use in the United States was US $208.9 billion in 2020 and is expected to rise to US $246 billion by 2030 [2]. With advances in cancer diagnosis and treatment and increased survivorship rates, there is a high prevalence of distressing pain, with a pooled prevalence rate of 40% [3]. Cancer pain can persist for months and even years, significantly eroding the quality of life [4-7]. The American Society of Clinical Oncology Practice guideline advocates patient-driven self-management as the primary pain management strategy [8-10]. However, managing cancer pain is a complex and multifaceted experience that poses numerous challenges for patients. These challenges include fear of opioid addiction, insufficient knowledge or skills, and a lack of health care professional supervision [11,12]. In the rapidly advancing digital technology era, it is worth considering leveraging mHealth apps to support patients with evidence-based resources, addressing these concerns, and empowering them with self-management skills to meet personal and social needs [13,14].

The mobile health (mHealth) app is a promising tool for supporting patients in self-managing pain and improving health outcomes due to its popularity, convenience, accessibility, personalization, and cost containment [15,16]. Studies have shown that mHealth interventions could improve medication adherence, self-management engagement, and health outcomes [17-19]. A mixed methods study suggested that individuals with advanced illnesses could greatly benefit from mHealth monitoring systems, which offer continuous patient assessment and critical symptom review information to optimize health outcomes [20]. According to IQVIA's digital health trends report in 2021, over 350,000 health apps were available in app stores, comprising 47% of all apps, with an increase of about 250 apps per day [21]. However, a content review in 2020 identified only 119 designed for patients with cancer among the thousands available in major mobile app marketplaces [22]. Moreover, a review conducted in 2017 identified 46 apps geared toward clinicians for palliative care guidelines, advance care planning, pharmaceutical tools, and sharing the latest news and opinions related to palliative care [23]. However, the same research team identified only 25 palliative care apps designed specifically for patients or families [24]. Another systematic review in 2021 found that only 101 out of 1189 apps included symptom-tracking features for patients with cancer [25]. However, research on mHealth in cancer pain management is limited. To date, only one systematic review has been done to evaluate the effectiveness of mHealth in managing cancer pain [26]. Reviews of acceptability and end users' experiences have not been reported. Therefore, this review aims to bridge the research gap and understand the acceptability, effectiveness, and roles of mHealth apps in supporting cancer pain self-management.

Methods

Theoretical Preparation

An integrative review is a broad review method that includes data from various research designs to comprehensively understand a phenomenon or interest [27]. The conducting of an integrative review involves a broad, flexible, and interpretative approach in six key steps [28,29]: (1) formulate purpose or review questions; (2) systematically search and select qualitative studies, quantitative studies, and mixed methods studies using predetermined criteria; (3) perform quality appraisal of included studies; (4) narratively and interpretively analyze and synthesize findings; (5) synthesize key and themes and offer insights into the broad implications of the findings; and (6) disseminate plans of findings to audience with diverse interests and applications.

Literature Searching Strategies

As indicated in the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram, this integrative review identified eligible studies by conducting literature searches in an electronic database and manual reference tracking of eligible articles. With the assistance of librarians, we conducted a literature search using both keywords and controlled vocabulary searches in CINAHL, PubMed, PsycINFO, Scopus, and Embase. Keywords and controlled vocabulary used for the search include "cancer patients" or "oncology patients" or "patients with cancer" or "cancer survivors" AND "mobile application" or "mobile app" or "mHealth app" or "mHealth application" or "eHealth app" or "eHealth application" or mHealth or eHealth or "cellular phone" or "cell phone" or "smartphone" AND "pain self-management" or "pain management" or "pain self-care" or "pain relief" or "pain control" or "pain reduction." Search strategies and results are detailed in 2 appendices (Multimedia Appendices 1 and 2). Upon identifying the included articles, a manual reference tracking method was implemented to identify any additional eligible studies.

Study Selection Criteria

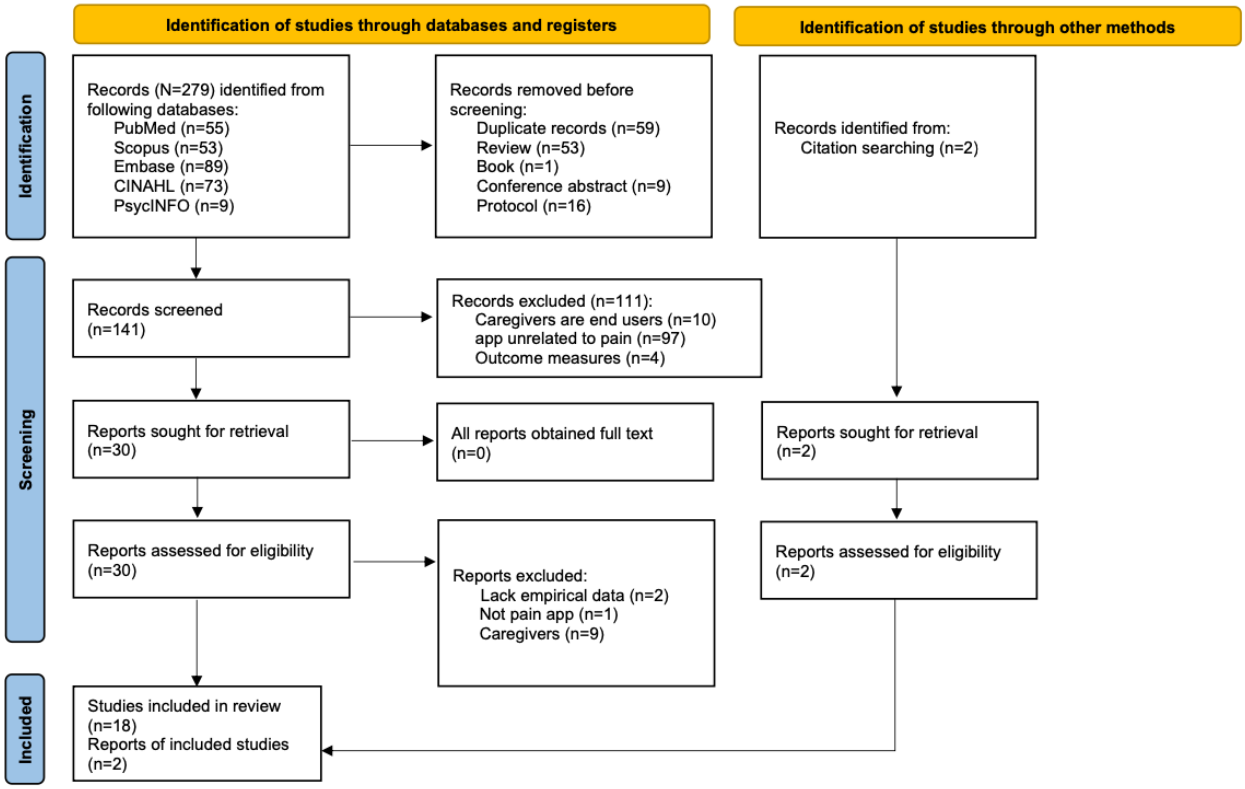
Studies were eligible for inclusion if they met the following criteria: (1) original empirical studies using qualitative, quantitative, or mixed methods design; (2) the study population included individuals of all ages diagnosed with any type of cancer; (3) assessing the effectiveness, acceptability, and cancer patients' experiences of mobile apps in cancer pain assessment and management; (4) primary outcomes focused on pain-related health outcomes, app acceptability, or users' experience; and (5) English-written, peer-reviewed, full-text articles published between 2013 and 2023.

Studies were excluded if (1) apps were exclusively intended for health professionals and caregivers as end users; (2) patients could not use the app independently for pain management; and (3) apps were used solely for delivering interventions, such as videoconferencing apps.

Study Selection Procedure

This integrative review used the PRISMA 2020 flow diagram to report the study selection process (Figure 1). The first author and an experienced librarian conducted the literature search. A total of 279 studies were retrieved by applying the searching strategies as mentioned above. The EndNote program was used to organize the studies and exclude duplicates. Two researchers reviewed the title and abstract of the studies to determine eligibility. The initial screening excluded non-English, nonempirical data, conference abstracts, book documents, reviews, and protocols. Of the 30 remaining papers, 2 researchers did the second round of screening by independently reviewing the full text of each study to determine the eligibility. Discrepancies between reviewers were resolved by discussion and consensus. A total of 18 studies were included after the second-round screening. In addition, 2 studies were identified by tracking the reference lists of included studies, bringing the total number of included articles to 20.

Figure 1. The PRISMA flow diagram. This figure provides the details of (1) identification source (the database searches and reference tracking), (2) the stepwise screening of 281 records (initial screening and abstract, title, and full text), and (3) included 20 records in the review. PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.



Quality Assessment

The quality assessment of 20 included studies was conducted using the Johns Hopkins Research Evidence Appraisal Tool, which is commonly used to appraise the qualities of various study designs [30]. The tool consists of distinct checklists to evaluate the quality of quantitative, qualitative, and mixed methods studies. Each checklist includes questions that facilitate determining the quality rated as A, B, or C (high, good, or low quality) and the level of evidence ranging from I (randomized controlled trial [RCT]) to III (nonexperimental-qualitative). Low-quality studies were excluded. Quantitative studies are assessed by a 14-item checklist based on factors including sample size, result consistency, control measures, conclusiveness, and literature review depth. A low-quality quantitative study refers to little evidence with inconsistent results, an insufficient sample size for the study design, and the inability to draw meaningful conclusions. Qualitative studies are assessed using a 13-item checklist emphasizing transparency, diligence, verification, self-reflection, participant-driven inquiry, and insightful interpretation. Low-quality qualitative studies exhibit a lack of clarity and coherence in reporting, lack of transparency in reporting methods, poor interpretation of data, and offer little insight into the phenomena of interest. Mixed methods studies require separate appraisals of both the quantitative and qualitative components and how well the design addresses the research questions. Low quality in mixed methods studies refers to good to low quality of separate quantitative and qualitative components, low relevance of study design, poor levels of integration of data or results, and no consideration of limits of integration. The first author and corresponding author conducted quality assessments and ratings.

Data Extraction and Synthesis

The authors conducted a comprehensive and iterative review of the included studies to extract overarching findings, and the results were reported following the PRISMA guideline (Multimedia Appendix 3). Table 1 presents the characteristics of the included studies. Due to the heterogeneity of the included studies, a narrative content analysis was conducted to analyze

the extracted quantitative and qualitative data. The synthesis captured the patients' acceptability of mHealth apps, the effectiveness of the targeted outcomes, the features of the apps from the patient's perspective, and how these features are achieved in the apps (Table 2).

Table 1. Basic information of each included study.

Study	Country	Cancer pain context	Cancer type	Age group (years)	Sample size, n
Yang et al [31]	China	Breakthrough pain	General	I[a]: mean 51.10 (SD 8.98), C[b]: mean 53.96 (SD 8.58)	58
Wilkie et al [32]	United States	General	General	Mean 68.4 (SD 14)	234
Hunter et al [33]	United States	General	General, with ALL[c] (71%)	I: mean 12.25 (SD 3.58), C: mean 11.86 (SD 3.44)	48
Jibb et al [34]	United States	General	General	Mean 14.2 (SD 1.7)	40
Stinson et al [35]	United States	General	General	S1[d]: mean 13.1 (SD 2.9), S2: mean 14.8 (SD 2.8)	106
Villegas et al [36]	Spain	Breakthrough pain	General	Mean 56.95 (SD 10.53)	21
Oldenmenger et al [37]	Netherlands	General	General	Mean 59 (SD 11, range 25-76)	84
Tiozzo et al [38]	Italy	General	Hematologic or solid tumors	Mean 9.1 (SD 5.4, range 0-21)	124
Fu et al [39]	United States	Lymphedema-related pain	Breast cancer	Mean 56.7 (SD 10.6)	120
Lin et al [40]	China	Oral pain	Head and neck cancer	I: mean 49.29 (SD 11.53), C: mean 50.03 (SD 9.21)	64
Salmani et al [41]	Iran	Abdominal pain	Colorectal cancer	Mean 57.18 (SD 17.47)	17
Jibb et al [42]	United States	General	General	Range 12-17	20
Hochstenbach et al [43]	Netherlands	General	General	Mean 53 (SD 15)	11
Bernier Carney et al [44]	United States	Chemotherapy-related pain	General	Median 8 (IQR 6-12)	19
Simon et al [45]	Netherlands	Chemotherapy-related pain	General	Mean 7.33 (SD 5)	27
Jibb et al [46]	United States	General	General	Mean 14.8 (SD 2.1)	16
Fortier et al [47]	United States	General	General	Mean 12.33 (SD 3.42)	12
Azizoddin et al [48]	United States	Advanced pain	Advanced cancer	Adults	14
Alberts et al [49]	Canada	General	Survivors of childhood cancer	I: mean 43.1 (SD 6.9), C: mean 45.0 (SD 10.1)	87
Stinson et al [50]	United States	General	General	S1: mean 13.9 (SD 1.9), S2: mean 13.4 (SD 2.9), S3: mean 13.2 (SD 2.3)	47

[a]I: intervention.

[b]C: control.

[c]ALL: acute lymphoblastic leukemia.

[d]S: study.

Table 2. Methodology and main findings of included studies.

Study	Study design	Outcome measures	Findings	Appraisal	
				L[a]	Q[b]
Yang et al [31]	RCT[c]	• Primary: app effectiveness • Secondary: feasibility	• Pain remission • Breakthrough pain and adverse reactions reduced • QoL[d] improved • Patients were satisfied	I	A
Wilkie et al [32]	Stepped-wedge RCT	• Primary: pain intensity and analgesic adherence • Secondary: pain misconception	• The intervention effect was not significant. • 62% of patients viewed the videos on pain misconception	I	A
Hunter et al [38]	Quant[e]	• Daily pain assessment • Pain intensity	• 61.2% completed daily dairy • Reports of daily average pain were not significant, but fewer moderate to severe pain	III	A
Jibb et al [35]	QED[f]	• Primary: feasibility test • Secondary: pain intensity, interference, and QoL	• Trends in improvements in pain intensity, pain interference, and QoL • Mean adherence to pain reporting was 68.8 (SD 38.1%)	II	A
Stinson et al [39]	Quant	• Construct reliability and validity	• The correlation between pain reports on the app and recall was moderate to high (0.43-0.68) • The app's internal consistency is 0.96	III	A
Villegas et al [36]	QED	• Daily pain assessment • Usability (System Usability Scale)	• Adherence: 61.9% (n=13/21) used the app daily during the 30-day study • Breakthrough pain was less frequent • Usability: the mean score of System Usability Scale was 85.77/100 (SD 12.09)	II	A
Oldenmenger et al [40]	Quant	• Feasibility in a pain diary, pain education, and eConsult	• Pain intensity decreased • Patients completed the diary for at least 65% of the days • Monitoring of pain via the Internet is feasible	III	A
Tiozzo et al [41]	Quant	• Pain intensity and characteristics • App usage satisfaction	• Significant pain relief • 94 (75.8%) reported pain at least once per month • Most patients were satisfied with the app	III	A
Fu et al [33]	RCT	• Pain reduction and QoL	• Significant benefits for breast cancer to manage chronic pain soreness, arm and hand swelling, heaviness, and impaired limb mobility	I	A
Lin et al [37]	QED	• Pain level and EORCT-QoL	• Significant pain relief in the app group at T2 and T3 and significantly higher QoL at T3[g]	II	A
Salmani et al [42]	Quant	• Usability evaluation	• Average score: 8.03 out of 9 • Overall reaction: 7.94 • Screen design and layout: 8.18 • Systems information: 7.97 • Learnability: 7.98 • System feature: 8.12	III	B
Jibb et al [43]	Qual[h]	• Perceptions of adolescent acceptability, satisfaction, and suggestions for improvement	• Enjoy using app • Endorse pain advice • Facilitate communication with providers • Therapeutic benefit • Improved awareness of pain	III	A
Hochstenbach et al [46]	MMR[i]	• App feasibility and patients' experience	• Learnability (4.8/5), usability (4.8/5), and desirability (4.6/5) • Patients were pleased with the simplicity and different components	III	B

Study	Study design	Outcome measures	Findings	Appraisal	
				L[a]	Q[b]
Bernier Carney et al [50]	MMR	• Pain severity and distress • Qualitative pain descriptions	• Children are willing to describe their ambulatory pain experiences on a game-based mobile app through quantitative reports and qualitative description	III	B
Simon et al [47]	MMR	• App adherence and feasibility • Barriers and facilitators of implementation	• 63% (N=17) used daily for 3 weeks • Three facilitators: technical functioning, impact on pain care, and user-friendliness of the app • Three barriers: technical problems with daily reminders, content and functionalities, and user-friendliness	III	B
Jibb et al [45]	Qual	• Efficiency • Ease of use and understanding • Utility • Acceptability • Usability	• The time to complete the pain assessment was 4.3 minutes • Easy to use and understand • Endorse design gamification and customizability • Valued content and navigation	III	A
Fortier et al [48]	MMR	• Content and usability • Patients' satisfaction	• Highly satisfied with the program • The 3D Avatar design was attractive • The skills training was useful	III	A
Azizoddin et al [44]	Qual	• Review wireframes of the content and its delivery	• Primary themes: (1) clarity, (2) visual appeal, (3) usefulness, and (4) engagement	III	A
Alberts et al [34]	RCT	• Feasibility and acceptability • Pain intensity and interference	• 90.3% (n=28) wore the device >50% of the trial (mean 21.8/30 days [SD 5.9]). 74.2% (23/31) were satisfied with the device. • Average pain relieved but not significant. • Facilitators: easy to use, beneficial, learning new ways, increased awareness, appreciation	I	A
Stinson et al [49]	MMR	• Usability, feasibility, compliance, and satisfaction	• Appealing to adolescents • Endorsed game-based and virtual reward systems • High compliance • Likable, easy to use, not bothersome	III	A

[a]L: the level of evidence ranging from I (randomized controlled trial) to III (nonexperimental-qualitative).

[b]Q: quality rated as A, B, or C (high, good, or low quality).

[c]RCT: randomized controlled trials.

[d]QoL: quality of life.

[e]Quant: quantitative.

[f]QED: quasi-experimental design.

[g]The researcher collected data at four time points: before treatment (T0), and the second week (T1), the first month (T2), and the second month (T3) after the start of treatment.

[h]Qual: qualitative.

[i]MMR: mixed methods research.

Results

Included Studies Characteristics

The characteristics and methodology of the 20 studies were detailed in Tables 1 and 2. The included studies involved quantitative design (RCT (n=4) [31,32,39,49], quasi-experimental (n=3) [34,36,40], prospective (observational, cohort; n=4) [33,35,37,38], cross-sectional design (n=1) [41], qualitative design (n=3) [42,46,48], and mixed methods design (n=5) [43-45,47,50]). The sample sizes ranged from 11 to 234. A total of 12 studies were conducted in North America (US [n=11] and Canada [n=1]), Europe (the Netherlands [n=3], Spain [n=1], and Italy [n=1]), and Asia (China [n=2] and Iran [n=1]).

Quality of Studies

In total, 20 studies included were assessed based on the evidence level and overall quality. Table 2 shows that 4 studies were rated as IA, 3 as IIA, and 8 as IIIA. All of these studies were considered high quality despite different study designs (levels of evidence). In addition, 5 studies were rated as IIIB, which were deemed to be of good quality. No studies were rated as low quality.

Pain Context

As shown in Table 1, 6 studies were conducted to assess pain management experiences in specific cancer subgroups or among patients with similar pain subcategories, such as management of oral pain in head and neck cancer [40], abdominal pain in colorectal cancer [41], lymphedema-related chronic pain in breast cancer [39], advanced cancer pain [48], and breakthrough pain [31,36]. In addition, 8 studies were conducted in specific age subgroups other than adult patients with cancer. These included studies involving children [33,38,45,47], school-age children [44], adolescents [34,42], and adult survivors of childhood cancer [49]. Four studies investigated pain management under specific cancer treatment, such as chemotherapy [44,45,47] or concurrent radiotherapy [40].

App Context

All pain management apps were developed by multidisciplinary teams that consisted of medical oncologists, palliative care nurse specialists, researchers, and app developers. These apps were created in one of 3 formats, that are (1) dedicated pain management apps (n=7), (2) pain management as the primary module in a comprehensive self-care app (n=5), and (3) pain management integrated into an existing app (n=2). For instance, "Pain Buddy" is a dedicated cancer pain app that helps children with pain self-management [33,47]. "ColorectAlong" is a comprehensive self-care app that includes pain management as one of its 8 components [41]. In addition, Villegas et al [36] detailed the integration of pain management features into an existing app.

App Feasibility

Overall, 18 studies assessed the feasibility of using mHealth apps in cancer pain management by evaluating usability, acceptability, fidelity, learnability, satisfaction, and desirability. Results indicated that the real-time pain assessment was efficient, valid, and reliable [33,35-38,42,43,45,46,49,50]. One study reported that the average time to complete the pain assessment was 4.3 (SD 3.5) minutes [46]. Another study observed a moderate to vigorous (0.43-0.68) correlation between weekly pain average ratings recorded on the app and retrospective weekly average pain ratings, indicating a high level of internal consistency over 2 weeks (standard Cronbach α=0.96) [35]. Using mHealth apps to monitor or manage pain was acceptable for most patients with cancer [33,37,38,43,45,49]. In total, 5 studies reported adherence to pain assessment during the trial periods (ranging from 2 weeks to 1 month) and found that 61.9%-76.8% of participants completed daily pain assessment. The remaining participants used the app for a shorter period but for at least half of the trial days (minimum 7 days) [33,37,38,43,45]. One study reported that 81% of participants wore the wearable device throughout the 30-day trial period [49]. Another study showed that 18 participants (45%) continued to complete pain assessments and receive treatment advice beyond the study trial [33]. The overall satisfaction with the apps was high, with an average score of 8.9 out of 10 [38], 8.0 out of 9 [41], 4.8 out of 5 [43], and 85.8 out of 100 [36]. Qualitative studies indicated participants' positive experiences with pain apps, such as likable, enjoyable, easy to use, and not bothersome to complete [42,50].

> *It is really appealing to the eye. The color, the theme, and the font are good. And it's not really that hard to understand. The vocabulary is really straightforward, and all of the things are on it. The multiple-choice questions and the [visual analog scale] sliders are really easy to use.* [42]

Effectiveness of mHealth Apps on Cancer Pain Self-Management

A total of 17 studies demonstrated the effectiveness of the mHealth app in supporting pain outcomes and self-management. Of the 9 studies evaluating the efficacy of pain-related outcomes, 6 reported significant effects of mHealth apps on pain remission (pain severity and pain intensity), improvement in pain medication adherence, reduction of adverse reaction and occurrence of breakthrough pain, and improvement in quality of life [31,34,36,37,39,40]. Two studies observed patterns of pain reduction, although not statistically significant (P>.05) [32,49]. In contrast, 1 study showed no significant difference in average daily pain reduction between the intervention and control groups but noted fewer instances of moderate to vigorous pain in the intervention group [33].

In addition, 8 studies included qualitative data from users' perspectives on cancer pain self-management apps and detailed descriptions of the app. Overall, patients highly valued and emphasized the significant roles of mHealth apps in their daily pain management. A content analysis was conducted to comprehensively understand users' perspectives and how these app features were implemented. The 3 primary features of pain apps in assisting patients with cancer pain self-management were identified and summarized as (1) pain monitoring, tracking, and reminder, (2) pain education facilitation, (3) pain support coordination.

Pain Monitor, Tracker, and Reminder

The primary benefit for patients using mobile apps was monitoring, tracking, and reminding them of their pain management. These apps measured patients' pain in their daily lives and provided continuous real-time monitoring of pain trends, enabling patients to adjust their medication and management plan [42,47].

> *It is helpful when the physician calls. I look at the graph to get a good picture. It also gives justification that I'm not exaggerating my pain* [43]

Pain apps offer various features to assist patients in managing their pain. Three apps provided daily reminders and precisely measured and recorded the cause, severity, intensity, location, nature, type, and duration of pain and the frequency of breakthrough pain, medication taking, and adverse reactions [31,37,40]. The apps used various instruments to assess pain, including the numerical rating scale (NRS) [31,37,45], the Ospedale Pediatrico Bambino Gesù tool [38], the visual analog scale [40,45], the pain diary [37], effective pain descriptors [38], free-text responses [44], and body maps for pinpoint location of pain [31,44,46]. In addition, the app Pain Buddy directed children to pinpoint the pain on an avatar and included a drawing feature to specify the location of the pain in more detail [44].

Pain Education Facilitator

The second common function of the apps was to provide pain education. Patients noted that pain educational modules in these apps improved their awareness and interest in pain management. There were 2 studies that reported 100% of users actively engaging in the learning module [42,43].

> *I thought the pain help ideas were awesome. They would suggest activities like relaxation and breathing. And when you click on it and, there is someone talking to you, walking you through it. Like how to relax* [42]

Most pain apps incorporate psychoeducational modules that focus on promoting comprehensive pain knowledge and self-management skills based on clinical or World Health Organization guidelines [31,32,36,37,39-43,45,47,49]. A total of 3 apps included features to evaluate patients' misunderstanding of pain and offered customized information concerning the fundamental causes of pain and the appropriate treatment methods [31,32,37]. In addition, 7 apps included features for self-management skill training, such as medicine instructions, music relaxation treatment or acupuncture [31,36,41], cognitive and behavioral skills training [47], breathing exercises [49], and step-by-step lymphatic exercises for patients with breast cancer [39]. Children users also found the skills training helpful, with belly breathing and distraction techniques being their favorite skills [47]. The trial results of the app, which focused on educating users about oral mucositis knowledge and care skills, showed that the group with access to these resources experienced significantly lower pain levels [40].

Notably, pain apps served as a distraction for many app users, as reported by participants who found it "fun to do" and a "positive challenge" to monitor their pain without constantly focusing on their pain [43,48]. The efficacy of these apps in providing distraction stemmed from the primary principles of app design, emphasizing the engagement and enjoyment of users while considering the unique characteristics and needs of different age groups. For instance, children found the design of 3D Avatar more attractive and enjoyable and were motivated to earn "coins" to customize the program further [46,47]. Adolescents, on the other hand, may be more inclined to gamification elements, including role-playing, badge acquisition, and point and leveling system [35,46], and were attracted by the "appealing to the eye" elements, such as color scheme, fonts, and graphics used [35,42].

Pain Support Coordinator

Patients reported that pain apps strengthened their partnerships with health care providers [36,41,42,45]. In particular, some apps offered the feature allowing patients to receive direct phone calls from health care providers when they reported experiencing severe pain, and self-management strategies were ineffective [45]. These features permitted direct and effective communication between patients and the health care team and were highly valued among patients [36]. Further, patients found pain apps enabled more efficient conversations with their health care professionals during office visits by providing precise symptom patterns and notes recorded in daily assessments [41,42].

> *We don't want to call the hospital all the time. With the app, you get the sense that pain is being monitored, and they call us when we report high pain scores. That is very comforting. It gives you the sense that you're being taken care of* [45]

Most apps (n=8) allow open dyad access, providing both the patient and their health care provider access to the apps [31,32,36,40,42,43,45,47]. In cases where patients reported a high pain score (over 5/10 or NRS≥4) or moderate to vigorous side effects, the system would automatically remind patients to take medication. One hour later, the system automatically reminded patients to reassess pain [31]. If a high pain score (over 5/10 or NRS≥4) was still reported, an email alert or clinical alarm was sent to clinical professionals to notify the uncontrolled pain condition [36,37]. The medical team could contact the patient directly for alarms when necessary [36,42]. In 2 RCTs, patients assigned to the app group with this alarm feature reported significantly lower frequencies of breakthrough pain and higher medication adherence than the control group and were more likely to promptly detect and address pain exacerbation [31,36]. Clinicians also acknowledged that the app improves treatments' safety, adherence, and effectiveness for managing breakthrough pain [36].

The eConsult module was embedded in several apps, with an email-like format or social network links, facilitating question-and-answer communication between patients and professionals [32,37,40]. Participants who frequently used eConsults found they were beneficial for pain management [37]. For instance, participants could obtain assistance from this module in setting pain management goals, devising action plans, and identifying in-time coping strategies for breakthrough pain when patients reported high scores [31,36,41]. Outpatients, in particular, appreciated the app's therapeutic benefit, as it allowed them to receive real-time support and efficient pain management advice without being constrained by time and space limitations [42,48].

> *It's a fantastic idea. As one who was living in constant pain, I was not the one to call the doctor. If I had this resource available. Things maybe would have changed for me a lot faster than they (doctors) did* [48]

Discussion

Principal Findings

This integrative review synthesizes the current evidence of mHealth apps supporting pain outcomes and self-management of patients with cancer. Overall, mHealth apps offer significant benefits for managing cancer pain and serve as multi-model interventions that provide critical features such as monitoring, tracking, reminders, education facilitation, and support coordination. These findings offer evidence-based insights into effectively leveraging the advantage of mHealth apps in supporting the pain self-management of patients with cancer.

Implications

mHealth apps are effective and acceptable in alleviating pain and supporting self-management, as indicated by participants' feedback and feasibility data. This finding added complementary evidence to systematic reviews that focused exclusively on experimental studies [24,51]. However, integrating these apps into standard clinical care components, such as electronic health records, has posed significant challenges due to factors such as inconsistent app availability [51], standard application programming interface development and adoption in the early stages [52], and security concerns [53]. Despite these challenges, nursing, medical scientists, and other stakeholders have been urged to direct their efforts toward creating a long-term strategic plan for developing and implementing eHealth services, and promoting equitable, accessible, and affordable health care [54]. With the outbreak of COVID-19, videoconferencing apps such as Skype, Zoom, Facebook, 2-way text messages, and other online platforms have played a critical role in transitioning offline health care services to an online format [55]. For example, the electronic patient-reported outcomes (ePROs) system significantly improved health care efficiency, enhanced patient-doctor interaction, and optimized medical resource use [56,57]. Cancer self-management apps have the potential to provide easy access to cancer health care services by eliminating time and space limitations, giving patients greater autonomy and control, and providing more precise or personalized strategies [58]. Further studies are necessary to explore the integration of ePROs into mHealth apps for managing cancer pain.

mHealth apps were promising tools to empower patients with cancer with the necessary knowledge and skills and enable them to manage their pain actively. Patients endorsed and mainly engaged with the educational modules, training programs, and coping strategies offered by pain apps. This finding was consistent with previous studies that patients expressed their educational needs in areas of opioid analgesics, long-term survivorship, and relying on professional service or guidance rather than nonprofessional sources like nonprofessional internet pages or television programs [59,60]. Compared with traditional paper-format educational material, mHealth apps lower the learning threshold and burden while increasing engagement through numerous attractive features such as straightforward, enjoyable, informative, accessible, and personalized designs [61]. Apps can also customize education modules based on patient's health literacy and ability to use the app over time with the advancement of artificial intelligence and big data [58]. Also, mHealth apps have emerged as a promising tool for translating evidence-based practice into the home setting. The included apps in this review reported that all educational modules were developed based on World Health Organization or clinical or evidence-based guidelines. This approach promotes science communication to the public and facilitates more informed decision-making [62].

mHealth apps could also serve as a distraction technique, drawing patients' attention away from the mental processing of pain through attractive features. Patients always viewed their cancer journey as life trauma, especially when accompanied by long-term unpleasant pain experiences [63]. Distraction, a simple nonpharmacological technique, is increasingly being examined as an effective intervention. It is applied primarily in pediatric oncological procedure pain [64], during needle insertion or lumbar puncture [65], or subcutaneous port access [66]. With the advancement of digital technology, virtual reality distraction has shown promise in reducing self-reported pain in patients with breast cancer [67], as well as children and adolescents with cancer [66]. Therefore, the app design should fully consider incorporating engagement techniques, that cater to diverse end users' characteristics, and pain care demands.

mHealth apps had the potential to bridge gaps in health care access and facilitate interaction and communication with health care providers. Previous studies highlighted patients' challenges in accessing professional survivorship care after completing clinical treatment [68,69]. mHealth app serves as an eConsult medical chatbot, allowing patients to seek professional consultation to address daily minor issues or confusion, thus minimizing unnecessary clinic visits [70]. In addition, mHealth apps with alert systems can help identify urgent situations that require immediate health intervention. Implementing ePROs within clinical practice enables remote monitoring and early detection of severe and worsening symptoms [71]. One clinical randomized trial illustrated that web-based ePRO tools are feasible and acceptable among patients with advanced cancer without increasing clinical burden [72]. However, it would be valuable to explore clinical staff's perspectives on whether these applications increase or reduce their workload. Furthermore, with the tracking function of pain apps, patients can precisely describe the trends and characteristics of their pain, enhancing the quality of communication and problem-solving efficiency between patients and professionals in pain management. The findings from another systematic review also supported the idea that practical technology tools can help strengthen communication and partnership between patients and providers [73]. For instance, a 2-way text message strategy has been shown to be effective in improving engagement and adherence in the survivorship management program, although it may require the research team to commit more effort to respond to text messages personally [74,75]. Building effective partnerships with providers is one core skill in promoting successful cancer pain self-management [76]. Further studies are needed to evaluate the cost-effectiveness of such interaction designs within the health care system across different contexts and populations.

Limitations

Several limitations exist in this review that warrant careful consideration. Small sample sizes and short study durations in the preliminary usability testing may limit the generalization of the study findings. There is a risk of a skewed perception of the intervention's success or survivor bias, particularly if individuals who did not adhere to long-term trials or lost interest in the app interventions were not adequately represented in the data. Further research is imperative to establish causality and generalizability of the findings. It is important to note that the review primarily focuses on patients' perspectives. Additional reviews focusing on insights from other stakeholders, such as health care professionals and caregivers, are needed to achieve a more holistic understanding before disseminating findings into clinical practice. The heterogeneity of the included studies,

including varied designs, pain measures, and app functionalities, along with limited evidence from RCTs, inherently constrains the level of analysis and evidence. Despite meticulous search efforts, limitations in search strategy, including search terms, databases used, inclusion and exclusion criteria, and the fast-paced development of technology, could lead to unintentionally omitting relevant new apps. Finally, while the review predominantly draws from academic literature, a more comprehensive understanding could be gained by incorporating insights from market app stores and usage reports.

Conclusion

Overall, mHealth apps are effective and acceptable in supporting the self-management of cancer pain. They offer a promising approach for patients to monitor, track, and manage their pain and receive multimodel interventions to promote pain self-management. These findings provide evidence-based insights for leveraging the features of mHealth apps in supporting cancer pain self-management. More high-quality studies are needed on the effectiveness of digital technology–based interventions for cancer pain self-management and to identify the facilitators and barriers to their implementation in real-world practice.

Acknowledgments

We would like to express our sincere gratitude to all the authors who contributed to this manuscript's conceptualization, methodology, writing, reviewing, and editing. This work was made possible by the Oncology Nursing Foundation, which provided research support through the RE01 grant in 2019. No specific grant number was assigned. The corresponding author, WX, received this research support, which was crucial in conducting this review and would cover potential publication fees.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Searching Strategies.
[DOCX File , 16 KB - mhealth_v12i1e53652_app1.docx]

Multimedia Appendix 2
Details of search database, syntax, and results.
[DOCX File , 22 KB - mhealth_v12i1e53652_app2.docx]

Multimedia Appendix 3
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOCX File , 35 KB - mhealth_v12i1e53652_app3.docx]

References

1. Miller KD, Nogueira L, Devasia T, Mariotto AB, Yabroff KR, Jemal A, et al. Cancer treatment and survivorship statistics, 2022. CA Cancer J Clin 2022;72(5):409-436 [FREE Full text] [doi: 10.3322/caac.21731] [Medline: 35736631]
2. Mariotto AB, Enewold L, Zhao J, Zeruto CA, Yabroff KR. Medical care costs associated with cancer survivorship in the United States. Cancer Epidemiol Biomarkers Prev 2020;29(7):1304-1312 [FREE Full text] [doi: 10.1158/1055-9965.EPI-19-1534] [Medline: 32522832]
3. Evenepoel M, Haenen V, De Baerdemaecker T, Meeus M, Devoogdt N, Dams L, et al. Pain prevalence during cancer treatment: a systematic review and meta-analysis. J Pain Symptom Manage 2022;63(3):e317-e335. [doi: 10.1016/j.jpainsymman.2021.09.011] [Medline: 34563628]
4. Börjeson S, Starkhammar H, Unosson M, Berterö C. Common symptoms and distress experienced among patients with colorectal cancer: a qualitative part of mixed method design. Open Nurs J 2012;6:100-107 [FREE Full text] [doi: 10.2174/1874434601206010100] [Medline: 22977653]
5. Jones D, Zhao F, Brell J, Lewis MA, Loprinzi CL, Weiss M, et al. Neuropathic symptoms, quality of life, and clinician perception of patient care in medical oncology outpatients with colorectal, breast, lung, and prostate cancer. J Cancer Surviv 2015;9(1):1-10 [FREE Full text] [doi: 10.1007/s11764-014-0379-x] [Medline: 25023039]
6. Révész D, van Kuijk SMJ, Mols F, van Duijnhoven FJB, Winkels RM, Hoofs H, et al. Development and internal validation of prediction models for colorectal cancer survivors to estimate the 1-year risk of low health-related quality of life in multiple domains. BMC Med Inform Decis Mak 2020;20(1):54 [FREE Full text] [doi: 10.1186/s12911-020-1064-9] [Medline: 32164641]
7. Vardy J, Dhillon H, Pond G, Renton C, Dodd A, Zhang H, et al. Fatigue in people with localized colorectal cancer who do and do not receive chemotherapy: a longitudinal prospective study. Ann Oncol 2016;27(9):1761-1767 [FREE Full text] [doi: 10.1093/annonc/mdw252] [Medline: 27443634]

8. Paice JA, Portenoy R, Lacchetti C, Campbell T, Cheville A, Citron M, et al. Management of chronic pain in survivors of adult cancers: American Society of Clinical Oncology clinical practice guideline. J Clin Oncol 2016;34(27):3325-3345. [doi: 10.1200/JCO.2016.68.5206] [Medline: 27458286]
9. Lim CYS, Laidsaar-Powell RC, Young JM, Kao SC, Zhang Y, Butow P. Colorectal cancer survivorship: a systematic review and thematic synthesis of qualitative research. Eur J Cancer Care (Engl) 2021;30(4):e13421. [doi: 10.1111/ecc.13421] [Medline: 33733545]
10. McCorkle R, Ercolano E, Lazenby M, Schulman-Green D, Schilling LS, Lorig K, et al. Self-management: enabling and empowering patients living with cancer as a chronic illness. CA Cancer J Clin 2011;61(1):50-62 [FREE Full text] [doi: 10.3322/caac.20093] [Medline: 21205833]
11. Bilodeau K, Tremblay D, Durand MJ. Gaps and delays in survivorship care in the return-to-work pathway for survivors of breast cancer-a qualitative study. Curr Oncol 2019;26(3):e414-e417 [FREE Full text] [doi: 10.3747/co.26.4787] [Medline: 31285687]
12. Bae KR, Cho J. Changes after cancer diagnosis and return to work: experience of Korean cancer patients. BMC Cancer 2021;21(1):86 [FREE Full text] [doi: 10.1186/s12885-021-07812-w] [Medline: 33478405]
13. Kiza A, Cong X. Adults' self-management of chronic cancer and noncancer pain in people with and without cognitive impairment: a concept analysis. Pain Manag Nurs 2021;22(1):69-73. [doi: 10.1016/j.pmn.2020.08.007] [Medline: 33132039]
14. Lorig KR, Holman HR. Self-management education: history, definition, outcomes, and mechanisms. Ann Behav Med 2003;26(1):1-7. [doi: 10.1207/S15324796ABM2601_01] [Medline: 12867348]
15. Birkhoff SD, Smeltzer SC. Perceptions of smartphone user-centered mobile health tracking apps across various chronic illness populations: an integrative review. J Nurs Scholarsh 2017;49(4):371-378. [doi: 10.1111/jnu.12298] [Medline: 28605151]
16. Anderson K, Burford O, Emmerton L. Mobile health apps to facilitate self-care: a qualitative study of user experiences. PLoS One 2016;11(5):e0156164 [FREE Full text] [doi: 10.1371/journal.pone.0156164] [Medline: 27214203]
17. Hou I, Lin H, Shen S, Chang K, Tai H, Tsai A, et al. Quality of life of women after a first diagnosis of breast cancer using a self-management support mHealth app in Taiwan: randomized controlled trial. JMIR Mhealth Uhealth 2020;8(3):e17084 [FREE Full text] [doi: 10.2196/17084] [Medline: 32130181]
18. Rock CL, Thomson CA, Sullivan KR, Howe CL, Kushi LH, Caan BJ, et al. American Cancer Society nutrition and physical activity guideline for cancer survivors. CA Cancer J Clin 2022;72(3):230-262 [FREE Full text] [doi: 10.3322/caac.21719] [Medline: 35294043]
19. Singleton AC, Raeside R, Hyun KK, Partridge SR, Di Tanna GL, Hafiz N, et al. Electronic health interventions for patients with breast cancer: systematic review and meta-analyses. J Clin Oncol 2022;40(20):2257-2270 [FREE Full text] [doi: 10.1200/JCO.21.01171] [Medline: 35500200]
20. Castillo Padrós MR, Pastor N, Altarriba Paracolls J, Mosquera Peña M, Pergolizzi D, Salvador Vergès À. A smart system for remote monitoring of patients in palliative care (HumanITcare platform): mixed methods study. JMIR Form Res 2023;7:e45654 [FREE Full text] [doi: 10.2196/45654] [Medline: 37256664]
21. Digital health trends 2021. URL: https://www.iqvia.com/insights/the-iqvia-institute/reports/digital-health-trends-2021 [accessed 2021-07-22]
22. Charbonneau DH, Hightower S, Katz A, Zhang K, Abrams J, Senft N, et al. Smartphone apps for cancer: a content analysis of the digital health marketplace. Digit Health 2020;6:2055207620905413 [FREE Full text] [doi: 10.1177/2055207620905413] [Medline: 32110428]
23. Meghani SH, MacKenzie MA, Morgan B, Kang Y, Wasim A, Sayani S. Clinician-targeted mobile apps in palliative care: a systematic review. J Palliat Med 2017;20(10):1139-1147. [doi: 10.1089/jpm.2017.0070] [Medline: 28557549]
24. MacKenzie Greenle M, Morgan B, Sayani S, Meghani SH. Identifying mobile apps targeting palliative care patients and family members. J Palliat Med 2018;21(10):1380-1385. [doi: 10.1089/jpm.2018.0157] [Medline: 30312126]
25. Lu DJ, Girgis M, David JM, Chung EM, Atkins KM, Kamrava M. Evaluation of mobile health applications to track patient-reported outcomes for oncology patients: a systematic review. Adv Radiat Oncol 2021;6(1):100576 [FREE Full text] [doi: 10.1016/j.adro.2020.09.016] [Medline: 33073061]
26. Hernandez Silva E, Lawler S, Langbecker D. The effectiveness of mHealth for self-management in improving pain, psychological distress, fatigue, and sleep in cancer survivors: a systematic review. J Cancer Surviv 2019;13(1):97-107. [doi: 10.1007/s11764-018-0730-8] [Medline: 30635865]
27. Broome ME. Integrative literature reviews for the development of concepts. In: Rodgers BL, Knafl KA, editors. Concept Development in Nursing. Phialdelphia, PA: WB Saunders Company; 2000:231-250.
28. Whittemore R, Knafl K. The integrative review: updated methodology. J Adv Nurs 2005;52(5):546-553. [doi: 10.1111/j.1365-2648.2005.03621.x] [Medline: 16268861]
29. Souza MTD, Silva MDD, Carvalho RD. Integrative review: what is it? How to do it? Einstein (Sao Paulo) 2010;8(1):102-106 [FREE Full text] [doi: 10.1590/S1679-45082010RW1134] [Medline: 26761761]
30. Buccheri RK, Sharifi C. Critical appraisal tools and reporting guidelines for evidence-based practice. Worldviews Evid Based Nurs 2017;14(6):463-472. [doi: 10.1111/wvn.12258] [Medline: 28898556]

31. Yang J, Weng L, Chen Z, Cai H, Lin X, Hu Z, et al. Development and testing of a mobile app for pain management among cancer patients discharged from hospital treatment: randomized controlled trial. JMIR Mhealth Uhealth 2019;7(5):e12542 [FREE Full text] [doi: 10.2196/12542] [Medline: 31144672]
32. Wilkie DJ, Yao Y, Ezenwa MO, Suarez ML, Dyal BW, Gill A, et al. A stepped-wedge randomized controlled trial: effects of eHealth interventions for pain control among adults with cancer in hospice. J Pain Symptom Manage 2020;59(3):626-636 [FREE Full text] [doi: 10.1016/j.jpainsymman.2019.10.028] [Medline: 31711969]
33. Hunter JF, Acevedo AM, Gago-Masague S, Kain A, Yun C, Torno L, et al. A pilot study of the preliminary efficacy of pain buddy: a novel intervention for the management of children's cancer-related pain. Pediatr Blood Cancer 2020;67(10):e28278 [FREE Full text] [doi: 10.1002/pbc.28278] [Medline: 32743950]
34. Jibb LA, Stevens BJ, Nathan PC, Seto E, Cafazzo JA, Johnston DL, et al. Implementation and preliminary effectiveness of a real-time pain management smartphone app for adolescents with cancer: a multicenter pilot clinical study. Pediatr Blood Cancer 2017;64(10). [doi: 10.1002/pbc.26554] [Medline: 28423223]
35. Stinson JN, Jibb LA, Nguyen C, Nathan PC, Maloney AM, Dupuis LL, et al. Construct validity and reliability of a real-time multidimensional smartphone app to assess pain in children and adolescents with cancer. Pain 2015;156(12):2607-2615. [doi: 10.1097/j.pain.0000000000000385] [Medline: 26580680]
36. Villegas F, Martínez-Borba V, Suso-Ribera C, Castilla D, Zaragoza I, García-Palacios A, et al. Characterizing breakthrough cancer pain using ecological momentary assessment with a smartphone app: feasibility and clinical findings. Int J Environ Res Public Health 2021;18(11):5991 [FREE Full text] [doi: 10.3390/ijerph18115991] [Medline: 34204871]
37. Oldenmenger WH, Baan MAG, van der Rijt CCD. Development and feasibility of a web application to monitor patients' cancer-related pain. Support Care Cancer 2018;26(2):635-642 [FREE Full text] [doi: 10.1007/s00520-017-3877-3] [Medline: 28929433]
38. Tiozzo E, Fondi S, Biagioli V, Piccinelli E, Alibrandi F, Gawronski O, et al. Electronic assessment and tracking of pain at home: a prospective study in children with hematologic or solid tumors. J Pediatr Oncol Nurs 2021;38(2):82-93. [doi: 10.1177/1043454220975443] [Medline: 33269620]
39. Fu MR, Axelrod D, Guth AA, Scagliola J, Rampertaap K, El-Shammaa N, et al. A web- and mobile-based intervention for women treated for breast cancer to manage chronic pain and symptoms related to lymphedema: results of a randomized clinical trial. JMIR Cancer 2022;8(1):e29485 [FREE Full text] [doi: 10.2196/29485] [Medline: 35037883]
40. Lin T, Wang Y, Huang C. Effects of a mobile oral care app on oral mucositis, pain, nutritional status, and quality of life in patients with head and neck cancer: a quasi-experimental study. Int J Nurs Pract 2022;28(4):e13042. [doi: 10.1111/ijn.13042] [Medline: 35224811]
41. Salmani H, Nahvijou A, Sheikhtaheri A. Smartphone-based application for self-management of patients with colorectal cancer: development and usability evaluation. Support Care Cancer 2022;30(4):3249-3258. [doi: 10.1007/s00520-021-06754-0] [Medline: 34984548]
42. Jibb LA, Stevens BJ, Nathan PC, Seto E, Cafazzo JA, Johnston DL, et al. Perceptions of adolescents with cancer related to a pain management app and its evaluation: qualitative study nested within a multicenter pilot feasibility study. JMIR Mhealth Uhealth 2018;6(4):e80 [FREE Full text] [doi: 10.2196/mhealth.9319] [Medline: 29625951]
43. Hochstenbach LMJ, Zwakhalen SMG, Courtens AM, van Kleef M, de Witte LP. Feasibility of a mobile and web-based intervention to support self-management in outpatients with cancer pain. Eur J Oncol Nurs 2016;23:97-105 [FREE Full text] [doi: 10.1016/j.ejon.2016.03.009] [Medline: 27456381]
44. Bernier Carney KM, Jung SH, Iacob E, Lewis M, Linder LA. Communication of pain by school-age children with cancer using a game-based symptom assessment app: a secondary analysis. Eur J Oncol Nurs 2021;52:101949 [FREE Full text] [doi: 10.1016/j.ejon.2021.101949] [Medline: 33813185]
45. Simon JDHP, Schepers SA, Grootenhuis MA, Mensink M, Huitema AD, Tissing WJE, et al. Reducing pain in children with cancer at home: a feasibility study of the KLIK pain monitor app. Support Care Cancer 2021;29(12):7617-7626 [FREE Full text] [doi: 10.1007/s00520-021-06357-9] [Medline: 34136954]
46. Jibb LA, Cafazzo JA, Nathan PC, Seto E, Stevens BJ, Nguyen C, et al. Development of a mHealth real-time pain self-management app for adolescents with cancer: an iterative usability testing study [formula: see text]. J Pediatr Oncol Nurs 2017;34(4):283-294. [doi: 10.1177/1043454217697022] [Medline: 28376666]
47. Fortier MA, Chung WW, Martinez A, Gago-Masague S, Sender L. Pain buddy: a novel use of m-health in the management of children's cancer pain. Comput Biol Med 2016;76:202-214 [FREE Full text] [doi: 10.1016/j.compbiomed.2016.07.012] [Medline: 27479493]
48. Azizoddin DR, Adam R, Kessler D, Wright AA, Kematick B, Sullivan C, et al. Leveraging mobile health technology and research methodology to optimize patient education and self-management support for advanced cancer pain. Support Care Cancer 2021;29(10):5741-5751 [FREE Full text] [doi: 10.1007/s00520-021-06146-4] [Medline: 33738594]
49. Alberts NM, Leisenring WM, Flynn JS, Whitton J, Gibson TM, Jibb L, et al. Wearable respiratory monitoring and feedback for chronic pain in adult survivors of childhood cancer: a feasibility randomized controlled trial from the Childhood Cancer Survivor Study. JCO Clin Cancer Inform 2020;4:1014-1026 [FREE Full text] [doi: 10.1200/CCI.20.00070] [Medline: 33147073]

50. Stinson JN, Jibb LA, Nguyen C, Nathan PC, Maloney AM, Dupuis LL, et al. Development and testing of a multidimensional iPhone pain assessment application for adolescents with cancer. J Med Internet Res 2013;15(3):e51 [FREE Full text] [doi: 10.2196/jmir.2350] [Medline: 23475457]
51. Shi N, Wong AKC, Wong FKY, Sha L. Mobile health application-based interventions to improve self-management of chemotherapy-related symptoms among people with breast cancer who are undergoing chemotherapy: a systematic review. Oncologist 2023;28(4):e175-e182 [FREE Full text] [doi: 10.1093/oncolo/oyac267] [Medline: 36801955]
52. Lobach DF, Boxwala A, Kashyap N, Heaney-Huls K, Chiao AB, Rafter T, et al. Integrating a patient engagement app into an electronic health record-enabled workflow using interoperability standards. Appl Clin Inform 2022;13(5):1163-1171 [FREE Full text] [doi: 10.1055/s-0042-1758736] [Medline: 36516969]
53. Aljedaani B, Babar MA. Challenges with developing secure mobile health applications: systematic review. JMIR Mhealth Uhealth 2021;9(6):e15654 [FREE Full text] [doi: 10.2196/15654] [Medline: 34152277]
54. Global strategy on digital health 2020-2025. 2021. URL: https://www.who.int/docs/default-source/documents/gs4dhdaa2a9f352b0445bafbc79ca799dce4d.pdf [accessed 2024-05-30]
55. Helou S, El Helou E, Abou-Khalil V, Wakim J, El Helou J, Daher A, et al. The effect of the COVID-19 pandemic on physicians' use and perception of telehealth: the case of Lebanon. Int J Environ Res Public Health 2020;17(13):4866 [FREE Full text] [doi: 10.3390/ijerph17134866] [Medline: 32640652]
56. Patt D, Wilfong L, Hudson KE, Patel A, Books H, Pearson B, et al. Implementation of electronic patient-reported outcomes for symptom monitoring in a large multisite community oncology practice: dancing the Texas two-step through a pandemic. JCO Clin Cancer Inform 2021;5:615-621 [FREE Full text] [doi: 10.1200/CCI.21.00063] [Medline: 34085537]
57. Haleem A, Javaid M, Singh R, Suman R. Telemedicine for healthcare: capabilities, features, barriers, and applications. Sens Int 2021;2:100117 [FREE Full text] [doi: 10.1016/j.sintl.2021.100117] [Medline: 34806053]
58. Bohr A, Memarzadeh K. The rise of artificial intelligence in healthcare applications. In: Artificial Intelligence in Healthcare. London, United Kingdom: Academics Press (Elsevier); 2020:25-60.
59. Stergiou-Kita M, Pritlove C, Holness DL, Kirsh B, van Eerd D, Duncan A, et al. Am I ready to return to work? Assisting cancer survivors to determine work readiness. J Cancer Surviv 2016;10(4):699-710. [doi: 10.1007/s11764-016-0516-9] [Medline: 26816271]
60. Tyson DM, Chavez MN, Lubrano B, Lake P, Gutierrez A, Marshall VK, et al. Understanding cancer survivors' educational needs about prescription opioid medications: implications for cancer education and health literacy. J Cancer Educ 2021;36(2):215-224. [doi: 10.1007/s13187-021-01957-9] [Medline: 33428119]
61. Santos-Vijande M, Gómez-Rico M, Molina-Collado A, Davison RM. Building user engagement to mhealth apps from a learning perspective: relationships among functional, emotional and social drivers of user value. Journal of Retailing and Consumer Services 2022;66:102956 [FREE Full text] [doi: 10.1016/j.jretconser.2022.102956]
62. National Academies of Sciences, Engineering, and Medicine, Health and Medicine Division, Board on Population Health and Public Health Practice, Roundtable on Population Health Improvement. Advancing the Science to Improve Population Health: Proceedings of a Workshop. Washington, DC: National Academies Press (US); 2016.
63. Knott V, Zrim S, Shanahan EM, Anastassiadis P, Lawn S, Kichenadasse G, et al. Returning to work following curative chemotherapy: a qualitative study of return to work barriers and preferences for intervention. Support Care Cancer 2014;22(12):3263-3273. [doi: 10.1007/s00520-014-2324-y] [Medline: 25066834]
64. Bukola IM, Paula D. The effectiveness of distraction as procedural pain management technique in pediatric oncology patients: a meta-analysis and systematic review. J Pain Symptom Manage 2017;54(4):589-600.e1 [FREE Full text] [doi: 10.1016/j.jpainsymman.2017.07.006] [Medline: 28712986]
65. Tran Thi TH, Konara Mudiyanselage SP, Huang M. Effects of distraction on reducing pain during invasive procedures in children with cancer: a systematic review and meta-analysis. Pain Manag Nurs 2022;23(3):281-292. [doi: 10.1016/j.pmn.2021.12.002] [Medline: 35031216]
66. Hundert AS, Birnie KA, Abla O, Positano K, Cassiani C, Lloyd S, et al. A pilot randomized controlled trial of virtual reality distraction to reduce procedural pain during subcutaneous port access in children and adolescents with cancer. Clin J Pain 2021;38(3):189-196 [FREE Full text] [doi: 10.1097/AJP.0000000000001017] [Medline: 34974512]
67. Bani Mohammad E, Ahmad M. Virtual reality as a distraction technique for pain and anxiety among patients with breast cancer: a randomized control trial. Palliat Support Care 2019;17(1):29-34. [doi: 10.1017/S1478951518000639] [Medline: 30198451]
68. Kennedy F, Haslam C, Munir F, Pryce J. Returning to work following cancer: a qualitative exploratory study into the experience of returning to work following cancer. Eur J Cancer Care (Engl) 2007;16(1):17-25 [FREE Full text] [doi: 10.1111/j.1365-2354.2007.00729.x] [Medline: 17227349]
69. Tiedtke C, de Rijk A, Dierckx de Casterlé B, Christiaens M, Donceel P. Experiences and concerns about 'returning to work' for women breast cancer survivors: a literature review. Psychooncology 2010;19(7):677-683. [doi: 10.1002/pon.1633] [Medline: 19823971]
70. Scerri A, Morin KH. Using chatbots like ChatGPT to support nursing practice. J Clin Nurs 2023;32(15-16):4211-4213. [doi: 10.1111/jocn.16677] [Medline: 36880216]

71. Basch E, Rocque G, Mody G, Mullangi S, Patt D. Tenets for implementing electronic patient-reported outcomes for remote symptom monitoring during cancer treatment. JCO Clin Cancer Inform 2023;7:e2200187 [FREE Full text] [doi: 10.1200/CCI.22.00187] [Medline: 36857630]
72. Generalova O, Roy M, Hall E, Shah SA, Cunanan K, Fardeen T, et al. Implementation of a cloud-based electronic patient-reported outcome (ePRO) platform in patients with advanced cancer. J Patient Rep Outcomes 2021;5(1):91 [FREE Full text] [doi: 10.1186/s41687-021-00358-2] [Medline: 34524558]
73. ElKefi S, Asan O. How technology impacts communication between cancer patients and their health care providers: a systematic literature review. Int J Med Inform 2021;149:104430 [FREE Full text] [doi: 10.1016/j.ijmedinf.2021.104430] [Medline: 33684711]
74. Psihogios AM, Li Y, Butler E, Hamilton J, Daniel LC, Barakat LP, et al. Text message responsivity in a 2-way short message service pilot intervention with adolescent and young adult survivors of cancer. JMIR Mhealth Uhealth 2019;7(4):e12547 [FREE Full text] [doi: 10.2196/12547] [Medline: 30998225]
75. Mendes-Santos C, Nunes F, Weiderpass E, Santana R, Andersson G. Development and evaluation of the usefulness, usability, and feasibility of iNNOV breast cancer: mixed methods study. JMIR Cancer 2022;8(1):e33550 [FREE Full text] [doi: 10.2196/33550] [Medline: 35166682]
76. Te Boveldt N, Vernooij-Dassen M, Leppink I, Samwel H, Vissers K, Engels Y. Patient empowerment in cancer pain management: an integrative literature review. Psychooncology 2014;23(11):1203-1211. [doi: 10.1002/pon.3573] [Medline: 24816749]

Abbreviations

ePRO: electronic patient-reported outcomes
mHealth: mobile health
NRS: numerical rating scale
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial

Edited by L Buis; submitted 27.10.23; peer-reviewed by E Ross, F Bastos; comments to author 23.01.24; revised version received 22.02.24; accepted 23.04.24; published 18.07.24.

Please cite as:
Wu W, Graziano T, Salner A, Chen MH, Judge MP, Cong X, Xu W
Acceptability, Effectiveness, and Roles of mHealth Applications in Supporting Cancer Pain Self-Management: Integrative Review
JMIR Mhealth Uhealth 2024;12:e53652
URL: https://mhealth.jmir.org/2024/1/e53652
doi:10.2196/53652
PMID:

Review

Development and Use of Mobile Messaging for Individuals With Musculoskeletal Pain Conditions: Scoping Review

Nigel Armfield[1,2,3,4], PhD; Rachel Elphinston[1,2,4,5], PhD; Jenna Liimatainen[1,2,4], BExPhys (Hons); Simone Scotti Requena[6], BSc (Hons); Chloe-Emily Eather[1,2,4,7], BPsySc (Hons); Sisira Edirippulige[8], PhD; Carrie Ritchie[1], PhD; Sarah Robins[1], BPhty; Michele Sterling[1,2,4], PhD

[1]RECOVER Injury Research Centre, The University of Queensland, Brisbane, Australia

[2]National Health and Medical Research Council (NHMRC) Centre for Research Excellence in Better Outcomes for Compensable Injury, Brisbane, Australia

[3]Centre for Health Services Research, Faculty of Medicine, The University of Queensland, Brisbane, Australia

[4]STARS Education and Research Alliance, Surgical Treatment and Rehabilitation Service, The University of Queensland and Metro North Health, Brisbane, Australia

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

[6]Centre for Mental Health, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, Australia

[7]School of Health and Rehabilitation Sciences, The University of Queensland, St Lucia, Australia

[8]Centre for Online Health, Centre for Health Services Research, Faculty of Medicine, The University of Queensland, Brisbane, Australia

Corresponding Author:
Nigel Armfield, PhD
RECOVER Injury Research Centre
The University of Queensland
Level 7, Surgical, Treatment and Rehabilitation Service
296 Herston Road
Brisbane, 4029
Australia
Phone: 61 733655560
Email: N.R.Armfield@uq.edu.au

Abstract

Background: Population studies show that musculoskeletal conditions are a leading contributor to the total burden of healthy life lost, second only to cancer and with a similar burden to cardiovascular disease. Prioritizing the delivery of effective treatments is necessary, and with the ubiquity of consumer smart devices, the use of digital health interventions is increasing. Messaging is popular and easy to use and has been studied for a range of health-related uses, including health promotion, encouragement of behavior change, and monitoring of disease progression. It may have a useful role to play in the management and self-management of musculoskeletal conditions.

Objective: Previous reviews on the use of messaging for people with musculoskeletal conditions have focused on synthesizing evidence of effectiveness from randomized controlled trials. In this review, our objective was to map the musculoskeletal messaging literature more broadly to identify information that may inform the design of future messaging interventions and summarize the current evidence of efficacy, effectiveness, and economics.

Methods: Following a prepublished protocol developed using the *Joanna Briggs Institute Manual for Evidence Synthesis*, we conducted a comprehensive scoping review of the literature (2010-2022; sources: PubMed, CINAHL, Embase, and PsycINFO) related to SMS text messaging and app-based messaging for people with musculoskeletal conditions. We described our findings using tables, plots, and a narrative summary.

Results: We identified a total of 8328 papers for screening, of which 50 (0.6%) were included in this review (3/50, 6% previous reviews and 47/50, 94% papers describing 40 primary studies). Rheumatic diseases accounted for the largest proportion of the included primary studies (19/40, 48%), followed by studies on multiple musculoskeletal conditions or pain sites (10/40, 25%), back pain (9/40, 23%), neck pain (1/40, 3%), and “other” (1/40, 3%). Most studies (33/40, 83%) described interventions intended to promote positive behavior change, typically by encouraging increased physical activity and exercise. The studies evaluated a range of outcomes, including pain, function, quality of life, and medication adherence. Overall, the results either favored messaging

interventions or had equivocal outcomes. While the theoretical underpinnings of the interventions were generally well described, only 4% (2/47) of the papers provided comprehensive descriptions of the messaging intervention design and development process. We found no relevant economic evaluations.

Conclusions: Messaging has been used for the care and self-management of a range of musculoskeletal conditions with generally favorable outcomes reported. However, with few exceptions, design considerations are poorly described in the literature. Further work is needed to understand and disseminate information about messaging content and message delivery characteristics, such as timing and frequency specifically for people with musculoskeletal conditions. Similarly, further work is needed to understand the economic effects of messaging and practical considerations related to implementation and sustainability.

International Registered Report Identifier (IRRID): RR2-10.1136/bmjopen-2021-048964

(JMIR Mhealth Uhealth 2024;12:e55625) doi:10.2196/55625

KEYWORDS

musculoskeletal; pain; SMS text messaging; mobile health; mHealth; intervention design; design; scoping review; musculoskeletal pain; development; mobile messaging; behavior change; efficacy; effectiveness; messaging; implementation; sustainability; mobile phone

Introduction

Background

Musculoskeletal conditions, those affecting the bones, muscles, and joints, are recognized as a global public health problem, although the prevalence and burden of healthy life lost are difficult to estimate with certainty because population studies are few [1]. Where representative studies have been conducted, they have consistently shown high prevalence of musculoskeletal conditions that increases with age and has a greater burden on female than male individuals [2-5]. In the *Health Survey for England* 2018, a total of 17% of adults reported having a long-term musculoskeletal condition (19.5% female vs 14.2% male), with prevalence increasing with age (4.7% at the ages of 16-24 years vs 39% at the age of ≥85 years). A total of 80% of people who reported having a long-term musculoskeletal condition also reported chronic pain (pain for >3 months), with 34.8% reporting pain that highly interfered with their life activities [4,5]. The Australian *National Health Survey* 2017 to 2018 reported that 29% of Australians were living with a chronic musculoskeletal condition (age standardized; adults aged ≥45 years: 51%; 55.3% female vs 47.3% male) [3]. musculoskeletal conditions were the second leading contributor to total burden of healthy life lost, equal to the burden of cardiovascular disease (13% of total burden in disability-adjusted life years), second only to cancer (18% of total burden) [2]. Prioritizing the delivery of effective treatments is necessary to address the substantial burden of musculoskeletal conditions.

With the ubiquity of consumer devices such as smartphones and tablets, technology may have a useful role to play in the management and self-management of musculoskeletal conditions; potentially improve accessibility of health care; and, in some circumstances, ease health system pressures. The use of technology for providing health-related activities is typically described as “digital health” and, more specifically, “mobile health” (mHealth) when referring to the use of mobile devices. While still a relatively new field, mHealth already has a considerable literature base, with examples of its use across most health disciplines and across the continuum of care from health promotion and prevention [6,7] to screening and diagnosis [8,9], therapy [10,11], and self-management [12,13] to cancer survivorship and palliative care [14,15]. While mHealth shows promise in improving aspects of health care, evidence to date is mixed, and caution is needed in interpreting the clinical value of mHealth for patients [16].

In this review, we focused on the development and use of mHealth for individuals with musculoskeletal pain conditions and specifically on health-related interactions that use text messaging as the delivery mechanism (SMS text messaging or messages provided via app-based push notifications), either alone or alongside another intervention. As one of the mobile technologies that have been established for longer, text messaging is familiar, easy to use, convenient, low cost, and available to anyone with a mobile device [17]. Messaging can be used as a vehicle to promote behavior change and guide self-management through prompts, reinforcement, reminders, activity recording, feedback, and adaptivity to the individual [17,18]. The effectiveness of messaging interventions has been assessed for a wide range of health problems, such as medication adherence and lifestyle change in diabetes; encouraging abstinence in smoking cessation; and, more recently, to encourage prevention behaviors during the COVID-19 pandemic [17,19-21].

In total, 2 previous reviews have explored the effectiveness of text messaging–based interventions for musculoskeletal conditions [22,23]. In a broad review of 19 randomized controlled trials (RCTs; 1086 participants) [23], 5 studies involved aspects of messaging [24-28], with 4 studies reporting improvements in pain [25-28] and functional disability [24-27] favoring digital interventions but not specifically favoring the messaging components [23]. A second review focused specifically on the effectiveness of text messaging–delivered interventions included 11 RCTs (1607 participants) [22]. Of the included studies, 5 assessed text messaging as an adjunct to usual care on treatment adherence and found improvements favoring text messaging [29-33]. In a further 5 RCTs, the effectiveness of text messaging as 1 component of a complex intervention was assessed [34-38], finding small but inconsistent effects on pain, functioning, adherence, and quality of life. In 1 RCT, text messaging was compared to telephone counseling, and similar effects on functioning were reported [39].

Objectives

These previous reviews focused on intervention effectiveness and synthesized data from RCTs only. The findings of observational studies have not been synthesized, and these studies may contain useful information to inform and, ultimately, improve the effectiveness and adoption of future musculoskeletal interventions delivered using text messaging. Furthermore, important characteristics of interventions, such as the configuration of digital content, method of presentation, dose, frequency, and preferences, have not been synthesized. Consequently, to inform the design, development, and evaluation of future messaging interventions for people with musculoskeletal pain, we need to explore the literature using a wider lens. Therefore, in this study focused on individuals with musculoskeletal pain conditions, we had three aims: (1) to map the literature related to the use of mobile messaging; (2) to identify information that could be useful in the design of future messaging interventions; and (3) to explore and summarize the findings on efficacy, effectiveness, and economics derived from previous experimental and observational messaging studies.

Methods

We designed and conducted this review according to a preregistered and published protocol [40] developed using the *Joanna Briggs Institute Manual for Evidence Synthesis* [41] and the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) [42] guidelines. The methods are described in full in the published protocol and summarized in brief in the following sections.

Review Questions

Research question (RQ) 1 was as follows: In the context of musculoskeletal pain conditions, for which individuals, with which problems, and for what purpose, has messaging on mobile devices been used (eg, medication reminders, alerts, education, motivation, prevention, and data collection)?

RQ 2 was as follows: What information exists to guide the development of mobile messaging for musculoskeletal pain conditions (eg, frequency of messages, length of messages, duration of the intervention, and theoretical basis)?

RQ 3 was as follows: How have patients' preferences been included in the design of a study, and how have their preferences been assessed?

RQ 4 was as follows: What methods have been used to evaluate the use of mobile messaging for musculoskeletal pain conditions (eg, how were outcomes assessed and what processes were involved)?

RQ 5 was as follows: Does the literature support the efficacy, effectiveness, and economics of messaging on mobile devices for individuals with musculoskeletal pain conditions?

Inclusion Criteria

Participants

We included studies on adult participants with acute or chronic musculoskeletal pain conditions.

Concept

The concepts of interest were the development or evaluation of patient-focused health-related messaging (eg, SMS text messaging and app push notifications) provided on mobile devices such as smartphones and tablets.

Context

We included articles that described messaging used in any setting either as a primary intervention or as an adjunct to other interventions. We excluded studies focused on spinal cord injury, traumatic brain injury, moderate to severe orthopedic injuries, surgical patients, and conditions related to mobile phone overuse. We also excluded studies focused on health conditions primarily unrelated to the bones, muscles, and connective tissue (eg, diabetes, asthma, cancer, and stroke).

Data Sources

We searched PubMed, CINAHL (via EBSCOhost), Embase, and PsycINFO (via APA PsycNET) using a strategy that combined controlled-vocabulary and free-text search terms related to messaging and musculoskeletal concepts. Because of resource limitations, we were unable to include gray literature in our searches.

Search Strategy

The search strategy is described in detail in the published protocol [40], and the search queries are provided again in this paper in Multimedia Appendix 1. The search strategy was developed through discussion among the team and an iterative process of pilot searches. The final searches were conducted by SSR. Because of resource limitations, we restricted our searches to articles published in English, and because the area of digital health is a rapidly changing field, we limited our searches to articles published in the previous 10 years.

Study Selection

We exported search results to EndNote (version X9; Clarivate Analytics) and Covidence (Veritas Health Innovation) [43] for duplicate removal and to manage the screening, selection, and record-keeping processes. We conducted study selection in 3 phases. First, using the predefined inclusion and exclusion criteria, 2 independent reviewers (from a pool of 7; SSR, JL, CEE, RE, CR, SR, and NA) screened the titles and abstracts to identify candidate articles for inclusion and to discard irrelevant articles. Second, 2 reviewers from the same pool reviewed the full text of each candidate article. Third, we searched the reference lists of the included papers to identify any further articles. At all stages, conflicts were resolved using a third reviewer from our pool.

Data Extraction

Data were extracted by one reviewer (JL) and independently confirmed by 2 others (NA and CEE). Data were extracted using predefined extraction forms, as described in the protocol [40].

Synthesis and Reporting

We described the results of the study selection process using a PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram [44], with findings reported in accordance with the PRISMA-ScR checklist [42]. For each

of our 5 review questions, we structured our findings using tables adapted from the Joanna Briggs Institute manual [41] refined as necessary at synthesis stage [40]. We then developed a narrative summary of the evidence for each of our review questions.

Protocol Deviations

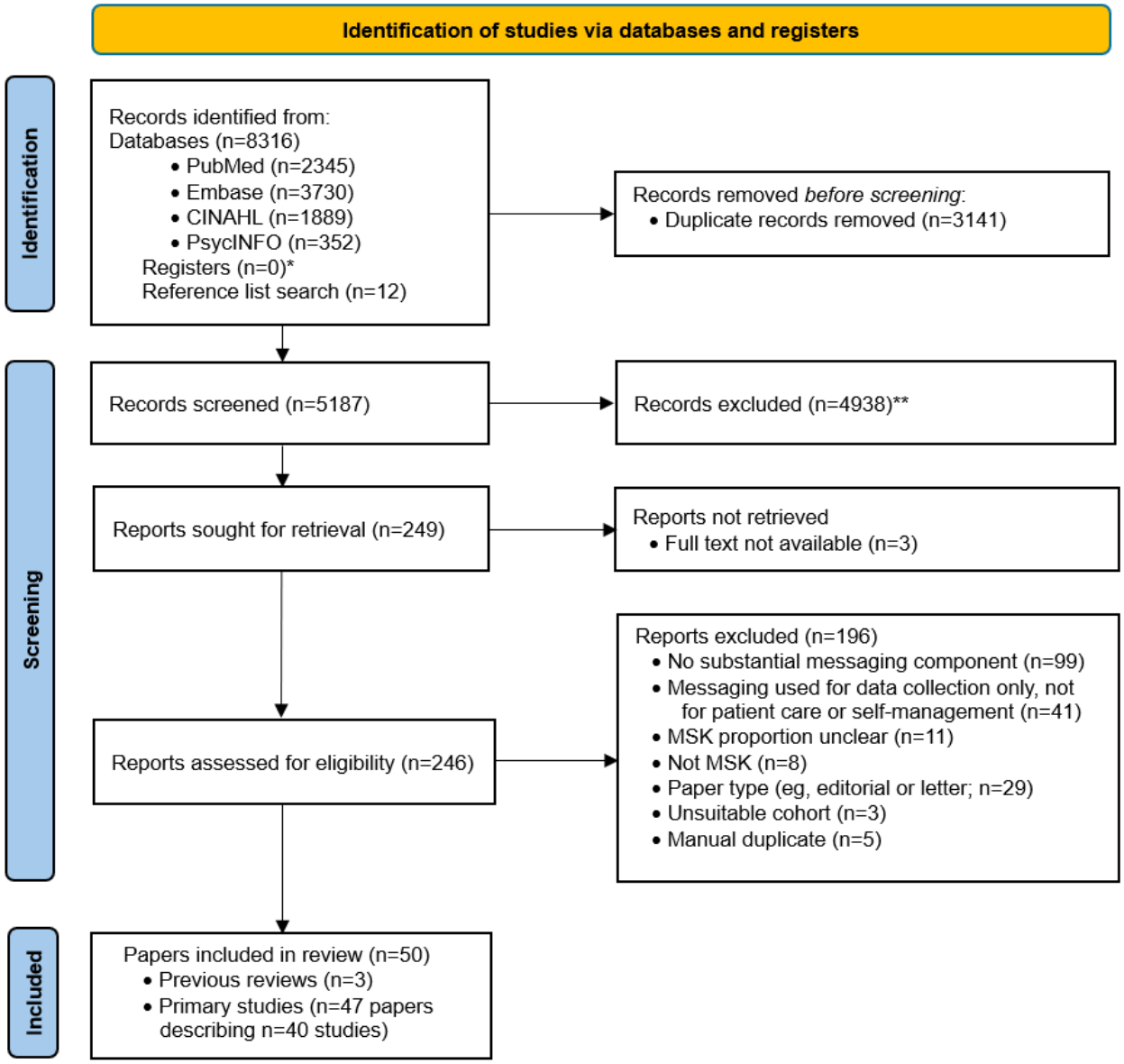
There were 4 minor protocol deviations. First, we excluded studies that described the use of mobile messaging to collect data in cases in which those data were not subsequently used to inform care or self-management (eg, studies that simply tested the feasibility of using text messaging to collect data and studies that used text messaging as a data collection method to model recovery trajectories). Second, we included study protocols associated with evaluation studies if they provided useful information about messaging design and development. Third, we classified the level of development of the country in which the study was conducted using the Human Development Index (HDI) [45]. Finally, we reran our searches in 2022 and, therefore, included studies from a 12-year period rather than the originally specified 10 years.

Results

Overview

Literature searches were conducted in August 2020 and repeated in May 2022. In this section, we present the combined results of both searches. We identified a total of 8328 papers (published in 2010-2022) for screening, of which 50 (0.6%) were included in this review. A PRISMA flowchart of the article selection process is shown in Figure 1.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart—article selection process. *No registers were searched; **No automation tools were used; MSK: musculoskeletal.



We identified 3 previous systematic reviews, 2 (67%) of which we had already found while developing the protocol for this review [22,23] and 1 (33%) that was new [46]. One review focused specifically on the effects of text messaging for managing musculoskeletal pain conditions [22], while the remainder focused more broadly on digital health or mHealth for musculoskeletal conditions but covering some aspects of messaging [23,46]. The previous reviews were conducted in Australia, the United Kingdom, and the Netherlands, all countries classed as *very highly developed* according to their HDI. The characteristics of the reviews are shown in Table 1, and the findings are shown in Table 2. We did not identify any previous reviews related to design aspects of messaging for musculoskeletal pain conditions.

We included 47 papers describing 40 primary studies (22/40, 55% experimental; 16/40, 40% observational; and 2/40, 5% mixed methods). In total, 10% (4/40) of the experimental and observational studies had associated or embedded qualitative or mixed methods studies. The results of 5% (2/40) of the studies were multiply reported, and 8% (3/40) of the studies had either an associated design paper or a protocol paper containing design information. A total of 18 countries were represented, with the United States publishing the largest number of studies (9/40, 23%) followed by Australia (6/40, 15%) and Denmark (4/40, 10%). By HDI, most primary studies were conducted in *very highly developed* countries (36/40, 90%), 8% (3/40) were conducted in *highly developed* countries, and 3% (1/40) were conducted in a country of *medium development*. No studies were reported from countries of *low development*.

At the time of our search, 70% (35/50) of the previous reviews and primary studies had been published in the 3 years before our search. The characteristics of the primary studies are shown in Table 3 and Figure 2 [24,26,27,30-32,35-38,47-83].

Table 1. Characteristics of review papers related to messaging for people with musculoskeletal (MSK) pain conditions.

Study, year; type	Country[a] (HDI[b])	Review focus	Studies and sample size	MSK condition focus	Primary outcomes	Messaging method		Adjunct
						SMS text messaging	Push notifications	
Fritsch et al [22], 2020; SR[c]	Australia (VH[d])	Effects of text messaging for managing MSK pain	7 RCTs[e]; n=1181[f]	Any acute or chronic MSK[f]	Pain, function, adherence, and QoL[g]	✓	✓	Both
Hewitt et al [23], 2020; SR	United Kingdom (VH)	Digital health in the management of MSK conditions	19 RCTs; n=3361; 5 RCTs (n=1086) related to messaging	Any MSK condition excluding postsurgical management and pain related to computer use	Pain and functional disability; in addition, catastrophizing, self-efficacy, QoL, and coping strategies	✓	✓	Both
Seppen et al [46], 2020; ScR[h]	The Netherlands (VH)	Asynchronous mHealth[i] interventions for RA[j]	10 studies; n=1214; 3 RCTs (n=266) related to messaging	RA	Medication compliance and sitting time	✓		Both

[a]On the basis of the lead author's affiliation.

[b]HDI: Human Development Index [45].

[c]SR: systematic review.

[d]VH: very high.

[e]RCT: randomized controlled trial.

[f]Review included surgical studies; we report the subgroup of nonsurgical studies or participants in this table.

[g]QoL: quality of life.

[h]ScR: scoping review.

[i]mHealth: mobile health.

[j]RA: rheumatoid arthritis.

Table 2. Findings of review papers related to messaging for people with musculoskeletal (MSK) pain conditions.

Study, year	Individuals, problems, and purpose	Design-related information	Outcomes assessed and review findings

Fritsch et al [22], 2020—effects of text messaging for managing MSK pain	• Review included 7 RCTs[a] on patients with MSK pain conditions (3 with RA[b], 1 with chronic widespread pain, 1 with upper- or lower-limb MSK injuries, 1 with frozen shoulder, and 1 with knee pain) [30-32,35-38,47]
	• Messaging used to support behavior change. Most studies targeted physical activity or medication compliance.
		• Messaging features varied across studies. Examples include individualization to patient goals, timing, frequency, duration, directionality, and other intervention characteristics.
		• The included studies provided little or limited description of the theoretical frameworks underpinning the interventions.
		• Patient preferences were not described.
			• Clinical outcomes such as pain, function, disability, exercise adherence, QoL[c], satisfaction with health care services, confidence in treatment, self-efficacy, and anthropometric measures
			• Findings:
			  • Text messaging+UC[d] vs UC
			  • No difference on pain [30]
			  • Equivocal or no difference on function [30,32]
			  • Equivocal or no difference on unscheduled appointments [31]
			  • Increase in calls to nurses [31]
			• Messaging as part of the intervention vs any treatment:
			  • Pain: decrease [37,38]; equivocal or no difference [35,47]
			  • Function: equivocal or no difference [35,47]; increase [36,37]
			  • Exercise adherence: increase in self-reported adherence; equivocal or no difference on assessor-reported adherence [36]
			  • QoL: equivocal or no difference [31]
			  • SF-36[e] MCS[f]: increase [35,37,47]
			  • SF-36 PCS[g]: increase [37]; equivocal or no difference [35,47]
			• Comparison of messaging vs phone counseling
			  • Patient feedback and AEs[h]: assessed in 7 studies; AEs reported in 3 studies unrelated to messages

Hewitt et al [23], 2020—digital health for managing MSK conditions	• Aspects of messaging were described in each of the following: 3 studies on self-management of back pain [24,25,27], 1 digitally delivered multidisciplinary pain program for back pain [28], and 1 conservative digital care program for knee pain [26].	• Not described	• Pain or function assessed via RCTs. • Messaging (along with phone calls or email reminders) was described in the context of "additional efforts to encourage engagement" or "additional forms of support." • Review concluded that "additional forms of support" may be linked to positive outcomes (including improvement in pain and function); however, variability in messaging intervention characteristics hinders conclusions regarding effectiveness specific to messaging.

Seppen et al [46], 2020—asynchronous mHealth[i] interventions for RA	• Included 3 RCTs assessing the effectiveness of SMS text message reminders for medication adherence [32] and reducing sitting time [37,48].	• Not described
		• Some studies incorporated patients' preferences; participants could select reminder frequency (1-5 per week) [37,48].
			• Messaging not evaluated directly; rather, patient outcomes relevant to the primary objective were assessed, such as medication compliance [32] and sedentary time [37,48].
			• Findings included the following:
			  • Increase in medication compliance[j] [32]
			  • Reduced sitting time [37]

[a]RCT: randomized controlled trial.

[b]RA: rheumatoid arthritis.

[c]QoL: quality of life.

[d]UC: usual care.

[e]SF-36: 36-item Short-Form Health Survey.

[f]MCS: Mental Component Summary.

[g]PCS: Physical Component Summary.

[h]AE: adverse event.

[i]mHealth: mobile health.

[j]19-item Compliance Questionnaire on Rheumatology, incorrectly described as the 9-item Compliance Questionnaire on Rheumatology in the review by Seppen et al [46].

Table 3. Characteristics of primary studies related to messaging for people with musculoskeletal (MSK) pain conditions.

Study, year	Country[a] (HDI[b])	Design	Primary aim				Messaging method		Adjunct
			Provide information	Behavior change	Data collection	Design	SMS text messaging	Push notifications	
Rheumatic diseases									
Kristjánsdóttir et al [35,47], 2013	Norway (VH[c])	Experimental		✓			✓		Yes
Theiler et al [49], 2016	Switzerland (VH)	Observational	✓				✓		No
Thomsen et al [48], 2016	Denmark (VH)	Experimental		✓			✓		Yes
Mecklenburg et al [26], 2018	The United States (VH)	Experimental		✓			✓	✓	Yes
Molinari et al [50], 2018	Spain (VH)	Experimental		✓			✓		No
Nordgren et al [51], 2018, and Demmelmaier et al [75], 2015	Sweden (VH)	Observational, mixed methods study (stand-alone, associated, or embedded within a trial)		✓			✓		Yes
Timmers et al [52], 2018	The Netherlands (VH)	Experimental	✓					✓	Yes
Wang et al [38], 2018	Australia (VH)	Experimental		✓			✓		Yes
Bartholdy et al [53], 2019	Denmark (VH)	Experimental		✓			✓		No
Ravn Jakobsen et al [76], 2018	Denmark (VH)	Observational				✓[d]	✓		No
Geuens et al [77], 2019	Belgium (VH)	Mixed methods study (stand-alone, associated, or embedded within a trial)				✓[d]		✓	No
Ji et al [54], 2019	China (H[e])	Observational		✓				✓	No
Mary et al [32], 2019	The United States (VH)	Experimental		✓			✓		Yes
Støme et al [55], 2019	Norway (VH)	Observational		✓				✓	Yes
Thomsen et al [37], 2017, and Thomsen et al [56], 2020	Denmark (VH)	Experimental		✓			✓		Yes
Zaslavsky et al [57], 2019	The United States (VH)	Observational		✓			✓		Yes
Kuusalo et al [31], 2020	Finland (VH)	Experimental			✓		✓		Yes
Nelligan et al [78], 2020 (qualitative study [stand-alone, associated, or embedded within a trial]), Nelligan et al [79], 2019 (qualitative study [stand-alone, associated, or embedded within a trial]), and Nelligan et al [58], 2019 (experimental)	Australia (VH)	Experimental and qualitative (stand-alone, associated, or embedded within a trial)		✓		✓[f]	✓		Yes

Study, year	Country[a] (HDI[b])	Design	Primary aim				Messaging method		Adjunct
			Provide information	Behavior change	Data collection	Design	SMS text messaging	Push notifications	
Pelle et al [59], 2020, and Pelle et al [80], 2019	The Netherlands (VH)	Experimental		✓		✓[f]		✓	No
Multiple MSK conditions									
Newell [60], 2012	Germany (VH)	Experimental		✓			✓		Yes
Taylor et al [61], 2012	Australia (VH)	Experimental		✓			✓		Yes
Gandy et al [62], 2016	Australia (VH)	Observational		✓			✓		Yes
Jamison et al [63], 2017	The United States (VH)	Experimental		✓				✓	Yes
Johnson et al [81], 2017	The United States (VH)	Observational				✓[d]	✓		Yes
Lambert et al [36], 2017	Australia (VH)	Experimental		✓				✓	Yes
Lo et al [64], 2018	China (H)	Observational		✓				✓	Yes
Frei et al [65], 2019	Switzerland (VH)	Mixed methods study (stand-alone, associated, or embedded within a trial)		✓				✓	Yes
Anan et al [66], 2021	Japan (VH)	Experimental	✓	✓				✓[g]	Yes
Bailey et al [67], 2020	The United States (VH)	Observational		✓			✓	✓	Yes
Low back pain									
Dekker-van Weering et al [68], 2015	The Netherlands (VH)	Observational		✓				✓	Yes
Chhabra et al [24], 2018	India (M[h])	Experimental		✓				✓	Yes
Rabbi et al [69], 2018	The United States (VH)	Observational		✓				✓	No
Selter et al [70], 2018	The United States (VH)	Observational		✓				✓	No
Hasenöhrl et al [71], 2020	Austria (VH)	Observational and qualitative (stand-alone, associated, or embedded within a trial)		✓				✓	Yes
Shebib et al [27], 2019	The United States (VH)	Experimental		✓				✓	Yes
Almhdawi et al [72], 2020	Jordan (H)	Experimental		✓				✓	Yes
Nordstoga et al [73], 2020 (qualitative [stand-alone, associated, or embedded within a trial]), and Mork and Bach [82], 2018 (observational; protocol)	Norway (VH)	Observational and qualitative (stand-alone, associated, or embedded within a trial)		✓		✓[f]		✓	Yes

Study, year	Country[a] (HDI[b])	Design	Primary aim				Messaging method		Adjunct
			Provide information	Behavior change	Data collection	Design	SMS text messaging	Push notifications	
Fritsch et al [83], 2021	Australia (VH)	Observational				✓[f]	✓		No
Neck									
Lee et al [74], 2017	Korea (VH)	Experimental		✓			✓		Yes
Frozen shoulder									
Chen et al [30], 2017	Taiwan (VH)[d]	Experimental		✓			✓		Yes

[a]On the basis of the lead author's affiliation.

[b]HDI: Human Development Index [45].

[c]VH: very high.

[d]Mobile health design paper.

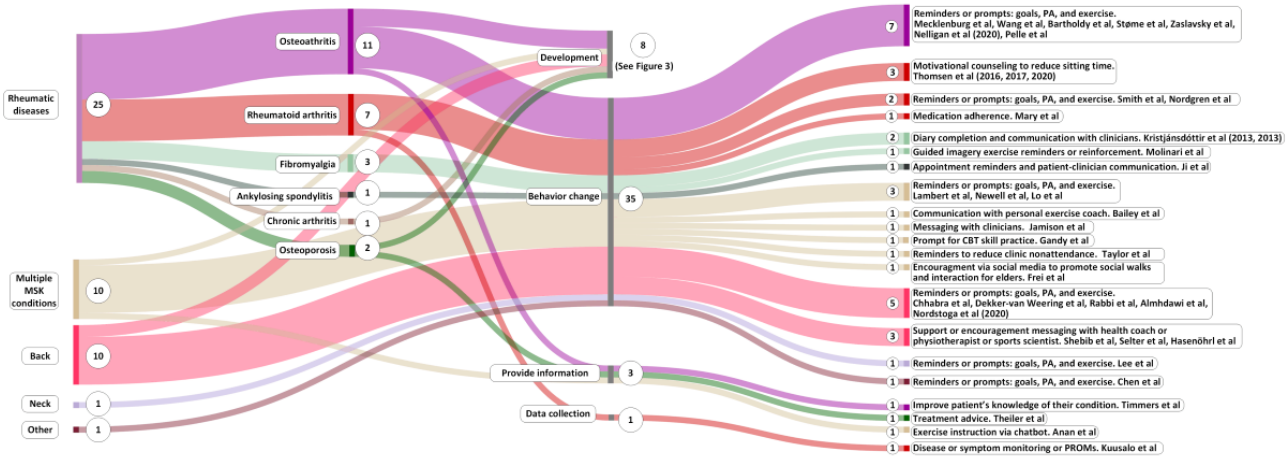
[e]H: high.

[f]Messaging-specific design paper.

[g]Messaging provided using a social media app.

[h]M: medium.

Figure 2. Overview of 47 papers describing 40 primary studies by condition, purpose, and role of messaging. The circled numbers represent the number of papers. CBT: cognitive behavioral therapy; MSK: musculoskeletal; PA: physical activity; PROM: patient-reported outcome measure.



RQ 1: Individuals, Problems, and Purpose

Previous Reviews

In the previous reviews [22,23,46] (Tables 1 and 2), the most commonly reported messaging interventions were for people with rheumatoid arthritis (RA) and back pain. For RA, messaging was used to monitor medication and disease activity [31] and improve medication adherence [32,46] and for reminders to reduce daily sitting time [37,46]. For people with back pain, messaging was used mostly as a component of self-management, with approaches focused on education and behavior change strategies [24,25], supportive messages provided by a health coach during periods of low engagement with a digital self-management program [27], and motivating messages sent as part of a multidisciplinary pain program [28].

Other studies described uses of messaging for people with knee pain, systemic lupus erythematosus, frozen shoulder, chronic widespread pain, and limb injuries or conditions. For knee pain, one study reported a lifestyle intervention focused on behavior change [22,38], and another reported participation reminders and app-based messaging with a personal coach as part of an exercise, education, or cognitive behavioral therapy (CBT) or weight loss or psychosocial support program [23,26]. For frozen shoulder, reminder, encouragement, and education messages were used to promote exercise compliance and improve shoulder function [30]. For chronic widespread pain, a CBT intervention used SMS text message diary completion prompts, with those diary entries then informing the treatment used by a therapist [35]. For limb injuries and conditions, messaging was used to promote adherence to a home exercise program in one study [36].

Primary Studies

Rheumatic diseases accounted for the largest proportion of the included primary studies (19/40, 48%), followed by studies on

multiple musculoskeletal conditions or pain sites (10/40, 25%), back pain (9/40, 23%), neck pain (1/40, 3%), and "other" (1/40, 3%; Table 3).

Rheumatic Diseases

Of the 19 rheumatic disease–related studies, 8 (42%) focused on osteoarthritis [26,38,52,53,55,57-59,78-80], 5 (26%) focused on RA [31,32,37,48,51,56,75], 2 (11%) focused on fibromyalgia [35,47,50], 2 (11%) focused on osteoporosis [49,76], and 1 (5%) each focused on ankylosing spondylitis [54] and chronic arthritis [77].

Of these 19 studies, 14 (74%) described the use of messaging to promote behavior change with the intention of improving levels of physical activity, assisting weight loss, improving sleep, or reducing stress [26,32,35,37,38,47,48,50,51,53-59,75,78-80]. A total of 11% (2/19) of the studies described messaging for providing information [49,52], and 5% (1/19) described the use of messaging to collect data for disease monitoring and guide clinical care [31]. In total, 26% (5/19) of the studies described aspects of design and development of messaging systems for people with knee osteoarthritis [79,80], osteoporosis [76], and chronic arthritis [77]. The design and development aspects are described in later sections.

Osteoarthritis Studies

Of the 8 studies on osteoarthritis, 2 (25%) focused on behavior change based on personalized goals. In the first study, which proposed personalized goals based on machine learning, participants were sent daily push notifications to remind them of their goals together with an interesting fact or answer to a frequently asked question [59,80]. Similarly, the second study used messaging to provide reminders to complete individualized physician-assigned goals and tasks, for which participants also used messaging to provide confirmation, or otherwise, that they had completed their personalized goals [55].

A total of 4 (50%) of studies focused on physical activity and exercise behavior change for people with knee osteoarthritis: of those, 1 (25%) used messages to decrease inactive behavior in people with knee osteoarthritis [53] and another (25%) used targeted personalized motivational reinforcement messages based on previous and current physical activity for people with osteoarthritis and sleep disturbance [57]. In the third study, which had an experimental design, the authors also explored patient attitudes and experiences of a self-directed digital health intervention incorporating automated messages to support strengthening exercises [78,79]. The fourth study, in which 77% of participants had knee osteoarthritis, described a digital care program that sent participants reminder messages if they did not engage with the program at the required intensity and also allowed participants to communicate with their health coach using messaging [26].

A single study focused on providing information for people with knee osteoarthritis, where messages were used to improve patients' knowledge about their condition and treatment options before consultation with their specialist as part of shared decision-making [52].

A further study focused on knee osteoarthritis prevention, describing a self-management lifestyle intervention for young to middle-aged rural-dwelling women that incorporated messaging to provide key behavior reminders [38].

RA Studies

Of the 5 studies on RA, 2 (40%) used message reminders as part of a motivational counseling intervention to reduce sitting time [37,48,56], and 1 (20%) focused on physical activity behavior change with messaging used for coaching, prompts, reminders, and monitoring of physical activity program adherence [51,75]. A further study assessed the effects of text messages on medication adherence [32]. One study collected data using text or app-based messaging for symptom or disease monitoring and patient-reported outcome measures [31].

In a study that recruited women with chronic widespread pain (80% met the American College of Rheumatology criteria for fibromyalgia), text messaging was used to prompt diary completion and allow participants to exchange short messages with their therapist. The diary information was used by therapists to inform patient care [35,47]. A second guided imagery study also focused on people with fibromyalgia used text messaging to remind participants to practice their imaging exercises together with randomly selected reinforcement messages [50].

A study on patients with osteoporosis and nontraumatic fractures used text messaging to provide patients with treatment advice based on a validated fracture assessment tool and assessed whether the advice provided subsequently changed primary care physician management of their fracture [49].

Finally, one study described the use of social media messaging (WeChat) for people with ankylosing spondylitis, with messaging used for appointment reminders, for communication between physicians and patients, to record follow-up information, and for patients to provide feedback [54].

Multiple Musculoskeletal Conditions or Pain Sites

A total of 10 studies focused on multiple musculoskeletal conditions or pain sites (n=1, 10% each on the neck or back [64], neck, shoulder, or back [66], and chronic knee or low back pain [LBP] [67]). A total of 50% (5/10) of the studies recruited participants with a range of musculoskeletal problems typically seen in the general population [36,60,61,65], and 20% (2/10) of the studies recruited adults with chronic pain but not pain exclusively of musculoskeletal origin [62,63]. A further study focused on chronic musculoskeletal pain in veterans [81].

Of these 10 studies, 9 (90%) described behavior change interventions [36,60-63,65-67,81], and 1 (10%) was focused on providing information [66].

For neck and back pain, one study described the use of an artificial intelligence–enabled app that implemented evidence-based guidelines for self-management, with messaging provided within the app to remind participants to exercise and provide contact with the treating team [64]. A second study on workers with neck, shoulder, or back pain also described the use of artificial intelligence, wherein a chatbot provided messages with exercise instructions and suggestions for symptom improvement [66]. One study focused on chronic knee

or LBP described a digital care program incorporating sensors and an app that allowed participants to communicate with a personal coach via SMS text messaging and app-based messaging [67].

Another 20% (2/10) of the studies included adults with chronic pain but not exclusively pain of musculoskeletal origin [62,63]. The first included patients being treated by a hospital-based pain management service for a range of conditions (LBP; cervical or upper-extremity, lower-extremity, abdominal or pelvic, and head or face pain; and multiple pain sites, with pain of ≥4 on a 0-10 scale). Participants used an app that incorporated reminders to complete daily assessments and also provided 2-way messaging [63]. The second study, with similar wide-ranging pain sites, used automated text messaging to prompt skill practice as part of an internet-delivered CBT program for chronic pain [62].

Regarding patients attending hospital physiotherapy services for a range of musculoskeletal problems, 10% (1/10) of the studies examined whether SMS text messaging could increase home exercise compliance [36]. In this study, compliance with exercises was encouraged via motivational SMS text messages sent by the physiotherapist. Similarly, the use of messaging to encourage home exercise compliance was described in a study on patients with musculoskeletal problems attending a chiropractic clinic [60].

In the physiotherapy outpatient setting, the use of SMS text message reminders to reduce clinic nonattendance was described in 10% (1/10) of the studies [61].

A total of 20% (2/10) of the studies focused on specific populations. The first, a community-based study, aimed to improve the physical activity of older adults (aged ≥60 years, most of whom had musculoskeletal problems) and used social media messaging (WhatsApp) to inform participants of scheduled walks and promote social interaction between participants [65]. The second study focused on a chronic musculoskeletal pain program in veterans and used behavior change messaging for stress management and adoption of healthy sleep practices and to increase engagement and retention in the program [81].

Back Pain

A total of 20% (8/40) of the studies described behavior change interventions [24,27,68-73,82], and 5% (2/40) described the design and development (described in a later section) [82,83]. Of the 8 behavior change studies, of these 4 (50%) described the use of individual or personalized messaging for physical activity goal reminders and reinforcement [24], encouragement messages and physical activity suggestions [69], motivational notifications for self-management [73,82], and individual activity level–based feedback messages provided on a PDA to encourage behavior change [68]. A total of 13% (1/8) of the studies described a self-management app with notifications to encourage walk breaks and posture exercises [72].

A total of 38% (3/8) of the studies described the use of 1- or 2-way messaging with a health coach, physiotherapist, or sports scientist for support, encouragement, and participation reminders as part of self-management programs [27,70,71].

Neck Pain

Only 3% (1/40) of the studies focused specifically on neck pain. This study described a behavior change intervention for office workers with chronic neck pain incorporating weekly messages about caring for their pain with information about the importance of exercise and to provide encouragement to complete prescribed exercises [74].

Other Conditions

A total of 3% (1/40) of the studies, on patients with frozen shoulder recruited from an orthopedic outpatient clinic, used messaging to provide reminders, encouragement, and education to promote shoulder exercise compliance [30].

RQs 2 and 3: Design and Development and Patient Preferences

Overview

In this section, we report findings related to the design and development of messaging interventions. Because patient preferences, where accommodated, were generally addressed through participatory or co-design, we have reported the results of review questions 2 and 3 together. The findings are presented in three groups: (1) information found in papers specifically focused on the design and development of messaging interventions, (2) information found in mHealth design papers where some aspect of messaging was described alongside other mHealth functions, and (3) incidental design and development information found in papers that reported the results of messaging or mHealth interventions. The design-specific papers are shown in Table 4 and Figure 3 [76-83].

Table 4. Papers focused on messaging design and development and patient preferences.

Study, year	Role of messaging	Design process (theory, method, and outcomes)
Johnson et al [81], 2017—describes the participatory design and pilot study of an mHealth[a] self-management program for veterans with chronic MSK[b] pain	• Tailored messages were an optional component intended to increase engagement and retention.	• Participatory design involving a panel of veteran advisors, experts, and end users (number not specified). Input sought through interviews, focus groups, and usability testing but not described in detail. • Messages were described as targeting behaviors, with message content and schedules matched to the participant's stage of change based on the transtheoretical model of health behavior change [84]. The process through which the message content and schedules were derived was not described. Example messages included the following: "As a Veteran, you likely know many people who have or had pain. Think about one of them who could inspire you to manage your pain. Stress can make people more prone to pain. If you lower your stress, you can help lower your pain. See PAC activity Get the Facts [short-url]."
Mork and Bach [82], 2018 (protocol)—describes the components and architecture of an app-based self-management decision support system for LBP[c] (self-BACK)	• Messaging (via push notifications) used within the app to encourage physical activity	• Authors stated that focus groups and iterative testing and development with patients, health professionals, and researchers were part of the development process without further detail. • Structured intervention mapping [85], behavior change theories [86], and normalization process theory [87] • During the development process, patients and health professionals (eg, physiotherapists and psychologists) were interviewed on their experience managing LBP. Educational content was reviewed by clinicians and researchers. • Patient case data (baseline information, physical activity monitoring, and weekly patient-reported health and adherence outcomes) were used to generate motivational notifications to encourage physical activity. Little messaging-specific design information was provided.
Ravn Jakobsen et al [76], 2018—describes the participatory design and development of an mHealth app for women with newly diagnosed osteoporosis	• Messaging used to communicate the results of a bone density scan (DXA[d]) to women and coordinate their follow-up appointment with their general practitioner.	• A participatory design [88] was used. The team consisted of researchers, women, physicians, other health care professionals, and app designers. • The iterative participatory app design process was somewhat unclear and described as commencing with 3 workshops (first, to generate ideas; second, to review wireframe designs; and third, to discuss the overall design content), followed by the creation of the design, feedback from users, development of a prototype, laboratory tests and feedback, adjustment, and final development. Messaging-specific design and development considerations were not described.

Nelligan et al [79], 2019—comprehensive description of the identification of behavior change targets; design of SMS text message library to support adherence to home exercise for people with knee OA[e]	• SMS text message–based intervention
	• Automated behavior change messages to promote exercise, with adaptive messages triggered by participant responses
		• Phase 1: theoretical rationale and application to inform the intervention
		  • SMS text messaging was selected as the mode of delivery based on literature describing it as a scalable, effective, efficient, and affordable way to promote adherence to health behaviors [19,89-94]. The authors used a previous scoping review [95] that mapped barriers and facilitators against the Theoretical Domains Framework [96]. Furthermore, the COM-B[f] framework for understanding health behavior [97] and the BCW[g] [97,98] were used throughout this phase. Previous work was used to identify messaging intervention functions appropriate for the SMS text messaging format [99].
		• Phase 2: development of SMS text messaging functions and a message library
		  • The SMS text messaging functionality was guided by the literature [100]. SMS text messages were automated and adaptive. Participants' self-reported exercise adherence triggered a BCT[h].
		  • The content of the messages was codeveloped by 7 academics, 4 physiotherapists, and 1 person with knee OA. The authors based their messaging frequency on previous literature, which, while inconsistent, suggests that 3 messages per week tapered over time was appropriate [92,100]. Examples are provided in appendixes accompanying the authors' article [79] and in Table 5.

Study, year	Role of messaging	Design process (theory, method, and outcomes)
Geuens et al [77], 2019—identified feature preferences and motivations for a hypothetical self-management app for chronic arthritis	• Messaging as a component in an app • To provide medication or postural reminders	• Authors referred to the PSD[i] model [101]. • Structured interviews to identify patient preferences for features in a hypothetical app and their motivations for selecting those features • Limited messaging-specific information provided; however, reminders were rated highest in terms of desired features (medication and also posture). Praise and reward messages were considered less important, and social interaction features were rated the lowest.
Pelle et al [80], 2019—describes the theoretical framework and iterative design of an app for OA self-management	• App was developed through an iterative design process that comprised medical researchers, physicians, physical therapists, patient representatives, and app developers.	• Iterative development process involving researchers, health professionals, app designers, and patient representatives over 3-week "sprints" of development; user testing; reiteration; and, finally, pilot-testing • After a review of the literature and consensus meetings, it was determined that motivation enhancement techniques such as reminders could increase the intervention effect. • The Fogg Behavior Model [102], persuasive design [103], and daily push notifications to remind users of their goal and provide education on OA
Fritsch et al [83], 2021—described the co-design process for a bank of evidence-based messages for an LBP self-management text messaging intervention	• Iterative codevelopment to identify relevant domains, content sources, frequency, appropriate timing, and a series of evidence-based messages for self-management of LBP	• Behavior change methodology [104] previously used by Redfern et al [105,106] that links BCTs to frameworks such as information-motivation-behavior, theory of reasoned action, theory of planned behavior, social cognitive theory, control theory, and operant conditioning. • 2-phase process previously used to develop messages in cardiology [106] conducted with consumers, researchers, and clinicians (n=39) to generate 82 messages • Phase 1: development of concept and content with 15 consumers, clinicians, and researchers over 2 workshops to determine messaging features. In the workshop, it was decided that 4 weekly messages would be sent across the domains to provide education, motivation, or behavior change. Timing of appropriate messages for LBP self-management (9 AM, 12:30 PM, 4 PM, and 6 PM) was drawn from the literature [90,105]. • Messages were subsequently drafted by 2 researchers and 2 consumer representatives and were then reviewed by 2 researchers with expertise in behavior change. • Phase 2: iterative web-based review phase beginning with experts, then followed by consumers. Each message was reviewed by at least 2 participants in each round. Experts provided a score (mean 8.3/10) for appropriateness of content with consideration to current evidence and the likelihood of clinical effectiveness. Messages with a score of <8/10 (34%) were revised and then assessed by people with lived experience with LBP. These consumers scored each text messages on utility of content, understanding, and language acceptability. Messages with a score of <12/15 (31%) were revised according to feedback. Most frequently, consumer feedback focused on making the content more specific and less technical and including more examples.

[a]mHealth: mobile health.

[b]MSK: musculoskeletal.

[c]LBP: low back pain.

[d]DXA: dual-energy x-ray absorptiometry.

[e]OA: osteoarthritis.

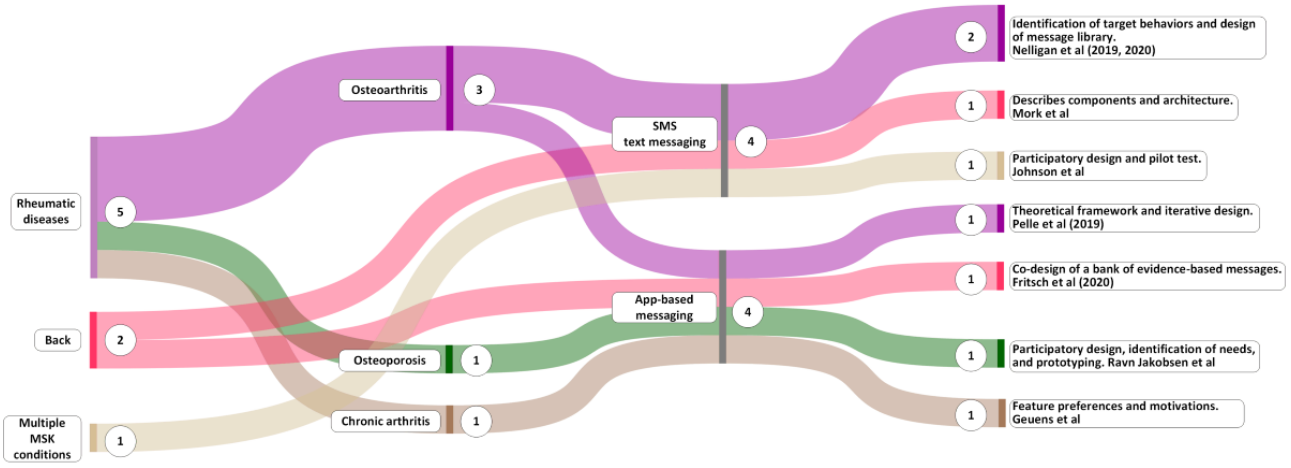
[f]COM-B: Capability, Opportunity, and Motivation–Behavior.

[g]BCW: Behavior Change Wheel.

[h]BCT: behavior change technique.

[i]PSD: Persuasive System Design.

Figure 3. Overview of 8 papers describing aspects of design and development of messaging. The circled numbers represent the number of papers. MSK: musculoskeletal.



Papers Focused Specifically on Messaging Design and Development

A total of 4% (2/47) of the papers comprehensively described the design and development of SMS text messaging interventions for knee osteoarthritis [79] and back pain [83].

An SMS Text Messaging Intervention to Support Home Exercise Adherence for People With Knee Osteoarthritis

In 2019, Nelligan et al [79] comprehensively described a formal two-phase process to (1) identify behavior change targets and (2) design a library of SMS text messages to support adherence to home exercises for people with knee osteoarthritis. The development was guided by the recommended steps for developing text messaging–based programs for health behavior change published by Abroms et al [100] in 2015.

The first phase of development, comprising 3 stages, focused on target behavior, barriers, facilitators, and behavior change techniques using the Behavior Change Wheel framework [97,98]. Stage 1 drew on the literature to define the problem in behavioral terms, explaining the behavioral target and context and the barriers and facilitators for people with knee osteoarthritis in terms of participating in exercise mapped to domains in the Theoretical Domains Framework [96]. Barriers and facilitators relevant to the target behaviors were organized using the Capability, Opportunity, and Motivation–Behavior model for behavior change [97]. Stage 2 mapped barriers and facilitators to select intervention functions and behavior change techniques appropriate for implementation using SMS text messaging [93]. In stage 3, behavior change techniques for each function were identified from the Behavior Change Technique Taxonomy (version 1) [99].

The second phase involved the development of SMS text messaging functionality, specifically, a message library of content and determination of message frequency and level of interaction. Messaging content was derived by taking each barrier- or facilitator-linked behavior change technique identified in the first phase and constructing a relevant SMS text message. Message content was derived with input from 12 participants (1 person with knee osteoarthritis, 7 researchers, and 4 physiotherapists). In total, 3 authors derived the final message bank. A fourth author reviewed the final SMS text message wording to ensure that it was consistent with the Behavior Change Wheel mapping process and the identified behavior change techniques. The final message bank was organized into a 24-week schedule, assessed using literacy tools for readability, and tested by the authors for functionality and errors.

Author-provided examples of the mapping process and resulting SMS text message content for example barriers and facilitators are shown in Table 5.

Table 5. Example barrier and facilitator mapping process, abridged from Nelligan et al [79].

	COM-B[a] category [97]	TDF[b] domain [96]	Intervention function [99]	BCT[c] [99]	Resulting SMS text message content
Barrier mapping					
Forgetfulness	Psychological capability	10—memory, attention, and decision processes	Training	8.3—habit formation	"[Name], it can be hard to remember. We suggest making the exercises a habit. Set aside the same time each day to do them. It's much harder to forget when something is a daily routine."
Facilitator mapping					
Prioritizing exercise	Psychological capability	14—behavioral regulation	Enablement	10.9—self-reward	"Did you prioritize your exercises this week and get them done? Then reward yourself, [name]! Sticking to an exercise program for this long is a real accomplishment that deserves celebration."

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

[b]TDF: Theoretical Domains Framework.

[c]BCT: behavior change technique.

A Messaging Self-Management Intervention for LBP

In 2019, Fritsch et al [83] described the co-design process used to derive a bank of evidence-based lifestyle-focused messages for an LBP self-management text messaging intervention.

The authors used an iterative 2-phase co-design approach based on a framework used to design prevention messages for patients with cardiovascular disease previously published by Redfern et al [106] in 2014.

Phase 1 consisted of two 2-hour workshops intended to develop the concept, initial content, and messages. Workshop participants were researchers, clinicians with specific knowledge related to LBP, and consumer representatives from the support group Musculoskeletal Australia. At the first workshop, participants identified important domains relevant to LBP (exercise, education, mood, use of care, sleep, and medication) through reference to an evidence-based consumer resource (*Managing your pain: An A-Z guide*; Musculoskeletal Australia). The second workshop was focused on identifying sources of content for messages and duration, frequency, and timing of messages. Identified sources of content were relevant peer-reviewed literature, Australian and international clinical practice guidelines for LBP, and consumer group patient educational resources. Message frequency (4 messages per week) and timing (9 AM, 12:30 PM, 4 PM, and 6 PM) were based on previous work in coronary heart disease [107]. The development team considered that an intervention program duration of 12 weeks would be appropriate, with *exercise* domain messages being sent twice per week (emphasizing the importance of remaining active) and 1 message sent per week for each of the other domains.

This phase of the development process was also informed by previous work on factors related to engagement, perceived usefulness, behavior change, and delivery preferences for patients with coronary heart disease [105].

Following the workshops, a team comprising 2 researchers and 2 consumer representatives drafted evidence-based behavior change messages following the same theoretical approach by Redfern et al [106]. Messages were focused on education, motivation, or behavior change in the domains of *providing information or encouragement*; *prompting about consequences, intention formation, monitoring self-behavior, and barrier identification*; *advice about setting graded tasks*; and *strategies aimed at relapse prevention* and *the use of prompting and cues*. The team drafted an initial set of 82 positively phrased messages (by domain: 40 *exercise* messages, 10 *education* messages, 10 *mood* messages, 8 *use of care* messages, 7 *sleep* messages, and 7 *medication* messages) to take forward to the second phase of development.

In the second phase, the authors used a web-based survey of leaders in the field of LBP management to assess the appropriateness of the message content, gather opinions on the likelihood that the messages would be clinically effective, and make recommendations for message content improvement. The mean score for the messages from the expert review was 8.30/10. Messages with a score of <8/10 (34%) were modified in response to accompanying feedback. Subsequently, consumers scored each text message on utility of content, understanding, and language acceptability. Text messages with a consumer review score of <12/15 (31%) were revised according to feedback (mean score 12.5/15 points). Most frequently, consumer feedback focused on making the content more specific and less technical and including more examples.

Papers Describing the Design and Development of Messaging Within mHealth Apps

A total of 9% (4/47) of the papers described the design and development of more general mHealth interventions, where those interventions contained some use of messaging (alongside other mHealth features) for people with knee or hip osteoarthritis [80] and back pain [82], pain self-management for veterans [81], and women newly diagnosed with osteoporosis [76]. In total, 2% (1/47) of the papers focused on feature preferences for an app to support the self-management of chronic arthritis [77].

In each case, the design of the overall intervention was typically well described; however, the design of the content, timing, and frequency of the messaging components was not described in detail (Table 4). Because these papers provided little useful

messaging-specific design and development information, we do not discuss them any further.

Incidental Design and Development Information Contained in Papers Reporting the Results of mHealth Interventions

We found little useful design-related information contained within the papers describing results of interventions. Typically, the papers described the purpose and behavior of the messaging component within their intervention, but the design processes used to determine message content, timing, and frequency were described incidentally or not at all (studies shown in Tables 6 and 7) [30,35,36,38,47,50,60,62,63,65,66,72,74]. For example, one paper provided examples of messages intended to provide encouragement, education, or motivation but provided no explanation of how these were derived [30]. Similarly, some papers (4/47, 9%) made a passing reference to co-design processes involving patients and clinicians but provided limited detail [37,48,55,56].

Some papers (17/47, 36%) described the use of messaging *adaptivity* (ie, dynamic system-initiated changes to the delivery of messaging to personalize content, frequency, or timing of messages based on automated or manual triggers) or *individualization*. Triggers for adaptivity included self-reported exercise adherence [79], automated physical activity data derived from wearables [57,68,73], self-reported data [54,66,71], personalized goals [37,48,55,56], and manual adaptivity triggers initiated by study personnel [31,63] and health coaches [26,27,67,70]. However, in these papers, no substantial detail was provided on the design considerations or processes related to the development of the intervention's adaptive behavior.

Table 6. Messaging-specific intervention studies—efficacy and effectiveness.

Study, year	Objective	Duration and sample size (n)	Outcomes favoring messaging	Equivocal outcomes
Newell [60], 2012	Experimental study; for patients receiving chiropractic exercise advice, does text messaging with their practitioner, compared with no text messaging, improve exercise compliance?	4 weeks (32)	• Self-reported exercise compliance (NRS[a])	• Patient-perceived practitioner care (NRS)
Taylor et al [61], 2012	Experimental study; for patients attending outpatient physical therapy clinics, do SMS text message reminders, compared with no reminders, reduce clinic nonattendance?	1 day (679)	• Nonattendance at outpatient physiotherapy appointments (proportion)	• Appointment attendance (proportion) • Appointment cancellation (proportion)
Gandy et al [62], 2016	Observational study; for patients receiving an internet-delivered CBT[b] program for pain, is the addition of message skill practice prompts, compared with no prompts, feasible and effective?	8 weeks (195)	• Acceptability of SMS text messages (Likert scale)	• Treatment satisfaction (Likert scale) • Pain-related disability (RMDQ[c]) • Depression (PHQ-9[d]) • Anxiety symptoms (GAD-7[e]) • Pain intensity (WBPQ[f])
Theiler et al [49], 2016	Observational study; for patients with osteoporosis, do SMS text message reminders, compared with no reminders, improve adherence to drug therapy?	2 months (399)	• Engagement with health care providers (proportion)	• None
Chen et al [30], 2017	Experimental study; for patients with frozen shoulder, are reminder, encouragement, and educational messages delivered via mobile phone, compared with no messages, effective to increase exercise adherence and physical functioning?	2 weeks (66)	• Patient-reported compliance with shoulder exercises[g] • Range of motion in forward flexion and internal and external rotation (goniometry) • Patient satisfaction with SMS text messaging intervention (Likert scale)	• Shoulder function (Simple Shoulder Test) • Shoulder abduction (goniometry) • Shoulder pain (VAS[h])
Jamison et al [63], 2017	Experimental study; for patients with chronic pain using an app to record their progress, does 2-way supportive messaging, compared with no messaging, increase use or improve measures of pain or mood?	3 months (105)	• Patient perceptions (more appealing, easier to use, easier to navigate, and less bothersome) • Favored controls: participant perceptions of the responsiveness of providers to their reports	• Frequency of use • Pain (BPI[i]) • Activity interference (PDI[j]) • Mood (HADS[k])
Timmers et al [52], 2018	Experimental study; for patients with knee osteoarthritis, does delivering education via an interactive mobile app, compared with standard education, increase patients' knowledge of their illness and treatment options?	7 days (213)	• Actual knowledge[l] • Perceived knowledge[l] • Patient satisfaction (NRS)	• None
Bartholdy et al [53], 2019	Experimental study; for patients with knee osteoarthritis, do messages containing information and advice about the importance of performing daily activity, compared with no messages, lead to improved levels of activity?	6 weeks (38)	• None	• Time spent physically inactive, standing, and moving (accelerometry) • Self-reported change in physical activity[m] • Pain severity, quality of life, and disability[n]
Mary et al [32], 2019	Experimental study; for patients with rheumatoid arthritis, compared with standard pharmacist consultation, does a 15-min pharmacist-led counseling session or message reminders improve methotrexate adherence?	6 months (96)	• Medication adherence (CQR-19[o]) • Patient satisfaction (Likert scale)	• Medication adherence (GS[p] and MPR[q]) • Disease activityr

Study, year	Objective	Duration and sample size (n)	Outcomes favoring messaging	Equivocal outcomes
Kuusalo et al [31], 2020	Experimental study; for patients with rheumatoid arthritis, does using automated messages for enhanced monitoring, compared with routine care, improve disease activity and remission and quality of life?	6 months (166)	• Physical functioning at 6 months after randomization (SF-36[r]) • Health care resource use (nurse telephone contact)	• Patients' confidence in treatment (VAS) • Physical and mental health–related quality of life (SF-36) • Physical functioning at 12 months after randomization (SF-36) • Disease activity[s] • Group rates of remission (proportion) • Health care resource use[t]
Anan et al [66], 2021	Experimental study; for workers with neck and shoulder stiffness and pain or LBP[u], does an AI[v]-assisted interactive health promotion system that operates through a mobile messaging app, compared with usual workplace exercise routine, lead to an improvement in musculoskeletal symptoms?	12 weeks (94)	• Pain intensity (Likert scale) • Perceived symptom improvement (Likert scale)	• None

[a]NRS: numeric rating scale.

[b]CBT: cognitive behavioral therapy.

[c]RMDQ: Roland-Morris Disability Questionnaire.

[d]PHQ-9: 9-item Patient Health Questionnaire.

[e]GAD-7: 7-item Generalized Anxiety Disorder Scale.

[f]WBPQ: Wisconsin Brief Pain Questionnaire.

[g]Calculated as days answered "yes" to exercise/total days in the intervention.

[h]VAS: visual analog scale.

[i]BPI: Brief Pain Inventory.

[j]PDI: Pain Disability Inventory.

[k]HADS: Hospital Anxiety and Depression Scale.

[l]Customized scale (actual perceived level was measured on a 0-36 scale, or perceived level was measured on a 0-25 scale).

[m]Customized scale for change in self-reported physical activity (included no change, less time, or 0-3.5 more times compared to baseline).

[n]Knee Injury and Osteoarthritis Outcome Score.

[o]CQR-19: Compliance Questionnaire on Rheumatology.

[p]GS: Girerd score.

[q]MPR: medication possession ratio.

[r]SF-36: 36-item Short-Form Health Survey.

[s]Disease Activity Score–28 for Rheumatoid Arthritis, Health Assessment Questionnaire, erythrocyte sedimentation rate, and C-reactive protein.

[t]Except nurse telephone contact.

[u]LBP: low back pain.

[v]AI: artificial intelligence.

Table 7. Mobile health (mHealth) studies with an embedded messaging component.

Study, year	Duration and sample size (n)	Intervention and messaging features	Summary of findings
Kristjánsdóttir et al [35,47], 2013—short- and long-terms effects of a smartphone-based intervention with diaries and therapist feedback to reduce catastrophizing and increase functioning in women with chronic widespread pain	4 weeks (140)	Experimental study: • Following a 4-week inpatient rehabilitation program, the study randomized participants to either a smartphone intervention or no smartphone intervention (controls). Follow-up occurred immediately after the intervention at 5 and 11 months. Smartphone intervention: • Initial in-person session with a nurse to discuss functioning, health-related behavior goals, support needs, values, and value-based activities. • Online web-based diaries completed 3 times a day on a smartphone covering pain interference, feelings and thoughts related to avoidance, catastrophizing and acceptance, planned and previous practice of self-management activities, and daily value-based and practical activities. • Daily written situational feedback from a therapist based on the information entered in the diaries • Audio files with guided mindfulness exercises SMS text messaging was used to prompt participants to complete their diaries and notify them when therapist feedback had been provided.	Reported between-group effects: • Small effects on catastrophizing (PCS[a]) and value-based living (CPVI[b]) immediately after the intervention. Effect was nonsignificant at the 5-month follow-up. • Moderate effect on acceptance (CPAQ[c]) immediately after the intervention and at the 5-month follow-up. • Moderate effect on sleep disturbance (VAS[d]) and functioning and symptom severity [69] at the 5-month follow-up • No effect on pain • No significant between-group differences at the 11-month follow-up
Dekker-van Weering et al [68], 2015—pilot study of an activity-based feedback system for people with LBP[e]	15 days (17)	• Observational study; participants' daily activity was measured using a body-worn sensor. • Real-time, hourly, personalized feedback was tailored to the individuals' objectively measured activity level (eg, to discourage movement, to encourage movement, or a neutral message).	• Encouraging feedback led to an increase in PA[f]. • Discouraging feedback led to a decrease in PA. • Greater participant response to feedback messages was associated with decreased pain scores.
Demmelmaier et al [75], 2015, and Nordgren et al [51], 2018—short- and longer-term (2-year) evaluation of an outsourced program to encourage PA in people with RA[g]	12 months (191)	• Observational study • Intervention: community-based exercise, support groups to facilitate behavior changes and feedback from physical therapists • 2 messages each week were sent to collect data on how often the participant engaged in circuit training and moderately intense exercise.	• Patients perceived the use of professional coaches and text messages to support the adoption of physical exercise as helpful. • While improvements in self-reported physical activity, the proportion of participants who maintained increased physical activity, decreased significantly during year 2 of the study. Grip strength and quality of life reduced significantly during year 1 and 2 of the intervention. Reductions in activity limitation, systolic blood pressure and waist circumference were observed during second year. With most other health improvements sustained during year 1 and 2 of the study • Participants reported that the text messages were a good reminder to engage in exercise (rated 4/5 on perceived value).

Study, year	Duration and sample size (n)	Intervention and messaging features	Summary of findings
Thomsen et al [37], 2016; Thomsen et al [48], 2017; and Thomsen et al [56], 2020—evaluating the effect of motivational interviewing and messages on sitting time in patients with RA	4 months; follow-up: 10 and 22 months (150)	• Experimental study; patients randomized to 3 individual motivational counseling sessions and messages aimed to reduce sedentary behavior (intervention) versus no contact and instructions to maintain usual lifestyle (controls). • Manually created individual tailored messaging was used to remind participants of goals that they had set in their individual counseling sessions. Participants selected the frequency and timing of messages.	Reported between-group effects: • Reduction in sitting time of –2.2 hours per day (95% CI –2.72 to –1.69) favoring the intervention • Secondary measures, including fatigue, pain, self-efficacy, and HRQoL[h], also favored the intervention.
Lambert et al [36], 2017—evaluating whether patients with MSK[i] conditions have better adherence to home exercises when content is delivered via an app-based intervention compared to paper handouts	4 weeks (77)	• Experimental study; participants randomized to receive home exercise program information via an app together with phone calls and motivational messages (intervention) versus paper handouts (controls) • All participants were prescribed a 4-week exercise program.	Reported between-group effects: • Small significant differences in adherence to the exercise program (NRS[j]; 1.3/11 points, 95% CI 0.2-2.3) and function (PSFS[k]; NRS 0.9/11 points, 95% CI 0.1-1.7) • There were no significant differences in disability, patient satisfaction, perceived global impression of change, or assessor-reported adherence.
Lee et al [74], 2017—assessed the effectiveness of app-based exercises supported by weekly messages in office workers with chronic neck pain	8 weeks (20)	• Small experimental pilot study in the workplace; participants randomized to receive prescribed exercises via a smartphone app (intervention) versus receiving a brochure and pain education (controls) • Both groups received weekly education and encouragement messages.	Reported between-group effects: • Statistically significant difference in pain intensity (VAS; 0-10) and functional disability (NDI[l]; expressed as a percentage); note: despite randomization, compared with controls, the intervention group had higher pain intensity at baseline (mean VAS score 5.20, SD 2.19 vs 4.02, SD 1.75) and a much higher NDI (mean 26.8, SD 9.68 vs 17.70, SD 9.20). • No between-group differences in the secondary outcomes of strength, fear avoidance, and quality of life (SF-36[m])
Chhabra et al [24], 2018—assessed the effect of a smartphone app on pain and function in patients with chronic LBP	12 weeks (93)	• Experimental study; participants randomized to receive daily activity goals (back and aerobic exercise) in addition to written prescriptions (medication and recommended level of PA) provided through an app (intervention) versus written prescriptions only (controls) • Activity goals were personalized based on participants' health status, activities of daily living, and daily activity progress. Automated reinforcement messages were delivered via app push notifications.	• Reported between-group effects: no significant difference in pain (NRS) and significant difference in disability (MODI[n]) favoring the intervention
Lo et al [64], 2018—assessed the feasibility of an AI[o]-embedded mHealth app for chronic neck and back pain in promoting self-management	Unclear (161)	• Observational study • Intervention: educational content including information on the pathophysiology of neck and back pain and principles of exercise for management of pain and coping strategies. Information was pushed via messages to participants' social media accounts.	• Pretest-posttest increase in time spent on rehabilitation exercises (custom questionnaire) • Mean "self-reported improvement" of 65% (0-100 scale) • Pretest-posttest reduction in pain from a median of 6 (IQR 5-8) to 4 (IQR 3-6; NRS 0-10) • Perceived usability was 73/100 (cutoff for "acceptable" was 68/100; SUS[p])
Mecklenburg et al [26], 2018—assessed the efficacy of a remotely delivered digital care program for chronic knee pain	12 weeks (162)		

Study, year	Duration and sample size (n)	Intervention and messaging features	Summary of findings
		• Experimental study; participants randomized to receive involved sensor-guided exercise therapy, psychoeducation, cognitive and behavioral therapy, and behavioral monitoring via the "Hinge Digital Care Program" (intervention) versus 3 digital education sessions and TAU[q] (controls) • The app included a coach and peer support discussion via messaging. Message or email reminders were sent if participants did not appear to engage at the recommended intensity of the program.	Reported between-group effects: • Significant difference in pain and physical functioning (KOOS[r]), pain (VAS), and stiffness (VAS) favoring the intervention • Interest in and the likelihood of needing surgery decreased, and patients' understanding of their condition improved. • Estimated surgery cost savings of US $4340 over 1 year and US $7900 over 5 years for participants who completed the digital care program compared to controls
Molinari et al [50], 2018—assessed the efficacy of using guided imagery to have patients with fibromyalgia picture their best possible selves	4 weeks (80)	• Experimental study; participants randomized to receive "Best Possible Self," a web-based app multimedia system to support patients through guided imagery (intervention) versus "Daily Activities" (active controls) • The active control condition was not well described. • Participants in both arms received 2 reminders each week via SMS text messaging prompting them to practice the guided imagery exercise.	Reported between-group effects: • Postintervention improvements in depression, positive affect, and self-efficacy favoring "Best Possible Self" • At the 30-month follow-up, there was improved optimism and negative affect favoring "Best Possible Self."
Rabbi et al [69], 2018—evaluate the feasibility and acceptability of a personalized app for PA recommendations for adults with chronic pain	5 weeks (10)	• Observational study • Intervention: MyBehaviourCBP app, which generated PA recommendations based on sensor-detected PA. Recommendations were contextualized to the environment (road names), and new suggestions were continuations of the users' repeated behaviors (eg, "Take walking break near Thompson St for 24minutes today"). • Study comprised a 1-week period with no recommendations, 2 weeks with generic recommendations provided by an expert, and 2 weeks with automated recommendations.	• Participants found the dynamic recommendations easier to adopt than the static generic recommendations. All participants found the recommendations "helpful." • Walking duration during the dynamic phase was greater than in the static phase (+4.9 min/d); no significant differences in pain ("Likert" scale; 0-10) and nonwalking exercise (min) were found. • Qualitative feedback included that participants wanted notifications in the moment and adaptivity in relation to the weather or weekend days in addition to information related to the relationship between pain and activity levels.
Selter et al [70], 2018—described patient engagement and perceived utility and assessed the validity of a smartphone app module to quantify the functional status for people with chronic LBP	12 weeks (93)	• Observational study • Intervention: physical therapy program using the Limbr app involving 3 daily self-reports of pain and activity level and chat-based health coaching • Health coaches monitored data and sent participants messages to provide support and remind them to interact with the program. • Participants with low engagement (eg, only 1-2 interactive components per week) were sent weekly emails containing visual feedback on their use.	• High level of attrition (38% completion rate); engagement was reported as "high amongst completers." • Depending on the type of self-report, 21%-32% interacted with the app. • 76% of patients found that daily notifications helped them remember to complete their exercises, and 71% found that they helped them complete the daily surveys.
Wang et al [38], 2018—evaluated the effectiveness of a community-based self-management lifestyle program for young to middle-aged women with knee pain living in rural Australia	12 months (649)	• Experimental study; participants randomized to receive 1 group session, monthly SMS text messages, 1 phone coaching session, and a program manual (intervention) versus 1 session of general women's health education (controls) • Program intended to improve lifestyle and prevent weight gain.	Reported between-group effects: • Overall, no difference in the risk of knee pain worsening over 12 months • For women who had knee pain at baseline (WOMAC[s]; 35% of participants), there was a lower risk of knee pain worsening over 12 months favoring the intervention, although this effect was only statistically significant for women with a BMI of ≥25 kg/m2 (OR[t] 0.28, 95% CI 0.09-0.87).

Study, year	Duration and sample size (n)	Intervention and messaging features	Summary of findings
Frei et al [65], 2019—assessed the effectiveness and feasibility and participant perceptions of a community-based PA intervention	12 months (29)	• Observational study • Intervention: the intervention facilitates and encourages participants to arrange walking groups within their local area. • The messaging app, WhatsApp, was used to facilitate communication between the participants and study team.	• Increased the minutes that participants engaged in moderate- to vigorous-intensity activity; no significant changes in step count • 76% of participants reported that they attempted to recruit their peers to participate in the intervention. • 62.4% of participants sent messages. • Participants continued to organize walking groups via WhatsApp after the study team ceased their involvement.
Ji et al [54], 2019—described the design and preliminary evaluation of an interactive mHealth tool designed to help with the management and self-management of ankylosing spondylitis	13.3 months (1201)	• Observational study • Intervention: app designed to provide patient education on disease management and assist patients with medication adherence (SpAMS[u]) • The tool consisted of a patient and physician portal and was linked to the social media app WeChat to allow for communication between physicians and patients, collect follow-up data, and obtain patient feedback.	• Improvement in the proportion of patients with inactive disease or low disease activity from baseline to a mean follow-up time of 13.3 months (57.2%-79.2%) • Problems solved using SpAMS avoided 29.1% of clinic visits. • Average savings of 5.3 hours per patient in travel time and US $51 per person in personal expenses (15% of Chinese monthly disposable income) on physicians.
Shebib et al [27], 2019—evaluated the efficacy of a digital care program for patients with LBP	12 weeks (177)	• Experimental study; participants randomized to receive a remotely administered digital care program that involved cognitive behavioral therapy, sensor-guided exercise therapy, education, symptom tracking, and unlimited personal coaching (intervention) versus 3 digital education articles and TAU (controls)	Reported between-group effects: improvements favoring the intervention in pain (MvK[v]: mean –16.4, 95% CI –22 to –10.9; VAS: mean –16, 95% CI –22.5 to –9.4), disability (MvK: mean –13, 95% CI –19.3 to –6.7; ODI[w]: mean –4.1, 95% CI –6.5 to –1.8), impact on daily life (VAS: mean –11/8, 95% CI –19.3 to –4.3), and understanding of their condition and treatment options (0-4; mean 0.5, 95% CI 0.2-0.7) and decreased interest in back surgery (mean –0.4, 95% CI –0.7 to –0.1)
Støme et al [55], 2019—feasibility study Investigating the acceptability, usability, and utility of a mobile app supporting goal achievement in patients with OA[x]	12 weeks (12)	• Observational study • Intervention: Vett app sent participants personal reminders to complete tasks that aligned with their PA, weight loss, and stress reduction goals. • Participants were assigned 2 to 3 weekly physician-developed tasks and self-monitored their progress or received individualized feedback.	• Primary reported outcome was mean goal achievement, which had a pretest-posttest improvement of 48%. • Mean user satisfaction was 81/100, and technical usability was 80/100 to 84/100.
Zaslavsky et al [57], 2019—pilot study that assessed the feasibility and preliminary efficacy of a self-management mHealth intervention aimed at improving sleep among older adults with OA and disturbed sleep	19 weeks (24)	• Observational study • PA feedback based on wearable (Fitbit) data • Participants received weekly personalized messages with motivational feedback in relation to their step count data. • Participants who maintained or increased their step count received reinforcing messages. Those with declining step counts received encouraging messages. • Participants also received motivational interviewing geared toward discussing the participants' goals and strategies to facilitate behavior change.	• Small pretest-posttest improvements in mean insomnia (ISI[y]; 1.2 points, 95% CI 2.45-0.05) and ASD[z] (2.5 points, 95% CI 0.9-4.1) and self-reported overall sleep quality (derived from sleep diaries; 0.3 points, 95% CI 0.02-0.58) • Nonsignificant improvements in step count, pain intensity, pain-related disability, self-efficacy, and sleep diary data and variables
Almhdawi et al [72], 2020—assessed the efficacy of an mHealth smartphone app in patients with LBP	6 weeks (39)	• Experimental study; participants randomized to receive evidence-based instructions, therapeutic exercises, and reminders (intervention) versus instructions about nutrition (controls) • Both arms received the app; the intervention group received reminders for walk breaks, posture, and exercises.	

Study, year	Duration and sample size (n)	Intervention and messaging features	Summary of findings
			Reported between-group effects: • Significant reductions (Cohen *d*) in pain intensity (VAS; 1.71, 0-11) and pain-related disability (ODI; 1.08) and improvements in physical quality of life (SF-12[aa] PCS[ab]; 1.18) • No significant differences in mental quality of life (SF-12 MCS[ac]); depression, anxiety, and stress symptoms (DASS-21[ad]); sleep quality (PSQI[ae]); and self-reported PA (IPAQ[af])
Bailey et al [67], 2020—evaluated the efficacy of a digital care program in patients with chronic knee and back pain	12 weeks (10,264)	• Observational study • Intervention: Hinge Health app, which delivered education, sensor-guided exercise therapy (using a Bluetooth wearable sensor), behavioral health support, and 1:1 health coaching • Patients were assigned a health coach, and communication occurred via SMS text messaging, email, or app-based messaging.	• 78% completed the program, with 69.6% achieving minimally important change in pain (20 points or 30% from baseline; VAS). • Greater reduction in pain scores was associated with increasing levels of engagement in exercise therapy and participant-to-coach interactions. • Significant reduction in the proportion of participants categorized as having depressive (PHQ-9≥5) or anxiety (GAD-7≥5) symptoms at 11 weeks compared with baseline (depression decreased by 57.9%, and anxiety decreased by 58.3%) • Mean 1-year surgery likelihood score (subjective self-report response to the following question: “What do you think are the chances you’ll have [back/knee] surgery in the next year, in %?”; 0%-100%) decreased by 67.4% with respect to baseline.
Hasenöhrl et al [71], 2020—evaluated the feasibility and acceptance of orthopedists prescribing therapeutic exercises via a smartphone app to patients with nonspecific back pain	4 weeks (prestest-posttest assessment: 27 and semistructured interview: 16)	• Small observational study with a qualitative component • Intervention: individual physician-selected exercises sent via in-app messaging (n=27 participants). The physician could provide encouragement and mental support or unlock new exercises. • Qualitative component: interviews and thematic analysis with a random sample of 16 of the 27 participants (research question not well described)	• Reduction in mean hip circumference (–1.54, SD 2.75 cm) • Reduction in back pain (ODI; mean –2.67, SD 4.99) • Quality of life (SF-36): improved physical functioning (+5, SD 11.9); improved bodily pain (+14.8, SD 7.8); vitality (+7.2, SD 14.8) • Participants reported that they would have preferred 2-way messaging
Nelligan et al [78], 2020 (qualitative), and Nelligan et al [58], 2021 (RCT[ag])—explored the experiences and attitudes of patients with knee OA who participated in an mHealth intervention to support exercise	24 weeks (16)	• Qualitative study (n=16 participants) embedded in an RCT with a targeted recruitment of n=206 • Participants were randomized to receive website+SMS text messaging adherence support+home exercises (intervention) versus website only (controls). • The website contained educational information (OA and exercise), PA recommendations, and prescription of knee-strengthening exercises. • If participants adhered to the exercise program, they received a positive reinforcement message. If participants did not adhere, they were asked to select a barrier. All participants received behavior change techniques to assist with exercise adherence.	

Study, year	Duration and sample size (n)	Intervention and messaging features	Summary of findings
			• Five themes were reported: (1) technology was easy to use, (2) facilitators to exercise participation (credible information, website features, exercises that could be done unsupervised, and freedom to adapt exercises to suit needs), (3) sense of support and accountability (SMS text messaging served as a good reminder to engage in exercise, was easy to use, and held them accountable to weekly exercise; SMS text message tone and automation could trigger guilt or shame; and inability to contact someone when needed), (4) positive outcomes (symptom improvement, self-management confidence, and encouragement of active living), and (5) suggestions for real-world application (preference for provision by a health professional and should be subsidized or low cost). • Primary outcomes favored in the intervention group in the RCT: decrease in pain scores (NRS; mean difference=1.6, 95% CI 0.9-2.22); increase in function (WOMAC; mean difference=5.2, 95% CI 1.9-8.5) • Most secondary outcomes favored the intervention, which included KOOS pain, function in sport and recreation, ASES[ah] pain and function subscales, AQOL-6D[ai], and overall satisfaction (Likert scale). Changes in PASE[aj], ASES function, and SEE[ak] were similar between groups. • Average participant message response rate was 73% (SD 7.5%), and 8% opted out. Patient perceptions (7-item Likert scale): mean perceived usefulness was 5.3 (SD 1.8), and mean agreement with message frequency was 5.3 (SD 1.7). Adverse events: 15.3% (intervention) vs 6.3% (control); a greater portion of the intervention group had knee pain (9.6%) compared to those in the control group (1.3%); a similar proportion used cointerventions throughout the study period.
Nordstoga et al [73], 2020, and Mork and Bach [82], 2018 (protocol)—evaluated the usability and acceptability of an mHealth intervention, selfBACK, in patients with LBP	4 weeks (16)	• Observational study • Intervention: the smartphone app provided participants with weekly self-management plans with content related to PA, flexibility exercises, and patient education. • Behavior change techniques were incorporated into the app (eg, goal setting, feedback, monitoring, information about health consequences, and prompts). • Motivational notification messages were sent to the participants' smartphones.	• Participants received an average of 1.8 notifications per day. • Participants opened 42% of the notifications; of those opened, 90% were liked, and 8% were disliked; notifications of goal attainment were most frequently liked by participants. • There was a lack of consensus on the frequency and appropriateness of motivational notifications. • Motivational reminders served as facilitators of the intervention. • 50% of the participants found the motivational messages useful. • 30% of the participants found the notifications to be irrelevant and not functioning properly (eg, unsynchronized).
Pelle et al [59], 2020—investigated the effect of an mHealth intervention on secondary health care use in people with hip and knee OA	6 months (427)		

Study, year	Duration and sample size (n)	Intervention and messaging features	Summary of findings
		• Experimental study; participants assigned to receive a self-management app (Dr Bart mHealth app) intended to support goal setting and education and enhance motivation, with daily push notifications providing reminders on selected goals and educational information (intervention) versus TAU (controls) • TAU consisted of any treatments initiated by participants.	Reported between-group effects: • No difference in knee- or hip-related OA secondary health care use • Significant group differences favoring the intervention were found between baseline and the 6-month follow-up for symptoms (mean difference=2.6, 95% CI 0.4-4.9), pain (mean difference=3.5, 95% CI 0.9-6.0), and activities of daily living (mean difference=2.9, 95% CI 0.2-5.6; HOOS[al] and KOOS). • No differences were found in any other outcome measures.

[a]PCS: Pain Catastrophizing Scale.
[b]CPVI: Chronic Pain Values Inventory.
[c]CPAQ: Chronic Pain Acceptance Questionnaire.
[d]VAS: visual analog scale.
[e]LBP: low back pain.
[f]PA: physical activity.
[g]RA: rheumatoid arthritis.
[h]HRQoL: health-related quality of life.
[i]MSK: musculoskeletal.
[j]NPRS: numeric rating scale.
[k]PSFS: Patient-Specific Functional Scale.
[l]NDI: Neck Disability Index.
[m]SF-36: 36-item Short-Form Health Survey.
[n]MODI: modified Oswestry Disability Index.
[o]AI: artificial intelligence.
[p]SUS: System Usability Scale.
[q]TAU: treatment as usual.
[r]KOOS: Knee Injury and Osteoarthritis Outcome Score.
[s]WOMAC: Western Ontario and McMaster Universities Osteoarthritis Index.
[t]OR: odds ratio.
[u]SpAMS: Smartphone Spondyloarthritis Management System.
[v]MvK: modified Von Korff scales.
[w]ODI: Oswestry Disability Index.
[x]OA: osteoarthritis.
[y]ISI: Insomnia Severity Index.
[z]ASD: acceptance of sleep difficulties.
[aa]SF-12: 12-item Short-Form Health Survey.
[ab]PCS: Physical Component Summary.
[ac]MCS: Mental Component Summary.
[ad]DASS-21: Depression, Anxiety, and Stress Scale–21.
[ae]PSQI: Pittsburgh Sleep Quality Index.
[af]IPAQ: International Physical Activity Questionnaire.
[ag]RCT: randomized controlled trial.
[ah]ASES: Arthritis Self-Efficacy Scale.
[ai]AQOL-6D: Assessment of Quality of Life.
[aj]PASE: Physical Activity Scale for the Elderly.
[ak]SEE: Self-Efficacy for Exercise.
[al]HOOS: Hip Injury and Osteoarthritis Outcome Score.

RQs 4 and 5: Evaluation Methods, Efficacy, Effectiveness, and Economics

To avoid repetition, the findings of review questions 4 and 5 are reported together. A total of 28% (11/40) of the studies directly compared the use of messaging with an alternative; a further 60% (24/40) of the studies evaluated mHealth interventions with embedded use of messaging.

Studies Comparing the Use of Messaging With an Alternative

Of the 11 studies that directly compared messaging to an alternative, 9 (82%) had an experimental design and 2 (18%) were observational. In most cases, the comparator or control condition was no messaging or treatment as usual, with outcome measures varying by the intent of the intervention. Of these 11 studies, 3 (27%) [30-32] were described in the previous review on the effectiveness of text messaging interventions on the management of musculoskeletal pain [22], and the remainder were not, likely because they did not meet the inclusion criteria or were published later [49,52,53,60-63,66].

Overall, the outcomes either favored the messaging condition or were equivocal.

Examples of outcomes favoring messaging interventions included improved knowledge of the illness and the available treatment options and physical activity for knee osteoarthritis [52,53], improved medication adherence and physical functioning for RA [31,32], improved attendance to outpatient physiotherapy [61] and engagement with general practitioner [49], and improved exercise compliance for frozen shoulder [30] and mixed musculoskeletal conditions [60]. However, despite participants sometimes reporting messaging as *acceptable* [62] or *appealing* [63], and while improved pain intensity was found in participants with neck and shoulder pain and LBP [66], some studies (4/40, 10%) reported equivocal findings for important patient outcomes such as time spent physically active [53], pain [53,63], and quality of life [31,53].

In no studies did the primary outcome favor the control condition. In only one study, a secondary outcome (clinician responsiveness) favored the control condition. In this study, patients with chronic pain recorded their progress using an app, with intervention recipients also having access to messaging with their clinician (controls could report progress but had no messaging). Control participants perceived their clinicians to be more responsive to their progress reports [63].

No studies reported economic outcomes.

The studies are summarized in Table 6.

mHealth Studies With Embedded Messaging Components

A total of 24 studies evaluated mHealth interventions containing some form of embedded messaging component (n=11, 46% experimental; n=11, 46% observational; n=1, 4% observational with a qualitative component; and n=1, 4% qualitative embedded within an experimental study).

The results of efficacy and effectiveness were mixed, but because messaging was embedded within a larger mHealth intervention, it was not possible to isolate the messaging-specific effects from the overall intervention effects.

A total of 8% (2/24) of the studies reported economic outcomes—avoided surgery costs associated with a digital education program for chronic knee pain, in which messaging was used for coaching or peer support and program engagement reminders [26], and reduced travel time associated with a self-management mHealth tool for ankylosing spondylitis, in which social media messaging was used for communication between physicians and patients [54].

While it was not possible to isolate messaging-specific effects, these studies are included for completeness and summarized in Table 7.

Discussion

Principal Findings

To our knowledge, this is the first study to comprehensively map how mobile messaging has been used in the treatment and self-management of musculoskeletal conditions. We mapped the conditions and purposes for which messaging has been used and the approaches used to design and develop messaging interventions and summarized the evidence of efficacy, effectiveness, and economics from both experimental and observational studies. Our intent was to draw together all the available relevant information to help inform the future design of messaging interventions for people with musculoskeletal conditions and identify research gaps.

While previous reviews in this area are few, this work builds on 3 existing syntheses of the effectiveness of messaging interventions for people with musculoskeletal conditions. One review focused specifically on text messaging interventions for musculoskeletal pain [22]. The review included studies across a range of musculoskeletal problems and included both studies in which messaging was added to and compared with usual care (findings of positive effects only on exercise and medication adherence) and studies in which messaging was a component of a larger intervention (reporting some small effects on pain intensity, function, care-seeking behavior, exercise and medication adherence, and quality of life). Overall, the quality of the evidence was low. The 2 other reviews focused more generally on digital health for managing musculoskeletal conditions [23] and mHealth interventions for people with RA [46].

In this review, all the included studies that assessed intervention efficacy or effectiveness (on pain [30,53,63,66], function [30,31,63], disability [53], adherence to the intervention [30,60,63], physical activity levels [53], appointment attendance [49,61], health care contact [31], mood [63], quality of life [31,53], remission [31], and disease activity [31,32]) reported either equivocal findings or findings favoring messaging.

The notable absence of studies reporting negative outcomes may suggest publication bias. The lack of economic studies is also concerning; no messaging-specific studies reported economic outcomes. While, of the 40 studies, 2 (5%) digital or mHealth studies with messaging components did report

economic outcomes, including avoided surgery costs and reduced travel time [26,54], the embedded nature of messaging means that it is not possible to attribute the observed savings specifically to the messaging components. While a previous review has shown messaging to be cost-effective in some circumstances, there is no information on the economic effects on musculoskeletal pain conditions; in cases in which messaging interventions are shown to be effective, further studies should be conducted to assess their economic effects [22,108].

We identified studies describing the use of messaging across a range of musculoskeletal conditions, with rheumatic diseases representing almost half (19/40, 48%) of the included studies, of which two-thirds (13/19, 68%) focused on osteoarthritis and RA. Somewhat surprisingly given its high population prevalence, back pain was represented by less than a quarter of primary studies (9/40, 23%). A further quarter of the studies (10/40, 25%) addressed multiple musculoskeletal conditions, but most (30/40, 75%) targeted single musculoskeletal conditions and pain sites despite evidence that musculoskeletal conditions often do not occur in isolation (eg, in Australia, 64% of people with back pain and 74% of people with arthritis have at least one other chronic condition) [3].

More than 80% of the included primary studies (34/40, 85%) focused on the use of prompts and reminder messages to foster positive behavior change at the individual level, most commonly in combination to encourage movement (eg, to increase physical activity, reduce sitting time, and improve compliance with prescribed exercise); compliance with prescribed medication; the practice of coping skills; and the meeting of personal goals. While a small number of studies (10/40, 25%) described the use of unidirectional or 2-way messaging with a health coach or exercise or sports scientist for support and encouragement, more studies (11/40, 28%) described the use of automated and unidirectional messaging, which, while economical on resources, may limit effectiveness in fostering behavior change.

While most studies (34/40, 85%) focused on influencing individual behavior change, there appears to be limited research into the use of messaging to improve treatment or self-management at the broader system level (eg, to improve health care processes, handover communication, and continuity of care between providers). One study on RA used SMS text messaging–based monitoring of medication adherence and disease progression to inform follow-up nurse contact but found no difference in the primary outcome of remission [31]. A second study on a digital health platform for ankylosing spondylitis management consisting of a patient and physician portal and 2-way chat via social media reported improvements in the proportion of patients with inactive disease and an avoidance of 29.1% of in-person clinic visits [54]. Future studies could focus on addressing the gaps in knowledge on more process- or system-focused interventions.

Overall, we found limited information about messaging design. One study (1/40, 3%) used development processes previously described by Redfern et al [106], which, while originally focused on cardiovascular event prevention messaging, have since been more widely adopted and adapted in the co-design of text messaging interventions, including for diabetes prevention [109], endometriosis support [110], and support after breast cancer treatment [111]. Similarly, one study (1/40, 3%) used guidance for the development and testing of messaging in health behavior change developed by Abroms et al [100].

Only 4% (2/47) of the papers provided comprehensive information about message design and development [79,83] and highlighted the importance of taking a formal approach and having a theoretical underpinning and meaningful consumer involvement. Beyond these 2 papers, design was typically poorly described or not described at all, and many projects appeared to leapfrog from concept to implementation with a limited or absent design phase, perhaps not recognizing the importance of formal design for subsequent adoption.

While some papers (8/47, 17%) did describe elements of participatory design or co-design, some papers (4/47, 9%) provided limited detail or had limited consumer involvement, or in some cases, consumers were involved after the design had already been conceived by researchers and clinicians. In most cases, we found that papers described the theoretical basis underlying their intended behavior change well, but consistent with a previous review [22], we found few detailed descriptions of the process through which the content, timing, and frequency of the messages were derived. This remains an important weakness in the musculoskeletal literature specifically and has been identified more generally by others [100,106]. Further work should be conducted to elicit preferences regarding these processes from people across the spectrum of musculoskeletal conditions.

The strengths of this review include our comprehensive search strategy and the inclusion of a wide range of studies and designs, providing a rich map of the literature expanding the insights provided by previous effectiveness-focused reviews. A limitation is that we focused specifically on the use of messaging in patient care and self-management, and it is possible that there are other messaging applications relevant to people with musculoskeletal conditions. However, we made this review as broad as possible within available resources. Because a comprehensive synthesis was time-consuming and we last conducted the searches in 2022, there may be important, more recent studies that we missed in this review. Similarly, resources did not allow us to review the gray literature.

Conclusions

In conclusion, messaging has been used for the care and self-management of a range of musculoskeletal conditions with generally favorable outcomes reported. Nonetheless, there are areas that should be addressed by future research to improve the quality of intervention design, which will hopefully translate into uptake and sustainability. First, preferences related to messaging content, timing, and frequency should be further explored specifically among people with musculoskeletal conditions, eliminating the reliance on information from other disciplines. Second, teams should incorporate digital intervention design expertise, follow formal design processes, and clearly describe design considerations and processes used. Finally, in cases in which messaging interventions are shown to be effective, further studies should be conducted to assess

their economic effects and practical considerations related to implementation and sustainability.

Authors' Contributions

MS, NA, and RE conceptualized the study. SSR and NA developed, pilot-tested, and refined the search strategy. SSR conducted the final searches. SSR, JL, CEE, RE, CR, SR, and NA screened the articles. JL, CEE, and NA extracted and confirmed the data. All authors contributed to data synthesis, manuscript development, and critical review. All authors approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOCX File , 20 KB - mhealth_v12i1e55625_app1.docx]

Multimedia Appendix 2
PRISMA-ScR Checklist.
[PDF File (Adobe PDF File), 103 KB - mhealth_v12i1e55625_app2.pdf]

References

1. Maher C, Ferreira G. Time to reconsider what Global Burden of Disease studies really tell us about low back pain. Ann Rheum Dis 2022 Mar;81(3):306-308 [FREE Full text] [doi: 10.1136/annrheumdis-2021-221173] [Medline: 34583922]
2. Australian burden of disease study: impact and causes of illness and death in Australia 2018. Australian Government, Australian Institute of Health and Welfare. 2021. URL: https://www.aihw.gov.au/reports/burden-of-disease/abds-impact-and-causes-of-illness-and-death-in-aus/summary [accessed 2023-05-10]
3. Chronic musculoskeletal conditions. Australian Government, Australian Institute of Health and Welfare. 2024. URL: https://www.aihw.gov.au/reports/chronic-musculoskeletal-conditions/musculoskeletal-conditions/contents/summary [accessed 2024-06-30]
4. Musculoskeletal health: trends, risk factors and disparities in England. Office for Health Improvement & Disparities, United Kingdom Government. URL: https://fingertips.phe.org.uk/static-reports/musculoskeletal-conditions/musculoskeletal-health-trends-risk-factors-and-disparities-in-england.html [accessed 2024-06-30]
5. Data file to accompany the published report: musculoskeletal health: trends, risk factors and disparities in England. United Kingdom Government. URL: https://assets.publishing.service.gov.uk [accessed 2024-07-10]
6. Early J, Gonzalez C, Gordon-Dseagu V, Robles-Calderon L. Use of mobile health (mHealth) technologies and interventions among community health workers globally: a scoping review. Health Promot Pract 2019 Nov;20(6):805-817. [doi: 10.1177/1524839919855391] [Medline: 31179777]
7. Fernández CE, Maturana CA, Coloma SI, Carrasco-Labra A, Giacaman RA. Teledentistry and mHealth for promotion and prevention of oral health: a systematic review and meta-analysis. J Dent Res 2021 Aug;100(9):914-927. [doi: 10.1177/00220345211003828] [Medline: 33769123]
8. Giebel GD, Gissel C. Accuracy of mHealth devices for atrial fibrillation screening: systematic review. JMIR Mhealth Uhealth 2019 Jun 16;7(6):e13641 [FREE Full text] [doi: 10.2196/13641] [Medline: 31199337]
9. Wood CS, Thomas MR, Budd J, Mashamba-Thompson TP, Herbst K, Pillay D, et al. Taking connected mobile-health diagnostics of infectious diseases to the field. Nature 2019 Feb;566(7745):467-474 [FREE Full text] [doi: 10.1038/s41586-019-0956-2] [Medline: 30814711]
10. Matis J, Svetlak M, Slezackova A, Svoboda M, Šumec R. Mindfulness-based programs for patients with cancer via ehealth and mobile health: systematic review and synthesis of quantitative research. J Med Internet Res 2020 Nov 16;22(11):e20709 [FREE Full text] [doi: 10.2196/20709] [Medline: 33196452]
11. Nussbaum R, Kelly C, Quinby E, Mac A, Parmanto B, Dicianno BE. Systematic review of mobile health applications in rehabilitation. Arch Phys Med Rehabil 2019 Jan;100(1):115-127. [doi: 10.1016/j.apmr.2018.07.439] [Medline: 30171827]
12. Aminuddin HB, Jiao N, Jiang Y, Hong J, Wang W. Effectiveness of smartphone-based self-management interventions on self-efficacy, self-care activities, health-related quality of life and clinical outcomes in patients with type 2 diabetes: a systematic review and meta-analysis. Int J Nurs Stud 2021 Apr;116:103286. [doi: 10.1016/j.ijnurstu.2019.02.003] [Medline: 30827741]
13. Li R, Liang N, Bu F, Hesketh T. The effectiveness of self-management of hypertension in adults using mobile health: systematic review and meta-analysis. JMIR Mhealth Uhealth 2020 Mar 27;8(3):e17776 [FREE Full text] [doi: 10.2196/17776] [Medline: 32217503]

14. Bienfait F, Petit M, Pardenaud R, Guineberteau C, Pignon A. Applying m-health to palliative care: a systematic review on the use of m-health in monitoring patients with chronic diseases and its transposition in palliative care. Am J Hosp Palliat Care 2020 Jul;37(7):549-564. [doi: 10.1177/1049909119885655] [Medline: 31773969]
15. Hernandez Silva E, Lawler S, Langbecker D. The effectiveness of mHealth for self-management in improving pain, psychological distress, fatigue, and sleep in cancer survivors: a systematic review. J Cancer Surviv 2019 Feb;13(1):97-107. [doi: 10.1007/s11764-018-0730-8] [Medline: 30635865]
16. Rowland SP, Fitzgerald JE, Holme T, Powell J, McGregor A. What is the clinical value of mHealth for patients? NPJ Digit Med 2020 Jan 13;3:4 [FREE Full text] [doi: 10.1038/s41746-019-0206-x] [Medline: 31970289]
17. Willcox JC, Dobson R, Whittaker R. Old-fashioned technology in the era of "bling": is there a future for text messaging in health care? J Med Internet Res 2019 Dec 20;21(12):e16630 [FREE Full text] [doi: 10.2196/16630] [Medline: 31859678]
18. Balaskas A, Schueller SM, Cox AL, Doherty G. Ecological momentary interventions for mental health: a scoping review. PLoS One 2021 Mar 11;16(3):e0248152 [FREE Full text] [doi: 10.1371/journal.pone.0248152] [Medline: 33705457]
19. Hall AK, Cole-Lewis H, Bernhardt JM. Mobile text messaging for health: a systematic review of reviews. Annu Rev Public Health 2015 Mar 18;36:393-415 [FREE Full text] [doi: 10.1146/annurev-publhealth-031914-122855] [Medline: 25785892]
20. Kitsiou S, Paré G, Jaana M, Gerber B. Effectiveness of mHealth interventions for patients with diabetes: an overview of systematic reviews. PLoS One 2017 Mar 01;12(3):e0173160 [FREE Full text] [doi: 10.1371/journal.pone.0173160] [Medline: 28249025]
21. Pink SL, Stagnaro MN, Chu J, Mernyk JS, Voelkel JG, Willer R. The effects of short messages encouraging prevention behaviors early in the COVID-19 pandemic. PLoS One 2023 Mar 14;18(4):e0284354 [FREE Full text] [doi: 10.1371/journal.pone.0284354] [Medline: 37058445]
22. Fritsch CG, Ferreira PH, Prior JL, McLachlan AJ, Ferreira ML. Effects of using text message interventions for the management of musculoskeletal pain: a systematic review. Pain 2020 Nov;161(11):2462-2475. [doi: 10.1097/j.pain.0000000000001958] [Medline: 32569088]
23. Hewitt S, Sephton R, Yeowell G. The effectiveness of digital health interventions in the management of musculoskeletal conditions: systematic literature review. J Med Internet Res 2020 Jun 05;22(6):e15617 [FREE Full text] [doi: 10.2196/15617] [Medline: 32501277]
24. Chhabra HS, Sharma S, Verma S. Smartphone app in self-management of chronic low back pain: a randomized controlled trial. Eur Spine J 2018 Nov;27(11):2862-2874. [doi: 10.1007/s00586-018-5788-5] [Medline: 30324496]
25. Irvine AB, Russell H, Manocchia M, Mino DE, Cox Glassen T, Morgan R, et al. Mobile-web app to self-manage low back pain: randomized controlled trial. J Med Internet Res 2015 Jan 02;17(1):e1 [FREE Full text] [doi: 10.2196/jmir.3130] [Medline: 25565416]
26. Mecklenburg G, Smittenaar P, Erhart-Hledik JC, Perez DA, Hunter S. Effects of a 12-week digital care program for chronic knee pain on pain, mobility, and surgery risk: randomized controlled trial. J Med Internet Res 2018 Apr 25;20(4):e156 [FREE Full text] [doi: 10.2196/jmir.9667] [Medline: 29695370]
27. Shebib R, Bailey JF, Smittenaar P, Perez DA, Mecklenburg G, Hunter S. Randomized controlled trial of a 12-week digital care program in improving low back pain. NPJ Digit Med 2019 Jan 07;2:1 [FREE Full text] [doi: 10.1038/s41746-018-0076-7] [Medline: 31304351]
28. Toelle TR, Utpadel-Fischler DA, Haas KK, Priebe JA. App-based multidisciplinary back pain treatment versus combined physiotherapy plus online education: a randomized controlled trial. NPJ Digit Med 2019 May 03;2:34 [FREE Full text] [doi: 10.1038/s41746-019-0109-x] [Medline: 31304380]
29. Brix LD, Bjørnholdt KT, Thillemann TM, Nikolajsen L. The effect of text messaging on medication adherence after outpatient knee arthroscopy: a randomized controlled trial. J Perianesth Nurs 2019 Aug;34(4):710-716. [doi: 10.1016/j.jopan.2018.11.011] [Medline: 30852173]
30. Chen HC, Chuang TY, Lin PC, Lin YK, Chuang YH. Effects of messages delivered by mobile phone on increasing compliance with shoulder exercises among patients with a frozen shoulder. J Nurs Scholarsh 2017 Jul;49(4):429-437. [doi: 10.1111/jnu.12308] [Medline: 28632975]
31. Kuusalo L, Sokka-Isler T, Kautiainen H, Ekman P, Kauppi MJ, Pirilä L, et al. Automated text message-enhanced monitoring versus routine monitoring in early rheumatoid arthritis: a randomized trial. Arthritis Care Res (Hoboken) 2020 Mar;72(3):319-325 [FREE Full text] [doi: 10.1002/acr.23846] [Medline: 30740935]
32. Mary A, Boursier A, Desailly Henry I, Grados F, Séjourné A, Salomon S, et al. Mobile phone text messages and effect on treatment adherence in patients taking methotrexate for rheumatoid arthritis: a randomized pilot study. Arthritis Care Res (Hoboken) 2019 Oct;71(10):1344-1352. [doi: 10.1002/acr.23750] [Medline: 30192070]
33. Smith OJ, Stewart CJ, Rastogi N, Abdaal A, Hachach-Haram N, Kerstein R, et al. Evaluation of the effect of an innovative automated text messaging service on patient experience in day-case hand trauma surgery. J Plast Reconstr Aesthet Surg 2018 Mar;71(3):450-451. [doi: 10.1016/j.bjps.2017.09.018] [Medline: 29066059]
34. Campbell KJ, Louie PK, Bohl DD, Edmiston T, Mikhail C, Li J, et al. A novel, automated text-messaging system is effective in patients undergoing total joint arthroplasty. J Bone Joint Surg Am 2019 Jan 16;101(2):145-151. [doi: 10.2106/JBJS.17.01505] [Medline: 30653044]

35. Kristjánsdóttir Ó, Fors EA, Eide E, Finset A, Stensrud TL, van Dulmen S, et al. A smartphone-based intervention with diaries and therapist feedback to reduce catastrophizing and increase functioning in women with chronic widespread pain. Part 2: 11-month follow-up results of a randomized trial. J Med Internet Res 2013 Mar 28;15(3):e72 [FREE Full text] [doi: 10.2196/jmir.2442] [Medline: 23538392]
36. Lambert TE, Harvey LA, Avdalis C, Chen LW, Jeyalingam S, Pratt CA, et al. An app with remote support achieves better adherence to home exercise programs than paper handouts in people with musculoskeletal conditions: a randomised trial. J Physiother 2017 Jul;63(3):161-167 [FREE Full text] [doi: 10.1016/j.jphys.2017.05.015] [Medline: 28662834]
37. Thomsen T, Aadahl M, Beyer N, Hetland ML, Løppenthin K, Midtgaard J, et al. The efficacy of motivational counselling and SMS reminders on daily sitting time in patients with rheumatoid arthritis: a randomised controlled trial. Ann Rheum Dis 2017 Sep;76(9):1603-1606 [FREE Full text] [doi: 10.1136/annrheumdis-2016-210953] [Medline: 28584189]
38. Wang Y, Lombard C, Hussain SM, Harrison C, Kozica S, Brady SR, et al. Effect of a low-intensity, self-management lifestyle intervention on knee pain in community-based young to middle-aged rural women: a cluster randomised controlled trial. Arthritis Res Ther 2018 Apr 17;20(1):74 [FREE Full text] [doi: 10.1186/s13075-018-1572-5] [Medline: 29665829]
39. Park KH, Song MR. The effects of postdischarge telephone counseling and short message service on the knee function, activities of daily living, and life satisfaction of patients undergoing total knee replacement. Orthop Nurs 2017;36(3):229-236 [FREE Full text] [doi: 10.1097/NOR.0000000000000332] [Medline: 28363197]
40. Scotti Requena S, Sterling M, Elphinston RA, Ritchie C, Robins S, R Armfield N. Development and use of mobile messaging for individuals with musculoskeletal pain conditions: a scoping review protocol. BMJ Open 2021 Jul 12;11(7):e048964 [FREE Full text] [doi: 10.1136/bmjopen-2021-048964] [Medline: 34253673]
41. Peters M, Godfrey CM, McInerney P, Munn Z, Tricco A, Khalil H. Chapter 11: scoping reviews. In: Aromataris E, Munn Z, editors. JBI Manual for Evidence Synthesis. Adelaide, Australia: Joanna Briggs Institute; 2020.
42. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
43. Covidence systematic review software. Veritas Health Innovation. URL: https://www.covidence.org/ [accessed 2024-06-30]
44. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Int J Surg 2010;8(5):336-341 [FREE Full text] [doi: 10.1016/j.ijsu.2010.02.007] [Medline: 20171303]
45. Human development insights. United Nations Development Programme. URL: https://hdr.undp.org/data-center/country-insights#/ranks [accessed 2024-06-30]
46. Seppen BF, den Boer P, Wiegel J, Ter Wee MM, van der Leeden M, de Vries R, et al. Asynchronous mHealth interventions in rheumatoid arthritis: systematic scoping review. JMIR Mhealth Uhealth 2020 Nov 05;8(11):e19260 [FREE Full text] [doi: 10.2196/19260] [Medline: 33151161]
47. Kristjánsdóttir OB, Fors EA, Eide E, Finset A, Stensrud TL, van Dulmen S, et al. A smartphone-based intervention with diaries and therapist-feedback to reduce catastrophizing and increase functioning in women with chronic widespread pain: randomized controlled trial. J Med Internet Res 2013 Jan 07;15(1):e5 [FREE Full text] [doi: 10.2196/jmir.2249] [Medline: 23291270]
48. Thomsen T, Aadahl M, Beyer N, Hetland ML, Løppenthin K, Midtgaard J, et al. Motivational counselling and SMS-reminders for reduction of daily sitting time in patients with rheumatoid arthritis: a descriptive randomised controlled feasibility study. BMC Musculoskelet Disord 2016 Oct 18;17(1):434 [FREE Full text] [doi: 10.1186/s12891-016-1266-6] [Medline: 27756265]
49. Theiler R, Freystaetter G, Platz A, Meier C, Mellinghoff HU. Capture the fracture by SMS. Osteologie 2017 Dec 23;25(04):283-286. [doi: 10.1055/s-0037-1619024]
50. Molinari G, García-Palacios A, Enrique Á, Roca P, Fernández-Llanio Comella N, Botella C. The power of visualization: back to the future for pain management in fibromyalgia syndrome. Pain Med 2018 Jul 01;19(7):1451-1468. [doi: 10.1093/pm/pnx298] [Medline: 29294081]
51. Nordgren B, Fridén C, Demmelmaier I, Bergström G, Lundberg IE, Nessen T, et al. An outsourced health-enhancing physical activity program for people with rheumatoid arthritis: study of the maintenance phase. J Rheumatol 2018 Aug;45(8):1093-1100. [doi: 10.3899/jrheum.171002] [Medline: 29717033]
52. Timmers T, Janssen L, Pronk Y, van der Zwaard BC, Koëter S, van Oostveen D, et al. Assessing the efficacy of an educational smartphone or tablet app with subdivided and interactive content to increase patients' medical knowledge: randomized controlled trial. JMIR Mhealth Uhealth 2018 Dec 21;6(12):e10742 [FREE Full text] [doi: 10.2196/10742] [Medline: 30578185]
53. Bartholdy C, Bliddal H, Henriksen M. Effectiveness of text messages for decreasing inactive behaviour in patients with knee osteoarthritis: a pilot randomised controlled study. Pilot Feasibility Stud 2019 Sep 7;5:112 [FREE Full text] [doi: 10.1186/s40814-019-0494-6] [Medline: 31516729]
54. Ji X, Wang Y, Ma Y, Hu Z, Man S, Zhang Y, et al. Improvement of disease management and cost effectiveness in Chinese patients with ankylosing spondylitis using a smart-phone management system: a prospective cohort study. Biomed Res Int 2019 Feb 25;2019:2171475 [FREE Full text] [doi: 10.1155/2019/2171475] [Medline: 30931322]

55. Støme LN, Pripp AH, Kværner JS, Kvaerner KJ. Acceptability, usability and utility of a personalised application in promoting behavioural change in patients with osteoarthritis: a feasibility study in Norway. BMJ Open 2019 Jan 28;9(1):e021608 [FREE Full text] [doi: 10.1136/bmjopen-2018-021608] [Medline: 30696666]
56. Thomsen T, Aadahl M, Beyer N, Hetland ML, Løppenthin KB, Midtgaard J, et al. Sustained long-term efficacy of motivational counseling and text message reminders on daily sitting time in patients with rheumatoid arthritis: long-term follow-up of a randomized, parallel-group trial. Arthritis Care Res (Hoboken) 2020 Nov;72(11):1560-1570. [doi: 10.1002/acr.24060] [Medline: 31507095]
57. Zaslavsky O, Thompson HJ, McCurry SM, Landis CA, Kitsiou S, Ward TM, et al. Use of a wearable technology and motivational interviews to improve sleep in older adults with osteoarthritis and sleep disturbance: a pilot study. Res Gerontol Nurs 2019 Jul 01;12(4):167-173 [FREE Full text] [doi: 10.3928/19404921-20190319-02] [Medline: 30901479]
58. Nelligan RK, Hinman RS, Kasza J, Crofts SJ, Bennell KL. Effects of a self-directed web-based strengthening exercise and physical activity program supported by automated text messages for people with knee osteoarthritis: a randomized clinical trial. JAMA Intern Med 2021 Jun 01;181(6):776-785 [FREE Full text] [doi: 10.1001/jamainternmed.2021.0991] [Medline: 33843948]
59. Pelle T, Bevers K, van der Palen J, van den Hoogen FH, van den Ende CH. Effect of the dr. Bart application on healthcare use and clinical outcomes in people with osteoarthritis of the knee and/or hip in the Netherlands; a randomized controlled trial. Osteoarthritis Cartilage 2020 Apr;28(4):418-427 [FREE Full text] [doi: 10.1016/j.joca.2020.02.831] [Medline: 32119972]
60. Newell D, MChiro RB. Increasing compliance toward home exercise in chiropractic patients using SMS texting: a pilot study. Clin Chiropr 2012 Dec;15(3-4):107-111. [doi: 10.1016/j.clch.2012.10.042]
61. Taylor NF, Bottrell J, Lawler K, Benjamin D. Mobile telephone short message service reminders can reduce nonattendance in physical therapy outpatient clinics: a randomized controlled trial. Arch Phys Med Rehabil 2012 Jan;93(1):21-26. [doi: 10.1016/j.apmr.2011.08.007] [Medline: 22000821]
62. Gandy M, Fogliati VJ, Terides MD, Johnston L, Nicholson Perry K, Newall C, et al. Short message service prompts for skills practice in internet-delivered cognitive behaviour therapy for chronic pain - are they feasible and effective? Eur J Pain 2016 Sep;20(8):1288-1298. [doi: 10.1002/ejp.853] [Medline: 27072189]
63. Jamison RN, Jurcik DC, Edwards RR, Huang CC, Ross EL. A pilot comparison of a smartphone app with or without 2-way messaging among chronic pain patients: who benefits from a pain app? Clin J Pain 2017 Aug;33(8):676-686 [FREE Full text] [doi: 10.1097/AJP.0000000000000455] [Medline: 27898460]
64. Lo WL, Lei D, Li L, Huang DF, Tong KF. The perceived benefits of an artificial intelligence-embedded mobile app implementing evidence-based guidelines for the self-management of chronic neck and back pain: observational study. JMIR Mhealth Uhealth 2018 Nov 26;6(11):e198 [FREE Full text] [doi: 10.2196/mhealth.8127] [Medline: 30478019]
65. Frei A, Dalla Lana K, Radtke T, Stone E, Knöpfli N, Puhan MA. A novel approach to increase physical activity in older adults in the community using citizen science: a mixed-methods study. Int J Public Health 2019 Jun;64(5):669-678. [doi: 10.1007/s00038-019-01230-3] [Medline: 30937463]
66. Anan T, Kajiki S, Oka H, Fujii T, Kawamata K, Mori K, et al. Effects of an artificial intelligence-assisted health program on workers with neck/shoulder pain/stiffness and low back pain: randomized controlled trial. JMIR Mhealth Uhealth 2021 Sep 24;9(9):e27535 [FREE Full text] [doi: 10.2196/27535] [Medline: 34559054]
67. Bailey JF, Agarwal V, Zheng P, Smuck M, Fredericson M, Kennedy DJ, et al. Digital care for chronic musculoskeletal pain: 10,000 participant longitudinal cohort study. J Med Internet Res 2020 May 11;22(5):e18250 [FREE Full text] [doi: 10.2196/18250] [Medline: 32208358]
68. Dekker-van Weering MG, Vollenbroek-Hutten MM, Hermens HJ. A pilot study - the potential value of an activity-based feedback system for treatment of individuals with chronic lower back pain. Disabil Rehabil 2015;37(24):2250-2256. [doi: 10.3109/09638288.2015.1019009] [Medline: 25738912]
69. Rabbi M, Aung MS, Gay G, Reid MC, Choudhury T. Feasibility and acceptability of mobile phone-based auto-personalized physical activity recommendations for chronic pain self-management: pilot study on adults. J Med Internet Res 2018 Oct 26;20(10):e10147 [FREE Full text] [doi: 10.2196/10147] [Medline: 30368433]
70. Selter A, Tsangouri C, Ali SB, Freed D, Vatchinsky A, Kizer J, et al. An mHealth app for self-management of chronic lower back pain (Limbr): pilot study. JMIR Mhealth Uhealth 2018 Sep 17;6(9):e179 [FREE Full text] [doi: 10.2196/mhealth.8256] [Medline: 30224333]
71. Hasenöhrl T, Windschnurer T, Dorotka R, Ambrozy C, Crevenna R. Prescription of individual therapeutic exercises via smartphone app for patients suffering from non-specific back pain : a qualitative feasibility and quantitative pilot study. Wien Klin Wochenschr 2020 Mar;132(5-6):115-123 [FREE Full text] [doi: 10.1007/s00508-020-01616-x] [Medline: 32060724]
72. Almhdawi KA, Obeidat DS, Kanaan SF, Oteir AO, Mansour ZM, Alrabbaei H. Efficacy of an innovative smartphone application for office workers with chronic non-specific low back pain: a pilot randomized controlled trial. Clin Rehabil 2020 Oct;34(10):1282-1291. [doi: 10.1177/0269215520937757] [Medline: 32602362]

73. Nordstoga AL, Bach K, Sani S, Wiratunga N, Mork PJ, Villumsen M, et al. Usability and acceptability of an app (SELFBACK) to support self-management of low back pain: mixed methods study. JMIR Rehabil Assist Technol 2020 Sep 09;7(2):e18729 [FREE Full text] [doi: 10.2196/18729] [Medline: 32902393]
74. Lee J, Lee M, Lim T, Kim T, Kim S, Suh D, et al. Effectiveness of an application-based neck exercise as a pain management tool for office workers with chronic neck pain and functional disability: a pilot randomized trial. Eur J Integr Med 2017 Jun;12:87-92. [doi: 10.1016/j.eujim.2017.04.012]
75. Demmelmaier I, Lindkvist Å, Nordgren B, Opava CH. "A gift from heaven" or "this was not for me". A mixed methods approach to describe experiences of participation in an outsourced physical activity program for persons with rheumatoid arthritis. Clin Rheumatol 2015 Mar;34(3):429-439. [doi: 10.1007/s10067-014-2723-y] [Medline: 24941928]
76. Ravn Jakobsen P, Hermann AP, Søndergaard J, Wiil UK, Clemensen J. Development of an mHealth application for women newly diagnosed with osteoporosis without preceding fractures: a participatory design approach. Int J Environ Res Public Health 2018 Mar 13;15(2):330 [FREE Full text] [doi: 10.3390/ijerph15020330] [Medline: 29438343]
77. Geuens J, Geurts L, Swinnen TW, Westhovens R, Vanden Abeele V. Mobile health features supporting self-management behavior in patients with chronic arthritis: mixed-methods approach on patient preferences. JMIR Mhealth Uhealth 2019 Mar 25;7(3):e12535 [FREE Full text] [doi: 10.2196/12535] [Medline: 30907737]
78. Nelligan RK, Hinman RS, Teo PL, Bennell KL. Exploring attitudes and experiences of people with knee osteoarthritis toward a self-directed eHealth intervention to support exercise: qualitative study. JMIR Rehabil Assist Technol 2020 Nov 26;7(2):e18860 [FREE Full text] [doi: 10.2196/18860] [Medline: 33242021]
79. Nelligan RK, Hinman RS, Atkins L, Bennell KL. A short message service intervention to support adherence to home-based strengthening exercise for people with knee osteoarthritis: intervention design applying the behavior change wheel. JMIR Mhealth Uhealth 2019 Oct 18;7(10):e14619 [FREE Full text] [doi: 10.2196/14619] [Medline: 31628786]
80. Pelle T, Bevers K, van der Palen J, van den Hoogen FH, van den Ende CH. Development and evaluation of a tailored e-self-management intervention (dr. Bart app) for knee and/or hip osteoarthritis: study protocol. BMC Musculoskelet Disord 2019 Aug 31;20(1):398 [FREE Full text] [doi: 10.1186/s12891-019-2768-9] [Medline: 31472687]
81. Johnson SS, Levesque DA, Broderick LE, Bailey DG, Kerns RD. Pain self-management for veterans: development and pilot test of a stage-based mobile-optimized intervention. JMIR Med Inform 2017 Oct 17;5(4):e40 [FREE Full text] [doi: 10.2196/medinform.7117] [Medline: 29042341]
82. Mork PJ, Bach K, selfBACK Consortium. A decision support system to enhance self-management of low back pain: protocol for the selfBACK project. JMIR Res Protoc 2018 Jul 20;7(7):e167 [FREE Full text] [doi: 10.2196/resprot.9379] [Medline: 30030208]
83. Fritsch CG, Ferreira PH, Prior JL, Vesentini G, Schlotfeldt P, Eyles J, et al. TEXT4myBACK - the development process of a self-management intervention delivered via text message for low back pain. Arch Rehabil Res Clin Transl 2021 Apr 27;3(2):100128 [FREE Full text] [doi: 10.1016/j.arrct.2021.100128] [Medline: 34179764]
84. Prochaska JO, Velicer WF. The transtheoretical model of health behavior change. Am J Health Promot 1997;12(1):38-48. [doi: 10.4278/0890-1171-12.1.38] [Medline: 10170434]
85. Bartholomew LK, Parcel GS, Kok G. Intervention mapping: a process for developing theory- and evidence-based health education programs. Health Educ Behav 1998 Oct;25(5):545-563. [doi: 10.1177/109019819802500502] [Medline: 9768376]
86. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011 Apr 23;6:42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
87. McEvoy R, Ballini L, Maltoni S, O'Donnell CA, Mair FS, Macfarlane A. A qualitative systematic review of studies using the normalization process theory to research implementation processes. Implement Sci 2014 Jan 02;9:2 [FREE Full text] [doi: 10.1186/1748-5908-9-2] [Medline: 24383661]
88. Simonsen J, Robertson T. Routledge International: Handbook of Participatory Design. Milton Park, UK: Routledge; 2012.
89. Buchholz SW, Wilbur J, Ingram D, Fogg L. Physical activity text messaging interventions in adults: a systematic review. Worldviews Evid Based Nurs 2013 Aug;10(3):163-173. [doi: 10.1111/wvn.12002] [Medline: 23746267]
90. Chow CK, Redfern J, Thiagalingam A, Jan S, Whittaker R, Hackett M, et al. Design and rationale of the tobacco, exercise and diet messages (TEXT ME) trial of a text message-based intervention for ongoing prevention of cardiovascular disease in people with coronary disease: a randomised controlled trial protocol. BMJ Open 2012 Jan 19;2(1):e000606 [FREE Full text] [doi: 10.1136/bmjopen-2011-000606] [Medline: 22267690]
91. de Jongh T, Gurol-Urganci I, Vodopivec-Jamsek V, Car J, Atun R. Mobile phone messaging for facilitating self-management of long-term illnesses. Cochrane Database Syst Rev 2012 Dec 12;12(12):CD007459 [FREE Full text] [doi: 10.1002/14651858.CD007459.pub2] [Medline: 23235644]
92. Head KJ, Noar SM, Iannarino NT, Grant Harrington N. Efficacy of text messaging-based interventions for health promotion: a meta-analysis. Soc Sci Med 2013 Nov;97:41-48. [doi: 10.1016/j.socscimed.2013.08.003] [Medline: 24161087]
93. Iribarren SJ, Brown W3, Giguere R, Stone P, Schnall R, Staggers N, et al. Scoping review and evaluation of SMS/text messaging platforms for mHealth projects or clinical interventions. Int J Med Inform 2017 May;101:28-40 [FREE Full text] [doi: 10.1016/j.ijmedinf.2017.01.017] [Medline: 28347445]

94. Kannisto KA, Koivunen MH, Välimäki MA. Use of mobile phone text message reminders in health care services: a narrative literature review. J Med Internet Res 2014 Oct 17;16(10):e222 [FREE Full text] [doi: 10.2196/jmir.3442] [Medline: 25326646]
95. Dobson F, Bennell KL, French SD, Nicolson PJ, Klaasman RN, Holden MA, et al. Barriers and facilitators to exercise participation in people with hip and/or knee osteoarthritis: synthesis of the literature using behavior change theory. Am J Phys Med Rehabil 2016 May;95(5):372-389. [doi: 10.1097/PHM.0000000000000448] [Medline: 26945211]
96. Cane J, O'Connor D, Michie S. Validation of the theoretical domains framework for use in behaviour change and implementation research. Implement Sci 2012 Apr 24;7:37 [FREE Full text] [doi: 10.1186/1748-5908-7-37] [Medline: 22530986]
97. Michie S, Atkins L, West R. The Behaviour Change Wheel: A Guide to Designing Interventions. Sutton, Surrey: Silverback Publishing; 2014.
98. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and evaluating digital interventions to promote behavior change in health and health care: recommendations resulting from an international workshop. J Med Internet Res 2017 Jun 29;19(6):e232 [FREE Full text] [doi: 10.2196/jmir.7126] [Medline: 28663162]
99. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95 [FREE Full text] [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
100. Abroms LC, Whittaker R, Free C, Mendel Van Alstyne J, Schindler-Ruwisch JM. Developing and pretesting a text messaging program for health behavior change: recommended steps. JMIR Mhealth Uhealth 2015 Dec 21;3(4):e107 [FREE Full text] [doi: 10.2196/mhealth.4917] [Medline: 26690917]
101. Oinas-Kukkonen H, Harjumaa M. Persuasive systems design: key issues, process model and system features. In: Routledge Handbook of Policy Design. Milton Park, UK: Routledge; 2018.
102. Fogg BJ. A behavior model for persuasive design. In: Proceedings of the 4th International Conference on Persuasive Technology. 2009 Presented at: Persuasive '09; April 26-29, 2009; Claremont,CA. [doi: 10.1145/1541948.1541999]
103. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JE. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012 Nov 14;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
104. Abraham C, Michie S. A taxonomy of behavior change techniques used in interventions. Health Psychol 2008 May;27(3):379-387. [doi: 10.1037/0278-6133.27.3.379] [Medline: 18624603]
105. Redfern J, Santo K, Coorey G, Thakkar J, Hackett M, Thiagalingam A, et al. Factors influencing engagement, perceived usefulness and behavioral mechanisms associated with a text message support program. PLoS One 2016 Oct 14;11(10):e0163929 [FREE Full text] [doi: 10.1371/journal.pone.0163929] [Medline: 27741244]
106. Redfern J, Thiagalingam A, Jan S, Whittaker R, Hackett ML, Mooney J, et al. Development of a set of mobile phone text messages designed for prevention of recurrent cardiovascular events. Eur J Prev Cardiol 2014 Apr;21(4):492-499. [doi: 10.1177/2047487312449416] [Medline: 22605787]
107. Chow CK, Redfern J, Hillis GS, Thakkar J, Santo K, Hackett ML, et al. Effect of lifestyle-focused text messaging on risk factor modification in patients with coronary heart disease: a randomized clinical trial. JAMA 2015;314(12):1255-1263. [doi: 10.1001/jama.2015.10945] [Medline: 26393848]
108. Gentili A, Failla G, Melnyk A, Puleo V, Tanna GL, Ricciardi W, et al. The cost-effectiveness of digital health interventions: a systematic review of the literature. Front Public Health 2022 Aug 11;10:787135 [FREE Full text] [doi: 10.3389/fpubh.2022.787135] [Medline: 36033812]
109. MacPherson M, Cranston K, Johnston C, Locke S, Jung ME. Evaluation and refinement of a bank of SMS text messages to promote behavior change adherence following a diabetes prevention program: survey study. JMIR Form Res 2021 Aug 27;5(8):e28163 [FREE Full text] [doi: 10.2196/28163] [Medline: 34448713]
110. Sherman KA, Pehlivan MJ, Singleton A, Hawkey A, Redfern J, Armour M, et al. Co-design and development of EndoSMS, a supportive text message intervention for individuals living with endometriosis: mixed methods study. JMIR Form Res 2022 Dec 09;6(12):e40837 [FREE Full text] [doi: 10.2196/40837] [Medline: 36485029]
111. Singleton A, Raeside R, Partridge SR, Hayes M, Maka K, Hyun KK, et al. Co-designing a lifestyle-focused text message intervention for women after breast cancer treatment: mixed methods study. J Med Internet Res 2021 Jun 14;23(6):e27076 [FREE Full text] [doi: 10.2196/27076] [Medline: 34125072]

Abbreviations

CBT: cognitive behavioral therapy
HDI: Human Development Index
LBP: low back pain
mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
RA: rheumatoid arthritis
RCT: randomized controlled trial
RQ: research question

Edited by L Buis; submitted 18.12.23; peer-reviewed by A Gholamrezaei, D Geagea; comments to author 22.04.24; revised version received 29.04.24; accepted 12.06.24; published 14.08.24.

Please cite as:
Armfield N, Elphinston R, Liimatainen J, Scotti Requena S, Eather CE, Edirippulige S, Ritchie C, Robins S, Sterling M
Development and Use of Mobile Messaging for Individuals With Musculoskeletal Pain Conditions: Scoping Review
JMIR Mhealth Uhealth 2024;12:e55625
URL: https://mhealth.jmir.org/2024/1/e55625
doi:10.2196/55625
PMID:

Review

Current Status of Barriers to mHealth Access Among Patients With Stroke and Steps Toward the Digital Health Era: Systematic Review

Atsadaporn Niyomyart[1], RN, LLB, MSN, PhD; Suebsarn Ruksakulpiwat[2], RN, MMed, PhD; Chitchanok Benjasirisan[2], RN, MS; Lalipat Phianhasin[2], RN, MS, AGPCNP-BC; Kabtamu Nigussie[3], BSc, MSc; Sutthinee Thorngthip[4], RN, MS; Gazi Shamita[5], BS, MS; Jai Thampakkul[6], BS; Lidya Begashaw[7], BS

[1]Ramathibodi School of Nursing, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok, Thailand

[2]Department of Medical Nursing, Faculty of Nursing, Mahidol University, Bangkok, Thailand

[3]Department of Psychiatry, College of Health and Medical Sciences, Haramaya University, Harar, Ethiopia

[4]Department of Nursing Siriraj Hospital, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand

[5]Department of Dermatology, School of Medicine, Case Western Reserve University, Cleveland, OH, United States

[6]Case School of Engineering, Case Western Reserve University, Cleveland, OH, United States

[7]Jack, Joseph and Morton Mandel School of Applied Social Sciences, Case Western Reserve University, Cleveland, OH, United States

Corresponding Author:
Suebsarn Ruksakulpiwat, RN, MMed, PhD
Department of Medical Nursing, Faculty of Nursing, Mahidol University
2 Wang Lang Road, Siriraj, Bangkok Noi
Bangkok, 10700
Thailand
Phone: 66 984782692
Email: suebsarn25@gmail.com

Abstract

Background: Mobile health (mHealth) offers significant benefits for patients with stroke, facilitating remote monitoring and personalized health care solutions beyond traditional settings. However, there is a dearth of comprehensive data, particularly qualitative insights, on the barriers to mHealth access. Understanding these barriers is crucial for devising strategies to enhance mHealth use among patients with stroke.

Objective: This study aims to examine the recent literature focusing on barriers to mHealth access among patients with stroke.

Methods: A systematic search of PubMed, MEDLINE, Web of Science, and CINAHL Plus Full Text was conducted for literature published between 2017 and 2023. Abstracts and full texts were independently screened based on predetermined inclusion and exclusion criteria. Data synthesis was performed using the convergent integrated analysis framework recommended by the Joanna Briggs Institute.

Results: A total of 12 studies met the inclusion criteria. The majority were qualitative studies (about 42%), followed by mixed methods (25%), pilot studies (about 17%), nonrandomized controlled trials (about 8%), and observational studies (about 8%). Participants included patients with stroke, caregivers, and various health care professionals. The most common mHealth practices were home-based telerehabilitation (30%) and poststroke mHealth and telecare services (20%). Identified barriers were categorized into two primary themes: (1) at the patient level and (2) at the health provider-patient-device interaction level. The first theme includes 2 subthemes: health-related issues and patient acceptability. The second theme encompassed 3 subthemes: infrastructure challenges (including software, networking, and hardware), support system deficiencies, and time constraints.

Conclusions: This systematic review underscores significant barriers to mHealth adoption among patients with stroke. Addressing these barriers in future research is imperative to ensure that mHealth solutions effectively meet patients' needs.

(JMIR Mhealth Uhealth 2024;12:e54511) doi:10.2196/54511

KEYWORDS

digital health; mHealth; barrier; stroke; systematic review; mobile phones

Introduction

Background

Stroke, a leading cause of disability and mortality worldwide, necessitates immediate and ongoing interventions for optimal recovery [1]. While mobile health (mHealth) technologies offer promising solutions for chronic disease management, understanding and addressing the unique barriers faced by patients with stroke is crucial for ensuring their equitable access and optimal use [2,3]. While previous research has explored mHealth adoption in various populations, the cognitive impairments [4], rehabilitation needs [5], and potential technology literacy limitations of patients with stroke [6] require distinct consideration. mHealth, encompassing the use of mobile devices, applications, and wireless communication devices, offers promising avenues to deliver personalized health care solutions to patients with stroke beyond the confines of traditional health care settings [3,7].

The health care landscape has evolved with the proliferation of mobile devices and the availability of high-speed internet connectivity. Consequently, mHealth platforms have emerged as tools that can potentially bridge the gap between health care providers and patients, enabling continuous monitoring, real-time communication, and targeted interventions [8,9]. In the context of stroke, where timely interventions and ongoing support are crucial, mHealth can revolutionize poststroke care by providing patients with access to rehabilitation exercises, medication reminders, educational resources, and even telemedicine consultations [3,6,10]. Despite the promise of mHealth, realizing its benefits for patients with stroke is contingent upon understanding and mitigating the barriers that hinder its widespread adoption and use. Previous research has underscored the importance of identifying these barriers to ensure equitable access and optimal use of mHealth services among patients with stroke [11]. While existing literature has explored mHealth adoption in various populations, the unique challenges faced by patients with stroke deserve particular attention due to the nature of their condition, potential cognitive impairments, and the necessity of tailored interventions. The period between 2017 and 2023 has witnessed substantial advancements in mHealth technologies, health care policies, and the prevalence of mobile device usage among diverse demographic groups. Consequently, there is a need to assess the current status of mHealth access among patients recovering from stroke within this evolving landscape. By systematically reviewing the literature and synthesizing recent findings, this study aims to delineate the barriers that impede access to mHealth services among patients with stroke, analyze their impact, and propose actionable recommendations for stakeholders in the health care ecosystem.

In light of the growing importance of digital health and its potential benefits for patients with stroke, investigating the barriers to mHealth access is a critical step toward ensuring that these advancements are inclusive and patient centered. By identifying these barriers and proposing strategies to overcome them, this study seeks to contribute to the advancement of stroke care in the digital health era.

Objective

This study seeks to evaluate recent literature, focusing on the notable barriers to accessing mHealth services within the population of patients with stroke. Through this assessment, the study aims to offer recommendations aimed at effectively addressing these barriers and propelling progress in this critical domain of health care.

Methods

Identify Relevant Studies

We adhered to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [12] while conducting this systematic review. This approach guided the presentation of the flow diagram depicting the identification, screening, exclusion, and inclusion of literature. To identify studies published between 2017 and 2023 that reported barriers to mHealth access among the patients with stroke, we systematically searched 4 electronic databases: PubMed, MEDLINE, Web of Science, and CINAHL Plus Full Text, on September 1, 2023. Our search involved the combination of terms: *(Stroke* or Cerebrovascular Accident* or CVA* or Cerebrovascular Apoplexy or Brain Vascular Accident* or Cerebrovascular Stroke* or Apoplexy or Cerebral Stroke* or Acute Stroke* or Acute Cerebrovascular Accident* or Brain Infarction* or Brain Infarct* or Anterior Circulation Brain Infarction or Brain Venous Infarction* or Venous Brain Infarction* or Anterior Cerebral Circulation Infarction or Posterior Circulation Brain Infarction or Cerebral Infarction* or Cerebral Infarct* or Left Hemisphere, Infarction, Cerebral or Subcortical Infarction* or Posterior Choroidal Artery Infarction or Anterior Choroidal Artery Infarction or Hemorrhagic Stroke* or Subarachnoid Hemorrhagic Stroke* or Intracerebral Hemorrhagic Stroke* or Ischemic Stroke* or Ischaemic Strok* or Cryptogenic Ischemic Stroke* or Cryptogenic Stroke* or Cryptogenic Embolism Stroke* or Wake-up Stroke* or Acute Ischemic Stroke* or Embolic Stroke* or Cardioembolic Stroke* or Cardio-embolic Stroke* or Thrombotic Stroke* or Acute Thrombotic Stroke*) AND (Tele-Referral* or Virtual Medicine or Tele Intensive Care or Tele ICU or Mobile Health or mHealth or Telehealth or eHealth or Remote Consultation or Teleconsultation* or Telenursing or Telepathology or Teleradiology or Telerehabilitation* or Remote Rehabilitation* or Virtual Rehabilitation*) AND (Barrier*)* using Boolean operators. Additionally, we manually searched the reference lists of the included studies to ensure inclusivity. All identified references were cataloged using EndNote.

Study Selection

We conducted a 3-step selection process. First, we screened titles and abstracts to identify eligible studies. Then, we assessed the full text to determine relevance. Finally, inclusion criteria were applied to ensure that only studies aligned with our objectives were included. Conversely, exclusion criteria were used to eliminate literature not pertinent to the review (see Textbox 1).

Textbox 1. Inclusion and exclusion criteria.

Inclusion criteria:

- The study included patients with stroke aged 18 years or older. The study may also encompass other populations, such as caregivers and health care teams, but patients with stroke must be included
- Original studies using quantitative, qualitative, or mixed methods approaches
- Studies that identified barriers to accessing mobile health (mHealth) services among patients with stroke (all types of strokes are eligible)
- In this study, mHealth is identified as the practice of medicine and public health supported by mobile devices. It encompasses the use of mobile devices, such as smartphones, tablet computers, personal digital assistants, and others, for health services, information, and data collection [13]
- Studies conducted in various settings, including inpatient, outpatient, or home environments
- Studies published in the English language
- Studies published between January 2017 and December 2023

Exclusion criteria:

- Studies that did not involve patients were excluded.
- Exclusion of conference proceedings, abstracts, review articles, theoretical papers, protocols, dissertations, letters to the editor, opinion (viewpoint) pieces, statement papers, government documents, or working papers

Data Extraction

For this review, we used a standardized data extraction chart that included the following data points for each study: Reference, country, year, study design, total sample size, target population, participant age (years), main study aim, main findings, presence of mHealth in included studies, key barriers to accessing mHealth, and further research implications.

Data Synthesis

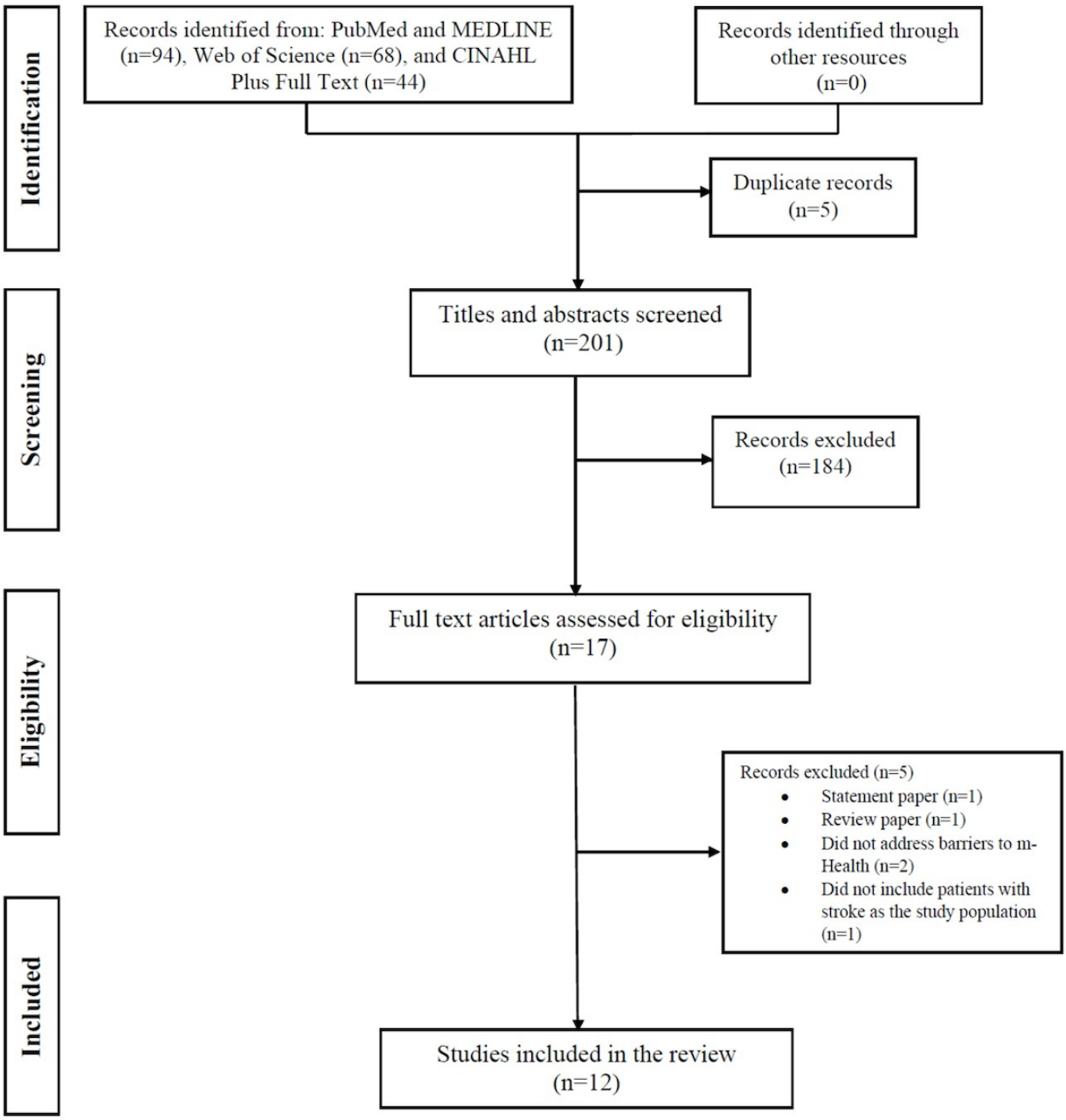
For the data synthesis of the included studies in this review, we used the convergent integrated analysis framework recommended by the Joanna Briggs Institute (JBI) for systematic reviews. In this process, themes will be derived from the key findings of the included studies by analyzing both commonalities and distinctions among the primary findings related to key barriers in accessing mHealth. Furthermore, subthemes will be extracted as necessary, aligning with the specific focus of the corresponding findings, akin to the methodology used by qualitative researchers [14].

Results

Search Results

Following the PRISMA guidelines [12], we initially identified a total of 206 articles. Out of these, 94 were obtained from PubMed and Medline, 68 from Web of Science, and 44 from CINAHL Plus Full Text. No additional articles were found from other sources. After a thorough review, we identified and removed 5 duplicate articles. Subsequently, the remaining articles underwent screening based on their titles and abstracts, following the inclusion and exclusion criteria (Textbox 1). At this stage, 184 articles did not meet the inclusion criteria and were excluded, leaving us with 17 articles eligible for full-text screening. During the full-text screening phase, 5 articles were excluded. Of these, 1 was identified as a statement paper, 1 was a review paper, 2 did not address barriers to mHealth, and 1 did not include patients with stroke as the study population. Consequently, 12 studies were included in the review (Figure 1).

Figure 1. Flowchart diagram displaying the selection method of qualified studies.



Data Summary

Multimedia Appendix 1 [15-26] provides a summary of each included study, including reference, country, year, study design, total sample size, target population, participant age (years), main study aim, main findings, presence of mHealth in included studies (optional), key barriers to accessing mHealth, and implications for further research.

Description of Included Studies

Table 1 shows that all included studies were published between 2017 and 2023, with the most publications in 2023 (n=5, 42%) and 2022 (n=3, 25%). Among these, one publication was found each in 2017, 2018, 2020, and 2021. Most of the included studies were conducted in the United States (n = 4, 33%). Other countries include the United Kingdom (n=2, 17%), Hong Kong (China) (n=2, 17%), Brazil (n=1, 8%), and the Netherlands (n=1, 8%). In terms of study design, qualitative studies were the most popular (n=5, 42%), followed by mixed methods (n=3, 25%), pilot studies (n=2, 17%), nonrandomized control trials (n=1, 8%), and observational studies (n=1, 8%). Among the patients with stroke included in the study, there was no specific stroke type identified (n=9, 64%), ischemic stroke (n=2, 14%), hemorrhagic stroke (n=2, 14%), and chronic hemiplegic stroke (n=1, 7%).

Table 1. The characteristics of the included studies.

Characteristics	Values, n[a] (%)
Publication year	
2023	5 (42)
2022	3 (25)
2021	1 (8)
2020	1 (8)
2018	1 (8)
2017	1 (8)
Country	
United States	4 (33)
United Kingdom	2 (17)
Hong Kong (China)	2 (17)
Brazil	1 (8)
The Netherlands	1 (8)
Singapore	1 (8)
Republic of Korea	1 (8)
Study design	
Qualitative study	5 (42)
Mixed methods	3 (25)
Pilot study	2 (17)
Nonrandomized controlled trial	1 (8)
Observational study	1 (8)
Target population	
Patients with stroke	
Not specify stroke type	9 (64)
Ischemic stroke	2 (14)
Hemorrhagic stroke	2 (14)
Chronic hemiplegic stroke	1 (7)
Health care professionals and other	
Caregiver	3 (14)
Clinician	1 (14)
Rehabilitation therapist	1 (14)
Occupational therapist	1 (14)
Nurse	1 (14)
Total sample size	
1-50	9 (75)
>50-100	1 (8)
>100-200	1 (8)
>200	1 (8)
mHealth[b] in included studies	
A home-based telerehabilitation [15-17]	3 (30)
A poststroke mHealth and telecare service [18,19]	2 (20)

Characteristics	Values, n[a] (%)
The remote physical exercise program [20]	1 (10)
A home-based web-based clinic [21]	1 (10)
Self-administered VR[c] telerehabilitation [22]	1 (10)
The Homecare Arm Rehabilitation System (MERLIN) [23]	1 (10)
Remote CCT[d] and metacognitive strategy (MST[e]) training [24]	1 (10)

[a]The number of included studies in which one study may include more than one characteristic (eg, target population, mobile health, etc).

[b]mHealth: mobile health.

[c]VR: virtual reality.

[d]CCT: computerized cognitive.

[e]MST: metacognitive strategy.

Additionally, caregivers (n=3, 43%), clinicians (n=1, 14%), rehabilitation therapists (n=1, 14%), occupational therapists (n=1, 14%), and nurses (n=1, 14%) were included in the study. Sample sizes ranged from 1 to 50 (n=9, 75%), over 50 to 100, over 100 to 200, and over 200 (all n=1, 8%). The most common mHealth practices in the included studies were home-based telerehabilitation (n=3, 30%) and poststroke mHealth and telecare service (n=2, 20%). The rest of mHealth consisted of remote physical exercise, a home-based web-based clinic, self-administered virtual reality (VR) telerehabilitation, and home care arm rehabilitation, as well as remote cognitive (CCT) and metacognitive strategy training (MST; all n=1, 10%).

Description of Barriers to mHealth Access in Patients With Stroke

Multimedia Appendix 1 [15-26] provides a summary of the barriers to mHealth accessibility experienced by patients with stroke. Having used the convergent integrated analysis framework recommended by JBI for systematic reviews [14], we can discern two primary themes: (1) at the patient-level and (2) at the health provider-patient-device interaction level. The first theme included 2 subthemes: health-related barriers and patient acceptability. The second theme encompassed 3 subthemes: infrastructure (inclusive of software, networking, and hardware), deficiencies in support systems, and constraints on available time (Table 2).

Table 2. Barriers to mobile health (mHealth) access.

Study	Barriers to mHealth Themes						
	The patient-level		The health provider-patient-device interaction level				
	Health-related barriers	Patient acceptability	Infrastructure			Deficiencies in support systems	Constraints on available time
			Software	Networking	Hardware		
Dodakian et al [15]	✓				✓		✓
Tyagi et al [26]			✓	✓	✓		
Dunne et al [25]		✓					✓
Torriani-Pasin et al [20]	✓					✓	✓
Lam et al [21]				✓			
Morse et al [22]		✓		✓	✓	✓	
Spits et al [23]		✓				✓	
Bhattacharjya et al [16]			✓			✓	✓
Chung et al [17]		✓					
Jaywant et al [24]		✓		✓	✓	✓	
Ramaswamy et al [18]			✓		✓		
Wong et al [19]			✓			✓	
Number of included studies, n (%)	2 (17)	5 (42)	4 (33)	4 (33)	5 (42)	6 (50)	4 (33)

The Patient Level

Health-Related Barriers

Our results indicate that health-related issues are among the barriers to mHealth access for patients with stroke. A study evaluating a home-based telerehabilitation system in patients with chronic hemiparetic stroke found that fatigue due to illness prevented patients from using the Home-Based Telerehabilitation Program [15]. Similarly, a clinical trial study from Brazil aimed at determining adherence and barriers to attending a remote physical exercise program for individuals after stroke shows that health-related factors such as a lack of motor skills, physical fitness, exercise-related pain, and other constraints are significant barriers preventing them from using the remote physical exercise program [20]. For instance, in a qualitative study conducted with survivors of stroke experiencing partial visual loss, caregivers and occupational therapists revealed that patients believed their lack of confidence and fear of using technology prevented them from learning new things, and their readiness to embrace technology was a barrier to telerehabilitation [25].

Patient Acceptability

Five studies in our review have highlighted patient acceptability as the primary barrier to mHealth [17,22-25]. For example, a mixed-method study identified facilitators and barriers to using self-administered VR telerehabilitation, suggesting that a lack of experience or confidence with technology hindered access to self-administered VR telerehabilitation [22]. Moreover, a pilot study found that patients with stroke, especially older adults with stroke, may have limited digital literacy skills, making it challenging to use mHealth tools effectively. Additionally, a study points out that participants' attitudes and acceptance of mHealth interventions may vary; some individuals may be hesitant or skeptical about using technology for health care purposes [24].

The Health Provider-Patient-Device Interaction Level

Infrastructure

Software

The complexity of mHealth software can pose challenges for patients in its usage [16,18,19,26]. For instance, a qualitative study conducted in Singapore explored the perceived barriers and facilitators of telerehabilitation by patients with stroke, caregivers, and rehabilitation therapists. The study noted that teletherapists encountered difficulties conducting comprehensive patient assessments remotely [26]. Similarly, patients perceived limitations in the variety and scope of rehabilitation exercises available through telerehabilitation, attributing this to issues related to the interface and design of the remote platform [26]. In line with this, another study aimed to explore potential mHealth apps to aid survivors of stroke with poststroke care and determine how demographic variables affect app preferences. This study found that the complexity of app usage acted as a barrier, preventing some users from using these apps effectively [18]. Considering apps on outdated smartphones or operating systems can result in challenges for patients when running and installing complex apps. In addition, there is a noticeable mismatch between the technical complexity of the apps and the users' capacities, as they lack the knowledge to address issues like nonfunctionality, showdowns, freezing, and crashes in the apps [27,28].

Network

In our review, 4 included studies highlighted the significance of network issues as a critical barrier to mHealth access [21,22,24,26]. Of these, 2 studies emphasized the importance of reliable internet access for effective remote interventions. Poor internet connectivity or limited access to high-speed internet in certain areas can impede the successful implementation of mHealth interventions [24,26]. Furthermore, the high cost and limited availability of internet access to support the equipment in survivors of stroke's homes can prevent them from accessing mHealth services [21,22]. Patients and health care professionals mainly rely on an internet-based system where health care providers must promptly update patient information. On the other hand, patients can use the apps to receive health updates and discuss their symptoms with health care providers effectively [29,30]. Thus, the absence of a digital option can restrict internet access and cause challenges, both at the patient's house and within the health care facility. However, adopting an offline mode can make it difficult to transfer data, which typically relies on web-based connectivity.

Hardware

Hardware malfunctions, such as those in mobile phones or computers, are considered significant factors affecting the delivery of mHealth services [15,18,22,24,26]. For instance, in a pilot study aiming to implement a home-based Telerehabilitation Program that included arm motor therapy games, therapeutic arm exercises, remote stroke education, and videoconferencing, hardware malfunctions limited patients' access to the intervention [15]. Moreover, a qualitative study highlighted equipment setup–related difficulties as a barrier to mHealth access, with patients with stroke encountering challenges in setting up and using the required equipment [26]. For example, getting ready for video-recorded exercise can be a challenge among patients with stroke when setting up the iPad, sensors, and monitoring equipment for heart and blood pressure [31]. In addition, specific details, such as the correct connection of the limb sensor node to assess the patient's range of motion and troubleshoot unexpected hardware problems, may not be straightforward, even though general instructions are provided. Another study focusing on remote interventions, which often rely on technology, noted that one potential barrier could be the accessibility of the necessary technology (eg, smartphones, computers, etc) for individuals with chronic stroke, especially those who may not be familiar with or have easy access to such devices [24].

Deficiencies in Support Systems

Lack of support systems is a significant barrier to mHealth access, as indicated by 6 included studies in our review [16,19,20,22-24]. For instance, a qualitative research study investigated the experiences of 13 survivors of stroke and health care providers regarding the use of a poststroke telecare service [19]. The study found that the lack of general guidelines for operating a telecare service, technical issues, and limited human

resources presented challenges to patients in using the service (eg, difficulties in joining Zoom meetings and troubleshooting) [19]. Another study also emphasized the importance of adequate training and support for participants using mHealth tools. This study recommended initial in-person sessions to familiarize participants with the technology, indicating that additional support might be necessary for some individuals [24]. In a mixed methods study aimed at identifying facilitators and barriers to the use of self-administered VR telerehabilitation, it was mentioned that a lack of an exercise companion and a safe environment for exercising prevented some individuals from using self-administered VR telerehabilitation [20].

Constraints on Available Time

In our study, we have found that constraints on available time are considered a barrier to mHealth, as mentioned in 4 published studies [15,16,20,25]. For example, 1 included study stated that conflicts with other medical appointments prevent patients with stroke from completing a home-based telerehabilitation program [15]. Additionally, work commitments are also a factor preventing patients with from using mHealth at home [20]. This issue is not limited to patients alone; it also affects health care providers. Qualitative research revealed that time-related issues are prevalent among providers, such as occupational therapists who reported having limited time to spend with survivors of stroke when providing mHealth interventions [25]. Additionally, a prior study highlighted the importance of the design and development stages of mHealth in reducing disruptions to health care providers' established workflows [32]. The study emphasized the transformation of functionalities into practical health tools, aiming to facilitate the integration of mHealth into their existing structured workflows, ultimately influencing increased engagement in the adoption of mHealth.

Discussion

The study aimed to explore recent literature to uncover possible barriers to mHealth access among patients with stroke. Although mHealth offers an important option for accessing needed health services, and people with stroke are currently using it for various purposes, this study identified several barriers inhibiting its use. One such barrier is health-related issues. Prior studies have shown that 25%-85% of survivors of stroke report experiencing fatigue, regardless of the severity of their condition [33,34]. As fatigue develops over time, it can result in the inability to perform basic activities, such as dressing and eating, and eventually limit more complex activities like shopping and preparing meals [35]. Considering that the study included complex telerehabilitation activities (eg, games, therapeutic programs, education, and videoconferencing), access to the program among patients with stroke could be limited due to fatigue. It is also important to consider a patient's age when examining fatigue, whether they are young or old [36-38]; this finding aligns with our study, which included patients aged 40 and older who had experienced strokes.

Despite the rise of mHealth to connect patients to health services outside of a clinical setting, our study found that patients' readiness and acceptance were barriers to accessing mHealth. Similar to previous studies, patients with stroke reported that visual and physical impairments hindered their access to mHealth, as they lacked the confidence to communicate electronically and were concerned about making mistakes during remote interactions [25,39]. These impairments, affecting sight and physical activity, may contribute to reduced readiness for mHealth. Additionally, digital health literacy should be considered, as patients with stroke, particularly those in the older population, may have difficulty obtaining, interpreting, and evaluating health information through digital sources [40]. A similar finding was observed in another study, where older patients with stroke had less access to phone or video telephone visits due to a lack of digital literacy and access, as well as a lack of experience using such technology [41]. In terms of acceptability, patients with stroke who perceived m-Health as useless (e.g., only making phone calls) restricted their access to mHealth applications, such as mobile home-based exercise programs [42].

One of the main obstacles for mHealth is the availability of adequate infrastructure to provide effective and comprehensive care to patients with stroke. The review identified the limitations of the current software in conducting patient assessments and delivering a wide range of rehabilitation exercises. Since the main goal of rehabilitation is to improve the patient's physical impairments, a health care provider needs to perform physical exams to determine and evaluate the intervention. Without such contact through mHealth, the health care provider's assessment ability is compromised, and the assessment accuracy might be reduced. However, previous studies have proposed remote methods to assess patients with stroke, such as internet-based telerobotic devices, videoconferencing, camera-based artificial intelligence (AI) models with wearable sensors, and remote Fugl-Meyer Assessment protocols [43-46]. The validity of the proposed methods was still low, and additional equipment and advanced software are required to capture the more accurate physical examination details.

The studies in the review also highlighted the narrow range of exercises the patients could perform through mHealth. A systematic review of home-based technologies for stroke rehabilitation revealed technology that increases the variety of exercise, such as games, telerehabilitation, robotic devices, VR devices, and tablets. This review also pointed out the limitations of each technology, which is consistent with our findings that the equipment requires proper guidance in setting up and using to achieve the therapeutic goals [47].

Besides having the proper software and hardware for mHealth services for patients with stroke, the studies also highlighted poor or limited internet connectivity as a major barrier to implementing mHealth. Many mHealth interventions require timely monitoring, access to the database, or health care. Some studies have explored alternative solutions to enhance mHealth delivery in areas with limited or nonexistent internet access. For example, some studies have used Firebase, a local offline database, or caching the user interface and assets to help patients stay connected to the information if the patients lose an internet connection [48-50]. Others have integrated communities or schools as a link between larger hospitals and patients who cannot access the internet [51,52]. Nevertheless, internet connectivity remains a critical obstacle for mHealth, as it was

considered a super social determinant of health that affects other aspects of health equity [53]. Therefore, promoting affordable and reliable internet access would be beneficial, especially for those who face disadvantages in accessing health care.

Our review reported deficiencies in support systems in using mHealth, including the need for general guidelines in using the services, technical issues, additional support, the lack of an exercise companion, and a safe environment. The lack of general guidelines or instructions in using a relative innovation can influence patients' or caregivers' decision to use and continue use. Our findings are congruent with previous longitudinal studies of the continued use of mHealth apps, indicating that users will not be motivated to continue service if they do not understand how the system works based on their initial interactions [54]. Along with a limited support system, most of the participants in our studies are notably older adults, and a prior study revealed that older adults, in general, have a lower rate of adopting mHealth [55]. The adoption of mHealth is not a simple process and requires not only technical assistance but also human support to enhance their experience and interaction with mHealth [56,57]; thus, concise and clear guidelines are essential in facilitating and engaging their usage. Another study also points out that mobile health for older adult patients used an aging barrier framework to explore usability problems and reported that participants have difficulties understanding the navigation structure of the apps and overseeing important text, buttons, and icon elements [58]. Therefore, clear guidelines and instructions should be considered when developing mHealth for survivors of stroke.

Regarding technical issues, it can be solved or reduced by additional support. Our findings align with a study evaluating patients' experience with the usability of a diabetes mHealth system. Technical issues related to functionality or operation can impact the overall usability of apps, such as challenges in deleting or inputting glucose values, resulting in difficulties in accurately interpreting value ranges (51). Furthermore, participants who encounter technical issues may struggle to navigate through different functions; for instance, participants have reported experiencing navigation difficulties while reporting glucose diary values. Prior studies have also highlighted the significance of app navigation for patients with stroke who face challenges in managing their medication [28,59]. Thus, navigation within apps is crucial as if it is difficult and user-unfriendly, it can impact patients' ability to carry out their tasks effectively [28,59].

For safety concerns, an mHealth system used to guide exercise for patients with cardiac disease developed safe algorithms to detect and warn of risky situations during exercise [60]. However, the lack of an exercise companion was not mentioned in previous studies. This can be considered for mHealth, especially exercise programs that may create live animations to exercise with patients.

In this study, researchers identified the constraint of available time as a major obstacle to accessing mHealth services among patients with stroke. This encompassed challenges such as patients' struggles with time management, inefficient use of available time, and the timely availability of caregivers [16,25,61,62]. These findings are consistent with prior studies investigating the barriers and incentives related to mHealth access among this demographic [63,64]. The effective management of time is a crucial element in the success of any mHealth intervention designed for patients with stroke [62].

The qualitative studies revealed that successful use of mHealth necessitates ongoing engagement with a caregiver, which is particularly crucial for older patients, those with more pronounced physical impairments, or those with lower digital health literacy levels. However, the effectiveness of mHealth was hampered by the limited availability of caregivers, posing challenges to sustaining quality rehabilitation approaches in home-based programs [16]. Effective time management can prove advantageous for both patients with stroke and caregivers when using telerehabilitation systems. Conversely, time constraints have been linked to decreased productivity and diminished quality within mHealth systems. Ultimately, the efficiency and efficacy of telerehabilitation hinge on the optimal use of scarce resources, notably time [65,66].

According to our findings, mHealth plays an essential role in managing and rehabilitating patients with stroke, improving and advancing the quality of care and patient outcomes, and increasing accessibility to health care resources. However, using mHealth in stroke care raises substantial concerns regarding data privacy and security. Survivors of stroke often rely on mobile applications, assisted devices for telerehabilitation, and remote monitoring tools to track their progress, communicate with health care providers, and access educational materials [24]. The data generated and transmitted through these platforms may sometimes include sensitive medical information, such as personal health records and mobility metrics. As a result, discussing mHealth in the context of stroke care must address these patient privacy and security concerns. Data transmission and storage are the most important aspects of patient privacy and security. As mobile health heavily relies on technology, cloud infrastructure, and apps, many potential risks and vulnerabilities emerge. Surprisingly, Müthing et al [67] uncovered a concerning reality within the mHealth app landscape. Their study revealed that many mHealth apps currently available lack adequate privacy and security safeguards. Even among apps certified by trusted organizations or widely adopted by the health care community, 89% were observed to transmit information digitally; alarmingly, 66% of this data was not encrypted. Moreover, there is a notable concern regarding the potential for data to be misplaced, stolen, or lost, primarily stemming from the mobile nature of the devices used in mHealth. The common habit of multiple family members sharing mobile devices like smartphones and tablets can exacerbate these concerns [68]. These risk concerns underscore the imperative requirement for solid security measures within the mHealth environment. In addition, many stakeholders have rights and responsibilities concerning an individual's medical records and the information they contain. To effectively minimize potential security risks for users in the future, it is crucial to systematically identify all parties involved and those who might be considered data "custodians" [68,69]. Informed consent, privacy policy, and access control are critical areas of concern when using mHealth. Informed consent serves as the

gateway to data sharing, granting individuals or their legal representatives the authority to specify when and with whom their personal information can be shared. However, consumers often find themselves unaware of the extensive data collection and analysis methods used by mHealth services and the extent of data sharing with third parties. This issue is exacerbated by the fact that only (183/600, 30.5%) of widely used mobile health apps have privacy policies and that most apps could negatively affect users due to privacy and security violations [69]. Moreover, many mHealth industry privacy policies often resemble lengthy academic articles with language at a university level, making them daunting for the average user to navigate [68]. Therefore, to address these concerns effectively, a set of proactive steps is required, particularly when considering the needs of older adults, the majority of patients with stroke. The future focus should be on simplifying and updating existing privacy policies to make them user-friendly, age-friendly, and compliant with regulations. Collaborating and providing active support for developing a robust regulatory framework for mHealth apps are essential. This framework should empower all users, including older adults, to control their technology and decide who can access information on their mobile devices. Significantly enhancing user privacy, with specific attention to the unique requirements of older adults, is paramount. Furthermore, promoting public technology literacy and creating awareness about responsible and cautious mobile device usage through educational initiatives is vital for empowering users and ensuring the secure adoption of mHealth solutions.

Certain limitations regarding the systematic review need to be considered. First, the ages of participants in the included studies are all over 40, with most having a median age of 55 years or older. Given the paper's focus on improving outcomes for survivors of stroke through mHealth, the older age range—whether due to sampling bias or the characteristics of patients with stroke—might negatively impact the user experience compared to younger generations. This could be attributed to varying levels of exposure to technology. Another potential limitation is the number of participants included in some of the studies. Most (7 out of 12) of the studies included fewer than 20 participants, and 10 out of 12 included fewer than 100 participants. Given the accessibility of current technology, smaller sample sizes could lead to a greater possibility of outliers, which might not adequately represent the general population of survivors of stroke, thus, introducing unintended bias. Another limiting factor is the criterion of including only English-language papers. Because stroke does not discriminate by language, including only papers written in English could omit potentially useful research regarding mHealth and survivors of stroke. Finally, this study focused solely on barriers to mHealth access in patients with stroke with a limited number of 12 included studies. To increase generalizability, future research should explore a broader range of contributing factors.

Steps Toward the Digital Health Era

The Digital Health Era is transforming the health care paradigm through various key initiatives. One of the most significant breakthroughs in the acceptance of mHealth, telemedicine, and remote patient monitoring, which greatly improve health care accessibility by enabling remote consultations and timely interventions, particularly in remote regions, is the widespread adoption of mobile devices and advanced communication technologies [70,71]. Our study has extensively explored this aspect. The integration of wearable technology and health apps also plays a crucial role in empowering individuals to actively manage their health [72-74]. These tools monitor vital signs and activity levels and provide real-time feedback, promoting a proactive approach to wellness [75-77]. Additionally, big data analytics and AI have transformed diagnostics, treatment strategies, and predictive health care [78-80]. AI algorithms analyze vast amounts of patient data, identifying patterns, and offering valuable insights to aid health care providers in making informed decisions [78-80]. The Digital Health Era emphasizes a focus on preventive care and personalized medicine, where health care providers leverage data-driven insights to create tailored treatment plans based on an individual's genetics, lifestyle, and specific health requirements [81,82]. Overall, the evolution toward the Digital Health Era is reshaping health care through initiatives such as telemedicine, wearable technology, big data analytics, and a focus on preventive care, personalized medicine, and user-centered design. These interventions must take into account the realities and constraints of the intended users. The advancements in different domains lay the foundation for an interlinked, data-driven health care system centered on patient requirements. Technology is vital in this evolution, enhancing patient outcomes, boosting productivity, and revolutionizing the health care experience. This ongoing amalgamation has the potential to establish a future where technology and health care converge effortlessly, leading to a more comprehensive and convenient health care environment for all.

Conclusions

This review has highlighted significant barriers to mHealth access among patients with stroke, emphasizing the gap between mHealth's potential and practical use. Key challenges include health-related issues, patient acceptance, infrastructure challenges, support system deficiencies, and time constraints. These findings point to the urgent need for user-friendly mHealth solutions and robust support mechanisms. Addressing these issues is critical to ensure that mHealth technologies are not only available but are also effectively adopted by survivors of stroke. Future research should focus on overcoming these barriers to enable mHealth to fully support poststroke rehabilitation and care. The goal must be to transform these obstacles into opportunities for innovation, ensuring mHealth's role as a cornerstone of patient-centered health care.

Acknowledgments

We would like to express our gratitude to Mahidol University, Thailand, for covering the Article Publishing Charge.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Summary data.
[DOCX File , 72 KB - mhealth_v12i1e54511_app1.docx]

Multimedia Appendix 2
Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist.
[PDF File (Adobe PDF File), 67 KB - mhealth_v12i1e54511_app2.pdf]

References

1. GBD 2019 Stroke Collaborators. Global, regional, and national burden of stroke and its risk factors, 1990-2019: a systematic analysis for the global burden of disease study 2019. Lancet Neurol 2021;20(10):795-820 [FREE Full text] [doi: 10.1016/S1474-4422(21)00252-0] [Medline: 34487721]
2. Burns SP, Terblanche M, MacKinen A, DeLaPena C, Fielder JDP. Smartphone and mHealth use after stroke: results from a pilot survey. OTJR (Thorofare N J) 2022;42(2):127-136. [doi: 10.1177/15394492211068851] [Medline: 35075926]
3. Sarfo FS, Ovbiagele B. Mobile health for stroke: a promising concept for research and practice. Mhealth 2017;3:4 [FREE Full text] [doi: 10.21037/mhealth.2017.02.01] [Medline: 28300225]
4. D'Souza CE, Greenway MRF, Graff-Radford J, Meschia JF. Cognitive impairment in patients with stroke. Semin Neurol 2021;41(1):75-84. [doi: 10.1055/s-0040-1722217] [Medline: 33418591]
5. Kim KT, Chang WK, Jung YS, Jee S, Sohn MK, Ko SH, et al. Unmet needs for rehabilitative management in common health-related problems negatively impact the quality of life of community-dwelling stroke survivors. Front Neurol 2021;12:758536 [FREE Full text] [doi: 10.3389/fneur.2021.758536] [Medline: 35002922]
6. Verma A, Towfighi A, Brown A, Abhat A, Casillas A. Moving towards equity with digital health innovations for stroke care. Stroke 2022;53(3):689-697 [FREE Full text] [doi: 10.1161/STROKEAHA.121.035307] [Medline: 35124973]
7. Cao W, Kadir AA, Wang Y, Wang J, Dai B, Zheng Y, et al. Description of apps targeting stroke patients: a review of apps store. Digit Health 2023;9:20552076231181473 [FREE Full text] [doi: 10.1177/20552076231181473] [Medline: 37342095]
8. Rowland SP, Fitzgerald JE, Holme T, Powell J, McGregor A. What is the clinical value of mHealth for patients? NPJ Digit Med 2020;3:4 [FREE Full text] [doi: 10.1038/s41746-019-0206-x] [Medline: 31970289]
9. Abaza H, Marschollek M. mHealth application areas and technology combinations*. A comparison of literature from high and low/middle income countries. Methods Inf Med 2017;56(7):e105-e122 [FREE Full text] [doi: 10.3414/ME17-05-0003] [Medline: 28925418]
10. Thurston C, Bezuidenhout L, Humphries S, Johansson S, von Koch L, Häger CK, et al. Mobile health to promote physical activity in people post stroke or transient ischemic attack - study protocol for a feasibility randomised controlled trial. BMC Neurol 2023;23(1):124 [FREE Full text] [doi: 10.1186/s12883-023-03163-0] [Medline: 36978045]
11. Adcock AK, Haggerty T, Crawford A, Espinosa C. mHealth impact on secondary stroke prevention: a scoping review of randomized controlled trials among stroke survivors between 2010-2020. Mhealth 2022;8:19 [FREE Full text] [doi: 10.21037/mhealth-21-27] [Medline: 35449509]
12. Moher D, Liberati A, Tetzlaff J, Altman D, PRISMA Group. Reprint--preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Phys Ther 2009;89(9):873-880. [Medline: 19723669]
13. Istepanian RSH. Mobile health (m-Health) in retrospect: the known unknowns. Int J Environ Res Public Health 2022;19(7):3747 [FREE Full text] [doi: 10.3390/ijerph19073747] [Medline: 35409431]
14. Moola S, Munn Z, Tufanaru C, Aromataris E, Sears K, Sfetcu R. Chapter 7: Systematic Reviews of Etiology and Risk. Australia: The Joanna Briggs Institute; 2017.
15. Dodakian L, McKenzie AL, Le V, See J, Pearson-Fuhrhop K, Burke Quinlan E, et al. A home-based telerehabilitation program for patients with stroke. Neurorehabil Neural Repair 2017;31(10-11):923-933 [FREE Full text] [doi: 10.1177/1545968317733818] [Medline: 29072556]
16. Bhattacharjya S, Linares I, Langan J, Xu W, Subryan H, Cavuoto LA. Engaging in a home-based exercise program: a mixed-methods approach to identify motivators and barriers for individuals with stroke. Assist Technol 2023;35(6):487-496. [doi: 10.1080/10400435.2022.2151663] [Medline: 36441082]
17. Chung J, Kim CT, Lee J, Lee E, Jung KI, Yoo WK, et al. Factors affecting participation in telerehabilitation using transcranial direct current stimulation for patients with poststroke paralysis in South Korea. Telemed J E Health 2023;29(9):1404-1411. [doi: 10.1089/tmj.2022.0452] [Medline: 36695668]
18. Ramaswamy S, Gilles N, Gruessner AC, Burton D, Fraser MA, Weingast S, et al. User-centered mobile applications for stroke survivors (MAPPS): a mixed-methods study of patient preferences. Arch Phys Med Rehabil 2023;104(10):1573-1579. [doi: 10.1016/j.apmr.2023.05.009] [Medline: 37295706]

19. Wong AKC, Bayuo J, Wong FKY, Kwok VWY, Tong DWK, Kwong MK, et al. Sustaining telecare consultations in nurse-led clinics: perceptions of stroke patients and advanced practice nurses: a qualitative study. Digit Health 2023;9:20552076231176163 [FREE Full text] [doi: 10.1177/20552076231176163] [Medline: 37214656]
20. Torriani-Pasin C, Palma GCDS, Makhoul MP, Antonio BDA, Lara ARF, da Silva TA, et al. Adherence rate, barriers to attend, safety, and overall experience of a remote physical exercise program during the COVID-19 pandemic for individuals after stroke. Front Psychol 2021;12:647883 [FREE Full text] [doi: 10.3389/fpsyg.2021.647883] [Medline: 34305716]
21. Lam SKY, Chau JPC, Lo SHS, Siow EKC, Lee VWY, Shum EWC, et al. User engagement in the development of a home-based virtual multidisciplinary stroke care clinic for stroke survivors and caregivers: a qualitative descriptive study. Disabil Rehabil 2022;44(20):5983-5989. [doi: 10.1080/09638288.2021.1955305] [Medline: 34297647]
22. Morse H, Biggart L, Pomeroy V, Rossit S. Exploring perspectives from stroke survivors, carers and clinicians on virtual reality as a precursor to using telerehabilitation for spatial neglect post-stroke. Neuropsychol Rehabil 2022;32(5):707-731. [doi: 10.1080/09602011.2020.1819827] [Medline: 32942950]
23. Spits AH, Rozevink SG, Balk GA, Hijmans JM, van der Sluis CK. Stroke survivors' experiences with home-based telerehabilitation using an assistive device to improve upper limb function: a qualitative study. Disabil Rehabil Assist Technol 2024;19(3):730-738. [doi: 10.1080/17483107.2022.2120641] [Medline: 36137223]
24. Jaywant A, Mautner L, Waldman R, O'Dell MW, Gunning FM, Toglia J. Feasibility and acceptability of a remotely delivered executive function intervention that combines computerized cognitive training and metacognitive strategy training in chronic stroke. Int J Environ Res Public Health 2023;20(9):5714 [FREE Full text] [doi: 10.3390/ijerph20095714] [Medline: 37174232]
25. Dunne S, Close H, Richards N, Ellison A, Lane AR. Maximizing telerehabilitation for patients with visual loss after stroke: interview and focus group study with stroke survivors, carers, and occupational therapists. J Med Internet Res 2020;22(10):e19604 [FREE Full text] [doi: 10.2196/19604] [Medline: 33095179]
26. Tyagi S, Lim DSY, Ho WHH, Koh YQ, Cai V, Koh GCH, et al. Acceptance of tele-rehabilitation by stroke patients: perceived barriers and facilitators. Arch Phys Med Rehabil 2018;99(12):2472-2477.e2. [doi: 10.1016/j.apmr.2018.04.033] [Medline: 29902469]
27. Mitchell J, Shirota C, Clanchy K. Factors that influence the adoption of rehabilitation technologies: a multi-disciplinary qualitative exploration. J Neuroeng Rehabil 2023;20(1):80 [FREE Full text] [doi: 10.1186/s12984-023-01194-9] [Medline: 37340496]
28. Lobo EH, Abdelrazek M, Frølich A, Rasmussen LJ, Livingston PM, Islam SMS, et al. Detecting user experience issues from mHealth apps that support stroke caregiver needs: an analysis of user reviews. Front Public Health 2023;11:1027667 [FREE Full text] [doi: 10.3389/fpubh.2023.1027667] [Medline: 37304084]
29. Pan X, Zhou X, Yu L, Hou L. Switching from offline to online health consultation in the post-pandemic era: the role of perceived pandemic risk. Front Public Health 2023;11:1121290 [FREE Full text] [doi: 10.3389/fpubh.2023.1121290] [Medline: 37261233]
30. Huang N, Yan Z, Yin H. Effects of online–offline service integration on e‐healthcare providers: a quasi‐natural experiment. Production and Operations Management 2021;30(8):2359-2378. [doi: 10.1111/poms.13381]
31. Koh GCH, Yen SC, Tay A, Cheong A, Ng YS, De Silva DA, et al. Singapore tele-technology aided rehabilitation in stroke (STARS) trial: protocol of a randomized clinical trial on tele-rehabilitation for stroke patients. BMC Neurol 2015;15:161 [FREE Full text] [doi: 10.1186/s12883-015-0420-3] [Medline: 26341358]
32. Wu N, Gong E, Wang B, Gu W, Ding N, Zhang Z, et al. A smart and multifaceted mobile health system for delivering evidence-based secondary prevention of stroke in rural China: design, development, and feasibility study. JMIR Mhealth Uhealth 2019;7(7):e13503 [FREE Full text] [doi: 10.2196/13503] [Medline: 31325288]
33. Cumming TB, Packer M, Kramer SF, English C. The prevalence of fatigue after stroke: a systematic review and meta-analysis. Int J Stroke 2016;11(9):968-977. [doi: 10.1177/1747493016669861] [Medline: 27703065]
34. Alghamdi I, Ariti C, Williams A, Wood E, Hewitt J. Prevalence of fatigue after stroke: a systematic review and meta-analysis. Eur Stroke J 2021;6(4):319-332 [FREE Full text] [doi: 10.1177/23969873211047681] [Medline: 35342803]
35. Vollertsen J, Björk M, Norlin AK, Ekbladh E. The impact of post-stroke fatigue on work and other everyday life activities for the working age population - a registry-based cohort study. Ann Med 2023;55(2):2269961 [FREE Full text] [doi: 10.1080/07853890.2023.2269961] [Medline: 37851842]
36. Lerdal A, Bakken LN, Kouwenhoven SE, Pedersen G, Kirkevold M, Finset A, et al. Poststroke fatigue--a review. J Pain Symptom Manage 2009;38(6):928-949 [FREE Full text] [doi: 10.1016/j.jpainsymman.2009.04.028] [Medline: 19811888]
37. Mead GE, Graham C, Dorman P, Bruins SK, Lewis SC, Dennis MS, UK Collaborators of IST. Fatigue after stroke: baseline predictors and influence on survival. Analysis of data from UK patients recruited in the international stroke trial. PLoS One 2011;6(3):e16988 [FREE Full text] [doi: 10.1371/journal.pone.0016988] [Medline: 21445242]
38. Snaphaan L, van der Werf S, de Leeuw FE. Time course and risk factors of post-stroke fatigue: a prospective cohort study. Eur J Neurol 2011;18(4):611-617. [doi: 10.1111/j.1468-1331.2010.03217.x] [Medline: 20868465]
39. Torriani-Pasin C, Palma GCDS, Makhoul MP, Antonio BDA, Lara ARF, da Silva TA, et al. Adherence rate, barriers to attend, safety, and overall experience of a remote physical exercise program during the COVID-19 pandemic for individuals after stroke. Front Psychol 2021;12:647883 [FREE Full text] [doi: 10.3389/fpsyg.2021.647883] [Medline: 34305716]

40. Norman CD, Skinner HA. eHealth literacy: essential skills for consumer health in a networked world. J Med Internet Res 2006;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
41. Naqvi IA, Cohen AS, Kim Y, Harris J, Denny MC, Strobino K, et al. Inequities in telemedicine use among patients with stroke and cerebrovascular diseases: a tricenter cross-sectional study. Neurol Clin Pract 2023;13(2):e200148 [FREE Full text] [doi: 10.1212/CPJ.0000000000200148] [Medline: 37064589]
42. Mahmood A, Blaizy V, Verma A, Stephen Sequeira J, Saha D, Ramachandran S, et al. Acceptability and attitude towards a mobile-based home exercise program among stroke survivors and caregivers: a cross-sectional study. Int J Telemed Appl 2019;2019:5903106 [FREE Full text] [doi: 10.1155/2019/5903106] [Medline: 31186627]
43. Kim J, Sin M, Kim WS, Min YS, Kim W, Park D, et al. Remote assessment of post-stroke elbow function using internet-based telerobotics: a proof-of-concept study. Front Neurol 2020;11:583101 [FREE Full text] [doi: 10.3389/fneur.2020.583101] [Medline: 33343489]
44. Razfar N, Kashef R, Mohammadi F. Automatic post-stroke severity assessment using novel unsupervised consensus learning for wearable and camera-based sensor datasets. Sensors (Basel) 2023;23(12):5513 [FREE Full text] [doi: 10.3390/s23125513] [Medline: 37420682]
45. Yang CL, Simpson LA, Eng JJ. A pilot study for remote evaluation of upper extremity motor function after stroke: the arm capacity and movement test (ArmCAM). Am J Occup Ther 2023;77(1):7701205020. [doi: 10.5014/ajot.2023.050020] [Medline: 36706274]
46. Rowe V, Blanton S, Aycock D, Hayat M, Ali S. Remote delivery of the fugl-meyer assessment for the upper extremity: a pilot study to assess feasibility, reliability, and validity. Arch Rehabil Res Clin Transl 2023;5(2):100261 [FREE Full text] [doi: 10.1016/j.arrct.2023.100261] [Medline: 37312985]
47. Chen Y, Abel KT, Janecek JT, Chen Y, Zheng K, Cramer SC. Home-based technologies for stroke rehabilitation: a systematic review. Int J Med Inform 2019;123:11-22 [FREE Full text] [doi: 10.1016/j.ijmedinf.2018.12.001] [Medline: 30654899]
48. Inupakutika D, Kaghyan S, Akopian D, Chalela P, Ramirez AG. Facilitating the development of cross-platform mHealth applications for chronic supportive care and a case study. J Biomed Inform 2020;105:103420 [FREE Full text] [doi: 10.1016/j.jbi.2020.103420] [Medline: 32275956]
49. Nikoviotis R, Ringas D. Self-care diabetes: a diabetes self-management application based on the seven self-management areas (AADE7) of the American association of diabetes educators. 2021 Presented at: Proceedings of the 25th Pan-Hellenic Conference on Informatics; 2022 Feb 22; United States p. 477-482. [doi: 10.1145/3503823.3503910]
50. Khan MM, Anwar R, Tanve FA, Shakil D, Banik M, Gupta SK. Development of web and mobile based smart online healthcare system. : IEEE; 2021 Presented at: 2021 IEEE 12th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON); 2021 Dec 01; USA. [doi: 10.1109/uemcon53757.2021.9666657]
51. Rono H, Bastawrous A, Macleod D, Mamboleo R, Bunywera C, Wanjala E, et al. Effectiveness of an mHealth system on access to eye health services in Kenya: a cluster-randomised controlled trial. Lancet Digit Health 2021;3(7):e414-e424 [FREE Full text] [doi: 10.1016/S2589-7500(21)00083-2] [Medline: 34167763]
52. Gizaw Z, Astale T, Kassie GM. What improves access to primary healthcare services in rural communities? A systematic review. BMC Prim Care 2022;23(1):313 [FREE Full text] [doi: 10.1186/s12875-022-01919-0] [Medline: 36474184]
53. Sieck CJ, Sheon A, Ancker JS, Castek J, Callahan B, Siefer A. Digital inclusion as a social determinant of health. NPJ Digit Med 2021;4(1):52 [FREE Full text] [doi: 10.1038/s41746-021-00413-8] [Medline: 33731887]
54. Vaghefi I, Tulu B. The continued use of mobile health apps: insights from a longitudinal study. JMIR Mhealth Uhealth 2019;7(8):e12983 [FREE Full text] [doi: 10.2196/12983] [Medline: 31469081]
55. Aranha M, James K, Deasy C, Heavin C. Exploring the barriers and facilitators which influence mHealth adoption among older adults: a literature review. Gerontechnology 2021;20(2):1-16. [doi: 10.4017/gt.2021.20.2.424.06]
56. Zakerabasali S, Ayyoubzadeh SM, Baniasadi T, Yazdani A, Abhari S. Mobile health technology and healthcare providers: systemic barriers to adoption. Healthc Inform Res 2021;27(4):267-278 [FREE Full text] [doi: 10.4258/hir.2021.27.4.267] [Medline: 34788907]
57. Hughes CML, Padilla A, Hintze A, Raymundo TM, Sera M, Weidner S, et al. Developing an mHealth app for post-stroke upper limb rehabilitation: feedback from US and ethiopian rehabilitation clinicians. Health Informatics J 2020;26(2):1104-1117 [FREE Full text] [doi: 10.1177/1460458219868356] [Medline: 31566456]
58. Wildenbos GA, Jaspers MWM, Schijven MP, Dusseljee-Peute LW. Mobile health for older adult patients: using an aging barriers framework to classify usability problems. Int J Med Inform 2019;124:68-77. [doi: 10.1016/j.ijmedinf.2019.01.006] [Medline: 30784429]
59. Ameer K, Ali K. iPad use in stroke neuro-rehabilitation. Geriatrics (Basel) 2017;2(1):2 [FREE Full text] [doi: 10.3390/geriatrics2010002] [Medline: 31011013]
60. Salvi D, Ottaviano M, Muuraiskangas S, Martínez-Romero A, Vera-Muñoz C, Triantafyllidis A, et al. An m-health system for education and motivation in cardiac rehabilitation: the experience of HeartCycle guided exercise. J Telemed Telecare 2018;24(4):303-316 [FREE Full text] [doi: 10.1177/1357633X17697501] [Medline: 28350282]
61. Torriani-Pasin C, Domingues VL, de Freitas TB, Silva TAD, Caldeira MF, Júnior RPA, et al. Adherence rate, barriers to attend, safety and overall experience of a physical exercise program via telemonitoring during COVID-19 pandemic for

individuals with parkinson's disease: a feasibility study. Physiother Res Int 2022;27(4):e1959 [FREE Full text] [doi: 10.1002/pri.1959] [Medline: 35633094]
62. Adcock AK, Haggerty T, Crawford A, Espinosa C. mHealth impact on secondary stroke prevention: a scoping review of randomized controlled trials among stroke survivors between 2010-2020. Mhealth 2022;8:19 [FREE Full text] [doi: 10.21037/mhealth-21-27] [Medline: 35449509]
63. Marzolini S. Including patients with stroke in cardiac rehabilitation: barriers and facilitators. J Cardiopulm Rehabil Prev 2020;40(5):294-301. [doi: 10.1097/HCR.0000000000000540] [Medline: 32868656]
64. Reicherzer L, Wirz M, Wieber F, Graf ES. Facilitators and barriers to health enhancing physical activity in individuals with severe functional limitations after stroke: a qualitative study. Front Psychol 2022;13:982302 [FREE Full text] [doi: 10.3389/fpsyg.2022.982302] [Medline: 36337530]
65. Grau-Pellicer M, Lalanza JF, Jovell-Fernández E, Capdevila L. Impact of mHealth technology on adherence to healthy PA after stroke: a randomized study. Top Stroke Rehabil 2020;27(5):354-368. [doi: 10.1080/10749357.2019.1691816] [Medline: 31790639]
66. Kamal A, Khoja A, Usmani B, Magsi S, Malani A, Peera Z, et al. Effect of 5-minute movies shown via a mobile phone app on risk factors and mortality after stroke in a low- to middle-income country: randomized controlled trial for the stroke caregiver dyad education intervention (Movies4Stroke). JMIR Mhealth Uhealth 2020;8(1):e12113 [FREE Full text] [doi: 10.2196/12113] [Medline: 32012080]
67. Müthing J, Jäschke T, Friedrich CM. Client-focused security assessment of mHealth apps and recommended practices to prevent or mitigate transport security issues. JMIR Mhealth Uhealth 2017;5(10):e147 [FREE Full text] [doi: 10.2196/mhealth.7791] [Medline: 29046271]
68. Galvin HK, DeMuro PR. Developments in privacy and data ownership in mobile health technologies, 2016-2019. Yearb Med Inform 2020;29(1):32-43 [FREE Full text] [doi: 10.1055/s-0040-1701987] [Medline: 32823298]
69. Bhuyan SS, Kim H, Isehunwa OO, Kumar N, Bhatt J, Wyant DK, et al. Privacy and security issues in mobile health: current research and future directions. Health Policy and Technology 2017;6(2):188-191 [FREE Full text] [doi: 10.1016/j.hlpt.2017.01.004]
70. Fatehi F, Taylor M, Caffery LJ, Smith AC. Telemedicine for clinical management of adults in remote and rural areas. Revolutionizing Tropical Medicine: Point-of-Care Tests, New Imaging Technologies and Digital Health 2019:439-461. [doi: 10.1002/9781119282686.ch26]
71. Muruganandam N, Sindhu E, Safraz M. The conceptualization and implementation of a patient monitoring system based on the internet of things for use in telemedicine. : IEEE; 2023 Presented at: 2023 International Conference on Artificial Intelligence and Knowledge Discovery in Concurrent Engineering (ICECONF); 2023 Jan 05; Chennai, India. [doi: 10.1109/iceconf57129.2023.10084203]
72. Scheid JL, Reed JL, West SL. Commentary: is wearable fitness technology a medically approved device? Yes and No. Int J Environ Res Public Health 2023;20(13):6230 [FREE Full text] [doi: 10.3390/ijerph20136230] [Medline: 37444078]
73. Sqalli MT, Al-Thani D. Towards an integrated health and wellness program using human-IoT interaction. 2020 Presented at: 2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT); 2020 Feb 02; Doha, Qatar. [doi: 10.1109/iciot48696.2020.9089538]
74. Uddin M, Syed-Abdul S. Introduction to digital health approach and mHealth applications for participatory health. Digital Health Elsevier 2021:1-6. [doi: 10.1016/b978-0-12-820077-3.00001-8]
75. Verma K, Preity P, Ranjan R. An insight into wearable devices for smart healthcare technologies. : IEEE; 2023 Presented at: 2023 13th International Conference on Cloud Computing, Data Science & Engineering (Confluence); 2023 Jan 19; Noida, India. [doi: 10.1109/confluence56041.2023.10048809]
76. Alves M, Pereira TC, Costa E, Magalhães B, Vázquez-Justo E. Advances in tourism, technology and systems: selected papers from ICOTTS 2022. In: Vital Signs-Health IoT Smartwatch. Portugal: Springer; 2023:619-631.
77. Wasekar RR, Junzal S, Sathavane GV. Health monitoring devices system and technology. ECS Trans 2022;107(1):15915-15922. [doi: 10.1149/10701.15915ecst]
78. Balasubramani S, Renjith P, Ramesh K, George AH, George AS, Shahul A. A detailed analysis of many different artificial intelligence-based healthcare systems. In: Handbook of Research on Thrust Technologies? Effect on Image Processing: IGI Global. USA: IGI Global; 2023:56-72.
79. Mina A. Big data and artificial intelligence in future patient management. How is it all started? Where are we at now? Quo tendimus? Adv Lab Med 2020;1(3):20200014 [FREE Full text] [doi: 10.1515/almed-2020-0014] [Medline: 37361493]
80. Miner GD, Miner LA, Burk S, Goldstein M, Nisbet R, Walton N. Practical data analytics for innovation in medicine. In: Building Real Predictive and Prescriptive Models in Personalized Healthcare and Medical Research Using AI, ML, and Related Technologies. USA: Academic Press; 2023.
81. Barrera-Vázquez OS, Rivero-Segura NA, Gomez-Verjan JC. Digital health and physiomics. In: Principles of Genetics and Molecular Epidemiology. Mexico: Springer; 2022:111-116.
82. Xiong S, Lu H, Peoples N, Duman EK, Najarro A, Ni Z, et al. Digital health interventions for non-communicable disease management in primary health care in low-and middle-income countries. NPJ Digit Med 2023;6(1):12 [FREE Full text] [doi: 10.1038/s41746-023-00764-4] [Medline: 36725977]

Abbreviations

AI: artificial intelligence
CCT: computerized cognitive training
JBI: Joanna Briggs Institute
mHealth: mobile health
MST: metacognitive strategy training
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
VR: virtual reality

Edited by L Buis; submitted 13.11.23; peer-reviewed by J Wang, T Salzmann; comments to author 20.02.24; revised version received 02.04.24; accepted 25.06.24; published 22.08.24.

Please cite as:
Niyomyart A, Ruksakulpiwat S, Benjasirisan C, Phianhasin L, Nigussie K, Thorngthip S, Shamita G, Thampakkul J, Begashaw L
Current Status of Barriers to mHealth Access Among Patients With Stroke and Steps Toward the Digital Health Era: Systematic Review
JMIR Mhealth Uhealth 2024;12:e54511
URL: https://mhealth.jmir.org/2024/1/e54511
doi:10.2196/54511
PMID:

Review

Real-World Accuracy of Wearable Activity Trackers for Detecting Medical Conditions: Systematic Review and Meta-Analysis

Ben Singh[1], PhD; Sebastien Chastin[2], PhD; Aaron Miatke[2], MSc; Rachel Curtis[2], PhD; Dorothea Dumuid[2], PhD; Jacinta Brinsley[2], PhD; Ty Ferguson[2], PhD; Kimberley Szeto[2], PhD; Catherine Simpson[2], MSc; Emily Eglitis[2], MSc; Iris Willems[2], MSc; Carol Maher[2], PhD

[1]Allied Health & Human Performance, University of South Australia, Adelaide, Australia

[2]Department of Rehabilitation Sciences and Physiotherapy, Ghent University, Ghent, Belgium

Corresponding Author:
Ben Singh, PhD
Allied Health & Human Performance
University of South Australia
Corner of North Terrace and Frome Road
Adelaide, 5001
Australia
Phone: 61 1300301703
Email: ben.singh@unisa.edu.au

Abstract

Background: Wearable activity trackers, including fitness bands and smartwatches, offer the potential for disease detection by monitoring physiological parameters. However, their accuracy as specific disease diagnostic tools remains uncertain.

Objective: This systematic review and meta-analysis aims to evaluate whether wearable activity trackers can be used to detect disease and medical events.

Methods: Ten electronic databases were searched for studies published from inception to April 1, 2023. Studies were eligible if they used a wearable activity tracker to diagnose or detect a medical condition or event (eg, falls) in free-living conditions in adults. Meta-analyses were performed to assess the overall area under the curve (%), accuracy (%), sensitivity (%), specificity (%), and positive predictive value (%). Subgroup analyses were performed to assess device type (Fitbit, Oura ring, and mixed). The risk of bias was assessed using the Joanna Briggs Institute Critical Appraisal Checklist for Diagnostic Test Accuracy Studies.

Results: A total of 28 studies were included, involving a total of 1,226,801 participants (age range 28.6-78.3). In total, 16 (57%) studies used wearables for diagnosis of COVID-19, 5 (18%) studies for atrial fibrillation, 3 (11%) studies for arrhythmia or abnormal pulse, 3 (11%) studies for falls, and 1 (4%) study for viral symptoms. The devices used were Fitbit (n=6), Apple watch (n=6), Oura ring (n=3), a combination of devices (n=7), Empatica E4 (n=1), Dynaport MoveMonitor (n=2), Samsung Galaxy Watch (n=1), and other or not specified (n=2). For COVID-19 detection, meta-analyses showed a pooled area under the curve of 80.2% (95% CI 71.0%-89.3%), an accuracy of 87.5% (95% CI 81.6%-93.5%), a sensitivity of 79.5% (95% CI 67.7%-91.3%), and specificity of 76.8% (95% CI 69.4%-84.1%). For atrial fibrillation detection, pooled positive predictive value was 87.4% (95% CI 75.7%-99.1%), sensitivity was 94.2% (95% CI 88.7%-99.7%), and specificity was 95.3% (95% CI 91.8%-98.8%). For fall detection, pooled sensitivity was 81.9% (95% CI 75.1%-88.1%) and specificity was 62.5% (95% CI 14.4%-100%).

Conclusions: Wearable activity trackers show promise in disease detection, with notable accuracy in identifying atrial fibrillation and COVID-19. While these findings are encouraging, further research and improvements are required to enhance their diagnostic precision and applicability.

Trial Registration: Prospero CRD42023407867; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=407867

(JMIR Mhealth Uhealth 2024;12:e56972) doi:10.2196/56972

KEYWORDS

wearable activity trackers; disease detection; atrial fibrillation; COVID-19 diagnosis; meta-analysis; wearables; wearable tracker; tracker; detection; monitoring; physiological; diagnostic tool; tool; tools; Fitbit; atrial; COVID-19; wearable

Introduction

As health care budgets around the world continue to soar, the need for cost-effective interventions that both reduce health care costs and improve patient outcomes has never been more urgent [1]. Early detection of medical conditions offers a pathway to achieve these goals, enabling prompt intervention during acute medical events or even pre-emptive action before such events occur [2]. Wearable activity monitors are emerging as a potential tool in this evolving landscape.

In recent years, wearable activity trackers have become ubiquitous tools, widely adopted for tracking and enhancing physical activity and other lifestyle behaviors, helping to mitigate the risk of chronic diseases [3]. These devices measure a plethora of activity metrics such as steps taken, distance covered, energy expenditure, physical activity intensities, and sleep patterns [4]. The scientific literature has witnessed a surge in original studies and systematic reviews and meta-analyses, focused on determining the reliability and validity of activity trackers for measuring activity levels [5,6] and their effectiveness in intervening in daily activity patterns and downstream health outcomes [7-12]. These studies have shown that interventions using consumer-based wearable activity trackers can increase physical activity participation and lead to significant improvements in health outcomes, across a range of populations [7-12]. As wearable technology has progressed, wearable activity trackers offer increasing potential to move beyond activity metrics and aid in the early identification of diseases and other medical events.

Rapid technological advancements have significantly extended the capabilities of contemporary consumer-grade wearable activity trackers such as Fitbits and Apple Watches [13]. Modern wearables incorporate sophisticated sensors capable of monitoring a wide array of physiological parameters beyond just movement including heart rate, blood oxygen levels, sleep quality, and stress markers [14]. While this expanded functionality holds promise for disease detection and monitoring, the evidence supporting the use of consumer wearables for such applications remains limited. For example, the systematic review by Alban-Cadena et al [15] evaluated wearable sensors for monitoring Parkinson disease–related gait impairments and symptoms such as tremors, bradykinesia, and dyskinesia. However, most included studies were very small (10-20 participants) and were conducted in controlled laboratory environments using specialized setups such as multi-sensor accelerometer arrays worn on the ankles and spine. While offering the potential for home-based rehabilitation, the generalizability of these findings to widely adopted, consumer-oriented wearable trackers designed for real-world, free-living conditions is unclear.

Other recent systematic reviews have evaluated the accuracy of wearable tracking devices for detecting specific health conditions such as arrhythmias [16], cardiovascular disease [17,18], and COVID-19 [19]. However, these reviews have notable limitations. Most included studies were conducted in controlled laboratory settings, limiting the generalizability of their findings to real-world, free-living conditions [16,17,19]. Additionally, these reviews focused narrowly on individual clinical outcomes, preventing comparisons of wearables' detection accuracy across different medical conditions and events. For example, the narrative syntheses highlighted wearables' potential as complementary tools for detecting cardiovascular conditions such as arrhythmias, atrial fibrillation, myocardial infarction, and heart failure [16,17]. The meta-analysis of Lee et al [18] of 26 studies found wearable devices had a pooled sensitivity of 94.80% and specificity of 96.96% for atrial fibrillation detection. In contrast, Cheong et al [19] reported lower diagnostic accuracy for COVID-19 detection, with area under the curve (AUC) values ranging from 75% to 94.4% and sensitivity and specificity ranging from 36.5% to 100% and 73% to 95.3%, respectively. Notably, all but one review [18] used narrative synthesis approaches [16,17,19], limiting their ability to quantify detection accuracy, and preventing readers from comparing detection accuracy across conditions reported in the respective reviews.

As wearable technology rapidly evolves, with frequent introductions of new and more advanced devices, the scientific evidence base for disease detection is growing, encompassing a wider range of medical conditions and events. Consequently, there is now sufficient data to warrant a comprehensive systematic review with meta-analyses, allowing quantitative comparisons of wearables' detection accuracy across various conditions in real-world settings.

Our systematic review and meta-analysis aim to fill this crucial gap by comprehensively assessing the reliability and accuracy of consumer-grade wearable activity trackers for detecting and monitoring a wide range of medical conditions and events in free-living, real-world settings. Unlike previous reviews that relied on narrative synthesis approaches, our quantitative meta-analyses will allow for robust comparisons of wearables' diagnostic performance across diverse conditions and events. By rigorously evaluating evidence from studies conducted in real-world contexts, our review will provide evidence to guide the responsible and effective implementation of wearable technology for early detection and continuous health monitoring by researchers, health care providers, policy makers, technology companies, and other stakeholders. As consumer adoption of wearables continues to rise rapidly worldwide, our comprehensive synthesis will assist in harnessing their potential while mitigating risks and ensuring appropriate use.

Methods

Protocol and Registration

The protocol for this systematic review was prospectively registered on PROSPERO (ID CRD42023407867) and this paper is reported according to PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) [20] guidelines.

Selection Criteria and Search Strategy

The inclusion criteria are summarized in Table S1 in Multimedia Appendix 1. The inclusion criteria were developed using the population, exposure, outcomes, and study type criteria as follows: adult population (aged 18 years or older) in free-living

conditions, that have not been recruited based on a specific health condition or diagnosis; use of a wearable activity tracker (eg, Fitbit, Apple Watch, or a research-grade accelerometer) for the detection of any disease or medical event (eg, atrial fibrillation, the onset of infectious disease, and falls). To be eligible, the wearable activity tracker had to be able to detect movement behavior (ie, include an accelerometer), but could also include other types of sensors (eg, light sensor and temperature sensor). The wearable activity tracker had to consist of a single device worn on a single body location (eg, on the wrist or chest, not across both); studies needed to assess the actual diagnosis of a medical condition or occurrence of events that had clinical relevance (eg, falls). Eligible studies are needed to report an outcome related to diagnostic accuracy, such as specificity or sensitivity of the device for early detection of disease or medical events. Examples could include but were not limited to, providing effect estimates of overall diagnostic accuracy (%), sensitivity (%), and specificity (%) with 95% CIs; and validation studies conducted under free-living conditions that were reported in a peer-reviewed journal study were included. This included secondary analyses conducted within the context of observational studies, experimental studies, or quasi-experimental studies. Both consumer-initiated studies, where existing consumers who had purchased their own wearables were invited to join a study, and researcher-initiated studies, where researchers recruited participants and provided them with wearables, were included, as they represent 2 complementary real-world contexts in which wearable devices are often implemented for disease detection and monitoring. Studies were included only if they evaluated wearable devices provided by health care providers or researchers as part of a formal monitoring program, and the detection of a specific clinical event or disease was a prespecified outcome measure of the study. Studies examining consumer-driven self-tracking with personal wearables outside of a health care context were excluded. The following were also excluded: studies involving children or adolescents, studies examining symptoms within people known to have a specific disease, wearable devices that cannot track activity levels (eg, continuous glucose monitors), studies evaluating an array of wearable sensors worn at multiple body locations (eg, watch plus skin patch) or pedometers, studies measuring the association between an exposure and an outcome (eg, using odds ratios, relative risk, and hazard ratios), lab- or hospital-based studies, and conference abstracts or dissertations.

Ten databases were searched (CINAHL, Cochrane Library, Embase via OVID, MEDLINE via OVID, Emcare via OVID, JMIR Publications, ProQuest Central, ProQuest Nursing and Allied Health Source, PsycINFO, and Scopus) using subject heading, keyword, and MeSH (Medical Subject Headings) term searches for terms related to “wearable device” and “detection” (see Table S2 in Multimedia Appendix 1 for the full search strategy). We intentionally used broad search terms to ensure a comprehensive capture of the evidence base, including all types of medical conditions and events, without restricting our search to predefined diagnostic or event outcomes. Database searches were limited to peer-reviewed journal studies published in English from inception to April 1, 2023.

Data Management and Extraction

Search results were imported into ASReview (version 2.0; ASReview Community), an open-source software artificial intelligence tool designed for screening studies for systematic reviews. Title or abstract screening was conducted in ASReview by paired independent reviewers (BS and DD, RC, TF, JB, IW, KS, CS, AM, or EE). The software uses an active learning algorithm that iteratively selects the most relevant studies for inclusion based on the initial judgments made by the research team. The screening was stopped when 100 consecutive nonrelevant studies were screened. Following title or abstract screening, results were then imported to EndNote X9 (Clarivate) where duplicates were removed and then exported into Covidence (Veritas Health Innovation) for full-text screening, data extraction, and risk of bias scoring which was completed in duplicate by paired independent reviewers (BS and DD, RC, TF, JB, IW, KS, CS, AM, or EE), with disagreements resolved by discussion.

Data were extracted in duplicate by paired independent reviewers (BS and DD, RC, TF, JB, IW, KS, CS, AM, or EE) using a standardized extraction form in Covidence. The risk of bias in the included reviews was assessed by 2 independent reviewers in duplicate using the Joanna Briggs Institute (JBI) Critical Appraisal Checklist for Diagnostic Test Accuracy Studies. Studies were rated out of nine for the following items: (1) enrollment of consecutive or random sample, (2) the avoidance of a case-control design, (3) inappropriate exclusions, (4) the interpretation of index test results, (5) the prespecification of thresholds, (6) reference standard classification, (7) the interpretation of reference standard, (8) timing of tests, and (9) analysis.

Data Synthesis and Analysis

For each meta-analysis, data were combined at the study level. Separate meta-analyses were performed for (1) COVID-19 detection, (2) atrial fibrillation or arrhythmia detection, and (3) fall detection. Outcomes of interest were analyzed and data were pooled using sensitivity (%), specificity (%), AUC (%), accuracy (%), and positive predictive value (PPV), with 95% CIs as the effects measures. Sensitivity (%) denotes the percentage of individuals with the disease or condition correctly identified by the test, while specificity (%) represents the percentage of those without the disease or condition correctly identified as negative. The AUC (%) quantifies the test’s overall diagnostic accuracy, ranging from 0% to 100%, with higher values indicating better performance. Accuracy (%) reflects the proportion of all tests accurately classified, and PPV (%) indicates the likelihood that a positive test result correlates with the disease or condition being tested for. If 95% CIs were not reported in a study, they were calculated based on available data, using recommended formulas [21]. Publication bias was evaluated using funnel plots of effect sizes and standard errors and evaluating for asymmetries or missing sections within the plot, for meta-analyses that involved more than 10 studies. The Cochran Q test was used to assess statistical heterogeneity and the I^2 statistic was used to quantify the proportion of the overall outcome attributed to variability. The following cut-off values for the I^2 statistic were used: 0% to 29%=no heterogeneity; 30%

to 49%=moderate heterogeneity; 50% to 74%=substantial heterogeneity; and 75% to 100%=considerable heterogeneity [22]. Subgroup analyses were undertaken to evaluate device type (Fitbit, Apple watch, Oura ring, and others) for outcomes that had at least 2 studies in each subgroup. Sensitivity analyses for the meta-analysis were performed by removing the study with the lowest sensitivity, specificity, AUC, accuracy, or PPV. All meta-analyses were performed using Stata/MP (version 16; StataCorp).

The overall level of evidence was graded using the Oxford Centre for Evidence-Based Medicine 2011 Levels of Evidence, as follows: grade A: consistent level 1 studies (ie, individual randomized controlled trials); B: consistent level 2 (ie, individual cohort studies) or 3 studies (ie, individual case-control studies) or extrapolations from level 1 studies; C: level 4 studies (ie, case series) or extrapolations from level 2 or 3 studies; or D: level 5 (ie, expert opinion without explicit critical appraisal) evidence or inconsistent or inconclusive studies of any level [23]. Each outcome of interest was assigned a "Grade of Recommendation" based on meeting these criteria.

Deviations From the Registered Protocol

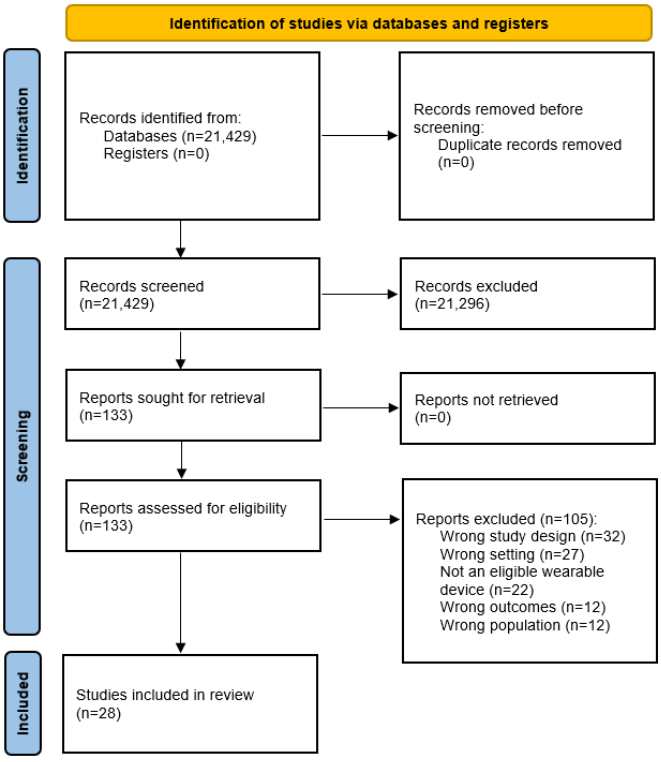
We planned to use the Effective Public Health Practice Project Quality Assessment Tool to assess study quality and risk of bias. However, during data extraction and quality assessment, we opted to use the JBI Critical Appraisal Checklist for Diagnostic Test for Accuracy Studies, as this instrument was more relevant to the included studies. Further, we were unable to conduct subgroup analyses for the type of wearable for atrial fibrillation and fall detection, due to an insufficient number of studies.

Results

Overview

Of the 21,429 records identified following the database search, 28 were eligible (see Figure 1 for PRISMA flowchart including reasons for exclusions; see Table S3 in Multimedia Appendix 1 for a complete list of full texts that were excluded during the final stage of screening, with reasons). An overview of all included study's characteristics is shown in Table S4 in Multimedia Appendix 1. There was a total of 1,226,801 participants (median sample size 264, IQR 96-8338; range 29-455,699). Median participant age was 47.3 (IQR 36.6-66), between 28.6 and 78.3, years and 21 (75%) studies involved female and male participants (gender was not reported in 7 (25%) studies). A total of 16 (57%) studies evaluated COVID-19, 5 (18%) studies evaluated atrial fibrillation, 3 (11%) studies assessed a broad range of cardiac arrhythmias, 3 (11%) studies assessed falls, and 1 (3.6%) study assessed viral symptoms. The devices used in the studies were Fitbit (n=6), Apple Watch (n=6), Oura ring (n=3), a combination of various devices (ie, studies that used a combination of the Apple Watch, Fitbit, Garmin, and other devices; n=7), Empatica E4 (n=1), Dynaport MoveMonitor (n=2), Samsung Galaxy Watch (n=1), and other or not specified (n=2). The median score for the JBI Critical Appraisal Checklist for Diagnostic Test Accuracy Studies was 6 (IQR 5-7; range 1-9) out of 9 (Table S5 in Multimedia Appendix 1).

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



There was sufficient data in the included studies to conduct meta-analyses for the following clinimetrics: (1) COVID-19 detection (accuracy, %; sensitivity, %; AUC, %; and specificity, %), (2) atrial fibrillation detection (PPV, %; sensitivity, %; and specificity, %), and (3) falls detection (sensitivity, %; and specificity, %).

Meta-Analysis Results

COVID-19 Detection

Meta-analysis results of AUC, accuracy, sensitivity, and specificity for COVID-19 detection are shown in Figure 2. Meta-analyses of 9 studies showed a pooled AUC of 80.15% (95% CI 71.03%-89.27%) and 5 studies had a pooled accuracy of 87.54% (95% CI 81.57%-93.51%). Pooled sensitivity from 8 studies was 79.53% (95% CI 67.73%-91.33%), and 7 studies showed a pooled specificity of 76.79% (95% CI 69.44%-84.13%).

Subgroup analysis for device type for sensitivity and specificity are shown in Figures S6 and S7 in Multimedia Appendix 1, respectively. A summary of sensitivity and specificity for the different devices is shown in Figure 3. Overall, the Fitbit had a sensitivity and specificity of 75.39% and 90.60%, respectively, the Oura ring had a sensitivity and specificity of 80.47% and 72.60%, respectively, and combined devices had a sensitivity and specificity of 82.69% and 74.62%, respectively.

The results of sensitivity analyses are shown in Figure S3 in Multimedia Appendix 1. Following the removal of the worst-performing study, AUC was 84.10%, accuracy was 88.65%, sensitivity was 85.62%, and specificity was 78.57%.

Grade of recommendation: (B) consistent level 2 studies supporting the use of wearable activity trackers for the detection of COVID-19.

Figure 2. Meta-analysis of accuracy, sensitivity, AUC, and specificity of wearable activity trackers for detection of COVID-19. AUC: area under the curve.

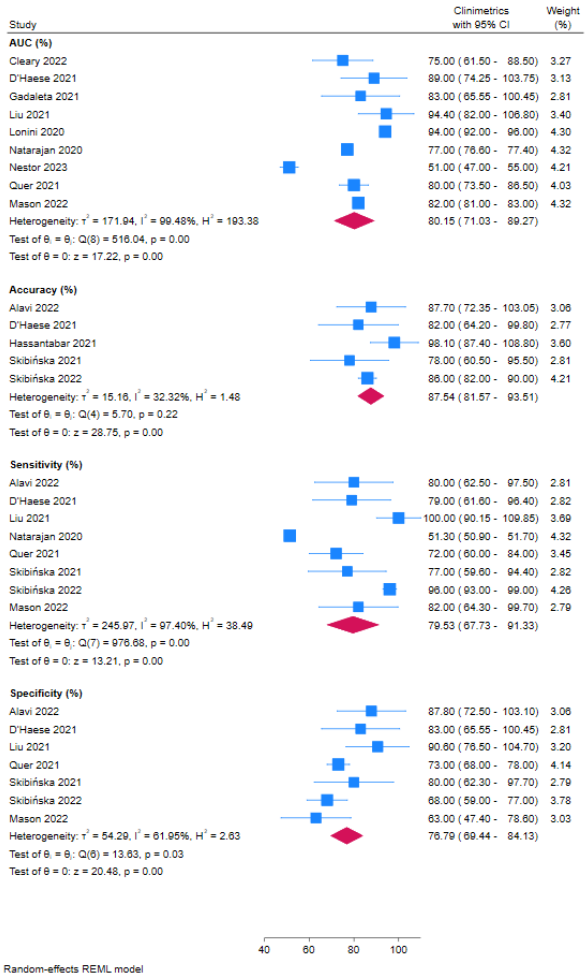
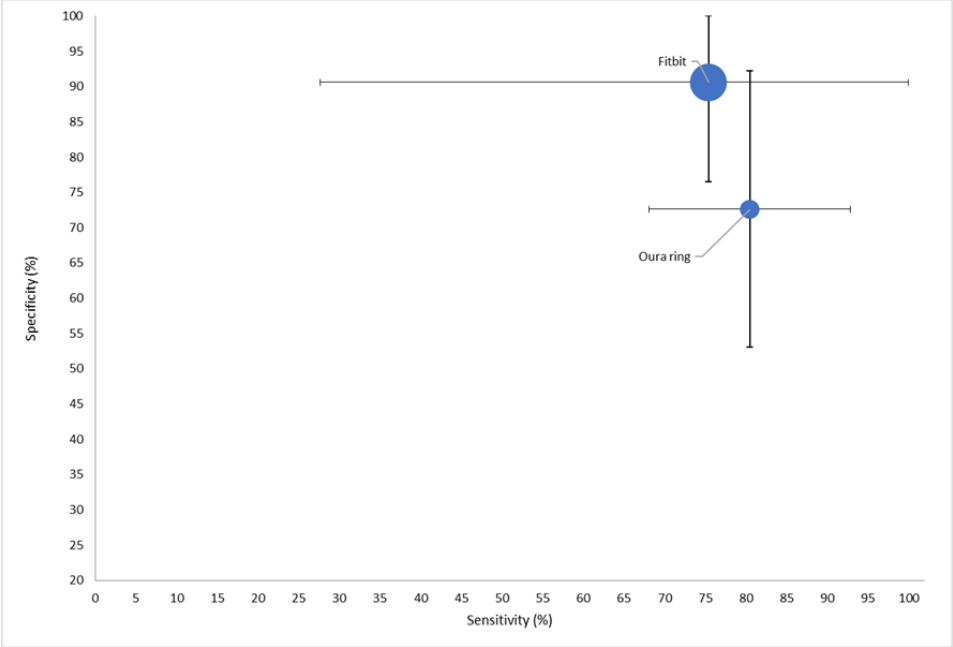


Figure 3. Overview of sensitivity and specificity for the different devices for COVID-19 detection.



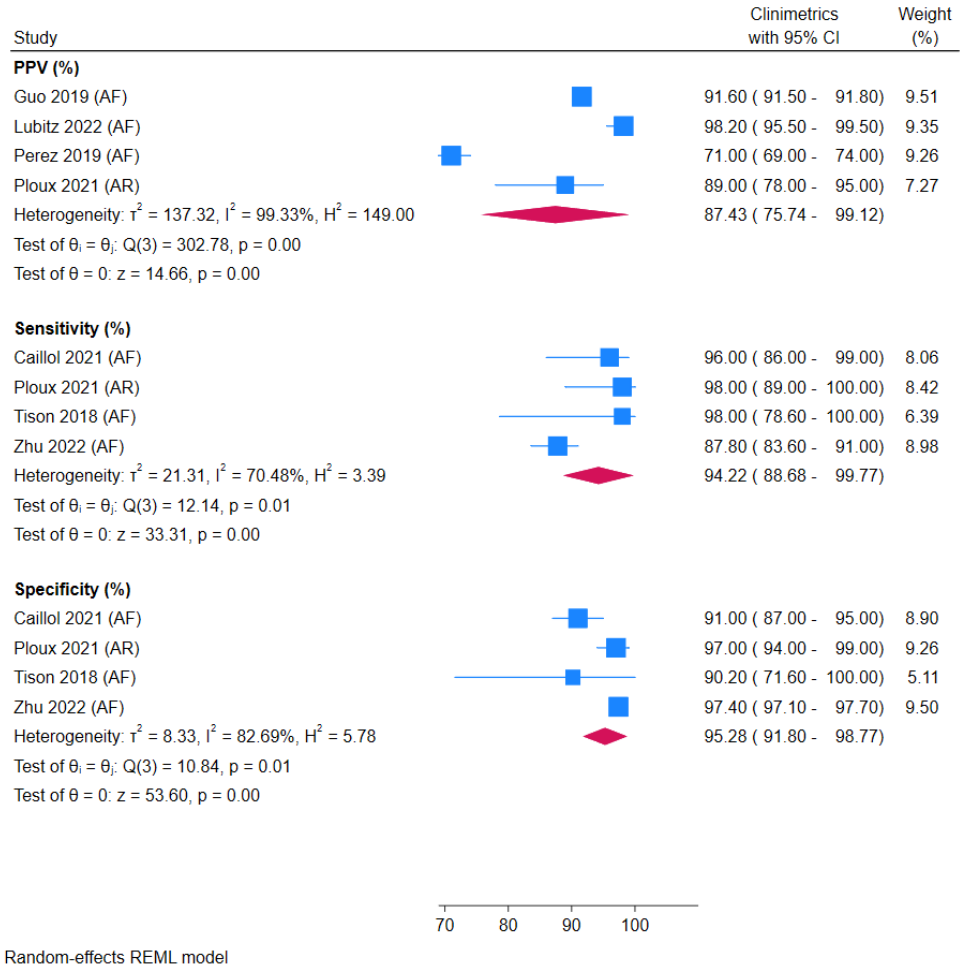
Atrial Fibrillation Detection

Pooled analyses of PPV, sensitivity, and specificity for atrial fibrillation detection are shown in Figure 4. Meta-analysis of 4 studies showed a combined PPV of 87.43% (95% CI 75.74%-99.12%). Pooled sensitivity was 94.22% (95% CI 88.68%-99.77%; 4 studies) and pooled specificity was 95.28% (95% CI 91.80%-98.77%; 4 studies).

The results of sensitivity analyses are shown in Figure S4 in Multimedia Appendix 1. Following the removal of the worst-performing study, PPV was 93.64%, sensitivity was 97.28%, and specificity was 95.55%.

Grade of recommendation: (B) consistent level 2 studies supporting the use of wearable activity trackers for the detection of atrial fibrillation.

Figure 4. Meta-analysis of PPV, sensitivity, and specificity of wearable activity trackers for detection of AF and AR. AF: atrial fibrillation; AR: arrhythmia; PPV: positive predictive value.

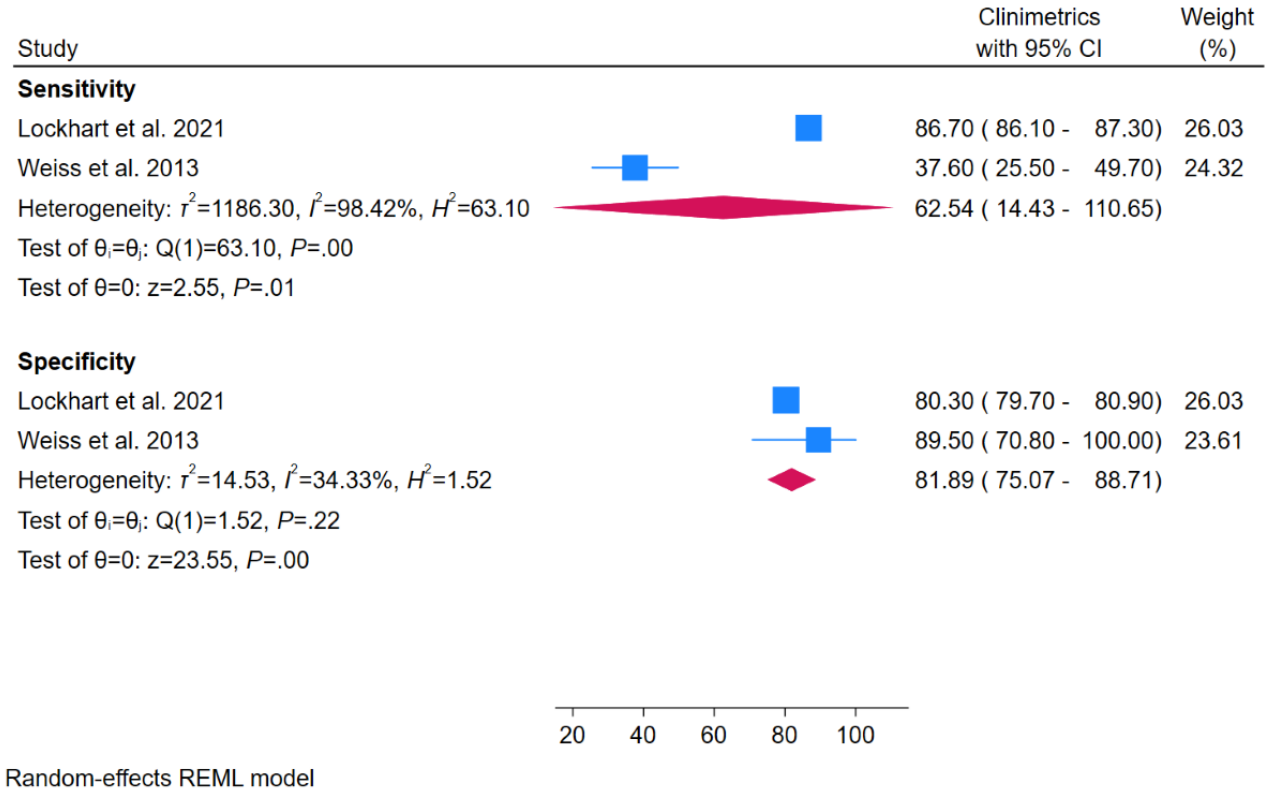


Falls Detection

Meta-analysis results of sensitivity and specificity for fall detection are shown in Figure 5. Meta-analyses of 2 studies showed a specificity of 62.54% (95% CI 14.43%-100%) and a sensitivity of 81.89% (95% CI 75.07%-88.17%). There was an insufficient number of studies for subgroup analyses of device type and sensitivity analyses for fall detection.

Grade of recommendation: (D) inconsistent or inconclusive studies of any level for the use of wearable activity trackers to predict falls.

Figure 5. Meta-analyses of sensitivity and specificity of wearable activity trackers for detection for fall detection.



Discussion

Principal Findings

In this study, we set out to systematically review and meta-analyze the current evidence regarding wearable activity trackers' ability to detect medical conditions and events under free-living conditions. To date, the majority of studies have focused on the detection of COVID-19, with a smaller number of studies focused on cardiac conditions and falls. For COVID-19 detection, the devices generally demonstrated good sensitivity and specificity. The most promising results were found for the detection of atrial fibrillation, for which the wearables showed high sensitivity and specificity. Whereas, for fall detection, the present findings devices showed moderate sensitivity but lower specificity. These findings indicate that while these devices are becoming more dependable for monitoring specific health conditions, their performance varies depending on the condition being detected.

The current body of evidence on the diagnostic potential of wearable activity trackers is notably skewed toward COVID-19 detection, a focus that is understandable given the pandemic's global impact and the consequent urgent need for monitoring solutions. Researching the feasibility of detecting COVID-19 through wearables holds appeal due to the availability of widely used reference standards. Rapid and polymerase chain reaction tests, widely used, allow for easy self-reporting of COVID-19 diagnoses by many individuals. In contrast, accessing a reliable gold standard for other health outcomes poses significant challenges. However, what was surprising to note is the limited number of studies exploring these trackers for other health conditions, especially given that numerous wearables advertise features such as sleep apnea detection—a topic noticeably absent in our findings. Our extensive database search identified only a handful of studies each related to cardiac issues and falls. This gap in the literature is striking considering the wide array of health conditions that could theoretically be monitored using wearable technology, given their ability to capture data related to heart rate, movement, skin temperature, and more. Such capabilities would suggest that a broad spectrum of medical conditions could be measured, spanning cardiovascular and respiratory conditions to neurological and psychological disorders. It is important to note that we intentionally focused on the accuracy of data collected in free-living conditions (with a view to understanding current-day diagnostic capabilities). We note numerous laboratory-based studies that were excluded (eg, [24] and [25]) suggesting that a wider range of diagnostic outcomes may become available in the future. Furthermore, many studies were excluded because they focused on monitoring symptoms in people with a known diagnosis (eg, seizures in people with epilepsy [26], and freezing gait in Parkinson disease [27]) which was outside the scope of this study, but highlights wearable activity trackers' potential for medical condition monitoring.

This study revealed that wearable activity trackers demonstrate moderate to high sensitivity and specificity for COVID-19 detection. It is interesting to compare our results with those for other COVID-19 screening tests. A systematic review by Mistry et al [28] on lateral flow devices (LFD) tests (also known as rapid antigen tests) evaluated 24 papers across 8 different LFD brands, covering over 26,000 test results. Their findings indicated that sensitivity ranged from 37.7% to 99.2% and specificity ranged from 92.4% to 100% [28]. Comparatively,

this study's pooled sensitivity for wearable-detected COVID-19 was 79.5% (range 51.3%-100%), which is in line with the LFD results. However, our specificity of 76.8% (range 63%-90.6%) was slightly lower. According to UK government guidelines, the benchmarks for COVID-19 workplace screening are ≥68% for sensitivity and ≥97% for specificity [29]. This suggests that while wearable activity monitor detection meets the sensitivity criterion, it falls short on specificity.

The most promising results were observed for the detection of atrial fibrillation, with figures that compare favorably to other clinical tests. For example, the sensitivity and specificity of a 12-lead electrocardiogram for detecting atrial fibrillation have previously been shown to range between 93% and 97% [30,31], which appears similar to our sensitivity and specificity of 94.2% and 95.3%, respectively. Over the course of 2022-2023, major brands, such as Fitbit [32], Apple Watch [33], Garmin [34], and Samsung [35], received approval from the US Food and Drug Administration for their atrial fibrillation detection features. The relatively higher accuracy in identifying cardiac arrhythmias as compared to COVID-19 is perhaps expected, given that cardiac functions can be deduced from wearables' optical heart-rate sensors. In contrast, COVID-19 detection usually requires intricate algorithms that amalgamate multiple data points [36,37].

While wearable activity trackers demonstrated effectiveness in detecting cardiac arrhythmia and COVID-19, our meta-analysis revealed that their accuracy in detecting falls was only moderate. The devices were generally effective in identifying actual falls, with a sensitivity of 81.9%. However, they also generated a significant number of false positives, as evidenced by a lower specificity of 62.5%. This aligns with existing literature on the subject [38,39]. It is crucial to note that our review specifically focused on the performance of these devices in real-world conditions among the general population. Most existing studies on fall detection with wearables have been conducted in controlled laboratory settings using simulated falls, where accuracy has generally been higher [38,39]. The false positives in fall detection are likely due to the devices relying on accelerometry data, which can misinterpret other rapid downward movements as falls. Further research is needed to refine the algorithms used in these devices to improve their performance in fall detection. Future studies might incorporate additional metrics, such as rapid changes in heart rate or galvanic skin response, which may accompany a fall, to enhance accuracy.

This study offers several significant strengths, including being the first systematic review and meta-analysis focused on the real-world accuracy of wearable activity trackers in detecting medical conditions and events. The review analyzed a robust data set from 28 studies, involving over 1 million participants, enabling a comprehensive meta-analysis of various outcomes. Instead of limiting our focus to specific diagnostic outcomes, we examined a broad range of medical conditions. Our search strategy was exceptionally thorough, encompassing 10 databases and reviewing over 21,000 studies to capture a wide array of diagnostic outcomes. Methodologically, we adhered to the PRISMA 2020 guidelines, which included conducting sensitivity and subgroup analyses, as well as evaluating the certainty of the evidence.

Study limitations must be acknowledged. There was considerable heterogeneity in the designs of included studies, such as their reference standards, diagnostic tests, and sample characteristics. Given the size of the current evidence, there were too few studies to conduct separate subgroup analyses based on specific device models or software versions. Our review included both researcher-initiated and consumer-initiated studies to provide a comprehensive assessment of wearable activity trackers in real-world settings. Researcher-initiated studies typically involved smaller sample sizes and controlled participant recruitment, while consumer-initiated studies often had larger sample sizes and reflected more naturalistic use patterns. While this combination enhances the generalizability of our findings, it also introduces heterogeneity. We acknowledge this as a limitation and suggest that future research should consider these differences when interpreting results. Additionally, our review only identified studies in the domains of COVID-19, cardiovascular conditions, and falls as eligible. While laboratory-based studies are being conducted for event detection in other health domains (such as stress and respiratory conditions), our focus was intentionally on studies conducted in free-living conditions. This approach offers insights into the wearables' event detection capabilities in real-world settings, as opposed to artificial (eg, laboratory) conditions.

Clinical Implications

The use of wearable activity trackers for detecting medical events is an emerging field with both significant promise and challenges. Wearable activity trackers demonstrate comparable ability to detect COVID-19 and atrial fibrillation compared with other clinical tests such as lateral flow tests and electrocardiograms. However, wearables offer the additional advantage of continuous, real-time monitoring for conditions requiring constant surveillance. As such, they may empower patients to take a more proactive role in their health care by giving them immediate feedback and data about their condition. They may also contribute to improved surveillance and resource planning for health care systems, which could be particularly useful in times of epidemics or pandemics.

Certain wearable device features excel at detecting specific medical events. For COVID-19, devices combining heart rate monitors, skin temperature sensors, and accelerometers proved effective by detecting deviations from an individual's baseline across multiple physiological parameters. In contrast, for atrial fibrillation detection, Food and Drug Administration-approved devices relied on optical heart rate sensors providing photoplethysmography data, capable of identifying irregular heart rhythms characteristic of arrhythmias. Fall detection primarily uses accelerometer data, with wrist-worn placement crucial for sensing sudden deceleration and impact forces. However, false positives persist due to nonfall rapid movements. Looking ahead, integrating multiple sensors can enhance accuracy across various medical conditions. Yet, fundamental sensor limitations may remain. Aligning device capabilities with specific use cases and recognizing sensor shortcomings

will inform future research and benchmarking efforts amid evolving technology.

As consumer wearables gradually morph from being lifestyle tools to over-the-counter medical instruments, they present a range of challenges, including concerns about data privacy and security, which will require stringent protective measures. Furthermore, as wearable devices become increasingly sophisticated in detecting medical conditions, such as atrial fibrillation, they offer both benefits and pitfalls. On the positive side, these devices have the potential to identify asymptomatic atrial fibrillation episodes. This is enormously beneficial, since currently, stroke is the first manifestation in at least 25% of atrial fibrillation-related stroke cases [40]. Early detection could therefore lead to timely intervention and stroke prevention. However, health care professionals have reported an uptick in patient consultations triggered by atrial fibrillation alerts from wearables, resulting in a surge of medical tests, such as electrocardiograms, to confirm diagnoses [41]. While some clinicians see this as an advancement in patient-initiated health care, others question the necessity of such screening, particularly in patient subgroups where atrial fibrillation may have a relatively benign prognosis [42]. Moreover, the use of wearables can generate both false positives and negatives, potentially causing unnecessary anxiety, diagnostic tests, and treatments, or giving users a false sense of security.

Future Research

Our review reveals that the current peer-reviewed evidence base concerning the event detection capabilities of consumer wearable activity trackers in free-living conditions is limited to COVID-19, cardiac function, and falls. This was somewhat surprising, given the potential of these devices to diagnose numerous other conditions. Our findings indicate a significant gap in the current literature, which was not apparent in previous reviews that typically focused on specific conditions and did not highlight the lack of studies across a broader range of conditions. Considering the diverse array of sensors incorporated in modern wearable activity trackers, these devices offer considerable potential for detecting and monitoring medical events across an extensive spectrum of health conditions into the future. This may include respiratory conditions, neurological disorders, mental health, stress and fatigue, and even environmental and allergic reactions. This will require research across the product design continuum, from algorithm training to laboratory testing and free-living testing. This will be made all the more challenging by the rapid pace at which new devices and models are released into the market. In the future, our meta-analysis could be updated to provide insight into the accuracy of such diagnostics by condition, device, and population.

Conclusions

This study provides a comprehensive overview of the current state of evidence regarding the diagnostic capabilities of consumer wearable activity trackers in real-world settings. While the devices show promise in detecting conditions, such as COVID-19 and atrial fibrillation, with moderate to high sensitivity and specificity, their performance in detecting falls is moderate, highlighting the need for further refinement of detection algorithms. The existing literature is notably skewed toward COVID-19, leaving a significant gap in our understanding of how these devices can be used for a broader range of health issues. This gap, which was not apparent in previous reviews, underscores the necessity for future research to expand the scope of conditions studied. As wearable technology continues to evolve, it is crucial to address the challenges posed by false positives and negatives, data privacy, and security concerns. This will ensure that the rapid advancements in this field can be matched by robust scientific validation, enabling these devices to realize their full potential as tools for health care monitoring and intervention.

Acknowledgments

This project received no specific funding. CM is supported by a Medical Research Future Fund Emerging Leader Grant (GNT1193862).

Data Availability

All data generated or analyzed during this study are included in this published article and Multimedia Appendix 1.

Authors' Contributions

All authors contributed to the review protocol. BS, SC, AM, RC, DD, JB, TF, KS, CS, EE, IW, and CM designed the search strategy and selected studies. BS, AM, RC, DD, JB, TF, KS, CS, EE, and IW extracted the data. BS analyzed the data. BS, SC, and CM drafted the manuscript. All authors contributed to the drafting of the review. All authors revised the manuscript critically for important intellectual content. All authors approved the final version of the article. All authors had access to all the data in the study and could take responsibility for the integrity of the data and the accuracy of the data analysis. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Supplementary materials.
[DOCX File , 123 KB - mhealth_v12i1e56972_app1.docx]

Multimedia Appendix 2
Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) checklist.
[PDF File (Adobe PDF File), 46 KB - mhealth_v12i1e56972_app2.pdf]

References

1. Watson SI, Sahota H, Taylor CA, Chen YF, Lilford RJ. Cost-effectiveness of health care service delivery interventions in low and middle income countries: a systematic review. Glob Health Res Policy 2018;3:17 [FREE Full text] [doi: 10.1186/s41256-018-0073-z] [Medline: 29930989]
2. Wang W, Yan Y, Guo Z, Hou H, Garcia M, Tan X, et al. All around suboptimal health—a joint position paper of the suboptimal health study consortium and European association for predictive, preventive and personalised medicine. EPMA J 2021;12(4):403-433 [FREE Full text] [doi: 10.1007/s13167-021-00253-2] [Medline: 34539937]
3. Natalucci V, Marmondi F, Biraghi M, Bonato M. The effectiveness of wearable devices in non-communicable diseases to manage physical activity and nutrition: where we are? Nutrients 2023;15(4):913 [FREE Full text] [doi: 10.3390/nu15040913] [Medline: 36839271]
4. Shin G, Jarrahi MH, Fei Y, Karami A, Gafinowitz N, Byun A, et al. Wearable activity trackers, accuracy, adoption, acceptance and health impact: a systematic literature review. J Biomed Inform 2019;93:103153 [FREE Full text] [doi: 10.1016/j.jbi.2019.103153] [Medline: 30910623]
5. Kooiman TJM, Dontje ML, Sprenger SR, Krijnen WP, van der Schans CP, de Groot M. Reliability and validity of ten consumer activity trackers. BMC Sports Sci Med Rehabil 2015;7:24 [FREE Full text] [doi: 10.1186/s13102-015-0018-5] [Medline: 26464801]
6. Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. Int J Behav Nutr Phys Act 2015;12:159 [FREE Full text] [doi: 10.1186/s12966-015-0314-1] [Medline: 26684758]
7. Ferguson T, Olds T, Curtis R, Blake H, Crozier AJ, Dankiw K, et al. Effectiveness of wearable activity trackers to increase physical activity and improve health: a systematic review of systematic reviews and meta-analyses. Lancet Digit Health 2022;4(8):e615-e626 [FREE Full text] [doi: 10.1016/S2589-7500(22)00111-X] [Medline: 35868813]
8. Tang MSS, Moore K, McGavigan A, Clark RA, Ganesan AN. Effectiveness of wearable trackers on physical activity in healthy adults: systematic review and meta-analysis of randomized controlled trials. JMIR Mhealth Uhealth 2020;8(7):e15576 [FREE Full text] [doi: 10.2196/15576] [Medline: 32706685]
9. Szeto K, Arnold J, Singh B, Gower B, Simpson CEM, Maher C. Interventions using wearable activity trackers to improve patient physical activity and other outcomes in adults who are hospitalized: a systematic review and meta-analysis. JAMA Netw Open 2023;6(6):e2318478 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.18478] [Medline: 37318806]
10. Laranjo L, Ding D, Heleno B, Kocaballi B, Quiroz JC, Tong HL, et al. Do smartphone applications and activity trackers increase physical activity in adults? Systematic review, meta-analysis and metaregression. Br J Sports Med 2021;55(8):422-432. [doi: 10.1136/bjsports-2020-102892] [Medline: 33355160]
11. Singh B, Zopf EM, Howden EJ. Effect and feasibility of wearable physical activity trackers and pedometers for increasing physical activity and improving health outcomes in cancer survivors: a systematic review and meta-analysis. J Sport Health Sci 2022;11(2):184-193 [FREE Full text] [doi: 10.1016/j.jshs.2021.07.008] [Medline: 34314878]
12. Davergne T, Pallot A, Dechartres A, Fautrel B, Gossec L. Use of wearable activity trackers to improve physical activity behavior in patients with rheumatic and musculoskeletal diseases: a systematic review and meta-analysis. Arthritis Care Res (Hoboken) 2019;71(6):758-767. [doi: 10.1002/acr.23752] [Medline: 30221489]
13. Bayoumy K, Gaber M, Elshafeey A, Mhaimeed O, Dineen EH, Marvel FA, et al. Smart wearable devices in cardiovascular care: where we are and how to move forward. Nat Rev Cardiol 2021;18(8):581-599 [FREE Full text] [doi: 10.1038/s41569-021-00522-7] [Medline: 33664502]
14. Shei RJ, Holder IG, Oumsang AS, Paris BA, Paris HL. Wearable activity trackers-advanced technology or advanced marketing? Eur J Appl Physiol 2022;122(9):1975-1990 [FREE Full text] [doi: 10.1007/s00421-022-04951-1] [Medline: 35445837]
15. Albán-Cadena AC, Villalba-Meneses F, Pila-Varela KO, Moreno-Calvo A, Villalba-Meneses CP, Almeida-Galárraga DA. Wearable sensors in the diagnosis and study of Parkinson's disease symptoms: a systematic review. J Med Eng Technol 2021;45(7):532-545. [doi: 10.1080/03091902.2021.1922528] [Medline: 34060967]
16. Pay L, Yumurtaş A, Satti DI, Hui JMH, Chan JSK, Mahalwar G, et al. Arrhythmias beyond atrial fibrillation detection using smartwatches: a systematic review. Anatol J Cardiol 2023;27(3):126-131 [FREE Full text] [doi: 10.14744/AnatolJCardiol.2023.2799] [Medline: 36856589]
17. Huang JD, Wang J, Ramsey E, Leavey G, Chico TJA, Condell J. Applying artificial intelligence to wearable sensor data to diagnose and predict cardiovascular disease: a review. Sensors (Basel) 2022;22(20):8002 [FREE Full text] [doi: 10.3390/s22208002] [Medline: 36298352]

18. Lee S, Chu Y, Ryu J, Park YJ, Yang S, Koh SB. Artificial intelligence for detection of cardiovascular-related diseases from wearable devices: a systematic review and meta-analysis. Yonsei Med J 2022;63(Suppl):S93-S107 [FREE Full text] [doi: 10.3349/ymj.2022.63.S93] [Medline: 35040610]
19. Cheong SHR, Ng YJX, Lau Y, Lau ST. Wearable technology for early detection of COVID-19: a systematic scoping review. Prev Med 2022;162:107170 [FREE Full text] [doi: 10.1016/j.ypmed.2022.107170] [Medline: 35878707]
20. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
21. Mackinnon A. A spreadsheet for the calculation of comprehensive statistics for the assessment of diagnostic tests and inter-rater agreement. Comput Biol Med 2000;30(3):127-134. [doi: 10.1016/s0010-4825(00)00006-8] [Medline: 10758228]
22. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al. Cochrane handbook for systematic reviews of interventions. Cochrane. 2023. URL: http://www.training.cochrane.org/handbook [accessed 2024-08-06]
23. Oxford Centre for Evidence-Based Medicine: levels of evidence (March 2009).: University of Oxford; 2022. URL: https://www.cebm.ox.ac.uk/resources/levels-of-evidence/oxford-centre-for-evidence-based-medicine-levels-of-evidence-march-2009 [accessed 2024-08-06]
24. Johansson D, Ohlsson F, Krýsl D, Rydenhag B, Czarnecki M, Gustafsson N, et al. Tonic-clonic seizure detection using accelerometry-based wearable sensors: a prospective, video-EEG controlled study. Seizure 2019;65:48-54 [FREE Full text] [doi: 10.1016/j.seizure.2018.12.024] [Medline: 30611010]
25. Roberts DM, Schade MM, Mathew GM, Gartenberg D, Buxton OM. Detecting sleep using heart rate and motion data from multisensor consumer-grade wearables, relative to wrist actigraphy and polysomnography. Sleep 2020;43(7):zsaa045 [FREE Full text] [doi: 10.1093/sleep/zsaa045] [Medline: 32215550]
26. Kusmakar S, Karmakar CK, Yan B, Obrien T, Muthuganapathy R, Palaniswami M. Automated detection of convulsive seizures using a wearable accelerometer device. IEEE Trans Biomed Eng 2019;66(2):421-432. [doi: 10.1109/TBME.2018.2845865] [Medline: 29993501]
27. Ahlrichs C, Samà A, Lawo M, Cabestany J, Rodríguez-Martín D, Pérez-López C, et al. Detecting freezing of gait with a tri-axial accelerometer in Parkinson's disease patients. Med Biol Eng Comput 2016;54(1):223-233. [doi: 10.1007/s11517-015-1395-3] [Medline: 26429349]
28. Mistry DA, Wang JY, Moeser ME, Starkey T, Lee LYW. A systematic review of the sensitivity and specificity of lateral flow devices in the detection of SARS-CoV-2. BMC Infect Dis 2021;21(1):828 [FREE Full text] [doi: 10.1186/s12879-021-06528-3] [Medline: 34407759]
29. UK Government. Technologies validation group: using tests to detect COVID-19. 2021. URL: https://www.gov.uk/guidance/technologies-validation-group-using-tests-to-detect-covid-19 [accessed 2021-10-19]
30. Welton NJ, McAleenan A, Thom HH, Davies P, Hollingworth W, Higgins JP, et al. Screening strategies for atrial fibrillation: a systematic review and cost-effectiveness analysis. Health Technol Assess 2017;21(29):1-236 [FREE Full text] [doi: 10.3310/hta21290] [Medline: 28629510]
31. Kvist LM, Vinter N, Urbonaviciene G, Lindholt JS, Diederichsen ACP, Frost L. Diagnostic accuracies of screening for atrial fibrillation by cardiac nurses versus radiographers. Open Heart 2019;6(1):e000942 [FREE Full text] [doi: 10.1136/openhrt-2018-000942] [Medline: 30997131]
32. Google. New Fitbit feature makes AFib detection more accessible. Google Blog. 2022. URL: https://blog.google/products/fitbit/irregular-heart-rhythm-notifications/ [accessed 2023-10-03]
33. Kritz F. Apple Watches have two new FDA-cleared health applications. 2022. URL: https://array.aami.org/content/news/apple-watches-have-two-new-fda-cleared-health-applications [accessed 2023-10-03]
34. Malik A. Garmin launches a new FDA-cleared ECG app for the Venu 2 Plus. 2023. URL: https://tinyurl.com/57y86ccf [accessed 2023-10-03]
35. Samsung announces FDA-cleared irregular heart rhythm notification for Galaxy Watch. Samsung Newsroom US. 2023. URL: https://news.samsung.com/us/fda-cleared-irregular-heart-rhythm-notification-for-galaxy-watch/2023 [accessed 2023-10-03]
36. Natarajan A, Su HW, Heneghan C. Assessment of physiological signs associated with COVID-19 measured using wearable devices. NPJ Digit Med 2020;3(1):156 [FREE Full text] [doi: 10.1038/s41746-020-00363-7] [Medline: 33299095]
37. Hassantabar S, Stefano N, Ghanakota V, Ferrari A, Nicola GN, Bruno R, et al. CovidDeep: SARS-CoV-2/COVID-19 test based on wearable medical sensors and efficient neural networks. IEEE Trans Consumer Electron 2021;67(4):244-256. [doi: 10.1109/tce.2021.3130228]
38. Chen M, Wang H, Yu L, Yeung EHK, Luo J, Tsui K, et al. A systematic review of wearable sensor-based technologies for fall risk assessment in older adults. Sensors (Basel) 2022;22(18):6752 [FREE Full text] [doi: 10.3390/s22186752] [Medline: 36146103]
39. Warrington DJ, Shortis EJ, Whittaker PJ. Are wearable devices effective for preventing and detecting falls: an umbrella review (a review of systematic reviews). BMC Public Health 2021;21(1):2091 [FREE Full text] [doi: 10.1186/s12889-021-12169-7] [Medline: 34775947]
40. Hindricks G, Potpara T, Dagres N, Arbelo E, Bax JJ, Blomström-Lundqvist C, et al. 2020 ESC guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery

(EACTS): the task force for the diagnosis and management of atrial fibrillation of the European Society of Cardiology (ESC) developed with the special contribution of the European Heart Rhythm Association (EHRA) of the ESC. Eur Heart J 2021;42(5):373-498. [doi: 10.1093/eurheartj/ehaa612] [Medline: 32860505]

41. Ding EY, Svennberg E, Wurster C, Duncker D, Manninger M, Lubitz SA, et al. Survey of current perspectives on consumer-available digital health devices for detecting atrial fibrillation. Cardiovasc Digit Health J 2020;1(1):21-29 [FREE Full text] [doi: 10.1016/j.cvdhj.2020.06.002] [Medline: 32924024]
42. Brandes A, Stavrakis S, Freedman B, Antoniou S, Boriani G, Camm AJ, et al. Consumer-led screening for atrial fibrillation: frontier review of the AF-SCREEN international collaboration. Circulation 2022;146(19):1461-1474 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.121.058911] [Medline: 36343103]

Abbreviations

AUC: area under the curve
JBI: Joanna Briggs Institute
LFD: lateral flow device
MeSH: Medical Subject Headings
PPV: positive predictive value
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by L Buis; submitted 31.01.24; peer-reviewed by Z Li, T Yano, J Claggett, F Keusch; comments to author 22.04.24; revised version received 03.05.24; accepted 26.06.24; published 30.08.24.

Please cite as:
Singh B, Chastin S, Miatke A, Curtis R, Dumuid D, Brinsley J, Ferguson T, Szeto K, Simpson C, Eglitis E, Willems I, Maher C
Real-World Accuracy of Wearable Activity Trackers for Detecting Medical Conditions: Systematic Review and Meta-Analysis
JMIR Mhealth Uhealth 2024;12:e56972
URL: https://mhealth.jmir.org/2024/1/e56972
doi:10.2196/56972
PMID:

Review

The Use of Noncommercial Parent-Focused mHealth Interventions for Behavioral Problems in Youth: Systematic Review

Katherine I Magnuson[1], PsyD; Kexin Li[1], PsyD; Grace Beuley[1], MSCP; Stacy R Ryan-Pettes[1], PhD

Department of Psychology and Neuroscience, Baylor University, Waco, TX, United States

Corresponding Author:
Katherine I Magnuson, PsyD
Department of Psychology and Neuroscience
Baylor University
Baylor Sciences Building
101 Bagby Ave
Waco, TX, 76706
United States
Phone: 1 (254) 710 2961
Email: katherine.magnuson@childrens.com

Abstract

Background: The rates of substance use among adolescents are alarmingly high, and current treatment options lack integration of parent-focused interventions, despite evidence that effective parenting practices can mediate treatment outcomes for adolescents involved in substance use. Accessibility and other barriers to parental interventions may be mitigated through mobile health (mHealth); however, few mHealth platforms target substance use behaviors for adolescents through the implementation of behavioral parent training strategies.

Objective: This study seeks to review current mHealth platforms within empirical literature that are designed to increase effective parenting through behavioral parent training techniques. Because of the paucity of mHealth modalities that use parenting strategies to target substance use in adolescents, the objective was expanded to include mHealth platforms addressing behavior problems among youth, given that parent-targeted treatments for these clinical presentations overlap with those for adolescent substance use. Overall, the systematic review was conducted to inform the development of mHealth apps for parents of youth involved in substance use, improve accessibility, and better align with parental needs.

Methods: This systematic review was conducted using the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) method to select relevant articles across several databases. Each study was assessed for relevance and inclusion. Each study was reviewed for demographics, delivery medium, intervention status as stand-alone treatment or as an enhancement to treatment, mobile device used, mental health condition targeted, intervention type, underlying intervention theory, behavior change theory applied in design, behavior change techniques, parent training techniques, youth outcomes, parent outcomes, visual design, content, and features.

Results: Overall, 11 studies were included. Nearly all studies (9/11, 82%) predominantly sampled female caregivers. Most of the studies (6/11, 55%) integrated social learning theory. Only a few of the studies (2/11, 18%) discussed the embedded behavior change theories, whereas all the studies (11/11, 100%) used at least one behavior change technique to encourage change in parental behaviors. Many of the studies (7/11, 64%) tailored design features to the end user. Of the various behavioral parent training techniques, nearly all studies (10/11, 91%) included the skill of strengthening the parent-child relationship. A preliminary evaluation of treatment outcomes suggests a positive impact of parent-targeted mHealth interventions. When reported, the effect sizes for treatment ranged from Cohen *d*=0.38 to Cohen *d*=1.58 for youth and from Cohen *d*=0.13 to Cohen *d*=2.59 for parents.

Conclusions: Although features and techniques were referenced, only a few of the studies provided specific information related to behavior change theory (2/11, 18%), visual design (2/11, 18%), and the translation of parent-targeted interventions to mHealth platforms. Such information would be useful for the development of mHealth apps. Preliminary outcomes for youth and parents are encouraging, but future studies should consider conducting a meta-analysis as the body of studies grows to determine aggregate statistical findings.

(JMIR Mhealth Uhealth 2024;12:e51273) doi:10.2196/51273

KEYWORDS

behavioral parent training; mobile health; mHealth; mobile app; adolescent; substance use; child mental health condition; mobile phone

Introduction

Background

Adolescent substance use occurs at alarming rates in the United States, with approximately 4.3 million youths using illicit substances in 2019 [1]. Despite evidence indicating that 1.1 million of these youths needed substance use treatment, <1% obtained treatment [1]. For the few youth who receive substance use treatment, parent-focused interventions, shown to improve parenting practices that mediate adolescent outcomes, are often a missing component [2-4]. This is concerning because there are limited resources and pathways of access for parents of adolescents involved in substance use to receive parenting resources or support [5-7].

This inability to access parent-focused interventions may be related to both a lack of availability of these interventions [8,9] and logistical, personal, or systemic barriers to treatment engagement [10-15]. Nonetheless, less frequently acknowledged is that the currently available treatment options for parents of adolescents involved in substance use may not embody the type of treatment that these parents desire. Recent research showed that, among parents of youth in treatment for substance use, the majority (72%) perceived a need for parent-focused services related to parenting their adolescent child after substance use treatment; when aftercare was offered via mobile phone, this figure increased to 91% [16]. One interpretation of these findings is that parents are not currently receiving support through their preferred medium.

Taken together, these findings highlight the need for greater access to strategies for engaging in effective parenting of children with a history of substance use, and leveraging mobile health (mHealth) may help address this service gap. Unfortunately, while the development of mHealth apps is moving at a rapid pace in most fields of health care, it lags in the area of adolescent substance use [17]. There is only 1 published study of an mHealth app for parents of youth involved with substances [18]. However, this app focuses on delivering mindfulness interventions and excludes a focus on behavioral parent training. Given the demonstrated benefits of behavioral parenting approaches in curtailing adolescent substance use [4,19] and the potential advantages that mHealth apps offer in broadening access and reach, it is surprising that more attention has not been paid to developing an mHealth intervention specifically for parents of youth involved in substance use.

This systematic review seeks to evaluate mHealth apps in empirical literature designed to increase effective parenting through behavioral parent training techniques for behavior problems in their child. Given the overlap in behavioral parent training interventions for behavior problems and substance use [20-23], the results of this review could inform the development of future parent-focused mHealth apps for parents of youth involved in substance use, improving accessibility and matching parental desires for treatment mediums [24,25].

Behavioral Parenting Practices and Adolescent Substance Use

Parenting practices shape the development and outcome of a child [26]. The literature is replete with results showing that ineffective parenting practices such as poor monitoring and supervision, inconsistent discipline, poor limit setting, and low positive parenting are associated with a range of behavior problems [27], including substance use disorders [28,29]. On the basis of the plethora of research demonstrating the importance of effective parenting practices, evidence-based treatments designed to treat behavior problems among adolescents, including substance use, are heavily steeped in addressing ineffective parenting using behavioral parent training [20,22]. Broadly, behavioral parent training is an evidence-based approach to helping parents apply behavioral strategies to improve their child's behavior and increase positive family interactions; it is also referred to as parent management training and parenting training [30].

mHealth Apps for Behavioral Parent Training

After conducting a literature review, Jones et al [31] concluded that behavioral parent training is a strong fit for transfer to technological mediums such as smartphone apps. We concur and argue that behavioral parent training is compatible for translation to mHealth because key parenting strategies in the behavioral parent training protocols can be aided with smartphone apps that include design features tapping into general principles of behavior change to promote parenting behaviors. Specifically, app features such as routine prompts and timely notifications with tips may promote consistent implementation of rules, facilitate limit setting, and support the use of consistent discipline. In fact, prompting through push notifications aligns with behavior change theories in mHealth that emphasize the use of reminders to enact skills and the integration of motivational support [32,33] to foster the consistent use of learned, effective parenting practices; for example, the use of encouragement may include periodic messages that remind parents of a learned parenting skill and encourage them to continue using the skill.

Objectives

The original aim of this study was to systematically review available noncommercial mHealth apps for parents of youth involved in substance use. However, the limited literature on mHealth apps providing parental intervention to target adolescent substance use made this aim challenging. In an effort to continue to explore and review this subject despite the scarcity of research, the objective of this study expanded. In particular, the focus shifted slightly to a systematic review of mHealth apps that provide behavioral parent training or components of behavioral parent training to enhance the use of effective parenting for behavior problems in youth. This shift broadened the search for current mHealth apps in the literature, while also maintaining relevance and applicability to parent-targeted mHealth interventions for adolescent substance use. More

specifically, the behavioral parenting interventions that have been implemented to target youth behavior problems significantly overlap with those used to intervene on adolescent substance use. The large overlap in treatment content may be related to the notion that adolescent substance use is often conceptualized through a broader lens of behavioral problems in youth [20-23]. Therefore, the results of this review could generalize to the development of future parent-focused mHealth apps for parents of youth involved in substance use [24,25].

This study sought to answer four main research questions:

1. What are the general characteristics of behavioral parent training apps under development?
2. What is the empirical evidence underlying behavioral parent training apps under development?
3. What are the main parenting strategies covered in behavioral parent training apps under development?
4. What implications do the characteristics, empirical evidence, and parenting strategies evidenced in current behavioral parent training apps have on the development of an mHealth app for parents of youth with behavior problems involved in substance use?

To answer these questions, we summarized the major design elements, features, content, and theoretical foundations of the evaluated apps and paralleled these with the components of substance use treatment to provide recommendations for the design of mHealth apps for parents of adolescents who use substances. In contrast to existing studies, these objectives enhance knowledge about apps tailored specifically for parents of adolescents who use substances.

While prior studies have reviewed mHealth apps based on behavior change theory and techniques, they included a narrow focus on these factors [33-35] and did not review behavioral parenting practices. Some studies have reviewed commercial parenting apps [17,36], apps for specific groups of parents (eg, fathers with low-income status and new parents [36-38]), or apps for parents with adult children [39]. However, these studies did not review apps designed to teach behavioral parenting skills to address behavior problems exhibited by their child. Indeed, several reviews provide information about the effectiveness of technology-based interventions for behavior change and for parents of children with emotional or behavioral issues [40-42]. However, these reviews included a mix of dated mediums alongside mHealth apps, including websites, software, videoconferencing services, and SMS text messaging.

Methods

Literature Search

The search was conducted electronically in English between June and September 2019, again in March 2021, and once more in October 2022. No restrictions on the date or year of article publication were imposed in the original 2019 search, and the 2021 and 2022 searches were limited to materials published in the time since the prior searches. The following databases were used: PsycINFO, MEDLINE (PubMed), Google Scholar, Scopus, Web of Science, and WorldCat. References from selected articles and past literature review articles were also examined to identify potential sources that may have met our criteria for this review [39,40,43,44].

The following mobile technology search terms were used: *mobile phone*, *mHealth*, *eHealth*, *SMS*, *text messaging*, *mobile application*, *tablet*, *smartphone*, and *cell phone*. The following parent treatment search terms were used: *parent training*, *intervention*, *treatment*, *parent management training*, *parent-child interaction*, and *behavioral training*. The following mental health search terms were used: *behavior*, *attention-deficit/hyperactivity disorder*, *autism spectrum disorder*, *posttraumatic stress disorder*, *trauma*, *psychological*, and *disorders*. These terms were entered into databases using various search combinations, including *(mobile phone OR cell phone OR smartphone OR tablet) AND (parent train* OR treat* OR parent management train*) AND (behav* OR trauma OR disorder OR attent* OR psycholog* OR autism) AND (SMS* OR text messag* OR application OR mHealth OR eHealth)*.

The search conducted in October 2022 to update the results used the original search terms with date restricted to the years since the search conducted in March 2021 (ie, 2021-2022). In the search update, searches in 3 databases were modified to limit the number of results for relevance. Specifically, in MEDLINE (PubMed), the search was limited to clinical trials and randomized controlled trials; in Scopus, it was limited to articles; and in Web of Science, additional search criteria—*adol* OR child* OR parent* OR caregiver OR mother OR father OR youth*—were applied to filter out irrelevant results.

Study Selection and Eligibility Criteria

Due to the paucity of studies in this field, the titles and abstracts identified from the search process included both peer-reviewed feasibility or acceptability articles and conference proceedings. Articles were screened against predefined inclusion criteria (Textbox 1) by 3 reviewers (SRR-P, KIM, and KL), who independently conducted the search and met afterward to integrate the search results and make joint decisions about inclusion and exclusion for each record.

Textbox 1. Criteria for eligibility.

Inclusion criteria

- The study investigated parent-targeted interventions to influence child mental health conditions (defined as the presence of adverse behavioral and emotional symptoms that may be contributing to psychological difficulties). These conditions may include disruptive behaviors and conduct disorder symptoms, substance use, attention-deficit/hyperactivity disorder symptoms, trauma symptoms, and autism spectrum disorder symptoms; however, developmental, language, speech, and motor delays were excluded because these may not always be directly related to psychological symptoms.
- The study provided data on the efficacy or effectiveness of the intervention.
- The study provided data on either parent or child outcomes.
- The interventions only used mobile or tablet devices (studies were excluded if they involved the use of websites or computers in any capacity).
- The intervention content, such as specific parenting skills, was delivered via SMS text messaging or mobile apps (as opposed to professionals delivering interventions via mobile devices).
- Either stand-alone treatments or enhancements to existing treatments were included if the intervention involved more than simple reminders to attend regular treatment, based on the rationale that even enhanced treatment components may serve as stand-alone interventions with further research development.
- Studies involving biological parents, nonbiological parents, and foster caregivers were included.
- Studies that involved interventions targeting parents of children ranging in age from 2 to 18 years were included, based on the rationale that regardless of differences in implementations depending on the age of the child the basic principles of certain effective parenting practices (eg, parental monitoring) remain consistent.
- The articles or conference papers were in English.

Identification and Description of Study Characteristics

Study Characteristics Assessed

Each article selected for the review was assessed for various characteristics, including demographics, delivery medium, intervention status as stand-alone treatment or as an enhancement to treatment, mobile device used, mental health condition targeted, intervention type, underlying intervention theory, behavior change theory applied in design, behavior change techniques, parent training techniques, youth outcomes, parent outcomes, visual design, content, and features. Each of these characteristics was operationalized according to this review's context (Textbox 2).

Textbox 2. Operationalization of the characteristics and features reviewed.

Characteristics and operationalization

- Delivery medium: the method used to deliver the intervention on the mobile device, which included the use of a mobile app, electronic monitoring wristbands, and the use of smartphone or tablet features such as SMS texting, video calls, and video recordings
- Stand-alone treatment: the intervention is administered solely via the mobile device without being administered alongside, or in conjunction with, in-person treatment
- Enhancement to treatment: the intervention is administered in person, and the mobile device is used as a supplemental feature of treatment
- Mobile device used: the type of mobile device used to deliver the intervention
- Mental health condition targeted: the adverse behavioral or emotional symptoms exhibited by the children of the population of parents studied
- Type of intervention used, incorporated, or adapted: the parent-targeted intervention used in the research study that can be fully used, shortened, selectively used, or adapted, with the primary skills being implemented
- Underlying intervention theory: the theoretical foundation of the parent-targeted intervention
- Behavior change theory applied in design: a method for understanding how variations in treatments or interventions can lead to changes in behavior (Hekler et al [45])
- Behavior change techniques: a range of 26 methods used in the design of the mobile intervention to change an individual's behavior [46]; the definitions of these 26 techniques can be found in the taxonomy developed by Abraham and Michie [46]
- Youth outcomes: changes in youth symptoms or behaviors after parent-targeted intervention is administered
- Parent outcomes: changes in parent behaviors after parent-targeted intervention is administered
- Visual design: assessment of the visual quality and look and feel of the program, including aesthetics, layout, and size [43]
- Content: assessment of the material provided and learned in the program, including the use of evidence-based content, quality of information provision, completeness and conciseness, and clarity about the program's purpose [43]
- Features: assessment of different aspects used in the design of the mobile intervention

These characteristics were first assessed through careful reading of each article by the lead author (KIM). If the relevant elements could not be identified through reading the article, the references of the article were reviewed to determine whether they were included in the preliminary work. The authors of 2 (18%) of the 11 studies were contacted to inquire whether further research surrounding the initial study had been conducted. One author responded to the inquiry. To identify behavior change techniques, visual design qualities, content, and theoretical foundations, the methodologies outlined in the following subsections were used.

Identification of Behavior Change Techniques

Interventions were evaluated for the types and number of behavior change techniques using the taxonomy of behavior change techniques developed by Abraham and Michie [46].

Assessment of Visual Design and Content

Visual design and content were evaluated using Enlight (MindTools.io), a 5-point rating system ranging from 1=*very poor* to 5=*very good*, developed for the assessment of eHealth interventions [43].

Identification and Assessment of Theoretical Foundations

Theoretical foundations of the treatments and mobile apps were assessed through the implementation of a theory coding scheme [47]. This coding scheme outlines various steps for classifying the presence of the use of theory in interventions [47]. In this review, theoretical foundations were coded as present based on their mention in the article or its references (ie, item 1 [47]). When referenced or mentioned by the study, the presence of a theoretical foundation was coded. When assessing behavior change theory in the design of the mobile interventions, it was noted when a theory was not specifically mentioned but only alluded to in the study. Specifically, when a study mentioned the use of theory without specifying the name of the theory or its characteristics, it was marked accordingly.

Results

Demographic Information and Designs of Reviewed Studies

The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) method was used to conduct the systematic search [48,49] (Figure 1), which included the PRISMA checklist that can be found in Multimedia Appendix 1 [50]. Through this process, a total of 11 studies were included in this review, and each was assessed for demographic information (Table 1). The earliest studies reviewed were published in 2014 [51,52]. Of the 11 studies, 7 (64%) were randomized controlled trials. The sample sizes ranged from 10 to 371 participants. Most of the parents included were mothers (ranging from 77% to 100%); however, in the study by May et al [50], the intervention was delivered to fathers. The parental ages ranged from 18 to ≥50 years. The target children's ages ranged from 2 to 18 years. Each study recruited participants from a range of settings, including primary care clinics (1/11, 9%), community health agencies (4/11, 36%), social services (2/11, 18%), juvenile justice centers (1/11, 9%), early education agencies (1/11, 9%), community parenting support groups (2/11, 18%), autism organizations and intervention centers (1/11, 9%), child psychiatrist (1/11, 9%), schools (1/11, 9%), and social media platforms (3/11, 27%). Families were included if the child had "externalizing behavior problems," "disruptive behaviors," "symptoms of conduct disorder," "autism spectrum disorder," or "attention-deficit/hyperactivity disorder."

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

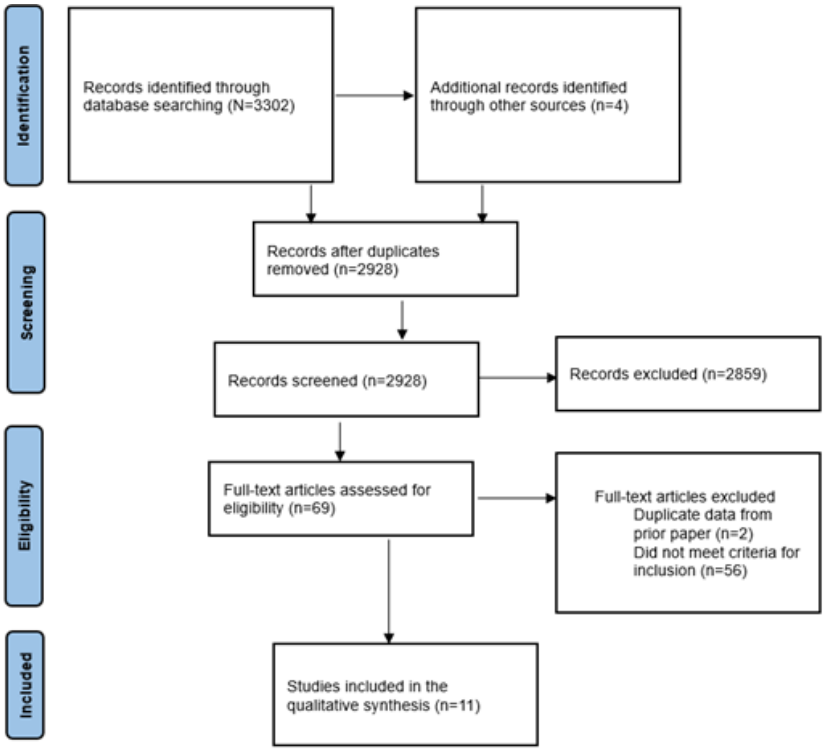


Table 1. Demographic information of reviewed studies.

Study, year	Caregiver identification[a] (%)	Child sex (%)	Parent age, y (%)	Child age (%)	Recruitment settings	Parental race or ethnicity (%)	Socioeconomic status (annual income; US $), (%)	Caregiver composition (%)	Treatment length
Breitenstein et al [53], 2016	Mother (94.9)	Female (57)	30-39 (63.3)	2-5 y (NR[b])	Primary care clinic	African American (64.6)	>20,000 (65.8)	Not married (60.8)	12 wk
Feil et al [54], 2018	Female (77)	NR	Mean 44.7, SD 10.08 (NR)	8-12 y (NR)	Community parenting groups and social media	White (89)	>25,000 (29); 25,000-50,000 (32)	Two-adult household (63)	4 wk
Hemdi and Daley [55], 2016	Mother (100)	NR	Mean 32.9 (NR)	Mean 63.18 mo (NR)	Autism organizations and intervention centers	NR	NR	Married (90.62)	4 sessions
Jones et al [51], 2014	Female (71)	Male (57)	Mean 35 y (NR)	Mean 5.57 y (NR)	Schools and community health agencies	Ethnic minority (57)	NR (low income[c]; 100)	Single (57)	8-12 sessions
Lefever et al [56], 2017	Mother (100)	Male (56)	Mean 28.91 y (NR)	Mean 4.56 y (NR)	Social service agencies, early education agencies, and community health agencies	Hispanic (46); African American (33)	Mean 18,608, SD 15,835	NR	8 sessions
Mason et al [57], 2021	Female (90.4)	Female (67)	Mean 45.6 y (NR)	Mean 15.2 y (NR)	Community health agencies	White (84.6)	NR (low income[d]; 100)	NR	4 wk
May et al [50], 2021	Father (100)	NR	Mean 42 y (NR)	4-11 y (78)	Community parenting groups and social media	NR	NR (financial difficulty; 39)	Two-parent household (71)	16 wk
Pina et al [52], 2014	Mother (80)	NR	Mean 38.4 y (NR)	NR (K-12[e])	NR	NR	NR	Two-adult household (100)	2 wk
Schaeffer et al [58], 2022	Female (100)	Male (55.9)	Mean 39.4 y (NR)	Mean 14.6 y (NR)	Social media and juvenile justice centers	White (76.5); Hispanic or Latinx (14.7)	10,001-20,000 (17.6); 20,001-30,000 (14.7); 50,001-60,000 (14.7); ≥60,000 (26.5)	Sole adult household (44.1); 2-parent household (29.4)	12 wk
Sonne et al [59], 2016	NR	Male (69.2)	NR	Mean 9.3 y (NR)	Community health agencies and child psychiatrist	NR	NR	NR	4 wk
Sullivan et al [60], 2019	Mother (95)	Male (55)	Mean 50 y (NR)	Mean 8.9 y (NR)	Social service agencies	White (95)	NR	Married (50); single (25)	10 wk

[a]Caregiver identification aligns with the report in the respective articles; identification as a mother or father should not assume gender.

[b]NR: not reported.

[c]Jones et al [51] define low income as the "adjusted gross income did not exceed 150% of the federal poverty limit, which takes into account both income and number of residents in the home."

[d]Mason et al [57] did not provide specific financial ranges; however, the sample was described as "low income."

[e]Pina et al [52] did not provide the ages of the children but specified that they were in grades K-12.

Of the 11 studies, 4 (36%) did not report the race or ethnicity of the sample. Among studies that reported race or ethnicity, the majority of the participants (ranging from 79% to 95%) were either from ethnic minority families (3/11, 27%) [51,53,56] or White (4/11, 36%) [54,57,58,60]. Information pertaining to socioeconomic status followed a similar pattern. Of the 11 studies, 5 (45%) included families who were identified as coming from a lower socioeconomic background [51,53,56-58],

2 (18%) included participants experiencing financial stress [50,54], and 4 (36%) either did not report income [52,55,59] or did not provide any socioeconomic information [60].

Of the 8 studies that reported the marital or partnership status of the parents, caregivers, or legal guardians, 5 (62%) reported that the majority (ranging from 50% to 90.6%) of the participants came from a 2-parent household (married couple or cohabiting couple).

Treatments were implemented for time periods ranging from 2 to approximately 16 weeks. Of the 11 interventions, 4 (36%) served as an enhancement to treatment, and 7 (64%) served as stand-alone treatments. Most of the studies (7/11, 64%) used a mobile app as a delivery medium.

Theoretical Frameworks

Transferring treatment to a digital platform requires consideration of both the intervention's theoretical framework and the theoretical frameworks that promote behavior change within a mobile platform.

Theoretical Framework of the Interventions

All studies (11/11, 100%) identified for this review drew from empirically based or evidence-based parent management training curricula, including the Chicago Parent Program, Behavioral Parent Training, Multisystemic Therapy, Behavioral Model Training, Helping the Noncompliant Child, and The Incredible Years Program. It is understood that the most prominent theoretical frameworks for these treatments include behaviorism (operant principles), the ecological systems framework developed by Bronfenbrenner [61], social learning theory, and the coercion model.

Of the 11 studies, 6 (55%) [51,53,56,57,59,60] explicitly discussed social learning theory, 2 (18%) [51,53] discussed the coercion model, 1 (9%) [58] discussed the social ecological framework developed by Bronfenbrenner [61], and 1 (9%) [55] discussed the transactional model of stress as the main theoretical framework for the parenting interventions included in the apps. Of the 11 studies, 4 (36%) did not expressly mention the coercion model as a theoretical framework, but this model was implied through information related to app content [52,57-59]; for example, these studies described implementing skills such as parental communication, effective parent-child interaction, monitoring, and limit setting to improve parental responses to child behaviors considered problematic, which is a tenet central to the coercion model [62]. Of the 11 studies, 2 (18%) did not expressly mention the guiding theoretical framework, and nor did they provide enough information to make inferences about the theoretical framework [50,54]. However, these studies did reference several behavioral parent training skills that are drawn from multiple interventions (eg, parent-child interaction therapy and parent management training).

Across all studies, when provided, there was a general mention of the theoretical framework for the interventions. While all studies (11/11, 100%) named the parenting skills used, only a few (4/11, 36%) provided comprehensive and specific information about specific parenting strategies (eg, examples of the language used or a description of the applications of skills to in vivo situations) [51,52,57,58]. These studies discuss examples of the applications of individual parenting skills to daily life situations [51,58] or specific language used to deliver the skill [52,57].

Theoretical Frameworks Promoting Behavior Change Within a Mobile Platform

Only 2 (18%) of the 11 studies described the use of behavior change theories in the design of the mobile intervention for parents [51,53], which included social cognitive theory [53] and self-determination theory [51]. Details about these frameworks were found by reviewing preliminary, formative research [31,51,53]. No other study provided information about behavior change theories having guided the mobile intervention design. Of note, Schaeffer et al [58] have a manuscript in preparation that aims to describe the development of the mobile app. As such, this manuscript under preparation may allude to the behavior change theories that underlie the mobile app development. Nonetheless, the lack of behavior change theory implementation in mobile interventions is consistent with findings from past reviews [33,34,63], suggesting that designing mobile phone–based interventions without a theoretical foundation for behavior change within design is a common practice across different niches in the mobile intervention literature.

However, an evaluation of the studies using the taxonomy of behavior change techniques [46] revealed that these techniques were frequently used. The number of behavior change techniques included in the interventions ranged from 2 [49] to 9 [51]. The most used behavior change techniques within the apps included providing instruction (9/11, 82%), prompting practice (9/11, 82%), and prompting self-monitoring of a behavior (8/11, 73%). Taken together, the studies seem to have implemented some behavior change techniques widely, but the techniques were not guided by a stated behavior change theory in most of the studies (9/11, 82%; Table 2).

Table 2. Characteristics and features of reviewed studies.

Study, year	Sample, n; design	Delivery medium	Standalone treatment or an enhancement to treatment	Mobile device used	Mental health condition targeted	Type of intervention used, incorporated, or adapted	Underlying intervention theory	Behavior change theory applied in design	Behavior change techniques	Parent training skills used in intervention	Youth outcomes	Parent outcomes
Breitenstein et al [53], 2016	79; RCT[a]	MA[b]	Standalone treatment	Tablet	Behavior problems	Chicago Parent Program	CM[c], SLT[d], and SCT[e]	SCT (theory was specified in a referenced article or preliminary work that was referenced)	CR[f], GE[g], M or D[h], OSC[i], PF[j], PI[k], PP[l], and SM[m]	Parent-child relationship; clear expectations and rules; rewards and incentives; setting behavior goals; and effective requests	No significant change in child behavior problems	Improvement in parental warmth (Cohen *d*=0.31); improvement in parental self-efficacy (Cohen *d*=0.13), and improvement in parental follow-through on skills (Cohen *d*=0.18)
Feil et al [54], 2018	42; RCT	MA	Standalone treatment	Smartphone	Conduct and antisocial behaviors	Behavioral parent training skills	—[n]	—	FU[o], GE, PP, RG[p], SGS[q], and SM	Clear expectations and rules; rewards and incentives; setting behavior goals	—	No significant change in parenting behaviors
Hemdi and Daley [55], 2016	62; RCT	MA (the app used was an existing messenger app, not a newly developed one)	Enhancement to treatment	Smartphone	Autism spectrum disorder	Psychoeducation intervention	DABCX[r] and TMS[s]	—	PIC[t], PIN[u], and SM	Parent-child relationship	Improvement in hyperactivity (Cohen *d*=–1.58)	Reduction in parenal stress (Cohen *d*=–0.98); reduction in parental depression (Cohen *d*=–2.05)
Jones et al [51], 2014	15; RCT	SPE[v]	Enhancement to treatment	Smartphone	Disruptive behavior disorders	Helping the Noncompliant Child	SLT and CM	SDT[w] (theory was specified in a referenced article or preliminary work that was referenced)	BI[x], CR, GE, M or D, OSC, PF, PI, PIN, PP, and SM	Parent-child relationship; clear expectations and rules; rewards and incentives; effective requests; praise; planned ignoring; modeling	Improvements in intensity of disruptive behaviors (Cohen *d*=0.99); improvements in presence of disruptive behaviors (Cohen *d*=0.54)	Improvement in parental engagement and generalization of parenting skills (Cohen *d*=0.88); increased participation in midweek check-ins (Cohen *d*=2.59); increased completion of home practice (Cohen *d*=0.63)

Study, year	Sample, n; design	Delivery medium	Standalone treatment or an enhancement to treatment	Mobile device used	Mental health condition targeted	Type of intervention used, incorporated, or adapted	Underlying intervention theory	Behavior change theory applied in design	Behavior change techniques	Parent training skills used in intervention	Youth outcomes	Parent outcomes
Lefever et al [56], 2017	371; RCT	SMS text messaging	Enhancement to treatment	Mobile phone	Behavior problems	Parent Child Interaction module of SafeCare	SLT and EST[y]	—	GE, PF, PI, and PP	Parent-child relationship; clear expectations and rules; rewards and incentives; praise; modeling	Improvement in cooperative behavior (Cohen *d*=0.38)	Increase in observation of parenting skills use (Cohen *d*=0.68); improvement in responsive parenting skills (Cohen *d*=0.35); growth of use of parenting skills (Cohen *d*=0.28)
Mason et al [57], 2021	52; RCT	SMS text messaging	Standalone treatment	Mobile phone	Substance use	Behavioral parent training skills	SLT, SCT, and CM	—	IF[z], PIC, RG, and PI	Parent-child relationship; effective requests; monitoring	Decrease in depressive symptoms (Cohen *d*=–0.63); decrease in anxiety symptoms (Cohen *d*=–0.57)	Improvement in parental relations (Cohen *d*=0.41); improvement in parenting skills (Cohen *d*=0.51)
May et al [50], 2021	184; pilot study	SMS text messaging	Standalone treatment	Mobile phone	Autism spectrum disorder	Behavioral parent training skills[aa]	—	—	PIN and GE	Parent-child relationship	—	Improvement in parent-child relationship
Pina et al [52], 2014	10; pilot study	MA; EDA wristband[ab]	Standalone treatment	Mobile phone; tablet	Attention-deficit/hyperactivity disorder	Parental behavioral therapy[aa]	SLT and TTC[ac]	—	GE, IF, IRM[ad], PI, PP, and SM	Parent-child relationship; clear expectations and rules; setting behavior goals; effective requests; praise; planned ignoring; modeling	—	—

Study, year	Sample, n; design	Delivery medium	Standalone treatment or an enhancement to treatment	Mobile device used	Mental health condition targeted	Type of intervention used, incorporated, or adapted	Underlying intervention theory	Behavior change theory applied in design	Behavior change techniques	Parent training skills used in intervention	Youth outcomes	Parent outcomes
Schaeffer et al [58], 2022	72; RCT	MA	Standalone treatment	Smartphone	Conduct problems	Multisystemic therapy	EST	— (app did not explicitly state theory but suggested the presence of a theory without providing content indicating that the theory was used in this study)	PP, SM, IF, M or D, PI, PIN, RG, SGS, and CR	Parent-child relationship; clear expectations and rules; rewards and incentives; setting behavior goals; effective requests; monitoring; modeling	Decrease in substance use, delinquency, and status offenses (Cohen *d*=0.54-0.84)	Improvement in discipline consistency (Cohen *d*=0.44); improvement in rule clarity (Cohen *d*=0.32)
Sonne et al [59], 2016	11; pilot study	MA	Standalone treatment	Smartphone	Attention-deficit/hyperactivity disorder	The Incredible Years Program[aa]	SLT	— (app did not explicitly state theory but suggested the presence of a theory without providing content indicating that the theory was used in this study)	IF, PI, PP, SGS, and SM	Parent-child relationship; clear expectations and rules; rewards and incentives; setting behavior goals; monitoring	Reduction in inattention (Cohen *d*=0.73); improvement in conduct-related behaviors (Cohen *d*=1.02); improvement in youth sleep (Cohen *d*=0.67)	Improvement in parental frustration

Study, year	Sample, n; design	Delivery medium	Standalone treatment or an enhancement to treatment	Mobile device used	Mental health condition targeted	Type of intervention used, incorporated, or adapted	Underlying intervention theory	Behavior change theory applied in design	Behavior change techniques	Parent training skills used in intervention	Youth outcomes	Parent outcomes
Sullivan et al [60], 2019	45; pilot study	MA	Enhancement to treatment	Smartphone	Trauma	RPC[ae] and TIPS[af]	AT[ag], CBT[ah], CDT[ai], SLT, and RT[aj]	— (app did not explicitly state theory but suggested the presence of a theory without providing content indicating that the theory was used in this study)	CR, M or D, OSC, PI, PIN, PP, and SM	Parent-child relationship; clear expectations and rules; praise; planned ignoring	Increase in youth prosocial behavior (Cohen *d*=0.40)	Improvement in parental self-efficacy (Cohen *d*=0.41)

[a]RCT: randomized controlled trial.

[b]MA: mobile app.

[c]CM: coercion model.

[d]SLT: social learning theory.

[e]SCT: social cognitive theory.

[f]CR: providing contingent rewards.

[g]GE: providing general encouragement.

[h]M or D: behavior modeled or demonstrated by a professional.

[i]OSC: opportunities to view social change.

[j]PF: providing feedback.

[k]PI: providing instruction.

[l]PP: prompting practice.

[m]SM: self-monitoring of specific behavior.

[n]Not applicable (not reported or not able to draw from study information).

[o]FU: providing follow-up prompts.

[p]RG: prompting a review of current goals.

[q]SGS: specific goal setting.

[r]DABCX: Double ABCX Model of Stress.

[s]TMS: Transactional Model of Stress.

[t]PIC: providing information on consequences of behaviors.

[u]PIN: providing information.

[v]SPE: smartphone enhancements, including SMS text messaging, video calls, alarms, and skills videos.

[w]SDT: self-determination theory.

[x]BI: barrier identification.

[y]EST: ecological systems theory.

[z]IF: prompting intention formation.

[aa]Intervention design was based on user and professional feedback but drew on elements of the mentioned intervention.

[ab]EDA: electrodermal activity.

[ac]TTC: transtheoretical change theory.

[ad]IRM: prompting identification as a role model.

[ae]RPC: Resource Parent Curriculum (National Child Traumatic Stress Network).

[af]TIPS: trauma-informed parenting skills.

[ag]AT: attachment theory.

[ah]CBT: cognitive behavioral theory.

[ai]CDT: child development theory.

[aj]RT: resilience theory.

Design Elements

Features

Assessment of all the studies suggested the presence of 5 features: tailoring intervention content, push notifications, tracking of behaviors, modeling skills through video demonstration, and reward systems. The reviewed studies varied in their implementation of these features. First, many of the studies (7/11, 64%) included options for the end user to tailor mobile app intervention content or features. Options to tailor intervention content included defining individualized behavioral goals such as completing household chores, following a bedtime routine, returning home by curfew, and completing homework (4/7, 57%) [53,54,58,59]; creating a schedule for when to use the parenting skills provided in the app (eg, choosing when to engage in particular modules, creating a routine for parents, and allowing ongoing access to psychoeducation; 3/7, 43%) [53,59,60]; delivering just-in-time interventions according to individual stress level (2/7, 29%) [52,58]; selecting rewards or contingencies that they think their child would value (4/7, 57%) [53,54,58,59]; and receiving psychoeducation tailored to individual circumstances (eg, how to intervene when the child is in a risky situation and using time-outs with children who have experienced trauma; 4/7, 57%) [51,53,58,60]. Several studies also offered opportunities to tailor features of the app (6/11, 54%). Options to tailor app features included choosing icons, avatars, and profile photos that embody the user (2/11, 18%%) [54,60]; filming oneself practicing skills with the youth (1/11, 9%) [51]; and integration of the user's name in the delivery medium (4/11, 36%) [54,57-59]. Notably, only a few of the studies (2/11, 18%) provided comprehensive visual examples or a description of the treatment content and mobile platform. As a result, other design features may be embedded in the apps but have not been identified in this review.

Second, most of the mobile interventions (8/11, 73%) included push notifications and SMS text messages to prompt practice of strategies or provide reinforcement and encouragement [50-52,55-59].

Third, some of the studies (3/11, 27%) included a mechanism for tracking youth behaviors [54,58,59], such as completing steps in a routine, monitoring the youth's location, and assessing the completion of positive behaviors. Behaviors were tracked either by parents [60] or by both parents and children [54,58,59], and they were logged by adding events to a log sheet [54,58] or by moving through a checklist in situ [59]. Behavioral tracking (ie, assessing and following the behaviors of the youth concerned) was implemented through the mobile intervention in each of these studies.

Fourth, some of the studies (4/11, 36%) included videos modeling parent-child interactions and other parenting skills [51,53,58,60].

Finally, nearly half of the studies (5/11, 45%) featured a reward system for either the parent or the child [53,54,58-60] that was implemented through the mobile intervention. For children, rewards included points [54,58] and stickers [59], while parents earned completion badges and certificates [53] or accessories for an avatar family [60].

Content

Broadly, the mobile intervention content related to behavioral parent training skills included strengthening the parent-child relationship (10/11, 91%; the exception was the study by Feil et al [54]), setting clear expectations and rules (8/11, 73%; the exceptions were the studies by May et al [50], Mason et al [57], and Hemdi and Daley [55]), the establishment of rewards and incentives (6/11, 55%) [51,53,54,56,58,59], setting behavioral goals for the youth (5/11, 45%) [52-54,58,59], the use of effective communication and requests (5/11, 45%) [51-53,57,58], praising desired behaviors (4/11, 36%) [51,52,56,60], modeling effective behaviors (4/11, 36%) [51,52,56,58], planned and active ignoring (4/11, 36%) [51,52,56,60], and the implementation of monitoring and supervision (3/11, 27%) [57-59]. Although an indication of behavioral parent training skills can be gleaned from the description of the intervention, it is challenging to determine the exact number of these skills. This difficulty stems from a lack of detailed information in the articles regarding the specific skills provided in the mobile intervention. Of note, some of the studies (2/11, 18%) [51,56] involved enhancements to in-person treatment delivery, suggesting that additional skills were likely provided and discussed through the technology, although they were not explicitly mentioned in the manuscripts.

Ideally, a review of app content includes an assessment across 4 domains: evidence-based content, quality of information provision, completeness and conciseness, and clarity about the program's purpose [43]. To fully implement this evaluation, studies must provide comprehensive information, including examples of content across the intervention (eg, specific messages designed for the end user, video dialogue, and prompts used to encourage practice). Unfortunately, most of the studies (10/11, 90%) included in this review did not include enough information for a thorough review of app content across these 4 dimensions. In fact, only 1 (9%) of the 11 studies [52] allowed for a partial evaluation of content according to the Enlight domains.

For 8 (73%) of the 11 studies, specific, direct content was not described [50,51,53-56,58,60]. Consequently, the content could not be evaluated against the Enlight domains.

Of the 11 studies, 3 (27%) provided specific examples of content [52,57,59]. However, the information was not comprehensive in that it consisted of brief sample treatment statements and lacked psychoeducation related to the skills being implemented. For 2 (67%) [57,59] of these 3 studies, none of the Enlight domains could be evaluated. With regard to the third study [52], some Enlight domains were assessed (ie, clear and concise goals, quality information necessary to obtain these goals, and clarity regarding the purpose and target population of the program). The evaluation showed that the content fulfilled these domains at *good* levels.

Visual Design

Ideally, a review of visual design includes the evaluation of aesthetics, layout, and size [43]. As with content evaluation, studies must provide a comprehensive collection of visualization of the mobile app. Because this assessment examines font consistency, the harmony of colors used throughout, and the size of the layout on the mobile device [43], a comprehensive set of visualizations would include examples of various pages in the app design and include examples of color, font, images, and treatment content. Most of the studies (6/11, 54%) included in this review provided few visualization examples, which created difficulty in conducting a full assessment of the visual design used throughout the app.

Of the 11 studies, 4 (36%) were not evaluated for visual design because either the interventions were SMS text message based [50,56,57] or no visual information was provided in the article [55]. Among the remaining 7 studies, 6 (86%) included examples of the mobile intervention with snapshots of select screens, rather than a visual design sample of the app in its entirety [52-54,58-60]. Of note, 1 (17%) of these 6 studies included visualizations of the mobile app through a website, rather than within the published study article [58]. Because of the scarcity of visual examples among the studies included in this review, evaluation based on the Enlight criteria was completed for the components of the intervention the researchers selected to present, rather than the app in totality. The results showed that the aesthetics classifications included *not attractive* (1/6, 17%), *fair* (1/6, 17%), *attractive* (2/6, 33%), and *very attractive* (2/6, 33%). The layout classifications ranged from *fair* (1/6, 17%) and *good* (4/6, 67%) to *very good* (1/6, 17%). The size qualities ranged from *fair* (1/6, 17%) and *good* (2/6, 33%) to *very good* (3/6, 50%).

Most of the studies (4/11, 36%) implemented muted colors on certain screens or activities for parent-directed content [52,53,59,60]. For parent- and child-directed or child-only–directed content, colors were brighter than those seen on parent-directed content screens [54,59]. Using brighter colors with children aligns with the robust literature on the preference of younger children for brighter, more saturated colors over more muted colors [64-66]. The depiction of families was an overwhelmingly common visual design element; for example, the studies included photos of families on home pages who resembled the families using the app [53], actual photos of the families themselves [54], or the integration of the names of the children [58].

Treatment Outcomes

Youth Outcomes

Of the 11 studies, 8 (73%) provided youth outcomes. Of these 8 studies, 7 (88%) indicated substantial improvement in the youth. Of these 7 studies, 6 (86%) were randomized controlled trials and indicated that the youth showed greater improvement in the technology-enhanced group when compared to the control group [51,53,55-58]. Overall, the youth in these randomized trials exhibited a decrease in behavioral and mood-related problems [51,55,56,58]; for example, Lefever et al [56] reported that youth with parents in the intervention condition demonstrated a significant improvement in cooperative behavior with a small to medium effect size (Cohen d=0.38); Hemdi and Daley [55] reported significant improvement in hyperactivity for youth in the intervention condition with a large effect size (Cohen d=−1.54); Jones et al [51] reported significant improvements in the intensity (Cohen d=0.99) and presence of disruptive behaviors (Cohen d=0.54) with medium to large effect sizes for youth in the intervention condition; Mason et al [57] reported significant small to medium effect sizes, demonstrating a decrease in depressive (Cohen d=−0.63) and anxiety (Cohen d=−0.57) symptoms; and Schaeffer et al [58] reported significant decreases in substance use, delinquency, and status offenses for youth in the intervention condition with medium to large effect sizes (ranging from Cohen d=0.54 to Cohen d=0.84).

Of the 8 studies, 2 (25%) were pilot studies [59,60]. Although these studies did not use a randomized group as a comparison, both reported similar improvements in youth behavior problems; for example, Sonne et al [59] reported a significant reduction in inattention at a medium effect size (Cohen d=0.73), improvement in conduct-related behaviors at a large effect size (Cohen d=1.02), and improvement in youth sleep at a medium effect size (Cohen d=0.67). Sullivan et al [60] reported an increase in youth prosocial behavior at a small effect size (Cohen d=0.40).

Parent Outcomes

Most of the studies (10/11, 91%) reported parent outcomes. Many of the studies (9/11, 81%) reported parental improvements when using mobile technology. Among the 7 studies using a randomized controlled trial design, all (7/7, 100%) reported parental improvements in the technology intervention groups compared to their respective control groups. Specifically, Breitenstein et al [53] reported an improvement with small to medium effect sizes in parental warmth (Cohen d=0.31), parental self-efficacy (Cohen d=0.13), and parental follow-through on skills (Cohen d=0.18). Hemdi and Daley [55] reported large effect sizes for reduction in parental stress (Cohen d=−0.98) and parental depression (Cohen d=−2.05) among parents in the mobile intervention group. Lefever et al [56] reported a medium effect size in the observation of parenting skills use (Cohen d=0.68), a small to medium effect size in the improvement in responsive parenting skills (Cohen d=0.35), and a small effect size in the growth of use of parenting skills (Cohen d=0.28). Jones et al [51] reported a greater improvement in parental engagement and the generalization of parenting skills for the parents using a mobile intervention with weekly attendance (Cohen d=0.88), participating in midweek check-ins (Cohen d=2.59), and the completion of home practice (Cohen d=0.63), reflecting medium to large effect sizes. Schaeffer et al [58] reported small to medium effects in the improvement of discipline consistency (Cohen d=0.44) and rule clarity and structure (Cohen d=0.32). Mason et al [57] reported small to medium effect sizes in the improvements of parent relations (Cohen d=0.41) and parenting skills (Cohen d=0.51), reflecting medium effect sizes. Finally, and of note, while Feil et al [54] reported a small to medium effect size in the reduction of negative parenting behaviors, the researchers note that this

finding is insignificant and did not report this coefficient. Therefore, this reported finding is not included in Table 2.

Of the 10 studies, 3 (30%) were pilot studies that incorporated parent outcomes. Although these studies did not use a randomized group as a comparison, they described a small effect size in the improvement of parental self-efficacy (Cohen *d*=0.41) [60], significant improvements in parental frustration [60], and improvements in parent-child relationship [50] among the participants engaging in the mobile intervention.

Discussion

Overview

This study systematically reviewed noncommercial mHealth apps that provide behavioral parent training or components of behavioral parent training for parents of children with behavior problems. This study had the specific goals of summarizing (1) general characteristics, (2) theoretical frameworks and empirical evidence, and (3) parenting strategies. The broad aim for this review was to use the results to inform the development of an app for parents of teens who are involved in substance use behaviors.

Use of Theory

This review found that all studies (11/11, 100%) included in this review used parent training interventions that are theoretically grounded. However, there was a paucity of clear information outlining the theoretical framework for the components designed within the mobile apps. While many of the apps were based on face-to-face parent management training interventions that have well-established theoretical frameworks (eg, Helping the Noncompliant Child, Parent-Child Interaction Therapy, and Parent Management Training), the studies often only referenced the interventions [56,57,59,60] or reported specific parenting skills without specifying the originating intervention [50,54,55]. Determining relevant theoretical frameworks required deductive reasoning based on a knowledge of the named intervention or parenting skill. Using this expertise, we found that many of the apps (6/11, 55%) used social learning theory [67], while only a few (4/11, 36%) [51-53,59] indicated the use of the coercion model [62], either by mentioning this model by name [51,53] or by describing the benefits of the parenting skills that were selected for inclusion in the apps [52,59].

Given that parent training interventions can draw from different theoretical frameworks beyond social learning theory and the coercion model [68] and that the full in-person treatment programs were not transferred to the mobile apps in the reviewed studies, implicit communication of the theoretical framework through only naming the originating intervention or specific skills is not sufficient. It is important to clearly state the relevant theoretical framework for the content transferred to mobile devices because the inclusion of a particular intervention does not guarantee that its principles are embodied in the mobile platform. A deeper analysis of whether specific interventions are consistent with the theory in terms of mobile app features could not be performed due to a lack of information.

Behavior Change

Although behavior change theory is a vital component of mobile interventions [33,34,62,69], the results of this review reveal that most of the studies (9/11, 81%) did not explicitly refer to behavior change theory. The absence of behavior change theory as a framework for app design in many of the reviewed studies (9/11, 81%) may be due to at least 3 reasons. First, many of the studies (8/11, 72%) failed to provide sufficient information about the content and development of the mobile intervention, making it challenging to understand the detailed study characteristics. Second, the studies may have relied on behavior change theories for the originating curriculum due to the well-established programs on which the interventions are based. However, this overlooks the challenges of transferring the interventions to mobile devices. Many parent management training curricula incorporate behavior change theories that consider factors such as personal motivation, social support, and perceived barriers and benefits of behavior change. While similar theories may be used in mHealth interventions, there are additional considerations, such as the need to focus more on technology-specific behavior change theories; for example, the technology acceptance model [70] focuses on how individuals perceive and adopt novel technologies, and it may be a suitable theory to embed in mHealth development. The lack of prior designs centered on the individual's perspective remains a gap. Third, there may have been a general oversight in including a coherent behavior change theory in the intervention's design. Given the impact of behavior change theory on the effectiveness of interventions [33,34,69,71], its inclusion is crucial for the development of effective parent-targeted mobile interventions.

Although behavior change theories were not commonly cited in the reviewed studies, behavior change techniques were used. The most used techniques were providing instruction, prompting practice, and self-monitoring [46]. The findings of this study are consistent with prior reviews showing that self-monitoring was the most frequently used behavior change technique in mobile interventions across different populations [33,35]. In this review, self-monitoring described tracking both the implementation of parental skills and behaviors and the presence of desired behaviors in children. The limitations of the studies in terms of behavior change theory and techniques will be discussed further while summarizing the design elements (refer to the next subsection).

Mobile Intervention Design

In this review, we evaluated the design of mobile interventions used in the included studies based on their content, visual design, and features. While all studies incorporated particular parenting skills (11/11, 100%), comprehensive information regarding the implementation of specific skills was not included in many of the studies (4/11, 36%). When information was provided, there was limited detail on the selection and integration of the skills into the mobile platform. This lack of information made it challenging to critically assess these interventions and consider them in their entirety for use with different populations, such as parents of adolescents recovering from substance use.

The visual design and features of the mobile interventions in the included studies were designed to be personalized. Each mobile intervention had its own way of promoting personalization through visual design and features; for example, users could customize the visual design by selecting icons, fonts, and colors that were personally appealing [53,54,59]. This practice of personalizing the layout and design has also been noted in previous literature reviews [72], suggesting consistent focus across populations and areas of study. The features of these mobile interventions also facilitated individuality, including avatar families [60], tailored messages [56], and the option to choose specific skills to practice based on individual needs [53].

In addition, the use of badges, rewards, logs, and tokens as a feature in the reviewed mobile interventions was a common pattern This feature was also noted in previous reviews as prevalent [72]. The use of rewards aimed at positively reinforcing desired behaviors in both parents and adolescents and encouraged individualization. Positive reinforcement, in which a stimulus increases the frequency of a particular behavior, is a well-established behavior change technique [73]. When implemented within a structured framework, positive reinforcement can be effective in promoting desired behaviors [73].

Treatment Outcomes

The preliminary findings of the reviewed studies indicate potential for positive parent and child outcomes after the use of a behavioral parent training app [50,51,55-60], but further research is necessary to support these findings. Most of the studies that reported on youth and parent outcomes (8/11, 72%) used interventions that were grounded in well-established theoretical frameworks [51,53,55-60], suggesting that theory-driven interventions may play a critical role in outcomes after behavioral parent training delivered through mobile devices.

To optimize the effectiveness of behavioral parenting apps, future studies should incorporate behavior change theories in the design and development process. The limited information available in previous studies on the content and development of parent-targeted interventions within mobile platforms makes it challenging to identify the behavior change theory applied, if any (the exceptions were the studies by Jones et al [31], Jones et al [74], May et al [75], Breitenstein et al [53], Schaeffer et al [58], Jones et al [51], and May et al [50]). Thus, further empirical research is necessary to determine the influence of behavior change theory on the outcomes of behavioral parent training delivered through mHealth apps.

Implications for mHealth Parenting Apps to Address Teen Substance Use

The development of behavioral parent training apps for parents of children with mental health difficulties is still in the preliminary stages, with available apps developed primarily for parents of younger children. However, there is a need for apps for parents of adolescents with conduct problems, including substance use. These apps are crucial because effective parenting strategies are related to decreased levels of substance use [2,3], and engagement in mobile platforms may be helpful for parents with difficulty accessing treatment in the community. Studies have revealed that it is challenging for parents to both access and engage in evidence-based treatments with behavioral parenting strategies in community settings [76,77].

While mHealth platforms provide accessibility, customization is key to fostering engagement. Thankfully, most of the apps (7/11, 63%) in this review of mHealth apps for delivering behavioral parent training included features such as customization and personalization, which are considered good practice [78,79]. However, the reviewed apps lacked integration within the app and between the app and smartphones. This limitation is due to the predominant focus on parents of younger children. However, as parents spend less time with their children, and key parenting strategies broaden to include monitoring and supervision during adolescence [80], the integration of features is likely to become increasingly important for behavioral parent training apps. To address this gap, it is recommended that a more comprehensive integration of app design and mobile phone features occur for apps targeting parents of adolescents; for example, location-based reminders could be used to track a teen's location and send reminders or prompts relevant to their current location to the parent because parents often either rely on youth reports of their location or do not check in on potential location changes. Specifically, if the teen is at a location where they are likely to use substances, the app could send a reminder for the parent to have a preplanned conversation with their teen or to check up on their whereabouts. In addition, geofencing could be used to set up web-based boundaries around specific locations, sending an alert or reminder when the teen enters an *off-limits* location and offering suggestions for which parenting strategies to use to address the infraction. Alternatively, an alert with scripted language that is consistent with effective praise [21] could be sent to the parent if the teen's movements suggest that an off-limits area was avoided so that the parent can engage in providing praise in the moment because offering immediate praise and feedback is key for changing behavior [73]. Notably, Schaeffer et al [58] used similar location-based strategies to encourage the monitoring and supervision of adolescents.

In addition, an app could be designed to use GPS data to generate reports that summarize the adolescent's whereabouts over time. These reports could serve as personalized feedback with recommendations to the parents. This sort of analysis could help parents discern patterns and areas that require the use of certain parenting strategies. Moreover, push notifications could be integrated to remind parents to use a specific parenting strategy at a designated time or in response to a trigger that was defined by the personalized feedback report. Rich integration of features may bring social learning theory and the coercion model to the forefront, resulting in the potential integration of behavior change theories and techniques. If so, this integration could address limitations in most current noncommercial apps.

Limitations and Strengths

It is important to consider the results of this review within the context of a few limitations. First, this review only looked at a specific area of mHealth research: mobile phone–based

interventions that were designed to provide behavioral parenting practices to parents of children with mental health difficulties. Studies on mobile interventions targeting parenting for parents of children with medical issues were not included in this review. Given the overlap in issues related to ineffective parenting between parents of children with behavior problems and parents of children with chronic medical conditions such as asthma and obesity [81], future studies may benefit from reviewing parenting apps designed for both groups for a broader view. Second, this review was based on a new field of research. As such, there was a paucity of studies available for review. As a result, we included studies involving children of different ages, spanning different developmental windows and parenting needs. Notably, 2 (18%) of the 11 studies provided insight into current behavioral parenting apps for parents of teens [57,58]. Third and last, because of the small pool of studies, synthesis of data to obtain overall effect sizes was not possible. Therefore, as the field grows, the effectiveness of parent-targeted mobile interventions for parents of youth with mental health issues should be empirically assessed using statistical analyses to develop a meta-analysis. The garnering of data will provide more robust evidence of the effectiveness of these interventions in a population of youth with mental health difficulties.

Despite these limitations, this review has some strengths that make it valuable for understanding current noncommercial parenting apps for informing the development of similar apps for related problems in childhood and adolescence. By focusing on behavioral parenting apps for parents of children and adolescents with mental health difficulties, this review provides targeted and relevant information for developers who are interested in designing an app using parenting practices that are well established for other behavior problems occurring in youth, including substance use [2-4,7,82]. Finally, the results of this review provide clear information about current practices and patterns so that future research can more closely align the development of apps with design features that may increase treatment engagement [72,83] and, hopefully, buttress outcomes.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[PDF File (Adobe PDF File), 79 KB - mhealth_v12i1e51273_app1.pdf]

References

1. Key substance use and mental health indicators in the United States: results from the 2020 National Survey on Drug Use and Health. Substance Abuse and Mental Health Services Administration (SAMHSA). URL: https://www.samhsa.gov/data/sites/default/files/reports/rpt35325/NSDUHFFRPDFWHTMLFiles2020/2020NSDUHFFR1PDFW102121.pdf [accessed 2024-04-29]
2. Henderson CE, Rowe CL, Dakof GA, Hawes SW, Liddle HA. Parenting practices as mediators of treatment effects in an early-intervention trial of multidimensional family therapy. Am J Drug Alcohol Abuse 2009 Aug 13;35(4):220-226 [FREE Full text] [doi: 10.1080/00952990903005890] [Medline: 20180674]
3. Winters KC, Lee S, Botzet A, Fahnhorst T, Nicholson A. One-year outcomes and mediators of a brief intervention for drug abusing adolescents. Psychol Addict Behav 2014 Jun;28(2):464-474 [FREE Full text] [doi: 10.1037/a0035041] [Medline: 24955669]
4. Allen ML, Garcia-Huidobro D, Porta C, Curran D, Patel R, Miller J, et al. Effective parenting interventions to reduce youth substance use: a systematic review. Pediatrics 2016 Dec;138(2):e20154425 [FREE Full text] [doi: 10.1542/peds.2015-4425] [Medline: 27443357]
5. Godley SH, Garner BR, Passetti LL, Funk RR, Dennis ML, Godley MD. Adolescent outpatient treatment and continuing care: main findings from a randomized clinical trial. Drug Alcohol Depend 2010 Jul 01;110(1-2):44-54 [FREE Full text] [doi: 10.1016/j.drugalcdep.2010.02.003] [Medline: 20219293]
6. Henderson CE, Wevodau AL, Henderson SE, Colbourn SL, Gharagozloo L, North LW, et al. An independent replication of the adolescent-community reinforcement approach with justice-involved youth. Am J Addict 2016 Apr 18;25(3):233-240 [FREE Full text] [doi: 10.1111/ajad.12366] [Medline: 26992083]
7. Stanger C, Ryan SR, Scherer EA, Norton GE, Budney AJ. Clinic- and home-based contingency management plus parent training for adolescent cannabis use disorders. J Am Acad Child Adolesc Psychiatry 2015 Jun;54(6):445-53.e2 [FREE Full text] [doi: 10.1016/j.jaac.2015.02.009] [Medline: 26004659]
8. Priester MA, Browne T, Iachini A, Clone S, DeHart D, Seay KD. Treatment access barriers and disparities among individuals with co-occurring mental health and substance use disorders: an integrative literature review. J Subst Abuse Treat 2016 Feb;61:47-59 [FREE Full text] [doi: 10.1016/j.jsat.2015.09.006] [Medline: 26531892]
9. Wilson DB, Olaghere A, Kimbrell CS. Implementing juvenile drug treatment courts: a meta-aggregation of process evaluations. J Res Crime Delinq 2019 Feb 14;56(4):605-645. [doi: 10.1177/0022427819826630]

10. Ashford RD, Brown AM, Curtis B. Systemic barriers in substance use disorder treatment: a prospective qualitative study of professionals in the field. Drug Alcohol Depend 2018 Aug 01;189:62-69. [doi: 10.1016/j.drugalcdep.2018.04.033] [Medline: 29883870]
11. Hogue A, Henderson CE, Becker SJ, Knight DK. Evidence base on outpatient behavioral treatments for adolescent substance use, 2014-2017: outcomes, treatment delivery, and promising horizons. J Clin Child Adolesc Psychol 2018;47(4):499-526 [FREE Full text] [doi: 10.1080/15374416.2018.1466307] [Medline: 29893607]
12. Wisdom JP, Cavaleri M, Gogel L, Nacht M. Barriers and facilitators to adolescent drug treatment: youth, family, and staff reports. Addict Res Theory 2010 Nov 30;19(2):179-188. [doi: 10.3109/16066359.2010.530711]
13. Gearing RE, Schwalbe CS, Short KD. Adolescent adherence to psychosocial treatment: mental health clinicians' perspectives on barriers and promoters. Psychother Res 2012;22(3):317-326. [doi: 10.1080/10503307.2011.653996] [Medline: 22313513]
14. Gulliver A, Farrer L, Chan JK, Tait RJ, Bennett K, Calear AL, et al. Technology-based interventions for tobacco and other drug use in university and college students: a systematic review and meta-analysis. Addict Sci Clin Pract 2015 Feb 24;10:5 [FREE Full text] [doi: 10.1186/s13722-015-0027-4] [Medline: 25928221]
15. Haqanee Z, Peterson-Badali M, Skilling T. Making “what works” work: examining probation officers’ experiences addressing the criminogenic needs of juvenile offenders. J Offender Rehabil 2014 Dec 20;54(1):37-59. [doi: 10.1080/10509674.2014.980485]
16. Curtis BL, Ashford RD, Magnuson KI, Ryan-Pettes SR. Comparison of smartphone ownership, social media use, and willingness to use digital interventions between generation Z and millennials in the treatment of substance use: cross-sectional questionnaire study. J Med Internet Res 2019 Apr 17;21(4):e13050 [FREE Full text] [doi: 10.2196/13050] [Medline: 30994464]
17. Li K, Magnuson KI, Beuley G, Davis L, Ryan-Pettes SR. Features, design, and adherence to evidence-based behavioral parenting principles in commercial mHealth parenting apps: systematic review. JMIR Pediatr Parent 2023 Jun 01;6:e43626 [FREE Full text] [doi: 10.2196/43626] [Medline: 37261886]
18. Smith JM, Bright KS, Mader J, Smith J, Afzal AR, Patterson C, et al. A pilot of a mindfulness based stress reduction intervention for female caregivers of youth who are experiencing substance use disorders. Addict Behav 2020 Apr;103:106223. [doi: 10.1016/j.addbeh.2019.106223] [Medline: 31838440]
19. Kuntsche S, Kuntsche E. Parent-based interventions for preventing or reducing adolescent substance use - a systematic literature review. Clin Psychol Rev 2016 Apr;45:89-101. [doi: 10.1016/j.cpr.2016.02.004] [Medline: 27111301]
20. Bo A, Hai AH, Jaccard J. Parent-based interventions on adolescent alcohol use outcomes: a systematic review and meta-analysis. Drug Alcohol Depend 2018 Oct 01;191:98-109. [doi: 10.1016/j.drugalcdep.2018.05.031] [Medline: 30096640]
21. Forgatch MS, Patterson GR. Parent management training-Oregon model: an intervention for antisocial behavior in children and adolescents. In: Weisz JR, Kazdin AE, editors. Evidence-Based Psychotherapies for Children and Adolescents. 2nd edition. New York, NY: The Guilford Press; 2010:159-177.
22. Horigian VE, Anderson AR, Szapocznik J. Family-based treatments for adolescent substance use. Child Adolesc Psychiatr Clin N Am 2016 Oct;25(4):603-628 [FREE Full text] [doi: 10.1016/j.chc.2016.06.001] [Medline: 27613341]
23. Kazdin AE. Problem-solving skills training and parent management training for oppositiona defiant disorder and conduct disorder. In: Weisz JR, Kazdin AE, editors. Evidence-Based Psychotherapies for Children and Adolescents. 2nd edition. New York, NY: The Guilford Press; 2010:211-226.
24. Nesvåg S, McKay JR. Feasibility and effects of digital interventions to support people in recovery from substance use disorders: systematic review. J Med Internet Res 2018 Aug 23;20(8):e255 [FREE Full text] [doi: 10.2196/jmir.9873] [Medline: 30139724]
25. Marcu G, Ondersma SJ, Spiller AN, Broderick BM, Kadri R, Buis LR. The perceived benefits of digital interventions for behavioral health: qualitative interview study. J Med Internet Res 2022 Mar 30;24(3):e34300 [FREE Full text] [doi: 10.2196/34300] [Medline: 35353045]
26. Hoskins DH. Consequences of parenting on adolescent outcomes. Societies 2014 Sep 18;4(3):506-531. [doi: 10.3390/soc4030506]
27. Pinquart M. Associations of parenting dimensions and styles with externalizing problems of children and adolescents: an updated meta-analysis. Dev Psychol 2017 May;53(5):873-932. [doi: 10.1037/dev0000295] [Medline: 28459276]
28. Fosco GM, Stormshak EA, Dishion TJ, Winter CE. Family relationships and parental monitoring during middle school as predictors of early adolescent problem behavior. J Clin Child Adolesc Psychol 2012 Mar;41(2):202-213 [FREE Full text] [doi: 10.1080/15374416.2012.651989] [Medline: 22417193]
29. Van Ryzin MJ, Fosco GM, Dishion TJ. Family and peer predictors of substance use from early adolescence to early adulthood: an 11-year prospective analysis. Addict Behav 2012 Dec;37(12):1314-1324 [FREE Full text] [doi: 10.1016/j.addbeh.2012.06.020] [Medline: 22958864]
30. Chronis AM, Chacko A, Fabiano GA, Wymbs BT, Pelham WE. Enhancements to the behavioral parent training paradigm for families of children with ADHD: review and future directions. Clin Child Fam Psychol Rev 2004 Mar;7(1):1-27. [doi: 10.1023/b:ccfp.0000020190.60808.a4] [Medline: 15119686]

31. Jones DJ, Forehand R, Cuellar J, Kincaid C, Parent J, Fenton N, et al. Harnessing innovative technologies to advance children's mental health: behavioral parent training as an example. Clin Psychol Rev 2013 Mar;33(2):241-252 [FREE Full text] [doi: 10.1016/j.cpr.2012.11.003] [Medline: 23313761]
32. Mao X, Zhao X, Liu Y. mHealth app recommendation based on the prediction of suitable behavior change techniques. Decis Support Syst 2020 May;132:113248. [doi: 10.1016/j.dss.2020.113248]
33. Zhao J, Freeman B, Li M. Can mobile phone apps influence people's health behavior change? An evidence review. J Med Internet Res 2016 Oct 31;18(11):e287 [FREE Full text] [doi: 10.2196/jmir.5692] [Medline: 27806926]
34. Cho YM, Lee S, Islam SM, Kim S. Theories applied to m-health interventions for behavior change in low- and middle-income countries: a systematic review. Telemed J E Health 2018 Feb 13;24(10):727-741. [doi: 10.1089/tmj.2017.0249] [Medline: 29437546]
35. Payne HE, Lister C, West JH, Bernhardt JM. Behavioral functionality of mobile apps in health interventions: a systematic review of the literature. JMIR Mhealth Uhealth 2015 Feb;3(1):e20 [FREE Full text] [doi: 10.2196/mhealth.3335] [Medline: 25803705]
36. Virani A, Duffett-Leger L, Letourneau N. Parenting apps review: in search of good quality apps. Mhealth 2019;5:44 [FREE Full text] [doi: 10.21037/mhealth.2019.08.10] [Medline: 31620471]
37. Davis DW, Logsdon MC, Vogt K, Rushton J, Myers J, Lauf A, et al. Parent education is changing: a review of smartphone apps. MCN Am J Matern Child Nurs 2017;42(5):248-256. [doi: 10.1097/NMC.0000000000000353] [Medline: 28639997]
38. Trahan MH, Morley RH, Nason EE, Rodrigues N, Huerta L, Metsis V. Virtual reality exposure simulation for student veteran social anxiety and PTSD: a case study. Clin Soc Work J 2021;49(2):220-230 [FREE Full text] [doi: 10.1007/s10615-020-00784-7] [Medline: 33487778]
39. Chi NC, Demiris G. A systematic review of telehealth tools and interventions to support family caregivers. J Telemed Telecare 2014 Dec 04;21(1):37-44. [doi: 10.1177/1357633x14562734]
40. Baumel A, Muench F, Edan S, Kane JM. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. J Med Internet Res 2019 Sep 25;21(9):e14567 [FREE Full text] [doi: 10.2196/14567] [Medline: 31573916]
41. Al Falahi A, Glazebrook C, Shokraneh F. The effectiveness of online parenting programmes in promoting parents' and adolescents' mental wellbeing: a systematic review protocol. F1000Res 2019 Nov 1;8:1842. [doi: 10.12688/f1000research.20975.1]
42. Flujas-Contreras JM, García-Palacios A, Gómez I. Technology-based parenting interventions for children's physical and psychological health: a systematic review and meta-analysis. Psychol Med 2019 Aug;49(11):1787-1798. [doi: 10.1017/S0033291719000692] [Medline: 30977462]
43. Baumel A, Pawar A, Mathur N, Kane JM, Correll CU. Technology-assisted parent training programs for children and adolescents with disruptive behaviors. J Clin Psychiatry 2017 Oct 25;78(8):e957-e969. [doi: 10.4088/jcp.16r11063]
44. Silva ME, Graham F, Levack W, Hay-Smith EJ. Persuasive technology and behaviour change in parent-focused eHealth interventions supporting child health: a scoping review protocol. NZJP 2019 Mar 01;47(1):36-48. [doi: 10.15619/nzjp/47.1.05]
45. Hekler EB, Michie S, Pavel M, Rivera DE, Collins LM, Jimison HB, et al. Advancing models and theories for digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):825-832 [FREE Full text] [doi: 10.1016/j.amepre.2016.06.013] [Medline: 27745682]
46. Abraham C, Michie S. A taxonomy of behavior change techniques used in interventions. Health Psychol 2008 May;27(3):379-387. [doi: 10.1037/0278-6133.27.3.379] [Medline: 18624603]
47. Michie S, Prestwich A. Are interventions theory-based? Development of a theory coding scheme. Health Psychol 2010 Jan;29(1):1-8. [doi: 10.1037/a0016939] [Medline: 20063930]
48. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ 2009 Jul 21;339(jul21 1):b2535 [FREE Full text] [doi: 10.1136/bmj.b2535] [Medline: 19622551]
49. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
50. May CD, St George JM, Lane S. From presence to participation: engagement with an SMS program for fathers of children on the autism spectrum. J Child Fam Stud 2020 Nov 07;30(1):29-37. [doi: 10.1007/s10826-020-01845-8]
51. Jones DJ, Forehand R, Cuellar J, Parent J, Honeycutt A, Khavjou O, et al. Technology-enhanced program for child disruptive behavior disorders: development and pilot randomized control trial. J Clin Child Adolesc Psychol 2014 Aug 07;43(1):88-101 [FREE Full text] [doi: 10.1080/15374416.2013.822308] [Medline: 23924046]
52. Pina L, Rowan K, Johns P, Roseway A, Hayes G, Czerwinski M. In situ cues for ADHD parenting strategies using mobile technology. In: Proceedings of the 8th International Conference on Pervasive Computing Technologies for Healthcare. 2014 Presented at: PervasiveHealth '14; May 20-23, 2014; Oldenburg, Germany p. 17-24 URL: https://dl.acm.org/doi/10.4108/icst.pervasivehealth.2014.254958 [doi: 10.4108/icst.pervasivehealth.2014.254958]

53. Breitenstein SM, Fogg L, Ocampo EV, Acosta DI, Gross D. Parent use and efficacy of a self-administered, tablet-based parent training intervention: a randomized controlled trial. JMIR Mhealth Uhealth 2016 Apr 20;4(2):e36 [FREE Full text] [doi: 10.2196/mhealth.5202] [Medline: 27098111]
54. Feil EG, Sprengelmeyer PG, Leve C. A randomized study of a mobile behavioral parent training application. Telemed J E Health 2018 Jun;24(6):457-463 [FREE Full text] [doi: 10.1089/tmj.2017.0137] [Medline: 29232180]
55. Hemdi A, Daley D. The effectiveness of a psychoeducation intervention delivered via WhatsApp for mothers of children with autism spectrum disorder (ASD) in the kingdom of Saudi Arabia: a randomized controlled trial. Child Care Health Dev 2017 Nov 20;43(6):933-941. [doi: 10.1111/cch.12520] [Medline: 28929582]
56. Lefever JE, Bigelow KM, Carta JJ, Borkowski JG, Grandfield E, McCune L, et al. Long-term impact of a cell phone-enhanced parenting intervention. Child Maltreat 2017 Nov 28;22(4):305-314. [doi: 10.1177/1077559517723125] [Medline: 28845676]
57. Mason MJ, Coatsworth JD, Russell M, Khatri P, Bailey S, Moore M, et al. Reducing risk for adolescent substance misuse with text-delivered counseling to adolescents and parents. Subst Use Misuse 2021 May 13;56(9):1247-1257. [doi: 10.1080/10826084.2021.1910709] [Medline: 33985404]
58. Schaeffer CM, Dimeff LA, Koerner K, Chung J, Kelley-Brimer A, Kako N, et al. A smartphone app for parental management of adolescent conduct problems: randomized clinical trial of iKinnect. J Clin Child Adolesc Psychol 2022 May 31:1-15. [doi: 10.1080/15374416.2022.2070851] [Medline: 35640058]
59. Sonne T, Müller J, Marshall P, Obel C, Grønbæk K. Changing family practices with assistive technology: MOBERO improves morning and bedtime routines for children with ADHD. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. 2016 Presented at: CHI '16; May 7-12, 2016; San Jose, CA p. 152-164 URL: https://dl.acm.org/doi/10.1145/2858036.2858157 [doi: 10.1145/2858036.2858157]
60. Sullivan AD, Breslend NL, Strolin-Goltzman J, Bielawski-Branch A, Jorgenson J, Deaver AH, et al. Feasibility investigation: leveraging smartphone technology in a trauma and behavior management-informed training for foster caregivers. Child Youth Serv Rev 2019 Jun;101:363-371. [doi: 10.1016/j.childyouth.2019.03.051]
61. Bronfenbrenner U. Ecological systems theory. In: Bronfenbrenner U, editor. Making Human Beings Human: Bioecological Perspectives on Human Development. Thousand Oaks, CA: Sage Publications; 1992:106-173.
62. Patterson GR. The early development of coercive family process. In: Reid JB, Patterson GR, Snyder J, editors. Antisocial Behavior in Children and Adolescents: A Developmental Analysis and Model for Intervention. New York, NY: American Psychological Association; 2002:25-44.
63. Ritterband LM, Thorndike FP, Cox DJ, Kovatchev BP, Gonder-Frederick LA. A behavior change model for internet interventions. Ann Behav Med 2009 Aug;38(1):18-27 [FREE Full text] [doi: 10.1007/s12160-009-9133-4] [Medline: 19802647]
64. Boyatzis CJ, Varghese R. Children's emotional associations with colors. J Genet Psychol 1994 Mar;155(1):77-85. [doi: 10.1080/00221325.1994.9914760] [Medline: 8021626]
65. Child IL, Hansen JA, Hornbeck FW. Age and sex differences in chilåçdren's color preferences. Child Develop 1968 Mar;39(1):237. [doi: 10.2307/1127374]
66. Gyu “Phillip” Park J. Correlations between color attributes and children's color preferences. Color Res Appl 2013 Mar 03;39(5):452-462. [doi: 10.1002/col.21801]
67. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. Psychol Rev 1977;84(2):191-215. [doi: 10.1037/0033-295X.84.2.191]
68. Scott S, Dadds MR. Practitioner review: when parent training doesn't work: theory-driven clinical strategies. J Child Psychol Psychiatry 2009 Dec 05;50(12):1441-1450. [doi: 10.1111/j.1469-7610.2009.02161.x] [Medline: 19754503]
69. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
70. Hu PJ, Chau PY, Sheng OR, Tam KY. Examining the technology acceptance model using physician acceptance of telemedicine technology. Manag Inf Syst 2015 Dec 02;16(2):91-112. [doi: 10.1080/07421222.1999.11518247]
71. Michie S, Johnston M. Theories and techniques of behaviour change: developing a cumulative science of behaviour change. Health Psychol Rev 2012 Mar;6(1):1-6. [doi: 10.1080/17437199.2012.654964]
72. Jeminiwa RN, Hohmann NS, Fox BI. Developing a theoretical framework for evaluating the quality of mHealth apps for adolescent users: a systematic review. J Pediatr Pharmacol Ther 2019 Jul;24(4):254-269 [FREE Full text] [doi: 10.5863/1551-6776-24.4.254] [Medline: 31337988]
73. Kazdin AE. Behavior Modification in Applied Settings. 7th edition. Long Grove, IL: Waveland Press; 2012.
74. Jones DJ, Forehand R, McKee LG, Cuellar J, Kincaid C. Behavioral parent training: is there an "app" for that? Behav Ther (N Y N Y) 2010 Apr;33(4):72-77 [FREE Full text] [Medline: 22199418]
75. May CD, St George JM, Fletcher RJ, Dempsey I, Newman LK. Coparenting competence in parents of children with ASD: a marker of coparenting quality. J Autism Dev Disord 2017 Oct 7;47(10):2969-2980. [doi: 10.1007/s10803-017-3208-z] [Medline: 28688075]

76. Chacko A, Jensen SA, Lowry LS, Cornwell M, Chimklis A, Chan E, et al. Engagement in behavioral parent training: review of the literature and implications for practice. Clin Child Fam Psychol Rev 2016 Dec;19(3):204-215. [doi: 10.1007/s10567-016-0205-2] [Medline: 27311693]
77. Weisenmuller C, Hilton D. Barriers to access, implementation, and utilization of parenting interventions: considerations for research and clinical applications. Am Psychol 2021 Jan;76(1):104-115. [doi: 10.1037/amp0000613] [Medline: 32134281]
78. Cao H, Jiang S, Lv M, Wu T, Chen W, Zhang J. Effectiveness of the Alfalfa app in warfarin therapy management for patients undergoing venous thrombosis prevention and treatment: cohort study. JMIR Mhealth Uhealth 2021 Mar 02;9(3):e23332 [FREE Full text] [doi: 10.2196/23332] [Medline: 33650976]
79. Dennison L, Morrison L, Conway G, Yardley L. Opportunities and challenges for smartphone applications in supporting health behavior change: qualitative study. J Med Internet Res 2013 Apr;15(4):e86 [FREE Full text] [doi: 10.2196/jmir.2583] [Medline: 23598614]
80. Barber BK, Maughan SL, Olsen JA. Patterns of parenting across adolescence. New Dir Child Adolesc Dev 2005;2005(108):5-16. [doi: 10.1002/cd.124] [Medline: 16121893]
81. Carr A. The effectiveness of family therapy and systemic interventions for child‐focused problems. J Fam Ther 2008 Dec 23;31(1):3-45. [doi: 10.1111/j.1467-6427.2008.00451.x]
82. Stanger C, Budney AJ, Kamon JL, Thostensen J. A randomized trial of contingency management for adolescent marijuana abuse and dependence. Drug Alcohol Depend 2009 Dec 01;105(3):240-247 [FREE Full text] [doi: 10.1016/j.drugalcdep.2009.07.009] [Medline: 19717250]
83. Doherty G, Coyle D, Sharry J. Engagement with online mental health interventions: an exploratory clinical study of a treatment for depression. In: Proceedings of the 2012 SIGCHI Conference on Human Factors in Computing Systems. 2012 Presented at: CHI '12; May 5-10, 2012; Austin, TX p. 1421-1430 URL: https://dl.acm.org/doi/10.1145/2207676.2208602 [doi: 10.1145/2207676.2208602]

Abbreviations

mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by L Buis; submitted 26.07.23; peer-reviewed by J Egunjobi; comments to author 01.04.24; revised version received 20.05.24; accepted 28.05.24; published 24.09.24.

Please cite as:
Magnuson KI, Li K, Beuley G, Ryan-Pettes SR
The Use of Noncommercial Parent-Focused mHealth Interventions for Behavioral Problems in Youth: Systematic Review
JMIR Mhealth Uhealth 2024;12:e51273
URL: https://mhealth.jmir.org/2024/1/e51273
doi:10.2196/51273
PMID:39316435

Review

Evaluating the Adoption of mHealth Technologies by Community Health Workers to Improve the Use of Maternal Health Services in Sub-Saharan Africa: Systematic Review

Chiyembekezo Kachimanga[1], MBBS, MPhil; Haules Robbins Zaniku[2,3], BSc; Titus Henry Divala[2], PhD; Johannes C.F Ket[4]; Joia S Mukherjee[5], MD; Daniel Palazuelos[5], MD; Alexandra V Kulinkina[6,7], PhD; Ibukun-Oluwa Omolade Abejirinde[8,9], PhD; Thomas van den Akker[1,10], PhD

[1]Athena Institute, Vrije Universiteit Amsterdam, Amsterdam, Netherlands
[2]Kamuzu University of Health Sciences, Blantyre, Malawi
[3]Ministry of Health, Neno, Malawi
[4]Medical Library, Vrije Universiteit Amsterdam, Amsterdam, Netherlands
[5]Partners In Health, Boston, MA, United States
[6]Swiss Tropical and Public Health Institute, Allschwil, Switzerland
[7]University of Basel, Basel, Switzerland
[8]Women College Hospital Institute for Health System Solutions and Virtual Care, Toronto, ON, Canada
[9]Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada
[10]Department of Obstetrics and Gynaecology, Leiden University Medical Center, Leiden, Netherlands

Corresponding Author:
Chiyembekezo Kachimanga, MBBS, MPhil
Athena Institute
Vrije Universiteit Amsterdam
De Boelelaan 1105
Amsterdam, 1081 HV
Netherlands
Phone: 31 887247777
Email: c.kachimanga@vu.nl

Abstract

Background: Limited information exists on the impact of mobile health (mHealth) use by community health workers (CHWs) on improving the use of maternal health services in sub-Saharan Africa (SSA).

Objective: This systematic review addresses 2 objectives: evaluating the impact of mHealth use by CHWs on antenatal care (ANC) use, facility-based births, and postnatal care (PNC) use in SSA; and identifying facilitators and barriers to mHealth use by CHWs in programs designed to increase ANC use, facility-based births, and PNC use in SSA using a sociotechnical system approach.

Methods: We searched for articles in 6 databases (MEDLINE, CINAHL, Web of Science, Embase, Scopus, and Africa Index Medicus) from inception up to September 2022, with additional articles identified from Google Scholar. After article selection, 2 independent reviewers performed title and abstract screening, full-text screening, and data extraction using Covidence software (Veritas Health Innovation Ltd). In addition, we manually screened the references lists of the included articles. Finally, we performed a narrative synthesis of the outcomes.

Results: Among the 2594 records retrieved, 10 (0.39%) studies (n=22, 0.85% articles) met the inclusion criteria and underwent data extraction. The studies were published between 2012 and 2022 in 6 countries. Of the studies reporting on ANC outcomes, 43% (3/7) reported that mHealth use by CHWs increased ANC use. Similarly, of the studies reporting on facility-based births, 89% (8/9) demonstrated an increase due to mHealth use by CHWs. In addition, in the PNC studies, 75% (3/4) showed increased PNC use associated with mHealth use by CHWs. Many of the studies reported on the importance of addressing factors related to the social environment of mHealth-enabled CHWs, including the perception of CHWs by the community, trust, relationships, digital literacy, training, mentorship and supervision, skills, CHW program ownership, and the provision of incentives. Very few studies reported on how program goals and culture influenced mHealth use by CHWs. Providing free equipment, accessories,

and internet connectivity while addressing ongoing challenges with connectivity, power, the ease of using mHealth software, and equipment maintenance support allowed mHealth-enabled CHW programs to thrive.

Conclusions: mHealth use by CHWs was associated with an increase in ANC use, facility-based births, and PNC use in SSA. Identifying and addressing social and technical barriers to the use of mHealth is essential to ensure the success of mHealth programs.

Trial Registration: PROSPERO CRD42022346364; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=346364

(JMIR Mhealth Uhealth 2024;12:e55819) doi:10.2196/55819

KEYWORDS

maternal health; antenatal care; postnatal care; facility-based births; sub-Saharan Africa; mobile health; mHealth; review; narrative synthesis; mobile phone

Introduction

Background

Sub-Saharan Africa (SSA) continues to have the highest maternal morbidity and mortality globally [1,2]. The region contributes only 15% of the world's population [3] while accounting for 70% of all maternal deaths [4]. In 2020, SSA had a maternal mortality ratio (MMR) of 545 maternal deaths per 100,000 live births and a 1 in 40 lifetime risk of maternal death [5]. These estimates are significantly higher than in any other region of the world [5].

Between 2000 and 2015, the MMR decreased in many regions, including SSA [6]. Unfortunately, recent trends in the MMR have not been promising; it has either increased or remained the same between 2016 and 2020 [7,8]. There are projections that the MMR may stay the same or increase by 2030 [9]. As such, there is a need for innovative approaches to reduce the MMR.

One way to accelerate the reduction in the MMR is to improve the coverage of maternal health services [7]. This includes improving the use of maternal health services within the continuum of care, that is, ≥4 antenatal care (ANC) contacts, facility-based births attended by skilled attendants, and early postnatal care (PNC) [7]. Providing high-quality care along the maternal health continuum has been shown to reduce maternal mortality [10-14]. However, the use of available services remains a significant challenge in SSA; for example, the use of ANC and PNC among women of reproductive ages between 15 to 24 years in 28 SSA countries was only 55% and 40%, respectively [15]. Studies in SSA showed lower rates of facility-based births, with women in rural areas having lower rates of facility-based births than those in urban areas [16,17]. Using data from the most recent surveys in SSA countries, Wan et al [18] and Straneo et al [19] found that only 7 out of 10 pregnant women give birth in health facilities in the region.

The impact of community health workers (CHWs) on increasing the use of maternal health services has been established [20]. Working collaboratively with communities, health facilities, national ministries of health (MOHs), and international health agencies, CHWs advocate for improved care and reduce cultural and other barriers preventing women from accessing maternal health services [21]. In addition, CHWs provide education, identify and refer women seeking maternal health services to health facilities, and may offer case management for selected health conditions [22]. Therefore, CHWs are an essential component in reducing maternal deaths, increasing the use of maternal health services, and eventually achieving the United Nations' Sustainable Development Goal 3.

To improve the efficiency of the tasks and responsibilities carried out by CHWs and help support improvements in clinical outcomes, mobile health (mHealth) technologies are increasingly being introduced to CHW programs [23]. mHealth is the use of mobile and wireless technologies in health care [24]. In general, evidence has shown that mHealth can improve outcomes in patients with chronic diseases, tuberculosis, and HIV infection [25]. In maternal health, mHealth has been shown to improve the coverage of ANC, facility-based births, and PNC [26,27]. However, previous reviews have not specifically examined the impact of mHealth tools when used by CHWs as the primary implementers. Specifically for CHWs, mHealth has been used to train them, improve their performance and retention, support data collection, support patient adherence to medication, and provide clinical decision support [28-32]. There is no review on mHealth use by CHWs to improve the maternal health continuum of care in SSA.

Objectives

Although some of the results in the aforementioned studies hold promise regarding the use of mHealth by CHWs in general, there is a lack of robust evidence on the impact of mHealth use by CHWs to improve the use of services along the maternal health continuum of care in SSA, especially when compared with CHWs who do not use mHealth. This review aims to provide evidence synthesis on the impact of mHealth use by CHWs in SSA to improve the coverage of maternal health services in comparison with CHWs not using mHealth. In addition, it examines the factors that support or hinder the successful implementation of mHealth for the improved use of maternal health services. These aims are captured in 2 objectives. First, we assessed the impact of mHealth use by CHWs on ANC use, facility-based births, and PNC use in SSA, comparing the outcomes with those achieved by CHWs not using mHealth. Second, we reviewed the facilitators and barriers to mHealth use by CHWs in programs designed to increase ANC use, facility-based births, and PNC use.

Methods

Overview

This systematic review adheres to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist (Multimedia Appendix 1) and guidelines [33]. As this is a mixed methods systematic review, we synthesized and integrated findings from both quantitative and qualitative studies to provide a more comprehensive understanding of the research question. We registered the review with PROSPERO (CRD42022346364), and the protocol was published previously [34].

Eligibility Criteria

The review included studies involving eligible women of reproductive age (15-49 y) using care across the maternal health continuum: pregnant women using ANC at any gestational age, pregnant women accessing intrapartum care at health care facilities, and women accessing PNC up to 42 days after giving birth regardless of the mode of giving birth. We included all studies that reported the use of mHealth by CHWs to improve the use of these 3 services. CHWs were included if they met the definition set by the World Health Organization (WHO): "health workers based in communities...who are either paid or volunteer, who are not professionals, and who have fewer than 2 years training but at least some training" [35]. For intervention studies included in the review, the comparator was CHW programs that were not using mHealth. We included ANC visits, facility-based births, and PNC visits as outcomes. We also collected qualitative data about facilitators and barriers to mHealth use by CHWs as described in the included studies. Further details and review criteria are outlined in the published protocol [34].

Search Strategy and Data Sources

We searched 6 databases (Scopus, MEDLINE, CINAHL, Web of Science, Embase, and Africa Index Medicus) from inception up to September 2022. To develop the concepts for search terms, we combined the following concepts: "women accessing maternal health services (pregnancy OR childbirth OR postnatal care) AND mHealth AND community health workers AND SSA countries." We included all known synonyms and related terms identified from the literature. We adapted the search terms to each database. Due to challenges in reproducibility, we used Google Scholar as part of reference checking [34]. For Google Scholar, we developed search terms mirroring these major concepts and searched the first 1000 results for any new articles that met the inclusion criteria but were not captured by the other databases. The search strategy, including search terms for Google Scholar, is presented in Multimedia Appendix 2. We included randomized controlled trials (RCTs), quasi-experimental studies, nonexperimental quantitative studies, qualitative studies, and mixed methods studies that met the inclusion criteria. We excluded reviews and other summary-type articles, policy documents, commentaries, abstracts and conference proceedings, case reports, and protocols. Manual searches were conducted by reviewing the references lists of systematic reviews and all included articles. The included articles were limited to the SSA region as defined by the World Bank [36]. To reduce language bias, ensure the identification of all relevant studies regardless of publication time, and allow the generalization of findings to SSA, we did not limit the search by language or year of publication.

Data Extraction

JCFK retrieved all studies from the electronic databases. CK and HRZ independently screened the titles and abstracts of the extracted articles, independently assessed full-text articles for inclusion, and conducted manual searches. An audit log was kept throughout the process, including documentation of reasons for exclusion. CK and HRZ performed the data extraction. TvdA resolved all discrepancies. Covidence software (Veritas Health Innovation Ltd) was used for the screening and data extraction [37].

We extracted information on the authors; publication year; study designs; country and geographic scope; the type and scope of work performed by CHWs; and mHealth characteristics, including platforms used, devices used, delivery methods, and the functions of mHealth. We also extracted the study results based on the outcomes as well as the facilitators and barriers to mHealth use.

Risk-of-Bias Assessment, Analysis, and Synthesis

All included studies underwent a risk-of-bias assessment conducted by CK and HRZ using the Mixed Methods Assessment Tool [38]. Regardless of the results of the risk-of-bias assessment, we included all studies in data extraction, analysis, and synthesis. Due to the heterogeneity of the studies, we conducted a narrative synthesis using the 3 steps proposed by Popay et al [39]. Narrative synthesis involves synthesizing data using words and text, rather than statistical methods that are often used for qualitative studies or when meta-analysis is not possible due to heterogeneity in the study designs or outcomes.

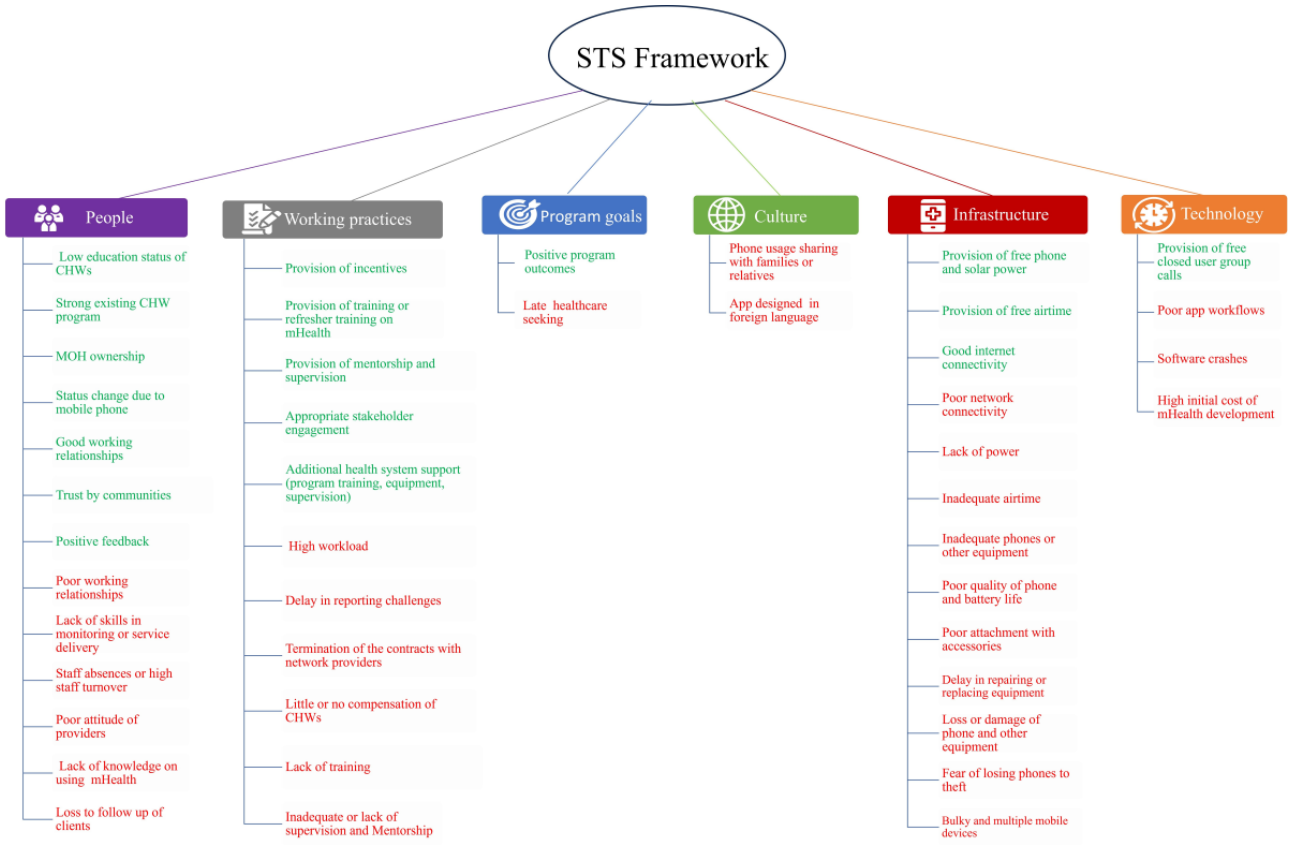
First, a preliminary synthesis based on the review objectives was performed. For the impact of mHealth use by CHWs on the use of maternal health services, we described the direction and size of the impact on ANC attendance, facility-based births, and PNC attendance, with the results tabulated. For facilitators and barriers to mHealth use, we inductively conducted a thematic analysis to identify barriers and facilitators to mHealth use by CHWs [40]. After thematic analysis, we synthesized and reflected on the results using the sociotechnical system (STS) framework developed by Davis et al [41]. The STS framework is not specific to CHWs or mHealth; however, it provides a useful lens for understanding how various dimensions—such as (1) people, (2) working practices, (3) program goals, (4) culture, (5) infrastructure, and (6) technology—influence the implementation and use of new technologies. The first 4 cover the social dimension of mHealth, while the last 2 cover the technology dimension of mHealth [42]. Applying the STS framework allowed a comprehensive and up-to-date analysis of the barriers and facilitators of both the technical and social systems of mHealth, rather than focusing on technology alone as has been done in other digital health studies [43,44]. Further discussion on the dimensions of the STS is presented in the

Results section, while Figure 1 outlines details of the 6 dimensions of the STS framework.

Second, we explored within- and between-study relationships to describe the variability in the results before integrating and synthesizing the findings.

Finally, we assessed the robustness of the synthesis by reflecting on the methodology and the results in the Discussion section.

Figure 1. Facilitators and barriers mapped according to the sociotechnical system (STS) framework. mHealth: mobile health.

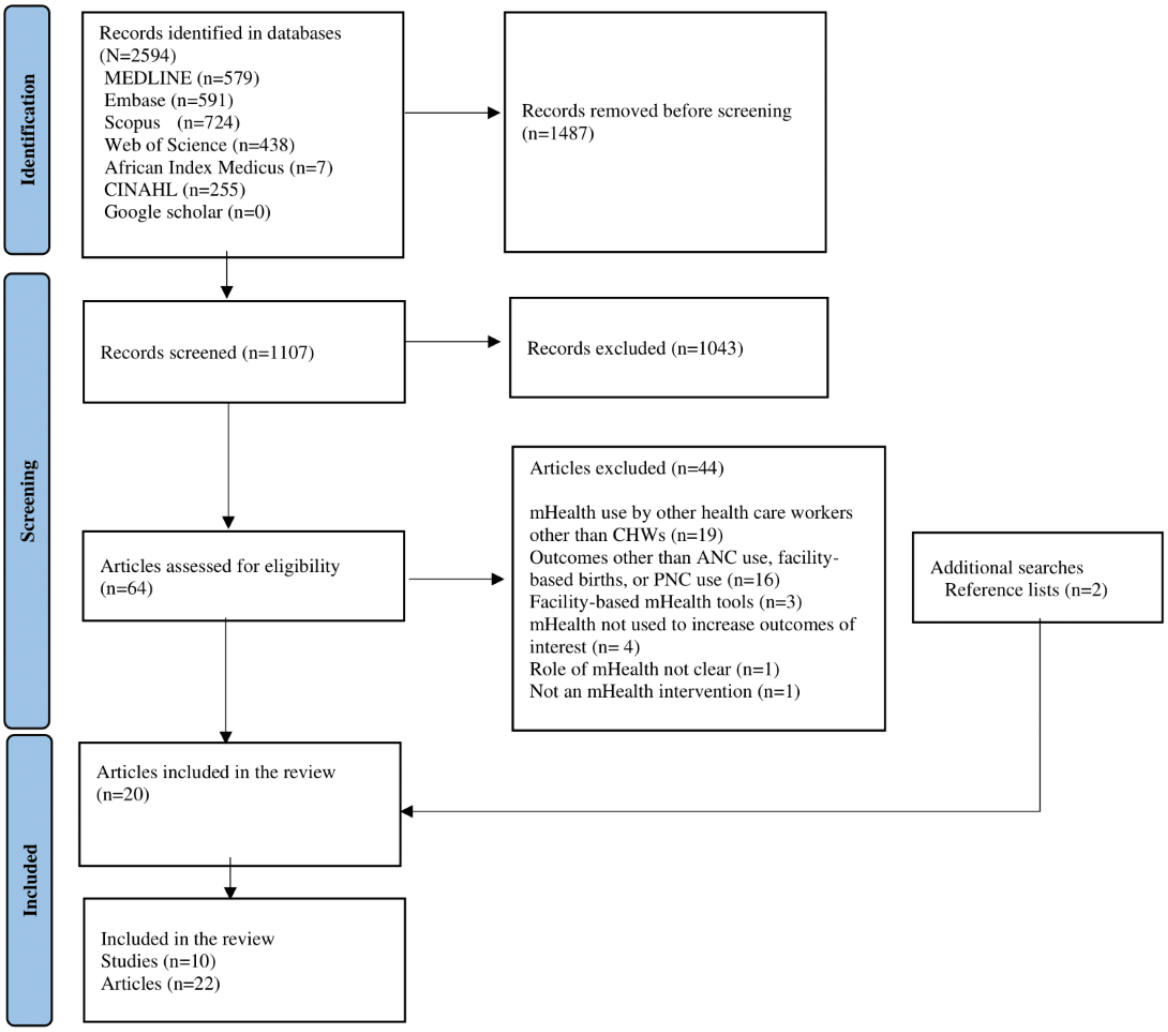


Results

Summary of the Studies

We retrieved 2594 records from all databases (refer to Figure 2 for the PRISMA flow diagram). After removing 1487 (57.32%) duplicates from the 2594 records, 1107 (42.68%) articles underwent title and abstract screening. Of these 1107 articles, 1043 (94.22%) were excluded, leaving 64 (5.78%) articles. After full-text screening, 20 (31%) of the 64 articles were included. We found 2 additional articles from reference searching, resulting in a final number of 10 studies comprising 22 articles. The reasons for exclusion of 44 articles after full-text screening included mHealth use by health care workers other than CHWs (n=19, 43%); outcomes other than ANC use, facility-based births, or PNC use (n=16, 36%); facility-based mHealth tools (n=3, 7%); mHealth not used to improve the outcomes of interest (n=4, 9%); unclear role of mHealth (n=1, 2%); and not an mHealth intervention (n=1, 2%; Multimedia Appendix 3).

Figure 2. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram. ANC: antenatal care; CHW: community health worker; mHealth: mobile health; PNC: postnatal care.



Characteristics of the Studies

The 10 studies were published between 2012 and 2022 (refer to the Risk-of-Bias Assessment subsection for the study designs used). Of the 10 studies, 3 (30%) were published in Tanzania [45-47]; 2 (20%) each in Ethiopia [48,49] and Uganda [50,51]; and 1 (10%) each in Rwanda [52], Mozambique [53], and Kenya [54]. The mHealth program for Rwanda was implemented nationally, while the other mHealth programs were implemented either as a pilot or at a subnational level. Half of the mHealth platforms used were mobile apps (5/10, 50%) [45-47,49,53], followed by SMS text messaging–based platforms (4/10, 40%) [48,50,52,54] and voice calls (1/10, 10%) [51].

Risk-of-Bias Assessment

All 22 included articles were assessed for risk of bias using the Mixed Methods Assessment Tool (Multimedia Appendix 4 [55-58]). Of the 4 qualitative articles, 3 (75%) were assessed as low risk of bias [55,56,59], and 1 (25%) had an unclear risk of bias [57]. Of the 4 RCTs, 1 (25%) had an unclear risk of bias [53], while 3 (75%) had a high risk of bias [45,48,51]. Of the 7 quantitative nonrandomized articles, 3 (43%) had a low risk of bias [46,52,60], while 4 (57%) had a high risk of bias [50,58,61,62]. Of the 2 descriptive articles, 1 (50%) had a low risk of bias [47], and 1 (50%) had a high risk of bias [63]. Finally, of the 5 mixed methods articles, 1 (20%) had a low risk of bias [54], while 4 (80%) had a high risk of bias [49,64-66].

Impact of mHealth Use by CHWs on the Use of Maternal Health Services

Overview

The 10 included studies (13/22, 59% articles) reported at least 1 outcome of interest (Table 1). We present the findings based on the outcome of interest.

Table 1. Quantitative studies reporting on the impact of mobile health (mHealth) use by community health workers (CHWs).

Study, year; outcomes of interest	Study design	Country (context)	Intervention description	mHealth description	Main findings
Sevene et al [53], 2020; ANC[a] and facility-based births	Cluster randomized controlled trial	Mozambique (subnational, rural, and urban districts)	• Intervention: CHWs made home visits to pregnant women to provide education, danger signs identification, referrals, birth preparedness, and PNCb; CHWs also measured blood pressure and administered drugs for severe hypertensive disease; other interventions included transport support and health talks at health facilities • Control: there were 6 control clusters where CHWs conducted home visits without using mHealth; no other interventions were provided	• Type: Mobile app (PIERS[c] On the Move) • Use: the app was used for clinical decision support during the CHW home visit; CHWs used the app to identify pregnant women with danger signs (using pictograms), and if danger signs were present, they would refer the client to health facilities; if no danger signs were present, CHWs measured blood pressure, and if women met predefined criteria for referral, they would be referred immediately to an appropriate health facility	• ANC: no difference. In comparison to control clusters, no difference in ≥4 ANC visits (48.6% vs 42.5%, aOR[d] 1.57, 99% CI 0.97-2.52) • Facility-based births: no difference. No differences in facility-based births (67.3% vs 74.2%, aOR 0.80, 99% CI 0.28-2.61; *P*=.71) or births at comprehensive emergency obstetric and neonatal care facilities (11.3% vs 13%, aOR 0.85, 99% CI 0.27-2.62; *P*=.70)
Hackett et al [45], 2018; facility-based births	Cluster randomized controlled trial	Tanzania (pilot and rural district)	• Intervention: in 16 clusters, CHWs were trained in integrated maternal, neonatal, and child health; CHWs also conducted household visits to educate and refer clients to care during pregnancy and the postpartum period; mHealth app was used during the home visits • Control: CHWs trained in integrated maternal, neonatal, and child health but used paper-based tools in 16 control clusters	• Type: mobile app (developed using CommCare, an open-source platform designed specifically for use by frontline health workers) • Use: CHWs used the app to register pregnant women, counsel pregnant women, identify danger signs, flag women who needed referrals, and create reminders for CHWs to follow up with the women they referred	• Facility-based births: increase in use. A majority of pregnant women (74%) gave birth in transit or at the facility in the intervention villages compared with those in the control villages (62%); the odds of facility delivery were 2 times the odds between the intervention and control group (OR[e] 1.96, CI 1.21-3.19; *P*=.01)
Atnafu et al [48], 2017; ANC and facility-based births	Cluster randomized controlled trial	Ethiopia (subnational, rural, and urban districts)	• Intervention: the intervention targeted pregnant women in 2 clusters. Partial: higher cadre CHWs (ie, HEWs[f]) in 1 cluster were provided mHealth, while lower cadre volunteer CHWs were responsible for community mobilization, education, and referrals but were not provided mobile phones. Full: in 1 cluster, HEWs were provided mHealth, and volunteer CHWs were provided regular mobile phones for voice calls • Control: no mHealth use by CHWs; CHWs performed home visits and community mobilization	• Type: SMS text messaging based (FrontlineSMS) and regular mobile phone for voice calls • Use: mHealth was used for SMS text messaging reminders to CHWs to follow up with pregnant women at 14, 24, 30, and 36 weeks of pregnancy to encourage them to attend all ANC visits and give birth at health facilities; in addition, voice calls were used to arrange referrals and communication between HEWs and volunteer CHWs; other functions included data collection and reporting as well as supply chain management	• ANC: increase in the use of ANC. High ANC attendance at baseline at both intervention and control sites; significant increase in ≥4 ANC visits at intervention sites in comparison with control sites (partial intervention: 45.3%-59.8%; *P*<.001; full intervention: 15.8%-31.5%; *P*<.001; control sites: 24.5%-23.3%) • Facility-based births: decrease in home births. Home births decreased at all intervention sites (partial intervention: 61.6%-33.7%; full intervention: 50.7%-35.8%; control sites: 72.8%-58.5%)

Study, year; outcomes of interest	Study design	Country (context)	Intervention description	mHealth description	Main findings
Ayiasi et al [51], 2016; ANC and facility-based births	Cluster randomized controlled trial	Uganda (pilot, 2 rural districts)	• Intervention: existing CHWs (called village health teams) conducted 2 ANC home visits and 1 PNC home visit to provide standardized educational messages for maternal and newborn care in 8 clusters; in addition, each village health team had unlimited calls with health care workers for clinical consultation • Control: village health teams in 8 health clusters provided usual community mobilization	• Type: voice calls • Use: mobile phones were used for clinical consultation between CHWs and providers and to arrange for referrals from the communities to health facilities	• ANC: no significant differences in the rates of ANC visits. A majority of the women (85%) in the intervention clusters made ≥3 ANC visits compared with those in the control clusters (71%); aOR 1.82, 95% CI 0.65-5.09; *P*=.26) • Facility-based births: higher at intervention sites than at control sites. Three times higher (90%) at the intervention sites than at the control sites (28%); the intervention increased the odds of facility-based births by 18-fold (OR 17.94, 95% CI 6.3-51.4; *P*<.001)
Webber et al [46], 2022; Webber et al [64], 2019; ANC, facility-based births, and PNC	Interrupted time series	Tanzania (subnational, rural)	• Intervention: CHWs were trained to educate pregnant women on the importance of ANC, giving birth at facilities, and identifying danger signs using an mHealth app; other interventions implemented included birth kit distribution at 34 to 36 weeks' gestation and transport support • Control: CHWs who were not using mHealth	• Type: mobile app (based on community health toolkit) • Use: the mobile health app supported CHWs in providing education on the importance of attending ANC visits, giving birth at facilities, and identifying danger signs	• Early results for facility-based births: suggested an increase in facility-based births. Rates of facility-based births increased (Bunda town from 87% to 93%, Bunda rural from 70% to 84%, and Tarime from 48% to 67%) between 2015 and 2016 • Final evaluation for ANC: no difference. No difference in ≥4 ANC visits after introducing mHealth (immediate change: OR 1.19, 95% CI 0.93-1.51; *P*=.17; gradual effect: OR 1.02, 95% CI 0.99-1.05; *P*=.24) • Final evaluation for facility-based births: increase in use. There was an increase in facility-based births from 71.8% at baseline to 85.1% after the intervention, with an immediate increase in the odds of facility-based births (OR 1.51, 95% CI 1.14-2.01; *P*=.004) and a small gradual effect (OR 1.03, 95% CI 1.00-1.07; *P*=.06). • Final evaluation for PNC: no difference. No significant immediate change and a decline in the trend of PNC use after introducing mHealth (immediate change: OR 1.07, 95% CI 0.61-1.89; *P*=.81; gradual effect: OR 0.92, 95% CI 0.86-0.98; *P*=.01)
Hategeka et al [52], 2019; ANC, facility-based births, and PNC	Interrupted time series	Rwanda (national program, rural and urban districts)			

Study, year; outcomes of interest	Study design	Country (context)	Intervention description	mHealth description	Main findings
			• Intervention: CHWs with mHealth supported pregnant women with education, follow-up, and linkage to care across the maternal health continuum. The intervention was stratified: Facilities in 20 districts received the usual support. Facilities in 10 districts received health system strengthening by 2 NGOs[g] (including ongoing training for CHWs and equipment provision to health facilities) • Control: CHWs not using mHealth	• Type: SMS text messaging based • Use: the open-source RapidSMS platform was used to facilitate communication between the health system and CHWs, facilitate clinical appointments by using reminders, support clinical decisions by providing information on what CHWs should do during an emergency, and facilitate referrals in emergencies by linking CHWs with the ambulance drivers; other functions of the RapidSMS system included registries or vital events tracking, data collection and reporting, and use an electronic health record system	• Rwanda Demographic and Health Survey (2014-2015) data in 29 of the 30 districts: For ANC, no difference. No change in any ANC visits (immediate change: −1.00, 95% CI −2.30 to 0.29 and gradual effect: −0.04, 95% CI −0.14 to 0.06). No change in ≥4 ANC visits (immediate change: −1.69, 95% CI −9.94 to 6.55 and gradual effect: −0.40, 95% CI −1.09 to 0.27). No change in first trimester ANC visits (immediate change: −3.80, 95% CI −13.66 to 6.05 and gradual effect: −0.62, 95% CI −1.43 to 0.19). For facility-based births, no difference. No change in facility-based births; (immediate change: −1.79, 95% CI −6.16 to 2.58 and gradual effect: −0.13, 95% CI −0.49 to 0.22) • Routine health facility data in all 30 districts: For ANC, no difference. No change in all 30 districts (*P*=.51 and *P*=.70 for supported districts and *P*=.38 and *P*=.50 for nonsupported districts). Facility-based births: a gradual increase in facility-based births. At the supported sites, there was no immediate change in facility-based births, but there was a change in gradual effect (estimate: 0.015 facility-based births per 1000 population, 95% CI 0.007-0.023; *P*<.001); no change in nonsupported sites. For PNC, an immediate increase in PNC visits. Change in supported facilities (immediate increase of 0.11 PNC visits per 1000 population, 95% CI 0.033-0.179; *P*=.007) and no change in trend; no change in the rate of PNC visits in nonsupported districts (*P*=.13)
Asiki et al [50], 2018; facility-based births	Nested cohort study	Uganda (pilot, rural)	• Intervention: in 13 clusters, CHWs registered pregnancies and made monthly visits to pregnant women to relay SMS text messages and track outcomes • Control: in 13 villages, CHWs followed pregnancies monthly using paper-based forms	• Type: SMS text messaging based • Use: mHealth supported CHWs to register pregnancies. In addition, each month, SMS text messaging reminders were sent to CHWs regarding which pregnant women they should visit to relay targeted messages on timely and safe ANC and facility-based births	• Facility-based births: decreases in-home births. Intervention home delivery was 9.2%, and control home delivery was 22.4%; after controlling for confounders, the intervention arm had lower odds of home births (aOR 0.38, 95% CI 0.15-0.97)
Mushamiri et al [54], 2015; ANC		Kenya (subnational, rural)			

Study, year; outcomes of interest	Study design	Country (context)	Intervention description	mHealth description	Main findings
	Mixed methods; the quantitative component used a nonrandomized control group study design		• Intervention: CHWs provided a community-based package of care and followed up with pregnant women using mHealth • Control: CHWs provided a community-based package of care and did not use mHealth	• Type: SMS text messaging–based platform • Use: SMS text messaging reminders were sent to CHWs to remind pregnant women 3 days before the ANC visit once they were registered in the mHealth program; reminders were sent twice if the women failed to attend the ANC visit; reminders were sent up to 18 months after giving birth	• ANC: mHealth increased the use of ANC. In comparison to the women not followed by CHWs using mHealth, the women in the intervention group had 3 times the odds of making more ANC visits, even after adjusting for HIV infection (aOR 2.58, 95% CI 1.10-6.01)
Nigussie et al [49], 2021; ANC, facility-based births, and PNC	Mixed methods	Ethiopia (rural and urban, subnational)	• Intervention: CHWs registered, referred, followed up, and tracked pregnant women along the maternal health continuum of care; CHWs and their supervisors used mHealth to support their tasks	• Type: mobile app • Use: once clients were registered in the app, the system sent notifications and reminders to CHWs via the app to visit pregnant women and remind them to visit a health facility for ANC and PNC; these reminders were also sent through SMS text messaging to CHWs and clients; in addition, the program supported with free caller user group, facilitating referrals from the community to the facility	• ANC: suggests increase in use. Pregnancy registration in the first and second trimesters increased between the third quarter of 2017 and the second quarter of 2018 (eg, first trimester registration increased from approximately 0% to approximately 10%), with a corresponding decline in registration in the third and fourth trimesters • Facility-based births: suggests increase in use. Facility-based births increased from approximately 100 per quarter to >900 in the second quarter of 2018 • PNC: suggests increase in use. PNC visits increased from approximately 100 per quarter to >700 in the second quarter of 2018
Fulcher et al [47], 2021 [65]; facility-based births, and PNC	Descriptive process evaluation	Tanzania (rural and urban, subnational)	• Intervention: CHWs registered and enrolled pregnant women, as well as conducted 3 ANC home visits and 3 PNC visits using mHealth; other supporting interventions included community savings, transport support, and stakeholder engagement	• Type: Mobile app (Mangologic app) • Use: the app helped CHWs to know when to conduct home visits, identify pregnant women with danger signs and refer them to care, follow up with the women within 3 days of referral, and coordinate referrals with health facilities; the app was also used for data collection	• Early findings for facility-based births: suggests increase in use. Among 13,231 births, 75% gave birth at hospitals in comparison with a baseline of 35%. Early findings for PNC: suggests increase in use. PNC attendance at intervention sites was 88% in comparison with a baseline of 36% from the demographic health survey • Final evaluation for facility-based births: suggests increase in use. Health facility births increased from 60%, 70%, 80%, and 80% between years 1 and 4. Final evaluation for PNC: suggests increase in use. The number of women who attended PNC visits increased 60%, 60%, 70%, and 80% in years 1, 2, 3, and 4, respectively

[a]ANC: antenatal care.

[b]PNC: postnatal care.

[c]PIERS: preeclampsia integrated estimate of risk.

[d]aOR: adjusted odds ratio.

[e]OR: odds ratio.

[f]HEW: health extension worker.

[g]NGO: nongovernmental organization.

ANC Outcomes

Of the 10 studies, 7 (70%; 8/22, 36% articles) [46,48,49,51-54,60] reported at least 1 ANC outcome. On the basis of outcomes, only 2 (29%) of the 7 studies reported on ANC visits in the first trimester [49,52]. None of the studies reported on ≥8 ANC contacts. Of the 7 studies, 3 (43%) [48,49,54] showed an association between mHealth use by CHWs and increased ANC use. On the basis of platforms used, the studies that showed an association used SMS text messaging–based platforms (2/7, 29%) [48,54] and a mobile app (1/7, 14%) [49].

The observational study conducted in Ethiopia by Nigussie et al [49], which had a high risk of bias, suggested increased ANC use after mHealth use by CHWs. Using a mobile app, the study showed an increase in ANC contacts in the first trimester by approximately 10%.

A total of 3 RCTs (n=1, 33% with an unclear risk of bias [53]; n=2, 67% with a high risk of bias [48,51]) reported on ANC outcomes. Atnafu et al [48] used an SMS text messaging–based platform to remind CHWs to follow up pregnant women at 14, 24, 30, and 36 weeks of pregnancy in Ethiopia. The study found a significant increase in the proportion of women with ≥4 ANC visits at the intervention sites in comparison with the control sites (cluster with higher cadre CHWs using an SMS text messaging–based platform vs lower cadre CHWs who were not provided mobile phones: 45.3% to 59.8%; *P*<.001; cluster with higher cadre CHWs using an SMS text messaging–based platform vs lower cadre CHWs using mHealth for voice calls: 15.8% to 31.5%, *P*<.001; control sites [no mHealth]: 24.5% to 23.3%). The RCT conducted by Sevene et al [53] in Mozambique involved a clinical decision support mHealth app primarily used to support CHWs to screen for hypertension and make referrals to health facilities for care when they enrolled pregnant women during initial home visits or scheduled community follow-ups of pregnant women. As a secondary outcome, the proportion of women with ≥4 ANC visits was not statistically different between the intervention and control sites. Finally, the study conducted by Ayiasi et al [51] in Uganda, where CHWs used voice calls for consultations with health care workers during home visits, showed no significant difference in the number of women who attended ≥3 ANC visits between the intervention and control sites.

Three quasi-experimental studies (four articles) reported on ANC outcomes. Using an SMS text messaging–based mHealth platform in Kenya, Mushamiri et al [54] reported positive findings when comparing pregnant women receiving appointment reminders from CHWs using mHealth and pregnant women receiving CHW care without mHealth. In the study, women receiving care from CHWs using mHealth and starting ANC in the second trimester had 3 times the odds of attending ANC visits (adjusted odds ratio [OR] 2.58, 95% CI 1.10-6.01) than women receiving care from CHWs not using mHealth [54].

An SMS text messaging–based study (reported by 2 articles, 50% of the four articles that used quasi-experimental study designs) [52,60] used an interrupted time series design to evaluate a nationally implemented RapidSMS platform in Rwanda that enabled 2-way communication between CHWs and health care workers, facilitating clinical appointments of pregnant women by using reminders, supporting the clinical decisions by providing information on what CHWs should do during an emergency, and facilitating referrals during emergencies. After scaling the RapidSMS platform countrywide, 10 districts received health system–strengthening support from 2 nongovernmental organizations (NGOs; ongoing training provided to CHWs and equipment provided to health facilities). By contrast, the rest of the districts received the usual support from the Rwanda MOH. Using Rwanda Demographic and Health Survey (2014-2015) data in 29 of the 30 districts of Rwanda, Hategeka et al [52] found no change in any ANC visits, ANC visits in the first trimester, or ≥4 ANC visits. Using routinely collected health facility data in 461 health facilities, Ruton et al [60] found no change in ANC visits in all 30 districts of Rwanda. Finally, a quasi-experimental study with a time series design by Webber et al [46] used an mHealth app in Tanzania to educate women on the importance of attending maternal health services. The findings showed no significant differences in ≥4 ANC visits after introducing mHealth.

Facility-Based Births

Of the 10 studies, 9 (90%; 12/22, 55% articles) reported on facility-based births. Of these 9 studies, 8 (89%) found an association between mHealth use by CHWs and an increase in facility-based births or a reduction in home births. Of these 8 studies, 3 (38%) [48,49,54] used an SMS text messaging–based platform, 4 (50%) [45-47,49] used a mobile app, and 1 (12%) used voice calls [51].

A process evaluation study with a low risk of bias implemented in Tanzania by Fulcher et al [47], in which CHWs used a mobile app to increase demand for facility services by pregnant women, showed an increase in facility-based births from year 1 to year 4 of implementation (from 60% to 90%). Nigussie et al [49] also found an increase in facility-based births after the implementation of mHealth by CHWs.

In an RCT with a high risk of bias conducted in Tanzania, Hackett et al [45] compared the impact of the mHealth app in intervention clusters where CHWs were using the app and control clusters where CHWs were not using mHealth. The odds of facility-based births in the intervention clusters were double those in the control clusters (OR 1.96, 95% CI 1.21-3.19). The RCTs conducted by Atnafu et al [48] and Ayiasi et al [51] also reported a reduction in home births (home births decreased at all intervention sites [cluster with higher cadre CHWs using an SMS text messaging–based platform vs lower cadre CHWs not provided with mHealth: 61.6% to 33.7%; cluster with higher cadre CHWs using an SMS text messaging–based platform vs

lower cadre CHWs using mHealth for voice calls: 50.7% to 35.8%; control sites: 72.8% to 58.5%]) and an increase in facility-based births (the intervention increased the odds of facility-based births by 18-fold [OR 17.94, 95% CI 6.3-51.4; *P*<.001]), respectively. However, Sevene et al [53] found no change in facility-based births between the intervention and control clusters.

Two quasi-experimental studies in which CHWs used mobile apps found a positive impact of mHealth on improving facility-based birth rates and reducing home birth rates. The study conducted in Tanzania by Webber et al [46], which had a low risk of bias, found that mHealth increased the odds of facility births (immediate increase: OR 1.51, 95% CI 1.14-2.01; *P*=.004; gradual effect: OR 1.03, 95% CI 1.00-1.07; *P*=.06). Asiki et al [50], who conducted a study in Uganda that had a high risk of bias, compared the impact of an SMS text messaging–based platform on CHWs using SMS text messaging reminders to follow up on pregnancy outcomes and CHWs not using mHealth. After controlling for confounders, mHealth reduced the odds of home births (adjusted OR 0.38, 95% CI 0.15-0.97). Two quasi-experimental articles based on the RapidSMS study in Rwanda showed mixed results. Hategeka et al [52] found no change in facility-based births. Ruton et al [60] found no change in the 20 districts that were not supported by the 2 NGOs, while there was a change in the 10 supported districts (gradual effect at the 10 supported sites: 0.015 facility-based births per 1000 population per month, 95% CI 0.007-0.023; *P*<.001), signifying the extra role played by embedding mHealth into broader health system–strengthening initiatives.

PNC Outcomes

Of the 10 studies, 2 (20%) observational studies [47,49] and 2 (20%) quasi-experimental studies reported on the impact of mHealth use by CHWs on PNC visits [46,60]. Of these 4 studies, 3 (75%) showed a positive association (n=2, 67% used a mobile app [47,49], and n=1, 33% used SMS text messaging [60]).

Fulcher et al [47] found that the mHealth app used by CHWs increased any PNC visits from 60% to 80% within 4 years of implementing the program. Nigussie et al [49] also showed an increase in any PNC visits after mHealth use by CHWs. In Rwanda, Ruton et al [60] reported an increase of 100% in PNC visits within a year of starting mHealth in the 10 districts that received extra NGO support (immediate increase of 0.11 PNC visits per 1000 population, 95% CI 0.033-0.179). However, the rate of PNC visits remained the same in the 20 districts not receiving health system strengthening. Finally, the study conducted in Tanzania by Webber et al [46] showed no impact of the mHealth app on PNC visits.

Facilitators and Barriers to mHealth Use

Overview

Of the 10 studies, 8 (80%; 14/22, 64% articles) reported on facilitators and barriers to mHealth uptake (Table 2; Figure 1). We will discuss facilitators and barriers simultaneously and in alignment with the 6 dimensions of the STS framework developed by Davis et al [41]: program goals, people, culture, working practices, infrastructure, and technology. The definitions of the dimensions are provided in each subsection that follows.

Table 2. Studies reporting on the barriers and facilitators to mobile health (mHealth) use by community health worker (CHWs).

Study, year	Article, year	Study design	mHealth description	Study aims and findings
Sevene et al [53], 2020	Boene et al [66], 2021	Mixed methods	Mobile app	• Facilitators: training on mHealth, refresher training, mentorship and supervision, improved status as health care workers through smartphone use, trust of communities, and good working relationships with clients • Barriers: poor battery life, difficulty in securely attaching accessories, lack of power to charge mobile phones, and poor network connectivity
Atnafu et al [48], 2017	Atnafu [61], 2015; Atnafu and Bisrat [62], 2015	Cross-sectional	SMS text messaging based	• Facilitators: good working relations with supervisors, as well as availability of free mobile phone and free airtime • Barriers: poor network connectivity, lack of power to charge mobile phones, an inadequate number of mobile phones or other equipment, loss or damage of mobile phone and other equipment, and inadequate airtime
Ayiasi et al [51], 2016	Ayiasi et al [55], 2015	Qualitative	Voice calls	• Facilitators: smartphones improved their status as health care workers, trust of communities, good working relations with supervisors, and positive supervisor feedback • Barriers: staff absences, lack of power to charge mobile phones, poor network connectivity, poor attitude of health care workers, and poor relationships with clients
Webber et al [46], 2022	Webber et al [64], 2019; Webber et al [57], 2020	Mixed methods; qualitative	Mobile app	• Facilitators: free mobile phone, airtime, and solar chargers • Barriers: poor network connectivity, lack of power to charge mobile phones, inadequate airtime, poor app navigation. and poor app workflows
Hategeka et al [52], 2019	Ngabo et al [63], 2012; Musabyimana et al [56], 2018; Mwendwa [59], 2016; Mwendwa [58], 2018	Descriptive, qualitative; qualitative; cross-sectional	SMS text messaging based	• Facilitators: MOHa ownership, incentives, strong existing CHW program, additional health system–strengthening activities, appropriate stakeholder engagement, high education level of CHWs, training, positive feedback, positive program outcomes, trust of communities, good working relations with supervisors, and good network connectivity • Barriers: high initial cost of development, illiteracy, poor network connectivity, lack of power to charge mobile phones, inadequate airtime, high workload due to both paper- and smartphone-based data entry, poor organization of training, app in a foreign language, inadequate supervision, an inadequate number of mobile phones or other equipment, poor relationships with clients, loss or damage of mobile phone and other equipment, and no stipend or salary for CHWs
Nigussie et al [49], 2021	Nigussie et al [49], 2021	Mixed methods	Mobile app	• Facilitators: appropriate stakeholder engagement and MOH ownership, as well as additional health system–strengthening activities • Barriers: high workload due to both paper- and smartphone-based data entry, mobile phone–sharing culture, a fear of losing mobile phones to theft, delay in reporting the challenges of mHealth, a lack of skills in monitoring service delivery, loss or damage of mobile phone and other equipment, poor network connectivity, inadequate airtime, delay in repairing or replacing equipment, and burden of carrying multiple mobile devices

Study, year	Article, year	Study design	mHealth description	Study aims and findings
Fulcher et al [47], 2021	Fulcher et al [47], 2021	Descriptive process evaluation	Mobile app	• Barriers: poor network connectivity, high loss to follow-up of clients, high attrition of CHWs, and software crashes
Mushamiri et al [54], 2015	Mushamiri et al [54], 2015	Mixed methods	SMS text messaging based	• Facilitator: free closed user group calls and good network connectivity • Barriers: contract termination with a network provider and late health-seeking behavior

[a]MOH: Ministry of Health.

People

The STS identifies this dimension as encompassing the users and stakeholders of a system and their characteristics. The included articles explored facilitators and barriers affecting CHWs, who are the main stakeholders, as well as their surrounding environment (eg, communities and supervisors), and their characteristics (eg, attitudes, behavior, and skills).

An unintended benefit yet powerful facilitator of mHealth use was the effect of the mHealth devices on the status of CHWs. mHealth improved the social status of CHWs because they were perceived as being recognized by the formal health system [55,66]. Perceived higher social status improved community trust, another facilitator identified in multiple studies [55,59,66]. Other facilitators identified included higher education level of CHWs [58,59], strong existing CHW program [63], MOH ownership of CHW programs [49,63], and positive feedback from supervisors [55,58,59].

The social dynamics within and outside the CHW program also positively or negatively impacted mHealth use. Good working relationships and positive feedback between CHWs, their supervisors, health care workers, and communities facilitated mHealth use [55,59,62,66]. By contrast, poor relationships among CHWs and communities and health care workers were singled out as barriers to mHealth use [49,55,56]. Other barriers identified in the studies included a lack of skills in monitoring service delivery by CHW supervisors [49], high staff absences and turnover [47,55], CHW illiteracy [59], poor attitude by facility-based staff [55], and high loss to follow-up of clients [47].

Working Practices

Processes refer to how practices and procedures are organized to support the system's uptake. The included articles described the influence of systems, practices, and procedures designed to support CHWs in effectively using mHealth. Additional existing health system–strengthening activities facilitated positive outcomes in some mHealth-enabled CHW programs [49]; for example, additional training, provision of extra equipment and supplies, and supervision in Rwanda's RapidSMS program facilitated mHealth use by CHWs [60]. In addition, engaging multiple and appropriate stakeholders, including telecommunication companies, facilitated the use of mHealth [49,63]. In a study in Kenya, suspending a contract with a mobile communication provider for a few weeks was one of the most significant barriers to SMS text messaging–based mHealth rollout [54].

Several other processes were reported in the studies. Adequate training and refresher training on mHealth [58,66], strong mentorship and supervision of CHWs [66], and enhancing the education levels of CHWs facilitated mHealth use [59]. By contrast, a high CHW workload [49], poor reporting systems [49,59], a lack of training or poor training organization [59], and inadequate supervision [59] were reported as barriers to mHealth use. The provision of incentives or salaries to CHWs was also mentioned, with regular incentives or salaries as a facilitator [63] and a lack of incentives or salaries or low incentives or salaries as a barrier to mHealth use [56].

Program Goals

This dimension explores how program performance and outcomes influence the uptake of a system. Very few studies reported on the impact of program goals on mHealth uptake by CHWs, and only 1 facilitator and 1 barrier were identified. Positive program outcomes (eg, he role played by CHWs in reducing mortality rates) in the RapidSMS program in Rwanda reinforced the use of mHealth [56,58,59]. The study conducted in Kenya reported late presentation to ANC as a barrier [54]. This factor had a detrimental impact on the mHealth-driven goals of ANC visits within the scope of this study.

Culture

In the STS framework, this dimension examines the influence of users' and stakeholders' beliefs, norms, and values within a system. Among the included studies, there was a lack of literature addressing how culture affected the use of mHealth. None of the studies reported on cultural factors that might facilitate mHealth adoption. The studies conducted in Ethiopia [49,62] identified the sharing of an mHealth-enabled mobile phone with other family members, a common norm, as a barrier to mHealth use because the mobile phone was often unavailable to CHWs when needed. In the study by Atnafu and Bisrat [62], more than a third of the CHWs reported sharing the mobile phone with other family members. In addition, the studies based on the RapidSMS platform in Rwanda cited the use of a foreign language (ie, English) in the app as a cultural barrier [58,59].

Infrastructure

This dimension encompasses the assets of a system. Among the included studies, mHealth equipment and internet connectivity were frequently noted as influential factors in the uptake of mHealth. Free mobile phones [57,62] as well as free airtime or reimbursement of airtime costs [57,62] were facilitators of mHealth use. The provision of solar power for charging devices was also reported as a facilitator [57]. Finally, reliable internet

connectivity was mentioned as a facilitator of mHealth use [54,63].

Most of the studies focused on barriers associated with mHealth devices and related equipment. The most commonly reported challenges included loss or damage of mobile phones and other equipment [49,62,63], delays in repairing or replacing equipment [49], the burden of carrying multiple mobile devices [49], a fear of losing mobile phones to theft [49], an inadequate number of mobile phones or other equipment [56,62], poor quality of mobile phones and battery life [66], and difficulty in securely attaching mobile phone accessories [66]. In many settings, especially rural settings, the lack of electricity was commonly reported as a barrier [55,57,59,61-64,66]. Finally, poor mobile network connectivity [47,49,55,57,59,61,62,64,66] and inadequate airtime [49,59,61,64] were also mentioned as barriers to mHealth use.

Technology

This dimension focuses on how mHealth software influences the uptake of mHealth use. Free closed user group calls facilitated the use of a voice-based mHealth platform; no other facilitators were identified [54]. However, several technology-related barriers were reported, including poor app workflows [57,64] and frequent app software crashes [47]. The study by Ngabo et al [63] reported the high cost of mHealth technology as a barrier.

Discussion

Principal Findings

This review is the first to synthesize evidence on the use of mHealth by CHWs to improve the use of maternal health services in SSA. While the results of most of the studies (8/9, 89%) supported that mHealth use by CHWs increase facility-based births, they are mixed for ANC and PNC use. For ANC and PNC, only approximately half (3/7, 43%) and three quarters of the studies (3/4, 75%) showed that mHealth use by CHWs increased the use of these services, respectively. On the basis of the intervention descriptions, mHealth use by CHWs may have increased use by creating demand for these services. As shown previously, the demand created by mHealth use is possible through multiple pathways. mHealth use by CHWs may have increased the knowledge of good maternal health practices, leading to behavior change toward health facility use for care [47]. This is particularly important for studies that use mHealth apps. In addition, primarily through SMS text messaging–based platforms, reminders may have encouraged pregnant women's attendance at health services [67]. It is also possible that mHealth may have increased demand by increasing satisfaction and trust in CHWs and the health system and may also have improved adherence to the practices used by CHWs to increase demand for health services [68].

This review adds evidence on the impact of mHealth use on the use of maternal health services. Previous reviews have focused on the impact of mHealth use on maternal health outcomes but did not distinguish the primary users of mHealth. A previous systematic review conducted by Gayesa et al [26] in low- and lower–middle-income countries found that mHealth use increased the odds of facility-based births and PNC use [26]. Similarly, Wagnew et al [69] reported that SMS text messaging–based mHealth increased the use of ≥4 ANC visits as well as facility-based births in low- and middle-income countries. Rahman et al [70] and Sondaal et al [27] also report positive effects of mHealth use on ANC coverage and facility-based births in low- and middle-income countries. However, our study is unique because it presents the effect of mHealth use specifically by CHWs, a target group not explored in the other studies. In addition, this is the first review to focus on SSA specifically.

We found mixed results on the impact of mHealth use by CHWs on ANC visits. Of the 7 studies that reported on ANC attendance, 3 (43%) showed that mHealth use by CHWs may increase ANC use. While the SMS text messaging–based mHealth studies in Kenya and Ethiopia [48,54] and the mobile app study in Ethiopia [49] found increases in the overall number of ANC visits or ≥4 ANC visits, all other studies (4/7, 57%) found no effect. We also observed that many studies reported ≥4 ANC visits as an outcome. Reporting on ≥4 ANC visits may be attributed to the previous WHO recommendation, according to which 4 ANC visits were deemed adequate [71,72]. Only 2 (29%) of the 7 studies reported on ANC visits in the first trimester, and none of the studies reported on ≥8 ANC contacts as recommended by recent WHO guidelines [73]. As observed in many settings, especially in SSA, women start ANC attendance late [74,75], which may affect the number of ANC visits; therefore, we suggest designing mHealth programs to specifically focus on encouraging early ANC attendance (Textbox 1). Designing mHealth to support women to start ANC attendance earlier may have 2 advantages. First, mHealth may help identify pregnancies early through decision support and referrals to health facilities. Second, mHealth may support the provision of high-quality, community-based ANC by CHWs [76]. This approach would ensure that mHealth supports the recommended ANC contacts as well as the quality and outcomes of these community contacts.

Textbox 1. A summary of recommendations for mobile health (mHealth) implementation in sub-Saharan Africa (SSA).

Maternal health continuum of care outcomes

- Antenatal care (ANC): design mHealth programs measuring the impact of mHealth on ANC attendance in the first trimester and ≥8 ANC contacts.
- Postnatal care: design more studies to measure the impact of mHealth on postnatal care.

Coverage of mHealth

- National scale-up of mHealth programs is required in settings where mHealth has been shown to work but is being implemented as a pilot or at the subnational level.
- Where national scale-up of mHealth is desired, consider adding health system–strengthening activities in addition to mHealth intervention.
- Scale up mHealth platforms that have been shown to work in other settings and countries in SSA.

Choice of mHealth platform

- Consider SMS text messaging–based platforms and mobile apps.

Sociotechnical system dimension requiring further data

- Design studies to measure the influence of culture and program goals on mHealth use.

We found very few studies (4/10, 40%) reporting on PNC outcomes compared with the studies reporting on ANC (7/10, 70%) and facility-based birth (9/10, 90%) outcomes. Of the 4 studies, 1 (25%) SMS text messaging–based study [60] and 2 (50%) mobile app studies [47,49] found increases in PNC visits after implementing mHealth programs. We suggest more studies designed to specifically show the effect of mHealth on PNC visits because providing care during the postnatal period reduces maternal and neonatal deaths as well as complications [7] (Textbox 1).

The findings from this review also have implications for the scale-up of mHealth programs and mHealth platforms of choice in SSA. Of the 10 included studies, only 1 (10%), which was conducted in Rwanda, was implemented nationally [52], while the rest (n=9, 90%) were implemented as pilots or at the subnational level. In addition, the majority of the studies (9/10, 90%) used either SMS text messaging or an mHealth app. As the results of this review show the benefits of mHealth use by CHWs on maternal health outcomes, we suggest a national scale-up in pilot or subnational programs; alternatively, new programs may consider the scale-up of mHealth from the beginning. As a choice of platform, this review has shown that SMS text messaging–based platforms or mHealth apps may be used as the platforms of choice (Textbox 1).

This review has also identified facilitators and barriers to mHealth use by CHWs across the 6 dimensions of the STS framework. Most of the studies reported facilitators and barriers with regard to people, working practices, infrastructure, and technology among the 6 dimensions. One of the common findings concerned the perceived improvement in CHWs' status when they started using mHealth, as well as improved trust. This finding is echoed across other studies in the literature [77,78] and may be an essential reason for introducing mHealth in limited-resource settings. A recent review on the use of mHealth by CHWs, specifically smart devices, also found that mHealth improves CHW status [79]. As echoed by Perry et al [31], improving the status of CHWs and increasing their recognition by the formal health system is a crucial enabler for successful CHW programs, and mHealth may provide the pathway to achieve this. In addition, paying attention to the social environment of CHWs, including relationships, plays a vital role in the success of mHealth programs [80].

The findings from the processes, infrastructure, and technology dimensions reinforce the importance of strengthening CHW programs and health service infrastructure before the introduction of mHealth or as part of its implementation. Critical components such as MOH ownership and stakeholder engagement, as well as standardized and robust systems such as training and the provision of incentives or salaries to CHWs need to be considered to ensure the success of mHealth in many settings. Other studies in the literature have also emphasized the need to build systems and appropriate governance to address barriers related to mHealth equipment and evolving mHealth technologies [81,82]. Therefore, we recommend that the design and implementation of mHealth programs in SSA should include health system–strengthening activities to maximize the impact of mHealth tools (Textbox 1).

We need further studies across the culture and goals dimensions of the STS framework. First, very few of the included studies reported on the influence of culture on mHealth use by CHWs in SSA. Incorporating various aspects of culture, such as the local language, into mHealth improves its acceptability, usability, and effectiveness [83,84]. Unfortunately, even in the literature, there are limited studies reporting on the impact of culture on mHealth use [84,85], and the results were similar in this review [58,59,62]. More studies should be designed to explore the impact of culture on mHealth use. Second, further research is needed on the impact of clearly defined goals as enablers and barriers to mHealth (Textbox 1).

Limitations

This review has some limitations that should be considered when interpreting the findings. First, it focused on mHealth interventions by CHWs for women of reproductive age using ANC, giving birth at health facilities, and using PNC within 42 days in SSA. While this limits generalizability to other

populations, settings, and maternal health outcomes, it allowed for a targeted examination of mHealth impact on key services across the continuum of care in a region with high maternal mortality.

Second, the review included studies that implemented mHealth alongside other health system–strengthening activities, making it difficult to isolate the effect of mHealth alone. However, this reflects the real-world implementation of mHealth as a tool to enhance CHW service delivery within broader health systems, rather than as a stand-alone solution.

Third, the review may have missed some relevant studies by excluding gray literature. In addition, not assessing publication bias could mean that the included studies are skewed toward positive findings. However, a comprehensive search of 6 databases and reference checking was conducted, and all studies meeting inclusion criteria from these sources were included.

Finally, studies were excluded if they lacked sufficient information on the mHealth intervention, which could introduce selection bias. However, this was necessary to ensure that the review could meaningfully synthesize and interpret how mHealth was used to impact outcomes.

Recommendations

This study has implications for program implementation, policy, and research. Although our recommendations focus on mHealth implementation in SSA, we hope some lessons can be applied to other settings. We present a summary of recommendations in Textbox 1.

Conclusions

The study found evidence that mHealth use by CHWs results in an increase in facility-based births. Although the results are mixed, approximately half of the studies (3/7, 43%) that reported on ANC and 75% (3/4) of studies that reported PNC attendance showed that mHealth use by CHWs increased the use of these services. We found limited studies (2/7, 29%) measuring the impact of mHealth on increasing ANC visits in the first trimester and no study reported on ≥8 ANC visits. On the basis of the STS framework, most of the studies explored barriers and facilitators across the people, processes and procedures, building and infrastructure, and technology dimensions. More studies are needed on the culture and goals dimensions of the STS to better understand the impact and uptake of mHealth for improving maternal health outcomes.

Authors' Contributions

CK, THD, TvdA, and IOOA were responsible for conceptualization and study design. CK, HRZ, JCFK, IOOA, and AVK were responsible for data collection. CK wrote the first draft. CK, THD, JCFK, AVK, JSM, DP, IOOA, and TvdA were responsible for the critical review and final approval of the manuscript.

Conflicts of Interest

DP also works for Brigham and Women's Hospital and Harvard Medical School and has done short-term contract work with Vanna Health, the World Health Organization, PIVOT, and Last Mile Health. All other authors declare no conflicts of interest.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOCX File , 34 KB - mhealth_v12i1e55819_app1.docx]

Multimedia Appendix 2
Search strategy.
[DOCX File , 68 KB - mhealth_v12i1e55819_app2.docx]

Multimedia Appendix 3
Excluded studies and reasons for exclusion.
[DOCX File , 74 KB - mhealth_v12i1e55819_app3.docx]

Multimedia Appendix 4
Risk-of-bias assessments.
[DOCX File , 37 KB - mhealth_v12i1e55819_app4.docx]

References

1. Geller SE, Koch AR, Garland CE, MacDonald EJ, Storey F, Lawton B. A global view of severe maternal morbidity: moving beyond maternal mortality. Reprod Health 2018 Jun 22;15(Suppl 1):98 [FREE Full text] [doi: 10.1186/s12978-018-0527-2] [Medline: 29945657]
2. Lawrence ER, Klein TJ, Beyuo TK. Maternal mortality in low and middle-income countries. Obstet Gynecol Clin North Am 2022 Dec;49(4):713-733. [doi: 10.1016/j.ogc.2022.07.001] [Medline: 36328676]

3. Population (sub-Saharan Africa & world). The World Bank Group. 2022. URL: https://data.worldbank.org/indicator/SP.POP.TOTL?end=2021&locations=ZG-1W&start=2021&view=map [accessed 2022-09-25]
4. Maternal mortality. World Health Organization. URL: https://www.who.int/news-room/fact-sheets/detail/maternal-mortality [accessed 2024-04-27]
5. Trends in maternal mortality 2000 to 2020: estimates by WHO, UNICEF, UNFPA, World Bank Group and UNDESA/Population Division. World Health Organization. URL: https://www.who.int/publications/i/item/9789240068759 [accessed 2024-04-29]
6. Stanton ME, Kwast BE, Shaver T, McCallon B, Koblinsky M. beyond the safe motherhood initiative: accelerated action urgently needed to end preventable maternal mortality. Glob Health Sci Pract 2018 Aug 08;6(3):408-412. [doi: 10.9745/ghsp-d-18-00100]
7. Ending preventable maternal mortality (EPMM): a renewed focus for improving maternal and newborn health and well-being. World Health Organization. 2021. URL: https://www.who.int/publications/i/item/9789240040519 [accessed 2024-04-29]
8. Strategies towards ending preventable maternal mortality (EPMM). World Health Organization. 2015. URL: https://www.paho.org/en/documents/strategies-toward-ending-preventable-maternal-mortality-epmm [accessed 2024-04-29]
9. Zarocostas J. Global maternal mortality rates stagnating. Lancet 2023 Feb 25;401(10377):632. [doi: 10.1016/S0140-6736(23)00385-9] [Medline: 36842436]
10. Zuñiga JA, García A, Kyololo OB, Hamilton-Solum P, Kabimba A, Milimo B, et al. Increasing utilisation of skilled attendants at birth in sub-Saharan Africa: a systematic review of interventions. Int J Nurs Stud 2021 Aug;120:103977. [doi: 10.1016/j.ijnurstu.2021.103977] [Medline: 34144356]
11. Merdad L, Ali MM. Timing of maternal death: levels, trends, and ecological correlates using sibling data from 34 sub-Saharan African countries. PLoS One 2018;13(1):e0189416 [FREE Full text] [doi: 10.1371/journal.pone.0189416] [Medline: 29342157]
12. Campbell OM, Graham WJ. Strategies for reducing maternal mortality: getting on with what works. Lancet 2006 Oct;368(9543):1284-1299. [doi: 10.1016/s0140-6736(06)69381-1]
13. Lassi ZS, Mansoor T, Salam RA, Das JK, Bhutta ZA. Essential pre-pregnancy and pregnancy interventions for improved maternal, newborn and child health. Reprod Health 2014;11 Suppl 1(Suppl 1):S2 [FREE Full text] [doi: 10.1186/1742-4755-11-S1-S2] [Medline: 25178042]
14. Lassi ZS, Kumar R, Mansoor T, Salam RA, Das JK, Bhutta ZA. Essential interventions: implementation strategies and proposed packages of care. Reprod Health 2014;11 Suppl 1(Suppl 1):S5 [FREE Full text] [doi: 10.1186/1742-4755-11-S1-S5] [Medline: 25178110]
15. Bain LE, Aboagye RG, Dowou RK, Kongnyuy EJ, Memiah P, Amu H. Prevalence and determinants of maternal healthcare utilisation among young women in sub-Saharan Africa: cross-sectional analyses of demographic and health survey data. BMC Public Health 2022 Apr 05;22(1):647 [FREE Full text] [doi: 10.1186/s12889-022-13037-8] [Medline: 35379198]
16. Udo IE, Doctor HV. Trends in health facility births in sub-Saharan Africa: an analysis of lessons learned under the millennium development goal framework. Afr J Reprod Health 2016 Sep;20(3):108-117. [doi: 10.29063/ajrh2016/v20i3.15] [Medline: 29553200]
17. Doctor HV, Nkhana-Salimu S, Abdulsalam-Anibilowo M. Health facility delivery in sub-Saharan Africa: successes, challenges, and implications for the 2030 development agenda. BMC Public Health 2018 Jun 19;18(1):765 [FREE Full text] [doi: 10.1186/s12889-018-5695-z] [Medline: 29921275]
18. Wan S, Jin B, Kpordoxah MR, Issah A, Yeboah D, Aballo J, et al. A descriptive analysis of the coverage of newborn care services among women who delivered in health facilities in 17 sub-Saharan African countries. BMC Pregnancy Childbirth 2023 Apr 17;23(1):256. [doi: 10.1186/s12884-023-05592-8]
19. Straneo M, Hanson C, van den Akker T, Afolabi BB, Asefa A, Delamou A, et al. Inequalities in use of hospitals for childbirth among rural women in sub-Saharan Africa: a comparative analysis of 18 countries using Demographic and Health Survey data. BMJ Glob Health 2024 Jan 22;9(1):e013029 [FREE Full text] [doi: 10.1136/bmjgh-2023-013029] [Medline: 38262683]
20. Kachimanga C, Dunbar EL, Watson S, Cundale K, Makungwa H, Wroe EB, et al. Increasing utilisation of perinatal services: estimating the impact of community health worker program in Neno, Malawi. BMC Pregnancy Childbirth 2020 Jan 06;20(1):22 [FREE Full text] [doi: 10.1186/s12884-019-2714-8] [Medline: 31906939]
21. Hodgins S, Kok M, Musoke D, Lewin S, Crigler L, LeBan K, et al. Community health workers at the dawn of a new era: 1. Introduction: tensions confronting large-scale CHW programmes. Health Res Policy Syst 2021 Oct 12;19(Suppl 3):109 [FREE Full text] [doi: 10.1186/s12961-021-00752-8] [Medline: 34641886]
22. Glenton C, Javadi D, Perry HB. Community health workers at the dawn of a new era: 5. Roles and tasks. Health Res Policy Syst 2021 Oct 12;19(Suppl 3):128 [FREE Full text] [doi: 10.1186/s12961-021-00748-4] [Medline: 34641903]
23. Agarwal S, Perry HB, Long LA, Labrique AB. Evidence on feasibility and effective use of mHealth strategies by frontline health workers in developing countries: systematic review. Trop Med Int Health 2015 Aug;20(8):1003-1014 [FREE Full text] [doi: 10.1111/tmi.12525] [Medline: 25881735]
24. Istepanian RS. Mobile health (m-Health) in retrospect: the known unknowns. Int J Environ Res Public Health 2022 Mar 22;19(7):3747 [FREE Full text] [doi: 10.3390/ijerph19073747] [Medline: 35409431]

25. Marcolino MS, Oliveira JA, D'Agostino M, Ribeiro AL, Alkmim MB, Novillo-Ortiz D. The impact of mHealth interventions: systematic review of systematic reviews. JMIR Mhealth Uhealth 2018 Jan 17;6(1):e23 [FREE Full text] [doi: 10.2196/mhealth.8873] [Medline: 29343463]
26. Gayesa RT, Ngai FW, Xie YJ. The effects of mHealth interventions on improving institutional delivery and uptake of postnatal care services in low-and lower-middle-income countries: a systematic review and meta-analysis. BMC Health Serv Res 2023 Jun 09;23(1):611 [FREE Full text] [doi: 10.1186/s12913-023-09581-7] [Medline: 37296420]
27. Sondaal SF, Browne JL, Amoakoh-Coleman M, Borgstein A, Miltenburg AS, Verwijs M, et al. Assessing the effect of mHealth interventions in improving maternal and neonatal care in low- and middle-income countries: a systematic review. PLoS One 2016;11(5):e0154664 [FREE Full text] [doi: 10.1371/journal.pone.0154664] [Medline: 27144393]
28. Early J, Gonzalez C, Gordon-Dseagu V, Robles-Calderon L. Use of mobile health (mHealth) technologies and interventions among community health workers globally: a scoping review. Health Promot Pract 2019 Nov;20(6):805-817. [doi: 10.1177/1524839919855391] [Medline: 31179777]
29. Odendaal WA, Anstey Watkins J, Leon N, Goudge J, Griffiths F, Tomlinson M, et al. Health workers' perceptions and experiences of using mHealth technologies to deliver primary healthcare services: a qualitative evidence synthesis. Cochrane Database Syst Rev 2020 Mar 26;3(3):CD011942 [FREE Full text] [doi: 10.1002/14651858.CD011942.pub2] [Medline: 32216074]
30. Krah EF, de Kruijf JG. Exploring the ambivalent evidence base of mobile health (mHealth): a systematic literature review on the use of mobile phones for the improvement of community health in Africa. Digit Health 2016;2:2055207616679264 [FREE Full text] [doi: 10.1177/2055207616679264] [Medline: 29942576]
31. Perry HB, Chowdhury M, Were M, LeBan K, Crigler L, Lewin S, et al. Community health workers at the dawn of a new era: 11. CHWs leading the way to "Health for All". Health Res Policy Syst 2021 Oct 12;19(Suppl 3):111 [FREE Full text] [doi: 10.1186/s12961-021-00755-5] [Medline: 34641891]
32. Winters N, Langer L, Geniets A. Scoping review assessing the evidence used to support the adoption of mobile health (mHealth) technologies for the education and training of community health workers (CHWs) in low-income and middle-income countries. BMJ Open 2018 Jul 30;8(7):e019827 [FREE Full text] [doi: 10.1136/bmjopen-2017-019827] [Medline: 30061430]
33. Page MJ, Moher D, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. BMJ 2021 Mar 29;372:n160 [FREE Full text] [doi: 10.1136/bmj.n160] [Medline: 33781993]
34. Kachimanga C, Divala TH, Ket JC, Kulinkina AV, Zaniku HR, Murkherjee J, et al. Adoption of mHealth technologies by community health workers to improve the use of maternal health services in sub-Saharan Africa: protocol for a mixed method systematic review. JMIR Res Protoc 2023 May 04;12:e44066 [FREE Full text] [doi: 10.2196/44066] [Medline: 37140981]
35. What do we know about community health workers? A systematic review of existing reviews. World Health Organization. 2020. URL: https://www.who.int/publications/i/item/whatdo-we-know-about-community-health-workers-a-systematic-review-of-existing-reviews#:~:text=This%20human%20resource%20group%20has,inequities%20in%20health%20status%20and [accessed 2024-04-29]
36. Sub-Saharan Africa. The World Bank Group. 2022. URL: https://www.worldbank.org/en/region/afr/overview#:~:text=According%20to%20the%20region's%20most,poised%20to%20exacerbate%20this%20fragility [accessed 2024-07-13]
37. Covidence systematic review software. Veritas Health Innovation. URL: https://www.covidence.org/ [accessed 2024-04-29]
38. Hong QN, Pluye P, Fàbregues S, Bartlett G, Boardman F, Cargo M, et al. Improving the content validity of the mixed methods appraisal tool: a modified e-Delphi study. J Clin Epidemiol 2019 Jul;111:49-59.e1 [FREE Full text] [doi: 10.1016/j.jclinepi.2019.03.008] [Medline: 30905698]
39. Popay J, Roberts H, Sowden A, Petticrew M, Arai L, Rodgers M, et al. Guidance on the conduct of narrative synthesis in systematic reviews: a product from the ESRC methods programme. Lancaster University. 2006. URL: https://tinyurl.com/2uu24jzs [accessed 2024-04-29]
40. Nowell LS, Norris JM, White DE, Moules NJ. Thematic analysis: striving to meet the Trustworthiness criteria. Int J Qual Methods 2017 Oct 02;16(1):160940691773384. [doi: 10.1177/1609406917733847]
41. Davis MC, Challenger R, Jayewardene DN, Clegg CW. Advancing socio-technical systems thinking: a call for bravery. Appl Ergon 2014 Mar;45(2):171-180. [doi: 10.1016/j.apergo.2013.02.009] [Medline: 23664481]
42. Hughes HP, Clegg CW, Bolton LE, Machon LC. Systems scenarios: a tool for facilitating the socio-technical design of work systems. Ergonomics 2017 Oct;60(10):1319-1335 [FREE Full text] [doi: 10.1080/00140139.2017.1288272] [Medline: 28277171]
43. Rahimi B, Nadri H, Lotfnezhad Afshar H, Timpka T. A systematic review of the technology acceptance model in health informatics. Appl Clin Inform 2018 Dec;9(3):604-634 [FREE Full text] [doi: 10.1055/s-0038-1668091] [Medline: 30112741]
44. Jacob C, Sezgin E, Sanchez-Vazquez A, Ivory C. Sociotechnical factors affecting patients' adoption of mobile health tools: systematic literature review and narrative synthesis. JMIR Mhealth Uhealth 2022 May 05;10(5):e36284 [FREE Full text] [doi: 10.2196/36284] [Medline: 35318189]

45. Hackett K, Lafleur C, Nyella P, Ginsburg O, Lou W, Sellen D. Impact of smartphone-assisted prenatal home visits on women's use of facility delivery: results from a cluster-randomized trial in rural Tanzania. PLoS One 2018 Jun 18;13(6):e0199400 [FREE Full text] [doi: 10.1371/journal.pone.0199400] [Medline: 29912954]
46. Webber G, Chirangi B, Magatti N, Mallick R, Taljaard M. Improving health care facility birth rates in Rorya District, Tanzania: a multiple baseline trial. BMC Pregnancy Childbirth 2022 Jan 27;22(1):74 [FREE Full text] [doi: 10.1186/s12884-022-04408-5] [Medline: 35086508]
47. Fulcher IR, Nelson AR, Tibaijuka JI, Seif SS, Lilienfeld S, Abdalla O, et al. Improving health facility delivery rates in Zanzibar, Tanzania through a large-scale digital community health volunteer programme: a process evaluation. Health Policy Plan 2021 Feb 16;35(10):1-11. [doi: 10.1093/heapol/czaa068] [Medline: 33263749]
48. Atnafu A, Otto K, Herbst CH. The role of mHealth intervention on maternal and child health service delivery: findings from a randomized controlled field trial in rural Ethiopia. Mhealth 2017 Sep;3:39 [FREE Full text] [doi: 10.21037/mhealth.2017.08.04] [Medline: 29184891]
49. Nigussie ZY, Zemicheal NF, Tiruneh GT, Bayou YT, Teklu GA, Kibret ES, et al. Using mHealth to improve timeliness and quality of maternal and newborn health in the primary health care system in Ethiopia. Glob Health Sci Pract 2021 Jul 30;9(3):668-681. [doi: 10.9745/ghsp-d-20-00685]
50. Asiki G, Newton R, Kibirige L, Kamali A, Marions L, Smedman L. Feasibility of using smartphones by village health workers for pregnancy registration and effectiveness of mobile phone text messages on reduction of homebirths in rural Uganda. PLoS One 2018 Jun 27;13(6):e0198653 [FREE Full text] [doi: 10.1371/journal.pone.0198653] [Medline: 29949593]
51. Mangwi Ayiasi R, Kolsteren P, Batwala V, Criel B, Orach CG. Effect of village health team home visits and mobile phone consultations on maternal and newborn care practices in Masindi and Kiryandongo, Uganda: a community-intervention trial. PLoS One 2016 Apr 21;11(4):e0153051 [FREE Full text] [doi: 10.1371/journal.pone.0153051] [Medline: 27101379]
52. Hategeka C, Ruton H, Law MR. Effect of a community health worker mHealth monitoring system on uptake of maternal and newborn health services in Rwanda. Glob Health Res Policy 2019 Mar 22;4(1):8 [FREE Full text] [doi: 10.1186/s41256-019-0098-y] [Medline: 30949586]
53. Sevene E, Sharma S, Munguambe K, Sacoor C, Vala A, Macuacua S, CLIP Mozambique Working Group. Community-level interventions for pre-eclampsia (CLIP) in Mozambique: a cluster randomised controlled trial. Pregnancy Hypertens 2020 Jul;21:96-105 [FREE Full text] [doi: 10.1016/j.preghy.2020.05.006] [Medline: 32464527]
54. Mushamiri I, Luo C, Iiams-Hauser C, Ben Amor Y. Evaluation of the impact of a mobile health system on adherence to antenatal and postnatal care and prevention of mother-to-child transmission of HIV programs in Kenya. BMC Public Health 2015 Feb 07;15(1):102 [FREE Full text] [doi: 10.1186/s12889-015-1358-5] [Medline: 25886279]
55. Mangwi Ayiasi R, Atuyambe LM, Kiguli J, Garimoi Orach C, Kolsteren P, Criel B. Use of mobile phone consultations during home visits by Community Health Workers for maternal and newborn care: community experiences from Masindi and Kiryandongo districts, Uganda. BMC Public Health 2015 Jun 18;15(1):560 [FREE Full text] [doi: 10.1186/s12889-015-1939-3] [Medline: 26084369]
56. Musabyimana A, Ruton H, Gaju E, Berhe A, Grépin KA, Ngenzi J, et al. Assessing the perspectives of users and beneficiaries of a community health worker mHealth tracking system for mothers and children in Rwanda. PLoS One 2018 Jun 7;13(6):e0198725 [FREE Full text] [doi: 10.1371/journal.pone.0198725] [Medline: 29879186]
57. Webber G, Chirangi B, Magatti N. Experiences of a multiple intervention trial to increase maternity care access in rural Tanzania: focus group findings with women, nurses and community health workers. Womens Health (Lond) 2020 Nov 04;16:1745506520969617 [FREE Full text] [doi: 10.1177/1745506520969617] [Medline: 33147116]
58. Mwendwa P. What encourages community health workers to use mobile technologies for health interventions? Emerging lessons from rural Rwanda. Dev Policy Rev 2017 Sep 11;36(1):111-129. [doi: 10.1111/dpr.12275]
59. Mwendwa P. Assessing the fit of RapidSMS for maternal and new-born health: perspectives of community health workers in rural Rwanda. Dev Pract 2015 Dec 15;26(1):38-51 [FREE Full text] [doi: 10.1080/09614524.2016.1112769]
60. Ruton H, Musabyimana A, Gaju E, Berhe A, Grépin KA, Ngenzi J, et al. The impact of an mHealth monitoring system on health care utilization by mothers and children: an evaluation using routine health information in Rwanda. Health Policy Plan 2018 Oct 01;33(8):920-927 [FREE Full text] [doi: 10.1093/heapol/czy066] [Medline: 30169638]
61. Atnafu A. Assessment of voluntary community health workers participation and contribution in mHealth intervention. Ethiop J Health Dev 2015;29(3):154-159 [FREE Full text]
62. Atnafu A, Bisrat A. Asessment of health extension workers (HEWs) knowledge, attitude and performance in mHealth intervention in Guraghe Zone of Ethiopia: cross-sectional study. Ethiop J Health Dev 2015;29(3):160-169 [FREE Full text]
63. Ngabo F, Nguimfack J, Nwaigwe F, Mugeni C, Muhoza D, Wilson DR, et al. Designing and implementing an innovative SMS-based alert system (RapidSMS-MCH) to monitor pregnancy and reduce maternal and child deaths in Rwanda. Pan Afr Med J 2012;13:31 [FREE Full text] [Medline: 23330022]
64. Webber GC, Chirangi BM, Magatti NJ. Challenges and successes of distributing birth kits with misoprostol to reduce maternal mortality in rural Tanzania. Afr J Reprod Health 2019 Sep;23(3):68-78. [doi: 10.29063/ajrh2019/v23i3.7] [Medline: 31782633]

65. Battle JD, Farrow L, Tibaijuka J, Mitchell M. mHealth for safer deliveries: a mixed methods evaluation of the effect of an integrated mobile health intervention on maternal care utilization. Healthc (Amst) 2015 Dec;3(4):180-184. [doi: 10.1016/j.hjdsi.2015.10.011] [Medline: 26699340]
66. Boene H, Valá A, Kinshella MW, La M, Sharma S, Vidler M, CLIP Mozambique Working Group. Implementation of the PIERS on the move mHealth application from the perspective of community health workers and nurses in rural Mozambique. Front Glob Womens Health 2021 May 3;2:659582 [FREE Full text] [doi: 10.3389/fgwh.2021.659582] [Medline: 34816216]
67. Bossman E, Johansen MA, Zanaboni P. mHealth interventions to reduce maternal and child mortality in sub-Saharan Africa and Southern Asia: a systematic literature review. Front Glob Womens Health 2022 Aug 25;3:942146 [FREE Full text] [doi: 10.3389/fgwh.2022.942146] [Medline: 36090599]
68. Kabongo EM, Mukumbang FC, Delobelle P, Nicol E. Explaining the impact of mHealth on maternal and child health care in low- and middle-income countries: a realist synthesis. BMC Pregnancy Childbirth 2021 Mar 09;21(1):196 [FREE Full text] [doi: 10.1186/s12884-021-03684-x] [Medline: 33750340]
69. Wagnew F, Dessie G, Alebel A, Mulugeta H, Belay YA, Abajobir AA. Does short message service improve focused antenatal care visit and skilled birth attendance? A systematic review and meta-analysis of randomized clinical trials. Reprod Health 2018;15(1):191 [FREE Full text] [doi: 10.1186/s12978-018-0635-z] [Medline: 30466453]
70. Rahman MO, Yamaji N, Nagamatsu Y, Ota E. Effects of mHealth interventions on improving antenatal care visits and skilled delivery care in low- and middle-income countries: systematic review and meta-analysis. J Med Internet Res 2022 Apr 22;24(4):e34061 [FREE Full text] [doi: 10.2196/34061] [Medline: 35451987]
71. Villar J, Ba'aqeel H, Piaggio G, Lumbiganon P, Miguel Belizán J, Farnot U, WHO Antenatal Care Trial Research Group. WHO antenatal care randomised trial for the evaluation of a new model of routine antenatal care. Lancet 2001 May 19;357(9268):1551-1564 [FREE Full text] [doi: 10.1016/s0140-6736(00)04722-x] [Medline: 11377642]
72. WHO antenatal care randomized trial: manual for the implementation of the new model. World Health Organization. 2002. URL: https://iris.who.int/handle/10665/42513 [accessed 2024-04-29]
73. WHO recommendations on antenatal care for a positive pregnancy experience. World Health Organization. 2016. URL: https://iris.who.int/bitstream/handle/10665/259947/WHO-RHR-18.02-eng.pdf [accessed 2024-04-29]
74. Mchenga M, Burger R, von Fintel D. Examining the impact of WHO's focused antenatal care policy on early access, underutilisation and quality of antenatal care services in malawi: a retrospective study. BMC Health Serv Res 2019 May 08;19(1):295 [FREE Full text] [doi: 10.1186/s12913-019-4130-1] [Medline: 31068183]
75. Jiwani SS, Amouzou-Aguirre A, Carvajal L, Chou D, Keita Y, Moran AC, et al. Timing and number of antenatal care contacts in low and middle-income countries: analysis in the Countdown to 2030 priority countries. J Glob Health 2020 Jun;10(1):010502 [FREE Full text] [doi: 10.7189/jogh.10.010502] [Medline: 32257157]
76. Kayentao K, Ghosh R, Guindo L, Whidden C, Treleaven E, Chiu C, et al. Effect of community health worker home visits on antenatal care and institutional delivery: an analysis of secondary outcomes from a cluster randomised trial in Mali. BMJ Glob Health 2023 Mar;8(3):e011071 [FREE Full text] [doi: 10.1136/bmjgh-2022-011071] [Medline: 36948531]
77. Madon S, Amaguru JO, Malecela MN, Michael E. Can mobile phones help control neglected tropical diseases? Experiences from Tanzania. Soc Sci Med 2014 Feb;102:103-110 [FREE Full text] [doi: 10.1016/j.socscimed.2013.11.036] [Medline: 24565147]
78. Khan NU, Rasheed S, Sharmin T, Ahmed T, Mahmood SS, Khatun F, et al. Experience of using mHealth to link village doctors with physicians: lessons from Chakaria, Bangladesh. BMC Med Inform Decis Mak 2015 Aug 05;15:62 [FREE Full text] [doi: 10.1186/s12911-015-0188-9] [Medline: 26242574]
79. Greuel M, Sy F, Bärnighausen T, Adam M, Vandormael A, Gates J, et al. Community health worker use of smart devices for health promotion: scoping review. JMIR Mhealth Uhealth 2023 Feb 22;11:e42023 [FREE Full text] [doi: 10.2196/42023] [Medline: 36811947]
80. LeBan K, Kok M, Perry HB. Community health workers at the dawn of a new era: 9. CHWs' relationships with the health system and communities. Health Res Policy Syst 2021 Oct 12;19(Suppl 3):116 [FREE Full text] [doi: 10.1186/s12961-021-00756-4] [Medline: 34641902]
81. Kruse C, Betancourt J, Ortiz S, Valdes Luna SM, Bamrah IK, Segovia N. Barriers to the use of mobile health in improving health outcomes in developing countries: systematic review. J Med Internet Res 2019 Oct 09;21(10):e13263 [FREE Full text] [doi: 10.2196/13263] [Medline: 31593543]
82. Gromisch ES, Turner AP, Haselkorn JK, Lo AC, Agresta T. Mobile health (mHealth) usage, barriers, and technological considerations in persons with multiple sclerosis: a literature review. JAMIA Open 2021 Jul;4(3):ooaa067 [FREE Full text] [doi: 10.1093/jamiaopen/ooaa067] [Medline: 34514349]
83. Garner SL, Koch H, George CE, Hitchcock J, Norman G, Green G, et al. Cross cultural team collaboration: integrating cultural humility in mHealth development and research. Inform Health Soc Care 2021 Dec 02;46(4):345-354. [doi: 10.1080/17538157.2021.1895168] [Medline: 33779468]
84. Maglalang DD, Yoo GJ, Ursua RA, Villanueva C, Chesla CA, Bender MS. "I don't have to explain, people understand": acceptability and cultural relevance of a mobile health lifestyle intervention for filipinos with type 2 diabetes. Ethn Dis 2017;27(2):143-154. [doi: 10.18865/ed.27.2.143] [Medline: 28439185]

85. Hilty DM, Crawford A, Teshima J, Nasatir-Hilty SE, Luo J, Chisler LS, et al. Mobile health and cultural competencies as a foundation for telehealth care: scoping review. J. technol. behav. sci 2021 Feb 15;6(2):197-230. [doi: 10.1007/s41347-020-00180-5]

Abbreviations

ANC: antenatal care
CHW: community health worker
mHealth: mobile health
MMR: maternal mortality ratio
MOH: ministry of health
NGO: nongovernmental organization
OR: odds ratio
PNC: postnatal care
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
SSA: sub-Saharan Africa
STS: sociotechnical system
WHO: World Health Organization

Edited by L Buis; submitted 26.12.23; peer-reviewed by TAR Sure, S Lee, YT Kuan; comments to author 26.03.24; revised version received 24.04.24; accepted 09.07.24; published 24.09.24.

Please cite as:
Kachimanga C, Zaniku HR, Divala TH, Ket JC, Mukherjee JS, Palazuelos D, Kulinkina AV, Abejirinde IOO, Akker TVD
Evaluating the Adoption of mHealth Technologies by Community Health Workers to Improve the Use of Maternal Health Services in Sub-Saharan Africa: Systematic Review
JMIR Mhealth Uhealth 2024;12:e55819
URL: https://mhealth.jmir.org/2024/1/e55819
doi:10.2196/55819
PMID:39316427

Review

Data Preprocessing Techniques for AI and Machine Learning Readiness: Scoping Review of Wearable Sensor Data in Cancer Care

Bengie L Ortiz[1*], PhD; Vibhuti Gupta[2*], PhD; Rajnish Kumar[1], PhD; Aditya Jalin[1*], BTECH, MRES; Xiao Cao[1], MS; Charles Ziegenbein[1,3], MS; Ashutosh Singhal[2], PhD; Muneesh Tewari[4,5,6,7,8], MD, PhD; Sung Won Choi[1,5*], MD, MS

[1]Department of Pediatrics, Hematology and Oncology Division, Michigan Medicine, University of Michigan Health System, Ann Arbor, MI, United States

[2]School of Applied Computational Sciences, Meharry Medical College, Nashville, TN, United States

[3]Autonomous Systems Research Department, Peraton Labs, Basking Ridge, NJ, United States

[4]Department of Biomedical Engineering, College of Engineering, University of Michigan, Ann Arbor, MI, United States

[5]Rogel Comprehensive Cancer Center, University of Michigan, Ann Arbor, MI, United States

[6]VA Ann Arbor Healthcare System, Ann Arbor, MI, United States

[7]Center for Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, United States

[8]Department of Internal Medicine, University of Michigan, Ann Arbor, MI, United States

[*]these authors contributed equally

Corresponding Author:
Vibhuti Gupta, PhD
School of Applied Computational Sciences, Meharry Medical College
3401 West End Ave #260
Nashville, TN, 37208
United States
Phone: 1 (615) 327 567
Email: vgupta@mmc.edu

Abstract

Background: Wearable sensors are increasingly being explored in health care, including in cancer care, for their potential in continuously monitoring patients. Despite their growing adoption, significant challenges remain in the quality and consistency of data collected from wearable sensors. Moreover, preprocessing pipelines to clean, transform, normalize, and standardize raw data have not yet been fully optimized.

Objective: This study aims to conduct a scoping review of preprocessing techniques used on raw wearable sensor data in cancer care, specifically focusing on methods implemented to ensure their readiness for artificial intelligence and machine learning (AI/ML) applications. We sought to understand the current landscape of approaches for handling issues, such as noise, missing values, normalization or standardization, and transformation, as well as techniques for extracting meaningful features from raw sensor outputs and converting them into usable formats for subsequent AI/ML analysis.

Methods: We systematically searched IEEE Xplore, PubMed, Embase, and Scopus to identify potentially relevant studies for this review. The eligibility criteria included (1) mobile health and wearable sensor studies in cancer, (2) written and published in English, (3) published between January 2018 and December 2023, (4) full text available rather than abstracts, and (5) original studies published in peer-reviewed journals or conferences.

Results: The initial search yielded 2147 articles, of which 20 (0.93%) met the inclusion criteria. Three major categories of preprocessing techniques were identified: data transformation (used in 12/20, 60% of selected studies), data normalization and standardization (used in 8/20, 40% of the selected studies), and data cleaning (used in 8/20, 40% of the selected studies). Transformation methods aimed to convert raw data into more informative formats for analysis, such as by segmenting sensor streams or extracting statistical features. Normalization and standardization techniques usually normalize the range of features to improve comparability and model convergence. Cleaning methods focused on enhancing data reliability by handling artifacts like missing values, outliers, and inconsistencies.

Conclusions: While wearable sensors are gaining traction in cancer care, realizing their full potential hinges on the ability to reliably translate raw outputs into high-quality data suitable for AI/ML applications. This review found that researchers are using various preprocessing techniques to address this challenge, but there remains a lack of standardized best practices. Our findings suggest a pressing need to develop and adopt uniform data quality and preprocessing workflows of wearable sensor data that can support the breadth of cancer research and varied patient populations. Given the diverse preprocessing techniques identified in the literature, there is an urgency for a framework that can guide researchers and clinicians in preparing wearable sensor data for AI/ML applications. For the scoping review as well as our research, we propose a general framework for preprocessing wearable sensor data, designed to be adaptable across different disease settings, moving beyond cancer care.

(JMIR Mhealth Uhealth 2024;12:e59587) doi:10.2196/59587

KEYWORDS

machine learning; artificial intelligence; preprocessing; wearables; mobile phone; cancer care

Introduction

Background

According to the US Food and Drug Administration, digital health is categorized as *mobile health* (mHealth), health information technology, wearable devices, telehealth, personalized medicine, and telemedicine [1]. Digital health has revolutionized health care by offering the potential for continuous and noninvasive monitoring of human physiological parameters, such as heart rate, sleep, and activity levels, to facilitate the early detection and prevention of life-threatening diseases [2]. Digital health consists of collecting, analyzing, storing, and sharing health care data by harnessing the power of technology, including smartphone apps, wearable sensors, telemedicine, the Internet of Medical Things, etc [3]. Due to the widespread use of mHealth technologies and routine use of wearable sensors (eg, smartwatches), the person-generated health data have become promising data sources for biomedical research [4].

Indeed, the integration of wearable sensors into cancer care has opened new pathways for remote monitoring, enabling health care providers to gather a wealth of real-time data from patients [5-7]. These wearables capture an array of physiological parameters, including skin temperature [8], offering insights into the patient's response to cancer treatment, quality of life, and overall well-being [9]. These continuous streams of data have the potential to transform cancer care by providing an improved understanding of patient conditions outside of the hospital setting, potentially improving clinical outcomes. Nevertheless, transforming raw data into meaningful analysis and insights presents numerous challenges, making standardized workflows for data preprocessing essential.

Data preprocessing involves a series of steps designed to clean and refine data to ensure its reliability and suitability for analysis using artificial intelligence and machine learning (AI/ML) techniques. The preprocessing steps help transform raw sensor data, which can be noisy and inconsistent, into a clean, structured format suitable for AI/ML models to process [10-12]. Without standardization in these procedures, there is a risk that subsequent data analysis might be based on flawed information, leading to uninterpretable data, a lack of generalizability, and erroneous conclusions. Typical preprocessing steps to make sensor data AI/ML ready include data cleaning (eg, noise reduction, outlier detection, and handling missing data) [13,14], data integration (eg, combining data sources and aligning time stamps), data transformation (eg, windowing and normalization) [15], dimensionality reduction (eg, feature selection), and data labeling (eg, annotating).

AI/ML's scope has become an amazing supportive tool for digital health [16,17] since its potential evolution to exploit meaningful relationships in biomedical data sets that can be used for diagnosis, prediction, and treatments [18-21]. AI/ML techniques have become popular in biometrics extraction mobile apps smart systems, such as eye disease detection [22-24], atrial fibrillation [25], heart rate monitoring [26], etc. In addition, a summary of the actual cancer statistics and its future directions is provided in the study by Moher et al [27].

Within the integration of electronic health record technology [26] in digital medicine, wearable monitoring devices have earned an important and crucial role for all people in the biomedical area (eg, patients, medical staff, and biomedical researchers). Oncology divisions have ultimately contemplated the importance of incorporating mHealth monitoring while conducting clinical cancer trials [1]. Moreover, multiple types of cancer disease detection using AI/ML techniques are a crucial factor considering its alarming impact rates on the population [27]. The mHealth integration on cancer applications for the development of AI/ML solutions has become popular in recent years [28]. However, the importance of data quality has not been highlighted while considering the design and development of prediction models. Building high-quality data is a critical step while applying AI/ML algorithms in mHealth and wearable studies; however, the emphasis on enriching the data quality is very limited in these studies, especially in oncology. Misclassifications, misdiagnoses, and wrong predictions can be avoided, and the whole mHealth system feasibility can be improved by enriching the data quality.

Goals of Our Review

This study aims to explore the use of wearable sensors for continuous monitoring of key physiological parameters in cancer care. We systematically reviewed the literature by identifying and assessing preprocessing workflows that are essential for transforming raw, noisy, and often inconsistent wearable sensor data into reliable and structured formats suitable for subsequent AI/ML modeling. By examining the current landscape of these practices, our research aims to improve wearable sensor data quality, specifically for cancer care, ensuring that downstream

data analyses and interpretations are rigorous and reproducible. Given the diverse preprocessing techniques identified in the literature, there is an urgency for a framework that can guide researchers and clinicians in preparing wearable sensor data for AI/ML applications. This paper proposes a framework designed to be adaptable across different continuous monitoring applications.

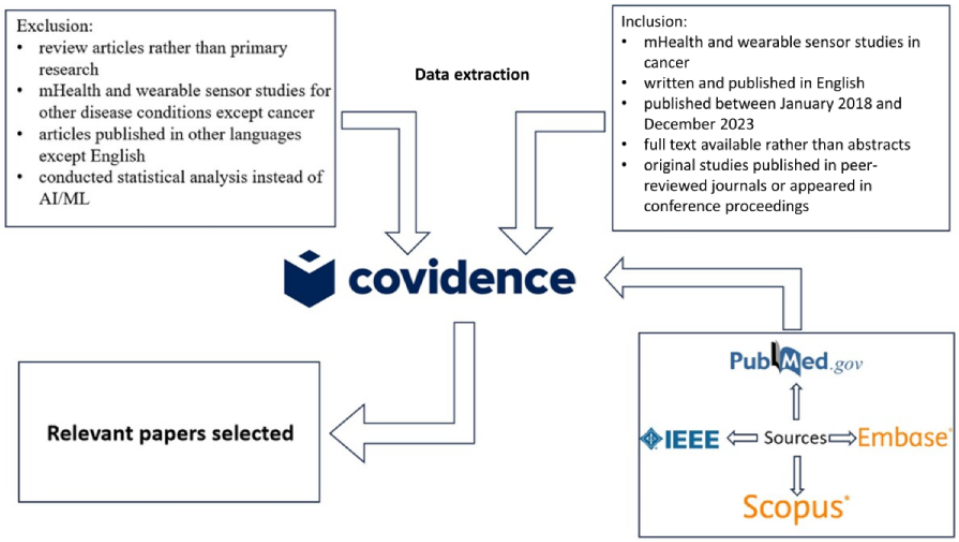
Methods

Search Strategy

We conducted a scoping review of articles written in English using the following literature databases: IEEE Xplore, PubMed, Embase, and Scopus, while following the PRISMA-ScR (Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews) guidelines [29]. We have used Covidence (Veritas Health Innovation Ltd) [30] for identification and screening stages. The search was performed on December 31, 2023, using the search queries shown in Multimedia Appendix 1. We selected full peer-reviewed publications from the last 5 years (from January 2018 to December 2023), focusing on preprocessing techniques used on wearable sensor data to ensure their readiness for AI/ML applications for different cancer populations. Searches were developed using 3 key concepts: wearable devices, AI/ML, and cancer. Controlled vocabulary and keywords were selected for the specific databases.

Figure 1 shows an illustration of the study selection process for this paper. The identified studies meeting the inclusion criteria were subsequently organized based on the major themes identified.

Figure 1. Illustration of the study selection process. AI/ML: artificial intelligence and machine learning; mHealth: mobile health.



Inclusion Criteria

Our results with the search query presented in Multimedia Appendix 1 were first imported into Covidence for screening. The title and abstracts of the resulting studies were screened to identify the studies related to preprocessing techniques for wearable sensor data in cancer. After identifying the eligible studies, additional inclusion exclusion criteria were applied to retrieve the primary studies of our review (Figure 2 in the *Results* section). Studies were eligible if they fulfilled the following inclusion criteria in our review: (1) mHealth and wearable sensor studies in cancer, (2) written and published in English, (3) published between January 2018 and December 2023, (4) full text available rather than abstracts, and (5) original studies published in peer-reviewed journals or appeared in conference proceedings. PRISMA-ScR checklist is provided in Multimedia Appendix 2.

Exclusion Criteria

Studies were not eligible if they fulfilled the following exclusion criteria in our review: (1) review articles rather than primary research, (2) mHealth and wearable sensor studies for other disease conditions except cancer, (3) articles published in other languages except English, and (4) conducted statistical analysis instead of AI/ML.

Data Extraction and Evaluation

The data were extracted from all studies meeting our inclusion criteria for the review and organized into tables containing each study's information (eg, authors' name, title, and year of publication), wearable sensor data collected in cancer studies (eg, activity data, physiological parameters, including steps, sleep, heart rate, blood oxygen saturation, and temperature), preprocessing techniques (eg, time segmentation, data filtering, data transformation, and imputation), wearable devices (eg, Fitbit [Google LLC], Empatica [Empatica Inc, and Actigraphy), type of AI/ML methods applied (eg, neural networks, decision trees, K-Nearest Neighbors, Supporting Vector Machine, and regressors), sample size (eg, number of participants; Table 1). The data for all selected studies were extracted independently by 3 authors (BLO, VG, and SWC) by mutual agreement, and discrepancies were resolved by discussion with other coauthors (RK, AJ, XC, and CZ). The outcomes from the data extraction part were finally evaluated independently by each author.

Table 1. Summary of eligible studies.

Reference	Cancer type	Sample size, N	Wearable sensor	Physiological parameter	Preprocessing procedure	Preprocessing category	AI/ML[a] techniques
Liu et al [30], 2023	Terminal cancer	40	Garmin VivoSmart 4	Steps, HR[b], sleep status, and blood oxygen saturation (measured during sleep time)	Missing data imputation	Data cleaning	LR[c], SVM[d], DT[e], RF[f], KNN[g], AdaBoost[h], and XGBoost[i]
Zhao et al [31], 2022	Breast cancer	4	Fuschia Band prototype	Accelerometer and gyroscope readings	Peak detection and fast Fourier transform	Data transformation	KNN
Moscato et al [32], 2022	Multiple types of cancer	21	Empatica E4 wristband	Photoplethysmography signals, skin temperature, accelerometer readings, and electrodermal activity	Different- order Butterworth filtering with different cutoff frequencies and data normalization	Data cleaning and normalization and standardization	SVM, RF, MLP[j], log, and AdaBoost
Yang et al [33], 2021	Terminal cancer	60	Actigraphy device XB40ACT	Activity level, angle, and spin	Zero padding and shortening the time series	Data transformation	LSTM[k]
Huang et al [34], 2023	Terminal cancer	78	Actigraphy device XB40ACT	Activity level, angle, and spin	Time Segmentation and zero padding	Data transformation	LSTM, bidirectional-LSTM, transformer, and GRU[l]
Cos et al [35], 2021	Pancreatic cancer	28	Fitbit inspire HR	Step count, HR, and sleep time–series data	One-hot encoding standardization and dimensionality reduction	Data transformation	RF, GBT[m], KNN, SVM with linear kernel, and LR with L1 penalty
Davoudi et al [36], 2021	Multiple types of cancer	27	ActiGraph GT3X	Accelerometer Readings and oxygen consumption	Bias reduction, data localization, and vector magnitude calculation	Data cleaning and transformation	RF, GBT, KNN, SVM with linear kernel, and LR with L1 penalty
Liu et al [37], 2020	Multiple types of cancer	3	Fitbit Alta	HR data and activity data	Missing data imputation and data standardization	Data cleaning and normalization and standardization	Hidden Markov models
Tedesco et al [38], 2021	Multiple types of cancer	2291	ActiGraph GT3X+	Steps taken, time in light, sedentary, moderate, vigorous activities, energy expenditure, etc.	Data standardization and missing data imputation	Data cleaning and normalization and standardization	AdaBoost
Dong et al [39], 2021	Pancreatic cancer	10	ActiGraph devices	Accelerometer, light, and inclinometer	Time window segmentation	Data transformation	GRL[n]
Patel et al [40], 2023	Multiple types of cancer	50	Actiwatch	Rest-activity, sleep, and routine clinical variables	Missing data imputation with averaging technique	Data cleaning	Penalized (regularized) regression models
Asghari [41], 2021	Colorectal cancer	400	IoMT[o] smart devices	Vital signs that were sensed through biomedical sensors	Cleaning inconsistencies and noise and Dimensionality reduction	Data cleaning and transformation	J48, SMO[p], MLP, and NB[q] methods
Rossi et al [42], 2021	Multiple types of cancer	52	PGHD[r] (VivoFit)	Daily steps	Temporal segmentations	Data transformation	LR
Vets et al [43], 2023	Breast cancer	10	ActiGraph wGT3X-BT	Accelerometer readings	Counts threshold and data normalization	Data transformation and normalization and standardization	Pretrained MLM[s]

Reference	Cancer type	Sample size, N	Wearable sensor	Physiological parameter	Preprocessing procedure	Preprocessing category	AI/ML[a] techniques
Feng et al [44], 2023	Prostate cancer	47	Google health, Fitbit, or Apple health	Step counts	Time window segmentation	Data transformation	LR
van den Eijnden et al [45], 2023	Multiple types of cancer	125	Elan sensor (wristband)	Activity features, activity counts, acceleration data, as well photoplethysmography signal	Features calculation, data dimensionality reduction and numerical to categorical data transformation, and standardization	Data transformation and normalization and standardization	LR, KNN, DT, RF, support vector regression, and XGBoost
S et al [46], 2020	Breast cancer	201	Cyrcadia breast monitor	Temperature readings	Removing outliers and missing data, duplicates removal, and data normalization	Data cleaning and normalization and standardization	DT, SVM, RF, and back propagation NN[t]
Barber et al [47], 2022	Gynecologic cancer	34	Fitbit Alta HR	Steps, HR, and intensity of physical activity	Data standardization and normalization	Data normalization and standardization	LR, RF, GBT, and XGBoost
Jacobsen et al [48], 2023	Blood cancer	79	Wearable-based RPM[u]	Time-series data recorded from biosensors	Dimensionality reduction	Data transformation	NN
Li et al [49], 2023	Multiple types of cancer	201	IMU[v] sensor nodes, and Heal Force PC-60NW	HR and inertial measurements	Interval scaling method and z score standardization	Data normalization and standardization	MMDF[w], XGBoost, LGBM[x], RF, AdaBoost, and GBT

[a]AI/ML: artificial intelligence and machine learning.
[b]HR: heart rate.
[c]LR: logistic regression.
[d]SVM: support vector machine.
[e]DT: decision tree.
[f]RF: random forest.
[g]KNN: k-nearest neighbors.
[h]AdaBoost: adaptive boosting trees.
[i]XGBoost: extreme gradient boosting trees.
[j]MLP: multilayer perceptron.
[k]LSTM: long short-term memory.
[l]GRU: gated recurrent unit.
[m]GBT: gradient boosted trees.
[n]GRL: graph representation learning.
[o]IoMT: Internet of Medical Things.
[p]SMO: sequential minimal optimization.
[q]NB: naïve Bayes.
[r]PGHD: patient-generated health data.
[s]MLM: machine learning model.
[t]NN: neural network.
[u]RPM: remote patient monitoring.
[v]IMU: inertial measurement unit.
[w]MMDF: multimodel decision fusion.
[x]LGBM: light gradient boosting machine.

Results

Overview

We identified 2147 studies in the initial extraction phase (n=248, 11.55% from PubMed; n=428, 19.93% from Scopus; n=996, 46.39% from IEEE Xplore; and n=475, 22.12% for Embase, including Embase, Embase Classic, MEDLINE, and PubMed-not-MEDLINE). A total of 173 (8.06%) duplicate articles were removed to produce 1974 (91.94%) for title and abstract screening. We conducted a thorough screening of titles and abstracts, which resulted in the exclusion of 1820 (92.2%) articles that did not meet the inclusion criteria. Following this

screening, we identified 154 (7.8%) articles for which we performed a full-text review to assess their eligibility for inclusion in our study in more detail. In the final screening, 20 (13%) of these 154 articles met our inclusion criteria and were considered for this scoping review, as shown in Figure 2. The workflow diagram for the systematic identification of scientific literature is shown in Figure 2. The geographical distribution of these studies is mapped in Figure 3, highlighting most research from the United States. These constituted 35% (7/20) of the selected publications. Terminal cancer research was reported from Taiwan.

In terms of publication years, our analysis revealed an uptick in the frequency of papers related to mHealth and wearables in cancer. Our review coincides with the emergence of the COVID-19 pandemic, during which there was a surge in research interest within the biomedical sciences, particularly related to the use of wearable technology in remote monitoring of patients with cancer. The distribution of publications during this period suggested that in the years 2020 to 2022 combined, approximately one-quarter of the selected studies were published, accounting for 25% (5/20) of our data set. The majority were distributed between the years 2021 to 2023, which collectively contributed to 75% (15/20) of the data quality improvement strategies for wearable data preprocessing in cancer care settings. In fact, 40% (8/20) of all selected studies were published in 2023 alone, marking a substantial rise and interest in this research domain.

Our findings reported the use of wearable technology across a diverse range of cancer types. Predominantly, the category encompassing “multiple types of cancer” accounted for 40% (8/20) of the studies in this area. The remainder of the research was distributed among specific types of cancer, with each category’s contribution detailed as follows: breast cancer (3/20, 15%), terminal cancer (3/20, 15%), pancreatic cancer (2/20, 10%), blood cancer (1/20, 5%), colorectal cancer (1/20, 5%), prostate cancer (1/20, 5%), and gynecologic cancer (1/20, 5%). In addition, the recent literature indicated a trend toward increased adoption of wearable technology for cancer surveillance, signifying a growing recognition of the potential benefits that wearables may offer in continuous patient monitoring across heterogeneous cancer types.

The initial database search yielded 2147 studies, of which 20 (0.93%) met the inclusion criteria after screening and full-text review (Figure 2). The included studies applied preprocessing techniques to wearable sensor data from a range of cancer populations, including breast, colorectal, gynecologic, and blood cancers, as well as multiple other types of cancer. The most commonly used wearable devices were actigraphy sensors and consumer-grade fitness trackers, which provided data on physical activity, sleep, heart rate, and other physiological parameters.

Various preprocessing approaches are used in each of the identified themes. The most common data transformation approaches included fast Fourier transform [31], time-series segmentation [33,34,39], and statistical feature calculation [30,35,45]. However, for the data normalization techniques, *z* score standardization and min-max normalization were the most frequently reported scaling methods [32,37,43,46,49] and for the data cleaning, imputation [30,37,40], outlier removal [36,46], and artifact filtering [32,41] approaches were used. Notably, 25% (5/20) of the studies combined multiple preprocessing techniques from different categories, suggesting that a comprehensive approach to data preparation may be beneficial [32,36,38,45,46]. However, there was significant heterogeneity in the specific techniques used and their implementations across studies, highlighting a lack of standardized preprocessing pipelines for wearable sensor data in cancer care.

The preprocessing techniques were applied to support a range of AI/ML applications, including treatment response prediction [35,42], symptom monitoring [44,47], and survival analysis [33,34]. The most common ML algorithms were random forests, support vector machines, and deep learning models, such as long short-term memory networks. However, few studies directly compared the impact of different preprocessing approaches on model performance, making it difficult to draw conclusions about optimal techniques.

Figure 2. PRISMA-ScR (Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews) diagram for a scoping review of biomedical scientific literature. ML: machine learning.

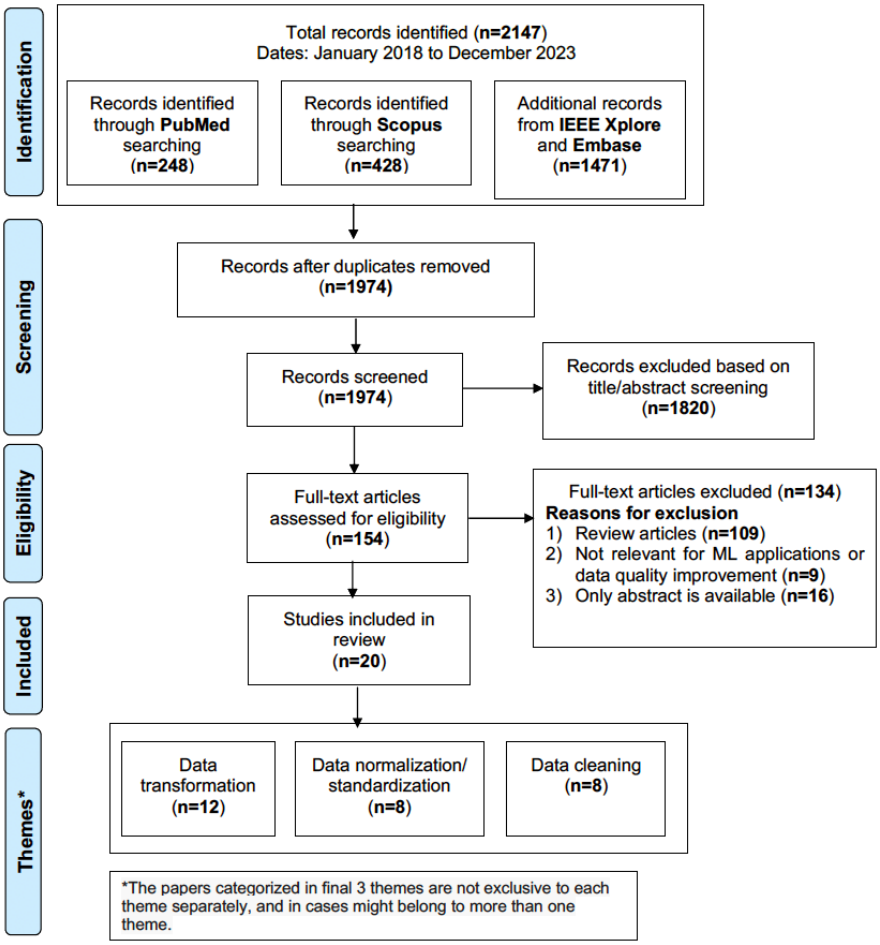
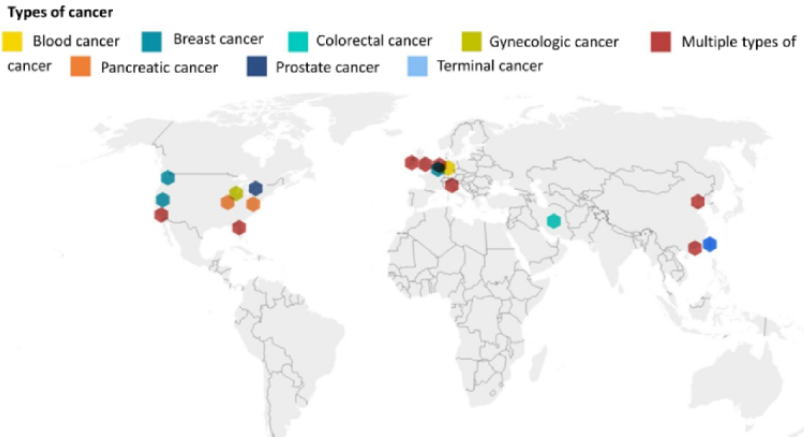


Figure 3. Relevant references by geographical location.



Major Themes Identified

Three major themes were identified, as outlined in Table 1: (1) data normalization and standardization (8/20, 40% of papers), (2) data transformation (12/20, 60% of papers), and (3) data cleaning (8/20, 40% of papers). These were subcategorized based on the preprocessing techniques. Data transformation comprises studies related to dimensionality reduction, data feature calculation, variable transformation, or domain transformation. Data normalization and standardization included data standardization or data normalization. The data cleaning category included data filtering, outliers' removal, imputation techniques, missing data, and duplicate removal. Multiple selected work categories were required to combine preprocessing tasks encompassing the previous 3 mentioned categories while addressing data quality issues [30-49], which are presented in Tables 1 and 2.

Table 2. A summary of relevant preprocessing elements on selected published works.

Reference	Time resolution	Exclusion criteria	Missing data imputation technique	Features extracted	Outcomes
Liu et al [30], 2023	Each day was a data point	Days with no wearable device data uploaded	Linear interpolation	A combination of basic demographic data, clinical assessment data, and wearable device data	Death event prediction
Zhao et al [31], 2022	Data were sent at a rate of 4 times per s	Determine whether an exercise is completed correctly or incorrectly	Not applicable	Statistical gyroscopic-based features obtained from all 3 axes (x, y, and z)	Rehabilitation
Moscato et al [32], 2022	A 2-min time window before the beginning of each session was created	Feature pairing was tested by Pearson correlation coefficient >0.9	Linear interpolation	12 features from the HRV[a] analysis, 5 features from the photoplethysmography morphological analysis, 17 features from the electrodermal activity, 3 features from the temperature, and 2 features from the activity index	Pain assessment
Yang et al [33], 2021	An average value of 20 timesteps within total time shortened to <500 timesteps	Time series >500 timesteps	Zero paddings until the maximum length of the time series was reached	Physical activity, angle, and spin	Survival prediction
Huang et al [34], 2023	A mean of 20 timesteps was chosen as the average value for 3 time frames (12, 24, and 48 h)	Properly designed patients' admission criteria	Zero padding was used to reach the maximum length of the time series	Physical activity, angle, and spin and the clinical data from patients were also considered	Survival prediction
Cos et al [35], 2021	Biobehavioral rhythmic features were computed for the entire tested period, and statistical and semantic features were generated daily	Biobehavioral rhythmic features were excluded due to the dimensions	Data-level and feature-level	First- and second-order statistical features from the daily step count, HR[b], and sleep time–series data	Pancreatectomy treatment outcomes from patients activity
Davoudi et al [36], 2021	Extracted relevant features from a 16-s window; data were eventually smoothed with a 30-s running average window	Data length <4 min	Not applicable	Time and frequency domain features	Physical activity recognition and energy expenditure estimation
Liu et al [37], 2020	Disaggregating the 15-min step count data and simulating the 1-min step count time series	Nonwear days were identified and removed before the analysis	Thresholding	Statistics from HR metrics and activity levels	Algorithm validation
Tedesco et al [38], 2021	Not provided	Wear time per day was <600 min	Feature mean	Statistical features from (1) demographics, (2) self-report health and lifestyle, (3) wearable data, and (4) laboratory tests	Cancer- specific mortality prediction
Dong et al [39], 2021	1-min epoch to aggregate and synchronize the raw actigraphy data	9.5 h window size for accelerometer data to fit models	Not applicable	Time and frequency domain features from actigraphy and laboratory tests	Salivary cortisol levels on in patients with pancreatic cancer
Patel et al [40], 2023	Numerical continuous variables involving sleep-wake times were entered in the 24 h format	Data were excluded from the 1-h period before and after going to bed	Average values	Sleep-based features and sleep-wake transitional-related features	Exploratory machine learning study
Asghari [41], 2021	Not provided	Data inconsistencies removal	Not applicable	Demographics, clinical features, and wearable data	Diagnostic prediction on CRC[c] older adults

Reference	Time resolution	Exclusion criteria	Missing data imputation technique	Features extracted	Outcomes
Rossi et al [42], 2021	Three distinct types of temporal segments for weekly observations	Periods before admission	Majority class	Activity or steps related features and clinical data	Postsurgery complications
Vets et al [43], 2023	Acceleration data's sampling rate was 30 Hz	Unknown data were discarded from further analysis	Spline interpolation	Statistical parameters from accelerometer readings	Rehabilitation study
Feng et al [44], 2023	A window of 48 h following step count decline	A decline of 1000 steps or more as a binary predictor among participants	Thresholding	Step counts calculated on different time windows	Physical activity monitoring on active treatment
van den Eijnden et al [45], 2023	The data were stored at 1-s intervals	Early stopping algorithm	Not applicable	For health dot sensor: RR[d], activity level (actlevel); for Elan wristband: statistical parameters from HR, and frequency domain features	Recovery scores
S et al [46], 2020	Temperature profiles had values from 16 sensors gathered for 1 d at every 5-min interval	Out-of-range temperature data discrimination	Not applicable	Linear and nonlinear features from the time-series temperature data	Introductory paper
Barber et al [47], 2022	Each day was considered an observation	Discrimination of days was applied to unscheduled contacts	Not applicable	Fatigue, physical function, anxiety, mean daily HR, daily steps, sleep, and time-related features	Feasibility and events prediction
Jacobsen et al [48], 2023	Raw signals were acquired with a frequency of >30 Hz; calculated parameters were stored with a rate of 1 Hz	Data points reduction due to interruptions	Not applicable	Noninvasive monitoring of vital signs and physical activity; SCC[e] events	Clinical complications during treatment
Li et al [49], 2023	Sampling frequency was 200 Hz for IMU[f]; the HR was stored at a sampling frequency was 1 Hz	Feature selection for redundancy removal	Majority class	HR metrics, physical activity parameters, Blood Mass Index, and blood oxygen statistical values	Physical fitness assessment

[a]HRV: heart rate variability.

[b]HR: heart rate.

[c]CRC: colorectal cancer.

[d]RR: respiratory rate.

[e]SCC: serious clinical complications.

[f]IMU: inertial measurement unit.

Data Transformation

Zhao et al [31] reported a proof-of-concept for postoperative rehabilitation in a small cohort of 4 patients with breast cancer, using a prototype that used peak detection and Fourier transform by switching time domain points of the 3D axis to a predetermined frequency. Yang et al [33] hypothesized that wristband actigraphy monitoring devices could predict in-hospital death of end-stage multiple types of patients with cancer during the hospitalization period admissions. To avoid variations in each patient's data length, zero padding was used until the maximum length of the time series was reached [33]. Scoring systems, such as the Palliative Prognostic Index and Palliative Performance Scale, were considered for fitting machine learning models (MLMs) [33]. Huang et al [34] reported a comparison study between the results of wearable-based activity monitoring with traditional prognostic tools for patients with end-stage cancer. In total 3 different time frames were segmented for preprocessing [34]. A mean of 20 timesteps was selected as the average value for each of the 3 different time frames (48, 24, and 12 h) [34]. Zero padding was used in the study by Huang et al [34], making it applicable to data transformation. Cos et al [35] used a wearable device to predict treatment outcomes in patients with pancreatic cancer, standardizing data before using ML methods.

Dong et al [39] proposed a general predictive modeling process that used actigraphy data to predict underlying salivary cortisol levels using graph representation learning. The raw sensor data were preprocessed using time window segmentation to reduce noise in the data [39]. Rossi et al [42] focused on predicting postdischarge oncologic surgical complications and their impact

on patient outcomes. There were 3 distinct types of temporal segments for each patient. They considered observations up to the second week after discharge, treating each week as a distinct observation [42].

Feng et al [44] evaluated the feasibility of daily step count monitoring and the association between step counts and treatment-emergent symptoms in patients with prostate cancer. As shown in Table 1, the preprocessing technique could be summarized as follows: (1) a decline of 1000 steps or more as a binary predictor and (2) time window segmentation [44]. Jacobsen et al [48] impacted medical literature by proposing self-supervised contrastive learning methods for hematological malignancy treatments. Noninvasive monitoring of vital signs and physical activity was recorded within serious clinical complications in the input data set [48]. Data downsampling was the selected preprocessing technique to eliminate physical interruptions [48]. These studies collectively illustrated diverse data transform methods, such as feature selection, time segmentation, domain transformation, and time windowing, to enhance wearable device data quality, making them more suitable for AI/ML modeling aimed at predicting patient outcomes in cancer care. In addition, these findings have leveraged a range of wearable technologies and AI/ML methods to advance cancer care. Techniques, such as peak detection and Fourier transform have been used for data preprocessing, supporting applications that include postoperative rehabilitation, physical activity classification, prediction of treatment outcomes, and assessment of cancer-specific mortality. These studies highlight the potential of integrating high-dimensional wearable data with clinical information to enhance patient monitoring and prognosis.

Data Normalization and Standardization

Barber et al [47] assessed the feasibility of postoperative intervention for patients with gynecologic cancer in a manner similar to Zhao et al [31], incorporating patient-reported outcomes and wearable activity data and also opting for standardization and normalization of preprocessing methods. Finally, Li et al [49] proposed a method using multimodel decision fusion based on multisource data for physical fitness assessment for patients with cancer. They enriched the raw data by using Baseline, Synthetic Minority Over-sampling Technique, random oversampling, adaptive synthetic oversampling, and Mahalanobis Distance and Boundary Constraints. The interval scaling method and *z* score standardization after segmentation are the common methods in the study by Li et al [49]. These additional investigations used tailored data preprocessing approaches to further refine the quality of wearable device data for subsequent analysis (eg, data partitioning for training and testing).

Moscato et al [32] proposed an automatic pain assessment for patients with cancer (21 in total) by using the Empatica wristband. Because all physiological signals were recorded at different sampling rates, different-order Butterworth filtering with different cutoff frequencies was the data enrichment selected method [32]. Each pulse was normalized with the *z* score procedure and processed with an automated algorithm that detects pulses suitable for heart rate variability analysis and derived metrics [32]. Liu et al [37] aimed to develop an unsupervised personalized sleep-wake identification algorithm using multistage data to explore the benefits of incorporating heart rate metrics and actigraphy data in these types of algorithms for the general population. After nonwear exclusion, there were 14 participants whose data qualified for analysis; 5 (36%) had high cholesterol, 6 (43%) participants had hypertension, 3 (21%) had cancer, 2 (14%) had diabetes mellitus, and 1 (7%) have had a stroke. They preprocessed the step count data, and 2 schematic ML-based models were designed by following the Markov model's fundamentals. To facilitate the fusion of step count and heart rate data in the models, downscaling was used to deal with the multigranularity data [37]. In addition, imputation techniques were implemented. Tedesco et al [38] explored the prediction of cancer-specific mortality over a 2- to 7-year period using a data set from a longitudinal study of 2291 70-year-old Swedish patients, integrating wearable and electronic health record data. They applied standardization and normalization preprocessing techniques within imputation.

Vets et al [43] aimed to determine the accuracy of a pretrained laboratory-based MLM to distinguish functional from nonfunctional arm motions through home interventions of survivors from breast cancer populations. From the accelerometer data, functional activity was defined using two separate methods: (1) the counts threshold method, and (2) a pretrained MLM [43]. Activity counts were calculated from the raw acceleration data [43]. The outcome "total minutes active" was calculated as the sum of the 1-second epochs where the count threshold exceeded 1 [43]. Data normalization was the final step before fitting AI/ML models. van den Eijnden et al [45] created a model that predicted continuous recovery scores (regressors) in perioperative care in the hospital and at home for objective oncology-based decision-making. They preprocessed data by obtaining a balanced split in which they equally divided the demographic predictors and surgery type into 2 groups by splitting the patients 10,000 times [45]. Finally, authors standardized features by scaling the data to a normal distribution with a mean of 0 and a unit variance [45]. S et al [46] introduced a noninvasive wearable device developed as an adjunct to current modalities to assist in the detection of breast tissue abnormalities in any type of breast tissue. In the study, data normalization and outliers' removal were the data transformation methods to enrich the quality of the collected temperature data.

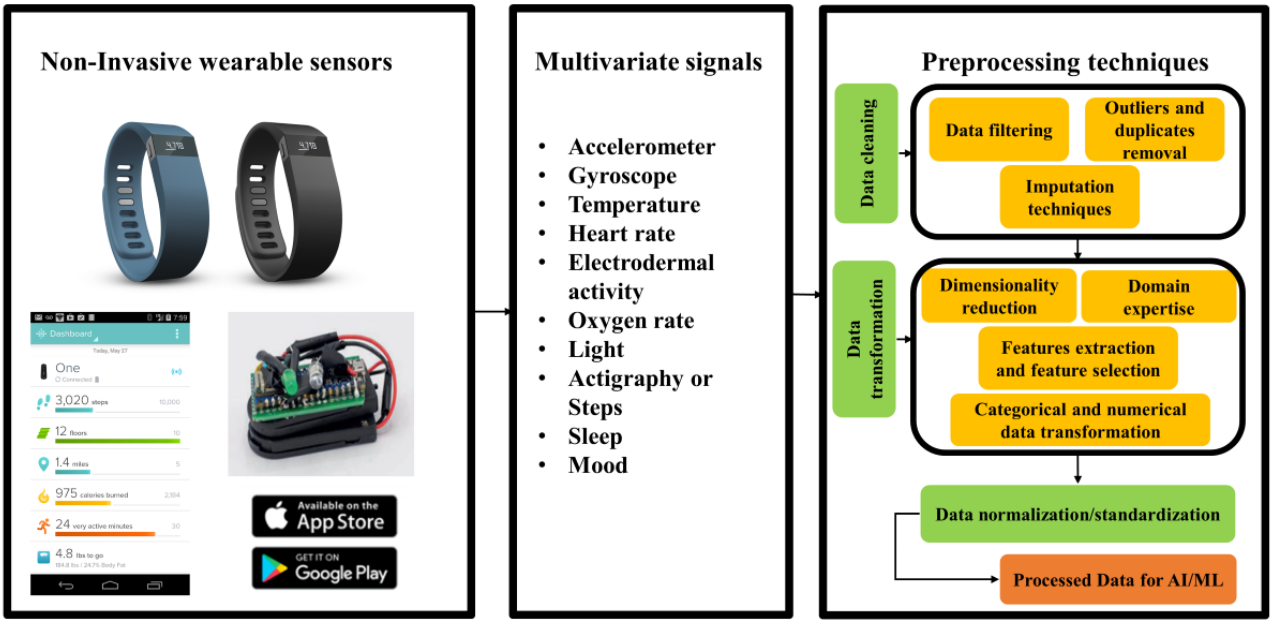
Data Cleaning

Liu et al [30] aimed to investigate the potential of using wearable devices and AI/ML to predict death events among patients with terminal cancer. To improve the model training, the authors used imputation techniques [30]. The data set was a combination of demographic, clinical, and wearable device data [30]. Davoudi et al [36] conducted a study comparing various accelerometer placements in classifying physical activity and associated energy expenditure among older adults. Of the 93 participants who completed the study, 27 (29%) were identified with a range of cancer diagnoses. Raw data were cleaned using bias reduction and eventually transformed by activity location and vector magnitude calculation [36]. Similarly, Patel et al

[40] sought to enhance prognostic tools by combining ML analysis of actigraphy, sleep data, and routine clinical data with a missing data imputation technique within averaging. Asghari [41] proposed an internet of things–based predicting model to predict colorectal cancer in older adults. The data preprocessing phase was required to clean the sensed medical internet of things data from the inconsistencies and the noises for the data mining phase [41]. Outliers' removal was the initial step selected for preprocessing.

Accordingly, we proposed a generalized preprocessing framework that comprises all 3 major data preprocessing themes (Figure 4), reflecting the core elements that were consistently reported across studies.

Figure 4. A general framework for data preprocessing techniques used to make noninvasive data collected from mobile health and wearable sensor artificial intelligence and machine learning (AI/ML) ready in cancer monitoring applications.



Discussion

Principal Findings

In this paper, we conducted a scoping review of the preprocessing techniques applied to wearable sensor data in cancer care. Our findings revealed a significant rise in the use of wearable sensors for patient monitoring, along with an increase in preprocessing methods for data analysis over the past 5 years. This likely stemmed from recent advancements in sensor technology, greater emphasis on personalized and remote patient care, the rising prevalence of big data analytics in health care, and increasing recognition of real-time health data for precision oncology.

Data transformation emerged as the most reported preprocessing technique, representing approximately 60% (12/20) of the literature findings. Most studies relied on data from commercially available products, except a study by Zhao et al [31], which assessed a prototype's efficiency in a small cohort. While published studies describing preprocessing methods for wearable devices are growing, the diagnoses being studied remain sparse and generally limited to single disease types or settings.

The physiological data captured from wearables are typically noisy, contain missing values, have outliers, redundant features, and erroneous measurements [50,51]. On the basis of the literature review in this paper, we found that various data cleaning procedures are used to clean the wearable sensor data, including data smoothing techniques (ie, moving average and exponential moving average) to reduce short-term signal artifacts and remove noise, removing duplicate entries, detection and removal of erroneous measurements due to sensor malfunctioning or losing contact of the sensor with skin or wearing the watch on incorrect body location, and outlier removal. The outlier removal for wearable data [52] in the reviewed studies consists of the range inspection of physiological parameter values with the clinically relevant range or developing a threshold using statistical techniques to detect outliers. Finally, missing data imputation is a critical component of data cleaning due to their ability to handle complex missing patterns as demonstrated in wearable-based data [53-57].

Our review suggests that the data cleaning procedures should be carefully inspected and applied based on the data captured from the wearables, as the captured data will produce false conclusions and predictions without proper data cleaning procedures, which is not acceptable in clinical research. In addition, the outliers' removal should be based on data behavior and domain knowledge, as a region of anomaly is often within the boundaries of normal patterns of physiological data; for example, for the heart rate data, the normal behavior might evolve, which can be considered anomalous behavior, and the removal of data points leads to the loss of critical data. A

generalized, automated, and adaptive data cleaning procedure is required for the wearable data to address the issues that arise due to improper data cleaning.

Time-series segmentation is the most used data transformation technique in wearable research identified in the review, necessitated by the multivariate nature of the data and varying sampling rates. Segmentation can be based on study outcomes, such as daily, hourly, or minute-by-minute intervals. Our review indicates that the optimal time window size for segmentation must be determined through experimentation to achieve the best performance results. This window size varies across different cancer cohorts and should be tailored to the specific data set rather than relying solely on literature. The granularity of time segmentation also affects feature extraction. For instance, summary statistics like mean, median, SD, and minimum, and maximum differ when calculated for daily versus hourly or minute-by-minute windows. The reviewed literature [58-60] also explores additional feature types, including frequency domain features and linear and nonlinear features.

Data compliance is another major challenge in wearable studies and has a profound impact on the study outcomes. Physiological data captured from wearables are highly variable [61] and have high noncompliance rates by the participants. The participants' compliance determines the validity of the data collected from the wearables and their utility. Different thresholds are established for various parameters, such as daily wear time or step counts to filter or preprocess the data [62-64]. This scoping review suggests that we should strive to develop algorithms for standardizing the physiological metrics collected, which includes establishing thresholds for data inclusion based on compliance, filtering data based on adequate wearable wear time in study participants undergoing cancer per day and per week, percentage of days on which wearable was worn by the participants, inclusion and exclusion of data due to participant wearable synchronization issues, etc. ML techniques can be exploited to automate the data compliance assessments for different data extracted in different types of cancer.

Finally, data normalization is critical to developing AI/ML-ready data for the wearable studies. The data scaling helps not only in building efficient and accurate MLMs but also removes the effect of different scales and ranges in the model prediction. Our review suggests that researchers should identify the appropriate normalization technique for their study and understand the data distribution and model results before and after applying these techniques.

In summary, this scoping review identified 3 main categories of preprocessing techniques: data transformation, data normalization and standardization, and data cleaning, that have been applied to wearable sensor data in cancer care. While these techniques are commonly used to prepare data for AI/ML analysis, there is a lack of standardization in their implementation and limited evidence of their comparative effectiveness. Moreover, wearable sensor data are highly unstructured, complex, and messy because it is generated continuously and with high frequency (thousands of observations per second), leading to rich streams of time-series data. Thus, there is an urgent need to develop novel preprocessing procedures and frameworks, enhancing data quality and data readiness for AI/ML applications in cancer research. Future work should focus on developing validated preprocessing pipelines and benchmarking their impact on AI/ML model performance across diverse cancer populations and wearable devices. By providing a generalizable framework, we aim to accelerate the development of AI/ML models in not only cancer care but also potentially other areas of health care that leverage wearable sensor data. Researchers and clinicians can adapt this framework to their specific needs, promoting standardization while allowing for necessary customization.

Preprocessing Techniques for General mHealth Applications

Preprocessing techniques have been a considerable topic of interest in the research community within its integration with the mHealth concept [65-67]. For example, cardiovascular diseases and diabetes are 2 conditions that have benefited from mHealth tools. In a study by Qaisar et al [68], an efficient method for the diagnosis of arrhythmia based on electrocardiogram inputs was proposed. The method combined multivariate processing, wavelet decomposition, frequency content-based subband coefficient selection, and ML techniques for preprocessing. In a study by Efat et al [69], a smart health monitoring tool for patients with diabetes was introduced. The objective of the authors was to use continuous sensor monitoring and processing with neural networks to provide a continuous evaluation of the patient's health risk status by considering the patients' noninvasive biometric data [69]. To improve data quality, the authors used data transformation. Photoplethysmography has been used for blood pressure monitoring by incorporating the mHealth concept [70]. The authors collected photoplethysmography signal data from smartphones and passed them through a high-pass filter with a cutoff frequency of 0.5 Hz. To filter out unwanted peaks and create a smooth signal, a moving average filter with a span of 5 data points was applied to the signals before peak detection was performed [70]. Peak detections were implemented by finding the local maximum values in the signals [70]. The incorporation of mHealth technology has brought several efficient alternatives for health care engineering. In addition, it becomes a challenging factor while addressing data quality issues. The general health care sector has experienced irregularities in converting raw data to suitable formats, there is not an exceptional case in cancer monitoring.

Proposed Preprocessing Framework

To address the challenges and limitations identified in the reviewed literature, we propose a general preprocessing framework to develop AI/ML-ready data for mHealth cancer monitoring applications. Figure 4 summarizes this framework for noninvasive physiological monitoring data analysis. While our framework is conceptually applied within the setting of general oncology monitoring to fit AI/ML models, it could also be applied in other disease settings by following the key elements and steps of data preprocessing techniques.

Our proposed framework (Figure 4) synthesizes the best practices identified in this review, offering a standardized approach to preprocessing wearable sensor data. The

framework's strength lies in its flexibility and broad applicability. While the framework was developed based on cancer care applications, its fundamental components, data cleaning, data transformation, and data normalization and standardization, are relevant to a wide range of chronic diseases that can benefit from continuous monitoring via wearable sensors. By extracting raw wearable-based data from a real-world scenario, as shown in this paper using the cancer care setting, researchers should be able to reproduce available preprocessing solutions to other settings that leverage wearable sensor data. For instance, the data cleaning techniques identified in cancer studies, such as handling missing data and removing artifacts, are equally crucial in preprocessing data for heart disease or diabetes monitoring. Similarly, the data transformation methods, including feature extraction and dimensionality reduction, can be adapted to extract relevant biomarkers for various conditions. The framework's emphasis on data normalization and standardization ensures that regardless of the specific disease context, the preprocessed data will be suitable for AI/ML applications.

Data captured from wearable sensors (eg, sleep parameters, heart rate, and steps) are unique in that they are collected passively, nonobtrusively, and continuously in real-world settings [71]. For cancer applications, the identification of noninvasive biomarkers is an attractive tool for possibly predicting clinical outcomes [72]. However, current challenges of applying AI/ML techniques in the cancer research setting include data quality issues, data dimensionality, diverse data types, dynamic evolution of disease states, lack of labeled data, frequent and irregular data sparsity, and data integration issues [73]. Noninvasive wearables, such as fitness trackers, smartwatches, and many medical monitoring devices, are built using standardized design and manufacturing processes. These standard processes pertain to aspects like how data are sampled (sampling rate), how the wearables are constructed (structural aspects), and how complex the devices are. Because of these standardized methods, wearable devices can operate in a manner that captures and provides data frequently, often in real time. This continuous stream of data means that wearables are consistently generating much information. Wearable technologies are still in their infancy in cancer research because they have not been widely implemented on patients diagnosed with oncology diseases. In addition, they still face challenges in being effectively used for cancer research because of difficulties in data collection, limited types of data captured, and the scattered nature of the data storage.

Strengths and Limitations of the Review and Preprocessing Techniques

Our review provides a valuable synthesis of current preprocessing practices for wearable sensor data in cancer applications and highlights key opportunities for standardization and future research. By transparently reporting our methods and potential biases, we aim to support the interpretability and trustworthiness of our findings. Prior research has primarily focused on ML methods rather than emphasizing on standardized preprocessing techniques to make the data AI/ML ready. Key strengths and limitations are summarized in Multimedia Appendix 3. In addition, we point out potential factors that may influence the validity of our scoping review.

First, despite our comprehensive search strategy across multiple databases, it is possible that some relevant studies were not captured, particularly if they were published in nonindexed journals or as gray literature. However, we believe the risk of missing significant preprocessing methodologies is low given the breadth of our search and focus on peer-reviewed articles.

Second, categorizing preprocessing techniques required some subjective interpretation, as nomenclature was not always consistent across studies. We mitigated this by having multiple authors independently classify techniques and resolve discrepancies through discussion. Nonetheless, some overlap between categories may remain. The framework we proposed offers a generalizable taxonomy but should be further validated and refined as the field evolves.

Third, our analysis was limited to assessing the reported preprocessing workflows in each study. Without access to the underlying data sets and code, we could not directly compare the effectiveness or reproducibility of different techniques. Quantitative benchmarking of preprocessing methods on standardized wearable data sets would be a valuable direction for future work to provide more objective guidance for researchers.

Conclusions

Herein, we conducted a scoping review of preprocessing techniques by focusing exclusively on enhancing raw data from wearables before fitting AI/ML models. Recently, there has been a worldwide interest in the data quality improvement elements in the biomedical area. Our review identified 3 different preprocessing categories applicable to cancer care. Data preprocessing plays a fundamental role in the knowledge discovery from analyzing cancer-related data, especially when data are captured from wearables. A general framework within conventional preprocessing tasks, including data cleaning, data transformation, and data normalization and standardization, has been proposed with a detailed preprocessing pipeline well described. However, due to the diversity of oncology diseases, we validated the availability of significant challenges in preprocessing technique implementation for AI/ML readiness. These methods can bring significant research outcomes across the enhancement of wearable data while addressing data quality issues through different data sets with diverse specifications. The general preprocessing framework proposed in this study represents a significant step toward standardizing the preparation of wearable sensor data for AI/ML applications. While developed in the context of cancer care, its principles are broadly applicable and adaptable to other chronic diseases requiring continuous monitoring. Future research should focus on validating and refining this framework across diverse health care contexts, potentially leading to more efficient and effective use of wearable sensor data in precision medicine.

Acknowledgments

The authors would like to thank Mrs Kate Saylor for her productive collaboration throughout the methodology development of this work. VG is supported by a National Heart, Lung, and Blood Institute (NHLBI) grant (K24HL156896). SWC is supported by the National Cancer Institute and NHLBI grants (R01CA249211 and K24HL156896). BLO is supported by an NHLBI training grant (T32HL007622).

Data Availability

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Authors' Contributions

Data extraction was performed by 3 authors (BLO, VG, and SWC) by mutual agreement, and discrepancies were resolved by discussion with other coauthors (RK, XC, AS, AJ, and CZ). The outcomes from the themes' categorization part were finally evaluated independently by each author. All listed authors have reviewed and contributed to the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search queries.
[DOCX File , 17 KB - mhealth_v12i1e59587_app1.docx]

Multimedia Appendix 2
PRISMA-ScR (Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews) checklist.
[DOCX File , 108 KB - mhealth_v12i1e59587_app2.docx]

Multimedia Appendix 3
Strengths and Limitations of Preprocessing Approaches.
[DOCX File , 16 KB - mhealth_v12i1e59587_app3.docx]

References

1. What is digital health? U.S. Food & Drug Administration. URL: https://www.fda.gov/medical-devices/digital-health-center-excellence/what-digital-health [accessed 2023-10-17]
2. Dunn J, Runge R, Snyder M. Wearables and the medical revolution. Per Med 2018 Sep;15(5):429-448. [doi: 10.2217/pme-2018-0044] [Medline: 30259801]
3. Kasoju N, Remya NS, Sasi R, Sujesh S, Soman B, Kesavadas C, et al. Digital health: trends, opportunities and challenges in medical devices, pharma and bio-technology. CSIT 2023 Apr 11;11:11-30. [doi: 10.1007/s40012-023-00380-3]
4. Nittas V, Mütsch M, Ehrler F, Puhan MA. Electronic patient-generated health data to facilitate prevention and health promotion: a scoping review protocol. BMJ Open 2018 Aug 10;8(8):e021245 [FREE Full text] [doi: 10.1136/bmjopen-2017-021245] [Medline: 30099392]
5. Tyler J, Choi SW, Tewari M. Real-time, personalized medicine through wearable sensors and dynamic predictive modeling: a new paradigm for clinical medicine. Curr Opin Syst Biol 2020 Apr;20:17-25 [FREE Full text] [doi: 10.1016/j.coisb.2020.07.001] [Medline: 32984661]
6. Flora C, Tyler J, Mayer C, Warner DE, Khan SN, Gupta V, et al. High-frequency temperature monitoring for early detection of febrile adverse events in patients with cancer. Cancer Cell 2021 Sep 13;39(9):1167-1168 [FREE Full text] [doi: 10.1016/j.ccell.2021.07.019] [Medline: 34388378]
7. Kumar R, Fu J, Ortiz BL, Cao X, Shedden K, Choi SW. Dyadic and individual variation in 24-hour heart rates of cancer patients and their caregivers. Bioengineering (Basel) 2024 Jan 18;11(1):95 [FREE Full text] [doi: 10.3390/bioengineering11010095] [Medline: 38247972]
8. Zhong Y, Li T, Fong S, Li X, Tallón-Ballesteros AJ, Mohammed S. A novel pre-processing method for enhancing classification over sensor data streams using subspace probability detection. In: Proceedings of the 16th International Conference on Hybrid Artificial Intelligent Systems. 2021 Presented at: HAIS 2021; September 22-24, 2021; Bilbao, Spain. [doi: 10.1007/978-3-030-86271-8_4]
9. Zhong Y, Fong S, Hu S, Wong R, Lin W. A novel sensor data pre-processing methodology for the internet of things using anomaly detection and transfer-by-subspace-similarity transformation. Sensors (Basel) 2019 Oct 18;19(20):4536 [FREE Full text] [doi: 10.3390/s19204536] [Medline: 31635371]

10. Tawakuli A, Kaiser D, Engel T. Synchronized preprocessing of sensor data. In: Proceedings of the IEEE International Conference on Big Data. 2020 Presented at: Big Data 2020; December 10-13, 2020; Atlanta, GA. [doi: 10.1109/bigdata50022.2020.9377900]
11. Gupta S, Gupta A. Dealing with noise problem in machine learning data-sets: a systematic review. Procedia Comput Sci 2019;161:466-474. [doi: 10.1016/j.procs.2019.11.146]
12. Emmanuel T, Maupong T, Mpoeleng D, Semong T, Mphago B, Tabona O. A survey on missing data in machine learning. J Big Data 2021;8(1):140 [FREE Full text] [doi: 10.1186/s40537-021-00516-9] [Medline: 34722113]
13. D E, K U, S A, GR A, M M, P G. An innovative non-invasive approach to personalized nutrition and metabolism monitoring using wearable sensors. In: Proceedings of the International Conference on Intelligent Technologies for Sustainable Electric and Communications Systems. 2023 Presented at: iTech SECOM 2023; December 18-19, 2023; Coimbatore, India. [doi: 10.1109/itechsecom59882.2023.10435320]
14. Ramesh AN, Kambhampati C, Monson JR, Drew PJ. Artificial intelligence in medicine. Ann R Coll Surg Engl 2004 Sep;86(5):334-338 [FREE Full text] [doi: 10.1308/147870804290] [Medline: 15333167]
15. Mitchell TM. Machine Learning. New York City, NY: McGraw-Hill Education; 1997.
16. Mitchell TM. Machine learning and data mining. Commun ACM 1999 Nov 1;42(11):30-36. [doi: 10.1145/319382.319388]
17. Beam AL, Kohane IS. Big data and machine learning in health care. JAMA 2018 Apr 03;319(13):1317-1318. [doi: 10.1001/jama.2017.18391] [Medline: 29532063]
18. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. N Engl J Med 2019 Apr 04;380(14):1347-1358. [doi: 10.1056/NEJMra1814259] [Medline: 30943338]
19. Askarian B, Ho P, Chong JW. Detecting cataract using smartphones. IEEE J Transl Eng Health Med 2021 Apr 20;9:3800110 [FREE Full text] [doi: 10.1109/JTEHM.2021.3074597] [Medline: 34786216]
20. Askarian B, Tabei F, Tipton GA, Chong JW. Novel keratoconus detection method using smartphone. In: Proceedings of the IEEE Healthcare Innovations and Point of Care Technologies. 2019 Presented at: HI-POCT 2019; November 20-22, 2019; Bethesda, MD. [doi: 10.1109/hi-poct45284.2019.8962648]
21. Askarian B, Tabei F, Askarian A, Chong JW. An affordable and easy-to-use diagnostic method for keratoconus detection using a smartphone. In: Proceedings of SPIE 2018. 2018 Presented at: SPIE 2018; January 30-February 1, 2018; San Francisco, CA. [doi: 10.1117/12.2293765]
22. Shoushan MM, Reyes BA, Rodriguez AM, Chong JW. Non-contact HR monitoring via smartphone and webcam during different respiratory maneuvers and body movements. IEEE J Biomed Health Inform 2021 Feb;25(2):602-612. [doi: 10.1109/jbhi.2020.2998399]
23. Muhsen IN, ElHassan T, Hashmi SK. Artificial intelligence approaches in hematopoietic cell transplantation: a review of the current status and future directions. Turk J Haematol 2018 Aug 03;35(3):152-157 [FREE Full text] [doi: 10.4274/tjh.2018.0123] [Medline: 29880463]
24. Cirillo D, Valencia A. Big data analytics for personalized medicine. Curr Opin Biotechnol 2019 Aug;58:161-167 [FREE Full text] [doi: 10.1016/j.copbio.2019.03.004] [Medline: 30965188]
25. Siegel RL, Miller KD, Wagle NS, Jemal A. Cancer statistics, 2023. CA Cancer J Clin 2023 Jan;73(1):17-48 [FREE Full text] [doi: 10.3322/caac.21763] [Medline: 36633525]
26. Racioppi A, Dalton T, Ramalingam S, Romero K, Ren Y, Bohannon L, et al. Assessing the feasibility of a novel mHealth app in hematopoietic stem cell transplant patients. Transplant Cell Ther 2021 Feb;27(2):181.e1-181.e9 [FREE Full text] [doi: 10.1016/j.jtct.2020.10.017] [Medline: 33830035]
27. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
28. Babineau J. Product review: covidence (systematic review software). J Can Health Lib Assoc 2014 Aug 01;35(2):68-71. [doi: 10.5596/c14-016]
29. Fan C, Chen M, Wang X, Wang J, Huang B. A review on data preprocessing techniques toward efficient and reliable knowledge discovery from building operational data. Front Energy Res 2021 Mar 29;9:1-17. [doi: 10.3389/fenrg.2021.652801]
30. Liu JH, Shih CY, Huang HL, Peng JK, Cheng SY, Tsai JS, et al. Evaluating the potential of machine learning and wearable devices in end-of-life care in predicting 7-day death events among patients with terminal cancer: cohort study. J Med Internet Res 2023 Aug 18;25:e47366 [FREE Full text] [doi: 10.2196/47366] [Medline: 37594793]
31. Zhao Y, Adams CM, Davis T, Zhao J, O'Rourke N, Peng H, et al. A wearable device for postoperative breast cancer rehabilitation with machine learning for motion tracking. In: Proceedings of the IEEE Global Humanitarian Technology Conference. 2022 Presented at: GHTC 2022; September 8-11, 2022; Santa Clara, CA. [doi: 10.1109/ghtc55712.2022.9910983]
32. Moscato S, Orlandi S, Giannelli A, Ostan R, Chiari L. Automatic pain assessment on cancer patients using physiological signals recorded in real-world contexts. In: Proceedings of the 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society. 2022 Presented at: EMBC 2022; July 11-15, 2022; Glasgow, UK. [doi: 10.1109/embc48229.2022.9871990]

33. Yang TY, Kuo PY, Huang Y, Lin HW, Malwade S, Lu LS, et al. Deep-learning approach to predict survival outcomes using wearable actigraphy device among end-stage cancer patients. Front Public Health 2021 Dec 9;9:730150 [FREE Full text] [doi: 10.3389/fpubh.2021.730150] [Medline: 34957004]
34. Huang Y, Roy N, Dhar E, Upadhyay U, Kabir MA, Uddin M, et al. Deep learning prediction model for patient survival outcomes in palliative care using actigraphy data and clinical information. Cancers (Basel) 2023 Apr 10;15(8):2232 [FREE Full text] [doi: 10.3390/cancers15082232] [Medline: 37190161]
35. Cos H, Li D, Williams G, Chininis J, Dai R, Zhang J, et al. Predicting outcomes in patients undergoing pancreatectomy using wearable technology and machine learning: prospective cohort study. J Med Internet Res 2021 Mar 18;23(3):e23595 [FREE Full text] [doi: 10.2196/23595] [Medline: 33734096]
36. Davoudi A, Mardini MT, Nelson D, Albinali F, Ranka S, Rashidi P, et al. The effect of sensor placement and number on physical activity recognition and energy expenditure estimation in older adults: validation study. JMIR Mhealth Uhealth 2021 May 03;9(5):e23681 [FREE Full text] [doi: 10.2196/23681] [Medline: 33938809]
37. Liu J, Zhao Y, Lai B, Wang H, Tsui KL. Wearable device heart rate and activity data in an unsupervised approach to personalized sleep monitoring: algorithm validation. JMIR Mhealth Uhealth 2020 Aug 05;8(8):e18370 [FREE Full text] [doi: 10.2196/18370] [Medline: 32755887]
38. Tedesco S, Andrulli M, Larsson M, Kelly D, Timmons S, Alamäki A, et al. Investigation of the analysis of wearable data for cancer-specific mortality prediction in older adults. In: Proceedings of the 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society. 2021 Presented at: EMBC 2021; November 1-5, 2021; Virtual Event. [doi: 10.1109/embc46164.2021.9630370]
39. Dong G, Boukhechba M, Shaffer KM, Ritterband LM, Gioeli DG, Reilley MJ, et al. Using graph representation learning to predict salivary cortisol levels in pancreatic cancer patients. J Healthc Inform Res 2021 Apr 21;5(4):401-419 [FREE Full text] [doi: 10.1007/s41666-021-00098-4] [Medline: 35419511]
40. Patel SD, Davies A, Laing E, Wu H, Mendis J, Dijk DJ. Prognostication in advanced cancer by combining actigraphy-derived rest-activity and sleep parameters with routine clinical data: an exploratory machine learning study. Cancers (Basel) 2023 Jan 13;15(2):503 [FREE Full text] [doi: 10.3390/cancers15020503] [Medline: 36672452]
41. Asghari P. A diagnostic prediction model for colorectal cancer in elderlies via internet of medical things. Int J Inf Technol 2021;13(4):1423-1429 [FREE Full text] [doi: 10.1007/s41870-021-00663-5] [Medline: 34155483]
42. Rossi LA, Melstrom LG, Fong Y, Sun V. Predicting post-discharge cancer surgery complications via telemonitoring of patient-reported outcomes and patient-generated health data. J Surg Oncol 2021 Apr;123(5):1345-1352 [FREE Full text] [doi: 10.1002/jso.26413] [Medline: 33621378]
43. Vets N, de Groef A, Verbeelen K, Devoogdt N, Smeets A, van Assche D, et al. Assessing upper limb function in breast cancer survivors using wearable sensors and machine learning in a free-living environment. Sensors (Basel) 2023 Jul 02;23(13):6100 [FREE Full text] [doi: 10.3390/s23136100] [Medline: 37447951]
44. Feng G, Parthipan M, Breunis H, Timilshina N, Soto-Perez-de-Celis E, Mina DS, et al. Daily physical activity monitoring in older adults with metastatic prostate cancer on active treatment: feasibility and associations with toxicity. J Geriatr Oncol 2023 Sep;14(7):101576. [doi: 10.1016/j.jgo.2023.101576] [Medline: 37421787]
45. van den Eijnden MA, van der Stam JA, Bouwman RA, Mestrom EH, Verhaegh WF, van Riel NA, et al. Machine learning for postoperative continuous recovery scores of oncology patients in perioperative care with data from wearables. Sensors (Basel) 2023 May 02;23(9):4455 [FREE Full text] [doi: 10.3390/s23094455] [Medline: 37177659]
46. S VS, Royea R, Buckman KJ, Benardis M, Holmes J, Fletcher RL, et al. An introduction to the cyrcadia breast monitor: a wearable breast health monitoring device. Comput Methods Programs Biomed 2020 Dec;197:105758. [doi: 10.1016/j.cmpb.2020.105758] [Medline: 33007593]
47. Barber EL, Garg R, Strohl A, Roque D, Tanner E. Feasibility and prediction of adverse events in a postoperative monitoring program of patient-reported outcomes and a wearable device among gynecologic oncology patients. JCO Clin Cancer Inform 2022 Mar;6(1):e2100167 [FREE Full text] [doi: 10.1200/CCI.21.00167] [Medline: 35427184]
48. Jacobsen M, Gholamipoor R, Dembek TA, Rottmann P, Verket M, Brandts J, et al. Wearable based monitoring and self-supervised contrastive learning detect clinical complications during treatment of hematologic malignancies. NPJ Digit Med 2023 Jun 02;6(1):105 [FREE Full text] [doi: 10.1038/s41746-023-00847-2] [Medline: 37268734]
49. Li J, Wang Z, Zhao H, Qiu S, Zhang K, Shi X, et al. Physical fitness assessment for cancer patients using multi-model decision fusion based on multi-source data. IEEE Trans Emerg Top Comput Intell 2023 Aug;7(4):1290-1300. [doi: 10.1109/tetci.2022.3221129]
50. Jovanov E. Preliminary analysis of the use of smartwatches for longitudinal health monitoring. In: Proceedings of the 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2015 Presented at: EMBC 2015; August 25-29, 2015; Milan, Italy. [doi: 10.1109/embc.2015.7318499]
51. Larradet F, Niewiadomski R, Barresi G, Caldwell DG, Mattos LS. Toward emotion recognition from physiological signals in the wild: approaching the methodological issues in real-life data collection. Front Psychol 2020 Jul 15;11:1111 [FREE Full text] [doi: 10.3389/fpsyg.2020.01111] [Medline: 32760305]

52. Sunny JS, Patro CP, Karnani K, Pingle SC, Lin F, Anekoji M, et al. Anomaly detection framework for wearables data: a perspective review on data concepts, data analysis algorithms and prospects. Sensors (Basel) 2022 Jan 19;22(3):756 [FREE Full text] [doi: 10.3390/s22030756] [Medline: 35161502]
53. Batista GE, Monard MC. An analysis of four missing data treatment methods for supervised learning. Appl Artif Intell 2003;17(5-6):519-533. [doi: 10.1080/713827181]
54. van Buuren S, Groothuis-Oudshoorn K. mice: multivariate imputation by chained equations in R. J Stat Softw 2011;45(3):1-67 [FREE Full text] [doi: 10.18637/jss.v045.i03]
55. Stekhoven DJ, Bühlmann P. MissForest--non-parametric missing value imputation for mixed-type data. Bioinformatics 2012 Jan 01;28(1):112-118. [doi: 10.1093/bioinformatics/btr597] [Medline: 22039212]
56. Gao X, Shi F, Shen D, Liu M. Task-induced pyramid and attention GAN for multimodal brain image imputation and classification in Alzheimer's disease. IEEE J Biomed Health Inform 2022 Jan;26(1):36-43. [doi: 10.1109/jbhi.2021.3097721]
57. Liu M, Li S, Yuan H, Ong ME, Ning Y, Xie F, et al. Handling missing values in healthcare data: a systematic review of deep learning-based imputation techniques. Artif Intell Med 2023 Aug;142:102587. [doi: 10.1016/j.artmed.2023.102587] [Medline: 37316097]
58. Tang S, Chappell GT, Mazzoli A, Tewari M, Choi SW, Wiens J. Predicting acute graft-versus-host disease using machine learning and longitudinal vital sign data from electronic health records. JCO Clin Cancer Inform 2020 Feb;4:128-135 [FREE Full text] [doi: 10.1200/CCI.19.00105] [Medline: 32083957]
59. Richman JS, Moorman JR. Physiological time-series analysis using approximate entropy and sample entropy. Am J Physiol Heart Circ Physiol 2000 Jun;278(6):H2039-H2049 [FREE Full text] [doi: 10.1152/ajpheart.2000.278.6.H2039] [Medline: 10843903]
60. Fu TC. A review on time series data mining. Eng Appl Artif Intell 2011 Feb;24(1):164-181. [doi: 10.1016/j.engappai.2010.09.007]
61. Johnston W, Judice PB, Molina García P, Mühlen JM, Lykke Skovgaard E, Stang J, et al. Recommendations for determining the validity of consumer wearable and smartphone step count: expert statement and checklist of the INTERLIVE network. Br J Sports Med 2021 Jul;55(14):780-793 [FREE Full text] [doi: 10.1136/bjsports-2020-103147] [Medline: 33361276]
62. Faust L, Purta R, Hachen D, Striegel A, Poellabauer C, Lizardo O, et al. Exploring compliance: observations from a large scale Fitbit study. In: Proceedings of the 2nd International Workshop on Social Sensing. 2017 Presented at: SocialSens'17; April 18-21, 2017; Pittsburgh, PA. [doi: 10.1145/3055601.3055608]
63. Hardcastle SJ, Jiménez-Castuera R, Maxwell-Smith C, Bulsara MK, Hince D. Fitbit wear-time and patterns of activity in cancer survivors throughout a physical activity intervention and follow-up: exploratory analysis from a randomised controlled trial. PLoS One 2020 Oct 19;15(10):e0240967 [FREE Full text] [doi: 10.1371/journal.pone.0240967] [Medline: 33075100]
64. Chan A, Chan D, Lee H, Ng CC, Yeo AH. Reporting adherence, validity and physical activity measures of wearable activity trackers in medical research: a systematic review. Int J Med Inform 2022 Apr;160:104696 [FREE Full text] [doi: 10.1016/j.ijmedinf.2022.104696] [Medline: 35121356]
65. Condry MW, Quan XI. Digital health innovation, informatics opportunity, and challenges. IEEE Eng Manag Rev 2021 Jun;49(2):81-88. [doi: 10.1109/emr.2021.3054330]
66. Triantafyllidis A, Kondylakis H, Katehakis D, Kouroubali A, Koumakis L, Marias K, et al. Deep learning in mHealth for cardiovascular disease, diabetes, and cancer: systematic review. JMIR Mhealth Uhealth 2022 Apr 04;10(4):e32344 [FREE Full text] [doi: 10.2196/32344] [Medline: 35377325]
67. Aledhari M, Razzak R, Qolomany B, Al-Fuqaha A, Saeed F. Biomedical IoT: enabling technologies, architectural elements, challenges, and future directions. IEEE Access 2022 Mar 14;10:31306-31339. [doi: 10.1109/access.2022.3159235]
68. Qaisar SM, Mihoub A, Krichen M, Nisar H. Multirate processing with selective subbands and machine learning for efficient arrhythmia classification. Sensors (Basel) 2021 Feb 22;21(4):1511 [FREE Full text] [doi: 10.3390/s21041511] [Medline: 33671583]
69. Efat MI, Rahman S, Rahman T. IoT based smart health monitoring system for diabetes patients using neural network. In: Proceedings of the Second EAI International Conference on Cyber Security and Computer Science. 2020 Presented at: ICONCS 2020; February 15-16, 2020; Dhaka, Bangladesh. [doi: 10.1007/978-3-030-52856-0_47]
70. Tabei F, Gresham JM, Askarian B, Jung K, Chong JW. Cuff-less blood pressure monitoring system using smartphones. IEEE Access 2020 Jan 08;8:11534-11545. [doi: 10.1109/access.2020.2965082]
71. Cho S, Ensari I, Weng C, Kahn MG, Natarajan K. Factors affecting the quality of person-generated wearable device data and associated challenges: rapid systematic review. JMIR Mhealth Uhealth 2021 Mar 19;9(3):e20738 [FREE Full text] [doi: 10.2196/20738] [Medline: 33739294]
72. Fuse K, Uemura S, Tamura S, Suwabe T, Katagiri T, Tanaka T, et al. Patient-based prediction algorithm of relapse after allo-HSCT for acute Leukemia and its usefulness in the decision-making process using a machine learning approach. Cancer Med 2019 Sep;8(11):5058-5067 [FREE Full text] [doi: 10.1002/cam4.2401] [Medline: 31305031]
73. Gupta V, Braun TM, Chowdhury M, Tewari M, Choi SW. A systematic review of machine learning techniques in hematopoietic stem cell transplantation (HSCT). Sensors (Basel) 2020 Oct 27;20(21):6100 [FREE Full text] [doi: 10.3390/s20216100] [Medline: 33120974]

Abbreviations

AI/ML: artificial intelligence and machine learning
mHealth: mobile health
MLM: machine learning model
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by L Buis; submitted 16.04.24; peer-reviewed by R Matovu, L Guo; comments to author 17.05.24; revised version received 12.06.24; accepted 27.08.24; published 27.09.24.

Please cite as:
Ortiz BL, Gupta V, Kumar R, Jalin A, Cao X, Ziegenbein C, Singhal A, Tewari M, Choi SW
Data Preprocessing Techniques for AI and Machine Learning Readiness: Scoping Review of Wearable Sensor Data in Cancer Care
JMIR Mhealth Uhealth 2024;12:e59587
URL: https://mhealth.jmir.org/2024/1/e59587
doi:10.2196/59587
PMID:38626290

Review

Usability Assessment Methods for Mobile Apps for Physical Rehabilitation: Umbrella Review

Sylvia Hach[1*], BA, BSc, MA, PhD; Gemma Alder[1*], BSc, MSc, PhD; Verna Stavric[1*], BSc, MHSc, PhD; Denise Taylor[1*], Certificate Health Economics, Grad Dip Physiotherapy, MSc, PhD; Nada Signal[1*], BHSc, MHSc, PhD

Health and Rehabilitation Research Institute, Faculty of Health and Environmental Sciences, Auckland University of Technology, Auckland, New Zealand

[*]all authors contributed equally

Corresponding Author:
Sylvia Hach, BA, BSc, MA, PhD
Health and Rehabilitation Research Institute
Faculty of Health and Environmental Sciences
Auckland University of Technology
90 Akoranga Drive
Northcote
Auckland, 0627
New Zealand
Phone: 64 2108184489
Email: sylvia.hach@googlemail.com

Abstract

Background: Usability has been touted as one determiner of success of mobile health (mHealth) interventions. Multiple systematic reviews of usability assessment approaches for different mHealth solutions for physical rehabilitation are available. However, there is a lack of synthesis in this portion of the literature, which results in clinicians and developers devoting a significant amount of time and effort in analyzing and summarizing a large body of systematic reviews.

Objective: This study aims to summarize systematic reviews examining usability assessment instruments, or measurements tools, in mHealth interventions including physical rehabilitation.

Methods: An umbrella review was conducted according to a published registered protocol. A topic-based search of PubMed, Cochrane, IEEE Xplore, Epistemonikos, Web of Science, and CINAHL Complete was conducted from January 2015 to April 2023 for systematic reviews investigating usability assessment instruments in mHealth interventions including physical exercise rehabilitation. Eligibility screening included date, language, participant, and article type. Data extraction and assessment of the methodological quality (AMSTAR 2 [A Measurement Tool to Assess Systematic Reviews 2]) was completed and tabulated for synthesis.

Results: A total of 12 systematic reviews were included, of which 3 (25%) did not refer to any theoretical usability framework and the remaining (n=9, 75%) most commonly referenced the ISO framework. The sample referenced a total of 32 usability assessment instruments and 66 custom-made, as well as hybrid, instruments. Information on psychometric properties was included for 9 (28%) instruments with satisfactory internal consistency and structural validity. A lack of reliability, responsiveness, and cross-cultural validity data was found. The methodological quality of the systematic reviews was limited, with 8 (67%) studies displaying 2 or more critical weaknesses.

Conclusions: There is significant diversity in the usability assessment of mHealth for rehabilitation, and a link to theoretical models is often lacking. There is widespread use of custom-made instruments, and preexisting instruments often do not display sufficient psychometric strength. As a result, existing mHealth usability evaluations are difficult to compare. It is proposed that multimethod usability assessment is used and that, in the selection of usability assessment instruments, there is a focus on explicit reference to their theoretical underpinning and acceptable psychometric properties. This could be facilitated by a closer collaboration between researchers, developers, and clinicians throughout the phases of mHealth tool development.

Trial Registration: PROSPERO CRD42022338785; https://www.crd.york.ac.uk/prospero/#recordDetails

(JMIR Mhealth Uhealth 2024;12:e49449) doi:10.2196/49449

KEYWORDS

usability; quality evaluation; mobile health; physical exercise; rehabilitation; overview; umbrella review; psychometrics

Introduction

The development of mobile health (mHealth) [1,2] solutions has seen exponential growth in recent times, driven particularly by the global pandemic [3,4]. mHealth has been heralded as a tool to provide access to quality rehabilitation input for patients outside of the time they are able to spend with clinicians [5] and for patients in geographically remote areas [6]. Furthermore, similar to the observed trend of increased health information seeking on the internet [7], the democratization of access to rehabilitation could be achieved by individuals actively seeking stand-alone mHealth solutions.

However, there is also increasing awareness that mHealth solutions available to clinicians and their patients often lack quality evaluations [8,9]. Many mHealth solutions only have short-term (<30 days) data from small sample sizes to support their effectiveness [10]. Moreover, only limited standardized outcome measures are typically used [11,12].

Usability is one key aspect commonly included in the evaluation of mHealth solutions [5,9,11]. It has been touted as a determiner of the success of mHealth interventions [13]. Usability is often delineated from two related concepts: (1) the concept of utility that captures a system's ability to meet user needs [14] and (2) user experience is commonly understood as a broader concept of the experience of using an mHealth solution and may include measures of user beliefs [15]. However, usability may or may not be part of how user experience is captured, and many different definitions of usability appear in the literature [15-17].

The diversity in definitions of usability is mirrored by the diversity in usability models or frameworks. The 5 most commonly cited models of usability are that of ISO9241-11 [18] and its revision [19]; ISO/IEC25010 [20]; Nielsen's usability model [21]; and, in the context of health in particular, the People At the Centre of Mobile Application Development (PACMAD) model [14,22]. These models identify factors such as efficiency, or the resources expended to achieve a task; effectiveness, the level of accuracy and completeness of a task achieved using a mobile solution; and satisfaction or positive user interaction while operating the mobile solution as components of usability. The key difference between the PACMAD and the aforementioned frameworks is that these and other factors such as errors are seen as arising from 3 different sources: the user themselves, the task, and the context. This could be argued to be of particular importance for mHealth, where users may experience limitations such as perceptual or cognitive (aging) barriers [23]. These additionally impact on task demands and therefore represent an important consideration in the design of mHealth tools.

Usability assessment has been included in several good practice guidelines for the development of mHealth solutions [24-28], as well as in many evaluation frameworks [29,30], and can be regarded as a crucial step for evaluation at different stages of the typical mHealth development cycles. To date, however, no accepted standard for the assessment of usability of mHealth solutions exists. This means that researchers and developers of mHealth are faced with difficult decisions when designing mHealth evaluation procedures that strike the balance between responsiveness, reliability, and validity and are unable to compare existing solutions for the purpose of innovating. Further, clinicians are unable to be guided in their prescription of mHealth solutions, and there are significant barriers for consumers to engage with existing solutions.

Numerous systematic reviews have explored usability assessment approaches for various mHealth solutions in the context of physical rehabilitation. However, there is a lack of synthesis in this area of the literature. This may contribute to clinicians and developers needing to devote a significant amount of time and effort in analyzing and summarizing a large body of systematic reviews. An umbrella review can act as "a means for a rapid review of the evidence to address a broad and high-quality evidence base" [31]. Specifically, an umbrella review allows for a broader scope than individual systematic reviews that may focus on individual treatment options or individual conditions [32-34]. Hence, the aim of this umbrella review was to provide a "user-friendly" summary of the use of usability assessment instruments, or measurement tools, for researchers, clinicians, and consumers of mHealth irrespective of the specific area of application (eg, diabetes, tuberculosis, and sleep). Specifically, the objective was to summarize systematic reviews that investigated usability assessment instruments in mHealth interventions including those related to physical exercise rehabilitation. It is envisaged that such a summary will first aid researchers, developers, and clinicians to gain an overview of usability assessment instruments without needing to explore primary literature. Second, the presented summary may aide the development of mHealth usability assessment standards.

Methods

Overview

The umbrella review protocol was developed based on the *Cochrane Handbook for Systematic Reviews of Interventions* [33] and other relevant methodology sources [34] and was registered with PROSPERO (CRD42022338785). StArt (State of the Art through Systematic Review) software [35] was used for the first- and second-level screening of result datasets and extracting relevant information.

Inclusion Criteria

Based on the objectives of the study, the following inclusion criteria were formulated: (1) articles published between January 1, 2015, and April 27, 2023 (the date range reflected the launch of Apple ResearchKit in 2015, which accelerated mHealth development and research [36]); (2) containing data on human participants; (3) with the "unit of searching" [33] being "systematic reviews" [37,38] in order to reduce the effect of cumulative bias that may arise when including nonsystematic reviews; (4) examining usability assessment instruments of

mobile apps for health professionals and for health care consumers; and (5) published in the English language to enable all contributing authors to perform screening, extraction, and synthesis of the search results. No post hoc modifications were made to the inclusion criteria. Systematic reviews of usability assessment instruments of other (mobile) solutions such as wearables, sensors, virtual reality, blockchain, Internet of Things, simulated data, or solutions for health care professionals only were excluded.

Search Methods and Search Terms

The following databases were searched with a combination of the search terms mobile application*, mobile app, usab*, usab* criteria, usab* evaluat*, systematic review, mhealth, mobile health, and physical exercise: PubMed, Cochrane, IEEE Xplore, Epistemonikos, Web of Science, and CINAHL Complete, combined using Boolean operators OR and AND and customized for each database in accordance with their filtering specifications. The result sets were imported into StArt [35]. The full search syntax for each database are presented in Table S1 in Multimedia Appendix 1.

Data Collection and Analysis

A preliminary search of existing systematic reviews was conducted before finalizing the search terms in order to scope the extent and type of existing evidence [33]. The subsequent final search terms produced a result set that was more refined in focus and feasible in terms of the size of the expected result set. Following the removal of duplicates, 2-level screening was performed: title and abstract screening was performed by the primary author (SH), and a randomly selected subset of articles (118/1479, approximately 8%) was screened by a second author (VS; κ=0.87). Second-level, full-text screening was performed by the primary author (SH) using StArt for data extraction from the final result set. A data extraction form including basic reference details, as well as information such as population of interest and interventions studied, was discussed and agreed on by 3 authors (SH, GA, NS) before data extraction (see review protocol PROSPERO CRD42022338785 for more detail).

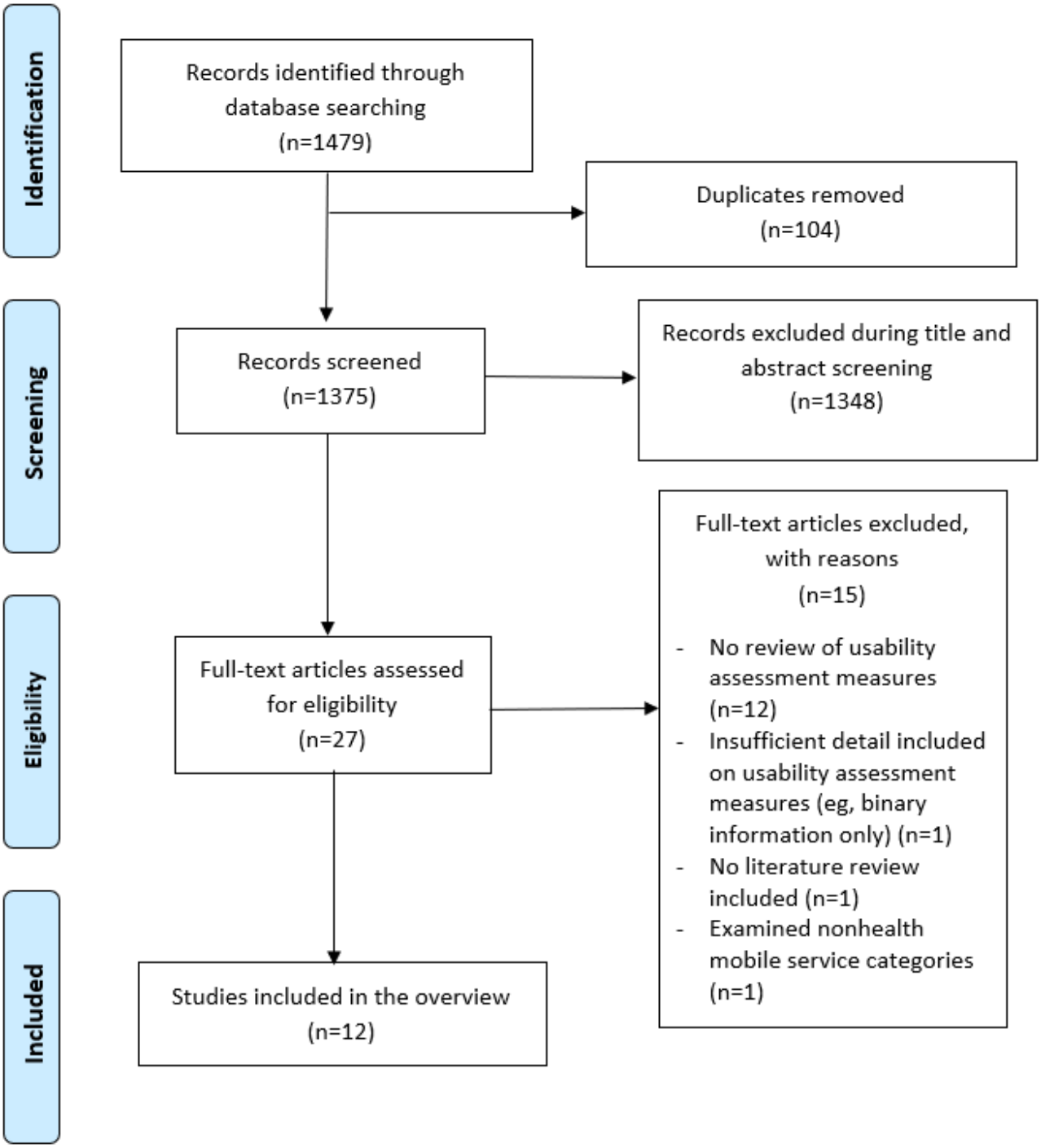
Quality assessment was completed using AMSTAR 2 (A Measurement Tool to Assess Systematic Reviews 2; Institute for Clinical Evaluative Sciences) [39] by the primary author (SH) and a second author (VS) separately (κ=0.823). Any disagreement was discussed and resolved via consensus. In line with recommendations by Shea et al [39], a discussion to determine AMSTAR 2 critical domains for this umbrella review occurred among 2 authors (SH, NS). Criteria 2, 4, and 7 were retained on the premise of constituting critical criteria as defined by the original publication [39]. The original critical criteria 9, 11, 13, and 15 were classified as noncritical for the purpose of this umbrella review due to pertaining to meta-analytic steps that none of the included systematic reviews performed. Instead, the following criteria were classified as critical: criterion 5 due to the variety of study designs and target user groups and/or clinical contexts included within the systematic reviews; and criterion 16 due to the context of mHealth usability, where the borders between academic enquiry and commercialization are more blurred and funding could constitute a significant source of bias and/or conflict of interest. A summary rating was produced according to recommendations by Shea et al [39].

Finally, to gauge potential skewing of the data caused by significant overlap of primary studies contained within the systematic reviews included in this umbrella review [40], overlap assessment was achieved via citation matrix [41,42] for the systematic reviews including the System Usability Scale (SUS) as an exemplar. The SUS was chosen because it is one of the most well-known instruments [43] and preliminary searches of the literature demonstrated its frequency of use and reference.

Results

The initial database search returned 1479 results, which were reduced to 1375 after removal of duplicates (see Figure 1). Title and abstract screening resulted in 27 articles being included for full-text screening. A total of 15 of the full-text articles retrieved (see Table S2 in Multimedia Appendix 1) were ineligible because they did not review usability assessment measures, include sufficient detail on usability assessment instruments (eg, including binary information only), include a literature review, or examine nonhealth mobile service categories (see Figure 1).

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



A total of 12 systematic reviews examining usability assessment instruments were included. Data were extracted (see Table S3 in Multimedia Appendix 1) as per the registered protocol. Across the systematic reviews included, there was coverage of primary studies from the start of records to 2020. Three of the systematic reviews included examined usability assessment instruments within a specific target user group (eg, users with diabetes [44] and users living with a mental health concern [45,46]). The remaining 9 systematic reviews [13,47-54] focused on usability assessment instruments used across different target user populations. Usability models or frameworks referenced included ISO [20] (referenced in [13,44,48,49]), Nielsen [21] (referenced in [45]), and the framework by the Canadian Institutes of Health Research and the Mental Health Commission Canada [55] (referenced in [47]). Three (25%) of the systematic reviews [46,50,51] included in this umbrella review did not refer to any theoretical framework (see Table S3 in Multimedia Appendix 1).

The systematic reviews included identified a total of 32 usability assessment instruments (see Table 1) and a further 66 custom-made usability assessment instruments as well as hybrid custom-made instruments (see Table S4 in Multimedia Appendix 1). The most commonly referenced usability assessment instrument was the SUS [56], followed by the IBM Computer Usability Satisfaction Questionnaire [57] and the Usefulness, Satisfaction, and Ease of Use (USE) Questionnaire [58].

Table 1. Overview of usability assessment scales identified by reviews included within this umbrella review.

Assessment scale	Reference	Systematic review identifying scale	Count	Psychometric properties as identified by systematic reviews included in this umbrella review						
				Internal consistency (Cronbach α)	Reliability (intraclass correlation)	Content validity	Structural validity	Cross-cultural validity	Criterion, convergent, concurrent, discriminant validity	Responsiveness
App adaptation Abbott's scale	[59]	Nouri et al [50]	1	NR[a]	NR	NR	NR	NR	NR	NR
After Scenario Questionnaire	[60]	Inal et al [45]	1	NR	NR	NR	NR	NR	NR	NR
App adaptation Brief DISCERN	[59]	Nouri et al [50]	1	NR	NR	NR	NR	NR	NR	NR
App adaptation CRAAP checklist	[61]	Nouri et al [50]	1	NR	NR	NR	NR	NR	NR	NR
Ease of Use and Usefulness Scale (EUUS)	[62]	Kien et al [53]	1	NR	NR	NR	NR	NR	NR	NR
Enlight	[63]	Azad-Khaneghah et al [47]	1	NR	NR	NR	NR	NR	NR	NR
Health Information Technology Usability Evaluation Scale (Health-ITUES)	[64]	Azad-Khaneghah et al [47], Muro-Culebras et al [51]	2	0.85-0.92	No	Expert panel and factor analysis	Exploratory and confirmatory factor analysis	No	Correlation with the Post-Study System Usability Questionnaire (PSSUQ)	Statistically significant difference was demonstrated with the intervention group
Health IT Usability Evaluation Model (Health-ITUEM)	[65]	Nouri et al [50], Vera et al [48]	2	NR	NR	NR	NR	NR	NR	NR
App adaptation Health-Related Website Evaluation Form (HRWEF)	[66]	Nouri et al [50]	1	NR	NR	NR	NR	NR	NR	NR
App adaptation Health On the Net (HON) code	[59]	Nouri et al [50]	1	NR	NR	NR	NR	NR	NR	NR
IBM Computer Usability Satisfaction Questionnaire	[57]	Azad-Khaneghah et al [47], Georgsson [44], Ng et al [46], Wakefield et al [52], Zapata et al [13]	5	0.89	No	Expert panel	No	NR	No	No
ISOMETRIC	[67]	Azad-Khaneghah et al [47]	1	NR	NR	NR	NR	NR	NR	NR
iSYScore index	[68]	Muro-Culebras et al [51]	1	No	No	Expert panel	No	NR	No	No

Assessment scale	Reference	Systematic review identifying scale	Count	Psychometric properties as identified by systematic reviews included in this umbrella review						
				Internal consistency (Cronbach α)	Reliability (intraclass correlation)	Content validity	Structural validity	Cross-cultural validity	Criterion, convergent, concurrent, discriminant validity	Responsiveness
App adaptation Kim Model	[69]	Nouri et al [50]	1	NR	NR	NR	NR	NR	NR	NR
Measurement Scales for Perceived Usefulness and Perceived Ease of Use	[70]	Muro-Culebras et al [51]	1	0.97 (usefulness), 0.91 (ease of use)	No	Focus group	Exploratory factor analysis	No	Convergent and discriminant validity	No
Mobile App Rating Scale (MARS)	[71]	Muro-Culebras et al [51], Nouri et al [50], Vera et al [48]	3	0.90	0.79	Expert panel	No	No	No	No
Mobile App Rating Scale (user version) (uMARS)	[72]	Muro-Culebras et al [51], Nouri et al [50]	2	0.90	0.66 (1-2 mo), 0.70 (3 mo)	Expert panel and focus groups	No	No	No	No
NASA Task Load Index (TLX)	[73]	Zapata et al [13]	1	NR	NR	NR	NR	NR	NR	NR
NICE guidelines tool	[74]	Azad-Khaneghah et al [47]	1	NR	NR	NR	NR	NR	NR	NR
Perceived Useful and Ease of Use Questionnaire (PUEU)	[75]	Azad-Khaneghah et al [47], Inal et al [45]	2	NR	NR	NR	NR	NR	NR	NR
Post-Study System Usability Scale (PSSUS)/PSSUQ	[76]	Inal et al [45], Niknejad et al [54], Vera et al [48]	3	NR	NR	NR	NR	NR	NR	NR
Quality Assessment tool for Evaluating Medical Apps (QAEM)	[77]	Azad-Khaneghah et al [47]	1	NR	NR	NR	NR	NR	NR	NR
Quality of Experience (QOE)	[78]	Azad-Khaneghah et al [47], Nouri et al [50]	2	NR	NR	NR	NR	NR	NR	NR
Questionnaire for User Interaction Satisfaction 7.0 (QUIS)	[79]	Georgsson [44], Saeed et al [49]	2	NR	NR	NR	NR	NR	NR	NR
App adaptation Silberg score	[80]	Azad-Khaneghah et al [47], Nouri et al [50]	2	NR	NR	NR	NR	NR	NR	NR
Software Usability Measurement Inventory (SUMI)	[81]	Azad-Khaneghah et al [47]	1	NR	NR	NR	NR	NR	NR	NR

Assessment scale	Reference	Systematic review identifying scale	Count	Psychometric properties as identified by systematic reviews included in this umbrella review						
				Internal consistency (Cronbach α)	Reliability (intraclass correlation)	Content validity	Structural validity	Cross-cultural validity	Criterion, convergent, concurrent, discriminant validity	Responsiveness
System Usability Scale (SUS)	[56]	Azad-Khaneghah et al [47], Georgsson [44], Inal et al [45], Muro-Culebras et al [51], Ng et al [46], Niknejad et al [54], Nouri et al [50], Vera et al [48], Wakefield et al [52], Zapata et al [13]	10	0.911	No	Focus group	Exploratory and confirmatory factor analysis	No	No	No
Telehealth Usability Questionnaire (TUQ)	[82]	Georgsson [44], Inal et al [45], Niknejad et al [54]	3	NR	NR	NR	NR	NR	NR	NR
Telemedicine Satisfaction and Usefulness Questionnaire (TSUQ)	[83]	Wakefield et al [52]	1	0.96 (video visits), 0.92 (use and impact)	No	Expert panel	Exploratory factor analysis	No	Significant discriminant validity (Hispanic vs non-Hispanic)	No
The mHealth App Usability Questionnaire for interactive mHealth apps (patient version) (MAUQ)	[84]	Muro-Culebras et al [51]	1	0.895, 0.829, 0.900	No	Expert panel	Exploratory factor analysis	No	Correlation with PSSUQ and SUS	No
The mHealth App Usability Questionnaire for standalone mHealth apps (patient version) (MAUQ)	[84]	Muro-Culebras et al [51]	1	0.847, 0.908, 0.717	No	Expert panel	Exploratory factor analysis	No	Correlation with PSSUQ and SUS	No
Usefulness, Satisfaction, and Ease of Use (USE) Questionnaire	[58]	Azad-Khaneghah et al [47], Inal et al [45], Kien et al [53], Ng et al [46]	4	NR	NR	NR	NR	NR	NR	NR

[a]NR: not reported as part of the systematic reviews included in this umbrella review.

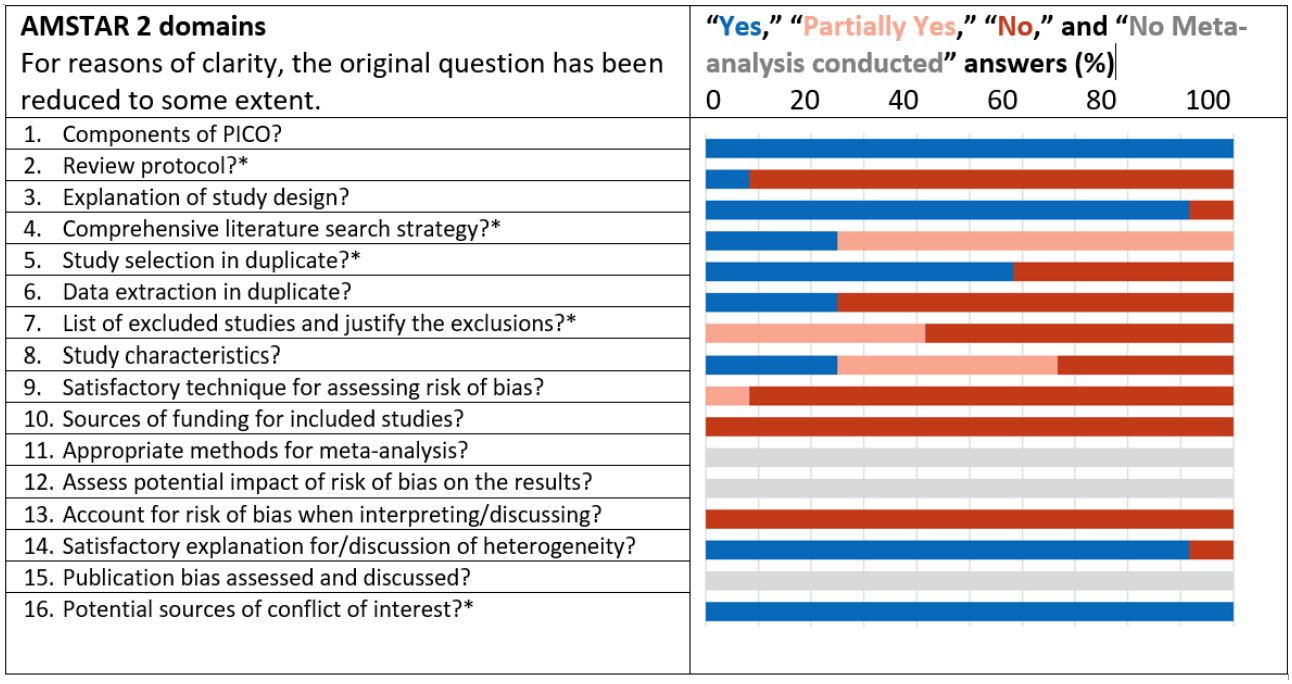
Data regarding the psychometric properties of 9 (28%) instruments [56,57,64,70-72,83,84] were included in the systematic reviews as detailed in Table 1. Internal consistency was generally good across these instruments, content validity was provided through expert panel or focus groups [56,57,64,70,71,83,84], and exploratory and/or confirmatory factor analyses were used in evidence of structural validity [56,64,70,83,84]. Details of convergent validity were included for 3 instruments [64,70,84] (see Table 1). Importantly, there was no evidence of reliability, responsiveness, or cross-cultural validity assessment for the usability assessment instruments referenced most often (ie, SUS, IBM Computer Usability Satisfaction Questionnaire, and USE Questionnaire).

Further, 8 (67%) of the systematic reviews [13,44-46,48-50,54] referred to usability assessment methods other than assessment scales. These included focus groups, heuristic evaluation,

think-aloud protocols, and other methods (see Table S5 in Multimedia Appendix 1).

Quality assessment of the systematic reviews using AMSTAR 2 revealed that 8 (67%) articles [13,44-46,48-50,54] exhibited at least 2 critical weaknesses (see Figure 2), 3 (25%) systematic reviews [47,51,52] were affected by 1 critical weakness, and 1 (8%) review [53] had only noncritical weaknesses. The most frequently unfulfilled assessment criteria included the sources of funding enquiry for the included studies (AMSTAR criterion 10), accounting for risk of bias when interpreting results (AMSTAR criterion 13), use of a satisfactory technique for assessing risk of bias (AMSTAR criterion 9), and inclusion of a review protocol (AMSTAR criterion 2; see Table S6 in Multimedia Appendix 1).

Figure 2. Overview of methodological quality of reviews according to AMSTAR 2 (A Measurement Tool to Assess Systematic Reviews 2). * denotes critical criterion as determined for this umbrella review.



Finally, visualization of citation overlap for systematic reviews including primary studies using the SUS showed minimal overlap with 4 (10%) of 41 primary studies included in 2 of the systematic reviews (see Table S7 in Multimedia Appendix 1). With the exception of the citation of the original publication of the SUS instrument [56], all other references included in the overview were unique to one of the systematic reviews included.

Discussion

Principal Findings

The exponential growth of research evidence related to the effectiveness of mobile solutions for rehabilitation [85-88] and the proliferation of technological solutions that afford new modes of treatment delivery [89,90] underscore the critical need for high-quality mHealth usability evaluation. Usability attributes such as efficiency, learnability, and memorability [21] are particularly important to consider for mHealth users who may face challenges due to neurological compromise [91], age-related issues [23], or limited technology experience [13]. This umbrella review aimed to summarize usability assessment instruments for mHealth researchers, clinicians, and consumers to guide the development, assessment, and selection of high-quality mHealth tools.

The review identified, first, significant diversity and common use of custom-made instruments when usability assessment instruments were employed to evaluate mHealth tools for rehabilitation. Second, there was a notable lack of theoretical grounding for selection of the assessment of usability. Third, a scarcity of psychometric data for widely used instruments for mHealth usability assessment was evident in the systematic reviews included.

Heterogeneity of Instruments, Including Nonstandardized Instruments

Regarding the first critical point, a wide range of different instruments for the assessment of usability was evident across the systematic reviews included. This range included adaptations of preexisting usability assessment instruments for the context of mobile apps [59,66] as well as assessment instruments, such as the Mobile App Rating Scale (MARS) [72], specifically designed for usability assessment of mHealth tools. In addition, both completely custom-made instruments and hybrids [92] of preexisting instruments with custom elements were prevalent in the mHealth usability literature.

Although the use of hybrid assessment instruments and adaptations of preexisting assessment instruments may increase flexibility and thereby possibly improve the experience for respondents, the fact that most studies are limited in sample size prevents validation of hybrid and adapted instruments [51]. Alternative approaches to increasing flexibility and improving respondent experience while ensuring psychometric integrity are needed instead. A good example of this may be seen in the creation of a hybrid version of the SUS with the inclusion of

pictorial elements, which increased respondent motivation [92]. Importantly, acceptable validity, consistency, and sensitivity were also evidenced, allowing future users of the hybrid measure to place greater trust in the quality of the data.

Theoretical Underpinning

Second, and similar to what has been found to be the case for individual-level studies assessing the usability of specific mHealth tools [47], this review revealed that some systematic reviews examining the broader literature related to usability assessment lacked connection to theoretical models of usability. This observation resonates with previous criticisms of the quality of reviews of health-related mobile apps [8] as well as research exploring technology adoption in fields beyond mHealth [93]. The latter exposed a reliance on a wide array of theoretical models of technology adoption in the literature and in some cases several within one review. To address this, it has been suggested that generic models for different service categories (eg, information and transaction) be developed [93]. A theoretically grounded, generic guide for mHealth usability assessment could similarly promote broader adoption and enhance comparison of usability across studies and use cases.

Psychometric Properties and Psychometric Testing

Third, systematic reviews included in our overview also reported significant limitations regarding the psychometric properties of preexisting instruments. For example, the MARS tool, which has been put forward as an instrument for standardized use in mHealth usability assessment [51], lacks structural validity. Moreover, other constructs such as internal consistency and criterion validity have been documented as significant areas of future work for measuring the implementation of interventions [53], with usability assessment playing a significant role.

Although consistent with previous research, this umbrella review did not specifically search for psychometric evaluations of usability assessment instruments; instead, it relied on summaries of psychometric evaluations presented as part of the included systematic reviews. As a result, it is likely that psychometric evaluation of other instruments is available. For example, psychometric evaluation of the popular USE Questionnaire [58] is available and, consistent with our observation, has been shown to be affected by a lack of reliability and validity [94]. Furthermore, outside of the academic literature, there is a still greater portion of mHealth solutions on the market that likely will not have undergone empirical evaluation of usability.

Although some of the acceptable psychometric information was referenced for the SUS [95], both the IBM Usability Satisfaction Questionnaire and the USE Questionnaire appear to lack reliability assessment. Reliability, or the freedom of measurement error [96], may be regarded as crucial with regard to any metrics that are gathered after, rather than during, a user's interaction with an application. The inability to separate true change in users' estimate of the usability of mHealth tools from random variation, or measurement error, originating from recall bias [9,97,98], for example, means that mHealth tool iterations [99] are unable to be evaluated appropriately.

Moreover, the widespread use of custom-made and hybrid assessment instruments leads to the loss of the original instrument's integrity and compromises its already-documented psychometric strengths [100]. Consequently, establishing the validity of results from individual usability investigations becomes challenging, and comparison across studies is difficult. Hence, there is an urgent need to assess the accuracy and appropriateness [101] of individual usability assessment instruments to capitalize on the promise of mHealth tools in rehabilitation [5,6].

Another important psychometric aspect of usability assessment instruments that the systematic reviews included in this umbrella review highlight as missing from the published literature is responsiveness. mHealth development usually involves iterative design and testing cycles [30,99] with associated formative and summative usability evaluation [45]. Across the life of mHealth development, iterative cycles are likely to span different stages of development and be undertaken in different clinical contexts [54,102]. Integrating usability assessment into this process requires instruments that are generic enough to capture user responses to a wide variety of mHealth strategies but also fine-grained enough to possess sufficient responsiveness [96].

Finally, with regard to the argument of lacking psychometric assessment, none of the preexisting mHealth usability assessment instruments referenced as part of the literature included in this umbrella review appear to have been informed by a breadth of cultural perspective or undergone cross-cultural validity testing. Given the global potential of mHealth to address inequities in access to and outcomes from rehabilitation [5,6], it is particularly important to establish cross-cultural validity of the usability assessment instruments employed in mHealth development. In addition, with the pervasiveness of technology, there is a certain element of unpredictability of the context in which mHealth tools will be trialed and used "in the wild" [103,104]. For that reason, an alternative argument could be made for innovative, culturally responsive methodology for mHealth tool design including usability testing [105]. A key difference in such attempts is user participation at multiple stages of development and responsiveness to expanding the stages of development as guided by stakeholders. This process likely includes constant negotiation and may be resource heavy but is arguably needed if the aim is to create mHealth solutions impacting indigenous outcomes, for example [74,106].

Considering the identified issues, including lack of theoretical grounding, common use of custom-made assessment instruments, and the scarcity of psychometric data for widely used mHealth usability assessment instruments, multimethod usability assessment appears paramount. This is consistent with recommendations made by a number of research groups [13,44,102,107] and reinforces the argument often advanced in favor of Ecological Momentary Assessment approaches, which are recognized for their advantage over retrospective assessment [97]. It is therefore proposed that standards be developed that specify the time points in the mHealth life cycle at which usability assessment is completed, with an emphasis on what methods to use. Moreover, these standards should mandate that individual assessment instruments are grounded in a theoretical framework and possess a minimum threshold for psychometric properties [53,108].

Recommendations

The establishment of a universal usability scoring system or algorithm would further facilitate the integration of these assessments into an overall framework [109]. It has been observed that, at present, less than half of existing evaluation frameworks include such a scoring system, but that such systems could support funding decisions [29] and advance the vision of prescribable mHealth apps [10]. Although technological advancement often outpaces academic enquiry necessitating new approaches to mHealth evaluation frameworks [110], usability factors are enduring [16] and investing resources into establishing standards will therefore be valuable.

Limitations

In the context of an area of practice where the lines between commercial and academic work are blurred and usability assessment constitutes a common practice in the global commercial environment [111], this umbrella review is limited to only including English language systematic reviews published within the academic literature indexed in the databases included. Furthermore, the quality of the included systematic reviews was found to be limited, and the fit of the AMSTAR 2 tool with methodological papers is not perfect. However, AMSTAR 2 could be argued to be more detailed than instruments developed for umbrella reviews specifically [31], and, in line with the AMSTAR 2 recommendations [39], the authors modified the list of critical criteria to reflect the specific aim of the overview. Finally, with regard to the review's methodology, 2 limitations are of note. First, although the search syntax for this umbrella review included the keyword "physical exercise," for pragmatic reasons, no validation step was included to confirm that all mHealth tools examined as part of the primary studies included within the systematic reviews included a physical exercise component. Regardless, the observations presented here are valid for mHealth tools for rehabilitation overall and provide valuable guidance to developers, researchers, and clinicians. Second, for practical reasons, data selection could only be performed by the primary author (SH) with a subset of articles being screened by a second author (VS). However, agreement on study selection was high (>80%), supporting the quality of the review.

Conclusions

There is considerable variety in approaches to and instruments for the assessment of usability in mHealth for rehabilitation, many of which lack theoretical foundation. Clinicians are therefore advised to critically evaluate mHealth literature and solutions, paying particular attention to the population in which usability testing was performed and the specific usability assessment instruments were employed. Future research efforts should be focused on producing high-quality systematic reviews and psychometric evaluations of usability assessment instruments. A collaborative effort between researchers, designers, and developers is essential to establish mHealth tool development standards. These standards should emphasize the incorporation of usability assessment instruments underpinned by a robust theoretical base. This umbrella review represents a valuable reference tool in this endeavor. Inclusion of multimethod usability assessment within the wider mHealth development cycle could also be part of these standards, which will ensure that we can capitalize on the widely heralded promise of mHealth to promote access to and outcomes from rehabilitation.

Acknowledgments

The authors thank the wider team of researchers and clinicians at the AUT Research Innovation Centre for workshop and input, and Exsurgo for valuable conversations on usability from a commercial perspective.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary files.
[DOCX File , 153 KB - mhealth_v12i1e49449_app1.docx]

References

1. Istepanian RSH, Alanzi T. Mobile health (m-health): evidence-based progress or scientific retrogression. In: Feng DD, editor. Biomedical Information Technology 2nd ed. Amsterdam, the Netherlands: Elsevier; 2022:717-734.
2. Istepanian RSH, Laxminarayan S, Pattichis CS. M-health: emerging mobile health systems. In: Mecheli-Tzanakou E, editor. Topics in Biomedical Engineering. New York City, NY: Springer; 2006.
3. Cao J, Lim Y, Sengoku S, Guo X, Kodama K. Exploring the shift in international trends in mobile health research from 2000 to 2020: bibliometric analysis. JMIR Mhealth Uhealth 2021;9(9):e31097 [FREE Full text] [doi: 10.2196/31097] [Medline: 34494968]
4. mHealth app economics 2017—current status and future trends in mobile health. Research 2 Guidance. 2017 Nov. URL: https://research2guidance.com/wp-content/uploads/2017/10/1-mHealth-Status-And-Trends-Reports.pdf [accessed 2024-08-28]

5. Price M, Yuen EK, Goetter EM, Herbert JD, Forman EM, Acierno R, et al. mHealth: a mechanism to deliver more accessible, more effective mental health care. Clin Psychol Psychother 2014;21(5):427-436 [FREE Full text] [doi: 10.1002/cpp.1855] [Medline: 23918764]
6. Beratarrechea A, Lee AG, Willner JM, Jahangir E, Ciapponi A, Rubinstein A. The impact of mobile health interventions on chronic disease outcomes in developing countries: a systematic review. Telemed J E Health 2014;20(1):75-82 [FREE Full text] [doi: 10.1089/tmj.2012.0328] [Medline: 24205809]
7. Chu JT, Wang MP, Shen C, Viswanath K, Lam TH, Chan SSC. How, when and why people seek health information online: qualitative study in Hong Kong. Interact J Med Res 2017;6(2):e24 [FREE Full text] [doi: 10.2196/ijmr.7000] [Medline: 29233802]
8. Grundy QH, Wang Z, Bero LA. Challenges in assessing mobile health app quality: a systematic review of prevalent and innovative methods. Am J Prev Med 2016;51(6):1051-1059. [doi: 10.1016/j.amepre.2016.07.009] [Medline: 27659122]
9. Nussbaum R, Kelly C, Quinby E, Mac A, Parmanto B, Dicianno BE. Systematic review of mobile health applications in rehabilitation. Arch Phys Med Rehabil 2019;100(1):115-127. [doi: 10.1016/j.apmr.2018.07.439] [Medline: 30171827]
10. Byambasuren O, Sanders S, Beller E, Glasziou P. Prescribable mHealth apps identified from an overview of systematic reviews. NPJ Digit Med 2018;1(1):1-12 [FREE Full text] [doi: 10.1038/s41746-018-0021-9] [Medline: 31304297]
11. Maramba I, Chatterjee A, Newman C. Methods of usability testing in the development of eHealth applications: a scoping review. Int J Med Inform 2019;126:95-104. [doi: 10.1016/j.ijmedinf.2019.03.018] [Medline: 31029270]
12. Fiedler J, Eckert T, Wunsch K, Woll A. Key facets to build up eHealth and mHealth interventions to enhance physical activity, sedentary behavior and nutrition in healthy subjects - an umbrella review. BMC Public Health 2020;20(1):1605 [FREE Full text] [doi: 10.1186/s12889-020-09700-7] [Medline: 33097013]
13. Zapata BC, Fernández-Alemán JL, Idri A, Toval A. Empirical studies on usability of mHealth apps: a systematic literature review. J Med Syst 2015;39(2):1. [doi: 10.1007/s10916-014-0182-2] [Medline: 25600193]
14. Harrison R, Flood D, Duce D. Usability of mobile applications: literature review and rationale for a new usability model. J Interact Sci 2013;1(1):1. [doi: 10.1186/2194-0827-1-1]
15. Weichbroth P. Usability of mobile applications: a systematic literature study. IEEE Access 2020;8:55563-55577. [doi: 10.1109/access.2020.2981892]
16. Albert B, Tullis T. Introduction. In: Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics. Amsterdam, the Netherlands: Elsevier; 2013:1-14.
17. Nielsen J. Defining usability. In: User Experience Re-mastered: Your Guide to Getting the Right Design. Amsterdam, the Netherlands: Elsevier; 2019:1-22.
18. Ergonomics of human-system interaction—part 11: usability: definitions and concepts. ISO. 2022. URL: https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-2:v1:en [accessed 2024-08-28]
19. Bevan N, Carter J, Harker S. ISO 9241-11 revised: what have we learnt about usability since 1998? In: Lecture Notes in Computer Science (Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics 9169). Berlin, Germany: Springer; 2015:143-151.
20. ISO/IEC 25010. ISO 25000 software and data quality. ISO. 2022. URL: https://iso25000.com/index.php/en/iso-25000-standards/iso-25010 [accessed 2024-08-28]
21. Nielsen J. Usability Engineering. San Francisco, CA: Morgan Kaufmann; 1994.
22. Alturki R, Gay V. Usability testing of fitness mobile application: methodology and quantitative results. 2017 Presented at: 7th International Conference on Computer Science, Engineering & Applications (ICCSEA 2017); September 23-24, 2017; Copenhagen, Denmark. [doi: 10.5121/csit.2017.71108]
23. Wildenbos GA, Peute L, Jaspers M. Aging barriers influencing mobile health usability for older adults: a literature based framework (MOLD-US). Int J Med Inform 2018;114:66-75. [doi: 10.1016/j.ijmedinf.2018.03.012] [Medline: 29673606]
24. National Health Service. Digital assessment questionnaire V2.1. NHS Digital. 2018. URL: https://developer-wp-uks.azurewebsites.net/wp-content/uploads/2018/09/Digital-Assessment-Questions-V2.1-Beta-PDF.pdf [accessed 2024-08-28]
25. Guidance on applying human factors and usability engineering of medical devices including drug-device combination products in Great Britain. MHRA. 2021. URL: https://assets.publishing.service.gov.uk/media/60521d98d3bf7f0455a6e61d/Human-Factors_Medical-Devices_v2.0.pdf [accessed 2024-09-24]
26. ORCHA. 2022. URL: https://orchahealth.com/about-us/unique-approach/ [accessed 2024-08-28]
27. Health navigator app library. New Zealand Ministry of Health. 2022. URL: https://healthify.nz/app-library [accessed 2024-08-28]
28. Health applications assessment guidance. New Zealand Ministry of Health. 2017. URL: https://www.tewhatuora.govt.nz/health-services-and-programmes/digital-health/other-digital-health-initiatives/health-applications-assessment-guidance/ [accessed 2024-08-28]
29. Moshi MR, Tooher R, Merlin T. Suitability of current evaluation frameworks for use in the health technology assessment of mobile medical applications: a systematic review. Int J Technol Assess Health Care 2018;34(5):464-475. [doi: 10.1017/S026646231800051X] [Medline: 30201060]

30. Bonten TN, Rauwerdink A, Wyatt JC, Kasteleyn MJ, Witkamp L, Riper H, et al. Online guide for electronic health evaluation approaches: systematic scoping review and concept mapping study. J Med Internet Res 2020;22(8):e17774 [FREE Full text] [doi: 10.2196/17774] [Medline: 32784173]
31. Aromataris E, Fernandez R, Godfrey CM, Holly C, Khalil H, Tungpunkom P. Summarizing systematic reviews: methodological development, conduct and reporting of an umbrella review approach. Int J Evid Based Healthc 2015;13(3):132-140. [doi: 10.1097/XEB.0000000000000055] [Medline: 26360830]
32. Pollock M, Fernandes RM, Pieper D, Tricco AC, Gates M, Gates A, et al. Preferred Reporting Items for Overviews of Reviews (PRIOR): a protocol for development of a reporting guideline for overviews of reviews of healthcare interventions. Syst Rev 2019;8(1):335 [FREE Full text] [doi: 10.1186/s13643-019-1252-9] [Medline: 31870434]
33. Pollock M, Fernandes RM, Becker LA, Pieper D, Hartling L. Chapter V: overviews of reviews. Cochrane Handbook for Systematic Reviews of Interventions. 2021. URL: https://training.cochrane.org/handbook/current/chapter-v#_Ref524711112 [accessed 2024-08-28]
34. Hunt H, Pollock A, Campbell P, Estcourt L, Brunton G. An introduction to overviews of reviews: planning a relevant research question and objective for an overview. Syst Rev 2018;7(1):39 [FREE Full text] [doi: 10.1186/s13643-018-0695-8] [Medline: 29490699]
35. Silva C, Zamboni A, Hernandes E, Di TA, Belgamo A, Fabbri S. State of the art through systematic review. LaPES. 2022. URL: https://www.lapes.ufscar.br/resources/tools-1/start-1 [accessed 2024-08-28]
36. Davis TL, DiClemente R, Prietula M. Taking mHealth forward: examining the core characteristics. JMIR Mhealth Uhealth 2016;4(3):e97 [FREE Full text] [doi: 10.2196/mhealth.5659] [Medline: 27511612]
37. Chandler J, Cumpston M, Thomas J, Higgins J, Deeks J, Clarke M. Chapter I: introduction. In: Higgins JPT, Green S, Chandler J, Cumpston M, Page M, Welch V, editors. Cochrane Handbook for Systematic Reviews of Interventions Version 6. Chichester, United Kingdom: Wiley-Blackwell; 2022.
38. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ 2009;339:b2700 [FREE Full text] [doi: 10.1136/bmj.b2700] [Medline: 19622552]
39. Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. BMJ 2017;358:j4008 [FREE Full text] [doi: 10.1136/bmj.j4008] [Medline: 28935701]
40. Gates M, Gates A, Guitard S, Pollock M, Hartling L. Guidance for overviews of reviews continues to accumulate, but important challenges remain: a scoping review. Syst Rev 2020;9(1):254 [FREE Full text] [doi: 10.1186/s13643-020-01509-0] [Medline: 33148319]
41. Bougioukas KI, Vounzoulaki E, Mantsiou CD, Savvides ED, Karakosta C, Diakonidis T, et al. Methods for depicting overlap in overviews of systematic reviews: an introduction to static tabular and graphical displays. J Clin Epidemiol 2021;132:34-45. [doi: 10.1016/j.jclinepi.2020.12.004] [Medline: 33309886]
42. Hennessy EA, Johnson BT. Examining overlap of included studies in meta-reviews: guidance for using the corrected covered area index. Res Synth Methods 2020;11(1):134-145 [FREE Full text] [doi: 10.1002/jrsm.1390] [Medline: 31823513]
43. Lewis JR. The system usability scale: past, present, and future. Int J Hum Comput Interact Taylor & Francis 2018;34(7):577-590. [doi: 10.1080/10447318.2018.1455307]
44. Georgsson M. A review of usability methods used in the evaluation of mobile health applications for diabetes. Stud Health Technol Inform 2020;273:228-233. [doi: 10.3233/SHTI200645] [Medline: 33087617]
45. Inal Y, Wake JD, Guribye F, Nordgreen T. Usability evaluations of mobile mental health technologies: systematic review. J Med Internet Res 2020;22(1):e15337 [FREE Full text] [doi: 10.2196/15337] [Medline: 31904579]
46. Ng MM, Firth J, Minen M, Torous J. User engagement in mental health apps: a review of measurement, reporting, and validity. Psychiatr Serv 2019;70(7):538-544 [FREE Full text] [doi: 10.1176/appi.ps.201800519] [Medline: 30914003]
47. Azad-Khaneghah P, Neubauer N, Miguel Cruz A, Liu L. Mobile health app usability and quality rating scales: a systematic review. Disabil Rehabil Assist Technol 2021;16(7):712-721. [doi: 10.1080/17483107.2019.1701103] [Medline: 31910687]
48. Vera F, Noël R, Taramasco C. Standards, processes and instruments for assessing usability of health mobile apps: a systematic literature review. Stud Health Technol Inform 2019;264:1797-1798. [doi: 10.3233/SHTI190653] [Medline: 31438349]
49. Saeed N, Manzoor M, Khosravi P. An exploration of usability issues in telecare monitoring systems and possible solutions: a systematic literature review. Disabil Rehabil Assist Technol 2020;15(3):271-281. [doi: 10.1080/17483107.2019.1578998] [Medline: 30794009]
50. Nouri R, Kalhori SRN, Ghazisaeedi M, Marchand G, Yasini M. Criteria for assessing the quality of mHealth apps: a systematic review. J Am Med Inform Assoc 2018;25(8):1089-1098 [FREE Full text] [doi: 10.1093/jamia/ocy050] [Medline: 29788283]
51. Muro-Culebras A, Escriche-Escuder A, Martin-Martin J, Roldán-Jiménez C, De-Torres I, Ruiz-Muñoz M, et al. Tools for evaluating the content, efficacy, and usability of mobile health apps according to the consensus-based standards for the selection of health measurement instruments: systematic review. JMIR Mhealth Uhealth 2021;9(12):e15433 [FREE Full text] [doi: 10.2196/15433] [Medline: 34855618]

52. Wakefield BJ, Turvey CL, Nazi KM, Holman JE, Hogan TP, Shimada SL, et al. Psychometric properties of patient-facing ehealth evaluation measures: systematic review and analysis. J Med Internet Res 2017;19(10):e346 [FREE Full text] [doi: 10.2196/jmir.7638] [Medline: 29021128]
53. Kien C, Schultes MT, Szelag M, Schoberberger R, Gartlehner G. German language questionnaires for assessing implementation constructs and outcomes of psychosocial and health-related interventions: a systematic review. Implement Sci 2018;13(1):150 [FREE Full text] [doi: 10.1186/s13012-018-0837-3] [Medline: 30541590]
54. Niknejad N, Ismail W, Bahari M, Nazari B. Understanding telerehabilitation technology to evaluate stakeholders' adoption of telerehabilitation services: a systematic literature review and directions for further research. Arch Phys Med Rehabil 2021;102(7):1390-1403. [doi: 10.1016/j.apmr.2020.12.014] [Medline: 33484693]
55. Canadian Institutes of Health Research (CIHR). Mental health apps: how to make an informed choice. Mental Health Commission of Canada (MHCC). 2016. URL: https://mentalhealthcommission.ca/resource/mental-health-apps-how-to-make-an-informed-choice-two-pager/ [accessed 2024-09-24]
56. Brooke J. SUS: a 'Quick and Dirty' usability scale. In: Usability Evaluation In Industry. Boca Raton, FL: CRC Press; 1996:207-212.
57. Lewis JR. IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. Int J Hum-Comput Interact 1995;7(1):57-78. [doi: 10.1080/10447319509526110]
58. Lund A. Measuring usability with the USE questionnaire. Usability Interface 2001;8(2):3-6 [FREE Full text]
59. van Singer M, Chatton A, Khazaal Y. Quality of smartphone apps related to panic disorder. Front Psychiatry 2015;6:96 [FREE Full text] [doi: 10.3389/fpsyt.2015.00096] [Medline: 26236242]
60. Lewis JR. An after-scenario questionnaire for usability studies: psychometric evaluation over three trials. ACM SIGCHI Bulletin 1991 Oct;23(4):79. [doi: 10.1145/126729.1056077]
61. McNiel P, McArthur EC. Evaluating health mobile apps: information literacy in undergraduate and graduate nursing courses. J Nurs Educ 2016;55(8):480. [doi: 10.3928/01484834-20160715-12] [Medline: 27459438]
62. Huis in 't Veld RMHA, Kosterink SM, Barbe T, Lindegård A, Marecek T, Vollenbroek-Hutten MMR. Relation between patient satisfaction, compliance and the clinical benefit of a teletreatment application for chronic pain. J Telemed Telecare 2010;16(6):322-328. [doi: 10.1258/jtt.2010.006006] [Medline: 20798426]
63. Baumel A, Faber K, Mathur N, Kane JM, Muench F. Enlight: a comprehensive quality and therapeutic potential evaluation tool for mobile and web-based ehealth interventions. J Med Internet Res 2017;19(3):e82 [FREE Full text] [doi: 10.2196/jmir.7270] [Medline: 28325712]
64. Yen P, Wantland D, Bakken S. Development of a customizable health IT usability evaluation scale. AMIA Annu Symp Proc 2010;2010:917-921 [FREE Full text] [Medline: 21347112]
65. Brown W, Yen P, Rojas M, Schnall R. Assessment of the health IT usability evaluation model (Health-ITUEM) for evaluating mobile health (mHealth) technology. J Biomed Inform 2013;46(6):1080-1087 [FREE Full text] [doi: 10.1016/j.jbi.2013.08.001] [Medline: 23973872]
66. Taki S, Campbell KJ, Russell CG, Elliott R, Laws R, Denney-Wilson E. Infant feeding websites and apps: a systematic assessment of quality and content. Interact J Med Res 2015;4(3):e18 [FREE Full text] [doi: 10.2196/ijmr.4323] [Medline: 26420339]
67. Gediga G, Hamborg KC, Düntsch I. The IsoMetrics usability inventory: An operationalization of ISO 9241-10 supporting summative and formative evaluation of software systems. Behaviour & Information Technology 1999;18(3):151-164. [doi: 10.1080/014492999119057]
68. Grau I, Kostov B, Gallego JA, Grajales Iii F, Fernández-Luque L, Sisó-Almirall A. Assessment method for mobile health applications in Spanish: the iSYScore index. Semergen 2016;42(8):575-583. [doi: 10.1016/j.semerg.2015.12.001] [Medline: 26879598]
69. Jin M, Kim J. Development and evaluation of an evaluation tool for healthcare smartphone applications. Telemed J E Health 2015;21(10):831-837. [doi: 10.1089/tmj.2014.0151] [Medline: 26431261]
70. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly 1989;13(3):319. [doi: 10.2307/249008]
71. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
72. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and validation of the user version of the mobile application rating scale (uMARS). JMIR Mhealth Uhealth 2016;4(2):e72 [FREE Full text] [doi: 10.2196/mhealth.5849] [Medline: 27287964]
73. Hart SG. Nasa-task load index (NASA-TLX); 20 years later. Proc Human Fact Ergon Soc Annu Meet 2006;50(9):904-908 [FREE Full text] [doi: 10.1177/154193120605000909]
74. McMillan B, Hickey E, Patel MG, Mitchell C. Quality assessment of a sample of mobile app-based health behavior change interventions using a tool based on the National Institute of Health and Care Excellence behavior change guidance. Patient Educ Couns 2016;99(3):429-435 [FREE Full text] [doi: 10.1016/j.pec.2015.10.023] [Medline: 26607787]

75. Price M, Sawyer T, Harris M, Skalka C. Usability evaluation of a mobile monitoring system to assess symptoms after a traumatic injury: a mixed-methods study. JMIR Ment Health 2016;3(1):e3 [FREE Full text] [doi: 10.2196/mental.5023] [Medline: 26753673]
76. Lewis JR. Psychometric evaluation of the post-study system usability questionnaire: the PSSUQ. Proc Hum Fact Soc Annu Meet 2016;36(16):1259-1260. [doi: 10.1177/154193129203601617]
77. Loy JS, Ali EE, Yap KY. Quality assessment of medical apps that target medication-related problems. J Manag Care Spec Pharm 2016;22(10):1124-1140 [FREE Full text] [doi: 10.18553/jmcp.2016.22.10.1124] [Medline: 27668561]
78. Martínez-Pérez B, de la Torre-Díez I, Candelas-Plasencia S, López-Coronado M. Development and evaluation of tools for measuring the quality of experience (QoE) in mHealth applications. J Med Syst 2013;37(5):9976. [doi: 10.1007/s10916-013-9976-x] [Medline: 24014255]
79. Chin JP, Diehl VA, Norman KL. Development of an instrument measuring user satisfaction of the human-computer interface. 1988 Presented at: Conference on Human Factors in Computing Systems—Proceedings Association for Computing Machinery; May 15-19, 1988; Washington, DC p. 213-218. [doi: 10.1145/57167.57203]
80. Silberg WM, Lundberg GD, Musacchio RA. Assessing, controlling, and assuring the quality of medical information on the Internet: caveant lector et viewor—let the reader and viewer beware. JAMA 1997;277(15):1244-1245. [doi: 10.1001/jama.1997.03540390074039]
81. Kirakowski J, Corbett M. SUMI: the software usability measurement inventory. Br J Educ Technol 1993;24(3):210-212. [doi: 10.1111/j.1467-8535.1993.tb00076.x]
82. Parmanto B, Pulantara IW, Schutte JL, Saptono A, McCue MP. An integrated telehealth system for remote administration of an adult autism assessment. Telemed J E Health 2013;19(2):88-94. [doi: 10.1089/tmj.2012.0104] [Medline: 23230821]
83. Bakken S, Grullon-Figueroa L, Izquierdo R, Lee N, Morin P, Palmas W, et al. Development, validation, and use of English and Spanish versions of the telemedicine satisfaction and usefulness questionnaire. J Am Med Inform Assoc 2006;13(6):660-667 [FREE Full text] [doi: 10.1197/jamia.M2146] [Medline: 16929036]
84. Zhou L, Bao J, Setiawan IMA, Saptono A, Parmanto B. The mHealth app usability questionnaire (MAUQ): development and validation study. JMIR Mhealth Uhealth 2019;7(4):e11500 [FREE Full text] [doi: 10.2196/11500] [Medline: 30973342]
85. Suso-Martí L, La Touche R, Herranz-Gómez A, Angulo-Díaz-Parreño S, Paris-Alemany A, Cuenca-Martínez F. Effectiveness of telerehabilitation in physical therapist practice: an umbrella and mapping review with meta-meta-analysis. Phys Ther 2021;101(5):pzab075 [FREE Full text] [doi: 10.1093/ptj/pzab075] [Medline: 33611598]
86. Marcolino MS, Oliveira JAQ, D'Agostino M, Ribeiro AL, Alkmim MBM, Novillo-Ortiz D. The impact of mHealth interventions: systematic review of systematic reviews. JMIR Mhealth Uhealth 2018;6(1):e23 [FREE Full text] [doi: 10.2196/mhealth.8873] [Medline: 29343463]
87. Mönninghoff A, Kramer JN, Hess AJ, Ismailova K, Teepe GW, Tudor Car L, et al. Long-term effectiveness of mHealth physical activity interventions: systematic review and meta-analysis of randomized controlled trials. J Med Internet Res 2021;23(4):e26699 [FREE Full text] [doi: 10.2196/26699] [Medline: 33811021]
88. Elavsky S, Knapova L, Klocek A, Smahel D. Mobile health interventions for physical activity, sedentary behavior, and sleep in adults aged 50 years and older: a systematic literature review. J Aging Phys Act 2019;27(4):565-593 [FREE Full text] [doi: 10.1123/japa.2017-0410] [Medline: 30507266]
89. Direito A, Carraça E, Rawstorn J, Whittaker R, Maddison R. mHealth technologies to influence physical activity and sedentary behaviors: behavior change techniques, systematic review and meta-analysis of randomized controlled trials. Ann Behav Med 2017;51(2):226-239. [doi: 10.1007/s12160-016-9846-0] [Medline: 27757789]
90. Koumpouros Y, Georgoulas A. Inform Health Soc Care 2020;45(2):168-187. [doi: 10.1080/17538157.2019.1656208] [Medline: 31743060]
91. Rabinowitz AR, Juengst SB. Introduction to topical Issue on mHealth for brain injury rehabilitation. J Head Trauma Rehabil 2022;37(3):131-133. [doi: 10.1097/HTR.0000000000000794] [Medline: 35703894]
92. Baumgartner J, Ruettgers N, Hasler A, Sonderegger A, Sauer J. Questionnaire experience and the hybrid system usability scale: using a novel concept to evaluate a new instrument. Int J Hum Comput Stud 2021;147:102575. [doi: 10.1016/j.ijhcs.2020.102575]
93. Ovčjak B, Heričko M, Polančič G. Factors impacting the acceptance of mobile data services—a systematic literature review. Comput Human Behav 2015;53:24-47. [doi: 10.1016/j.chb.2015.06.013]
94. Gao M, Kortum P, Oswald FL. Psychometric evaluation of the USE (Usefulness, Satisfaction, and Ease of use) questionnaire for reliability and validity. 2018 Presented at: Proceedings of the Human Factors and Ergonomics Society Annual Meeting; Human Factors and Ergonomics Society Annual Meeting; October 1-5, 2018; Philadelphia, PA p. 1414-1418. [doi: 10.1177/1541931218621322]
95. Peres SC, Pham T, Phillips R. Validation of the System Usability Scale (SUS): SUS in the Wild. Proc Hum Factors Ergon Soc Annu Meet 2013;57(1):192-196. [doi: 10.1177/1541931213571043]
96. McDowell I. The theoretical and technical foundations of health measurement. In: Measuring Health. New York, NY: Oxford University Press; 2006.

97. Demers M, Winstein CJ. A perspective on the use of ecological momentary assessment and intervention to promote stroke recovery and rehabilitation. Top Stroke Rehabil 2021;28(8):594-605. [doi: 10.1080/10749357.2020.1856557] [Medline: 33272137]
98. Boyd K, Bond R, Vertesi A, Dogan H, Magee J. How people judge the usability of a desktop graphic user interface at different time points: is there evidence for memory decay, recall bias or temporal bias? Interact Comput 2019;31(2):230. [doi: 10.1093/iwc/iwz019]
99. Alwashmi MF, Hawboldt J, Davis E, Fetters MD. The iterative convergent design for mobile health usability testing: mixed methods approach. JMIR Mhealth Uhealth 2019;7(4):e11656 [FREE Full text] [doi: 10.2196/11656] [Medline: 31025951]
100. Switzer GE, Wisniewski SR, Belle SH, Dew MA, Schultz R. Selecting, developing, and evaluating research instruments. Soc Psychiatry Psychiatr Epidemiol 1999;34(8):399-409. [doi: 10.1007/s001270050161] [Medline: 10501709]
101. Hughes DJ. Psychometric validity: establishing the accuracy and appropriateness of psychometric measures. In: Irwing P, Booth T, Hughes DJ, editors. The Wiley Handbook of Psychometric Testing: A Multidisciplinary Reference on Survey, Scale and Test Development. Chichester, United Kingdom: Wiley Blackwell; 2018:751-779.
102. Zein S, Salleh N, Grundy J. A systematic mapping study of mobile application testing techniques. J Syst Softw 2016;117:334-356. [doi: 10.1016/j.jss.2016.03.065]
103. Brown B, Reeves S, Sherwood S. Into the wild: challenges and opportunities for field trial methods. 2011 Presented at: CHI '11: the SIGCHI Conference on Human Factors in Computing Systems; May 7-12, 2011; Vancouver, BC p. 1657-1666. [doi: 10.1145/1978942.1979185]
104. Kjeldskov J, Stage J. New techniques for usability evaluation of mobile systems. Int J Hum-Comput Stud 2004;60(5-6):599-620. [doi: 10.1016/j.ijhcs.2003.11.001]
105. Rolleston AK, Bowen J, Hinze A, Korohina E, Matamua R. Collaboration in research: weaving Kaupapa Māori and computer science. AlterNative 2021;17(4):469-479. [doi: 10.1177/11771801211043164]
106. Stowell E, Lyson M, Saksono H, Wurth R, Jimison H, Pavel M, et al. Designing and evaluating mHealth interventions for vulnerable populations: a systematic review. 2018 Presented at: CHI '18: the 2018 CHI Conference on Human Factors in Computing Systems; April 21-26, 2018; Montreal, QC, Canada. [doi: 10.1145/3173574.3173589]
107. Bernhaupt R, Mihalic K, Obrist M. Usability evaluation methods for mobile applications. In: Handbook of Research on User Interface Design and Evaluation for Mobile Technology. Hershey, PA: IGI Global; 2011:745-758.
108. Lewis CC, Mettert KD, Dorsey CN, Martinez RG, Weiner BJ, Nolen E, et al. An updated protocol for a systematic review of implementation-related measures. Syst Rev 2018;7(1):66 [FREE Full text] [doi: 10.1186/s13643-018-0728-3] [Medline: 29695295]
109. Oyebode O, Alqahtani F, Orji R. Using machine learning and thematic analysis methods to evaluate mental health apps based on user reviews. IEEE Access 2020;8:111141-111158. [doi: 10.1109/access.2020.3002176]
110. Quintana Y, Torous J. A framework for evaluation of mobile apps for youth mental health. Homewood Research Institute. 2020. URL: https://hriresearch.com/publication/a-framework-for-the-evaluation-of-mobile-apps-for-youth-mental-health-research-report/ [accessed 2024-09-24]
111. Hosseiniravandi M, Kahlaee AH, Karim H, Ghamkhar L, Safdari R. Home-based telerehabilitation software systems for remote supervising: a systematic review. Int J Technol Assess Health Care 2020;36(2):113-125. [doi: 10.1017/S0266462320000021] [Medline: 32151291]

Abbreviations

AMSTAR 2: A Measurement Tool to Assess Systematic Reviews 2
mHealth: mobile health
PACMAD: People At the Centre of Mobile Application Development
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
StArt: State of the Art through Systematic Review
SUS: System Usability Scale
USE: Usefulness, Satisfaction, and Ease of Use

Edited by L Buis; submitted 29.05.23; peer-reviewed by T Davergne, K Harrington, S Hoppe-Ludwig, S Nataletti; comments to author 11.02.24; revised version received 04.05.24; accepted 30.07.24; published 04.10.24.

Please cite as:
Hach S, Alder G, Stavric V, Taylor D, Signal N
Usability Assessment Methods for Mobile Apps for Physical Rehabilitation: Umbrella Review
JMIR Mhealth Uhealth 2024;12:e49449
URL: https://mhealth.jmir.org/2024/1/e49449
doi:10.2196/49449
PMID:

Review

Sensors for Smoking Detection in Epidemiological Research: Scoping Review

Giuliana Favara[1], PhD; Martina Barchitta[1], PhD; Andrea Maugeri[1], PhD; Roberta Magnano San Lio[1], PhD; Antonella Agodi[1], PhD

Department of Medical and Surgical Sciences and Advanced Technologies "GF Ingrassia", University of Catania, Catania, Italy

Corresponding Author:
Antonella Agodi, PhD
Department of Medical and Surgical Sciences and Advanced Technologies "GF Ingrassia"
University of Catania
Via Santa Sofia 87
Catania, 95123
Italy
Phone: 39 0953782183
Email: agodia@unict.it

Abstract

Background: The use of wearable sensors is being explored as a challenging way to accurately identify smoking behaviors by measuring physiological and environmental factors in real-life settings. Although they hold potential benefits for aiding smoking cessation, no single wearable device currently achieves high accuracy in detecting smoking events. Furthermore, it is crucial to emphasize that this area of study is dynamic and requires ongoing updates.

Objective: This scoping review aims to map the scientific literature for identifying the main sensors developed or used for tobacco smoke detection, with a specific focus on wearable sensors, as well as describe their key features and categorize them by type.

Methods: According to the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) protocol, an electronic search was conducted on the PubMed, MEDLINE, and Web of Science databases, using the following keywords: ("biosensors" OR "biosensor" OR "sensors" OR "sensor" OR "wearable") AND ("smoking" OR "smoke").

Results: Among a total of 37 studies included in this scoping review published between 2012 and March 2024, 16 described sensors based on wearable bands, 15 described multisensory systems, and 6 described other strategies to detect tobacco smoke exposure. Included studies provided details about the design or application of wearable sensors based on an elastic band to detect different aspects of tobacco smoke exposure (eg, arm, wrist, and finger movements, and lighting events). Some studies proposed a system composed of different sensor modalities (eg, Personal Automatic Cigarette Tracker [PACT], PACT 2.0, and AutoSense).

Conclusions: Our scoping review has revealed both the obstacles and opportunities linked to wearable devices, offering valuable insights for future research initiatives. Tackling the recognized challenges and delving into potential avenues for enhancement could elevate wearable devices into even more effective tools for aiding smoking cessation. In this context, continuous research is essential to fine-tune and optimize these devices, guaranteeing their practicality and reliability in real-world applications.

(JMIR Mhealth Uhealth 2024;12:e52383) doi:10.2196/52383

KEYWORDS

smoking; tobacco smoke; smoke exposure; cigarette smoking; wearable sensor; public health

Introduction

According to the World Health Organization, tobacco smoke poses a major public health issue, causing approximately 8 million deaths annually worldwide [1]. It is a prominent contributor to noncommunicable diseases, such as cardiovascular diseases, chronic respiratory diseases, cancer, and diabetes, accounting for 1 out of every 6 deaths caused by these diseases [2,3]. In line with the sustainable development goals, reducing tobacco use is an essential requirement to achieve progress in the prevention and control of noncommunicable diseases, as well as for monitoring tobacco control efforts [4]. However,

smoking cessation is often hindered by low awareness of the health risks related to tobacco smoke [2]. In order to tackle this problem, there are multiple strategies available to aid individuals in their journey to quit smoking, and it is crucial to establish regular data collection on smoking habits and individual exposures [5,6].

A number of conventional approaches (eg, portable puff instruments, self-report questionnaires, and ecological momentary assessments) have been employed to monitor smoking habits in epidemiological research [2]. However, these methods have shown limited efficacy in accurately detecting smoke exposures, due to memory biases and underreporting by individuals. For these reasons, they do not offer valuable tools to support effective public health interventions [7]. More recently, wearable sensors have emerged as a potential approach for detecting smoking exposure in individuals. Broadly speaking, these sensors are usually intended to be worn on the body and measure various physiological or environmental parameters and behaviors of smoke exposure. It is intriguing that these devices comprise a blend of distinct sensor modalities (such as electrical, inertial, and acoustic) and a system of multiple sensors [8,9]. However, it is important to note that this field of research is still evolving, indicating that there is currently no single wearable device that exhibits high accuracy in detecting smoking events in all situations, isolating puffs and smoke inhalations, or evaluating smoke exposure.

In this context, the systematic review conducted by Imtiaz et al [2] provides a summary of recent innovative approaches (ie, individual and multisensor combinations, various body locations, and signal processing methodologies) of cutting-edge wearable sensors designed for monitoring cigarette smoking in real-world conditions, including studies published from 1990 to 2019. However, the following years have been marked by the COVID-19 pandemic, a globally impactful event that may have influenced the interest and potential applications of new technologies for monitoring lifestyles. Additionally, given that sensor use is a rapidly evolving field, it is necessary to consistently provide updates in the scientific literature.

Considering this, our scoping review was undertaken to offer an updated summary of studies published until March 2024, presenting the application of wearable sensors for monitoring cigarette smoking and smoke exposure, and focusing on both single-sensor and multisensor approaches.

Methods

The review methodology was employed by following these steps: (1) identifying the research question (RQ); (2) defining the search strategy and protocol; (3) conducting a literature search; and (4) collecting data from the included studies.

Thus, the RQ of our systematic review was as follows: what is the state of the art regarding the development and use of sensors (particularly wearable sensors) for monitoring smoking habits in epidemiological studies? Additionally, what are the main types of sensors and how do they differ in terms of modalities and potential applications?

This scoping review was conducted according to the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) protocol (Multimedia Appendix 1) [10]. Two authors (GF and AM) independently conducted a literature search and selected potentially relevant articles from inception to March 10, 2024, using the PubMed, MEDLINE, and Web of Science databases. The electronic search strategy consisted of the following keywords: (“biosensors” OR “biosensor” OR “sensors” OR “sensor” OR “wearable”) AND (“smoking” OR “smoke”).

Articles were included in the scoping review if they reported studies describing, developing, or applying sensors for the detection of tobacco smoke exposure, with the potential to be wearable. Thus, articles were included if they met the following criteria: (1) were written in English; (2) involved human participants; (3) described, developed, or applied sensors that could be wearable; (4) analyzed active or passive exposure to tobacco smoke; (5) were conducted in a laboratory or the community; and (6) described protocols according to the RQ of the scoping review. By contrast, articles were excluded if they met the following criteria: (1) did not fit into the RQ of the scoping review; (2) reported studies on sensors that were not or potentially not wearable; (3) did not analyze active or passive exposure to tobacco smoke; (4) presented only the biochemical or biomolecular applications of sensors; and (5) were letters, comments, abstracts, editorials, reviews, systematic reviews, or meta-analyses.

By using a data abstraction form designed for this scoping review, 2 authors (GF and AM) collected the following information from all the included studies: article characteristics (eg, first author, publication year, and country of origin), study characteristics (eg, study design), sensor specifications (eg, type of wearable sensors, parameters that were measured, accuracy, data availability, and autonomous or integrated system), and applications proposed.

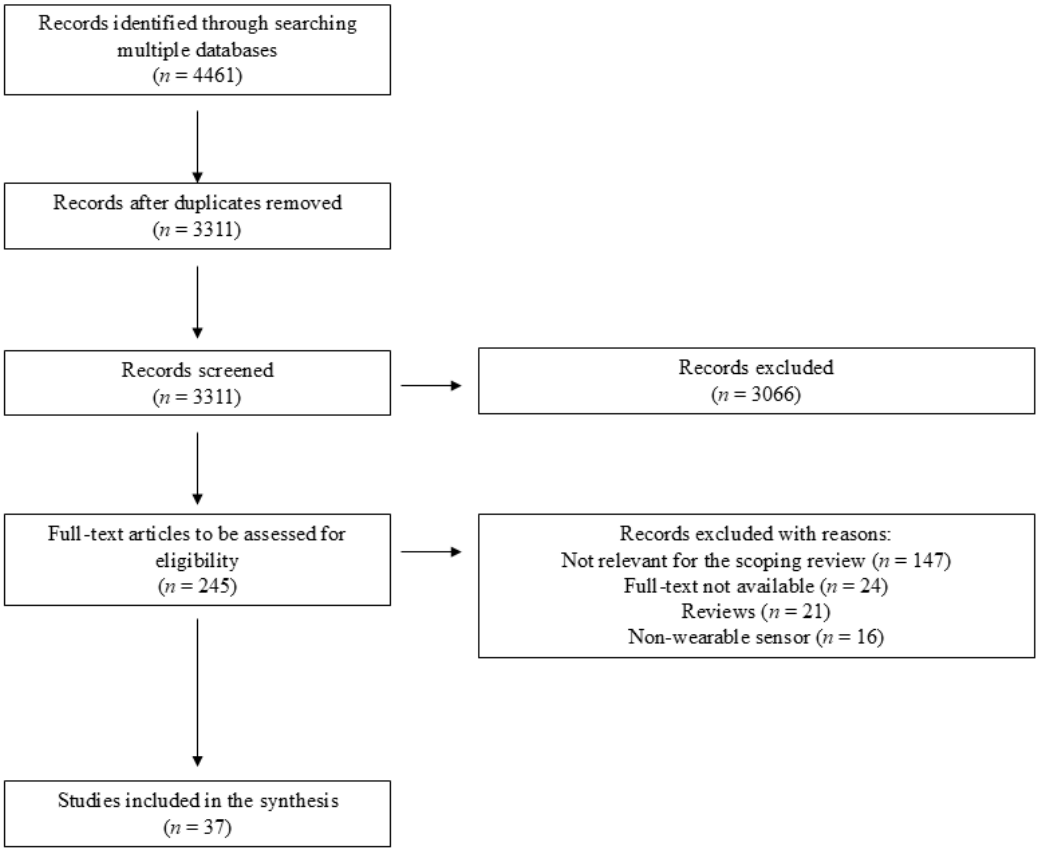
Diverse opinions between investigators were resolved through discussion between the 2 authors or by consulting a third author (AA). The studies were categorized based on the types of wearable devices employed, with particular attention given to distinguishing between single-device applications and combinations or systems of these devices.

Results

Study Selection

Following the removal of duplicates, a total of 3311 articles were initially identified through the literature search. Through the screening process of titles and abstracts, 3066 articles were excluded, leaving 245 full-text articles that were thoroughly evaluated for eligibility. Among these, 208 were excluded for the following reasons: 147 were not relevant for the objective of the current scoping review, 24 were without full text, 21 were reviews, and 16 used nonwearable sensors. Thus, the remaining 37 studies were considered eligible and were included in this scoping review (Figure 1).

Figure 1. Flowchart illustrating the selection of studies included in the scoping review.



Characteristics of the Studies Included

The vast majority of studies included in this scoping review originated from the United States (n=27). In particular, 16 studies described single wearable devices or devices in combination, 15 studies described multisensor systems, and 6 studies described other innovative strategies for the detection of cigarette smoke exposure. All of the studies included in this review were conducted within the general population, with a particular emphasis on individuals who smoke.

In the following sections, we grouped the studies included by the type of wearable sensor considered, as follows: (1) a single elastic band or a combination of elastic bands; (2) multisensor systems (ie, “Personal Automatic Cigarette Tracker [PACT],” “PACT 2.0,” and “AutoSense”); and (3) other alternative strategies.

Wearable Sensors Based on Elastic Bands

Sixteen studies focused on detailing the design or application of wearable sensors [8,11-14] or a combination thereof [15-23] to detect different aspects of tobacco smoke exposure (Table 1). While numerous studies commonly suggested sensors mounted in fixed locations, these approaches did not permit a comprehensive determination of tobacco smoke exposure among specific subgroups of individuals [24]. In this context, tobacco test strips are among the commercially available options for assessing nicotine levels in urine and saliva samples [25].

Table 1. Summary of studies exploring wearable sensors based on elastic bands.

Study	Country	Type of wearable	Application	Sensor modality	Accuracy	Data availability	Autonomous/integrated
Cole et al [26], 2021	United States	Smartwatch and smartphone app	To validate a smartwatch for the examination of smoking temporal patterns	Interpuff intervals and puff duration	—[a]	User and researcher	Integrated with a smartphone app
Dar [19], 2018	Israel	Smartwatch and wristbands	To investigate whether smoker monitoring through the SmokeBeat app would result in a reduction in smoking	Hand-to-mouth gestures	—	User and researcher	Integrated with a smartphone app
Horvath et al [20], 2021	United States	Smartband and smartphone	To propose a protocol for aiding daily smokers in their attempt to quit smoking	Hand-to-mouth gestures	—	User and researcher	Integrated with a smartphone app
Joyce et al [21], 2021	United States	Smartwatch or smartband	To assess the viability, acceptability, and efficacy of a smoking cessation program utilizing smartwatch technology and incentive-based strategies	Hand-to-mouth gestures	—	User and researcher	Integrated with a smartphone app
Lopez-Meyer et al [18], 2013	United States	Wristband	To describe a sensor for monitoring gestures in cigarette smokers	Hand-to-mouth gestures	Sensitivity: 90%	User and researcher	Data logger unit
Maguire et al [15], 2022	United States	Smartwatch and finger elastic band	To develop a smoking detection system to classify presmoking behaviors	Arm, wrist, and finger movements	Accuracy: 80.6%	User and researcher	Integrated with a smartphone app
Morriscey et al [22], 2018	United States	Smartwatch and smartphone app	To determine the sensitivity and specificity of the smartphone app	Hand-to-mouth gestures	Sensitivity: 22.5%-41.7%	User and researcher	Integrated with a smartphone app
Parate et al [12], 2014	United States	Wristband	To design a mobile solution to capture changes in the orientation of a person's arm	Hand-to-mouth gestures	Accuracy: 95.7%; precision: 91%; recall: 81%	User and researcher	Integrated with a smartphone app
Quintana et al [27], 2018	United States	Silicone wristband	To assess the effectiveness of silicone wristbands as personal nicotine samplers	Nicotine level	—	Researcher	Autonomous
Raiff et al [13], 2014	United States	Wrist, elbow, and shoulder bands	To develop and test the ability of inertial sensors to detect cigarette smoking	Arm movements	—	User and researcher	Integrated with a tablet
Schnall et al [23], 2022	United States	Smartwatch and smartphone app	To assess the feasibility and efficacy of the Lumme Quit Smoking mobile app	Hand-to-mouth gestures	—	User and researcher	Integrated with a smartphone app
Senyurek et al [14], 2019	United States	Chest band	To develop and test a machine learning model for smoking recognition	Breathing signals	Accuracy: 80.0%	—	—
Skinner et al [17], 2018	United Kingdom	Smartwatch	To present a system for passive detection of cigarette smoking	Hand-to-mouth gestures	—	—	Integrated with a smartphone
Tai et al [11], 2020	United States	Forearm band	To detect nicotine levels in the sweat of subjects inhaling cigarette smoke	Nicotine level	—	Researcher	Integrate with a computer
Takur et al [8], 2022	India	Wristband	To develop a modeling framework for smoking recognition	Hand-to-mouth gestures	Accuracy: 98.7%	User and researcher	Data logger unit integrated with a computer
Zhai et al [16], 2020	Belgium	A system that includes: an electric lighter, wristband, and chest patch	To reveal temporal patterns of smoking behaviors	Hearth rate, breathing, and hand-to-mouth gestures	—	User and researcher	Integrated with a smartphone app

[a]Not applicable.

Nonetheless, achieving a personalized and real-time assessment of individual exposure remains a challenge, emphasizing the need for further efforts to develop tailored preventive strategies accordingly. In an effort to address this concern, some authors presented a strategy to bridge the gap between fixed tobacco sensors and nicotine test strips. In particular, they proposed a noninvasive and wearable forearm band designed for continuous and real-time monitoring of nicotine levels in human sweat after nicotine inhalation. The device comprised a flexible electrode array connected to an electronic circuit, demonstrating favorable sensitivity and stability in both smokers and nonsmokers alike [11].

A recent study introduced a smoking cessation system aimed at predicting presmoking movements, such as reaching for a pack of cigarettes or lighting a cigarette, which are associated with smoking behaviors. This system utilizes an accelerometer embedded in a smartwatch to capture arm and wrist movements, along with a wearable finger sensor to measure the bending angle of the user's index finger. Their results showed that a model integrating data from both the smartwatch and finger sensor achieved higher accuracy in classifying presmoking activities compared to the model relying solely on the smartwatch. These findings lay the groundwork for developing an effective smoking cessation strategy that utilizes the combined input from these devices [15].

The authors who designed the "RisQ" mobile solution made additional endeavors to differentiate smoking episodes by validating a solution based on a wristband capable of detecting arm and wrist movements during smoking, as well as hand-to-mouth gestures, utilizing a machine learning model. Specifically, the authors compared the performance of 2 different algorithms in accurately identifying hand-to-mouth gestures, puffs, and smoking events. The authors proposed a pipeline aimed at effectively and promptly detecting smoking gestures in real time, yielding favorable accuracy, precision, and recall values [12]. In this context, a study conducted in 2020 assessed recurring patterns of smoking behaviors in real-life settings. The authors proposed a system comprising a wristband (capturing arm movements and skin impedance data), a chest patch (gathering electrocardiogram [ECG] information), and an electric lighter (detecting smoking events) to collect data regarding the contextual aspects in which individuals smoke. Notably, the study revealed intriguing differences in temporal patterns, encompassing weekly, daily, and time-of-day variations, as well as variations in emotional states during smoking episodes [16].

Some researchers introduced an innovative use of wearable devices for the real-time detection of smoking activity. They developed a wearable wristband that integrated 6-axial inertial sensors to gather data on physical activities, such as walking, running, walking upstairs, and walking downstairs, as well as smoking events. Additionally, they leveraged the collected data to create a machine learning model capable of distinguishing smoking activity from various daily activities, demonstrating promising performance [8]. Similarly, another study aimed to assess the capability of 4 inertial sensors (positioned on the wrist, elbow, and arm) to capture arm movements. Interestingly, the authors compared 2 distinct algorithms, namely support vector machine (SVM) and edge-detection–based learning. The results highlighted that the SVM model had superior performance in recognizing smoking events and interpuff intervals [13].

Similarly, a study published in 2019 compared the performance of various machine learning models with a focus on detecting smoke inhalations, utilizing a wearable chest band to capture breathing signals. The findings indicated that novel deep learning approaches may offer a more accurate method for detecting smoke inhalations than conventional machine learning models [14]. A validation study presented a system designed for the detection of passive cigarette smoking. The system consisted of a smartwatch equipped with a combination of accelerometers and gyroscope sensors, enabling the detection of hand movements associated with cigarette smoking [17]. Similarly, some authors validated a wristband that transmitted hand-to-mouth gestures to a receiver positioned on the user's chest. Interestingly, the proposed sensor exhibited good sensitivity and provided a methodology to differentiate hand-to-mouth gestures originating from smokers [18].

Although mobile apps for smoking cessation are becoming more widely accessible, their effectiveness is yet to be proven. In this context, a pilot trial being carried out in the United States is evaluating the potential of utilizing a smartband and smartphone for real-time monitoring, detecting smoking, and delivering concise mindfulness interventions to diminish smoking. The trial protocol seeks to provide insights into the practicability of employing the combined use of a smartband and smartphone [20]. An additional pilot study investigated SmokeBeat, an inventive app tailored for use with smartwatches and wristbands. This app processes information and utilizes embedded sensors in wearables to identify hand-to-mouth gestures in real time. The study's authors highlighted that the SmokeBeat algorithm accurately identified more than 80% of smoking episodes, and participants in the experimental group demonstrated a notable decrease in smoking rates throughout the 30-day trial [19]. SmokeBeat was also employed to assess the feasibility, acceptability, and effectiveness of a smoking cessation program designed for low-income pregnant smokers, utilizing smartwatch technology and incentive-based strategies. Reductions in smoking were noted in both the control and intervention groups across all pilot studies. While the utilization of the SmokeBeat program did not lead to a significant improvement in cessation rates, its feasibility and acceptability were deemed moderately high [21]. Similarly, another study demonstrated SmokeBeat's good sensitivity and specificity in cigarette detection following a learning period [22]. Remarkably, a pilot study indicated outstanding feasibility and acceptability in utilizing the "Lumme Quit Smoking" mobile app in conjunction with a smartwatch, aiming to enhance smoking cessation outcomes among individuals with HIV [23].

In this context, a pilot feasibility study was recently conducted to study temporal patterns and characteristics of smoking among adult smokers in a controlled laboratory setting. The authors compared the agreement of recording smoking characteristics by comparing the Polar M600 smartwatch that recorded accelerometer data to identify the durations of puff and interpuff intervals using Automated Smoking Perception and Recording

(ASPIRE) software and the pocket Clinical Research Support System (CReSS) topography device that uses video observation. The results suggested that the ASPIRE approach is more accurate than the CReSS method for passively monitoring smoking behavior. Moreover, the ASPIRE approach was more accurate than the CReSS method for measuring puff and interpuff intervals. In fact, the ASPIRE approach consistently produced a higher count of puffs and maintained more stable durations of interpuff intervals in comparison to the CReSS method, aligning both methods with the visually observed puff count. After filtering out implausible data from the CReSS method, both the ASPIRE approach and CReSS method provided consistent results for both puff duration and interpuff intervals [26]. A different study has proposed that uncomplicated silicone wristbands could serve as potential passive samplers for monitoring tobacco product exposure in children. Additionally, they have emerged as a noteworthy strategy for epidemiological and intervention studies. The researchers investigated the efficacy of 2 silicone wristbands for sampling personal nicotine levels in children, which were worn for durations of 7 days and 2 days, respectively. To achieve this, they compared the nicotine levels detected by the wristbands to urinary cotinine, a nicotine metabolite, measured in the urine of children exposed to contaminants in tobacco smoke or vapor from electronic cigarettes, as well as those living in nonsmoking households. The nicotine detected in the wristbands worn for 2 days exhibited a high correlation with urinary cotinine concentration, mirroring the correlation found in wristbands worn for 7 days. Moreover, the nicotine amounts recorded in the wristbands for both 2 and 7 days were significantly correlated [27].

Multisensor System: PACT

Five studies used the PACT to monitor tobacco smoke exposure (Table 2). The PACT is a comprehensive system composed of various sensor modalities [2,28-31]. Its primary objective is to track smoking episodes by detecting hand-to-mouth gestures that precede smoke inhalations. The PACT system comprises several components, including a chest module, a wrist or forearm band, an instrumented lighter, and a data logger. Specifically, the PACT system consists of the following elements. First, wearable respiratory inductance plethysmograph (RIP) sensors that are mounted in abdominal and thoracic bands. These sensors capture changes in breath volume, which are indicative of the expansion and contraction of the subject's lungs. Second, a radio frequency (RF) proximity sensor that is used to detect hand-to-mouth gestures. It consists of a transmitter positioned on the wrist and a receiver positioned on the chest. This sensor records the proximity of the hand to the mouth during smoking episodes. Third, a portable data logger that captures and stores the signals from the RIP sensors and RF proximity sensors. Additionally, the PACT system includes a self-report button that allows users to manually register each smoking event [2,32]. This feature enables users to input and record their smoking activities within the PACT system, providing an additional means of tracking and monitoring their smoking behavior [2,32].

Table 2. Studies exploring the Personal Automatic Cigarette Tracker system.

Study	Country	Type of wearable	Application	Sensor modality	Accuracy	Data availability	Autonomous/integrated
Lopez-Meyer et al [29], 2013	United States	Personal Automatic Cigarette Tracker (PACT) system that includes: thoracic, abdominal, and wrist bands; a chest receiver; a portable plethysmograph; and a data logger	To describe the development of a noninvasive wearable sensor system	Hand-to-mouth gestures and breathing	Precision: 87%; recall: 80%	Researcher	Autonomous data logger unit
Lopez-Meyer et al [30], 2012	United States	PACT system that includes: thoracic, abdominal, and wrist bands; a chest receiver; a portable plethysmograph; and a data logger	To apply a machine learning model for identifying cigarette smoke inhalations from wearable sensor data	Hand-to-mouth gestures and breathing	Precision: >87%; recall: >80%	Researcher	Autonomous data logger unit
Patil et al [28], 2013	United States	PACT system that includes: thoracic, abdominal, and wrist bands; a chest receiver; a portable plethysmograph; and a data logger	To apply a machine learning model for detecting cigarette smoke inhalations from respiratory signals	Hand-to-mouth gestures and breathing	Accuracy: 80%	Researcher	Autonomous data logger unit
Patil et al [32], 2014	United States	PACT system that includes: thoracic, abdominal, and wrist bands; a chest receiver; a portable plethysmograph; and a data logger	To analyze the factors affecting the output quality of the abdominal and thoracic bands	Hand-to-mouth gestures and breathing	F-score: 94%	Researcher	Autonomous data logger unit
Sazonov et al [31], 2013	United States	PACT system that includes: thoracic, abdominal, and wrist bands; an airflow sensor; a chest receiver; a portable plethysmograph; and a data logger	To describe the prototype of the sensor system and preliminary results of initial testing	Hand-to-mouth gestures, breathing, and oral and nasal airflow	—[a]	Researcher	Autonomous data logger unit

[a]Not applicable.

A validation study found that the accuracy of the classifier depends on the signals recorded, with abdominal breathing and hand gestures playing a significant role in detecting smoke inhalations. Additionally, the authors evaluated the impact of anthropometric measures on the quality of data captured by the PACT system. They suggested that the BMI and posture of individuals may influence the quality of smoking breathing signals [32]. Furthermore, another study demonstrated the feasibility of detecting smoke inhalation using the PACT system, indicating that each individual has unique characteristics in their response during smoking [28].

In the context of the PACT system, a study demonstrating the feasibility of automatically recognizing smoke inhalations was conducted. By applying the SVM algorithm, the authors found that breathing patterns exhibited individual characteristics. Subject-dependent models showed higher precision and recall values compared to subject-independent models [29]. In a further study, the same authors reported the findings of a subject-independent model for detecting smoke inhalations using data collected through the PACT system. In particular, the SVM model achieved high precision and recall values for detecting cigarette smoke inhalations [30].

A laboratory study proposed an intriguing application of the PACT system for reliable monitoring of smoking episodes in real-life settings. The findings revealed that smoking breathing patterns exhibit individual characteristics and display a strong correlation with hand-to-mouth gestures. As a result, the authors suggested that the PACT system can be employed to automatically assess daily smoking habits and evaluate the efficacy of behavioral and pharmacological interventions [31].

Multisensor System: PACT 2.0

Some studies (n=4) introduced the PACT system version 2.0, an enhanced iteration designed for automated real-time monitoring of smoking behavior (Table 3) [33-36]. Interestingly, the PACT 2.0 system eliminates the need for manual input from users as all smoking events are automatically detected and recorded by the sensors. In addition to capturing smoking data, the PACT 2.0 system collects supplementary information on smoking behaviors, including GPS location data and ECG data. It further offers users real-time feedback on their smoking behavior through a smartphone interface. This advanced system comprises a combination of components, including a chest module, a wrist or forearm module, an instrumented lighter, and a dedicated smartphone app.

Table 3. Studies exploring the Personal Automatic Cigarette Tracker 2.0 system.

Study	Country	Type of wearable	Application	Sensor modality	Accuracy	Data availability	Autonomous/integrated
Imtiaz et al [33], 2017	Switzerland	Personal Automatic Cigarette Tracker (PACT) 2.0 system that includes: a wristband, chest band, and smart lighter	To describe and validate a multisensory wearable system for monitoring cigarette smoking behavior	Hand-to-mouth gestures, breathing, lighting events, and heart rate	—[a]	Researcher	Autonomous data logger unit
Imtaiz et al [34], 2019	United States	PACT 2.0 system that includes: 2 cables and adhesive electrodes	To describe a novel method to identify smoking events	Heart rate, breathing, and hand-to-mouth gestures	Sensitivity: 87%; F-score: 79%	Researcher	Autonomous data logger unit
Senyurek et al [35], 2019	United States	PACT 2.0 system that includes: a wristband and smart lighter	To develop a robust sensor-based monitoring solution to detect smoking events	Hand-to-mouth gestures and lighting events	Accuracy: 84%; F-score: 91%	User and researcher	Integrated with a smartphone app
Senyurek et al [36], 2019	United States	PACT 2.0 system that includes: a wristband and smart lighter	To describe a novel method to identify smoking events	Hand-to-mouth gestures and lighting events	Accuracy: 84%; F-score: 91%	User and researcher	Integrated with a smartphone app

[a]Not applicable.

The chest module of the PACT 2.0 system captures various types of data, including breathing patterns recorded by inductive and bioimpedance respiratory sensors, cardiac activity measured by an ECG sensor, chest movement monitored by a 3-axis accelerometer, hand-to-mouth proximity detected by an RF receiver, and geospatial information obtained through a GPS receiver. On the other hand, the hand module records hand-to-mouth gestures using an inertial measurement unit (IMU) integrated with an RF transmitter. Additionally, the hand module serves as a pedometer, measuring the user's steps. The instrumented lighter within the PACT system is capable of detecting when the user is lighting a cigarette, as well as monitoring the smoking process and detecting when the cigarette is extinguished. To further enhance the user experience, the system is accompanied by a smartphone app that offers real-time feedback on smoking behavior. This includes information, such as the number of cigarettes smoked, the timing of each smoking event, and the associated location [2].

In a study conducted on a sample of 40 smokers in real-life settings, the PACT 2.0 system was developed and validated for monitoring cigarette smoking. The system demonstrated high acceptability and reliability, serving as an effective platform for detecting smoking behaviors [33]. In another investigation, the authors put forth changes in heart rate, along with breathing signals and body motion, as specific indicators of cigarette smoking, by employing the chest module of the PACT 2.0 system for this purpose. Utilizing an SVM model, they successfully developed an automated detection system for smoking events, achieving a high level of accuracy [34].

In another study, a combination of an instrumented lighter and a wrist IMU from the PACT 2.0 system was used in a group of smokers. The results indicated that integrating the IMU and instrumented lighter holds potential for studying smoking behavior in natural settings, resulting in higher accuracy values for the SVM classifier [35]. In a subsequent study, the temporal regularity of hand gestures was identified as a novel approach for detecting smoking events using the wrist IMU and PACT 2.0 lighter. Interestingly, this study revealed a high level of regularity in hand-to-mouth gestures during smoking episodes [36].

Multisensor System: AutoSense

In recent advancements, there have been proposals to leverage multiple wearable sensors in order to enhance accuracy, simplify signal detection, and combine various sensor modalities. Notably, 6 studies have put forth the utilization of AutoSense technology to monitor cigarette exposure and offer real-time feedback to individuals (Table 4) [37-42]. This technology employs a combination of wearable sensors that capture physiological signals associated with cardiovascular, respiratory, and thermoregulatory activities.

Table 4. Studies exploring the AutoSense system.

Study	Country	Type of wearable	Application	Sensor modality	Accuracy	Data availability	Autonomous/integrated
Battalio et al [39], 2021	United States	AutoSense system that includes: a chest band, wristbands, and electrocardiogram (ECG) electrodes	To investigate whether the delivery of a prompt could perform stress management and smoking behaviors	Hearth rate, breathing, and hand-to-mouth gestures	—[a]	User and researcher	Integrated with a smartphone
Chatterjee et al [41], 2016	United States	AutoSense system that includes: a chest band, wristbands, and ECG electrodes	To estimate cigarette craving during smoking abstinence	Hearth rate, breathing, and hand-to-mouth gestures	—	User and researcher	Integrated with a smartphone
Chatterjee et al [40], 2020	United States	AutoSense system that includes: a chest band, wristbands, and ECG electrodes	To automatically detect smoking "opportunity context"	Hearth rate, breathing, and hand-to-mouth gestures	—	User and researcher	Integrated with a smartphone
Hernandez et al [38], 2021	United States	AutoSense system that includes: a chest band, wristbands, and ECG electrodes	To deliver mindfulness-based strategies in real-time among individuals attempting to quit smoking	Hearth rate, breathing, and hand-to-mouth gestures	—	User and researcher	Integrated with a smartphone
Nakajima et al [37], 2020	United States	AutoSense system that includes: a chest band, wristbands, and ECG electrodes	To examine relationships between stress and smoking behavior and lapse among smokers motivated to quit smoking	Hearth rate, breathing, and hand-to-mouth gestures	Sensitivity: 80%	User	Integrated with a smartphone
Saleheen et al [42], 2015	United States	AutoSense system that includes: a chest band, wristbands, and ECG electrodes	To propose and evaluate a new model for detecting a smoking lapse	Hearth rate, breathing, and hand-to-mouth gestures	—	User and researcher	Integrated with a smartphone

[a]Not applicable.

AutoSense integrates various wearable sensors, including (1) a chest band equipped with a RIF sensor to capture respiration patterns and lung volume; (2) two chest ECG electrodes to measure electrical heart activity; and (3) inertial sensors placed on each wrist to detect movement patterns, hand-to-mouth gestures, and changes in body posture. All sensors continuously transmitted the collected data to a mobile phone for analysis and monitoring [2,42].

In a noteworthy study, the use of AutoSense technology was explored to identify stress states (cStress) and to predict first lapse smoking episodes (puffMarker) during a clinical study of smoking cessation. These findings suggested that heart rate and cStress could serve as useful predictors of smoking lapse [37]. Furthermore, this scoping review included a protocol that focused on recruiting motivated individuals who wanted to quit smoking. The protocol outlined a microrandomized controlled trial that would utilize the AutoSense system to deliver mindfulness strategies during the quit smoking attempt [38]. Another trial conducted on 75 smokers who expressed a desire to quit smoking and wore the AutoSense system, provided valuable insights for enhancing just-in-time stress management interventions aimed at preventing smoking relapse [39].

In this scenario, another notable study used data collected by the AutoSense system to identify common contexts of smoking "opportunity" (eg, smoking lapse, overeating or binge drinking, etc), which can either discourage or encourage individuals from engaging in adverse daily-life behaviors [40]. In a related study, the AutoSense system was employed to investigate the phenomenon of cigarette craving during smoking abstinence. Interestingly, individuals who reported high cravings experienced higher levels of stress during the hours of the day when the craving was elevated, compared to those with low cravings [41].

Moreover, a research group proposed the application of the AutoSense system in a population of abstinent smokers to detect the occurrence of a first lapse. By using an SVM classifier, the authors demonstrated the ability to detect the timing of the first lapse in smoking cessation within real-life settings [42].

Other Strategies

In our scoping review, we identified 6 studies that proposed alternative sensor modalities for the detection of smoking behavior (Table 5). One intriguing approach involves leveraging the distinct acoustic characteristics of smoking breaths, which differ from those of nonsmoking breaths. This opens up the potential for utilizing specialized sensors to gain novel insights into assessing smoking in real-life situations. To illustrate, a study explored the use of a wearable acoustic sensor designed to capture smoke-related sounds. The sensor, consisting of a microphone attached to the throat using an adhesive, was employed to record breath sounds. The authors further developed

an algorithm capable of automatically distinguishing between smoking and nonsmoking breaths, offering a promising advancement in smoking detection [43].

Table 5. Studies proposing alternative strategies for the detection of smoking behavior.

Study	Country	Type of wearable	Application	Sensor modality	Accuracy	Data availability	Autonomous/integrated
Cheng et al [44], 2019	China	Surface acoustic wave (SAW) sensor	To develop a sensor to absorb ambient tobacco markers	Gas	—[a]	Researcher	Autonomous
Echebarria et al [43], 2017	United Kingdom	Adhesive microphone	To monitor the acoustic properties of smoking breaths	Breathing acoustic signals	Sensitivity: >70%; specificity: >90%	Researcher	Autonomous
Gurtner et al [45], 2018	New Zealand	Wearable camera	To monitor children's exposure to smoking	Images	—	Researcher	Autonomous
Imtiaz et al [46], 2020	United States	Wearable camera–based sensor	To capture images of cigarette smoking episodes	Images	—	User and researcher	Integrated with a smartphone app
Qiao et al [47], 2019	China	Photoacoustic spectroscopy (PAS)-based carbon dioxide (CO_2) sensor	To demonstrate the applicability of a PAS sensor	Gas	—	—	Autonomous
Rahman et al [48], 2022	Australia	Chemoresistive sensor	To demonstrate the effectiveness of a sensor for nicotine detection	Gas	—	User and researcher	Integrated with a smart electronic device

[a]Not applicable.

With recent advancements in computer vision, there is an opportunity to explore novel approaches using images and video analysis for the detection of smoking episodes. This emerging field offers promising instruments to complement existing methods. Thus, a wearable camera for capturing various aspects of cigarette smoking behaviors (ie, smoking actions, smoking environment, and social interactions) was developed and tested. In this regard, the authors determined that positioning the camera on the eyeglass temple yielded the best results in terms of capturing images related to cigarette smoking while minimizing image blurriness across different body locations [46]. In a similar vein, some authors evaluated the extent and nature of the exposure of children living in households with a smoker, using image data. In this study, children wore wearable cameras around their necks for 4 days, which automatically took pictures every 7 seconds. The wearable camera proposed appears to have high utility for studying health behaviors in private spaces, including smoking episodes both in private spaces (ie, home and cars) and outdoors [45]. Interestingly, in a separate study, a surface acoustic wave (SAW) sensor specifically was designed to detect cigarette smoke in real time by adsorbing ambient tobacco markers [44]. Finally, another study demonstrated the effectiveness of a highly sensitive photoacoustic spectroscopy–based CO_2 trace gas sensor. This sensor holds the potential for detecting carbon dioxide (CO_2) levels resulting from cigarette smoking, among other real-world applications [47]. These advancements in sensor technologies provide promising avenues for further research in the detection and analysis of smoking behaviors.

Another study proposed a chemoresistive sensor for real-time monitoring of nicotine vapor from e-cigarettes in the air. In particular, the authors proposed a vanadium dioxide (VO_2)-based nicotine sensor integrated with an epidermal near-field communication (NFC) interface that enables battery-free operation and data transmission to smart electronic devices to record and store sensor data to detect nicotine at ambient concentrations [48].

Discussion

Principal Results and Comparison With Prior Work

The real-time monitoring of smoking behavior presents a significant challenge for public health research and the development of effective smoking-cessation interventions. However, advancements in mobile health (mHealth) technologies offer promising opportunities to address this challenge [13,16]. Through our scoping review, we have identified 37 studies published between 2012 and 2024 (March 10) that explore the use of wearable devices for assessing tobacco smoke exposure. These studies highlight the ongoing evolution of research in this field. Currently, research in the field predominantly focuses on the utilization of single or combined sensors to detect behaviors associated with smoking episodes. The inclusion of wearable sensors, particularly those integrated into elastic bands, has emerged as a noninvasive and mobile approach that enables continuous monitoring of smoking exposures. These advancements in sensor technology offer valuable insights for the development of targeted preventive measures and policies. In general, traditional smoking cessation programs, whether delivered in-person or via a smartphone, have proven efficacy but are constrained in their reach and utilization. In contrast, digital solutions exhibit greater potential for widespread accessibility, and there is evidence suggesting

their efficacy. However, only a limited number of scientifically validated apps have been developed for commercial purposes, limiting their full potential for reach. Moreover, most digital solutions for smoking cessation have predominantly relied on a single form of technology, such as text messaging or an app. Mobile apps hold great promise in supporting patients in health care and promoting healthy behavioral changes. However, the success of these apps is largely determined by their features, influencing patients' attitudes toward their use. In this context, the varied range of wearable sensors and studies highlighted in our scoping review have the potential to provide individuals with valuable approaches that fulfill both educational and motivational purposes. Furthermore, these wearable sensors can be smoothly integrated into evidence-based smoking cessation initiatives, thereby improving the overall effectiveness of such programs and encouraging a heightened interest in participating in cessation programs [49,50].

The studies included in our analysis demonstrate the efficacy of wearable sensors in monitoring smoking exposure. These sensors, deployed in various wearable bands, facilitate the detection of hand-to-mouth gestures, arm movements, and smoke inhalation during smoking episodes [12-14]. These sensors have also been employed to classify presmoking episodes [15] and evaluate temporal and emotional patterns associated with smoking behaviors [16]. Of particular interest is the use of a wearable forearm band, which allows for the monitoring of nicotine levels in sweat. This novel approach addresses the limitations of existing smoke monitoring devices and provides the potential capability to assess tobacco smoke exposures [11].

Overall, our findings highlight the ongoing challenges in achieving a personalized assessment of tobacco smoke for individuals in real-life settings. This suggests the need for further efforts to implement the proposed applications and enable a more precise evaluation of tobacco smoke exposure. It is crucial to consider personal factors, such as gender, age, and BMI, as well as contextual information including location, activity, and social context, in order to deeply characterize smoking patterns.

The focus of research in this field is shifting toward the utilization of complex systems comprising various types of sensors. For example, the integration of accelerometers and gyroscope sensors in a wearable smartwatch, along with wrist and chest bands, has demonstrated potential in detecting both active and passive tobacco smoking [17,18]. Some studies have explored noninvasive wearable devices that continuously collect real-time data on multiple physiological and environmental parameters, such as breathing patterns, chest movement, hand-to-mouth gestures, and lighting events. Among these innovative systems, the PACT system stands out, incorporating a comprehensive array of sensor modalities. This includes a chest module, a wrist or forearm band, an instrumented lighter, and a data logger. The findings from these studies have indicated that anthropometric measures of individuals can impact the signals collected by the PACT system. Additionally, they have suggested that breathing patterns show individual characteristics and are closely associated with hand-to-mouth gestures [28,29,31,32].

Several studies have introduced an advanced version of the PACT system, known as PACT 2.0, which offers several advantages and additional features. This upgraded system collects additional information, including GPS data, bioimpedance data, and heart activity, providing users with real-time feedback through a mobile app. Remarkably, the PACT 2.0 system has undergone development and validation to ensure its effectiveness in monitoring tobacco smoking in real-life situations, achieving a high level of acceptability [33]. Furthermore, it has been found that heart rate parameters can serve as an additional indicator of cigarette smoking, adding to the system's capabilities [34]. The implementation of the PACT 2.0 system has yielded promising results among smokers, presenting opportunities for various applications in the field of public health research [35]. Notably, data collected through the PACT 2.0 system have revealed that hand-to-mouth gestures exhibit a high degree of regularity during smoking events. This finding suggests that the system can serve as a valuable tool for accurately identifying smoking episodes amidst various daily activities [36].

One notable advantage of these approaches is their ability to monitor without relying on self-reporting by smokers. This eliminates potential biases caused by underreporting. Moreover, the use of the PACT system only requires the cooperation of subjects in wearing the system, offering the potential to capture smoke exposure parameters that are not accessible through other methods. In addition, the information recorded by the PACT system can enhance our understanding of the health consequences associated with smoke exposure by evaluating the relation between biomarkers (eg, blood levels of CO_2 and cotinine) and individual smoking behaviors. The PACT system also provides a unique advantage in its ability to capture comprehensive data on the complete breathing cycle during smoking, including puff, smoke inhalation, smoking apnea, and smoke exhalation. This is in contrast to traditional puff topography devices that only measure and analyze the air drawn through a cigarette during a puff. This comprehensive data can offer valuable insights into smoking behavior and its effects on health outcomes.

However, it is important to acknowledge some limitations of these systems. One such limitation is the relatively large size of the sensors, which are typically mounted on garments or vests. To overcome this limitation, there is a need to significantly miniaturize the sensors and integrate them into a fully wearable device, ensuring greater comfort and convenience for users. Overall, while these systems show promise in improving the monitoring of smoke exposure, it is essential to address the limitations by developing more compact and integrated sensor solutions for enhanced usability and effectiveness.

In this context, the AutoSense system emerges as a valuable combination that can capture a range of information, including heart rate, respiration, skin conductance, and physical activity. Its versatility opens up promising applications in various settings such as health care, workplaces, and communities. AutoSense not only enables the collection of physiological measurements but also offers particular potential in assessing stress response.

The real-time transmission of data to a smartphone allows for the monitoring of physiological responses to real-life stressors and the continuous estimation of stress levels. This capability positions AutoSense as a valuable tool for evaluating behaviors associated with stress, such as drinking, smoking, physical activity, movement patterns, and conversations. Furthermore, it provides insights into physical, behavioral, and mental health conditions. Looking ahead, ongoing research in this field holds the promise of developing effective prevention and intervention strategies. The ability to deliver these strategies directly on smartphones aligns with the vision of mHealth, empowering individuals to proactively manage their health and well-being. By harnessing the potential of AutoSense, we can anticipate advancements that contribute to improved health outcomes and enhanced quality of life.

Some authors thoroughly discussed user behavior, comfort levels, and compliance with the use of wearable sensors in real-world settings [8,13,33,37,46]. To address this, brief acceptability questionnaires were administered to gauge users' personal experiences with the wearable sensors. For instance, the results from these assessments indicated a strong inclination among participants to continue wearing the wearable camera for any subsequent multiday experiments [46]. This positive sentiment also encompassed other wearable accessories, such as elastic bands [13], and the PACT 2.0 system [33], highlighting that these devices were not only comfortable but also considered acceptable for prolonged use beyond the laboratory setting. Participants expressed contentment with the ease of wear for these technologies over extended periods.

As denoted in our work, there has been significant interest in sensors as a potential solution to address challenges related to the detection of tobacco smoke exposures. However, it is essential to highlight that no single wearable method has demonstrated 100% accuracy in detecting smoking events under all circumstances, with certain technologies that may be effective in specific environments and others that may not yield satisfactory results in the same context. To address this issue, it is evident that no single device currently offers a complete and accurate solution. Moreover, sensor responses are often influenced by ambient factors, such as motion and clothing. Although we described various single-sensor and multisensor approaches, our work denoted a lack of comparative analyses of their respective efficacies. To date, indeed, there has been no comprehensive survey or comparison study conducted on these approaches, elucidating the strengths and limitations of sensing technologies and their applicability in real-world settings. Furthermore, there has been limited evaluation of the underlying detection algorithms and their comparative accuracy.

In our research, only the study conducted by Imtiaz et al [46] addressed specific ethical considerations related to wearable camera research. They recommended that researchers involved in such studies actively protect the rights, privacy, dignity, and well-being of the individuals being studied. Emphasizing the importance of voluntary informed consent, research efforts should strive to ensure that participants willingly agree to take part. Additionally, research participants should be informed about the level of anonymity and confidentiality guaranteed during the publication and dissemination of findings, along with the potential for data reuse. To adhere to these ethical principles, participants throughout the study were instructed to remove the sensor system in situations where privacy was expected, such as restrooms, during activities causing discomfort (like sports and water-related events), and upon requests by individuals in their vicinity [46].

In general, wearable sensors described in our work are commonly characterized by their lightweight, mobile, and convenient design. However, as of today, there is a notable gap in the literature regarding a comprehensive analysis of their cost-effectiveness. While the advantages of wearables, such as ease of use and adaptability to various settings, are frequently acknowledged, there remains a critical need for research that delves into the economic aspects of implementing these technologies. Understanding the cost-effectiveness of wearable sensors is crucial for widespread adoption and integration into various domains, including health care and environmental monitoring.

Future studies could explore not only the initial investment and production costs but also the long-term economic benefits, considering factors such as durability, maintenance, and the potential impact on health outcomes or performance improvements. A thorough cost-effectiveness analysis would provide valuable insights for decision-makers, researchers, and industries looking to leverage wearable sensor technologies while considering the economic implications of their adoption.

Limitations

Our findings indicate that wearable devices for detecting cigarette smoking and assessing smoke exposure are still in the early stage of development and require further advancements. Therefore, it is crucial to acknowledge certain limitations when considering the implications of our review. First, the included studies evaluated various behavioral and physiological characteristics, such as hand gestures, breathing, and lighting, using wearable sensors during smoking episodes. However, there is a need for future research to explore the integration of wearable chemical sensors for detecting smoking and measuring exposure to both traditional and electronic cigarette smoke.

Second, the included studies did not extensively explore how external environmental factors might impact the accuracy and reliability of these wearable sensors. A thorough investigation into how factors, such as ambient smoke, weather conditions, and other pollutants, could influence sensor readings would provide valuable insights. Third, it is important to note that our review primarily focused on wearable device development and validation, involving a limited number of participants, which included both smokers and nonsmokers. Fourth, the considerable heterogeneity between the studies precluded the possibility of conducting a meta-analytical approach, thereby hindering the ability to synthesize data, as well as to offer a quantitative perspective on the efficacy or cost-effectiveness of the wearable sensors explored. On the one hand, this represents a limitation of existing studies, but on the other hand, it represents an important perspective for future research.

Finally, the potential for biases in the selection of studies, which could impact the representativeness and generalizability of the

obtained results, needs to be considered. Addressing these biases could enhance the overall robustness and reliability of our scoping review's conclusions.

Conclusions

Our scoping review highlights existing evidence about single-sensor or multisensor wearables in the context of tobacco smoke detection. While their potential is evident, further advancements and investigations are necessary to deeply evaluate their full potential and assist both individuals and health care professionals in addressing smoking-related health issues, as well as designing effective public health strategies.

Overall, wearable sensors may hold great promise in assisting individuals in their journey to quit smoking by offering real-time feedback on their exposures and smoking behaviors, empowering them to make informed decisions and take control of their habits. Through our scoping review, we have identified both the challenges and possibilities associated with wearable devices, which can serve as valuable guidance for future research endeavors. By addressing the identified pitfalls and exploring the potential avenues of improvement, wearable devices may become even more valuable tools in supporting smoking cessation efforts. In this scenario, ongoing research is crucial to refine and optimize these devices, ensuring their practicality and reliability in real-world settings.

Indeed, further research is needed to enhance the accuracy and usability of these devices, enabling researchers and health care professionals to gain a better understanding of the health implications related to smoking and to develop effective preventive interventions. In addition, future studies should examine factors such as comfort, adherence, and acceptability of wearable sensors. In fact, while many wearable sensors have been validated and tested in controlled experimental settings, there is a scarcity of research on their accuracy and applicability in real-life contexts.

Acknowledgments

This work was partially funded by the Department of Medical and Surgical Science and Advanced Technologies "GF Ingrassia," University of Catania, Italy. The works of MB, RMSL, and AA have been supported by the Ministry of University and Research (MUR) in the framework of the National Recovery and Resilience Plan (PNRR) Mission 4, Component 2, Investment 1.5 under project SAMOTHRACE (Sicilian Micro and Nano Technology Research and Innovation Center).

Authors' Contributions

Conceptualization and design – GF, MB, AM, AA
Data collection – GF, AM
Data analysis – GF, AM, RMSL
Interpretation – GF, MB, AM, RMSL, AA
Writing – GF, MB, AM, RMSL, AA

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[PDF File (Adobe PDF File), 134 KB - mhealth_v12i1e52383_app1.pdf]

References

1. Tobacco. WHO. URL: https://www.who.int/news-room/fact-sheets/detail/tobacco [accessed 2024-10-14]
2. Imtiaz MH, Ramos-Garcia RI, Wattal S, Tiffany S, Sazonov E. Wearable sensors for monitoring of cigarette smoking in free-living: A systematic review. Sensors (Basel) 2019 Oct 28;19(21):4678 [FREE Full text] [doi: 10.3390/s19214678] [Medline: 31661856]
3. International Statistical Classification of Diseases and Related Health Problems (ICD). WHO. URL: https://www.who.int/classifications/classification-of-diseases [accessed 2024-10-14]
4. WHO global report on trends in prevalence of tobacco use 2000-2025, fourth edition. WHO. URL: https://www.who.int/publications/i/item/9789240039322 [accessed 2024-10-14]
5. Hartmann-Boyce J, Lancaster T, Stead LF. Print-based self-help interventions for smoking cessation. Cochrane Database Syst Rev 2014 Jun 03(6):CD001118. [doi: 10.1002/14651858.CD001118.pub3] [Medline: 24888233]
6. Michie S, Brown J, Geraghty AWA, Miller S, Yardley L, Gardner B, et al. Development of StopAdvisor: A theory-based interactive internet-based smoking cessation intervention. Transl Behav Med 2012 Sep;2(3):263-275. [doi: 10.1007/s13142-012-0135-6] [Medline: 24073123]
7. Shiffman S, Gwaltney CJ, Balabanis MH, Liu KS, Paty JA, Kassel JD, et al. Immediate antecedents of cigarette smoking: an analysis from ecological momentary assessment. J Abnorm Psychol 2002 Nov;111(4):531-545. [doi: 10.1037//0021-843x.111.4.531] [Medline: 12428767]

8. Thakur SS, Poddar P, Roy RB. Real-time prediction of smoking activity using machine learning based multi-class classification model. Multimed Tools Appl 2022;81(10):14529-14551 [FREE Full text] [doi: 10.1007/s11042-022-12349-6] [Medline: 35233178]
9. Naresh V, Lee N. A review on biosensors and recent development of nanostructured materials-enabled biosensors. Sensors (Basel) 2021 Feb 05;21(4):1109 [FREE Full text] [doi: 10.3390/s21041109] [Medline: 33562639]
10. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
11. Tai L, Ahn CH, Nyein HYY, Ji W, Bariya M, Lin Y, et al. Nicotine monitoring with a wearable sweat band. ACS Sens 2020 Jun 26;5(6):1831-1837. [doi: 10.1021/acssensors.0c00791] [Medline: 32429661]
12. Parate A, Chiu M, Chadowitz C, Ganesan D, Kalogerakis E. RisQ: Recognizing smoking gestures with inertial sensors on a wristband. MobiSys 2014 Jun;2014:149-161 [FREE Full text] [doi: 10.1145/2594368.2594379] [Medline: 26688835]
13. Raiff BR, Karataş Ç, McClure EA, Pompili D, Walls TA. Laboratory validation of inertial body sensors to detect cigarette smoking arm movements. Electronics (Basel) 2014 Feb 27;3(1):87-110 [FREE Full text] [doi: 10.3390/electronics3010087] [Medline: 25553255]
14. Senyurek VY, Imtiaz MH, Belsare P, Tiffany S, Sazonov E. A comparison of SVM and CNN-LSTM based approach for detecting smoke inhalations from respiratory signal. Annu Int Conf IEEE Eng Med Biol Soc 2019 Jul;2019:3262-3265. [doi: 10.1109/EMBC.2019.8856395] [Medline: 31946581]
15. Maguire G, Chen H, Schnall R, Xu W, Huang M. Smoking cessation system for preemptive smoking detection. IEEE Internet Things J 2022 Mar;9(5):3204-3214 [FREE Full text] [doi: 10.1109/jiot.2021.3097728] [Medline: 36059439]
16. Zhai D, Simoes-Capela N, Schiavone G, Raedt WD, Van Hoof C. Reveal temporal patterns of smoking behavior in real life using data acquired through automatic tracking systems. Annu Int Conf IEEE Eng Med Biol Soc 2020 Jul;2020:6005-6008. [doi: 10.1109/EMBC44109.2020.9175363] [Medline: 33019340]
17. Skinner AL, Stone CJ, Doughty H, Munafò M. StopWatch: The preliminary evaluation of a smartwatch-based system for passive detection of cigarette smoking. Nicotine Tob Res 2019 Jan 04;21(2):257-261 [FREE Full text] [doi: 10.1093/ntr/nty008] [Medline: 29373720]
18. Lopez-Meyer P, Patil Y, Tiffany T, Sazonov E. Detection of hand-to-mouth gestures using a RF operated proximity sensor for monitoring cigarette smoking. Open Biomed Eng J 2013;9:41-49 [FREE Full text] [doi: 10.2174/1874120701307010041] [Medline: 23723954]
19. Dar R. Effect of real-time monitoring and notification of smoking episodes on smoking reduction: A pilot study of a novel smoking cessation app. Nicotine Tob Res 2018 Nov 15;20(12):1515-1518. [doi: 10.1093/ntr/ntx223] [Medline: 29126209]
20. Horvath M, Grutman A, O'Malley SS, Gueorguieva R, Khan N, Brewer JA, et al. Smartband-based automatic smoking detection and real-time mindfulness intervention: Protocol for a feasibility trial. JMIR Res Protoc 2021 Nov 16;10(11):e32521 [FREE Full text] [doi: 10.2196/32521] [Medline: 34783663]
21. Joyce CM, Saulsgiver K, Mohanty S, Bachireddy C, Molfetta C, Steffy M, et al. Remote patient monitoring and incentives to support smoking cessation among pregnant and postpartum Medicaid members: Three randomized controlled pilot studies. JMIR Form Res 2021 Sep 30;5(9):e27801 [FREE Full text] [doi: 10.2196/27801] [Medline: 34591023]
22. Morriscey C, Shephard A, van Houdt A, Kerr D, Barrett SP. Using 'smart' technology to aid in cigarette smoking cessation: Examining an innovative way to monitor and improve quit attempt outcomes. J. Smok Cessat 2018 Sep 25;14(3):149-154. [doi: 10.1017/jsc.2018.33]
23. Schnall R, Liu J, Alvarez G, Porras T, Ganzhorn S, Boerner S, et al. A smoking cessation mobile app for persons living with HIV: Preliminary efficacy and feasibility study. JMIR Form Res 2022 Aug 18;6(8):e28626 [FREE Full text] [doi: 10.2196/28626] [Medline: 35980739]
24. Liu Y, Antwi-Boampong S, BelBruno JJ, Crane MA, Tanski SE. Detection of secondhand cigarette smoke via nicotine using conductive polymer films. Nicotine Tob Res 2013 Sep;15(9):1511-1518 [FREE Full text] [doi: 10.1093/ntr/ntt007] [Medline: 23482719]
25. Achilihu H, Feng J, Wang L, Bernert JT. Tobacco use classification by inexpensive urinary cotinine immunoassay test strips. J Anal Toxicol 2019 Mar 01;43(2):149-153 [FREE Full text] [doi: 10.1093/jat/bky075] [Medline: 30395272]
26. Cole CA, Powers S, Tomko RL, Froeliger B, Valafar H. Quantification of smoking characteristics using smartwatch technology: Pilot feasibility study of new technology. JMIR Form Res 2021 Feb 05;5(2):e20464 [FREE Full text] [doi: 10.2196/20464] [Medline: 33544083]
27. Quintana PJE, Hoh E, Dodder NG, Matt GE, Zakarian JM, Anderson KA, et al. Nicotine levels in silicone wristband samplers worn by children exposed to secondhand smoke and electronic cigarette vapor are highly correlated with child's urinary cotinine. J Expo Sci Environ Epidemiol 2019 Oct;29(6):733-741. [doi: 10.1038/s41370-019-0116-7] [Medline: 30728487]
28. Patil Y, Lopez-Meyer P, Tiffany S, Sazonov E. Detection of cigarette smoke inhalations from respiratory signals using reduced feature set. Annu Int Conf IEEE Eng Med Biol Soc 2013;2013:6031-6034. [doi: 10.1109/EMBC.2013.6610927] [Medline: 24111114]

29. Lopez-Meyer P, Tiffany S, Patil Y, Sazonov E. Monitoring of cigarette smoking using wearable sensors and support vector machines. IEEE Trans Biomed Eng 2013 Jul;60(7):1867-1872. [doi: 10.1109/TBME.2013.2243729] [Medline: 23372073]
30. Lopez-Meyer P, Tiffany S, Sazonov E. Identification of cigarette smoke inhalations from wearable sensor data using a Support Vector Machine classifier. Annu Int Conf IEEE Eng Med Biol Soc 2012;2012:4050-4053. [doi: 10.1109/EMBC.2012.6346856] [Medline: 23366817]
31. Sazonov E, Lopez-Meyer P, Tiffany S. A wearable sensor system for monitoring cigarette smoking. J Stud Alcohol Drugs 2013 Nov;74(6):956-964 [FREE Full text] [doi: 10.15288/jsad.2013.74.956] [Medline: 24172124]
32. Patil Y, Tiffany S, Sazonov E. Understanding smoking behavior using wearable sensors: relative importance of various sensor modalities. Annu Int Conf IEEE Eng Med Biol Soc 2014;2014:6899-6902. [doi: 10.1109/EMBC.2014.6945214] [Medline: 25571582]
33. Imtiaz M, Ramos-Garcia R, Senyurek V, Tiffany S, Sazonov E. Development of a multisensory wearable system for monitoring cigarette smoking behavior in free-living conditions. Electronics (Basel) 2017 Dec 28;6(4):104 [FREE Full text] [doi: 10.3390/electronics6040104] [Medline: 29607211]
34. Imtiaz MH, Senyurek VY, Belsare P, Tiffany S, Sazonov E. Objective detection of cigarette smoking from physiological sensor signals. Annu Int Conf IEEE Eng Med Biol Soc 2019 Jul;2019:3563-3566. [doi: 10.1109/EMBC.2019.8856831] [Medline: 31946648]
35. Senyurek V, Imtiaz M, Belsare P, Tiffany S, Sazonov E. Cigarette smoking detection with an inertial sensor and a smart lighter. Sensors (Basel) 2019 Jan 29;19(3):570 [FREE Full text] [doi: 10.3390/s19030570] [Medline: 30700056]
36. Senyurek VY, Imtiaz MH, Belsare P, Tiffany S, Sazonov E. Smoking detection based on regularity analysis of hand to mouth gestures. Biomed Signal Process Control 2019 May;51:106-112 [FREE Full text] [doi: 10.1016/j.bspc.2019.01.026] [Medline: 30854022]
37. Nakajima M, Lemieux AM, Fiecas M, Chatterjee S, Sarker H, Saleheen N, et al. Using novel mobile sensors to assess stress and smoking lapse. Int J Psychophysiol 2020 Dec;158:411-418 [FREE Full text] [doi: 10.1016/j.ijpsycho.2020.11.005] [Medline: 33189770]
38. Hernandez LM, Wetter DW, Kumar S, Sutton SK, Vinci C. Smoking cessation using wearable sensors: Protocol for a microrandomized trial. JMIR Res Protoc 2021 Feb 24;10(2):e22877 [FREE Full text] [doi: 10.2196/22877] [Medline: 33625366]
39. Battalio SL, Conroy DE, Dempsey W, Liao P, Menictas M, Murphy S, et al. Sense2Stop: A micro-randomized trial using wearable sensors to optimize a just-in-time-adaptive stress management intervention for smoking relapse prevention. Contemp Clin Trials 2021 Oct;109:106534. [doi: 10.1016/j.cct.2021.106534] [Medline: 34375749]
40. Chatterjee S, Moreno A, Lizotte SL, Akther S, Ertin E, Fagundes CP, et al. SmokingOpp: Detecting the smoking 'opportunity' context using mobile sensors. Proc ACM Interact Mob Wearable Ubiquitous Technol 2020 Mar;4(1):4 [FREE Full text] [doi: 10.1145/3380987] [Medline: 34651096]
41. Chatterjee S, Hovsepian K, Sarker H, Saleheen N, al'Absi M, Atluri G, et al. mCrave: Continuous estimation of craving during smoking cessation. Proc ACM Int Conf Ubiquitous Comput 2016 Sep;2016:863-874 [FREE Full text] [doi: 10.1145/2971648.2971672] [Medline: 27990501]
42. Saleheen N, Ali AA, Hossain SM, Sarker H, Chatterjee S, Marlin B, et al. puffMarker: A multi-sensor approach for pinpointing the timing of first lapse in smoking cessation. Proc ACM Int Conf Ubiquitous Comput 2015 Sep;2015:999-1010 [FREE Full text] [Medline: 26543927]
43. Echebarria ITU, Imtiaz SA, Peng M, Rodriguez-Villegas E. Monitoring smoking behaviour using a wearable acoustic sensor. Annu Int Conf IEEE Eng Med Biol Soc 2017 Jul;2017:4459-4462 [FREE Full text] [doi: 10.1109/EMBC.2017.8037846] [Medline: 29060887]
44. Cheng C, Huang S, Yang C, Tang K, Yao D. Detection of cigarette smoke using a surface-acoustic-wave gas sensor with non-polymer-based oxidized hollow mesoporous carbon nanospheres. Micromachines (Basel) 2019 Apr 24;10(4):276 [FREE Full text] [doi: 10.3390/mi10040276] [Medline: 31022928]
45. Gurtner M, Gage R, Thomson G, Jaine R, Stanley J, Smith M, et al. Are children smoke-free at home? Using wearable cameras to study children's exposure to smoking and smoking paraphernalia in private spaces. Child Care Health Dev 2019 Mar;45(2):306-309. [doi: 10.1111/cch.12631] [Medline: 30548456]
46. Imtiaz MH, Hossain D, Senyurek VY, Belsare P, Tiffany S, Sazonov E. Wearable egocentric camera as a monitoring tool of free-living cigarette smoking: A feasibility study. Nicotine Tob Res 2020 Oct 08;22(10):1883-1890 [FREE Full text] [doi: 10.1093/ntr/ntz208] [Medline: 31693162]
47. Qiao S, Qu Y, Ma Y, He Y, Wang Y, Hu Y, et al. A sensitive carbon dioxide sensor based on photoacoustic spectroscopy with a fixed wavelength quantum cascade laser. Sensors (Basel) 2019 Sep 26;19(19):4187 [FREE Full text] [doi: 10.3390/s19194187] [Medline: 31561611]
48. Rahman MA, Cai L, Tawfik SA, Tucker S, Burton A, Perera G, et al. Nicotine sensors for wearable battery-free monitoring of vaping. ACS Sens 2022 Jan 28;7(1):82-88. [doi: 10.1021/acssensors.1c01633] [Medline: 34877860]
49. Patrick H, Fujii CA, Glaser DB, Utley DS, Marler JD. A comprehensive digital program for smoking cessation: Assessing feasibility in a single-group cohort study. JMIR Mhealth Uhealth 2018 Dec 18;6(12):e11708 [FREE Full text] [doi: 10.2196/11708] [Medline: 30563807]

50. Cobos-Campos R, Cordero-Guevara JA, Apiñaniz A, de Lafuente AS, Bermúdez Ampudia C, Argaluza Escudero J, et al. The impact of digital health on smoking cessation. Interact J Med Res 2023 Mar 15;12:e41182 [FREE Full text] [doi: 10.2196/41182] [Medline: 36920468]

Abbreviations

ASPIRE: Automated Smoking Perception and Recording
CReSS: Clinical Research Support System
ECG: electrocardiogram
IMU: inertial measurement unit
mHealth: mobile health
PACT: Personal Automatic Cigarette Tracker
RF: radio frequency
RIP: respiratory inductance plethysmograph
RQ: research question
SVM: support vector machine

Edited by L Buis; submitted 01.09.23; peer-reviewed by P Belsare, S Mitra, K Bold; comments to author 22.11.23; revised version received 16.01.24; accepted 24.05.24; published 30.10.24.

Please cite as:
Favara G, Barchitta M, Maugeri A, Magnano San Lio R, Agodi A
Sensors for Smoking Detection in Epidemiological Research: Scoping Review
JMIR Mhealth Uhealth 2024;12:e52383
URL: https://mhealth.jmir.org/2024/1/e52383
doi:10.2196/52383
PMID:

Review

Mobile Apps for the Personal Safety of At-Risk Children and Youth: Scoping Review

Camille Bowen-Forbes[1], BSc, MPH, PhD; Tilovatul Khondaker[1], MBBS, MPH, MD; Tania Stafinski[1], PhD; Maliheh Hadizadeh[1], BSc, MSc, PhD; Devidas Menon[1], MHSA, PhD

Health Technology and Policy Unit, School of Public Health, Edmonton, AB, Canada

Corresponding Author:
Camille Bowen-Forbes, BSc, MPH, PhD
Health Technology and Policy Unit
School of Public Health
Edmonton Clinic Health Academy
11405 St 87 Avenue NW
Edmonton, AB, T6G 1C9
Canada
Phone: 1 587 712 2086
Email: bowenfor@ualberta.ca

Abstract

Background: Personal safety is a widespread public health issue that affects people of all demographics. There is a growing interest in the use of mobile apps for enhancing personal safety, particularly for children and youth at risk, who are among the most vulnerable groups in society.

Objective: This study aims to explore what is known about the use of mobile apps for personal safety among children and youth identified to be "at risk."

Methods: A scoping review following published methodological guidelines was conducted. In total, 5 databases (Scopus, SocINDEX, PsycINFO, Compendex, and Inspec Archive) were searched for relevant scholarly articles published between January 2005 and October 2023. The gray literature was searched using Google and Google Scholar search engines. The results were reported using the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) guidelines. For summarizing the features and users' experiences of the apps, a published framework for evaluating the quality of mobile health apps for youth was used.

Results: A total of 1986 articles were identified, and 41 (2.1%) were included in the review. Nine personal safety apps were captured and categorized into 4 groups based on the goals of the apps, as follows: dating and sexual violence prevention (n=4, 44% of apps), bullying and school violence prevention (n=2, 22% of apps), self-harm and suicide prevention (n=2, 22% of apps), and homeless youth support (n=1, 11% of apps). Of the 41 articles, 25 (61%) provided data solely on app descriptions and features, while the remaining 16 (39%) articles provided data on app evaluations and descriptions. Outcomes focused on app engagement, users' experiences, and effectiveness. Four articles reported on app use, 3 (75%) of which reported relatively high app use. Data on users' experience were obtained from 13 studies. In general, participants found the app features to be easy to use and useful as educational resources and personal safety tools. Most of the views were positive. Negative perceptions included redundancy of app features and a lack of usefulness. Five apps were evaluated for effectiveness (n=2, 40% dating and sexual violence prevention; n=2, 40% self-harm and suicide prevention; and n=1, 20% bullying and school violence prevention) and were all associated with a statistically significant reduction (P=.001 to .048) in harm or risk to participants at the 95% CI.

Conclusions: Although many personal safety apps are available, few studies have specifically evaluated those designed for youth. However, the evidence suggests that mobile safety apps generally appear to be beneficial for reducing harm to at-risk children and youth without any associated adverse events. Recommendations for future research have been made to strengthen the evidence and increase the availability of effective personal safety apps for children and youth.

(JMIR Mhealth Uhealth 2024;12:e58127) doi:10.2196/58127

KEYWORDS

children; youth; personal safety apps; smartphones; mobile apps; violence; bullying; suicide prevention; youth support; homeless support; mobile phone

Introduction

Background

Interpersonal violence is a global public health and human rights challenge, having effects at the family, community, and national levels, with impacts reverberating across generations [1]. Physical violence, psychological violence, verbal abuse, and sexual assault or harassment are common forms of interpersonal violence [2]. More specific examples are child abuse, dating violence, domestic violence, human trafficking, stalking, hazing, bullying, and older person abuse [3]. It is a leading cause of not only physical and psychological harm but also early mortality and is therefore a significant threat to personal safety [1,4]. Interpersonal violence, therefore, has considerable societal consequences, including significant economic burden due to health care provision, loss of productivity, criminal justice involvement, and antiviolence campaigns and interventions [4,5]. The total economic cost of violence has been estimated to be in the billions for many countries worldwide, including Canada [6,7], the United States [8], and the United Kingdom [5,9].

While every age demographic experiences interpersonal violence, children and youth constitute a particularly important sector. This is because such violence, which can reoccur throughout an individual's life, has enduring consequences, increasing one's lifelong vulnerability to a myriad of emotional and physical health problems and negative health behaviors such as substance misuse and risky sexual behaviors [1,10,11]. In 2020, it was estimated that 1 billion children (1 out of every 2 children worldwide) experience some form of violence each year. When aggregated across billions of people, the effects of violence against children can have detrimental effects on economic development [12]. It has been reported that individuals between 12 and 34 years of age are at the highest risk for sexual assault [13]. In a study involving 8629 participants in the United States, violent childhood experiences were reported to double the risk of experiencing intimate partner violence (IPV) in adulthood for women and double the risk of IPV perpetration in men [14]. Studies such as this show that the impacts on younger populations can be more devastating, affecting individuals, families, communities, and society as a whole. It has been shown that children and youth are among the sectors of society that are at greatest risk of violence, sexual abuse, and going missing [15,16]. Among other factors, youth and young adults are at increased risk of victimization, as they are more likely to be single, have lower income, and engage to greater extents in nighttime activities [4].

In addition to harm from older adults, children and youth are also at risk from their peers through incidents such as dating violence, sexual violence, and bullying [10,17,18]. A Youth Risk Behavior Survey conducted in 2019 among high school students in the United States revealed that 25% of students reported bullying victimization and approximately 12% reported dating violence (physical or sexual) [10]. Furthermore, interpersonal violence is one of the main causes of death among adolescents and young adults in most countries worldwide [1].

Youth personal safety is not only impacted by interpersonal violence but also by the risk of self-harm. Mental health issues such as depression, anxiety, and suicide pose significant threats to the lives and well-being of individuals between 10 and 24 years of age worldwide [19]. The World Health Organization describes children as persons aged <18 years and youth as persons between 15 and 24 years of age [19,20]. With people aged <25 years accounting for 42% of the world population and being among the most vulnerable [21], measures to protect their lives and well-being are of utmost importance.

Addressing the issue of violence requires a multifaceted approach involving various levels of society. Mobile technology greatly expands the possible range of available options for addressing these issues [22]. Increasingly, health and human service organizations, policy makers, as well as practitioners across the world have recognized the potential of smartphone apps in helping to address social issues including interpersonal violence and mental health issues at both individual and community levels [5]. The global increase in smartphone ownership makes this option even more potentially useful. The number of smartphone users worldwide has continuously increased from approximately 1 billion in 2014 to 4.88 billion in 2024 and is forecast to reach 6.4 billion by 2029 [23]. There were almost 7 billion smartphone mobile network subscriptions worldwide in 2023, and this number is expected to exceed 7.7 billion by 2028 [24]. Smartphone apps are a particularly important avenue for addressing youth's issues, as youth tend to be more open to technological services compared to the more traditional approaches [25]. The prevalence of smartphones among the younger demographic is well known [26], with smartphones being ubiquitous among youth and young adults. As of January 2024, a significant 98% of Gen Z (people born between 1997 and 2012) own a smartphone [27]. In April 2022, a significant 87% of teenagers between 12 and 17 years of age in Canada were using smartphones; half of the children between 7 and 11 years of age and 39% of children between 2 and 6 years of age were reported to use a mobile device [28]. Similarly, in the United States, 88% of teenagers aged between 13 and 18 years owned a smartphone in 2021; among younger kids between 8 and 12 years of age, ownership of tablets (57%) was more prevalent than smartphone ownership (43%) [29].

Widespread access to mobile phones opens up opportunities for their use as tools to mitigate the risk of harm to children and youth, improving outcomes in instances when such incidences occur. There has been a growing interest in the use of mobile apps for enhancing personal safety; however, there is a lack of evidence on the use and effectiveness of such apps that are specifically geared toward protecting children and youth. Most of the literature on mobile apps has been focused on health and fitness [4,30-38]. There are also some studies on personal safety

apps, but most of them either cover a wide age demographic, are focused on sexual violence against women (with no youth focus), or are focused solely on app development with no associated evaluation [5,39-43]. Furthermore, evidence on apps intended for autonomous use by children and youth is lacking in the literature. Ford et al [5] published an overview of smartphone apps available in the United Kingdom. Of the 86 apps included in the study, 52% targeted the general population, 26% targeted women, and 13% targeted families. None of the studies specifically targeted youth. Nonetheless, that research found that app functionality included providing an alarm (22%), sending alerts to predesignated contacts (71%), providing evidence capture (34%), and offering educational information (26%). More than 70% of apps had a user rating of at least 4 out of 5. Key aspects included positive consequences of app use, technical issues, dissatisfaction with the financial cost of some features, and ethical issues [5]. The effectiveness of the apps was not evaluated.

Most of the literature on personal safety apps is focused on preventing sexual violence or domestic violence, particularly against women. This is not surprising, considering the high prevalence of sexual violence victimization in women globally. For example, in Canada, the rate of IPV was >3 times higher among women and girls compared to men and boys in 2022 [15]. Doria et al [39] identified 3 themes in their review on women's experience with safety apps: security, accessibility, and knowledge. Although there was no evaluation of effectiveness, a common thread among most of the app users was their view that the apps were acceptable, user-friendly, and useful [15]. The review highlighted the potential of smartphone interventions to become a valuable tool for preventing sexual violence in women. Sumra et al [44] conducted a systematic review that included 136 smartphone apps that targeted domestic violence prevention. They found that over two-thirds of the apps (71%) were released between 2020 and 2022, with almost a half of them (46%) being from northeast America. Five app categories were described: emergency assistance (44%), avoidance (21%), informative (21%), legal information (7%), and self-assessment (5%) [44], which were similar to those identified by other researchers [45]. Unique features among the apps included geo-fences, shake-based alert, accelerometer-based alert, alert auto cancelation, anonymous communication, and data encryption [44]. None of the apps had automated alerts or used artificial intelligence to help potential survivors. There was no focus on youth and no evaluation of effectiveness. A 2016 systematic app search for intimate partner and sexual violence prevention and response apps found that, of the 132 unique apps identified, 66% targeted adults, 24% targeted the general population, 27% targeted young adults, 10% targeted teens, and 2% targeted children aged <12 years. However, the app categories were not mutually exclusive, and the specific apps were not identified. As a result, it is impossible to determine what proportion of the apps specifically targeted the younger demographic or to identify them [13]. The apps were found to vary greatly in quality, and sharing information or resources was the primary purpose of most of the apps (76%).

Draughon Moret et al [13], who were experienced forensic examiners, reported that there were only a few apps that they would use as clinicians or recommend to their patients after a physical or sexual assault. The apps focused largely on education and information sharing; therefore, it was thought that they may not successfully meet their desired goal. In addition, they experienced difficulty in finding the apps, as searches for violence prevention and response apps yielded many disturbing apps (zombie-killing games, dating sims, etc), which could potentially retraumatize patients. Furthermore, there was a lack of quality and evidence base among the apps [13].

Reviews focused on sexual violence or domestic violence prevention have found that most of the apps addressed emergencies, with a large proportion of apps focusing on avoidance or education [44-46]. They concluded that further research on app development should focus on automation, making better use of artificial intelligence, speech recognition, and pitch detection to assist in live analysis of the situation and for accurately generating emergency alerts [44]. Other recommendations for further research include a greater focus on app efficacy, sustainability, and data security [45].

Despite widespread access to mobile apps and the growing interest in their use for enhancing personal safety, there is a lack of evidence on the use and effectiveness of such apps that are specifically geared toward protecting children and youth.

Objective

This review aimed to understand what is known about the use of mobile apps for personal safety among at-risk children and youth.

Methods

Overview

A scoping review was conducted following published methodological guidelines by Arksey and O'Malley [47]. They comprise the following 6 steps: identifying the research question; consulting with stakeholders (an optional step in the framework); identifying relevant studies; selecting studies; charting the data; and collating, summarizing, and reporting the results [47]. A scoping review was conducted, as this type of review is particularly useful for mapping the scope, range, and character of the literature and identifying any potential gaps in the body of knowledge on a given topic [48]. Multimedia Appendix 1 [49] provides the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) checklist [49]. No protocol for this review was previously published. The term "at-risk children and youth" refers to those who are in physical or mental danger [50].

The main categories of the theoretical framework developed by Jeminiwa et al [51] for evaluating the quality of mobile health (mHealth) apps for adolescent users were used to provide an overview of app features. The framework has 5 categories (technical quality; engagement; support system; autonomy; and safety, privacy, and trust). However, the authors of this paper modified it to include "esthetics" (included as a subcategory of "engagement" in the framework by Jeminiwa et al [51]) as a distinct category to cover layout, graphics, and visual appeal. In addition, a "subjective quality" category was added to cover

concepts such as usefulness and recommendability. The modifications were guided by the features of the validated Mobile App Rating Scale [52,53], and "personal safety" was also added by the authors to capture features such as self-tracking and a panic button (Multimedia Appendix 2 [51,52]).

Identifying the Research Question

The research question was as follows: "What is known about the use of mobile apps to ensure personal safety among at-risk children and youth?"

Consulting With Stakeholders

To inform the research, a police service division in Alberta planning to develop a personal safety app for at-risk children and youth provided information on important issues to consider. Topics such as app features, use, users' perceptions, and effectiveness were discussed. Issues related to privacy and security were also discussed. A computing science professor from the University of Saskatchewan with expertise in the development of apps also offered insights into key aspects of personal safety apps for youth.

Identifying Relevant Studies

Both peer-reviewed and gray literature sources were included in this review. As non–peer-reviewed sources (eg, reports and app-specific websites) can provide valuable insights and perspectives that may not be captured solely through peer-reviewed literature, these sources were included. In particular, they provided useful information on the characteristics of the various apps. With support from an experienced research librarian, a search strategy for scholarly literature was developed and tested iteratively. In total, 5 databases were searched: Scopus, SocIndex (EBSCO platform), PsycINFO (Ovid platform), Compendex, and Inspec Archive (both Engineering Village platforms). The searches were performed from July 19 to July 30, 2023, using combinations of relevant terms, such as "at-risk," "youth," "children," "safety," and "mobile application." Keywords included "homeless teenagers," "runaway children," "abandoned children," "street youth," "school-aged," "Indigenous youth," "poor children," "juvenile offenders," "LGBTQ+," "sexually abused teenagers," "domestic violence," "protect," "prevent," "safety app," "mobile-based," and "smartphone." Adjustments to the search strategy across different databases were made due to database-specific indexing or features. For example, both APA PsycINFO (Ovid platform) and SocINDEX (Ebsco platform) include extensive but differing controlled vocabularies for children who are abused, fostered, homeless, or neglected and their care. APA PsycINFO uses terms such as "foster care," "child neglect," and "protective services," whereas SocINDEX uses "foster home care," "child abuse," and "child protection services." Where possible, equivalent free-text terms were used across all the databases. Syntax was adjusted according to the specifications of each database or platform. All citations were imported into EndNote version 9.3.3 (Clarivate Analytics, Inc), and duplicates were removed. For the gray literature, Google and Google Scholar were searched using similar terms to those applied to the peer-reviewed strategy.

Study Selection

Included articles satisfied the following criteria: (1) participants were at-risk children or youth; (2) the article focused on mobile apps designed for personal safety; (3) the children and youth had autonomous control of the app; (4) the article was published between 2005 and 2023, as the use of mobile apps for safety applications has been fairly recent; and (5) the evaluation study assessed app users' experience, app engagement, or app impact. Articles intended to be used for characterizing the apps did not need to be evaluation studies. As the terms "children" and "youth" are variously described in the literature, no strict age limits were applied for inclusion; rather, if the target or study population was described using descriptors for children and youth such as "teenagers," "adolescents," or "college students," the study was included. Due to the paucity of available articles, studies focused on participants not strictly considered "at risk" were also included, as long as they focused on children and youth. If ≥1 of the abovementioned criteria were not satisfied for a given app, the articles were excluded. Bullying prevention apps focusing solely on cyberbullying were outside of the scope of this project and were therefore excluded, as were articles not available in English.

For the peer-reviewed literature search, 2 researchers (CB-F and TK) independently screened the titles and abstracts of the identified articles. For quality assurance, a portion of the articles was reviewed by both researchers. Conflicts were resolved through discussion. In cases of disagreement, a third researcher (DM) arbitrated. For the gray literature search, the same 2 researchers systematically searched Google and Google Scholar using similar keywords to those applied to the peer-reviewed search and scanned the first 50 "hits" generated from applying the search terms. Excel software (Microsoft Corporation) was used for data management.

Charting the Data

Information collected from papers was extracted using a standard template. The data extracted included the following elements: goal of app, operating system, date launched, provider or developer, target users, general description, features, app funding, study aim, study type, study period, methods, participants, outcomes measured, findings, facilitators and barriers to app use, app limitations, and conclusions and recommendations. The data extraction tables were piloted and revised as necessary. To ensure consistency in data extraction, CB-F and TK each independently extracted data from a single article and then reviewed each other's work to establish a consistent approach to charting. The researchers met several times during the screening process to ensure a consistent data charting approach.

Collating, Summarizing, and Reporting the Results

The apps were categorized into 4 groups based on app goals and target populations. A descriptive analytical approach was then used to summarize the findings. This involved using common analytical frameworks for summarizing different aspects of the included articles and collecting standard information from each of them [47]. For example, to summarize app features and users' perceptions, a modified version of the

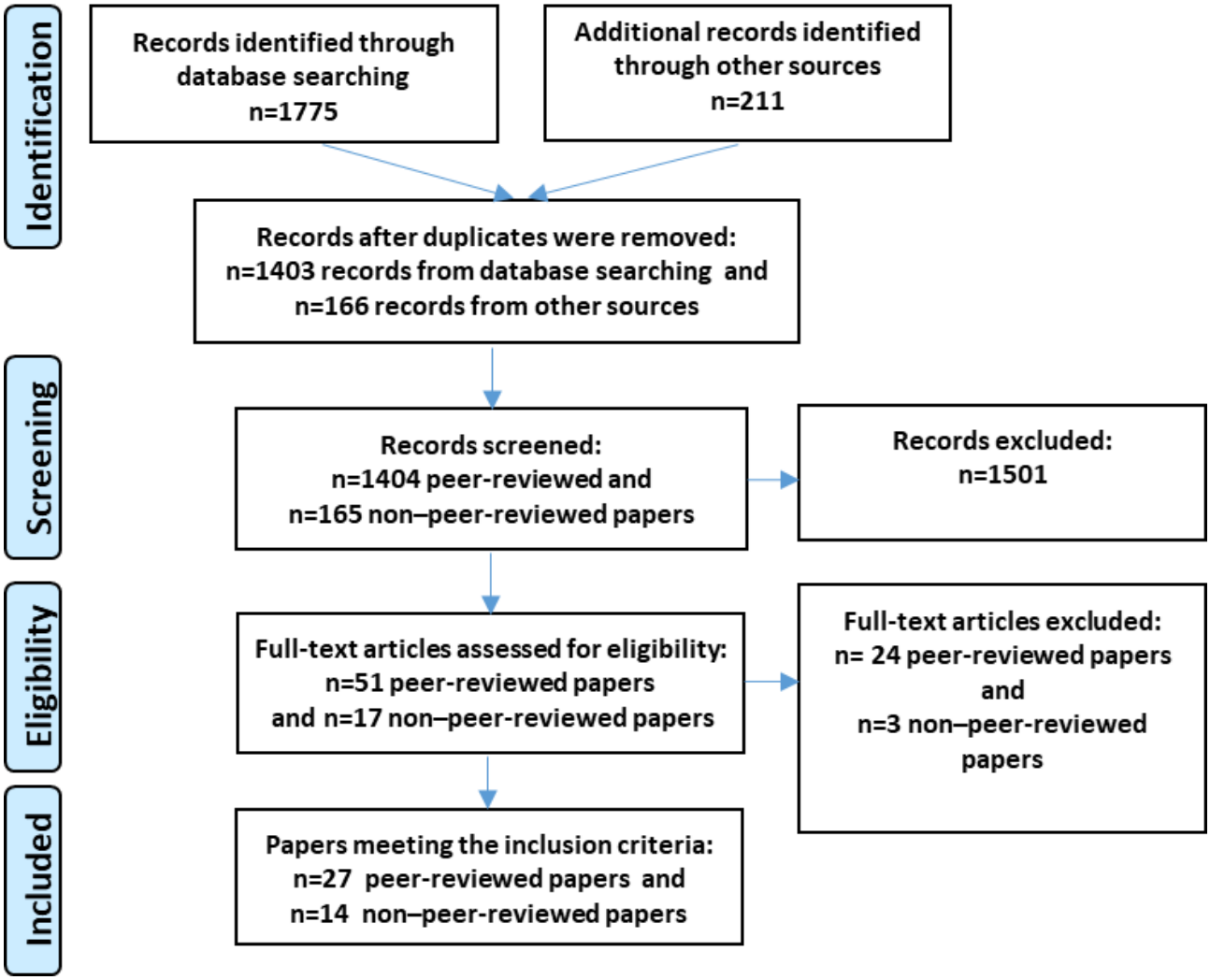
framework developed by Jeminiwa et al [51] was used. For app features, 6 categories were captured: engagement; esthetics; support system; personal safety; autonomy; and safety, privacy, and trust. Four categories were captured for users' perception: engagement; esthetics; safety, privacy, and trust; and subjective quality. Evidence on the effectiveness of the apps was organized by outcomes, such as IPV and other sexual violence, school violence and bullying, and suicide ideation and suicide risk. Data on app evaluation were summarized in Microsoft Excel spreadsheets.

Results

Results of Literature Search

A total of 1986 articles were identified through peer-reviewed (n=1775, 89.37% articles) and gray literature searches (n=211, 10.62% articles). Most of the identified literature on mobile apps primarily focused on health and fitness [4,30-37]. There were also some studies on personal safety apps; however, most either covered a broad or older demographic, focused on sexual violence against women, or solely addressed app development without evaluation [5,38-42]. Consequently, of the 1986 articles, only 68 (3.42%) were eligible for full-text screening. Finally, 41 articles (n=27, 66% peer-reviewed and n=14, 34% non–peer-reviewed) met the inclusion criteria and were included for data extraction. Collectively, these 41 articles provided data on the features and evaluation of 9 apps that met our inclusion criteria. Several studies reported on various aspects or phases of the app development and evaluation process in different articles. For example, for 1 app, acceptability and impact were captured in 2 separate articles [54,55]. The non–peer-reviewed literature primarily provided detailed information on app characterization, including descriptions, features, and functionalities. The results of the screening and selection process are presented in the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram (Figure 1).

Figure 1. PRISMA (Preferred Reporting Item for Systematic Reviews and Meta-Analyses) flow diagram of study selection. Additional records identified through other sources include both non–peer-reviewed and peer-reviewed articles. Reasons for exclusion of studies include the following: not focused on children or youth; not focused on personal safety apps; no information on app use, users' perception, or app impact; published before 2005; and non-English.



Basis of App Development and Stakeholder Engagement

All 9 apps identified through the included articles are summarized in Table 1, and further details are provided in Multimedia Appendix 3 [54-91]. In addition to the language used in the respective articles to describe the type of app, the descriptors from the typology of crime prevention apps by Wood et al [56] were also used. For example, some apps were described as decision aid apps, which are apps designed to help individuals in making decisions based on high-quality evidence [56]. The use of a relevant theory as a basis for development was noted for 3 (33%) of the 9 apps: ambivalent sexism theory and romantic love myths for Liad@s (Universitat de València) [57,58], cognitive behavior theory and dialectic behavior therapy for BlueIce (Oxford Health NHS), and acceptance and commitment therapy for iBobbly (Black Dog Institute). In addition, myPlan (Johns Hopkins University School of Nursing) was developed based on literature on empowerment, internet safety decision aid, and safety planning [59]; +FORT (AXEL; University of Montréal) was developed based on a coordinated sequence of 4 mechanisms of action central to problem-solving [60]; and organizations focused on domestic and sexual violence were consulted in the development of uSafeUS (University of New Hampshire) [61]. The development of the YTH Street Connect prototype app (Santa Clara University Frugal Innovation Hub) was based on information obtained from formative research on homeless or unstably housed youth and mHealth apps and in consultation with homeless or unstably housed service providers [62]. Collaborators from multiple sectors were involved in the development of most of the included apps (7/9, 77%). They included universities (7/9, 77%), schools (2/9, 22%), target users (7/9, 77%), parents (3/9, 33%), companies or organizations (3/9, 33%), and mental health professionals (2/9, 22%). In total, 7 (77%) of the 9 apps are available in English only, while 1 (11%) is available in English and French (+FORT) and 1 (11%) is available in Spanish (Liad@s). In total, 6 (66%) of the 9 apps are currently available, 5 (56%) of which may be freely downloaded and 1 app, 1 (11%) is available only by prescription from child mental health services [54,55]. Of the 3 apps that are not currently available, 2 (67%) were simply prototype apps (YTH StreetConnect) [62] and Circle of 6 [Co6] (Youth Tech Health; Kliq) [63]) and 1 (33%; iBobbly) was recently discontinued [92]. On the basis of an official statement of the First Nations team at the Black Dog Institute (iBobbly developers) in New South Wales, Australia, iBobbly was decommissioned in response to evolving community needs, with the focus now being on providing and recommending best-in-market products (email, November 12, 2023).

Table 1. Overview of app features.

App features and descriptors	Homeless youth support	Dating and sexual violence prevention				Bullying and school violence prevention		Self-harm and suicide prevention	
	YTH Street-Connect [62]	Circle of 6 [63]	Liad@s [57,64]	myPlan [59,65-67]	uSafeUS [61,68,93]	+FORT [60,69,91]	uSafeHS [70,71]	BlueIce [54,55,72,73]	iBobbly [74-78]
Engagement									
Customizable features				✓	✓		✓	✓	✓
Activities for youth			✓					✓	✓
Gamified			✓				✓		
Aesthetics									
Appealing design	✓			✓		✓	✓	✓	✓
Support system									
Educational content	✓	✓	✓	✓	✓	✓	✓	✓	✓
Resource locator	✓				✓		✓		
Decision aid or personalized action plan	✓			✓	✓				✓
Youth and admin platforms	✓						✓	✓	
Diary or self-checks						✓		✓	✓
Trusted contacts	✓	✓						✓	
Personal safety									
Panic button					✓				
Self-tracking	✓	✓			✓				
Fake call or text		✓			✓				
Emergency call or text	✓	✓		✓				✓	✓
Incident reporting					✓		✓		
Autonomy									
Youth controlled	✓	✓	✓	✓	✓	✓	✓	✓	✓
Free			✓	✓	✓	✓			
Safety, privacy, and trust									
Android and iOS	✓	✓	✓	✓	✓	✓	✓	✓	✓
PIN[a] or password protected				✓	✓		✓	✓	✓

[a]PIN: personal identification number.

App Characteristics

The apps were categorized into 4 groups based on app goals and target populations: homeless youth support (1 app), dating and sexual violence prevention (4 apps), bullying and school violence prevention (1 app), and self-harm and suicide prevention (2 apps). A more detailed description is provided in Multimedia Appendix 3. All 9 apps were designed for the autonomous use of youth. They all featured goal-specific educational content and were available on both iOS and Android devices.

Homeless Youth Support

YTH StreetConnect is a decision aid mobile phone app developed in 2016 with the goal of connecting homeless or unstably housed youth in Santa Clara County, California, United States, to health and vital resources [62]. YTH StreetConnect Pro is a companion tablet app for providers who serve these youth [62]. Both are discussed as a single app for the purpose of this review. YTH StreetConnect has features such as a location-based database, interactive mapping, and emergency hotlines. The app helps youth locate services using visual enhancements. Youth have access to sexual health information and weekly health tips via SMS text messaging. YTH StreetConnect Pro features include a referral function and a medical questionnaire to assess clients' homelessness vulnerability and sexual risk [62].

Dating and Sexual Violence Prevention

Four dating and sexual violence prevention apps were identified, 3 (75%) of which were developed in the United States (Co6, myPlan, and USafeUS) and 1 (Liad@s) in Spain [57,61,63,66-68,79,93]. Liad@s targets adolescents, USafeUS targets college or university students, and Co6 and myPlan

target people at risk of sexual violence. Two of the apps have decision aid or personalized action plan features (myPlan and uSafeUS), and 1 app has an interactive game format (Liad@s). Two apps have an emergency text or a fake-a-call or text feature (Co6 and USafeUS, respectively), and 1 (USafeUS) app has an incident reporting feature. Co6, which was a pilot app, is not currently available on the market [80].

Bullying and School Violence Prevention

Two apps were identified in the bullying and school violence prevention category: +FORT (Canada) [60,69,81] and uSafeHS (University of New Hampshire; United States) [70,71]. +FORT, developed in Quebec, was first available in French and was subsequently made available in English. Both apps target high school students, with +FORT aimed at preventing bullying victimization. uSafeHS aims to prevent school violence in general, including bullying (Multimedia Appendix 3). +FORT allows youth to journal their bullying victimization experiences and compiles the information in simple graphs, which youth may use to enhance their safety awareness [60,81]. uSafeHS has an administrative platform, incident reporting, and an interactive game feature to facilitate social and emotional learning [70,71].

Self-Harm and Suicide Prevention

BlueIce [54,55,72,73,82-85] and iBobbly [74-78,86,87] are self-harm and suicide prevention apps that were identified. BlueIce targets young people attending Child and Adolescent Mental Health Services across the United Kingdom and aims to help them reduce urges to self-harm. iBobbly targeted Aboriginal and Torres Strait Islander Australians aged ≥15 years (youth, in particular) and aimed to tackle suicide prevention in a culturally appropriate way. Both have emergency call or text, diary or self-check, and activities that youth may engage in as features.

Characteristics of App Evaluation Studies

A total of 14 studies conducted between 2013 and 2022 were identified and reported in 16 peer-reviewed papers. In total, 15 papers reported on experimental study designs (3 randomized controlled trials [RCTs] [59,74,76], 9 single-arm design [54,55,61-63,65,66,70,79], 2 quasi-experimental design [57,64], and 1 that used single-arm and quasi-experimental designs in 2 phases [60]). One paper reported on a cross-sectional study [93]. All the studies included self-reported data. The studies were conducted in the United States (5 apps and 9 papers [59,61-63,65,66,70,79,93]), Australia [74,76], Canada [60], Spain [57,64], and the United Kingdom [54,55] (1 app each).

The 14 studies were conducted across several different settings as follows: 6 (43%) in college or universities [59,61,63,65,66,79,93]; 3 (21%) in high schools [60,70], and 1 (7%) each in mental health services for children [54,55], childcare homes [57], Aboriginal and Torres Strait Islander communities [74,76], the homeless youth context [62], and dating locations [63].

The duration of app use ranged from 20 to 40 minutes in 1 single-arm qualitative study [65] to 12 months in an RCT. For 7 (50%) of the 14 studies (corresponding to 6 apps), the apps were used for at least 6 weeks.

Characteristics of Participants

The reviewed studies included >3800 children and youth, but demographics were not consistently reported. Reported average age ranged from 14 to 27 years; in 1 study, one-third were aged ≥26 years [93], and in another study, the age range was 19 to 29 years [74]. Where information on sex was reported, only 1 study had more men than women [57] (in a childcare home), and in 2 studies evaluating dating and sexual violence prevention apps, only women were enrolled [59,63]. Only 3 studies captured information on sexual orientation [62,63,79]. Where reported, the proportion of White participants ranged from 33% (2/6) [62] to 71.3% (122/171) [59,62].

In addition to children and youth, other stakeholders (eg, parents, app administrators, and police officers) were included as participants in the evaluation of 4 apps [54,55,60,61,70].

Study Outcomes and Measurement Approaches

Table 2 provides a summary of the outcomes and associated measurement tools used in the included studies. Outcomes comprised app engagement (4/9, 44% apps), users' experiences (9/9, 100% apps), and effectiveness of the app (7/9, 78% apps; Multimedia Appendix 4 [54,55,57,59-66,70,74,76,79,92,93]). Effect measures were safety behavior, bystander behavior, sexism, IPV, self-harm, suicide risk, depression, anxiety, and bullying victimization. A wide range of tools was used for measuring outcomes. Self-reported measures were used for all 3 types of outcomes reported (app use: 2/4, 50% apps, users' experience: 9/9, 100% apps, and effectiveness: 7/7, 100% apps). Validated tools were used in the evaluation of effectiveness of 86% (6/7) of the apps. For the remaining app (+FORT), even though no validated tool was used for measuring effectiveness, the app itself uses items adapted from the validated Olweus Bullying Questionnaire for logging information about users' bullying experiences to help users identify more effective strategies to deal with bullying victimization [60]. Objective measurements were used for measuring app use for 2 (50%) out of 4 apps and for measuring the effectiveness of 1 (14%) out of 7 apps. For example, for the BlueIce app, the change in users' frequency of self-harming was determined by comparing historical clinical data with the self-reports of postintervention self-harming incidence [46]. For 3 apps, the general internet use of app users was also assessed, and self-reported, unvalidated measures were used [54,55,62,70]. Further details on study outcomes and measurement approaches are provided in Multimedia Appendix 4.

Table 2. Study outcomes and measurement approaches.

Outcomes and tool name and/or description	Mode of measurement	App	Study
General internet use			
Question, questionnaire, survey, or interview	Self-reported and unvalidated	iBobbly, uSafeHS, and YTH StreetConnect	[62,70,74]
App engagement (downloads, frequency of use, features used, etc)			
Mobile device download data	Objectively measured	iBobbly	[74]
Administrative dashboard analytics	Objectively measured	usSafeHS	[70]
Question, questionnaire, survey, or interview	Self-reported and unvalidated	Circle of 6	[63]
Question, questionnaire, survey, or interview	Self-reported and unvalidated	uSafeUS	[93]
Users' experience (perception, feasibility, acceptability, or subjective value)			
Question, questionnaire, survey, or interview	Self-reported and unvalidated	BlueIce, Circle of 6, +FORT, Liad@s, iBobbly, myPlan, uSafeHS, uSafeUS, and YTH StreetConnect	[54,57,59-63,65,66,70,74,79,93]
Effectiveness			
Decisional conflict: Decisional Conflict Scale (modified)	Self-reported and validated	myPlan	[59]
Safety behaviors: question, questionnaire, survey, or interview (number of safety behaviors tried on app)	Self-reported and validated	myPlan	[59]
Bystander behavior: Banyard's Bystander Scale	Self-reported and validated	Circle of 6	[63]
Intention to help: 10-item modified Intention to Help Scale	Self-reported and validated	Circle of 6	[63]
Sexism (hostile sexism and benevolent sexism): Ambivalent Sexism Inventory-Adolescents	Self-reported and validated	Liad@s	[57,64]
Ambivalence toward men: Ambivalence toward Men Inventory	Self-reported and validated	Liad@s	[57]
Distortions about romantic love: Myths, Fallacies, and Erroneous Beliefs about the Ideal of Romantic Love Scale	Self-reported and validated	Liad@s	[57]
Sexual victimization: 11 item-Revised Sexual Experience Survey	Self-reported and validated	Circle of 6	[63]
Intimate partner violence: Composite Abuse Scale	Self-reported and validated	myPlan	[59]
Intimate partner violence: traumatic brain injury (questions or questionnaire)	Self-reported and unvalidated	Liad@s	[57]
Intimate partner violence: digital abuse (questions or questionnaire)	Self-reported and unvalidated	myPlan	[59]
Intimate partner violence: reproductive coercion (questions or questionnaire)	Self-reported and unvalidated	myPlan	[59]
Substance use: frequency of alcohol use	Self-reported and unvalidated	Circle of 6 and myPlan	[59,63]
Substance use: frequency of getting drunk	Self-reported and unvalidated	myPlan	[59]
Substance use: frequency of binge drinking	Self-reported and unvalidated	myPlan	[59]
Substance use: any marijuana use	Self-reported and unvalidated	myPlan	[59]
Substance use: any drug use other than marijuana	Self-reported and unvalidated	myPlan	[59]
Substance use: feeling of intoxication	Self-reported and unvalidated	Circle of 6	[63]
Self-harm: clinical data	Objectively measured	BlueIce	[63]
Self-harm: question, questionnaire, survey, and interview	Self-reported and unvalidated	BlueIce	[63]
Bullying victimization: Multidimensional Peer-Victimization Scale	Self-reported and validated	+FORT	[60]
Depression: Center for Epidemiologic Studies Depression Scale Revised	Self-reported and validated	myPlan	[59]
Depression: Mood and Feelings Questionnaire	Self-reported and validated	BlueIce	[54,55]
Depression: Patient Health Questionnaire 9	Self-reported and validated	iBobbly	[74,76]
Anxiety: Revised Child Anxiety and Depression Score	Self-reported and validated	BlueIce	[54,55]

Outcomes and tool name and/or description	Mode of measurement	App	Study
Psychological distress: Kessler Psychological Distress Scale	Self-reported and validated	iBobbly	[74,76]
Impulsivity: Barratt Impulsivity Scale	Self-reported and validated	iBobbly	[76]
Behavior: Strengths and Difficulties Questionnaire	Self-reported and validated	BlueIce	[54,55]
Suicide risk: Center for Epidemiologic Studies Depression Scale	Self-reported and validated	myPlan	[59]
Suicidal ideation: Depressive Symptom Inventory–Suicidality Subscale	Self-reported and validated	iBobbly	[74,76]

App Engagement

Four studies reported data on the use of 4 apps: 2 on dating and sexual violence prevention (Co6 [63] and uSafeUS [93]), 1 on bullying and school violence prevention (uSafeHS) [60], and 1 on self-harm or suicide prevention (iBobbly) [74]. Multimedia Appendices 4 and 5 [54,55,57,59-66,70,74,76,79,92,93] provide details on the characteristics and the findings of the included evaluation studies. On the basis of app download and use, the findings of 2 studies indicated that there is a high level of app use among high school students (uSafeUS) [70] and Aboriginal and Torres Strait Islander youth (iBobbly) [74]. Two studies involving college students found generally low app engagement [63,93]. The findings indicated that younger app users had higher engagement levels than older users, and women were more engaged than men [93]. The reasons for low app use among college students included the perceived redundancy with existing smartphone features, college women's discomfort with group messaging (Co6) [63], and the opinion among older college students that the app was not relevant to their needs (uSafeUS) [61].

App Users' Experience

Data on users' experience of the apps were obtained from 13 studies, with all 4 app categories being covered (Tables 3 and 4; Multimedia Appendices 4 and 5) [54,57,59-63,65,66,70,74,79,93]. In general, participants liked the app features and perceived the apps to be easy to use and effective. One dating and sexual violence prevention app (myPlan), 1 self-harm and suicide prevention app (iBobbly), and 1 bullying and school violence prevention app (+FORT) were perceived to be judgment free by youth [60,66,74]. Among the 9 apps, the features perceived to be the most useful included map features, personalized action plan, mood diary, and bullying prevention strategies. The appealing features, confidentiality, accessibility, ease of use, and useful resources were common facilitators of app use among youth. In addition, the judgment-free nature of some apps facilitated their use. By contrast, the repetitive, redundant nature and lack of specific resource information were among the reported barriers to app use.

Table 3. Youth's perceptions of mobile personal safety apps.

Users' perception and descriptors	Homeless youth support	Dating and sexual violence prevention				Bullying and school violence prevention		Self-harm and suicide prevention	
	YTH StreetConnect [62]	Circle of 6 [63]	Liad@s [57]	myPlan [59,65,66,79]	USafeUS [61,93]	+FORT [60]	uSafeHS [70]	BlueIce [54,55]	iBobbly [74]
Engagement									
Accessible						✓		✓	✓
Easy to use	✓	✓		✓	✓		✓	✓	
Fun or enjoyable	✓								
Favorable features	✓	✓		✓	✓		✓	✓	
Aesthetics									
Appealing design	✓			✓					
Safety, privacy, and trust									
Private or confidential								✓	✓

Table 4. Youth's perceptions of the usefulness of mobile personal safety apps.

Users' perception and descriptors	Homeless youth support	Dating and sexual violence prevention				Bullying and school violence prevention		Self-harm and suicide prevention	
	YTH StreetConnect [62]	Circle of 6 [63]	Liad@s [57]	myPlan [59,65,66,79]	USafeUS [61,93]	+FORT [60]	uSafeHS [70]	BlueIce [54,55]	iBobbly [74]
Subjective quality									
Useful	✓	—[a]	—	✓	—	✓	—	—	✓
Effective	✓	Mixed	✓	✓	✓	✓	✓	✓	✓
Judgment free or shame free	—	—	—	✓	—	✓	—	—	✓
Would use again	✓	—	—	—	—	—	✓	✓	—
Worth recommending	—	Mixed	—	—	—	—	—	✓	✓
Most useful or helpful features	Map feature (Resource Finder)	—	—	Personalized action plan and Myth section	—	Bullying prevention strategies and journal	—	Mood diary, mood lifter, and emergency numbers	—
Preferred features	—	Location feature	—	Danger assessment tool	Customizable and educational and off-campus resources	—	Customizable	Mood diary	—
Nonpreferred features	Lack of specific resource information	Redundant or unnecessary	Not helpful	Lack of information on emotional abuse	Not helpful	—	—	Repetitive and not challenging	—

[a]No or not reported.

Both the youth and service providers who participated in evaluating the homeless youth support app, YTH StreetConnect [62], enjoyed using the app and considered it to be accessible, appropriate, and confidential for locating services. None of the evaluation studies analyzed the privacy or confidentiality of the apps.

For the dating and sexual violence prevention apps, the views of youth and other stakeholders, including college campus administrators and crisis center advocates, were captured [59,61,63,65,66,79,93]. Participants generally had positive perceptions of the apps' sexual violence–related resources and supporting features. For example, Lindsay et al [66] found that women survivors of dating violence who attended college found myPlan to be "useful, innovative, and effective" in conveying information regarding dating violence and relationship safety. In particular, the "My Plan" (personalized safety plan) feature was found to be the most useful feature [66]. By contrast, Debnam and Kumodzi [79] found that among participants who represented a sexually diverse group of adolescents, there was an intolerance to gender-conforming language in the app, which targeted females only. They strongly believed that men can be victims too, and that conversely, women can also be perpetrators. Consequently, participants recommended that the app be modified to reflect a more inclusive group to users with diverse sexual orientations and to have a greater focus on safety dilemmas faced by youth, such as emotional or psychological abuse and power imbalance induced by age difference. Because of that research, a lesbian, gay, bisexual, transgender, queer (LGBTQ) version of the myPlan app was developed and later released [79]. In addition, Potter et al [61] found that while most college students and other stakeholders who used uSafeUS agreed on the need of mobile apps to protect against sexual violence, most of the graduate students (particularly older students) as well as commuter students indicated that they did not perceive the app as being applicable or relevant to their needs.

Across the 2 studies that evaluated bullying and school violence prevention apps among high school students, participants found the app features to be favorable and felt that the apps were useful personal safety tools [60,70]. For example, all the survivors of bullying who used +FORT felt that it may be a beneficial tool, with 1 participant stating, "We talked about it [bullying] during three hours at school and I learned more about it with Stronger than Bullying (as the app was initially called) in five minutes" [60]. uSafeHS users felt that the app could serve as a useful tool for high school students, with all participants who completed the gamified social emotional learning educational modules expressing that their knowledge had improved [70].

Both self-harm and suicide prevention apps were found to be acceptable and helpful to users [54,55,74,76]. In particular, iBobbly was considered culturally appropriate by Aboriginal and Torres Strait Islander youth [74,76]. Although 2 users of BlueIce reported initial concerns that seeing their mood diary full of negative days might not help them, they felt that overall, self-monitoring was beneficial [55].

Evidence by Outcome

The effect of app use on substance use; decisional conflict; safety behaviors; protective behavior; sexism; ambivalence toward men; love myths; sexual violence; bullying victimization; and mental health issues, such as depression, anxiety, self-harm, and suicide risk, were evaluated in 6 studies (8 articles). All the studies that evaluated effectiveness reported positive outcomes associated with app use in at least 1 outcome measure.

Substance Use

Two studies reported on substance use, but only 1 assessed the impact of the app on this outcome. A 12-month RCT study involving college women found that there was a reduction in the frequency of alcohol use, getting drunk, binge drinking, and nonmarijuana drug use over time in both the intervention and control groups. However, only the reduction in drunkenness frequency achieved statistical significance (P=.001), but there was no significant difference between both groups. Interestingly, there was a slight increase in marijuana use in both groups over time [59]. Although the feasibility and acceptability of Co6 among college women who drink alcohol were assessed, its effects on alcohol and the risk of sexual violence were not assessed. Rather, how app users perceived the app as a sexual violence risk reduction tool was assessed [63].

Decisional Conflict and Safety Behaviors

One study that evaluated decisional conflict and safety behaviors in college women reported statistically significantly greater improvement in preparedness to make better safety decisions in the intervention group compared to the control group [59]. There were immediate improvements in all decisional conflict subscales in both groups. In particular, participants in the intervention group were statistically significantly better able to weigh the risks and benefits of different safety options compared to those in the control group (P=.02). It was found that the number of helpful safety behaviors used on the app increased over time, although there was no statistically significant difference between the intervention and the control groups. There was a statistically significant association between the number of safety behaviors tried and IPV reduction in the intervention group only (P<.001) [59].

Protective Behavior

Four studies that evaluated protective behavior or had themes surrounding that topic found that youth were generally willing to help their friends in risky situations [63,65,79,93]. In 1 study, users of Co6 app expressed almost 3 times more protective behavior in sexually aggressive situations toward friends compared to strangers at 2-month follow-up [63]. Of note, they also expressed greater intention to help friends than strangers at the start of the study. In another study, the myPlan app helped friends of survivors of IPV to understand abusive relationships better and helped them to better understand the severity of violence, identify resource options, and know possible ways to intervene [65]. In evaluating the reasons for downloading uSafeUS, it was found that 90% of college women felt confident that it would provide sufficient resources to help them support a friend who disclosed that they had been sexually assaulted [93]. In another study on the myPlan app, while adolescents expressed willingness to help protect their friends who experience dating violence, they also described the moral distress they experienced regarding protecting themselves over their friends in risky situations [79].

Sexism, Ambivalence Toward Men, and Love Myths

Sexism was measured in 2 quasi-experimental studies evaluating the Liad@s app. One involved a 2-week intervention involving residents of childcare homes in Spain, who ranged in age from 11 to 18 years [57], and the other involved a 2-hour intervention involving high school students aged 13.9 years, on average [64]. Ambivalence toward men and love myths (distortions about romantic love) were measured only in the study involving residents of childcare homes in Spain [57]. Across the 2 studies, the app was found to be effective in reducing sexism, ambivalence toward men, and love myths. Participants experienced a statistically significant reduction in hostile sexism (P=.009) and benevolent sexism post intervention relative to pre intervention (P<.001), with greater reductions observed in these variables in the intervention group compared to the control group. The difference was statistically significant only in the study involving high school students [64]. Similarly, there was a significant reduction in ambivalence toward men and distortions of romantic love postintervention relative to pre intervention among residents in a childcare home (P=.02 and P<.001, respectively) [57]. A statistically significant pre-post difference in distortions of romantic love was observed in the intervention group only. There were no significant gender differences in sexism or myths about romantic love. A statistically significant decrease in hostile sexism with increasing age was observed, and there was also a decrease in paternal resentment with age [57].

IPV and Other Sexual Violence

One RCT study that evaluated the impact of a personal safety app (myPlan) on IPV among college-going female survivors of IPV between 18 and 24 years of age found that after 12 months of using the app, there was a statistically significant decrease in IPV in both intervention and control groups in all 4 subscales measured (Composite Abuse Scale, traumatic brain injury–related IPV, digital abuse, and reproductive coercion) [59]. The intervention group, however, experienced a statistically significantly greater reduction in reproductive coercion compared to the control group (P=.02). In 1 mixed methods study that evaluated sexual victimization among college women, 23% (10/44) of participants who used Co6 reported sexual victimization at 2-month follow-up, which involved unwanted sexual contact (5/10, 50%), completed rape (3/10, 30%), and attempted rape (2/10, 20%) [63]. Most of the perpetrators were friends or acquaintances 44% (4/9), while 33% (3/9) were strangers and 22% (2/9) were their boyfriends. The participants had mixed views on whether the app made them feel safer from sexual violence. App use was low due to perceived redundancy with existing smartphone features and college women's discomfort with group messaging [63] (Multimedia Appendices 4 and 5).

School Violence and Bullying Victimization

One study that reported on bullying victimization among high school students found a 2-fold reduction in bullying victimization, which occurred after 4 to 6 weeks of app use (P<.001). The reduction in victimization was 16 times greater for the intervention group compared to the control group, who did not receive the app. None of the 5 parents or educators involved in the study believed that the app could jeopardize or conflict with existing services [60]. One study that gathered participants' input and feedback on the development and testing of a school violence prevention app (uSafeHS) found that the app was well received by youth and appeared to be a useful tool in streamlining all services for homeless or unstably housed youth and their service providers. The impact on safety was, however, not evaluated [70].

Depression

All 3 studies that reported on depression (2 on self-harm or suicide prevention apps and 1 on a dating and sexual violence prevention app) found a statistically significant reduction among youth who used personal safety apps (P<.001 to .02) [55,59,76]. For 1 app (myPlan), there was no difference between the intervention and control arm, whereas for another (iBobbly), the difference between arms was statistically significant. Interestingly, a follow-up study involving participants of the initial iBobbly RCT study found a nonstatistically significant reduction in depression over time [74]. Of note, the sample size of the follow-up study was much smaller than that of the initial study (13 compared to 61).

Anxiety, Psychological Distress, and Impulsivity

Two studies (3 articles) on self-harm or suicide prevention apps that reported on anxiety-related outcomes found statistically significant reductions in anxiety and psychological distress over time [55,74,76]. One noncontrolled study on BlueIce found a statistically significant reduction in overall scores across all 5 Revised Child Anxiety and Depression Score subscales: panic disorder, separation anxiety disorder, generalized anxiety disorder, social anxiety disorder, and obsessive compulsive disorder (P<.001). One RCT on iBobbly reported statistically significant reductions in psychological distress (P=.02), which was statistically different from the control (waitlist) arm. However, a follow-up report on iBobbly found nonstatistically significant reductions in psychological distress and impulsivity [74].

Self-Harm

A pre-post mixed methods study on a self-harm and suicide prevention app found a reduction in self-harming incidents among 33 youth between 12 and 17 years of age who attended mental health services in the United Kingdom [54,55]. The app helped individuals to not act on their urges to self-harm, with 15% (4/26) of those who self-harmed before the study stopping that practice, and a further 58% (15/26) self-harmed less frequently after using the app [55]. A total of 308 incidents of self-harm were prevented during the study, based on historical clinical data and self-reported rates after app use. In total, 27% (7/26) of participants had no reduction in self-harming behavior. No app user felt that the app would increase their thoughts of self-harming, and no adverse events were reported [54].

Suicide Ideation and Suicide Risk

The 2 studies that evaluated suicide risk–related outcomes among youth reported a reduction after using personal safety apps [59,74]. For 1 dating and sexual violence prevention app, which was used for 12 months, there was a reduction in suicide risk, which was significantly greater in the intervention group relative to the control group (P=.048) [59]. For a self-harm or suicide prevention app, which was used for 6 weeks, the reduction in suicide ideation was nonsignificant. In addition, a third study, which evaluated the safety of another self-harm or suicide prevention app, found that no clinician withdrew any participant from the study because of escalated or emergent risk of suicide planning or attempt [55]. The effectiveness of the apps was attributed to several factors, including the provision of distractions and emotional outlets, tracking and recognizing mood patterns, identifying triggers for negative emotions, gaining new perspectives and coping strategies, improving interpersonal communications, and quick access to emergency numbers [54,74].

Discussion

Principal Findings

This scoping review identified, categorized, and characterized mobile apps used for personal safety among at-risk children and youth and summarized the findings on app use (engagement), users' perception, and effectiveness. To the best of our knowledge, this is the first review that provides this type of information on personal safety apps intended to be used autonomously by children and youth. Six (86%) of the 7 studies that evaluated effectiveness reported positive outcomes associated with app use in at least 1 outcome measure, with statistically significant reductions in drunkenness frequency [59], sexism, ambivalence toward men, love myths [57,64], IPV [59], bullying victimization [57], depression [55,59,76], anxiety [54,55], and suicide risk [59] reported over time. In addition, statistically significant differences between intervention and control arms were reported for sexism [57], reproductive coercion (a measure of IPV) [59], and suicide risk [59]. Furthermore, no study reported an increase in harm to participants. These are promising results, which suggest that mobile personal safety apps may be a viable tool for enhancing the safety of children and youth.

From a global personal safety perspective, 100% (13/13) of the included studies were conducted in high-income countries, a reflection of the concentration of mobile app development in such countries. This is highly disproportionate to the high global prevalence and trends in sexual violence, depression, and anxiety in low-income countries [94-96]. A study of global crime patterns during the period 2006 to 2019 revealed that African and Latin American countries experienced the highest levels of various types of crime, followed by Asian countries. Intermediate or relatively low levels of most types of crime were reported for European, North American, and Australian countries [97]. The abovementioned findings demonstrate that more research needs to be undertaken in low- and

middle-income countries (LMICs), particularly in areas where mobile phone app use is known to be prevalent. Ding et al [98] had similar findings from their review of mHealth and youth mental health. Furthermore, Madonsela et al [99], in a scoping review on the development and use of mHealth interventions in LMICs, identified only 6 relevant studies from 5 countries. Only 2 of the studies were focused on smartphone apps, and 1 study involved multiple intervention types, including smartphone interventions. Only 1 study that involved a SMS text messaging intervention was focused on self-harm or suicide. The authors concluded that more research is needed to build the evidence base in LMICs to develop this field [99]. Decker et al [100] have made progress in increasing research in LMICs through their RCT to evaluate the efficacy of a culturally and linguistically adapted version of the myPlan app used by women at risk of and experiencing IPV in Nairobi, Kenya. The study, however, had a short duration (3-month follow-up) and was not focused on youth [100].

Considering that only 9 apps met the inclusion criteria, this review suggests that despite the growing number of personal safety apps available on the market, very few are specifically geared toward autonomous use by children and youth. This represents a gap in the mobile personal safety app industry. Many apps are available for use by parents and guardians to track their children, and these have their purpose. However, with a high premium placed on their autonomy, youth generally prefer to use apps that they completely control. As it is for mHealth apps [100], the field of personal safety apps offers both opportunities and risks. For any app, but particularly for personal safety apps geared toward autonomous youth control, it is important that the development be undertaken with target user engagement and for the app to be based on scientific evidence and be well validated. The review found that for most of the apps (8/9, 89%), relevant theories or expertise were used in their development. This indicates that, in general, measures are being taken to ensure that personal safety apps for youth are developed based on sound theories and evidence.

A common feature among the 9 apps is the inclusion of educational content. Interestingly, for 2 apps—1 bullying and school violence prevention app [70] and 1 self-harm and suicide prevention app [74]—youth desired more educational content. In contrast, for the evaluation of 1 dating and sexual violence prevention app, college women found the explanatory video for first-time users “cumbersome” and suggested using dialogue boxes with brief instructions that pop up [61]. The review findings suggest that while younger youth generally have positive perceptions of personal safety apps and highly value them, older youth tend to find them less valuable. As victimization oftentimes begins during adolescence and prevails into adulthood [10], and with the appetite for personal safety apps apparently lessening with age, it may be prudent for more research and development to be focused on the younger segment of the youth demographic [10].

The features perceived to be the most useful included map features (homeless youth support), personalized action plan and myth-debunking sections (dating and sexual violence protection), bullying prevention strategies (bullying and school violence prevention), mood diary, and emergency numbers (self-harm and suicide prevention). The danger assessment tool featured in the myPlan app is unique among the included apps. This user-preferred app feature is a validated tool which provides both numerical and graphical displays of the assessed risk for repeated severe IPV [59,65,66]. Such a feature is potentially very valuable in sexual violence prevention. Common facilitators of app use among youth were the appealing features, confidentiality, accessibility, ease of use, and useful resources. However, the repetitive, redundant nature and lack of specific resource information were barriers to app use. For example, for the YTH StreetConnect app, homeless youth found the lack of specific information such as the number of available beds available at a given shelter to be a barrier to app use [62]. Features that youth suggested for app development or improvement included ambiguous name and branding, an easy delete option, a panic button option, bystander-focused intervention for reducing dating violence, and gamification (Multimedia Appendix 5 [54,55,57,59-66,70,74,76,79,92,93]). As target users’ perception is of utmost importance to app development, these findings are of relevance for app developers.

In 5 (38%) of the 13 included evaluation studies, the period of app engagement did not extend beyond 6 weeks [57,60,63,64,74,76]. In 1 study on users’ experience, app use was as short as 20 to 40 minutes [65]. The longest period reported was 12 months in an RCT study [59]. This reveals that a high proportion of the evaluation studies involved very short app engagement periods, which leaves us to wonder to what extent the results can be related to the real-life context. In particular, no conclusion can be drawn about the long-term effectiveness of personal safety apps for children and youth based on this review. Further studies involving longer app engagement periods and follow-up times need to be undertaken to shed light on effectiveness in the long term.

Regarding measurement outcomes, self-reported measures were used for all but 3 distinct outcomes measured (2 on app use and 1 on self-harming frequency) among all the included studies. While self-reported measures are a valuable tool in public health research, they can inadvertently be affected by various biases, particularly information bias. The use of validated tools, however, can markedly strengthen the validity of self-reported measurement tools. Most of the included studies incorporated the use of validated measures for evaluating effectiveness (6/7, 86% apps). However, only 2 (22%) of the 9 validated measures used were specifically geared toward youth, namely, the Multidimensional Peer-Victimization Scale [60] and the Revised Child Anxiety and Depression Score [54,55]. It may be beneficial for future research to focus on developing and validating tools for evaluating mobile personal safety apps specifically for youth to improve the validity of such research involving youth. In addition, the outcome measurements directly related to victimization or risky behavior were generally based only on a reduction in frequency. Although these measures are useful, the incorporation of measures of the nature or severity of victimization and risky behavior would only serve to improve upon the quality of research conducted.

Limitations

This review was conducted in accordance with best practices in conducting scoping reviews. However, some limitations exist. Only articles written in English were included; therefore, some relevant articles could possibly have been missed. In addition, a systematic app search was not a part of the search process. Such a search could have potentially identified apps that could have directed the search for evaluation studies. As apps with no available information on users' experience, app engagement, or impact were outside the inclusion criteria, we believe that a targeted database search along with a search of the gray literature was sufficient and effective in yielding relevant articles. A broad range of study designs and methodologies were included; hence, no quality assessment of included articles was done. However, quality assessment is not a requirement of scoping reviews. The short durations of app use for most of the studies limit the strength of the review findings. In addition, all the evaluation studies used self-reported data, which are more subject to bias than objectively determined measures. The inclusion of validated tools in evaluating all the apps, however, counteracts this limitation to some extent. It is challenging to realistically compare the use, users' perception, and impact of the apps due to the variability of outcome measures and methodologies. Finally, the generalizability of the findings is limited due to the heterogeneity in participant characteristics across different apps and studies, variations in intervention durations, a small number of apps within each category, and the small sample sizes in several studies.

Implications for App Development and Future Research

This review sheds light on youth's experience and perception of personal safety apps geared toward their use. As target users' perception is of critical importance to app development, these findings are of relevance for app developers. The features perceived to be the most useful included maps, personalized action plans, mood diaries, violence prevention strategies, and emergency numbers. Appealing features, confidentiality, accessibility, ease of use, and useful resources motivated youth to use the apps. Of note, the inclusion of validated danger assessment tools such as the one included in the myPlan app [59,65,66] may be a very useful feature for personal safety apps in general. With the constant evolution of technology, the information included in apps can quickly become obsolete. Therefore, co-design methodologies are essential for increasing and sustaining youth engagement as well as increasing the likelihood of universal acceptability [101]. App developers should, therefore, ensure that youth collaborate in app design and development and that app information is regularly updated, including specific details on helpful resources for at-risk youth. In addition, it may be prudent for more research and development to focus on the younger segment of the youth demographic, given that this group of youth appears to value personal safety apps more compared to older ones [10].

All 9 included apps featured educational content. Inclusion of a test-retest knowledge assessment would be a useful feature to provide basic data on the effectiveness of the apps in increasing knowledge [13] and should be considered in developing these apps. A noteworthy, but not surprising, finding of this review is that youth are generally willing to help their friends in risky situations [63,65,79,93]. This information can be harnessed by app developers. It is potentially very useful to routinely develop companion apps for personal protection apps that target friends of at-risk youth or include features in the apps that allow them to be customized for friends of at-risk youth.

Despite the availability of personal safety apps, the scoping literature review found that there were not many that specifically target youth. Furthermore, all the included studies were conducted in high-income countries, and a high proportion of the evaluation studies had short durations. Most of the evaluation measures were self-reported, and the validated tools were generally not specifically geared toward youth. In addition, some of the outcome measurements were based only on a reduction in frequency. Importantly, the review found that the interest in personal safety apps appears to diminish with age. In light of the abovementioned findings, to better serve the global youth population and to more robustly determine effectiveness, including over the long term, future studies should be conducted as follows: (1) increase focus on development of apps that target youth, particularly the younger demographic; (2) conduct longitudinal studies to determine long-term effectiveness; (3) conduct studies in LMICs; (4) incorporate objective outcome measures into studies; (5) develop standardized measures for evaluating the effectiveness of apps specifically geared toward youth; and (6) include outcome measures that are focused on extent or severity (in addition to frequency of occurrence).

Implications for Policy Making

Smartphone personal safety apps cannot be considered the panacea for violence against children and youth. Nonetheless, the limited studies available suggest that if personal safety apps are designed based on strong evidence, integrated appropriately into existing interventions, and used effectively, they have the potential to serve as valuable tools for personal safety and, by extension, global health. As is true for personal safety apps for the older demographic, few studies have associated evidence on effectiveness, and among those that do, numerous limitations reduce their generalizability. In addition, there was no focus on privacy and confidentiality in the included evaluation studies. Strong evidence on the effectiveness and security of personal safety apps is needed for them to be fit for integration into interventions used in school, clinical, police, community services, and other settings. As has been recommended for mHealth apps, stringent standards for providing personal safety apps should be established and incorporated into the submission processes used by app stores. In addition, it is imperative that experts in the various fields (education, health care, social security, etc) play a more central role in developing, recommending, and distributing these apps. Furthermore, systematic frameworks to facilitate the translation of personal safety apps into schools, clinical settings, etc would be required.

Conclusions

The results of this scoping review indicate that mobile personal safety apps generally seem to be effective in reducing harm to at-risk children and youth, with no associated adverse events.

Although the findings are promising, several factors limit the robustness of the evidence. Recommendations for future research to improve upon the current state of evidence and availability of effective personal safety apps for children and youth have been made, such as the development of apps that specifically target youth, undertaking studies in LMICs, conducting longitudinal studies, and incorporating objective outcome measures into studies such as the number and nature of reports of victimization to authorities and pre-post professional psychological assessments of risk for self-harm or suicide. Recommendations for app development include incorporating features such as maps, personalized action plans, mood diaries, violence prevention strategies, test-retest knowledge assessments, and validated danger assessment tools. Another recommendation is for app developers to develop companion personal safety apps that target friends of at-risk youth or include features that allow them to be customized for friends' use, in light of the willingness of youth to help their friends in risky situations. Strong evidence on the effectiveness and security of personal safety apps is needed for them to be fit for integration into interventions used in school, clinical, police, community services, and other settings. There is yet a far way to go in that regard.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist.
[DOCX File , 26 KB - mhealth_v12i1e58127_app1.docx]

Multimedia Appendix 2
The framework by Jeminiwa et al [51] modified for evaluating the quality of mobile personal safety apps for youth.
[DOCX File , 19 KB - mhealth_v12i1e58127_app2.docx]

Multimedia Appendix 3
Characteristics of the mobile personal safety apps for youth.
[DOCX File , 47 KB - mhealth_v12i1e58127_app3.docx]

Multimedia Appendix 4
Characteristics of the studies evaluating mobile personal safety apps for youth.
[DOCX File , 70 KB - mhealth_v12i1e58127_app4.docx]

Multimedia Appendix 5
Findings of the studies evaluating mobile personal safety apps for youth.
[DOCX File , 61 KB - mhealth_v12i1e58127_app5.docx]

References

1. Mercy JA, Hillis SD, Butchart A, Bellis MA, Ward CL, Fang X, et al. Interpersonal violence: global impact and paths to prevention. In: Mock CN, Nugent R, Kobusingye O, Smith KR, editors. Injury Prevention and Environmental Health. Third Edition. Washington, DC: The International Bank for Reconstruction and Development / The World Bank; 2017:71-96.
2. Xue J, Hu R, Zhang W, Zhao Y, Zhang B, Liu N, et al. Virtual reality or augmented reality as a tool for studying bystander behaviors in interpersonal violence: scoping review. J Med Internet Res 2021 Feb 15;23(2):e25322 [FREE Full text] [doi: 10.2196/25322] [Medline: 33587044]
3. Interpersonal violence prevention. Texas A and M University. URL: https://studentlife.tamu.edu/education/ivp/ [accessed 2024-05-23]
4. Maxwell L, Sanders A, Skues J, Wise L. A content analysis of personal safety apps: are they keeping us safe or making us more vulnerable? Violence Against Women 2020 Feb 11;26(2):233-248. [doi: 10.1177/1077801219832124] [Medline: 30854941]
5. Ford K, Bellis MA, Judd N, Griffith N, Hughes K. The use of mobile phone applications to enhance personal safety from interpersonal violence - an overview of available smartphone applications in the United Kingdom. BMC Public Health 2022 Jun 09;22(1):1158 [FREE Full text] [doi: 10.1186/s12889-022-13551-9] [Medline: 35681167]
6. Gabor T. Costs of crime and criminal justice responses. Public Safety Canada. 2015. URL: https://www.publicsafety.gc.ca/cnt/rsrcs/pblctns/2015-r022/2015-r022-en.pdf [accessed 2024-05-22]
7. An estimation of the economic impact of violent victimization in Canada, 2009. Government of Canada. URL: https://justice.gc.ca/eng/rp-pr/cj-jp/victim/rr14_01/p0.html [accessed 2024-05-22]

8. Waters HR, Hyder AA, Rajkotia Y, Basu S, Butchart A. The costs of interpersonal violence--an international review. Health Policy 2005 Sep 08;73(3):303-315. [doi: 10.1016/j.healthpol.2004.11.022] [Medline: 16039349]
9. Senior M, Fazel S, Tsiachristas A. The economic impact of violence perpetration in severe mental illness: a retrospective, prevalence-based analysis in England and Wales. The Lancet Public Health 2020 Feb;5(2):E99-106. [doi: 10.1016/s2468-2667(19)30245-2]
10. Basile KC, Clayton HB, DeGue S, Gilford JW, Vagi KJ, Suarez NA, et al. Interpersonal violence victimization among high school students - youth risk behavior survey, United States, 2019. MMWR Suppl 2020 Aug 21;69(1):28-37 [FREE Full text] [doi: 10.15585/mmwr.su6901a4] [Medline: 32817605]
11. Midei AJ, Matthews KA. Interpersonal violence in childhood as a risk factor for obesity: a systematic review of the literature and proposed pathways. Obesity Rev 2011 Apr 27;12(5):e159-e172. [doi: 10.1111/j.1467-789x.2010.00823.x]
12. Global status report on preventing violence against children 2020. World Health Organization. 2020. URL: https://iris.who.int/bitstream/handle/10665/332394/9789240004191-eng.pdf?sequence=1 [accessed 2024-05-29]
13. Draughon Moret J, Todd A, Rose L, Pollitt E, Anderson J. Mobile phone apps for intimate partner and sexual violence prevention and response: systematic search on app stores. JMIR Form Res 2022 Feb 08;6(2):e28959 [FREE Full text] [doi: 10.2196/28959] [Medline: 35133285]
14. Whitfield CL, Anda RF, Dube SR, Felitti VJ. Violent childhood experiences and the risk of intimate partner violence in adults: assessment in a large health maintenance organization. J Interpers Violence 2003 Feb 01;18(2):166-185. [doi: 10.1177/0886260502238733]
15. Phoenix J, Francis BJ. Police risk assessment and case outcomes in missing person investigations. Police J 2022 Apr 25;96(3):390-410. [doi: 10.1177/0032258X221087829]
16. Missing persons. Edmonton Police Service. URL: https://www.edmontonpolice.ca/CommunityPolicing/PersonalPropertyCrimes/MissingPersons [accessed 2023-10-10]
17. Armitage R. Bullying in children: impact on child health. BMJ Paediatrics Open 2021 Mar 11;5:e000939. [doi: 10.1136/bmjpo-2020-000939]
18. Glew G, Rivara F, Feudtner C. Bullying: children hurting children. Pediatr Rev 2000 Jun 1;21(6):183-190. [doi: 10.1542/pir.21.6.183]
19. Adolescent and young adult health. World Health Organization. 2024 Oct 2. URL: https://www.who.int/news-room/fact-sheets/detail/adolescents-health-risks-and-solutions [accessed 2024-10-14]
20. Adolescent health in the South-East Asia region. World Health Organization. URL: https://www.who.int/southeastasia/health-topics/adolescent-health [accessed 2024-03-21]
21. Coming of age: adolescent health. World Health Organization. URL: https://www.who.int/news-room/spotlight/coming-of-age-adolescent-health [accessed 2024-10-14]
22. Brignone L, Edleson JL. The dating and domestic violence app rubric: synthesizing clinical best practices and digital health app standards for relationship violence prevention smartphone apps. Int J Hum Comput Interact 2019 Mar 29;35(19):1859-1869. [doi: 10.1080/10447318.2019.1574100]
23. Number of smartphone users worldwide from 2014 to 2029. Statista. URL: https://www.statista.com/forecasts/1143723/smartphone-users-in-the-world [accessed 2024-05-24]
24. Number of smartphone mobile network subscriptions worldwide from 2016 to 2023, with forecasts from 2023 to 2028. Statista. URL: https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/ [accessed 2024-05-24]
25. Aguirre RT, McCoy MK, Roan M. Development guidelines from a study of suicide prevention mobile applications (apps). J Technol Human Serv 2013 Jul;31(3):269-293. [doi: 10.1080/15228835.2013.814750]
26. Zilka GC. A lways with them: smartphone use by children, adolescents, and young adults—characteristics, habits of use, sharing, and satisfaction of needs. Univ Access Inf Soc 2018 Oct 16;19(1):145-155. [doi: 10.1007/s10209-018-0635-3]
27. How many people own smartphones? (2024-2029). What's the Big Data. 2024 Jan 31. URL: https://whatsthebigdata.com/smartphone-stats/ [accessed 2024-07-01]
28. Mobile internet usage in Canada. Statista. 2024. URL: https://www.statista.com/study/43940/mobile-usage-in-canada/ [accessed 2024-10-05]
29. At what age do kids start getting smartphones? Marketing Charts. 2022 Apr 25. URL: https://www.marketingcharts.com/demographics-and-audiences/teens-and-younger-225502 [accessed 2022-04-25]
30. Anderson K, Burford O, Emmerton L. Mobile health apps to facilitate self-care: a qualitative study of user experiences. PLoS One 2016 May 23;11(5):e0156164 [FREE Full text] [doi: 10.1371/journal.pone.0156164] [Medline: 27214203]
31. Peng W, Kanthawala S, Yuan S, Hussain SA. A qualitative study of user perceptions of mobile health apps. BMC Public Health 2016 Nov 14;16(1):1158 [FREE Full text] [doi: 10.1186/s12889-016-3808-0] [Medline: 27842533]
32. Sama PR, Eapen ZJ, Weinfurt KP, Shah BR, Schulman KA. An evaluation of mobile health application tools. JMIR Mhealth Uhealth 2014 May 01;2(2):e19 [FREE Full text] [doi: 10.2196/mhealth.3088] [Medline: 25099179]
33. Vaghefi I, Tulu B. The continued use of mobile health apps: insights from a longitudinal study. JMIR Mhealth Uhealth 2019 Aug 29;7(8):e12983 [FREE Full text] [doi: 10.2196/12983] [Medline: 31469081]

34. Muntaner-Mas A, Martinez-Nicolas A, Lavie CJ, Blair SN, Ross R, Arena R, et al. A systematic review of fitness apps and their potential clinical and sports utility for objective and remote assessment of cardiorespiratory fitness. Sports Med 2019 Mar 1;49:587-600. [doi: 10.1007/s40279-019-01084-y]
35. Park M, Yoo H, Kim J, Lee J. Why do young people use fitness apps? Cognitive characteristics and app quality. Electron Commer Res 2018 Jan 29;18:755-761. [doi: 10.1007/s10660-017-9282-7]
36. West JH, Hall PC, Hanson CL, Barnes MD, Giraud-Carrier C, Barrett J. There's an app for that: content analysis of paid health and fitness apps. J Med Internet Res 2012 May 14;14(3):e72 [FREE Full text] [doi: 10.2196/jmir.1977] [Medline: 22584372]
37. Wells C, Spry C. An Overview of Smartphone Apps: CADTH Horizon Scan. Ottawa, ON: Canadian Agency for Drugs and Technologies in Health; 2022.
38. Kai OJ. Developing a real time digital emergency personal safety app to assist in emergency situations using best time. Universiti Tunku Abdul Rahman. 2019 Jan. URL: http://eprints.utar.edu.my/3443/1/fyp_IA_2019_OJK_1504447.pdf [accessed 2023-08-15]
39. Doria N, Ausman C, Wilson S, Consalvo A, Sinno J, Boulos L, et al. Women's experiences of safety apps for sexualized violence: a narrative scoping review. BMC Public Health 2021 Dec 30;21:2330. [doi: 10.1186/s12889-021-12292-5]
40. Kanagaraj SA, Arjun G, Shahina A. Cheeka: a mobile application for personal safety. In: Proceedings of the IEEE International Conference on Collaborative Computing: Networking, Applications and Worksharing. 2013 Presented at: CollaborateCom 2013; October 20-23, 2013; Austin, TX. [doi: 10.4108/icst.collaboratecom.2013.254079]
41. McGrath S. Mobile apps can enhance personal safety efforts. Stud Aff Today 2016 Mar 23;19(1):6. [doi: 10.1002/say.30186]
42. Oksiutycz A, Lubinga E. Factors affecting the adoption of personal safety apps among millennials in Johannesburg, South Africa. S Afr J Inf Manag 2021;23(1):a1246. [doi: 10.4102/sajim.v23i1.1246]
43. O'Campo P, Velonis A, Buhariwala P, Kamalanathan J, Hassan MA, Metheny N. Design and development of a suite of intimate partner violence screening and safety planning web apps: user-centered approach. J Med Internet Res 2021 Dec 21;23(12):e24114 [FREE Full text] [doi: 10.2196/24114] [Medline: 34931998]
44. Sumra M, Asghar S, Khan KS, Fernández-Luna JM, Huete JF, Bueno-Cavanillas A. Smartphone apps for domestic violence prevention: a systematic review. Int J Environ Res Public Health 2023 Mar 23;20(7):5246. [doi: 10.3390/ijerph20075246]
45. Eisenhut K, Sauerborn E, García-Moreno C, Wild V. Mobile applications addressing violence against women: a systematic review. BMJ Glob Health 2020 Apr 22;5(4):e001954. [doi: 10.1136/bmjgh-2019-001954]
46. Sinha S, Shrivastava A, Paradis C. A survey of the mobile phone-based interventions for violence prevention among women. Adv Soc Work 2020 Mar 25;19(2):493-517. [doi: 10.18060/22526]
47. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
48. Mak S, Thomas A. Steps for conducting a scoping review. J Grad Med Educ 2022;14(5):565-567. [doi: 10.4300/jgme-d-22-00621.1]
49. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Sep 04;169(7):467. [doi: 10.7326/M18-0850]
50. Etzion D, Romi S. Typology of youth at risk. Child Youth Serv Rev 2015 Dec;59:184-195. [doi: 10.1016/j.childyouth.2015.10.017]
51. Jeminiwa RN, Hohmann NS, Fox BI. Developing a theoretical framework for evaluating the quality of mHealth apps for adolescent users: a systematic review. J Pediatr Pharmacol Ther 2019 Jul;24(4):254-269 [FREE Full text] [doi: 10.5863/1551-6776-24.4.254] [Medline: 31337988]
52. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
53. Terhorst Y, Philippi P, Sander LB, Schultchen D, Paganini S, Bardus M, et al. Validation of the Mobile Application Rating Scale (MARS). PLoS One 2020 Nov 2;15(11):e0241480 [FREE Full text] [doi: 10.1371/journal.pone.0241480] [Medline: 33137123]
54. Grist R, Porter J, Stallard P. Acceptability, use, and safety of a mobile phone app (BlueIce) for young people who self-harm: qualitative study of service users' experience. JMIR Ment Health 2018 Feb 23;5(1):e16 [FREE Full text] [doi: 10.2196/mental.8779] [Medline: 29475823]
55. Stallard P, Porter J, Grist R. A smartphone app (BlueIce) for young people who self-harm: open phase 1 pre-post trial. JMIR Mhealth Uhealth 2018 Jan 30;6(1):e32 [FREE Full text] [doi: 10.2196/mhealth.8917] [Medline: 29382632]
56. Wood MA, Ross S, Johns D. Primary crime prevention apps: a typology and scoping review. Trauma Violence Abuse 2021 Jan 22;23(4):1093-1110. [doi: 10.1177/1524838020985560]
57. Navarro-Pérez JJ, Oliver A, Carbonell Á, Schneider BH. Effectiveness of a mobile app intervention to prevent dating violence in residential child care. Psychosoc Interv 2020;29(2):59-66. [doi: 10.5093/pi2020a3]
58. Liad@s of the University allows for reduce between 6% and 25% the sexism in teenagers. Universitat de València. 2017 Nov 24. URL: https://www.uv.es/uvweb/college/en/profile/

liad-s-university-allows-reduce-6-25-sexism-teenagers-1285950309813/Novetat.html?id=1286026137921&plantilla=UV/Page/TPGDetaill [accessed 2023-03-21]
59. Glass NE, Clough A, Messing JT, Bloom T, Brown ML, Eden KB, et al. Longitudinal impact of the myPlan app on health and safety among college women experiencing partner violence. J Interpers Violence 2022 Jul 12;37(13-14):NP11436-NP11459. [doi: 10.1177/0886260521991880] [Medline: 33576291]
60. Ouellet-Morin I, Robitaille MP. Stronger than bullying, a mobile application for victims of bullying: development and initial steps toward validation. In: Campbell M, Bauman S, editors. Reducing Cyberbullying in Schools: International Evidence-Based Best Practices. Cambridge, MA: Academic Press; 2018:159-174.
61. Potter SJ, Moschella EA, Demers JM, Lynch M. Using mobile technology to enhance college sexual violence response, prevention, and risk reduction efforts. J Technol Human Serv 2021 Jun 01;40(1):25-46. [doi: 10.1080/15228835.2021.1929665]
62. Sheoran B, Silva CL, Lykens JE, Gamedze L, Williams S, Ford JV, et al. YTH StreetConnect: development and usability of a mobile app for homeless and unstably housed youth. JMIR Mhealth Uhealth 2016 Jul 14;4(3):e82 [FREE Full text] [doi: 10.2196/mhealth.5168] [Medline: 27417882]
63. Blayney JA, Jenzer T, Read JP, Livingston JA, Testa M. Enlisting friends to reduce sexual victimization risk: there's an app for that… but nobody uses it. J Am Coll Health 2018;66(8):767-773. [doi: 10.1080/07448481.2018.1446439] [Medline: 29488831]
64. Navarro-Pérez JJ, Carbonell Á, Oliver A. The effectiveness of a psycho-educational app to reduce sexist attitudes in adolescents. Revista de Psicodidáctica (English ed.) 2019;24(1):9-16. [doi: 10.1016/j.psicoe.2018.07.002]
65. Alhusen J, Bloom T, Clough A, Glass N. Development of the MyPlan safety decision app with friends of college women in abusive dating relationships. J Technol Hum Serv 2015 Aug 12;33(3):263-282. [doi: 10.1080/15228835.2015.1037414]
66. Lindsay M, Messing JT, Thaller J, Baldwin A, Clough A, Bloom T, et al. Survivor feedback on a safety decision aid smartphone application for college-age women in abusive relationships. J Technol Hum Serv 2013 Oct;31(4):368-388. [doi: 10.1080/15228835.2013.861784]
67. Empowering decisions for a safe path forward. myPlan. URL: https://myplanapp.org [accessed 2023-08-15]
68. uSafeUS mobile app homepage. uSafeUS Mobile App. URL: https://usafeus.org/ [accessed 2023-08-15]
69. +Fort: stronger than bullying. Google Play. URL: https://play.google.com/store/apps/details?id=com.sidekick.ciusss.plusfort&hl=en_CA&gl=US&pli=1 [accessed 2023-08-15]
70. Potter SJ, Moschella-Smith EA, Lynch M. Building a high school violence prevention app to educate and protect students. J Res Technol Educ 2022 Aug 12;56(2):110-128. [doi: 10.1080/15391523.2022.2110336]
71. uSafeHS. URL: https://www.usafeapp.org [accessed 2023-08-15]
72. BlueIce app. Oxford Health NHS Foundation Trust. URL: https://www.oxfordhealth.nhs.uk/blueice/ [accessed 2024-05-03]
73. Stallard P, Porter J, Grist R. Safety, acceptability, and use of a smartphone app, BlueIce, for young people who self-harm: protocol for an open phase I trial. JMIR Res Protoc 2016 Nov 16;5(4):e217 [FREE Full text] [doi: 10.2196/resprot.6525] [Medline: 27852560]
74. Tighe J, Shand F, McKay K, Mcalister TJ, Mackinnon A, Christensen H. Usage and acceptability of the iBobbly app: pilot trial for suicide prevention in aboriginal and Torres Strait islander youth. JMIR Ment Health 2020 Dec 01;7(12):e14296 [FREE Full text] [doi: 10.2196/14296] [Medline: 33258782]
75. Digital tools and apps. Black Dog Institute. URL: https://www.blackdoginstitute.org.au/resources-support/digital-tools-apps/ibobbly/ [accessed 2023-08-21]
76. Tighe J, Shand F, Ridani R, Mackinnon A, De La Mata N, Christensen H. Ibobbly mobile health intervention for suicide prevention in Australian indigenous youth: a pilot randomised controlled trial. BMJ Open 2017;7:e013518. [doi: 10.1136/bmjopen-2016-013518]
77. Tighe J, Shand F, Christensen H. The iBobbly app suicide prevention Kimberley trial 2013 – 2015 community report. Black Dog Institute. URL: https://www.blackdoginstitute.org.au/wp-content/uploads/2020/04/17-08-09_res_ibobblysummarydocument_website_final.pdf?sfvrsn=2 [accessed 2023-08-20]
78. Povey J, Mills PP, Dingwall KM, Lowell A, Singer J, Rotumah D, et al. Acceptability of mental health apps for Aboriginal and Torres Strait Islander Australians: a qualitative study. J Med Internet Res 2016 Mar 11;18(3):e65 [FREE Full text] [doi: 10.2196/jmir.5314] [Medline: 26969043]
79. Debnam KJ, Kumodzi T. Adolescent perceptions of an interactive mobile application to respond to teen dating violence. J Interpers Violence 2019 Jan 02;36(13-14):6821-6837. [doi: 10.1177/0886260518821455]
80. Henry A. Circle of 6 for iPhone and Android prevents violence, gives you a way out of dangerous situations. Lifehacker. 2012 Sep 07. URL: https://lifehacker.com/circle-of-6-for-iphone-and-android-prevents-violence-g-5941286 [accessed 2023-08-14]
81. +Fort: stronger than bullying. Axel. URL: https://centreaxel.com/en/projects/stronger-than-bullying/ [accessed 2023-08-15]
82. Greenhalgh I, Tingley J, Taylor G, Medina-Lara A, Rhodes S, Stallard P. Beating Adolescent Self-Harm (BASH): a randomised controlled trial comparing usual care versus usual care plus a smartphone self-harm prevention app (BlueIce) in young adolescents aged 12–17 who self-harm: study protocol. BMJ Open 2021;11:e049859. [doi: 10.1136/bmjopen-2021-049859]

83. Smartphone app BlueIce could help reduce self-harm. The National. 2019 Aug 1. URL: https://www.thenational.scot/news/17807482.smartphone-app-blueice-help-reduce-self-harm/ [accessed 2024-03-05]
84. Tingley J, Greenhalgh I, Stallard P. Technology matters: BlueIce – using a smartphone app to beat adolescent self‐harm. Child Adolesc Ment Health 2020 Jun 11;25(3):192-194. [doi: 10.1111/camh.12397]
85. 'Prescribed' smartphone app offers hope to young people who self-harm. University of Bath. URL: https://www.bath.ac.uk/case-studies/prescribed-smartphone-app-offers-hope-to-young-people-who-self-harm/ [accessed 2023-08-15]
86. iBobbly: a social and emotional wellbeing app for aboriginal and Torres Strait Islander peoples. Black Dog Institute. URL: https://www.blackdoginstitute.org.au/wp-content/uploads/2020/08/iBobbly-brochure.pdf [accessed 2024-03-05]
87. Shand FL, Ridani R, Tighe J, Christensen H. The effectiveness of a suicide prevention app for indigenous Australian youths: study protocol for a randomized controlled trial. Trials 2013 Nov 20;14:396 [FREE Full text] [doi: 10.1186/1745-6215-14-396] [Medline: 24257410]
88. Circle of 6. Devpost. URL: https://devpost.com/software/circle-of-6 [accessed 2024-05-02]
89. Past project: circle of 6. Youth Tech Health. URL: https://yth.org/projects/circle-of-6/ [accessed 2023-03-21]
90. Liad@s. Google Play. URL: https://play.google.com/store/apps/details?id=es.uv.artec.Liados&hl=en_GB&pli=1 [accessed 2024-03-05]
91. +Fort: stronger than bullying. Google Play. URL: https://play.google.com/store/apps/details?id=com.sidekick.ciusss.plusfort&hl=en_CA&gl=US&pli=1 [accessed 2023-08-15]
92. iBobbly [Discontinued]. e-Mental Health in Practice. URL: https://www.emhprac.org.au/directory/ibobbly/ [accessed 2023-11-10]
93. Potter SJ, Moschella EA, Smith D, Draper N. Exploring the usage of a violence prevention and response app among community college students. Health Educ Behav 2020 May 26;47(1_suppl):44S-53S. [doi: 10.1177/1090198120910995]
94. Abrahams N, Devries K, Watts C, Pallitto C, Petzold M, Shamu S, et al. Worldwide prevalence of non-partner sexual violence: a systematic review. The Lancet 2014 May 10;383(9929):1648-1654. [doi: 10.1016/s0140-6736(13)62243-6]
95. Borumandnia N, Khadembashi N, Tabatabaei M, Alavi Majd H. The prevalence rate of sexual violence worldwide: a trend analysis. BMC Public Health 2020 Nov 30;20:1835. [doi: 10.1186/s12889-020-09926-5]
96. COVID-19 Mental Disorders Collaborators. Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. Lancet 2021 Nov 06;398(10312):1700-1712 [FREE Full text] [doi: 10.1016/S0140-6736(21)02143-7] [Medline: 34634250]
97. van Dijk J, Nieuwbeerta P, Joudo Larsen J. Global crime patterns: an analysis of survey data from 166 countries around the world, 2006–2019. J Quant Criminol 2021 Mar 20;38:793-827. [doi: 10.1007/s10940-021-09501-0]
98. Ding X, Wuerth K, Sakakibara B, Schmidt J, Parde N, Holsti L, et al. Understanding mobile health and youth mental health: scoping review. JMIR Mhealth Uhealth 2023 Jun 16;11:e44951 [FREE Full text] [doi: 10.2196/44951] [Medline: 37220197]
99. Madonsela S, Ware LJ, Scott M, Watermeyer J. The development and use of adolescent mobile mental health (m-mhealth) interventions in low- and middle-income countries: a scoping review. S Afr J Psychol 2023 Jul 08;53(4):471-483. [doi: 10.1177/00812463231186260]
100. Decker MR, Wood SN, Hameeduddin Z, Kennedy SR, Perrin N, Tallam C, et al. Safety decision-making and planning mobile app for intimate partner violence prevention and response: randomised controlled trial in Kenya. BMJ Glob Health 2020;5:e002091. [doi: 10.1136/bmjgh-2019-002091]
101. Malloy JA, Partridge SR, Kemper JA, Braakhuis A, Roy R. Co-design of digital health interventions for young adults: protocol for a scoping review. JMIR Res Protoc 2022 Oct 24;11(10):e38635 [FREE Full text] [doi: 10.2196/38635] [Medline: 36279167]

Abbreviations

Co6: Circle of 6
IPV: intimate partner violence
LGBTQ: lesbian, gay, bisexual, transgender, queer
LMIC: low- and middle-income country
mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews
RCT: randomized controlled trial

Edited by L Buis; submitted 07.03.24; peer-reviewed by C Emezue, A Hosny, XX Wang, S Price; comments to author 15.05.24; revised version received 26.06.24; accepted 23.09.24; published 05.11.24.

Please cite as:
Bowen-Forbes C, Khondaker T, Stafinski T, Hadizadeh M, Menon D
Mobile Apps for the Personal Safety of At-Risk Children and Youth: Scoping Review
JMIR Mhealth Uhealth 2024;12:e58127
URL: https://mhealth.jmir.org/2024/1/e58127
doi:10.2196/58127
PMID:

Evaluating the Impact of a Daylight-Simulating Luminaire on Mood, Agitation, Rest-Activity Patterns, and Social Well-Being Parameters in a Care Home for People With Dementia: Cohort Study

Kate Turley[1], BSc; Joseph Rafferty[1], PhD; Raymond Bond[1], PhD; Maurice Mulvenna[1], PhD; Assumpta Ryan[2], PhD; Lloyd Crawford[3]

[1]School of Computing, Ulster University, Cromore Rd, Belfast, United Kingdom

[2]School of Nursing and Pandemic Science, Ulster University, Belfast, United Kingdom

[3]Chroma Lighting, Belfast, United Kingdom

Corresponding Author:
Kate Turley, BSc
School of Computing, Ulster University, , Cromore Rd, Belfast, , United Kingdom

Abstract

Background: Living with a diagnosis of dementia can involve managing certain behavioral and psychological symptoms. Alongside cognitive decline, this cohort expresses a suppression in melatonin production which can negatively influence their alignment of sleep or wake timings with the 24 hour day and night cycle. As a result, their circadian rhythms become disrupted. Since daylight has the capacity to stimulate the circadian rhythm and humans spend approximately 90% of their time indoors, research has shifted toward the use of indoor lighting to achieve this same effect. This type of lighting is programmed in a daylight-simulating manner; mimicking the spectral changes of the sun throughout the day. As such, this paper focuses on the use of a dynamic lighting and sensing technology used to support the circadian rhythm, behavioral and psychological symptoms, and well-being of people living with dementia.

Objective: This study aimed to understand how dynamic lighting, as opposed to static lighting, may impact the well-being of those who are living with dementia.

Methods: An ethically approved trial was conducted within a care home for people with dementia. Data were collected in both quantitative and qualitative formats using environmentally deployed radar sensing technology and the validated QUALIDEM (Quality of Life for People With Dementia) well-being scale, respectively. An initial 4 weeks of static baseline lighting was used before switching out for 12 weeks of dynamic lighting. Metrics were collected for 11 participants on mood, social interactions, agitation, sense of feeling, and sleep and rest-activity over a period of 16 weeks.

Results: Dynamic lighting showed significant improvement with a moderate effect size in well-being parameters including positive affect (P=.03), social isolation (P=.048), and feeling at home (P=.047) after 5 - 10 weeks of dynamic lighting exposure. The results also highlight statistically significant improvements in rest-activity–related parameters of interdaily stability (P<.001), intradaily variation (P<.001), and relative amplitude (P=.03) from baseline to weeks 5 - 10, with the effect propagating for interdaily stability at weeks 10 - 16 as well (P<.001). Nonsignificant improvements are also noted for sleep metrics with a small effect size; however, the affect in agitation does not reflect this improvement.

Conclusions: Dynamic lighting has the potential to support well-being in dementia, with seemingly stronger influence in earlier weeks where the dynamic lighting initially follows the static lighting contrast, before proceeding to aggregate as marginal gains over time. Future longitudinal studies are recommended to assess the additional impact that varying daylight availability throughout the year may have on the measured parameters.

(JMIR Mhealth Uhealth 2024;12:e56951) doi:10.2196/56951

KEYWORDS

digital health; dementia; dynamic lighting; sensors; circadian rhythm; daylight; wellbeing; mood; agitation; sleep; social wellbeing; care home; older adults; elderly; cardiac; psychological; monitoring

Introduction

Experiencing life with an atypical body clock can often materialize as a result of living with dementia. In fact, studies have shown that the relationship between atypical body clocks and dementia may even be bidirectional [1]. This means that the disruption of one's body clock may catalyze their progression of dementia and vice versa. The combined effect of this can lead to disjointed rest-activity patterns and poor sleep quality, which then impacts mood and well-being in the longer term [2,3].

At present, there are over 55 million people living with dementia around the world [4]. This is 55 million people living without a cure for their diagnosis. As a result of this, society places great importance on research which can either predict the diagnosis in advance or alleviate the symptoms after the fact in a preventative versus reactionary approach [5].

This paper focuses on the latter approach, exploring means to alleviate symptoms of dementia to improve well-being. To expand the above content, a disrupted body clock (aka circadian rhythm) is a common factor of natural ageing, and is heightened in people living with dementia [6]. This disruption lends itself to poor sleep quality. In turn, this misaligns the typical timing of one's sleep and wake cycle with the timing of sunset and sunrise which acts as our inherent, evolutionary alarm clock. Alongside this 24 hour cycle of light and dark, our bodies experience a diurnal cycle of the hormones melatonin and cortisol [7]. Melatonin is a hormone which encourages relaxation and sleep while cortisol is commonly named the "stress hormone" and governs the alerting response in the body. In a typical body clock aligned with sunlight cues, melatonin levels will begin to rise in the evening hours at the onset of darkness, and cortisol levels peak in the afternoon when sunlight is wholly available [8]. With dementia, the disrupted body clock suppresses this cycle of melatonin to cortisol and there becomes a discrepancy between the signals within the body, which should control our sleep-wake rhythm at these predefined times of the day. When sleep and hormone cycles are not in synchronization, the interlocking of these mechanisms becomes chaotic and thus a disrupted circadian rhythm and its symptoms ensue [8]. Since the circadian rhythm is responsible for controlling mood, body temperature, appetite, sleep-wake cycles, rest-activity levels among other factors, ensuring its synchronization is essential for supporting well-being for people living with dementia.

Objectively, it is known that light is the strongest zeitgeber (german for time-giver) for the human body and its circadian rhythm [9]. The circadian rhythm is governed from the suprachiasmatic nucleus in the brain, which is accessed via pathways from the eye. Within the eye, there are rods and cones which receive lighting spectra and translate this into a visual response. However, the discovery of the intrinsically photosensitive retinal ganglion cells (ipRGCs) in 2002 supported the theory that light delivers both photopic and melanopic contributions within human systems [10]. In other words, light is not only necessary for visual purposes (photopic) but also for aligning the circadian rhythm (melanopic). These ipRGCs exhibit varying sensitivity to different wavelengths of light, peaking at 480 nanometers (blue colored light) [10]. This means that the incorporation of dynamic lighting as opposed to static lighting may be critical for attuning the circadian rhythm with the analogous varying properties of daylight with which we have evolved.

Although, however inherent the need for lighting is to align the human circadian rhythm, the extent to which this applies in dementia is largely uncertain [11,12]. Typically, this is because the nature of the disease varies largely on factors such as the state of progression of dementia, alongside different factors that affect the response to lighting: chronotype, age, gender, and previous lighting exposures [13]. Additionally, the methods for obtaining the amount of lighting one is exposed to are commonly not accounted for, and the units of the lighting spectra trialled are far from systematic [14]. This makes comparison of lighting trials and studies within dementia cohorts difficult, resulting in a critical gap in knowledge within this field. This study aims to address some of these gaps by monitoring aspects of well-being and correlating this with commonly reported lighting metrics.

In order to monitor the above, it is critical to collect information on the daily activity and well-being of people living with dementia [15]. As such, this study makes use of integrated, environmentally deployed sensors to enrich this dataset in a nonintrusive manner. This setup allows for the formation of a technology, which can deliver circadian-aligned lighting and simultaneously monitor any resultant changes to well-being. Additionally, validated questionnaires are used to collect information on certain behavioral symptoms, which the sensors cannot themselves generate. Collaboratively, this information will provide informatics in response to dynamic lighting systems for people living with dementia. This study therefore hypothesizes that all measured parameters of well-being will benefit from exposure to dynamic lighting as opposed to static lighting over time.

Methods

Study Design

A combined lighting and sensing intervention was carried out from March to July 2023 for a period of 16 weeks, whereby participants with dementia experienced a change in their received lighting output. The technology is designed to deliver dynamic lighting to support the circadian rhythm of people living with dementia.

Simultaneously, it is designed to monitor the change to certain aspects of their circadian rhythm and well-being that are expected to be impacted by changes in the lighting spectra. The trial invoked a baseline-intervention strategy on the same group of individuals to obtain repeated measures at several key points throughout this study.

Ethical Considerations

This study has been approved by the Office for Research Ethics Committees Northern Ireland under Integrated Research Application System ID 311547. Recruitment of participants was managed via the care home manager acting as gatekeeper. With consideration of the inclusion criteria, a dementia-friendly information leaflet was circulated by the care home manager to

the residents and family members of people living with dementia. A similar leaflet was delivered to the care staff. Those who were interested in participating were invited to a presentation afternoon in the care home where 2 members of the research team brought the luminaire or sensor prototype to the care home and explained this study's requirements for all involved parties.

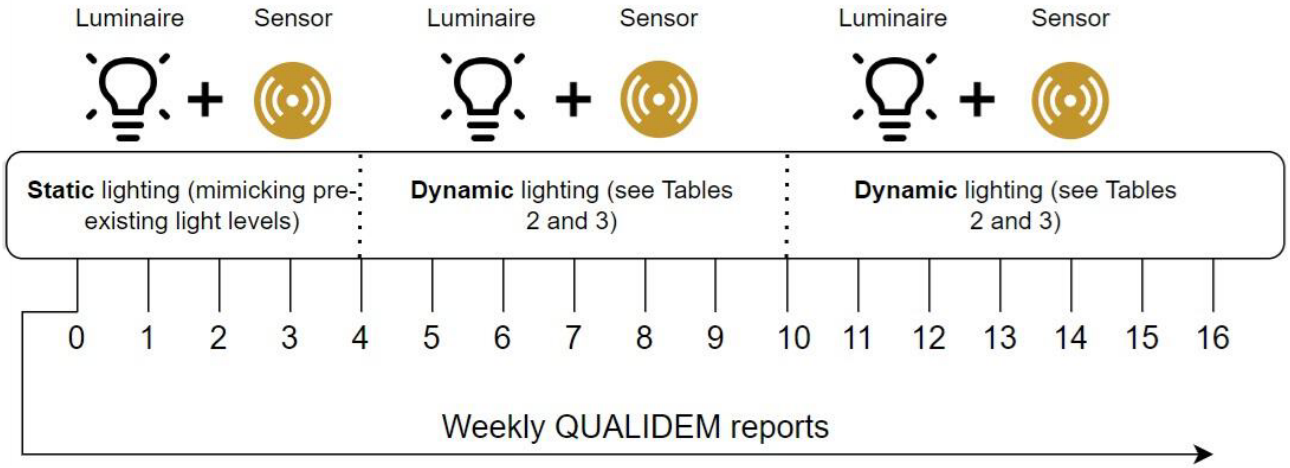
Care staff were shown the questionnaire and user guide and offered to take part. After this, for those who were willing, initial consent was obtained and reaffirmed verbally throughout this study. In addition to care staff providing informed consent for their contribution to questionnaires, family members were asked to provide proxy consent on behalf of the selected residents recruited by the care home manager. For the residents living in this unit, all of their families possessed the legal right to consent to matters on their behalf that relate to their everyday life and well-being. This was communicated to the researchers by the care home manager. Since all participants had moderate-to-severe stages of progression of dementia and the fact that this was a pilot study requiring only passive interactions with the technology, it was deemed appropriate by family members, researchers, and the care home manager to gain initial consent via proxy at this stage of this study. A paper copy of a consent form was given to the participant's proxy to be signed. If consent has been given, the participants data will be collected and analyzed by the research team. If at any point a participant proxy retracts their consent or the resident with dementia indicates they do not wish to be part of this study (verbally or otherwise), the data up until that point will have been collected and analyzed, and any data collected after this may be stored but not used for analysis, or exempt from collection altogether if requested. Proxies for participants were informed should they wish, to opt in and opt out at any point by informing the care manager, care staff, or chief investigator.

Study Protocol

This study was implemented using the following protocol as seen in Figure 1.

- Weeks 1 - 4 : baseline lighting and static lighting
- Weeks 5 - 16: dynamic lighting

Figure 1. Baseline lighting was static in nature and was used as a control for 4 weeks. Dynamic lighting was introduced from weeks 5-16 inclusive. Sensor metrics were collected 24/7 for all 16 weeks with QUALIDEM reports completed on a weekly basis. QUALIDEM: Quality of Life for People With Dementia.



The baseline lighting was used as a control for 4 weeks to generate differences in well-being attributed to the dynamic lighting. After these 4 weeks, the lighting program was changed to exhibit the dynamic lighting spectra. The period of 12 weeks to trial the dynamic lighting was chosen due to the fact that it is estimated to take approximately 2 - 5 weeks to realign a circadian rhythm, therefore providing ample opportunity to witness potential progression over time [16]. As such, a baseline (weeks 1 - 4), midpoint (weeks 5 - 10), and end of intervention (weeks 10 - 16) approach for analysis was used on the results to observe the progression in well-being throughout the trial. Alongside the sensing metrics collected, the trial required insight into additional well-being parameters that could not be collected via the sensors. As such, a validated well-being questionnaire (QUALIDEM) was also completed by care staff on a weekly basis [17]. The QUALIDEM scale represents the "Quality of Life for People With Dementia" and is a validated measure of quality of life and well-being for people living with dementia [17]. It is used to understand the behavior and attitude of people living with dementia on a weekly basis. The scale measures 10 aspects of quality of life; (A) care relationship, (B) positive affect, (C) negative affect, (D) restless tense behavior, (E) positive self image, (F) social relationships, (G) social isolation, (H) feeling at home, (I) occupation, and (J) other questions for further research. For this study, the latter two parameters (I and J) are presented within the report but excluded from analysis as these are not factors anticipated to be impacted by the circadian rhythm in response to lighting. The QUALIDEM scale is to be completed by care staff on behalf of the residents with dementia. Although there may be unintentional bias on reporting scores, it was deemed suitable for multiple care staff to complete the QUALIDEM scales due to its high Inter Rater Reliability score [18]. This also facilitated the reduction in caregiver workload and burden during the week.

Each resident was free to enjoy their daily routines normally and no constraints on their activity was requested. The sole hypothesis tested in the trial was the changes to various aspects

of well-being under exposure to static versus dynamic lighting spectra.

Hardware and Architecture

The bespoke device deployed within the care home can be seen in Figure 2. It has been designed to integrate both lighting and sensing components of the technology in a feedback setup. The lighting output is produced by the dimmable, tuneable-white LED boards which can change in intensity and color temperature. The integrated sensor is a radar sensor, which can track movements on a frame-by-frame basis at fine-grained resolution. The sensor and luminaire have bidirectional communication pathways with a cloud storage, enabling data-driven feedback to inform the lighting. This means that the luminaire is bespoke in the fact that it has the capacity to actuate changes in the lighting spectra based on data-driven insights. For example, if a resident is observed to frequent their bed during the day for a nap, it could be possible to create a more "alerting" lighting environment (more blue output at higher intensities) to help aid their wakefulness in the day and resultantly promote their sleep during the night. However it should be noted that this pilot study is comparing a predetermined dynamic spectra (consistent with the daylight spectra) to a static lighting spectra in order to generate an initial foundation for how lighting impacts well-being for people living with dementia in order to ensure that the future data-driven actuation will be suitable for each individual.

The device communicates via Bluetooth protocol to the next device or node in the network, creating a mesh network of the lighting and sensing devices. This mesh then interacts with an edge node operating as a Bluetooth low energy gateway and transferring data to a 3rd party cloud platform. From here, the data is accessed via WebSockets (IETF) and consumed by the backend logic for processing and visualization. The entire architecture can be seen in Figure 3.

The data collected include both lighting and sensor data harvested from the network, in 15 minute intervals and real time, respectively. Due to the range of Bluetooth protocol, several repeater nodes had to be included at regular distances to bridge the communication between devices. Figure 3 demonstrates the Internet of Things architecture for the home implementation. The real time sensor data is filtered and consumed by the metric generator. This generator invokes algorithms which search for the activity levels for the previous 15-minute interval alongside the sleep-wake metrics for each day and night (see further detail in the Sensing Component section). In addition, the live location is also assessed and reported to the influx database for access on the visualization platform.

Figure 2. Lighting and sensing device. Device includes (A) diffuser, (B) tuneable white LED quadrants, and (C) radar sensor mounted on a base plate.

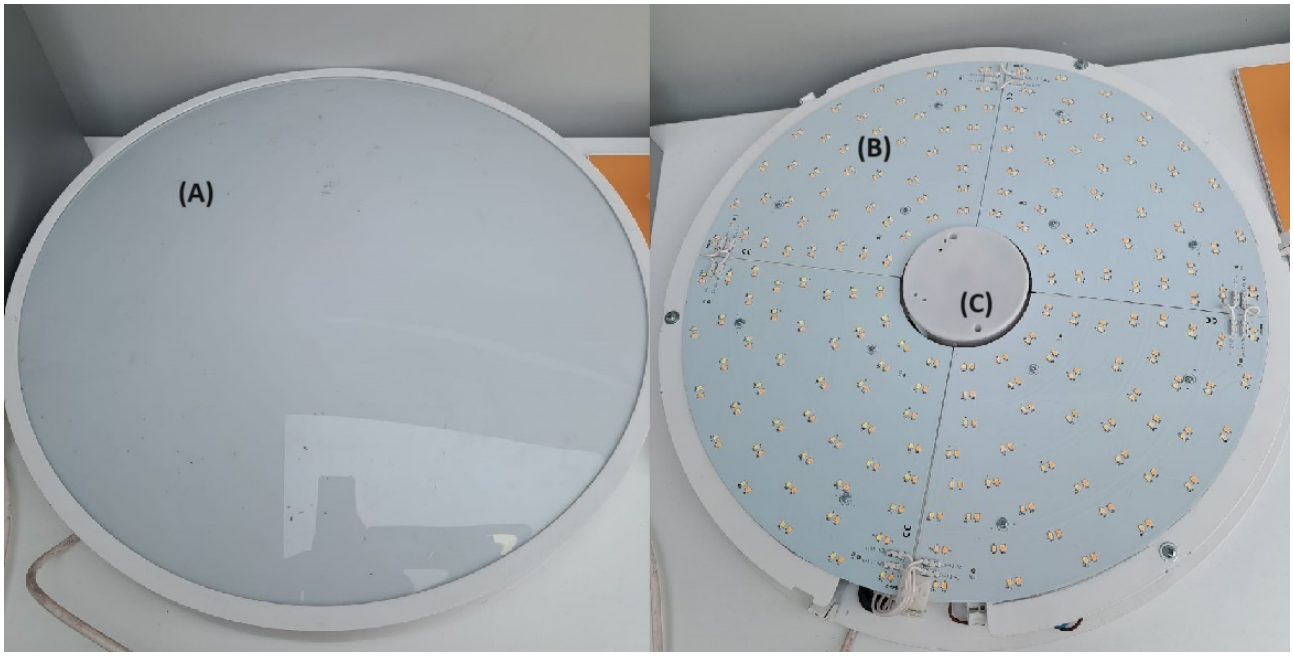
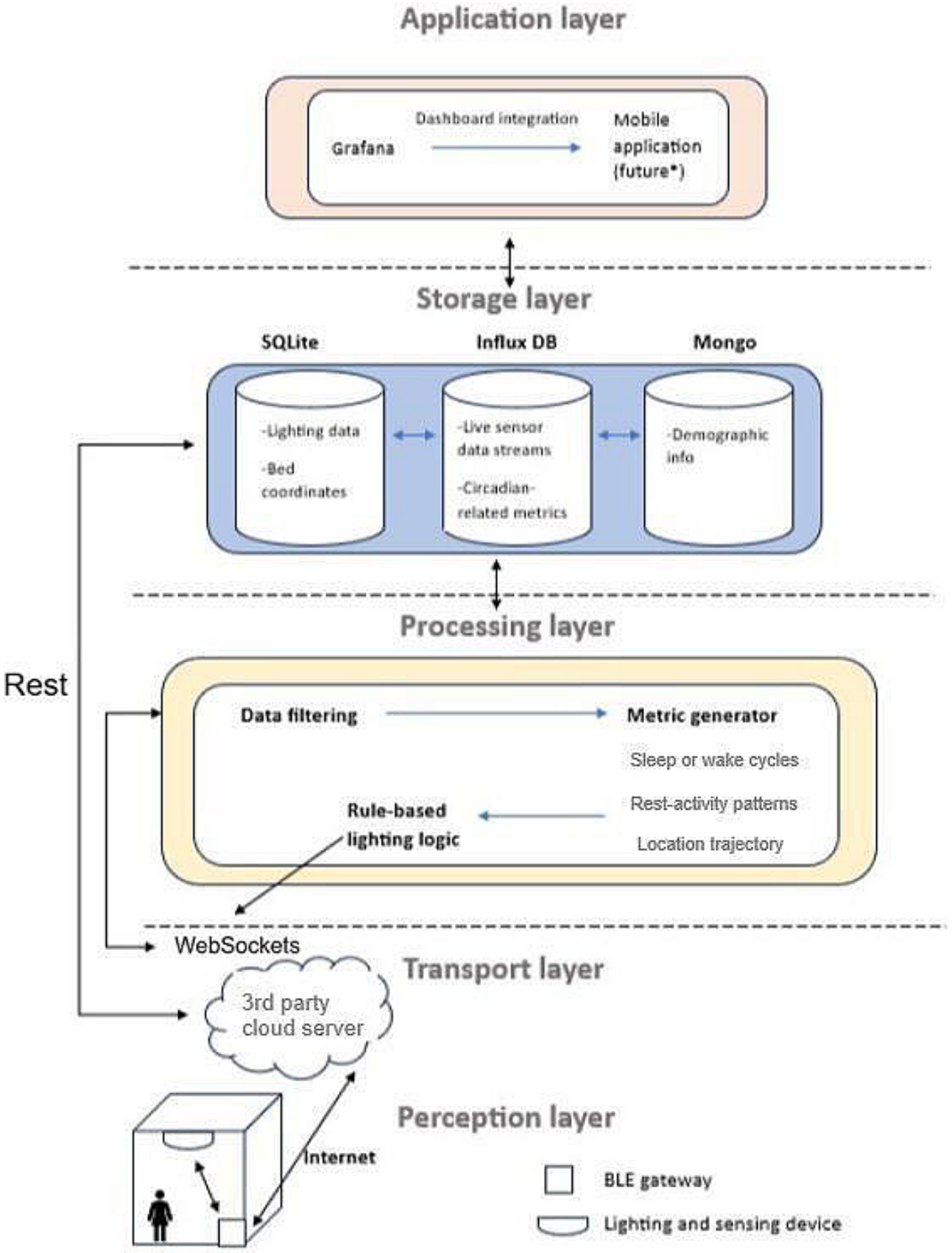


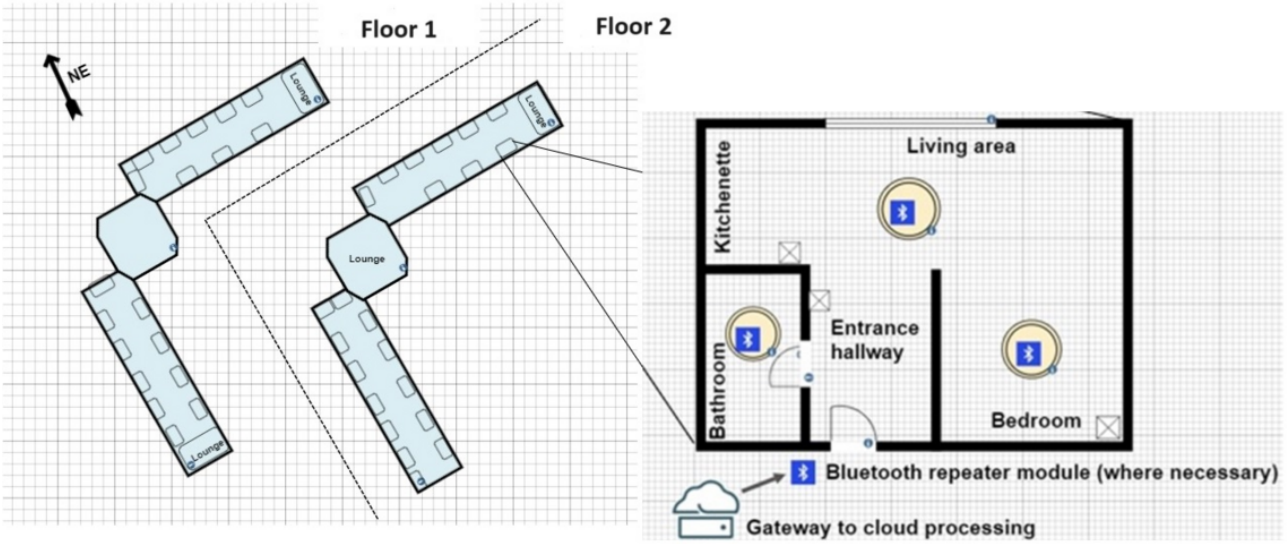
Figure 3. IoT architecture of smart care environment. Data is transferred across the network to storage and accessed by algorithms performing metric generation for visualization. BLE: Bluetooth low energy; DB: database; IoT: internet of Things.



Site

The technology was trialled in a care home within the United Kingdom. The site accommodates up to 42 residents with a dedicated wing for dementia care. Various rooms were included in this study between the upper and lower corridors, totaling 11 rooms. Five rooms faced northeast, 3 rooms faced southwest, 2 rooms faced northwest, and 1 room faced southeast. Each room is an independent living quarter for every person with dementia, with a shared common area for each wing where one can dine and watch television, as shown in Figure 4. All rooms experienced a similar deployment, with a device in each of the kitchenette and living area, bedroom, and bathrooms. The common areas were also equipped with several dynamic lighting devices but with the sensors removed to compensate for those present within these areas who had not consented to data collection.

Figure 4. Floor plan of the care home. Individual flats contained a kitchenette, living area, bathroom and bedroom with access to communal lounges or common areas in each wing. Further, 3 lighting and sensing devices were placed in each flat. NE: northeast.



Participants

Eleven residents participated in the intervention as seen in Table 1; all with a diagnosis of dementia. Each individual's type of dementia and Mini-Mental State Examination was not logged, however all participants were known to exhibit a moderate-severe state of progression. Participants were included based on the criteria that they reside at the care home and had a diagnosis of dementia. Participants were excluded on the basis that they experience schizophrenia, epilepsy, restless leg syndrome, serious eye disease such as retinitis pigmentosa or cataracts or blindness, mania, historical head injuries, diabetes, or any physical disability which would restrict the participant to their bed. These criteria were assessed by the care home manager and relevant participants selected for recruitment. Upon recruitment, 13 participants were identified and consented to this study. Eleven participants were included in the final trial due to the layout of the building and restriction in the Bluetooth range between large distances with concrete walls between. During this study, participants could live their lives as they would prior to the lighting trial. The care home has a scheduled time for breakfast, lunch, and dinner with the option to have a staggered start to accommodate all residents; aside from this, residents are given autonomy to settle in any area of the care home wing including a connected outdoor or indoor walkway, their flatlets, common areas, or their neighbor's flatlets. At night, residents are encouraged by care staff back to their flatlets and into bed and are checked on in 2 hourly intervals.

Table . Participant details.

Variable		Details
Participants, n		11
Female, n		11
Male, n		0
Age (years), mean (SD)		87.45 (2.05)
Weight (kg), mean (SD)		57.45 (1.72)
Medications, n		
	Diazepam (anxiety)	2
	Memantine (dementia)	9
	Risperidone (antipsychotic)	6
	Donepezil (dementia)	4
	Circadin (sleep)	1
	Sertraline (antidepressant)	4
	Mirtazapine (depression and anxiety)	1
	Melatonin (sleep)	1
	Citalopram (low mood)	1

Lighting Component

First, the dynamic lighting was designed to align with the photopic lighting requirements known for the ageing population from existing lighting guidance [19]. These are typically reported as illuminance at eye level measured in lux alongside measurements for the correlated color temperature of the lighting reported in Kelvin. Second, the lighting was required to be of significant spectra to ensure a melanopic response would be invoked [14,20,21]. This response is most recently reported in terms of either circadian stimulus (CS) or the α-opic Melanopic Equivalent Daylight Illuminance (M-EDI) lux [20,21]. The former is defined as the "effectiveness of the spectrally weighted irradiance at the cornea from threshold (CS=0.1) to saturation (CS=0.7)" [20]. The latter refers to International System of Units–compliant metrics as supported by the International Commission on Illumination, which converts multiple spectral inputs to ipRGC-relevant quantities [21]. We have used this to obtain additional values of M-EDI lux according to the LED spectral distribution and melanopic ratio of the lighting. The LED board is bespoke and has been designed to optimize the melanopic response.

In this trial, the illuminance was measured using a lux meter and the color temperature was obtained from the luminaire network. The calculations for the melanopic metrics were completed using the spectral power distribution values from our LED datasheet [22] and the lux intensities from Table 2, with the aid of open access 3rd party software for both CS and M-EDI lux calculations [20,23]. The melanopic ratio does not change throughout this study. The previously installed static lighting for the care home's bathrooms, bedrooms, kitchennette and living space, and common areas exhibit mean illuminance of 99 (SD 49), 75 (SD 12), 59.33 (SD 15.41), and 147 (SD 60.42) lux, respectively. In contrast, the dynamic intervention metrics are shown in both Tables 2 and 3.

The intervention lighting was programmed to simulate the varying properties of the sunlight cycle throughout the day as depicted in Figure 5. Warmer color temperatures were present in the early mornings and evenings with cooler color temperature in the middle of the day. The capacity to switch on or off the lighting was maintained by the residents or the care staff according to preference and was also logged in the network. This then allows for metrics on actual lighting exposures to be calculated.

Table . Photopic lighting intervention parameters.

Intensity (lux)	Correlated colour temperature (K)	Output (%)[a]	Time
300	3350	90	7 AM to 8 AM
550	4000	100	8 AM to 8:30 AM
550	5200	100	8:30 AM to 9 AM
550	6500	100	9 AM to 2:30 PM
300	4200	90	2:30 PM to 3 PM
260	3400	90	3 PM to 4 PM
300	2700	90	4 PM to 6 PM

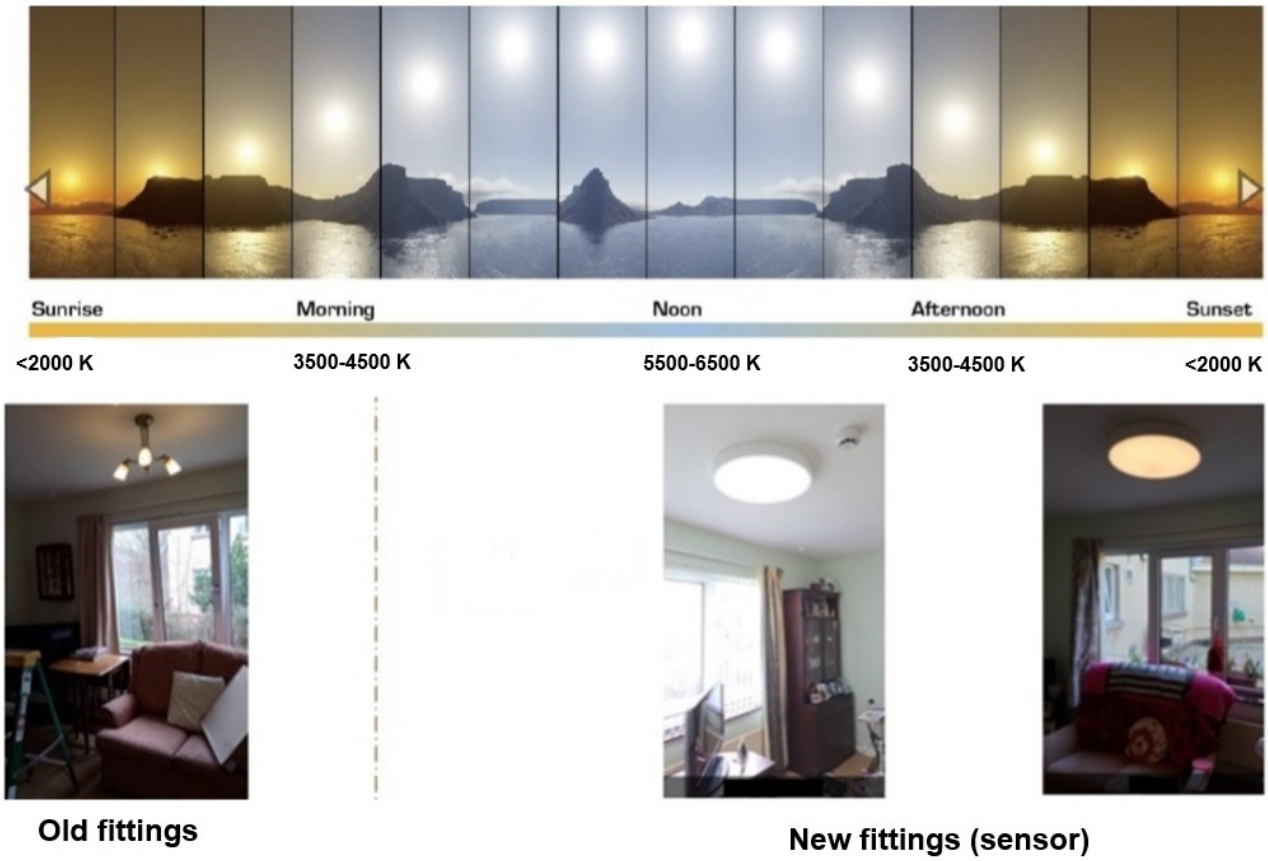
[a]The output is from the luminaire manufacturer and is only expressed in percentages. The actual values are not available. The “intensity (lux)” column states what the authors can measure at eye level that reflects what the percentages give us at this amount of distance from the lighting.

Table . Melanopic lighting intervention parameters.

Intensity (lux)	Circadian stimulus	M-EDI[a] lux
260	0.404	331
300	0.431	382
550	0.532	700

[a]M-EDI Melanopic Equivalent Daylight Illuminance.

Figure 5. Lighting before and after intervention highlighting the variation in spectra throughout the day.



Sensing Component

The sensor is an environmental sensor integrated inside the luminaire and concealed by the diffuser. It is a depth sensing radar sensor, which can perceive the averaged 3D coordinates of an individual in a privacy-friendly manner. The sensor is deployed to track the location, rest-activity, and night activity metrics of the residents with dementia; all parameters expected to be influenced by changes to the circadian rhythm. Frame-by-frame x-y translations were captured at 8 Hz and the total magnitude calculated over 15-minute intervals. This was

normalized by the maximum total translation in the current 15-minute window in order to give activity levels as a percentage. As such, the pseudocode below Figures 6 and 7 demonstrates the type of information collected about each resident.

Figure 6. Algorithm 1: activity metrics.

Algorithm 1: Activity metrics

```
Data: rooms
Result: activity 15min(activity 15, max activity 15, scaled activity)
for room in rooms do
    Get last 15 minutes of x, y data;
    Isolate by single sensor contribution;
    Resample at 1s resolution;
    Assess intervals of sensor overlap (if any);
    if overlap exists then
        Average out the overlapping contributions;
        Combine data;
    else
        Combine data;
    Calculate and store the activity magnitudes per second: |activity|/s;
    activity.append(|activity|/s);
    Calculate activity per 15 minutes:;
    activity_15 = sum(activity);
    Calculate max activity per 15 mins:;
    max_activity = max(activity);
    max_activity 15 = (60 × 15) × max activity;
    Scale the 15 minute activity data to the maximum activity for that time interval:;
    scaled_activity = activity 15 / max_activity 15;
    Return activity 15min(activity 15, max activity 15, scaled activity);
```

Figure 7. Algorithm 2: sleep metrics.

Algorithm 2: Sleep metrics

```
Data: rooms
Result: bed coords(x min, x max, y min, y max, z min, z max)
for room in rooms do
    Collect 1 week of x, y, z position data;
    Isolate between hours of 11pm-5am;
    Calculate mode of x, y, z ± n*ϑ;

Return sleep metrics(sleep start, sleep wake, no of disturbances, bout length));

for room in rooms do
    Pull bed_coords;
    Get last 24 hours of x, y, z position data;
    Generate 1 minute windows of average x, y, z values;
    Assess if each window within bed coord boundaries;
    Put status flag as 'in bed' or 'out of bed' based on above criteria;
    Re-write rules if < 2 consecutive 'out of bed' occurrences present as likely to be noise;

Return bed_status('in bed' — 'out of bed');

for room in rooms do
    Pull bed_status for last 24 hours;
    Isolate from 18:00 (reasonable minimum bed time);
    If > 10 minutes of consecutive 'in bed' occurrences:;
        sleep_start = start time of window;
    Isolate from 05:00 (reasonable minimum wake time);
    If > 10 minutes of consecutive 'out of bed' occurrences:;
        sleep_wake = start time of window;
    Get bed_status between sleep_start and sleep_wake;
    Count consecutive 'out of bed' occurrences and their durations:;
        no_of_disturbances = total no of bed leaves;
        bout_length = average length of time spent 'out of bed' during these hours;

Return sleep_metrics(sleep start, sleep wake, no of disturbances, bout length);
```

Data Analysis

Data analyses was undertaken using SPSS software [24]. All datasets were tested for normality using the Shapiro-Wilk test since the cohort was less than 50 participants. If the data was parametric, paired samples *t* tests (2-tailed) were used to determine statistical significance. If the data exhibited a nonparametric distribution, the Wilcoxon-Signed rank test was used between related samples. Effect sizes were calculated using Cohen *d* and Cliff δ for parametric and nonparametric data, respectively. All measured parameters of well-being were determined parametric with significant deviation from normal and detected solely for 2 parameters of the QUALIDEM scale (E and H: positive self-image and feeling at home). These were therefore analyzed according to the nonparametric methodology highlighted above. Due to the heterogeneity in the sleep metrics obtained, the median of the datasets was calculated to allow for less skew that would arise from taking an average.

Results

Lighting

The lighting metrics were harvested from the network to estimate how much time a resident spent under the dynamic lighting intervention over the duration of the trial. This can be seen in Figure 8. The blue line represents the amount of time spent in their rooms. The orange line indicates that while in their rooms, this is the average percentage of time that the lights were switched on and they were in fact receiving the dynamic lighting intervention. Green lines represent the percentage exposure to dynamic lighting. Typically, the care staff note that common room lighting is generally switched on during the hours of 8 am to 8 pm and all resident lighting is switched on upon wake up time and altered at their discretion throughout the day.

A group analysis of Figure 8 can be seen in Figure 9. The amount of time spent in their rooms generally decreased meaning the amount of time assumed spent in common areas

increased. Although both the blue and orange lines are decreasing, they are converging.

Assuming the times not spent in their rooms were considered to be under common area fully dynamic lighting, the general percentage exposure to dynamic lighting from baseline to weeks 5 - 10 to weeks 10 - 16 increased from the following:

- Baseline: 0% exposure as fully static lighting
- Weeks 5 - 10: 37.48% exposure (SD 2.77%)
- Weeks 10 - 16: 53.08% exposure (SD 4.07%).

Figure 8. Lighting exposure per participant. Blue represents the percentage time in sensor areas, orange is the percentage time while in these areas with lighting switched on, and green represents the percentage time of exposure to dynamic lighting.

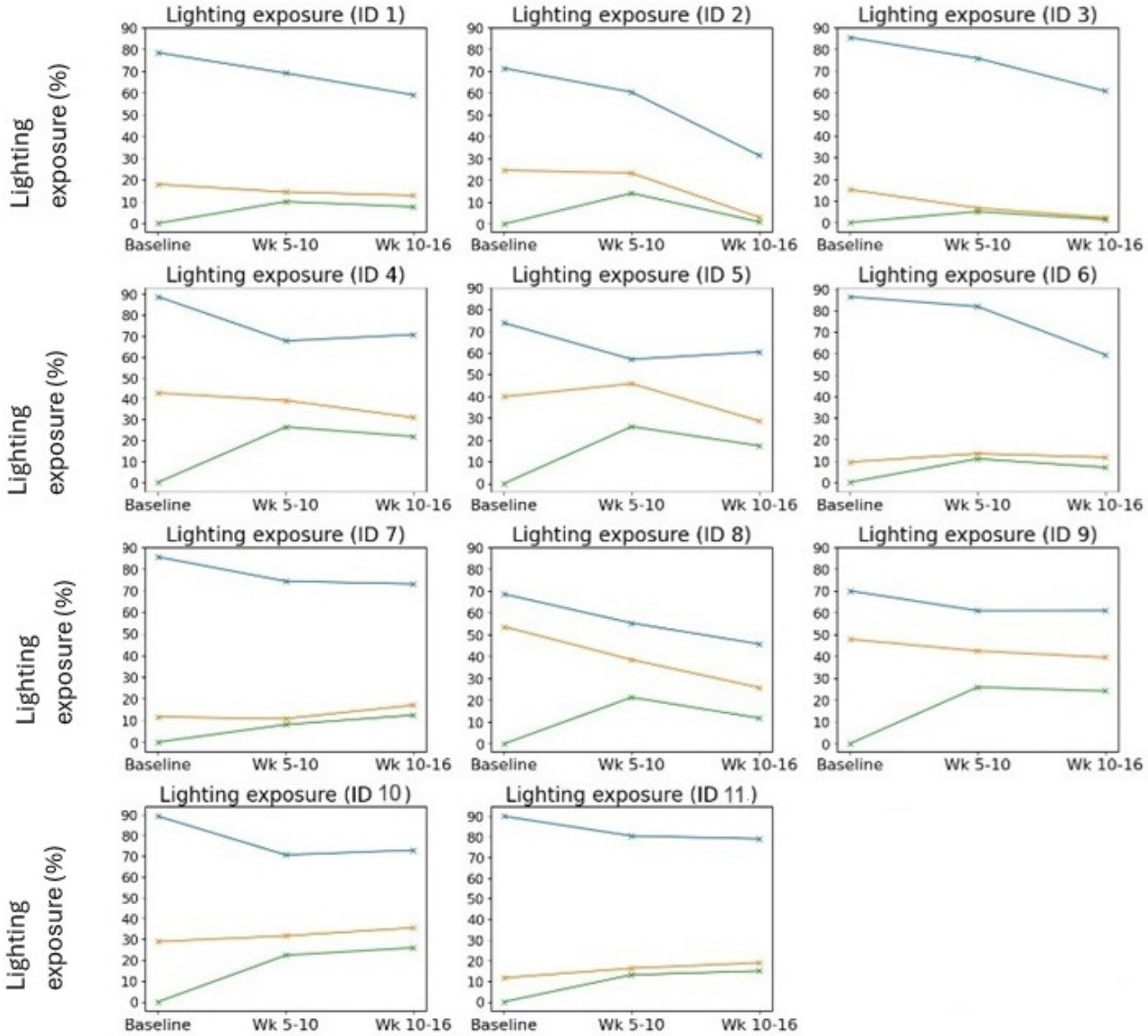
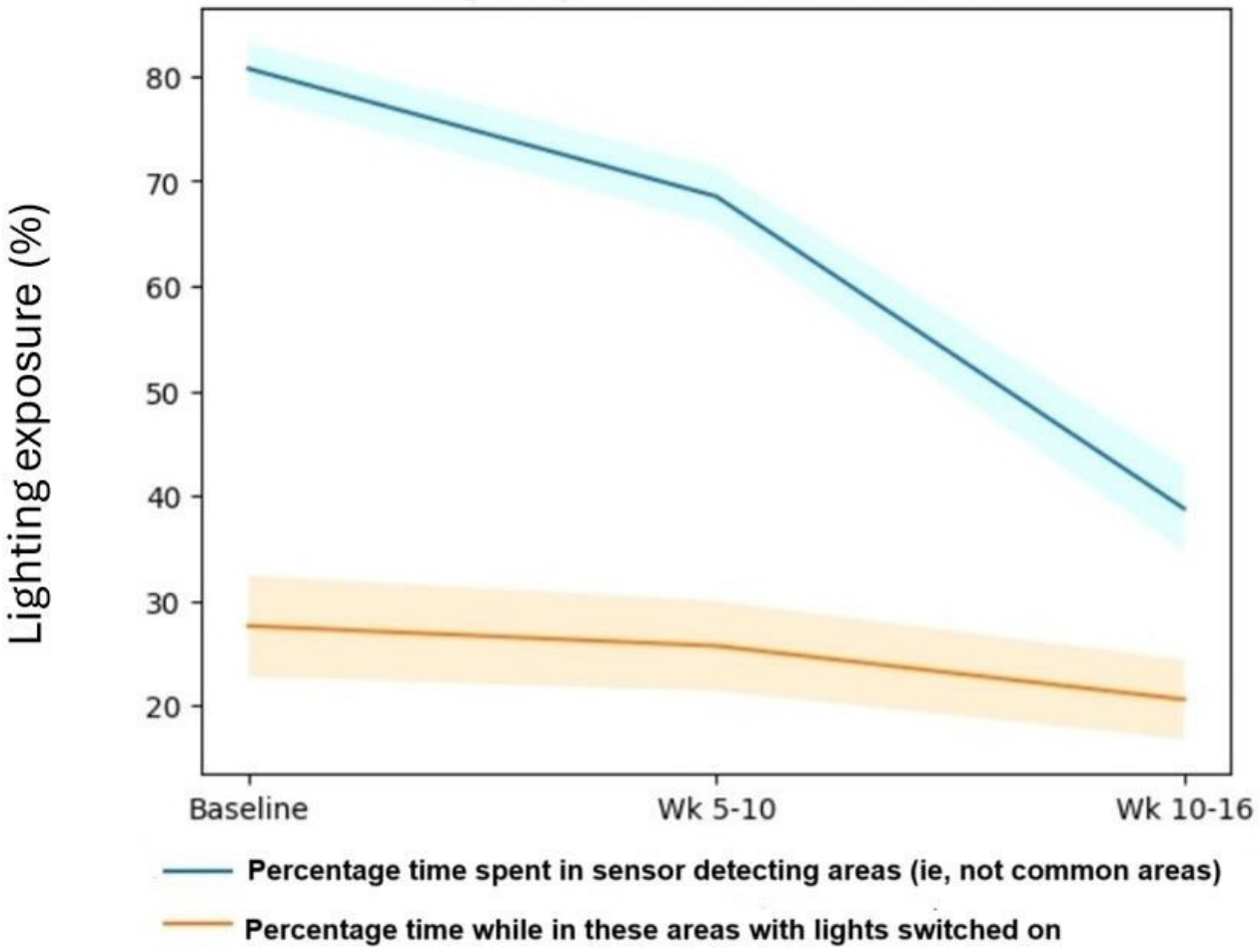


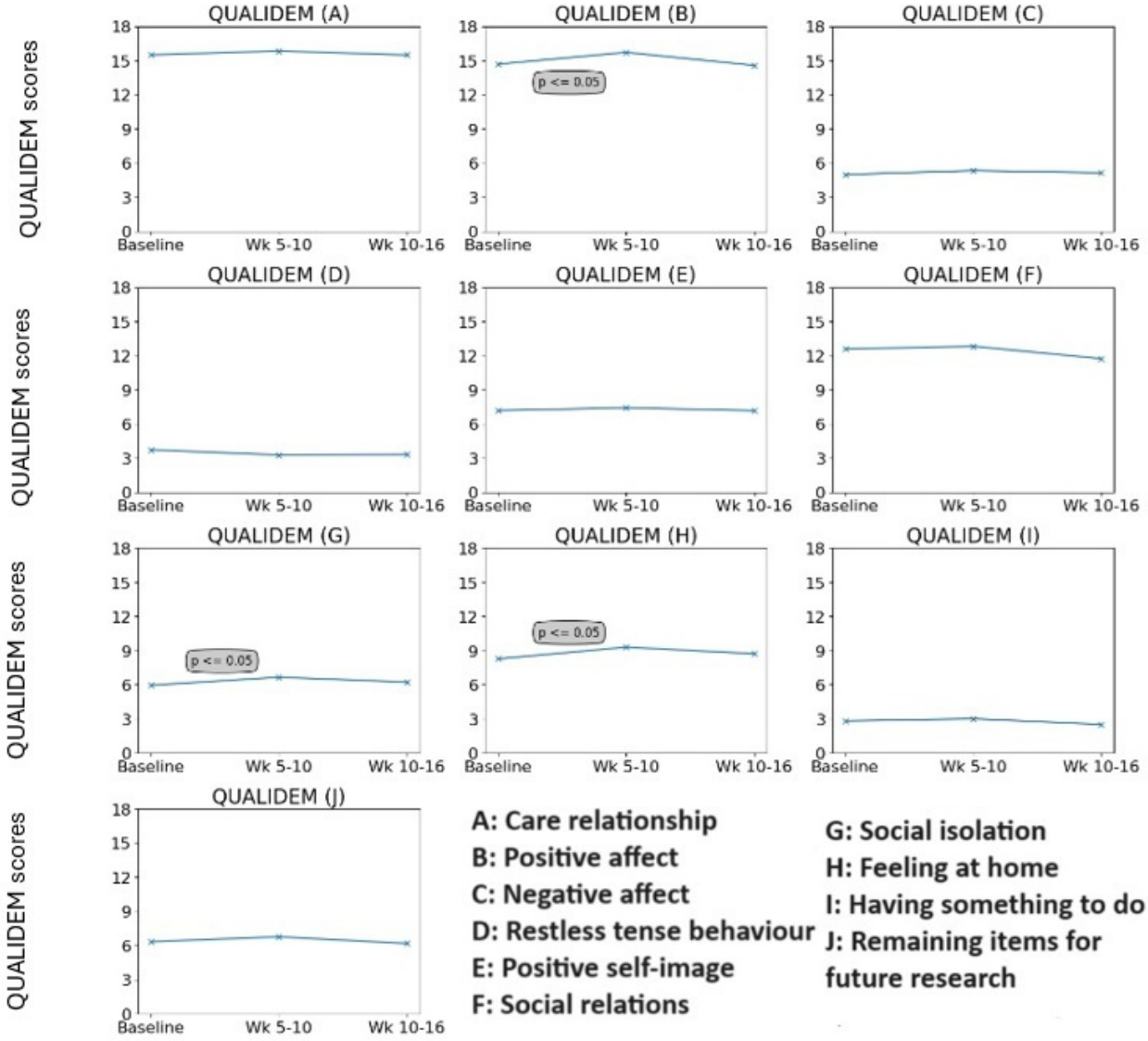
Figure 9. Time spent with lights on as a group trend. The lighter lines indicate error boundaries.



QUALIDEM

Figure 10 shows the QUALIDEM parameters analyzed independently in accordance with the documentation. The results show that from baseline to weeks 5 - 10, there was a statistically significant improvement in B, G, and H with values of *P*=.03, *P*=.048, and *P*=.047 and a moderate effect size. This corresponds to statistically significant improvements in positive affect, social isolation, interdaily stability (IS), and feeling at home.

Figure 10. QUALIDEM scores as a group analyses during both baseline (static lighting) and weeks 5-10/10-16 (dynamic lighting). Note items I and J are not analyzed in this study. Larger values indicate improvement. QUALIDEM: Quality of Life for People with Dementia.



Activity

Commonly reported metrics used in rest-activity studies are the outputs of IS, intradaily variation (IV), and relative amplitude (RA) [25]. IS is an indicator for the synchronization to light and other metabolic cues that inform the body clock. It takes a value between 0 and 1 with the larger number indicating better synchronization. IV gives an indication of body clock disruption and sleep efficacy, with possible values between 0 and 2. The lower the value the better sleep efficacy. RA gives an indication of the robustness of the daily rest-activity rhythm, with higher values equating to a more robust cycle [25].

On a group level, the IS and IV both experience a significant improvement from baseline to weeks 5 - 10 with values of $P \leq .001$ for both variables (Figure 11). The IS also reflects this from baseline to weeks 10 - 16 with a significance of $P \leq .001$. The IV increases from baseline to weeks 10 - 16 but not significantly with a P value of .22.

Figure 12 shows the individual RA. From baseline to weeks 5 - 10, there was a significant improvement in the RA, which means the participant experiences a more robust 24 hour rhythm. The P value for this period was .03. The baseline to weeks 10 - 16 was not significant, but it follows a general upward trend on a group analysis.

An assessment was also carried out of activity levels during typical sundowning hours. The 4 hours before sunset (8 PM, 9 PM, and 10 PM for baseline, weeks 5 - 10, and weeks 10 - 16, respectively) were analyzed as seen in Figure 13. The activity levels increased significantly from baseline to weeks 5 - 10 ($P \leq .001$) but then decreased significantly from baseline to weeks 10 - 16 ($P = .01$). If we take the sunset activity as an analogy to activity during the sundowning hours in dementia, we can see that activity during sundowning hours seems to have increased at weeks 5 - 10 and decreased at weeks 10 - 16.

Figure 11. Box plots for rest-activity metrics as a group analysis, including interdaily stability, intradaily variation, and relative amplitude.

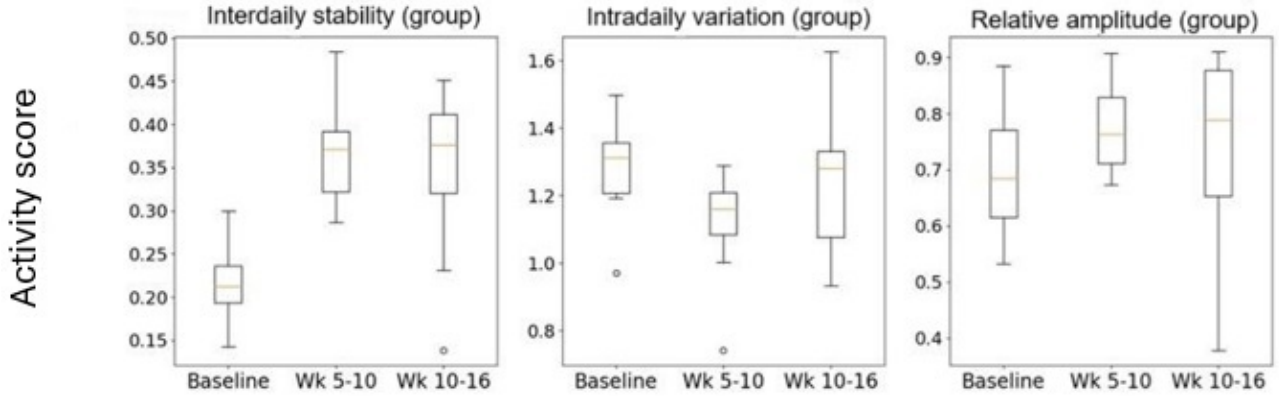


Figure 12. Line plots for rest-activity metrics for an individual analysis, including interdaily stability, intradaily variation, and relative amplitude.

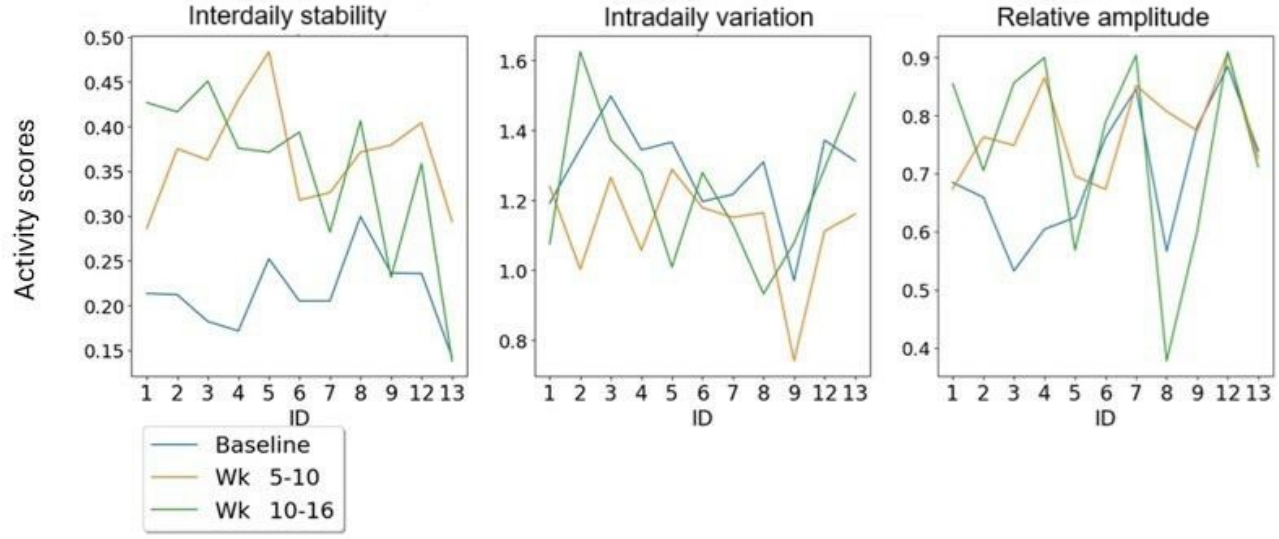
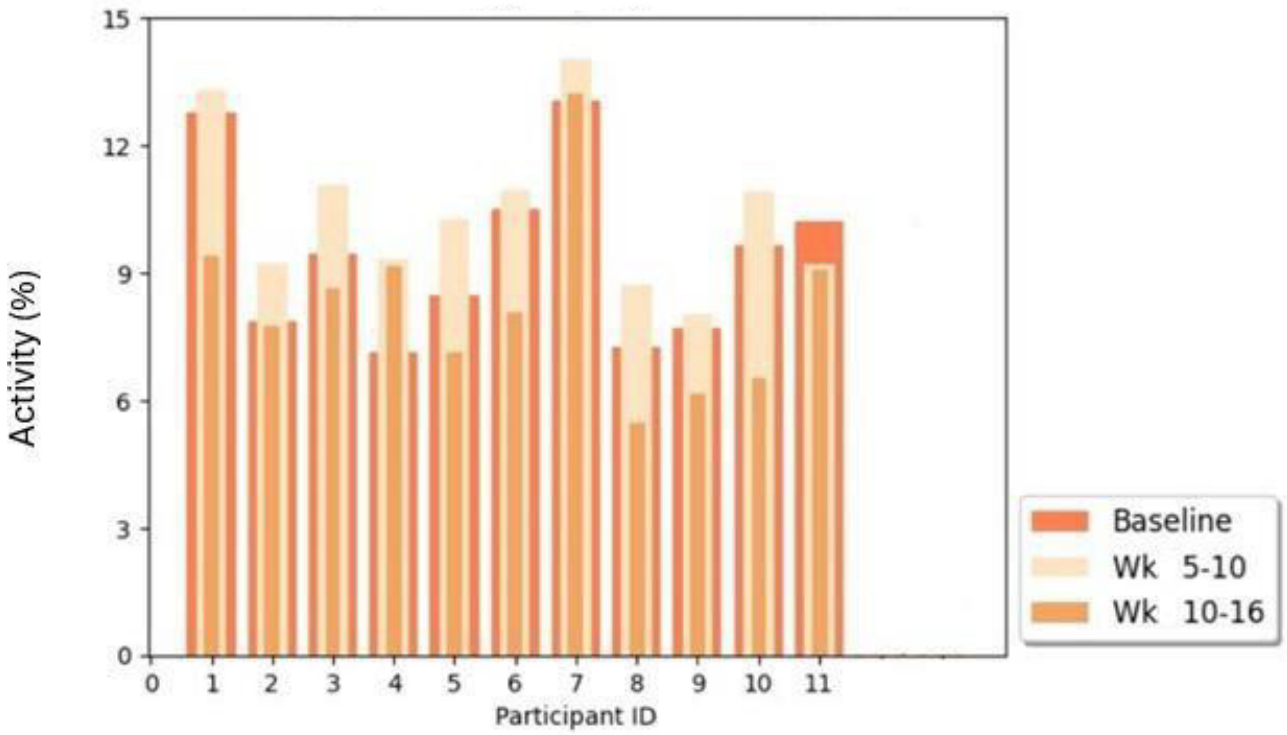


Figure 13. This is the percentage activity levels in the 4 hours before sunset during the trial under different lighting exposures. These data were analyzed with a view to compare with sunset activity or sundowning.



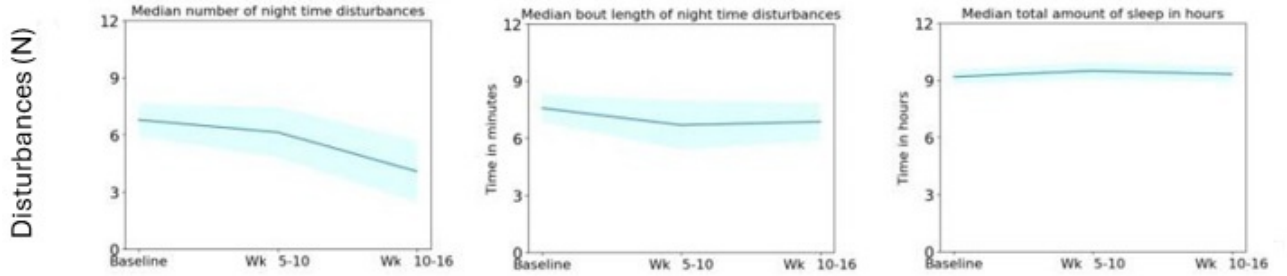
Sleep

Figure 14 shows the median number of sleep disturbances as a group decreased from both baseline to weeks 5 - 10 and from baseline to weeks 10 - 16; however, this did not reach statistical significance.

It should be noted that the highest number of disturbance counts throughout the night on average happened to participant 5 who is the sole one on sleeping medication. If the paired *t* test is repeated for the group without participant 5's data, the reduction in sleep disturbances becomes statistically significant from baseline to weeks 10 - 16.

Figure 14 shows the median length of time that these disturbances lasted for (bout length) also decreased from both baseline to weeks 5 - 10 and baseline to weeks 10 - 16; however, this did not reach statistical significance.

Figure 14. This is the sleep-related metrics recorded on a group level, highlighting the median number of night-time disturbances, bout length of these night-time disturbances, and the total amount of sleep in hours.



Summary

Table 4 shows the *P* values and effect sizes for the statistically significant well-being parameters. Using G*Power (Heinrich Heine University Düsseldorf) software to deduce the required sample size for a medium effect size (0.5), the optimum number of participants to recruit for a future study would be 34.

Table . *P* values and effect sizes for statistically significant well-being parameters.

	P values	Cohen *d*/ Cliffδ**	Summary
QUALIDEM[a] (B)[b]			
Wk 5 - 10	.03	0.639	Significant with medium effect
QUALIDEM (G)[c]			
Wk 5 - 10	.048	0.552	Significant with medium effect
QUALIDEM (H)[d]			
Wk 5 - 10	.047	0.289**	Significant with high overlap
Activity (IS)[e]			
Wk 5 - 10	<.001	2.601	Significant with large effect
Activity (IS)			
Wk 10 - 16	<.001	1.517	Significant with large effect
Activity (IV)[f]			
Wk 5 - 10	<.001	1.314	Significant with large effect
Activity (RA)[g]			
Wk 5 - 10	.03	0.621	Significant with medium effect

[a]QUALIDEM: Quality of Life for People With Dementia.

[b]B: positive affect.

[c]G: social isolation.

[d]H: feeling at home.

[e]IS: interdaily stability.

[f]IV: intradaily variation.

[g]RA: relative amplitude.

Discussion

Principal Findings

This study was conducted to determine the effect of dynamic lighting on well-being for people who are living with dementia.

Social Isolation

Previous studies have highlighted that the exposure to dynamic lighting is not well documented [26-29]. The results from this study highlight that the amount of exposure to dynamic lighting as a group increased from baseline to weeks 5 - 10 and 10 - 16 by 37.48% and 53.08% respectively. Moreover, the amount of times spent in the common areas increased in comparison to the amount of time spent within their own living quarters. Interestingly, by looking at the QUALIDEM scale for parameter "G" as seen in Figure 10, significant improvements to social isolation were found at weeks 5 - 10. This may indicate that exposure to dynamic over static lighting may be responsible for the positive implications found for social isolation. This is in agreement with similar studies conducted by Sust et al [30] who found that residents with dementia experienced improved communications and frequency of participation in activities under dynamic lighting compared to static baseline lighting. Their results were particularly noticed under higher illuminances than delivered in this study (300 - 2200 lux), but with dynamic changes to color parallel with those within this study [30].

The reduction in social isolation within our study continued into weeks 10 - 16 from baseline but did not reach statistical significance. The reason for the subdued progress may be due to the contrast in lighting within a "circadian synchronization period." A "circadian synchronization period" refers to the amount of time it takes for the body to adapt to changes in the circadian rhythm; commonly reported as a period of 2 - 5 weeks [16]. From baseline to weeks 5 - 10, the contrast in lighting transitions is from static lighting to dynamic lighting, whereas there is no contrast introduced in weeks 5 - 10 to weeks 10 - 16; it is simply a continuance of the dynamic lighting parameters. This dampening in the significant improvement from weeks 5 - 10 to nonsignificant improvement in weeks 10 - 16 is interesting to consider when comparing with studies by Wong et al [31]. They deciphered that the exposure to previous lighting histories exhibits a huge impact on an individual's response to lighting [31]. For example, Wong et al [31], found that there is a greater response to lighting interventions upon prior exposure to dark stimuli than bright stimuli, perhaps providing an example of how the larger the contrast in lighting between circadian synchronization periods, the larger the expected impact to the alleviation of certain symptoms. It could be that upon extending our study past 16 weeks, we would have continued to observe a slower paced improvement to social isolation. A further indicator that this might be the case is provided by a study by Figueiro et al [32] who found that symptoms such as agitation, sleep quality, and depression after exposure to dynamic lighting experienced greater alleviation as time progressed in the 25 week study [32].

Sleep

There is a recurring theme with a lot of our findings in that the most significant improvements are observed from baseline to weeks 5 - 10 of dynamic lighting exposure. In most parameter's cases, this improvement from baseline continues into weeks 10 - 16 but at a much lesser and nonsignificant effect than the initial phase. Consider the sleep metrics within this study; the total amount of sleep in hours as highlighted in Figure 14 experiences a progressive increase from baseline to weeks 5 - 10 and 10 - 16. Again, this improvement follows a steeper incline in the first phase of this study in comparison to the second phase. The aforementioned study by Figueiro et al [32] also demonstrates that between the first 3 weeks and 9 weeks of dynamic lighting exposure there is an improvement from baseline, but with the former interval marginally exceeding the 9 week interval. Since weeks 5 - 10 and 10 - 16 within our study refer to the first 5 and 11 weeks of dynamic lighting exposure, respectively, our findings would align with the findings from this study [32]. However, it should be noted that there are other studies which have monitored the impact to sleep after exposure to dynamic lighting and found improvements become progressively larger as the exposure interval increases. For example, Hjetland et al [28] found improvements to proxy-rated sleep after 16 and 24 weeks but not after 8 weeks. One reason these results may contrast is that both our study and Figueiro et al [32] placed dynamic lighting in every vicinity that a resident may frequent within a care home, however Hjetland et al [28] solely placed luminaires in the living space. This could suggest that the absence of a positive implication in sleep parameters witnessed by Hjetland et al [28] after 8 weeks of dynamic lighting could be due to the fact that the exposure amounts were not sufficient enough and therefore the impact to the sleep response would take longer to synchronize. Further exploring the sleep data in our study, Figure 14 observes that the number of sleep disturbances reduced from baseline to both phases of the dynamic lighting exposure. Interestingly, it should be noted that participant 5 was the sole resident on sleeping medication and that if this participant's data are removed from the cohort, the reduction in sleep disturbances at the 10 - 16 week dynamic lighting interval becomes significant. This could indicate that people with dementia on sleep medication may require more tailored dynamic lighting to reflect upon their intensified circadian disruption. This is in alignment with studies suggesting that intense sleep disruption leads to a more rapid cognitive decline and more disrupted circadian rhythms acting in a self-depreciating cycle [1].

Sundowning-Analogous Behavior and Agitation

In terms of sundowning, for most of the participants in our study the percentage of sundowning-analogous activity as seen in Figure 11 demonstrates that after 10 - 16 weeks, the majority of participants experience a nonsignificant reduction in sundowning compared to baseline lighting. The sole participants who experienced an increase in this sundowning behavior at this time interval were participants 4 and 7. Interestingly, when calculating the median wake times for residents, these same participants were the latest to wake with times commencing from 8:25 AM and 9:29 AM, respectively. During these times, it is likely that the first lighting these residents will receive is

of maximum CS and of blue (peak melanopic) wavelength as seen in Tables 2 and 3, as opposed to a general build up to this level the other residents will have experienced at wake times prior to 8 AM. Considering a study by Colenda et al [33] used bright light therapy for participants immediately as they woke, they found that only 1 of 5 participants experienced positive implications on sleep and agitation measurements [33]. Although a much stronger illuminance was used in their study, it should be considered that timings for peak melanopic lighting immediately upon waking may not be of maximum benefit for well-being.

In addition to this, sundowning seems to be a parameter that progressively improves over time. Again looking at Figure 11, it seems sundowning occurrences seem to increase after a period of 5 - 10 weeks before decreasing after a period of 10 - 16 weeks. Similar studies by Baandrup and Jennum [34] and Saidane et al [35] have monitored agitation and found no change to agitation after 4 weeks of dynamic lighting and a reduction in the frequency of agitation after 6 months of dynamic lighting intervention [29,34]. This suggests that longer exposure durations may be needed to positively impact this parameter of well-being in dementia. Additionally, a study by Burns et al [36] observed that the reduction in agitation was significantly correlated to the increase in day length, and so may explain why reductions in activity levels during sundowning hours were larger in later weeks as they occurred over the month of June when the summer solstice occurs [36]. In contrast, the QUALIDEM reported agitation scores (parameter D in Figure 10) highlights a nonsignificant increase in agitation from baseline to weeks 5 - 10 and 10 - 16; however, these results exhibit a nonsignificant decrease in agitation from weeks 5 - 10 to 10 - 16 of dynamic lighting exposure. Once again, the pattern of larger impacts to symptoms becoming present in the transition from static to dynamic lighting phases compared to marginally extended dynamic lighting phases is apparent. The results of the parameter of agitation are complex to analyze due to the contrasting behaviors in sundowning-analogous and QUALIDEM-reported metrics in our study. Reassuringly, it is commonly reported in the literature that the parameter of agitation in dementia in response to lighting is not well understood. In a 2014 lighting review by Forbes et al [37], they found no significant findings for bright light therapies reducing agitation with one study even contributing to a worsening of symptoms in 5 participants [37,38].

Positive Affect and Feeling at Home

Our study found significant improvements in both the well-being parameters of “feeling at home” and “positive affect” during weeks 5 - 10 of dynamic lighting as seen in Figure 10. Positive affect can be likened to positive mood and similar results after exposure to dynamic lighting have been published by Bromundt et al [39], Figueiro et al [32], Kolberg et al [40], and van Lieushout-van Dal et al [41]. It is interesting to consider whether several parameters within the QUALIDEM scale influence the progression of other parameters; for instance does positive affect influence social isolation or sense of feeling at home? Nevertheless, the positive impact on sense of feeling at home (parameter H in Figure 10) has been replicated in other qualitative studies which have researched the opinion on lighting interventions in dementia [42]. The impact to this parameter of well-being after exposure to dynamic lighting may be less surprising, as is seems unanimous in studies designed to reflect on the environmental opinion of dynamic lighting systems that it is better received than that of static lighting [43,44].

Rest-Activity

Our study also monitored the changes to rest-activity rhythms upon exposure to dynamic lighting, reporting findings on IS, IV, and RA (see Figures 2 and 11). All three activity metrics experience significant improvements from baseline to weeks 5 - 10, with IS reporting significance from baseline to weeks 10 - 16 as well. Since these metrics demonstrate adherence of circadian rhythm stability and robustness to light or dark stimuli, it is interesting to consider how the transition to dynamic lighting has impacted them. For example, if we look at Figure 8, which highlights the individual exposures to dynamic lighting, we observe that from weeks 5 - 10 to weeks 10 - 16 for participants 2, 5, and 8, that they experienced the largest reduction in exposure to dynamic lighting, indicating a suboptimal percentage exposure to dynamic lighting. Consequently, if we look at Figure 10 in the RA plot, we can see that there are spikes at lower values than the trend (green line) for participants 2, 5, and 8. This indicates that insufficient exposure to dynamic lighting can significantly impact the robustness of the circadian rhythm.

Furthermore, these significant improvements in rest-activity upon exposure to dynamic lighting are in agreement with studies undertaken by Arden-van Delft et al [27] who concluded that there may be stabilization in activity after using dynamic lighting over a period of 5.5 months in a recurring control-intervention approach with phases spanning between 28 and 42 days [27]. Other studies such as Baandrup et al [34] found that there was no significant changes to rest-activity after exposure to dynamic lighting over 4 weeks [34]. The reasons for the larger improvement in rest-activity witnessed in our study may be due to the fact the dynamic lighting was provided in a completely ambient fashion even in external common areas. However, it may also be attributed to the fact that rather ironically our activity results are being provided by sensors which are environmentally deployed within their living quarters only (not ambient in common areas), and they are therefore not able to capture rest-activity at times when participants are in the common areas.

Summary

It seems possible that prolonged use of dynamic lighting for people living with dementia would lead to aggregated improvements to their well-being over time. In this study, the first phase after baseline (weeks 5 - 10) experienced a 37.48% increase in dynamic lighting exposure, while the transition from phase 1 (weeks 5 - 10) to phase 2 (weeks 10 - 16) experienced an additional exposure of 15.6%, which is less than half of the initial static to dynamic transition. Perhaps the lack of statistical significance continued into weeks 10 - 16 can be explained by the fact that the dynamic lighting exposure amounts have not increased as much as they did in the first instance; from static (no dynamic exposure) to weeks 5 - 10. Since this effect of larger statistical significance is demonstrated in QUALIDEM

parameters, rest-activity, and sleep at weeks 5 - 10 and not at weeks 10 - 16, it seems plausible that insufficient exposure to dynamic lighting could be the cause.

Study Limitations and Future Work

An overall viewpoint of the technology was communicated by the care home manager who supported its use and benefits. One suggestion was for a central control unit to be introduced whereby the care staff could have control over the light switches in resident's rooms. This was suggested so that the residents could obtain the majority exposure to the dynamic lighting but may cause issues in implementing due to the fact that it could take the autonomy away from the residents with dementia. The lack of sensors within the light fittings in common areas may have limited the amount of activity data that could be attributed to each individual throughout the day. In future studies, it would be beneficial to request the consent of the use of sensors in these areas and to introduce sensor processing techniques to distinguish between residents within these areas. In addition, the number of participants within this study was small, and undoubtedly a larger cohort (34 participants) would have provided more applicable effect sizes. Implementing a randomized controlled trial to definitively distinguish cause and effect would be the ideal future pathway for research.

Another limitation to this study was that the 16 week trial period took place from March to July which can be considered a "sunny period"; hence, there is the possibility that the changes observed were due to the natural daylight impact. However, the authors carefully considered the 4 week baseline period in order to ensure that the changes that each individual exhibited could always be compared to a baseline for that individual. In this way, even if all months are "sunny periods," there would be a comparison between static indoor and dynamic indoor lighting. There is a similar argument to be made for the different sun exposures in each room. Since each individual is their own control, these differences are present in both the baseline and intervention measurements and are therefore considered to have minimal impact on any changes observed over the baseline and intervention transition. However as this is a pilot study, future work involves the extension of this study into the winter months to account for external seasonal differences as well.

Additionally, there will be differences in daylight exposure between the equinox in March and the summer solstice in July. The impact of this was not measured but the effect this may have had on study participants cannot be ignored. Again, the authors recommended elongating this study across multiple seasons to gain insight into any potential impact this may administer.

The care home that took part in this study was chosen due to its location within an accessible geographic region for the research team to attend when necessary. Due to the small number of people living with dementia in these facilities, the authors chose to recruit all residents who met the inclusion criteria in favor of random sampling. As a result, the trial was conducted using convenience sampling. The authors note that random sampling may have provided a better representation of the cohort of people living with dementia, but due to the piloting of the trial and the challenges involved in the recruitment of vulnerable adults the authors decided to include all participants. Another limitation was that the type of dementia was not logged. Since several lighting studies have demonstrated that the impact to well-being may be more prominent depending on the type of dementia, this is a parameter that should be logged in all future studies of this type. The authors appreciate that logging this information may provide additional insight into these results.

Conclusion

A 16-week study trialling dynamic lighting for supporting well-being in a care home for residents with dementia was carried out from March to July 2023. The results found that the use of dynamic lighting over static lighting statistically and significantly improved positive affect, social isolation, sense of feeling at home, and rest-activity within 5 weeks of initial exposure. Nonstatistically significant improvements to sleep amount and disturbances were also recorded and extended even 12 weeks into this study; however, improvements were not observed for the well-being parameter of agitation.

Acknowledgments

The authors would like to acknowledge the participants, families and staff from the care home for helping to facilitate this research. They would also like to express thanks to the Royal Commission for the Exhibition of 1851 in their support for this research.

Conflicts of Interest

None declared.

References

1. Li P, Gao L, Gaba A, et al. Circadian disturbances in Alzheimer's disease progression: a prospective observational cohort study of community-based older adults. Lancet Healthy Longevity 2020 Dec;1(3):e96-e105. [doi: 10.1016/S2666-7568(20)30015-5]
2. Wirz-Justice A, Skene DJ, Münch M. The relevance of daylight for humans. Biochem Pharmacol 2021 Sep;191:114304. [doi: 10.1016/j.bcp.2020.114304] [Medline: 33129807]
3. Cho E, Kim S, Hwang S, et al. Factors associated with behavioral and psychological symptoms of dementia: prospective observational study using actigraphy. J Med Internet Res 2021 Oct 29;23(10):e29001. [doi: 10.2196/29001] [Medline: 34714244]

4. World Health Organisation. Dementia. 2023. URL: https://www.who.int/news-room/fact-sheets/detail/dementia [accessed 2024-11-01]
5. Livingston G, Huntley J, Sommerlad A, et al. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. Lancet 2020 Aug 8;396(10248):413-446. [doi: 10.1016/S0140-6736(20)30367-6] [Medline: 32738937]
6. Musiek ES, Xiong DD, Holtzman DM. Sleep, circadian rhythms, and the pathogenesis of Alzheimer disease. Exp Mol Med 2015 Mar 13;47(3):e148. [doi: 10.1038/emm.2014.121] [Medline: 25766617]
7. Nair NP, Schwartz G, Kin N, et al. Melatonin and cortisol circadian rhyhtms in Alzheimer's disease patients and normal elderly subjects. In: Biological Clocks: Mechanisms and Applications: Elsevier Health Sciences; 1998:357-360.
8. Sato M, Matsuo T, Atmore H, Akashi M. Possible contribution of chronobiology to cardiovascular health. Front Physiol 2013;4:409. [doi: 10.3389/fphys.2013.00409] [Medline: 24478711]
9. Grabe S, Mahammadov E, Olmo MD, Herzel H. Synergies of multiple zeitgebers tune entrainment. Front Netw Physiol 2021;1:803011. [doi: 10.3389/fnetp.2021.803011] [Medline: 36925578]
10. Schlangen LJM, Price LLA. The lighting environment, its metrology, and non-visual responses. Front Neurol 2021;12:624861. [doi: 10.3389/fneur.2021.624861] [Medline: 33746879]
11. Goudriaan I, van Boekel LC, Verbiest MEA, van Hoof J, Luijkx KG. Dementia enlightened?! a systematic literature review of the influence of indoor environmental light on the health of older persons with dementia in long-term care facilities. Clin Interv Aging 2021;16:909-937. [doi: 10.2147/CIA.S297865] [Medline: 34079240]
12. Torrington JM, Tregenza PR. Lighting for people with dementia. Light Res Technol 2007 Mar;39(1):81-97. [doi: 10.1177/1365782806074484]
13. Figueiro MG, Plitnick B, Rea MS. The effects of chronotype, sleep schedule and light/dark pattern exposures on circadian phase. Sleep Med 2014 Dec;15(12):1554-1564. [doi: 10.1016/j.sleep.2014.07.009]
14. Brown TM. Melanopic illuminance defines the magnitude of human circadian light responses under a wide range of conditions. J Pineal Res 2020 Aug;69(1):e12655. [doi: 10.1111/jpi.12655] [Medline: 32248548]
15. Meiland F, Innes A, Mountain G, et al. Technologies to support community-dwelling persons with dementia: a position paper on issues regarding development, usability, effectiveness and cost-effectiveness, deployment, and ethics. JMIR Rehabil Assist Technol 2017 Jan 16;4(1):e1. [doi: 10.2196/rehab.6376] [Medline: 28582262]
16. Al-Karawi D, Jubair L. Bright light therapy for nonseasonal depression: meta-analysis of clinical trials. J Affect Disord 2016 Jul 1;198:64-71. [doi: 10.1016/j.jad.2016.03.016] [Medline: 27011361]
17. Ettema TP, Dröes R, de Lange J, Mellenbergh GJ, Ribbe MW. QUALIDEM: development and evaluation of a dementia specific quality of life instrument—validation. Int J Geriat Psychiatry 2007 May;22(5):424-430. [doi: 10.1002/gps.1692]
18. Dichter M. Inter-rater reliability of the Qol-measure QUALIDEM. BMC Nurs 2015 Dec;14(S1). [doi: 10.1186/1472-6955-14-S1-S12]
19. Cadbury B, Monaghan N. Lighting guided 9: lighting for communal residential buildings. The Society of Light and Lighting (CIBSE). 2022. URL: https://www.cibse.org/knowledge-research/knowledge-portal/lg9-lighting-for-communal-residential-buildings-2022 [accessed 2024-11-01]
20. Figueiro MG, Gonzales K, Pedler D. Designing with circadian stimulus. lighting design and applications (LD+A) the magazine of the illuminating engineering society of north america (IESNA) (2016). Lighting Research Center. 2016. URL: https://www.lrc.rpi.edu/resources/newsroom/LDA_CircadianStimulus_Oct2016.pdf [accessed 2024-11-05]
21. Brown TM, Brainard GC, Cajochen C, et al. Recommendations for daytime, evening, and nighttime indoor light exposure to best support physiology, sleep, and wakefulness in healthy adults. PLOS Biol 2022 Mar;20(3):e3001571. [doi: 10.1371/journal.pbio.3001571] [Medline: 35298459]
22. Lumileds. Luxeon 2835 color line. 2021. URL: https://lumileds.com/wp-content/uploads/2021/10/DS236-luxeon-2835-color-line-datasheet.pdf [accessed 2024-11-01]
23. Price L. User guide to the α-opic toolbox for implementing CIE S 026. CIE: International Commission on Illumination. 2020. URL: https://files.cie.co.at/CIE%20S%20026%20alpha-opic%20Toolbox%20User%20Guide.pdf [accessed 2024-11-01]
24. IBM SPSS statistics for windows, armonk, NY. : IBM Corp; 2021. URL: https://www.ibm.com/products/spss-statistics [accessed 2024-11-01]
25. Li J, Somers VK, Lopez-Jimenez F, Di J, Covassin N. Demographic characteristics associated with circadian rest-activity rhythm patterns: a cross-sectional study. Int J Behav Nutr Phys Act 2021 Aug 18;18(1):107. [doi: 10.1186/s12966-021-01174-z] [Medline: 34407852]
26. van Lieshout-van Dal E, Snaphaan L, Bongers I. Biodynamic lighting effects on the sleep pattern of people with dementia. Build Environ 2019 Mar;150:245-253. [doi: 10.1016/j.buildenv.2019.01.010]
27. Aarden-van Delft N, Peeters M, Snaphaan L. Biodynamic lighting and functional disability; a single-case experimental design in three community dwelling people with dementia. Appl Sci (Basel) 2021;11(20):9433. [doi: 10.3390/app11209433]
28. Hjetland GJ, Kolberg E, Pallesen S, et al. Ambient bright light treatment improved proxy-rated sleep but not sleep measured by actigraphy in nursing home patients with dementia: a placebo-controlled randomised trial. BMC Geriatr 2021 May 17;21(1):312. [doi: 10.1186/s12877-021-02236-4] [Medline: 34001024]

29. Jingen L, Lopez-Jimenez F, et al. An explorative investigation of the effect of naturalistic light on agitation-associated behaviour in nursing home residents with dementia: a pilot study. Health Environ Res Des J 2022;16(2):1-9. [doi: 10.1177/19375867221146154]
30. Sust CA, Dehoff P, Lang D, Lorenz D. Improved quality of life for resident dementia patients: St Katharina research project in Vienna. Zumbotel tech report. Zumbotel. 2012. URL: https://www.zumtobel.com/PDB/teaser/EN/Study_health_and_care_stkatharina.pdf [accessed 2024-11-05]
31. Wong KY, Dunn FA, Berson DM. Photoreceptor adaptation in intrinsically photosensitive retinal ganglion cells. Neuron 2005 Dec 22;48(6):1001-1010. [doi: 10.1016/j.neuron.2005.11.016] [Medline: 16364903]
32. Figueiro MG, Sahin L, Kalsher M, Plitnick B, Rea MS. Long-term, all-day exposure to circadian-effective light improves sleep, mood, and behavior in persons with dementia. J Alzheimers Dis Rep 2020 Aug 4;4(1):297-312. [doi: 10.3233/ADR-200212] [Medline: 33024938]
33. Colenda CC, Cohen W, McCall WV, Rosenquist PB. Phototherapy for patients with Alzheimer disease with disturbed sleep patterns: results of a community-based pilot study. Alzheimer Dis Assoc Disord 1997 Sep;11(3):175-178. [doi: 10.1097/00002093-199709000-00011] [Medline: 9305504]
34. Baandrup L, Jennum PJ. Effect of a dynamic lighting intervention on circadian rest-activity disturbances in cognitively impaired, older adults living in a nursing home: a proof-of-concept study. Neurobiol Sleep Circadian Rhythms 2021 Nov;11:100067. [doi: 10.1016/j.nbscr.2021.100067] [Medline: 34095610]
35. Saidane HA, Rasmussen T, Andersen K, Iversen HK, West AS. An explorative investigation of the effect of naturalistic lght on agitation-associated behavior in nursing home residents with dementia: a pilot study. HERD 2023 Apr;16(2):146-154. [doi: 10.1177/19375867221146154] [Medline: 36573255]
36. Burns A, Allen H, Tomenson B, Duignan D, Byrne J. Bright light therapy for agitation in dementia: a randomized controlled trial. Int Psychogeriatr 2009 Aug;21(4):711-721. [doi: 10.1017/S1041610209008886] [Medline: 19323872]
37. Forbes D, Blake CM, Thiessen EJ, Peacock S, Hawranik P. Light therapy for improving cognition, activities of daily living, sleep, challenging behaviour, and psychiatric disturbances in dementia. Cochrane Database Syst Rev 2014 Feb 26;2014(2):CD003946. [doi: 10.1002/14651858.CD003946.pub4] [Medline: 24574061]
38. Lyketsos CG, Lindell Veiel L, Baker A, Steele C. A randomized, controlled trial of bright light therapy for agitated behaviors in dementia patients residing in long-term care. Int J Geriatr Psychiatry 1999 Jul;14(7):520-525. [Medline: 10440971]
39. Bromundt V, Wirz-Justice A, Boutellier M, et al. Effects of a dawn-dusk simulation on circadian rest-activity cycles, sleep, mood and well-being in dementia patients. Exp Gerontol 2019 Sep;124:110641. [doi: 10.1016/j.exger.2019.110641] [Medline: 31252161]
40. Kolberg E, Hjetland GJ, Thun E, et al. The effects of bright light treatment on affective symptoms in people with dementia: a 24-week cluster randomized controlled trial. BMC Psychiatry 2021 Jul 28;21(1):377. [doi: 10.1186/s12888-021-03376-y] [Medline: 34320937]
41. van Lieshout-van Dal E, Snaphaan L, de Kort Y, Bouwmeester S, Bongers I. Impact of dynamic light exposure on sleep-wake pattern and BPSD in people with dementia living at home. Des for Health 2023 Mar 28;7(1):64-81. [doi: 10.1080/24735132.2023.2183622]
42. van Lieshout-van Dal E, Snaphaan L, Bouwmeester S, de Kort Y, Bongers I. Testing a single-case experimental design to study dynamic light exposure in people with dementia living at home. Appl Sci (Basel) 2021;11(21):10221. [doi: 10.3390/app112110221]
43. Aarts MPJ, Aries MBC, Straathof J, Hoof JV. Dynamic lighting systems in psychogeriatric care facilities in the Netherlands: a quantitative and qualitative analysis of stakeholders' responses and applied technology. Indoor Built Environ 2015 Aug;24(5):617-630. [doi: 10.1177/1420326X14532387]
44. Figueiro MG, Steverson B, Heerwagen J, et al. The impact of daytime light exposures on sleep and mood in office workers. Sleep Health 2017 Jun;3(3):204-215. [doi: 10.1016/j.sleh.2017.03.005] [Medline: 28526259]

Abbreviations

CS: circadian stimulus
ipRGC: intrinsically photosensitive retinal ganglion cell
IS: interdaily stability
IV: intradaily variation
M-EDI: Melanopic Equivalent Daylight Illuminance
QUALIDEM: Quality of Life for People With Dementia
RA: relative amplitude

Edited by L Buis; submitted 31.01.24; peer-reviewed by L Rossi, R Nilforooshan; revised version received 17.09.24; accepted 11.10.24; published 29.11.24.

Please cite as:
Turley K, Rafferty J, Bond R, Mulvenna M, Ryan A, Crawford L
Evaluating the Impact of a Daylight-Simulating Luminaire on Mood, Agitation, Rest-Activity Patterns, and Social Well-Being Parameters in a Care Home for People With Dementia: Cohort Study
JMIR Mhealth Uhealth 2024;12:e56951
URL: https://mhealth.jmir.org/2024/1/e56951
doi:10.2196/56951

Automated Pain Spots Recognition Algorithm Provided by a Web Service–Based Platform: Instrument Validation Study

Corrado Cescon[1], PhD; Giuseppe Landolfi[2], MSc; Niko Bonomi[2], MSc; Marco Derboni[3], MSc; Vincenzo Giuffrida[3], MSc; Andrea Emilio Rizzoli[3], PhD; Paolo Maino[4], MD; Eva Koetsier[4], MD, PhD; Marco Barbero[1], PhD

[1]Rehabilitation Research Laboratory 2rLab, Department of Business Economics, Health and Social Care, University of Applied Sciences and Arts of Southern Switzerland, Via Violino 11, Manno, Switzerland

[2]Institute of Systems and Technologies for Sustainable Production, University of Applied Sciences and Arts of Southern Switzerland, Lugano, Switzerland

[3]IDSIA Dalle Molle Institute for Artificial Intelligence, USI-SUPSI, Lugano, Switzerland

[4]Pain Management Center, Division of Anaesthesiology, Department of Acute Medicine, Neurocenter of Southern Switzerland, Regional Hospital of Lugano, Lugano, Switzerland

Corresponding Author:
Corrado Cescon, PhD
Rehabilitation Research Laboratory 2rLab, Department of Business Economics, Health and Social Care, University of Applied Sciences and Arts of Southern Switzerland, , Via Violino 11, Manno, , Switzerland

Abstract

Background: Understanding the causes and mechanisms underlying musculoskeletal pain is crucial for developing effective treatments and improving patient outcomes. Self-report measures, such as the Pain Drawing Scale, involve individuals rating their level of pain on a scale. In this technique, individuals color the area where they experience pain, and the resulting picture is rated based on the depicted pain intensity. Analyzing pain drawings (PDs) typically involves measuring the size of the pain region. There are several studies focusing on assessing the clinical use of PDs, and now, with the introduction of digital PDs, the usability and reliability of these platforms need validation. Comparative studies between traditional and digital PDs have shown good agreement and reliability. The evolution of PD acquisition over the last 2 decades mirrors the commercialization of digital technologies. However, the pen-on-paper approach seems to be more accepted by patients, but there is currently no standardized method for scanning PDs.

Objective: The objective of this study was to evaluate the accuracy of PD analysis performed by a web platform using various digital scanners. The primary goal was to demonstrate that simple and affordable mobile devices can be used to acquire PDs without losing important information.

Methods: Two sets of PDs were generated: one with the addition of 216 colored circles and another composed of various red shapes distributed randomly on a frontal view body chart of an adult male. These drawings were then printed in color on A4 sheets, including QR codes at the corners in order to allow automatic alignment, and subsequently scanned using different devices and apps. The scanners used were flatbed scanners of different sizes and prices (professional, portable flatbed, and home printer or scanner), smartphones with varying price ranges, and 6 virtual scanner apps. The acquisitions were made under normal light conditions by the same operator.

Results: High-saturation colors, such as red, cyan, magenta, and yellow, were accurately identified by all devices. The percentage error for small, medium, and large pain spots was consistently below 20% for all devices, with smaller values associated with larger areas. In addition, a significant negative correlation was observed between the percentage of error and spot size (R=−0.237; P=.04). The proposed platform proved to be robust and reliable for acquiring paper PDs via a wide range of scanning devices.

Conclusions: This study demonstrates that a web platform can accurately analyze PDs acquired through various digital scanners. The findings support the use of simple and cost-effective mobile devices for PD acquisition without compromising the quality of data. Standardizing the scanning process using the proposed platform can contribute to more efficient and consistent PD analysis in clinical and research settings.

(JMIR Mhealth Uhealth 2024;12:e53119) doi:10.2196/53119

KEYWORDS

pain drawing; image processing; body charts; scan; pain; draw; drawing; scanner; scanners; app; apps; applications; device; devices; image; images; smartphone; smartphones; scale; musculoskeletal; body chart; accuracy; reliability; accurate; reliable; picture; pictures; mobile phone

Introduction

Musculoskeletal pain is a frequent problem that affects a significant portion of the population and can have a major impact on quality of life [1]. Understanding the causes and mechanisms underlying musculoskeletal pain is crucial for the development of effective treatments and enhancement of patient outcomes. Moreover, investigating musculoskeletal pain contributes to the advancement of our understanding of anatomy, physiology, and pain mechanisms, with potential implications for comprehending and managing pain [2,3].

There are several methods for measuring muscle pain, including self-report measures, behavioral measures, and physiological measures [4]. Self-report measures involve asking the person to rate his or her level of pain on a scale, such as the Visual Analog Scale or the Numeric Rating Scale [5]. A promising way of evaluating pain using drawings is known as the Pain Drawing Scale [6-10].

Digital technologies have had a significant influence on the evolution of pain drawings (PDs), with different applications in the field of medical apps [11].

The fields of application of PDs include diagnosis of low back pain disorders; paresthesias evoked by implanted neurological stimulators; depiction of orofacial pain, such as headaches and toothaches; and evaluation of users of electric wheelchairs with pain located in the back, buttocks region, and so forth.

Body charts can also illustrate other types of sensory experiences such as numbness, tingling, hypoesthesia, or allodynia [12].

While digitally acquired PDs offer advantages, many studies demonstrate sophisticated analyses of scanned or digitized pen-and-paper PDs, showcasing the versatility of digital image processing. This capability enables the digitization and analysis of extensive collections of pen-and-paper pain diagrams, making it adaptable to various settings and needs.

In this technique, the individuals are instructed to color the area where they are experiencing pain, and the picture is then rated on a scale based on the amount of pain depicted [13]. This can be a useful tool for measuring pain in individuals who have difficulty verbalizing their pain experience, such as young children or nonverbal individuals [14]. However, it is important to keep in mind that PDs can be subjective and may be influenced by factors such as the person's cultural background or level of education [15]. PDs typically consist of body charts with different views of the human body (dorsal, ventral, and side) or subportions (head, hand, etc), and patients are instructed to color with a marker the area where they experience pain. The body charts can be divided into regions such as the Margolis regions [16]. This technique is used to describe and categorize the location of musculoskeletal pain in the body. The regions are based on anatomical divisions of the body, including the neck, upper extremities, low back, and lower extremities [17].

The purpose of the Margolis regions is to provide a standardized and easily understood way to describe the location of musculoskeletal pain, which can help with diagnosis and treatment planning [18,19].

PDs can be analyzed in a variety of ways, depending on the purpose of the analysis and the method used to create the drawings. The most common method of analyzing PDs is the measurement of the size of the pain region. This can be done using computer software or manual measurement techniques [20].

There are different software programs available for analyzing PDs [21-25]. These programs can be used to perform both qualitative and quantitative analyses of PDs, depending on the specific software and the features it offers [26]. Some researchers also introduced sex-specific body charts in order to facilitate the communication of pain for women [17,27,28].

Some of the features offered by pain analysis software may include image digitization (allowing the conversion of traditional paper drawings into digital format for analysis), image scaling (allowing the adjustment of the size of the PDs to match a reference scale), image analysis (using algorithms to automatically identify and quantify features of the PD, such as the size and shape of the pain region), and data visualization (displaying the results of the analysis in a clear and easy-to-understand format, such as graphs or heat maps).

Submitting paper PDs to patients is simpler than using drawing applications running on tablets. Anyway, the use of PDs is not indicated in patients with vision impairment or in preschool children, although some studies investigated the application in teenagers. Being a self-assessed measure the patients should not have cognitive impairments, or diseases including misperception of their body.

While digital drawings can be easily edited and manipulated, and the tools available on a tablet can offer a wider range of color options and effects, paper PDs are largely used in clinical settings. This preference stems from the fact that many patients feel more comfortable using the pen-on-paper approach rather than digital devices [29-31].

There are different methods for scanning PDs, including using a flatbed scanner, a device that scans flat, thin documents placed on a glass window; a handheld scanner, a portable device that can scan images while being moved over them; a drum scanner, a high-end scanner that uses a rotating cylinder to capture the image; a multifunctional printer scanner, a printer that also includes a scanner function; and a virtual scanner, a software that can use a camera to scan images.

To date, there is no standardization in scanning PDs. The existing softwares for PD acquisition work with specific body charts and do not allow a direct comparison between using the same drawing. The aim of this study was to evaluate the accuracy of PD analysis performed by a web platform using different digital scanners. The objective of this study was to demonstrate that simple and relatively cheap mobile devices can be used to acquire PDs without loss of information.

Methods

Ethical Considerations

We did not involve patients, subjects, or animals. The data set was generated through a computer simulation; thus, there was no need to have ethical approval.

Sketch Your Pain Platform

The proposed analysis was performed using a web platform. The main features of this distributed web application (currently available on a local server [32]) are as follows:

1. *Knowledge-base management*: the platform allows the collection of patient's data (biometric, pain history, applied therapies, diagnoses, etc).
2. *PD acquisition*: PDs can be uploaded both digitally and from paper (see details in the following section).
3. *Basic PD analysis*: each pain spot is analyzed individually (ie, number of pixels, barycenter, etc).
4. *Smart analysis services*: the platform provides a plug-in–based mechanism that allows the implementation of additional analysis within the platform. In this way, researchers can apply specific innovative tools to the PDs stored in the database [33].

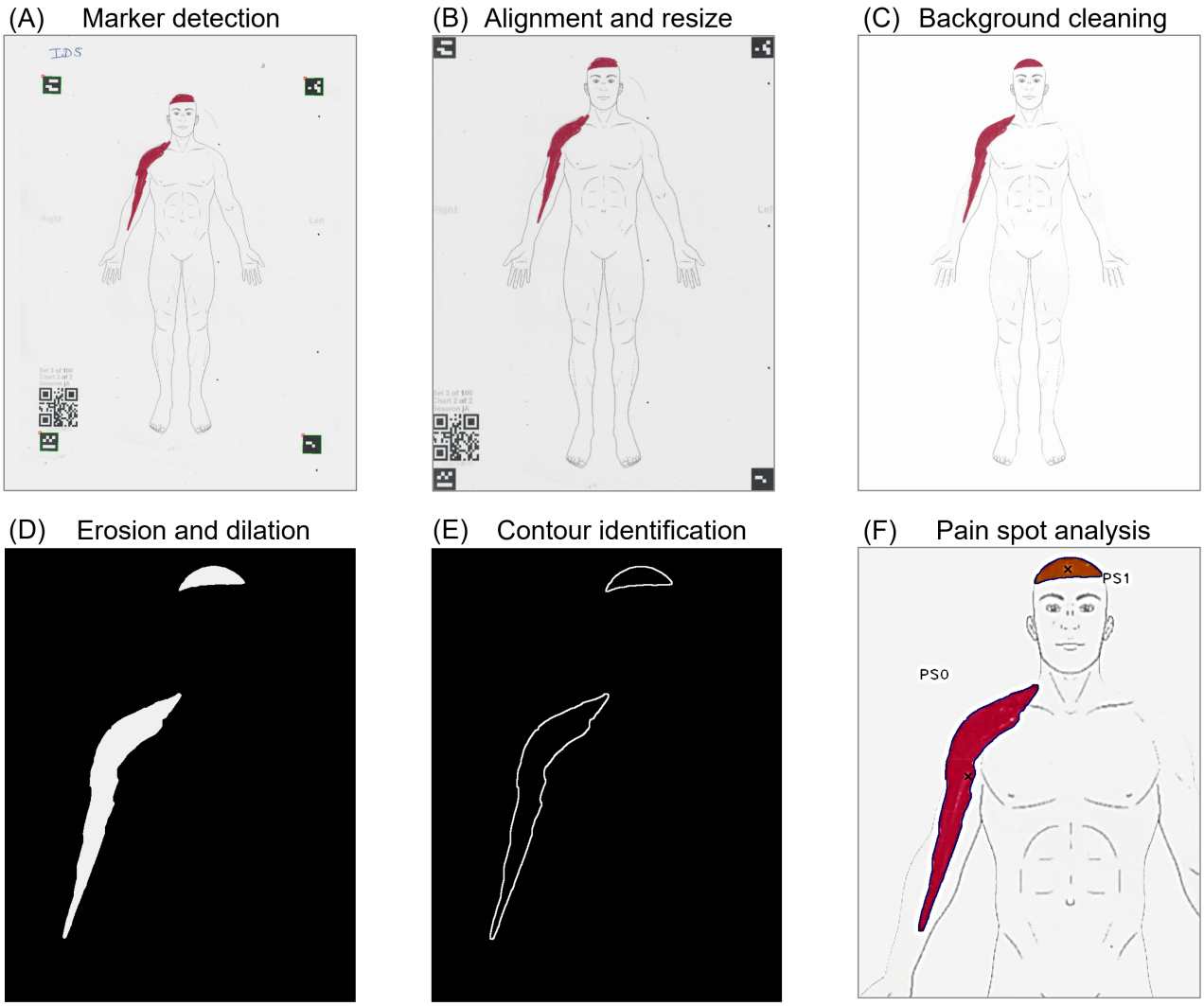
A paper PD can be imported in two ways: (1) it can be digitally imported by using the specific acquisition tool that allows for drawing directly on a tablet, using a digital pen on the touch screen (available on a local server [34]), or (2) it can be manually imported as a PDF file (the platform allows one to download a PDF file including empty body charts with a unique QR code, which can be printed, filled manually using a color marker, and scanned as a PDF file).

When a PD is generated digitally, it is already aligned with the body charts, while for the paper drawings, the process is more complex and can be summarized as follows. The body chart and all related information are identified thanks to the QR code, including the information of the protocol, subject ID, gender, and view of the body chart, and are stored in a database. The platform code and data are stored on a local server. The patient names and sensitive data are anonymized using codes that are available only to the operators.

The scanned image is aligned and cropped using the 4 markers at the corners as pivots (Figure 1A).

The image is resized in order to have the same number of pixels of the body charts stored in the platform (ie, 2048×1536 pixels; Figure 1B). The areas outside the body chart are removed using a mask image. This step also allows for the removal of all possible out-of-body staining errors (Figure 1C). The pain spots (that should be drawn in color and not in any shade of gray) are identified and isolated from the body chart by computing the SD of each pixel (in this way, the SD of black pixels [0, 0, 0] and white pixels [255, 255, 255] is equal to 0, while a red pixel [255, 0, 0] has an SD of 147.2). The optimal threshold for the minimum SD (based on preliminary tests) that works best in extracting pain spots from the body chart with different conditions of light and colors is 25. In this way, the color image is converted into a Boolean matrix where ones correspond to pixels with pain. An algorithm for extrusion and subsequent erosion is applied to the Boolean image in order to fill possible gaps that can happen when the user is using a sharp marker (Figure 1D). All contours of the pain spots and the potential holes in them are identified by means of the Canny edge detection algorithm [35]. The pain spots whose contours contain fewer than 20 points/pixels are removed. Likewise, the holes present in the spots, whose contours contain less than 15 points/pixels, are removed. Further, the pain spots smaller than 100 pixels are removed. Likewise, the holes smaller than 150 pixels are removed (Figure 1E). The individual pain spots are identified by an image segmentation algorithm, and for each spot, the area in pixels and the coordinates of the centroid are computed. The final result of the process is shown in Figure 1F.

Figure 1. Pain spot detection process. (A) The 4 markers at the corners and the QR code are identified. (B) The pain drawing is aligned and scaled. (C) The areas outside the body chart are removed. (D) Pain drawings are separated from the background and eroded in order to correct the imperfections due to pen drawing. (E) Pain spot contours are identified and small holes are removed. (F) Each pain spot is analyzed to extract area and position. PS: pain spot.



Generation of Artificial PDs

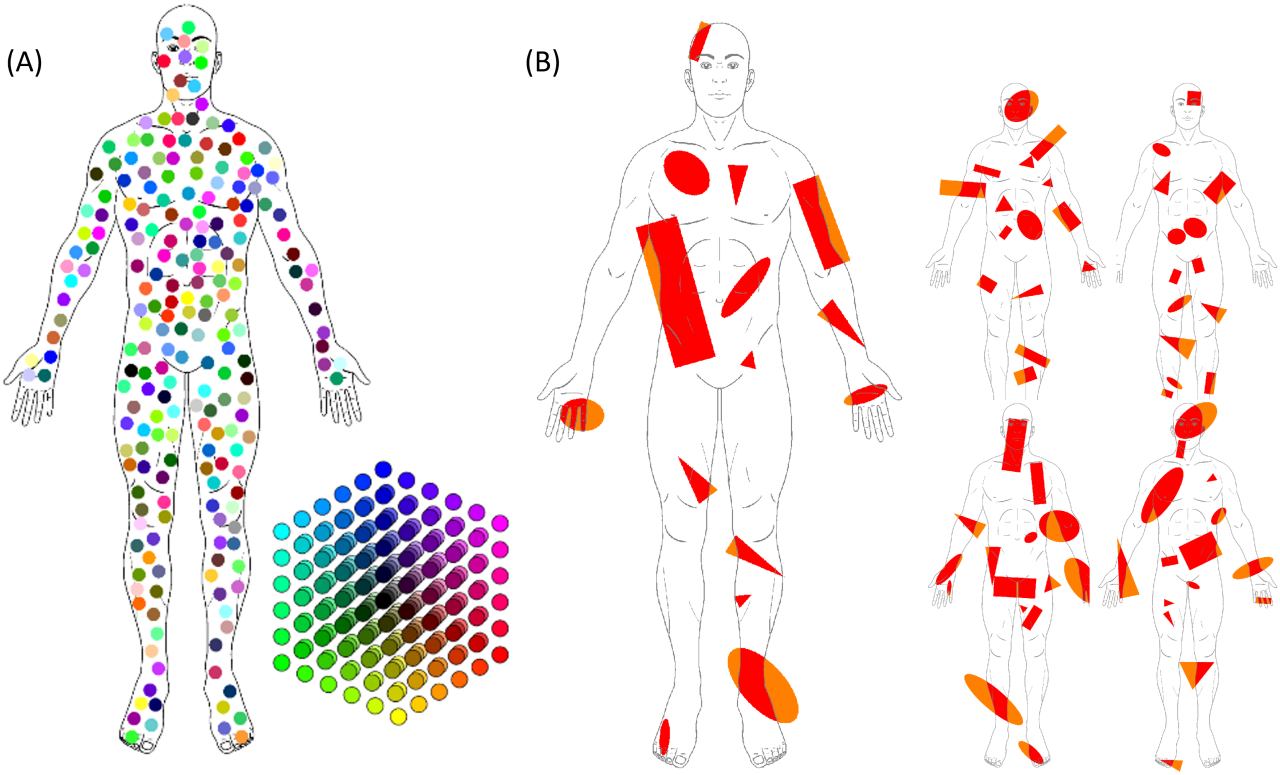
For the present protocol, 2 sets of PDs were generated with a homemade MATLAB (MathWorks) code. We decided to test the platform by using artificial PDs in order to have complete control of the process and of the analysis. For each of the pain spots generated randomly, we had information on pain location (barycenter of the pain spot), area in square pixels, and shade of color in the red, green, and blue (RGB) scale, and with these data, we could assess the performance of each of the scanning devices. A preliminary study was conducted on the platform using different scanning devices on PDs of real patients with similar results [36].

The body chart selected was a male frontal body chart, representing the contours of a full male body in frontal view (dimensions: 1536×2048 pixels).

Color Analysis

The first artificial PD was generated, adding 216 colored circles, which were 33 pixels in diameter, within the body chart map. The circles were randomly positioned within the body chart in order to be nonoverlapping and not touching each other. The colors were chosen in order to uniformly span the RGB color cube, using 6 different intensities for each color. Since the color depth was defined on a range from 0 to 255 (1 byte), the values of each color were 0, 51, 102, 153, 204, and 255. In this way, the total number of colors was 6^3=216, ranging from black (0, 0, 0) to white (255, 255, 255) and including all combinations of RGB (ie, [0, 0, 0], [0, 0, 51], [0, 0, 102],…[51, 0, 0],…[255, 255, 255]; Figure 2A).

Figure 2. Representation of the artificial pain drawings generated with MATLAB. (A) A total of 216 colored circles with a 30-pixel diameter were randomly located within the area of the body chart. The colors were uniformly distributed in the RGB color cube. (B) Five body charts with randomly generated shapes. RGB: red, green, and blue.



Area and Location Analysis

The second set of PDs was composed of 5 artificial PDs generated by adding several red shapes (ellipses, rectangles, and triangles) to the same body chart; the shapes were generated with random sizes, orientations, and positions and could overlap and be partially outside the body chart mask (Figure 2B). The red color was chosen mainly because it can be easily associated with pain in a body chart; in addition, in the RGB cube, the red color is located in one of the vertices (ie, it has the highest SD value among triplets of RGB values, together with yellow, magenta, cyan, blue, and green), and it is easy to find red pens or markers in common shops.

Each of the 2 sets was printed in color, using the same printer (Sharp MX-7580) to print 11 copies; markers were added at the 4 corners, and a QR code was added at the bottom left side. The markers and the QR code were included in order to allow the platform to align the images and add the PDs to the internal database.

Selection of Scanning Devices

The 11 sets of drawings were then scanned using different devices and apps (Table 1).

We selected 3 flatbed scanners with different sizes and prices: 1 professional office printer or scanner that was available in our university (Sharp MX-4070; price about US $5000), 1 portable flatbed scanner (Canon Lide 220; ~US $50), and a home printer or scanner (HP Envy 4500; ~US $300).

In addition, we selected 3 smartphones with different price ranges: iPhone 12 (~US $1000), Samsung Galaxy (~US $400), and Ulefone Armor (~US $100). All the 3 devices were using the same app for scanning images (vFlat scan), in order to compare only the hardware of the devices.

Moreover, for the cheapest smartphone, we selected 6 free apps available in the android apps Google Play repository. The apps were selected according to their popularity and ranking based on users' comments.

For each scanner, a PDF file was generated including the corresponding set of images (1 with colored circles and 5 with red shapes). The PDF files were uploaded in the sketch your pain platform [32].

Table . List of devices used to scan the artificial pain drawings[a].

Type	Device model	Resolution	Price (US $)	App
Flatbed	Sharp MX-4070	300 dpi[b]	~5000	—[c]
Flatbed	Canon Lide 220	300 dpi	~50	—
Flatbed	HP Envy 4500	300 dpi	~300	—
Smartphone	iPhone 12	12 Mpx[d]	~1000	vFlat Scan
Smartphone	Samsung Galaxy S10 Lite	32 Mpx	~400	vFlat Scan
Smartphone	Ulefone Armor-X	13 Mpx	~100	vFlat Scan
Smartphone	Ulefone Armor-X	13 Mpx	~100	TapScanner
Smartphone	Ulefone Armor-X	13 Mpx	~100	Simple Scan
Smartphone	Ulefone Armor-X	13 Mpx	~100	Fast Scanner
Smartphone	Ulefone Armor-X	13 Mpx	~100	CamScanner
Smartphone	Ulefone Armor-X	13 Mpx	~100	TurboScan

[a]The flatbed devices generated the PDF files using proprietary software, while smartphones needed an app to generate the PDF using the camera.

[b]dpi: dots per inch.

[c]Not available.

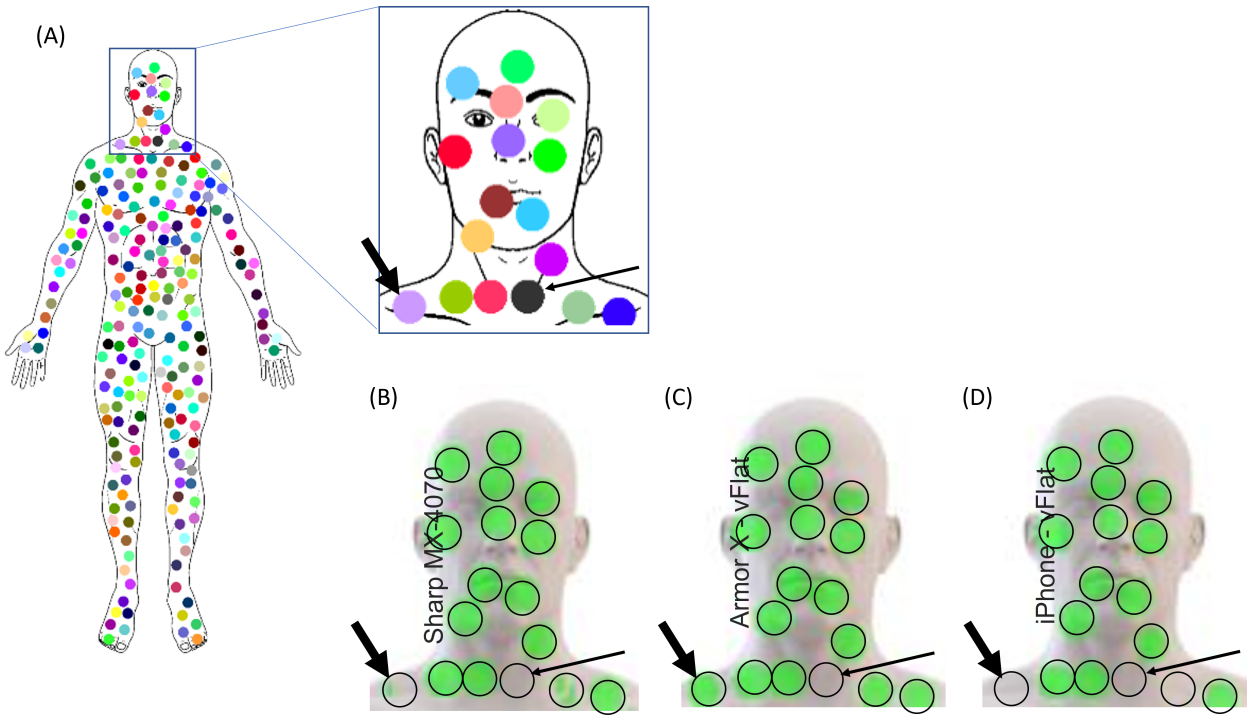
[d]Mpx: megapixels.

Image Processing

The original area of each pain spot generated with MATLAB was computed as well as the coordinates of the centroid of each pain spot. The sketch your pain platform identified the QR codes and allowed the processing of each identified pain spot, providing area in pixels and coordinates of the centroid.

For the analysis of colors, we analyzed whether the platform was able to identify a pain spot corresponding to each of the locations where colored circles were generated. If the area of the identified pain spot was larger than a fixed threshold (90% of the theoretical area; eg, 450 out of 500 pixels), then the pain spot was counted (Figure 3).

Figure 3. Examples of identification of pain spots from colored circles. (A) The original drawing and the output of the platform algorithm for 3 different devices are shown. The green color represents an identified pain spot. The purple circle on the bottom left corner (indicated with the thick arrow) was not identified by the iPhone (D) and Sharp scanner (B), but it was identified by the Armor phone (C). The black circle on the neck (indicated with the thin arrow) was not identified by any of the devices.



Statistical Analysis

The variables used for the statistics were the area (A) of pain spots in square pixels and the coordinates of the centroid (x, y) of each pain spot in pixels. The variables computed for each scanning device were compared with the variables computed for the corresponding pain spots on the original artificial PDs generated with MATLAB. The percentage area error (E) was computed as the difference between the 2 areas divided by the area computed on original PD and expressed as a percentage. The distance (D) between the theoretical centroid of the pain spot computed on original image and the centroid of the pain spot identified by each device was also computed and expressed in pixels.

Intraclass correlation coefficient (ICC) estimates (and their 95% CIs) of pain area and barycenter coordinates were calculated using MATLAB and a 1-way mixed-effects model.

In addition, standard error of measurement (SEM) and minimal clinical differences were computed for pain area and barycenter position.

Descriptive statistics is presented with box and whisker plots with median and IQR values.

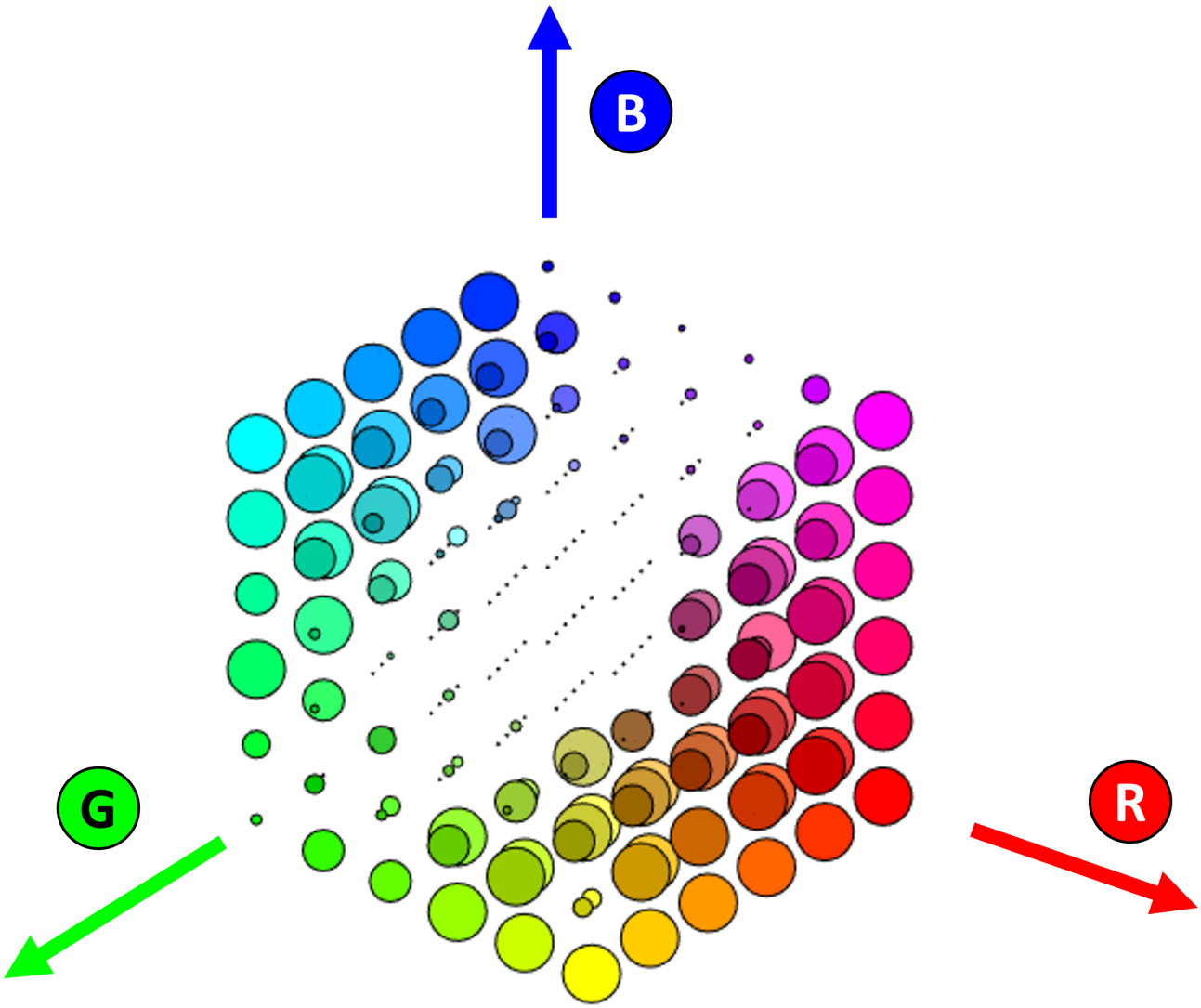
Results

Color Analysis

A Boolean table with the results of pain spot identification was generated with 216 lines (1 for each color) and 11 columns (1 for each scanner). A graphical representation of the table is represented in Figure 4.

Each color is represented by a circle in the 3D color cube, and the size of each circle is proportional to the number of scanners that were able to identify that color (ie, if the area of the identified pain spot was larger than 90% of the printed circle). The maximum circle diameter was set to two-thirds of the spacing between adjacent circles. In this way, it is easy see which colors are best for PDs. As expected, the colors with identical values in RGB triplets (ie, shades of gray, black, and white) were never identified by the software, and colors with low values of SD in RGB triplets (eg, low-saturation colors) were not identified with most devices. The colors that were identified with all devices are located close to the corners of the color cube (ie, high-saturation colors, such as red, cyan, magenta, and yellow).

Figure 4. Representation of the performance of the algorithm in identifying different color circles. The middle diagonal (from white to black where RGB components are equal) and the colors located close to the central diagonal are not identified by the algorithm, while colors such as red, magenta, and yellow are identified by all devices. RGB: red, green, and blue.



Area and Location Analysis

Figure 5 shows the distribution of areas of the pain spots generated artificially and randomly distributed on each of the 5 body charts. The pain spots were divided into 3 categories according to their area in square pixels ($A<50^2$: small; $50^2 \leq A<100^2$: medium; and $A \geq 100^2$: large).

The ICC for pain area was 0.99, with a 95% CI 0.99 - 0.99 ($F_{74,750}$=1.44e+04). In addition, the ICC and CI values for barycenter coordinates were above 0.99 (x-coordinate: $F_{84,930}$=4.55e+05; y-coordinate: $F_{103,843}$=5.12e+05).

Table 2 shows the SEM and minimal detectable change values for each device compared with the theoretical value.

Figure 6A shows the percentage error of pain extent for each device for the 3 categories of pain spot areas. For all devices, the percentage error was below 20% for small, medium, and big pain spots, with lower values associated with bigger areas. A significant negative correlation was observed between percentage of error and spot size (R=−0.237; P=.04; Figure 6B).

Figure 7 shows the percentage error of distance between the theoretical location of the centroid of each pain spot and the location of the centroid of the identified pain spot. The distribution of the distances was always below 5 pixels except for the Armor device with the TurboScan app (11th column).

Figure 5. Representation of the distribution of shapes according to their size. The 3 colors are used to represent the 3 categories that were used for further analysis The image in the legend shows the thresholds used to divide the categories (as square shapes).

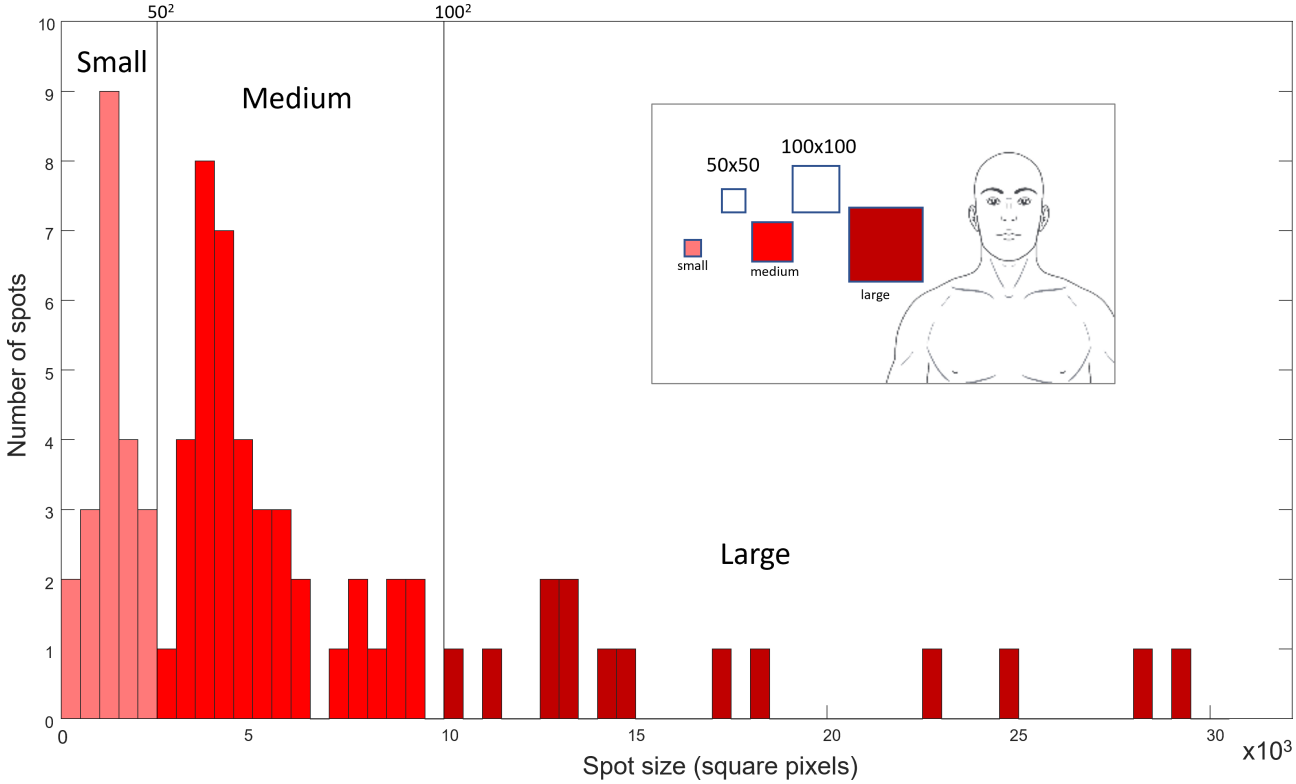


Table . Standard error of measurement and minimal detectable change for the identification of pain area and for the barycenter distance for each device compared with the theoretical value.

Device model	Area SEM[a] (pixels^2)	Area MDC[b] (pixels^2)	Barycenter distance SEM (pixels)	Barycenter distance MDC (pixels)
Sharp MX-4070	262	513	1.5	2.9
Canon Lide 220	348	682	1.5	3.0
HP Envy 4500	219	429	1.5	3.0
iPhone—vFlat	134	262	1.0	2.1
Galaxy S10 lite—vFlat	172	336	1.4	2.6
Armor X—vFlat	182	356	1.3	2.5
Armor X—TapScanner	247	485	2.1	4.2
Armor X—Simple Scan	251	491	2.6	5.1
Armor X—Fast Scanner	287	563	2.4	4.8
Armor X—CamScanner	227	444	2.0	3.8
Armor X—TurboScan	505	989	4.0	7.8

[a]SEM: standard error of measurement.

[b]MDC: minimal detectable change.

Figure 6. (A) Distribution of errors in identifying the pain spot areas expressed in percentages. The 3-color box and whisker plots for each device represent the distribution of area error for each of the 3 categories (small, medium, and large pain spots). (B) Correlation between percentage of error and spot size (the regression line is indicated as a dashed line).

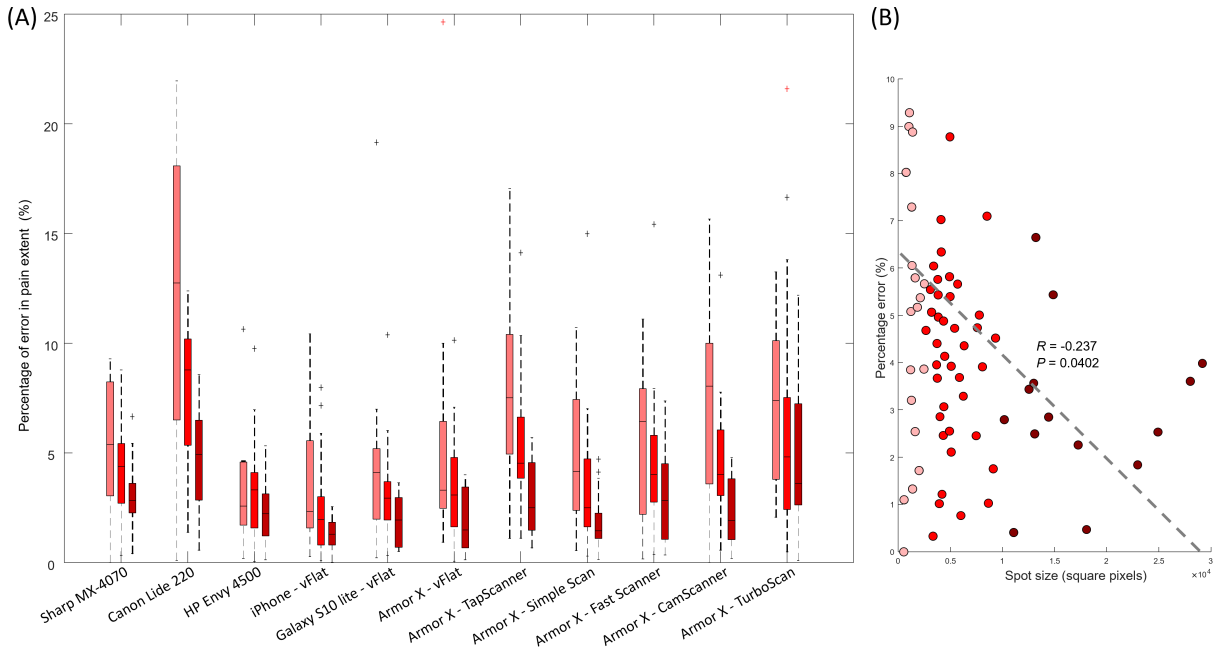
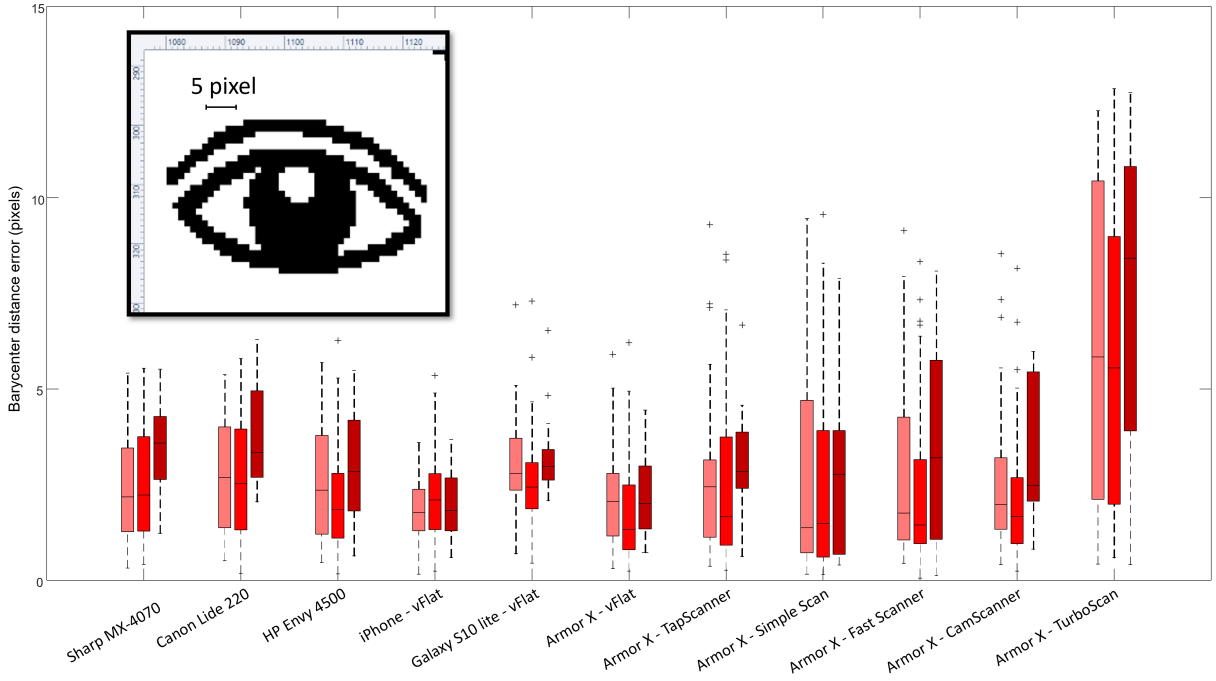


Figure 7. Distribution of errors in the location of the pain spots expressed in pixels. The 3-color box and whisker plots for each device represent the distribution of barycenter distance error for each of the 3 categories (small, medium, and large pain spots). The image of the eye shows the actual dimensions of 5 pixels in the full body chart (paper dimensions: 2048×1536 pixels).



Discussion

Principal Findings

The results of this study showed that the pen-and-paper drawings can be imported and processed with negligible differences using different devices. To our knowledge, there were no other studies focusing on the acquisition of paper pen drawings using mobile devices. Before conducting this study, we were largely using our platform, asking patients to use the red marker because we suspected that blue markers would be mistaken as black. As expected, indeed the red color is the best choice for multiple reasons: the red color is associated with the inflammatory process; thus, it is easy for a patient to visualize their own pain as a red spot (eg, compared with green or blue). Moreover, in shops, markers labeled as “red” are very similar to the theoretical

value of (255, 0, 0), while markers labeled as "green," or "blue" can have different darker or lighter shades.

Surprisingly, in our results, blue and green were identified only with few devices (mostly flatbed scanners). One of the reasons could be the ink of the printer, whose color was slightly different from what was expected. The illumination of the room could also have an impact, since the light emitted by different bulbs or neon could carry different wavelengths in different proportions. The light sensors of cameras and scanners could have different sensitivities to different wavelengths, and probably the red-light sensors have higher sensitivities.

When observing the printed page, we noticed that the blue ink was slightly darker than what we observed on the PC monitor, but it was difficult to objectively evaluate which one was correct, as we did not have a gold standard for each color. In summary, the results confirmed our initial prediction: the red marker is the optimal choice for PDs. However, in cases where the patient does not have a red marker available, we recommend using a flatbed scanner to generate the PDF for import to the platform. This approach helps minimize the bias resulting from external factors such as lights or photo LED sensitivity.

Regarding the accuracy of the pain spot identification, all the devices showed similar performance when using red color. The cheapest flatbed scanner showed larger errors probably due to the distortion of the image. When observing the digital image, we noticed that the proportions were slightly distorted (maybe due to the calibration of the motor or due to friction of the transmission chain). For this reason, the alignment process of the algorithm could not perfectly align the 4 markers in the corner; thus, the pain spot location had larger errors and extensions.

As expected, the smaller pain spots showed larger percentage errors but no differences in the barycenter location error, because of the distortion of pain spots due to the erosion process. The choice of the app for the mobile device had a significant effect on the percentage errors. In particular, the app named V-Flat showed better results than the others (when installed on the cheapest mobile device), leading to results that were comparable with high-rank mobile devices. The V-Flat app includes an algorithm that recognizes the corners of the paper and compensates for the distortions of the camera and even the distortions due to bent paper. For this reason, the results were as good as flatbed professional scanners (<5% of error). In general, the errors in identification of pain extent (<5%) and pain location (<5 pixels) were much lower than the precision of a subject in drawing or identifying their own pain and thus were lower than the minimum clinical significance of PDs [17,37,38].

Limitations

This study has some limitations that may introduce bias into the results. First, the acquisitions made with mobile devices were not conducted under controlled lighting conditions. Although all acquisitions took place during daylight hours without direct sunlight on the paper, variations in the time of day and weather conditions could have affected the colors identified by the devices. In addition, we did not calibrate the "white level" of the mobile phone camera. While some apps offered advanced camera settings for optimizing virtual scanners, we chose to use the default settings to maintain as close to a real-life environment as possible. As a result, we did not test under unusual lighting conditions (eg, candlelight, colored lamps, neon lights).

Another limitation is that we could not directly compare the performance of our platform with other existing methods, since the body charts are specific for each existing app.

Furthermore, we tested only 1 printer to print all the artificial PDs, which introduces potential bias. The colored circles in the first part of the study were positioned within the body chart but in different locations. This variation in placement could impact the results, as colors farther from the center of the paper may experience greater distortion due to misalignment. However, this approach was necessary to avoid printing an excessive number of papers.

The decision of using artificial PDs is due to the fact that we conducted 2 studies in parallel. The first study involved the use of actual PDs generated by humans [36], while in this study, we wanted to investigate different colors in order to investigate all the RGB space. In addition, the location of pain spots in PDs generated by humans depends on the pathology of patients, while in this case, we preferred to have a uniform distribution of pain spots with a priori known sizes and locations. Both studies showed that the platform has excellent results, but in this study, we were able to quantify the error since we knew the theoretical pain spot areas.

The sample size of mobile devices and flatbed scanners is small. However, the objective of the study was not to provide an exhaustive sample of devices but rather to demonstrate that even inexpensive devices are sufficient for accurately acquiring paper PDs.

Similarly, the sample of scan apps is also limited, and some of them are no longer available for free as of manuscript submission. The app market is continuously evolving, with new apps being released regularly. Our intention was to find a selection of free virtual scanners available in the Play Store, and again, our aim was to show that various apps perform similarly.

Conclusions

The system was already tested in real clinical settings and was shown to be easy to implement, easy to use, and well accepted. The acquisition of paper PDs using the proposed platform has been demonstrated to be robust and reliable across a wide range of scanning devices. The accuracy of pain extent and location analysis consistently falls within the error measurement range of PDs. The use of the proposed algorithm will enable the use of PD analysis in various clinical settings.

Authors' Contributions

CC participated in the conceptualization, methodology, software, data curation, and writing—original draft preparation. GL participated in the software and validation. NB led the visualization and investigation. VG participated in the investigation. MD contributed to data curation. EK participated in writing—review and editing. PM did the supervision. AER participated in project administration. MB participated in the conceptualization, funding acquisition, and writing—reviewing and editing.

Conflicts of Interest

None declared.

References

1. Puntillo F, Giglio M, Paladini A, et al. Pathophysiology of musculoskeletal pain: a narrative review. Ther Adv Musculoskelet Dis 2021;13:1759720X21995067. [doi: 10.1177/1759720X21995067] [Medline: 33737965]
2. Ekman EF, Koman LA. Acute pain following musculoskeletal injuries and orthopaedic surgery: mechanisms and management. Instr Course Lect 2005;54:21-33. [Medline: 15948432]
3. Graven-Nielsen T. Mechanisms and manifestations in musculoskeletal pain: from experimental to clinical pain settings. Pain 2022 Nov 1;163(Suppl 1):S29-S45. [doi: 10.1097/j.pain.0000000000002690] [Medline: 35984370]
4. Roulin MJ, Ramelet AS. Pain indicators in brain-injured critical care adults: an integrative review. Aust Crit Care 2012 May;25(2):110-118. [doi: 10.1016/j.aucc.2011.10.002] [Medline: 22104632]
5. Shafshak TS, Elnemr R. The visual analogue scale versus numerical rating scale in measuring pain severity and predicting disability in low back pain. J Clin Rheumatol 2021 Oct 1;27(7):282-285. [doi: 10.1097/RHU.0000000000001320] [Medline: 31985722]
6. Galer BS, Jensen MP. Development and preliminary validation of a pain measure specific to neuropathic pain: the Neuropathic Pain Scale. Neurology 1997 Feb;48(2):332-338. [doi: 10.1212/wnl.48.2.332] [Medline: 9040716]
7. Gracely RH, Kwilosz DM. The Descriptor Differential Scale: applying psychophysical principles to clinical pain assessment. Pain 1988 Dec;35(3):279-288. [doi: 10.1016/0304-3959(88)90138-8] [Medline: 3226757]
8. McCaffery M, Pasero C. Pain Clinical Manual, 2nd edition: Mosby; 1999.
9. Melzack R, Katz J. McGill Pain Questionnaire. In: Encyclopedia of Pain: Springer; 1971:1792-1794. [doi: 10.1007/978-3-642-28753-4]
10. Wong DL, Hockenberry MJ, Wilson D, Winkelstein ML, Schwartz P. Wong's Essentials of Pediatric Nursing, 6th edition: Mosby; 2001.
11. Shaballout N, Neubert TA, Boudreau S, Beissner F. From paper to digital applications of the pain drawing: systematic review of methodological milestones. JMIR Mhealth Uhealth 2019 Sep 5;7(9):e14569. [doi: 10.2196/14569] [Medline: 31489841]
12. Schmid AB, Ridgway L, Hailey L, et al. Factors predicting the transition from acute to persistent pain in people with 'sciatica': the FORECAST longitudinal prognostic factor cohort study protocol. BMJ Open 2023 Apr 5;13(4):e072832. [doi: 10.1136/bmjopen-2023-072832] [Medline: 37019481]
13. Suvinen TI, Kemppainen P, Le Bell Y, Kauko T, Forssell H. Assessment of pain drawings and self-reported comorbid pains as part of the biopsychosocial profiling of temporomandibular disorder pain patients. J Oral Facial Pain Headache 2016 Oct;30(4):287-295. [doi: 10.11607/ofph.1589] [Medline: 27792795]
14. Unruh A, McGrath P, Cunningham JS, Humphreys P. Children s drawings of their pain. Pain 1983 Dec;17(4):385-392. [doi: 10.1016/0304-3959(83)90170-7] [Medline: 6664684]
15. Grunnesjö M, Bogefeldt J, Blomberg S, Delaney H, Svärdsudd K. The course of pain drawings during a 10-week treatment period in patients with acute and sub-acute low back pain. BMC Musculoskelet Disord 2006 Dec 11;7:65. [doi: 10.1186/1471-2474-7-65] [Medline: 16901354]
16. Margolis RB, Tait RC, Krause SJ. A rating system for use with patient pain drawings. Pain 1986 Jan;24(1):57-65. [doi: 10.1016/0304-3959(86)90026-6] [Medline: 2937007]
17. Barbero M, Moresi F, Leoni D, Gatti R, Egloff M, Falla D. Test-retest reliability of pain extent and pain location using a novel method for pain drawing analysis. Eur J Pain 2015 Sep;19(8):1129-1138. [doi: 10.1002/ejp.636] [Medline: 25565607]
18. Balasch-Bernat M, Dueñas L, Aguilar-Rodríguez M, et al. The spatial extent of pain is associated with pain intensity, catastrophizing and some measures of central sensitization in people with frozen shoulder. J Clin Med 2021 Dec 28;11(1):154. [doi: 10.3390/jcm11010154] [Medline: 35011895]
19. Ginzburg BM, Merskey H, Lau CL. The relationship between pain drawings and the psychological state. Pain 1988 Nov;35(2):141-146. [doi: 10.1016/0304-3959(88)90221-7] [Medline: 3237428]
20. Türp JC, Kowalski CJ, O'Leary N, Stohler CS. Pain maps from facial pain patients indicate a broad pain geography. J Dent Res 1998 Jun;77(6):1465-1472. [doi: 10.1177/00220345980770061101] [Medline: 9649175]
21. Ali SM, Lau WJ, McBeth J, Dixon WG, van der Veer SN. Digital manikins to self-report pain on a smartphone: a systematic review of mobile apps. Eur J Pain 2021 Feb;25(2):327-338. [doi: 10.1002/ejp.1688] [Medline: 33113241]

22. Boudreau SA. Visualizing and quantifying spatial and qualitative pain sensations. Scand J Pain 2022 Oct 26;22(4):681-683. [doi: 10.1515/sjpain-2022-0098] [Medline: 36136613]
23. Kanellopoulos AK, Kanellopoulos EK, Dimitriadis Z, et al. Novel software for pain drawing analysis. Cureus 2021 Dec;13(12):e20422. [doi: 10.7759/cureus.20422] [Medline: 35047261]
24. Neubert TA, Dusch M, Karst M, Beissner F. Designing a tablet-based software app for mapping bodily symptoms: usability evaluation and reproducibility analysis. JMIR Mhealth Uhealth 2018 May 30;6(5):e127. [doi: 10.2196/mhealth.8409] [Medline: 29848470]
25. Shaballout N, Aloumar A, Neubert TA, Dusch M, Beissner F. Digital pain drawings can improve doctors' understanding of acute pain patients: survey and pain drawing analysis. JMIR Preprints. Preprint posted online on Jun 27, 2018. [doi: 10.2196/preprints.11412]
26. Dixit A, Lee M. Quantification of digital body maps for pain: development and application of an algorithm for generating pain frequency maps. JMIR Form Res 2022 Jun 24;6(6):e36687. [doi: 10.2196/36687] [Medline: 35749160]
27. Galve Villa M, Palsson TS, Cid Royo A, Bjarkam CR, Boudreau SA. Digital pain mapping and tracking in patients with chronic pain: longitudinal study. J Med Internet Res 2020 Oct 26;22(10):e21475. [doi: 10.2196/21475] [Medline: 33104012]
28. Egsgaard LL, Christensen TS, Petersen IM, Brønnum DS, Boudreau SA. Do gender-specific and high-resolution three dimensional body charts facilitate the communication of pain for women? A quantitative and qualitative study. JMIR Hum Factors 2016 Jul 20;3(2):e19. [doi: 10.2196/humanfactors.5693] [Medline: 27440737]
29. Campbell N, Ali F, Finlay AY, Salek SS. Equivalence of electronic and paper-based patient-reported outcome measures. Qual Life Res 2015 Aug;24(8):1949-1961. [doi: 10.1007/s11136-015-0937-3] [Medline: 25702266]
30. Noyes JM, Garland KJ. Computer- vs. paper-based tasks: are they equivalent? Ergonomics 2008 Sep;51(9):1352-1375. [doi: 10.1080/00140130802170387] [Medline: 18802819]
31. Touvier M, Méjean C, Kesse-Guyot E, et al. Comparison between web-based and paper versions of a self-administered anthropometric questionnaire. Eur J Epidemiol 2010 May;25(5):287-296. [doi: 10.1007/s10654-010-9433-9] [Medline: 20191377]
32. SYP Dashboard. SUPSI - Sketch Your Pain. 2020. URL: https://syp.spslab.ch [accessed 2024-08-19]
33. Luque-Suarez A, Falla D, Barbero M, et al. Digital pain extent is associated with pain intensity but not with pain-related cognitions and disability in people with chronic musculoskeletal pain: a cross-sectional study. BMC Musculoskelet Disord 2022 Jul 30;23(1):727. [doi: 10.1186/s12891-022-05700-3] [Medline: 35906575]
34. Sketch Your Pain—tablet. SUPSI - Sketch Your Pain. 2020. URL: https://syp.spslab.ch/tablet [accessed 2024-08-19]
35. Canny J. A computational approach to edge detection. IEEE Trans Pattern Anal Mach Intell 1986 Nov;8(6):679-698. [Medline: 21869365]
36. Barbero M, Cescon C, Schneebeli A, et al. Reliability of the pen-on-paper pain drawing analysis using different scanning procedures. J Pain Symptom Manage 2024 Feb;67(2):e129-e136. [doi: 10.1016/j.jpainsymman.2023.10.019] [Medline: 37898312]
37. Abichandani D, Barbero M, Cescon C, et al. Can people with chronic neck pain recognize their own digital pain drawing? Pain Physician 2020 Mar;23(2):E231-E240. [Medline: 32214308]
38. Leoni D, Falla D, Heitz C, et al. Test-retest reliability in reporting the pain induced by a pain provocation test: further validation of a novel approach for pain drawing acquisition and analysis. Pain Pract 2017 Feb;17(2):176-184. [doi: 10.1111/papr.12429] [Medline: 26992099]

Abbreviations

ICC: intraclass correlation coefficient
PD: pain drawing
RGB: red, green, and blue
SEM: standard error of measurement

Edited by L Buis; submitted 26.09.23; peer-reviewed by M Estai, P Gazerani; revised version received 22.04.24; accepted 13.05.24; published 27.08.24.

Please cite as:
Cescon C, Landolfi G, Bonomi N, Derboni M, Giuffrida V, Rizzoli AE, Maino P, Koetsier E, Barbero M
Automated Pain Spots Recognition Algorithm Provided by a Web Service–Based Platform: Instrument Validation Study
JMIR Mhealth Uhealth 2024;12:e53119
URL: https://mhealth.jmir.org/2024/1/e53119
doi:10.2196/53119

The Goldilocks Dilemma on Balancing User Response and Reflection in mHealth Interventions: Observational Study

Lyndsay A Nelson[1,2], PhD; Andrew J Spieker[3], PhD; Lauren M LeStourgeon[1,2], MPH; Robert A Greevy Jr[3], PhD; Samuel Molli[1,2], BS; McKenzie K Roddy[1,2], PhD; Lindsay S Mayberry[1,2,4], MS, PhD

[1]Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, United States

[2]Center for Health Behavior and Health Education, Vanderbilt University Medical Center, Nashville, TN, United States

[3]Department of Biostatistics, Vanderbilt University Medical Center, Nashville, TN, United States

[4]Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, United States

Corresponding Author:
Lyndsay A Nelson, PhD
Department of Medicine, Vanderbilt University Medical Center, , Nashville, TN, , United States

Abstract

Background: Mobile health (mHealth) has the potential to radically improve health behaviors and quality of life; however, there are still key gaps in understanding how to optimize mHealth engagement. Most engagement research reports only on system use without consideration of whether the user is reflecting on the content cognitively. Although interactions with mHealth are critical, cognitive investment may also be important for meaningful behavior change. Notably, content that is designed to request too much reflection could result in users' disengagement. Understanding how to strike the balance between response burden and reflection burden has critical implications for achieving effective engagement to impact intended outcomes.

Objective: In this observational study, we sought to understand the interplay between response burden and reflection burden and how they impact mHealth engagement. Specifically, we explored how varying the response and reflection burdens of mHealth content would impact users' text message response rates in an mHealth intervention.

Methods: We recruited support persons of people with diabetes for a randomized controlled trial that evaluated an mHealth intervention for diabetes management. Support person participants assigned to the intervention (n=148) completed a survey and received text messages for 9 months. During the 2-year randomized controlled trial, we sent 4 versions of a weekly, two-way text message that varied in both reflection burden (level of cognitive reflection requested relative to that of other messages) and response burden (level of information requested for the response relative to that of other messages). We quantified engagement by using participant-level response rates. We compared the odds of responding to each text and used Poisson regression to estimate associations between participant characteristics and response rates.

Results: The texts requesting the most reflection had the lowest response rates regardless of response burden (high reflection and low response burdens: median 10%, IQR 0%-40%; high reflection and high response burdens: median 23%, IQR 0%-51%). The response rate was highest for the text requesting the least reflection (low reflection and low response burdens: median 90%, IQR 61%-100%) yet still relatively high for the text requesting medium reflection (medium reflection and low response burdens: median 75%, IQR 38%-96%). Lower odds of responding were associated with higher reflection burden (P<.001). Younger participants and participants who had a lower socioeconomic status had lower response rates to texts with more reflection burden, relative to those of their counterparts (all P values were <.05).

Conclusions: As reflection burden increased, engagement decreased, and we found more disparities in engagement across participants' characteristics. Content encouraging moderate levels of reflection may be ideal for achieving both cognitive investment and system use. Our findings provide insights into mHealth design and the optimization of both engagement and effectiveness.

(JMIR Mhealth Uhealth 2024;12:e47632) doi:10.2196/47632

KEYWORDS

engagement; mobile phone; text messaging; messaging; SMS; diabetes; diabetic; mobile health; mHealth; technology; user response; users; quality of life; engagement; mHealth management; management; socioeconomic; effectiveness; support person; support worker; support persons; text message; text messages; reflection; behavior change

Introduction

Background

Mobile health (mHealth) is transforming health delivery as a highly convenient and effective approach for supporting individuals with chronic conditions [1-3]. Delivered via phones, tablets, and wearables, mHealth provides education, motivation, monitoring, and other forms of support to improve health behaviors. SMS text messaging is one form of mHealth that is uniquely poised to benefit everyone, including people who are older, are disadvantaged, and are from traditionally minoritized racial or ethnic backgrounds [4-6]. A critical factor influencing mHealth effectiveness is users' engagement or interaction with the technology, which is typically measured via system use [7-9]. Across the mHealth literature, engagement tends to be highly variable [10,11], which has spurred a whole body of research that aims to understand predictors of engagement, including user characteristics and intervention features (eg, intervention duration and frequency of sending content) [10-14]. However, very little research has attended to the type of mHealth content that users are expected to engage with [15] and, more specifically, how the content may be requesting more or less cognitive reflection.

The primary goal in having users engage with mHealth content is health behavior change. With respect to mHealth interventions, there is a hyperfocus on wanting the user to interact with the technology (eg, responding to a text message), with less consideration of whether the user is reflecting on the content cognitively (eg, reflecting on past behavior and planning future behavior) [8]. Although interaction with the technology is a critical measure, there is a growing consensus that cognitive investment is also important for meaningful behavior change in many types of mHealth interventions [16-18]. Notably, content may be designed in a way that represents a low response burden, thereby easily eliciting a response (ie, producing high engagement), but such content may not evoke the necessary cognitive reflection required to change behavior [18]. Alternatively, content that is designed to encourage deeper reflection may overwhelm users, which risks them disengaging completely. Understanding how to strike the balance between response and reflection has critical implications for effective engagement (ie, engagement needed to impact outcomes) [19]. To our knowledge, no studies have explored the association between reflection demands and the degree of interaction with an mHealth tool. Understanding the interplay between reflection burden and response burden will help guide the design of interventions seeking to strike this balance.

Objective

Our team previously developed an mHealth intervention (delivered via text messages and phone calls) called *Family/Friend Activation to Motivate Self-care* (FAMS) [20,21]. FAMS is a diabetes self-management intervention that targets persons with type 2 diabetes and provides the option for persons with diabetes to invite a support person to also receive text messages. We recently evaluated FAMS in a randomized controlled trial (RCT) [22], and during routine monitoring in the first few weeks of the trial, we observed a low response rate to text messages among support persons. Because support persons' engagement with the text messages was an optional component of the intervention, we determined that this was an opportunity to explore how changing the content of these texts might impact response rates without compromising our ability to evaluate FAMS' effects. Over the course of the RCT [22], we used a pragmatic approach to vary both the reflection burden and the response burden of the two-way text messages sent to support persons assigned to the intervention. In this observational study, our primary goal was to explore how these variations would impact users' engagement with text messages, as measured via response rates. We also explored support persons' characteristics that were associated with response rates for each type of text and described the different responses to each text.

Methods

Study Design and Eligibility

This study was conducted as part of the FAMS 2.0 RCT. The trial design, intervention details, and outcomes for persons with diabetes and support persons were published [22-24]. For the trial, dyads comprising a person with diabetes and their support person were randomized to FAMS or a control condition. We recruited persons with diabetes who were receiving care for type 2 diabetes at Vanderbilt University Medical Center primary care clinics. Enrolling persons with diabetes were asked to invite a support person to participate with them and receive text messages; however, support person invitation and enrollment were not required. We defined a *support person* as any family member or friend with whom the person with diabetes would feel comfortable talking about diabetes management and health goals. Eligible support persons were aged ≥18 years, could speak and read English, and had a mobile phone separate from that of the person with diabetes. The only exclusion criterion was the inability to receive and respond to a text after training. For this study, we analyzed data from support persons in dyads that were randomly assigned to the intervention group (FAMS).

Ethical Considerations

The Vanderbilt University Institutional Review Board approved all study procedures (institutional review board number: 200398; approved April 8, 2020), and the trial was registered on ClinicalTrials.gov (trial number: NCT04347291).

Procedure

While enrolling persons with diabetes into the trial, research assistants collected contact information for a potential support person. A research assistant then contacted potential support persons to verify interest and eligibility, obtain verbal informed consent, and ask support persons to complete a baseline survey. Surveys were completed by phone with a research assistant, on the web via an emailed link, or via a mailed paper copy, per participants' preferences. All survey data were stored in REDCap (Research Electronic Data Capture; Vanderbilt University), a secure web-based platform that supports data capture for research studies [25,26]. In addition to collecting data on sociodemographic characteristics, surveys asked support persons to choose the time of day they wanted to receive text

messages. Relevant survey responses were transferred from REDCap to our technology partner, PerfectServe, using an automated application programming interface. PerfectServe used participant information to tailor, schedule, and send text messages to support persons for 9 months. Support persons could earn a total of US $120 for completing all study surveys (through 15 mo for the larger RCT). There was no compensation for receiving or responding to text messages.

The Intervention

Persons with diabetes received daily text message support and monthly coaching sessions, during which they set behavioral diabetes self-management goals (as detailed by Mayberry et al [22]). Support persons received text messages that were designed to increase dialogue about and facilitate their involvement in the diabetes self-management of the persons with diabetes; a one-way message was sent 3 to 4 times per week, and a two-way message (also known as an *interactive text message*) was sent once per week. One-way messages were either a general text message about providing diabetes self-management support or a text message tailored to the identified diabetes goals of the persons with diabetes. Two-way messages asked support persons about how they supported the health of the persons with diabetes. Support persons who replied to the two-way text received an automated response, thanking them for their answer.

Although an individual support person's intervention experience lasted 9 months (36 wk), intervention delivery for the trial lasted 2 years. Over the course of those 2 years, we varied both the response burden (the level of information requested for the response relative to that of other messages) and the reflection burden (the level of cognitive reflection requested relative to that of other messages) of the weekly two-way text messages, which were sent to support persons in 6 fixed periods (ie, waves). The waves coincided with the weeks of the trial; they did not coincide with the weeks of each individual support person's intervention experience. Figure 1 includes the content for each version of the text message, the weeks of the trial when each text was sent (ie, calendar time), and the respective waves. We started the trial (wave 1) by sending a text message that was high in both reflection burden and response burden (*high/high*). In wave 2, we tested a text that was low in both reflection burden and response burden (*low/low*), and then in wave 3, we tested a text that involved medium reflection burden and low response burden (*medium/low*). In wave 4, we retested the *high/high* message to help determine if the point at which the text was sent during the trial impacted engagement. In wave 5, we sought to delineate the relative impacts of reflection burden and response burden; therefore, we tested a text message that was high in reflection burden and low in response burden (*high/low*). Finally, we closed out the trial by retesting the *low/low* text message (wave 6). The decisions about what messages to test were made iteratively based on response rates to the prior message, with the goal of learning how much reflection we could request while still achieving a relatively high response rate.

Figure 1. The four versions of the two-way text message. The content references an example person with diabetes named Tom.

Reflection burden	Response burden	Content	Weeks of trial when each text was sent	Wave
High	High	"This week is done! Reflect on how you supported Tom this week. Reply with what went well or what could go better next week."	1-24; 73-85	1; 4
Low	Low	"Please text back Yes or No. Were you able to support Tom with his health goals this week?"	25-48; 96-113	2; 6
Medium	Low	"How confident do you feel supporting Tom with his health goals? Please provide a rating from 1 to 5 where 1=not so confident and 5=very confident."	49-72	3
High	Low	"Take a minute to think about how you supported Tom this week with his health goals. Consider what went well and what could go better. Reply OK when done."	86-95	5

Of note, each support person only received the versions of the two-way text message that were sent during their 36-week trial participation, with most (120/148, 81.1%) receiving 2 or 3 different versions and no participants receiving the same message in 2 separate waves. Because this analysis used data from support persons only, we refer to them as *participants* henceforth.

Measures

Sociodemographic and Relationship Characteristics

We collected self-reported data on age, gender, race, ethnicity, socioeconomic status (measured based on education [ie, years in school] and annual household income), and health literacy (assessed via the Brief Health Literacy Screen [27]). In addition, we asked whether participants were cohabitating with the

persons with diabetes and the frequency with which they provided diabetes-specific helpful involvement to the persons with diabetes at baseline, as assessed via the Family and Friend Involvement in Adults' Diabetes (FIAD) helpful subscale, support person version [28].

Engagement

We operationalized engagement by using response rate (ie, two-way messages responded to divided by the two-way messages sent, for each participant).

Analyses

Statistical Analysis Overview

All statistical analyses were performed by using R version 4.2.1 (R Foundation for Statistical Computing). We described participant characteristics via means and SDs or via frequencies and percentages, as appropriate. Except for when examining temporality, message waves that included the same version of the two-way text message were grouped together. Because this study was exploratory, we did not perform sample size calculations.

Overall Engagement by Text Message Version

For each version of the text message, we determined the proportion of two-way text messages sent to support persons that received a response by study week (ie, calendar time). We also generated summary statistics (means, medians, and first and third quartiles) for response rates at the participant level; reporting both mean and median provides more detailed information on the distribution of data. If participants withdrew during their intervention experience, we calculated their response rates based on the data available prior to their withdrawal. To account for repeated measures within participants, we used generalized estimating equations with a working independence correlation structure and a logistic link function to compare the odds of responding to two-way text messages across the four versions.

Participant Characteristics and Engagement

We used Poisson regression to estimate associations (as incidence rate ratios) between participant characteristics and text message response rates for each version of the text message. We included the number of two-way messages sent to a participant as an offset term in order to account for variation in the number of messages sent to each participant in a given wave; therefore, the exponentiated coefficients from the Poisson regression model compared response rates on a per-message basis. Participant characteristics included age, race and ethnicity (non-Hispanic White vs minoritized race or ethnicity), gender, education (years), annual household income (≥US $50,000 per year), health literacy (Brief Health Literacy Screen), whether the persons with diabetes and support persons were cohabitating, and self-reported baseline helpful involvement (FIAD). Further, we multiplied participants' age by 10 to allow for easier interpretation of the results. Especially in regression models, it can be difficult to interpret the association between age and an outcome when the change in the outcome is based on a single-year change in age (ie, the coefficients end up being too small). Scaling the age variable in this way allowed us to interpret the findings in a more meaningful way, that is, we compared groups that differed in age by 1 decade rather than 1 year.

For this analysis, we excluded 5 participants who were missing all baseline data. However, missing covariate values were otherwise addressed via multiple imputation by chained equations (M=500 iterations).

Types of Responses

We characterized the responses to each version of the two-way text message. For the *low/low*, *medium/low*, and *high/low* texts, we reported the frequency of responses based on what the respective text requested (eg, "Yes," "No," "1," "2," "3," "4," or " 5"). For the *high/high* text messages, 2 team members reviewed responses and categorized each as being either high effort or low effort. High-effort responses included comments on what went well that week, comments on what could go better next week, or both, and they referred to a diabetes self-management behavior such as diet, exercise, stress management, or communication (eg, "[He] and I got out several times this week walking after work. Our biggest problem is watching portion size when we are eating. Always continue to work on that."). A low-effort response consisted of only a brief phrase that did not reference a diabetes self-management behavior (eg, "Things went well" and "We were on vacation this week") or did mention a behavior but was unclear as to what went well or what could go better next week (eg, "Walking").

Results

Participant Characteristics

In the trial, of the 150 support person participants who were enrolled and randomized to receive the FAMS intervention, 2 withdrew before the intervention started. The remaining 148 were included in the analyses (Table 1). The mean age was 50.3 (SD 14.7) years; 28.4% (42/148) of participants were men, and 33.1% (49/148) reported a minoritized racial or ethnic background. The mean length of education was 14.9 (SD 2.5) years, and 31.1% (46/148) of participants had annual household incomes of <US $50,000. Over half (84/148, 56.8%) were spouses or partners of the persons with diabetes, and 70.3% (104/148) were cohabitating with the persons with diabetes. Further, 9 participants withdrew at some point during the intervention; the analyses below reflect their engagement during the time they participated.

Table . Participant characteristics (N=148).

Characteristic			Value
Age[a] (y), mean (SD)			50.3 (14.7)
Gender [b] , n (%)			
	Men		42 (28.4)
	Women		101 (68.2)
Race and ethnicity, n (%)			
	Non-Hispanic White		92 (62.2)
	Non-Hispanic Black		29 (19.6)
	Other non-Hispanic races		12 (8.1)
	Hispanic		8 (5.4)
	Missing		7 (4.7)
Socioeconomic status			
	Education[c] (y), mean (SD)		14.9 (2.5)
	Annual household income (US $), n (%)		
		<35,000	25 (16.9)
		35,000-49,999	21 (14.2)
		50,000-74,999	22 (14.9)
		75,000-99,999	24 (16.2)
		≥100,000	40 (27)
		Missing or unknown	16 (10.8)
	Health literacy (BHLS[d,e]), mean (SD)		13.7 (1.5)
Relationship variables			
	Relationship type, n (%)		
		Spouse or partner	84 (56.8)
		Parent	15 (10.1)
		Son or daughter	22 (14.9)
		Grandchild	7 (4.7)
		Friend	11 (7.4)
		Other	4 (2.7)
		Missing	5 (3.3)
	Helpful involvement (FIAD[f,g]), mean (SD)		2.7 (0.9)
	Cohabitating with person with diabetes[h], n (%)		104 (70.3)

[a]10 participants did not report their date of birth (ie, age).

[b]5 participants did not provide data on gender.

[c]8 participants did not report years of education.

[d]BHLS: Brief Health Literacy Screen.

[e]5 participants did not have data for the BHLS measure.

[f]FIAD: Family and Friend Involvement in Adults' Diabetes.

[g]5 participants did not have data for the FIAD helpful subscale.

[h]8 participants did not have data about cohabitating with the persons with diabetes.

Overall Engagement by Text Message Type

Figure 2 presents the proportion of two-way text messages that received a response within each week of the trial (ie, by calendar time). Notably, text message response rates for waves 1 and 4 (both were *high/high* message waves) were comparable, as were those for waves 2 and 6 (both were *low/low* message waves). Table 2 includes descriptive statistics for the overall and

participant-level response rates for each version of the two-way text. The median response rates for the *high/high*, *medium/low*, and *low/low* messages were 23% (IQR 0%-51%), 75% (IQR 38%-96%), and 90% (IQR 61%-100%), respectively. When we kept reflection burden high but lowered response burden in the *high/low* message, the median response rate (10%, IQR 0%-40%) was closest to that for the *high/high* message, suggesting that reflection burden was responsible for the lower response rates seen with the *high/high* message.

Figure 2. Text message response rates by week across each wave. Response data were excluded for the first 4 weeks and the last 5 weeks of the trial when <5 individuals were receiving the intervention.

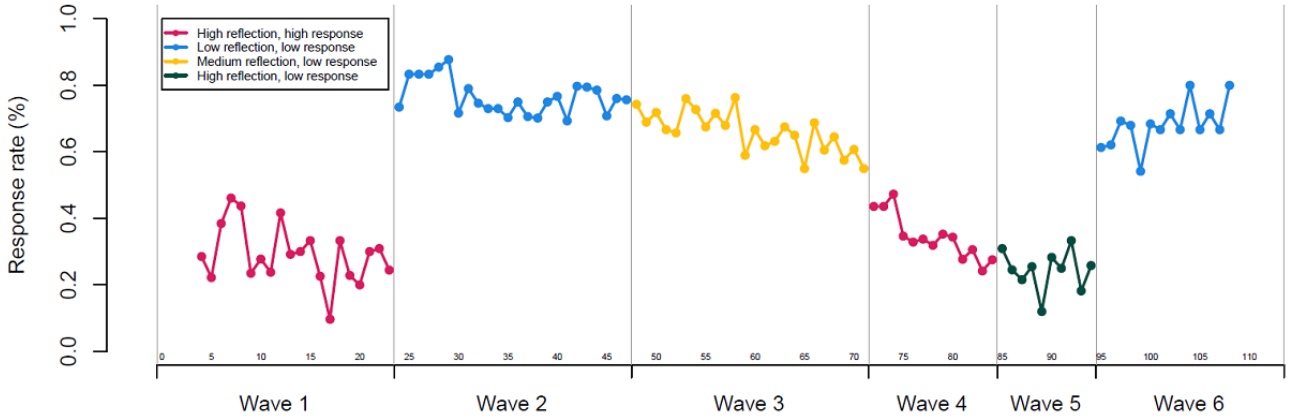


Table . Response rates for each version of the two-way text message.

Reflection burden and response burden	Participants included in analysis, n	Participant-specific response rate	
		%, mean (SD)	%, median (IQR)
High and high	127	33 (33)	23 (0-51)
Low and low	123	74 (32)	90 (61-100)
Medium and low	125	65 (35)	75 (38-96)
High and low	55	26 (32)	10 (0-40)

We also compared the odds of responding to the four versions of the text message (Table 3). When compared to the *high/high* message, the odds of responding to the *low/low* message was 53% (95% CI 42%-65%) higher, the odds of responding to the *medium/low* message was 40% (95% CI 31%-49%) higher, and the odds of responding to the *high/low* message was 7.5% (95% CI 0.7%-14%) lower. All other pairwise comparisons (Table 3) indicated decreasing odds of responding at increasing levels of reflection burden.

Table . Comparison of text message response rates by text message version. Included are odds ratios (ORs) and 95% CIs, along with *P* values.

Comparison		OR (95% CI)	*P* value
Relative to high reflection burden and high response burden			
	Low reflection burden and low response burden	1.53 (1.42-1.65)	<.001
	Medium reflection burden and low response burden	1.40 (1.31-1.49)	<.001
	High reflection burden and low response burden	0.93 (0.86-0.99)	.03
Relative to low reflection burden and low response burden			
	Medium reflection burden and low response burden	0.92 (0.86-0.98)	.01
	High reflection burden and low response burden	0.60 (0.55-0.66)	<.001
Relative to medium reflection burden and low response burden			
	High reflection burden and low response burden	0.66 (0.61-0.72)	<.001

Participant Characteristics Associated With Odds of Responding

Table 4 presents estimated incident rate ratios, along with 95% CIs and *P* values, from multivariate Poisson regression models that were used to identify participant characteristics predictive of response rate. Younger participants, participants who were not cohabitating with the persons with diabetes, and participants who had a lower socioeconomic status had lower response rates to both the *high/high* message and the *medium/low* message compared to those of older participants, participants who were cohabitating, or participants who had a higher socioeconomic status, respectively. The only characteristic associated with response rates for the *low/low* message was gender, such that men had lower response rates. Further, the only characteristic associated with response rates for the *high/low* message was age, such that younger age was associated with lower response rates. Across message versions, younger participants had lower response rates to any message with more burden than the *low/low* message, and participants who were not cohabitating with the persons with diabetes had lower response rates to the higher-burden messages than those of participants who were cohabitating. Race, ethnicity, health literacy, and baseline helpful involvement provided to the person with diabetes did not show patterns indicating the prediction of response rates to any message version.

Table . Participant characteristics predicting text message response rates for each version of the text. Presented are estimated incident rate ratios (IRRs) and 95% CIs, along with *P* values.[a]

Predictor	High reflection/high response		Low reflection/low response		Medium reflection/low response		High reflection/low response	
	IRR (95% CI)	*P* value	IRR (95% CI)	*P* value	IRR (95% CI)	*P* value	IRR (95% CI)	*P* value
Age (y × 10)	1.14 (1.06-1.23)	.004[b]	0.99 (0.95-1.03)	.67	1.08 (1.04-1.14)	.001[b]	1.28 (1.09-1.51)	.002[b]
Race and ethnicity	1.07 (0.85-1.35)	.57	0.96 (0.84-1.09)	.51	0.94 (0.81-1.09)	.40	0.60 (0.36-1.01)	.052
Gender (men)	0.62 (0.49-0.78)	<.001[b]	0.78 (0.68-0.89)	<.001[b]	0.89 (0.78-1.02)	.10	1.28 (0.83-1.98)	.26
Education	0.98 (0.94-1.02)	.34	1.00 (0.98-1.03)	.77	0.96 (0.94-0.99)	.007[b]	0.99 (0.90-1.09)	.87
Income	0.71 (0.57-0.89)	.003[b]	1.05 (0.91-1.21)	.51	1.14 (0.98-1.33)	.10	0.77 (0.46-1.27)	.30
BHLS[c]	1.04 (0.97-1.12)	.30	1.04 (1.00-1.09)	.06	1.02 (0.98-1.06)	.24	1.02 (0.86-1.21)	.80
Cohabitating	1.37 (1.09-1.73)	.007[b]	1.03 (0.90-1.19)	.64	1.21 (1.05-1.40)	.008[b]	1.63 (0.93-2.85)	.09
FIAD[d]	1.05 (0.94-1.17)	.41	1.00 (0.93-1.07)	.99	0.99 (0.93-1.06)	.79	0.90 (0.71-1.14)	.37

[a]A total of 5 support persons without baseline characteristics were excluded from this analysis: 2 were excluded from the models for the *high/high*, *low/low*, and *medium/low* messages; 1 was excluded from the models for the *high/high* and *medium/low* messages; 1 was excluded from the models for the *high/high*, *low/low*, and *high/low* messages; and 1 was excluded from the models for the *low/low* and *medium/low* messages.

[b]$P<.05$.

[c]BHLS: Brief Health Literacy Screen.

[d]FIAD: Family and Friend Involvement in Adults' Diabetes.

Types of Responses

In this section, we report on engagement at the text message level (vs the participant level). For the *high/high* messages, 1429 texts were sent, and 445 responses were received. Further, 13 responses were excluded from the analysis because the content was uninterpretable or was not relevant to the two-way text prompt. The reviewers categorized each response into the high- or low-effort response category, with 98.6% (responses: 426/432) agreement; of the 426 texts agreed upon, 350 (82.2%) were categorized as high-effort responses, and 76 (17.8%) were categorized as low-effort responses. For the *low/low* texts, 1791 texts were sent, and 1341 responses were received; almost all of the responses were "yes" responses (n=1239, 92.4%), while only 97 (7.2%) were "no" responses, and 5 (0.4%) were considered "other" responses. For the *medium/low* texts, 1847 texts were sent, and 1218 responses were received; the most common response was a "5" response (n=681, 55.9%), followed by a "4" response (n=338, 27.8%) and then a "3" response (n=144, 11.8%). Lastly, for the *high/low* texts, 446 texts were sent, and 109 responses were received; most of the responses received were "OK" responses, as requested in the message (n=86, 78.9%).

Discussion

Principal Results

Despite the potential of mHealth to enhance self-management support and quality of life, there are still key gaps in understanding how to optimize mHealth engagement [16,17,19]. Most engagement research reports only on system use without consideration of the cognitive reflection done in the process of

engaging with the content [16,29]. Ideally, we want to encourage reflection that results in meaningful behavior change, but it is unclear how much we can request, with respect to reflection, before users disengage. We varied the reflection and response burdens of two-way text messages to examine how these variations impacted users' engagement, as assessed via response rates. We found, generally, that as the reflection burden of the message increased, participants' engagement decreased. Importantly, when the same version of the text was sent at different points in the trial, participants' engagement was consistent, suggesting that the message itself was key for response rates. The response rates for the *high/low* message were similar to those for the *high/high* message, and this supports reflection burden (vs response burden) being the primary driver of lower engagement. We also found evidence that as the reflection burden of the message increased, there were more disparities in engagement across participant characteristics. This finding helps inform who we may lose with content requiring more from users and thus how to ensure mHealth interventions do not widen existing health disparities.

Research focused on promoting mHealth engagement has proliferated in recent years, with the primary goal of increasing system use [30]. Across such studies, we see evidence that content that is more personalized; uses simpler, nontechnical language; and is empowering results in higher engagement [10,19,31-35]. Less research has compared specific types of content and has rarely tested different types within the same study. An exception is a recent study by Klimis et al [15], wherein they used machine learning to demonstrate that text messages with informative (providing health facts or education) and instructional (providing tips or recommendations) message intents were associated with increased engagement, while notification messages that addressed noneducational matters (eg, welcome and exit messages) were associated with reduced engagement [15]. Our study targeted a two-way message and varied the levels of reflection and response burdens in that message. By adjusting the reflection level specifically, we gained unique insight into how engagement with each message variation may ultimately influence behavior change [17,18]. The main way in which our study differs from others in this area of research is that we looked beyond system use as the sole dimension of mHealth engagement. Our goal was not necessarily to see which message resulted in the highest response rate but rather to determine how much reflection we could request from users and achieve a level of interaction that suggested that they were still invested in the content.

Although our results show generally that engagement decreases with more reflection, the nuances in our findings allow us to provide unique recommendations around mHealth design. For instance, it may be best to alternate through content with different levels of reflection burden. Although users were more likely to respond to content that was lower in reflection burden, nearly all (350/426, 82.2%) of the responses that we received to the high-reflection messages included a high-effort level of reflection. The act of asking people to reflect stimulates internal thoughts that are difficult to measure without a response [36] but may still occur among some persons who do not respond. Alternating content may help promote periodic responding and reflecting throughout an intervention experience. Another option involves using an adaptive intervention to tailor the content based on each person's responsiveness. That is, everyone could start receiving content with a high reflection burden, but if a person's response rate starts to drop, they could then switch to content with a moderate reflection burden. Finally, especially in situations where there is limited flexibility with the mHealth functionality, researchers may consider sending the *medium/low* message to all participants, given that the content encouraged a moderate level of reflection (more than the *low/low* message) yet still yielded a high response rate.

Limitations

Our study has several limitations to acknowledge. For instance, our results are based on an SMS text messaging intervention, which is a specific form of mHealth. It is possible that users would have responded differently if the content was delivered via an app or wearable technology. Importantly, compared to apps and other internet-dependent technologies, SMS text messaging is both lower in cost and more easily accessed, and it tends to have higher rates of engagement [4,37]. In addition, this study recruited persons with diabetes and their support persons from a specific region in Middle Tennessee. We acknowledge that the findings may not be generalizable to other types of individuals who are living in other locations. Relatedly, the content asked about how the support persons supported the health of the persons with diabetes, and engagement may differ when asking about a user's own health; however, the marked differences in engagement across message types support broader applications. Another limitation of our work is that we restricted our assessment of engagement to a behavioral measure (ie, responding to the text) and did not have a way to assess participants' cognitive investment or experience with each version of the text. Based on our analysis of participants' responses to the *high/high* message, it appeared that responders were cognitively engaged, but we were not able to compare cognitive engagement across the other messages. In addition, the sample size for the *high/low* message analysis (n=55) was considerably smaller compared to those for the other message analyses, which was due to testing the *high/low* message during only 1 wave near the end of the trial when fewer participants were enrolled. Relatedly, we did not have the time or a sufficient number of participants toward the end of the trial to test the *medium/low* and *high/low* text messages in a second wave, and we do not know for certain whether engagement with these texts could be impacted by temporality; however, as engagement with the *high/high* and *low/low* texts remained similar across multiple waves, it is unlikely. The ordering of the text messages was variable across participants, and due to the observational nature of this study, we cannot determine the extent that ordering may have impacted results; however, the average response rates and trends across message versions and waves provide general insights on how these variations may impact engagement. Finally, we did not assess the impact of engagement on outcomes, as this fell outside the scope of our study; however, other studies in digital health have examined this association [38,39].

Conclusions

In order for individuals to benefit from mHealth and achieve desired effects on outcomes, engagement with the mHealth tool is needed. Our results help elucidate how truly complex the nature of engagement is. Although our past work and that of others have demonstrated the importance of behaviorally interacting with mHealth interventions [9,31,40,41], this measure represents one piece of a larger puzzle. Engagement may be best conceptualized as including both a behavioral dimension and a cognitive dimension. Balancing these dimensions may be what is ultimately needed to achieve effective engagement for impacting intended outcomes. Our study contributes to a growing body of research that encourages a more nuanced approach to studying engagement that goes beyond measuring system use. We hope that our findings help advance the field of mHealth and inform intervention design, with the goal of optimizing both engagement and effectiveness.

Acknowledgments

This research was funded by the National Institutes of Health (NIH), National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), through grant R01-DK119282. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. MKR is funded by the National Center for Advancing Translational Sciences (NIH; grant KL2TR002245).

Authors' Contributions

LAN, AJS, LSM, RAG Jr, and MKR significantly contributed to the conception of the study and planned analyses. LAN wrote the manuscript. AJS analyzed the quantitative data, and RAG Jr oversaw data analyses. LML and SM analyzed responses to the *high/high* message. LSM was the principal investigator of the Family/Friend Activation to Motivate Self-care (FAMS) study. All authors were involved in data interpretation and manuscript revision and approved the final version submitted for publication.

Conflicts of Interest

None declared.

References

1. Fagherazzi G, Ravaud P. Digital diabetes: perspectives for diabetes prevention, management and research. Diabetes Metab 2019 Sep;45(4):322-329. [doi: 10.1016/j.diabet.2018.08.012] [Medline: 30243616]
2. Fan K, Zhao Y. Mobile health technology: a novel tool in chronic disease management. Intelligent Medicine 2022 Feb;2(1):41-47. [doi: 10.1016/j.imed.2021.06.003]
3. Li R, Liang N, Bu F, Hesketh T. The effectiveness of self-management of hypertension in adults using mobile health: systematic review and meta-analysis. JMIR Mhealth Uhealth 2020 Mar 27;8(3):e17776. [doi: 10.2196/17776] [Medline: 32217503]
4. Willcox JC, Dobson R, Whittaker R. Old-fashioned technology in the era of “Bling”: is there a future for text messaging in health care? J Med Internet Res 2019 Dec 20;21(12):e16630. [doi: 10.2196/16630] [Medline: 31859678]
5. Anderson-Lewis C, Darville G, Mercado RE, Howell S, Di Maggio S. mHealth technology use and implications in historically underserved and minority populations in the United States: systematic literature review. JMIR Mhealth Uhealth 2018 Jun 18;6(6):e128. [doi: 10.2196/mhealth.8383] [Medline: 29914860]
6. Nelson LA, Greevy RA, Spieker A, et al. Effects of a tailored text messaging intervention among diverse adults with type 2 diabetes: evidence from the 15-month REACH randomized controlled trial. Diabetes Care 2021 Jan;44(1):26-34. [doi: 10.2337/dc20-0961] [Medline: 33154039]
7. Murray E, Hekler EB, Andersson G, et al. Evaluating digital health interventions: key questions and approaches. Am J Prev Med 2016 Nov;51(5):843-851. [doi: 10.1016/j.amepre.2016.06.008] [Medline: 27745684]
8. Cole-Lewis H, Ezeanochie N, Turgiss J. Understanding health behavior technology engagement: pathway to measuring digital behavior change interventions. JMIR Form Res 2019 Oct 10;3(4):e14052. [doi: 10.2196/14052] [Medline: 31603427]
9. Nelson LA, Spieker AJ, Mayberry LS, McNaughton C, Greevy RA. Estimating the impact of engagement with digital health interventions on patient outcomes in randomized trials. J Am Med Inform Assoc 2021 Dec 28;29(1):128-136. [doi: 10.1093/jamia/ocab254] [Medline: 34963143]
10. Nelson LA, Coston TD, Cherrington AL, Osborn CY. Patterns of user engagement with mobile- and web-delivered self-care interventions for adults with T2DM: a review of the literature. Curr Diab Rep 2016 Jul;16(7):66. [doi: 10.1007/s11892-016-0755-1] [Medline: 27255269]
11. Killikelly C, He Z, Reeder C, Wykes T. Improving adherence to web-based and mobile technologies for people with psychosis: systematic review of new potential predictors of adherence. JMIR Mhealth Uhealth 2017 Jul 20;5(7):e94. [doi: 10.2196/mhealth.7088] [Medline: 28729235]
12. Nouri SS, Adler-Milstein J, Thao C, et al. Patient characteristics associated with objective measures of digital health tool use in the United States: a literature review. J Am Med Inform Assoc 2020 May 1;27(5):834-841. [doi: 10.1093/jamia/ocaa024] [Medline: 32364238]

13. Buck B, Chander A, Ben-Zeev D. Clinical and demographic predictors of engagement in mobile health vs. clinic-based interventions for serious mental illness. J Behav Cogn Ther 2020 Apr;30(1):3-11. [doi: 10.1016/j.jbct.2020.03.004]
14. Nelson LA, Spieker A, Greevy R, LeStourgeon LM, Wallston KA, Mayberry LS. User engagement among diverse adults in a 12-month text message-delivered diabetes support intervention: results from a randomized controlled trial. JMIR Mhealth Uhealth 2020 Jul 21;8(7):e17534. [doi: 10.2196/17534] [Medline: 32706738]
15. Klimis H, Nothman J, Lu D, et al. Text message analysis using machine learning to assess predictors of engagement with mobile health chronic disease prevention programs: content analysis. JMIR Mhealth Uhealth 2021 Nov 10;9(11):e27779. [doi: 10.2196/27779] [Medline: 34757324]
16. Torous J, Michalak EE, O'Brien HL. Digital health and engagement-looking behind the measures and methods. JAMA Netw Open 2020 Jul 1;3(7):e2010918. [doi: 10.1001/jamanetworkopen.2020.10918] [Medline: 32678446]
17. Yeager CM, Benight CC. If we build it, will they come? issues of engagement with digital health interventions for trauma recovery. Mhealth 2018 Sep 11;4:37. [doi: 10.21037/mhealth.2018.08.04] [Medline: 30363749]
18. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267. [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
19. Yardley L, Spring BJ, Riper H, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
20. Mayberry LS, Berg CA, Greevy RA, et al. Mixed-methods randomized evaluation of FAMS: a mobile phone-delivered intervention to improve family/friend involvement in adults' type 2 diabetes self-care. Ann Behav Med 2021 Mar 16;55(2):165-178. [doi: 10.1093/abm/kaaa041] [Medline: 32706852]
21. Mayberry LS, Berg CA, Harper KJ, Osborn CY. The design, usability, and feasibility of a family-focused diabetes self-care support mHealth intervention for diverse, low-income adults with type 2 diabetes. J Diabetes Res 2016 Nov 7;2016:7586385. [doi: 10.1155/2016/7586385] [Medline: 27891524]
22. Mayberry LS, El-Rifai M, Nelson LA, et al. Rationale, design, and recruitment outcomes for the Family/Friend Activation to Motivate Self-care (FAMS) 2.0 randomized controlled trial among adults with type 2 diabetes and their support persons. Contemp Clin Trials 2022 Nov;122:106956. [doi: 10.1016/j.cct.2022.106956] [Medline: 36208719]
23. Roddy MK, Spieker AJ, Nelson LA, et al. Well-being outcomes of a family-focused intervention for persons with type 2 diabetes and support persons: main, mediated, and subgroup effects from the FAMS 2.0 RCT. Diabetes Res Clin Pract 2023 Oct;204:110921. [doi: 10.1016/j.diabres.2023.110921] [Medline: 37742801]
24. Nelson LA, Spieker AJ, Greevy RAJ, et al. Glycemic outcomes of a family-focused intervention for adults with type 2 diabetes: main, mediated, and subgroup effects from the FAMS 2.0 RCT. Diabetes Res Clin Pract 2023 Dec;206:110991. [doi: 10.1016/j.diabres.2023.110991] [Medline: 37925077]
25. Harris PA, Taylor R, Minor BL, et al. The REDCap Consortium: building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208. [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
26. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research Informatics support. J Biomed Inform 2009 Apr;42(2):377-381. [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
27. Wallston KA, Cawthon C, McNaughton CD, Rothman RL, Osborn CY, Kripalani S. Psychometric properties of the Brief Health Literacy Screen in clinical practice. J Gen Intern Med 2014 Jan;29(1):119-126. [doi: 10.1007/s11606-013-2568-0] [Medline: 23918160]
28. Mayberry LS, Berg CA, Greevy RAJ, Wallston KA. Assessing helpful and harmful family and friend involvement in adults' type 2 diabetes self-management. Patient Educ Couns 2019 Jul;102(7):1380-1388. [doi: 10.1016/j.pec.2019.02.027] [Medline: 30922622]
29. Short CE, DeSmet A, Woods C, et al. Measuring engagement in eHealth and mHealth behavior change interventions: viewpoint of methodologies. J Med Internet Res 2018 Nov 16;20(11):e292. [doi: 10.2196/jmir.9397] [Medline: 30446482]
30. Druce KL, Dixon WG, McBeth J. Maximizing engagement in mobile health studies: lessons learned and future directions. Rheum Dis Clin North Am 2019 May;45(2):159-172. [doi: 10.1016/j.rdc.2019.01.004] [Medline: 30952390]
31. Cao W, Milks MW, Liu X, et al. mHealth interventions for self-management of hypertension: framework and systematic review on engagement, interactivity, and tailoring. JMIR Mhealth Uhealth 2022 Mar 2;10(3):e29415. [doi: 10.2196/29415] [Medline: 35234655]
32. Alkhaldi G, Modrow K, Hamilton F, Pal K, Ross J, Murray E. Promoting engagement with a digital health intervention (HeLP-Diabetes) using email and text message prompts: mixed-methods study. Interact J Med Res 2017 Aug 22;6(2):e14. [doi: 10.2196/ijmr.6952] [Medline: 28829328]
33. Nahum-Shani I, Rabbi M, Yap J, et al. Translating strategies for promoting engagement in mobile health: a proof-of-concept microrandomized trial. Health Psychol 2021 Dec;40(12):974-987. [doi: 10.1037/hea0001101] [Medline: 34735165]
34. Wei Y, Zheng P, Deng H, Wang X, Li X, Fu H. Design features for improving mobile health intervention user engagement: systematic review and thematic analysis. J Med Internet Res 2020 Dec 9;22(12):e21687. [doi: 10.2196/21687] [Medline: 33295292]

35. Shan R, Sarkar S, Martin SS. Digital health technology and mobile devices for the management of diabetes mellitus: state of the art. Diabetologia 2019 Jun;62(6):877-887. [doi: 10.1007/s00125-019-4864-7] [Medline: 30963188]
36. Irvine L, Melson AJ, Williams B, et al. Real time monitoring of engagement with a text message intervention to reduce binge drinking among men living in socially disadvantaged areas of Scotland. Int J Behav Med 2017 Oct;24(5):713-721. [doi: 10.1007/s12529-017-9666-z] [Medline: 28702758]
37. Mayberry LS, Jaser SS. Should there be an app for that? the case for text messaging in mHealth interventions. J Intern Med 2018 Feb;283(2):212-213. [doi: 10.1111/joim.12687] [Medline: 28881402]
38. Delaney T, Mclaughlin M, Hall A, et al. Associations between digital health intervention engagement and dietary intake: a systematic review. Nutrients 2021 Sep 20;13(9):3281. [doi: 10.3390/nu13093281] [Medline: 34579158]
39. Mclaughlin M, Delaney T, Hall A, et al. Associations between digital health intervention engagement, physical activity, and sedentary behavior: systematic review and meta-analysis. J Med Internet Res 2021 Feb 19;23(2):e23180. [doi: 10.2196/23180] [Medline: 33605897]
40. Li Y, Guo Y, Hong YA, et al. Dose-response effects of patient engagement on health outcomes in an mHealth intervention: secondary analysis of a randomized controlled trial. JMIR Mhealth Uhealth 2022 Jan 4;10(1):e25586. [doi: 10.2196/25586] [Medline: 34982724]
41. Spieker AJ, Nelson LA, Rothman RL, et al. Feasibility and short-term effects of a multi-component emergency department blood pressure intervention: a pilot randomized trial. J Am Heart Assoc 2022 Mar;11(5):e024339. [doi: 10.1161/JAHA.121.024339] [Medline: 35195015]

Abbreviations

FAMS: Family/Friend Activation to Motivate Self-care
FIAD: Family and Friend Involvement in Adults' Diabetes
mHealth: mobile health
RCT: randomized controlled trial
REDCap: Research Electronic Data Capture

Edited by L Buis; submitted 31.03.23; peer-reviewed by T Delaney, X Liu; revised version received 29.11.23; accepted 30.11.23; published 19.01.24.

Please cite as:
Nelson LA, Spieker AJ, LeStourgeon LM, Greevy Jr RA, Molli S, Roddy MK, Mayberry LS
The Goldilocks Dilemma on Balancing User Response and Reflection in mHealth Interventions: Observational Study
JMIR Mhealth Uhealth 2024;12:e47632
URL: https://mhealth.jmir.org/2024/1/e47632
doi:10.2196/47632

Original Paper

A Smartphone Food Record App Developed for the Dutch National Food Consumption Survey: Relative Validity Study

Marga Ocké[1*], PhD; Ceciel Simone Dinnissen[1*], MSc; Coline van den Bogaard[1], BSc; Marja Beukers[1], BSc; José Drijvers[1], BSc; Eline Sanderman-Nawijn[1], MSc; Caroline van Rossum[1], PhD; Ido Toxopeus[1], PhD

National Institute for Public Health and the Environment, Bilthoven, Netherlands

[*]these authors contributed equally

Corresponding Author:
Ceciel Simone Dinnissen, MSc
National Institute for Public Health and the Environment
Antonie van Leeuwenhoeklaan 9
Bilthoven, 3721 MA
Netherlands
Phone: 31 088 689 8989
Email: ceciel.dinnissen@rivm.nl

Related Article:

This is a corrected version. See correction statement: https://mhealth.jmir.org/2024/1/e59530

Abstract

Background: In the Dutch National Food Consumption Survey, dietary intake has been assessed since 2003 through 24-hour dietary recalls using the GloboDiet software. A new self-administered smartphone food record app called DitEetIk! was developed for potential use in future surveys.

Objective: This study aims to evaluate the data collected using the DitEetIk! app and its relative validity for food group, energy, and nutrient intake compared with the previous dietary assessment method (GloboDiet 24-hour dietary recalls).

Methods: A total of 300 participants aged 18 to 79 years were recruited from a consumer panel. Participants were asked to keep a record of their consumption using the DitEetIk! app on 3 nonconsecutive days. Trained dietitians conducted a 24-hour dietary recall interview by telephone using the GloboDiet software (International Agency for Research on Cancer) regarding 1 of 3 DitEetIk! recording days. Nutrient intake was calculated using the NEVO database (version 2021/7.0). Relative validity was studied by comparing data from GloboDiet 24-hour dietary recalls and the DitEetIk app for the same day. Participants with implausible records, defined as days with energy intake of <0.6 or >3.0 basal metabolic rate, were excluded from the analyses. For 19 food groups and 29 nutrients, differences in median intake were assessed using the Wilcoxon signed rank test, and Spearman correlation coefficients were calculated. Bland-Altman plots with mean differences and 95% limits of agreement were created for energy intake and the contribution to energy intake from fat, carbohydrates, and protein.

Results: A total of 227 participants completed a combination of a DitEetIk! app recording day and a 24-hour dietary recall interview for the same day. Of this group, 211 participants (n=104, 49.3% men and n=107, 50.7% women) had plausible recording days. Of all recorded food items, 12.8% (114/894) were entered via food barcode scanning, and 18.9% (169/894) were searched at the brand level. For 31% (5/16) of the food groups, the median intake assessed using the DitEetIk! app was >10% lower than that assessed using 24-hour dietary recalls; this was the case for fruit (P=.005), added fats (P=.001), milk and milk products (P=.02), cereal products (P=.01), and sauces (P<.001). This was also the case for 14% (4/29) of the nutrients (all P<.001). Regarding mean intake, differences were generally smaller. Regarding energy intake, the mean difference and 95% limits of agreement were 14 kcal (−1096 to 1124). Spearman correlation coefficients between intake assessed using the DitEetIk! app and 24-hour dietary recalls ranged from 0.48 to 0.88 (median 0.78) for food groups and from 0.58 to 0.90 (median 0.72) for nutrients.

Conclusions: Compared with GloboDiet 24-hour dietary recalls, the DitEetIk! app assessed similar mean energy intake levels but somewhat lower median intake levels for several food groups and nutrients.

(JMIR Mhealth Uhealth 2024;12:e50196) doi:10.2196/50196

KEYWORDS

relative validity; smartphone food record; 24-hour dietary recall; mobile app; national food consumption surveys; smartphone; food; food consumption; app; diet; dietary intake; nutrients; survey; mobile phone

Introduction

Background

Many countries conduct national food consumption surveys as these are considered important instruments for prioritizing, developing, and evaluating food policies [1,2]. Food consumption survey data can be used to assess adherence to food-based dietary guidelines and obtain insight into the food consumption patterns of a population. After combination with other data sources such as food composition databases, food contamination occurrence data, and life cycle assessment data, the nutritional adequacy of the diet, dietary exposure assessment to potentially hazardous substances, and the environmental impact of dietary patterns can be assessed [3-5].

The use of national food consumption survey data for multiple purposes requires dietary assessment methods that allow all consumed food items to be reported with detailed characterizations. On the basis of European projects such as the European Food Consumption Survey Method [6] and European Food Consumption Validation [7], the 2014 guidance on the European Union Menu methodology by the European Food Safety Authority prescribes the use of food records for children and 24-hour dietary recalls for adults as dietary assessment methods in European national food consumption surveys [8]. It is advised that trained personnel is employed to administer the recall interviews or conduct a food record completion interview with the participants [8,9]. The requirements of trained personnel and detailed food descriptions, the large number of and continuously changing food items on the market, and the need to handle all possible reported food items make food consumption surveys costly [10]. Moreover, these requirements also pose a burden on the survey participants. It has been suggested that this burden has led to declining and possibly selective participation rates in national food consumption surveys [1].

In the past decades, various tools for self-administered 24-hour dietary recalls or food records have become available [11]. These digital tools have the potential to be less resource intensive. Many users prefer these applications over the traditional methods as they can be used where and when it is convenient [12]. Information and communications technology–based applications also enable the use of user-friendly support functionalities that were not feasible with interviewer-based methods. For example, food package barcode scanning using the camera function of a smartphone [11] combined with a comprehensive branded food database reduces the time burden of searching for a product through a long list of food items. However, without the help of a trained interviewer, it might be challenging for participants to report all food items consumed and describe and quantify them accurately. These developments warrant further exploration of whether interviewer-based dietary assessments in national food consumption surveys can be replaced with self-administered dietary assessments using digital food record applications.

In the Dutch National Food Consumption Survey (DNFCS), dietary intake has been assessed through 24-hour dietary recalls by trained interviewers using the GloboDiet software since 2003 [13-16]. A new food record app for dietary assessment called DitEetIk!, which uses self-administration, was developed for potential use in future national dietary surveys in the Netherlands. A smartphone food record was chosen over a self-administered digital 24-hour dietary recall because of the availability of a branded food database in the Netherlands [17]. Such a database allows for specific food identification and can be used most optimally when keeping a food diary throughout the day on a mobile phone with a camera function for barcode scanning. Moreover, the level of smartphone ownership and use in the Netherlands is high. In 2019, smartphones were present in 89% of Dutch households [18].

Objectives

To assess the suitability of the DitEetIk! app for future surveys, it is important to evaluate its quality and comparative validity against the method currently used in the DNFCS. Therefore, the aim of this study was to evaluate the level of detail regarding the food description obtained in the reported consumption in the DitEetIk! app and determine how well the DitEetIk! app is able to assess the daily intake of food groups, energy, and nutrients in comparison with dietitian-administered 24-hour dietary recalls using the GloboDiet software (International Agency for Research on Cancer) in adults. This study focused on systematic differences at the food group and nutrient levels, with random error being of secondary interest. To study the potential effects of the study design on the 24-hour dietary recall data, we also compared the GloboDiet 24-hour dietary recall data in this study with those of a matched population of the DNFCS 2019 to 2021 [19].

Methods

Recruitment

The intention was to collect data from 200 participants with sufficient variation in gender, age group, and educational level. To account for potential dropouts and invalid food recording data, 300 participants were recruited. Participants from a consumer panel of Kantar Netherlands were invited via email to take part in the study. Information regarding the privacy policy of the DitEetIk! app was provided. The sociodemographic characteristics of the panel members were known. Potential participants were eligible if they were aged between 18 and 79 years; were not institutionalized; did not participate in the DNFCS 2019 to 2021 [19] or the Eetmeter study [20]; did not use tube or parental feeding; and were able to use the DitEetIk! app on their smartphone, which had to run on the Android operating system version 7 or higher.

Ethical Considerations

The Medical Research Ethics Committee of Utrecht University evaluated that the study was not subject to the Medical Research Involving Human Subjects Act of the Netherlands (dossier 21/686). All study participants provided written informed consent. After completion of the study, participants received an incentive bonus (NIPOints to be exchanged for a gift card or coupon).

DitEetIk! App

Objective of the DitEetIk! App

In the Dutch language, "Dit eet ik" means "This is what I eat." The DitEetIk! app was developed specifically for the objectives of the DNFCS. The food description had to be specific enough to provide insights into the intake of nutrients, the exposure to chemicals relevant from a food safety point of view, and the environmental impact of the diet of the Dutch general population aged 1 to 79 years.

Development Process

The DitEetIk! app was developed using an agile approach. Various usability tests and focus group sessions were part of this process. The development period lasted approximately 3 years, with a team of app builders, data managers, dietitians, and nutritionists. Technical development was conducted by Dienst ICT Uitvoering in collaboration with National Institute for Public Health and the Environment for functional development and formative evaluation. Both are Dutch governmental organizations. Safety and General Data Protection Regulation issues were considered in the app development. The DitEetIk! app does not collect information that makes the participant identifiable. The user interface of the app is in Dutch (B1 level).

App Availability, Registration, and Instruction

Version 1.0 of the DitEetIk! app was developed for Android smartphones and is available on the Google Play store. Using the DitEetIk! app, a person can keep a food record for specific days upon invitation (ie, the DitEetIk! app can be used only after entering a participant number with matching gender and age). An instruction movie can be viewed at any time after registration. Moreover, to support participants, context-specific information and relevant sections of the instruction movie are available on each screen. Participants were instructed to record all food and drinks consumed from getting up one day to getting up the next day.

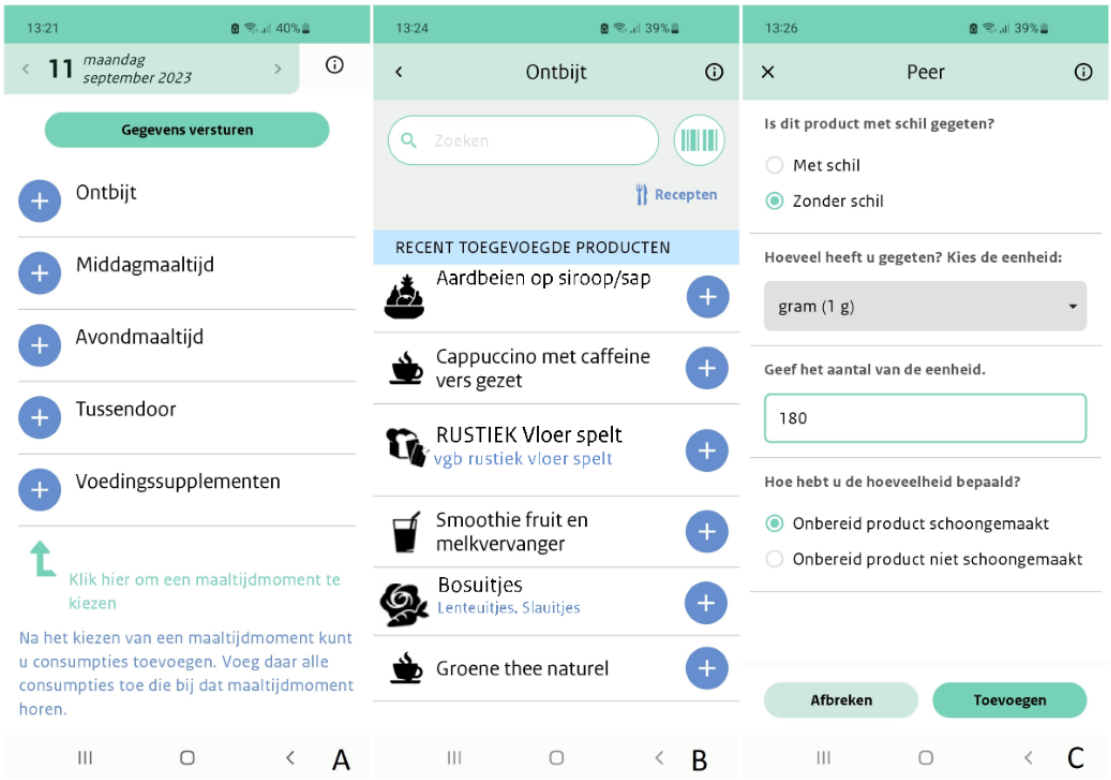
Push Notifications and Feedback

At several moments—before, during, and after the specified day—push notification messages are sent via the DitEetIk! app to the participants to help remind them of food recording and submission of the food record. The DitEetIk! app does not provide instant feedback to the participants regarding their food consumption as, for dietary monitoring, it is important not to influence the participants.

Main Menu

Food recording is performed via a main menu where 4 eating occasions can be selected (ie, breakfast, midday meal, evening meal, and in between meals). If applicable, dietary supplements can be filled out separately (Figure 1A). After choosing breakfast, midday meal, or evening meal as eating occasions, the time and place of consumption need to be registered; for the occasion "in between meals," time and place of consumption (eg, home or restaurant) are asked every time a food item is selected, whereas for dietary supplements, information regarding time and place of consumption is not asked for.

Figure 1. Screenshots of the DitEetIk! app depicting (A) eating occasions, (B) recently added products, and (C) food-specific follow-up questions.



Food Recording

The DitEetIk! app food list in this study included 140,781 food items, of which 3432 (2.44%) were generic and the rest were a selection of branded food items from the Dutch-branded food database Dutch-branded food database (Levensmiddelendatabank [LEDA]; download date: October 27, 2021) [17]. The selection consisted of 70.89% (137,349/193,742) of the branded food items in the LEDA database that were matched to a generic product. Generic food items can be found through text searching. Branded food items can be found by scanning the barcode or, for predefined food groups, by selecting the brand and specific branded product once a generic food item is chosen. If the scanned food items are not included in the LEDA database or are not matched to generic products, they are not recognized in the app, and the participant has to record their food via text search. Once a product is selected and added to an eating occasion, the food item can also be found via "recently added products" and can be selected again (Figure 1B). A recipe feature is available it allows participants to create mixed dishes indicating quantities of foods that were used for the prepared dish as a whole or save frequently consumed combinations of foods. Commonly used mixed dishes are also available as prepared generic food items.

Follow-Up Questions for Each Food Item

After a food item is chosen or scanned, food-specific follow-up questions are asked regarding the preparation method, consumption with or without skin or peel, and quantity consumed (Figure 1C). In case a preparation method with fat (eg, frying or deep-frying) is chosen, the participant is asked to specify the type of fat used. The follow-up questions are defined at the generic food item level. For this reason, each branded food item is linked to a comparable generic food item. Consumed amounts can be indicated via various options that always include the weight in grams or volume in milliliters and often the number or fraction of household measures, natural units, or commercial units. For user-defined recipes, the fraction of the total recipe can be indicated as the consumed portion.

Data Submission and Output

When all consumptions for a day have been recorded, the participant can submit the food record. Upon submission, questions regarding completeness are asked, as well as whether the day was special or not regarding consumption. The data collected using the DitEetIk! app can be downloaded as CSV files that include information on eating occasions with time, place, and registered food items and the answers to the follow-up questions for each consumption.

Study Design and Data Collection

Data were collected in the spring of 2022. Participants were asked to record all food, drinks, and dietary supplements consumed on 3 nonconsecutive days on the DitEetIk! app. The days were assigned by the DitEetIk! app in such a way that all days of the week were covered proportionally at the group level and that there were at least 7 days between each recording day. The day before each registration day, the participants received a push notification message on their smartphones. At the individual level, any combination of days could occur. In case the participant indicated that the day was not convenient for recording or did not record any food items, a replacement recording day was assigned automatically. This could be done maximally 3 times; thereafter, no new days were assigned, and the participant was excluded from further participation.

After 1 of the 3 DitEetIk! app recording days, participants were contacted by a trained dietitian for a 24-hour dietary recall interview by telephone. For the 24-hour dietary recall interviews, it was allowed to make an appointment with the participant; for practical reasons, they were not unannounced. The 24-hour dietary recalls were administered using the Dutch version of the GloboDiet software (version 2021-09-24). This software, which was previously called EPIC-Soft, has been described in detail elsewhere [16]. Briefly, the interview started by composing a quick list in which the participant was asked to roughly list all consumed food and drinks for 7 potential eating occasions with time and place of consumption. Food items were recalled starting from getting up in the morning until getting up the following day. In the second step, the interviewer specified each food item on the quick list using a series of follow-up questions applicable to that food item, for example, asking about preparation methods and, if relevant, the type of fat used. Mixed dishes could be entered as new individual recipes or as (adjustments of) standard recipes, which the software disaggregated into ingredients. Dietary supplements were explicitly asked about. Consumed amounts of the food items could be quantified in several ways: by means of quantities as shown in photos in a picture booklet with a series of 61 food photographs. or in household measures, units, and standard portions; by weight or volume; and by the proportion of a total recipe. Bread shapes were used to estimate the quantity of spreads. At various points, quality control of the data was incorporated into the GloboDiet software, for example, checks on missing quantities, probing questions on often forgotten food items, and checks on total intake of energy and macronutrients.

After the 3 food records were completed, the perceived usability of the DitEetIk! app was evaluated by the participants through a web-based questionnaire, the System Usability Scale (SUS) [21]. This is a widely used questionnaire for the evaluation of electronic devices and systems, including smartphone apps. It consists of 10 statements with response options on a 5-point Likert scale ranging from *strongly disagree* (1) to *strongly agree* (5). There are 5 positive statements alternated with 5 negative statements. The originally English-worded items were translated into Dutch, and the word “system” was replaced with “application” to make the questions more specific to the device.

Data Handling

For the DitEetIk! app data, intake per food item per day was calculated by multiplying the weight of the chosen portion or serving by the number of portions. If applicable, food density (in the case of estimates in household measures or milliliters), an edible fraction (in the case of the inedible part), weight change because of food preparation (amount estimated as unprepared food), and percentage of fat absorption were applied to the calculation of the amount of food in grams per day in its consumed state. Similar calculations were performed using the GloboDiet software for the 24-hour dietary recalls [16]. All food items reported in the GloboDiet 24-hour dietary recalls and the DitEetIk! app were categorized into the food groups mentioned in the Wheel of Five Dutch food-based dietary guidelines [22,23]. Subsequently, consumption of food groups per person per day was calculated.

For both GloboDiet and DitEetIk! data, intake of energy and nutrients per person per day was calculated by multiplying the consumed amount of food by the nutrient level per gram of food and adding the nutrient intake of all food items consumed in a day. Information on food composition was obtained from the NEVO database (version 2021/7.0) [24]. Dietary supplements were not considered in the calculation of nutrient intake.

All extremely high values in energy, nutrient, and food group intake in the 24-hour dietary recalls were evaluated using the same methodology as in the DNFCS [15]. The food items that contributed the most to the high intake values were checked for obvious errors. None of the extreme intake values were judged as unlikely.

The equations developed by Henry [25] were applied to calculate the estimated basal metabolic rate (BMR) using height and weight information provided by the participants in the DitEetIk! app. The average ratio of energy intake to BMR was calculated, as well as the percentage of extreme energy reporters (ie, those participants with a ratio of <0.6 or >3.0). Any day with such an implausible extreme of energy intake in the DitEetIk! app was excluded from further analyses.

Assessment of the Effect of Study Design on the GloboDiet 24-Hour Dietary Recall Results

The GloboDiet 24-hour dietary recall interview was always conducted after the recording in the DitEetIk! app. This might have influenced the results because of potential memory or learning effects [26]. The extent to which this occurred was estimated by comparing the results of the GloboDiet 24-hour dietary recalls in this study with the findings among participants in the DNFCS 2019 to 2021, in which 24-hour dietary recalls were collected using the same software. We only used the first GloboDiet 24-hour dietary recall interviews of these participants. Each participant in the DitEetIk! app evaluation study was matched with a participant in the DNFCS 2019 to 2021 based on characteristics associated with dietary intake, that is, age (5-year classes), gender, weight (10-kg classes), and educational level (3 classes). This provided matches for 86.3% (182/211) of the participants. For those without a match, the matching characteristics were relaxed, starting with educational level (28 matches) followed by age (1 match). For participants in the evaluation study with multiple possible matches in the DNFCS 2019 to 2021, a person with the same or the closest height was chosen (182 matches).

Statistical Analysis

For the analyses, only DitEetIk! data for the day that was recalled using the GloboDiet software were used. Frequency analyses were conducted to describe the population of DitEetIk! app users and the matched participants of the DNFCS 2019 to 2021 in terms of sociodemographic characteristics. For the items of the SUS and the total SUS score, means and SDs were calculated.

The medians, SDs, and IQRs of the food group, energy, and nutrient consumption assessed using both methods were calculated. Owing to skewed distributions, the nonparametric Wilcoxon signed rank test was used to test whether differences between the DitEetIk! app and the 24-hour dietary recalls for

food group and nutrient intake were symmetrical around 0. Differences were considered relevant if the median intake was >10% of the 24-hour dietary recall value. Only food groups for which the 75th percentile was >0 were reported. The analyses were repeatedly stratified by educational level (3 classes) and BMI (3 classes).

In addition, the number of consumers and median intake for consumers only were calculated for each food group for each method. The McNemar test was used to test whether being a consumer per food group differed significantly between both methods. In addition, the Wilcoxon signed rank test was used to test whether the distribution of food group consumption only differed systematically by method. Spearman rank correlation coefficients were calculated for food group consumption and also energy and nutrient intake assessed using both methods.

Bland-Altman plots were constructed for the intake of energy and energy percentage derived from fat, protein, and carbohydrates, plotting the difference in intake assessed using both methods against the mean intake for each participant. The derived 95% limits of agreement [27] were presented to provide information on the variation in individual relative validity.

To assess any design effects, a nonparametric Wilcoxon signed rank test was used to test whether there was a systematic difference between the 24-hour dietary recalls in the relative validity study and the DNFCS 2019 to 2021 regarding the intake of energy, nutrients, and food groups.

All statistical analyses were conducted using SAS (version 9.4; SAS Institute Inc). *P* values of <.05 were considered statistically significant, and 2-sided statistical tests were performed.

Results

Study Population Characteristics

Of the 3418 contacted people, 443 (12.96%) were willing to participate (Figure 2). A total of 300 people were invited to start the study based on their ability to use an Android smartphone and their sociodemographic characteristics. Of the 300 invited individuals, 227 (75.7%) completed a combination of a DitEetIk! app recording day and a 24-hour dietary recall interview for the same day. In total, 7% (16/227) of the participants were excluded because of an implausible ratio of energy intake to BMR of <0.6 or >3.0, resulting in 211 participants included in the analyses.

Approximately half (107/211, 50.7%) of the participants were women (Table 1). The study population consisted of more participants with a middle and higher educational level (81/211, 38.4% and 93/211, 44.1%, respectively) than those with a lower educational level (37/211, 17.5%). Of the 211 participants, 129 (61.1%) had a BMI of ≥25 kg/m^2. Fewer people in the highest age category of 60 to 79 years participated in the study (58/211, 27.5%) than people in the 2 lower age categories (76/211, 36% and 77/211, 36.5%). The 211 participants included in the DitEetIk! app study were matched with 211 participants of the DNFCS 2019 to 2021. They had similar distributions in terms of gender, age, educational level, and BMI.

Of the 211 participants, 207 (98.1%) completed 3 DitEetIk! app recording days. Most GloboDiet 24-hour dietary recall interviews were for the first or second DitEetIk! app recording day (78/211, 37% and 84/211, 39.8%, respectively). Most of the participants started recording their food consumption on the DitEetIk! app in the morning (105/211, 49.8%) or afternoon (61/211, 28.9%) and ended their recording the same day in the evening (95/211, 45%) or the next day in the morning (82/211, 38.9%). The time between starting and ending the food recording using the DitEetIk! app was between 8 and 20 hours for 46% (97/211) of the participants, whereas it took >20 hours for 33.2% (70/211) of the participants and <8 hours for 20.9% (44/211) of the participants (Multimedia Appendix 1).

Food identification for recording was performed via text searching of the food items at the generic level for 23.6% (211/894) of the products and choosing previously selected food items (175/894, 19.6%). Branded food items were recorded via text searching (169/894, 18.9%) and barcode scanning (114/894, 12.8%). The option to make an individual recipe was used for 1.8% (16/894) of the products, associated food items (fats and oils for frying and milk or sugar in tea or coffee) were used for 16.4% (147/894) of the products, and supplements were used for 6.9% (62/894) of the products. Without considering previously selected food items, branded food items were recorded 39.4% (283/719) of times (Multimedia Appendix 2).

The SUS questionnaire was used to evaluate the system usability of the DitEetIk! app. The questionnaire was completed by 98.1% (207/211) of the included participants. The mean score per question ranged from 1.7 (SD 0.8) to 3.7 (SD 0.9). The mean overall score of the SUS was 66.6 (SD 15.1).

Figure 2. Participant recruitment and selection flow. A plausible recording day was a recording day with a ratio of energy intake over BMR between 0.6 and 3.0. BMR: basal metabolic rate; EI: energy intake.

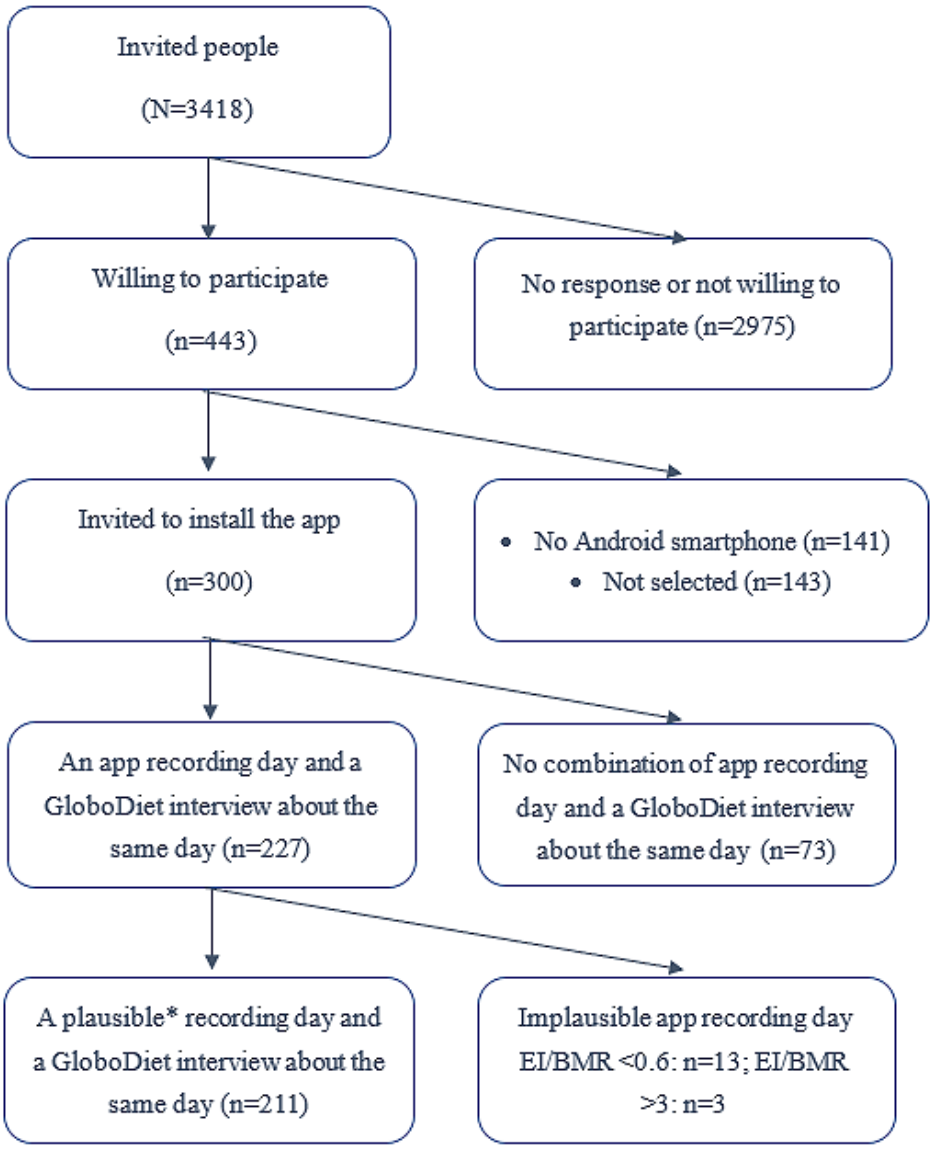


Table 1. Characteristics of the 211 participants in the DitEetIk! app study and matched participants from the Dutch National Food Consumption Survey (DNFCS) 2019 to 2021.

Characteristic	Evaluation study (n=211), n (%)	DNFCS 2019-2021 (n=211), n (%)
Gender		
Men	104 (49.3)	104 (49.3)
Women	107 (50.7)	107 (50.7)
Age category (y)		
18-39	76 (36)	75 (35.5)
40-59	77 (36.5)	78 (37)
60-79	58 (27.5)	58 (27.5)
Highest educational level attained		
Low[a]	37 (17.5)	31 (14.7)
Middle[b]	81 (38.4)	87 (41.2)
High[c]	93 (44.1)	93 (44.1)
BMI category (kg/m^2)		
<18.5	1 (0.5)	2 (0.9)
≥18.5 to ≤25	81 (38.4)	81 (38.4)
≥25 to ≤30	73 (34.6)	73 (34.6)
≥30	56 (26.5)	55 (26.1)

[a]Low educational level: primary education, lower vocational education, or advanced elementary education.

[b]Middle educational level: intermediate vocational education or higher secondary education.

[c]High educational level: higher vocational education and university.

Relative Validity for Food Groups

For 44% (7/16) of the food groups, no statistically significant differences between the median intake of food groups assessed using the DitEetIk! app food record and the 24-hour dietary recall were observed (Table 2). No statistically significant differences were observed for the median consumption of vegetables (*P*=.13), meat (*P*=.10), eggs (*P*=.44), nuts (*P*=.73), bread (*P*=.95), potatoes (*P*=.96), and snacks (*P*=.41).

Statistically significant differences of >10% between the 2 methods were observed for the median consumption of fruit (83 vs 130 g/d; *P*=.005), added fats (12 vs 17 g/d; *P*=.001), milk and milk products (219 vs 252 g/d; *P*=0.02), cereal products (6 vs 20 g/d; *P*=.01), and sauces (2 vs 22 g/d; *P*<.001). For all food groups except "Other," the DitEetIk! app assessed a lower median consumption and, therefore, relatively underestimated the consumption of food groups compared with the 24-hour dietary recalls. However, when looking at the mean rather than the median, for 7 food groups, higher values were observed in the DitEetIk! app. Spearman correlation coefficients ranged from 0.50 for the food group "Other" to 0.88 for the food groups *Potatoes*, *Sandwich spreads*, and *Snacks*. The median correlation coefficient was 0.78.

A significant difference in the number of consumers between the 2 methods was observed for the food groups *Fruit* (*P*=.002), *Added fats* (*P*=.02), *Cereal products* (*P*=.03), *Sauces* (*P*<.001), and *Other products* (*P*<.001; Table 3). This list includes 4 of the 5 food groups for which differences in median intake for all participants were statistically significant and of >10%. Focusing on the 5 food groups, the median intake of consumers differed significantly for *Added fats* and *Sauces* (median 15 and 24 grams per day in the DitEetIk! app vs 20 and 45 grams per day in the GloboDiet 24-hour dietary recalls; *P*=.06 and *P*=.004) but not for *Fruit*, *Milk and milk products*, and *Cereal products*, whereas for the food groups *Fish* (90 vs 115 g/d), *Cheese* (36 vs 45 g/d), *Sandwich spreads* (26 vs 20 g/d), and *Soups* (188 vs 50 g/d), significant differences of >10% were also observed in the median intake of consumers only.

Within the strata of educational level (3 classes), a few significant differences that were ≥10% for the median were observed for food groups (Multimedia Appendix 1). For sauces (low educational level *P*=.02; middle educational level *P*=.006; high educational level *P*<.001), the differences were consistent across the 3 educational levels. For milk and milk products, only those with a high educational level had a significantly lower median in the DitEetIk! app compared with the GloboDiet 24-hour dietary recalls (*P*=.01). Within the strata of BMI (3 levels), more differences were observed, but there was no consistent pattern in which differences were generally smaller or larger for persons in one of the BMI classes (Multimedia Appendix 2).

Table 2. The mean, SD, median, and IQR of consumption of food groups[a] as assessed using the DitEetIk! app and 24-hour dietary recalls for the same day and their correlation for the 211 participants with plausible energy intakes.

Food group	DitEetIk! app food record (g/d)		GloboDiet 24-hour dietary recall (g/d)		Wilcoxon signed rank test *P* value[b]	Spearman correlation coefficient
	Values, mean (SD)	Values, median (IQR)	Values, mean (SD)	Values, median (IQR)		
Vegetables	163 (200)	117 (31-226)	160 (144)	130 (50-240)	.13	0.76
Fruit	128 (186)	83 (0-188)	140 (146)	130 (0-217)	.005	0.79
Added fats	16 (16)	12 (3-24)	19 (15)	17 (6-29)	.001	0.54
Meat	103 (112)	73 (23-135)	92 (83)	75 (33-120)	.10	0.70
Eggs	17 (37)	0 (0-0)	17 (34)	0 (0-13)	.44	0.76
Nuts	15 (30)	0 (0-20)	15 (30)	0 (0-20)	.73	0.84
Milk (products)	264 (263)	219 (16-391)	288 (248)	252 (80-423)	.02	0.80
Cheese	33 (36)	30 (0-56)	39 (44)	31 (0-62)	.006	0.76
Bread	146 (113)	126 (70-199)	138 (88)	132 (70-180)	.95	0.85
Cereal products	67 (133)	6 (0-88)	74 (106)	20 (0-119)	.01	0.80
Potatoes	72 (119)	0 (0-128)	66 (104)	0 (0-120)	.96	0.88
Drinks	1888 (956)	1836 (1275-2311)	2097 (889)	1963 (1582-2539)	<.001	0.68
Sandwich spreads	15 (27)	0 (0-20)	12 (23)	0 (0-15)	.05	0.88
Snacks	91 (119)	52 (15-118)	83 (89)	56 (14-126)	.41	0.88
Sauces	21 (37)	2 (0-26)	33 (38)	22 (0-57)	<.001	0.60
Other	13 (52)	0 (0-10)	5 (12)	0 (0-5)	<.001	0.50

[a]Food groups are Wheel of Five food groups—main groups [23]. The food groups *Fish*, *Legumes*, and *Soups* were excluded as the 75th percentile was 0 for both methods. Table 3 provides more information on these food groups.

[b]Wilcoxon signed rank test (normal approximation) of the differences between intake assessed using the DitEetIk! app and the GloboDiet 24-hour dietary recalls for the same day.

Table 3. Number of consumers of a food group and median of consumed amount of a food group[a] for consumers only as assessed using the DitEetIk! app and the 24-hour dietary recalls (n=211).

Food group	DitEetIk! app food record (g/d)		GloboDiet 24-hour dietary recall (g/d)		Mc Nemar test *P* value[b]	Wilcoxon signed rank test *P* value[c]
	Consumers (n=211), n (%)	Consumers, median	Consumers (n=211), n (%)	Consumers, median		
Vegetables	175 (82.9)	146	181 (85.8)	147	.16	.23
Fruit	134 (63.5)	156	148 (70.1)	165	.002	.21
Added fats	173 (82)	15	186 (88.2)	20	.02	.06
Fish	25 (11.8)	90	26 (12.3)	115	.66	.04
Legumes	5 (2.4)	111	10 (4.7)	79	.06	.38
Meat	178 (84.4)	86	179 (84.8)	85	.81	.06
Eggs	52 (24.6)	50	61 (28.9)	50	.08	.83
Nuts	66 (31.3)	32	70 (33.2)	31	.32	.29
Milk and milk products	165 (78.2)	265	173 (82)	302	.06	.13
Cheese	138 (65.4)	36	145 (68.7)	45	.11	.03
Bread	194 (91.9)	136	198 (93.8)	140	.10	.71
Cereal products	113 (53.6)	70	124 (58.8)	98	.03	.05
Potatoes	87 (41.2)	150	90 (42.7)	142	.44	.91
Drinks	206 (97.6)	1848	211 (100)	1963	N/A[d]	<.001
Sandwich spreads	85 (40.3)	26	89 (42.2)	20	.25	.01
Soups	33 (15.6)	188	28 (13.3)	50	.10	<.001
Snacks	173 (82)	72	172 (81.5)	78	.76	.69
Sauces	116 (55)	24	143 (67.8)	45	<.001	.004
Other	126 (59.7)	5	59 (28)	12	<.001	.11

[a]Food groups are Wheel of Five food groups [23].

[b]McNemar test of the differences between the number of consumers of the DitEetIk! app and the GloboDiet 24-hour dietary recalls for the same day.

[c]Wilcoxon signed rank test (normal approximation) of the differences between intake assessed using the DitEetIk! app and the GloboDiet 24-hour dietary recalls for the same day for consumers of each food group.

[d]N/A: not applicable.

Relative Validity for Nutrients

For 41% (12/29) of the nutrients, the median intake assessed using the DitEetIk! app and the 24-hour dietary recalls did not differ significantly (Table 4). Statistically significant differences of >10% between the 2 methods were observed for the median intake of vitamin A (453 vs 515 µg retinol activity equivalents/d; *P*<.001), folate (equivalents; 241 vs 269 µg/d; *P*<.001), vitamin D (1.8 vs 2.1 µg/d; *P*<.001), and vitamin E (9.7 vs 11.7 mg/d; *P*<.001). For 8 nutrients, significant differences were between 5% and 10%, and for 5 nutrients, they were <5%. With the exceptions of mono- and disaccharides and magnesium, in all cases of statistically significant differences, the DitEetIk! app had lower values than the 24-hour dietary recalls. Spearman correlation coefficients between intake assessed using the DitEetIk! app and the 24-hour dietary recalls ranged from 0.55 for sodium to 0.9 for alcohol, with a median correlation coefficient of 0.72.

Compared with the expected energy intake, the mean underreporting using the DitEetIk! app was 19.9% (0.316/1.59; calculated as [measured-expected physical activity level]/[expected physical activity level]) versus 20.1% (0.321/1.59) using the 24-hour dietary recalls. At the individual level, 20.4% (43/211) of the participants could be considered to be underreporting and 0.5% (1/211) of the participants could be considered to be overreporting energy intake using the DitEetIk! app. For the 24-hour dietary recalls, 18% (38/211) of the participants could be considered to be underreporting, and none were overreporting.

Figure 3 shows the Bland-Altman plots for energy intake and for fat, carbohydrates, and protein expressed as a percentage of energy intake. The mean differences and 95% limits of agreement were 14 (−1096 to 1124) for energy in kilocalories, −2.6 (−16.2 to 11) for energy percentage derived from fat, 2.5 (−10.9 to 16) for energy percentage derived from carbohydrates, and −7.5 (−7.5 to 7.4) for energy percentage derived from protein.

Table 4. The mean, SD, median, and IQR of energy and nutrient intake per day as assessed using the DitEetIk! app and the 24-hour dietary recalls and their correlation (n=211).

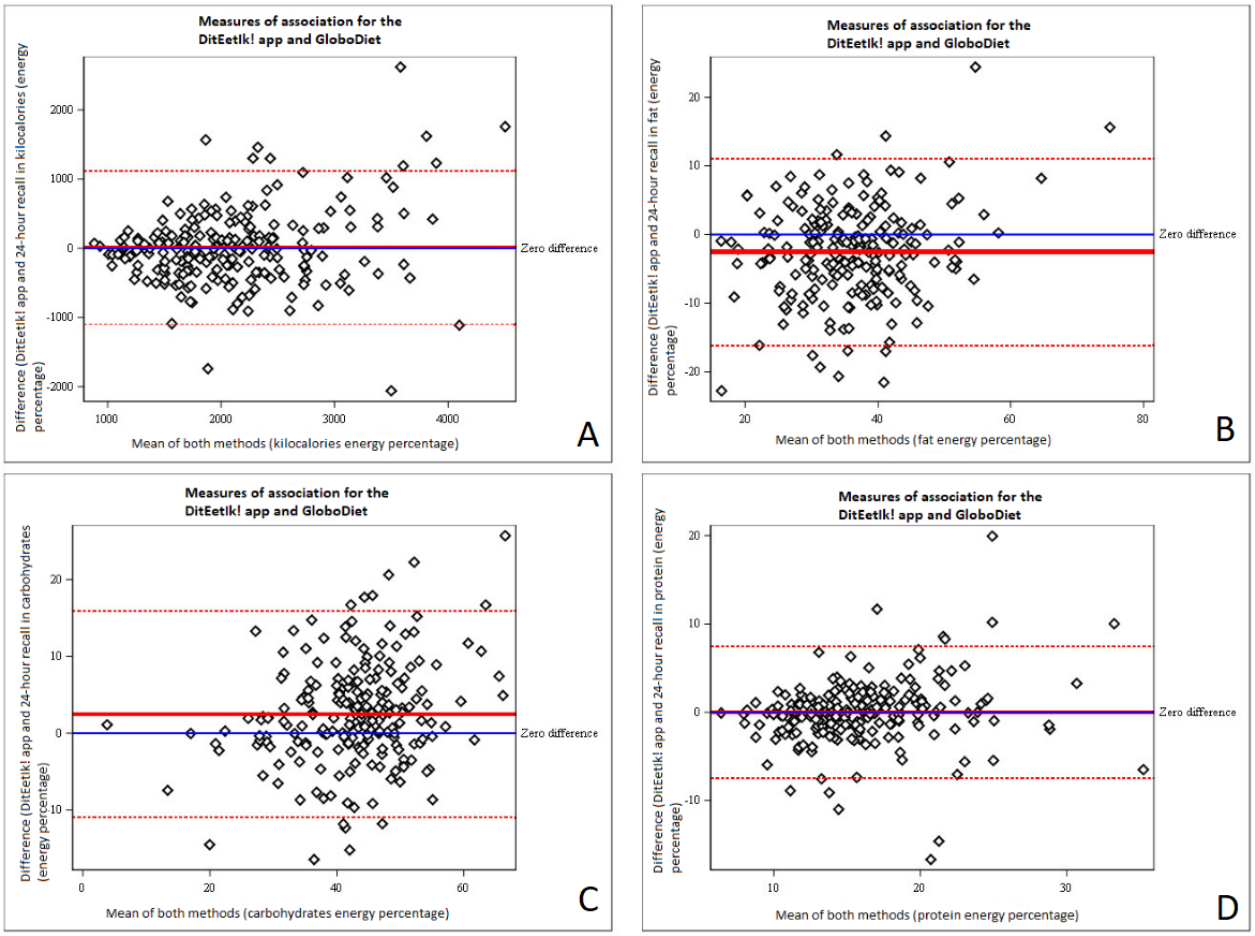
Nutrients	DitEetIk! app food record		GloboDiet 24-hour dietary recall		Wilcoxon signed rank test *P* value[a]	Spearman correlation coefficient
	Values, mean (SD)	Values, median (IQR)	Values, mean (SD)	Values, median (IQR)		
Energy (kcal)	2112 (795)	1994 (1534-2488)	2100 (675)	1994 (1614-2539)	.59	0.75
Fat (g)	83 (42)	76 (51-105)	88 (39)	81 (61-110)	.001	0.74
Saturated fatty acids (g)	29 (15)	27 (18-38)	31 (15)	28 (20-39)	.01	0.70
Protein (g)	82 (34)	77 (59-100)	82 (30)	77 (62-101)	.54	0.70
Vegetable protein (g)	33 (15)	30 (23-41)	33 (13)	30 (23-42)	.37	0.75
Carbohydrates (g)	230 (99)	215 (166-280)	215 (77)	208 (160-259)	.04	0.78
Mono- and disaccharides (g)	93 (53)	81 (56-120)	85 (43)	80 (56-111)	.01	0.80
Fiber (g)	21 (11)	20 (14-26)	21 (9)	20 (14-26)	.15	0.75
Alcohol (g)	10 (21)	0 (0-11)	9 (20)	0 (0-10)	.45	0.92
Water (g)	2681 (1012)	2570 (2048-3229)	2907 (955)	2775 (2285-3349)	<.001	0.68
Vitamin A (µg RAE[b])	714 (1007)	453 (286-685)	804 (1013)	515 (334-806)	<.001	0.74
Vitamin B1 (mg)	1.0 (0.6)	0.9 (0.7-1.3)	1.0 (0.5)	0.9 (0.7-1.3)	.21	0.63
Vitamin B2 (mg)	1.3 (0.6)	1.3 (0.9-1.6)	1.4 (0.6)	1.3 (1.0-1.8)	.002	0.73
Vitamin B3 (mg)	18.8 (10.8)	16.5 (10.8-24.4)	19.0 (9.3)	17.3 (11.5-24.4)	.17	0.70
Vitamin B6 (mg)	1.5 (0.7)	1.4 (1.0-1.9)	1.6 (0.7)	1.5 (1.1-1.9)	<.001	0.71
Folate (equivalents; µg)	266 (148)	241 (172-331)	286 (122)	269 (203-357)	<.001	0.67
Vitamin B12 (µg)	3.8 (3.2)	2.9 (2.0-4.8)	4.2 (3.4)	3.2 (2.2-5.1)	.001	0.78
Vitamin C (mg)	94 (123)	65 (32-108)	88 (74)	67 (33-128)	.09	0.77
Vitamin D (µg)	2.3 (2.2)	1.8 (0.9-2.9)	2.7 (2.5)	2.1 (1.2-3.5)	<.001	0.64
Vitamin E (mg)	11.4 (7.3)	9.7 (6.5-14.3)	13.1 (6.7)	11.7 (8.2-16.7)	<.001	0.58
Calcium (mg)	965 (451)	909 (658-1231)	1006 (452)	921 (664-1295)	.12	0.74
Iodine (µg)[c]	173 (92)	159 (111-213)	172 (79)	162 (116-209)	.42	0.80
Iron (mg)	10.3 (4.5)	9.3 (7.2-12.7)	10.2 (4.0)	9.4 (7.2-12.1)	.41	0.73
Magnesium (mg)	338 (126)	334 (249-403)	346 (118)	328 (264-411)	.04	0.72
Phosphorus (mg)	1504 (537)	1466 (1086-1864)	1506 (489)	1478 (1135-1790)	.63	0.72
Potassium (mg)	3202 (1202)	3147 (2314-3860)	3206 (1059)	3106 (2475-3886)	.38	0.72
Selenium (µg)	52 (44)	43 (31-58)	50 (32)	43 (31-63)	.60	0.68
Sodium (mg)[c]	2625 (1443)	2373 (1637-3096)	2410 (1040)	2283 (1694-2875)	.05	0.55
Zinc (mg)	10.1 (4.5)	9.4 (7.0-12.5)	10.3 (3.9)	10.0 (7.5-12.3)	.10	0.70

[a]Wilcoxon signed rank test (normal approximation) of the differences between intake assessed using the DitEetIk! app and the GloboDiet 24-hour dietary recalls for the same day.

[b]RAE: retinol activity equivalents.

[c]Sodium and iodine from food only.

Figure 3. Bland-Altman plot assessed using the DitEetIk! app and GloboDiet (GD) 24-hour dietary recalls (n=211). (A) Results for energy intake, (B) results for fat (energy percentage), (C) results for carbohydrates (energy percentage), and (D) results for protein (energy percentage).



Evaluation of the GloboDiet 24-Hour Dietary Recall Data

For almost all food groups, the consumption of participants in the DitEetIk! app evaluation study, assessed using the GloboDiet 24-hour dietary recall, and the consumption of matched participants selected from the DNFCS 2019 to 2021 did not differ significantly (Table 5). Statistically significant differences of >10% in median intake between the 2 methods were only observed for the consumption of milk and milk products (median 252 grams per day in the DitEetIk! app vs 282 grams per day in the GloboDiet 24-hour dietary recalls; *P*=.03) and bread (median 132 grams per day vs 105 grams per day; *P*=.03).

For all nutrients except vitamin B3, the intake of participants in the evaluation study, assessed using the GloboDiet 24-hour dietary recalls, and the intake of the matched participants selected from the DNFCS 2019 to 2021 did not differ significantly (Table 6).

Table 5. Comparison of consumption of food groups assessed using the GloboDiet 24-hour dietary recalls in the DitEetIk! app evaluation study and the first interview in the Dutch National Food Consumption Survey (DNFCS) 2019 to 2021 for a matched group of participants (n=211).

Food group[a]	DitEetIk! app evaluation study (g/d)		DNFCS 2019-2021 (g/d)		*P* value[b]
	Values, mean (SD)	Values, median (IQR)	Values, mean (SD)	Values, median (IQR)	
Vegetables	160 (144)	130 (50-240)	155 (140)	125 (53-217)	>.99
Fruit	140 (146)	130 (0-217)	124 (135)	108 (0-195)	.51
Added fats	19 (15)	17 (6-29)	22 (20)	18 (8-32)	.006
Fish	17 (57)	0 (0-0)	15 (44)	0 (0-0)	.91
Legumes	4 (20)	0 (0-0)	8 (36)	0 (0-0)	.09
Meat	92 (83)	75 (33-120)	88 (80)	77 (29-116)	.76
Eggs	17 (34)	0 (0-13)	16 (32)	0 (0-13)	.85
Nuts	15 (30)	0 (0-20)	19 (45)	0 (0-22)	.24
Milk and milk products	288 (248)	252 (80-423)	332 (267)	282 (150-484)	.03
Cheese	39 (44)	31 (0-62)	38 (39)	30 (0-62)	.85
Bread	138 (88)	132 (70-180)	117 (80)	105 (60-169)	.03
Cereal products	74 (106)	20 (0-119)	79 (108)	30 (0-122)	.57
Potatoes	66 (104)	0 (0-120)	69 (93)	0 (0-140)	.26
Drinks	2097 (889)	1963 (1582-2539)	2132 (914)	1958 (1468-2608)	.63
Sandwich spreads	12 (23)	0 (0-15)	18 (29)	0 (0-23)	.03
Soups	12 (46)	0 (0-0)	17 (66)	0 (0-0)	.57
Snacks	83 (89)	56 (14-126)	71 (79)	41 (10-114)	.24
Sauces	33 (38)	22 (0-57)	29 (44)	11 (0-36)	.14
Other	5 (12)	0 (0-5)	6 (16)	0 (0-5)	.75

[a]Food groups are Wheel of Five food groups [23].

[b]Wilcoxon signed rank test (normal approximation) of the differences between intake assessed using GloboDiet 24-hour dietary recalls in the DitEetIk! app evaluation study and the first interview with adults in the DNFCS 2019 to 2021.

Table 6. Comparison of intake of energy and selected nutrients assessed using the GloboDiet 24-hour dietary recalls in the DitEetIk! app evaluation study and the first interview with adults in the Dutch National Food Consumption Survey (DNFCS) 2019 to 2021 (n=211).

Nutrients	DitEetIk! app food record		GloboDiet 24-hour dietary recall		*P* value[a]
	Values, mean (SD)	Values, median (IQR)	Values, mean (SD)	Values, median (IQR)	
Energy (kcal)	2100 (675)	1994 (1614-2539)	2091 (761)	2036 (1506-2535)	.86
Fat (g)	88 (39)	81 (61-110)	89 (43)	81 (60-111)	.86
Saturated fatty acids (g)	31 (15)	28 (20-39)	32 (16)	30 (22-40)	.48
Protein (g)	82 (30)	77 (62-101)	80 (29)	77 (62-100)	.29
Vegetable protein (g)	33 (13)	30 (23-42)	32 (14)	30 (22-39)	.31
Carbohydrates (g)	215 (77)	208 (160-259)	217 (88)	203 (155-262)	.71
Mono- and disaccharides (g)	85 (43)	80 (56-111)	94 (55)	85 (60-112)	.21
Fiber (g)	21 (9)	20 (14-26)	22 (9)	20 (16-26)	.18
Alcohol (g)	9 (20)	0 (0-10)	8 (15)	0 (0-12)	.64
Water (g)	2907 (955)	2775 (2285-3349)	2950 (962)	2831 (2235-3481)	.75
Vitamin A (µg RAE[b])	804 (1013)	515 (334-806)	828 (1026)	570 (360-894)	.42
Vitamin B1 (mg)	1.0 (0.5)	0.9 (0.7-1.3)	1.0 (0.6)	0.9 (0.6-1.2)	.96
Vitamin B2 (mg)	1.4 (0.6)	1.3 (1.0-1.8)	1.4 (0.7)	1.4 (1.0-1.8)	.79
Vitamin B3 (mg)	19.0 (9.3)	17.3 (11.5-24.4)	17.1 (9.3)	15.7 (10.6-21.2)	.02
Vitamin B6 (mg)	1.6 (0.7)	1.5 (1.1-1.9)	1.5 (0.7)	1.4 (1.1-1.9)	.29
Folate (equivalents; µg)	286 (122)	269 (203-357)	272 (121)	243 (194-340)	.13
Vitamin B12 (µg)	4.2 (3.4)	3.2 (2.2-5.1)	4.3 (3.1)	3.6 (2.6-5.6)	.65
Vitamin C (mg)	88 (74)	67 (33-128)	82 (65)	65 (35-110)	.39
Vitamin D (µg)	2.7 (2.5)	2.1 (1.2-3.5)	2.7 (2.1)	2.2 (1.1-3.9)	.52
Vitamin E (mg)	13.1 (6.7)	11.7 (8.2-16.7)	13.5 (8.4)	12.1 (8.4-16.8)	.85
Calcium (mg)	1006 (452)	921 (664-1295)	1045 (480)	983 (702-1321)	.39
Iodine (µg)[c]	172 (79)	162 (116-209)	166 (75)	158 (112-210)	.42
Iron (mg)	10.2 (4.0)	9.4 (7.2-12.1)	10.6 (4.1)	10.0 (7.9-12.5)	.35
Magnesium (mg)	346 (118)	328 (264-411)	349 (128)	325 (250-426)	.86
Phosphorus (mg)	1506 (489)	1478 (1135-1790)	1493 (518)	1490 (1128-1781)	.67
Potassium (mg)	3206 (1059)	3106 (2475-3886)	3223 (1127)	3170 (2358-3768)	.99
Selenium (µg)	50 (32)	43 (31-63)	51 (46)	41 (30-57)	.40
Sodium (mg)[c]	2410 (1040)	2283 (1694-2875)	2321 (1042)	2125 (1562-2983)	.42
Zinc (mg)	10.3 (3.9)	10.0 (7.5-12.3)	10.5 (4.8)	10.2 (7.6-13.1)	.53

[a]Wilcoxon signed rank test (normal approximation) of the differences between intake assessed using the GloboDiet 24-hour dietary recalls in the DitEetIk! app evaluation study and the first interview with adults in the DNFCS 2019 to 2021.

[b]RAE: retinol activity equivalents.

[c]Sodium and iodine from food only.

Discussion

Principal Findings

Compared with GloboDiet 24-hour dietary recalls, the DitEetIk! app assessed similar mean levels of energy intake but somewhat lower median levels of intake for several food groups and nutrients. Differences were of >10% for fruit; added fats; cereal products; sauces; and vitamins A, D, and E and folate. Of all logged food items and beverages, most were selected via text searching, whereas the scanning functionality was used for approximately one-seventh of the food products.

Incomplete recording of consumed food items in the DitEetIk! app seems to have occurred for various food groups, such as fruit, added fats, cereal products, and sauces. During the GloboDiet 24-hour dietary recall, the trained interviewer specifically probes for easily forgotten food items [28], which

may explain this difference. Similarly, other studies evaluating mobile food record apps based on text searching reported food omissions, particularly of condiment food items [29]. In a study using wearable camera images as a reference, it was observed that the most forgotten food groups in the Australian Eat and Track app were savory sauces and condiments, vegetables, confectionery, fruit and dairy, and alternatives [30]. There are various options to stimulate complete recording. According to a review of smartphone dietary assessment tools, the most common feature to do so was to allow participants to review the records and make adjustments if information was missing or false [31]. This feature was also built into the DitEetIk! app; before submitting the food recording for one day, participants were shown an overview of reported food items and were asked whether this was complete. In addition, incorporating (more) probing questions for frequently omitted food items into DitEetIk! app could be a way to remind a participant to report such food items. This could be probing questions either linked to other food items (eg, salad dressing in the case of salads) or linked to eating occasions (eg, fruit in between meals). Alternatively, sending prompts when an eating moment is expected or when the DitEetIk! app has not been used for a certain period or allowing participants to explicitly state that they did not consume anything at a given eating occasion are also approaches with the potential to improve completeness [32].

In contrast to the results for median intake, for some food groups and nutrients, the mean intake was higher in the DitEetIk! app than in the 24-hour dietary recalls. For the food groups *Vegetables*, *Meat*, *Bread*, and *Potatoes*, this was caused by higher amounts recorded in the DitEetIk! app than those indicated in the 24-hour dietary recalls. Choosing unlikely high portion sizes in the DitEetIk! app was possible without a warning message in case the amount eaten was indicated in units rather than grams, whereas in the GloboDiet 24-hour dietary recalls, all indicated portions were converted into grams and checked against set improbable maximum values, and if needed, the interviewer was prompted to check with the participant whether the answer was correct [28]. A similar functionality could be considered for inclusion in the DitEetIk! app.

Of all logged food and beverages, most were selected via text searching. In the feedback given by participants in the remark field of the DitEetIk! app, they mentioned that finding the correct food item on the list was a challenging task. This disadvantage of extensive food item lists was also described in a systematic review [33]. This was probably the reason why the average SUS score for the DitEetIk! app was just below 70, the threshold that is generally considered “good” [34]. Only 12.8% (114/894) of all logged food items and drinks were scanned. This percentage was lower than expected based on experiences in a project in which approximately 50% of the food items were scanned using a commercial smartphone food record (personal communication by MO). The availability of branded food items in the DitEetIk! app was still limited, and these did not include food items from some supermarket chains or that were not matched to the generic food composition database. This may have affected the use of the scanning option. One could understand that participants stopped scanning barcodes after some failed attempts. Therefore, including more branded food items in the DitEetIk! app is highly recommended. In addition, a crowdsourcing function could be incorporated whereby users can contribute information on missing products, such as that developed for the FoodSwitch app [35]. The collected food product information can then be added to the database to ensure that the DitEetIk! app is supported by actual and complete product information. If more food items are scanned, food recording will probably be perceived as easier.

As described previously, based on the main findings, several options for improvement via additional DitEetIk! app functionalities can be formulated. However, one should also be careful not to burden participants with too frequent notifications, reminders, and prompts [33]. More insight on the impact of different features used in smartphone-based dietary assessment tools and the characteristics of these features on the respondents’ willingness and ability to record intake reliably and on the validity of the recorded dietary data is needed [31].

Strengths and Limitations

This study is one of the few food record validation studies (Zhang et al [29]) that report results for a rather comprehensive list of food groups and nutrients in a group of >200 men and women of various ages. An important limitation is that relative rather than objective validity was studied. Therefore, lower and higher values compared with GloboDiet 24-hour dietary recall values cannot be interpreted as underestimation or overestimation. However, the results on energy misreporting were included, which are not dependent on the subjective reporting of dietary intake. In the future, follow-up validation with doubly labeled water and excretion of nitrogen, potassium, and sodium in 24-hour urine is recommended. Another limitation is the large number of statistical tests that were conducted, which may have led to chance findings. Moreover, the study did not follow a crossover design, which might have caused a potential memory or learning effect in the 24-hour dietary recall data. However, comparing these data with those from the DNFCS 2019 to 2021 gave no indication that this occurred. People with a lower education and of higher age were included in the study population but were underrepresented. We cannot conclude whether the relative validity is similar for these population subgroups. This study did not focus on the potential selection bias of including only participants who had an Android smartphone. Such an evaluation is also important for use in a national food consumption survey. Potentially, developing an iOS version of the DitEetIk! app and offering an interview option to persons without a smartphone needs to be considered. In this study, the DitEetIk! app was described according to the recommendations of Eldridge et al [11], and the validity study was reported according to the guidance provided by Kirkpatrick et al [26].

Comparison With Prior Work

Although many different smartphone-based dietary assessment tools exist, only a few validation studies have been conducted. Burrows et al [36] concluded that their validity seems to be comparable with that of more traditional dietary assessment methods and that energy intake is often underreported. In a

review from 2013 to 2019, Zhang et al [29] identified 14 smartphone-based food records that were not image based. In the meta-analyses based on 11 tools, all of them underestimated energy intake, with a pooled effect of approximately −200 kcal and limits of agreement of 1918 kcal. The results for the DitEetIk! app fit with this picture, with a below-average mean underestimation and similar limits of agreement for energy intake. The relative validity results were also in line with those of Eetmeter, another Dutch app, although the food groups for which underestimation occurred partly differed [20]. Unlike the DitEetIk! app, Eetmeter shows energy and nutrient values for logged food items, which might influence reporting.

Conclusions

Compared with GloboDiet 24-hour dietary recalls, the DitEetIk! app assessed somewhat lower levels of intake for several food groups and nutrients. Therefore, adding functionalities to the DitEetIk! app that stimulate more complete food recording is important before using the app in national food consumption surveys. In addition, it is advisable to develop a functionality to warn participants when entering extremely large consumption amounts. Less participant burden and more detailed information about consumed food items can be obtained by stimulating the use of barcode scanning.

Acknowledgments

This study as well as the development of the DitEetIk! app was funded by the Dutch Ministry of Health, Welfare, and Sport. The authors would like to thank Dr Maarten Schipper for providing statistical advice and Dr Steffen Bruns, Henny Brants, and Marjolein de Jong for critically reviewing the manuscript.

Data Availability

The data presented in this paper are available upon request from the corresponding author. The data are not publicly available to comply with the requirements of the General Data Protection Regulation (European Union) 2016/679.

Conflicts of Interest

National Institute for Public Health and the Environment (RIVM), where the authors work, is the owner of the software.

Multimedia Appendix 1
The mean, SD, median, and IQR of consumption of food groups (g/d) by educational level as assessed using the DitEetIk! app and 24-hour dietary recalls for the same day and their correlation for 211 participants with plausible energy intakes.
[DOCX File , 29 KB - mhealth_v12i1e50196_app1.docx]

Multimedia Appendix 2
The mean, SD, median, and IQR of consumption of food groups (g/d) by BMI as assessed using the DitEetIk! app and 24-hour dietary recalls for the same day and their correlation for 211 participants with plausible energy intakes.
[DOCX File , 29 KB - mhealth_v12i1e50196_app2.docx]

References

1. De Keyzer W, Bracke T, McNaughton SA, Parnell W, Moshfegh AJ, Pereira RA, et al. Cross-continental comparison of national food consumption survey methods--a narrative review. Nutrients 2015 May 13;7(5):3587-3620 [FREE Full text] [doi: 10.3390/nu7053587] [Medline: 25984745]
2. Huybrechts I, Aglago EK, Mullee A, De Keyzer W, Leclercq C, Allemand P, et al. Global comparison of national individual food consumption surveys as a basis for health research and integration in national health surveillance programmes. Proc Nutr Soc 2017 Nov;76(4):549-567. [doi: 10.1017/S0029665117001161] [Medline: 28803558]
3. van Bussel LM, van Rossum CT, Temme EH, Boon PE, Ocké MC. Educational differences in healthy, environmentally sustainable and safe food consumption among adults in the Netherlands. Public Health Nutr 2020 Aug;23(12):2057-2067 [FREE Full text] [doi: 10.1017/S1368980019005214] [Medline: 32383426]
4. Ahuja JK, Moshfegh AJ, Holden JM, Harris E. USDA food and nutrient databases provide the infrastructure for food and nutrition research, policy, and practice. J Nutr 2013 Feb;143(2):241S-249S [FREE Full text] [doi: 10.3945/jn.112.170043] [Medline: 23269654]
5. Ocke MC, van Rossum CT, de Boer EJ. Dietary surveys: national food intake. In: Caballero B, Finglas PM, Toldrá F, editors. Encyclopedia of Food and Health. New York, NY: Academic Press; 2015. [doi: 10.1016/B978-0-12-384947-2.00237-3]
6. De Henauw S, Brants HA, Becker W, Kaic-Rak A, Ruprich J, Sekula W, EFCOSUM Group. Operationalization of food consumption surveys in Europe: recommendations from the European Food Consumption Survey Methods (EFCOSUM) project. Eur J Clin Nutr 2002 May;56 Suppl 2:S75-S88. [doi: 10.1038/sj.ejcn.1601431] [Medline: 12082520]

7. de Boer EJ, Slimani N, van 't Veer P, Boeing H, Feinberg M, Leclercq C, EFCOVAL Consortium. The European Food Consumption Validation Project: conclusions and recommendations. Eur J Clin Nutr 2011 Jul;65 Suppl 1:S102-S107. [doi: 10.1038/ejcn.2011.94] [Medline: 21731001]
8. European Food Safety Authority. Guidance on the EU Menu methodology. EFSA J 2014 Dec;12(12):3944 [FREE Full text] [doi: 10.2903/j.efsa.2014.3944]
9. Thompson FE, Subar AF. Dietary assessment methodology. In: Coulston AM, Boushey CJ, Ferruzzi M, editors. Nutrition in the Prevention and Treatment of Disease. 3rd edition. New York, NY: Academic Press; 2013:5-46.
10. Ocké MC, Slimani N, Brants H, Buurma-Rethans E, Casagrande C, Nicolas G, EFCOVAL Consortium. Potential and requirements for a standardized pan-European food consumption survey using the EPIC-Soft software. Eur J Clin Nutr 2011 Jul 06;65 Suppl 1(S1):S48-S57. [doi: 10.1038/ejcn.2011.87] [Medline: 21731005]
11. Eldridge AL, Piernas C, Illner AK, Gibney MJ, Gurinović MA, de Vries JH, et al. Evaluation of new technology-based tools for dietary intake assessment-an ILSI Europe dietary intake and exposure task force evaluation. Nutrients 2018 Dec 28;11(1):55 [FREE Full text] [doi: 10.3390/nu11010055] [Medline: 30597864]
12. Sharp DB, Allman-Farinelli M. Feasibility and validity of mobile phones to assess dietary intake. Nutrition 2014;30(11-12):1257-1266. [doi: 10.1016/j.nut.2014.02.020] [Medline: 24976425]
13. Dinnissen CS, Ocké MC, Buurma-Rethans EJ, van Rossum CT. Dietary changes among adults in the Netherlands in the period 2007-2010 and 2012-2016. Results from two cross-sectional national food consumption surveys. Nutrients 2021 Apr 30;13(5):1520 [FREE Full text] [doi: 10.3390/nu13051520] [Medline: 33946365]
14. Ocké MC, Hulshof KF, van Rossum CT. The Dutch national food consumption survey 2003. Arch Public Health 2006;63:227-241 [FREE Full text]
15. Van Rossum CTM, Buurma-Rethans EJM, Dinnissen CS, Beukers M, Brants HAM, Ocké MC. The diet of the Dutch; Results of the Dutch National Food Consumption Survey 2012-2016. RIVM, 2020-0083 2020 [FREE Full text]
16. Slimani N, Casagrande C, Nicolas G, Freisling H, Huybrechts I, Ocké MC, EFCOVAL Consortium. The standardized computerized 24-h dietary recall method EPIC-Soft adapted for pan-European dietary monitoring. Eur J Clin Nutr 2011 Jul;65 Suppl 1:S5-15. [doi: 10.1038/ejcn.2011.83] [Medline: 21731006]
17. Westenbrink S, van der Vossen-Wijmenga W, Toxopeus I, Milder I, Ocké M. LEDA, the branded food database in the Netherlands: data challenges and opportunities. J Food Compost Anal 2021 Sep;102:104044 [FREE Full text] [doi: 10.1016/j.jfca.2021.104044]
18. Arends-Tóth J. ICT-gebruik van huishoudens en personen. Centraal Bureau voor de Statistiek. URL: http://tinyurl.com/242xc8w5 [accessed 2023-01-04]
19. Van Rossum CTM, Buurma-Rethans EJM, Dinnissen CS, Beukers M, Brants HAM, Ocké MC. The diet of the Dutch; Results of the Dutch National Food Consumption Survey 2019-2021 on food consumption and evaluation with dietary guidelines. RIVM, 2022-0190 2023 [FREE Full text]
20. Ocké M, Dinnissen C, Stafleu A, de Vries J, van Rossum C. Relative validity of MijnEetmeter: a food diary app for self-monitoring of dietary intake. Nutrients 2021 Mar 30;13(4):1135 [FREE Full text] [doi: 10.3390/nu13041135] [Medline: 33808209]
21. Brooke J. Usability evaluation in industry. In: Brooke J, editor. SUS: A 'Quick and Dirty' Usability Scale. London, UK: CRC Press; 1996:189-194.
22. Schuurman R, Beukers M, van RC. Eet en drinkt Nederland volgens de Richtlijnen Schijf van Vijf? Resultaten van de voedselconsumptiepeiling 2012-2016. Utrecht, Netherlands: Rijksinstituut voor Volksgezondheid en Milieu; 2012.
23. Brink E, van Rossum C, Postma-Smeets A, Stafleu A, Wolvers D, van Dooren C, et al. Development of healthy and sustainable food-based dietary guidelines for the Netherlands. Public Health Nutr 2019 Sep;22(13):2419-2435 [FREE Full text] [doi: 10.1017/S1368980019001435] [Medline: 31262374]
24. Nederlands Voedingsstoffenbestand (NEVO). version 2019/6. Ministerie van Volksgezondheid, Welzijn en Sport. 2019. URL: http://www.rivm.nl/nevo_en/online/ [accessed 2024-01-04]
25. Henry CJ. Basal metabolic rate studies in humans: measurement and development of new equations. Public Health Nutr 2005 Oct;8(7A):1133-1152. [doi: 10.1079/phn2005801] [Medline: 16277825]
26. Kirkpatrick SI, Baranowski T, Subar AF, Tooze JA, Frongillo EA. Best practices for conducting and interpreting studies to validate self-report dietary assessment methods. J Acad Nutr Diet 2019 Nov;119(11):1801-1816. [doi: 10.1016/j.jand.2019.06.010] [Medline: 31521583]
27. Altman DG, Bland JM. Measurement in medicine: the analysis of method comparison studies. Statistician 1983 Sep;32(3):307 [FREE Full text] [doi: 10.2307/2987937]
28. Crispim SP, Nicolas G, Casagrande C, Knaze V, Illner A, Huybrechts I, et al. Quality assurance of the international computerised 24 h dietary recall method (EPIC-Soft). Br J Nutr 2013 Sep 04;111(3):506-515. [doi: 10.1017/s0007114513002766]
29. Zhang L, Misir A, Boshuizen H, Ocké M. A systematic review and meta-analysis of validation studies performed on dietary record apps. Adv Nutr 2021 Dec 01;12(6):2321-2332 [FREE Full text] [doi: 10.1093/advances/nmab058] [Medline: 34019624]

30. Chan V, Davies A, Wellard-Cole L, Lu S, Ng H, Tsoi L, et al. Using wearable cameras to assess foods and beverages omitted in 24 hour dietary recalls and a text entry food record app. Nutrients 2021 May 26;13(6):1806 [FREE Full text] [doi: 10.3390/nu13061806] [Medline: 34073378]
31. König LM, Van Emmenis M, Nurmi J, Kassavou A, Sutton S. Characteristics of smartphone-based dietary assessment tools: a systematic review. Health Psychol Rev 2022 Dec 21;16(4):526-550. [doi: 10.1080/17437199.2021.2016066] [Medline: 34875978]
32. Ziesemer K, König LM, Boushey CJ, Villinger K, Wahl DR, Butscher S, et al. Occurrence of and reasons for "missing events" in mobile dietary assessments: results from three event-based ecological momentary assessment studies. JMIR Mhealth Uhealth 2020 Oct 14;8(10):e15430 [FREE Full text] [doi: 10.2196/15430] [Medline: 33052123]
33. König LM, Attig C, Franke T, Renner B. Barriers to and facilitators for using nutrition apps: systematic review and conceptual framework. JMIR Mhealth Uhealth 2021 Apr 01;9(6):e20037 [FREE Full text] [doi: 10.2196/20037] [Medline: 34254938]
34. Bangor A, Staff T, Kortum P, Miller J, Staff T. Determining what individual SUS scores mean: adding an adjective rating scale. J Usability Stud 2009;4(3):114-123 [FREE Full text]
35. Dunford E, Trevena H, Goodsell C, Ng KH, Webster J, Millis A, et al. FoodSwitch: a mobile phone app to enable consumers to make healthier food choices and crowdsourcing of national food composition data. JMIR Mhealth Uhealth 2014 Aug 21;2(3):e37 [FREE Full text] [doi: 10.2196/mhealth.3230] [Medline: 25147135]
36. Burrows TL, Ho YY, Rollo ME, Collins CE. Validity of dietary assessment methods when compared to the method of doubly labeled water: a systematic review in adults. Front Endocrinol (Lausanne) 2019;10:850 [FREE Full text] [doi: 10.3389/fendo.2019.00850] [Medline: 31920966]

Abbreviations

BMR: basal metabolic rate
DNFCS: Dutch National Food Consumption Survey
LEDA: Dutch-branded food database (Levensmiddelendatabank)
SUS: System Usability Scale

Edited by L Buis; submitted 22.06.23; peer-reviewed by J Kohl, C Matthys; comments to author 25.08.23; revised version received 09.10.23; accepted 01.12.23; published 09.02.24.

Please cite as:
Ocké M, Dinnissen CS, van den Bogaard C, Beukers M, Drijvers J, Sanderman-Nawijn E, van Rossum C, Toxopeus I
A Smartphone Food Record App Developed for the Dutch National Food Consumption Survey: Relative Validity Study
JMIR Mhealth Uhealth 2024;12:e50196
URL: https://mhealth.jmir.org/2024/1/e50196
doi:10.2196/50196
PMID:38335009

Original Paper

Feasibility Study on Menstrual Cycles With Fitbit Device (FEMFIT): Prospective Observational Cohort Study

Anna-Lena Lang[1], MD; Rosa-Lotta Bruhn[2]; Maya Fehling[1], Dr med; Anouk Heidenreich[1]; Jonathan Reisdorf[1]; Ifrah Khanyaree[1], MSc; Maike Henningsen[2], Dr med, PD; Cornelius Remschmidt[1], Dr med, PD

[1]Data4Life gGmbH, Potsdam, Germany

[2]Faculty of Health, University Witten Herdecke, Witten Herdecke, Germany

Corresponding Author:
Anna-Lena Lang, MD
Data4Life gGmbH
c/o Digital Health Cluster im Hasso-Plattner-Institut
Potsdam, 14482
Germany
Phone: 49 15756025551
Email: annalena.lang.26@gmail.com

Abstract

Background: Despite its importance to women's reproductive health and its impact on women's daily lives, the menstrual cycle, its regulation, and its impact on health remain poorly understood. As conventional clinical trials rely on infrequent in-person assessments, digital studies with wearable devices enable the collection of longitudinal subjective and objective measures.

Objective: The study aims to explore the technical feasibility of collecting combined wearable and digital questionnaire data and its potential for gaining biological insights into the menstrual cycle.

Methods: This prospective observational cohort study was conducted online over 12 weeks. A total of 42 cisgender women were recruited by their local gynecologist in Berlin, Germany, and given a Fitbit Inspire 2 device and access to a study app with digital questionnaires. Statistical analysis included descriptive statistics on user behavior and retention, as well as a comparative analysis of symptoms from the digital questionnaires with metrics from the sensor devices at different phases of the menstrual cycle.

Results: The average time spent in the study was 63.3 (SD 33.0) days with 9 of the 42 individuals dropping out within 2 weeks of the start of the study. We collected partial data from 114 ovulatory cycles, encompassing 33 participants, and obtained complete data from a total of 50 cycles. Participants reported a total of 2468 symptoms in the daily questionnaires administered during the luteal phase and menses. Despite difficulties with data completeness, the combined questionnaire and sensor data collection was technically feasible and provided interesting biological insights. We observed an increased heart rate in the mid and end luteal phase compared with menses and participants with severe premenstrual syndrome walked substantially fewer steps (average daily steps 10,283, SD 6277) during the luteal phase and menses compared with participants with no or low premenstrual syndrome (mean 11,694, SD 6458).

Conclusions: We demonstrate the feasibility of using an app-based approach to collect combined wearable device and questionnaire data on menstrual cycles. Dropouts in the early weeks of the study indicated that engagement efforts would need to be improved for larger studies. Despite the challenges of collecting wearable data on consecutive days, the data collected provided valuable biological insights, suggesting that the use of questionnaires in conjunction with wearable data may provide a more complete understanding of the menstrual cycle and its impact on daily life. The biological findings should motivate further research into understanding the relationship between the menstrual cycle and objective physiological measurements from sensor devices.

(JMIR Mhealth Uhealth 2024;12:e50135) doi:10.2196/50135

KEYWORDS

women's health; menstrual cycle; premenstrual syndrome; PMS; mobile app; wearable device; sensor data; digital health

Introduction

An estimated 56% to 93% of women across the world experience recurrent painful periods [1-5]. Menstrual pain is debilitating for many women and has a major impact on their health-related quality of life [6-11]. One condition that can cause painful menstruation is endometriosis, which affects 10% to 15% of women of reproductive age [12]. In addition to severe pain, intermenstrual bleeding, painful periods (dysmenorrhea), painful intercourse (dyspareunia), painful bowel movements (dyschezia), and painful urination (dysuria), endometriosis can cause infertility, increase psychological distress, and can affect sexuality and relationships [12,13]. Despite its debilitating nature, it still takes around 6 to 8 years from symptom onset to diagnosis [14,15]. The gold standard for diagnosis is still laparoscopy, which is not only risky but also expensive. With the rise of gender equality movements, clinical research into menstrual health has slowly gained more attention in recent years, with researchers emphasizing the importance of timely diagnosis and treatment of menstrual cycle–related concerns and disorders [16]. Each individual has a different baseline of subjective pain experience and the availability of sufficient baseline data can enable a shift toward precision medicine in menstrual health [17]. Frequent and continuous data collection can provide an understanding of symptom variability, which is likely to be an important contributor to variability in treatment response [17]. However, traditional clinical trials to date have relied on infrequent in-person assessments and subjective retrospective data, failing to capture the daily changes in physical and mental well-being that occur over the course of the menstrual cycle [18,19]. In this regard, commercially available menstrual tracker apps offer new opportunities for research. Women can continuously track their menstrual health in digital diaries, with self-learning algorithms continuously improving predictions and educational content empowering users to increase their knowledge about menstrual health [20-22]. As a complement to subjective digital questionnaires, commercially available wearable technology can provide an easy way to collect continuous and objective real-world health data for women's health research purposes. In the example of endometriosis, the combined collection of subjective data from digital diaries and objective wearable devices could help distinguish between "normal" menstrual pain and pain associated with endometriosis, potentially speeding up the diagnosis process and avoiding unnecessary invasive tests.

Previous research, albeit with small sample sizes, has already shown that the collection of sensor data such as step count, heart rate, and sleep duration combined with self-reported menstrual cycle data can uncover interesting correlations and advance knowledge about menstrual health [23-26]. For example, wearables are already increasingly being used to evaluate alternative contraceptive methods and to predict the fertility window [26-29]. Interesting findings from previous research using wearables include an observed lower distal skin temperature (as measured with an Oura ring) during ovulation, as well as higher heart rate in the ovulatory, mid, and late luteal phases [24,25]. While previous studies show a potential effect of physical activity on the menstrual cycle and vice versa [30-32], none of them used wearable devices for daily activity tracking during the menstrual cycle. Similarly, there is a lack of studies using wearable data to analyze sleep behavior during the menstrual cycle. Current scientific knowledge on changes in sleep behavior across the menstrual cycle is conflicting, with some studies finding a decrease in subjective and objective sleep quality during the premenstrual phase and menses [33-36], while other studies did not find such correlations [24,37]. The ongoing Apple Women's Health Study [27], a mobile app–based longitudinal cohort study that includes both survey and sensor-based data, has not yet published results related to sensor data.

Combining subjective data from women's health questionnaires with objective sensor data from wearables can not only facilitate cycle tracking for the everyday consumer but also allow researchers and participants alike, to gain a deeper insight into the clinical changes during the menstrual cycle. Commercially available wrist-worn sensor devices, such as the Fitbit Inspire 2 device, are thereby not only much more affordable than research-grade sensor devices, but can be nearly as accurate [38,39]. However, real-world data collection studies conducted exclusively in the home can present difficulties in terms of retention and adherence to the study protocol [19,40]. To explore the feasibility of consistently collecting wearable and questionnaire data across multiple menstrual cycles, we conducted a 12-week feasibility study with 42 participants. Using the Fitbit Inspire 2 device and digital questionnaires within the Data4Life study app, our primary objective was to refine methods for collecting authentic menstrual cycle data in a real-world setting. This included assessing participant retention rates and gathering usability feedback on their engagement with wearable devices and digital questionnaires. In addition, our study explored potential clinical correlations between wearable and questionnaire data, with the aim of uncovering potential correlations with key parameters of the menstrual cycle.

Methods

Study Design

This digital prospective cohort study was conducted mainly online with enrollment between December 2021 and April 2022. This study was called the Feasibility Study of Menstrual Cycles With Fitbit Device (FEMFIT). Participants were recruited by their local gynecologist at a practice in Berlin, Germany. The owner of the practice, who is also a coauthor of this study (MH), received financial incentives for recruiting participants. During recruitment, participants received a token to access the FEMFIT study in the study app. During account creation, participants provided digital informed consent to share their data for research. All participants received a free Fitbit Inspire 2 (Fitbit International Limited) as an incentive after completing the study. Participants were asked to wear the device at all times throughout the study period of 3 menstrual cycles or 12 weeks. Dropout was defined as individuals who did not provide data for more than 14 days. This cut-off was deliberately set to manage participant dropout while ensuring the inclusion of data from individuals with shorter menstrual cycles.

The primary outcome variables of this study centered on assessing the feasibility of continuous menstrual cycle data collection using the Fitbit Inspire 2 device and digital questionnaires. Specifically, we focused on participant retention rates and usability feedback regarding engagement with these tools. In addition, secondary outcome variables included exploring potential clinical correlations between the collected wearable and questionnaire data and key menstrual cycle parameters.

Data Collection: Digital Questionnaires and Wearable Data

Prior to enrollment, participants completed a paper questionnaire to assess their digital literacy [41] (Multimedia Appendix 1).

Digital questionnaires (Multimedia Appendix 1; Figure S2 in Multimedia Appendix 2) were accessible within the app at varying intervals, with email notifications reminding participants of new questions. Email notifications reminded participants of newly available questions. At enrollment, participants provided demographic information and were asked for clinical information with a focus on women's health, including preexisting conditions, use of hormonal contraception, cycle regularity, and the first day of the last menses (Multimedia Appendix 1). After enrollment, weekly questionnaires focused on monitoring the duration of wearable device use duration and assessing mental and physical well-being (Multimedia Appendix 1). Notably, the specific results of the mental and physical well-being assessments are not presented in this publication. Triggered by the first day of their last menstrual cycle, from day 13 to day 5 of their menstrual cycle, participants were asked about the 12 typical symptoms of premenstrual syndrome (PMS): seclusion, irritability, swelling, anger, weight gain, joint pain, headache, confusion, dejection, bloating, anxiety, and tenderness [42]; participants were also asked about the severity of their bleeding. If a new menstrual cycle had begun, the first day of bleeding triggered the new cadence for all future questionnaires. Data from these daily questionnaires were used for correlation with the wearable data collected simultaneously.

In addition to the digital questionnaires, the Fitbit device recorded 3 parameters daily: average resting heart rate, total steps, and total sleep time. For this study, average resting heart rate was rounded to the nearest 5 beats per minute, sleep time was rounded to the nearest 10 minutes, and step count was rounded to the nearest 100 steps. This generalization was made to protect the anonymity of participants when analyzing deidentified donation records in this small feasibility study.

Definitions

The relevant phases of the menstrual cycle were defined as follows: the menses phase (MP; days 1 to 5), the ovulation day (OD; day 1 of the next MP minus 14) [43], the follicular phase (day 1 of MP to OD minus 1), the luteal phase (OD plus 1 to the menses), and the midluteal phase (OD plus 3 to OD plus 9).

In our analysis of PMS, only PMS symptoms from the days after OD to the end of MP were considered. Severe PMS was defined as a symptom severity rating of 3, equivalent to severe, on the visual analog scale. Participants were included in the severe PMS group if they reported severe PMS symptoms in at least 2 cycles. Participants were included in the no or low PMS group if they never reported PMS symptom severity higher than 1 on a severity scale from 0 to 3.

Data Storage and Wearable Data Integration

The nonprofit organization Data4Life provided the research infrastructure (study application, data storage, and analysis platform) for this study. All research data was stored in Data4Life's secure research environment on servers in Germany. Data4Life is certified by the German Federal Office for Information Security (BSI). For participants who allowed the use of cookies, we were able to analyze user behavior using the General Data Protection Regulation–compliant business analytics tool Matomo (Matomo; data stored in Europe only). In the web app, after entering their study token and providing digital study consent, participants were asked to connect to Fitbit to allow access to their wearable data via the Fitbit web application programming interface (API) [44]. After completing the OAuth 2.0 Authorization Code Grant Flow [45], the API refresh token was stored in end-to-end encrypted form for subsequent use for the duration of their participation in the study or until they withdrew their consent. This allowed the study's wearable parameters to be retrieved client-side via the respective activity and sleep endpoints, and stored in end-to-end encrypted form as Fast Healthcare Interoperability Resources Standard for Trial Use, version 3, [46] observation records each time participants logged into the web app to complete their questionnaires. As the API returns time series data over a few days, data gaps between donations were filled by comparing previously stored records with the data points returned by the API.

Statistical Analysis

All analyses were performed on the Data4Life analytics platform on a jupyterhub notebook running Python (version 3.10.4), using *pandas* (version 1.4.2), *matplotlib* (version 3.5.2), *seaborn* (version 0.11.2), and *numpy* (version 1.22.4). Descriptive statistics were used to report details of the study cohort, retention, and adherence measures. Continuous values were reported as mean with SD or median with IQR; categorical values were reported as numbers with percentages. For comparative analyses of clinical outcomes, we used the Mann-Whitney test [47] for nonnormally distributed data and the 2-tailed *t* test [48] for normally distributed data.

Ethical Considerations

The study was approved by the ethics committee of the Berlin Chamber of Physicians (Eth-11/22). Registration was open for people aged 18 years and older. All participants provided digital informed consent for study participation. Participation was voluntary. Participants were allowed to keep the Fitbit Inspire 2 device after the study ended. Email and password were required to log in to the study app as well as two-factor authentication via phone. Participants could access all study content through a web app on both desktop and mobile devices. All study data was end-to-end encrypted and pseudonymized. Only authorized researchers were provided access to the data on the Data4Life analytics platform. Data was stored exclusively on Data4Life data centers in Germany. Based on IT-Grundschutz

(ISO 27001), Data4Life is certified by the German Federal Office for Information Security (BSI).

Results

Study Setting and Participants

This feasibility study of combined wearable and questionnaire data collection on the menstrual cycle was conducted online with self-reporting of questionnaire data in a study app accompanied by the collection of wrist-worn sensor data from Fitbit Inspire 2 devices. A total of 42 cisgender women were recruited from December 2021 to April 2022 at a gynecological practice in Berlin and agreed to provide menstrual cycle data for 12 weeks (Figure 1).

One participant never requested a wearable device and therefore dropped out of the study before creating an account in the study app. Three participants never donated any questionnaire or wearable data (Figure S1 in Multimedia Appendix 2), leaving data from 38 participants who started to donate data in the study app. Of these, a total of 5 participants withdrew from the study within the first 2 weeks of study entry, 3 of whom never donated any wearable data, leaving valuable data from 33 participants for further analysis. Active participants (n=33) were on average 24.2 (SD 3.0) years old and predominantly identified as female, with 1 individual identifying as diverse (Table 1). Five participants (15.1%) reported preexisting mental health problems (Table 1). In terms of women's health issues, 2 (6.1%) participants reported a history of chlamydia infection and 1 person (3%) reported a history of human papillomavirus infection. At enrollment, the median time since the last menstruation was 17 (IQR 12-26) days.

Figure 1. Feasibility Study of Menstrual Cycles With Fitbit Device (FEMFIT) study concept. The FEMFIT study on collecting combined wearable and questionnaire data on the menstrual cycle was carried out online. Self-reporting of questionnaire data in a study app was accompanied by the collection of wrist-worn sensor data from Fitbit Inspire 2 devices. Weekly questionnaires asked about physical and mental well-being. From day 13 to day 5 of the menstrual cycle, participants answered daily questionnaires on bleeding severity and PMS symptoms. PMS: premenstrual syndrome.

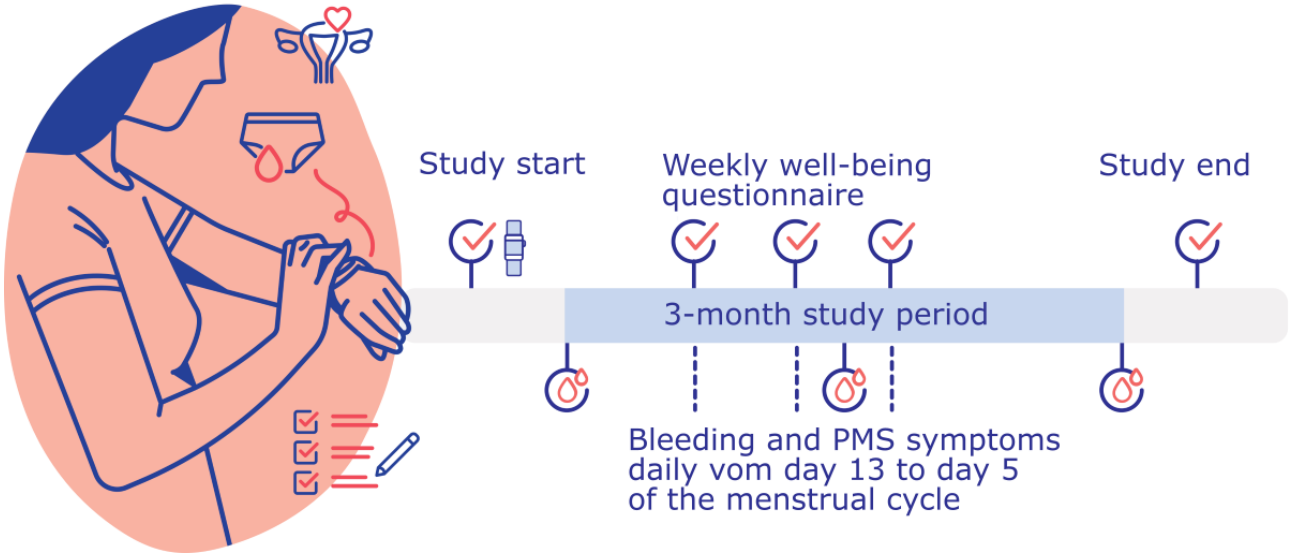


Table 1. Details on study participants (n=33).

Demographics of active study participants	Values
Age (years), mean (SD)	24.2 (3.0)
Gender, n (%)	
Female	32 (97)
Diverse	1 (3)
Preexisting health condition, n (%)	
Chlamydia	2 (6.1)
HPV[a]	1 (3)
Mental health	5 (15.1)
Hypothyroidism	1 (3)
Hyperthyroidism	1 (3)
Cycle regularity, n (%)	
Irregular	2 (6.1)
Hormonal contraception, n (%)	
Yes	4 (12.1)
Days since last period, median (IQR)	17 (12-26)

[a]HPV: human papillomavirus.

High Degree of Digital Literacy of Study Participants

Of the 42 people initially recruited, 25 (59.5%) people reported that they had never worn a wearable device before. Of those who had experience wearing a sensor device, 5 had worn an Apple Watch, and other devices worn included Fitbit (n=2), Garmin (n=1), Samsung (n=1), and Denver (n=1). The majority of participants (26/42, 61.9%) regularly used apps on their smartphones to track their food intake, physical activity, or menstrual cycle. The most commonly used menstrual cycle tracking apps were Flo (9/26, 34.6%) and Clue (4/26, 15.3%). At the start of the study, a standardized digital literacy questionnaire [41] showed that the participants were highly digitally literate with an average score of 37.1 (SD 3.1) out of 48.

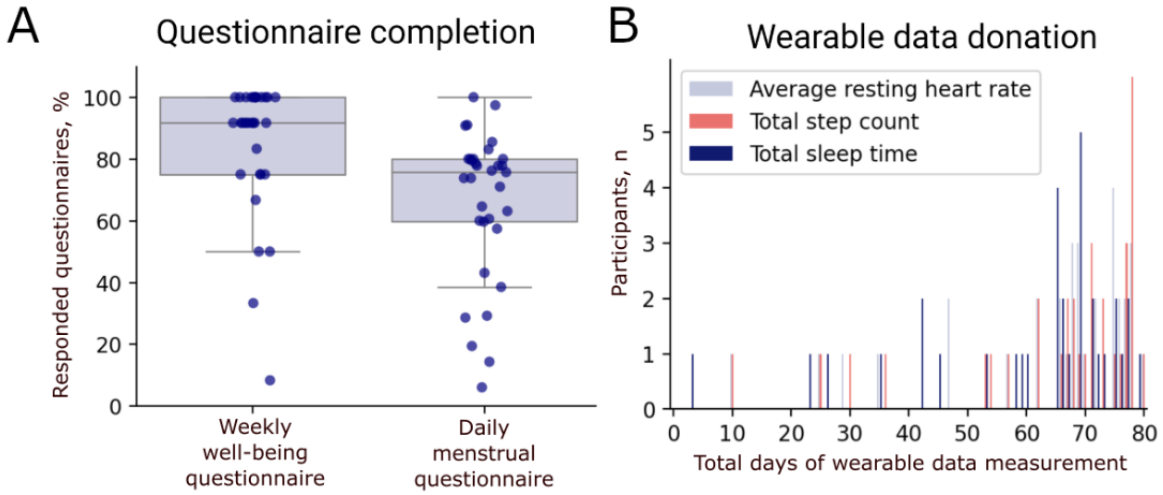
Increased User Engagement After Early Dropouts

Across all 41 participants who enrolled in the study and logged into the study app, the mean time spent in the study (measured from enrollment to last weekly or daily questionnaire delivery) was 63.3 (SD 33.0) days The reported reasons for not completing the study were too great a time commitment (n=1), a high technical barrier (n=2), and technical difficulties that made it impossible to complete the study (n=3).

Among participants who remained in the study for more than 14 days (n=33), questionnaire completion rates were high. The weekly physical and mental well-being questionnaires were completed on average 10.2 times (SD 2.6) out of 12, while 69.7% (n=23) of active participants completed at least 11 of the 12 weekly questionnaires. An average of 65.8% (SD 24.1%) of all questionnaires on menstrual bleeding and PMS symptoms, which were asked daily from day 13 to day 5 of the menstrual cycle, were completed (Figure 2A). Overall, 66.7% (n=22) of participants reported starting dates of at least 3 menstrual cycles, with a median cycle length of 28.0 days (IQR 24-31), and we collected data from a total of 50 complete ovulatory cycles. In general, logins to our study app occurred mostly after email notifications (Figure S3 in Multimedia Appendix 2), suggesting that frequent study reminders may increase retention.

The average number of days wearable data were provided was 65.2 (SD 17.0) days for step count, 64.2 (SD 17.1) days for heart rate, and 60.5 (SD 17.8) days for sleep time (Figure 2B). There were difficulties in collecting continuous wearable data over the entire study period, particularly for sleep time. None of the 33 active participants provided consecutive wearable data over the entire study period. The average number of consecutive days of data donation was 36.4 (SD 25.6) days for sleep time, 38.6 (SD 31.4) days for heart rate, and 52.6 (SD 28.3) days for step count. Wearable data was provided consecutively for at least 50 days by 11 participants and for more than 70 days by 3 participants. In the weekly study questionnaire, participants reported that they were wearing the device continuously for an average of 9 (SD 2.9) weeks out of 12. When participants reported that they did not wear the device continuously, off-times were stated to be longer than 3 hours per day.

Figure 2. Participant engagement. (A) Questionnaire completion rates of weekly and daily questionnaires. Questionnaires on physical and mental well-being were filled out once per week, whereas questions on premenstrual symptoms and bleeding severity were filled out daily from day 13 to day 5 of the menstrual cycle. The box plots show the median and IQR, with whiskers reflecting minimum and maximum values for outliers. Dots represent all data points. (B) Bar plot visualizing the total days of wearable data donation during the study period per participant. Bars are color-coded by the parameter assessed using the Fitbit Inspire 2 wearable device.



Fitbit Device and Study App Supported Subjective Health Monitoring

At the end of the study, 22 participants completed questions about their study experience. Most participants reported that the Fitbit device was very easy to use (n=22; mean rating 1.6, SD 0.6; 1=very easy, 5=very challenging) and did not interfere with their daily life (n=22; mean rating 1.2, SD 2; 1=very comfortable, 5=very disturbing). One person found wearing the Fitbit device at night very uncomfortable, whereas overall it was not perceived to be bothersome at night (n=22; mean rating 2.2, SD 1).

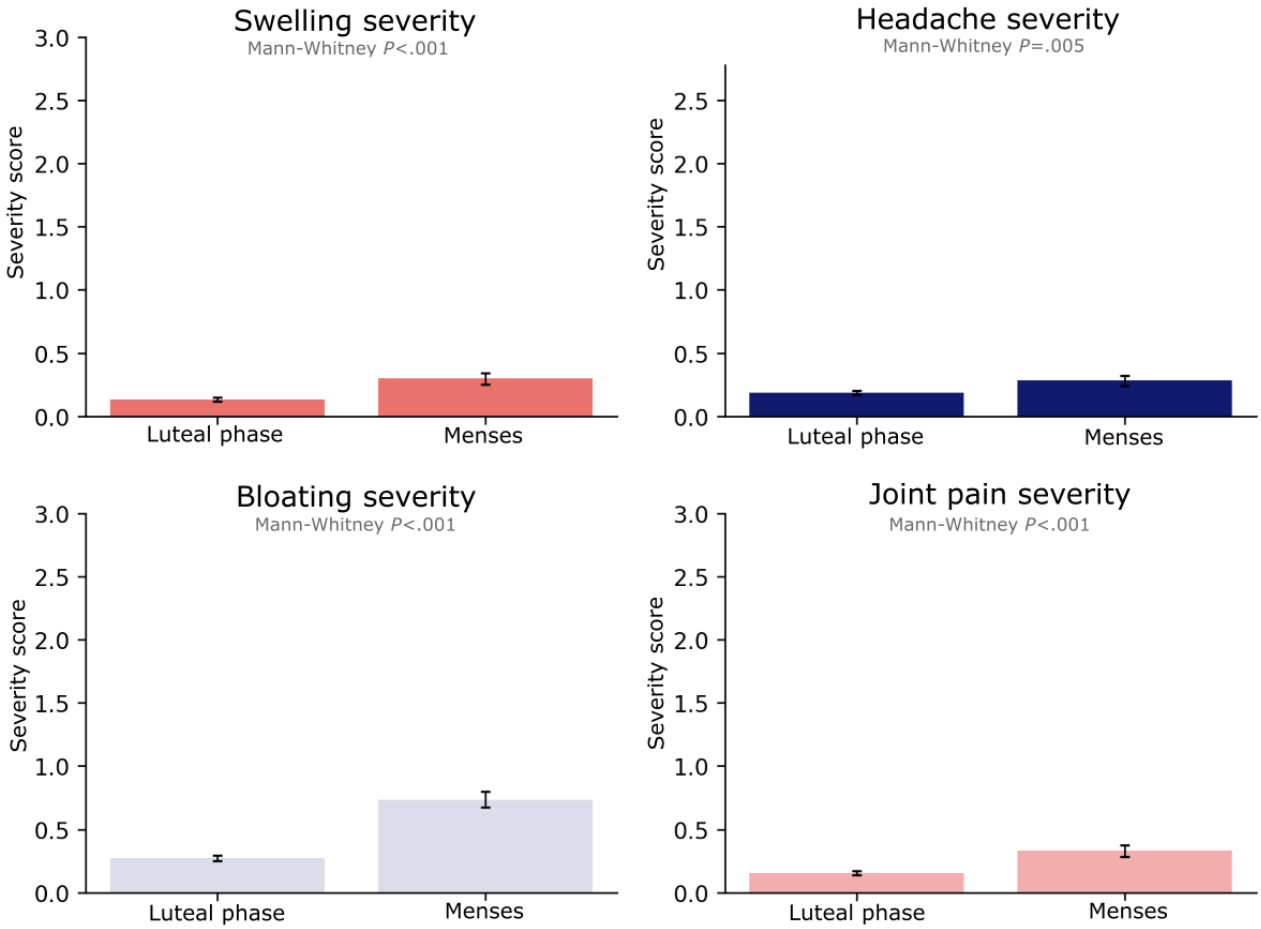
The majority of participants completing the end-of-study questionnaire (13/22, 59%) stated that wearing the device increased their physical activity over the study period. Step count data from all participants over the study period thereby showed an average step count of 9987 (SD 3856) steps per day in the last 2 weeks of study participation compared with an average of 9301 (SD 4076) steps per day in the first 2 weeks of study participation, but this observation was not significant. Overall, participants did not feel that the sensor tracking changed their sleep behavior (13/22, 59%). Five participants reported that wearing the device and answering the questionnaires made them feel stressed, while 9 participants reported that wearing the Fitbit device made them feel more in control. For 15 participants (67.9%), the study helped them to better associate symptoms such as mood swings, headaches, and pelvic pain with a particular phase of their menstrual cycle. Overall, the study seemed to improve participants' health awareness, and 21 of the 22 participants who completed the final questionnaire said they would take part in a similar study again.

Digital Symptom Questionnaires Revealed Patterns of PMS Symptoms Across the Menstrual Cycle

PMS-related symptoms and their severity on a scale of 0 to 3 were only assessed on potential PMS days during the luteal phase and menses. The majority of participants (22/33, 66.7%) reported severe PMS symptoms on at least 1 day of the menstrual cycle for at least 2 cycles. Severe PMS was reported by 18.2% (n=6) of participants in at most one of the cycles recorded in our study, and only 15.2% (n=5) did not experience any severe PMS symptoms during the study period.

In terms of mental health, dejection (average score 0.5, SD 0.9) and seclusion (average score 0.3, SD 0.7) were among the most commonly reported symptoms during the menses. The data suggested a notable impact on participants' mental health during the menses, with higher severity of seclusion (Mann-Whitney P=.03), dejection (Mann-Whitney P=.02), and irritability (Mann-Whitney P=.01) reported during the menses compared with the luteal phase (Figure S4 in Multimedia Appendix 2). Reported levels of anxiety, confusion, anger, and tenderness did not differ substantially between the luteal phase and menses (Figure S4 in Multimedia Appendix 2). In contrast to mental health symptoms, reported physical symptoms during menses were more severe. Participants experienced a significantly higher severity of bloating during the menses (mean 0.7, SD 1.1) compared with the luteal phase (mean 0.3, SD 0.7; Mann-Whitney P<.001; Figure 3). Similarly, the swelling was reported to be more severe during the menses (mean 0.3, SD 0.8; luteal phase: mean 0.1, SD 0.5; Mann-Whitney P<.001). In addition, participants reported more severe headaches during the menses compared with the luteal phase (mean 0.38, SD 0.7; luteal phase: mean 0.2, SD 0.67; Mann-Whitney P=.005) and more bothersome joint pain during the menses (mean 0.3, SD 0.8; luteal phase: mean 0.2, SD 0.5; Mann-Whitney P<.001). In terms of reported weight gain, there was a trend toward increased weight gain during the menses, although this increase did not reach statistical significance (Figure S4 in Multimedia Appendix 2).

Figure 3. Symptom severity over the menstrual cycle. Bar plots show the mean rating for symptom severity on a scale from 0 to 3 (0=no symptoms to 3=very strong symptom severity). Error bars are the SE. Mann-Whitney *P* values comparing groups are shown for each plot.



Combined Questionnaire and Wearable Data Delivered Interesting Insights Into the Menstrual Cycle

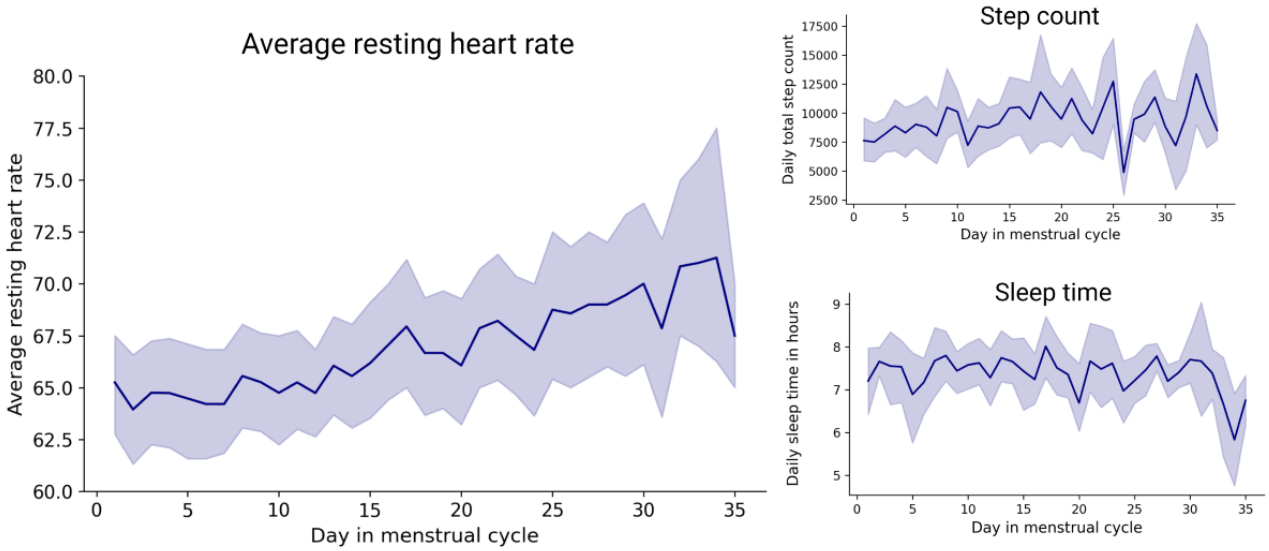
Participants provided a total of 23,424 symptom reports, with a total of 2468 symptoms reported in the daily questionnaires, which were administered during the luteal phase and menses. Heart rate sensor data was collected on a total of 2118 days, step count on 2153 days, and sleep time on 1996 days.

The average resting heart rate over the entire study period was 66.7 (SD 6.7). Participants slept an average of 7.3 (SD 1.6) hours while wearing the Fitbit device. With an average of 10,184 (SD 6120) steps per day, the FEMFIT cohort was quite physically active, with some participants walking >20,000 steps per day (153 days from 21 participants recorded with >20,000 steps; Figure S5 in Multimedia Appendix 2). When examining correlations between sensor data and questionnaire data on the menstrual cycle, we observed that mean resting heart rate appeared to be highest in the mid and late luteal phases relative to the time between menses and ovulation (Mann-Whitney $P<.001$; Figure 4, Figure S6 in Multimedia Appendix 2). We did not observe any significant differences in the change in step count (Mann-Whitney $P=.72$) or sleep time (Mann-Whitney $P=.58$) between the follicular and luteal phases (Figure 4).

In an exploratory analysis of the wearable data across participants with different PMS severity, participants with severe PMS (n=22) walked significantly fewer steps (mean daily steps10,283, SD 6277) during the luteal phase and menses compared with participants with no or low PMS (n=5; mean 11,694, SD 6458; Mann-Whitney $P<.001$). There was no significant difference in sleep duration (Mann-Whitney $P=.18$) or heart rate with PMS severity (Mann-Whitney $P=.72$).

To investigate whether hormonal contraception had an effect on step count, sleep time, or heart rate, we compared 4 participants using hormonal contraception with 29 participants not using hormonal contraception. Participants on hormonal contraception appeared to be less active (mean step count 7400, SD 4433; no hormonal contraception: mean 9800, SD 6203; Mann-Whitney $P<.001$). The average daily resting heart rate was significantly higher in the hormonal contraceptive group (mean 68.2, SD 7.7 vs mean 66.6, SD 6.6; Mann-Whitney $P<.001$). There was no difference in sleep behavior between the 2 groups.

Figure 4. Physical parameters measured by wearable devices over the menstrual cycle. Progression of daily average resting heart rate, step count, and sleep time over the menstrual cycle across all active study participants. Only cycles with a maximum length of 35 days were included. The dark lines show the mean and the shadowed lines show the 95% CI. The average resting heart rate was highest in the mid and late luteal phase relative to the time from menses to ovulation (Mann-Whitney $P<.001$); no significant difference was seen for sleep time or step count.



Discussion

Principal Findings

To date, only a few studies have assessed the menstrual cycle using both objective wearable and subjective questionnaire data [23,27,29,49,50]. Previous research has demonstrated potential effects of physical activity on the menstrual cycle and vice versa [30-32]. However, none of these studies used wearable devices to monitor daily activity levels, and there is a gap in studies using wearable data to examine sleep patterns across the menstrual cycle. Recognizing that participant retention is a significant challenge in digital cohort studies [19,40], our feasibility study was designed to primarily assess the effectiveness of study participation and adherence to the study protocol, highlighting challenges and opportunities for future large-scale studies in women's health. Secondarily, we focused on potential biological outcomes.

Our cohort was a group of young, mostly digitally literate, cisgender women with regular cycle lengths. Overall, retention was good among participants who remained in the study beyond the second week. The observed dropout rate of 21.4% (n=9) within the first 2 weeks of the study is consistent with rates reported in other digital cohort studies [51,52], and the overall median retention of 81.0 days was remarkably high [52]. Nevertheless, these dropouts may have introduced selection bias. Among the 5 dropouts who provided demographic data, we did not identify any discernible demographic patterns in relation to retention. The mean age of the dropouts was 24.8 (SD 3.2) years, which was not significantly different from the rest of the cohort. None of the dropouts had preexisting medical conditions and only 1 reported menstrual irregularity. In future large-scale studies, it would be advisable to collect demographic information directly at recruitment (in our case when the wearable device was distributed) to ensure comprehensive data on dropouts. This may contribute to a better understanding of the potential retention patterns and help to address selection bias. Among those who dropped out early, technical barriers were the most commonly reported problem. Although our study team provided access to technical support, future studies could potentially reduce dropout rates by further reducing barriers to accessing technical support or by conducting active outreach. The young age of participants, such as our cohort, is also known to be associated with lower retention [19,52]. Targeted engagement strategies such as increased outreach and communication with younger participants could counteract this [19]. For large-scale studies, continuous monitoring of data donations in an ongoing study can facilitate real-time strategies such as increased outreach to engage inactive participants, reduce dropouts, and ensure data completeness [19]. In 2 similar studies using sensor devices to investigate physiology surrounding the menstrual cycle [23,25], retention was generally not an issue, but retention efforts were not further highlighted in the publications.

In addition to the potential bias introduced by study dropouts, ensuring high data completeness is critical for investigating temporal patterns in menstrual cycles. Similar to our study, Maijala et al [23] reported issues with continuous data contribution across consecutive days. In particular, issues related to participant retention and data completeness could potentially undermine the integrity of the results. In our study, this was particularly challenging for sleep data. To mitigate these concerns, several strategies can be considered for future research. The comfort of wearable devices should be a priority in study design, as some participants in our study found the devices too uncomfortable. Data completeness may be further increased if a clear purpose for data collection is visible to participants, for example, having the clinical condition of interest in the study [52,53]. As this was a feasibility study without a specific disease focus, we expect retention and data completeness to be higher when applying a similar study setting

to a clinical condition with a high disease burden, such as endometriosis [54]. Recruitment strategies should also be tailored to improve data completeness. Although participants in our study were highly digitally literate, the vast majority had never worn a wearable device before. Recruiting participants who already have a sensor device and are accustomed to wearing it on a daily basis may increase data completeness. However, such targeted recruitment would introduce additional bias, as sensor devices are mostly owned by individuals from high-income households [55]. Although some participants in our study felt stressed by wearing the sensor devices, positive prior experience with wearing sensor devices as a selection criterion for study participation may further increase data completeness. In our study, participants received the wearable device as a free gift at the start of the study, and the incentive was not dependent on their questionnaire completion rate or the duration of wearing the device. Future studies may benefit from offering incentives upon successful completion of the entire study, thereby encouraging full participation [52,56]. Increasing compensation beyond just the Fitbit device could further improve retention [52]. In terms of questionnaire data completeness, it may be beneficial to reduce the number of daily questions. In our study, completion rates for daily questionnaires were lower than for weekly questionnaires, and 1 of the 9 people who dropped out within the first 2 weeks reported that the study was too time-consuming. These strategies can potentially help to improve participant retention and ensure more comprehensive and reliable data.

Despite challenges related to data completeness, our study successfully demonstrated the technical feasibility and value of combining wearable and questionnaire data. The information derived from the digital questionnaires helped to map the sensor data to the respective day of the menstrual cycle and to correlate the objective physiological data with the subjectively reported symptom data from the study app, leading to interesting biological insights. The average resting heart rate measured by wrist-worn wearable sensors changed over the course of the menstrual cycle, with an increase in the mid and late luteal phases compared with the menses. This observation supports the findings of a larger study by Shilaih et al [25], where the pulse rate measured using laboratory standard photoplethysmography technology showed a peak in the midluteal phase. Another study with a smaller sample size similarly found increased heart rate (measured during sleep) in the mid and late luteal phases compared with menses and ovulation [24]. In our feasibility study, we also found that people with more severe PMS symptoms had lower daily step counts. Due to the observational nature of our study, we cannot establish causality. This finding may suggest that individuals with more severe PMS symptoms may exercise less due to the impact of the symptoms on their physical activity. Alternatively, this finding may be consistent with the existing literature, where increased physical activity has been found to reduce PMS symptoms [57-62]. In attempting to fill the research gap regarding sleep patterns across the menstrual cycle, we did not make any notable discoveries when comparing sleep duration between different phases of the menstrual cycle, across PMS severity, or when comparing individuals using hormonal contraception with those who did not. It is worth noting that there were more issues with data completeness for sleep data compared with daytime data, highlighting the need for improved retention strategies to motivate participants to wear the devices consistently at night. Although of very small sample size, the observed differences in heart rate and physical activity in participants using hormonal contraception (n=4) also provided an interesting starting point for future investigation. These observations show that sensor devices may offer a great opportunity to study physical changes over the menstrual cycle and suggest that further investigation on a larger scale may be valuable. Overall, it is important to interpret the biological findings of this feasibility study with caution, given the limitations of the small sample size and issues with data completeness.

Limitations

Apart from the interesting results of this feasibility study, this study has several limitations that should be taken into account when interpreting the findings.

First, the study was conducted as a feasibility study with a small sample size. Any biological interpretations made from such a small sample size must be taken with caution and merely represent an interesting starting point for future large-scale research. Our cohort was also not representative, recruited from only 1 recruitment site, and lacked ethnic diversity.

We further had limitations in data accuracy. The amount of wearable data we collected was limited to 3 aggregated parameters, which severely limited our ability for quality filtering. As we did not collect continuous wearable data throughout the day, but rather one aggregated data point for each measurement of interest, we had no objective insight into how long the device was worn each day and relied only on subjective reporting of wear time in the weekly questionnaires. Not wearing the device consistently or correctly could have resulted in inaccurate data.

Furthermore, the study relied on self-reported questionnaire data, which may introduce reporting bias and errors in data collection. In addition, the wrist-worn wearable device used in our study may not be as accurate as research-grade devices in tracking changes in physiological parameters related to the menstrual cycle. One must consider potential confounding factors that may affect the measurement accuracy of the sensor devices such as body sweat and unusual movements [63].

Regarding the medical accuracy of possible correlations between wearable data and specific phases of the menstrual cycle, the design of our study only allowed the day of ovulation to be estimated based on the total cycle length. In future studies, additional at-home ovulation tests could ensure greater accuracy in assessing the different phases of the menstrual cycle [25].

Conclusions

In conclusion, this study demonstrated that an app-based approach to collecting combined wearable and questionnaire data on the menstrual cycle is technically feasible and provides interesting biological insights. Study participants had a high level of digital literacy, which may have supported adherence to the study protocol. With 9 of 42 participants dropping out

within the first 2 weeks, engagement efforts at baseline would need to be improved in future, larger studies to ensure long-term adherence to the study protocol. To further improve the quality and applicability of such large-scale studies, challenges with data completeness need to be addressed. This could include using more frequent reminders for study participation, using more user-friendly wearable technology, or providing incentives to encourage participants to maintain consistent data contributions over the entire study. Although the sample size was small, we discovered an increase in average resting heart rate over the menstrual cycle and found differences in step count by PMS symptom severity in our cohort. Ultimately, using digital studies to research the menstrual cycle can be feasible and has the potential to improve our understanding of women's health and inform the development of personalized health care approaches.

Acknowledgments

The authors want to thank all feasibility study of menstrual cycles with Fitbit device (FEMFIT) participants for their contribution to women's health research. The authors further want to acknowledge the work and dedication of all Data4Life employees working on this project. The authors especially want to thank the business intelligence (BI) and analytics team for helping with all inquiries concerning BI data and the development team for the technical implementation of the study. This study was fully funded by the Data4Life gGmbH.

Data Availability

The data sets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

MF and CR initiated the study and wrote the study protocol. ALL conducted all data analysis on wearable and questionnaire data. IK supported data analysis. JR set up wearable data collection from Fitbit devices. RLB and MH recruited participants and handed out pre- and poststudy questionnaires. RLB evaluated the questionnaire responses to surveys conducted before and after the study. ALL wrote the first version of the manuscript. All authors have contributed to writing the manuscript.

Conflicts of Interest

ALL, IK, JR, CR, and MF worked for Data4Life, the company that developed the study and study app as well as the underlying research infrastructure. RLB declares no competing interests. MH received financial incentives from Data4Life for recruiting participants for this study.

Multimedia Appendix 1
Study questionnaires.
[XLSX File (Microsoft Excel File), 22 KB - mhealth_v12i1e50135_app1.xlsx]

Multimedia Appendix 2
Supplementary figures.
[DOCX File , 618 KB - mhealth_v12i1e50135_app2.docx]

References

1. Santos LBD, Barbosa IR, de Macedo Dantas TH, Araujo CM, Dantas JH, Ferreira CWS, et al. Prevalence of primary dysmenorrhea and associated factors in adult women. Rev Assoc Med Bras (1992) 2022;68(1):31-36 [FREE Full text] [doi: 10.1590/1806-9282.20210341] [Medline: 35239934]
2. Hadjou OK, Jouannin A, Lavoue V, Leveque J, Esvan M, Bidet M. Prevalence of dysmenorrhea in adolescents in France: results of a large cross-sectional study. J Gynecol Obstet Hum Reprod 2022;51(3):102302 [FREE Full text] [doi: 10.1016/j.jogoh.2021.102302] [Medline: 34973478]
3. Bakhsh H, Algenaimi E, Aldhuwayhi R, AboWadaan M. Prevalence of dysmenorrhea among reproductive age group in Saudi women. BMC Womens Health 2022;22(1):78 [FREE Full text] [doi: 10.1186/s12905-022-01654-9] [Medline: 35305636]
4. Karout S, Soubra L, Rahme D, Karout L, Khojah HMJ, Itani R. Prevalence, risk factors, and management practices of primary dysmenorrhea among young females. BMC Womens Health 2021;21(1):392 [FREE Full text] [doi: 10.1186/s12905-021-01532-w] [Medline: 34749716]
5. Durand H, Monahan K, McGuire BE. Prevalence and impact of dysmenorrhea among university students in Ireland. Pain Med 2021;22(12):2835-2845 [FREE Full text] [doi: 10.1093/pm/pnab122] [Medline: 33822197]
6. Ju H, Jones M, Mishra G. The prevalence and risk factors of dysmenorrhea. Epidemiol Rev 2014;36:104-113 [FREE Full text] [doi: 10.1093/epirev/mxt009] [Medline: 24284871]

7. Schoep ME, Nieboer TE, van der Zanden M, Braat DDM, Nap AW. The impact of menstrual symptoms on everyday life: a survey among 42,879 women. Am J Obstet Gynecol 2019;220(6):569.e1-569.e7. [doi: 10.1016/j.ajog.2019.02.048] [Medline: 30885768]
8. Pogodina A, Dolgikh O, Astakhova T, Klimkina J, Khramova E, Rychkova L. Health-related quality of life and menstrual problems in adolescents. J Paediatr Child Health 2022;58(6):1028-1032. [doi: 10.1111/jpc.15895] [Medline: 35141982]
9. Sharma A, Taneja DK, Sharma P, Saha R. Problems related to menstruation and their effect on daily routine of students of a medical college in Delhi, India. Asia Pac J Public Health 2008;20(3):234-241. [doi: 10.1177/1010539508316939] [Medline: 19124317]
10. Ballagh SA, Heyl A. Communicating with women about menstrual cycle symptoms. J Reprod Med 2008;53(11):837-846. [Medline: 19097516]
11. Barnard K, Frayne SM, Skinner KM, Sullivan LM. Health status among women with menstrual symptoms. J Womens Health (Larchmt) 2003;12(9):911-919. [doi: 10.1089/154099903770948140] [Medline: 14670171]
12. Parasar P, Ozcan P, Terry KL. Endometriosis: epidemiology, diagnosis and clinical management. Curr Obstet Gynecol Rep 2017;6(1):34-41 [FREE Full text] [doi: 10.1007/s13669-017-0187-1] [Medline: 29276652]
13. Sinaii N, Plumb K, Cotton L, Lambert A, Kennedy S, Zondervan K, et al. Differences in characteristics among 1,000 women with endometriosis based on extent of disease. Fertil Steril 2008;89(3):538-545 [FREE Full text] [doi: 10.1016/j.fertnstert.2007.03.069] [Medline: 17498711]
14. Ghai V, Jan H, Shakir F, Haines P, Kent A. Diagnostic delay for superficial and deep endometriosis in the United Kingdom. J Obstet Gynaecol 2020;40(1):83-89. [doi: 10.1080/01443615.2019.1603217] [Medline: 31328629]
15. Nnoaham KE, Hummelshoj L, Webster P, d'Hooghe T, de Cicco Nardone F, de Cicco Nardone C, et al. Impact of endometriosis on quality of life and work productivity: a multicenter study across ten countries. Fertil Steril 2011;96(2):366-373.e8 [FREE Full text] [doi: 10.1016/j.fertnstert.2011.05.090] [Medline: 21718982]
16. Hennegan J, Winkler IT, Bobel C, Keiser D, Hampton J, Larsson G, et al. Menstrual health: a definition for policy, practice, and research. Sex Reprod Health Matters 2021;29(1):1911618 [FREE Full text] [doi: 10.1080/26410397.2021.1911618] [Medline: 33910492]
17. Izmailova ES, Wagner JA, Perakslis ED. Wearable devices in clinical trials: hype and hypothesis. Clin Pharmacol Ther 2018;104(1):42-52 [FREE Full text] [doi: 10.1002/cpt.966] [Medline: 29205294]
18. Snyder M, Zhou W. Big data and health. Lancet Digit Health 2019;1(6):e252-e254 [FREE Full text] [doi: 10.1016/S2589-7500(19)30109-8] [Medline: 33323249]
19. Zhang Y, Pratap A, Folarin AA, Sun S, Cummins N, Matcham F, et al. Long-term participant retention and engagement patterns in an app and wearable-based multinational remote digital depression study. NPJ Digit Med 2023;6(1):25 [FREE Full text] [doi: 10.1038/s41746-023-00749-3] [Medline: 36806317]
20. Li K, Urteaga I, Wiggins CH, Druet A, Shea A, Vitzthum VJ, et al. Characterizing physiological and symptomatic variation in menstrual cycles using self-tracked mobile-health data. NPJ Digit Med 2020;3:79 [FREE Full text] [doi: 10.1038/s41746-020-0269-8] [Medline: 32509976]
21. Wheeler MV, Tabor VH, Teel K, LaVigne M. The science of your cycle: evidence-based app design. Clue. 2015. URL: https://helloclue.com/articles/about-clue/science-your-cycle-evidence-based-app-design [accessed 2023-02-08]
22. Hantsoo L, Rangaswamy S, Voegtline K, Salimgaraev R, Zhaunova L, Payne JL. Premenstrual symptoms across the lifespan in an international sample: data from a mobile application. Arch Womens Ment Health 2022;25(5):903-910 [FREE Full text] [doi: 10.1007/s00737-022-01261-5] [Medline: 36018464]
23. Maijala A, Kinnunen H, Koskimäki H, Jämsä T, Kangas M. Nocturnal finger skin temperature in menstrual cycle tracking: ambulatory pilot study using a wearable Oura ring. BMC Womens Health 2019;19(1):150 [FREE Full text] [doi: 10.1186/s12905-019-0844-9] [Medline: 31783840]
24. Alzueta E, de Zambotti M, Javitz H, Dulai T, Albinni B, Simon KC, et al. Tracking sleep, temperature, heart rate, and daily symptoms across the menstrual cycle with the Oura ring in healthy women. Int J Womens Health 2022;14:491-503 [FREE Full text] [doi: 10.2147/IJWH.S341917] [Medline: 35422659]
25. Shilaih M, Clerck VD, Falco L, Kübler F, Leeners B. Pulse rate measurement during sleep using wearable sensors, and its correlation with the menstrual cycle phases, a prospective observational study. Sci Rep 2017;7(1):1294 [FREE Full text] [doi: 10.1038/s41598-017-01433-9] [Medline: 28465583]
26. Goodale BM, Shilaih M, Falco L, Dammeier F, Hamvas G, Leeners B. Wearable sensors reveal menses-driven changes in physiology and enable prediction of the fertile window: observational study. J Med Internet Res 2019;21(4):e13404 [FREE Full text] [doi: 10.2196/13404] [Medline: 30998226]
27. Mahalingaiah S, Fruh V, Rodriguez E, Konanki SC, Onnela JP, de Figueiredo Veiga A, et al. Design and methods of the Apple Women's Health Study: a digital longitudinal cohort study. Am J Obstet Gynecol 2022;226(4):545.e1-545.e29 [FREE Full text] [doi: 10.1016/j.ajog.2021.09.041] [Medline: 34610322]
28. Piwek L, Ellis DA, Andrews S, Joinson A. The rise of consumer health wearables: promises and barriers. PLoS Med 2016;13(2):e1001953 [FREE Full text] [doi: 10.1371/journal.pmed.1001953] [Medline: 26836780]

29. Zhu TY, Rothenbühler M, Hamvas G, Hofmann A, Welter J, Kahr M, et al. The accuracy of wrist skin temperature in detecting ovulation compared to basal body temperature: prospective comparative diagnostic accuracy study. J Med Internet Res 2021;23(6):e20710 [FREE Full text] [doi: 10.2196/20710] [Medline: 34100763]
30. Pearce E, Jolly K, Jones LL, Matthewman G, Zanganeh M, Daley A. Exercise for premenstrual syndrome: a systematic review and meta-analysis of randomised controlled trials. BJGP Open 2020;4(3):bjgpopen20X101032 [FREE Full text] [doi: 10.3399/bjgpopen20X101032] [Medline: 32522750]
31. Dehnavi ZM, Jafarnejad F, Goghary SS. The effect of 8 weeks aerobic exercise on severity of physical symptoms of premenstrual syndrome: a clinical trial study. BMC Womens Health 2018;18(1):80 [FREE Full text] [doi: 10.1186/s12905-018-0565-5] [Medline: 29855308]
32. Vaghela N, Mishra D, Sheth M, Dani VB. To compare the effects of aerobic exercise and yoga on premenstrual syndrome. J Educ Health Promot 2019;8:199 [FREE Full text] [doi: 10.4103/jehp.jehp_50_19] [Medline: 31867375]
33. Romans SE, Kreindler D, Einstein G, Laredo S, Petrovic MJ, Stanley J. Sleep quality and the menstrual cycle. Sleep Med 2015;16(4):489-495. [doi: 10.1016/j.sleep.2014.12.001] [Medline: 25747332]
34. Baker FC, Driver HS. Self-reported sleep across the menstrual cycle in young, healthy women. J Psychosom Res 2004;56(2):239-243. [doi: 10.1016/S0022-3999(03)00067-9] [Medline: 15016584]
35. Van Reen E, Kiesner J. Individual differences in self-reported difficulty sleeping across the menstrual cycle. Arch Womens Ment Health 2016;19(4):599-608. [doi: 10.1007/s00737-016-0621-9] [Medline: 26973332]
36. Hachul H, Andersen ML, Bittencourt LRA, Santos-Silva R, Conway SG, Tufik S. Does the reproductive cycle influence sleep patterns in women with sleep complaints? Climacteric 2010;13(6):594-603. [doi: 10.3109/13697130903450147] [Medline: 20001564]
37. Li DX, Romans S, De Souza MJ, Murray B, Einstein G. Actigraphic and self-reported sleep quality in women: associations with ovarian hormones and mood. Sleep Med 2015;16(10):1217-1224. [doi: 10.1016/j.sleep.2015.06.009] [Medline: 26429749]
38. Chevance G, Golaszewski NM, Tipton E, Hekler EB, Buman M, Welk GJ, et al. Accuracy and precision of energy expenditure, heart rate, and steps measured by combined-sensing fitbits against reference measures: systematic review and meta-analysis. JMIR Mhealth Uhealth 2022;10(4):e35626 [FREE Full text] [doi: 10.2196/35626] [Medline: 35416777]
39. Nissen M, Slim S, Jäger K, Flaucher M, Huebner H, Danzberger N, et al. Heart rate measurement accuracy of Fitbit Charge 4 and Samsung Galaxy Watch Active2: device evaluation study. JMIR Form Res 2022;6(3):e33635 [FREE Full text] [doi: 10.2196/33635] [Medline: 35230250]
40. Druce KL, Dixon WG, McBeth J. Maximizing engagement in mobile health studies: lessons learned and future directions. Rheum Dis Clin North Am 2019;45(2):159-172 [FREE Full text] [doi: 10.1016/j.rdc.2019.01.004] [Medline: 30952390]
41. Karrer K, Glaser C, Clemens C, Bruder C. Technikaffinität erfassen—der Fragebogen TA-EG. Der Mensch im Mittelpunkt technischer Systeme. 8. Berliner Werkstatt Mensch-Maschine-Systeme 2009;8:196-201 [FREE Full text] [doi: 10.30965/9783657793389_004]
42. Keye WR. Premenstrual Syndrome (PMS). Encycl Endocr Dis 2004:73-75. [doi: 10.1016/b0-12-475570-4/01067-2]
43. Lenton EA, Landgren BM, Sexton L. Normal variation in the length of the luteal phase of the menstrual cycle: identification of the short luteal phase. Br J Obstet Gynaecol 1984;91(7):685-689. [doi: 10.1111/j.1471-0528.1984.tb04831.x] [Medline: 6743610]
44. Fitbit development: web API. Fitbit. URL: https://dev.fitbit.com/build/reference/web-api/ [accessed 2023-06-20]
45. Fitbit development: authorization. Fitbit. URL: https://dev.fitbit.com/build/reference/web-api/developer-guide/authorization/ [accessed 2023-06-20]
46. FHIR Specification v3.0.2. HL7 FHIR. URL: https://hl7.org/fhir/STU3/ [accessed 2023-06-20]
47. McKnight PE, Najab J. Mann-Whitney U Test. In: Weiner IB, Craighead WE, editors. The Corsini Encyclopedia of Psychology, Volume 3. Hoboken, NJ: John Wiley & Sons; 2010.
48. Student. The probable error of a mean. Biometrika 1908;6(1):1-25. [doi: 10.1093/biomet/6.1.1]
49. Stein P, Falco L, Kuebler F, Annaheim S, Lemkaddem A, Delgado-Gonzalo R, et al. Digital womens health based on wearables and big data. Fertil Steril 2016;106(3):e113 [FREE Full text] [doi: 10.1016/j.fertnstert.2016.07.339]
50. Yu JL, Su YF, Zhang C, Jin L, Lin XH, Chen LT, et al. Tracking of menstrual cycles and prediction of the fertile window via measurements of basal body temperature and heart rate as well as machine-learning algorithms. Reprod Biol Endocrinol 2022;20(1):118 [FREE Full text] [doi: 10.1186/s12958-022-00993-4] [Medline: 35964035]
51. Torous J, Lipschitz J, Ng M, Firth J. Dropout rates in clinical trials of smartphone apps for depressive symptoms: a systematic review and meta-analysis. J Affect Disord 2020;263:413-419. [doi: 10.1016/j.jad.2019.11.167] [Medline: 31969272]
52. Pratap A, Neto EC, Snyder P, Stepnowsky C, Elhadad N, Grant D, et al. Indicators of retention in remote digital health studies: a cross-study evaluation of 100,000 participants. NPJ Digit Med 2020;3:21 [FREE Full text] [doi: 10.1038/s41746-020-0224-8] [Medline: 32128451]
53. Simblett S, Greer B, Matcham F, Curtis H, Polhemus A, Ferrão J, et al. Barriers to and facilitators of engagement with remote measurement technology for managing health: systematic review and content analysis of findings. J Med Internet Res 2018;20(7):e10480 [FREE Full text] [doi: 10.2196/10480] [Medline: 30001997]

54. Ajayi KV, Wachira E, Onyeaka HK, Montour T, Olowolaju S, Garney W. The use of digital health tools for health promotion among women with and without chronic diseases: insights from the 2017-2020 health information national trends survey. JMIR Mhealth Uhealth 2022;10(8):e39520 [FREE Full text] [doi: 10.2196/39520] [Medline: 35984680]
55. Holko M, Litwin TR, Munoz F, Theisz KI, Salgin L, Jenks NP, et al. Wearable fitness tracker use in federally qualified health center patients: strategies to improve the health of all of us using digital health devices. NPJ Digit Med 2022;5(1):53 [FREE Full text] [doi: 10.1038/s41746-022-00593-x] [Medline: 35469045]
56. Yu S, Alper HE, Nguyen AM, Brackbill RM, Turner L, Walker DJ, et al. The effectiveness of a monetary incentive offer on survey response rates and response completeness in a longitudinal study. BMC Med Res Methodol 2017;17(1):77 [FREE Full text] [doi: 10.1186/s12874-017-0353-1] [Medline: 28446131]
57. Kawabe R, Chen CY, Morino S, Mukaiyama K, Shinohara Y, Kato M, et al. The relationship between high physical activity and premenstrual syndrome in Japanese female college students. BMC Sports Sci Med Rehabil 2022;14(1):175 [FREE Full text] [doi: 10.1186/s13102-022-00569-0] [Medline: 36163183]
58. Samadi Z, Taghian F, Valiani M. The effects of 8 weeks of regular aerobic exercise on the symptoms of premenstrual syndrome in non-athlete girls. Iran J Nurs Midwifery Res 2013;18(1):14-19 [FREE Full text] [Medline: 23983722]
59. Witkoś J, Hartman-Petrycka M. The influence of running and dancing on the occurrence and progression of premenstrual disorders. Int J Environ Res Public Health 2021;18(15):7946 [FREE Full text] [doi: 10.3390/ijerph18157946] [Medline: 34360242]
60. Cicek G. The effect of regular aerobic exercises on premenstrual syndrome in sedentary women. Balt J Health Phys Act 2018;10(2):34-42. [doi: 10.29359/bjhpa.10.2.04]
61. El-Lithy A, El-Mazny A, Sabbour A, El-Deeb A. Effect of aerobic exercise on premenstrual symptoms, haematological and hormonal parameters in young women. J Obstet Gynaecol 2015;35(4):389-392. [doi: 10.3109/01443615.2014.960823] [Medline: 25279689]
62. Maged AM, Abbassy AH, Sakr HRS, Elsawah H, Wagih H, Ogila AI, et al. Effect of swimming exercise on premenstrual syndrome. Arch Gynecol Obstet 2018;297(4):951-959. [doi: 10.1007/s00404-018-4664-1] [Medline: 29350276]
63. Fuller D, Colwell E, Low J, Orychock K, Tobin MA, Simango B, et al. Reliability and validity of commercially available wearable devices for measuring steps, energy expenditure, and heart rate: systematic review. JMIR Mhealth Uhealth 2020;8(9):e18694 [FREE Full text] [doi: 10.2196/18694] [Medline: 32897239]

Abbreviations

API: application programming interface
FEMFIT: Feasibility Study of Menstrual Cycles With Fitbit Device
MP: menses phase
OD: ovulation day
PMS: premenstrual syndrome

Edited by L Buis; submitted 20.06.23; peer-reviewed by R Sun, A Silva, R AlSaad; comments to author 02.11.23; revised version received 26.11.23; accepted 24.01.24; published 12.03.24.

Please cite as:
Lang AL, Bruhn RL, Fehling M, Heidenreich A, Reisdorf J, Khanyaree I, Henningsen M, Remschmidt C
Feasibility Study on Menstrual Cycles With Fitbit Device (FEMFIT): Prospective Observational Cohort Study
JMIR Mhealth Uhealth 2024;12:e50135
URL: https://mhealth.jmir.org/2024/1/e50135
doi:10.2196/50135
PMID:38470472

Original Paper

Mobile Health App and Web Platform (eDOL) for Medical Follow-Up of Patients With Chronic Pain: Cohort Study Involving the French eDOL National Cohort After 1 Year

Noémie Delage[1], MD; Nathalie Cantagrel[2], MD; Sandrine Soriot-Thomas[3], MD; Marie Frost[4], MD; Rodrigue Deleens[5], MD; Patrick Ginies[6], MD; Alain Eschalier[7], Prof Dr Med; Alice Corteval[7], PharmD; Alicia Laveyssière[7], MSci; Jules Phalip[7,8], MOST; Célian Bertin[8], MD, PhD; Bruno Pereira[9], PhD; Chouki Chenaf[8], MD, PhD; Bastien Doreau[10], MSci; Nicolas Authier[8], Prof Dr Med; ePAIN[11]; Nicolas Kerckhove[8], PhD

[1]Centre d'évaluation et de Traitement de la douleur, CHU Clermont-Ferrand, Clermont-Ferrand, France
[2]Centre d'évaluation et de Traitement de la douleur, CHU Toulouse, Toulouse, France
[3]Centre de Recherche Clinique, CHU Amiens Picardie, Amiens, France
[4]Centre d'évaluation et de Traitement de la douleur, CHU Grenoble, Grenoble, France
[5]Centre d'évaluation et de Traitement de la douleur, CHU Rouen, Rouen, France
[6]Centre d'évaluation et de Traitement de la douleur, CHU Montpellier, Montpellier, France
[7]Analgesia Institute, Clermont-Ferrand, France
[8]Service de pharmacologie médicale, CHU Clermont-Ferrand, Clermont-Ferrand, France
[9]Direction de la recherche clinique et de l'innovation, CHU Clermont-Ferrand, Clermont-Ferrand, France
[10]Laboratoire d'Informatique, de Modélisation et d'Optimisation des Systèmes, Université Clermont Auvergne, Clermont-Ferrand, France
[11]See Acknowledgments, , France

Corresponding Author:
Nicolas Kerckhove, PhD
Service de pharmacologie médicale
CHU Clermont-Ferrand
58 rue Montalembert
Clermont-Ferrand, 63000
France
Phone: 33 473754833
Email: nkerckhove@chu-clermontferrand.fr

Abstract

Background: Chronic pain affects approximately 30% of the general population, severely degrades quality of life and professional life, and leads to additional health care costs. Moreover, the medical follow-up of patients with chronic pain remains complex and provides only fragmentary data on painful daily experiences. This situation makes the management of patients with chronic pain less than optimal and may partly explain the lack of effectiveness of current therapies. Real-life monitoring of subjective and objective markers of chronic pain using mobile health (mHealth) programs could better characterize patients, chronic pain, pain medications, and daily impact to help medical management.

Objective: This cohort study aimed to assess the ability of our mHealth tool (eDOL) to collect extensive real-life medical data from chronic pain patients after 1 year of use. The data collected in this way would provide new epidemiological and pathophysiological data on chronic pain.

Methods: A French national cohort of patients with chronic pain treated at 18 pain clinics has been established and followed up using mHealth tools. This cohort makes it possible to collect the determinants and repercussions of chronic pain and their evolutions in a real-life context, taking into account all environmental events likely to influence chronic pain. The patients were asked to complete several questionnaires, body schemes, and weekly meters, and were able to interact with a chatbot and use educational modules on chronic pain. Physicians could monitor their patients' progress in real time via an online platform.

Results: The cohort study included 1427 patients and analyzed 1178 patients. The eDOL tool was able to collect various sociodemographic data; specific data for characterizing pain disorders, including body scheme; data on comorbidities related to chronic pain and its psychological and overall impact on patients' quality of life; data on drug and nondrug therapeutics and their

benefit-to-risk ratio; and medical or treatment history. Among the patients completing weekly meters, 49.4% (497/1007) continued to complete them after 3 months of follow-up, and the proportion stabilized at 39.3% (108/275) after 12 months of follow-up. Overall, despite a fairly high attrition rate over the follow-up period, the eDOL tool collected extensive data. This amount of data will increase over time and provide a significant volume of health data of interest for future research involving the epidemiology, care pathways, trajectories, medical management, sociodemographic characteristics, and other aspects of patients with chronic pain.

Conclusions: This work demonstrates that the mHealth tool eDOL is able to generate a considerable volume of data concerning the determinants and repercussions of chronic pain and their evolutions in a real-life context. The eDOL tool can incorporate numerous parameters to ensure the detailed characterization of patients with chronic pain for future research and pain management.

Trial Registration: ClinicalTrials.gov NCT04880096; https://clinicaltrials.gov/ct2/show/NCT04880096

(JMIR Mhealth Uhealth 2024;12:e54579) doi:10.2196/54579

KEYWORDS

mHealth; mobile health; eHealth; self-monitoring; chronic pain; observational study

Introduction

Chronic pain affects approximately 30% of the general adult population [1,2] and was one of the top 5 leading causes of years lived with disability in 2016 [3], especially among older people [4]. Moreover, 60% of people with chronic pain are less able or unable to work, and 20% report having lost their job as a result [5]. The overall cost of chronic pain is estimated to be approximately €441 billion (US $468 billion) in Europe and US $560 to $635 billion in the United States [6-8]. At the same time, the market for analgesic drugs represented approximately US $68 billion in 2016, and an increase from 2% to 5% was forecast for 2021, with a further 5% increase by 2025 [9]. Unfortunately, available analgesics often have limited efficacy, with undesirable effects and without any real pharmacological innovation [10], despite prolific basic research [11].

One of the solutions could be based on better patient characterization to offer physicians decision aids in the management and initial choice of treatment for chronic pain. A preliminary step that would make it possible to identify patient trajectories and factors capable of predicting treatment success. This characterization should cover the entire biopsychosocial field. It should also include a temporal dimension, with real-life monitoring of different parameters, and subjective and objective markers of chronic pain. Indeed, the current assessment of patients with chronic pain provides only fragmentary data on their daily pain experiences due to memory bias [12]. This monitoring strategy is currently being developed by several research teams [13-18]. It is also necessary to make the patient an active participant in their own care, with the goal of improving compliance and adherence. Indeed, evidence suggests that the self-management of chronic diseases reduces hospitalizations, the use of emergency services, and overall care management [19,20].

The use of innovative digital technologies (eHealth) appears to be a solution. eHealth can improve the characterization and monitoring of pain symptoms, and the management of pain and associated comorbidities [21-23]. eHealth can also offer remote interventions without the difficulties inherent in long-distance travel or regular visits to the general practitioner or specialist [24]. Thus, recent publications have highlighted the urgent need to develop, assess, and use validated eHealth programs for chronic pain [13,14,18,25-31]. eHealth technologies encompass a wide range of diverse tools, offering potential applicability for chronic conditions [32,33]. These technologies can be categorized into 6 types: virtual visits, electronic health records, digital therapeutics, artificial intelligence and machine learning, wearable monitors, and mobile health (mHealth) apps [32]. This last category can vary in its content, which can generally be grouped into 3 objectives: education, monitoring, and treatment (often including self-management strategies), sometimes with a combination of these objectives [34]. However, there are many obstacles in putting this management into practice and monitoring it via digital tools, including the lack of training of patients in the management of their pathology and the use of digital tools, the lack of interest or training of some physicians for this new management integrating new digital technologies, the lack of integration of these digital strategies in medical management, the over-medicalization of pain that can result from the use of these digital tools, and the lack of time and information in health care structures [35].

Change is therefore necessary to integrate digital technology in the management of chronic diseases, including pain [36]. Among digital tools, mHealth offers significant opportunities for patients to foster communication with their practitioners or other patients (ie, through forums). mHealth offers distraction strategies, provides information and therapeutic education, promotes self-expression, improves access to health care, and facilitates social support. Moreover, mHealth can provide patients with greater convenience and more regular access to information about their conditions compared to traditional care. Although many studies have examined the use of innovative mHealth to assist in the management of chronic diseases, less attention has been paid to chronic pain [37]. However, the context is favorable, with growing interest from health agencies and funding organizations for these solutions. Moreover, in France, patient demand was evidenced by the results of a survey on compliance conducted in early 2017 by the French Institute of Public Opinion. The results showed a strong demand by patients for the development of digital tools for the self-management of their chronic diseases.

In response to this situation, the Analgesia Institute (Clermont-Ferrand, France) has recently created an mHealth

tool (called eDOL) for data collection and monitoring of chronic pain and its impacts on patient well-being. Following a feasibility study on the use and acceptability of eDOL Version 1 [38] and the implementation of Version 2, we provide in this article an overview of eDOL's ability to retrieve a significant amount of medical data of interest for physicians and researchers on the subject of chronic pain. This study was carried out after more than a year of existence of the eDOL national cohort, and it included more than 1000 patients. This study did not assess the impact of the eDOL tool on patients' medical management or on patients' pain symptoms and their consequences.

Methods

Ethical Considerations

The study was approved by the research ethics committee (Comité de Protection des Personnes Ouest II Angers; reference number: 2020-A02027-32) and the French Data Protection Authority (Commission Nationale de l'Informatique et des Libertés [CNIL]; n° 921059). The study has been registered at ClinicalTrials.gov (NCT04880096). The study was conducted in accordance with French laws and regulations on research on human beings and data protection, and in accordance with the Declaration of Helsinki [39].

Data Collection

Data were collected and managed using the eDOL tool (mobile app + web platform) developed by Bepatient (Paris, France) and hosted by Avenir Télématique (Villeneuve-d'Ascq, France). In accordance with the provisions relating to the confidentiality of information concerning, in particular, the people who took part in the research and the results obtained [40], individuals with direct access took all the necessary precautions to ensure the confidentiality of the information relating to the participants. These persons and the investigators themselves are subject to professional secrecy [41]. All data collected and transmitted to the sponsor (University Hospital of Clermont-Ferrand, France) were anonymized, and each patient had an individual numerical code. The head of research ensured that each patient was informed of which data were collected and that they did not object to their use or disclosure. Answers to questionnaires and medical data were transmitted in spreadsheet format (Excel 2013; Microsoft Corporation). All anonymized data were accessible to the biostatisticians (BD and BP), coordinator (ND), and project manager (NK). Only the investigators could access their patients' personal data to identify them. A dashboard linking patients' identities and individual numerical codes was available only on the investigators' professional interface on the eDOL web platform. The final database, which was used for statistical analyses, included only individual numerical codes to preserve anonymity.

Data Entry and Processing

To manage data processing, a data warehouse has been set up in the Laboratory of Informatics, Modelling and Optimization of the Systems (LIMOS) Mixed Unit of Research (UMR 6158) (Clermont Auvergne University, France).

This warehouse has made it possible to retrieve the anonymized data collected by BePatient regularly and automatically. The data were retrieved via a secure application programming interface (API). Files were saved and read to extract the data. These data were then cleaned according to a set of rules to remove erroneous data (patient input errors), integrate them into a database, and provide an interface to view and export the cohort data.

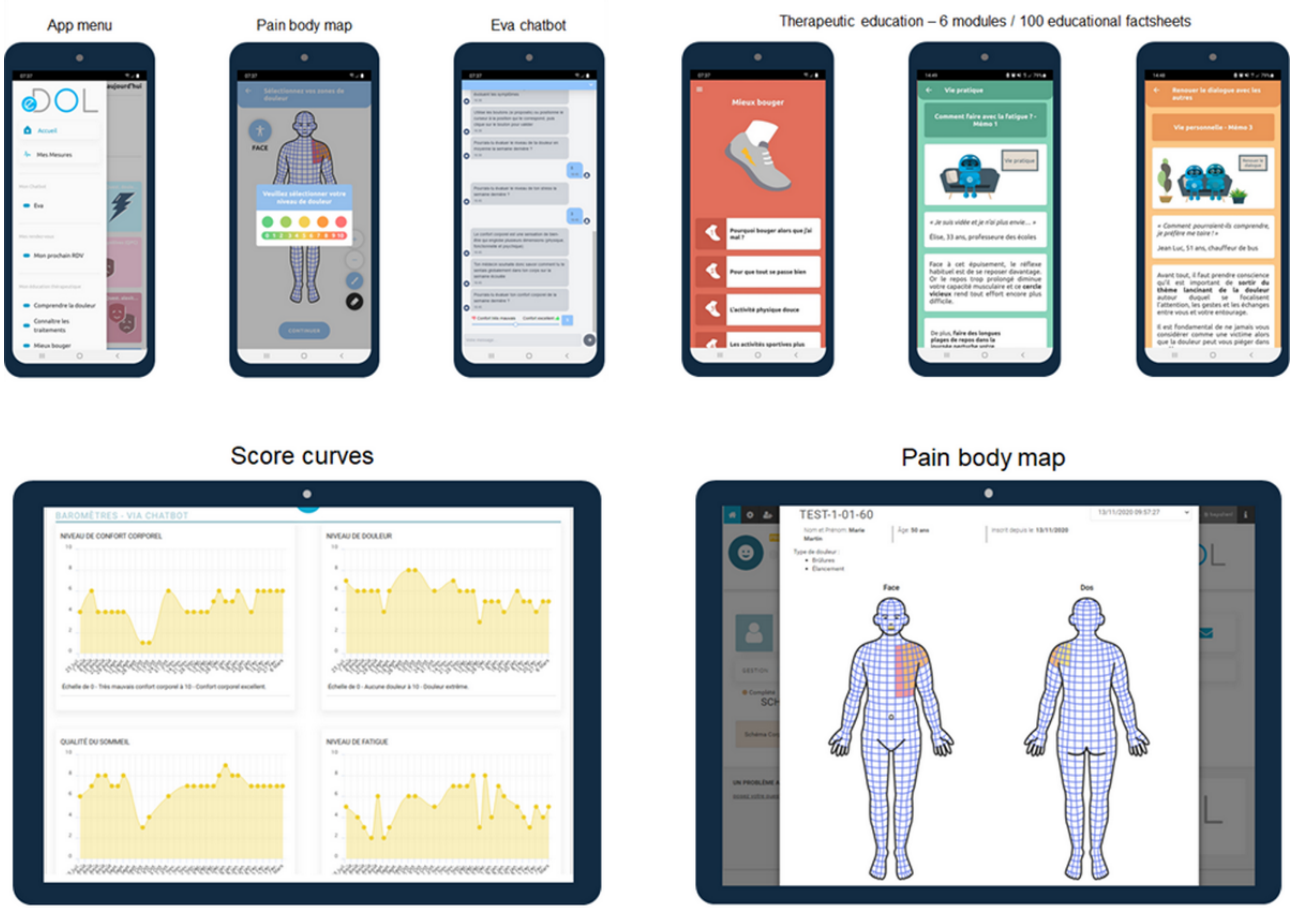
Study Design and Population

This cohort study provides a snapshot after 1 year of a national cohort of patients with chronic pain using the eDOL app. The characterization and real-life monitoring of patients from 18 pain clinics and follow-up by 80 investigators in France took place between September 14, 2021, and January 31, 2023. The study was offered to all physicians and allied health professionals in the investigating centers. Participation in the study was offered to patients with chronic pain, who were owners and regular users of a smartphone and who were followed up in a pain clinic. All adult (≥18 years old) patients able to read and understand French and provide consent to integrate the cohort study were included. Participants were free to withdraw their consent at any time by informing the sponsor. Each patient had access to the information document (paper or electronic) detailing the purpose, content, and conduct of the cohort. If they agreed to participate, they were asked to download the eDOL app and complete questionnaires using the eDOL app. The URL to access this app was sent by email from physicians to their patients. After downloading the app and creating their profiles, patients could accept the general terms and conditions of use and confirm that they agree to the use of their medical data in this study. Each patient made only 1 initial study visit during which the physician introduced the study to them, checked their eligibility, explained the eDOL tool, and gave them a brief training document on how to use the eDOL app. Participants completed several questionnaires and assessments over a period of 1 month (initial patient characterization) and then repeatedly (or not) quarterly, half-yearly, and yearly depending on the questionnaire. Throughout the duration of the study, physicians saw their patients at several follow-up visits (optional), according to the patient's health pathway, after the inclusion visit.

eDOL Digital Tool

eDOL is a digital health tool used for data collection and monitoring related to pain and its impacts on patient well-being. It comprises 3 main components: a smartphone app for patients, a web interface for health care professionals (Figure 1), and a data repository for researchers. The tool provides various features and modules to support patients in managing their chronic pain while allowing physicians to access and monitor real-time data for clinical and therapeutic purposes.

Figure 1. Interface of eDOL Version 2. Screenshots of the various modules of the eDOL smartphone app for patients (top) and the eDOL web interface for clinicians (bottom).



Patient Mobile App

The app allows patients to regularly input data regarding their pain and its effects and to access educational content. Patients were invited to complete various contents, but none of them were compulsory.

Questionnaires

Patients complete questionnaires and assessments on a regular basis. These questionnaires cover general aspects, specific chronic pain symptoms, and associated comorbidities. They are filled systematically at different intervals, such as once only, quarterly, half-yearly, or yearly. Some questionnaires address specific types of chronic pain. For a detailed list of all the questionnaires used, refer to the previous publication by Kerckhove et al [38].

Weekly Meters

The app facilitates the real-life monitoring of various parameters related to pathology. Patients can assess the evolution of their pain and its impacts using an 11-point numeric rating scale (from 0 to 10) for intensity assessment in different areas, such as pain, stress and anxiety, fatigue, sleep, morale, and bodily comfort.

Body Map

Each month, patients can use a body map to select and indicate the painful areas on their body, assigning corresponding intensity levels. The map provides front and back views (Figure 1).

Chatbot

Patients can download a chatbot (conversational agent called Eva) to integrate it into the app. The Eva chatbot has been developed in collaboration with the company Aliae (Metz, France) and with a medical group of the Analgesia Institute. The chatbot prompts participants to complete their meters on a weekly basis.

Educational Modules

The app offers educational modules that cover therapeutic education on pain and its broader aspects. These modules consist of informative sheets with images for better comprehension. They have been developed with the assistance of a group of physicians working in various pain clinics. Each module starts with a “What you need to know” section, defining the chapter’s scope, followed by educational sheets comprising an image corresponding to the subsection, a short testimonial or quote, an explanatory text, and a personalized image with a concluding statement. Modules include understanding pain, understanding treatments, moving better, improving fatigue and sleep, functioning in daily life, and managing emotions.

In the event of technical problems or questions regarding the app or the completion of questionnaires or meters, participants could refer to contact details provided via an email, and for each investigator center, a clinical research assistant was available by phone.

Web Interface for Health Care Professionals

The eDOL web platform enables health care professionals to access real-life patient data (questionnaire scores and graphical presentation of weekly meters) and input consultation data (clinical examinations, treatments, etc). The platform features an ergonomic dashboard with tabs for management (patient medical records), health measures (graphical display of weekly assessments), and questionnaires (all completed with scores and answers).

Big Data Repository for Researchers

The development of eDOL V2 involved the creation of a data repository in collaboration with the LIMOS. The data repository interfaces directly with eDOL V2 and automatically collects patient cohort data (via the app) and data from health care professionals (via the web platform) on a weekly basis. The transferred data are anonymized to comply with the French data protection authority for analysis purposes. Information on the data contained in the database is available to investigators on request. In the event of a clinical study requiring eDOL data, a written request describing the study must first be sent to the data controller (Analgesia Institute). The Analgesia Institute's Scientific Committee will assess the relevance of the project before giving its approval.

Study Outcomes

The primary objective of this study was to describe the tool's ability to retrieve a large amount of longitudinal medical data and its relevance to cohort patient follow-up.

The secondary objectives were to assess the characterization of participating patients, pain disorders, pain medications, and related comorbidities, and evaluate the acceptability of the eDOL app with a satisfaction survey (based on the Patient Satisfaction Questionnaire Short Form [42] and the Client Satisfaction Questionnaire [43,44]) for patients (10 questions). The satisfaction survey (in French language) was sent to each included patient via the eDOL tool. Response options for each question ranged from 0 (strongly disagree with the statement) to 10 (strongly agree with the statement). The questionnaire/weekly meter completion rate and center participation (inclusion rate) were also calculated.

As an exploratory study, the profile of patients adhering to the eDOL program was analyzed.

Statistics

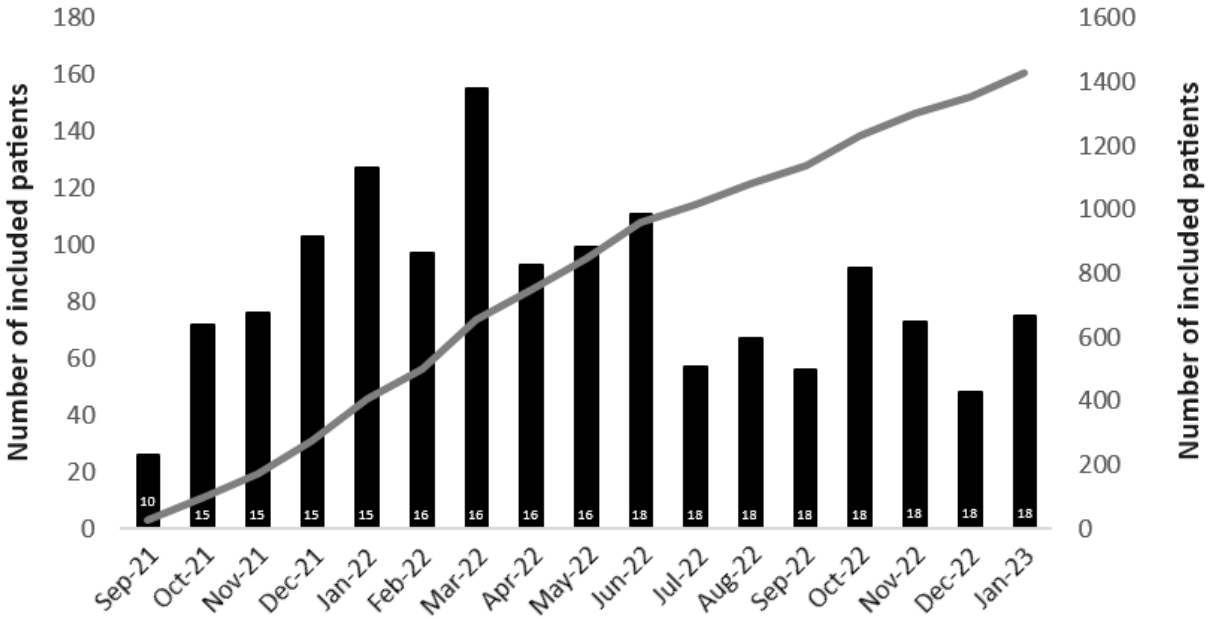
Patients have been described according to the following variables: compliance with eligibility criteria, epidemiological characteristics, and clinical and treatment characteristics. Categorical data are expressed as numbers and associated percentages, and continuous data are expressed as means with standard deviations or medians (25th-75th percentiles), according to the statistical distribution. The assumption of normality was studied using the Shapiro-Wilk test. Comparisons according to observance were conducted using the chi-square test or Fisher's exact test for binary variables, whereas comparisons concerning quantitative variables were performed using ANOVA or the Kruskal-Wallis test. When appropriate (omnibus P value <.05), a post-hoc test for 2×2 multiple comparisons was applied: Tukey-Kramer test after ANOVA or Dunn test after the Kruskal-Wallis test. Different analyses were carried out according to the level of use (compliance) of the eDOL app, which was defined by quartiles of percentage weekly meters completed during follow-up (Q1: <5%, Q2: <50%, and Q3: >50%). Statistical analyses were performed using Stata software (Version 15; StataCorp). All tests were 2-sided, with an α level set at 5% for statistical significance.

Results

Generalities

All the 18 pain clinics participating in the study included at least one patient. The median rate of inclusion per center was 5.5 patients/month, and an average of 84.5 (SD 31.9) patients were consistently included every month (Figure 2). During the study period, from September 14, 2021, to January 31, 2023, 1427 patients were included.

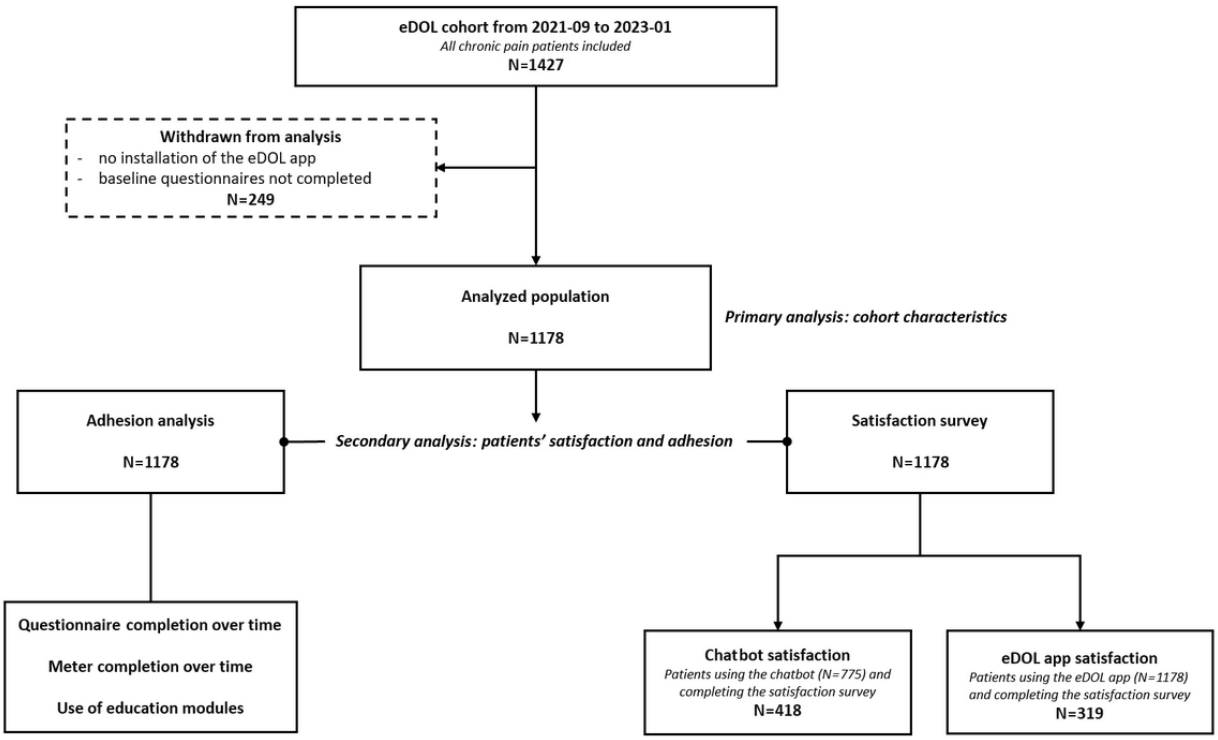
Figure 2. Recruitment rate by month and cumulative frequency. Rate of patient recruitment in the eDOL study is represented in terms of patient frequency/month (histogram) and cumulative frequency (curve). The numbers at the bottom of the histogram bars represent the number of active investigator centers.



Collected Data

Of the 1427 patients included in the study, 249 (17.4%) were not analyzable (no installation of the eDOL app or no completion of the baseline sociodemographic questionnaire). Thus, data from 1178 patients were analyzed (Figure 3). At the time of data recovery (February 1, 2023), 23.3% (275/1178) of patients had ≥12 months of follow-up, 30.1% (355/1178) had ≥9 to <12 months of follow-up, 20.5% (242/1178) had ≥6 to <9 months of follow-up, 11.5% (135/1178) had ≥3 to <6 months of follow-up, and 14.3% (168/1178) had <3 months of follow-up. The median follow-up time for patients in the cohort was 9.3 (IQR 9.1-9.5) months. eDOL's functionalities enable it to retrieve a large and varied range of biopsychosocial data (Multimedia Appendix 1). Below is a detailed overview of the sociodemographic and medical data collected, which allowed us to precisely characterize our cohort population.

Figure 3. Study flowchart. Graphical representation of the various stages in patient recruitment and analysis.



At baseline, participating patients were mostly middle-aged (median 49.0, IQR 48.3-49.7 years), women (937/1178, 79.5%), in a relationship (759/1178, 64.4%), with children (635/1178, 53.9%), nonsmoking (805/1096, 73.4%), with a university degree (581/1178, 49.3%), and in professional activity (732/1178, 62.1%), and a significant number were considered to be in a precarious state (468/980, 47.8%). Among patients in professional activity, 34.1% (250/732) were on sick leave due to their chronic pain (Multimedia Appendix 2 and Multimedia Appendix 3).

Most patients (906/1034, 87.6%) had moderate to severe pain (>5/10), of which 26.0% (236/906) had a high chronic pain interference score (called "high impact chronic pain" [45]). Moreover, 25.1% (296/1178) had at least two different types of chronic pain. Most patients (492/1178, 41.8%) had nociplastic pain, and the duration was longer than 5 years in more than 50% (601/1178, 51.0%) of patients. Most participants (667/1178, 56.6%) were at the beginning of their course in the pain clinic (≤6 months), with a median of 5.0 (IQR 3.0-7.0) months prior to inclusion.

Several patients (481/1094, 44.0%) described their chronic pain as permanent during both day and night (with painful paroxysms every day) and as inducing frequent nocturnal awakenings (995/1094, 90.9%) (Multimedia Appendix 3 and Multimedia Appendix 4). Moreover, body schemes were recovered for 73.1% (861/1178) of patients. Among these patients, 94.0% (809/861) had at least two different pain locations. The dorsolumbar, hypogastric, and cervical regions were the most frequent in terms of pain location, with over 40.0% (344/861) of patients affected (Multimedia Appendix 5).

The eDOL tool also enabled us to collect data on comorbidities related to chronic pain and its psychological and overall impact on patients' quality of life. High numbers of patients were considered to have kinesiophobia (753/1029, 73.2%), alexithymia (591/1038, 56.9%), degraded quality of life (599/1025, 58.5%), catastrophism (448/1056, 42.4%), and a possible cognitive disorder (774/1008, 76.8%). More than 60% (640/1033, 62.0%) of patients had impaired sleep, and 44.7% (457/1023) and 26.8% (274/1023) had proven anxiety and depressive disorders, respectively (Multimedia Appendix 3 and Multimedia Appendix 6).

With regard to analgesic treatments, 2643 treatments were prescribed for 839 patients (71.2%), with 64.8% (763/1178) of patients using nondrug analgesic treatments, including transdermal electroneurostimulation (480/763, 62.9%) and physical techniques such as physiotherapy (352/763, 41.6%) (Multimedia Appendix 3 and Multimedia Appendix 7). Interestingly, 23.3% (275/1178) of patients were not receiving treatment (pharmacological or nonpharmacological) for their pain at the time of inclusion. The drugs used by the patients were mainly antidepressants (369/839, 44.0%), followed by paracetamol (363/839, 43.2%), opioids (358/839, 42.7%), and antiepileptics (224/839, 26.7%). The therapeutic benefit was assessed by clinicians for 1407 prescribed treatments, and opioids and antidepressants had the best medical benefit (1282/1407, 91.1% of treatments were rated as very good or good) (Multimedia Appendix 8). In the same way, clinicians were able to report on treatment adverse events (AEs) and patient compliance. A total of 289 AEs were reported, and antidepressants and anticonvulsants were the most likely causes (73/289, 25.4% and 66/289, 23.0% of AEs, respectively) (Multimedia Appendix 8). Compliance was assessed by clinicians for 1192 prescribed treatments, and in 97.1% (1158/1192) of cases, compliance was very good or good.

Lastly, the eDOL tool enabled clinicians to record patients' medical history. A large majority of patients (981/1178, 83.3%) had at least one medical history (Multimedia Appendix 9). Among these patients, 60.9% (597/981) declared a traumatic life history (mainly death, accident, and divorce), 35.3% (346/981) reported a history of violence (mainly physical and psychological), 43.3% (425/981) reported a psychiatric history (mainly anxiety and depression), 37.9% (372/981) reported addictive behaviors (mainly tobacco), and 82.2% (806/981) reported ongoing illnesses (mainly rheumatologic, digestive, and neurological).

Real-Life and Longitudinal Data Collected

As previously reported, 82.6% (1178/1427) of included patients effectively used the eDOL tool in their daily life. On average, each questionnaire was completed by 86.7% (SD 2.0%) of participants at inclusion. Among the questionnaires repeated every 3 (2 questionnaires) or 6 months (7 questionnaires), this percentage decreased over time to 60.8% (SD 0.1%) at the 3-month follow-up, 42.8% (SD 3.6%) at the 6-month follow-up, 32.1% (SD 0.1%) at the 9-month follow-up, and 23.6% (SD 2.5%) at the 12-month follow-up (Multimedia Appendix 10).

The completion rate of the weekly assessments for the real-life monitoring of the different meters (pain, moral, anxiety, fatigue, sleep, bodily comfort, and physical activity) was 75.1% (885/1178) at patient inclusion. Among the 1007 patients with at least 3 months of follow-up, the completion rate decreased over the course of follow-up to 49.4% (497/1007) after 3 months of follow-up and then stabilized at 39.3% (108/275) for patients with at least 12 months of follow-up (Multimedia Appendix 10). An acceptable meter completion rate was defined as at least 50% completion of all theoretical meters over the follow-up period (1 meter/week). Univariate analysis was performed to characterize patients according to their percentage of weekly meter completion (<5%, ≥5% to <50%, and ≥50%). Univariate analysis showed very little difference between compliant and noncompliant patients. In fact, women and patients with an intermediate profession or no activity appeared to be slightly more compliant. Age, level of education, family situation, and alcohol or tobacco consumption had no impact on compliance. Surprisingly, the same observation was noted for pain characteristics. It was found that the length of time the pain had been present, the frequency or duration of painful paroxysms, the pain intensity and interference, and the presence of nocturnal awakenings or difficulty in falling asleep did not interact with patient compliance. There was no difference in the inclusion scores for the various questionnaires, depending on the level of compliance of patients. Finally, analysis of the type of pain and treatments received showed no difference between patients irrespective of their level of compliance.

The eDOL tool also includes educational modules on chronic pain and its repercussions. These modules were freely available to the patients but were not compulsory. Overall, 16.3% (192/1178) of patients viewed at least one educational module. The educational modules were viewed in their entirety by around 65.2% (minimum: 44.6%, maximum: 85.5%, according to the educational module) of patients who opened them. The most frequently viewed educational module was the one on "improving fatigue and sleep." Conversely, the educational module on "living with pain" was the least viewed (Table 1).

Table 1. Educational module use.

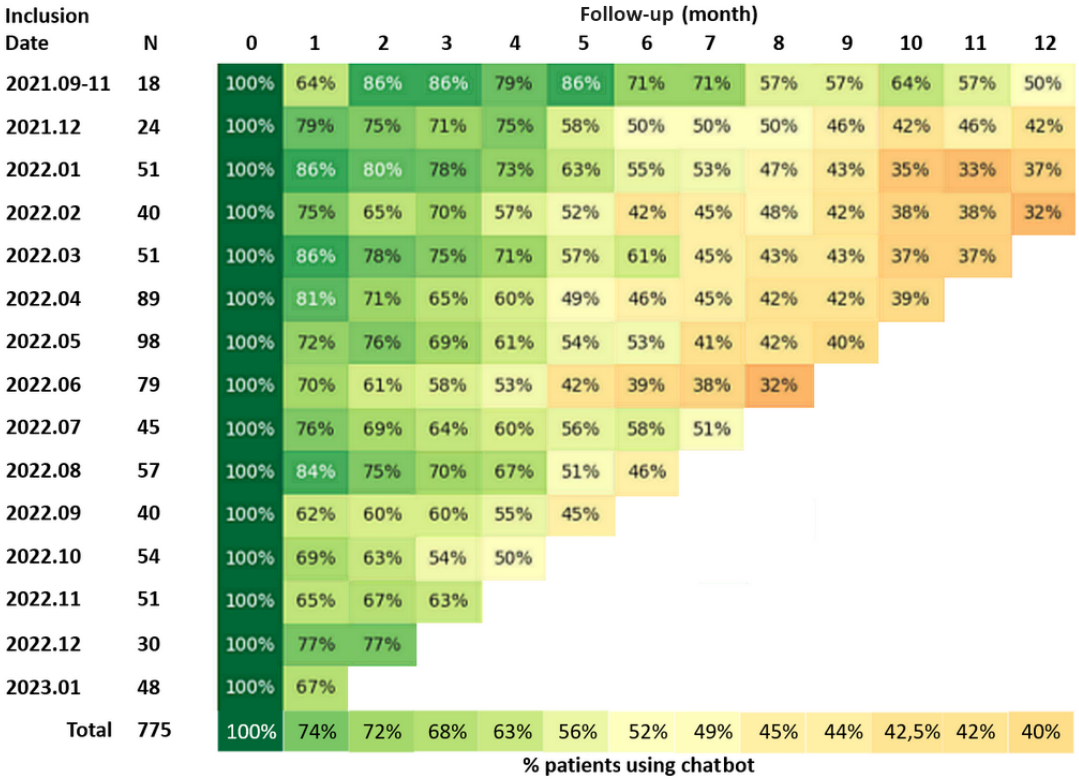
Educational module	Patients opening the module (N=1178), n (%)	Patients viewing the entire module (N=1178), n (%)
Understanding pain	188 (16.0)	118 (10.0)
Knowing about treatments	181 (15.4)	129 (10.9)
Moving better	173 (14.7)	148 (12.6)
Improving fatigue and sleep	369 (31.3)	272 (23.1)
Living with pain	159 (13.5)	71 (6.0)
Managing emotions	209 (17.7)	111 (9.4)

Satisfaction With the Use of the eDOL Tool

The satisfaction questionnaire was filled in by 27.1% (319/1178) of patients at the end of the study. These patients had similar characteristics to all the patients in the cohort (average age 49.4 years; 258/319, 80.1% female; 213/319, 66.9% with pain for more than 5 years; and 137/319, 42.8% with nociplastic pain). The median acceptability score was 7.2/10 (95% CI 6.9-7.7), with only 7.2% (23/319) of patients providing a rating less than 5/10. Moreover, 83.4% (267/319) of patients who responded wanted to participate in the further development of the eDOL tool. The items with the lowest scores were "I enjoy my exchanges with the Eva chatbot and I think it's a tool that can assist me on a daily basis" (mean 6.0, SD 3.1), "I think the Eva chatbot is a relevant tool for collecting my meters every week" (mean 6.7, SD 2.8), and "I believe that the information I have entered in eDOL enables my doctor to better understand my pain and improve its management" (mean 6.9, SD 2.6) (Multimedia Appendix 11).

The interest of adding a chatbot in the eDOL tool was also assessed. Among the 1178 participating patients, 775 (65.8%) downloaded the chatbot, and among these, all 775 (100%) used the chatbot at inclusion. This percentage halved (403/775, 52.0%) at the 6-month follow-up and appeared to stabilize at 40.0% (310/775) at the 12-month follow-up (Figure 4). The use of the chatbot allowed us to recover 9456 weekly meters and 12,300 conversations. With regard to satisfaction, of the 775 patients, 418 (53.9%) responded to the questionnaire, and among these patients, 230 (55.0%) had a satisfaction score of at least 7/10, which corresponds to high satisfaction, and only 35 (8.4%) had a score below 5/10, which corresponds to low satisfaction. The main positive points of the chatbot indicated by the participants were "easy to use," "responsive," "moral support," "converse with a positive message," and "availability." Conversely, the main difficulties related to the use of the chatbot indicated by the participants were "forgetfulness of use," "answers sometimes not adapted," and "repetitive."

Figure 4. Chatbot use over time. Visual representation of chatbot use among the 775 patients who downloaded the chatbot. Patients were classified according to their month of inclusion, and chatbot use (at least one conversation in the month) was assessed for each month of follow-up. A darker green color represents a higher utilization rate, and a darker red color represents a lower utilization rate.



Discussion

Principal Findings

This study presents the most extensive cohort of patients using an mHealth follow-up app and living with chronic pain, irrespective of the type of pain, worldwide. Indeed, studies in the literature on the use of an mHealth app in a cohort of patients have mainly focused on specific types of pain conditions, predominantly osteoarthritis and chronic low back pain [34]. Most studies have concentrated on digital interventions directed toward the patient-provider interface [46], which makes comparisons with the available literature challenging. This is also due to the typically shorter duration of such studies.

In this cohort study, 82.6% (1178/1427) of eligible patients consented to use the eDOL tool, and among them, about 86.7% (1021/1178) of participants used the eDOL app to complete different questionnaires during the initial phase. Thus, the rate is in line with the rate identified in our previous feasibility study (89.3%) [38] and the rate reported in the study by Stoppok et al [47] wherein 85.5% of participants expressed a willingness to use the app for more than a month when surveyed beforehand. Nevertheless, attrition appears to be high in internet interventions for chronic pain [48], and this has been reflected in our study, with about 60% (167/275, 60.7%) loss of users (according to weekly meter filling) among patients with at least 12 months of follow-up. A comprehensive American study [49] involving a large sample revealed that over half of the participants discontinued participation within the first week. Attrition is not solely confined to mHealth, and a systematic review of longitudinal studies on chronic spinal pain identified attrition as the most significant risk of bias [50]. Notably, attrition rates varied widely across these longitudinal studies, with a rate of 63% after 12 months in a study assessing pain after intensive care unit discharge [51] and 83% after 12 months in a study assessing pain self-management (chronic musculoskeletal pain and comorbid depression) [52]. Moreover, in the longitudinal study conducted by Bicego et al [53], attrition rates varied within the range of 11% to 22% 1 year after treatment.

With the loss of patient adherence over time, it is necessary to include a huge number of patients (inclusion of a minimum of 5000 patients is planned in the eDOL cohort) to have enough longitudinal data available. Despite this, the eDOL tool is already making it possible to collect a significant volume of data to characterize patients experiencing chronic pain, and it opens the door to future clinical studies to better understand the biopsychosocial aspect of chronic pain, the therapeutic management (drug or nondrug) implemented, the patient's care pathway (type of consultation, frequency of consultation, medical wandering, etc), and the cost of management (treatment prescriptions, consultations, work stoppage, etc). All these possibilities make the eDOL cohort an enormous potential source of data on chronic pain that is available to clinicians and researchers.

Limitations and Difficulties Encountered

The main limitation of our study was the attrition rate described previously. This attrition could influence the results of patient satisfaction with the eDOL tool, as only 30% responded to the

satisfaction questionnaire. However, it is difficult to know whether the patients who did not respond were dissatisfied, or simply forgot or did not feel like answering the questionnaire despite their effective satisfaction, especially since compliant patients and those who responded to the satisfaction questionnaire had similar characteristics to all the patients in the cohort. In our study, the chatbot was an attempt to add a social role to this tool in order to establish a good therapeutic alliance, as recommended for promoting engagement and reducing attrition in digital interventions [54]. However, its algorithms were trained in the United States, which might make it impersonal and irrelevant for the French population, which are characteristics identified as barriers to the use of such tools [55]. The lack of use of the chatbot could also be explained by patients' lack of interest in "chatting" with a robot, as such a robot has shown several limitations in its exchanges. One solution would be to replace it with a chat or message involving health care staff, who have already shown an interest in adhering to the digital tool [56]. However, this strategy inevitably requires additional time for caregivers to respond to patients' expectations, and such time is rarely available in view of caregivers' current workloads in pain clinics. An additional possibility is to add a forum for patient users to exchange information with each other. This strategy has been proven to be beneficial [31].

Educational modules have been integrated into the eDOL app to make it more informative for patients and thus increase their adherence. Nonetheless, patients made very little use of them. Their low use can be partially explained by the fact that the modules were not available on the homepage but were present on a dedicated page of the menu. Another explanation is that patients only received 2 smartphone notifications informing them of the existence of these modules in the app. We can hypothesize that patients might have had difficulties in finding them.

Additionally, in the absence of compulsory follow-up visits in the eDOL protocol, patient visits to the pain clinics were those scheduled as part of their usual care pathway (ie, not focused on the app itself); thus, some patients had visits 3, 6, or even 12 months apart. This may have acted as a hindrance to user engagement, as noted in telemedicine interventions [57]. Furthermore, although the use of an app and a mobile phone appears to be widely preferred over other interface options to access pain-related resources among individuals living with chronic pain, less than 60% of them expressed a desire to use online pain management tools permanently, according to a German survey [47]. This finding could also explain the significant attrition commonly observed in mHealth interventions.

Various strategies are available to mitigate this adherence issue observed within this study, including reducing the number of questionnaires and the number of questions per questionnaire, simplifying navigation (user experience/user interface design), adding activities to help patients manage their symptoms (therapeutic education for sleep, mindfulness for stress, positive psychology for morale, etc), ensuring greater involvement of the care team, etc. Our research team has already taken steps to mitigate attrition rates in the ongoing novel version of the eDOL tool (Version 3; since November 2023) by reducing the number of questionnaires, replacing some questionnaires with shorter ones (PROMIS [Patient-Reported Outcomes Measurement Information System] questionnaires), enhancing user experience by designing a new interface, and deactivating the chatbot. Additionally, the reminders to complete meters, body schemes, or questionnaires were changed into notifications instead of emails as patients did not solely rely on their phones to read their emails. The limited interaction with educational modules suggests that there is an opportunity to enhance user engagement and accessibility features within the app. The new design should facilitate increased engagement with therapeutic education activities by making them more prominently visible. Additional activities are currently being prepared and will be incorporated into the new version soon. Finally, to limit missing medical data, the new version of eDOL enables us to link our data with the national databases of the French health insurance system. This linkage will make it possible to obtain data on all the care received by patients in our cohort, including treatment reimbursements, doses and quantities of treatments prescribed, hospitalizations or consultations, sick leave benefits, etc. This linkage will also enable us to carry out medicoeconomic studies on the cost of chronic pain management.

The second limitation pertains to the representativeness of the sample in this cohort regarding the general population living with chronic pain. The participants in this study were recruited exclusively from pain clinics; however, only 3% of people living with chronic pain benefit from this type of health care setting in France. The vast majority seek assistance in primary care settings or in other specialized services such as rheumatology, neurology, and oncology. Nonetheless, the characteristics of the patients included in the eDOL cohort demonstrate resemblances to people living with chronic pain in Europe in terms of various factors such as age, pain intensity, duration of pain, and comorbidities [2,58-60]. These data reveal a higher proportion of women, which is not surprising as this difference in prevalence is well documented in the literature [61-63]. This female-to-male ratio is also observed in remote research studies utilizing mobile apps (median 56.9%) [49]. However, as in our feasibility study [38], this ratio is exacerbated in our cohort, reaching nearly 8/2. There are other examples of digital health studies showing a similar gender distribution [64-66].

A selection bias was evident, primarily due to the prerequisite of smartphone utilization, which inherently excludes individuals who lack access to or proficiency with this technology. This criterion could potentially result in the exclusion of older patients or patients with a more precarious status. Lastly, another limitation could be measurement bias, a phenomenon commonly encountered in observational studies [67]. It can emerge from factors such as recall period, selective recall, social desirability, or approach to sampling. Within our study, the recall period emerged as the primary potential source of risk. Given that all the inquiries focused on the present moment or, at the most, events occurring within the preceding 1 to 2 weeks, the likelihood of recall bias can be considered marginal.

Conclusion

Overall, the eDOL tool collected a large volume of medical and biopsychosocial data from 1178 patients over a short period of just over 1 year. The quantity of data collected demonstrated the effectiveness and interest of the eDOL tool with regard to the collection of medical data for research and for caregivers. This interest will continue to grow as more patients are included in the cohort, enabling clinical studies to be carried out on eDOL data to assess care pathways, management costs, patient trajectories, patient profiles, benefits or risks of treatments, and other aspects.

Acknowledgments

The ePAIN group includes the following investigators: Pascale Picard, Lola Vachon, Cécile Le Chapelin Frenay, Eric Serra, Johan Leclerc, Stéphane Foulon, Ludovic Douay, Nadine Attal, Gregory Tosti, Lola Ranaivoson, Andreia Murgu, Emmanuel Nolot, Annachiara Spagna, Serge Perrot, Anne-Priscille Trouvin, Françoise Laroche, Sinja Meyer, Nicolas Seytre, Julien Guerin, Dominique Perocheau, Doumit Chahwan, Christian Coichard-Guy, Andréa Passard, Caroline Maindet-Dominici, Hasan Hodaj, Pascale Vergne-Salle, Marc Rozette, Daniele Ranoux, Lucie Guillain, Nathalie Vercoustre, Samuel Seigneur, Anaïs Depouilly Lachatre, Christian Gov, Geneviève Demarquay, Manon Duraffourg, François Jedryka, Eric Viel, Anaïs Pages, Sophie Pouplin, Marianick Guidt, Marine Lanquetuit, Cécile Lestrade, Emmanuel Combe, Antoine Elyn, Marguerite D'ussel, Anne Coutaux, Chloé Lacoste, Anne Wyniecki, Julien Nizard, Emmanuelle Kuhn, Sebastien Abad, Julien Labarre, François Leger, Blandine Armand, Jean-Baptiste Lamorte, Alexandra Menet, Isabelle Berkelmans, Agathe Isselin, Christelle Creac'h, Béatrice Lietar, Aurélie Maxo, Mélanie Forestier, Philippe Nayme, Dominique Payre, Isabelle Fayolle Minon, Roland Peyron, Jean-Marie Amodeo, Irena Cussac, Xavier Moisset, Sakahlé Conde, Raphael Gonon Demoulian, Véronique Vesperini, Patrick Olive, and Fabienne Marcaillou.

Conflicts of Interest

None declared.

Multimedia Appendix 1
eDOL tool contents. Listing of all modules and functions integrated into the smartphone app and eDOL web interface.
[PNG File , 333 KB - mhealth_v12i1e54579_app1.png]

Multimedia Appendix 2
Sociodemographic characteristics of the cohort. Graphical representation of the characteristics of patients included in the eDOL cohort. Sex, age class, marital status, alcohol or tobacco consumption, and professional status are represented in percentage of patients.
[PNG File , 108 KB - mhealth_v12i1e54579_app2.png]

Multimedia Appendix 3
Supplementary data. All questionnaires and responses on characteristics of patients, chronic pain, and pain medications; and on biopsychosocial items related to chronic pain.
[DOCX File , 32 KB - mhealth_v12i1e54579_app3.docx]

Multimedia Appendix 4
Characteristics of chronic pain–related comorbidities.
[DOCX File , 14 KB - mhealth_v12i1e54579_app4.docx]

Multimedia Appendix 5
Distribution of pain locations.
[DOCX File , 16 KB - mhealth_v12i1e54579_app5.docx]

Multimedia Appendix 6
Characteristics of chronic pain and daily impact. Graphical representation of the characteristics of chronic pain and its daily impact defined by the assessment of pain duration, frequency, presence of paroxysms, impact on sleep, and type of chronic pain among all included patients.
[PNG File , 170 KB - mhealth_v12i1e54579_app6.png]

Multimedia Appendix 7

Characteristics of pain medications. Graphical representation of the characteristics of pain medications (medicines and nonmedicines) according to the percentage of patients treated among all included patients.
[PNG File , 96 KB - mhealth_v12i1e54579_app7.png]

Multimedia Appendix 8
Medical benefits and adverse events of pain medications.
[DOCX File , 14 KB - mhealth_v12i1e54579_app8.docx]

Multimedia Appendix 9
Characteristics of medical and lifestyle history. Graphical representation of the distribution of medical and lifestyle history among all included patients.
[PNG File , 501 KB - mhealth_v12i1e54579_app9.png]

Multimedia Appendix 10
Adherence to the eDOL tool. Graphical representation of patients' adherence to the eDOL tool reflected in the completion of their questionnaires (A) and weekly meters (B) at inclusion and at 3, 6, 9, and 12 months of follow-up.
[PNG File , 26 KB - mhealth_v12i1e54579_app10.png]

Multimedia Appendix 11
Acceptability questionnaire data. French and English versions of the acceptability questionnaire used to assess patient satisfaction.
[PNG File , 556 KB - mhealth_v12i1e54579_app11.png]

References

1. van Hecke O, Torrance N, Smith BH. Chronic pain epidemiology and its clinical relevance. Br J Anaesth 2013 Jul;111(1):13-18 [FREE Full text] [doi: 10.1093/bja/aet123] [Medline: 23794640]
2. Chenaf C, Delorme J, Delage N, Ardid D, Eschalier A, Authier N. Prevalence of chronic pain with or without neuropathic characteristics in France using the capture-recapture method: a population-based study. Pain 2018 Nov;159(11):2394-2402. [doi: 10.1097/j.pain.0000000000001347] [Medline: 30028790]
3. GBD 2016 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet 2017 Sep 16;390(10100):1211-1259 [FREE Full text] [doi: 10.1016/S0140-6736(17)32154-2] [Medline: 28919117]
4. Breivik H, Collett B, Ventafridda V, Cohen R, Gallacher D. Survey of chronic pain in Europe: Prevalence, impact on daily life, and treatment. European Journal of Pain 2012 Jan 11;10(4):287-287. [doi: 10.1016/J.EJPAIN.2005.06.009]
5. Attal N, Lanteri-Minet M, Laurent B, Fermanian J, Bouhassira D. The specific disease burden of neuropathic pain: results of a French nationwide survey. Pain 2011 Dec;152(12):2836-2843. [doi: 10.1016/j.pain.2011.09.014] [Medline: 22019149]
6. Societal impact of pain costs the EU up to 441 billion euros annually. The Societal Impact of Pain (SIP). 2017. URL: https://www.businesswire.com/news/home/20170519005141/en/Societal-Impact-of-Pain-Costs-the-EU-up-to-441-Billion-Euros-Annually [accessed 2024-04-14]
7. Gaskin DJ, Richard P. The economic costs of pain in the United States. J Pain 2012 Aug;13(8):715-724 [FREE Full text] [doi: 10.1016/j.jpain.2012.03.009] [Medline: 22607834]
8. Breivik H, Eisenberg E, O'Brien T, OPENMinds. The individual and societal burden of chronic pain in Europe: the case for strategic prioritisation and action to improve knowledge and availability of appropriate care. BMC Public Health 2013 Dec 24;13(1):1229 [FREE Full text] [doi: 10.1186/1471-2458-13-1229] [Medline: 24365383]
9. Global medicine spending and usage trends: outlook to 2025. IQVIA. 2021. URL: https://www.iqvia.com/insights/the-iqvia-institute/reports/global-medicine-spending-and-usage-trends-outlook-to-2025 [accessed 2021-04-29]
10. Crofford LJ. Adverse effects of chronic opioid therapy for chronic musculoskeletal pain. Nat Rev Rheumatol 2010 Apr;6(4):191-197. [doi: 10.1038/nrrheum.2010.24] [Medline: 20357788]
11. Mogil JS. Animal models of pain: progress and challenges. Nat Rev Neurosci 2009 Apr;10(4):283-294. [doi: 10.1038/nrn2606] [Medline: 19259101]
12. Heron KE, Smyth JM. Ecological momentary interventions: incorporating mobile technology into psychosocial and health behaviour treatments. Br J Health Psychol 2010 Feb;15(Pt 1):1-39 [FREE Full text] [doi: 10.1348/135910709X466063] [Medline: 19646331]
13. Jamison RN, Jurcik DC, Edwards RR, Huang C, Ross EL. A Pilot Comparison of a Smartphone App With or Without 2-Way Messaging Among Chronic Pain Patients: Who Benefits From a Pain App? Clin J Pain 2017 Aug;33(8):676-686 [FREE Full text] [doi: 10.1097/AJP.0000000000000455] [Medline: 27898460]

14. Dear BF, Gandy M, Karin E, Staples LG, Johnston L, Fogliati VJ, et al. The Pain Course: a randomised controlled trial examining an internet-delivered pain management program when provided with different levels of clinician support. Pain 2015 Oct;156(10):1920-1935 [FREE Full text] [doi: 10.1097/j.pain.0000000000000251] [Medline: 26039902]
15. Suso-Ribera C, Castilla D, Zaragozá I, Ribera-Canudas MV, Botella C, García-Palacios A. Validity, reliability, feasibility, and usefulness of pain monitor: a multidimensional smartphone app for daily monitoring of adults with heterogenous chronic pain. Clin J Pain 2018 Oct;34(10):900-908. [doi: 10.1097/AJP.0000000000000618] [Medline: 29659375]
16. Shadd JD, Ryan BL, Maddocks HL, McKay SD, Moulin DE. Neuropathic pain in a primary care electronic health record database. Eur J Pain 2015 May;19(5):715-721. [doi: 10.1002/ejp.594] [Medline: 25504680]
17. Minen MT, Jalloh A, Ortega E, Powers SW, Sevick MA, Lipton RB. User Design and Experience Preferences in a Novel Smartphone Application for Migraine Management: A Think Aloud Study of the RELAXaHEAD Application. Pain Med 2019 Feb 01;20(2):369-377 [FREE Full text] [doi: 10.1093/pm/pny080] [Medline: 29868895]
18. Jamison RN, Mei A, Ross EL. Longitudinal trial of a smartphone pain application for chronic pain patients: predictors of compliance and satisfaction. J Telemed Telecare 2018 Feb;24(2):93-100. [doi: 10.1177/1357633X16679049] [Medline: 27831496]
19. Stenner P, Cross V, McCrum C, McGowan J, Defever E, Lloyd P, et al. Self-management of chronic low back pain: four viewpoints from patients and healthcare providers. Health Psychol Open 2015 Jul;2(2):2055102915615337 [FREE Full text] [doi: 10.1177/2055102915615337] [Medline: 28070378]
20. Coleman MT, Newton KS. Supporting self-management in patients with chronic illness. Am Fam Physician 2005 Oct 15;72(8):1503-1510 [FREE Full text] [Medline: 16273817]
21. Rodríguez Sánchez-Laulhé P, Luque-Romero LG, Barrero-García F, Biscarri-Carbonero Á, Blanquero J, Suero-Pineda A, et al. An Exercise and Educational and Self-management Program Delivered With a Smartphone App (CareHand) in Adults With Rheumatoid Arthritis of the Hands: Randomized Controlled Trial. JMIR Mhealth Uhealth 2022 Apr 07;10(4):e35462 [FREE Full text] [doi: 10.2196/35462] [Medline: 35389367]
22. Anan T, Kajiki S, Oka H, Fujii T, Kawamata K, Mori K, et al. Effects of an artificial intelligence-assisted health program on workers with neck/shoulder pain/stiffness and low back pain: randomized controlled trial. JMIR Mhealth Uhealth 2021 Sep 24;9(9):e27535 [FREE Full text] [doi: 10.2196/27535] [Medline: 34559054]
23. Fatoye F, Gebrye T, Fatoye C, Mbada CE, Olaoye MI, Odole AC, et al. The clinical and cost-effectiveness of telerehabilitation for people with nonspecific chronic low back pain: randomized controlled trial. JMIR Mhealth Uhealth 2020 Jun 24;8(6):e15375 [FREE Full text] [doi: 10.2196/15375] [Medline: 32357128]
24. Anderson K, Burford O, Emmerton L. Mobile health apps to facilitate self-care: a qualitative study of user experiences. PLoS One 2016;11(5):e0156164 [FREE Full text] [doi: 10.1371/journal.pone.0156164] [Medline: 27214203]
25. Gogovor A, Visca R, Auger C, Bouvrette-Leblanc L, Symeonidis I, Poissant L, et al. Informing the development of an Internet-based chronic pain self-management program. Int J Med Inform 2017 Jan;97:109-119. [doi: 10.1016/j.ijmedinf.2016.10.005] [Medline: 27919369]
26. McGuire BE, Henderson EM, McGrath PJ. Translating e-pain research into patient care. Pain 2017 Feb 10;158(2):190-193. [doi: 10.1097/j.pain.0000000000000686] [Medline: 27525833]
27. Sundararaman LV, Edwards RR, Ross EL, Jamison RN. Integration of mobile health technology in the treatment of chronic pain: a critical review. Reg Anesth Pain Med 2017;42(4):488-498. [doi: 10.1097/AAP.0000000000000621] [Medline: 28570436]
28. Martin CL, Bakker CJ, Breth MS, Gao G, Lee K, Lee MA, et al. The efficacy of mobile health interventions used to manage acute or chronic pain: a systematic review. Res Nurs Health 2021 Feb;44(1):111-128. [doi: 10.1002/nur.22097] [Medline: 33341989]
29. Suso-Ribera C, Castilla D, Zaragozá I, Mesas Á, Server A, Medel J, et al. Telemonitoring in chronic pain management using smartphone apps: a randomized controlled trial comparing usual assessment against app-based monitoring with and without clinical alarms. Int J Environ Res Public Health 2020 Sep 09;17(18):17186568 [FREE Full text] [doi: 10.3390/ijerph17186568] [Medline: 32916983]
30. Ross EL, Jamison RN, Nicholls L, Perry BM, Nolen KD. Clinical integration of a smartphone app for patients with chronic pain: retrospective analysis of predictors of benefits and patient engagement between clinic visits. J Med Internet Res 2020 Apr 16;22(4):e16939 [FREE Full text] [doi: 10.2196/16939] [Medline: 32297871]
31. Mariano TY, Wan L, Edwards RR, Lazaridou A, Ross EL, Jamison RN. Online group pain management for chronic pain: preliminary results of a novel treatment approach to teletherapy. J Telemed Telecare 2021 May;27(4):209-216. [doi: 10.1177/1357633X19870369] [Medline: 31431133]
32. Solomon DH, Rudin RS. Digital health technologies: opportunities and challenges in rheumatology. Nat Rev Rheumatol 2020 Sep 24;16(9):525-535. [doi: 10.1038/s41584-020-0461-x] [Medline: 32709998]
33. Alhussein G, Hadjileontiadis L. Digital Health Technologies for Long-term Self-management of Osteoporosis: Systematic Review and Meta-analysis. JMIR Mhealth Uhealth 2022 Apr 21;10(4):e32557 [FREE Full text] [doi: 10.2196/32557] [Medline: 35451968]

34. Moreno-Ligero M, Moral-Munoz JA, Salazar A, Failde I. mHealth intervention for improving pain, quality of life, and functional disability in patients with chronic pain: systematic review. JMIR Mhealth Uhealth 2023 Feb 02;11:e40844 [FREE Full text] [doi: 10.2196/40844] [Medline: 36729570]
35. Gordon K, Rice H, Allcock N, Bell P, Dunbar M, Gilbert S, et al. Barriers to self-management of chronic pain in primary care: a qualitative focus group study. Br J Gen Pract 2017 Mar;67(656):e209-e217 [FREE Full text] [doi: 10.3399/bjgp17X688825] [Medline: 27993899]
36. Currie M, Philip LJ, Roberts A. Attitudes towards the use and acceptance of eHealth technologies: a case study of older adults living with chronic pain and implications for rural healthcare. BMC Health Serv Res 2015 Apr 16;15:162 [FREE Full text] [doi: 10.1186/s12913-015-0825-0] [Medline: 25888988]
37. Alexander JC, Joshi GP. Smartphone applications for chronic pain management: a critical appraisal. J Pain Res 2016;9:731-734 [FREE Full text] [doi: 10.2147/JPR.S119966] [Medline: 27713649]
38. Kerckhove N, Delage N, Cambier S, Cantagrel N, Serra E, Marcaillou F, et al. eDOL mHealth App and Web Platform for Self-monitoring and Medical Follow-up of Patients With Chronic Pain: Observational Feasibility Study. JMIR Form Res 2022 Mar 02;6(3):e30052 [FREE Full text] [doi: 10.2196/30052] [Medline: 35234654]
39. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA 2013 Nov 27;310(20):2191-2194. [doi: 10.1001/jama.2013.281053] [Medline: 24141714]
40. Article R5121-13 - Code de la santé publique. Légifrance. URL: https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000025788178 [accessed 2023-10-17]
41. Article 226-13 - Code pénal. Légifrance. URL: https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000006417945/ [accessed 2023-10-17]
42. Thayaparan AJ, Mahdi E. The Patient Satisfaction Questionnaire Short Form (PSQ-18) as an adaptable, reliable, and validated tool for use in various settings. Med Educ Online 2013 Jul 23;18:21747 [FREE Full text] [doi: 10.3402/meo.v18i0.21747] [Medline: 23883565]
43. Boß L, Lehr D, Reis D, Vis C, Riper H, Berking M, et al. Reliability and Validity of Assessing User Satisfaction With Web-Based Health Interventions. J Med Internet Res 2016 Aug 31;18(8):e234 [FREE Full text] [doi: 10.2196/jmir.5952] [Medline: 27582341]
44. Larsen DL, Attkisson CC, Hargreaves WA, Nguyen TD. Assessment of client/patient satisfaction: development of a general scale. Eval Program Plann 1979;2(3):197-207. [doi: 10.1016/0149-7189(79)90094-6] [Medline: 10245370]
45. Von Korff M, Scher AI, Helmick C, Carter-Pokras O, Dodick DW, Goulet J, et al. United States National Pain Strategy for Population Research: Concepts, Definitions, and Pilot Data. J Pain 2016 Oct;17(10):1068-1080 [FREE Full text] [doi: 10.1016/j.jpain.2016.06.009] [Medline: 27377620]
46. Valentijn PP, Tymchenko L, Jacobson T, Kromann J, Biermann CW, AlMoslemany MA, et al. Digital Health Interventions for Musculoskeletal Pain Conditions: Systematic Review and Meta-analysis of Randomized Controlled Trials. J Med Internet Res 2022 Sep 06;24(9):e37869 [FREE Full text] [doi: 10.2196/37869] [Medline: 36066943]
47. Stoppok P, Frewer A, Schweda A, Geiger S, Skoda E, Müßgens D, et al. Needs and Demands for eHealth Pain Management Interventions in Chronic Pain Patients. J Pers Med 2023 Apr 17;13(4):675 [FREE Full text] [doi: 10.3390/jpm13040675] [Medline: 37109061]
48. Buhrman M, Gordh T, Andersson G. Internet interventions for chronic pain including headache: A systematic review. Internet Interv 2016 May;4:17-34 [FREE Full text] [doi: 10.1016/j.invent.2015.12.001] [Medline: 30135787]
49. Pratap A, Neto EC, Snyder P, Stepnowsky C, Elhadad N, Grant D, et al. Indicators of retention in remote digital health studies: a cross-study evaluation of 100,000 participants. NPJ Digit Med 2020 Feb 17;3(1):21 [FREE Full text] [doi: 10.1038/s41746-020-0224-8] [Medline: 32128451]
50. Manderlier A, de Fooz M, Patris S, Berquin A. Modifiable lifestyle-related prognostic factors for the onset of chronic spinal pain: A systematic review of longitudinal studies. Ann Phys Rehabil Med 2022 Nov;65(6):101660 [FREE Full text] [doi: 10.1016/j.rehab.2022.101660] [Medline: 35351652]
51. Valsø Å, Rustøen T, Småstuen M, Puntillo K, Skogstad L, Schou-Bredal I, et al. Occurrence and characteristics of pain after ICU discharge: A longitudinal study. Nurs Crit Care 2022 Sep;27(5):718-727. [doi: 10.1111/nicc.12701] [Medline: 34382725]
52. Damush TM, Kroenke K, Bair MJ, Wu J, Tu W, Krebs EE, et al. Pain self-management training increases self-efficacy, self-management behaviours and pain and depression outcomes. Eur J Pain 2016 Aug;20(7):1070-1078. [doi: 10.1002/ejp.830] [Medline: 26849410]
53. Bicego A, Monseur J, Collinet A, Donneau A, Fontaine R, Libbrecht D, et al. Complementary treatment comparison for chronic pain management: A randomized longitudinal study. PLoS One 2021;16(8):e0256001 [FREE Full text] [doi: 10.1371/journal.pone.0256001] [Medline: 34358272]
54. Kassinopoulos O, Vasiliou V, Karekla M. Overcoming challenges in adherence and engagement digital interventions: The development of the ALGEApp for chronic pain management. Internet Interv 2023 Apr;32:100611 [FREE Full text] [doi: 10.1016/j.invent.2023.100611] [Medline: 36910302]

55. Fernandes LG, Devan H, Fioratti I, Kamper SJ, Williams CM, Saragiotto BT. At my own pace, space, and place: a systematic review of qualitative studies of enablers and barriers to telehealth interventions for people with chronic pain. Pain 2022 Feb 01;163(2):e165-e181. [doi: 10.1097/j.pain.0000000000002364] [Medline: 34433776]
56. Mariano TY, Wan L, Edwards RR, Jamison RN. Online teletherapy for chronic pain: A systematic review. J Telemed Telecare 2021 May;27(4):195-208. [doi: 10.1177/1357633X19871746] [Medline: 31488004]
57. Cascella M, Marinangeli F, Vittori A, Scala C, Piccinini M, Braga A, et al. Open Issues and Practical Suggestions for Telemedicine in Chronic Pain. Int J Environ Res Public Health 2021 Nov 25;18(23):12416 [FREE Full text] [doi: 10.3390/ijerph182312416] [Medline: 34886140]
58. Azevedo LF, Costa-Pereira A, Mendonça L, Dias CC, Castro-Lopes JM. Epidemiology of chronic pain: a population-based nationwide study on its prevalence, characteristics and associated disability in Portugal. J Pain 2012 Aug;13(8):773-783 [FREE Full text] [doi: 10.1016/j.jpain.2012.05.012] [Medline: 22858343]
59. Köppen P, Dorner TE, Stein KV, Simon J, Crevenna R. Health literacy, pain intensity and pain perception in patients with chronic pain. Wien Klin Wochenschr 2018 Jan;130(1-2):23-30 [FREE Full text] [doi: 10.1007/s00508-017-1309-5] [Medline: 29322378]
60. Fayaz A, Croft P, Langford RM, Donaldson LJ, Jones GT. Prevalence of chronic pain in the UK: a systematic review and meta-analysis of population studies. BMJ Open 2016 Jun 20;6(6):e010364 [FREE Full text] [doi: 10.1136/bmjopen-2015-010364] [Medline: 27324708]
61. Larsson C, Hansson EE, Sundquist K, Jakobsson U. Chronic pain in older adults: prevalence, incidence, and risk factors. Scand J Rheumatol 2017 Jul;46(4):317-325. [doi: 10.1080/03009742.2016.1218543] [Medline: 27885914]
62. Andrews P, Steultjens M, Riskowski J. Chronic widespread pain prevalence in the general population: a systematic review. Eur J Pain 2018 Jan;22(1):5-18. [doi: 10.1002/ejp.1090] [Medline: 28815801]
63. Jackson T, Thomas S, Stabile V, Han X, Shotwell M, McQueen K. Prevalence of chronic pain in low-income and middle-income countries: a systematic review and meta-analysis. Lancet 2015 Apr 27;385 Suppl 2:S10. [doi: 10.1016/S0140-6736(15)60805-4] [Medline: 26313056]
64. Pratap A, Renn BN, Volponi J, Mooney SD, Gazzaley A, Arean PA, et al. Using mobile apps to assess and treat depression in Hispanic and Latino populations: fully remote randomized clinical trial. J Med Internet Res 2018 Aug 09;20(8):e10130 [FREE Full text] [doi: 10.2196/10130] [Medline: 30093372]
65. Anguera JA, Jordan JT, Castaneda D, Gazzaley A, Areán P. Conducting a fully mobile and randomised clinical trial for depression: access, engagement and expense. BMJ Innov 2016 Jan;2(1):14-21 [FREE Full text] [doi: 10.1136/bmjinnov-2015-000098] [Medline: 27019745]
66. Crouthamel M, Quattrocchi E, Watts S, Wang S, Berry P, Garcia-Gancedo L, et al. Using a ResearchKit Smartphone App to Collect Rheumatoid Arthritis Symptoms From Real-World Participants: Feasibility Study. JMIR Mhealth Uhealth 2018 Sep 13;6(9):e177 [FREE Full text] [doi: 10.2196/mhealth.9656] [Medline: 30213779]
67. Althubaiti A. Information bias in health research: definition, pitfalls, and adjustment methods. J Multidiscip Healthc 2016;9:211-217 [FREE Full text] [doi: 10.2147/JMDH.S104807] [Medline: 27217764]

Abbreviations

AE: adverse event
LIMOS: Laboratory of Informatics, Modelling and Optimization of the Systems

Edited by L Buis; submitted 27.11.23; peer-reviewed by G Alhussein, E Ross; comments to author 23.01.24; revised version received 13.03.24; accepted 27.03.24; published 12.06.24.

Please cite as:
Delage N, Cantagrel N, Soriot-Thomas S, Frost M, Deleens R, Ginies P, Eschalier A, Corteval A, Laveyssière A, Phalip J, Bertin C, Pereira B, Chenaf C, Doreau B, Authier N, ePAIN, Kerckhove N
Mobile Health App and Web Platform (eDOL) for Medical Follow-Up of Patients With Chronic Pain: Cohort Study Involving the French eDOL National Cohort After 1 Year
JMIR Mhealth Uhealth 2024;12:e54579
URL: https://mhealth.jmir.org/2024/1/e54579
doi:10.2196/54579
PMID:38865173

Original Paper

Identifying Weekly Trajectories of Pain Severity Using Daily Data From an mHealth Study: Cluster Analysis

Claire L Little[1], BSc, MA; David M Schultz[2,3], PhD; Thomas House[4], PhD; William G Dixon[1,5], PhD; John McBeth[1,5,6], PhD

[1]Centre for Epidemiology Versus Arthritis, University of Manchester, Manchester, United Kingdom

[2]Department of Earth and Environmental Sciences, Centre for Atmospheric Science, University of Manchester, Manchester, United Kingdom

[3]Centre for Crisis Studies and Mitigation, University of Manchester, Manchester, United Kingdom

[4]Department of Mathematics, University of Manchester, Manchester, United Kingdom

[5]NIHR Manchester Musculoskeletal Biomedical Research Unit, Central Manchester University Hospitals NHS Foundation Trust, Manchester, United Kingdom

[6]School of Primary Care, Population Sciences and Medical Education, University of Southamptom, Southampton, United Kingdom

Corresponding Author:
John McBeth, PhD
School of Primary Care, Population Sciences and Medical Education
University of Southamptom
University Road
Southampton, SO17 1BJ
United Kingdom
Phone: 1 02380595000
Email: john.mcbeth@southampton.ac.uk

Abstract

Background: People with chronic pain experience variability in their trajectories of pain severity. Previous studies have explored pain trajectories by clustering sparse data; however, to understand daily pain variability, there is a need to identify clusters of weekly trajectories using daily pain data. Between-week variability can be explored by quantifying the week-to-week movement between these clusters. We propose that future work can use clusters of pain severity in a forecasting model for short-term (eg, daily fluctuations) and longer-term (eg, weekly patterns) variability. Specifically, future work can use clusters of weekly trajectories to predict between-cluster movement and within-cluster variability in pain severity.

Objective: This study aims to understand clusters of common weekly patterns as a first stage in developing a pain-forecasting model.

Methods: Data from a population-based mobile health study were used to compile weekly pain trajectories (n=21,919) that were then clustered using a k-medoids algorithm. Sensitivity analyses tested the impact of assumptions related to the ordinal and longitudinal structure of the data. The characteristics of people within clusters were examined, and a transition analysis was conducted to understand the movement of people between consecutive weekly clusters.

Results: Four clusters were identified representing trajectories of *no or low pain* (1714/21,919, 7.82%), *mild pain* (8246/21,919, 37.62%), *moderate pain* (8376/21,919, 38.21%), and *severe pain* (3583/21,919, 16.35%). Sensitivity analyses confirmed the 4-cluster solution, and the resulting clusters were similar to those in the main analysis, with at least 85% of the trajectories belonging to the same cluster as in the main analysis. Male participants spent longer (participant mean 7.9, 95% bootstrap CI 6%-9.9%) in the *no or low pain* cluster than female participants (participant mean 6.5, 95% bootstrap CI 5.7%-7.3%). Younger people (aged 17-24 y) spent longer (participant mean 28.3, 95% bootstrap CI 19.3%-38.5%) in the *severe pain* cluster than older people (aged 65-86 y; participant mean 9.8, 95% bootstrap CI 7.7%-12.3%). People with fibromyalgia (participant mean 31.5, 95% bootstrap CI 28.5%-34.4%) and neuropathic pain (participant mean 31.1, 95% bootstrap CI 27.3%-34.9%) spent longer in the *severe pain* cluster than those with other conditions, and people with rheumatoid arthritis spent longer (participant mean 7.8, 95% bootstrap CI 6.1%-9.6%) in the *no or low pain* cluster than those with other conditions. There were 12,267 pairs of consecutive weeks that contributed to the transition analysis. The empirical percentage remaining in the same cluster across consecutive weeks was 65.96% (8091/12,267). When movement between clusters occurred, the highest percentage of movement was to an adjacent cluster.

Conclusions: The clusters of pain severity identified in this study provide a parsimonious description of the weekly experiences of people with chronic pain. These clusters could be used for future study of between-cluster movement and within-cluster variability to develop accurate and stakeholder-informed pain-forecasting tools.

(JMIR Mhealth Uhealth 2024;12:e48582) doi:10.2196/48582

KEYWORDS

mobile health; mHealth; pain; cluster; trajectory; k-medoids; transition; forecast; mobile phone

Introduction

Background

Chronic pain (ie, pain lasting ≥3 months) is a common symptom of many long-term health conditions [1,2] and is associated with poor quality of life, poor health outcomes, and low participation in work and social activities [3,4]. There is substantial day-to-day variability in the severity of pain experienced [5,6], and people with chronic pain report that this variability leads to feelings of frustration and uncertainty about future pain [7,8]. Studies have identified associations between pain variability and response to treatment [9] as well as lower quality of life [10,11]. However, pain variability remains underestimated by researchers [12].

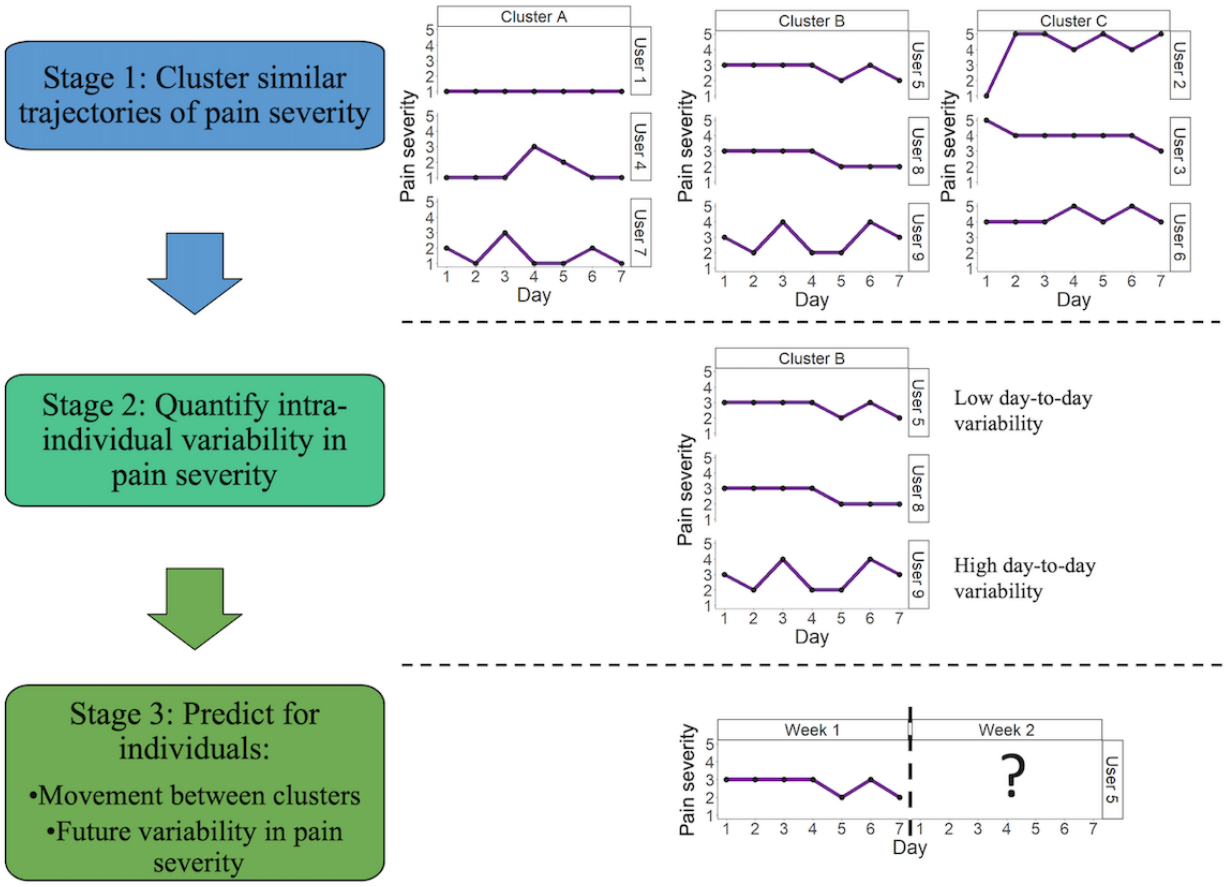
One way to explore pain variability is to cluster common pain trajectories and quantify movement between clusters. Previous studies have identified patterns of pain severity by clustering pain trajectories among individuals with chronic pain. These studies have often used sparse data on pain severity collected once per week [13], once per month [14], or less frequently [15]. These studies inform our understanding of longer-term experiences of chronic pain but not the day-to-day experience of pain severity that is important to patients. There is a need to extend this knowledge of pain clusters to within-week pain trajectories. Recent advances in mobile health (mHealth) methods that support the collection of data in the patients' own environments [16,17], often using their own devices (eg, smartphones and tablets) [18], offer the opportunity to capture daily pain severity data.

It is also possible to explore movement between clusters of pain data; for example, Rahman et al [19] used changes between pain severity scores (not necessarily day-to-day changes) to identify 2 clusters of low pain volatility and high pain volatility. The authors then predicted movement between these clusters at 6-month intervals. However, there is a need to explore movement between clusters on a shorter time frame.

Once identified, weekly pain trajectories could be forecast. People living with chronic pain have reported that a pain forecast would reduce unpredictability and could be used to support planning daily activities, such as shopping, chores, and social participation [20,21]. In a research prioritization study, 75% of the respondents to a survey said they would use a pain forecast and prioritized a model that predicted daily fluctuations (ie, relatively short-term variability) and pain flares (patterns across multiple days) [20].

We propose 3 stages to develop a pain forecast (Figure 1). Stage 1 identifies common weekly trajectories of pain severity using cluster analyses. Stage 2 investigates day-to-day variability in trajectories of pain severity for individuals within each cluster. These first 2 stages provide a better understanding of an individual's pain experiences. Stage 3 predicts for an individual their movement between clusters of pain severity across consecutive weeks and future within-cluster day-to-day variability. This study focuses on the first of these stages: clustering trajectories of pain severity. Understanding clusters of weekly trajectories is an important stage in this forecasting model to identify group-level associations that may be masked by population-level analysis.

Figure 1. Three stages to build a pain forecast using data from a mobile health study. Data used in this figure are for illustrative purposes only (to provide an example of how data may be used in the pipeline of developing a pain forecast). First, data are clustered to identify common trajectories of weekly pain severity. Second, the remaining variability is explored for each trajectory within a cluster. The process is repeated for each cluster. Third, movement between clusters on consecutive weeks and the amount of day-to-day variability are predicted for an individual. The process is repeated for each individual.



Once daily pain severity data are collected, there are several challenges to overcome in clustering these data for use in a pain-forecasting model. First, patient-generated health data are often collected on an ordinal scale. However, equal intervals between responses cannot be assumed, and using metric models to analyze ordinal data can lead to errors [22]. Second, data collected are longitudinal, and algorithms used for clustering should respect this longitudinal feature of the data. Third, clusters of pain severity that will be used in a pain-forecasting model should be interpretable to end users. To address these challenges, it is necessary to identify and use a suitable method for clustering patient-generated health data. Any assumptions made about the data should be tested in sensitivity analyses to ensure robustness. Observing substantial movement between clusters would suggest the feasibility of forecasting cluster movement in future work. Therefore, understanding the characteristics of individuals who contribute to different clusters and how these individuals move between clusters over time will aid end-user interpretability.

Objectives

The aim of this study was to understand pain severity clusters in people living with chronic pain. The specific objectives were to (1) use a suitable algorithm to identify the optimum number of clusters of pain trajectories, (2) conduct sensitivity analyses to test assumptions made when clustering data, (3) examine the characteristics of people within clusters, and (4) describe the movement of people between different clusters over time.

Methods

Data Source

This study is a secondary analysis of a population-based mHealth study, Cloudy with a Chance of Pain [16,23,24]. Study participants were recruited between January 2016 and January 2017 through advertisements on television, radio, and social media. Data collection ended in April 2017, with participants able to contribute data for between 0 and 15 months. The inclusion criteria were as follows: participants with chronic pain, aged ≥17 years, living in the United Kingdom, and owning an Android or iOS smartphone. Participants downloaded a co-designed mobile phone app; gave electronic consent; and provided demographic information, including their sex (male or female), year of birth (entered as free text), and pain conditions (selected from a list of predefined responses, eg, rheumatoid arthritis and fibromyalgia). Daily reports of 10 variables were collected, including pain severity. Participants were asked "How severe was your pain today?" They responded by selecting *no pain* (score=1), *mild pain* (score=2), *moderate pain* (score=3), *severe pain* (score=4), or *very severe pain* (score=5). Daily reports of other variables included mood, fatigue, and physical activity, but these were not included in

this secondary analysis. Data were collected locally on the smartphone, transferred to an external server where they were anonymized, and then returned to the researchers in anonymized form. Daily data could be contributed for a maximum of 15 months, with participants requested to track symptoms for 6 months. In total, 10,584 people downloaded the app and recorded their demographic information and at least 1 record of pain severity. Of the 10,584 participants, 8554 (80.82%) were female, with a mean age of 51 (SD 12.5) years. On average, these participants contributed pain severity data on 76 days (10,067/10,584, 95.12% of the participants contributed data on between 1 and 359 days). Previous analysis of these data classified participants as *highly engaged* (865/6370, 13.58%), *moderately engaged* (1384/6370, 21.73%), *less engaged* (2503/6370, 39.4%), and *tourists* (1618/6370, 25.4%) [24].

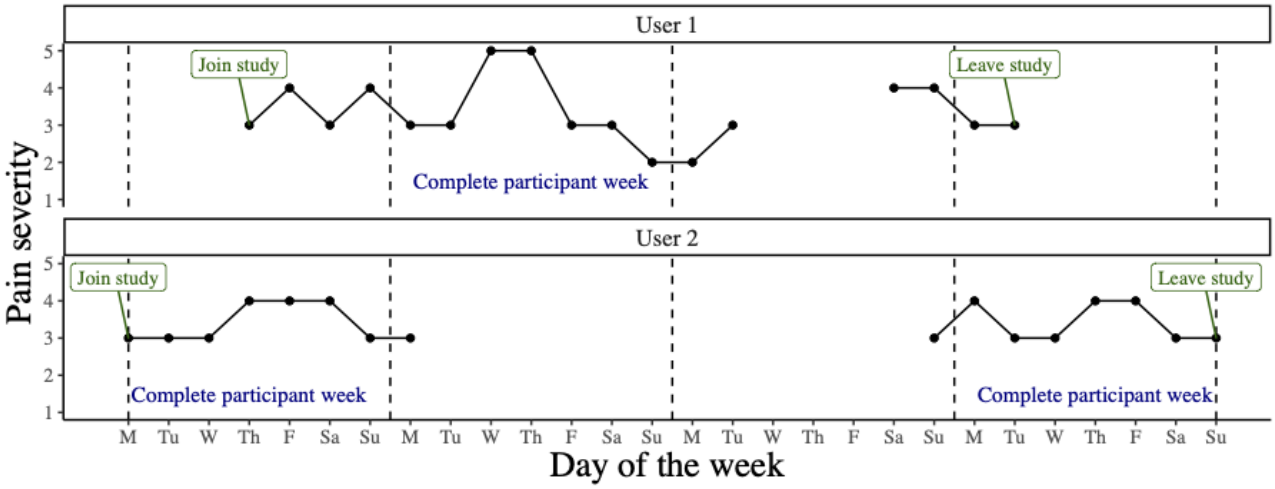
Ethical Considerations

Ethics approval for the Cloudy with a Chance of Pain mHealth study was obtained from the University of Manchester Research Ethics Committee (ethics/15522) and from the National Health Service Integrated Research Application System (23/NW/0716). Participants were required to provide electronic consent for study inclusion. Anonymized data were received by the research team. Further ethics approval was not required for the secondary analysis described in this study.

Data Preparation

For this study, weekly trajectories of pain severity data were used. To align data across multiple respondents, trajectories beginning on a Monday were identified. This alignment introduced a structure to the data based on the work week to mitigate the impact of individuals entering the study at different times and to deal with day-of-the-week effects. A complete participant week was defined as complete pain severity data contributed by a single participant during a single calendar week (Monday-Sunday; Figure 2). Pain severity data from a complete participant week were included in the analysis if (1) the participant had joined the study on or before the Monday, (2) the participant had remained in the study on or after the following Sunday, and (3) the participant had provided complete pain severity data (ie, 1 pain severity score on each of the 7 days). Multiple complete participant weeks could be included in the analysis for each participant (up to 64 weeks due to the length of the study).

Figure 2. Example selections of complete trajectory weeks for 2 participants. The participants join and leave the study at different times. One complete participant week from user 1 is included in the analysis. Two complete participant weeks from user 2 are included in the analysis.



Statistical Methods

Identifying the Optimal Number of Clusters

Previous studies have used a range of methods to cluster pain severity, including k-means clustering [25,26], hierarchical cluster analysis [27], growth mixture modeling [28-31], latent class growth analysis [13-15,32,33], multilevel latent class analysis [34], and group-based trajectory analysis [35-40]. Different approaches have different strengths and kinds of assumptions; for example, some may assume that clusters are internally homogeneous, while others may assume that the data overall follow a particular (eg, linear) form, are continuous, or are similar. In fact, clustering of ordinal and longitudinal data using a model that explicitly represents these features in a computationally inexpensive way remains a major unsolved methodological challenge. In this study, we chose a method that does not make strong assumptions about the form or generating mechanism of the data or the within-cluster variance and maintains the assumption that the data are ordinal in nature while having excellent computational performance and convergence properties. However, the assumption about the longitudinal nature of the data is relaxed.

To identify the optimal number of clusters, data were summarized in feature vectors, compared using the Manhattan (ℓ_1) distance measure, and clustered using an adaptation of the k-medoids algorithm, detailed herein. The feature vectors were 7D, with entries representing the pain severity data on each of the 7 days in a complete participant week. Using the data directly in this way ensured that feature vectors remained interpretable. The differences between feature vectors were found by calculating the Manhattan distance through entry-wise summation of absolute differences to respect the ordinal nature of the outcome variable. The implementation of the k-medoids algorithm used to cluster the feature vectors can be derived as follows. A k-medoids algorithm randomly assigns user-defined k feature vectors to be the cluster centers (or medoids) and then

iteratively (1) assigns each feature vector to the closest medoid and (2) recalculates the medoid of the clusters. The term *medoid* refers to the use of actual data points as the centers for the clusters [41]. Such use of observed data as centers for the clusters prevents outputs such as "pain severity of 3.2" that might arise if means are used and that are uninterpretable and erroneously assume an interval scale. To implement the k-medoids algorithm, the Clustering Large Applications (CLARA) program was used, which was specifically designed to be used with large data sets to reduce overall computation time [42,43].

A k-medoids algorithm requires a user-defined value for the number of clusters (k) in the data [41]. The implementation of the CLARA program was therefore repeated for values of k from 1 to 20. The output of the algorithm can be sensitive to the random feature vectors selected as the medoids in the first stage of the algorithm. The algorithm was therefore repeated 20 times, once for each value of k. At each iteration, the remaining variance within each cluster was calculated as the within-cluster sum of squares (WSS). The WSS calculates the total remaining distance between pairs of feature vectors in the same cluster. For each value of k, the iteration that returns the smallest value of WSS is selected and reported on a plot.

The optimal number of clusters was then selected using 3 criteria. First, from the plot of k against WSS, the optimal number of clusters was chosen visually using the elbow method [44]. While ideally a formal trade-off would be made between model complexity and goodness of fit, there is no clear method to use. Existing methods (eg, information criteria, silhouette method, and gap statistic) can suggest different numbers of clusters [45], possibly due to underpenalizing the complexity of data sets of the size used in this study. Therefore, the less formal elbow method allows us to be more explicit in the judgments we make to resolve the absence of an unambiguous method for learning cluster numbers from data. Second, clusters were required to contain 5% of the trajectories, similar to previous studies [13,36,46-49]. Third, cluster solutions were examined for clinical interpretability. For this measure, candidate solutions were examined to ensure meaningful differences between the cluster medoids. Furthermore, the distribution of the demographic data of participants contributing trajectories to each cluster were examined to ensure that the results reflected expected distributions.

Sensitivity Analyses

Overview

Three sensitivity analyses were conducted to test assumptions made in the main analysis, with the methodology behind choosing these being to modify assumptions made about the data and to see whether the broad conclusions were robust. Robust conclusions would be indicative of a strong model-independent signal in the data, even if modified assumptions led to a less interpretable output. The main analysis assumed that data were on an ordered scale but relaxed the assumption that the data were longitudinal.

Sensitivity Analysis 1

The first sensitivity analysis maintained the longitudinal nature of the data but implicitly assumed that the outcome variable was on a continuous scale. Feature vectors were compared using the Euclidean distance, which erroneously assumes regular intervals between values on the pain severity scale. However, the use of the Euclidean distance permits the use of the *KmL* package, which specifically clusters longitudinal data [50]. The *KmL* package is an adaptation of the k-means algorithm. The k-means algorithm is similar to k-medoids, but the center of each cluster is calculated using the mean of the feature vectors assigned to the cluster. The use of the *KmL* package, instead of the CLARA program, and the resulting use of mean trajectories rather than medoid trajectories were the only adaptions to this sensitivity analysis. The feature vectors, the 20 repetitions of the algorithm for each value of k, and the use of the elbow method to select k remained unchanged.

Sensitivity Analysis 2

The second sensitivity analysis relaxed assumptions about the longitudinal nature as well as the ordinal nature of the outcome variable. In this sensitivity analysis, the data were not assumed to be longitudinal, and the outcome variable was assumed to be unordered categorical data. A different feature vector was used that converted ordinal pain severity values into dummy variables using one-hot encoding. In this encoding, there were 35 binary categories, each representing a unique day and pain severity category. The feature was recorded as *1* if the pain severity score was seen on that day and *0* otherwise. In this way, 7 of the features were recorded as *1* for each complete participant week. The feature vectors were compared using the Jaccard distance, typically used for such vectors of binary data. The cluster analysis was then conducted using the CLARA program in the same manner as described in the main analysis.

Sensitivity Analysis 3

The third sensitivity analysis challenged the definition of a Monday-to-Sunday week when defining complete participant weeks. Instead, the following analysis was conducted for each day (*D*) in the week. Complete participant weeks were selected from the original data for each participant when there were pain severity data for each day in the *D*-to-*D*+6 week (eg, Wednesday-Tuesday week). On each new data set corresponding to a different *D*, clustering was conducted using the CLARA program at each value of k between 1 and 20, as described in the main analysis. Due to the adapted complete participant weeks, individuals may have contributed different numbers of weeks to the sensitivity analysis.

Calculations for Each Sensitivity Analysis

For each sensitivity analysis, the optimal number of clusters was calculated. Similar numbers and descriptions of the clusters would provide evidence that the conclusions from the main analysis are robust. Furthermore, for each cluster in the main analysis, the proportion of trajectories allocated to the same cluster in each sensitivity analysis was calculated. A high proportion would further suggest that the results are robust to the assumptions made by using the Manhattan distance and k-medoids algorithm in the main analysis.

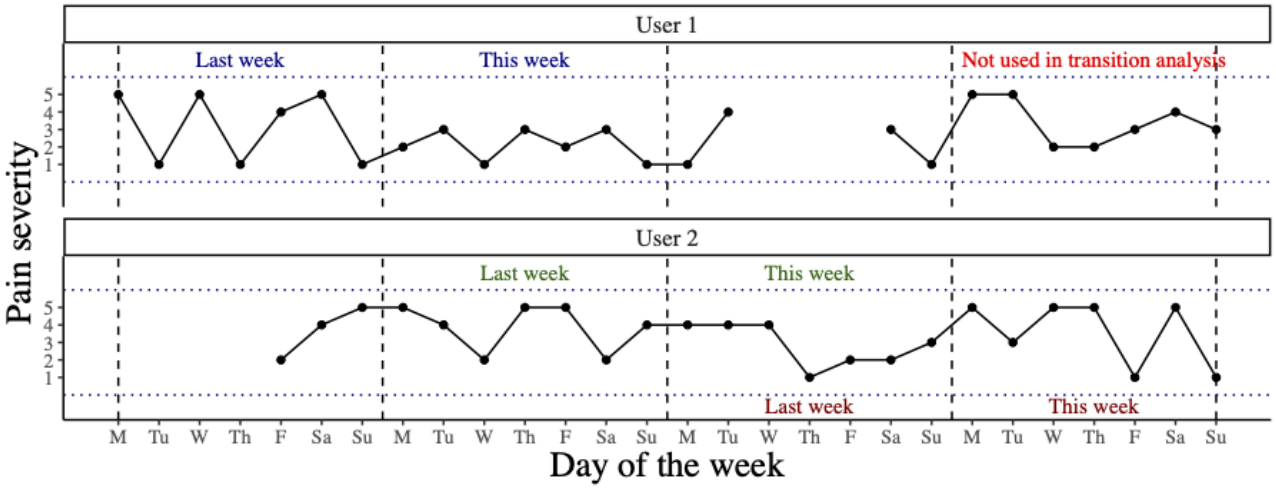
Description of Clusters

Information about the trajectories assigned to each of the clusters in the optimal solution was summarized. First, the number of trajectories assigned to each cluster was reported. Second, the clusters were visualized with a spaghetti plot of individual trajectories and the medoid of each cluster. Finally, the average proportion of time spent in each cluster by each participant was calculated. This information was summarized by calculating the mean proportion of time spent in each cluster across demographics (ie, age, sex, and chronic pain condition or conditions).

Transition Between Clusters

For the optimal solution of clusters in the main analysis, the transition of individuals between clusters on consecutive weeks was examined. To do this, a subset of the total data was used. Complete participant weeks (*this week*) were retained if the participant had also contributed a complete participant week in the directly preceding week (*last week*). A trajectory could be labeled as both *this week* and *last week* if there were both preceding and succeeding weeks for the individual (Figure 3). The demographic data of participants included in this transition analysis were compared to those included in the main analysis.

Figure 3. Example data from 2 participants highlighting how their data are used to examine transitions between clusters. User 1 provided data in 3 complete participant weeks. The first 2 are consecutive and therefore are used in the transition analysis. The final complete participant week is not used. User 2 provided 3 complete participant weeks. All three are used in the transition analysis. The middle week is labeled as both this week and last week in different pairings.



Each trajectory was assigned a cluster in the CLARA program of the k-medoids cluster algorithm. The transition probabilities were then calculated as follows. For all trajectories in each cluster *last week*, the percentage of trajectories that transitioned to each cluster *this week* were calculated. The resulting percentages are reported in a transition matrix.

Data were analyzed in R (version 4.1.2; R Foundation for Statistical Computing). The reporting of the analysis followed the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines [51].

Results

Data Source

There were 2807 participants who contributed 21,919 participant weeks of data to this analysis. The participants' mean age was 51.2 (SD 12.8) years, and 83.11% (2333/2807) were female. Table 1 reports the number of participants by age, sex, chronic pain condition, and the average number of participant weeks contributed to the analysis by members of the subgroup. Overall, older participants contributed a greater number of participant weeks than younger participants. Male participants contributed slightly more (8.1) participant weeks than female participants (7.7). Participants with osteoarthritis (9.1) and unspecific arthritis (9.0) contributed the highest number of participant weeks, and participants with chronic headache (6.0) contributed the fewest participant weeks. Comorbid conditions are described in Tables S1 and S2 in Multimedia Appendix 1.

Table 1. Demographic information of the participants who contributed to the analysis and the average number of participant weeks contributed by each subgroup (n=2807).

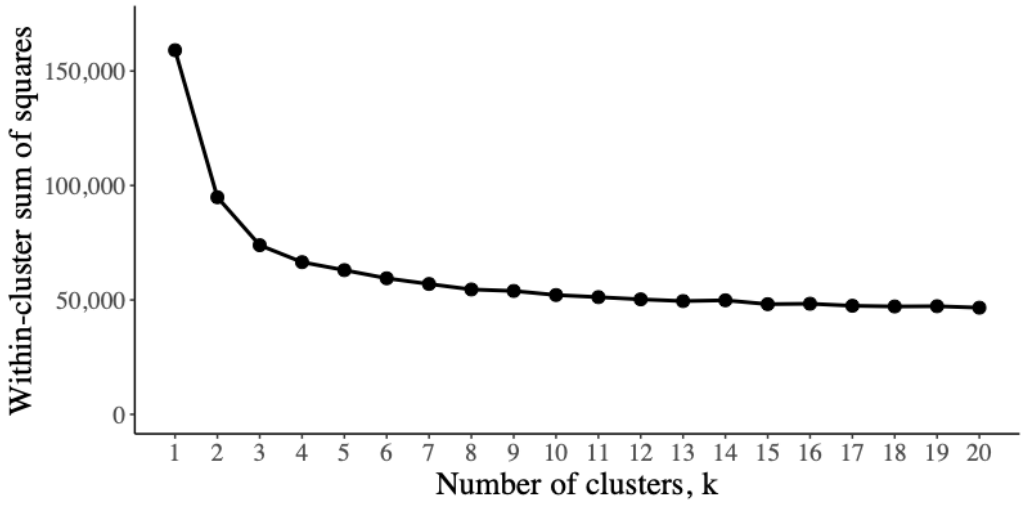
Demographic information	Participants, n (%)	Weekly trajectories contributed by participants, mean (SD)
Age group (y)		
17-24	67 (2.39)	5.2 (8.6)
25-34	255 (9.08)	5.6 (7.1)
35-44	508 (18.1)	6.8 (9.0)
45-54	755 (26.9)	7.5 (9.1)
55-64	788 (28.07)	8.6 (9.7)
65-86	434 (15.46)	9.9 (10.5)
Sex		
Female	2333 (83.11)	7.7 (9.4)
Male	474 (16.89)	8.1 (9.4)
Chronic pain condition[a]		
Rheumatoid arthritis	548 (19.52)	7.7 (8.6)
Osteoarthritis	975 (34.73)	9.1 (10.3)
Spondyloarthropathy	254 (9.05)	7.6 (8.6)
Gout	96 (3.42)	7.8 (10.6)
Unspecific arthritis	1028 (36.62)	9.0 (10.3)
Fibromyalgia	718 (25.58)	7.1 (8.9)
Chronic headache	274 (9.76)	6.0 (6.5)
Neuropathic pain	427 (15.21)	7.5 (9.7)
Other or no medical diagnosis	668 (23.8)	6.9 (8.9)

[a]Percentages exceed 100% because participants could report multiple chronic pain conditions.

Identifying the Optimal Number of Clusters

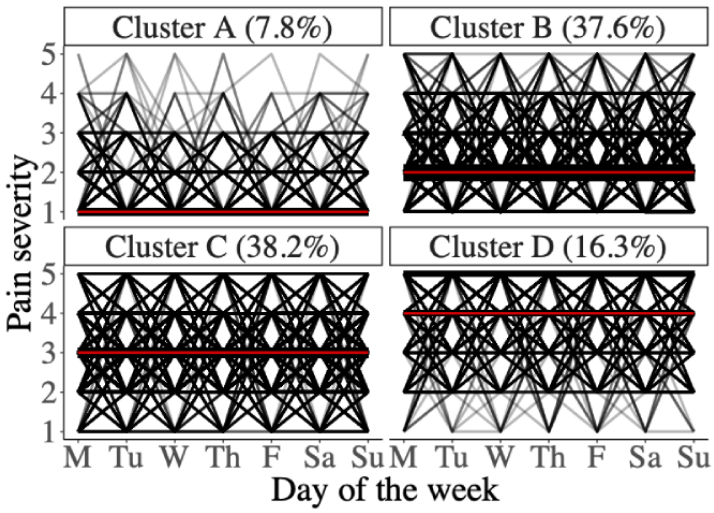
The results of the CLARA algorithm are shown herein. Figure 4 reports the remaining variability within clusters as the WSS at each value of k. There is an elbow at k=4, suggesting that most of the observed variability can be explained by a solution with 4 clusters, with diminishing returns for including further clusters in the solution. Four clusters reduced the WSS from 159,100 to 66,507; therefore, the clustering algorithm describes 58.2% of the variability in the data. Each cluster contained >5% of the pain trajectories. Therefore, 4 clusters provide an appropriate choice for these data.

Figure 4. Unexplained variability across different cluster (k) solutions. The within-cluster sum of squares indicates the remaining variance within clusters. An elbow at k=4 suggests an appropriate solution, with diminishing returns for the inclusion of further clusters.



The trajectories in each cluster are shown in the spaghetti plot in Figure 5. Trajectories are weighted such that thicker lines represent a higher number of trajectories following the path. The red line represents the medoid of the k-medoids algorithm. The clusters can be named by examining the medoid: A=*no or low pain*, B=*mild pain*, C=*moderate pain*, and D=*severe pain*. Of the 21,919 trajectories, cluster A contained 1714 (7.82%), cluster B contained 8246 (37.62%), cluster C contained 8376 (38.2%), and cluster D contained 3583 (16.35%).

Figure 5. Weighted spaghetti plot of trajectories assigned to each cluster. The weight (and transparency) of each path represents the number of trajectories following that path. The red line represents the medoid of the cluster. Cluster A=no or low pain, cluster B=mild pain, cluster C=moderate pain, and cluster D=severe pain. The percentage of trajectories assigned to each cluster is shown.

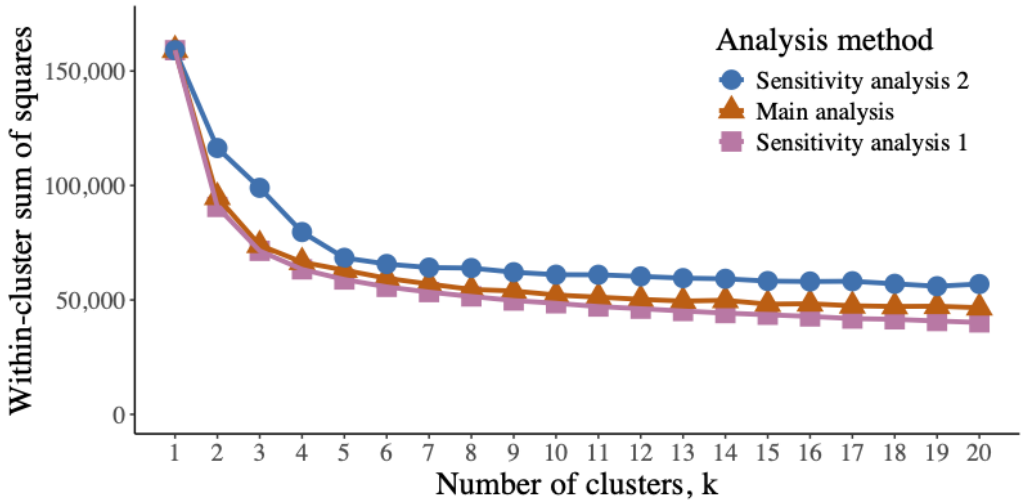


Sensitivity Analyses

Sensitivity Analysis 1 (KmL Algorithm and Euclidean Distance)

Full results are presented in Multimedia Appendix 2. The plot visualizing WSS against k for this analysis is similar to that of the main analysis and has an elbow at k=4 (Figure 6). The optimal 4-cluster solution describes 60% of the observed variability. The descriptions of the spaghetti plots (ie, cluster A=*no or low pain*, cluster B=mild pain, cluster C=*moderate pain*, and cluster D=*severe pain*) are the same as those in the main analysis, despite the use of a mean rather than a medoid to describe the average trajectory in each cluster. Of the 21,919 trajectories, 18,895 (86.2%) were assigned to the same cluster as in the main analysis, indicating similar results. Clusters B and C remain the largest clusters (8484/21,919, 38.71% and 8001/21,919, 36.5% trajectories, respectively), although cluster A is larger in this sensitivity analysis than in the main analysis (2493/21,919, 11.37% vs 1714/21,919, 7.82%).

Figure 6. Unexplained variability across different cluster (k) solutions for the main analysis and 2 sensitivity analyses. In the main analysis and sensitivity analysis 1, there is an elbow at k=4. In sensitivity analysis 2, there is an elbow at k=5.



Sensitivity Analysis 2 (CLARA Algorithm and Jaccard Distance)

Full results are presented in Multimedia Appendix 2. The plot of k against WSS for this analysis has an elbow at k=5 (Figure 6). However, 1 cluster contained 990 (4.52%) of the 21,919 trajectories in the 5-cluster solution, which did not meet the criteria for cluster sizes >5%, and therefore a 4-cluster solution remained optimal in this analysis. A 4-cluster solution describes 50% of the variability. Spaghetti plots of the 4-cluster solution show the same descriptions as those in the main analysis. In total, 20,197 (92.14%) of the 21,919 trajectories were assigned to the same cluster as in the main analysis.

Sensitivity Analysis 3 (Day of the Week)

Full results are presented in Multimedia Appendix 3. Each plot of WSS against k suggested an optimal solution at k=4. The proportions of trajectories assigned to each cluster in each 4-cluster solution are similar to those in the main analysis. The proportions in cluster A ranged between 7.72% (1717/22,255) and 7.9% (1769/22,404), cluster B between 37.49% (8273/22,067) and 37.71% (8393/22,255), cluster C between 38.04% (8523/22,404) and 38.43% (8481/22,067), and cluster D between 16.19% (3614/22,320) and 16.42% (3679/22,404). These results show that the main analysis is robust to the day of the week on which the trajectories begin.

Description of Clusters

The average proportions of time spent in different clusters across different characteristics (age, sex, and condition) are summarized in Table 2. The participants in the oldest age bracket (65-86 y) spent less time (mean 9.8, 95% bootstrap CI 7.7%-12.2%) in the *severe pain* cluster than those in the youngest age bracket (17-24 y; mean 28.3, 95% bootstrap CI 19%-38%). Female participants spent more time in the *severe pain* cluster (mean 18, 95% bootstrap CI 16.6%-19.3%) than male participants (mean 12.3, 95% bootstrap CI 10%-14.7%) and less time in the *lowest pain* cluster (female participants: mean 6.5, 95% bootstrap CI 5.8%-7.3%); male participants: mean 7.9, 95% bootstrap CI 6.1%-10%). Participants with fibromyalgia and neuropathic pain spent the most time in the *severe pain* cluster (mean 31.5, 95% bootstrap CI 28.8%-34.5%) and (mean 31.1, 95% bootstrap CI 27.2%-35.1%, respectively). Participants with rheumatoid arthritis spent the most time in the *lowest pain* cluster (mean 7.8, 95% bootstrap CI 6%-9.6%).

Table 2. Percentage of time spent in each cluster by baseline characteristic (for each characteristic, the average percentage of time spent in each cluster by members of the characteristic is reported).

	Time spent in cluster A (%), mean (95% CI)	Time spent in cluster B (%), mean (95% CI)	Time spent in cluster C (%), mean (95% CI)	Time spent in cluster D (%), mean (95% CI)
All	6.7 (6.0-7.5)	36.8 (35.4-38.2)	39.4 (38.0-40.9)	17.0 (15.9-18.2)
Age group (y)				
17-24	3.4 (0.8-6.9)	28.5 (19.9-38.1)	39.8 (30.6-49.3)	28.3 (19.0-38.0)
25-34	6.7 (4.3-9.4)	32.0 (27.6-36.5)	41.5 (36.9-46.4)	19.8 (15.5-24.2)
35-44	5.3 (3.9-6.9)	31.5 (28.4-34.6)	38.6 (35.2-42.1)	24.5 (21.2-27.9)
45-54	5.6 (4.5-6.9)	36.3 (33.6-39.0)	39.4 (36.7-42.1)	18.7 (16.5-21.0)
55-64	8.4 (6.8-10.1)	38.5 (35.9-41.0)	40.4 (37.9-43.1)	12.7 (10.9-14.7)
65-86	7.9 (6.1-9.9)	44.9 (41.6-48.4)	37.3 (33.9-40.6)	9.8 (7.7-12.2)
Sex				
Female	6.5 (5.8-7.3)	35.6 (34.1-37.2)	39.9 (38.4-41.4)	18.0 (16.6-19.3)
Male	7.9 (6.1-10.0)	42.5 (39.1-46.1)	37.2 (34.0-40.5)	12.3 (10.0-14.7)
Chronic pain condition[a]				
Rheumatoid arthritis	7.8 (6.0-9.6)	38.9 (35.7-42.0)	39.5 (36.5-42.5)	13.8 (11.5-16.2)
Osteoarthritis	5.4 (4.4-6.5)	34.7 (32.5-37.0)	42.6 (40.2-44.9)	17.2 (15.2-19.3)
Spondyloarthropathy	4.1 (2.5-6.1)	31.8 (27.5-36.2)	43.5 (38.9-47.9)	20.6 (16.3-25.0)
Gout	6.1 (2.4-10.4)	33.0 (25.5-40.1)	41.6 (34.2-48.9)	19.3 (12.9-26.2)
Unspecific arthritis	6.3 (5.2-7.5)	39.0 (36.6-41.3)	38.5 (36.3-40.7)	16.2 (14.3-18.2)
Fibromyalgia	1.7 (1.0-2.4)	19.1 (17.0-21.3)	47.7 (45.0-50.5)	31.5 (28.8-34.5)
Chronic headache	5.1 (3.1-7.4)	28.7 (24.6-33.2)	40.3 (35.9-44.7)	25.9 (21.5-30.4)
Neuropathic pain	3.3 (2.0-4.7)	23.6 (20.4-27.0)	42.0 (38.4-45.8)	31.1 (27.2-35.1)
Other or no medical diagnosis	6.7 (6.0-7.5)	36.8 (35.4-38.2)	39.4 (38.0-40.8)	17.0 (15.8-18.3)

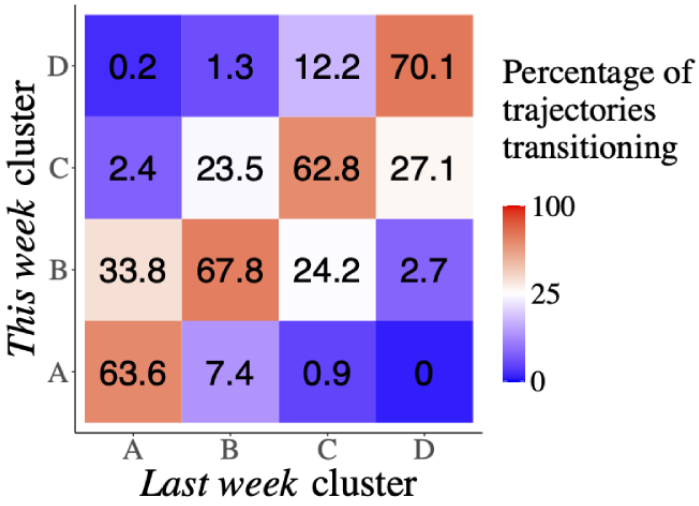
[a]Percentages exceed 100% because participants could report multiple chronic pain conditions.

Transition Between Clusters

There were 12,267 pairs of participant weeks from 1761 participants used in the transition analysis. The demographic data are compared to those in the main analysis in Multimedia Appendix 4. In general, a slightly higher proportion of older adults contributed to the transition analysis compared to the cluster analysis; for example, of the 2807 participants in the main analysis, 434 (15.46%) were aged 65 to 86 years, but of the 1761 participants in the transition analysis, 300 (17.04%) were older adults. There are no other differences in the demographics of participants contributing to the main analysis and the transition analysis.

The percentages of consecutive trajectories transitioning between clusters are shown in Figure 7. For each cluster, the highest percentages of trajectories in consecutive weeks remain in the same cluster, with the percentage values ranging between 62.76% (2948/4697) and 70.14% (1466/2090). On average, 65.96% (8091/12,267) of the trajectories remain in the same cluster. When individuals move between clusters, it is most frequently to an adjacent cluster. There is a very small percentage of consecutive weeks displaying movement between clusters ≥2 levels away.

Figure 7. Transition matrix of movement between clusters on consecutive weeks. For membership in each cluster for last week, the percentage of membership in each cluster for this week is shown. Random movement between clusters would suggest that each combination has a transition percentage of 25%. Blue squares represent transitions that have a higher-than-random percentage (>25%). Red squares represent transitions that have a lower-than-random percentage (<25%). A white square would have exactly the random percentage (25%).



Discussion

Principal Findings

This study identified and described clusters of weekly trajectories of pain severity in a large population-based mHealth study to address 4 objectives in turn. First, we reported that 4 clusters (A=*no or low pain*, B=*mild pain*, C=*moderate pain*, and D=*severe pain*) represented an optimal clustering solution for these data. In this solution, clusters B and C contained the greatest number of weekly pain trajectories.

Second, we conducted sensitivity analyses to identify whether the conclusions made about the first objective were robust to modified assumptions around the structure of the data. Two sensitivity analyses were conducted when the outcome variable was assumed to be (1) continuous and longitudinal and (2) categorical and not longitudinal. These analyses found that 4 clusters remained a suitable conclusion. A third sensitivity analysis found no differences in the clusters of trajectories starting on different days of the week.

Third, younger people and female participants contributed a greater number of trajectories to the *severe pain* cluster than older people and male participants, respectively. Participants with fibromyalgia and neuropathic pain contributed more trajectories to the *severe pain* cluster than those with other pain conditions, whereas participants with rheumatoid arthritis contributed more trajectories to the *no or low pain* cluster than those with other pain conditions.

Fourth, we examined transitions between clusters and found that 65.96% (8091/12,267) of the consecutive trajectories contributed to the same cluster. However, there was clear evidence of between-cluster movement with 34.04% (4176/12,267) of the consecutive trajectories assigned to different clusters. Between-cluster movement was most likely to a neighboring cluster; for example, moving from cluster 1 to cluster 2 was more common than moving from cluster 1 to cluster 3. This analysis demonstrates that overall, individuals tend to experience similar patterns of pain severity from week to week, although there are substantial experiences of increases or decreases in pain severity, thereby reflecting the lived experience of people with chronic pain having variability in symptoms and noting how pain can fluctuate between weeks.

People with chronic pain have highlighted a need to describe and predict the variability in the severity of their pain. Through clustering, this study has described 4 common experiences of pain severity, accounting for two-thirds of the observed variability. However, trajectories within each cluster are not homogeneous, and there remains within-cluster variation. To

better describe the individual weekly pain experience, future work should explore the remaining variability within clusters.

Comparison With Prior Work

Many studies have identified clusters of pain trajectories among individuals living with chronic pain. Some have focused on participants with 1 chronic pain condition, such as osteoarthritis [15,36-39,52-58], low back pain [13,14,27,59-65], other back pain [25,32,49,66], neck or shoulder pain [33,61,67,68], leg pain [29], knee pain [69], or foot pain [70], whereas others have identified clusters among a broader population, such as those with musculoskeletal pain [26,31,47,71,72] or general pain [48,73-75]. Clusters in these studies were described by the severity of pain (eg, *no pain*, *very low pain*, *mild pain*, *moderate pain*, *high pain*, and *severe pain*), the level of change in pain severity (eg, *persistent*, *ongoing*, *episodic*, *worsening*, *recovering*, and *fluctuating*), or a combination of these features.

These previous studies have often considered only sparse data, with relatively large time intervals between consecutive data points. Of those gathering data for at least 1 year (n=27 studies), data were collected >2 times in only 2 studies [13,67]. In these 2 studies, data were collected weekly for 1 year to explore the course of specific pain conditions (neck pain and low back pain). Kongsted et al [13] used 12 models to identify between 5 and 12 clusters in each model. Clusters were described by the severity of pain (eg, *moderate* and *severe*) and also by the temporal features of the trajectories (eg, *episodic*, *recovery*, and *ongoing*). Pico-Espinosa et al [67] identified 6 clusters of pain described as *small improvement*, *moderate improvement*, *persistent*, *large improvement*, *slightly fluctuating*, and *highly fluctuating*. The clusters identified in our analysis were described by the severity of pain, similar to clusters in studies with sparse data. Our clusters were unlikely to identify long-term disease development, as with trajectories over longer periods.

Similar to our study, some previous studies have used methods from the k-means and k-medoids family of algorithms. Knecht et al [25] used the *KmL* package to identify 2 clusters in *responders* and *nonresponders* groups. Weng et al [76] used a k-median algorithm to identify 4 clusters of pain severity: *slightly rise*, *completely drop*, *sudden rise*, and *steady group*. Both these studies identified trajectories with changing pain severity, while our study identified trajectories of weekly pain where the medoid was stable across the week.

In all aforementioned studies, the experiences of individuals were described by a single trajectory across the full duration of follow-up, whereas our study examined week-to-week transition between clusters. Kongsted et al [77] have previously examined week-to-week pain severity across a year, using predefined clusters. The authors identified that 41% and 21% of the respondents in 2 different data sets had stable pain over a year, defined as pain within 1 point of the mean pain value on an 11-point numerical rating scale. The remaining pain trajectories were classified as having a single episode of pain, being *episodic* or *fluctuating*. The transitions identified in our study suggest stability between 65.96% (8091/12,267) of the consecutive weeks. However, some individuals in our study may experience the other longer-term descriptions outlined by Kongsted et al [77]; for example, an individual might not transition out of a cluster for most of the year; yet they might experience only 1 episode. Future studies should further examine the movement between different pain states and identify the drivers of these transitions.

Strengths and Limitations

A number of strengths and limitations of this study should be considered. First, a strength was that participants could contribute daily data for up to 64 weeks. This frequent and granular data collection, enabled by mHealth, overcame limitations of sparse data collection in previous studies (as identified by Beukenhorst et al [78]). As a result, this study was able to analyze the weekly trajectories contributed by participants, determining common pain patterns among a population with chronic pain at a more granular scale than previously investigated.

Second, the analysis presented in this paper modeled weekly pain trajectories rather than individual people. In contrast to studies that assign each individual to a single cluster across the whole follow-up, individuals were able to transition between different pain clusters over time as their pain experience changed, and their condition developed. These transitions were observed in 34.04% (4176/12,267) of the consecutive weeks, and this flexibility can be used in future work to explore the mechanisms driving movement between clusters.

Third, assumptions about the ordinal and longitudinal form of the data were modified in sensitivity analyses. A 4-cluster solution was most suitable for each analysis, indicating a strong model-independent signal in the data and a more robust conclusion regarding the most suitable number of clusters. Furthermore, the assignment of trajectories to each cluster were similar in each analysis (at least 86% similarity), indicating further stability in the results. There were benefits to the use of both the k-medoids algorithm and the longitudinal adaptation of the k-means algorithm used in this analysis. First, neither of these methods requires parametric assumptions about the form of the data [50]. Second, no prior assumptions, including the shape of the trajectory, are required by the algorithms [79]. Therefore, this data-driven approach made limited assumptions about the form of the data.

There were also limitations to the study. The data used in this study were from a population-based study that represented the UK population. Cloudy with a Chance of Pain recruited participants from all UK postcodes, although male participants and those in the age brackets 17 to 34 years and ≥75 years were underrepresented in the study population [24]. Despite being a smaller population, older people and male participants contributed more trajectories on average and were more likely to contribute trajectories to a *less severe pain* category. As these clusters will be used in the development of a pain-forecasting model, clusters should be generalizable to the population with chronic pain, and there remains the possibility that different pain clusters and between-cluster transitions could be realized among those who contributed to the study did and those who did not. Although it is unlikely that our large study population would display pain clusters and transitions different from those of the population with chronic pain that would use a smartphone

tracking app, it remains a possibility that should be explored in future studies.

This analysis further selected participants by the requirement to provide a week of complete pain severity data, thus excluding missing data. There are reasons that data might be missing not at random, including missing due to severe pain, missing due to low pain severity, and missing due to stable pain, that result in repetitive score input and thus disengagement. The transition analysis also further selected participants by requiring 2 weeks of complete pain severity data. However, the age, sex, and chronic pain conditions of respondents in the main analysis and transition analysis (Multimedia Appendix 4) were similar to those in the full-study population (see the first supplementary in the study by Dixon et al [16]), suggesting that the included participants were representative of the study population.

There were limitations in the method used for clustering. First, the absence of parametric assumptions in either the k-medoids algorithm or the *KmL* package resulted in goodness-of-fit measures being inappropriate [79]. Therefore, the elbow method was used to select the optimum number of clusters. However, the use of the elbow method introduces subjectivity. Second, both the k-medoids algorithm and the *KmL* package require random starting values for the cluster centers, which can add volatility to the results. This volatility was mitigated by repeating the algorithms 20 times each and selecting the solution with the lowest remaining variability within clusters.

Conclusions

Previous research has highlighted a need to better understand pain variability experienced by individuals with chronic pain [20]. Feelings of uncertainty among people with chronic pain have led them to want to better understand the pain that they may experience in the future. Clustering weekly pain trajectories offers a first step to better understanding common experiences of pain severity. Once these common experiences are better described, they can be used in future work to predict movement between clusters.

There are limited methods available for clustering pain severity that respect the ordinal and longitudinal nature of patient-generated health data in a computationally inexpensive manner. The clustering method and subsequent sensitivity analyses presented in this paper suggest that the use of k-medoids is robust to assumptions about the data structure.

This study identified 4 distinct patterns of weekly pain severity: *no or low pain*, *mild pain*, *moderate pain*, and *severe pain*. These can be used to describe the short-term pain experiences of people with chronic pain. Future work is required to identify how these clusters can be used in a pain-forecasting model. First, there remains individual variability within clusters of pain severity. Participants in patient and public involvement studies have identified that fluctuations in pain severity should be forecast, and therefore within-cluster variability should be quantified to further understand the weekly pain experience of individuals. Second, the transition of individuals between clusters should be explored to identify the drivers of movement between pain clusters on an individual level. The clusters identified in this study and in future work to understand within-cluster variability and the drivers of movement between clusters will enable a future pain-forecasting model.

Acknowledgments

This study was supported by the Centre for Epidemiology Versus Arthritis (21755). Cloudy with a Chance of Pain was funded by Versus Arthritis (21225), with additional support from the Centre for Epidemiology Versus Arthritis (21755 and 20380). TH received funding from the Royal Society (INF/R2/180067) and the Alan Turing Institute for Data Science and Artificial Intelligence. The funders of the study had no role in study design, data interpretation, writing of the manuscript, or the decision to submit for publication.

Data Availability

The data sets generated and analyzed during this study are available from the corresponding author on reasonable request. Data management and analyses were performed in R (version 4.1.2; R Foundation for Statistical Computing). The code is available from the corresponding author on reasonable request.

Conflicts of Interest

WGD has received consultancy fees from Google, and DMS has received consultancy fees from Palta, both unrelated to this work. All other authors declare no other conflicts of interest.

Multimedia Appendix 1
Comorbid pain conditions.
[DOCX File , 14 KB - mhealth_v12i1e48582_app1.docx]

Multimedia Appendix 2
Sensitivity analyses 1 and 2.
[DOCX File , 444 KB - mhealth_v12i1e48582_app2.docx]

Multimedia Appendix 3
Sensitivity analysis 3.
[DOCX File , 170 KB - mhealth_v12i1e48582_app3.docx]

Multimedia Appendix 4
Demographic data of participants included in transition analysis.
[DOCX File , 19 KB - mhealth_v12i1e48582_app4.docx]

References

1. McQueenie R, Jani BD, Siebert S, McLoone P, McCowan C, Macdonald S, et al. Prevalence of chronic pain in LTCs and multimorbidity: a cross-sectional study using UK Biobank. J Multimorb Comorb 2021 Dec 21;11:26335565211005870 [FREE Full text] [doi: 10.1177/26335565211005870] [Medline: 35004337]
2. Wolfe F, Clauw DJ, Fitzcharles MA, Goldenberg DL, Katz RS, Mease P, et al. The American College of Rheumatology preliminary diagnostic criteria for fibromyalgia and measurement of symptom severity. Arthritis Care Res (Hoboken) 2010 May 23;62(5):600-610 [FREE Full text] [doi: 10.1002/acr.20140] [Medline: 20461783]
3. Steineck A, Bradford MC, O'Daffer A, Fladeboe KM, O'Donnell MB, Scott S, et al. Quality of life in adolescents and young adults: the role of symptom burden. J Pain Symptom Manage 2022 Sep;64(3):244-53.e2 [FREE Full text] [doi: 10.1016/j.jpainsymman.2022.05.017] [Medline: 35649460]
4. Dworkin RH, Turk DC, Farrar JT, Haythornthwaite JA, Jensen MP, Katz NP, et al. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. Pain 2005 Jan;113(1-2):9-19. [doi: 10.1016/j.pain.2004.09.012] [Medline: 15621359]
5. Schneider S, Junghaenel DU, Keefe FJ, Schwartz JE, Stone AA, Broderick JE. Individual differences in the day-to-day variability of pain, fatigue, and well-being in patients with rheumatic disease: associations with psychological variables. Pain 2012 Apr;153(4):813-822 [FREE Full text] [doi: 10.1016/j.pain.2012.01.001] [Medline: 22349917]
6. Wesolowicz DM, Bishop MD, Robinson ME. An examination of day-to-day and intraindividual pain variability in low back pain. Pain Med 2021 Oct 08;22(10):2263-2275. [doi: 10.1093/pm/pnab119] [Medline: 33822203]
7. Crowe M, Whitehead L, Seaton P, Jordan J, Mccall C, Maskill V, et al. Qualitative meta-synthesis: the experience of chronic pain across conditions. J Adv Nurs 2017 May;73(5):1004-1016. [doi: 10.1111/jan.13174] [Medline: 27679971]
8. Johnson LM, Zautra AJ, Davis MC. The role of illness uncertainty on coping with fibromyalgia symptoms. Health Psychol 2006 Nov;25(6):696-703 [FREE Full text] [doi: 10.1037/0278-6133.25.6.696] [Medline: 17100498]
9. Edwards RR, Dworkin RH, Turk DC, Angst MS, Dionne R, Freeman R, et al. Patient phenotyping in clinical trials of chronic pain treatments: IMMPACT recommendations. Pain 2016 Sep;157(9):1851-1871 [FREE Full text] [doi: 10.1097/j.pain.0000000000000602] [Medline: 27152687]
10. Tupper SM, Rosenberg AM, Pahwa P, Stinson JN. Pain intensity variability and its relationship with quality of life in youths with juvenile idiopathic arthritis. Arthritis Care Res (Hoboken) 2013 Apr;65(4):563-570 [FREE Full text] [doi: 10.1002/acr.21850] [Medline: 22972759]
11. Zakoscielna KM, Parmelee PA. Pain variability and its predictors in older adults: depression, cognition, functional status, health, and pain. J Aging Health 2013 Dec;25(8):1329-1339. [doi: 10.1177/0898264313504457] [Medline: 24122353]
12. Stone AA, Broderick JE, Goldman RE, Junghaenel DU, Bolton A, May M, et al. I. Indices of pain intensity derived from ecological momentary assessments: rationale and stakeholder preferences. J Pain 2021 Apr;22(4):359-370 [FREE Full text] [doi: 10.1016/j.jpain.2020.08.003] [Medline: 32947012]
13. Kongsted A, Kent P, Hestbaek L, Vach W. Patients with low back pain had distinct clinical course patterns that were typically neither complete recovery nor constant pain. A latent class analysis of longitudinal data. Spine J 2015 May 01;15(5):885-894 [FREE Full text] [doi: 10.1016/j.spinee.2015.02.012] [Medline: 25681230]
14. Dunn KM, Jordan K, Croft PR. Characterizing the course of low back pain: a latent class analysis. Am J Epidemiol 2006 Apr 15;163(8):754-761. [doi: 10.1093/aje/kwj100] [Medline: 16495468]
15. Bastick AN, Wesseling J, Damen J, Verkleij SP, Emans PJ, Bindels PJ, et al. Defining knee pain trajectories in early symptomatic knee osteoarthritis in primary care: 5-year results from a nationwide prospective cohort study (CHECK). Br J Gen Pract 2016 Jan;66(642):e32-e39 [FREE Full text] [doi: 10.3399/bjgp15X688129] [Medline: 26639946]
16. Dixon WG, Beukenhorst AL, Yimer BB, Cook L, Gasparrini A, El-Hay T, et al. How the weather affects the pain of citizen scientists using a smartphone app. NPJ Digit Med 2019 Oct 24;2:105 [FREE Full text] [doi: 10.1038/s41746-019-0180-3] [Medline: 31667359]
17. Check Hayden E. Mobile-phone health apps deliver data bounty. Nature 2016 Mar 24;531(7595):422-423. [doi: 10.1038/531422a] [Medline: 27008946]
18. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]

19. Rahman QA, Janmohamed T, Pirbaglou M, Clarke H, Ritvo P, Heffernan JM, et al. Defining and predicting pain volatility in users of the manage my pain app: analysis using data mining and machine learning methods. J Med Internet Res 2018 Nov 15;20(11):e12001 [FREE Full text] [doi: 10.2196/12001] [Medline: 30442636]
20. Little CL, Druce KL, Dixon WG, Schultz DM, House T, McBeth J. What do people living with chronic pain want from a pain forecast? A research prioritization study. PLoS One 2023 Oct 12;18(10):e0292968 [FREE Full text] [doi: 10.1371/journal.pone.0292968] [Medline: 37824568]
21. Flurey CA, Morris M, Richards P, Hughes R, Hewlett S. It's like a juggling act: rheumatoid arthritis patient perspectives on daily life and flare while on current treatment regimes. Rheumatology (Oxford) 2014 Apr;53(4):696-703 [FREE Full text] [doi: 10.1093/rheumatology/ket416] [Medline: 24357813]
22. Liddell TM, Kruschke JK. Analyzing ordinal data with metric models: what could possibly go wrong? J Exp Soc Psychol 2018 Nov;79:328-348. [doi: 10.1016/j.jesp.2018.08.009]
23. Schultz DM, Beukenhorst AL, Yimer BB, Cook L, Pisaniello HL, House T, et al. Weather patterns associated with pain in chronic-pain sufferers. Bull Am Meteorol Soc 2020 May 01;101(5):E555-E566. [doi: 10.1175/BAMS-D-19-0265.1]
24. Druce KL, McBeth J, van der Veer SN, Selby DA, Vidgen B, Georgatzis K, et al. Recruitment and ongoing engagement in a UK smartphone study examining the association between weather and pain: cohort study. JMIR Mhealth Uhealth 2017 Nov 01;5(11):e168 [FREE Full text] [doi: 10.2196/mhealth.8162] [Medline: 29092810]
25. Knecht C, Hartnack S, Sick B, Riner F, Schweinhardt P, Wirth B. A prospective observational study on trajectories and prognostic factors of mid back pain. BMC Musculoskelet Disord 2020 Aug 17;21(1):554 [FREE Full text] [doi: 10.1186/s12891-020-03534-5] [Medline: 32807140]
26. Crook J, Moldofsky H. The clinical course of musculoskeletal pain in empirically derived groupings of injured workers. Pain 1996 Oct;67(2-3):427-433. [doi: 10.1016/0304-3959(96)03151-x] [Medline: 8951938]
27. Axén I, Bodin L, Bergström G, Halasz L, Lange F, Lövgren PW, et al. Clustering patients on the basis of their individual course of low back pain over a six month period. BMC Musculoskelet Disord 2011 May 17;12:99 [FREE Full text] [doi: 10.1186/1471-2474-12-99] [Medline: 21586117]
28. Brennan PL, SooHoo S. Effects of mental health disorders on nursing home residents' nine-month pain trajectories. Pain Med 2020 Mar 01;21(3):488-500. [doi: 10.1093/pm/pnz177] [Medline: 31407787]
29. Ogollah RO, Konstantinou K, Stynes S, Dunn KM. Determining one-year trajectories of low-back-related leg pain in primary care patients: growth mixture modeling of a prospective cohort study. Arthritis Care Res (Hoboken) 2018 Dec;70(12):1840-1848. [doi: 10.1002/acr.23556] [Medline: 29575673]
30. Battaglia M, Garon-Carrier G, Brendgen M, Feng B, Dionne G, Vitaro F, et al. Trajectories of pain and anxiety in a longitudinal cohort of adolescent twins. Depress Anxiety 2020 May;37(5):475-484. [doi: 10.1002/da.22992] [Medline: 31944483]
31. Neupane S, Nygård CH, Prakash KC, von Bonsdorff MB, von Bonsdorff ME, Seitsamo J, et al. Multisite musculoskeletal pain trajectories from midlife to old age: a 28-year follow-up of municipal employees. Occup Environ Med 2018 Dec;75(12):863-870. [doi: 10.1136/oemed-2018-105235] [Medline: 30323013]
32. Dunn KM, Campbell P, Jordan KP. Long-term trajectories of back pain: cohort study with 7-year follow-up. BMJ Open 2013 Dec 11;3(12):e003838 [FREE Full text] [doi: 10.1136/bmjopen-2013-003838] [Medline: 24334157]
33. Hallman DM, Rasmussen CD, Jørgensen MB, Holtermann A. Time course of neck-shoulder pain among workers: a longitudinal latent class growth analysis. Scand J Work Environ Health 2018 Jan 01;44(1):47-57 [FREE Full text] [doi: 10.5271/sjweh.3690] [Medline: 29120478]
34. Vermunt JK. Latent class and finite mixture models for multilevel data sets. Stat Methods Med Res 2008 Feb;17(1):33-51. [doi: 10.1177/0962280207081238] [Medline: 17855746]
35. Hebert JJ, Abraham E, Wedderkopp N, Bigney E, Richardson E, Darling M, et al. Patients undergoing surgery for lumbar spinal stenosis experience unique courses of pain and disability: a group-based trajectory analysis. PLoS One 2019 Nov 7;14(11):e0224200 [FREE Full text] [doi: 10.1371/journal.pone.0224200] [Medline: 31697714]
36. Wieczorek M, Rotonda C, Coste J, Pouchot J, Saraux A, Guillemin F, et al. Trajectory analysis combining pain and physical function in individuals with knee and hip osteoarthritis: results from the French KHOALA cohort. Rheumatology (Oxford) 2020 Nov 01;59(11):3488-3498. [doi: 10.1093/rheumatology/keaa148] [Medline: 32375174]
37. Dai Z, Lu N, Niu J, Felson DT, Zhang Y. Dietary fiber intake in relation to knee pain trajectory. Arthritis Care Res (Hoboken) 2017 Sep;69(9):1331-1339 [FREE Full text] [doi: 10.1002/acr.23158] [Medline: 27899003]
38. Radojčić MR, Arden NK, Yang X, Strauss VY, Birrell F, Cooper C, et al. Pain trajectory defines knee osteoarthritis subgroups: a prospective observational study. Pain 2020 Dec;161(12):2841-2851 [FREE Full text] [doi: 10.1097/j.pain.0000000000001975] [Medline: 32639366]
39. Pan F, Tian J, Aitken D, Cicuttini F, Jones G. Predictors of pain severity trajectory in older adults: a 10.7-year follow-up study. Osteoarthritis Cartilage 2018 Dec;26(12):1619-1626 [FREE Full text] [doi: 10.1016/j.joca.2018.08.002] [Medline: 30121348]
40. Pan F, Tian J, Cicuttini F, Jones G. Metabolic syndrome and trajectory of knee pain in older adults. Osteoarthritis Cartilage 2020 Jan;28(1):45-52 [FREE Full text] [doi: 10.1016/j.joca.2019.05.030] [Medline: 31394191]

41. Kassambara A. Practical guide to cluster analysis in R: unsupervised machine learning. Statistical Tools for High-throughput Data Analysis. 2017. URL: http://www.sthda.com/french/ [accessed 2024-07-01]
42. Kaufman L, Rousseeuw P. Clustering large data sets. In: Gelsema ES, Kanal LN, editors. Pattern Recognition in Practice II. Amsterdam, The Netherlands: Elsevier; 1986:425-437.
43. Kaufman L, Rousseeuw PJ. Finding Groups in Data: An Introduction to Cluster Analysis. Hoboken, NJ: John Wiley & Sons; 1990.
44. Sarang P. Centroid-based clustering: clustering algorithms for hard clustering. In: Thinking Data Science. Cham, Switzerland: Springer; 2023.
45. Hardy A. On the number of clusters. Comput Stat Data Anal 1996 Nov 15;23(1):83-96. [doi: 10.1016/s0167-9473(96)00022-9]
46. Thomas MJ, Yu D, Nicholls E, Bierma-Zeinstra S, Conaghan PG, Stoner KJ, et al. Short-term recovery trajectories of acute flares in knee pain: a UK-Netherlands multicenter prospective cohort analysis. Arthritis Care Res (Hoboken) 2020 Dec;72(12):1687-1692 [FREE Full text] [doi: 10.1002/acr.24088] [Medline: 31600034]
47. Airila A, Hakanen JJ, Luukkonen R, Lusa S, Punakallio A, Leino-Arjas P. Developmental trajectories of multisite musculoskeletal pain and depressive symptoms: the effects of job demands and resources and individual factors. Psychol Health 2014;29(12):1421-1441. [doi: 10.1080/08870446.2014.945929] [Medline: 25096992]
48. James RJ, Walsh DA, Ferguson E. General and disease-specific pain trajectories as predictors of social and political outcomes in arthritis and cancer. BMC Med 2018 Apr 09;16(1):51 [FREE Full text] [doi: 10.1186/s12916-018-1031-9] [Medline: 29628016]
49. Canizares M, Rampersaud YR, Badley EM. Course of back pain in the Canadian population: trajectories, predictors, and outcomes. Arthritis Care Res (Hoboken) 2019 Dec;71(12):1660-1670. [doi: 10.1002/acr.23811] [Medline: 30637980]
50. Genolini C, Falissard B. KmL: a package to cluster longitudinal data. Comput Methods Programs Biomed 2011 Dec;104(3):e112-e121. [doi: 10.1016/j.cmpb.2011.05.008] [Medline: 21708413]
51. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. Ann Intern Med 2007 Oct 16;147(8):573-577 [FREE Full text] [doi: 10.7326/0003-4819-147-8-200710160-00010] [Medline: 17938396]
52. Verkleij SP, Hoekstra T, Rozendaal RM, Waarsing JH, Koes BW, Luijsterburg PA, et al. Defining discriminative pain trajectories in hip osteoarthritis over a 2-year time period. Ann Rheum Dis 2012 Sep;71(9):1517-1523. [doi: 10.1136/annrheumdis-2011-200687] [Medline: 22492782]
53. Wesseling J, Bastick AN, ten Wolde S, Kloppenburg M, Lafeber FP, Bierma-Zeinstra SM, et al. Identifying trajectories of pain severity in early symptomatic knee osteoarthritis: a 5-year followup of the cohort hip and cohort knee (CHECK) study. J Rheumatol 2015 Aug;42(8):1470-1477. [doi: 10.3899/jrheum.141036] [Medline: 26136492]
54. Nicholls E, Thomas E, van der Windt DA, Croft PR, Peat G. Pain trajectory groups in persons with, or at high risk of, knee osteoarthritis: findings from the knee clinical assessment study and the osteoarthritis initiative. Osteoarthritis Cartilage 2014 Dec;22(12):2041-2050 [FREE Full text] [doi: 10.1016/j.joca.2014.09.026] [Medline: 25305072]
55. Bastick AN, Verkleij SP, Damen J, Wesseling J, Hilberdink WK, Bindels PJ, et al. Defining hip pain trajectories in early symptomatic hip osteoarthritis--5 year results from a nationwide prospective cohort study (CHECK). Osteoarthritis Cartilage 2016 May;24(5):768-775 [FREE Full text] [doi: 10.1016/j.joca.2015.11.023] [Medline: 26854794]
56. Collins JE, Katz JN, Dervan EE, Losina EL. Trajectories and risk profiles of pain in persons with radiographic, symptomatic knee osteoarthritis: data from the osteoarthritis initiative. Osteoarthritis Cartilage 2014 May;22(5):622-630 [FREE Full text] [doi: 10.1016/j.joca.2014.03.009] [Medline: 24662734]
57. Johnson AJ, Vasilopoulos T, Booker SQ, Cardoso J, Terry EL, Powell-Roach K, et al. Knee pain trajectories over 18 months in non-Hispanic Black and non-Hispanic White adults with or at risk for knee osteoarthritis. BMC Musculoskelet Disord 2021 May 05;22(1):415 [FREE Full text] [doi: 10.1186/s12891-021-04284-8] [Medline: 33952243]
58. Atukorala I, Downie A, Pathmeswaran A, Deveza LM, Chang T, Zhang Y, et al. Short-term pain trajectories in patients with knee osteoarthritis. Int J Rheum Dis 2022 Mar;25(3):281-294. [doi: 10.1111/1756-185X.14267] [Medline: 35080125]
59. Tamcan O, Mannion AF, Eisenring C, Horisberger B, Elfering A, Müller U. The course of chronic and recurrent low back pain in the general population. Pain 2010 Sep;150(3):451-457. [doi: 10.1016/j.pain.2010.05.019] [Medline: 20591572]
60. Coenen P, Smith A, Paananen M, O'Sullivan P, Beales D, Straker L. Trajectories of low back pain from adolescence to young adulthood. Arthritis Care Res (Hoboken) 2017 Mar;69(3):403-412. [doi: 10.1002/acr.22949] [Medline: 27273901]
61. Ailliet L, Rubinstein SM, Hoekstra T, van Tulder MW, de Vet HC. Long-term trajectories of patients with neck pain and low back pain presenting to chiropractic care: a latent class growth analysis. Eur J Pain 2018 Jan;22(1):103-113. [doi: 10.1002/ejp.1094] [Medline: 28871611]
62. Downie AS, Hancock MJ, Rzewuska M, Williams CM, Lin CW, Maher CG. Trajectories of acute low back pain: a latent class growth analysis. Pain 2016 Jan;157(1):225-234. [doi: 10.1097/j.pain.0000000000000351] [Medline: 26397929]
63. Panken G, Hoekstra T, Verhagen A, van Tulder M, Twisk J, Heymans MW. Predicting chronic low-back pain based on pain trajectories in patients in an occupational setting: an exploratory analysis. Scand J Work Environ Health 2016 Jun 01;42(6):520-527 [FREE Full text] [doi: 10.5271/sjweh.3584] [Medline: 27537989]

64. Macedo LG, Maher CG, Latimer J, McAuley JH, Hodges PW, Rogers WT. Nature and determinants of the course of chronic low back pain over a 12-month period: a cluster analysis. Phys Ther 2014 Feb;94(2):210-221 [FREE Full text] [doi: 10.2522/ptj.20120416] [Medline: 24072729]
65. Kyrönlahti SM, Nygård CH, Neupane S. Trajectories of low back pain from midlife to retirement and functional ability at old age. Eur J Public Health 2022 Jun 01;32(3):497-503 [FREE Full text] [doi: 10.1093/eurpub/ckab191] [Medline: 34792114]
66. Enthoven WT, Koes BW, Bierma-Zeinstra SM, Bueving HJ, Bohnen AM, Peul WC, et al. Defining trajectories in older adults with back pain presenting in general practice. Age Ageing 2016 Nov;45(6):878-883. [doi: 10.1093/ageing/afw127] [Medline: 27515678]
67. Pico-Espinosa OJ, Côté P, Hogg-Johnson S, Jensen I, Axén I, Holm L, et al. Trajectories of pain intensity over 1 year in adults with disabling subacute or chronic neck pain. Clin J Pain 2019 Aug;35(8):678-685 [FREE Full text] [doi: 10.1097/AJP.0000000000000727] [Medline: 31149935]
68. Eriks-Hoogland IE, Hoekstra T, de Groot S, Stucki G, Post MW, van der Woude LH. Trajectories of musculoskeletal shoulder pain after spinal cord injury: identification and predictors. J Spinal Cord Med 2013 Nov 07;37(3):288-298. [doi: 10.1179/2045772313y.0000000168]
69. Haybatollahi SM, James RJ, Fernandes G, Valdes A, Doherty M, Zhang W, et al. Identifying multiple knee pain trajectories and the prediction of opioid and NSAID medication used: a latent class growth approach. Pain Pract 2022 Feb 25;22(2):210-221. [doi: 10.1111/papr.13082] [Medline: 34634169]
70. Marshall M, Blagojevic-Bucknall M, Rathod-Mistry T, Thomas MJ, Edwards JJ, Peat G, et al. Identifying long-term trajectories of foot pain severity and potential prognostic factors: a population-based cohort study. Arthritis Care Res (Hoboken) 2023 May 25;75(5):1123-1131 [FREE Full text] [doi: 10.1002/acr.24823] [Medline: 34806345]
71. Leino-Arjas P, Rajaleid K, Mekuria G, Nummi T, Virtanen P, Hammarström A. Trajectories of musculoskeletal pain from adolescence to middle age: the role of early depressive symptoms, a 27-year follow-up of the Northern Swedish Cohort. Pain 2018 Jan 19;159(1):67-74. [doi: 10.1097/j.pain.0000000000001065] [Medline: 28937577]
72. Oakman J, Neupane S, Kyrönlahti S, Nygård CH, Lambert K. Musculoskeletal pain trajectories of employees working from home during the COVID-19 pandemic. Int Arch Occup Environ Health 2022 Nov 08;95(9):1891-1901 [FREE Full text] [doi: 10.1007/s00420-022-01885-1] [Medline: 35674803]
73. Neupane S, Lallukka T, Pietiläinen O, Rahkonen O, Leino-Arjas P. Trajectories of multisite musculoskeletal pain in midlife: associations with common mental disorders. Eur J Pain 2020 Feb 24;24(2):364-373. [doi: 10.1002/ejp.1492] [Medline: 31612552]
74. Sun F, Park NS, Wardian J, Lee BS, Roff LL, Klemmack DL, et al. Predicting the trajectories of perceived pain intensity in southern community-dwelling older adults: the role of religiousness. Res Aging 2013 Nov 01;35(6):643-662 [FREE Full text] [doi: 10.1177/0164027512456402] [Medline: 24187410]
75. Wickrama KA, Lee TK, O'Neal CW. Heterogeneity in husbands' and wives' physical pain trajectories over mid-later years: biopsychosocial stratification and implications for later-life well-being. Gerontologist 2021 Aug 13;61(6):930-941 [FREE Full text] [doi: 10.1093/geront/gnab007] [Medline: 33453117]
76. Weng Y, Tian L, Tedesco D, Desai K, Asch SM, Carroll I, et al. Trajectory analysis for postoperative pain using electronic health records: a nonparametric method with robust linear regression and K-medians cluster analysis. Health Informatics J 2020 Jun 17;26(2):1404-1418 [FREE Full text] [doi: 10.1177/1460458219881339] [Medline: 31621460]
77. Kongsted A, Hestbæk L, Kent P. How can latent trajectories of back pain be translated into defined subgroups? BMC Musculoskelet Disord 2017 Jul 03;18(1):285 [FREE Full text] [doi: 10.1186/s12891-017-1644-8] [Medline: 28673341]
78. Beukenhorst AL, Schultz DM, McBeth J, Sergeant JC, Dixon WG. Are weather conditions associated with chronic musculoskeletal pain? Review of results and methodologies. Pain 2020 Apr 20;161(4):668-683. [doi: 10.1097/j.pain.0000000000001776] [Medline: 32195783]
79. Genolini C, Alacoque X, Sentenac M, Arnaud C. kml and kml3d: R packages to cluster longitudinal data. J Stat Softw 2015;65(4):1-34. [doi: 10.18637/jss.v065.i04]

Abbreviations

CLARA: Clustering Large Applications
mHealth: mobile health
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology
WSS: within-cluster sum of squares

Edited by L Buis; submitted 28.04.23; peer-reviewed by A Obbarius, Q Rahman; comments to author 27.07.23; revised version received 17.09.23; accepted 14.11.23; published 19.07.24.

Please cite as:
Little CL, Schultz DM, House T, Dixon WG, McBeth J
Identifying Weekly Trajectories of Pain Severity Using Daily Data From an mHealth Study: Cluster Analysis
JMIR Mhealth Uhealth 2024;12:e48582
URL: https://mhealth.jmir.org/2024/1/e48582
doi:10.2196/48582
PMID:39028557

Original Paper

Using Text Messaging Surveys in General Practice Research to Engage With People From Low-Income Groups: Multi-Methods Study

Elizabeth Sturgiss[1], BMed, MPH, PhD; Jenny Advocat[1], BA, MSW, GDipBA, PhD; Christopher Barton[2], BSc, MMedSc, PhD; Emma N Walker[2], BMedSci, BA, BSci, MD; Suzanne Nielsen[3], BPharm(Hons), MPS, PhD; Annemarie Wright[4,5], BAppSc, MMedSc, PhD; Tina Lam[3], BPsych(Hons), PhD; Nilakshi Gunatillaka[6], BSc(Hons), MIPH; Symrin Oad[7], BBmedSci; Christopher Wood[8], LLB/BActS (Hons)

[1]School of Primary and Allied Health Care, Monash University, Frankston, Australia

[2]Department of General Practice, School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia

[3]Monash Addiction Research Centre, Eastern Health Clinical School, Monash University, Melbourne, Australia

[4]Department of Health, Victorian State Government, Melbourne, Australia

[5]Australia and Melbourne School of Global and Population Health, University of Melbourne, Melbourne, Australia

[6]EPIC Research Unit, School of Primary and Allied Health Care, Monash University, Frankston, Australia

[7]School of Primary and Allied Health Care, Monash University, Melbourne, Australia

[8]Capital Health Network, Canberra, Australia

Corresponding Author:
Elizabeth Sturgiss, BMed, MPH, PhD
School of Primary and Allied Health Care
Monash University
Peninsula Campus
Moorooduc Highway
Frankston, 3199
Australia
Phone: 61 412233119
Email: liz.sturgiss@monash.edu

Abstract

Background: SMS text messages through mobile phones are a common means of interpersonal communication. SMS text message surveys are gaining traction in health care and research due to their feasibility and patient acceptability. However, challenges arise in implementing SMS text message surveys, especially when targeting marginalized populations, because of barriers to accessing phones and data as well as communication difficulties. In primary care, traditional surveys (paper-based and online) often face low response rates that are particularly pronounced among disadvantaged groups due to financial limitations, language barriers, and time constraints.

Objective: This study aimed to investigate the potential of SMS text message–based patient recruitment and surveys within general practices situated in lower socioeconomic areas. This study was nested within the Reducing Alcohol-Harm in General Practice project that aimed to reduce alcohol-related harm through screening in Australian general practice.

Methods: This study follows a 2-step SMS text message data collection process. An initial SMS text message with an online survey link was sent to patients, followed by subsequent surveys every 3 months for consenting participants. Interviews were conducted with the local primary health network organization staff, the participating practice staff, and the clinicians. The qualitative data were analyzed using constructs from the Consolidated Framework for Implementation Research.

Results: Out of 6 general practices, 4 were able to send SMS text messages to their patients. The initial SMS text message was sent to 8333 patients and 702 responses (8.2%) were received, most of which were not from a low-income group. This low initial response was in contrast to the improved response rate to the ongoing 3-month SMS text message surveys (55/107, 51.4% at 3 months; 29/67, 43.3% at 6 months; and 44/102, 43.1% at 9 months). We interviewed 4 general practitioners, 4 nurses, and 4 administrative staff from 5 of the different practices. Qualitative data uncovered barriers to engaging marginalized groups including limited smartphone access, limited financial capacity (telephone, internet, and Wi-Fi credit), language barriers, literacy issues,

mental health conditions, and physical limitations such as manual dexterity and vision issues. Practice managers and clinicians suggested strategies to overcome these barriers, including using paper-based surveys in trusted spaces, offering assistance during survey completion, and offering honoraria to support participation.

Conclusions: While SMS text message surveys for primary care research may be useful for the broader population, additional efforts are required to ensure the representation and involvement of marginalized groups. More intensive methods such as in-person data collection may be more appropriate to capture the voice of low-income groups in primary care research.

International Registered Report Identifier (IRRID): RR2-10.3399/BJGPO.2021.0037

(JMIR Mhealth Uhealth 2024;12:e55354) doi:10.2196/55354

KEYWORDS

SMS; data collection; research methods; disadvantaged population; priority populations; message; messages; messaging; disadvantaged; underserved; survey; surveys; digital divide; marginalized; access; accessibility; barrier; barriers; smartphone; smartphones; digital health; underrepresented; data collection; mobile phone; short message service

Introduction

Mobile phone ownership in countries with advanced economies is almost universal [1,2], and most phone users regularly send text messages through an SMS. [3] An SMS text message is one of the most frequently used channels of interpersonal mobile communication that enables real-time exchange of alphanumeric messages, commonly in packages of up to 160 characters [4], and may be delivered manually or through an automated system.

SMS text messages have been used for clinical and research purposes with promising results in terms of patient feasibility and acceptability. While the standard route for collecting patient data in health research has traditionally been paper-based surveys [5], SMS text message methods may offer equivalent, if not higher, response rates than paper-based methods [6,7], and patients express a preference for SMS text message–based surveys over paper-based alternatives [8], including in low-income settings [9]. When information from multiple time points is requested (as in the case of daily or weekly surveys), SMS text message–based methods may offer a greater chance of obtaining more reliable, complete data, as recall bias can be minimized [6]. In addition, SMS text messaging may assist in collecting data on stigmatized topics, with some evidence suggesting that participants are more likely to disclose issues, such as mental health and substance use information, when asked by non–paper-based methods (eg, SMS text message and internet surveys), rather than face-to-face interviews [10,11].

Surveys of patients in primary care often have low response rates, and this rate is even lower among disadvantaged groups [12]. This reduced participation in research occurs for a variety of reasons, including financial barriers to participation, language barriers, and lack of discretionary time [12]. A recent case study conducted in Australia found that SMS text message surveys of Arabic-literate participants recruited through a community group were successful in data collection while recognizing difficulties with translating materials into readily understandable Arabic resources [13].

Digital inclusion is an important consideration for improving research participation [14]. Digital technologies have been highlighted as a potential option for low-cost, scalable solutions for survey participation that could allow adequate representation of participants from disadvantaged groups. There is variation in access to digital technology through device type, and disadvantaged communities may find the internet most accessible through mobile phones [15]. A recent Australian study of surveys conducted with a culturally and linguistically diverse group using SMS text messages recommended further research on using SMS text messaging surveys with people from low-income groups [13]. There continue to be disparities in smartphone ownership, limited access to data or Wi-Fi, and language barriers that can impede participation in digital health initiatives [16]. While SMS text messaging surveys appear to be an attractive option for research data collection, more research is needed to determine if this would allow equitable participation to achieve digital inclusion [16].

We aimed to explore the use of SMS text messaging for recruitment and data collection in general practice research with a focus on patients who are socioeconomically disadvantaged.

Methods

Ethics Approval

The study was approved by the Monash University Human Research Ethics Committee (22865). Participants provided informed consent to participate. Data were deidentified.

Recruitment

This substudy was nested within an implementation trial of the REACH (Reducing Alcohol-Harm in General Practice) project [17,18]. The overall objective of the study was to explore the feasibility and acceptability of a tool kit to support the use of alcohol brief interventions in general practice. This paper focuses exclusively on the use of SMS text messages to collect data from patients through general practice.

This study was set in Melbourne, Australia. Within Australia, 83.6% of the population is estimated to visit a general practitioner (GP) at least once a year [19], and smartphone ownership is estimated to be 91% [20]. The general practices were located within the state of Victoria, which is Australia's second-largest jurisdiction by population. To be eligible for this study, the general practice had to be situated in a lower socioeconomic area [18]. There were 6 general practices located in lower income communities that participated in the REACH trial and were recruited for the implementation trial through the local primary health network (PHN) [18]. PHNs are

independent, primarily federally funded organizations across Australia that support primary health care providers, including general practices and commission services based on local unmet needs. They are similar in function to other primary care commissioning bodies in the United Kingdom and Canada [21].

SMS Text Message Data Collection

Each practice had its own process for sending the initial SMS text message, details of which are outlined in the Results section. Practices were paid an honorarium of US $658 (US $1=Aus $1.52) to cover some of the costs associated with sending the SMS text messages, including administrative time. The research program manager and PHN staff were also available to discuss any issues the practice had with sending SMS text messages and followed up practices that were having technological difficulties. This approach meant that the research team did not have access to patient names or phone numbers at any time.

There were 2 main steps to the SMS text message surveys. For the first step, a staff member in each practice was asked to send an SMS text message to all current patients aged 18 years or older through an SMS text message blast (where an SMS text message is sent out to a large group of numbers at one time). This initial email contained a link to an online Qualtrics survey (Textbox 1). Patients needed a smartphone to accept the SMS text message as well as mobile data or access to a Wi-Fi connection to complete the survey online. All surveys were in English.

Textbox 1. Initial SMS text messages sent to patients by their general practice.

"Hello, you are receiving this message as you recently visited <<Practice Name>>. We and Monash University are inviting you to complete a survey on alcohol and your health through <<Survey link>>. The survey is confidential, voluntary, and will help us improve the service we deliver to you. Thanks!"

To include patients who attended during the implementation of the REACH project, patients were eligible to complete the online survey if they had visited the general practice within the past 3 months. At the end of the online survey, patients could elect to give their details (name and mobile phone number) for further SMS text message surveys to be sent 3, 6, and 9 months to capture data about their alcohol use and whether they had recently consulted with their GP (Multimedia Appendix 1). We did not have formal checks to assess the authenticity of responses or remove duplicates within Qualtrics. However, the recruitment process (direct SMS text message from the patients' GP clinic) and lack of an incentive to provide multiple responses make it unlikely that we had bot-generated or multiple responses.

The second step involved 2-way SMS text message surveys sent to patients who submitted their details in the online Qualtrics survey that was then downloaded and stored in a password-protected secure drive. In a 2-way SMS text message survey, it is only possible to ask closed questions that can be answered with a number (Textbox 2). The 2-way SMS survey had 5 questions in total and each question had 5 possible responses that were indicated by the patient selecting a number between 1 and 5. It should be noted that patients did not have any contact with the research team between the 3-month, 2-way SMS text message surveys.

Textbox 2. Two-way SMS text message surveys were sent to patients from the research team who agreed to receive the surveys in Step 1.

Hello! This is the REACH team.

Thank you for agreeing to take part in our survey when you last visited your doctor. Text STOP to opt-out of this survey (Note–the respondent answers the SMS with the number that corresponds to their answer)

- Q1 How often do you have a drink containing alcohol?
  - Never (1)
  - Monthly or less (2)
  - Two to four times a month (3)
  - Two to three times a week (4)
  - Four or more times a week (5)
- Skip To: Q4 If the answer is (1), Never
- Q2 How many standard drinks do you have on a typical day when you are drinking?
  - One or two (1)
  - Three or four (2)
  - Five or six (3)
  - Seven to nine (4)
  - Ten or more (5)
- Q3 How often do you have six or more standard drinks on one occasion?
  - Never (1)
  - Less than monthly (2)
  - Monthly (3)
  - Weekly (4)
  - Daily or almost daily (5)
  - Q4 Have you spoken with your GP in the last 3 months?
  - No (1)
  - Yes, and we spoke about alcohol (2)
  - Yes, but we didn't talk about alcohol (3)
  - I can't remember (4)
- Q5 Which of the following describes you?
  - I'm drinking within safe limits (1)
  - I plan to cut down in the next 3 months (2)
  - I want to cut down but haven't decided when (3)
  - I should cut down but really don't want to (4)
  - I don't want to cut down (5)

Interviews and Qualitative Analysis

Toward the end of the REACH trial (May to August 2021), JA, an experienced qualitative research fellow, interviewed PHN staff, clinicians, and practice staff at each of the general practices about their experience of participating in the REACH project, including the SMS text message substudy reported here. JA is a qualitative researcher and an allied health clinician, with more than a decade of experience in primary care research. All staff and clinicians of the practices who were involved with REACH were invited to participate. Data collection concluded when the authors had interviewed participants from each practice and had sufficient data to explain the process of implementation.

The semistructured interview guide included questions about the benefits and barriers of using SMS text messages to communicate with patients. The interviews were conducted remotely due to COVID-19 restrictions (phone or videoconference), and lasted from 18 to 60 minutes, with most approximately 30 minutes. They were audio-recorded and professionally transcribed. The excerpts of the interviews

pertaining to SMS text messages were used in the analysis of this specific study. The elements related to SMS text messages were coded inductively using NVivo software (version 14, QSR International) and then organized according to constructs from the Consolidated Framework for Implementation Research (CFIR) [22] codebook to understand the findings in relationship to implementation factors, both in the inner and outer contexts of the practices.

JA verified the transcripts and coded the data, and a subset of investigators formed an analysis team who met regularly (JA, ES, SN, TL, NG, and CB) to discuss findings based on the CFIR factors.

Results

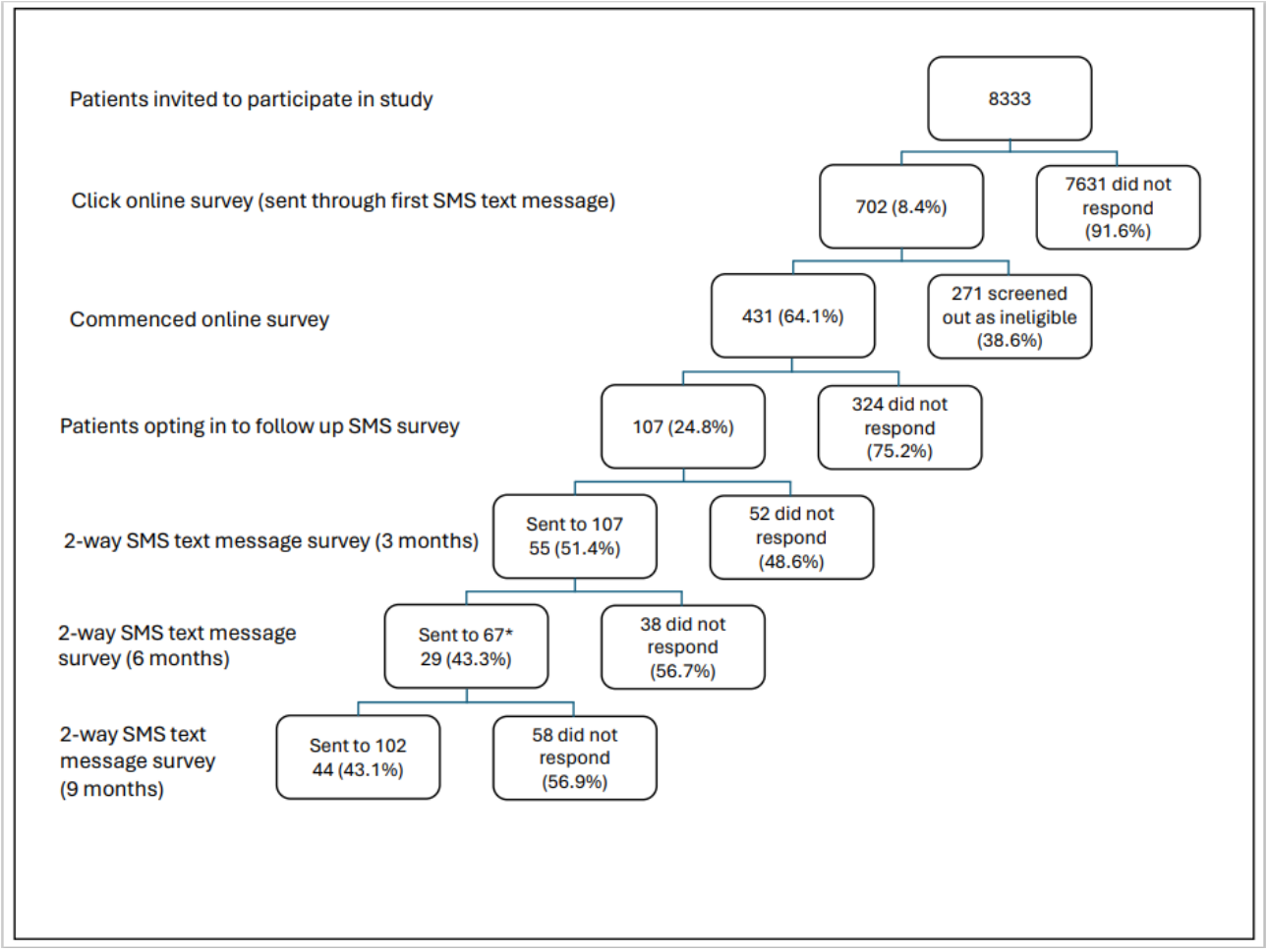
Overview

The practices involved in the REACH project were all located in lower socioeconomic areas and most practices estimated that more than 50% of their patients lived in low-income households [18]. Only 4 of the 6 participating practices were able to send the initial survey link by SMS text message to their patient cohort. One practice reported that IT problems meant they were not able to send SMS text messages to their patients at all, and they did not use SMS text messages to communicate with their patient group. A second practice felt that the SMS text message was inappropriate as their patient group was highly culturally and linguistically diverse. The practice preferred a poster in their waiting room with a QR code for the online survey and then they could assist patients to complete the survey or explain the survey, as needed. However, no patients from this practice completed the survey through the QR code.

Each of the 4 general practices that were able to send SMS text messages to their patients had very different internal processes and thus ended up sending the SMS text message to vastly different numbers of patients. One practice could send the SMS text message to each current patient's mobile number in their system, and a total of 5286 SMS text messages were sent, with 506 commencing the survey (9.6%). Another practice could only send the SMS text message to patients attending on the actual day due to an IT system issue, so they sent it to 500 patients over 3 days with 8 responses (1.6%). The third practice had software that did not allow any past patients to be contacted (usually not seen within 2 years). This practice sent an SMS text message to 2500 patients with 172 initial responses (6.9%). The final practice required patients to give their specific consent to receive the SMS text message, so they asked patients as they presented to the practice and only sent it to those who agreed, with a total of 50 sent across 2 days with 16 responses (32%).

Figure 1 shows the overall response rate at each step of the SMS text message study. In total, 8333 SMS text messages were sent by the practices, and 702 (8.4%) of SMS text messages received a response. To record a response, the patient had to click on the online link for the survey. This would require the mobile phone number for the patient on record to be current, and the mobile phone to have data or Wi-Fi to enable a response if they did click on the external link.

Figure 1. Number of SMS text messages sent and response rate at each step of the SMS survey. *The 6-month 2-way SMS text message was sent to 67 participants; 40 participants only received the 2-way SMS survey at 3 and 9 months.



A total of 431 patients completed at least some of the online surveys with a majority of the survey respondents being women, aged 45 years or older, and were not from a low-income group (Table 1).

Of the 431 survey respondents, 107 (24.8%) agreed to receive an SMS text message survey every 3 months. For the second step of the study using 2-way SMS text messages, the response rate improved to approximately 50% (216/431) for each survey (Figure 1).

Table 1. Respondents in the online survey. The survey link was sent through an initial SMS text message from the general practice (n=431).

Demographics	Participants, n (%)
Age (years; n=340)	
18-24	15 (4.4)
25-34	27 (7.9)
35-44	50 (14.7)
45-54	86 (25.3)
55-64	70 (20.6)
65-74	74 (21.8)
75-84	16 (4.7)
85 or older	2 (0.6)
Gender (n=340)	
Woman	227 (66.8)
Man	105 (30.9)
Nonbinary or third gender	3 (0.9)
Prefer to self-describe as transgender male or nonbinary trans	4 (1.2)
Prefer not to answer	1 (0.3)
Low-income status (n=335; multiple choices possible)	
Unemployed and looking for work	17 (5.1)
Receive a government pension	50 (14.9)
Health care cardholder	16 (4.8)
Live in a low-income household	8 (2.4)
None of these apply to me	236 (70.4)
Prefer not to say	8 (2.4)
Do you have a chronic disease? (n=344)	
Yes	141 (41.0)
No	195 (56.7)
Prefer not to say	8 (2.3)
Did you talk to your doctor or nurse about your alcohol intake during your last visit? (n=377)	
Yes	58 (15.4)
No	303 (80.4)
Do not recall	15 (4.0)
Prefer not to say	1 (0.3)
How often do you have a drink containing alcohol? (n=344)	
Never	60 (17.4)
Monthly or less	77 (22.4)
2-4 times a month	138 (40.1)
4 times a week or more	69 (20.1)

Qualitative Findings

We interviewed 12 representatives from the 5 practices. At least 2 staff members were interviewed from each practice. Our sample included 4 GPs, 4 practice nurses (PNs), and 4 administrative staff members, including PMs. Table 2 summarizes the demographic and professional characteristics of participants.

A total of 7 PHN staff were interviewed, detailed in Table 3.

Table 2. Interviewees from each of the 5 general practices.

Practice	Professional background
1	
Participant 1	General practitioner
Participant 2	Practice manager
Participant 3	Practice nurse
2	
Participant 1	General practitioner
Participant 2	Practice nurse
Participant 3	CEO
3	
Participant 1	General practitioner
Participant 2	Administrative staff member
4	
Participant 1	General practitioner
Participant 2	Practice nurse
5	
Participant 1	Practice manager
Participant 2	Nurse/care coordinator

Table 3. Interviewees from the primary health networks.

Number	Staff role
1	Practice relationship manager
2	Practice relationship manager
3	Continuous quality improvement program officer
4	Project coordinator
5	Continuous quality improvement program officer
6	Continuous quality improvement program officer
7	Manager

In the interviews, the PMs and clinicians gave details about their experience of using SMS text messages in their own practice. PHN staff described any issues that came up with practices they supported that were relevant to the SMS substudy. Qualitative results are organized with reference to CFIR, to assist in understanding factors that affected the implementation, both internal and external, of the practices (Textbox 3).

Textbox 3. Findings from qualitative data mapped to the Consolidated Framework for Implementation Research constructs.

Outer setting

- Needs and resources of the patient population, including language, tech ability, Wi-Fi access
- Competing demands from external policy, including those related to the COVID-19 vaccination roll-out

Inner setting

- Size of the general practice
- Networks and communication between team members
- Implementation climate
- Readiness for implementation of the SMS text message surveys

Characteristics of individuals

- Individual attitudes toward the SMS text message surveys

Innovation characteristics

- Complexity of the SMS text message process
- Relative advantage compared with usual ways of communicating with patients
- Cost-effective
- Modern forms of communication

Outer Setting

Needs and Resources of Those Served by the Organization

Patient demographics came up often during interviews with clinicians and practice staff, including language spoken, age, cultural background, and mental health status. Concerns about patient literacy were reported, especially by staff at clinics serving more diverse communities:

> *I think from the patient's perspective, there were some struggling, because we have 80% non-English speaking people here, so I think they struggled with the text messages.* [P1GP]

For patients whose primary language was other than English, it was common that they were neither able to read words written in their primary language, nor English.

> *…they're not translated at all, even the SMS messages we were able to send out to the patients, we didn't use them because it would just be them ringing us up saying, "We received a text message and we don't know what it is." So, it was just too much work, so we didn't go ahead with that either.* [P1PM]

Patient characteristics also came up in relation to the patient population served at one of the practices, which included people with dual diagnoses of complex mental illness and alcohol and other drug addictions. It was thought by the PM that patients in this practice would not respond to SMS text messages and that they would not have the technical skills to adapt to new systems.

> *not everyone is tech-savvy. I have to say, some of my patients still have the flip-flop phones, not smartphones. So if we send - it definitely can receive a text message, but it's not adaptable to QR codes, to links, those things…So for us, a good old paper that they can take and read and not print out I think is still the best…* [P5PM]

If patients did have a suitable phone, they may not have a phone plan with data or access to Wi-Fi. Practices did not report providing free access to Wi-Fi for patients within their practice. Low phone credit could also make returning an SMS text message survey problematic.

The age of patients also came up as a relevant factor for 2 of the practices (P4 and P5):

> *For some, the younger ones prefer technological things, like SMS, QR codes. But the older ones prefer something to read that's actually printed. Yeah. It's a good thing that either way, you'll be available for that.* [P5PM]

External Policy and Incentives

External policy and incentives is a broad CFIR construct that is related here to the inclusion of policy and regulations from the governmental level or external mandates. Given the timing of the study, one practice was rolling out the COVID-19 vaccination at the direction of the government at the same time they were trying to troubleshoot IT issues and did not have the resources to participate (from PHN2).

Inner Setting

Structural Characteristics

The most prominent structural characteristic mentioned by participants was the size of one of the clinics. Its small size, having only 1 GP, meant that they had not set up the SMS text message component of the clinical software.

> *We don't have a big amount to do, if you know what I mean. So, it's just something that we never really*

> *set up, because we didn't really need to use it. Like, when we go through our recalls, I just give it to reception and they just call each patient individually.* [P4PN]

Networks and Communication

Another practice found that networks and communication within the practice were not strong enough to maintain a "whole of team" approach. They were not, therefore, able to keep the SMS text message process in their institutional memory.

> *one of the other girls was involved, like organising the SMS' to be sent out, but she's actually left the clinic to take on another job.* [P2PN]

Implementation Climate

Here we examined the capacity for change of the practices. In total, 2 practices had work processes that lacked compatibility with the innovation. As 1 (PN) put it, their "IT guy" provided an "ongoing battle" to get it set up. (P4PN) The other practice noted that their clinical software was not compatible with what the study asked of them (P5PM).

Readiness for Implementation

Readiness for implementation refers to "tangible and immediate indicators of organizational commitment to its decision to implement an innovation." The level of resources that an organization allocates to implementation is an indication of readiness. The CFIR subcode "available resources" came up in the data through a discussion of cost. In total, 1 PHN participant noted that the use of SMS text messages costs more than email and might have been a barrier to practices participating in this substudy.

> *"sometimes they're not too keen on doing huge campaigns on SMS, because it actually costs them money to send the information. Whereas, via email would have been better, but it's just difficult to pull that sort of information from the clinical software systems."* [PHN1]

Characteristics of Individuals

An individual stage of change refers to an individual attitudes toward innovation. In total, 1 PHN participant noted that GPs are sometimes hesitant about change.

> *But a lot of them can be a bit hesitant to sending patients out anything, How will that look? Am I targeting my patients? They're a bit anxious.* [PHN6]

Innovation Characteristics

In total, 1 PHN worker noted that one of the practices found the complexity of the intervention a bit more than they were used to and it made participation difficult:

> *I did have a practice that was struggling a little bit with – because there was the SMS surveys to send out. So I had a practice that was – they weren't very used to using that sort of system. They just did phone recalls and reminders to patients, not SMS.*

A total of 2 "Innovation Characteristics" were found to drive participation, including relative advantage and another way of understanding cost. In total, 1 practice in particular understood the SMS text message substudy offered them a couple of advantages over their usual practice. First, the SMS text message approach was thought to be "proactive and opportunistic," enabling them to use a flexible approach to engage their "really passive" patients (P5PM). Second, the study was thought to provide improved access, giving the practice another alternative to offer patients:

> *I could give you this, I could send this to you via email. I can send it to you by text. So that adds to what we can do, what we can provide.* [P5PM]

For 1 PHN worker, the SMS study was a more modern and cost-effective approach to communicating with their patients:

> *I think they realised that it's another way of communicating with the patients that they really should have been using for a long time. So that's been quite a good benefit for that practice certainly, just getting them on board with a more modern way of working with their patients.* [PHN1]

Suggestions for Future Implementation

Practice managers and clinicians had several suggestions for how researchers could more successfully engage with patients who are from disadvantaged groups. These included preferentially using paper-based surveys that were administered at a place that patients already knew and trusted; having researchers available to assist patients in completing surveys if literacy, manual dexterity, or vision was a problem; and vouchers for patients who complete surveys (suggested US $6.6 [US $1=Aus $1.52]).

Discussion

Principal Findings

We found that sending an SMS text message through general practice did not lead to a high response rate from patients, but a 2-way SMS text message survey to patients who had answered the first SMS text message had a higher response rate. Our approach did not capture participants from more marginalized groups due to constraints related to technology and human factors.

We used 2 processes to explore the use of SMS text messages for data collection in general practice research (1) an SMS text message from the patient's own general practice that contained an external link to an online survey and (2) a 2-way SMS survey that was sent from the research team to patients who agreed to be contacted. We found most general practices experienced technical constraints in sending SMS text messages, such as limited software infrastructure. A researcher needed to be proactive in communicating with the clinic's professional and administrative staff to adapt each clinic's unique practice processes to the SMS text message survey procedure. We also used a poster in the waiting room of one practice at the practice's request; however, no patients used this QR code and it did not prove to be a practical strategy in this culturally and linguistically diverse patient group.

Only a small number of participants clicked the external link to the online survey within the initial SMS text message, but it was a comparable percentage to primary care surveys in general [12]. The ongoing 2-way SMS text message surveys were more successful as the patients had agreed to be contacted again and were able to answer questions within the SMS text message environment. Response rates remained over 50% for the 3-month SMS text messages, with no additional contact from the research team or the general practice.

CFIR provided a structure to tease out characteristics that influenced the successful uptake of the SMS text message study from the qualitative data [22]. The most relevant outer setting factors that impacted participation included the needs and resources of the patient populations, notably English language proficiency, age, and 1 practice, which worked primarily with addiction. The data further indicates that size, teamwork, and capacity for change all influenced the uptake of the SMS text message study.

While SMS text message surveys are an attractive option for primary care research, there are constraints to consider. These include a restriction in the type of questions that can be asked due to the limits of SMS text message length and closed question format. Furthermore, SMS text messages are unlikely to capture socioeconomically disadvantaged populations who have barriers, including phone infrastructure and hesitancy around unknown contact numbers. With a recent increase in SMS text message use in scams, this hesitance may increase in the future [23]. Researchers should also be aware that practices are likely to have their own policies or software limitations in sending SMS text messages to patients, and this is likely to influence recruitment and response rates, as seen in this study.

Comparison With Previous Work

The lower response rate in our study is still in line with the international literature where response rates for SMS text message data collection vary considerably, from 12.5% to 100%, with the lowest uptake (12.5%) in a drug and alcohol clinic [24], while response rates beyond 90% were seen a decade ago in general practice, especially when youth were involved [25-27]. A patient's willingness is likely dependent on a number of factors, including familiarity with texting, relationship with the individual recruiting them, health-related motivation or interest in research, and the presence of incentives [28]. Most longitudinal SMS text message surveys also see response rates typically decline, by an estimated 2%-13%, over the duration of the study, which is comparable to our study's decline of 8% from baseline to the 9-month follow-up [24,29-31]. In studies where participants did not respond to an initial SMS text message, high responsiveness was seen after reminder messages were sent [5,32]. Other studies have called participants who did not respond to prompts, which was an effective method of retaining participants in the study, although this does increase the research staff's time and requires consent for contact by phone call [8].

We kept our 2-way SMS text message survey as short as possible to improve response and completion rates [10]. Among patients who ceased surveys before completion or stated they would not be willing to complete a future text survey, a commonly cited reason was that the initial survey had too many questions [29,33]. More than half (52%) of patients in a Singaporean study stated they preferred surveys with 1-10 questions, and only 12% stated they would be willing to complete an SMS survey with more than 30 questions [29]. In 1 study, participants were asked what they would consider an acceptable number of SMS text messages to receive in 1 week from researchers, with a mean response of 4 (SD 3.7) [24], while most participants in another study felt 2 SMS text message questions per day was sufficient [9].

This study included about one-third of participants from a low-income group. The digital divide recognizes easier digital health access for the more advantaged patients [34], and patients have reported the cost of texting as a reason not to participate in an SMS text message trial [24]. The type of phone a patient uses may also influence completion, with patients who use smartphones more likely to complete SMS text message surveys than those using older, more basic phones [33]. With these recognized barriers, this study also highlighted the additional research effort that should be afforded to capture data from the most marginalized patients.

Limitations

A key limitation of this study is the absence of qualitative data from patients about their experience of SMS text message surveys. Although clinicians and PMs could draw on their broader experience, they may have made some assumptions about patient preferences that we could not verify with patients themselves. Our low response rate, while a finding in itself, limited what we were able to learn about SMS text message surveys in lower income populations. We used the general practice as a trusted source to deliver the initial SMS text message, with the assumption that this would increase the response rate, but this did not appear to facilitate SMS text message responses. We are also unable to report if SMS text messages were actually received, which may artificially lower the response rate. For example, this could occur if the patient had changed their phone number since they were last at the practice. Our quantitative data could not make direct comparisons with those from other groups, and further work is required to determine if other strategies may make SMS text message survey work feasible in lower income primary care populations.

Conclusions

While SMS text message survey methods offer a low-intensity option for research data collection, a general SMS text message survey is unlikely to capture participants from more marginalized groups. When recruiting patients through general practice, researchers need to consider the different practice protocols that may be in place for contacting patients by SMS text message as this greatly influences the potential of the method. To promote research participation from the most socially disadvantaged groups, paper-based, or researcher-facilitated surveys undertaken at a trusted location may yield more responses.

Acknowledgments

This research was funded by the Victorian Health Promotion Foundation Impact Research Grant.

Conflicts of Interest

AW was an employee at Victorian Health Promotion Foundation at the time of this research.

Multimedia Appendix 1
Reducing Alcohol-Harm in General Practice project online patient survey.
[DOCX File , 23 KB - mhealth_v12i1e55354_app1.docx]

References

1. Silver L. Smartphone ownership is growing rapidly around the world, but not always equally. Pew Research Center. 2019. URL: https://www.pewresearch.org/global/2019/02/05/smartphone-ownership-is-growing-rapidly-around-the-world-but-not-always-equally/ [accessed 2024-08-01]
2. Scott EE, van der Nagel E, Julian T. CCi digital futures 2014: the internet in Australia. Analysis and Policy Observatory. 2014. URL: https://apo.org.au/sites/default/files/resource-files/2014-12/apo-nid53125.pdf [accessed 2024-08-25]
3. Kohut A, Wike R, Horowitz JM, Simmons K, Poushter J, Barker C, et al. Global digital communication: texting, social networking popular worldwide. Pew Research Center. 2011. URL: https://www.pewresearch.org/global/wp-content/uploads/sites/2/2011/12/Pew-Global-Attitudes-Technology-Report-FINAL-December-20-2011.pdf [accessed 2024-08-01]
4. Hall AK, Cole-Lewis H, Bernhardt JM. Mobile text messaging for health: a systematic review of reviews. Annu Rev Public Health 2015;36:393-415 [FREE Full text] [doi: 10.1146/annurev-publhealth-031914-122855] [Medline: 25785892]
5. Duane S, Tandan M, Murphy AW, Vellinga A. Using mobile phones to collect patient data: lessons learned from the SIMPle study. JMIR Res Protoc 2017;6(4):e61 [FREE Full text] [doi: 10.2196/resprot.6389] [Medline: 28442451]
6. Shimoni N, Nippita S, Castaño PM. Best practices for collecting repeated measures data using text messages. BMC Med Res Methodol 2020;20(1):2 [FREE Full text] [doi: 10.1186/s12874-019-0891-9] [Medline: 31900108]
7. Sammut R, Griscti O, Norman I. Strategies to improve response rates to web surveys: a literature review. Int J Nurs Stud 2021;123:104058. [doi: 10.1016/j.ijnurstu.2021.104058] [Medline: 34454334]
8. Stockwell MS, Cano M, Jakob K, Broder KR, Gyamfi-Bannerman C, Castaño PM, et al. Feasibility of text message influenza vaccine safety monitoring during pregnancy. Am J Prev Med 2017;53(3):282-289 [FREE Full text] [doi: 10.1016/j.amepre.2017.03.014] [Medline: 28495223]
9. Chang T, Gossa W, Sharp A, Rowe Z, Kohatsu L, Cobb EM, et al. Text messaging as a community-based survey tool: a pilot study. BMC Public Health 2014;14:936 [FREE Full text] [doi: 10.1186/1471-2458-14-936] [Medline: 25201051]
10. West BT, Ghimire D, Axinn WG. Evaluating a modular design approach to collecting survey data using text messages. Surv Res Methods 2015;9(2):111-123 [FREE Full text] [Medline: 26322137]
11. Butler SF, Villapiano A, Malinow A. The effect of computer-mediated administration on self-disclosure of problems on the addiction severity index. J Addict Med 2009;3(4):194-203 [FREE Full text] [doi: 10.1097/ADM.0b013e3181902844] [Medline: 20161486]
12. Booker QS, Austin JD, Balasubramanian BA. Survey strategies to increase participant response rates in primary care research studies. Fam Pract 2021;38(5):699-702. [doi: 10.1093/fampra/cmab070] [Medline: 34213547]
13. Chan L, El-Haddad N, Freeman B, O'Hara BJ, Woodland L, Harris-Roxas B. A case study of an SMS text message community panel survey and its potential for use during the COVID-19 pandemic. JMIR Form Res 2021;5(11):e28929 [FREE Full text] [doi: 10.2196/28929] [Medline: 34612824]
14. Wiley K, Goulding A. The conceptualization of digital inclusion in government policy: a qualitative content analysis. In: Information for a Better World: Normality, Virtuality, Physicality, Inclusivity. Switzerland: Springer, Cham; 2023:527-543.
15. Humphry J. 'Second-Class' access: smartphone dependence and the mobile marketplace. In: Homelessness and Mobile Communication: Precariously Connected. Singapore: Springer Nature; 2022:65-91.
16. Osmanlliu E, Paquette J, Grenier AD, Lewis P, Bouthillier ME, Bédard S, et al. Fantastic perspectives and where to find them: involving patients and citizens in digital health research. Res Involv Engagem 2022;8(1):37 [FREE Full text] [doi: 10.1186/s40900-022-00374-6] [Medline: 35918730]
17. Sturgiss E, Gunatillaka N, Ball L, Lam T, Nielsen S, O'Donnell R, et al. Embedding brief interventions for alcohol in general practice: a study protocol for the REACH Project feasibility trial. BJGP Open 2021;5(4):BJGPO.2021.0037 [FREE Full text] [doi: 10.3399/BJGPO.2021.0037] [Medline: 33910916]
18. Sturgiss E, Advocat J, Lam T, Nielsen S, Ball L, Gunatillaka N, et al. Multifaceted intervention to increase the delivery of alcohol brief interventions in primary care: a mixed-methods process analysis. Br J Gen Pract 2023;73(735):e778-e788 [FREE Full text] [doi: 10.3399/BJGP.2022.0613] [Medline: 37666514]

19. Patient experiences.: Australian Bureau of Statistics; 2023. URL: https://www.abs.gov.au/statistics/health/health-services/patient-experiences/latest-release [accessed 2024-08-01]
20. Mobile nation 2022.: Deloitte; 2022. URL: https://www.deloitte.com/au/en/services/economics/perspectives/mobile-nation.html [accessed 2024-08-01]
21. Booth M, Boxall AM. Commissioning services and primary health networks. Aust J Prim Health 2016;22(1):3-4. [doi: 10.1071/PY15167] [Medline: 26907469]
22. Damschroder L, Hall C, Gillon L, Reardon C, Kelley C, Sparks J, et al. The Consolidated Framework for Implementation Research (CFIR): progress to date, tools and resources, and plans for the future. Implementation Sci 2015;10(Suppl 1):A12. [doi: 10.1186/1748-5908-10-s1-a12]
23. Lallie HS, Shepherd LA, Nurse JRC, Erola A, Epiphaniou G, Maple C, et al. Cyber security in the age of COVID-19: a timeline and analysis of cyber-crime and cyber-attacks during the pandemic. Comput Secur 2021;105:102248 [FREE Full text] [doi: 10.1016/j.cose.2021.102248] [Medline: 36540648]
24. Taylor HE, Deane FP, Blackman R. Feasibility of using short message service (SMS) to collect outcome data in an Australian residential alcohol and drug treatment service. Addict Disord Their Treat 2018;17(2):65-75 [FREE Full text]
25. Haller DM, Sanci LA, Patton GC, Sawyer SM. Text message communication in primary care research: a randomized controlled trial. Fam Pract 2009;26(4):325-330. [doi: 10.1093/fampra/cmp040] [Medline: 19542193]
26. Cooper G, Walker J, Harris D, Stewart R, Nicol D, Ogg M. Use of text messaging to audit early clinical outcome following vasectomy in primary care. Br J Gen Pract 2011;61(585):280-282 [FREE Full text] [doi: 10.3399/bjgp11X567135] [Medline: 21439186]
27. Corrigan MA, McHugh SM, Murphy RKJ, Dhillon P, Shah A, Hennessy I, et al. Improving surgical outpatient efficiency through mobile phone text messaging. Surg Innov 2011;18(4):354-357. [doi: 10.1177/1553350611403771] [Medline: 21521699]
28. Brueton VC, Tierney J, Stenning S, Harding S, Meredith S, Nazareth I, et al. Strategies to improve retention in randomised trials. Cochrane Database Syst Rev 2013(12):MR000032 [FREE Full text] [doi: 10.1002/14651858.MR000032.pub2] [Medline: 24297482]
29. Lee SSS, Xin X, Lee WP, Sim EJ, Tan B, Bien MPG, et al. The feasibility of using SMS as a health survey tool: an exploratory study in patients with rheumatoid arthritis. Int J Med Inform 2013;82(5):427-434. [doi: 10.1016/j.ijmedinf.2012.12.003] [Medline: 23332386]
30. Richmond SJ, Keding A, Hover M, Gabe R, Cross B, Torgerson D, et al. Feasibility, acceptability and validity of SMS text messaging for measuring change in depression during a randomised controlled trial. BMC Psychiatry 2015;15:68 [FREE Full text] [doi: 10.1186/s12888-015-0456-3] [Medline: 25886377]
31. Ríos-Bedoya CF, Hay C. Feasibility of using text messaging for unhealthy behaviors screening in a clinical setting: a case study on adolescent hazardous alcohol use. J Am Med Inform Assoc 2013;20(2):373-376 [FREE Full text] [doi: 10.1136/amiajnl-2011-000688] [Medline: 22759622]
32. Christie A, Dagfinrud H, Dale Ø, Schulz T, Hagen KB. Collection of patient-reported outcomes;--text messages on mobile phones provide valid scores and high response rates. BMC Med Res Methodol 2014;14:52 [FREE Full text] [doi: 10.1186/1471-2288-14-52] [Medline: 24735061]
33. Rai M, Moniz MH, Blaszczak J, Richardson CR, Chang T. Real-time data collection using text messaging in a primary care clinic. Telemed J E Health 2017;23(12):955-963. [doi: 10.1089/tmj.2017.0022] [Medline: 28537527]
34. Choy MA, Sturgiss E, Goodyear-Smith F, Smith GJ. Digital health tools and patients with drug use disorders: qualitative patient experience study of the electronic case-finding and help assessment tool (eCHAT). J Med Internet Res 2020;22(9):e19256 [FREE Full text] [doi: 10.2196/19256] [Medline: 32924959]

Abbreviations

CFIR: Consolidated Framework for Implementation Research
GP: general practitioner
PHN: primary health network
PM: practice manager
PN: practice nurse
REACH: Reducing Alcohol-Harm in General Practice

Edited by L Buis; submitted 13.12.23; peer-reviewed by H Fang, I Wilson; comments to author 29.02.24; revised version received 17.03.24; accepted 16.07.24; published 05.09.24.

Please cite as:
Sturgiss E, Advocat J, Barton C, Walker EN, Nielsen S, Wright A, Lam T, Gunatillaka N, Oad S, Wood C
Using Text Messaging Surveys in General Practice Research to Engage With People From Low-Income Groups: Multi-Methods Study
JMIR Mhealth Uhealth 2024;12:e55354
URL: https://mhealth.jmir.org/2024/1/e55354
doi:10.2196/55354
PMID:39235843

Original Paper

A Texting- and Internet-Based Self-Reporting System for Enhanced Vaccine Safety Surveillance With Insights From a Large Integrated Health Care System in the United States: Prospective Cohort Study

Debbie E Malden[1], MSci, DPhil; Julianne Gee[2], MPH; Sungching Glenn[1], MS; Zhuoxin Li[1], MS; Denison S Ryan[1], MPH; Zheng Gu[1], MS; Cassandra Bezi[1], MS; Sunhea Kim[1], MPH; Amelia Jazwa[2], MSPH; Michael M McNeil[2], MD, MPH; Eric S Weintraub[2], MPH; Sara Y Tartof[1,3], MPH, PHD

[1]Department of Research & Evaluation, Kaiser Permanente Southern California, Pasadena, CA, United States

[2]Immunization Safety Office, Centers for Disease Control and Prevention, Atlanta, GA, United States

[3]Kaiser Permanente Bernard J. Tyson School of Medicine, Pasadena, CA, United States

Corresponding Author:
Debbie E Malden, MSci, DPhil
Department of Research & Evaluation
Kaiser Permanente Southern California
100 S Los Robles Ave
Pasadena, CA, 91101
United States
Phone: 1 310 456 4324
Email: d.malden.11@aberdeen.ac.uk

Abstract

Background: SMS text messaging- and internet-based self-reporting systems can supplement existing vaccine safety surveillance systems, but real-world participation patterns have not been assessed at scale.

Objective: This study aimed to describe the participation rates of a new SMS text messaging- and internet-based self-reporting system called the Kaiser Permanente Side Effect Monitor (KPSEM) within a large integrated health care system.

Methods: We conducted a prospective cohort study of Kaiser Permanente Southern California (KPSC) patients receiving a COVID-19 vaccination from April 23, 2021, to July 31, 2023. Patients received invitations through flyers, SMS text messages, emails, or patient health care portals. After consenting, patients received regular surveys to assess adverse events up to 5 weeks after each dose. Linkage with medical records provided demographic and clinical data. In this study, we describe KPSEM participation rates, defined as providing consent and completing at least 1 survey within 35 days of COVID-19 vaccination.

Results: Approximately, 8% (164,636/2,091,975) of all vaccinated patients provided consent and completed at least 1 survey within 35 days. The lowest participation rates were observed for parents of children aged 12-17 years (1349/152,928, 0.9% participation rate), and the highest participation was observed among older adults aged 61-70 years (39,844/329,487, 12.1%). Persons of non-Hispanic White race were more likely to participate compared with other races and ethnicities (13.1% vs 3.9%-7.5%, respectively; $P<.001$). In addition, patients residing in areas with a higher neighborhood deprivation index were less likely to participate (5.1%, 16,503/323,122 vs 10.8%, 38,084/352,939 in the highest vs lowest deprivation quintiles, respectively; $P<.001$). Invitations through the individual's Kaiser Permanente health care portal account and by SMS text message were associated with the highest participation rate (19.2%, 70,248/366,377 and 10.5%, 96,169/914,793, respectively), followed by email (19,464/396,912, 4.9%) and then QR codes on flyers (25,882/2,091,975, 1.2%). SMS text messaging–based surveys demonstrated the highest sustained daily response rates compared with internet-based surveys.

Conclusions: This real-world prospective study demonstrated that a novel digital vaccine safety self-reporting system implemented through an integrated health care system can achieve high participation rates. Linkage with participants' electronic health records is another unique benefit of this surveillance system. We also identified lower participation among selected vulnerable populations, which may have implications when interpreting data collected from similar digital systems.

(JMIR Mhealth Uhealth 2024;12:e58991) doi:10.2196/58991

KEYWORDS

digital health; survey participation; vaccine safety monitoring; COVID-19 vaccines; vaccine; vaccine safety; vaccine monitoring; text; text message; USA; US; surveillance; internet based; survey; monitoring; cohort study; self-reporting; vaccination; COVID-19 vaccination; medical records; text-based surveys; survey; surveys; surveillance system; data collection; disparity; vulnerable; EHR; electronic health records; mobile phone

Introduction

Globally, there has been a rapid adoption of digital technologies, partly due to widespread access to mobile phones [1]. Mobile health (mHealth) has the power to revolutionize health care delivery and monitoring systems. Due to their ability to collect real-time self-reported symptom data from a large number of patients, mHealth systems have proven particularly useful in post licensure vaccine monitoring surveillance [2,3]. However, the broader applicability and significance of mHealth extends across multiple clinical domains, including chronic disease management, mental health support, and preventative care [4-7], underscoring the need to ensure their equitable implementation.

Despite this clear need, digital health care tools have been criticized for the underrepresentation of selected population groups, such as persons of different races or ethnicities or older age [8]. This phenomenon, known as the "digital divide" [9], impacts the generalizability of findings and risks exacerbating existing barriers to health care and health disparities [10,11]. Furthermore, in addition to enhancing equity, understanding and accounting for major drivers of participation also has important implications when interpreting data collected from mHealth systems. Without accounting for these systemic biases, epidemiological associations could be substantially altered, as demonstrated previously [12]. However, previous studies have lacked data with enough granularity within a closed system to identify population-level participation patterns.

These studies include several smartphone-based reporting systems for postvaccine reporting of adverse events (AEs) following immunizations [13-17]. However, since most systems rely on voluntary app-based smartphone enrollment, uptake has generally been low (often under 5% of vaccinated individuals) [3,18-22], contributing to a lack of representativeness. In addition, none of these studies have linked self-reported symptom data with individual-level electronic health record (EHR) data, limiting their ability to contextualize demographic or clinical information against that of the total vaccinated population and to validate clinical diagnoses. Evaluating participation in a similar digital system implemented across a large integrated health care system will improve the interpretation of data collected from SMS text messaging- and internet-based self-reporting systems and can inform the design of future systems. In January 2021, Kaiser Permanente Southern California (KPSC) developed the Kaiser Permanente Side Effect Monitor (KPSEM), a digital survey tool that allows patients to report on potential adverse events following COVID-19 vaccination. In this study, we aimed to assess KPSEM participation by demographic and clinical characteristics within a large integrated health care system.

Methods

Study Population

This prospective cohort study of all patients receiving COVID-19 vaccinations through KPSC from April 23, 2021, through July 31, 2023, monitored self-reported adverse events and solicited symptoms following vaccination.

Data Collection

Patients receiving COVID-19 vaccinations at a KPSC facility could join the system online by scanning a QR code available on study flyers and posters using their smartphone devices. Except for SMS text messages, all study recruitment and communication materials in this study were branded with standard Kaiser Permanente institutional affiliations, including the study flyer, emails, and portal messages. Patients who did not sign up on the day of vaccination received an invitation by SMS text message, email, or a notification through their online Kaiser Permanente health care portal account. This included patients with Kaiser Permanente membership who were vaccinated by an external provider and identified through insurance claims databases and other integrated sources. Parents or guardians provided consent and submitted responses on behalf of their child under 18 years of age or their legal dependent. Initially, the Kaiser Permanente health care portal message center and SMS text message invitations were prioritized over email if patients had online accounts and valid contact information on file since these channels were found to result in higher participation rates during a pilot study (unpublished data). However, in December 2021, all health care portal communication was stopped due to concern of it interfering with care delivery messages, and SMS text messages became the prioritized channel of communication for initial invites. During the consent procedure, participants were asked whether they would prefer to complete the survey exclusively by SMS text message rather than online. After providing consent, patients were sent surveys inquiring about solicited AEs and symptoms at regular intervals according to their preferred contact method for up to 5 weeks after each dose; daily for the first week, alternate days for the second week, then weekly for 3 additional weeks. Survey wording and timing of survey questions are provided in Tables S1 and S2 in Multimedia Appendix 1. Upon receipt of additional vaccine doses, the survey cycle was restarted. Participants could actively withdraw from the survey at any point and would not be recontacted for subsequent doses. Survey information was subsequently combined with EHR data for access to patients' demographic and clinical data across all care settings. Surveys were also made available in Spanish according to a patient's on-file preference.

Statistical Analysis

Selected characteristics were described among all patients who received a COVID-19 vaccination at KPSC over the study period

and among those opting to participate in KPSEM. Participation was defined as completing at least 1 survey within 35 days of their vaccination. Selected clinical and demographic characteristics of interest included age, sex, race or ethnicity, socioeconomic status, and the presence of chronic comorbidities within 1 year before the date of first vaccination within the study period. Prespecified *ICD-10* (*International Statistical Classification of Diseases, Tenth Revision*) diagnosis codes were used to define comorbidities (Tables S3 and S4 in Multimedia Appendix 1). The neighborhood deprivation index (NDI) was used as an indicator of community-level socioeconomic status [23]. Race or ethnicity was categorized using mutually exclusive self-determined categories of non-Hispanic race (White, Black, or Asian), ethnicity (Hispanic), other (not within race or ethnicity groupings), or unknown race or ethnicity. Differences in survey participation statistics, survey reminder preferences (ie, SMS text message, email, or Kaiser Permanente portal message reminders), and self-reported solicited AEs and symptoms were described across selected demographic and clinical characteristics. Similarly, differences in participation (response rates and withdrawal rates over time) were described by separate recruitment and reminder channels, irrespective of demographic and clinical characteristics. Participant characteristics were compared using the chi-square test. All analyses were done with SAS Enterprise Guide statistical software (version 7.1) and R (version 4.3.0, R Foundation for Statistical Computing). Results were reported according to the updated CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) guidelines [24] and the CHERRIES (Checklist for Reporting Results of Internet E-Surveys) [25].

Ethical Considerations

All study activities were reviewed by the Centers for Disease Control and Prevention (CDC) and completed in accordance with applicable Federal law and CDC policy. In addition, the study protocol was reviewed and approved by the KPSC institutional review board, which waived the requirement for informed consent (#12769).

Results

Study Population and Recruitment

Among 2,091,975 patients who were vaccinated between April 23, 2021, and July 31, 2023 (Table 1), 164,636 (7.9%) enrolled in the KPSEM system and completed at least 1 survey in the 35 days following the first dose of COVID-19 vaccination they received within the study period. This included patients who were vaccinated at KPSC facilities and KPSC members who received vaccination at non-KPSC facilities. Passive recruitment methods through patients scanning QR codes on flyers and posters at the time of vaccination enrolled only a small proportion of all vaccinated individuals (25,882/2,092,824, 1.2%; Figure 1 and Table S4 in Multimedia Appendix 1). Among persons contacted through their Kaiser Permanente health care portal account, 19.2% (70,248/366,377) joined and submitted at least 1 survey following their vaccination. Hence, this method was the most successful recruitment channel. SMS text message invitations resulted in a participation rate of 10.5% (96,169/914,793). Email invitations resulted in the lowest participation rate compared with the other active invitation channels (4.9%, 19,464/396,912). Among all individuals who received at least 1 digital invitation across all vaccine doses, participation was 13.1% (181,462/1,382,095; Table S4 in Multimedia Appendix 1).

Table 1. Characteristics of Kaiser Permanente Side Effect Monitor survey population receiving a COVID-19 vaccination from April 23, 2021, to July 31, 2023.

	All vaccinated patients, n			Participation rates[a]					
	First dose[b]	Second dose[b]	Third dose[b]	First dose [b]		Second dose[b]		Third dose[b]	
				n (%)	*P* value[c]	n (%)	*P* value[c]	n (%)	*P* value[c]
Total	2,091,975	1,087,240	352,086	164,636 (7.9)		56,268 (5.2)		26,054 (7.4)	
Continuous KP[d] membership[e]					<.001		<.001		<.001
Yes	1,738,566	962,252	333,236	148,141 (8.5)		53,545 (5.6)		25,462 (7.6)	
No	353,409	124,988	18,850	16,495 (4.7)		2723 (2.2)		592 (3.1)	
Age at vaccination (years)					<.001		<.001		<.001
0-4	35,594	28,895	12,467	2266 (6.4)		706 (2.4)		188 (1.5)	
5-11	133,418	112,032	35,673	8001 (6)		4301 (3.8)		1381 (3.9)	
12-17	152,928	92,191	23,789	1349 (0.9)		860 (0.9)		511 (2.1)	
18-30	274,968	106,297	17,353	10,599 (3.9)		2061 (1.9)		519 (3.0)	
31-40	265,379	99,747	16,253	19,032 (7.2)		4185 (4.2)		883 (5.4)	
41-50	268,120	104,102	17,036	20,834 (7.8)		5346 (5.1)		1127 (6.6)	
51-60	316,438	148,062	45,380	31,429 (9.9)		9831 (6.6)		4277 (9.4)	
61-70	329,487	188,122	78,429	39,844 (12.1)		15,000 (8.0)		8356 (10.7)	
71+	315,643	207,792	105,706	31,282 (9.9)		13,978 (6.7)		8812 (8.3)	
Sex					<.001		<.001		<.001
Female	1,135,626	586,217	191,269	102,445 (9.0)		34,433 (5.9)		15,700 (8.2)	
Male	956,349	501,023	160,817	62,191 (6.5)		21,835 (4.4)		10,354 (6.4)	
Race and ethnicity					<.001		<.001		<.001
Hispanic	775,176	397,035	109,292	42,638 (5.5)		13,403 (3.4)		5080 (4.6)	
Non-Hispanic Asian	253,029	134,827	49,590	16,675 (6.6)		5069 (3.8)		2285 (4.6)	
Non-Hispanic Black	174,759	105,912	37,813	13,033 (7.5)		5197 (4.9)		2669 (7.1)	
Other	117,099	59,152	18,778	8448 (7.2)		2729 (4.6)		1162 (6.2)	
Non-Hispanic White	581,531	310,958	120,255	76,451 (13.1)		27,869 (9.0)		14,078 (11.7)	
Unknown	190,381	79,356	16,358	7391 (3.9)		2001 (2.5)		780 (4.8)	
NDI quintile[f]					<.001		<.001		<.001
1 (least deprived)	352,939	184,330	69,342	38,084 (10.8)		13,384 (7.3)		6855 (9.9)	
2	448,763	232,809	80,942	42,978 (9.6)		14,744 (6.3)		6968 (8.6)	
3	496,499	258,213	84,071	38,341 (7.7)		13,172 (5.1)		5997 (7.1)	
4	465,142	242,381	71,573	28,461 (6.1)		9411 (3.9)		3997 (5.6)	
5 (most deprived)	323,122	167,771	45,780	16,503 (5.1)		5499 (3.3)		2216 (4.8)	
Unknown	5510	1736	378	269 (4.9)		58 (3.3)		21 (5.6)	

[a]Participation rate was calculated as: Participants with at least 1 documented report through KPSEM within 35 days following receipt of vaccination ÷ total vaccinated population × 100.

[b]COVID-19 vaccine dose is defined as a documented dose received within the defined study period, regardless of patient history of previous doses.

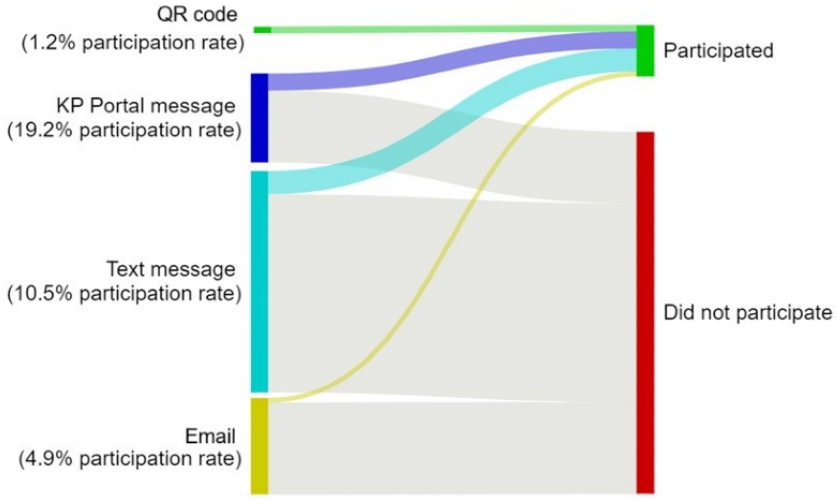
[c]Differences in participation rate were compared across demographic characteristics using the chi-square test for heterogeneity.

[d]KP: Kaiser Permanente.

[e]KP membership is defined as enrollment at least 1 year before vaccination and for 35 days following vaccination.

[f]Neighborhood Deprivation Index (NDI) was defined as the latest available NDI before COVID-19 vaccination.

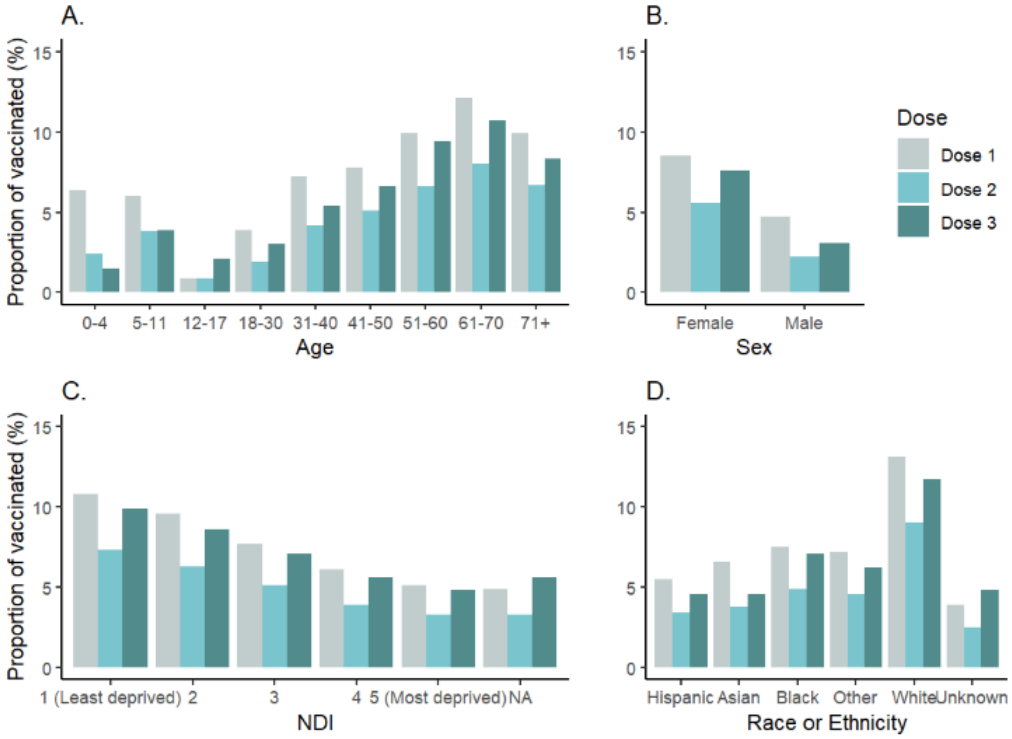
Figure 1. Participation rate by invitation channel among patients receiving a COVID-19 vaccination from April 23, 2021, to July 31, 2023. The participation rate was defined as follows: at least 1 documented report through Kaiser Permanente Side Effect Monitor within 35 days following receipt of vaccination ÷ total invited population × 100. KP: Kaiser Permanente.



Patterns of Participation

Females (102,445/1,135,626, 9%) were more likely to participate than males (62,191/956,349, 6.5%, *P*<.001; Table 1). Parents or guardians joining on behalf of their children demonstrated higher participation rates for younger children aged 0-4 years (2266/35,594, 6.4%) and 5-11 years (8001/133,418, 6%) than parents or guardians of children aged 12-17 years (1349/152,928, 0.9%; Figure 2). Low participation rates were also observed for young adults aged 18-30 years (10,599/274,968, 3.9%). The age group with the highest participation rates was adults aged 61-70 years (39,844/329,487, 12.1%). Participation also varied systematically by race or ethnicity and NDI; persons of non-Hispanic White race were more likely to participate compared with other races or ethnicities (13.1% vs 3.9%-7.5%, respectively, *P*<.001). Patients residing in areas with a higher NDI were less likely to participate (5.1%, 16,503/323,122 vs 10.8%, 38,084/352,939, in the highest vs lowest deprivation quintiles, respectively, *P*<.001). Overall, regardless of demographics, participation decreased for the second dose following dose 1 and then increased following the third dose (Figure 2).

Figure 2. Participation rate by dose and demographic characteristics among patients receiving a COVID-19 vaccination from April 23, 2021, to July 31, 2023. Participation rate was defined as follows: participants with at least 1 documented report through Kaiser Permanente Side Effect Monitor within 35 days following receipt of vaccination/total vaccinated population x 100. NDI: Neighborhood Deprivation Index.

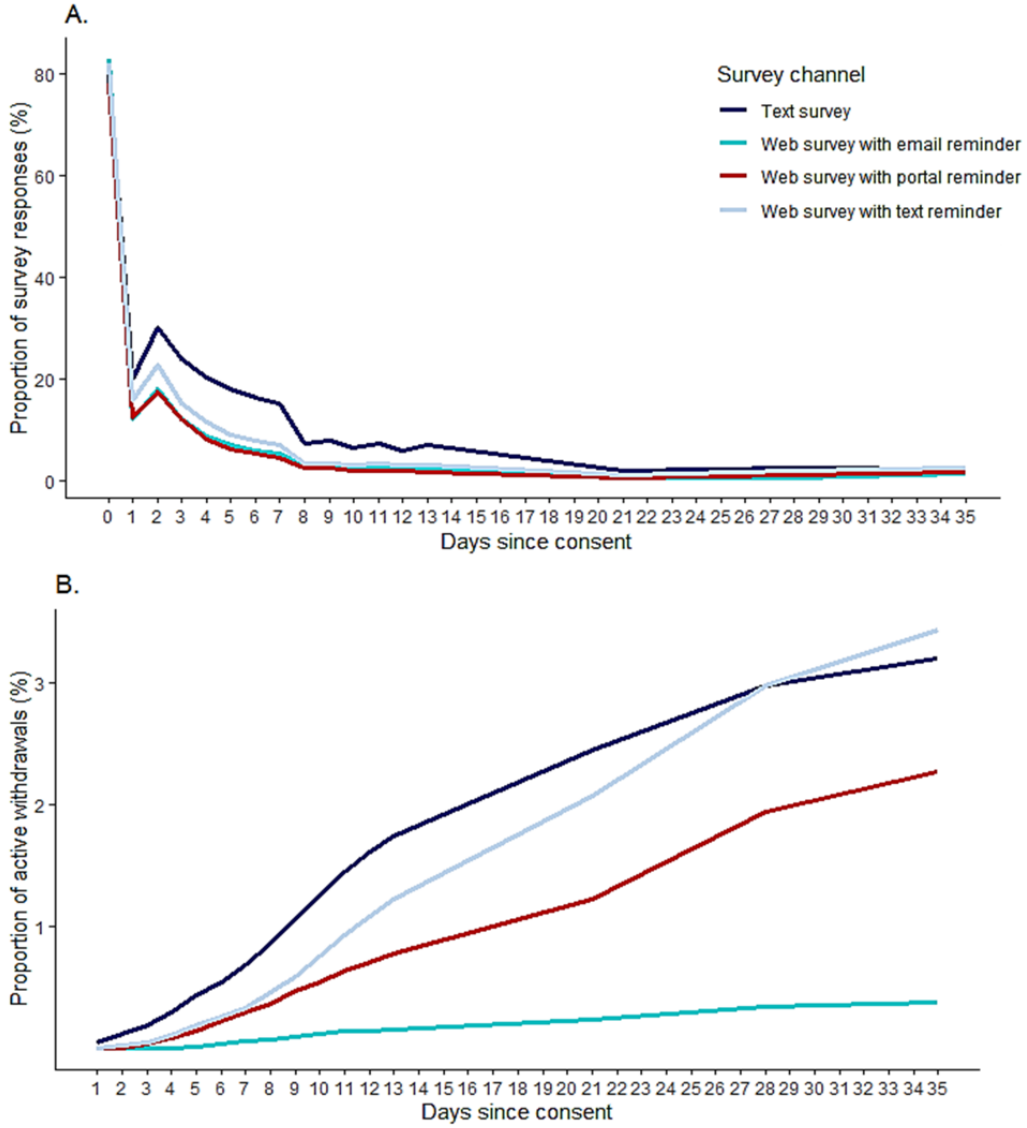


Survey response rates dropped during the first week of survey notifications, from over 80% of participants across all invite channels on the day of consent (day 0) to below 30% of all enrolled participants on day 2 of the survey (Figure 3A). After

the second day of the survey, the response rates consistently dropped through approximately week 3 of the survey and remained low until week 5 (<5% of enrolled participants). However, the sustained response rates differed by survey reminder channel, notably, SMS text message prompts and SMS text message surveys achieved consistently higher sustained responses compared with the other survey channels over 5 weeks. During the 35-day follow-up period, a cumulative proportion of around 3% of study participants withdrew from the study among those who were receiving SMS text message reminder notifications (Figure 3B). Withdrawals were less frequent over the same period among persons who opted to receive notifications through their Kaiser Permanente health care portal account (approximately 2%), and active withdrawals were lowest among persons receiving email reminders (<1%), although email reminders also had the lowest daily response rates (Figure 3A).

Initial reactions (local, systemic, and additional symptoms) were reported more frequently among persons who responded to multiple surveys during the 35-day follow-up period compared with participants who only responded to a single survey (Table S5 in Multimedia Appendix 1), as was the proportion reporting seeking medical care for symptoms; however, initial reports of reactions or solicited symptoms did not affect the likelihood of participation after subsequent vaccine doses (Table S6 in Multimedia Appendix 1).

Figure 3. Rate of survey response (A) and active withdrawals (B) by day since consent and survey channel among Kaiser Permanente Side Effect Monitor participants receiving a COVID-19 vaccination from April 23, 2021, to July 31, 2023. Participation was defined as at least 1 documented report through Kaiser Permanente Side Effect Monitor within 35 days following receipt of COVID-19 vaccination.



Discussion

Principal Findings

Overall, these findings provide evidence for the feasibility of a digital self-reporting system as a timely, flexible, and scalable supplement to existing vaccine safety surveillance systems. We observed a high participation rate among vaccinated patients compared with other digital reporting systems deployed to a general US population. However, we also observed differences in participation across neighborhood-level socioeconomic characteristics and race or ethnicity, suggesting the existence of barriers to participation in these systems. If self-reporting

digital health systems are to be used at scale, outreach efforts must be optimized to ensure that they accurately represent the underlying population. These findings also provide valuable information to improve the design of future digital self-reporting systems and to improve the interpretation and validity of the data collected through such systems.

There could be several reasons for the high participation rate of 30% observed in this study when compared with other digital vaccine safety self-reporting systems such as V-safe (CDC), which enrolled approximately 1-2% of all vaccinees in the United States during the national COVID-19 vaccination program [26]. First, owing to the integrated health care system setting (rather than a federal entity), our system may have had enhanced trust among participants [27,28]. An increase in awareness of the system by staff and patients could have also contributed to enhanced participation since a conscious effort was made to publicize the system through staff awareness sessions and internal communication platforms. Furthermore, recruitment invitations were sent to patients following receipt of multiple vaccine doses, further increasing their awareness of the system and the likelihood of participating with each new dose. In addition, by leveraging the expertise of trained digital teams, our system had many unique features designed to enhance survey usability and acceptance, such as the ability for enrollees to select their preferred survey methods (ie, SMS text message or internet-based surveys) and survey reminder preferences which may have improved participation and survey retention rates. Indeed, we observed that users had a strong preference for SMS text message surveys compared with other channels, and SMS text messages were associated with higher sustained survey response rates over 5 weeks. SMS text messages have proven to be an effective communication method to enhance trust and awareness of vaccine safety previously [29]. Furthermore, the preference for SMS text messaging–based systems over smartphone apps or targeted emails has been noted elsewhere [3,22], and most app-based reporting systems have demonstrated much lower participation rates compared with this study [3,18-20,22]. Previous work with patient usability surveys has identified barriers specifically related to smartphone apps, including lack of device memory, the need for software updates, and the additional time required to download apps [22]. Hence, not requiring a smartphone to participate in our system likely contributed to our observed high participation rates.

Akin to this study, other studies assessing the uptake of self-reporting systems for vaccine safety have identified the highest rates of participation among middle-aged adults and females [3,13,18-21,30]. In addition, we observed significant differences in participation by other demographic and socioeconomic characteristics not as well studied, most notably by NDI, where there was a clear inverse association between community-level socioeconomic deprivation and survey participation. Although general health care disparities predate the pandemic, they have been observed across many aspects of COVID-19–related care, including testing, vaccination, and severe outcomes, even in settings with no direct consumer health care costs such as prepaid insurance premiums [31]. As observed in this study, these disparities are also known to exist with respect to digital self-reporting systems [32]. However, although this area is less studied, encouragingly there is some evidence for reduced disparities for telehealth tools compared with in-person care [33]. The reasons underlying the observed disparities in digital system participation are unknown, but some have suggested that data literacy plays an important role [34]. Future research would benefit from understanding the barriers to participation in digital self-reporting systems, including eliminating gaps in digital literacy that could exacerbate health inequities.

As well as informing approaches for improved participation and potential targeted efforts for equitable uptake of future digital self-reporting systems, this study provides important insights that will aid in the interpretation of large-scale analysis using such data. This has not been feasible at scale previously due to the need for complete EHR data. For example, we identified that persons with immunocompromising conditions were more likely than the general vaccinated population to enroll in the system and, hence, may be overrepresented in similar studies. This is important when interpreting data from similar self-reporting vaccine safety monitoring systems since immunocompromised individuals may systematically differ with respect to reactions following vaccination [35]. In addition, most previous work has relied on enrollees inputting their demographic details and clinical information, which may adversely impact enrollment, reduce survey completion rates, and introduce data entry errors. In addition, one-third of smartphone-based vaccine safety app users in Germany reported that they found entering vaccination details difficult, often not knowing the vaccine's name [19]. This exposure misclassification issue was also associated with certain patient demographics, such as advanced age, hence introducing systematic bias. Furthermore, another important finding from this study was that participation appeared to be slightly higher following a third vaccine dose compared with the second dose across most population subgroups studied. This could have been due to selection bias since third doses were not enforced to the same extent as the primary vaccine series, and hence they were more likely to be received by so-called "early adopters" or older or multimorbid populations, both of which are more likely to take part in research projects.

Data from previous self-reporting vaccine safety systems could be biased by the propensity of persons who have experienced reactions or AEs to be more motivated to engage in self-reporting studies. Interestingly, in this study, individuals who answered multiple surveys within 35 days following receipt of a vaccine were more likely to report local or systemic reactions on their initial report after the same dose and were also more likely to report seeking medical care for their reactions compared with participants who only answered the survey once. Therefore, reactions themselves may be a motivating factor underlying sustained survey participation and hence may lead to overestimations of symptom duration [36]. In contrast, while it was true that certain groups of participants were more likely to enroll after receipt of multiple vaccine doses in this study (parents of young children, persons of white race, etc.), baseline symptom reports did not differ greatly between persons who took part after only 1 dose compared with participation after multiple doses. Hence, we found no evidence that reporting

reactogenicity to a previous vaccine dose influences participation after a subsequent vaccine dose. To date, no studies have investigated this phenomenon due to the need to access EHR for complete follow-up of a vaccinated cohort after multiple doses.

Public Health Implications and Future Directions

Overall, our study findings demonstrated the ability of an SMS text messaging- and internet-based self-reporting system to complement other existing vaccine safety surveillance systems. The findings can also inform the future design of similar digital systems that seek to optimize the direct involvement of the public as empowered stakeholders in scientific research [37]. Furthermore, the population sociodemographic of persons likely to participate in such systems will inform the interpretation of future clinical studies using similar tools. Given the rapid expansion of digital health technology, understanding the potential for biased outputs should remain a priority for health care researchers [34].

Limitations

There are some potential limitations to this study. First, the digital system was designed to operate almost entirely remotely, hence providing a real-world pragmatic overview of the scalability and uptake of our digital system. Consequently, patient usability was not formally assessed in a representative sample of users. However, previous studies incorporating usability surveys have generally lacked representativeness and were conducted among a small sample of participants due to feasibility constraints [22]. Second, although the long study period was a particular strength of this analysis, heightened media attention during the introduction of COVID-19 vaccines likely increased public awareness and possibly influenced participation during earlier months of the vaccine roll-out compared with later months. This could have explained the higher participation rate observed for older persons since they were eligible for vaccination earlier or among those invited through their Kaiser Permanente health care message center compared with other invitation channels because this contact route was halted after November 2021.

Conclusion

In this study, an SMS text messaging- and internet-based self-reporting vaccine safety system enabled reactions to be reported by vaccinees in real time within a large and diverse managed care population. Hence, these findings demonstrated the feasibility of digital systems as timely and scalable methods that could supplement existing vaccine safety surveillance platforms. Our findings also emphasized the importance of implementing such a system through a trusted health care system for higher participation rates and potential clinical follow-up. However, we observed lower participation among selected vulnerable populations, demonstrating the existence of barriers to participation in digital reporting tools. If self-reporting digital tools are to be used in public health, their reach must be optimized to ensure that their implementation is equitable.

Acknowledgments

This study was funded by the US Centers for Disease Control and Prevention but was solely conducted at Kaiser Permanente Southern California. The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention (CDC). Mention of a product or company name is for identification purposes only and does not constitute endorsement by the CDC.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Additional information.
[PDF File (Adobe PDF File), 299 KB - mhealth_v12i1e58991_app1.pdf]

References

1. Pew Research Center. Mobile fact sheet. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2024-01-31]
2. Zhou W, Pool V, Iskander JK, English-Bullard R, Ball R, Wise RP, et al. Surveillance for safety after immunization: vaccine adverse event reporting system (VAERS)--United States, 1991-2001. MMWR Surveill Summ 2003;52(1):1-24 [FREE Full text] [Medline: 12825543]
3. Bota AB, Bettinger JA, Sarfo-Mensah S, Lopez J, Smith DP, Atkinson KM, et al. Comparing the use of a mobile app and a web-based notification platform for surveillance of adverse events following influenza immunization: randomized controlled trial. JMIR Public Health Surveill 2023;9:e39700 [FREE Full text] [doi: 10.2196/39700] [Medline: 37155240]
4. Gordon WJ, Landman A, Zhang H, Bates DW. Beyond validation: getting health apps into clinical practice. NPJ Digit Med 2020;3:14 [FREE Full text] [doi: 10.1038/s41746-019-0212-z] [Medline: 32047860]
5. Leigh JW, Gerber BS, Gans CP, Kansal MM, Kitsiou S. Smartphone ownership and interest in mobile health technologies for self-care among patients with chronic heart failure: cross-sectional survey study. JMIR Cardio 2022;6(1):e31982 [FREE Full text] [doi: 10.2196/31982] [Medline: 35029533]

6. Oudin A, Maatoug R, Bourla A, Ferreri F, Bonnot O, Millet B, et al. Digital phenotyping: data-driven psychiatry to redefine mental health. J Med Internet Res 2023;25:e44502 [FREE Full text] [doi: 10.2196/44502] [Medline: 37792430]
7. Willis VC, Thomas Craig KJ, Jabbarpour Y, Scheufele EL, Arriaga YE, Ajinkya M, et al. Digital health interventions to enhance prevention in primary care: scoping review. JMIR Med Inform 2022;10(1):e33518 [FREE Full text] [doi: 10.2196/33518] [Medline: 35060909]
8. The Lancet Digital Health. Digital health equity for older populations. Lancet Digit Health 2023;5(7):e395 [FREE Full text] [doi: 10.1016/S2589-7500(23)00114-0] [Medline: 37391262]
9. Xiao A, Xu Z, Skare M, Qin Y, Wang X. Bridging the digital divide: the impact of technological innovation on income inequality and human interactions. Humanit Soc Sci Commun 2024;11:809. [doi: 10.1057/s41599-024-03307-8]
10. Ibrahim H, Liu X, Zariffa N, Morris AD, Denniston AK. Health data poverty: an assailable barrier to equitable digital health care. Lancet Digit Health 2021;3(4):e260-e265 [FREE Full text] [doi: 10.1016/S2589-7500(20)30317-4] [Medline: 33678589]
11. Kc S, Tewolde S, Laverty AA, Costelloe C, Papoutsi C, Reidy C, et al. Uptake and adoption of the NHS App in England: an observational study. Br J Gen Pract 2023;73(737):e932-e940 [FREE Full text] [doi: 10.3399/BJGP.2022.0150] [Medline: 37783512]
12. Pouwels K, Eyre DW, House T, Aspey B, Matthews PC, Stoesser N, COVID−19 Infection Survey Team. Improving the representativeness of UK's national COVID-19 infection survey through spatio-temporal regression and post-stratification. Nat Commun 2024;15(1):5340 [FREE Full text] [doi: 10.1038/s41467-024-49201-4] [Medline: 38914564]
13. Antonelli M, Penfold RS, Merino J, Sudre CH, Molteni E, Berry S, et al. Risk factors and disease profile of post-vaccination SARS-CoV-2 infection in UK users of the COVID symptom study app: a prospective, community-based, nested, case-control study. Lancet Infect Dis 2022;22(1):43-55 [FREE Full text] [doi: 10.1016/S1473-3099(21)00460-6] [Medline: 34480857]
14. Chapin-Bardales J, Gee J, Myers T. Reactogenicity following receipt of mRNA-based COVID-19 vaccines. JAMA 2021;325(21):2201-2202. [doi: 10.1001/jama.2021.5374] [Medline: 33818592]
15. Drury RE, O'Connor D. Symptom study app provides real-world data on COVID-19 vaccines. Lancet Infect Dis 2021;21(7):890-891 [FREE Full text] [doi: 10.1016/S1473-3099(21)00264-4] [Medline: 33930321]
16. Houhamdi L, Fournier PE. Smart apps for self-reporting clinical information. Lancet 2022;399(10335):1575-1576 [FREE Full text] [doi: 10.1016/S0140-6736(22)00453-6] [Medline: 35397852]
17. Drury RE, O'Connor D. Symptom study app provides real-world data on COVID-19 vaccines. Lancet Infect Dis 2021;21(7):890-891 [FREE Full text] [doi: 10.1016/S1473-3099(21)00264-4] [Medline: 33930321]
18. Menni C, Klaser K, May A, Polidori L, Capdevila J, Louca P, et al. Vaccine side-effects and SARS-CoV-2 infection after vaccination in users of the COVID symptom study app in the UK: a prospective observational study. Lancet Infect Dis 2021;21(7):939-949 [FREE Full text] [doi: 10.1016/S1473-3099(21)00224-3] [Medline: 33930320]
19. Nguyen MTH, Krause G, Keller-Stanislawski B, Glöckner S, Mentzer D, Ott JJ. Postmarketing safety monitoring after influenza vaccination using a mobile health app: prospective longitudinal feasibility study. JMIR Mhealth Uhealth 2021;9(5):e26289 [FREE Full text] [doi: 10.2196/26289] [Medline: 33960950]
20. Nguyen MTH, Ott JJ, Caputo M, Keller-Stanislawski B, Klett-Tammen CJ, Linnig S, et al. User preferences for a mobile application to report adverse events following vaccination. Pharmazie 2020;75(1):27-31 [FREE Full text] [doi: 10.1691/ph.2020.9734] [Medline: 32033630]
21. Rolfes L, Härmark L, Kant A, van Balveren L, Hilgersom W, van Hunsel F. COVID-19 vaccine reactogenicity - a cohort event monitoring study in the Netherlands using patient reported outcomes. Vaccine 2022;40(7):970-976 [FREE Full text] [doi: 10.1016/j.vaccine.2022.01.013] [Medline: 35067381]
22. Wilson K, Atkinson KM, Westeinde J, Bell C, Marty K, Fergusson D, et al. An evaluation of the feasibility and usability of a proof of concept mobile app for adverse event reporting post influenza vaccination. Hum Vaccin Immunother 2016;12(7):1738-1748 [FREE Full text] [doi: 10.1080/21645515.2016.1152434] [Medline: 26905396]
23. Messer LC, Laraia BA, Kaufman JS, Eyster J, Holzman C, Culhane J, et al. The development of a standardized neighborhood deprivation index. J Urban Health 2006;83(6):1041-1062 [FREE Full text] [doi: 10.1007/s11524-006-9094-x] [Medline: 17031568]
24. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of web-based and mobile health interventions. J Med Internet Res 2011;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
25. Eysenbach G. Improving the quality of web surveys: the checklist for reporting results of internet E-surveys (CHERRIES). J Med Internet Res 2004;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
26. Woo EJ, Gee J, Marquez P, Baggs J, Abara WE, McNeil MM, et al. Post-authorization safety surveillance of Ad.26.COV2.S vaccine: reports to the vaccine adverse event reporting system and v-safe, February 2021-February 2022. Vaccine 2023;41(30):4422-4430 [FREE Full text] [doi: 10.1016/j.vaccine.2023.06.023] [Medline: 37321898]
27. Pollard MS, Davis LM. Decline in trust in the centers for disease control and prevention during the COVID-19 pandemic. Rand Health Q 2022;9(3):23 [FREE Full text] [Medline: 35837520]
28. Bogart LM, Ojikutu BO, Tyagi K, Klein DJ, Mutchler MG, Dong L, et al. COVID-19 related medical mistrust, health impacts, and potential vaccine hesitancy among black Americans living with HIV. J Acquir Immune Defic Syndr 2021;86(2):200-207 [FREE Full text] [doi: 10.1097/QAI.0000000000002570] [Medline: 33196555]

29. Panickar R, Aziz Z, Mohd Sani N, Kamarulzaman A. The use of technology in vaccine safety communication: a systematic review of randomised controlled trials. Patient Educ Couns 2023;112:107707. [doi: 10.1016/j.pec.2023.107707] [Medline: 36989861]
30. Hall VJ, Foulkes S, Saei A, Andrews N, Oguti B, Charlett A, SIREN Study Group. COVID-19 vaccine coverage in health-care workers in England and effectiveness of BNT162b2 mRNA vaccine against infection (SIREN): a prospective, multicentre, cohort study. Lancet 2021;397(10286):1725-1735 [FREE Full text] [doi: 10.1016/S0140-6736(21)00790-X] [Medline: 33901423]
31. Jefferson C, Watson E, Certa JM, Gordon KS, Park LS, D'Souza G, NA-ACCORD Corona-Infectious-Virus Epidemiology Team (CIVET). Differences in COVID-19 testing and adverse outcomes by race, ethnicity, sex, and health system setting in a large diverse US cohort. PLoS One 2022;17(11):e0276742 [FREE Full text] [doi: 10.1371/journal.pone.0276742] [Medline: 36417366]
32. Qian L, Sy LS, Hong V, Glenn SC, Ryan DS, Morrissette K, et al. Disparities in outpatient and telehealth visits during the COVID-19 pandemic in a large integrated health care organization: retrospective cohort study. J Med Internet Res 2021;23(9):e29959 [FREE Full text] [doi: 10.2196/29959] [Medline: 34351865]
33. Palzes VA, Chi FW, Metz VE, Sterling S, Asyyed A, Ridout KK, et al. Overall and telehealth addiction treatment utilization by age, race, ethnicity, and socioeconomic status in California after COVID-19 policy changes. JAMA Health Forum 2023;4(5):e231018 [FREE Full text] [doi: 10.1001/jamahealthforum.2023.1018] [Medline: 37204804]
34. The Lancet Digital Health. Power to the people. Lancet Digital Health 2021;3(8):e462. [doi: 10.1016/s2589-7500(21)00147-3]
35. Chantasrisawad N, Puthanakit T, Tangsathapornpong A, Techasaensiri C, Phongsamart W, Suwanpakdee D, Study Team. Immunogenicity and reactogenicity of mRNA BNT162b2 COVID-19 vaccine among thai adolescents with chronic diseases. Vaccines (Basel) 2022;10(6):871 [FREE Full text] [doi: 10.3390/vaccines10060871] [Medline: 35746478]
36. Dos Santos G, Eckermann T, Martínez-Gómez X, Parra J, Nwoji U, Salamanca de la Cueva I. Enhanced safety surveillance of GSK's quadrivalent seasonal influenza vaccine in Germany and Spain (2021/2022 season) using an electronic patient-reported outcome system for vaccine safety remote monitoring. Influenza Other Respir Viruses 2023;17(3):e13098 [FREE Full text] [doi: 10.1111/irv.13098] [Medline: 36991538]
37. Cohen AB, Mathews SC, Dorsey ER, Bates DW, Safavi K. Direct-to-consumer digital health. Lancet Digit Health 2020;2(4):e163-e165 [FREE Full text] [doi: 10.1016/S2589-7500(20)30057-1] [Medline: 33328077]

Abbreviations

AE: adverse event
CDC: Centers for Disease Control and Prevention
CHERRIES: Checklist for Reporting Results of Internet E-Surveys (CHERRIES Checklist for Reporting Results of Internet E-Surveys
CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
EHR: electronic health record
***ICD-10*:** *International Statistical Classification of Diseases, Tenth Revision*
KPSC: Kaiser Permanente Southern California
KPSEM: Kaiser Permanente Side Effect Monitor
mHealth: mobile health
NDI: Neighborhood Deprivation Index

Edited by L Buis; submitted 29.03.24; peer-reviewed by C Tziraki, A Hause; comments to author 18.06.24; revised version received 29.07.24; accepted 12.08.24; published 11.10.24.

Please cite as:
Malden DE, Gee J, Glenn S, Li Z, Ryan DS, Gu Z, Bezi C, Kim S, Jazwa A, McNeil MM, Weintraub ES, Tartof SY
A Texting- and Internet-Based Self-Reporting System for Enhanced Vaccine Safety Surveillance With Insights From a Large Integrated Health Care System in the United States: Prospective Cohort Study
JMIR Mhealth Uhealth 2024;12:e58991
URL: https://mhealth.jmir.org/2024/1/e58991
doi:10.2196/58991
PMID:

Feasibility, Acceptability, Satisfaction, and Challenges of an mHealth App (e-ASCov) for Community-Based COVID-19 Screening by Community Health Workers in Rwanda: Mixed Methods Study

Abdou Y Omorou[1,2], MD, PhD; Pacifique Ndishimye[3,4], PhD; Bruno Hoen[1,5], MD, PhD; Léon Mutesa[6], MD, PhD; Prosper Karame[4], MSc; Ladislas Nshimiyimana[4], MSc; Simon Galmiche[7], MD, MPH; Hassan Mugabo[4], MSc; Janvier Murayire[8], MD; Muco Mugisha[9], MD; Marie Michele Umulisa[9]; Yvonne Delphine Nsaba Uwera[10], MSc; Clarisse Musanagabanwa[4], PhD; Noella Bigirimana[4]; Sabin Nsanzimana[11], MD, PhD; Francis Guillemin[1,2], MD, PhD; Jean Paul Rwabihama[8,11], MD, PhD

[1]UMR 1319 INSPIIRE, INSERM, Université de Lorraine, Vandoeuvre-Lès-Nancy, France

[10]Midwifery Department, School of Nursing and Midwifery, College of Medicine and Health Sciences, University of Rwanda, Kigali, Rwanda

[11]Ministry of Health Rwanda, Kigali, Rwanda

[2]CHRU Nancy, INSERM, Université de Lorraine, CIC Clinical Epidemiology, Vandoeuvre-Lès-Nancy, France

[3]Laboratory of Emerging and Infectious Diseases, School of Medicine, Dalhousie University, Halifax, NS, Canada

[4]Research, Innovation and Data Science Division, Rwanda Biomedical Centre, Kigali, Rwanda

[5]Direction de la recherche médicale, Institut Pasteur, Paris, France

[6]Centre for Human Genetics, College of Medicine and Health Sciences, University of Rwanda, Kigali, Rwanda

[7]Emerging Diseases Epidemiology Unit, Institut Pasteur, Université Paris Cité, Paris, France

[8]Department of Internal Medicine, College of Medicine and Health Sciences, University of Rwanda, Kigali, Rwanda

[9]Rinda Ubuzima, Kigali, Rwanda

Corresponding Author:
Abdou Y Omorou, MD, PhD
UMR 1319 INSPIIRE, INSERM, Université de Lorraine, , Vandoeuvre-Lès-Nancy, , France

Abstract

Background: Although at the base of the pyramid-shaped organization of the Rwandan health system, community health workers (CHWs) are central to the community-based management of disease outbreaks.

Objective: This mixed methods study aimed to explore the feasibility, acceptability, satisfaction, and challenges of a mobile health (mHealth) tool for community-based COVID-19 screening in Rwanda.

Methods: Two urban (Gasabo and Nyarugenge) and 2 rural (Rusizi and Kirehe) districts in Rwanda participated in the project (smartphone app for COVID-19 screening). A mixed methods approach was used to inform the feasibility (awareness and expectation), acceptability (use and perceived benefits), satisfaction, and challenges of the mHealth intervention. At the end of the project, CHWs were asked to complete a quantitative questionnaire on the use of and satisfaction with the app. Then, in-depth interviews and focus group discussions were organized with CHWs. A content analysis was performed on the transcripts.

Results: Overall, 383 CHWs were recruited and trained; 378 CHWs participated in the study. The mean age of CHWs was 36.7 (SD 6.6) to 45.3 (SD 9.9) years and most were women (237/378, 62.7%). More than 7000 people were registered with the use of the app and 20% were referred to a local COVID-19 testing facility. According to CHW reporting, the median number of people screened by each CHW ranged from 152 (IQR 70-276) for Nyarugenge to 24 (IQR 16-90) for Rusizi. COVID-19 positivity rates were higher in urban than rural districts: more than half of the CHWs in Gasabo reported a confirmed positive case versus only 2.4% for Kirehe and 15.4% for Rusizi. Despite the app being a novel tool, CHWs were well aware of the use of such a tool and had appropriate expectations. Acceptability and satisfaction were very high, with differences between urban and rural districts. Satisfaction was higher in Nyarugenge (72.8/100) and Gasabo (80.7/100) than in Kirehe (61.6/100) and Rusizi (64.5/100). More than 80% of the CHWs were willing to continue using the e-ASCov app, with the exception of CHWs in Kirehe (56.7%). The app was perceived as a tool to generate information on COVID-19, inform on the status of the pandemic, and help curb the spread of the pandemic in Rwanda. CHWs were satisfied with the app at all stages of its implementation in their districts.

Conclusions: In this proof-of-concept study, a smartphone app for screening COVID-19 was useful as an mHealth tool to be used by CHWs, with the potential to increase health system efficiency in an epidemic context. The context should be analyzed for generalization on a country-wide scale, both in case of an epidemic and to take into account certain conditions at the community level. Information is needed on the conditions of generalization and transferability of this type of app to other health conditions so that CHWs can be given their full place in a pyramidal health system.

(JMIR Mhealth Uhealth 2024;12:e50745) doi:10.2196/50745

KEYWORDS

community health workers; COVID-19 screening tool; COVID-19; SARS-CoV-2; screening; acceptability; feasibility; satisfaction; community based; LMIC; Africa; challenges; barriers; smartphone; proof-of-concept; mHealth; mobile health; apps; COVID-19 screening

Introduction

Since December 2019, the appearance of a new strain of coronavirus (SARS-Cov-2) has been at the origin of the current COVID-19 pandemic [1]. In Africa, the first case of COVID-19 was reported on February 14, 2020, in Egypt and on February 27, 2020, in Nigeria, sub-Saharan Africa. With lockdowns, curfews, closures of public places and businesses, and the suspension of international flights, the continent's public health leaders tried to contain the epidemic. Rwanda was greatly affected by the COVID-19 pandemic, with initially a majority of imported cases.

The Rwandan government, through the Ministry of Health and the Rwanda Biomedical Center, a national agency for developing and implementing public health programs, developed a nationwide plan to deal with the COVID-19 epidemic [2]. This action plan aimed to improve Rwanda's ability to prevent, rapidly diagnose, and effectively respond to the COVID-19 epidemic and was structured around prevention and management. Prevention activities were mainly based on epidemiological surveillance, human resources, patient transfer, hygiene reinforcement, social mobilization and coordination, and monitoring activities [3]. A multisectoral rapid response team was activated at central and local levels to respond to the pandemic [4].

The pyramidal organization of the Rwandan health system gives an important place to community health workers (CHWs) at the base of the pyramid. Rwanda has an important network of CHWs throughout the country. Their role is essential in the field of primary health care. These CHWs are the essential link between the population and the health care system at the community level. The activation of this network and the provision of simple and ergonomic tools for epidemic surveillance were seen as an effective and efficient way to combat this epidemic.

The experience of the Ebola epidemic in 3 sub-Saharan African countries (Guinea, Liberia, and Sierra Leone) showed that CHWs played a key role in controlling the epidemic, but it also highlighted the need to involve CHWs very early in the epidemic response to ensure optimal effectiveness [5,6]. Some health actors believed that with this COVID-19 pandemic, CHWs were more than ever indispensable in addressing the disease [7]. However, formal evidence was needed. Although long-term solutions to the pandemic would be based on the rollout of vaccines and effective treatments, digital technologies remained crucial in terms of detection during the pandemic.

In light of its strategic objectives to leverage opportunities and innovations of communication technology and essentially drive the transformation of Rwanda to a knowledge-based economy, Rwanda has embraced innovative technologies as part of the COVID-19 early case detection response [8]. In this regard, the Rwanda Biomedical Center and its partners (CIC de Nancy, and Institut Pasteur, Paris, France) launched the e-ASCov project (Utilisation des outils numériques par les agents de santé communautaire dans la stratégie d'endiguement et d'atténuation de l'épidémie de Covid-19 au Rwanda–une recherche-action). This project aimed to assess the involvement of CHWs and the use of digital tools (mobile health [mHealth] smartphone app) in managing the COVID-19 pandemic in Rwanda in 2020.

This study aimed to assess the feasibility (awareness and expectations), acceptability (use and perceived benefits), satisfaction, and challenges of an mHealth tool for community-based COVID-19 screening in Rwanda.

Methods

The e-ASCov Project

The e-ASCov project is an action-research project that enrolled, trained, and equipped CHWs in 4 administrative districts in Rwanda: 2 urban (Nyarugenge and Gasabo) and 2 rural (Rusizi and Kirehe). CHWs were equipped with a smartphone app (e-ASCov app) to detect and triage suspected COVID-19 cases at the community level and to propose referral to the health care system for appropriate management. A total of 400 CHWs were selected (100 per district) based on 2 criteria: age <50 years and availability and proficiency in the use of a smartphone. The project was carried out from November 2020 to December 2021. Phase 1 of the project emphasized training and equipping CHWs with the e-ASCov app, and facilitating them to use the app in their day-to-day activities. Phase 2 emphasized an evaluation of the implementation of the e-ASCov app for the management of the COVID-19 pandemic. A parallel mixed (quantitative and qualitative) method was used to highlight the implementation of the e-ASCov app for managing COVID-19 at the community level.

Ethical Considerations

This project was approved by the Rwanda National Ethics Committee (number 752/RNEC/2020) as well as the Rwanda

CHW Technical Working Group. The target population consisted of CHWs trained in the use of digital tools (e-ASCov app) for managing the COVID-19 pandemic. Informed consent was required from all participants before using the app, and they had the ability to opt out without any consequence. All data collected were anonymized and without the possibility of participant identification. No compensation was given to participants.

Data Collection

Quantitative Data

All CHWs were asked to complete a questionnaire at the end of the intervention (use of the e-ASCov app by CHWs). The questionnaire covered sociodemographic characteristics (sex, age, educational attainment, occupation, marital status, etc), the description of the intervention (number of persons screened, persons referred, confirmed cases), and satisfaction with the e-ASCov app. Satisfaction was measured on a 5-point Likert scale (very satisfied, satisfied, undecided/neutral, unsatisfied, very unsatisfied) that addressed the following dimensions: quality of training, availability of equipment, internet connectivity, simplicity/use of the tool, time needed to complete one record, getting technical assistance, delivery of service to clients, and continued use of the tool.

Qualitative Data

In-depth interviews (IDIs) and focus group discussions (FGDs) were conducted: 2 FGDs with CHWs and 2 with clients of CHWs per district (4 FGDs per district). Each FGD consisted of 8 participants, for a total of 64 CHWs and 64 clients. For the IDIs, a purposive sampling strategy was used to recruit key members of staff (who were not directly involved in using the app in the field) from the following 3 categories: COVID-19 taskforce members, data managers at central and district levels, and CHW supervisors at central and district levels. Qualitative data management and analysis involved verbatim transcription of interviews, translation, and thematic analysis. Interviews were conducted in Kinyarwanda, transcribed using Microsoft Word, and translated into English. The transcripts were cross-checked for quality by 2 analysts who compared audio records and scripts in Kinyarwanda, then with the English translation, considering what was directly said by the respondent. Each individual transcript was used for analysis.

Data Analysis

Qualitative and quantitative data were analyzed to address the question of feasibility (awareness and expectations), acceptability (use and perceived benefits), satisfaction, and challenges.

Quantitative Data

Sociodemographic characteristics are described with mean and SD for continuous and number (percentage) for categorical variables. We compared characteristics according to districts of participants by Student *t* test or chi-square test as applicable. Then we compared the number of persons screened with or without e-ASCov, persons referred to the district hospital, and confirmed cases by the district of the CHWs. A satisfaction score was calculated and normalized on a scale from 0 to 100 (highest satisfaction) for each dimension, and a global satisfaction score was calculated.

Qualitative Data

An inductive thematic analysis approach was used to analyze qualitative data in the following 6 areas: familiarization, coding (each code described the idea or the feeling expressed in the specific part of the text under the topic of discussion), generating themes and subthemes (ideas and patterns that came up repeatedly), reviewing themes, defining and naming themes, and writing up [9]. The analysis was conducted with Dedoose software (version 9.0.18), a web-based cross-platform for qualitative analysis and mixed methods research.

Results

Sample Characteristics

Of the 400 CWHs recruited, 378 CWHs completed the quantitative questionnaires (Gasabo region: n=93; Kirehe: n=97; Nyarugenge: n=90; Rusizi: n=98). The mean age of CHWs ranged from 36.7 (SD 6.6) to 45.3 (SD 9.9) years (Table 1). CHWs were mostly women (237/378, 62.7%) with primary or secondary educational attainment. CHWs were mostly unemployed in the districts of Gasabo (52/93, 55.9%) and Kirehe (71/97, 73.2%), whereas those from the 2 other districts were mostly employed (Nyarugenge: 53/90, 58.9%; Rusizi: 79/98, 80.6%). Most were experienced CHWs, with a mean number of years of experience of 7.3 to 10.1.

The IDIs involved 19 respondents: 5 CHW supervisors (2 at the central level and 3 at the district level), 5 CHWs from the COVID-19 taskforce (2 at the central level, 3 at the district level), 3 data managers from the district hospitals, 2 director generals of the hospitals, and 4 clinical directors at the district level. The FGDs involved 64 CHWs from all 4 districts included in this study.

Table . Sociodemographic characteristics of community health workers according to district in Rwanda.

Characteristics		Gasabo (n=93, 24.6%)	Kirehe (n=97, 25.7%)	Nyarugenge (n=90, 23.8%)	Rusizi (n=98, 25.9%)	*P* value[a]
Age (years)						<.001
	Participants, n (%)	93 (100.0)	97 (100.0)	90 (100.0)	98 (100.0)	
	Mean (SD)	36.7 (6.6)	43.6 (7.1)	40.5 (7.6)	45.3 (9.9)	
Sex, n (%)						.01
	Male	27 (29.0)	45(46.4)	26 (28.9)	43 (43.9)	
	Female	66 (71.0)	52 (53.6)	64 (71.1)	55 (56.1)	
Type of district, n (%)						<.001
	Rural	0 (0.0)	97 (100.0)	0 (0.0)	98 (100.0)	
	Urban	93 (100.0)	0 (0.0)	90 (100.0)	0 (0.0)	
Educational attainment, n (%)						<.001
	Primary	15 (16.1)	63 (64.9)	20 (22.2)	63 (64.3)	
	Secondary	52 (55.9)	28 (28.9)	54 (60.0)	26 (26.5)	
	TVET[b]	0 (0.0)	3 (3.1)	2 (2.2)	5 (5.1)	
	University	26 (28.0)	3 (3.1)	14 (15.6)	4 (4.1)	
Occupation, n (%)						<.001
	Employed	41 (44.1)	26 (26.8)	53 (58.9)	79 (80.6)	
	Unemployed	52 (55.9)	71 (73.2)	37 (41.1)	19 (19.4)	
Marital status, n (%)						<.001
	Divorced/separated	2 (2.2)	2 (2.1)	2 (2.2)	2 (2.0)	
	Married/living with the partner	64 (68.8)	90 (92.8)	62 (68.9)	86 (87.8)	
	Single/never married	26 (28.0)	2 (2.1)	15 (16.7)	3 (3.1)	
	Widowed	1 (1.1)	3 (3.1)	11 (12.2)	7 (7.1)	
Function, n (%)						<.001
	MCHC[c]	20 (21.5)	13 (13.4)	30 (33.3)	11 (11.2)	
	Binome	56 (60.2)	59 (60.8)	52 (57.8)	76 (77.6)	
	CEHO[d]	10 (10.8)	2 (2.1)	7 (7.8)	3 (3.1)	
	Prevention	7 (7.5)	23 (23.7)	1 (1.1)	8 (8.2)	
Years of experience						.002
	Participants, n (%)	93 (100.0)	97 (100.0)	90 (100.0)	98 (100.0)	
	Mean (SD)	7.3 (4.5)	9.6 (5.7)	8.6 (4.8)	10.1 (6.1)	

[a]*P* value for χ^2 test for qualitative variables or Student *t* test for quantitative variables.

[b]TVET: technical vocation education training.

[c]MCHC: maternal and child health coordinator.

[d]CEHO: community and environmental health officer.

Awareness of and Expectations for the e-ASCov App

Most respondents were aware of the e-ASCov app and believed that it was used to provide information on COVID-19. These respondents were located predominantly in urban districts (Nyarugenge and Gasabo).

> *e-ASCov is the application which is used by the community health workers to help people who are suspected of COVID-19 symptoms* [COVID-19 Taskforce, central level]

The CHWs, especially the respondents with the opportunity to use the app, considered it a tool to generate information on COVID-19.

> *Those who gave us training, they explained to us how to write a report on COVID-19 using forms for identified clients; after filling the forms, we could take them to the health center. That is how we worked* [CHW, Kirehe District, FGD]

> *About knowledge on e-ASCOv, I can tell you that we do not have any because the only training we received was on how to use the forms.* [CHW, Kirehe District]

The respondents expected the app to help inform on the status of the pandemic and help with community outreach on COVID-19. Even though some CHWs were not aware of the app and did not have any expectations, a number of them were expecting the e-ASCov app to help curb the spread of COVID-19. They also expected it to support an extension of the service to home-based care, provide support to cases, and improve CHWs knowledge and the app to be used by all CHWs.

> *What we have expected and still expect with current results is that the use of technology has impacted well in curbing the spread of COVID-19* [COVID-19 Taskforce, district level]

> *What we all expect is that the app contributes [to] the continuous fight against COVID-19, like getting information on time* [CHW supervisor, central level]

> *As I said, what we were expecting is to find out the status of the pandemic, and other measures that could be taken to prevent it because there is no other way to use it without the technology* [district director of health]

The CHWs' expectations were mostly that the app would provide an update on the status of the pandemic in their area. They thought that the detection of COVID-19 cases would lead to a decrease in the magnitude of the pandemic. In addition, some CHWs emphasized the app's contribution to early diagnosis and treatment and the opportunity to screen other health commodities as in their routine community-based health services. Some clients also considered that screening with the technology would reduce the burden of COVID-19.

> *What we expected was the results of using this application in helping Rwandans during the pandemic. We can say, like follow-up of the infected people before they get complications for them to receive early medical management, for the health system at central level to get information on time and all the people with the disease to be managed on time. I think that was what we expected and the results have been seen* [CHW, Gasabo District, FDG]

> *We expected to provide early and quick information if someone was having the signs and symptoms for the healthcare providers to provide early and timely care. Another thing that this application helped … and we expected … was data storage in a trusted way as we are moving from analogue to a digital world* [CHW, Nyarugenge District, FDG]

Use and Perceived Benefits of the App

According to CHW reporting, the median number of people screened by each CHW ranged from 152 (IQR 70-276) for Nyarugenge to 24 (IQR 16-90) for Rusizi (Table 2). Of these, few were referred to a health center for diagnostic confirmation, between 1 and 5 per CHW. In total, 48.4% and 47.4% of the CHWs sent more than 5 people screened to the health centers in Gasabo and Kirehe, respectively. Positivity rates were higher in urban than rural districts. For example, more than half of the CHWs in Gasabo reported a confirmed positive case, whereas this rate was only 23.7% for Kirehe and 15.4% for Rusizi. Only 2 CHWs from Gasabo declared a case of death.

A high number of respondents considered that the most perceived benefit of using the e-ASCOV app was the enhancement of information sharing about COVID-19, followed by timely decision-making and quick dissemination of information.

> *The first thing, information got to be known so fast and this helps to make measures* [district director of health]

> *The way of using e-ASCOV or technology in delivering services and information about COVID-19 pandemic is used to know information about the pandemic earlier, so that the information can reach … the decision makers, and the patients may get treated in a quick way because technology might be quick than the ordinary ways in delivering information.* [CHW supervisor, central level]

Some of the respondents believed that using e-ASCov app would benefit quick data retrieval and decision-making. Respondents also shared that the e-ASCov app would benefit in avoiding COVID-19 infection spread by enhancing information storage because it was paperless and time-saving.

> *The benefits that I have seen in our application that we visited every household and asked the questions in the application and found that he was seriously sick at home but did not know the symptoms of COVID-19 and I immediately advised the client to go to the health center and accompanied the client and treat him right away; that is the benefit I have seen.* [CHW from Gasabo District]

Table . Description of overall intervention indicators according to the district.

Description	Gasabo (n=93, 24.6%)		Kirehe (n=97, 25.7%)		Nyarugenge (n=90, 23.8%)		Rusizi (n=98, 25.9%)		*P* value[a]
	n (%)	Median (IQR)	n (%)	Median (IQR)	n (%)	Median (IQR)	n (%)	Median (IQR)	
Total number of persons screened	93 (100)	86.0 (15.0-160.0)	97 (100)	48.0 (15.0-86.0)	90 (100)	152.5 (70.0-276.0)	98 (100)	24.0 (16.0-90.0)	<.001
Number of clients referred	93 (100)	5.0 (0.0-15.0)	97 (100)	5.0 (0.0-15.0)	90 (100)	2.0 (0.0-8.0)	98 (100)	1.0 (0.0-4.0)	.005
Number of patients referred									<.001
None	32 (34.4)	—[b]	27 (27.8)	—	35 (38.9)	—	44 (44.9)	—	
1 - 5	16 (17.2)	—	24 (24.7)	—	26 (28.9)	—	35 (35.7)	—	
>5	45 (48.4)	—	46 (47.4)	—	29 (32.2)	—	19 (19.4)	—	
Number of clients confirmed positive among those referred	93 (100)	1.0 (0.0-4.0)	97 (100)	0.0 (0.0-0.0)	90 (100)	0.0 (0.0-2.0)	98 (100)	0.0 (0.0-0.0)	<.001
At least 1 positive case									<.001
No	42 (45.2)	—	74 (76.3)	—	55 (61.1)	—	83 (84.7)	—	
Yes	51 (54.8)	—	23 (23.7)	—	35 (38.9)	—	15 (15.3)	—	
Number of confirmed cases with severe symptoms transferred for further management	93 (100)	0.0 (0.0-0.0)	97 (100)	0.0 (0.0-0.0)	90 (100)	0.0 (0.0-0.0)	98 (100)	0.0 (0.0-0.0)	.006
At least 1 severe case									.006
No	73 (78.5)	—	88 (90.7)	—	72 (80)	—	91 (92.9)	—	
Yes	20 (21.5)	—	9 (9.3)	—	18 (20)	—	7 (7.1)	—	
Number of deaths among confirmed cases	93 (100)	0.0 (0.0-0.0)	97 (100)	0.0 (0.0-0.0)	90 (100)	0.0 (0.0-0.0)	98 (100)	0.0 (0.0-0.0)	.10
At least 1 death									.10
No	91 (97.8)	—	97 (100)	—	90 (100)	—	98 (100)	—	
Yes	2 (2.2)	—	0 (0)	—	0 (0)	—	0 (0)	—	

[a]*P* value for χ^2 test for qualitative variables or Student *t* test for quantitative variables.

[b]Not applicable.

Satisfaction With and Challenges of the App

The satisfaction of CHWs is presented in Table 3. Overall satisfaction was higher in Nyarugenge (72.8/100) and Gasabo (80.7/100) than Kirehe (61.6/100) and Rusizi (64.5/100). More than 80% of CHWs were satisfied with the training offered before deployment of the app in all districts. In terms of availability of tools and internet connection, satisfaction was lower in rural than urban districts. More than 80% of the CHWs were willing to continue using the app, with the exception of CHWs in Kirehe (56.7%).

Table . Satisfaction (very satisfied/satisfied) of community health workers according to district.

	Gasabo (n=93; 24.6%)		Kirehe (n=97; 25.7%)		Nyarugenge (n=90; 23.8%)		Rusizi (n=98; 25.9%)		*P* value[a]
	n (%)	Mean (SD)	n (%)	Mean (SD)	n (%)	Mean (SD)	n (%)	Mean (SD)	
Training provided in use of digital tool									.23
No	17 (18.3)	—[b]	8 (8.2)	—	11 (12.2)	—	13 (13.3)	—	
Yes	76 (81.7)	—	89 (91.8)	—	79 (87.8)	—	85 (86.7)	—	
Score: 0 - 100	93 (100)	81.7 (23.9)	97 (100)	85.1 (19.7)	90 (100)	80.8 (20.9)	98 (100)	79.1 (18.9)	.24
Availability of equipment (telephone, etc)									<.001
No	27 (29.0)	—	76 (78.4)	—	19 (21.1)	—	47 (48.0)	—	
Yes	66 (71.0)	—	21 (21.6)	—	71 (78.9)	—	51 (52)	—	
Score: 0 - 100	93 (100)	81.2 (23.8)	97 (100)	54.4 (18.8)	90 (100)	70.8 (24.4)	98 (100)	57.9 (25.0)	<.001
Internet connectivity									<.001
No	27 (29)	—	76 (78.4)	—	19 (21.1)	—	47 (48)	—	
Yes	66 (71)	—	21 (21.6)	—	71 (78.9)	—	51 (52)	—	
Score: 0 - 100	93 (100)	81.2 (23.8)	97 (100)	54.4 (18.8)	90 (100)	70.8 (24.4)	98 (100)	57.9 (25)	<.001
Simplicity/ease in use of the tool									<.001
No	26 (28)	—	74 (76.3)	—	12 (13.3)	—	19 (19.4)	—	
Yes	67 (72)	—	23 (23.7)	—	78 (86.7)	—	79 (80.6)	—	
Score: 0 - 100	93 (100)	80.6 (23.9)	97 (100)	55.9 (20.7)	90 (100)	75.6 (23.1)	98 (100)	69.4 (27.2)	<.001
Time used to fill and complete one record with the tool									<.001
No	27 (29)	—	76 (78.4)	—	19 (21.1)	—	47 (48)	—	
Yes	66 (71)	—	21 (21.6)	—	71 (78.9)	—	51 (52)	—	
Score: 0 - 100	93 (100)	81.2 (23.8)	97 (100)	54.4 (18.8)	90 (100)	70.8 (24.4)	98 (100)	57.9 (25)	<.001
Getting technical assistance/support									<.001
No	42 (45.2)	—	76 (78.4)	—	33 (36.7)	—	52 (53.1)	—	
Yes	51 (54.8)	—	21 (21.6)	—	57 (63.3)	—	46 (46.9)	—	
Score: 0 - 100	93 (100)	69.9 (28.4)	97 (100)	55.2 (20.7)	90 (100)	59.2 (25.8)	98 (100)	49 (29.4)	<.001
Delivery of services to the clients with the tool									<.001
No	26 (28)	—	74 (76.3)	—	12 (13.3)	—	19 (19.4)	—	
Yes	67 (72)	—	23 (23.7)	—	78 (86.7)	—	79 (80.6)	—	
Score: 0 - 100	93 (100)	80.6 (23.9)	97 (100)	55.9 (20.7)	90 (100)	75.6 (23.1)	98 (100)	69.4 (27.2)	<.001
Continued use of the tool									<.001
No	15 (16.1)	—	42 (43.3)	—	11 (12.2)	—	17 (17.3)	—	
Yes	78 (83.9)	—	55 (56.7)	—	79 (87.8)	—	81 (82.7)	—	
Score: 0 - 100	93 (100)	89.2 (19.6)	97 (100)	77.3 (24.5)	90 (100)	78.6 (27.7)	98 (100)	75.3 29.9	<.001
Global satisfaction score: 0 - 100	93 (100)	80.7 (19.4)	97 (100)	61.6 (15.4)	90 (100)	72.8 (19.2)	98 (100)	64.5 (21.6)	<.001

[a]*P* value of χ^2 test for qualitative variables or Student *t* test for quantitative variables.

[b]Not applicable.

Most of the respondents stated that the main challenges were poor internet access or a poor network, CHWs having a low level of operational knowledge, lack of access to smartphones, and the shortage of equipment. Challenges were the lack of CHWs' incentive and internet fees, use of a personal cell phone, low number of trained CHWs, limited access to electricity by CHWs, incompatibility of the app with CHWs' cell phones, inability to use the app, inability to edit and correct an error, older CHWs feeling excluded, and the inability to retrieve information. The CHWs viewed the lack of smartphones as a factor that impeded reporting via the e-ASCov app. The poor internet connectivity was also highlighted by most of the CHWs as a limiting factor for better access.

Discussion

Principal Findings

This project confirmed the central position of CHWs in an effective strategy to deal with a pandemic such as COVID-19 in a pyramidal health system such as Rwanda's. By including 2 districts, 2 rural and 2 urban, we were able to show that this type of system was possible in the health systems of countries where CHWs play a predominant role. Qualitative and quantitative results were consistent, and there was enthusiasm for the e-ASCov app. The smooth organization of the CHW network and its fully integrated nature into the Rwandan health system ensured that the app was easily deployed. In an epidemic context, CHWs were able to use this app as a screening and prevention tool to stem the spread of the epidemic in each district. The CHWs were able to use the app for screening suspected cases at the community level and were largely satisfied. However, there were some slight differences, particularly between urban and rural districts. Therefore, the conditions of implementation and transferability of such an app on a national scale need to be investigated in a broader context.

Comparison With Prior Studies

This type of tool has been used in the African context for other diseases [10,11]. Although the use of the app was good in all districts, it was easier in urban than rural districts. This difference, beyond the organization of the CHW system in participating districts, seems to be linked to the level of education and CHWs' experience in implementing this type of action. Indeed, in the urban districts, three-quarters of the CHWs had a level of education higher than primary school, whereas in the rural districts, the opposite was true. In addition, the smoothness of the internet connection in urban districts was also identified as a facilitating factor in the use of the app. Duclos et al [12] stated that mHealth implementation in rural health districts is expected to present considerable challenges, including technological barriers, organizational challenges, gender issues, confidentiality concerns, and unplanned aftereffects.

The successful implementation of the e-ASCov app at the community level opens up several possibilities. This type of app could be used in an epidemic context and in community-based monitoring of certain chronic diseases or maternal and child health [13]. The Rwanda CHW program was established in 1995. It aimed to strengthen essential maternal and child health care services through the education of pregnant women, health promotion and risk behavior reduction, follow-up, and linkages with health services. Rwanda has more than 50,000 CHWs (4 per village) with an organization of 3 CHWs (3 different dedicated functions) for each 100 to 150 households. One CHW is in charge of maternal and child health (identifying pregnant women, accompanying the regular monitoring of the pregnancy, and ensuring that the delivery takes place in a health facility). Two CHWs are responsible for referral and integrated community-based management of diseases (assessment, classification, and treatment) such as diarrheal diseases, pneumonia, malaria, and malnutrition in children younger than 5 years; provision of community-based contraceptives; directly observed treatment surveillance for tuberculosis; and prevention of noncommunicable diseases. The fourth CHW is specifically responsible for the hygiene of the village population. This organization allows for the use of connected tools for screening and monitoring several other diseases.

Although the implementation and experience of CHWs seem to support the good acceptability and feasibility of this type of tool, a positive perception of the population receiving the intervention is also important. This perception should be ensured by raising the awareness of the population, especially in rural areas, to improve their literacy in relation to digital health tools. For e-ASCov, part of the population was not sufficiently informed of the use of e-ASCov app. Those who were aware of it stated that the system provides useful information about COVID-19.

> *According to that technology of e-ASCOV, they did not use it. I think they were not even having it. I agree with what others have said...* [client, Kirehe District, FGD]

Limitations

The first limitation is that the e-ASCov project was implemented in only 4 of the country's 30 districts, which may restrict confidence in a full generalization of results. To minimize representation bias, we selected the 4 districts according to their rural or urban character and their geographical location. The 2 urban districts are located in the capital, where most of the cases of COVID-19 were identified (imported cases). The 2 other rural districts were located at the 2 extremities of the country: Rusizi in the western region bordering the Democratic Republic of Congo and Kirehe in the eastern region bordering Tanzania. Another limitation was the significant spread of the epidemic throughout the country, so the use of this type of app was less effective and less appropriate. Indeed, when the prevalence of the disease becomes very large and the strategy is to learn to live with the virus, community-based screening becomes less relevant. However, beyond COVID-19, the e-ASCov project reinforces the idea that digital tools such as smartphone apps can be of great use in community health. Despite these limitations, the e-ASCov project is one of the first projects to include more than 400 CHWs who were trained to use the app and who screened several thousand potential cases over a year. The results of this study have led the country's health authorities to consider rolling out the app nationwide.

Conclusion

Findings from this study show proof of concept that the use of mHealth tools by CHWs can enhance their participation and contribution to an effective and resilient health system in low- and middle-income countries. A study is needed of the generalization and transferability conditions of this type of app to other health conditions to enable CHWs to be given their full place in a pyramidal health system.

Acknowledgments

Many people worked together selflessly and enthusiastically for the success of the e-ASCov project. The e-ASCov project group warmly acknowledges the community health workers, members of the Rwanda Biomedical Centre, and members of the Rinda Ubuzima Association who contributed to the project. We especially thank Florian Le Goff from Kelindi, a French start-up, for the contextualization of the digital tool. The e-ASCov project was funded by the French Agency for Research on AIDS and Viral Hepatitis (ANRS-MIE) and the Embassy of France in Kigali (Rwanda).

Conflicts of Interest

None declared.

References

1. Holmes EC, Goldstein SA, Rasmussen AL, et al. The origins of SARS-Cov-2: a critical review. Cell 2021 Sep 16;184(19):4848-4856. [doi: 10.1016/j.cell.2021.08.017] [Medline: 34480864]
2. Coronavirus disease 2019, national preparness and response plan. Ministry of Health, Republic of Rwanda. 2020. URL: https://rbc.gov.rw/index.php?id=707&L=0 [accessed 2020-04-09]
3. Benimana TD, Lee N, Jung S, Lee W, Hwang SS. Epidemiological and spatio-temporal characteristics of COVID-19 in Rwanda. Glob Epidemiol 2021 Nov;3:100058. [doi: 10.1016/j.gloepi.2021.100058] [Medline: 34368752]
4. Musanabaganwa C, Cubaka V, Mpabuka E, et al. One hundred thirty-three observed COVID-19 deaths in 10 months: unpacking lower than predicted mortality in Rwanda. BMJ Glob Health 2021 Feb;6(2):e004547. [doi: 10.1136/bmjgh-2020-004547] [Medline: 33627363]
5. Miller NP, Milsom P, Johnson G, et al. Community health workers during the Ebola outbreak in Guinea, Liberia, and Sierra Leone. J Glob Health 2018 Dec;8(2):020601. [doi: 10.7189/jogh.08.020601] [Medline: 30023054]
6. Siekmans K, Sohani S, Boima T, Koffa F, Basil L, Laaziz S. Community-based health care is an essential component of a resilient health system: evidence from Ebola outbreak in Liberia. BMC Public Health 2017 Jan 17;17(1):84. [doi: 10.1186/s12889-016-4012-y] [Medline: 28095824]
7. Opinion: why community health workers matter now more than ever. Devex. 2020. URL: https://www.devex.com/news/sponsored/opinion-why-community-health-workers-matter-now-more-than-ever-96950 [accessed 2020-04-13]
8. Karim N, Jing L, Lee JA, et al. Lessons learned from Rwanda: innovative strategies for prevention and containment of COVID-19. Ann Glob Health 2021 Feb 25;87(1):23. [doi: 10.5334/aogh.3172] [Medline: 33665145]
9. Braun V, Clarke V. Thematic analysis. In: APA Handbook of Research Methods in Psychology, Vol 2: Research Designs: Quantitative, Qualitative, Neuropsychological, and Biological: American Psychological Association; 2012:57-71. [doi: 10.1037/13620-000]
10. Palazuelos D, Diallo AB, Palazuelos L, Carlile N, Payne JD, Franke MF. User perceptions of an mHealth medicine dosing tool for community health workers. JMIR Mhealth Uhealth 2013 Apr 4;1(1):e2. [doi: 10.2196/mhealth.2459] [Medline: 25100670]
11. Chang LW, Njie-Carr V, Kalenge S, Kelly JF, Bollinger RC, Alamo-Talisuna S. Perceptions and acceptability of mHealth interventions for improving patient care at a community-based HIV/AIDS clinic in Uganda: a mixed methods study. AIDS Care 2013;25(7):874-880. [doi: 10.1080/09540121.2013.774315] [Medline: 23452084]
12. Duclos V, Yé M, Moubassira K, et al. Situating mobile health: a qualitative study of mHealth expectations in the rural health district of Nouna, Burkina Faso. Health Res Policy Syst 2017 Jul 12;15(Suppl 1):47. [doi: 10.1186/s12961-017-0211-y] [Medline: 28722558]
13. Arnaert A, Ponzoni N, Debe Z, Meda MM, Nana NG, Arnaert S. Experiences of women receiving mhealth-supported antenatal care in the village from community health workers in rural Burkina Faso, Africa. Dig Health 2019 Dec 1;5:2055207619892756. [doi: 10.1177/2055207619892756] [Medline: 31832224]

Abbreviations

CHW: community health worker
FGD: focus group discussion
IDI: in-depth interview
mHealth: mobile health

Edited by L Buis; submitted 11.07.23; peer-reviewed by C Jacob, D Laborde; revised version received 08.04.24; accepted 25.04.24; published 14.10.24.

Please cite as:
Omorou AY, Ndishimye P, Hoen B, Mutesa L, Karame P, Nshimiyimana L, Galmiche S, Mugabo H, Murayire J, Mugisha M, Umulisa MM, Uwera YDN, Musanagabanwa C, Bigirimana N, Nsanzimana S, Guillemin F, Rwabihama JP
Feasibility, Acceptability, Satisfaction, and Challenges of an mHealth App (e-ASCov) for Community-Based COVID-19 Screening by Community Health Workers in Rwanda: Mixed Methods Study
JMIR Mhealth Uhealth 2024;12:e50745
URL: https://mhealth.jmir.org/2024/1/e50745
doi:10.2196/50745

Original Paper

Investigating Citizens' Acceptance of Contact Tracing Apps: Quantitative Study of the Role of Trust and Privacy

Grace Fox[1], PhD; Lisa van der Werff[1], PhD; Pierangelo Rosati[2], PhD; Theo Lynn[1], PhD

[1]Irish Institute of Digital Business, Dublin City University, Dublin, Ireland

[2]J.E. Cairnes School of Business & Economics, University of Galway, Galway, Ireland

Corresponding Author:
Theo Lynn, PhD
Irish Institute of Digital Business
Dublin City University
Collins Ave, Dublin9
Dublin
Ireland
Phone: 353 1 700 6873
Email: theo.lynn@dcu.ie

Abstract

Background: The COVID-19 pandemic accelerated the need to understand citizen acceptance of health surveillance technologies such as contact tracing (CT) apps. Indeed, the success of these apps required widespread public acceptance and the alleviation of concerns about privacy, surveillance, and trust.

Objective: This study aims to examine the factors that foster a sense of trust and a perception of privacy in CT apps. Our study also investigates how trust and perceived privacy influence citizens' willingness to adopt, disclose personal data, and continue to use these apps.

Methods: Drawing on privacy calculus and procedural fairness theories, we developed a model of the antecedents and behavioral intentions related to trust and privacy perceptions. We used structural equation modeling to test our hypotheses on a data set collected at 2 time points (before and after the launch of a national CT app). The sample consisted of 405 Irish residents.

Results: Trust in CT apps was positively influenced by propensity to trust technology (β=.074; *P*=.006), perceived need for surveillance (β=.119; *P*<.001), and perceptions of government motives (β=.671; *P*<.001) and negatively influenced by perceived invasion (β=−.224; *P*<.001). Perceived privacy was positively influenced by trust (β=.466; *P*<.001) and perceived control (β=.451; *P*<.001) and negatively influenced by perceived invasion (β=−.165; *P*<.001). Prelaunch intentions toward adoption were influenced by trust (β=.590; *P*<.001) and perceived privacy (β=.247; *P*<.001). Prelaunch intentions to disclose personal data to the app were also influenced by trust (β=.215; *P*<.001) and perceived privacy (β=.208; *P*<.001) as well as adoption intentions before the launch (β=.550; *P*<.001). However, postlaunch intentions to use the app were directly influenced by prelaunch intentions (β=.530; *P*<.001), but trust and perceived privacy only had an indirect influence. Finally, with regard to intentions to disclose after the launch, use intentions after the launch (β=.665; *P*<.001) and trust (β=.215; *P*<.001) had a direct influence, but perceived privacy only had an indirect influence. The proposed model explained 74.4% of variance in trust, 91% of variance in perceived privacy, 66.6% of variance in prelaunch adoption intentions, 45.9% of variance in postlaunch use intentions, and 83.9% and 79.4% of variance in willingness to disclose before the launch and after the launch, respectively.

Conclusions: Positive perceptions of trust and privacy can be fostered through clear communication regarding the need and motives for CT apps, the level of control citizens maintain, and measures to limit invasive data practice. By engendering these positive beliefs before launch and reinforcing them after launch, citizens may be more likely to accept and use CT apps. These insights are important for the launch of future apps and technologies that require mass acceptance and information disclosure.

(JMIR Mhealth Uhealth 2024;12:e48700) doi:10.2196/48700

KEYWORDS

privacy; trust; public health surveillance; contact tracing; mobile apps; adoption; information disclosure

Introduction

Background

The outbreak of COVID-19 and the ensuing global pandemic resulted in many governments undertaking digital government transformation [1] through the introduction of public health surveillance technologies including contact tracing (CT) apps [2,3]. As a result, and unlike previous health emergencies, governments had access to an unprecedented volume, variety, and velocity of location and health data [4]. The use of such data for epidemiological surveillance can aid in decision support, accelerate case identification, interrupt community transmission, and enable public health communication [5]. Notwithstanding these benefits, the pace at which these apps have been implemented and the level of surveillance they enable have raised ethical concerns [6] and fears around privacy and public trust [7].

The success of CT apps is dependent on uptake by large populations [8], and privacy-related concerns have been positioned as a critical barrier facing government-introduced CT apps [9]. Government-introduced CT apps differ from surveillance technologies and mobile health (mHealth) apps, as they combine both location-based data and electronic personal health information (ePHI). Both these contexts, in themselves, raise significant privacy concerns, particularly with respect to potential secondary use and government intrusion [10]. Thus, the COVID-19 pandemic presents a unique empirical context to explore citizens' perceptions of health surveillance using mobile apps that capture both location-based data and ePHI. Government-introduced CT apps constitute a new public health context. Although there is an established literature based on traditional CT, digital CT is an innovation that can only be fully explored during pandemics, and thus research opportunities are limited. Given the novel context for individuals and public health agencies, where the former engage with new or unfamiliar trust referents, it is particularly critical to explore how individuals marry competing beliefs about surveillance, trust, and government motives for introducing these technologies and how these beliefs influence their behavioral responses. Extant literature has demonstrated the importance of trust in the government in influencing CT adoption [11,12] and perceptions of CT technologies [13] and shown that privacy concerns represent a barrier to adoption [9].

Prior Work

This study builds upon important extant research focusing on the acceptance of CT apps to delve further into the role of privacy and trust and addresses 3 gaps in the literature. First, existing studies support the importance of trust in driving the acceptance of CT apps, but the approach to measuring trust and the trust referent under examination varies. For example, studies have found that high trust in the national government, the health care system, and science positively impacted willingness to use CT apps in Switzerland [14]. A US-based study found that trust in COVID-19 information positively influenced citizens' comfort with and acceptance of CT [8]. In this study, our emphasis is on technology-related trust, as opposed to trust in an individual or organization. Specifically, we examined trust perceptions regarding a specific technology, namely a CT app. In addition, we investigated the perceptual factors that shape trust, namely perceptions of government motivations, need for surveillance, and propensity to trust technology (PTTT).

Second, studies have investigated the influence of privacy on CT adoption, with many studies finding that privacy concerns reduce intentions toward adopt CT apps [11,14]. Several studies conclude that privacy represents a barrier to the success of CT apps, with respondents in several countries citing privacy concerns as a reason for not installing apps [9] or conversely, those with low privacy concerns are more likely to use CT apps [12]. However, the influence of privacy concerns on CT adoption intentions was weak in another study [15]. Although privacy concerns are the most common proxy for measuring privacy across many contexts, the negative connotation and failure to directly capture privacy suggest the need for more precise operationalization [16]. Thus, we focus on perceived privacy defined as "an individual's self-assessed state in which external agents have limited access to information about him or her" [16]. In this study, perceived privacy refers to a citizen's belief in the level of privacy afforded by the CT app. Perceived privacy influenced intentions toward CT apps in a recent Brazilian study, thus supporting its use [17]. To further our understanding of perceived privacy, this study investigated the role of trust, perceived control, and perceived intrusion in shaping citizens' perceptions of privacy.

Third, many CT studies are cross-sectional in nature, with the exception of a small number of longitudinal studies [12]. In addition, the literature focuses largely on whether citizens adopt CT apps or engage in behaviors recommended by CT apps such as staying at home [14,18]. This study broadens our understanding of citizens' acceptance of CT apps by examining 2 variables related to acceptance, namely intention to adopt or continue using the app and willingness to disclose personal information. These acceptance variables are measured before and after the app launch, thereby deepening our understanding of how privacy and trust influence intentions toward and use of CT apps.

We argue that understanding the determinants of success of CT apps is critical not only for future digital CT but also for other contexts that require rapid digital technology adoption by the population [19]. This paper proceeds with an overview of the hypothesized relationships and our research context. Our methodology, data analysis, and results are presented in the following sections. The *Discussion* section outlines the implications of this study. The paper concludes with the limitations and avenues for future research.

Theory

Overview

Privacy Calculus Theory (PCT) posits that before engaging in a behavior such as adopting a new technology or disclosing personal information, individuals will conduct a cognitive comparison of the costs and benefits associated with this behavior [20]. Individuals are likely to engage in the behavior for as long as the benefits outweigh the costs [20]. Thus, PCT has direct comparisons with the concept of calculus-based trust,

which underpins trust decisions when engaging with new or unfamiliar trust referents [21]. PCT has been operationalized in many contexts using a variety of belief-based variables that can be grouped into confidence beliefs and risk or privacy beliefs. An extension of PCT that holds considerable promise for understanding privacy and trust in the context of CT apps is the inclusion of the procedural fairness theory. Procedural fairness refers to an individual's perception that a particular activity is conducted fairly [22]. In the context of information privacy, fairness refers to the perception that personal data are collected and used fairly. Culnan and Armstrong [22] proposed that perceptions of procedural fairness can help citizens to "strike a balance between the competing forces of privacy and information use." Individuals' perceptions of the fairness of an organization's data collection and use practices can influence their decision-making related to technology use and information disclosure [23]. In this study, we investigated the drivers of trust and privacy through the procedural fairness lens, which suggests the importance of factors related to the legitimacy of data collection (ie, the need for government surveillance and perceptions of the government's motive for the app), the costs to the citizen (ie, perceived intrusion), and the level of autonomy and input citizens are afforded (ie, perceived control). Furthermore, the wider literature on fairness and trust suggests that individual differences in citizens' PTTT are likely to play an important role alongside procedural fairness perceptions in trust [24], particularly given the unfamiliar referent of the CT app. In addition, the theory of reasoned action (TRA) allows us to consider the influence of these perceptions on behavioral outcomes. The TRA argues that individuals' behaviors are determined by their beliefs, attitudes, and intentions [25]. We propose that individuals will express positive intentions toward downloading the app and disclosing information if they believe that the app demonstrates fairness.

Hypotheses

Trust is an important factor in the success of CT apps as demonstrated in recent studies [14]. Indeed, trust allows individuals to overcome concerns about uncertainty and fosters a willingness to engage in trust-related behaviors, such as disclosing information and engaging with technology [26]. In this study, our emphasis is on technology-related trust, which refers to individuals' beliefs that the technology in question will perform as expected [27]. Trust perceptions refer to the extent to which the CT app will consistently deliver the proposed services and act in citizens' best interest.

A significant body of theoretical work suggests that variables related to trust propensity are important drivers of trust perceptions [28], particularly in new and unfamiliar trust referents [29]. PTTT refers to a general tendency that is not specific to one trustee or situation but focuses on individuals' willingness to depend on technology across different contexts and technologies [27]. We propose that, in the current context, PTTT will positively influence trust in CT apps.

- *Hypothesis 1a: PTTT will have a positive association with trust in the app.*

Surveillance programs are often introduced following large-scale events such as terrorist attacks [30]. Given the public health emergency caused by the COVID-19 pandemic, the importance of surveillance technologies, such as CT apps, is clear. The link between surveillance and trust has long attracted discussion with Dutton et al [31], highlighting the existence of *trust tension* between the government's need to collect surveillance data and citizens' concerns about the excessive use of this information. They assert that developing trust is imperative for resolving this tension. As CT apps require the surveillance of large groups of people [32], citizens must understand the need for government surveillance in the general sense to build trust in a CT app. Need for government surveillance refers to individuals' perceptions that the government requires authority to access personal information using web-based means [33].

- *Hypothesis 1b: need for surveillance will have a positive association with trust in the app.*

Procedural fairness theory suggests that if individuals believe that the government's motivations to introduce the app are rooted in good intentions, such as reducing virus transmission, they will express higher trust in the app to perform consistently and with their best interests in mind. Indeed, a perception of benevolent motives is at the heart of theories regarding trustworthiness [28], and perceptions of trustworthiness at the government level are likely to trickle down to influence trust in related referents [34]. Accordingly, we posit that government motive will engender trust in the app.

- *Hypothesis 1c: government motive will have a positive association with trust in the app.*

Government surveillance technologies can garner negative opinions, which may lead individuals to alter their behaviors. Perceived intrusion is described as a "harmful incursion into the personal information space" [35]. This relates to procedural justice and individuals' perceptions of whether data are collected fairly in the CT app. Government surveillance technologies can be viewed as intrusive, but not all intrusions are considered harmful [35]. For example, the CT app may be viewed as intrusive, but as data are collected to reduce virus transmission, some individuals may not view this as harmful. Thus, only if individuals believe that the intrusiveness of the app is harmful to them, then their trust in the app is likely to be reduced.

- *Hypothesis 1d: perceived intrusion will have a negative association with trust in the app.*

Studies have asserted that privacy concerns represent a barrier to the success of CT apps [14,36]. However, there is a lack of research examining *if* citizens are likely to accept CT apps when they perceive that they provide some level of privacy. As perceived privacy refers to a perception that access to personal information by external agents is limited, the relevance of perceived intrusion as a privacy cost resulting from the use of an app is apparent. Indeed, the potential of CT apps to violate citizens' privacy has been raised [36]. Thus, if individuals believe that the app is intrusive in a harmful manner in their informational space, they are less likely to believe that the app affords them privacy.

- *Hypothesis 2a: perceived intrusion will have a negative association with perceived privacy.*

Perceived control is described as individuals' perceptions of their ability to control their personal information [37]. If individuals perceive that they maintain control over their information when transacting with a technology, they are more likely to feel comfortable that the technology will not act in a harmful manner [37], strengthening their perception of privacy in that context. A recent study supports the positive association between perceived control and perceived privacy of CT apps in Brazil [17]. We proposed that if individuals believe that they maintain control in the app, they will express higher levels of perceived privacy.

- *Hypothesis 2b: perceived control will have a positive association with perceived privacy.*

Finally, we argue that from a theoretical perspective, trust in the app will act as an uncertainty-reducing mechanism [38,39] and a heuristic that allows citizens to form privacy perceptions. Specifically, if citizens believe that the app will perform consistently and with their best interests in mind when using personal information, they will believe that the app provides some degree of privacy. This uncertainty reduction provides a foundation for facilitating other judgments of a technological artifact. Thus, trust in the app may influence perceptions of privacy.

- *Hypothesis 2c: trust in the app will have a positive association with perceived privacy.*

The success of CT apps is largely dependent on a critical mass of people downloading [40] and disclosing personal information. Thus, we draw on the TRA to consider 2 context-critical dependent variables: adoption intention and willingness to disclose personal information. Before the app launch, these variables were behavioral intentions. Adoption intention is described as an individual's internal subjective judgment of the probability that they will perform the behavior in question [25]. The willingness to disclose information is based on an individual's willingness to provide personal information when using the app [33]. Trust has repeatedly been identified as a driver of behavioral intentions across a range of contexts, including the acceptance of e-government technologies [41] and surveillance [30]. Indeed, behavioral operationalizations of trust often use behavioral intentions related to disclosure and reliance [42]. Empirical evidence suggests that perceived privacy influences intentions toward the use and data disclosure in general CT apps [17]. Although our study focuses on the privacy perception of a government-led CT app as opposed to general CT apps, we argue for similar effects. Finally, if individuals express high intentions toward downloading the app, we argue that they will be more willing to disclose personal information, as it is crucial to the app's functionality.

- *H3a-b: trust in the app before the launch (a) and perceived privacy (b) will have a positive association with adoption intentions before the launch.*
- *H4a-c: trust in the app before the launch (a), perceived privacy (b), and adoption intentions before the launch (c) will have a positive association with disclosure intentions before the launch.*

There have been calls for research to understand the perceptions of a technology before and after the launch [43]. In the context of CT apps, it is important to explore how both perceptions of privacy and trust influence individuals' behavioral intentions after launch. Thus, we examined both intentions after the launch. As some individuals may have already downloaded the app, adoption intentions are represented by future use intentions, which encompasses intentions to continue use among app users and intentions to adopt in the future among nonusers. TRA asserts that intentions will lead to behavior [44]. In other words, individuals' intentions to download the app before the launch will be positively related to their use intentions after the launch. We draw on the TRA to posit effects similar to those hypothesized for before the launch. We argue that trust perceptions regarding the app and perceived privacy will positively impact use intentions after the launch.

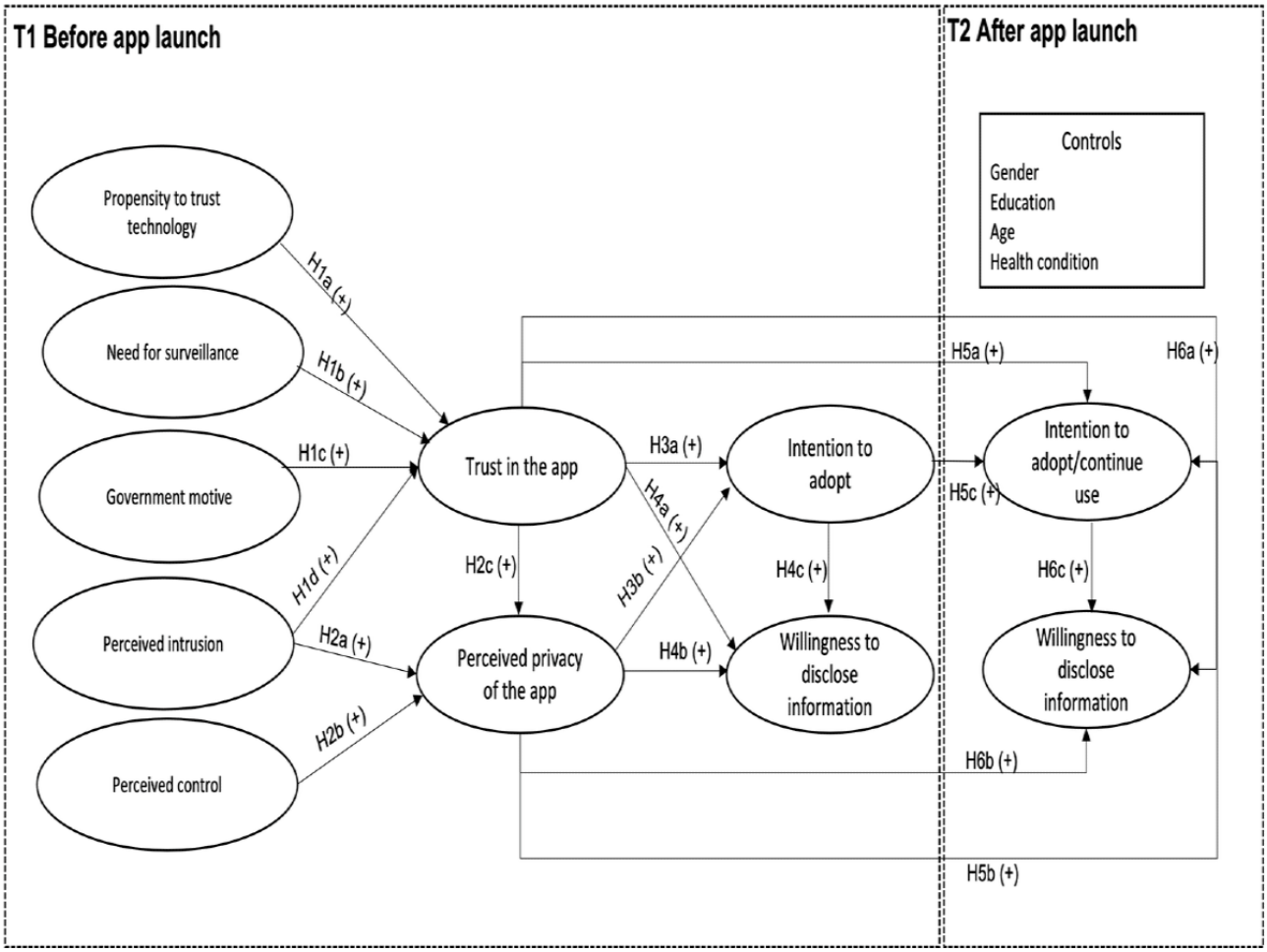
- *H5a-b: trust in the app (a) and perceived privacy (b) will have a positive association with intentions to use after the launch.*
- *H5c: adoption intentions before the launch will have a positive association with intentions to use after the launch.*

The influence of trust and privacy on the willingness to disclose a specific surveillance technology after the launch has not been explored. Again, we leverage the TRA intention-behavior link and our hypotheses before the launch and argue that trust and perceived privacy will positively impact willingness to disclose personal information after the launch. Finally, as was the case before the launch, we argue that if individuals express high intentions toward using the app after the launch, they will be more willing to disclose personal information.

- *H6a-b: trust in the app (a) and perceived privacy (b) will have a positive association with disclosure intentions after the launch.*
- *H6c: intentions to use after the launch will have a positive association with disclosure intentions after the launch.*

The hypotheses are depicted in Figure 1.

Figure 1. Research model.



Methods

Study Context

On July 6, 2020, the COVID Tracker app was launched in Ireland. COVID Tracker has 3 main features. CT uses Bluetooth and anonymous ID numbers to log phones within close contact for more than 15 minutes. It downloads the anonymous ID numbers of people who have tested positive and provides an alert if the user has been in close contact with those ID numbers. Check-in allows users to check for symptoms and seek health advice. The updates provide an overview of the daily COVID-19 figures. Within 48 hours of launch, 1 million people had downloaded the app [45]. By November 2021, the COVID Tracker app had over 1.7 million active users, representing 67% to 76% of the total possible Irish users [46].

Instrument Development

We used existing scales when developing our instrument with minor wording amendments to adapt the items to the context. We provide the full list of items in Table S1 in Multimedia Appendix 1. The survey at T1 included general constructs related to PTTT, the need for government surveillance, and situationally framed constructs related to the proposed app, namely government motive for introducing the app, trust in the app, perceived surveillance in the app, perceived control in the app, and perceived privacy in the app. The dependent variables included intention to download the app on launch and willingness to disclose personal information to use the app. At T2, the emphasis was on future behavioral intentions, namely the use of the app and information disclosure. At T2, we asked participants if they had downloaded the app, and app users' intentions to continue to use the app were examined and nonusers' intentions to download the app in the future were examined. These intentions were combined as use intentions for analysis. Willingness to disclose personal information to the app was examined across both users and nonusers of the app. In addition, gender, age, and education were used as the control variables. Respondents were asked if they had any health condition that left them particularly susceptible to COVID-19. This was also a control variable. Both surveys were piloted and tested among a small panel of survey design experiments, and several wording amendments were made. Respondents were asked to answer demographic and health questions first, followed by general perceptual constructs and control variables, the order of which was randomized. In the third section, participants were presented with a neutrally framed description of the proposed national CT app at T1, and a description of the launched app was presented at T2. The final section examined perceptions of the app, behavioral intentions, and willingness to disclose personal data, the order of which was randomized.

Ethical Considerations

Ethics approval was obtained from the university's research ethics committee before the launch of the survey (DCUREC/2020/096).

Recruitment

Qualtrics (Qualtrics International Inc) was used to host and administer the survey using their panel services. An attention check was included to screen for unengaged responses. A total of 1109 complete responses were received at T1 and were recontacted at T2. After 2 follow-up invitations, 405 responses were received at T2, achieving a response rate of 36.5%. Responses at T1 and T2 were, on average, 77 days apart. Incomplete responses and responses failing the attention check were removed using Qualtrics. The sample characteristics are illustrated in Table 1, along with the population characteristics as per the latest census at the time of data collection. Overall, the sample demographics were similar to the population characteristics of Ireland, as reported in the last census, and included respondents from the 26 counties within the country. Of the 405 respondents in T2, 202 had downloaded the app and 203 had not yet downloaded the app.

Table 1. Sample and population characteristics (N=405).

	Sample, n (%)	Population (%)[a,b]
Gender		
Man	180 (44.4)	49
Woman	225 (55.6)	51
Rather not say	0 (0)	N/A[c]
Age group (years)		
18-24	13 (3.2)	11
25-44	124 (30.6)	39
45-64	173 (42.7)	32
≥65	95 (23.5)	18
Employment		
Employed	186 (45.9)	45
Self-employed	26 (6.4)	8
Unemployed	36 (8.9)	6
Student	11 (2.7)	11
Unavailable for work	42 (10.4)	12
Retired	104 (25.7)	15
Education		
Secondary school	157 (38.8)	28
Trade	5 (1.2)	7
Diploma	32 (7.9)	12
Bachelor degree	133 (32.8)	27
Other qualification	64 (15.8)	14
Doctorate degree	14 (3.5)	1

[a]Population figures are based on data provided by the Irish Central Statistics Office in the latest population census at the time of data collection (ie, 2016).

[b]Employment and education figures include all people aged ≥15 years living in Ireland in 2016, whereas our sample only includes people aged ≥18 years.

[c]N/A: not applicable.

Results

Reliability and Validity Testing

Data analysis was performed using IBM AMOS (version 25.0). The proposed model comprising 11 constructs was examined using Confirmatory Factor Analysis with further detail provided in Table S2 in Multimedia Appendix 1. In total, 3 items were dropped from the PTTT because of their low loadings. The model indicated a good fit: c_{min}/df=1.805, comparative fit index=0.980, root mean square error of approximation=0.045, and standardized root mean squared residual=0.034. A test of equal specific bias was conducted to examine potential common method bias among the data [47]. This test demonstrated an unevenly distributed bias; thus, the specific bias construct was

retained for causal analysis to control for any effects because of method [48]. The validity and reliability of all the constructs were explored. Convergent validity was assessed by calculating the average variance extracted (AVE). As all the variables had AVE scores above 0.500, convergent validity was achieved [49]. Discriminant validity was tested by comparing the square root of the AVE with the interconstruct correlations. As the square root of AVE was higher than the interconstruct correlations, discriminant validity was achieved, as shown by the italicized diagonal values in Table 2. Reliability was assessed by calculating composite reliability for each construct. With composite reliability scores above 0.700, all constructs were reliable [50]. Further details on the validity testing are provided in the supplementary appendices available on the web.

Table 2. Validity and reliability statistics.

	Composite reliability	Average variance extracted	1	2	3	4	5	6	7	8	9	10	11
Need for Surveillance	0.868	0.526	*0.725* [a]	—[b]	—	—	—	—	—	—	—	—	—
Propensity to trust technology	0.870	0.626	0.183[c]	*0.791*	—	—	—	—	—	—	—	—	—
Perceived control	0.967	0.879	0.329[d]	0.210[d]	*0.938*	—	—	—	—	—	—	—	—
Willingness to disclose information (T1)	0.983	0.966	0.378[d]	0.271[d]	0.682[d]	*0.983*	—	—	—	—	—	—	—
Intention to adopt (T1)	0.990	0.970	0.342[d]	0.270[d]	0.666[d]	0.872[d]	*0.985*	—	—	—	—	—	—
Perceived intrusion	0.932	0.820	−0.181[e]	−0.109[e]	−0.471[d]	−0.508[d]	−0.397[d]	*0.906*	—	—	—	—	—
Trust in App	0.916	0.786	0.365[d]	0.336[d]	0.702[d]	0.814[d]	0.784[d]	−0.491[d]	*0.886*	—	—	—	—
Perceived Privacy in App	0.962	0.895	0.371[d]	0.259[d]	0.852[d]	0.801[d]	0.753[d]	−0.603[d]	0.852[d]	*0.946*	—	—	—
Intention to adopt or use (T2)	0.991	0.973	0.234[d]	0.167[e]	0.500[d]	0.612[d]	0.658[d]	−0.347[d]	0.574[d]	0.548[d]	*0.986*	—	—
Willingness to disclose information (T2)	0.985	0.970	0.309[d]	0.241[d]	0.537[d]	0.678[d]	0.666[d]	−0.401[d]	0.686[d]	0.646[d]	0.844[d]	*0.985*	—
Government motive	0.971	0.894	0.246[d]	0.332[d]	0.550[d]	0.682[d]	0.654[d]	−0.378[d]	0.799[d]	0.658[d]	0.494[d]	0.575[d]	*0.946*

[a]As the square root of AVE was higher than the interconstruct correlations, discriminant validity was achieved, as shown by the italicized values.

[b]Not available.

[c]Significance at 10% level.

[d]Significance at 1% level.

[e]Significance at 5% level.

Hypotheses Testing

The causal model was tested using Structural Equation Modeling in AMOS. The model indicated a good fit c_{min}/df=2.835, comparative fit index=0.985, root mean square error of approximation=0.067, and standardized root mean squared residual=0.021. H1a-d focused on the antecedents of trust in the app. H1a proposed a positive relationship between PTTT and trust. The data revealed a positive, significant relationship supporting hypothesis 1a (β=.074; *P*=.006). Hypothesis 1b posited that the perceived need for government surveillance would positively influence trust. This was also supported (hypothesis 1b: β=.119; *P*<.001). H1c posited that government motive would be positively related to trust. The data supported the hypothesis (hypothesis 1c: β=.671; *P*<.001). The negative relationship between perceived intrusion and trust was supported (hypothesis 1d: β=−.224; *P*<.001). The following set of hypotheses examined the antecedents of perceived privacy. Hypothesis 2a proposed a negative association between perceived intrusion and perceived privacy. These data supported hypothesis 2a (β=−.165; *P*<.001). We hypothesized that perceived control and trust would be positively related to perceived privacy. Both relationships were supported (hypothesis 2b: β=.451; *P*<.001; H2c: β=.466; *P*<.001).

In terms of T1 behavioral intentions, it was posited that perceived privacy and trust would positively influence the intention to adopt. Both relationships were supported (hypothesis 3a: β=.247; *P*<.001; hypothesis 3b: β=.590; *P*<.001). H4a-c proposed that trust, perceived privacy, and adoption intentions would positively influence willingness to disclose personal information. The data revealed that trust (β=.215; *P*<.001), perceived privacy (β=.208; *P*<.001), and adoption intentions (β=.550; *P*<.001) each positively influenced willingness to disclose. In terms of T2 behavioral intentions, hypothesis 5a to 5c proposed that trust, perceived privacy, and adoption intentions (T1) would all influence use intentions (T2). Both trust (β=.124; *P*=.15) and perceived privacy (β=.042; *P*=.60) had a positive but nonsignificant influence on intentions. T1 adoption intentions significantly influenced use intentions

supporting hypothesis 5c (β=.530; *P*<.001). Finally, trust, perceived privacy, and use intentions at T2 were proposed to positively influence the willingness to disclose at T2. Perceived privacy (β=.042; *P*=.40) had a nonsignificant influence, whereas trust and use intentions had significant relationships supporting hypothesis 6a and hypothesis 6c (hypothesis 6a: β=.250; *P*<.001; H6c: β=.655; *P*<.001). In terms of control variables, COVID-19 vulnerable illness had a significant negative effect on individuals' willingness to disclose at T1 (β=−.043; *P*=.009), and education had a positive effect on T2 use intentions (β=.075; *P*=.04).

The model explains 74.4% of variance in trust, 91% of variance in perceived privacy, 66.6% of variance in T1 adoption intentions, 45.9% of variance in T2 adoption intentions, and 83.9% and 79.4% of variance in willingness to disclose at T1 and T2. Bootstrapping using 2000 samples and a confidence level of 90% was conducted in AMOS to explore the indirect effects. The findings revealed that perceived privacy had a significant influence on T2 adoption intentions (β=.131; *P*=.001) and willingness to disclose at T2 (β=.127; *P*=.04). Similarly, trust had a significant influence on intention to download (β=.394; *P*<.001) and willingness to disclose at T2 (β=.386; *P*<.001). Further detail is provided in Table S3 in Multimedia Appendix 1.

Discussion

Principal Findings

This study focuses on understanding how citizens' beliefs shape their perceptions of privacy and trust to influence their acceptance of a CT app for COVID-19. Our study found that trust in the app was positively influenced by the PTTT, perceived government motive, and perceived need for government surveillance, whereas perceived intrusion had a negative influence. Perceived privacy was positively shaped by perceptions of control and trust and negatively shaped by perceived invasion. The study examined citizens' acceptance of CT app at 2 time intervals. Before launch, the intention to adopt the app was positively influenced by trust and perceived privacy, and willingness to disclose personal information to the app was influenced by adoption intentions, trust, and perceived privacy. However, postlaunch use intentions were influenced only by prelaunch adoption intentions, whereas willingness to disclose personal information was influenced by trust and postlaunch use intentions but not by perceived privacy. Although the insignificant results may suggest that perceived privacy is only important before launch, and the influence of trust on use intentions diminishes over time, post hoc bootstrapping analysis revealed that both perceived privacy and trust had significant indirect relationships with use intentions and willingness to disclose information at T2. This suggests that both perceptions play a role in influencing behavioral intentions before and after the launch.

Contributions

Studies have shown that trust and privacy are important factors in the success of health surveillance technologies such as CT apps [11]. Our study leverages the procedural fairness theory to understand how citizens' perceptions of trust and privacy emerge in the context of a CT app. This context is interesting, as the technology in question was introduced by the national government and backed by several organizations with the app's potential benefits extending to the public at large. Therefore, it is important to look beyond the role of a single organization in driving perceptions of fairness to consider a broader set of antecedents that drive perceptions of trust and privacy in this context. Indeed, as research has shown the importance of trust and privacy in the success of mHealth and health technologies introduced by health care organizations and indeed national health systems [51], our study contributes to the broader health technology literature.

The first contribution of our study is the deeper understanding of how trust is formed in this context. Lack of trust in the government has been identified as a barrier to CT app adoption [18]. Thus, it is important to provide governments and public health organizations with insights into how trust in CT can be fostered [8]. Our findings bolster assertions regarding the important role of fairness perceptions and suggest that citizens' trust perceptions regarding the app are formed based on their beliefs about the legitimacy of data collection, perceived autonomy, and perceived costs. Legitimacy is represented by citizens' perceptions of the need for government surveillance and perceptions of the government's motive for introducing the app, autonomy is captured by perceptions of control over one's information in the app, and perceived costs to the individual relate to perceptions of personal intrusion.

The second contribution of our study is the investigation of how perceptions of privacy are formed. Many studies have highlighted the negative impact of privacy concerns on CT adoption [11,52-54]. However, we argue that privacy can be seen as a factor driving adoption if citizens believe that apps can afford them with some level of privacy. Our findings demonstrate that citizens' privacy perceptions are shaped by trust in the app, which encompasses legitimacy perceptions and their perception of control offered by the app, and are negatively influenced by perceptions of intrusion. By highlighting the importance of fairness perceptions and elucidating the role of several perceptual factors at the governmental level (need for surveillance and government motives) and the app level (perceived intrusion and perceived control), which have been sparingly studied to date, our study advances our understanding of how privacy perceptions are developed in this context.

Understanding the factors driving CT app acceptance is paramount for future outbreaks [8]. The third contribution relates to understanding how citizens form intentions toward CT apps at different stages of the implementation process. Our study provides support for the influence of perceived privacy on individuals' intentions to adopt an app and willingness to disclose personal information before launch and support for an indirect influence over time on future use and willingness to disclose data. This finding supports many studies that found that privacy concerns negatively impacted adoption intentions toward mHealth [51]. In the context of a national CT app, if individuals perceive that the app offers a sufficient state of privacy, they will express positive intentions toward adoption and information disclosure before and after the app launch. Put

simply, perceived privacy can have a sustained positive influence on behavioral intentions.

Trust has been widely studied within the privacy and other domains to understand individuals' intentions to disclose information [42]. Trust in the app was found to influence individuals' adoption and disclosure intentions before launch, indirectly influencing use intentions after launch and directly influencing disclosure intentions after launch. These findings suggest that the influence of initial trust perceptions prevails over time and may operate as a heuristic for interacting with the app on an ongoing basis. The stability of trust perceptions and their ongoing influence are a relatively nascent topic, although some theorists have suggested the possibility of trust as a heuristic [55]. Our research provides empirical evidence for this phenomenon and offers further support for claims that initial trust perceptions might be relatively robust and long acting [56].

Implications for Practice

The findings of this study have several practical implications. First, the trust tension between public good and the intrusiveness of surveillance technologies has led researchers to emphasize the importance of effective trust-building strategies when introducing surveillance programs [31]. Indeed, citizens in the United States and Germany have expressed concerns regarding possible surveillance stemming from CT apps [9]. Our study shows that citizens' perceptions of trust and privacy can be influenced by fairness perceptions based on their beliefs regarding the need for surveillance and the government's motives for introducing surveillance technologies, the perceived control they are offered over their personal information and negatively influenced by their perceptions of the intrusiveness of these technologies. Thus, governments should focus on transparency in their public health surveillance efforts, including the involvement of data protection authorities and civil liberties advocates throughout the project life cycle, potentially through a privacy advisory committee [57]. This transparency should be extended to communication with citizens on the need and purpose of a technology while stressing the control they have over their personal information. Our research suggests that early communications that shape first impressions are particularly important. Such practices not only comply with data protection laws, such as the EU (European Union) General Data Protection Regulation, but also foster a sense of trust and ultimately influence the use of technology.

Second, our findings highlight the positive influence of privacy perceptions on adoption and disclosure. Thus, we argue that privacy should not be viewed as a barrier to new technologies, such as mHealth or CT apps, but rather as an important consideration throughout the design, implementation, and postlaunch stages. Designers should ensure compliance with the regulatory requirements for consent and control. Governments and other organizations charged with introducing new technologies should ensure that they clearly communicate their compliance with regulations and the considerations of individuals' personal data. Given that CT apps provide data on the location, copresence, and potentially ePHI of not only the focal person but also others that they have been in contact with, the principles of both necessity and proportionality would appear to be key. As per Ienca and Vayena [4], data collection must (1) be proportional to the seriousness of the public health threat, (2) be limited to what is necessary to achieve a specific public health objective, and (3) be scientifically justified. Policy makers and public health decision makers need to consider what communication and control mechanisms can be introduced to (1) build trust with the public and (2) repair trust, if necessary. This includes declaring what data will be collected and used while the app is live and by whom, confirming that data have been deleted, when no longer relevant (as is the case with COVID-19 data) or once the app is no longer required.

Third, in the context of technologies that require mass acceptance and willingness to share personal data, the focus cannot be placed solely on the number of downloads but must account for actual use and disclosure behaviors. Individuals' intentions to download CT apps influence their willingness to disclose personal information both before and after launch. Once they have downloaded the app, it is critical that decision makers encourage use and that the widespread use of the app is linked, through public communication, to successful intervention strategies so that the benefits to the individual and society are reinforced.

Limitations and Directions for Future Research

This study has several limitations. First, other factors may influence privacy and trust perceptions or moderate the relationships between trust and privacy and adoption. Although it is not possible to consider all potential antecedents and intervening variables, it would be interesting to explore the role of other prominent perceptions, such as perceived sensitivity, as these apps require users to disclose identifying information, health information, and location information, all of which are arguably sensitive. Second, although our study considers 2 important technology use outcomes, before and after the launch, this approach has limitations. First, the collection of data from the same respondents at multiple time points inevitably led to a drop in responses. Although we sent repeated invitations during the second phase of data collection, the final sample that completed both surveys was smaller. Although this is commonplace within this approach and our sample characteristics are similar to the broader population of Ireland, we acknowledge that a large sample would be ideal and stress the importance of considering the sample size when drawing inferences from our study. Second, our 2 time points did not allow us to take full advantage of the potential to model longitudinal change trajectories over time. Further work is needed to incorporate time more fully into our understanding of how privacy and trust influence adoption and use behaviors.

Third, our study relies on individuals' self-reported adoption and disclosure intentions. This approach is commonplace in the privacy and technology adoption literature streams, and it would not have been possible to study actual behaviors. However, we must acknowledge that intentions are not always matched with behaviors and that information disclosure is not always accurate or true. In other settings, it may be more feasible to collect more objective behavioral data, and we would encourage researchers to do so, particularly in settings where widespread adoption is

required for success. In addition, studies may go beyond our focus to understand disclosure behaviors at a deeper level and examine privacy-protective behaviors, such as withholding information or falsifying information. These protective behaviors are potentially dangerous in contexts such as CT apps because of the reliance on accurate data to track virus transmission.

Finally, our study explores a public health surveillance context where the focal person volunteers to participate and therefore has notice of the surveillance, control of their data and gives explicit consent. There are several conditions under which public health surveillance, including name reporting, may be undertaken without notice or explicit patient consent with well-established justifications in public health ethics, science, and law [58]. Even in the context of COVID-19, digital CT has not always been voluntary. In China, there is evidence of digital CT without notice or consent [59]. Furthermore, even when the focal person has notice and gives consent, contacts of the focal person have not given explicit consent. Although the primary focal person is subject to direct active surveillance, the secondary focal person is subject to passive indirect surveillance. In addition to the ethical issues that such practices raise, particularly where there is coordination and data exchange between private firms and the government [60], testing the theoretical framework developed in this study in this new context may provide a fruitful avenue of research. Similarly, aggregated anonymous spatiotemporal data sourced from commercial providers have been used as proxies for human movement and social interaction and as indicators of the effectiveness of social distancing interventions [61]. Although these data are currently anonymous, governments have already mandated access to identifiable data on the basis that the public interest overrides privacy rights [62]. This context may provide interesting insights and further extend our understanding of the limits of consumer acceptance of governmental health surveillance.

Conclusions

The COVID-19 pandemic was the first time governments implemented large-scale digital CT. Its success as a public health intervention depended on rapid technology adoption by a significant proportion of the population. Here, surveillance is active, and the target of government surveillance through COVID-19 digital CT apps is an active participant in sharing data with the government on their personal health status, their location, and often their social network. The opportunity to study such an empirical context is not only rare but also the time frame for research is limited. Understanding the formation of individuals' perceptions of trust and privacy in this context and how these perceptions influence their acceptance of digital CT apps is critical not only for informing the design of future digital CT initiatives but also for other situations that require rapid digital technology adoption by a significant proportion of society. If governments wish to leverage the power of digital technologies to control future public health threats, we recommend 3 principles to guide the design of both their surveillance initiatives and communications with the public—necessity, transparency, and proportionality—before and after the launch.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey items, validity, and reliability testing.
[PDF File (Adobe PDF File), 172 KB - mhealth_v12i1e48700_app1.pdf]

References

1. Eom SJ, Lee J. Digital government transformation in turbulent times: responses, challenges, and future direction. Gov Inf Q 2022 Apr;39(2):101690 [FREE Full text] [doi: 10.1016/j.giq.2022.101690] [Medline: 35291492]
2. Trang S, Trenz M, Weiger WH, Tarafdar M, Cheung CM. One app to trace them all? Examining app specifications for mass acceptance of contact-tracing apps. Eur J Inf Syst 2020 Jul 27;29(4):415-428. [doi: 10.1080/0960085x.2020.1784046]
3. Kolasa K, Mazzi F, Leszczuk-Czubkowska E, Zrubka Z, Péntek M. State of the art in adoption of contact tracing apps and recommendations regarding privacy protection and public health: systematic review. JMIR Mhealth Uhealth 2021 Jun 10;9(6):e23250 [FREE Full text] [doi: 10.2196/23250] [Medline: 34033581]
4. Ienca M, Vayena E. On the responsible use of digital data to tackle the COVID-19 pandemic. Nat Med 2020 Apr 27;26(4):463-464 [FREE Full text] [doi: 10.1038/s41591-020-0832-5] [Medline: 32284619]
5. Budd J, Miller BS, Manning EM, Lampos V, Zhuang M, Edelstein M, et al. Digital technologies in the public-health response to COVID-19. Nat Med 2020 Aug 07;26(8):1183-1192 [FREE Full text] [doi: 10.1038/s41591-020-1011-4] [Medline: 32770165]
6. Marabelli M, Vaast E, Li JL. Preventing the digital scars of COVID-19. Eur J Inf Syst 2021 Jan 19;30(2):176-192. [doi: 10.1080/0960085x.2020.1863752]
7. Mora L, Kummitha RK, Esposito G. Not everything is as it seems: digital technology affordance, pandemic control, and the mediating role of sociomaterial arrangements. Gov Inf Q 2021 Oct;38(4):101599 [FREE Full text] [doi: 10.1016/j.giq.2021.101599] [Medline: 36570778]

8. Van Fossen JA, Ropp JW, Darcy K, Hamm JA. Comfort with and willingness to participate in COVID-19 contact tracing: the role of risk perceptions, trust, and political ideology. Soc Sci Med 2022 Aug;306:115174 [FREE Full text] [doi: 10.1016/j.socscimed.2022.115174] [Medline: 35777171]
9. Altmann S, Milsom L, Zillessen H, Blasone R, Gerdon F, Bach R, et al. Acceptability of app-based contact tracing for COVID-19: cross-country survey study. JMIR Mhealth Uhealth 2020 Aug 28;8(8):e19857 [FREE Full text] [doi: 10.2196/19857] [Medline: 32759102]
10. Moon LA. Factors influencing health data sharing preferences of consumers: a critical review. Health Policy Technol 2017 Jun;6(2):169-187. [doi: 10.1016/j.hlpt.2017.01.001]
11. Lin J, Carter L, Liu D. Privacy concerns and digital government: exploring citizen willingness to adopt the COVIDSafe app. Eur J Inf Syst 2021 May 05;30(4):389-402. [doi: 10.1080/0960085x.2021.1920857]
12. Munzert S, Selb P, Gohdes A, Stoetzer LF, Lowe W. Tracking and promoting the usage of a COVID-19 contact tracing app. Nat Hum Behav 2021 Feb 21;5(2):247-255. [doi: 10.1038/s41562-020-01044-x] [Medline: 33479505]
13. Nunes N, Adamo G, Ribeiro M, R Gouveia BR, Rubio Gouveia E, Teixeira P, et al. Modeling adoption, security, and privacy of COVID-19 apps: findings and recommendations from an empirical study using the unified theory of acceptance and use of technology. JMIR Hum Factors 2022 Sep 14;9(3):e35434 [FREE Full text] [doi: 10.2196/35434] [Medline: 35862671]
14. von Wyl V, Höglinger M, Sieber C, Kaufmann M, Moser A, Serra-Burriel M, et al. Drivers of acceptance of COVID-19 proximity tracing apps in Switzerland: panel survey analysis. JMIR Public Health Surveill 2021 Jan 06;7(1):e25701 [FREE Full text] [doi: 10.2196/25701] [Medline: 33326411]
15. Fox G, Clohessy T, van der Werff L, Rosati P, Lynn T. Exploring the competing influences of privacy concerns and positive beliefs on citizen acceptance of contact tracing mobile applications. Comput Hum Behav 2021 Aug;121:106806. [doi: 10.1016/j.chb.2021.106806]
16. Dinev T, Xu H, Smith JH, Hart P. Information privacy and correlates: an empirical attempt to bridge and distinguish privacy-related concepts. Eur J Inf Syst 2017 Dec 19;22(3):295-316. [doi: 10.1057/ejis.2012.23]
17. Fox G, van der Werff L, Rosati P, Takako Endo P, Lynn T. Examining the determinants of acceptance and use of mobile contact tracing applications in Brazil: an extended privacy calculus perspective. J Assoc Inf Sci Technol 2021 Nov 10;73(7):944-967. [doi: 10.1002/asi.24602]
18. Shoji M, Ito A, Cato S, Iida T, Ishida K, Katsumata H, et al. Prosociality and the uptake of COVID-19 contact tracing apps: survey analysis of intergenerational differences in Japan. JMIR Mhealth Uhealth 2021 Aug 19;9(8):e29923 [FREE Full text] [doi: 10.2196/29923] [Medline: 34313601]
19. Hassandoust F, Akhlaghpour S, Johnston AC. Individuals' privacy concerns and adoption of contact tracing mobile applications in a pandemic: a situational privacy calculus perspective. J Am Med Inform Assoc 2021 Mar 01;28(3):463-471 [FREE Full text] [doi: 10.1093/jamia/ocaa240] [Medline: 33164077]
20. Laufer RS, Wolfe M. Privacy as a concept and a social issue: a multidimensional developmental theory. J Soc Issues 2010 Apr 14;33(3):22-42. [doi: 10.1111/j.1540-4560.1977.tb01880.x]
21. Lewicki RJ, Bunker BB. Developing and maintaining trust in work relationships. In: Kramer R, Tyler T, editors. Trust in Organizations: Frontiers of Theory and Research. Thousand Oaks, CA: SAGE Publications, Inc; 1996:114-139.
22. Culnan MJ, Armstrong PK. Information privacy concerns, procedural fairness, and impersonal trust: an empirical investigation. Org Sci 1999 Feb;10(1):104-115. [doi: 10.1287/orsc.10.1.104]
23. Xu H, Teo HH, Tan BC, Agarwal R. The role of push-pull technology in privacy calculus: the case of location-based services. J Manage Inf Syst 2014 Dec 08;26(3):135-174. [doi: 10.2753/mis0742-1222260305]
24. Searle R, Den Hartog DN, Weibel A, Gillespie N, Six F, Hatzakis T, et al. Trust in the employer: the role of high-involvement work practices and procedural justice in European organizations. Int J Hum Resour Manage 2011 Mar;22(5):1069-1092. [doi: 10.1080/09585192.2011.556782]
25. Fishbein M, Ajzen I. Belief, Attitude, Intention and Behaviour: An Introduction to Theory and Research. London, UK: Longman Higher Education; 1976.
26. McKnight DH, Choudhury V, Kacmar C. Developing and validating trust measures for e-commerce: an integrative typology. Inf Syst Res 2002 Sep;13(3):334-359. [doi: 10.1287/isre.13.3.334.81]
27. Mcknight DH, Carter M, Thatcher JB, Clay PF. Trust in a specific technology: an investigation of its components and measures. ACM Trans Manage Inf Syst 2011 Jul;2(2):1-25. [doi: 10.1145/1985347.1985353]
28. Mayer RC, Davis JH, Schoorman FD. An integrative model of organizational trust. Acad Manage Rev 1995 Jul;20(3):709-734. [doi: 10.2307/258792]
29. van der Werff L, Buckley F. Getting to know you: a longitudinal examination of trust cues and trust development during socialization. J Manage 2016 Jul 09;43(3):742-770. [doi: 10.1177/0149206314543475]
30. Thompson N, McGill T, Bunn A, Alexander R. Cultural factors and the role of privacy concerns in acceptance of government surveillance. J Assoc Inf Sci Technol 2020 Jun 25;71(9):1129-1142. [doi: 10.1002/asi.24372]
31. Dutton W, Guerra GA, Zizzo DJ, Peltu M. The cyber trust tension in e-government: balancing identity, privacy, security. Inf Polity 2005 Nov 22;10(1-2):13-23. [doi: 10.3233/ip-2005-0066]

32. Kim ES, Chung JB. Korean mothers' morality in the wake of COVID-19 contact-tracing surveillance. Soc Sci Med 2021 Feb;270:113673 [FREE Full text] [doi: 10.1016/j.socscimed.2021.113673] [Medline: 33453628]
33. Dinev T, Hart P, Mullen MR. Internet privacy concerns and beliefs about government surveillance – an empirical investigation. J Strat Inf Syst 2008 Sep;17(3):214-233. [doi: 10.1016/j.jsis.2007.09.002]
34. Lipponen J, Kaltiainen J, van der Werff L, Steffens NK. Merger-specific trust cues in the development of trust in new supervisors during an organizational merger: a naturally occurring quasi-experiment. Leadership Q 2020 Aug;31(4):101365. [doi: 10.1016/j.leaqua.2019.101365]
35. Xu H, Dinev T, Smith HJ, Hart P. Examining the formation of individual's privacy concerns: toward an integrative view. In: Proceedings of the International Conference on Information Systems 2008. 2008 Presented at: ICIS 2008; December 14-17, 2008; Paris, France.
36. Zimmermann BM, Fiske A, Prainsack B, Hangel N, McLennan S, Buyx A. Early perceptions of COVID-19 contact tracing apps in German-speaking countries: comparative mixed methods study. J Med Internet Res 2021 Feb 08;23(2):e25525 [FREE Full text] [doi: 10.2196/25525] [Medline: 33503000]
37. Malhotra NK, Kim SS, Agarwal J. Internet users' information privacy concerns (IUIPC): the construct, the scale, and a causal model. Inf Syst Res 2004 Dec;15(4):336-355. [doi: 10.1287/isre.1040.0032]
38. Colquitt JA, Lepine JA, Piccolo RF, Zapata CP, Rich BL. Explaining the justice-performance relationship: trust as exchange deepener or trust as uncertainty reducer? J Appl Psychol 2012 Jan;97(1):1-15. [doi: 10.1037/a0025208] [Medline: 21910516]
39. Oldeweme A, Märtins J, Westmattelmann D, Schewe G. The role of transparency, trust, and social influence on uncertainty reduction in times of pandemics: empirical study on the adoption of COVID-19 tracing apps. J Med Internet Res 2021 Feb 08;23(2):e25893 [FREE Full text] [doi: 10.2196/25893] [Medline: 33465036]
40. Yasaka TM, Lehrich BM, Sahyouni R. Peer-to-peer contact tracing: development of a privacy-preserving smartphone app. JMIR Mhealth Uhealth 2020 Apr 07;8(4):e18936 [FREE Full text] [doi: 10.2196/18936] [Medline: 32240973]
41. Venkatesh V, Thong JY, Chan FK, Hu PJ. Managing citizens' uncertainty in e-government services: the mediating and moderating roles of transparency and trust. Inf Syst Res 2016 Mar;27(1):87-111. [doi: 10.1287/isre.2015.0612]
42. McEvily B, Tortoriello M. Measuring trust in organisational research: review and recommendations. J Trust Res 2011 Apr;1(1):23-63. [doi: 10.1080/21515581.2011.552424]
43. Hoehle H, Aloysius JA, Goodarzi S, Venkatesh V. A nomological network of customers' privacy perceptions: linking artifact design to shopping efficiency. Eur J Inf Syst 2018 Jul 24;28(1):91-113. [doi: 10.1080/0960085x.2018.1496882]
44. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003 Sep;27(3):425-478. [doi: 10.2307/30036540]
45. Brennan C, McConnell D. HSE's Covid-19 tracing app passes 1m downloads. Irish Examiner. 2020 Jul 8. URL: https://www.irishexaminer.com/news/arid-31010089.html [accessed 2022-07-01]
46. O'Callaghan ME, Abbas M, Buckley J, Fitzgerald B, Johnson K, Laffey J, et al. Public opinion of the Irish "COVID tracker" digital contact tracing app: a national survey. Digit Health 2022 Mar 16;8 [FREE Full text] [doi: 10.1177/20552076221085065] [Medline: 35321018]
47. Serrano Archimi C, Reynaud E, Yasin HM, Bhatti ZA. How perceived corporate social responsibility affects employee cynicism: the mediating role of organizational trust. J Bus Ethics 2018 May 2;151(4):907-921. [doi: 10.1007/s10551-018-3882-6]
48. Gaskin J, Lim J. Model bias: AMOS plugin. Gaskination's Statistics. URL: http://statwiki.kolobkreations.com/ [accessed 2020-10-10]
49. Fornell C, Larcker DF. Evaluating structural equation models with unobservable variables and measurement error. J Mark Res 2018 Nov 28;18(1):39-50. [doi: 10.1177/002224378101800104]
50. Raykov T. Estimation of composite reliability for congeneric measures. Appl Psychol Meas 2016 Jul 27;21(2):173-184. [doi: 10.1177/014662169702100206]
51. Fox G, Connolly R. Mobile health technology adoption across generations: narrowing the digital divide. Inf Syst J 2018 Jan 29;28(6):995-1019. [doi: 10.1111/isj.12179]
52. Walrave M, Waeterloos C, Ponnet K. Adoption of a contact tracing app for containing COVID-19: a health belief model approach. JMIR Public Health Surveill 2020 Sep 01;6(3):e20572 [FREE Full text] [doi: 10.2196/20572] [Medline: 32755882]
53. Seto E, Challa P, Ware P. Adoption of COVID-19 contact tracing apps: a balance between privacy and effectiveness. J Med Internet Res 2021 Mar 04;23(3):e25726 [FREE Full text] [doi: 10.2196/25726] [Medline: 33617459]
54. Walrave M, Waeterloos C, Ponnet K. Reasons for nonuse, discontinuation of use, and acceptance of additional functionalities of a COVID-19 contact tracing app: cross-sectional survey study. JMIR Public Health Surveill 2022 Jan 14;8(1):e22113 [FREE Full text] [doi: 10.2196/22113] [Medline: 34794117]
55. McEvily B. Reorganizing the boundaries of trust: from discrete alternatives to hybrid forms. Org Sci 2011 Oct;22(5):1266-1276. [doi: 10.1287/orsc.1110.0649]
56. van der Werff L, Legood A, Buckley F, Weibel A, de Cremer D. Trust motivation: the self-regulatory processes underlying trust decisions. Org Psychol Rev 2019 Sep 11;9(2-3):99-123. [doi: 10.1177/2041386619873616]

57. Oliver N, Lepri B, Sterly H, Lambiotte R, Deletaille S, De Nadai M, et al. Mobile phone data for informing public health actions across the COVID-19 pandemic life cycle. Sci Adv 2020 Jun 05;6(23):eabc0764 [FREE Full text] [doi: 10.1126/sciadv.abc0764] [Medline: 32548274]
58. Lee LM, Heilig CM, White A. Ethical justification for conducting public health surveillance without patient consent. Am J Public Health 2012 Jan;102(1):38-44. [doi: 10.2105/ajph.2011.300297]
59. Mozur P, Zhong R, Krolik A. In coronavirus fight, China gives citizens a color code, with red flags. The New York Times. 2020 Mar 1. URL: https://www.nytimes.com/2020/03/01/business/china-coronavirus-surveillance.html [accessed 2023-03-01]
60. Parker MJ, Fraser C, Abeler-Dörner L, Bonsall D. Ethics of instantaneous contact tracing using mobile phone apps in the control of the COVID-19 pandemic. J Med Ethics 2020 Jul 04;46(7):427-431 [FREE Full text] [doi: 10.1136/medethics-2020-106314] [Medline: 32366705]
61. Endo PT, Silva I, Lima L, Bezerra L, Gomes R, Ribeiro-Dantas M, et al. #StayHome: monitoring and benchmarking social isolation trends in Caruaru and the Região Metropolitana do Recife during the COVID-19 pandemic. Rev Soc Bras Med Trop 2020 Jun 24;53:e20200271 [FREE Full text] [doi: 10.1590/0037-8682-0271-2020] [Medline: 32609249]
62. Tidy J. Coronavirus: Israel enables emergency spy powers. BBC News. 2020 Mar 17. URL: https://www.bbc.com/news/technology-51930681 [accessed 2023-03-01]

Abbreviations

AVE: average variance extracted
CT: contact tracing
ePHI: electronic personal health information
EU: European Union
mHealth: mobile health
PCT: Privacy Calculus Theory
PTTT: propensity to trust technology
TRA: theory of reasoned action

Edited by L Buis; submitted 10.05.23; peer-reviewed by Z Zrubka, M Shoji; comments to author 18.07.23; revised version received 20.10.23; accepted 06.12.23; published 18.01.24.

Please cite as:
Fox G, van der Werff L, Rosati P, Lynn T
Investigating Citizens' Acceptance of Contact Tracing Apps: Quantitative Study of the Role of Trust and Privacy
JMIR Mhealth Uhealth 2024;12:e48700
URL: https://mhealth.jmir.org/2024/1/e48700
doi:10.2196/48700
PMID:38085914

Original Paper

Investigating the Integration and the Long-Term Use of Smart Speakers in Older Adults' Daily Practices: Qualitative Study

Fangyuan Chang[1], PhD; Lin Sheng[1], MEng; Zhenyu Gu[1], PhD

Interaction Design Lab, School of Design, Shanghai Jiao Tong University, Shanghai, China

Corresponding Author:
Zhenyu Gu, PhD
Interaction Design Lab
School of Design
Shanghai Jiao Tong University
800 Dongchuan Road
Shanghai, 200240
China
Phone: 86 13167232872
Email: zygu@sjtu.edu.cn

Abstract

Background: As smart speakers become more popular, there have been an increasing number of studies on how they may benefit older adults or how older adults perceive them. Despite the increasing ownership rates of smart speakers among older adults, studies that examine their integration and the long-term use in older adults' daily practices are scarce.

Objective: This study aims to uncover the integration of smart speakers into the daily practices of older adults over the long term, contributing to an in-depth understanding of maintained technology use among this demographic.

Methods: To achieve these objectives, the study interviewed 20 older adults who had been using smart speakers for over 6 months. These semistructured interviews enabled participants to share their insights and experiences regarding the maintained use of smart speakers in the long term.

Results: We identified 4 dimensions of the long-term use of smart speakers among older adults, including functional integration, spatial integration, cognitive integration, and semantic integration. For the functional integration of smart speakers, the study reported different types of use, including entertainment, information collection, medication reminders, companionship, environment modification, and emergency calls. For the spatial integration of smart speakers, the study showed older adults' agency in defining, changing, and reshaping daily practices through the spatial organization of smart speakers. For the cognitive integration of smart speakers, the findings showed the cognitive processes involved in adapting to and incorporating smart speakers into daily habits and routines. For the semantic integration of smart speakers, the findings revealed that older adults' enjoyable user experience and strong bonds with the device contributed to their acceptance of occasional functional errors. Finally, the study proposed several suggestions for designers and developers to better design smart speakers that promote maintainable use behaviors among older adults.

Conclusions: On the basis of the findings, this study highlighted the importance of understanding how older adults use smart speakers and the practices through which they integrate them into their daily routines. The findings suggest that smart speakers can provide significant benefits for older adults, including increased convenience and improved quality of life. However, to promote maintainable use behaviors, designers and developers should consider more about the technology use contexts and the specific needs and preferences of older adults when designing these devices.

(JMIR Mhealth Uhealth 2024;12:e47472) doi:10.2196/47472

KEYWORDS

smart speaker; private home; older adults; long-term use; daily practices; smart speakers

Introduction

Along with the advancement of speech technology and artificial intelligence, smart speakers such as Google Home and Amazon Echo are becoming integral to households [1]. Equipped with smart voice assistants such as Google Assistant and Amazon Alexa, these devices respond to voice commands, facilitating activities like playing music, answering questions, setting reminders, and controlling smart home appliances. Their speech input and output features enhance accessibility, especially for individuals with limited mobility and vision [2,3].

In the realm of health care, the adoption of smart speakers has opened avenues for significant advancements, akin to the transformative impact of mobile phones. These devices present unique advantages for health research in out-of-hospital environments, offering opportunities for chronic disease management, passive identification of medical emergencies, detection of behavioral and cognitive changes, and remote monitoring of respiratory diseases impacting public health [4,5]. A range of successful pilot studies has demonstrated the positive impact of smart speakers on users requiring social care, and adults with learning differences, showcasing potential cost savings and improved well-being [6,7].

Despite the potential benefits, scholars have highlighted a notable gap in knowledge concerning older adults' experiences with smart speakers in long-term use, which poses challenges to assessing the long-term impact and implications of these devices on the well-being and quality of life of older adults. Specifically, while research has explored smart speaker use among various demographics, such as low-income populations [8,9], people with disabilities [10,11], parents [12,13], and children [14], studies specifically focusing on the older adult demographic remain relatively scarce [3,15]. Among the limited studies, the majority of the existing literature concentrates on the design of smart speakers for later life, encompassing discussions on effective conversational cues [16], privacy concerns [17], and the anthropomorphism of speakers [18]. Other studies have assessed the feasibility and usability of smart speakers for promoting active aging [19,20], investigated older adults' first impressions of smart speakers [19,21], and identified influential factors regarding older adults' attitudes toward smart speakers [22]. The most recent studies emphasized there is a missing knowledge about older adults' experiences with smart speakers in long-term use. This knowledge gap hinders the development of tailored interventions and policies to maximize the benefits of smart speakers for this demographic, ensuring their inclusion in the digital revolution and promoting healthy aging in the digital era [23-25]. Consequently, there is a pressing need for comprehensive exploration into older adults' long-term experiences with smart speakers, emphasizing their perceptions, attitudes, and the evolving nature of their interactions with this technology.

This study aims to understand how older adults routinely use smart speakers and integrate them into their daily lives. The outcomes can add to the emerging body of literature for a more comprehensive understanding of older adults' use of smart speakers in the long term. We conducted semistructured interviews with 20 older adults who used the smart speaker for over half a year, to answer the following research questions: "How does the use of smart speakers integrate into older adults' daily practices?" and "What design considerations can be generated from the long-term use of smart speakers by older adults?"

In summary, our study embarks on an exploration into the long-term experiences of older adults with smart speakers, filling a notable gap in the existing knowledge. While previous research has delved into aspects such as conversational cues, the feasibility of smart speakers for active aging, and the before-and-after adoption changes or the broader impact, the focus on the nuanced, day-to-day interactions of older adults with these devices over an extended period is a novel dimension that sets our study apart [26]. By unraveling the complexities and diverse experiences during older adults' long-term engagement with smart speaker technology, we aim to contribute not only to the effective design of age-friendly devices but also to the broader discourse on the role of technology in promoting well-being among older populations.

Methods

Study Design

A qualitative research study was conducted to investigate the experiences of older adults' long-term use of smart speakers in daily practices. Semistructured interviews were used to facilitate an in-depth exploration of older adults' experiences, even when the study involved a relatively small number of participants [27,28]. The study adhered to the guidelines and reporting standards outlined in the Consolidated Criteria for Reporting Qualitative Research (COREQ) checklist for qualitative studies [29].

Ethical Considerations

Ethical approval for the research was obtained from the research ethics committee of the Shanghai Jiao Tong University (H2022335I). Informed consent was obtained from all participants.

Participants and Recruitment

The research was conducted in Shanghai, China. To ensure diverse perspectives, participants were recruited from 2 distinct communities of older people—one situated in a bustling urban environment and the other in a rural area on the outskirts of the city. In the urban community, the researchers partnered with a community coordinator who had established relationships with older adults. In the rural community, the research team collaborated with the head of the older people's community, who was responsible for arranging community activities. The coordinator and the community head served as a liaison between the research team and the two communities to help in identifying suitable candidates.

The specific criteria for participant inclusion in the study remained consistent across both communities, requiring individuals to be aged 60 years or older and to have a minimum of 6 months of experience using a smart speaker in their homes. The time limit served to ensure that current practices of using

smart speakers were represented and to improve the reliability of experience recollection. Therefore, to be included in the study, participants needed to be aged 60 years or older and have had at least 6 months of experience using a smart speaker in their homes. Community members were thus excluded if they did not meet the age requirement, or had less than 6 months of experience with a smart speaker.

During the recruitment process, the research team first contacted the community coordinator or the community head, who shared the information about this research with all the members of the community. The older adults who met the inclusion criteria and showed their interest in participating in the study were recruited. The snowball approach was then used to identify more older adults using smart speakers for more than 6 months. Those who met the inclusion criteria were recruited as participants. The participants were compensated with gift cards (valued at around US $9) for interview participation. Ultimately, 20 older adults were recruited (Table 1). The sample size is in line with the recommendations by Guest et al [30] and Hennink et al [31], who pointed out that empirical data reached saturation within 20 interviews. Other qualitative literature about older adults' use of technology used a similar number of participants [24,32].

Table 1. Demographics of participants.

Label	Sex	Age (years)	Years of experience[a]	Technical proficiency[b]	Used smart speaker
OA1[c]	Female	72	2	Intermediate	Brand A
OA2	Female	66	0.5	Intermediate	Brand A
OA3	Female	69	1	Novice	Brand A
OA4	Female	71	1	Novice	Brand A
OA5	Male	73	3	Novice	Brand B
OA6	Male	67	2.5	Novice	Brand B
OA7	Male	68	0.5	Intermediate	Brand A
OA8	Female	61	0.5	Novice	Brand B
OA9	Female	73	2	Intermediate	Brand A
OA10	Female	66	2	Novice	Brand A
OA11	Male	77	1.5	Intermediate	Brand B
OA12	Female	65	2	Novice	Brand B
OA13	Male	73	1.5	Intermediate	Brand B
OA14	Female	64	1	Novice	Brand A
OA15	Female	73	1.5	Novice	Brand A
OA16	Male	73	1	Intermediate	Brand B
OA17	Male	62	0.5	Intermediate	Brand B
OA18	Male	64	1	Novice	Brand B
OA19	Female	69	2	Novice	Brand A
OA20	Female	70	1.5	Intermediate	Brand B

[a]The actual time that the participant starts to use the smart speaker regularly (self-reported).

[b]Individuals' ability to discern the usefulness of a smart speaker and their comfort in using it for various purposes (evaluated by the Artificial Intelligence Literacy Questionnaire proposed in [33], Multimedia Appendix 1).

[c]OA: older adult.

Data Collection

A literature review of technology use among older adults informed the development of a flexible and semistructured interview guide. The interview guide was developed in consultation with a professor with expertise in qualitative data collection and analysis. One preliminary interview was conducted with the included community coordinator (aged 58 years) to ensure the fluency of formal interviews (eg, whether questions are easy to understand). As the goal of our study was to better understand older adults' long-term use of smart speakers in home environments, the research questions were developed encompassing their practices, experiences, values, and expectations about using the smart speaker (Multimedia Appendix 2).

All interviews received written consent and were conducted in a private confidential environment such as the community activity room and private homes between November 2022 and September 2023. The participants were encouraged to give examples to contextualize their daily practices of using the smart speaker. The interviews ranged in length between 30 and 60 minutes. Participants who had limited knowledge or experience in using smart speakers had shorter interviews. All interviews

were audio-recorded. During the interview, we conducted regular summaries to ensure the validity of the collected data.

Data Analysis

All interviews were transcribed. The interview transcripts were reported in a way to ensure the anonymity of the participants. The collected data were interpreted by inductive content analysis [34], with a focus on how older adults use the smart speaker in their home environments, and why the speaker is used in the observed way. The transcripts were analyzed in NVivo (version 12; Lumivero) following the process proposed by Braun and Clarke [27]. After a data familiarization stage, 2 authors (FC and LS) separately extracted text segments related to the research goal and categorized them into codes. Applying an iterative process, the relationships between codes were analyzed and the subthemes were formed by grouping related segments together. The 2 authors compared the codes and subthemes and resolved any differences in coding through discussion. The assignment of text segments to the subcategories was repeatedly checked to see if they reflected the same meaning. Data saturation was determined when no new themes and relationships among the interview data were found [35]. As a result, a list of 4 themes and 12 subthemes, upon which the authors achieved an agreement, was generated [36].

Results

Overview

We identified 4 dimensions of the long-term use of smart speakers among older adults, including functional integration, spatial integration, cognitive integration, and semantic integration (Table 2).

Table 2. Identified themes, subthemes, and code examples related to the long-term use of smart speakers among older adults.

Themes and subthemes	Code examples
Functional integration	
Entertainment	• Using it for listening to music • Chatting with the small thing for fun • Telling some stories
Information collection	• Hear daily news • Weather information • A useful tool for knowing what is going on
Medication reminders	• The speaker can send out medication reminders • When I see it, I know, “Oh I need to take pills”
Companionship	• Enjoy hearing the speaker’s answers • The speaker is accompanying me, he is a friend • Love sitting in the coach, with her beside me
Environment modification	• Just need to say “Turn off the lights” • Connect it with the switch • Connect to the camera to show who is knocking on the door
Emergency calls	• Call for help if falling on the ground
Spatial integration	
Smart speaker location	• Decide the device’s location according to daily habits • Put it beside the medicine box for task convenience
Appliance connection	• Locate the smart speaker in the bedroom for device connection • Put the smart speaker beside the door for better internet • Put it on the bedside table for easily hearing the doorbells
Cognitive integration	
Learning and problem-solving	• Memorize the required voice commands • Handle the emerging technical hurdles
Mental adaptation	• Overcome skepticism and building trust • Realize the benefits of smart speakers for personal convenience
Semantic integration	
Enjoyable experience	• Save energy in memorizing things • Assist older adults in performing multitask operations
Affective bonds	• Feel a shared and caring presence through the technology use • Experience tranquility while reading with the smart speaker

Functional Integration of Smart Speakers

Functional integration relates to the ways in which older adults use smart speakers to meet various functional needs in their daily lives. This dimension focuses on the practical and functional aspects of smart speaker use and how they are integrated into daily practices. Our findings revealed the diversity of use genres that had been developed in older adults’ use of smart speakers in their living environments. Though most participants were using a smart speaker of the same brand, the ways of using the device were multiple, including entertainment, information collection, medication reminders, companionship, environment modification, as well as emergency calls.

The most commonly mentioned use genre was entertainment. Almost all participants said they use the smart speaker to listen to music.

> *...I often listen to songs [via the smart speaker]. Some songs could be used for our square dance.* [Female, age 69 years]

In addition to listening to music, 1 participant mentioned that he once used the smart speaker to recognize the name of a song.

> *...I was recalling the name of a song. My granddaughter told me if I sang the song’s melody, the speaker could tell me the song’s name.* [Male, age 73 years]

Similar to the example above, using the smart speaker for information collection was widely found in the interviews. Our participants used the smart speaker to gain information such as daily weather, local news, and safety tips related to the COVID-19 pandemic.

The smart speaker was also commonly used for medication reminders. The participants indicated that they had to always remember when to take medicines before using the smart speaker. With the help of their families or relatives, they set up the smart speaker to remind them to take medicines at a specific time.

> *...I have to take medicines every day.... Now the speaker can remind me of taking medicines.* [Female, age 66 years]

Besides, the participants emphasized that they benefited from the companionship of the device. The following example showed how older adults saw the smart speaker as a person who accompanied them, and enjoyed the conversation with the smart speaker.

> *...My smart speaker is a well-behaved little girl, just like my granddaughter. She wakes me up every morning. I say good morning to her sometimes. Sometimes I ask her some questions that I know the answers.... I am just curious about how she would reply to me, and enjoy hearing her answers.* [Male, age 73 years]

Another use genre was about environment modification. A few participants used the smart speaker to control other smart appliances in their living environments. For instance, 1 participant explained that the smart speaker helped overcome the inconvenience caused by switches that were in distant places.

> *The switch in my bedroom is beside the door.... I had to get out of bed, switch off the light, and then get into the bed in the darkness.... Now I just need to say "turn off the lights.*" [Female, age 72 years]

Finally, 1 participant mentioned that she had not used the smart speaker for emergency calls, but she knew she could call her daughters through the smart speaker or a "white button" when necessary.

> *I asked the smart speaker to call my daughter, her phone rang. I clicked the white button, her phone rang...[a white button, about 5 cm in diameter, is shown in her hand]*. [Female, age 61 years]

Spatial Integration of Smart Speakers

Spatial integration relates to the ways in which older adults use the physical space and placement of their smart speakers to shape and organize their daily routines. This dimension focuses on the physical aspects of smart speaker use. Our findings suggested that participants located their smart speakers in different living spaces, including the living room, the bedroom, the study, the bathroom, and the balcony. By spatially organizing the speaker, and connecting it with other smart appliances or daily tasks in these contexts, the participants built up the use contexts under which they are comfortable interacting with the smart speakers, and actively changed and reshaped their daily practices.

The following example showed how participants decided the locations of the smart speaker and formed new daily practices according to the purposes for using the smart speaker and the existing daily practices.

> *...I wanted to take the pills, I had to get water from the kitchen, so I put the speaker, together with the medicine box, in the corner of my kitchen now.* [Female, age 73 years]

Some participants incorporated the smart speaker by connecting it with the existing appliances in their living environments. In doing so, their practices of completing specific tasks were simplified.

> *...The light in my bedroom can be connected and controlled by the smart speaker, I thus put it in my bedroom. Now I just need to say "turn off the lights" when I want to sleep.* [Female, age 72 years]

Cognitive Integration of Smart Speakers

Cognitive integration relates to the way in which older adults adapt to and incorporate smart speakers into daily habits and routines from a cognitive perspective. Participants shared their experiences of familiarizing themselves with the smart speaker's user interface and functionalities. The impact of successful problem-solving was evident in the participants' narratives. Many acknowledged that navigating through and resolving issues not only enhanced their understanding of the device but also strengthened their overall intention to use the smart speaker in the long term. Overcoming learning or technical challenges contributed to a sense of accomplishment and increased confidence in using the smart speaker.

Almost all participants mentioned the challenges in learning voice commands and the overall functionality of the device, as well as resolving technical hurdles. They also mentioned the importance of adept problem-solving skills when facing these challenges. Some participants recounted instances of encountering glitches such as software malfunctions, intermittent connectivity issues, and occasional misinterpretations of voice commands by the smart speaker. One participant shared a specific experience, recalling:

> *There were times when it simply didn't respond, leaving me puzzled. I had to delve into troubleshooting methods to identify the issue.* [Female, age 64 years]

Another participant highlighted the initial difficulty in grasping the full range of functions, expressing:

> *Mastering the voice commands posed a bit of a challenge. Instead of saying "XX, I need..." you have to initiate with "Hi, XX."...took me some time to unravel the full range of functions it offered.* [Male, age 62 years]

In response to these challenges, older adults demonstrated resilience and resourcefulness in addressing technical issues. Problem-solving strategies ranged from seeking help from

family members or friends to experimenting with different emotional and cognitive aspects of smart speaker use. Some engaged in hands-on exploration, experimenting with different commands to enhance their functional and mental understanding. Additionally, participants discussed seeking guidance from families, internet-based resources, and user manuals to expedite the adaptation process. As 1 participant shared:

> *I watched some tutorial videos online and read the manual, and asked my family. It helped me get a better grasp of what the smart speaker could do.* [Female, age 65 years]

As smart speakers became integrated into daily routines, participants described making gradual adjustments. The learning curve transformed into a journey of mental adaptation, with older adults incorporating the device into activities such as setting reminders, checking the weather, or even incorporating it into their medication routines. One participant shared the experience of overcoming skepticism and building trust with the smart speaker over time.

> *I was skeptical at first...But as I gradually used it, I found myself relying on it more.* [Male, age 73 years]

The transformative power of smart speakers was evident in their ability to enhance social interactions during family gatherings. One user realized the potential of the device in this context, stating:

> *I realized the potential of using the smart speaker during family gatherings. Now, we use it to play music, share interesting facts, and even settle debates.* [Male, age 64 years]

Similarly, another user's experience highlighted the versatile nature of the smart speaker. Initially used for simple tasks like checking the weather and setting alarms, the device became a gateway to a multitude of possibilities.

> *Initially, I used it mainly for weather updates and setting alarms. But then, I discovered I could ask it to read audiobooks or provide cooking tips. It's like unlocking a treasure trove of possibilities.* [Female, age 70 years]

Semantic Integration of Smart Speakers

Semantic integration refers to the emotional and meaningful connections that older adults develop with their smart speakers over time. This dimension focuses on the emotional and cognitive aspects of smart speaker use. According to our data, although the smart speakers had occasional functional errors that annoyed the participants, all participants acknowledged, appreciated, and valued the benefits offered by using the smart speakers. The benefits included not only enjoyable experiences to complete daily tasks but also bonds with the device from an affective dimension. The enjoyable user experiences and the strong bonds with the smart speaker enhanced the participants' positive attitudes toward their device, making them feel that occasional functional errors were acceptable.

The participants claimed that the smart speaker enabled them enjoyable experiences such as helping them access more information and enjoy more forms of entertainment. Besides, some said the speaker empowered them to multitask and saved time and energy. Yet, looking solely at the daily tasks completed by smart speakers, participants thought the smart speakers were not unreplaceable. As 1 participant indicated:

> *Apparently the smart speaker makes the completion of some daily tasks easier...but if only looking at the tasks it supports, I feel it doesn't have that much uniqueness.* [Female, age 69 years]

The same person, however, emphasized that her affective bonds with the smart speaker meant a lot to her.

> *...what I feel uniqueness is the companionship offered by the smart speaker...by using it, I feel as if someone is experiencing every day together with me, and taking care of me.* [Female, age 69 years]

Other participants also emphasized the importance of companionship offered by the smart speaker. Some stated that they started to develop bonds with the smart speaker because of its capability for task completion, but the bonds were strengthened and maintained through the companionship over time, driving them to maintain their use behaviors.

> *...My son bought the smart speaker for me, because my friend in the community said it was convenient to listen to music.... Gradually [the speaker] becomes a "person" accompanying me.... I feel peace of mind when I am reading a book while he is just beside me, in the room together with me.* [Male, age 67 years]

The functional errors that emerged during the technology use were spontaneously mentioned by our participants. However, people looked at the errors with charity. Some even treated these errors as fun. For example, 1 participant mentioned that sometimes the smart speaker may be activated by the television sound.

> *Once, the leading actor said "I don't love you," the smart speaker replied "That is heartbreaking" [laughing].... I like it, I think these so-called "functional errors" are acceptable, but sometimes are also fun.* [Male, age 67 years]

In this case, the smart speaker was not viewed as a pure device whose fluency in function operations was prioritized. Instead, the participant treated the smart speaker as a social agent as its inappropriate responses activated by television was unexpected social actions (ie, conversations between different devices).

Discussion

Principal Findings

The study aimed to investigate how smart speakers are integrated into the daily practices of older adults in the long term. The findings of our study illuminate a nuanced understanding of how older adults seamlessly integrate smart speakers into their lives, encompassing functional, spatial, cognitive, and semantic dimensions. This integration aligns with existing literature, shedding light on the multifaceted benefits and challenges associated with the adoption of smart speaker technology among older populations.

Integrate Multiple: Functional, Spatial, Cognitive, and Semantic Technology Integration

The functional integration of smart speakers among older adults is characterized by a rich diversity of use genres, reflecting the adaptability of these devices to meet various needs. Entertainment emerged as a dominant use genre, with participants expressing a universal affinity for using smart speakers to listen to music, transforming the device into a musical companion for activities like square dancing. This aligns with previous studies emphasizing the role of smart speakers in enhancing leisure activities for older adults [37]. Beyond entertainment, the device's pivotal role in information collection, medication reminders, and environment modification resonates with literature, highlighting smart speakers as valuable tools for health management and home automation among older adults [38]. The potential for emergency calls, while not frequently used, aligns with the findings of studies emphasizing the importance of safety features for older adults [39]. The acceptance of smart speakers as reliable companions for various tasks aligns with the idea that these devices can address specific needs and preferences, contributing to the overall well-being of older users [40].

Spatial integration revealed how older adults strategically positioned smart speakers to shape and organize their daily routines. The participants allocated their speakers across diverse living spaces, from the living room to the bedroom, study, bathroom, and even the balcony. This spatial organization facilitated the creation of specific use contexts, where the smart speaker seamlessly blended into existing daily practices. This personalized approach aligns with literature emphasizing the importance of tailoring technology to older adults' physical environments [41]. The integration of smart speakers with daily routines, such as pill-taking in the kitchen or controlling bedroom lighting, showcases the adaptability of older adults in incorporating technology seamlessly into their daily lives, emphasizing the active role older adults play in shaping their living environments to accommodate and optimize the technology functionality [42,43].

Cognitive integration illuminated the adaptive processes through which older adults familiarized themselves with the smart speaker's functionalities and overcame learning challenges. The initial hurdles in mastering voice commands and comprehending the device's full range of functions were common experiences. Yet, the participants' narratives highlight their resilience and problem-solving acumen, aligning with studies that stress the importance of user support and educational resources in facilitating older adults' technology adoption [44]. The gradual adjustments and incorporation of the smart speaker into daily habits reinforce the idea that cognitive integration is an ongoing process, transforming initial challenges into a journey of mental adaptation [45].

Semantic integration delved into the emotional and meaningful connections older adults forged with their smart speakers. Despite occasional functional errors, participants universally acknowledged and valued the benefits offered by these devices. The smart speaker's role in facilitating enjoyable experiences, multitasking, and saving time and energy contributed to positive attitudes among older adults. The findings speak against the common discourse that technical issues and occasional functional errors could significantly diminish the overall positive impact and perceived value of smart speakers among older adults [46]. Instead, participants not only accepted these imperfections with understanding but also found moments of humor and enjoyment in the device's occasional quirks. This resilience toward technical glitches underscores the robust emotional and meaningful connections formed between older adults and their smart speakers [24,47], highlighting that the perceived benefits and companionship offered by the technology far outweigh occasional operational hiccups.

To sum up, the integration of smart speakers among older adults is a holistic process that involves functional use, spatial organization, cognitive adaptation, and semantic connection with the device. Understanding the interplay of these dimensions provides a comprehensive insight into how smart speakers become integral components of older adults' daily lives, offering not only practical functionalities but also emotional fulfillment and companionship.

Design Implications

Being Aware of the Heterogeneity in Older Adults' Technology Use

In our study, older adults used the smart speaker in different physical spaces, through different actions, and for different purposes. The underlying reasons could be the diversity of personalities [33], the differences in sociocultural and socioeconomic situations [48], and the disparities in individual digital literacy [49]. The heterogeneity in older adults' technology use may pose challenges for designers and developers to transform ideas and insights into concrete designs. Merely focusing on a specific issue met by older adults would limit the benefits of smart speakers to older adults. The complex entanglement of space, practices, and user needs that may affect technology use at different sites should be carefully considered. We thus suggest designers and developers to zoom out and get the big picture of older adults' daily practices before zooming in to define the design problem to solve.

Serving for the Existing Activities in Older Adults' Home Environments

Despite the discourse that smart speakers can support older adults to explore unexperienced activities, in our study, participants used the device to enhance the quality of their existing activities. For instance, the smart speaker supports medication reminding, turning on or off the lights, and music listening. Consequently, we suggest designers and developers to pay more attention to the issues or opportunities related to older adults' existing activities at home, rather than new activities that are not commonly experienced among this group of population.

Attaching Importance to Additional Devices Centered on the Smart Speaker

Our findings show that some additional devices connecting to the smart speakers were used by older adults. For instance, a "white button," designed for emergency calls was carried by a

participant all the time. The solution of connecting smart speakers to light switches in case of harm in darkness was also acknowledged. The 2 cases suggest that additional devices, especially the ones ensuring personal safety, have a market in the aging groups. Because many older adults have difficulties using the apps on mobile phones [3], we argue more additional devices centered on the smart speaker are needed.

Leaving Space for the Agency of Older Adults

Older adults are usually treated as inexpert technology users, and judged to be risky to do technology appropriation [50]. Yet, our findings show how participants spatially organized the smart speaker, and actively defined, changed, and reshaped their practices after using the device. In our cases, older adults created solutions that suit their specific needs [51]. These solutions were not predefined or preformatted by someone else, but specifically for older adults themselves [52]. We believe this may contribute to developing strong affective bonds between older adults and the device, driving their maintainable use behaviors, and promoting their resilient attitudes toward functional errors. Hence, we call for more studies about the association between older adults' agency in technology use and their affective bonds with technology. We also encourage designers and developers to carefully consider the space for the agency of older adults in terms of technology appropriation.

Getting to Know the Preferences of Communities

The most commonly mentioned genre of using the smart speaker was listening to music in our study. This is understandable as many participants are from a square dancing community. It shows that the community where older adults are involved can implicitly manifest their preferences toward the purposes of technology use. Therefore, we encourage designers and developers to shift partial attention from individuals to communities when investigating the preferences of older adults.

As smart speakers swiftly reach the mainstream, understanding the detailed, nuanced experiences of older adults using these devices becomes paramount. Our research delves into the intricacies of daily use. By unraveling the complexities and diverse experiences during older adults' long-term engagement with smart speaker technology, we aim to bridge the gap between technology and aging, offering insights that go beyond design considerations to impact policies and interventions. Moreover, our study seeks to contextualize the significance of understanding older adults' experiences within the broader sociotechnological landscape. The potential benefits of these devices in health care, chronic disease management, and emergency identification align with broader societal challenges in health care accessibility and aging populations. In an age where digital advancements redefine societal norms, the integration of smart speakers into the lives of older adults is not just a technological trend; it is a pivotal factor in ensuring their inclusion in the digital revolution. Consequently, this research does not merely contribute to the attractiveness of smart speakers for older users but serves as a critical voice in the ongoing dialogue about healthy aging in the digital era.

Conclusions

This paper focuses on the integration of smart speakers into older adults' daily practices. We aim to understand how older adults routinely use smart speakers and integrate them into their daily lives, which in turn enables us to propose suggestions for designers and developers. A total of 20 older adults who have used a smart speaker for over 6 months were interviewed. The findings demonstrate that older adults integrate smart speakers into daily use through functional integration, spatial integration, cognitive integration, and semantic integration. Based on the findings, we proposed several suggestions for designers and developers to better design smart speakers that promote maintainable use behaviors by older adults.

Limitations

Our findings must be evaluated within the context of several limitations. First, participants were recruited from 2 communities, potentially impacting the generalizability of the results. The risk of selection bias or unaccounted factors, such as the homogeneity of participant characteristics within the same facility, may have influenced interview responses. Yet, the study indeed highlighted the intricate nature of long-term technology use among older adults in these specific communities, offering valuable insights for designers and developers in the context of smart speaker design. Second, our study only investigated the long-term use of smart speakers, without paying much attention to the discontinuation of using smart speakers in the long term. The investigation of the discontinued use of smart speakers may reveal more insights into older adults' technology use. However, we believe that it is important to examine the integration of smart speakers in older adults' daily practices, as the findings may contribute to the maintainability of older adults' technology use.

Acknowledgments

The authors would like to thank the participants in the study for sharing their experiences and perceptions. We also thank the Ageing Dancing community and Bright community for making the study possible. This research was funded by a Chinese postdoctoral grant (2023M732231) and a Design School Innovation grant (WH102243005/201/004).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Artificial Intelligence Literacy Questionnaire (adapted).

[DOCX File , 12 KB - mhealth_v12i1e47472_app1.docx]

Multimedia Appendix 2
Interview schedule.
[DOCX File , 14 KB - mhealth_v12i1e47472_app2.docx]

References

1. Felix R. Smart speaker adoption continues to rise. Statista. 2020. URL: https://www.statista.com/chart/16597/smart-speaker-ownership-in-the-united-states/ [accessed 2024-01-30]
2. Motta I, Quaresma M. Understanding task differences to leverage the usability and adoption of Voice Assistants (VAs). In: Design, User Experience, and Usability: Design for Contemporary Technological Environments. 2021 Presented at: 10th International Conference, DUXU 2021, Held as Part of the 23rd HCI International Conference, HCII 2021; July 24–29, 2021; Virtual p. 483-502. [doi: 10.1007/978-3-030-78227-6_35]
3. Portet F, Vacher M, Golanski C, Roux C, Meillon B. Design and evaluation of a smart home voice interface for the elderly: acceptability and objection aspects. Pers Ubiquit Comput 2011;17(1):127-144. [doi: 10.1007/s00779-011-0470-5]
4. Brause SR, Blank G. Externalized domestication: smart speaker assistants, networks and domestication theory. Inf Commun Soc 2020;23(5):751-763 [FREE Full text] [doi: 10.1080/1369118x.2020.1713845]
5. Lopatovska I, Rink K, Knight I, Raines K, Cosenza K, Williams H, et al. Talk to me: exploring user interactions with the Amazon Alexa. J Librariansh Inf Sci 2018;51(4):984-997 [FREE Full text] [doi: 10.1177/0961000618759414]
6. Kim J, Shin E, Han K, Park S, Youn JH, Jin G, et al. Efficacy of smart speaker-based metamemory training in older adults: case-control cohort study. J Med Internet Res 2021;23(2):e20177 [FREE Full text] [doi: 10.2196/20177] [Medline: 33591276]
7. Sunshine J. Smart speakers: the next frontier in mHealth. JMIR Mhealth Uhealth 2022;10(2):e28686 [FREE Full text] [doi: 10.2196/28686] [Medline: 35188467]
8. Robinson S, Pearson J, Ahire S, Ahirwar R, Bhikne B, Maravi N, et al. Revisiting “hole in the wall” computing: private smart speakers and public slum settings. 2018 Presented at: CHI '18: CHI Conference on Human Factors in Computing Systems; April 21-26, 2018; Montreal, QC. [doi: 10.1145/3173574.3174072]
9. Pearson J, Robinson S, Reitmaier T, Jones M, Ahire S, Joshi A, et al. StreetWise: smart speakers vs human help in public slum settings. 2019 Presented at: CHI '19: CHI Conference on Human Factors in Computing Systems; May 4-9, 2019; Glasgow, UK. [doi: 10.1145/3290605.3300326]
10. Abdolrahmani A, Kuber R, Branham SM. "Siri talks at you": an empirical investigation of Voice-Activated Personal Assistant (VAPA) usage by individuals who are blind. 2018 Presented at: ASSETS '18: The 20th International ACM SIGACCESS Conference on Computers and Accessibility; October 22-24, 2018; Galway Ireland p. 249-258. [doi: 10.1145/3234695.3236344]
11. Abdolrahmani A, Storer KM, Roy ARM, Kuber R, Branham SM. Blind leading the sighted: drawing design insights from blind users towards more productivity-oriented voice interfaces. ACM Trans Access Comput 2020;12(4):1-35 [FREE Full text] [doi: 10.1145/3368426]
12. Wald R, Piotrowski JT, Araujo T, van Oosten JMF. Virtual assistants in the family home. Understanding parents’ motivations to use virtual assistants with their child(dren). Comput Human Behav 2023;139:107526 [FREE Full text] [doi: 10.1016/j.chb.2022.107526]
13. Garg R, Sengupta S. He is just like me: a study of the long-term use of smart speakers by parents and children. Proc ACM Interact Mob Wearable Ubiquitous Technol 2020;4(1):1-24 [FREE Full text] [doi: 10.1145/3381002]
14. Garg R, Cui H, Seligson S, Zhangqq B, Porcheron M, Clark L, et al. The last decade of HCI research on children and voice-based conversational agents. 2022 Presented at: CHI '22: CHI Conference on Human Factors in Computing Systems; 29 April-5 May, 2022; New Orleans, LA URL: https://dl.acm.org/doi/10.1145/3491102.3502016 [doi: 10.1145/3491102.3502016]
15. McTear M, Callejas Z, Griol D. The Conversational Interface: Talking to Smart Devices, Volume 6. Cham, Switzerland: Springer International Publishing; 2016.
16. Vtyurina A, Fourney A. Exploring the role of conversational cues in guided task support with virtual assistants. 2018 Presented at: CHI '18: CHI Conference on Human Factors in Computing Systems; April 21-26, 2018; Montreal, QC. [doi: 10.1145/3173574.3173782]
17. Lau J, Zimmerman B, Schaub F. Alexa, are you listening?: Privacy perceptions, concerns and privacy-seeking behaviors with smart speakers. Proc ACM Hum-Comput Interact 2018;2(CSCW):1-31 [FREE Full text] [doi: 10.1145/3274371]
18. Festerling J, Siraj I. Anthropomorphizing technology: a conceptual review of anthropomorphism research and how it relates to children's engagements with digital voice assistants. Integr Psychol Behav Sci 2022;56(3):709-738 [FREE Full text] [doi: 10.1007/s12124-021-09668-y] [Medline: 34811705]
19. Pradhan A, Lazar A, Findlater L. Use of intelligent voice assistants by older adults with low technology use. ACM Trans Comput-Hum Interact 2020;27(4):1-27 [FREE Full text] [doi: 10.1145/3373759]

20. Trajkova M, Martin-Hammond A. "Alexa is a toy": exploring older adults' reasons for using, limiting, and abandoning echo. 2020 Presented at: CHI '20: CHI Conference on Human Factors in Computing Systems; April 25-30, 2020; Honolulu, HI. [doi: 10.1145/3313831.3376760]
21. Brewer R, Pierce C, Upadhyay P, Park L. An empirical study of older adult's voice assistant use for health information seeking. ACM Trans Interact Intell Syst 2022;12(2):1-32 [FREE Full text] [doi: 10.1145/3484507]
22. Pradhan A, Findlater L, Lazar A. "Phantom friend" or "just a box with information": personification and ontological categorization of smart speaker-based voice assistants by older adults. Proc ACM Hum Comput Interact 2019;3(CSCW):1-21 [FREE Full text] [doi: 10.1145/3359316]
23. Park S, Kim B. The impact of everyday AI-based smart speaker use on the well-being of older adults living alone. Technol Soc 2022;71:102133 [FREE Full text] [doi: 10.1016/j.techsoc.2022.102133]
24. Kim S, Choudhury A. Exploring older adults' perception and use of smart speaker-based voice assistants: a longitudinal study. Comput Human Behav 2021;124(3):106914 [FREE Full text] [doi: 10.1016/j.chb.2021.106914]
25. McCloud R, Perez C, Bekalu MA, Viswanath K. Using smart speaker technology for health and well-being in an older adult population: pre-post feasibility study. JMIR Aging 2022;5(2):e33498 [FREE Full text] [doi: 10.2196/33498] [Medline: 35532979]
26. Bhattacherjee A, Premkumar G. Understanding changes in belief and attitude toward information technology usage: a theoretical model and longitudinal test. MIS Quarterly 2004;28(2):229-254 [FREE Full text] [doi: 10.2307/25148634]
27. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77-101 [FREE Full text] [doi: 10.1191/1478088706qp063oa]
28. Crouch M, McKenzie H. The logic of small samples in interview-based qualitative research. Soc Sci Inf 2016;45(4):483-499 [FREE Full text] [doi: 10.1177/0539018406069584]
29. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007;19(6):349-357 [FREE Full text] [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
30. Guest G, Bunce A, Johnson L. How many interviews are enough?: An experiment with data saturation and variability. Field methods 2016;18(1):59-82 [FREE Full text] [doi: 10.1177/1525822X05279903]
31. Hennink M, Kaiser BN. Sample sizes for saturation in qualitative research: a systematic review of empirical tests. Soc Sci Med 2022;292:114523 [FREE Full text] [doi: 10.1016/j.socscimed.2021.114523] [Medline: 34785096]
32. Jansons P, Fyfe J, Via JD, Daly RM, Gvozdenko E, Scott D. Barriers and enablers for older adults participating in a home-based pragmatic exercise program delivered and monitored by Amazon Alexa: a qualitative study. BMC Geriatr 2022;22(1):248 [FREE Full text] [doi: 10.1186/s12877-022-02963-2] [Medline: 35337284]
33. Wang B, Rau PLP, Yuan T. Measuring user competence in using artificial intelligence: validity and reliability of artificial intelligence literacy scale. Behav Inf Technol 2022;42(9):1324-1337 [FREE Full text] [doi: 10.1080/0144929x.2022.2072768]
34. Graneheim UH, Lindgren BM, Lundman B. Methodological challenges in qualitative content analysis: a discussion paper. Nurse Educ Today 2017;56:29-34 [FREE Full text] [doi: 10.1016/j.nedt.2017.06.002] [Medline: 28651100]
35. Fusch PI, Ness LR. Are we there yet? Data saturation in qualitative research. Walden Faculty Staff Publ 2015;20(9):1408-1416 [FREE Full text] [doi: 10.46743/2160-3715/2015.2281]
36. Carter N, Bryant-Lukosius D, DiCenso A, Blythe J, Neville AJ. The use of triangulation in qualitative research. Oncol Nurs Forum 2014;41(5):545-547 [FREE Full text] [doi: 10.1188/14.ONF.545-547] [Medline: 25158659]
37. Chung J, Gendron T, Winship J, Wood RE, Mansion N, Parsons P, et al. Smart speaker and ICT use in relationship with social connectedness during the pandemic: loneliness and social isolation found in older adults in low-income housing. Gerontologist 2023:gnad145 [FREE Full text] [doi: 10.1093/geront/gnad145] [Medline: 37880825]
38. Kim S. Exploring how older adults use a smart speaker-based voice assistant in their first interactions: qualitative study. JMIR Mhealth Uhealth 2021;9(1):e20427 [FREE Full text] [doi: 10.2196/20427] [Medline: 33439130]
39. Wei J, Dingler T, Kostakos V. Understanding user perceptions of proactive smart speakers. Proc ACM Interact Mob Wearable Ubiquitous Technol 2021;5(4):1-28 [FREE Full text] [doi: 10.1145/3494965]
40. Arnold A, Kolody S, Comeau A, Cruz AM. What does the literature say about the use of personal voice assistants in older adults? A scoping review. Disabil Rehabil Assist Technol 2024;19(1):100-111 [FREE Full text] [doi: 10.1080/17483107.2022.2065369] [Medline: 35459429]
41. Oudshoorn N. How places matter: telecare technologies and the changing spatial dimensions of healthcare. Soc Stud Sci 2012;42(1):121-142 [FREE Full text] [doi: 10.1177/0306312711431817] [Medline: 22530385]
42. Fischer B, Peine A, Östlund B. The importance of user involvement: a systematic review of involving older users in technology design. Gerontologist 2020;60(7):e513-e523 [FREE Full text] [doi: 10.1093/geront/gnz163] [Medline: 31773145]
43. Peine A, Rollwagen I, Neven L. The rise of the "innosumer"—rethinking older technology users. Technol Forecast Soc Change 2014;82:199-214 [FREE Full text] [doi: 10.1016/j.techfore.2013.06.013]
44. Portz JD, Bayliss EA, Bull S, Boxer RS, Bekelman DB, Gleason K, et al. Using the technology acceptance model to explore user experience, intent to use, and use behavior of a patient portal among older adults with multiple chronic conditions: descriptive qualitative study. J Med Internet Res 2019;21(4):e11604 [FREE Full text] [doi: 10.2196/11604] [Medline: 30958272]

45. Chang F, Östlund B, Kuoppamäki S. Domesticating social alarm systems in nursing homes: qualitative study of differences in the perspectives of assistant nurses. J Med Internet Res 2023;25:e44692 [FREE Full text] [doi: 10.2196/44692] [Medline: 37145835]
46. Peine A, Neven L. From intervention to co-constitution: new directions in theorizing about aging and technology. Gerontologist 2019;59(1):15-21 [FREE Full text] [doi: 10.1093/geront/gny050] [Medline: 29850812]
47. Bentley F, Luvogt C, Silverman M, Wirasinghe R, White B, Lottridge D. Understanding the long-term use of smart speaker assistants. Proc ACM Interact Mob Wearable Ubiquitous Technol 2018;2(3):1-24 [FREE Full text] [doi: 10.1145/3264901]
48. Ware P, Bartlett SJ, Paré G, Symeonidis I, Tannenbaum C, Bartlett G, et al. Using eHealth technologies: interests, preferences, and concerns of older adults. Interact J Med Res 2017;6(1):e3 [FREE Full text] [doi: 10.2196/ijmr.4447] [Medline: 28336506]
49. Arreola I, Morris Z, Francisco M, Connelly K, Caine K, White G. From checking on to checking in: designing for low socio-economic status older adults. 2014 Presented at: CHI '14: CHI Conference on Human Factors in Computing Systems; 26 April-1 May 2014; Toronto, ON p. 1933-1936. [doi: 10.1145/2556288.2557084]
50. Waycott J, Vetere F, Ozanne E. Building social connections: a framework for enriching older adults' social connectedness through information and communication technologies. In: Neves BB, Vetere F, editors. Ageing and Digital Technology: Designing and Evaluating Emerging Technologies for Older Adults. Singapore: Springer Nature Singapore; 2019:65-82.
51. Giaccardi E, Kuijer L, Neven L. Design for resourceful ageing : intervening in the ethics of gerontechnology. 2016 Presented at: 2016 Design Research Society Conference (DRS 2016); June 27-30, 2016; Brighton, UK URL: https://research.tue.nl/en/publications/design-for-resourceful-ageing-intervening-in-the-ethics-of-geront [doi: 10.21606/drs.2016.258]
52. Pirhonen J, Lolich L, Tuominen K, Jolanki O, Timonen V. "These devices have not been made for older people's needs" – older adults' perceptions of digital technologies in Finland and Ireland. Technol Soc 2020;62:101287 [FREE Full text] [doi: 10.1016/j.techsoc.2020.101287]

Abbreviations

COREQ: Consolidated Criteria for Reporting Qualitative Research

Edited by L Buis; submitted 23.03.23; peer-reviewed by R Meng, J Farmer; comments to author 06.07.23; revised version received 22.11.23; accepted 24.01.24; published 12.02.24.

Please cite as:
Chang F, Sheng L, Gu Z
Investigating the Integration and the Long-Term Use of Smart Speakers in Older Adults' Daily Practices: Qualitative Study
JMIR Mhealth Uhealth 2024;12:e47472
URL: https://mhealth.jmir.org/2024/1/e47472
doi:10.2196/47472
PMID:38345844

Original Paper

Doctors' Personal Preference and Adoption of Mobile Apps to Communicate with Patients in China: Qualitative Study

Dongjin Chen[1], PhD; Wenchao Han[2], MA; Yili Yang[3], MSc; Jay Pan[4], PhD

[1]Institute for Social Governance and Communication Innovation of Zhejiang, Communication University of Zhejiang, Hangzhou, China

[2]Center for Asia-Europe Study, Xi'an Jiaotong University, Xi'an, China

[3]HEOA Group, Institute for Healthy Cities and West China Research Center for Rural Health Development, Sichuan University, Chengdu, China

[4]West China School of Public Health and West China Fourth Hospital, Sichuan University, Chengdu, China

Corresponding Author:
Jay Pan, PhD
West China School of Public Health and West China Fourth Hospital
Sichuan University
No. 16, Section 3, Ren Min Nan Road
Chengdu, 610041
China
Phone: 86 28 85501096
Email: panjie.jay@scu.edu.cn

Abstract

Background: Different kinds of mobile apps are used to promote communications between patients and doctors. Studies have investigated patients' mobile app adoption behavior; however, they offer limited insights into doctors' personal preferences among a variety of choices of mobile apps.

Objective: This study aimed to investigate the nuanced adoption behaviors among doctors in China, which has a robust adoption of mobile apps in health care, and to explore the constraints influencing their selection of specific mobile apps. This paper addressed 3 research questions: (1) Which doctors opt to adopt mobile apps to communicate with patients? (2) What types of mobile apps do they choose? (3) To what degree do they exercise personal choice in adopting specific mobile apps?

Methods: We used thematic content analysis of qualitative data gathered from semistructured interviews with 11 doctors in Hangzhou, which has been recognized for its advanced adoption of mobile technology in social services, including health care services. The selection of participants was purposive, encompassing diverse departments and hospitals.

Results: In total, 5 themes emerged from the data analysis. First, the interviewees had a variety of options for communicating with patients via mobile apps, with the predominant ones being social networking apps (eg, WeChat) and medical platforms (eg, Haodf). Second, all interviewees used WeChat to facilitate communication with patients, although their willingness to share personal accounts varied (they are more likely to share with trusty intermediaries). Third, fewer than half of the doctors adopted medical platforms, and they were all from tertiary hospitals. Fourth, the preferences for in-person, WeChat, or medical platform communication reflected the interviewees' perceptions of different patient cohorts. Lastly, the selection of a particular kind of mobile app was significantly influenced by the doctors' affiliation with hospitals, driven by their professional obligations to fulfill multiple tasks assigned by the hospitals or the necessity of maintaining social connections with their colleagues.

Conclusions: Our findings contribute to a nuanced understanding of doctors' adoption behavior regarding specific types of mobile apps for patient communication, instead of addressing such adoption behavior of a wide range of mobile apps as equal. Their choices of a particular kind of app were positioned within a social context where health care policies (eg, limited funding for public hospitals, dominance of public health care institutions, and absence of robust referral systems) and traditional culture (eg, trust based on social connections) largely shape their behavioral patterns.

(JMIR Mhealth Uhealth 2024;12:e49040) doi:10.2196/49040

KEYWORDS

medical platforms; doctor-patient communication; social networking apps; thematic content analysis; China

Introduction

The widespread usage of smartphones, embraced by over 6 billion individuals, has facilitated the proliferation of mobile apps among users worldwide. Within this landscape, the integration of mobile apps to facilitate online medical services has emerged as one of the most rapidly evolving sectors in the health care industry. The term "mobile health", or "mHealth," is defined by the World Health Organization (WHO) as "medical and public health practice supported by mobile devices, such as mobile phones, patient monitoring devices, Personal Digital Assistants (PDAs), and other wireless devices" [1]. In addition to mHealth, eHealth [2], telehealth [3], and other newly emerged terms have also been used to describe the dynamic evolution of contemporary online health services. As demonstrated by numerous recent studies, the adoption of mobile apps in health services has immense potential to enhance doctor-patient communications, primarily through 2 types of mobile apps: online medical platforms and social networking apps. In this context, exploring the functioning mechanisms of these mobile apps has become imperative for identifying barriers to mHealth penetration, thereby offering insights to inform the targeted expansion of specific app types to improve the quality and efficiency of health care delivery procedures.

In China, more than 1 billion people used mobile phones in 2022 [4]. As the stressful relationship between health care providers and patients persists as a longstanding issue in the context of China's health care system [5], the adoption of mobile apps presents a promising strategy to mitigate this tension to some extent [6]. Recent studies have highlighted the benefits conferred by mobile apps, including promoting health care delivery procedures regardless of geographical barriers [7], contributing to the alleviation of urban-rural health care disparities by offering online medical consultation to those living in remote areas [8], and expanding the range of service areas for health care professionals who provide quality specialized care via mobile connections [9]. As such, the adoption of mobile apps proves to be a vital strategy for advancing both health care delivery and the doctor-patient relationship.

The adoption of mobile apps has brought numerous benefits, yet scholars have raised concerns regarding various issues encountered during implementation. Among these concerns, protecting patients' online personal information stands out as a significant security challenge [10]. In addition, health care professionals' use of social media apps might lead to an identity crisis, blurring the boundaries between their personal and professional lives [11]. Most studies have emphasized individual-level factors as major determinants of personal preference, yielding diverse findings on aspects such as social and economic returns [12], extrinsic and intrinsic motivations [13], their disciplines [7], their professional titles [14], anticipated rewards [15], and monetary incentives [13].

Despite different findings and conclusions, these studies were based on an implicit assumption that these doctors have full control of their choices, which means that doctors have the freedom to decide how to spend their spare time at their own will. However, in China, this assumption warrants scrutiny due to the well-known issue of excessive workload faced by health care professionals, leaving them with limited leisure time after extended hours [16].

Previous studies have identified the level of doctors' affiliated health care organizations as a potential factor that affects the adoption of web-based health information technologies. For example, Li et al [14] found that quality doctors from tertiary hospitals exhibit greater activity on web-based medical consultation platforms. However, the underlying mechanism of this influence requires further explanation, not to mention that some macrolevel influences remain to be explored, such as the impact of national health care policies. These influences are beyond doctors' personal choices. Nevertheless, a few exceptions from the current literature managed to analyze meso- and macrolevel factors, in addition to individual-level factors. Peng et al [17] applied social ecosystem theory to understand the determinants of physicians' online uptake and regarded their perceptions of law and regulation of online health as macrolevel factors. However, these meso- and macrolevel factors proposed for analysis primarily pertain to online health care services, thus neglecting the potential impacts of the particular contexts. Other scholars have conducted review studies describing the current status of China's mHealth industry, with emphasis of the overall impacts of the health care system, but have not provided details of the underlying mechanism [6].

To address the existing gap in the current literature, this study aimed to address the nuanced mobile app adoption behaviors of doctors in China and explore the factors influencing their choice of specific mobile apps, rather than merely examining general adoption behavior. We conducted thematic analysis of semistructured interviews with selected doctors in Hangzhou, focusing on these 3 questions: (1) Which doctors opt to adopt mobile apps to communicate with patients? (2) What type of mobile apps do they adopt? (3) To what degree do they exercise personal choice in adopting specific mobile apps?

Methods

Research Design

The inductive approach was used to facilitate exploratory research, and interviews were conducted to understand interviewees' inner thoughts about the issue being investigated [18]. Thematic content analysis was then used to uncover subtle concerns underlying their adoption behavior. This qualitative method has been used to explore health care services provision from the supply side in order to understand health care professionals' perception of internet hospitals [19], as well as their use of mHealth in practice to deliver health care services [20]. Thus, the self-reported narratives from interviewees provided direct responses to the set of research questions we previously proposed.

Study Setting

We conducted our study in Hangzhou, the provincial capital of Zhejiang Province. Hangzhou City is situated on the east coast of China, approximately a 2-hours ride from Shanghai. We selected this city as the research setting mainly out of 3

considerations. First, with a population of 12.20 million and a local gross domestic product (GDP) per capita of Renminbi (RMB) 149,900 (~US $21,122) [21], significantly higher than the national average of RMB 80,000 (~US $11,273) in 2021 [22], the city has a relatively affluent population. This means local residents are more likely to use their out-of-pocket money to pay the fee charged by medical platforms. Second, this city is renowned for its technical advancement and has attracted a lot of mobile app companies (eg, Taobao, the biggest online shopping company in China) to build their headquarters there. Third, this city is home to 3 hospitals ranked among China's top 50 hospitals, where highly skilled health care professionals are more likely to embrace advanced health information technology [23]. Therefore, this city presents an ideal research environment for investigating the current landscape of mobile app adoption in China's clinical settings and exploring the issues and concerns perceived by health care professionals during their practical interactions with mobile apps.

Data Collection

We conducted purposeful sampling to select participants for in-depth interviews. Given that investigating doctors' personal choice of mobile apps might involve disclosure of their or their patients' private information, securing cooperation from the doctors proved challenging.

We selected participants using the following criteria: First, they were under 50 years old, as older individuals are less likely to adopt new technology. Second, we tried to cover different levels of medical agencies, including local health community clinics and tertiary hospitals. Considering that doctors working in tertiary hospitals are more likely to use mobile apps to provide online medical consultation services [14], more than half of the interviewees were selected from tertiary hospitals. Third, we tried to cover doctors specializing in different areas, as well as with different professional titles. Fourth, as most medical agencies in China are public agencies, we selected our interviewees primarily from public hospitals.

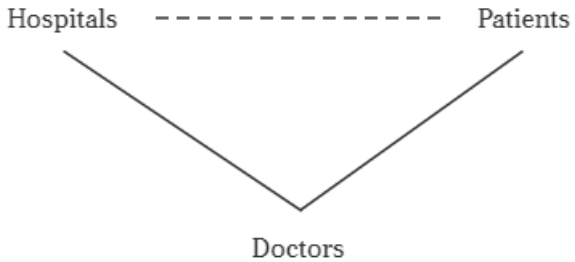
Interviews

Prior to interviews, the researchers sent the topic list to the interviewees so that they could prepare for the interviews ahead. All 11 interviews occurred within hospital settings, with the interviewers maintaining impartiality. Four of the researchers conducted the interviews, two of whom (authors DC and JP) possessed previous research experience with interviewees, while the other two (authors WH and YY) received training in interview skills. Field notes were taken during interviews, and audio recordings were made with the interviewees' consent. Each interview commenced with a brief introduction of the interviewer's research background and the research project. The interviewees were given the option to decline participation, although nobody dropped out. We conducted the interviews from June to December 2022, and each interview lasted for approximately 30-60 minutes. Interviewees received RMB 200-400 (~US $27-55), depending on the interview duration. Data saturation was achieved after interviewing 11 doctors.

Primary Model and Interview Questions

Different from previous studies focusing on a straight relationship between doctors and patients, we added another bilateral relationship between doctors and hospitals to our analyses, forming a basic framework to approach this research. Current studies have underscored the interconnectedness between offline and online health care services [15,24]. It is necessary to investigate institutional impacts posed by their affiliated hospitals, which are also the employers of their offline practice. We devised a triangular model to understand the dynamics between doctors, patients, and hospitals, as depicted in Figure 1. This model delineates 3 sets of dual relationships: doctor-patient, doctor-hospital, and patient-hospital. We constructed our interview questions based on this model and collected interviewees' responses to each question accordingly. The interview questions are listed in Multimedia Appendix 1.

Figure 1. The relationship triangle between doctors, patients, and hospitals.



Data Analysis

Two researchers (DC and WH) transcribed all the recorded files and reviewed them for accuracy. According to Braun and Clarke [25], both manual coding and software coding are applicable for thematic analysis. Drawing from the interviewees' spoken words, we used manual open coding to identify potential themes. The analysis included both inductive and deductive coding [26]. The primary model outlined in Figure 1 guided deductive coding, while transcribed data yielded unanticipated themes, necessitating inductive coding. The two researchers (DC and WH) read through the interview transcripts and applied both inductive coding and deductive coding to generate an individual code book independently. Next, they cross-referenced their respective code book, and any discrepancies were resolved through discussion with a third researcher (JP), ultimately reaching consensus on overarching themes. Following 11 interviews, no new themes emerged and saturation was reached. The researchers adhered to the COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist to report the empirical research process [27].

Ethical Considerations

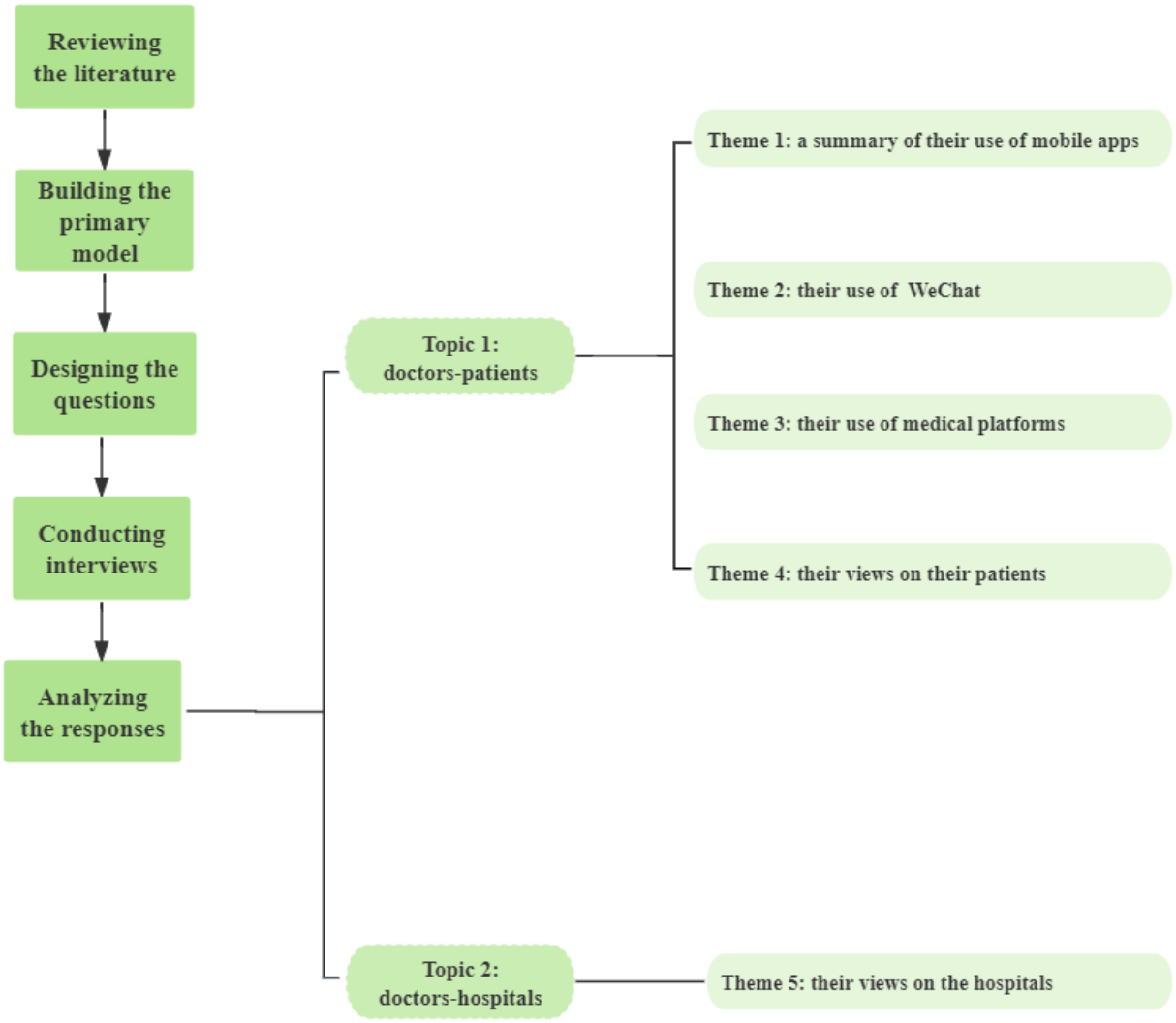
Ethical approval for this research was obtained from News and Communication Research Institute, Communication University of Zhejiang. Prior to interviews, the researchers assured the interviewees of the anonymization of their personal information, and all interviewees provided written informed consent.

Results

Research Process

Each step of the research process is presented in Figure 2. We extracted 5 themes from the data analysis, including a summary of the interviewees' mobile app usage, the specific use of 2 primary kinds of mobile apps, and the interviewees' views on patients and affiliated hospitals.

Figure 2. Stages of the research process.



Summary of Interviewees' Use of Apps

We interviewed a total of 11 participants, 9 (82%) of whom were from tertiary hospitals, with 2 (18%) from top-tier hospitals in China, ranked among the top 50 hospitals. Table 1 delineates the key characteristics of each interviewee. We added 1 category to the table to describe the varying degrees of difficulty in making an appointment with these doctors. Specifically, "very difficult" means that a waiting period of approximately 1 month is necessary for potential patients to compete for a possible slot either over the phone or through the online registration system, while "difficult" suggests a shorter waiting time of 1-2 weeks. For the remaining doctors, no appointment is required in advance, as they provide readily accessible services to patients on their clinical days.

Table 1. Characteristics of interviewees (N=11) collected from June to December 2022.

Characteristics	Interviewees, n (%)
Gender	
Male	8 (73)
Female	3 (27)
Age (years)	
30-40	3 (27)
40-50	8 (73)
Professional title	
Chief physician	2 (18)
Associate chief physician	5 (45)
Attending doctor	4 (36)
Type of affiliated hospital	
Public tertiary hospital	9 (82)
Community clinic	1 (9)
Private clinic	1 (9)
Difficulty in making an appointment	
Very difficult	3 (28)
Difficult	4 (36)
Not difficult	4 (36)
Department director or not	
Department director	3 (27)
Not department director	8 (73)
Affiliated department	
Surgery	2 (18)
Internal medicine	2 (18)
Traditional Chinese medicine	4 (36)
Gynecology	1 (9)
Dentistry	2 (18)

According to the interviewees, they used a variety of mobile apps for patient communication, with the predominant ones being social networking apps (eg, WeChat) and medical platforms (eg, Haodf). All interviewees used WeChat to communicate with patients, either through 1-to-1 chats or group chats, although their willingness to share their personal WeChat accounts with patients varied. Only 4 (36%) interviewees reported using medical platforms at the time of the interviews, with detailed use discussed in later sections.

In terms of communication style (the number of people engaged in the communication), 1-to-1 chats via WeChat involve 2 people. WeChat-facilitated group chats entail the participation of multiple individuals, commonly including 1 or several doctors, alongside a group of patients. These discussions typically focus on medical consultation for specific conditions or the dissemination of supportive information, such as guidance on appointment scheduling. The medical platforms enable doctors to have 1-to-1 conversations with anonymous patients, and these conversations or the feedback on them can be further reviewed by other end users whenever they access the same medical platforms. Meanwhile, patients' medical examination results and their identification information remain confidential.

The Use of WeChat

WeChat is the most widely embraced social media app in China. In addition to text-based messages, it supports multimedia communication through the sharing of images, recordings, and videos, alongside the integration of numerous miniapps. Table 2 presents a summary of the interviewees' use of WeChat to communicate with patients or potential patients. Notably, 1-to-1 chats were the most personalized and confidential mode of communication, involving interactions between 2 individuals.

Table 2. Interviewees' use of specific WeChat functions to communicate with patients from June to December 2022.

Interviewee ID	Specific WeChat function			
	One-to-one chat[a]	Group chat	WeChat Moments (friend circle)	WeChat short-term video
1	Likely	—[b]	Yes	—
2	Rare	—	—	—
3	Very likely	—	Yes	Yes
4	Very likely	Yes	Yes	Yes
5	Rare	—	—	—
6	Very likely	Yes	—	—
7	Likely	—	—	—
8	Rare	—	Yes	Yes
9	Rare	—	Yes	—
10	Likely	Yes	—	—
11	Very likely	—	Yes	—

[a]The frequency of 1-to-1 chats was measured based on the average number of patients with whom these interviewees had WeChat conversations in 1 week: <5 patients measured as "rare"; 6-8 patients as "likely"; ≥9 patients as "very likely".

[b]Not applicable.

Of the 11 doctors, 4 (36%) exhibited a propensity to share their personal WeChat accounts with patients very likely; these doctors were all from low-risk departments, including internal medicine, dentistry, or traditional Chinese medicine. In contrast, 4 (36%) doctors rarely did so due to safety concerns with potential patients.

> *Yes, I will share my WeChat account with patients. After I joined this private dental clinic, my patients were first allocated by the clinic. Later, as these patients found satisfaction with my services, they began to introduce my account to other people, and I have more and more patients.* [Interviewee 3]

Additionally, a few doctors adopted a discerning approach, depending on initial interactions with newly admitted patients, regarding assessments of medical compliance as a determining factor.

> *I will make an evaluation of patients when deciding to share my WeChat account with them. I select those patients who are reasonable and cooperative, with a good sense of medical compliance. I am very cautious against those picky patients. Usually, doctors do not have obligations to share their personal account with patients. We do so only for the sake of their conveniences.* [Interviewee 10]

Furthermore, all interviewees expressed a willingness to share their personal accounts with those introduced by well-acquainted intermediaries.

> *I will not share my WeChat account with patients. The rare exceptions are for some special patients, depending on whether patients' parents are easy to communicate with and their diseases need long-time follow-up. In addition, I will share my account with those who are introduced by my familiar colleagues, friends, and relatives. China is a society replete with social connections, and I cannot avoid these connections.* [Interviewee 8]

Group discussions are another frequently used communication form on the WeChat platform, especially for patients with chronic diseases requiring long-term care and occasional relapse. Some WeChat groups can be created by institutions, hospitals, or departments. One local clinic (the employer of interviewee 10) created WeChat groups including 2 or 3 family doctors to meet the needs of local residents. Alternatively, interviewees may initiate their own groups. Here is an interviewee's response.

> *In my WeChat account, I have ten WeChat patient groups. Each group has 500 patients. Those initially added to the group invited their friends and relatives, and the number of each WeChat group reached the maximum capacity of 500 very quickly.* [Interviewee 6]

Doctors' adoption of WeChat facilitates effective knowledge dissemination in the form of photos, articles, and short video demonstrations (eg, WeChat Moments and WeChat short-term video-sharing functions). Some doctors, such as a department chair (interviewee 8) from a newly established department, demonstrate high levels of proactivity in using video-sharing functions to promote their reputation in order to attract new patients.

The Use of Medical Platforms

In China, medical platforms serve multiple functions, primarily facilitating medical appointments, medical consultations, and online prescriptions. Except an interviewee from a private dental clinic, all remaining interviewees' affiliated hospitals have already offered online platforms to execute some basic functions. These hospitals have created their own WeChat official accounts to provide supporting services via integrated miniapps in WeChat, such as making medical appointments and checking medical exam results, or created their own medical platforms

to provide a wider range of services. Table 3 outlines the hospitals' involvement in mHealth, enhancing patients' convenience in scheduling appointments via mobile apps. However, considering this kind of functionality does not necessitate interactions between doctors and patients, our investigation primarily focused on functionalities reliant on patient-doctor interactions on medical platforms.

Table 3. Affiliated hospitals' involvement in mHealth[a] adoption from June to December 2022.

Involvement of interviewees' employers in mHealth	Specific functions	Interviewees (N=11), n (%)
mHealth functions built into the hospitals' official WeChat account	• Making medical appointments • Reviewing medical exam results	10 (91)[b]
Their own medical platform exclusively designed for 1 hospital	• Making medical appointments • Reviewing medical exam results • Preparing prescriptions • Having online consultations with both first-time and follow-up patients	1 (9); interviewee 5
Official partnering with a medical platform designed for many hospitals	• Making medical appointments • Reviewing medical exam results • Preparing prescriptions • Having online consultations with both first-time and follow-up patients	1 (9); interviewee 8
No mHealth app	N/A[c]	1 (9); interviewee 3[d]

[a]mHealth: mobile health.

[b]Although these interviewees' affiliated hospitals were involved in mHealth to varying degrees via official WeChat accounts, the interviewees were not involved, since there was no need to communicate with patients via accounts at the time of the interviews.

[c]N/A: not applicable.

[d]The interviewee was from a private dental clinic.

Based on our findings, interviewees' use of medical platforms to communicate with patients was not as frequent as their use of WeChat. Of the 11 doctors, only 4 (36%) were using medical platforms at the time of the interviews, including Haodf (the largest commercial medical platform in China), Nali (a platform having an official partnership with some hospitals), the affiliated hospital's official platform, and Gancao Doctor (a commercial platform focusing on traditional Chinese medicine). They were all from tertiary hospitals. Of these 4 doctors, 2 (50%) were from esteemed top-tier hospitals (interviewees 5 and 8) and the other 2 (50%) from average tertiary hospitals (interviewees 6 and 9). They used multiple medical platforms simultaneously, with Haodf being the most popular.

The motivations behind interviewees' choices of specific platforms varied. Building departmental reputation (interviewees 8 and 9) and providing more convenience to offline patients (interviewees 5, 6, and 9) were reported as 2 key incentives for interviewees who need to perform surgeries. These interviewees have too many offline patients, which makes it rather difficult for some of those patients to make a timely appointment in the offline service. As admitted by these interviewees, the volume of patients using medical platforms is quite limited compared to their offline patients, and some of the online patients are diverted from offline services to seek more further suggestions.

One department director (interviewee 9) highlighted that using a medical platform could enable doctors to earn an extra income, and encourages subordinates, particularly young doctors, to use medical platforms to provide medical consultations to patients and then to attract more online patients to offline services. A young attending doctor (interviewee 6) reported actively using a medical platform (Gancao Doctor) to prescribe traditional Chinese medicine to patients. As explained by this doctor, following the initial in-person medical consultation that occurs at the hospital, subsequent follow-ups can be conducted online, which significantly enhances the convenience for those living in remote areas and generates an additional income for the doctor.

Some interviewees had different reasons for not using medical platforms. Interviewees 1 and 4 had tried medical platforms but ultimately discontinued their usage due to the limited number of patients engaged on those platforms. Some interviewees were overwhelmed by their offline duties and, thus, were not willing to further sacrifice their precious spare time (interviewees 1-3 and 7). Interviewee 10 from a local clinic was not busy but was still reluctant to use mobile apps related to work beyond working hours, due to family obligations. Interviewees 3 and 11 regarded WeChat as an effective approach to maintaining their connections with their patients, and they perceived WeChat as a more convenient alternative to medical platforms.

Cross-comparing the interviewees' responses on whether use medical platforms presented a contradictory understanding of the possible influence of heavy workloads from offline services. A heavy workload can either discourage or encourage the adoption of a medical platform. A cross-sectional study found no statistically significant association between workloads and doctors' uptake of e-hospitals [28]. Our interviewees' responses highlighted the underlying reason. Their different choices reflected their views on how to maintain a balance between professional duties and personal life. More than half of the interviewees from average tertiary hospitals preferred focusing

solely on offline patients to preserve some personal time. Conversely, both interviewees from esteemed top-tier hospitals were extremely busy, leaving little opportunity to prioritize their personal life during spare moments. Despite these divergent choices, interviewees from tertiary hospitals unanimously agreed that personal time is severely limited.

Interviewees' Views on Patients

The interviewees had a mixed view on the strengths and weaknesses of different communication methods, and they all preferred face-to-face communication. In addition to more comprehensive examinations of patients' symptoms, they contended that such in-person communication would facilitate the detection of subtle signs and symptoms with greater precision. Another concern arising from online conversations was the heightened cautiousness that doctors exhibit during conversations, as indicated by interviewee 8's response. The reason is that all conversations are documented in written form and might serve as evidence in the case of patient complaints. However, they admitted that the strength of online mobile communication lies in its convenience to patients, such as enabling them to review their medical exams results and providing follow-up suggestions.

> *I have many years' working experience, and I have a strong sense of cautiousness when dealing with online communication risks. I will add some preconditions with uncertain conditions. I will phrase many sentences structured as "if…, you need to do…"* [Interviewee 8]

When communicating via mobile apps, patients might be viewed differently by doctors, depending on the app selected. A notable distinction is between the 1-to-1 conversation on WeChat and communication via medical platforms. Based on the aforementioned responses of the interviewees and observations of the content of the Haodf app, we collated the disparities between these 2 communication modalities, as shown in Table 4. This table indicates that 1-to-1 messaging on WeChat is more restricted to known patients than interactions on medical platforms; in the latter case, doctors can communicate with both known and unknown patients. Medical platforms are accessible to everyone on the internet, allowing interested parties to review feedback from other patients and search for their desired doctors.

Table 4. Differences between 2 kinds of communication modalities.

Feature	WeChat (1-to-1 chat)	Medical platform
Patients	Acquaintances	Known and unknown patients
Trust	Based on reliable intermediaries or multiple rounds of communications	Based on cumulative feedback from patients on the platform
Outsiders allowed or not	Closed to outsiders	Open to outsiders
Boundary between personal and professional lives	The lack of boundary between personal life and professional work	Being professional
Cashing in on knowledge	Unable to cash in on one's knowledge immediately	Able to cash in on one's knowledge immediately

Regarding WeChat-facilitated communication, it is noteworthy that this is an enclosed mode of communication that only occurs between well-acquainted individuals, thus safeguarding the privacy of conversations. As previously discussed, the interviewees had varying degrees of willingness to share their personal accounts with others. Of the 4 (36%) interviewees using medical platforms, 3 (75%) rarely shared their WeChat accounts with patients. In China, doctor-patient communication is intertwined with social connections (*guanxi*). All interviewees were willing to share their personal WeChat accounts with patients if they were introduced by well-acquainted intermediaries, such as colleagues, relatives, or friends in hometowns or other acquaintances who could provide mutual benefits in other domains.

Interviewees' Views on the Hospitals

Of the 11 interviewees invited to participate in our study, 3 (27%) were department directors, thus expressing attitudes reflective of those of their affiliated hospitals. As reported by these interviewees, the hospitals have never hindered employees from using mobile apps to connect with patients; instead, they have actively implemented supportive strategies to promote online mobile services, such as making online medical appointments via their official portals. Nevertheless, such affiliation with hospitals still imposes certain constraints on doctors' adoption of mobile apps.

The primary factor influencing doctors' reliance on mobile phones for improved work efficiency is the multitude of tasks assigned to them. Except an interviewee from a local community clinic, all remaining interviewees reported the issue of overworking in hospitals, dedicating an average of 50 hours per week to their duties. The heavy workload prompts some doctors to use medical platforms to enhance communication efficiency with patients. Additionally, busy daytime schedules necessitate the use of WeChat for coordination tasks within departments and hospitals. Moreover, doctors in public tertiary hospitals are burdened with research and academic responsibilities, adding to their workload. Here is 2 (18%) interviewees' feedback with regard to their typical working situations.

> *For every week, I have 2 surgery days, Monday and Wednesday, and I won't be able to get off work until 9:00 p.m. or 10:00 p.m. on these days. Sometimes, it will be as late as 1:00 a.m. and 2:00 a.m. I have 2 half days scheduled for outpatients, one with around 50 patients and the other with around 35 patients. I need to work almost every day, even for the weekend. Besides, I need to have 1 night staying in the hospital in each week.* [Interviewee 5]

> *In addition to being a doctor, I am also an associate professor in a medical school… For each school year, I teach 4-5 courses to students, and I am also an advisor to 4 graduate students (master students)… I am not only in need to publish academic articles for myself in order to get higher professional titles, I also need to push my graduate students to publish articles... Using WeChat for working happens very often, and I have to tell them not to bother me after 10:00 p.m.* [Interviewee 7]

Despite the various tasks assigned by their affiliated hospitals, health care professionals' compensation remains comparatively low in China. Among the interviewees working in public hospitals, more than half mentioned that their income is not satisfactory, especially compared to their counterparts in private hospitals with higher incomes. This point of view was shared by interviewee 3, who left a public hospital and joined a private dental clinic due to an unsatisfactory income. Under such circumstances, the adoption of medical platforms could increase doctors' income, such as interviewee 6, who managed to make extra money by prescribing medications to online patients on medical platforms.

Despite the challenging tasks for interviewees in tertiary hospitals, most of them were still reluctant to leave their affiliated hospitals. For example, 2 (18%) interviewees pointed out 2 major reasons that made it unlikely for them to disconnect their affiliation with their employers: the institutional support provided by the hospitals and the hospitals' reputation to attract more patients. As stated by interviewee 1, relying solely on medical platforms is not a feasible option for them, as it is their affiliation with hospitals that provides institutional support for their careers. Interviewee 2 emphasized the predominant role of hospitals' reputation in attracting new patients.

> *Online medical platforms are operated only for profits from doctors' work. Once the disputes between doctors and patients break out, these platforms could not do anything to protect doctors and they only turn in the online communication records between doctors and patients. Doctors could hardly count on these platforms to protect them.* [Interviewee 1]

> *In China, most people care more about hospitals than doctors. If Chinese people care more about doctors, that will be a positive impact upon doctors to build their personal reputation. However, doctors are indispensable with their hospitals in China. We cannot survive without our hospitals.* [Interviewee 2]

In addition to the formal affiliation, these interviewees also have strong informal connections with their employers, particularly with their colleagues within the same organization. According to the interviewees, most doctors value their connections with their colleagues and quite often receive requests from them to provide medical services to their friends. These connections serve as crucial intermediaries for new patients, helping identify the most appropriate medical expertise for specific conditions, as mentioned by interviewee 9. However, such strong ties might also become a burden for some doctors, as exemplified by interviewee 3, who chose to leave their former employer due to the perceived disturbance caused by these connections.

> *My colleagues often introduce patients to me. I will also introduce some patients to my colleagues. We cooperate with each other. We are very clear about the inside information with regard to which doctor is good at which disease. When we introduce some patients to other doctors, we make comprehensive evaluations on their medical skills, as well as their personality.* [Interviewee 9]

> *When I worked with a previous employer (a public tertiary hospital), the relationship between me and my colleagues was very close. We were like a family. Now my connections with my colleagues in this private clinic are straight working cooperation. My former colleagues were kind of disturbing for me. I rarely introduced my friends to other doctors, because I thought it would bring some extra burden for others. However, my former colleagues would bring many patients to me. I do not like this kind of relationship, and this is one of the reasons that I chose to leave my former employer.* [Interviewee 3]

Discussion

Principal Findings

According to the semistructured interview responses, we extracted 5 themes from the verbal data of the interviewees' narratives. First, doctors have a variety of options for communicating with patients via mobile apps, with the predominant ones being social networking apps (eg, WeChat) and medical platforms (eg, Haodf). Second, all interviewees use WeChat to facilitate communication with patients, although their willingness to share personal accounts varies; they are more likely to do so with reliable intermediaries. Third, fewer than half of the doctors, all from tertiary hospitals, have adopted medical platforms. Fourth, preferences for in-person, WeChat, or medical platform communication reflect doctors' views on different patient cohorts. Fifth, doctors' choices of a particular kind of mobile app are largely affected by their affiliation with their employers, which stems from their professional duties to accomplish multiple tasks assigned by hospitals or the necessity of maintaining social connections with their colleagues.

Our findings are consistent with previous research in some aspects. Several studies have documented the use of WeChat for improving doctor-patient communication for managing certain diseases, such as advanced pancreatic ductal adenocarcinoma in China [29]. The interviewees' reservation to share personal accounts has also been echoed in previous studies on the controversies arising from the use of personal social media for medical services [30]. Meanwhile, the finding that those adopting medical platforms are all from public tertiary hospitals is also consistent with prior results [31].

However, in contrast to previous studies, our study revealed that doctors' choices of mobile apps are more than just their personal choice. A heavy workload, as mentioned in theme 4; comparatively low incomes; and strong social connections

between colleagues, as mentioned in theme 5, demonstrate that there are strong bonds between interviewees and hospitals. These constraints were consistently observed across all interviewees from public tertiary hospitals. Previous studies [32,33] and media reports [34] have documented the popularity of these issues in China. Even though most interviewees in tertiary hospitals complain about the stressful workload and low income, few of them are willing to give up their positions in their hospitals. As discussed in theme 5, they rely on these tertiary hospitals to provide institutional support, such as attracting patients and safeguarding against potential disputes.

In terms of the particular type of mobile app to facilitate doctor-patient communication, most interviewees tend to choose the ones enabling them to complete various tasks assigned by their affiliated hospitals beyond regular working hours. In other words, their online communication with patients in mobile apps is an extension of their offline clinical service. They are bonded by not only their affiliated hospitals but also their social connections with others. Some studies from Western contexts have often highlighted doctors' privacy concerns regarding mobile communications [11]. However, such privacy concerns were not prominent among the interviewees, who share their personal social medial accounts with certain patients.

Within these bonds, it is difficult for doctors to reach their individual goals of increasing their income by using medical platforms, such as Haodf, or having more spare time by not using any mobile apps after work. The findings provide some support for the validity of the single-triangle model, demonstrating the impact of the doctor-hospital relationship on the patient-doctor relationship. However, this model cannot explain why these bonds are so influential. Thus, the model we proposed at the beginning warrants reassessment to enhance our comprehension of doctor-patient interactions facilitated by mobile apps within China's health care landscape.

Remodeling the Relationships Among Doctors, Patients, and Hospitals

Behind the possible influence of hospitals are macrolevel factors. In the revised model (Multimedia Appendix 2), the triangle involving doctors, patients, and hospitals serves as a projection from a bigger triangle involving state, society, and health care provision. This double-triangle model indicates that the equilibrium within the former triangle is influenced by the dynamics of the latter triangle.

The influence of the state represents governmental influence. Governmental involvement in China's social and health care spheres is pervasive, manifesting in regulations governing mobile businesses and the oversight of medical institutions. The most notable manifestation of the state's involvement can be seen in the current prevailing status of public hospitals [35]. However, governmental funding for these public hospitals is insufficient to support their long-term growth [36]. Under such circumstances, these hospitals need to accommodate increasing patient loads to sustain their development. This explains why the interviewees from public tertiary hospitals are obligated to work for prolonged hours in order to accomplish various tasks assigned by their employers.

The influence of society embodies China's social traditions, particularly its reliance on social connections for various activities. As members of this society, doctors also need to comply with these traditions. Our findings indicate a popular pattern among the interviewees to offer informal medical advice, particularly via WeChat, to patients introduced by colleagues, relatives, friends, or fellow townspeople. Previous studies suggest that social connections play a role in building patient-doctor trust in offline services [37], and our study provides evidence for trust building through communication via mobile apps.

The influence of health care provision refers to the general characteristics of China's health care service, with 2 being particularly important. First, patients in China are not constrained by stringent referral policies, which enables them to have more flexibility in choosing their doctors [37]. Second, the patients are more inclined to go to tertiary hospitals to seek medical services, even for mild ailments [38], which is connected to the first characteristic. In this context, doctors working in those hospitals (eg, interviewees 5 and 8) are more likely to use medical platforms as a means to promote their departmental or personal reputations to attract more patients.

Rethinking the Influence of Mobile Apps

Following the discussion of the significant role of hospitals in the doctor-patient relationship, a further question arises regarding the potential impact of mobile apps on the patient-doctor-hospital dynamic. In the extant literature, online health care, including web-based consultation, has emerged as a vital alternative solution to offline health care provision, especially in terms of addressing the flaws and deficiencies persistently embedded in offline services [17]. However, our results suggest that the impact of mobile apps might be more nuanced than we initially assumed. On the one hand, their impact on offline services might be misunderstood to some extent. Although newly emerged mobile apps are gaining increasing prominence, their potential is not strong enough to challenge the predominant position of offline services, especially in the context of China's health care system. On the other hand, they might exert indirect impacts on those involved in doctor-patient communications, which might be neglected. Doctors could leverage mobile apps to promote their personal reputation, rather than simply relying on their institutional prominence to attract new patients.

Online vs Offline Health Care Provision

In our study, we found that 4 of 11 interviewees (4 of 9 doctors from tertiary hospitals) have adopted medical platforms and taken advantage of medical platforms as a more convenient means to deliver medical services to patients or to enhance their reputation to attract more potential patients for in-person hospital visits. However, it is noteworthy that offline patients remain the primary focus. For complex and severe diseases, offline services provide more advantages, as face-to-face interactions and on-site medical examinations could provide more reliable information. In a relevant study on patients' selection of doctors from Haodf, Chen et al [39] reached a similar conclusion that online health services are well connected to offline services,

especially for those departments requiring medical tests and surgeries.

Additionally, as the foremost offline service providers, hospitals have never been passive observers of emerging medical platforms but have taken a series of actions to address such challenges posed by medical information technologies. These hospitals never discourage their doctors from using medical platforms or WeChat for medical consultation purposes; rather, they are becoming more proactive in the implementation of such mobile apps to enhance the accessibility of medical services to patients [40], especially since the outbreak of COVID-19 [9]. As reflected in Table 3, almost all interviewees' employers have created official WeChat accounts for the public to access, via which patients can make medical appointments or review their medical exam results. Alternatively, they could develop their own medical platforms with functionalities mirroring existing commercial medical platforms.

Based on our observations, an isomorphic pattern might better characterize the development of both online and offline services. The predominant role of public tertiary hospitals as the primary providers of offline services is likely to persist in the realm of online medical service provision. With the adoption of online services becoming a universal trend among all hospitals, the value of online platforms tends to mainly lie in their supportive functions.

Traditional Society vs Modern Mobile Technology (Individuals vs Institutions)

The choices that doctors make regarding mobile apps might be impacted by their affiliations with particular hospitals, and the future progress of online health care services might also be closely tied to these institutions. However, by facilitating interactions between doctors and patients outside conventional clinical settings, these apps provide more opportunities for doctors to communicate with patients and to promote their professional reputation.

The establishment of individual doctors' personal reputation as competent health professionals is crucial for their career development [41]. As reported by interviewee 2, patients often rely on their employers' reputation or doctors' professional titles to search for an appropriate doctor, especially for the initial medical visit. For those using medical platforms, interviewee 6 (an attending doctor) diverts some offline patients to online medical platforms, suggesting that a patient tends to choose them based on their personal reputation rather than their affiliation with an esteemed hospital or professional title. With the advent of medical platforms, patients can review feedback from other patients when searching for desired doctors. As Table 4 suggests, these doctors' medical skills and ethical values are more accessible to those outside of conventional clinical settings.

Social networking apps could also contribute to the dissemination of doctors' professional reputation. Using 1-to-1 communication with patients via WeChat, doctors can provide convenient medical consultation to patients, or they can use group discussions to facilitate information dissemination to a wider audience. As Table 2 indicates, 6 of 11 interviewees use WeChat Moment functions, and 3 of 11 use short-term video functions, sharing the success stories of their medical treatments.

Although both kinds of apps, medical platforms and social networking apps, can facilitate doctors' reputation, the choice of apps represents a different pattern of social interaction. As Table 4 suggests, social connections have been further used in social networking apps, while medical platforms are more reflective of modern society's norms of personal interactions. *Guanxi* is a prominent property of traditional societies, where interpersonal trust largely relies on well-established social connections [37,42]. With the help of WeChat, the traditional interaction pattern continues in doctor-patient communication.

In contrast, medical platforms are more likely to attract the registration of new patients than social networking apps. The practice of being open to strangers on medical platforms aligns with the norms of personal interactions in modern society. Patients can easily access detailed information under doctors' profiles to understand their professional skills and experiences, without having to rely on their friends or relatives in hospitals to obtain that information. Moreover, each doctor could maximize the utility of their professional skills by cashing in on their knowledge immediately, without having to depend on their social connections for potential benefits in the future.

In the modern era, medical platforms and social networking apps are among the most innovative creations. Through the assistance of medical platforms, patients are empowered to choose their preferred doctors by following guidelines on how to acquire detailed information about doctors' skill sets and experiences [43]. Our study suggests that doctors are also empowered as they can expand their personal reputation beyond hospital confines.

Limitations

Several limitations of our study should be acknowledged. First, it remains controversial whether the sample of doctors we selected for interviews was representative enough to reflect the general situation of all doctors in the medical community. Nevertheless, the qualitative research methodologies we adopted in this study did have great potential to investigate underresearched topics and population groups. This study provides meaningful implications for future studies that might consider incorporating quantitative methods to add more potent evidence in this field. Second, another issue arises due to the overwhelming number of doctors from tertiary hospitals. As suggested by previous studies, doctors from these hospitals are more inclined to using medical platforms, and residents tend to seek medical services from these hospitals even for nonsevere conditions in China. As such, the overrepresentation of doctors from these hospitals might not undermine the validity of findings from our study. However, it is highly suggested that future studies conduct more in-depth investigations to understand the distinct behavior patterns of doctors from tertiary hospitals as compared to their counterparts from community clinics. Third, the selection of Hangzhou might merit questions about the external validity of the findings. Hangzhou is the most dynamic city in terms of its acceptance of mobile apps, as mentioned in the *Methods* section. If doctors in Hangzhou have reservations

about choosing a particular kind of mobile app, doctors in other cities probably will have similar concerns.

Conclusion

In contrast to the previous studies focusing on a single type of mobile app, our study further revealed doctors' adoption patterns across a wide range of mobile platforms, and the findings suggest their choices are aimed at improving their communication with patients, depending on the specific needs in offline services. Despite constraints posed by their affiliation with health care institutes, doctors still retain some autonomy in the adoption of medical information technologies. This autonomy enables them to leverage their medical expertise for disseminating knowledge or providing convenience to patients through social networking apps or medical platforms. The proliferation of medical information technologies offers doctors an alternative avenue to promote their own professional reputation without the necessity of relying on institutional support.

Our findings suggest that doctors' adoption of mobile apps is affected by macrolevel factors, particularly institutional interventions stemming from hospital affiliations. The presence of a state-society-health care provision nexus plays a pivotal role in shaping the doctor-patient-hospital relationship. Such findings also provide meaningful implications to inform doctors' adoption behavior in other countries, especially lower- and middle-income countries, which are confronted with similar issues of transitioning from traditional to modern societies. By addressing these macrolevel constraints, strategies can be devised to enhance the quality and efficiency of medical service delivery through mobile apps.

Acknowledgments

This work was supported by the General Project of the Zhejiang Philosophy and Social Science Fund (grant 21NDJC111YB), the National Natural Science Foundation of China (grant 72374149), and the Double-First Class funds of Sichuan University for talents (grant SKSYL2023-03). The authors would like to thank Yibing Wang of the Communication University of Zhejiang for her valuable support in the revision of the draft.

Authors' Contributions

DC was responsible for conceptualization, investigation, resources, revising, and project administration; DC, JP, WH, and YY for methodology, validation, and formal analysis; DC, YY, and JP for writing—original draft preparation and writing—review and editing; and DC and JP for funding acquisition.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Semistructured interview questions.
[DOCX File , 19 KB - mhealth_v12i1e49040_app1.docx]

Multimedia Appendix 2
Double-triangle model to understand mobile communication between doctors and patients.
[DOCX File , 44 KB - mhealth_v12i1e49040_app2.docx]

References

1. Jacob C, Sanchez-Vazquez A, Ivory C. Social, organizational, and technological factors impacting clinicians' adoption of mobile health tools: systematic literature review. JMIR Mhealth Uhealth 2020 Feb 20;8(2):e15935 [FREE Full text] [doi: 10.2196/15935] [Medline: 32130167]
2. Global diffusion of eHealth: making universal health coverage achievable: report of the third global survey on eHealth. World Health Organization. 2016. URL: https://apps.who.int/iris/handle/10665/252529 [accessed 2023-01-03]
3. Dorsey ER, Topol EJ. State of telehealth. N Engl J Med 2016 Jul 14;375(2):154-161 [FREE Full text] [doi: 10.1056/NEJMra1601705] [Medline: 27410924]
4. China hits 1.05 billion internet users, 99.6% are cellphone users. CGTN. 2022 Aug 31. URL: https://news.cgtn.com/news/2022-08-31/China-hits-1-05-billion-internet-users-74-4-penetration-1cWikCwAMWk/index.html [accessed 2024-05-30]
5. Chen D, Zhou Q, Pratt CB, Su Z, Gu Z. Investigating the relationships between public health literacy and public trust in physicians in China's control of COVID-19: a cross-sectional study. Front Public Health 2021 Oct 28;9:758529 [FREE Full text] [doi: 10.3389/fpubh.2021.758529] [Medline: 34778189]
6. Sun L, Buijsen M. Mobile health in China: does it meet availability, accessibility, acceptability and quality standards? Health Policy Technol 2022 Sep;11(3):100660. [doi: 10.1016/j.hlpt.2022.100660]

7. Jiang X, Xie H, Tang R, Du Y, Li T, Gao J, et al. Characteristics of online health care services from China's largest online medical platform: cross-sectional survey study. J Med Internet Res 2021 Apr 15;23(4):e25817 [FREE Full text] [doi: 10.2196/25817] [Medline: 33729985]
8. Wang J, Chiu Y, Yu H, Hsu Y. Understanding a nonlinear causal relationship between rewards and physicians' contributions in online health care communities: longitudinal study. J Med Internet Res 2017 Dec 21;19(12):e427 [FREE Full text] [doi: 10.2196/jmir.9082] [Medline: 29269344]
9. Lian W, Wen L, Zhou Q, Zhu W, Duan W, Xiao X, et al. Digital health technologies respond to the COVID-19 pandemic in a tertiary hospital in China: development and usability study. J Med Internet Res 2020 Nov 24;22(11):e24505 [FREE Full text] [doi: 10.2196/24505] [Medline: 33141679]
10. Hsu J, Liu D, Yu YM, Zhao HT, Chen ZR, Li J, et al. The top Chinese mobile health apps: a systematic investigation. J Med Internet Res 2016 Aug 29;18(8):e222 [FREE Full text] [doi: 10.2196/jmir.5955] [Medline: 27573724]
11. DeCamp M, Koenig TW, Chisolm MS. Social media and physicians' online identity crisis. JAMA 2013 Aug 14;310(6):581-582 [FREE Full text] [doi: 10.1001/jama.2013.8238] [Medline: 23942675]
12. Guo S, Guo X, Fang Y, Vogel D. How doctors gain social and economic returns in online health-care communities: a professional capital perspective. J Manag Inf Syst 2017 Aug 17;34(2):487-519. [doi: 10.1080/07421222.2017.1334480]
13. Yang Y, Zhu X, Song R, Zhang X, Guo F. Not just for the money? An examination of the motives behind physicians' sharing of paid health information. J Inf Sci 2021 Feb 14;49(1):145-163. [doi: 10.1177/0165551521991029]
14. Li Y, Yan X, Song X. Provision of paid web-based medical consultation in China: cross-sectional analysis of data from a medical consultation website. J Med Internet Res 2019 Jun 03;21(6):e12126 [FREE Full text] [doi: 10.2196/12126] [Medline: 31162129]
15. Chen J, Hsu P, Chang Y, Shiau W, Lan Y. For free or paid? A comparison of doctors' intention to offer consulting services in eHealth. IMDS 2022 Jul 08;122(8):1816-1852. [doi: 10.1108/imds-05-2021-0336]
16. Du Y, Du Y, Yao N. Patient-provider relationships in China: a qualitative study on the perspectives of healthcare students and junior professionals. PLoS One 2020 Oct 21;15(10):e0240747 [FREE Full text] [doi: 10.1371/journal.pone.0240747] [Medline: 33085691]
17. Peng X, Li Z, Zhang C, Liu R, Jiang Y, Chen J, et al. Determinants of physicians' online medical services uptake: a cross-sectional study applying social ecosystem theory. BMJ Open 2021 Sep 16;11(9):e048851 [FREE Full text] [doi: 10.1136/bmjopen-2021-048851] [Medline: 34531212]
18. Helvi K, Mikkonen K. The Application of Content Analysis in Nursing Science Research. Switzerland: Springer Nature; 2020:41-48.
19. Lai Y, Chen S, Li M, Ung COL, Hu H. Policy interventions, development trends, and service innovations of internet hospitals in China: documentary analysis and qualitative interview study. J Med Internet Res 2021 Jul 20;23(7):e22330 [FREE Full text] [doi: 10.2196/22330] [Medline: 34283025]
20. Jahangir YT, Arora A, Liamputtong P, Nabi MH, Meyer SB. Provider perspectives on sexual health services used by Bangladeshi women with mHealth digital approach: a qualitative study. Int J Environ Res Public Health 2020 Aug 26;17(17):6195 [FREE Full text] [doi: 10.3390/ijerph17176195] [Medline: 32859017]
21. Zhejiang's economic and social development achievements since the 18th national congress of the communist party of China - Hangzhou chapter. Zhejiang Provincial Bureau of Statistics. 2022 Oct 11. URL: http://tjj.zj.gov.cn/art/2022/10/11/art_1229129214_5006018.html [accessed 2024-05-30]
22. The total economic volume is 114.4 trillion yuan, exceeding the world's per capita GDP level... China's economic highlights in 2021!. Government of China. 2022. URL: http://www.gov.cn/xinwen/2022-01/17/content_5668815.htm [accessed 2023-02-10]
23. Sun G. Fudan version of "2021 Chinese hospital rankings" released (with detailed rankings). Health News. 2022. URL: https://m.gmw.cn/baijia/2022-11/20/1303200341.html [accessed 2023-01-30]
24. Chang Y, Hsu P, Wang Y, Chang P. Integration of online and offline health services: the role of doctor-patient online interaction. Patient Educ Couns 2019 Oct;102(10):1905-1910. [doi: 10.1016/j.pec.2019.04.018] [Medline: 31279612]
25. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
26. Denneson LM, Cromer R, Williams HB, Pisciotta M, Dobscha SK. A qualitative analysis of how online access to mental health notes is changing clinician perceptions of power and the therapeutic relationship. J Med Internet Res 2017 Jun 14;19(6):e208 [FREE Full text] [doi: 10.2196/jmir.6915] [Medline: 28615152]
27. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: a synthesis of recommendations. Acad Med 2014;89(9):1245-1251. [doi: 10.1097/acm.0000000000000388]
28. Li P, Luo Y, Yu X, Mason E, Zeng Z, Wen J, et al. Readiness of healthcare providers for e-hospitals: a cross-sectional analysis in China before the COVID-19 period. BMJ Open 2022 Feb 21;12(2):e054169 [FREE Full text] [doi: 10.1136/bmjopen-2021-054169] [Medline: 35190431]
29. Tang H, Zhu Z, Ying J, You T, Ge H, Cheng Y, et al. The prognostic role of WeChat-based instant doctor-patient communication in patients with advanced pancreatic cancer. J Cancer Surviv 2023 Jul 18:Online ahead of print [FREE Full text] [doi: 10.1007/s11764-023-01429-7] [Medline: 37460857]

30. Ding X, Chen Y, Ding Z, Xu Y. Boundary negotiation for patient-provider communication via WeChat in China. Proc ACM Hum-Comput Interact 2019 Nov 07;3(CSCW):1-24 [FREE Full text] [doi: 10.1145/3359259]
31. Ma Q, Sun D, Cui F, Zhai Y, Zhao J, He X, et al. Impact of the internet on medical decisions of Chinese adults: longitudinal data analysis. J Med Internet Res 2020 Sep 03;22(9):e18481 [FREE Full text] [doi: 10.2196/18481] [Medline: 32880581]
32. Liu J, Yu W, Ding T, Li M, Zhang L. Cross-sectional survey on job satisfaction and its associated factors among doctors in tertiary public hospitals in Shanghai, China. BMJ Open 2019 Mar 01;9(3):e023823 [FREE Full text] [doi: 10.1136/bmjopen-2018-023823] [Medline: 30826758]
33. Hu Y, Zhang Z. Skilled doctors in tertiary hospitals are already overworked in China. Lancet Global Health 2015 Dec;3(12):e737. [doi: 10.1016/s2214-109x(15)00192-8]
34. Surgeon's using WeChat to connect with up to 2,000 patients becomes the new normal. Beijing Youth Daily. 2018 Aug 19. URL: http://www.xinhuanet.com/politics/2018-08/19/c_1123291045.htm [accessed 2024-05-23]
35. Chen D, Lang Y. The cream-skimming effect in China's health care services: a mixed methods study. Int J Health Plann Manage 2021 Jan 10;36(1):113-133. [doi: 10.1002/hpm.3071] [Medline: 32914460]
36. Yip W, Fu H, Chen AT, Zhai T, Jian W, Xu R, et al. 10 Years of health-care reform in China: progress and gaps in universal health coverage. Lancet 2019 Sep;394(10204):1192-1204. [doi: 10.1016/s0140-6736(19)32136-1]
37. Wu D, Lam T, Lam K, Zhou X, Sun K. Challenges to healthcare reform in China: profit-oriented medical practices, patients' choice of care and guanxi culture in Zhejiang province. Health Policy Plan 2017 Nov 01;32(9):1241-1247. [doi: 10.1093/heapol/czx059] [Medline: 28981666]
38. Wagstaff A, Lindelow M. Can insurance increase financial risk? The curious case of health insurance in China. J Health Econ 2008 Jul;27(4):990-1005. [doi: 10.1016/j.jhealeco.2008.02.002] [Medline: 18342963]
39. Chen Q, Xu D, Fu H, Yip W. Distance effects and home bias in patient choice on the internet: evidence from an online healthcare platform in China. China Econ Rev 2022 Apr;72:101757. [doi: 10.1016/j.chieco.2022.101757]
40. Wang F, Wang Z, Sun W, Yang X, Bian Z, Shen L, et al. Evaluating the quality of health-related WeChat public accounts: cross-sectional study. JMIR Mhealth Uhealth 2020 May 08;8(5):e14826 [FREE Full text] [doi: 10.2196/14826] [Medline: 32383684]
41. Ren D, Ma B. Effectiveness of interactive tools in online health care communities: social exchange theory perspective. J Med Internet Res 2021 Mar 12;23(3):e21892 [FREE Full text] [doi: 10.2196/21892] [Medline: 33709940]
42. Zhang S, Wu Y, Ma W, Cheng M, Li L. Seeking medical services through guanxi in China. Patient Educ Couns 2021 Jul;104(7):1789-1795. [doi: 10.1016/j.pec.2020.12.004] [Medline: 33386188]
43. Varshney U. Mobile health: four emerging themes of research. Decis Support Syst 2014 Oct;66:20-35. [doi: 10.1016/j.dss.2014.06.001]

Abbreviations

COREQ: Consolidated Criteria for Reporting Qualitative Research
GDP: Gross Domestic Product
mHealth: mobile health
RMB: Renminbi
PDAs: Personal Digital Assistants
WHO: World Health Organization

Edited by L Buis; submitted 16.05.23; peer-reviewed by L Chaofan; comments to author 05.12.23; revised version received 26.12.23; accepted 13.05.24; published 10.06.24.

Please cite as:
Chen D, Han W, Yang Y, Pan J
Doctors' Personal Preference and Adoption of Mobile Apps to Communicate with Patients in China: Qualitative Study
JMIR Mhealth Uhealth 2024;12:e49040
URL: https://mhealth.jmir.org/2024/1/e49040
doi:10.2196/49040
PMID:38857491

Qualitative Evaluation of mHealth Implementation for Infectious Disease Care in Low- and Middle-Income Countries: Narrative Review

Josephine Greenall-Ota[1*], MRes; H Manisha Yapa[2,3*], MBChB, DTM&H, MPH, PhD; Greg J Fox[4,5,6], PhD; Joel Negin[7], PhD

[1]Faculty of Science, University of Sydney, Sydney, NSW, Australia

[2]Sydney Infectious Diseases Institute, Faculty of Medicine and Health, University of Sydney, Science Rd, Sydney, NSW, Australia

[3]Westmead Hospital, Western Sydney Local Health District, Westmead, Sydney, NSW, Australia

[4]Faculty of Medicine and Health, University of Sydney, Sydney, NSW, Australia

[5]Woolcock Institute of Medical Research, Macquarie Park, Sydney, NSW, Australia

[6]Royal Prince Alfred Hospital, Sydney Local Health District, Camperdown, Sydney, NSW, Australia

[7]School of Public Health, Faculty of Medicine & Health, University of Sydney, Sydney, NSW, Australia

[*]these authors contributed equally

Corresponding Author:
H Manisha Yapa, MBChB, DTM&H, MPH, PhD
Sydney Infectious Diseases Institute, Faculty of Medicine and Health, University of Sydney, Science Rd, Sydney, NSW, Australia

Abstract

Background: Mobile health (mHealth) interventions have the potential to improve health outcomes in low- and middle-income countries (LMICs) by aiding health workers to strengthen service delivery, as well as by helping patients and communities manage and prevent diseases. It is crucial to understand how best to implement mHealth within already burdened health services to maximally improve health outcomes and sustain the intervention in LMICs.

Objective: We aimed to identify key barriers to and facilitators of the implementation of mHealth interventions for infectious diseases in LMICs, drawing on a health systems analysis framework.

Methods: We followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist to select qualitative or mixed methods studies reporting on determinants of already implemented infectious disease mHealth interventions in LMICs. We searched MEDLINE, Embase, PubMed, CINAHL, the Social Sciences Citation Index, and Global Health. We extracted characteristics of the mHealth interventions and implementation experiences, then conducted an analysis of determinants using the Tailored Implementation for Chronic Diseases framework.

Results: We identified 10,494 titles for screening, among which 20 studies met our eligibility criteria. Of these, 9 studies examined mHealth smartphone apps and 11 examined SMS text messaging interventions. The interventions addressed HIV (n=7), malaria (n=4), tuberculosis (n=4), pneumonia (n=2), dengue (n=1), human papillomavirus (n=1), COVID-19 (n=1), and respiratory illnesses or childhood infectious diseases (n=2), with 2 studies addressing multiple diseases. Within these studies, 10 interventions were intended for use by health workers and the remainder targeted patients, at-risk individuals, or community members. Access to reliable technological resources, familiarity with technology, and training and support were key determinants of implementation. Additional themes included users forgetting to use the mHealth interventions and mHealth intervention designs affecting ease of use.

Conclusions: Acceptance of the intervention and the capacity of existing health care system infrastructure and resources are 2 key factors affecting the implementation of mHealth interventions. Understanding the interaction between mHealth interventions, their implementation, and health systems will improve their uptake in LMICs.

(JMIR Mhealth Uhealth 2024;12:e55189) doi:10.2196/55189

KEYWORDS

mHealth; implementation; LMIC; infectious diseases; Tailored Implementation for Chronic Diseases; mobile phone; interventions; short messaging service; chronic disease; narrative review; implementation; barrier; mHealth intervention; infectious disease; screening; community; design; health system; SMS; app

Introduction

Mobile health (mHealth) technologies, defined by the World Health Organization (WHO) as "the use of mobile and wireless technologies to support health objectives," have the potential to improve health outcomes globally, including in low- and middle-income countries (LMICs) [1-6]. This is achievable through improving patient education, improving disease self-management, decreasing health care costs, and performing remote monitoring of patients, as reported in a recent systematic review of mHealth in LMICs [3]. In addition, mHealth can support preventative measures, facilitate disease management, or support health workers to strengthen the delivery of health care [7-10]. The WHO has highlighted the need to advance national digital health strategies that can facilitate universal health care [11]. The WHO's "Recommendations on digital interventions for health system strengthening" highlights that digital technologies, including mHealth, can directly address health system challenges by supporting more widespread coverage across population groups and improving the quality and affordability of health care [6]. This digital transformation of the health care system has been made possible by the widespread availability of affordable digital technology; currently, 95% of the world population has internet access [12].

mHealth interventions targeting infectious diseases care have the potential to greatly transform the health care landscape of LMICs, where infectious diseases still represent a substantial burden [13,14]. This is particularly important given health system challenges such as low health service utilization, poor adherence to clinical protocols among health workers, and geographic inaccessibility of health facilities [6].

The success and sustainability of mHealth interventions require overcoming context-specific barriers and enhancing facilitators of mHealth implementation; these factors must be considered prior to intervention design. The WHO's "Global Strategy on Digital Health 2020 - 2025" acknowledged the need to adapt digital health intervention implementation to unique national contexts, health situations, and trends, as well as a country's vision, available resources, and core values [11]. The WHO's "Recommendations on digital interventions for health system strengthening" further identified key implementation enablers including health content aligning with recommended practices, intervention functionality, and greater leadership and governance [6]. The WHO has emphasized that recognizing and addressing digital health implementation challenges uniquely faced by the least-developed countries is a large factor influencing the scalability and sustainability of emerging mHealth technologies [11].

mHealth interventions in LMICs have had limited success due to a range of health system factors not considered during the development and implementation of interventions [3-5,15-17]. Many mHealth interventions in LMICs remain as pilot studies that investigate feasibility, usability, and effectiveness, and they have not been scaled-up for integration within the broader health care system [17,18]. mHealth initiatives have often been developed for use in higher-resource health systems, with little consideration of differing contexts affecting implementation, such as social norms around a health-promoting behavior or access to resources [17]. Industry representatives, such as those from mobile phone providers, often push the scale-up of mHealth interventions rather than researchers, governments, or health workers [19]. This excludes crucial end user perspectives when developing mHealth interventions and risks having market-driven motives unrelated to health care encouraging the scale-up of interventions [19]. mHealth intervention teams also often fail to understand the relationship between users and mHealth technologies [2,19-21]. Previously reported barriers to the widespread adoption of mobile technologies that uniquely concern LMICs include poor mobile network coverage, limited health care workforce capacity, limited data access, or negative health worker and patient perceptions toward mHealth interventions [3-5,15-17].

Among the limited number of published reviews evaluating mHealth implementation, there is a lack of rigorous evaluation regarding the design and implementation of mHealth interventions to aid policymakers [3,16,20,22,23].

We therefore conducted a narrative review of the existing literature to understand the determinants of mHealth implementation for infectious diseases in LMICs. The review aims to consider the broader context, drawing on a comprehensive health systems analysis framework.

Methods

We searched MEDLINE, Embase, PubMed, CINAHL, the Social Sciences Citation Index, and Global Health. We collected studies that were the earliest available indexed in the above databases, up to and including May 31, 2023; the studies were exclusively in English.

Selection Criteria

Inclusion Criteria

1. Study population: We included individuals (of any age) with infectious diseases in LMICs. We took LMIC search terms from the Cochrane Effective Practice and Organisation of Care LMIC filters, defined according to the World Bank Classification (2022) [24]. We did not restrict the type of participants in the intervention (ie, we included health workers, patients, carers, general community members, and multiple types of participants).
2. Intervention: We defined mHealth interventions as per the WHO [6]. These interventions included SMS, electronic decision-support tools, educational tools, apps, and other strategies to improve health care delivery. We included interventions that used either mobile phones, smartphones, or tablet devices, conducted at any level of the health care system.
3. Comparator: We included studies where the current standard of care was a comparator, where applicable, in addition to studies without a control group listed.
4. Outcomes: We included qualitative and mixed methods studies that included a description of the mHealth intervention and implementation processes and reported on factors affecting implementation (eg, acceptability,

feasibility, essential resources) based on interviews or discussion groups.

Exclusion Criteria

We excluded formative research studies (ie, studies conducted before fully developing or implementing an intervention); study protocols; interventions involving computers or web-based health care (eg, websites); telehealth interventions (defined as consultation with a health worker via a mobile phone either through SMS or phone calls); and quantitative studies including randomized controlled trials, Likert scale surveys, and impact evaluations, as they did not provide in-depth reporting of qualitative factors affecting implementation. We excluded studies where mHealth was part of a larger complex intervention, studies from high-income countries, and studies that combined analysis of determinants across multiple countries where it was not possible to separate out findings from LMICs versus high-income countries.

Data Extraction and Analysis

Overview

The full search terms and strategy for the databases are detailed in Multimedia Appendix 1. Briefly, we included terms pertaining to LMICs, infectious diseases (eg, communicable disease), and mHealth terms (eg, mHealth, text message, mobile app). We did not include additional filters for qualitative versus quantitative studies.

Extracted Data

One reviewer (JGO) screened the titles and abstracts of the search output for relevant studies. As a next step, we conducted full-text screening. Where eligibility criteria were unclear, final consensus on article eligibility was based on discussions with another member of the author team (HMY).

We extracted characteristics of the mHealth intervention including intervention setting (country, LMIC status, health care setting); intervention design; content and purpose; target disease and population; and its quantitative impact on health outcomes originally targeted by the intervention, as reported in the included qualitative study. This was to aid our interpretation of how the qualitative implementation determinants we identified may have affected targeted health outcomes. The original quantitative impact evaluations were not sourced for this review. We also extracted data on details of the qualitative implementation study setting, study population, research question, data collection method, and study size, as well as broad implementation determinants considered by the authors.

Data Analysis and Reporting: Tailored Implementation for Chronic Diseases Framework

We performed a framework analysis of determinants affecting implementation based on the Tailored Implementation for Chronic Diseases (TICD) framework [25]. The TICD framework is a comprehensive checklist of determinants of clinical practice developed to inform implementation research projects that are tailored to local conditions [25,26]. The framework can be applied beyond contexts of chronic diseases, as the framework broadly focuses on health system components that determine quality of care. It identifies 7 key domains: guideline factors (clinical care guidelines or mHealth as a “guideline”); health worker factors; patient factors; professional interactions; incentives and resources; capacity for organizational change; and social, political, and legal factors. Its strength lies in its emphasis on ensuring tailoring to local conditions, which is valuable to consider for LMIC interventions, and inclusion of contextual (including political and legal) and patient factors affecting implementation. The TICD framework is a comprehensive health systems framework aligned with the systems thinking framework, which considers interactions with the broader context and patient needs in addition to the structural components of a health system [27].

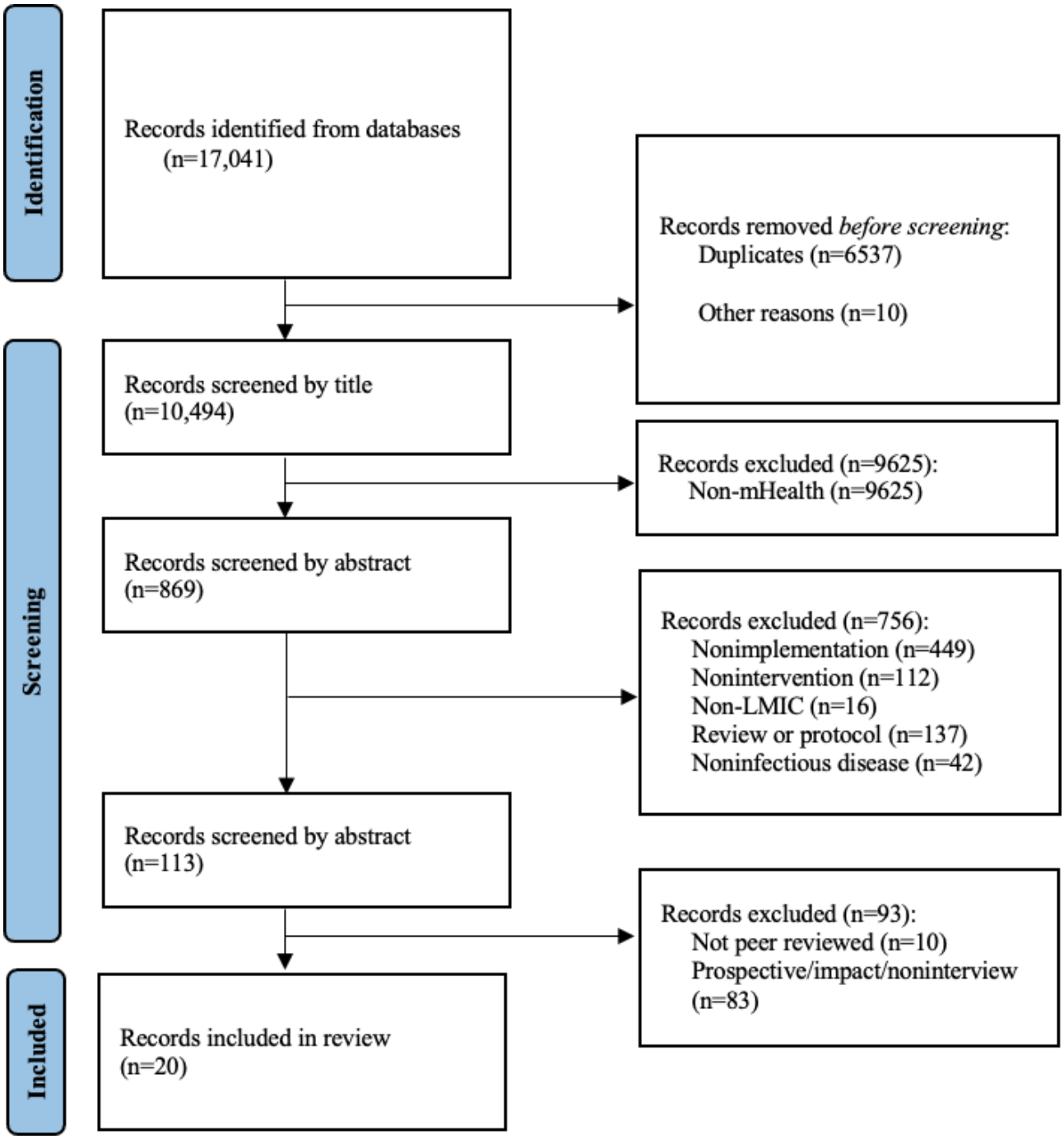
The framework was used to initially categorize the reported determinants of mHealth intervention implementation and was used to structure the reported findings in the results. Additional emergent themes, such as intervention design and forgetfulness, were extracted.

Results

Screening Results

A total of 17,041 records were initially identified. After removing 6537 duplicates, 10 non-English studies, and 9625 non–mHealth-related studies, a total of 869 studies underwent abstract screening in 2 rounds, each examining different criteria, as seen in Figure 1. A total of 20 studies were included for data extraction and final analysis.

Figure 1. PRISMA flow diagram. Screening strategy and PRISMA reported according to flow diagram for systematic reviews, including database searches. mHealth: mobile health; LMIC: low- and middle-income countries; PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.



Characteristics of Included Studies

Details of the intervention including study setting; intervention type and purpose; target disease and population; and quantitative impact evaluations (as reported in the included qualitative manuscripts) are presented in Table 1. Table 2 details qualitative studies analyzed in this review. Among the 20 included studies, 9 were apps and 11 were SMS interventions. Based on the World Bank 2022 Country Group by Income LMIC status [28], 8 were low income [8,9,29-34], 6 were lower middle income [10,35-39], and 6 were upper middle income [7,21,40-43]. Of the 20 studies, 7 focused on people living with or at risk of acquiring HIV [7,21,31,33,34,39,41]; the remaining studies targeted malaria [8,30,35,36], tuberculosis [31,38,42,43], pneumonia [10,30], dengue [37], a grouping of "respiratory illnesses" [9] or "childhood chronic infectious diseases" [29], HPV [40], or COVID-19 [32]. There were 2 studies that addressed multiple diseases [30,31]. Half (10/20) of the interventions targeted health workers [9,10,21,30,31,35,36,39,40,42,43], while the remainder were for patients, individuals at risk of disease, or general community members [7,8,21,29,32-34,37,38,41]. Most studies (11/20) involved community level health care [7,8,10,21,29,32,33,36,37,39,41].

Determinants of mHealth implementation are reported according to the 7 TICD domains and additional emergent themes.

Table . Characteristics of mHealth intervention among included studies.[a]

Country, author, year, reference	World Bank country income classification	Level of health care; health care context	mHealth intervention	Study design of original quantitative study/comparator	Content of the intervention	Purpose of the intervention	Target disease(s)	Target user(s)	Quantitative impact of mHealth intervention on targeted health outcome(s) as reported in the qualitative study
Argentina, Straw et al, 2023 [40]	Upper middle	Primary; primary hospital and primary health care centers consisting of community health workers	SMS	C-RCT[b]/standard of care	Reminder messages to women with HPV[c], and one to community health workers about women with no triage 60 days after positive HPV test	Improve Pap triage of HPV-positive women	HPV	HPV-positive women and health workers	Both acceptability of the intervention by HPV-tested women and its adoption by health workers were high: 15% increase in percentage of women with triage Pap after HPV result. Statistical significance not reported
Ghana, Ginsburg et al, 2016 [10]	Lower middle	Community; health centers and community-based health planning and service centers in rural Ghana	App (phone or tablet)	Not reported	Software-based breath counter and a pulse oximeter to count child's breaths, off-the-shelf reusable pediatric pulse oximeter to detect hypoxia	Improve pneumonia diagnosis and treatment and childhood illnesses in general in accordance with the Integrated Management of Childhood Illness guidelines	Pneumonia	Health workers	Not reported
Kenya, Jones et al, 2012 [35]	Lower middle	Primary; government dispensaries and health centers	SMS	C-RCT/control group	Pediatric outpatient malaria case management accompanied by "motivating" quotes to health worker's personal mobile phones. Two messages per day across five working days for the duration of the study (26 weeks)	Improve health worker's malaria case-management practices, specifically drug dispensing and management	Malaria	Health workers	Intention-to-treat analysis showed 24% improvement compared to baseline in correct antimalarial drug management immediately after the intervention, sustained effect of 25% six months later. Statistical significance not reported

Country, author, year, reference	World Bank country income classification	Level of health care; health care context	mHealth intervention	Study design of original quantitative study/comparator	Content of the intervention	Purpose of the intervention	Target disease(s)	Target user(s)	Quantitative impact of mHealth intervention on targeted health outcome(s) as reported in the qualitative study
Malawi, Ide et al, 2019 [29]	Lower	Community; village clinic in Northern Malawi	App	Not reported	App-directed assessment and management of the visit; the visit was documented in both the app and the village clinic register	Improve assessment, classification, and treatment of seriously ill children, facilitate disease monitoring and surveillance [44]	Childhood infectious disease outbreaks	Health surveillance assistants, caregivers	Not reported
Malawi, Kaunda-Khangamwa et al, 2018 [30]	Lower	Primary and tertiary; health facilities operated by government or private care (including tertiary hospitals)	SMS	C-RCT/control group with no messages	Twice-daily text message reminders on case management of malaria, pneumonia, and diarrhea sent to clinicians and drug dispensers	Improve case management of malaria, diarrhea, and pneumonia	Malaria, diarrhea, pneumonia	Health workers	Nonsignificant, 4% improvement in correct malaria case management. Statistical significance not reported
Mali, Mangam et al, 2016 [8]	Lower	Community; rural district community	SMS	3 pilot intervention villages against 3 nonintervention villages	Educate and instruct households about indoor residential spraying campaign	Lower malaria prevalence through preventative measures	Malaria	Households	Significantly lower among the mobile-messaging villages than the door-to-door mobilization villages (86% vs 96%, respectively; P=.02) and significantly lower structural preparedness in households mobilized through the mobile-messaging approach compared with the door-to-door approach (household and food items removed; 49% vs 75%, respectively; P=.03)

Country, author, year, reference	World Bank country income classification	Level of health care; health care context	mHealth intervention	Study design of original quantitative study/comparator	Content of the intervention	Purpose of the intervention	Target disease(s)	Target user(s)	Quantitative impact of mHealth intervention on targeted health outcome(s) as reported in the qualitative study
Mozambique, Nhavoto et al, 2017 [31]	Lower	Primary; health care centers providing intense ART[d] and tuberculosis care	SMS	RCT[e]/no comparator	Structured series of SMS text messages sent automatically based on appointments and scheduled drug pick-ups. Messages were sent 7 and 2 days before appointment or drug pick-up	Support retention in ART and tuberculosis treatment	HIV and tuberculosis	Patients and health workers	The majority of HIV patients (61/68, 90%) and the majority of tuberculosis patients (60/68, 88%) reported not having missed any appointments. Majority of the patients (HIV: 56/68, 82%; tuberculosis: 65/67, 97%) reported not having missed medication pick-up at any time
Myanmar, Win Han et al, 2021 [36]	Lower middle	Community; ICMV[f] managed by Myanmar's National Malaria Control Programme and its implementing partners	App (phone)	Mixed method/paper-based reporting control group	Malaria case-based data entered by ICMVs directly in the app on their mobile phones, which is instantly uploaded onto the dedicated District Health Information System 2 database	Enabled more accurate and complete data reported to improve integrated community malaria volunteers' malaria prevention, diagnosis, treatment and referral services	Malaria	Health workers	Not reported

Country, author, year, reference	World Bank country income classification	Level of health care; health care context	mHealth intervention	Study design of original quantitative study/comparator	Content of the intervention	Purpose of the intervention	Target disease(s)	Target user(s)	Quantitative impact of mHealth intervention on targeted health outcome(s) as reported in the qualitative study
Nepal, Bhattarai et al, 2019 [37]	Lower middle	Community; rural households in district highest in dengue outbreaks	SMS	Nonrandomized quasi-experimental design. Control group (standard of care), dengue prevention leaflet only group, dengue prevention leaflet with mobile SMS intervention group	Disseminate preventative health messages via mobile phone twice per week	Change in dengue preventative knowledge and practice of respondents	Dengue	Households	Dengue prevention leaflet and SMS intervention combination significantly improved dengue prevention knowledge (mean 32.7, SD 13.7 vs mean 13.3, SD 8.8) and practice (mean 27.9, SD 11.4 vs mean 4.9, SD 5.4) compared to without SMS (P<.001)
Pakistan, Mohammed et al, 2012 [38]	Lower middle	Secondary; hospital in low-income industrial area (free)	SMS	No control group	Daily reminders sent and patients were asked to respond after taking their medication. Motivational message, followed by a reminder to respond to the system	Improve tuberculosis drug adherence	Tuberculosis	Patients with tuberculosis	Mean response rate of 57% for all participants. The mean response rate fell from 62% during the first 10 days to 49% during the last 10 days across the 30-day intervention. Statistical significance not reported.

Country, author, year, reference	World Bank country income classification	Level of health care; health care context	mHealth intervention	Study design of original quantitative study/comparator	Content of the intervention	Purpose of the intervention	Target disease(s)	Target user(s)	Quantitative impact of mHealth intervention on targeted health outcome(s) as reported in the qualitative study
Russian Federation (Siberia), Hodges et al, 2022 [7]	Upper middle	Community; inpatient and outpatient settings in the tuberculosis referral hospital and AIDS center	App (phone)	Not reported	Daily patient check-ins or queries regarding stress, mood, and ART adherence; appointment reminders; tailored educational resources; access to HIV-related lab results; community message board for anonymous peer messaging; direct messaging with clinic care team members outside of the clinic	Enhance linkage of HIV/tuberculosis coinfected patients to HIV care and promote sustained engagement with and integration of HIV and tuberculosis care	HIV	People with HIV	Improved rates of linkage to care at the AIDS center, improved medication refill rates, reduced virologic failure at 6 months on ART. Related statistics were not reported.
Rwanda, Babili et al, 2023 [32]	Lower	Community; home-based care	SMS	Not reported	Automated check-in message sent daily for 14 days, contact tracing, and data centralization	Real-time remote monitoring and support of COVID-19 cases and contacts	COVID-19	Patients with COVID-19	Nonresponse rate 25% - 30%
South Africa, Adeagbo et al, 2021 [41]	Upper middle	Community; rural and periurban HIV hyperendemic area	App (tablet)	C-RCT/standard of care	Enable user to hear the “story” of a chosen character portrayed on the app	Improve HIV testing and linking with care in rural South Africa	HIV	Male participants at risk of HIV	83% consented to and used a home-based rapid HIV test, 33% received HIV testing for the first time in the annual HIV surveillance. Statistical significance not reported.

Country, author, year, reference	World Bank country income classification	Level of health care; health care context	mHealth intervention	Study design of original quantitative study/comparator	Content of the intervention	Purpose of the intervention	Target disease(s)	Target user(s)	Quantitative impact of mHealth intervention on targeted health outcome(s) as reported in the qualitative study
South Africa, Chaiyachati et al, 2013 [42]	Upper middle	Tertiary; decentralized MDR-TB[g] treatment centers in rural South Africa, part of larger MDR-TB hospital	App (phone)	Not reported	Reporting of adverse effects of MDR-TB, decision aids for triaging symptoms complaints, adherence questions, and a tool for tracing newly diagnosed tuberculosis patients or finding defaulters from tuberculosis treatment	Improving the acceptability feasibility of clinical monitoring and management of adverse events in patients receiving community-based MDR-TB treatment	MDR-TB	Health workers	Low user uptake: 27% of health workers submitted adverse event forms through the mHealth app
South Africa, Janssen et al, 2020 [21]	Upper middle	Community; community township–based HIV clinics [45]	App	Not reported	Video that shows the user how to conduct the oral self-test, guide on how to interpret test results, information about HIV and HIV transmission, HIV risk assessment with questions regarding a person's sexual behaviors, condom use, and alcohol and drug use	Support HIV self-testing and care	HIV	People presenting to HIV self-test clinics	Not reported

Country, author, year, reference	World Bank country income classification	Level of health care; health care context	mHealth intervention	Study design of original quantitative study/comparator	Content of the intervention	Purpose of the intervention	Target disease(s)	Target user(s)	Quantitative impact of mHealth intervention on targeted health outcome(s) as reported in the qualitative study
South Africa, Maraba et al, 2018 [43]	Upper middle	Primary; primary health clinics	App and SMS	Quantitative comparisons between preintervention and postintervention	Data collection, automate results delivery, display results, provide notifications, and directly provide results via results notification and via text	Reduce the time and effort required for tuberculosis data reporting, provide rapid and automatic access to Xpert MTB/RIF tuberculosis test results	Tuberculosis	Health workers	No statistically significant difference in results between paper-based system and mHealth-based system in terms of details documented, proportion on tuberculosis treatment, and time before results availability
Uganda, Ellington et al, 2021 [9]	Lower	Primary; primary health care facilities, one periurban and one rural	App	Not reported	Decision support tool, partially automated respiratory rate counter, educational videos, adapted respiratory assessment score to determine bronchodilator responsiveness	To improve diagnosis and treatment of acute lower respiratory infections in children <5 years of age	Acute lower respiratory illness	Health workers	Not reported
Uganda, Haberer et al, 2010 [33]	Lower	Community; outpatient HIV clinic at regional rural referral hospital	SMS and interactive voice recording	Not reported	Users were expected to respond to a question regarding HIV ART treatment adherence sent via SMS, unlocked via personal identification number prompt	Automated collection of weekly individual-level ART adherence data	HIV	Caregivers of HIV-positive patients	76% of the SMS cycles were not answered, meaning that no response was received to the greeting and prompt for the personal identification number necessary to respond to the adherence question

Country, author, year, reference	World Bank country income classification	Level of health care; health care context	mHealth intervention	Study design of original quantitative study/comparator	Content of the intervention	Purpose of the intervention	Target disease(s)	Target user(s)	Quantitative impact of mHealth intervention on targeted health outcome(s) as reported in the qualitative study
Uganda, Twimukye et al, 2021 [34]	Lower	Primary; HIV clinics from a periurban government health facility	SMS and interactive voice recording	RCT/standard of care	Interactive voice recording calls or SMS text message appointment reminders on or before the scheduled appointment date. Allows patients to report symptoms at the end of the scheduled call through a toll-free line	Promote adherence to ART for young adults	HIV	HIV-positive youth	Not reported
Zimbabwe, Venables et al, 2019 [39]	Lower middle	Community; health centers in rural settings	SMS	Not reported	HIV viral load testing results sent to health facilities or patient	Reduce the amount of waiting time for test results and improve adherence support	HIV	Health workers and patients	Median waiting time from reporting of the viral load result at the laboratory to starting enhanced adherence counseling was reduced from 47 days to 30 days compared to preintervention

[a]mHealth: mobile health.

[b]C-RCT: clustered randomized controlled trial.

[c]HPV: human papillomavirus.

[d]ART: antiretroviral therapy.

[e]RCT: randomized controlled trial.

[f]ICMV: integrated community malaria volunteers.

[g]MDR-TB: multidrug resistant tuberculosis.

Table . Characteristics of qualitative study design among included studies.

Country, author, year, reference	Study population	Research question	Data collection method (study size)	Reported qualitative factors affecting user uptake
Argentina, Straw et al, 2023 [40]	Health decision makers, health workers	Stakeholder's perception about mHealth[a] implementation strategy and factors affecting scale-up	Semistructured interviews (n=20)	• Knowledge of the strategy • Characteristics of the intervention (intervention source; design quality and access to knowledge and information; adaptability; complexity; and compatibility with norms, values, and existing workflows and systems; relative advantage; consideration of patient needs; relative priority; leadership engagement; external policies; cost)
Ghana, Ginsburg et al, 2016 [10]	Health administrators, health workers, caregivers	Feasibility, usability, and acceptability of the app in 6 health centers and 5 community-based health planning and services centers	In-depth interviews (n=69)	• Feasibility of the app: national-level support, integration into routine childhood care, electricity needs • Usability of the app: easy-to-use design, improvement over manual assessments, time constraints with full assessment • Acceptability, algorithm perception as accurate
Kenya, Jones et al, 2012 [35]	Health workers	Perceptions and experiences of health workers involved in the mHealth intervention	Interview (n=24)	• Perception of the intervention based on app design • Clinical importance of practice • Relationship to training, guidelines, and other interventions
Malawi, Ide et al, 2019 [29]	Health surveillance assistants, caregivers	Acceptability and impact of the app	Semistructured interviews (n=40)	• Health surveillance assistant and caregiver acceptability and beliefs: evidence strength and quality, tension for change, beliefs about the app • Technical and clinical characteristics of the app: impact on clinical assessments, learning curve, features of the app, relative advantage of the app over standard care • Technological infrastructure • Caregiver-health surveillance assistant relations

Country, author, year, reference	Study population	Research question	Data collection method (study size)	Reported qualitative factors affecting user uptake
Malawi, Kaunda-Khangamwa et al, 2018 [30]	Drug dispensers and health workers	Health worker perceptions of the messages received, possible mechanisms of action, and potential challenges to acting on the SMS reminders with the overarching goal of understanding the reasons why the intervention was ineffective and elucidating lessons learned	Semistructured interviews (n=50)	• Perceptions of the SMS intervention • Design of SMS intervention • Health facility resources: staff, finance • Communication between health workers
Mali, Mangam et al, 2016 [8]	Pilot village community members	Pilot effectiveness was investigated by evaluating structure preparedness	Interview (n=18), survey interview (n=673)	• Language literacy (writing and reading) • Perceptions and attitudes about the intervention due to familiarity with mobile phones • Enumeration
Mozambique, Nhavoto et al, 2017 [31]	Patients, health workers	Patient and health worker views on an mHealth intervention aiming to support retention in ART[b] and tuberculosis treatment in Mozambique	Semistructured interview (n=181)	• Perceptions of participants toward the SMS system
Myanmar, Win Han et al, 2021 [36]	Community malaria volunteers, malaria program stakeholders	Qualitative assessment of the sustainability prospects of the reporting system in the context of Myanmar's malaria elimination program	FGDs[c] (n=84); semistructured, in-depth interviews (n=14)	• User satisfaction, system access • Ownership, human resources, financial sustainability, system applicability (net benefits), policies and operation procedures • Technological system interoperability, system scalability, system relevance, system quality
Nepal, Bhattarai et al, 2019 [37]	All household heads or spouses, SMS recipients, key informants from stakeholder organizations (dengue focal points, public health officer)	Effectiveness, acceptability, and appropriateness of the mobile SMS intervention in improving behavior in dengue endemic areas of Nepal	In-depth interviews (n=13), survey interviews (n=300)	• Network • Mobile phone familiarity, existing practices • Financial cost • Organizational and social responsibility • Organizational readiness • Intervention design: entertainment value, informative nature, timing of messages • Perceived usefulness
Pakistan, Mohammed et al, 2012 [38]	Patients	Perceptions, acceptability, and engagement with an interactive SMS reminder system for patients with tuberculosis	Semistructured interviews (n=24)	• Literacy • Ownership of a mobile phone and familiarity with SMS messaging • Familiarity with tuberculosis treatment adherence • Technological problems
Russian Federation (Siberia), Hodges et al, 2022 [7]				

Country, author, year, reference	Study population	Research question	Data collection method (study size)	Reported qualitative factors affecting user uptake
	People with HIV treated at the AIDS center, people with HIV and tuberculosis treated at the referral hospital, clinical and nonclinical providers from the tuberculosis hospital and AIDS center	Process evaluation for adaptation, testing, and dissemination of the mHealth program	Unstructured group interviews (n=30), community message board sampling survey (n=47)	• Language adaptation of platform components • Server management and technological resources • Iteration of platform features • Program organizational practice integration • Communication about the intervention
Rwanda, Babili et al, 2023 [32]	Senior staff (policymakers, directors, and senior managers), technical teams (case managers and health workers supporting intervention implementation)	Rationale, perspectives, and experiences of stakeholders during mHealth intervention implementation and the intervention's scalability and adoptability	Semistructured 1-on-1 interviews (n=7)	• Local governance and policies • Local infrastructure • Intervention characteristics • End user characteristics • Local culture and communication
South Africa, Adeagbo et al, 2021 [41]	Men aged >15 years	Acceptability of and satisfaction with the intervention	In-depth interviews (n=20), qualitative survey interviews (n=232)	• Relevance and convenience • Familiarity of practice based on existing behavior • Ease of use
South Africa, Chaiyachati et al, 2013 [42]	Mobile health workers	Evaluate acceptability and feasibility of using the app to record and submit adverse event forms weekly; evaluate mobile health workers' perceptions throughout the pilot period	2 in-depth FGDs with health workers (group size n=5)	• Forgetfulness • Technological functionality • Responsibility of care • Ease of use
South Africa, Janssen et al, 2020 [21]	Nurses and health workers, patients	Affective dimensions of HIV self-testing using a smartphone app strategy	Interview (n=30), 1 FGD (group size n=6)	• Past interactions with health care system or experiences with HIV testing • Perception of the app
South Africa, Maraba et al, 2018 [43]	Tuberculosis patients, health workers	Feasibility, acceptability, and potential of an mHealth app to reduce initial loss to tuberculosis follow-up	Structured interviews (n=29)	• Perceived benefit (patients and health workers) • Technical difficulties • Proficiency in receiving text messages/understanding the intervention
Uganda, Ellington et al, 2021 [9]	Health facility administrators, primary health workers (nurses, clinical officers)	Health workers' perceptions of acceptability, usability, and feasibility of the app	Semistructured interviews (n=3), 3 FGDs (n=25)	• Acceptability and perceived benefit of the app, app usability, provider-patient relationship • Health facility resources, eg, human resources, supply chain • App integration into existing health system • Stakeholders buy-in
Uganda, Haberer et al, 2010 [33]	Caregivers, intervention participants	Participant impressions of the technologies	Qualitative interview (n=19)	

Country, author, year, reference	Study population	Research question	Data collection method (study size)	Reported qualitative factors affecting user uptake
				• Previous experience with mobile phones • Understanding how to respond to the interactive voice recording and SMS prompts • Network access and technological resources
Uganda, Twimukye et al, 2021 [34]	Young adults (aged 18 - 25 years) with HIV infection	Acceptability and feasibility of a mobile phone support tool to promote adherence to ART among young adults in a randomized controlled trial	In-depth interview (n=11), 1 FGD (group participants n=21).	• Perceptions toward app • Technical issues, access to mobile phones • Stigma
Zimbabwe, Venables et al, 2019 [39]	Patients, health workers	Patient and health care worker experiences and perceptions of the SMS intervention	In-depth interview (n=32), FGD (n=5)	• Patients' understanding of messages • Perceived benefits of text messages • Organizational personnel leading the intervention • Technology resources

[a]mHealth: mobile health.

[b]ART: antiretroviral therapy.

[c]FGD: focus group discussion.

Determinants of mHealth Implementation: TICD Framework

Guideline Factors

There were 6 studies that reported on how lack of clarity and understanding of the intervention hindered its uptake. For example, users did not respond to mHealth intervention text message prompts because they were uncertain about how to use the personal identification number, were uncertain about how to respond to the message, or received little to no information about the background of the intervention [7,30,33,39,40,43]. Compatibility of the intervention with users' past experiences with either the technology and its use in health care, or the intended change in health care practice, was a key facilitator that aided in implementation. Existing familiarity with the intended behavior promoted by the intervention—such as antiretroviral treatment adherence, HIV testing, or general interactions with the health care system—facilitated uptake [21,29,35,38,41]. The mHealth intervention in Babili et al's study—WelTel, assessed for COVID-19 case and contact management—was previously implemented for HIV epidemic virtual care, which facilitated its implementation as users were familiar with the platform and its functionality [32]. Similarly, health workers commented on how an app's alignment with existing practices of using the village clinic register increased the likelihood of engagement [29]. Existing mobile phone use or interest in the use of new technologies were additional facilitators, as users were already familiar with making appointments, making calls, or using SMS technology [8-10,32,33,36,37,42]. Trustworthiness of the source of the recommendation given by the mHealth intervention facilitated implementation, particularly when the mHealth intervention intended to aid health workers in improving their health care practice or disease management [10,29,32,35,40].

Individual Health Care Worker Professional Factors

mHealth being perceived as useful by end users in improving existing health care practices facilitated the uptake and integration of the intervention. For instance, after initial use of the intervention, a perception that it might improve existing clinical practices, improve patient engagement with services, or relieve strain on the health system were key for implementation [9,29-32,35,36,39,40,43]. Health workers in Ide et al's study perceived the app as advantageous over existing systems, as the intervention aided in conducting more accurate, error-free community case management of malaria, diarrhea, and pneumonia, which facilitated uptake [29].

Furthermore, for interventions for which the main users were health workers, attitudes toward the intervention were impacted by the perceived effect of the use of the intervention on the health worker's reputation. Ellington et al identified that the perceived loss of trust between patients and health workers in the health worker's ability to diagnose and treat patients due to their reliance on technology to deliver health care was a barrier to using the app [9]. In contrast, Ide et al commented on how the app facilitated perceived professionalism [29]. Twimukye et al commented on how the health worker's use of the app improved how the patient perceived the health worker's attention to detail and care [34].

Patient Factors

Patients perceiving the intervention as beneficial in improving health outcomes through increased convenience, awareness, or reminders facilitated implementation of the intervention [34,38,41,43]. Adeagbo et al commented on how the app's positive messages about HIV testing and adherence promoted users to adopt new behaviors by improving individual competency to make informed, healthy decisions concerning sexual health [41]. Language literacy was a barrier that limited engagement with 2 SMS intervention studies targeted at patients or community members [8,38].

Professional Interactions

Limited supervision and follow-up of the intervention implementation by the research team leading the intervention was noted as a barrier, as users requested feedback and confirmation of correct intervention use [30]. Lack of or limited access to training to provide necessary skills to effectively engage with the mHealth intervention was a barrier to implementation for health workers [9,36]. Access to professional training was a facilitator of implementation [10,29,40].

Incentives and Resources

Several studies reported that access to resources and essential infrastructure influenced implementation. Specifically, 6 studies reported that poor telecommunications networks, problems with electricity, a lack of phone coverage, limited staffing, and a lack of equipment to implement the behavior change were barriers to implementation [10,30,33,37,40,42]. Network problems were particularly prominent in geographically remote areas. Other barriers included technology that repeatedly malfunctioned (such as periodic freezing and system crashes) and limited access to technology support systems to troubleshoot technological problems [7,29,32-34,36,38,42,43]. Access to technological resources and support in case of malfunctions facilitated implementation [33].

Capacity for Organizational Change

Financial instability, existing patient overflow, incompatible technological equipment, and length of appointment time within the clinic hindered the implementation of mHealth interventions [7,9,30,37,39]. Kaunda-Khangamwa et al reported that 90% of the health worker respondents blamed existing high workloads and drug stockouts as factors discouraging health workers to respond to SMS reminders that promoted infectious disease case management [30]. Similarly, Ellington et al's study identified that the existing length of the appointment time was not compatible with mHealth intervention use as the time to complete a health assessment through the app took longer than the appointment duration [9]. Straw et al commented on the compatibility of the mHealth intervention with the existing organization functionality to facilitate normal workflow as a facilitator of implementation [40]. A lack of management and ownership of the intervention by health workers further contributed to a limited capacity for embedding the intervention within the health care facility [39].

Social, Political, and Legal Factors

One study commented that the costly nature of airing messages during the daytime and limited funding for the intervention were barriers to successful implementation [42]. Receiving national-level support on a political level including the Minister of Health or district leadership was a facilitator of implementation [9,10,32]. Babili et al commented on how the Rwandan government's support of digitization across all governmental sectors by offering resources for implementation aided the adoption of the digital health intervention [32]. Furthermore, shared responsibility and corporate social responsibility felt among the wider community was an enabler, as the charitable community assisted in reducing costs and improving uptake [37].

Emergent Themes

A novel factor not directly addressed in the TICD framework is the importance of considering app or SMS design features. Frequency of messages, language, and integration of local narratives to engage users were reported as affecting ease of use and user uptake [9,10,30,34,35,37,41,42]. mHealth interventions that adapted content toward the intervention context by using local proverbs, narratives, or language facilitated uptake of the intervention [7,30,35,41]. Moreover, features such as a user-friendly interface and a streamlined workflow facilitated implementation, while an intervention design that increased the workload of health workers was a barrier to implementation [32].

The study by Mangam et al, whose reminder SMS system replaced the existing door-to-door reminders, uniquely commented on how the absence of face-to-face reminders increased the rate of forgetfulness and patients ignoring the messaging, thereby affecting its implementation [8].

mHealth Impacts on Target Health Outcomes and How Implementation Determinants May Have Influenced These Outcomes

As described in Table 1, there were 8 studies that identified that mHealth had a positive impact on health outcomes or behavior, whether that was through increased HIV testing, reduced errors in drug and disease management, or improved disease prevention knowledge, results collection, or linkage to care for better clinical practice [7,29,31,35,37,39-41]. In 4 studies, the mHealth intervention did not lead to an improvement in the health outcome—responses to SMS prompts were low, preventative measures were worse than in the non-mHealth control group, or user uptake was low [8,33,38,42]. There were nonsignificant changes in health outcomes or behavior in 2 studies [30,43]. Quantitative health outcomes were not reported in 6 of the reviewed studies [9,10,21,32,34,36].

Improved outcomes may be explained by familiarity with the health behavior or technology [7,29,35,37,41], positive attitudes among health workers toward the technology [29,31,35], or ease of use of the mHealth technology [7,35,37,41]. Technological barriers, lack of familiarity with technology, and resource limitations [8,30,33,38,42,43] may have reduced engagement with the intervention or the participants' ability to

implement the behaviors enforced by the intervention, therefore diluting the intervention effect.

Discussion

Principal Findings

Findings from the 20 reviewed studies and categorization into the TICD framework were synthesized to deduce two overarching themes that influenced the successful implementation of mHealth initiatives in LMICs: (1) the acceptance of the intervention by patients and health workers (as well as on a sociopolitical level), regardless of the target user, and (2) the capacity of existing infrastructure and resources to implement the intervention, which was strongly tied to the health system's capacity for change. This relationship is visually depicted in Figure 2. The logic flow diagram in Figure 3 further represents these reported factors according to inputs required for mHealth interventions and the required processes for success.

Figure 2. Model of factors influencing implementation of mHealth interventions in LMICs. The two main factors are presented in large boxes: (1) acceptance of the intervention and (2) capacity of the existing infrastructure and resources to accommodate mHealth. Acceptance, in turn, was mainly influenced by perceived usefulness of the intervention, amount of training and communication, and previous experience with the guideline behavior and mHealth or technology. The health care system's capacity and infrastructure and resources were influenced by funding, network availability, and technological support. mHealth: mobile health.

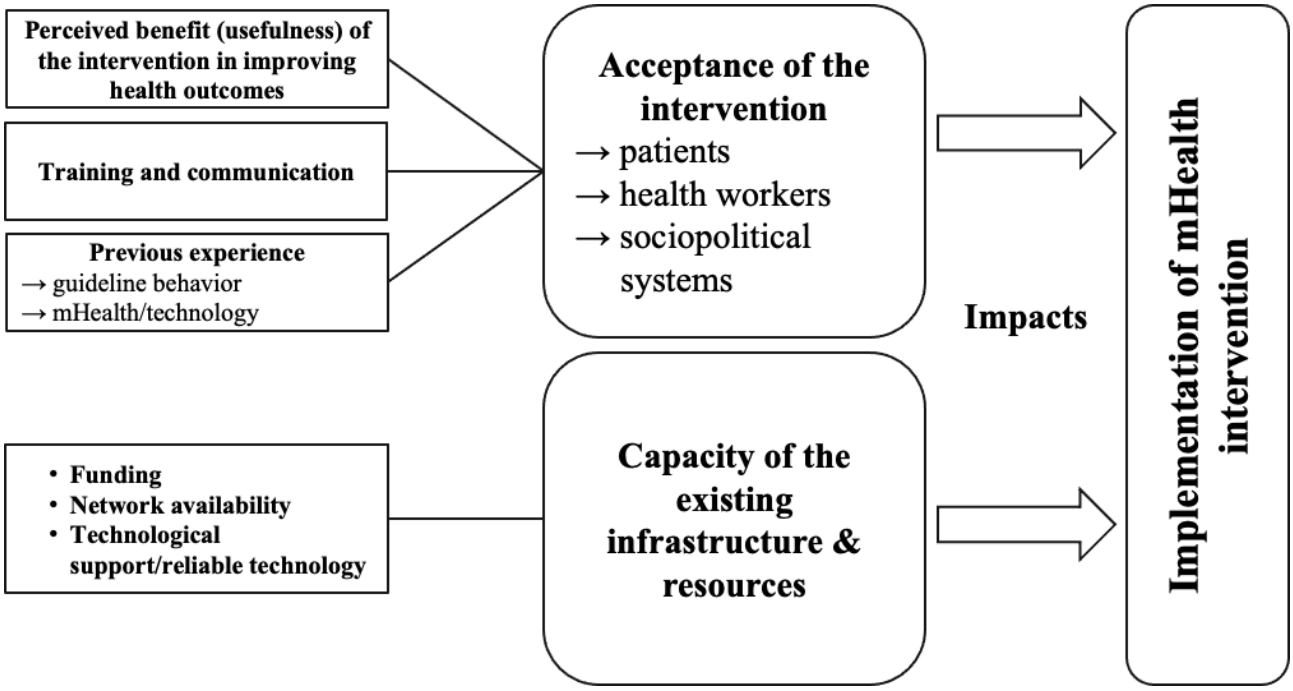
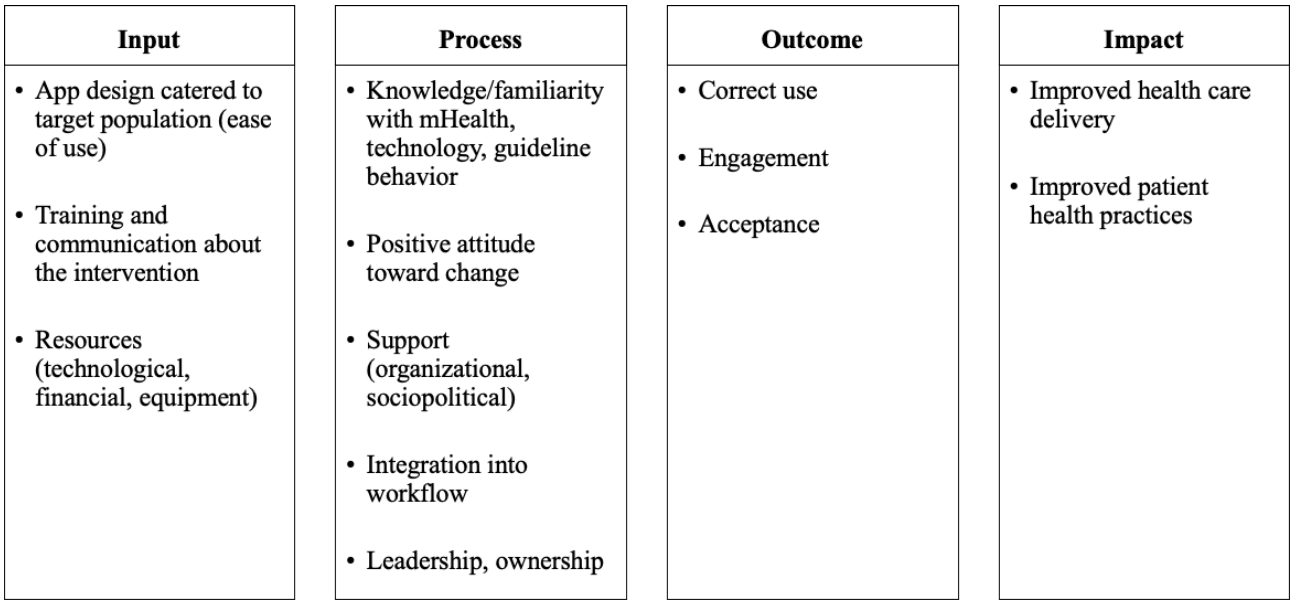


Figure 3. Logic diagram depicting determinants of successful mHealth implementation. Overview of mHealth implementation determinants from the mHealth design and resource input stage to the process of implementation and the desired outcomes and impact. mHealth: mobile health.



Acceptance of the Intervention

Overview

Acceptance of the intervention was strongly linked to uptake of the intervention, which facilitated implementation of the intervention and its potential to improve health outcomes. Patients and health workers were influenced by the perceived benefits of the intervention in improving health outcomes, the extent of training, and previous experience with the guideline behavior or mHealth.

Perceived Benefit of the Intervention

Perceived benefit of the intervention was influenced by ease of use and design of the intervention [9,10,30,32,34,35,37,41,42]. Where an app or SMS design engaged users in a way that improved health outcomes, patient engagement with health care services, or health care efficiency, participants were less likely to perceive the use of the mHealth technology as burdensome, facilitating successful implementation [9,29-32,34-36,38-41,43]. An intervention that provided a clear advantage compared to existing practice facilitated the perception of it being beneficial, as supported by the intervention's design and ease of use. These findings were consistent with Davis's Technology Acceptance Model and its application in health care, which suggests that perceived ease of use and usefulness both affect attitudes toward and use of technology [46,47]. Design considerations was an emergent factor unique to mHealth and technology-related interventions, emphasizing the value of iterative design of the mHealth intervention before implementation [9,10,30,34,35,37,41,42]. Further reviews reporting on mHealth in the context of antiretroviral therapy adherence and maternal health have identified the importance of intervention design, including tailoring SMS messages and the frequency of reminders [48,49].

Political support of the intervention ensured effective implementation; this included stakeholder buy-in and approval of the intervention, and support from health administrators on a district level [9,10,32]. The importance of understanding the need for the intervention to reduce disease burdens or improve health care services is supported by a previous review by Opoku, Stephani, and Quentin [50]. Therefore, regardless of the target user, perceived benefit of the intervention is crucial for implementation of the initiative, emphasizing the community-wide, integrated nature of mHealth interventions.

However, studies by Kaunda-Khangamwa et al and Mangam et al noted that, despite the users' positive attitudes toward the intervention, factors such as lack of communication regarding intervention use and limited resources to implement the behavioral guideline hindered implementation [8,30]. Perceived value of the intervention and acceptance alone therefore cannot guarantee successful implementation and positive outcomes from an intervention.

Training and Communication

Users who received communication and training on the intervention and its use before or during its implementation engaged well with the intervention and intended guideline practice [9,10,29,36]. Lack of awareness and clarity regarding how to engage with the intervention were consistently noted across studies with limited user engagement [30,33,43]. Sufficient training has been previously noted in a review as a contributing factor to mHealth intervention implementation, which is closely linked to the perceived ease of use [2]. For interventions aimed toward health workers, training was either provided by the intervention research team members or between health workers; follow-up on correct use was identified as being important for encouragement and continued intervention use [7,9,10,30,42].

Compatibility With Existing Health Care Practices and Social Norms

Familiarity With the Guideline Recommendation

Among studies that had high engagement with mHealth and improved health outcomes, existing familiarity with the guideline behavior, such as treatment adherence or HIV testing, was recurringly noted as a facilitator of implementation [7,37,39,41]. This suggests that successful mHealth initiatives complemented normative behavior, existing health care practices, and "new" practice (eg, interventions to improve treatment adherence), given their importance was already understood. This suggests mHealth has limited value in establishing new behavioral practices but rather is advantageous in complementing existing practices. Compatibility with existing behavior was strongly linked to the perceived benefit of the intervention; when the intended guideline was not yet an established practice or initially perceived as useful, mHealth interventions that encouraged this behavior were less likely to be accepted by the user [21,29,35,38,41]. Ide et al's study commented on how the mHealth intervention facilitated existing practices of childhood infectious disease management and improved reliability of the tests [29].

Although social norms (such as stigma) were only reported in 1 study [34], a number of studies commented on how users who were already comfortable sharing diagnoses or their health status were more engaged in the mHealth intervention, suggesting mHealth success is dependent on existing social norms and behaviors [34,39,43].

Attitudes and Familiarity With mHealth

Lack of familiarity with mobile phone use was a clear barrier to implementation; this barrier was particularly noted in studies with limited successes [8,43]. An existing understanding of the benefit of mHealth interventions—or previous positive experiences with mHealth—also affected user uptake. This was indirectly seen in mHealth intervention uptake being influenced by the perceived impact of use on provider-patient relationships [9,29,34]. This factor was conflicting across multiple studies, as Ide et al and Twimkukye et al commented on how the mHealth intervention facilitated perceived professionalism and improved provider-patient relationships, while Ellington et al noted the perceived decreased patient trust of the health worker to diagnose and treat the patient [9,29,34]. A systematic review of maternal health interventions also identified that technological literacy and previous experience of mHealth use were enablers of mHealth uptake, among a range of other factors [51,52].

Capacity of Existing Infrastructure and Resource Availability

The importance of capable infrastructure and resource availability have been noted by existing systematic reviews as key determinants of implementation [2,5,15,50-52]. Reviewed studies further confirmed this and noted the importance of staffing, network availability, technological support, and reliable technology in facilitating the implementation of the intervention [9,29,33,34,36-38,42,43]. These factors were tightly linked to social, political, and legal factors (such as limited staff funding or unstable network coverage to remote areas); these barriers often reflected the greater health care system's resourcefulness. Analysis of the factors influencing implementation therefore emphasized the interaction of health system components and the importance of considering the broader context beyond the health care system, as described in the systems thinking framework [27]. It is possible that due to the targeted, narrow nature of some of interventions, there was insufficient technological support or insufficient resources for successful implementation. Increased health data reporting on a health care system level could also contribute to improved resource allocation and policy decisions from sociopolitical organizations that could aid in mHealth implementation [53]. This further emphasizes the importance of understanding the value of the intervention in improving health outcomes across all stakeholders, as it could result in securing increased funding for the improved implementation of the initiative.

Further determinants of implementation included the system's capacity for change, such as how the intervention fit into existing appointment durations and organizational leadership structures [9,36]. This limited capacity for change could reflect the unstable foundation and support of the health care systems within these communities and indicates a potential lack of preparation for future changes or health challenges. Existing reviews on mHealth implementation have also commented on the importance of considering the existing health care system, such as government funding and capacity, when implementing mHealth interventions [2,15,16].

Strengths and Limitations

This review is valuable in its consideration of findings across a range of different LMIC settings in Africa and Asia, with a particular focus on periurban and rural areas. The diversity of study settings provides a broad range of factors to consider during implementation in different LMIC contexts. This review synthesized findings by drawing on a comprehensive health systems framework [25] and additional themes, further contributing to its novelty.

However, our review had several limitations. First, most studies (15/20) were conducted in Africa, limiting generalizability to other regions. Second, it was challenging to weigh the relative importance of each implementation determinant in each study, as the included studies were all qualitative. Regardless, the findings provide insights that quantitative results would not have been able to capture. Third, all studies were limited in that data were mainly from patients, health workers, administrators, or assistants, rather than from a sociopolitical level. Fourth, we did not source the primary quantitative impact evaluations of mHealth interventions. For further study, quantitative outcomes from impact studies can be better correlated with the specific determinants—and their respective strength of association—of implementation.

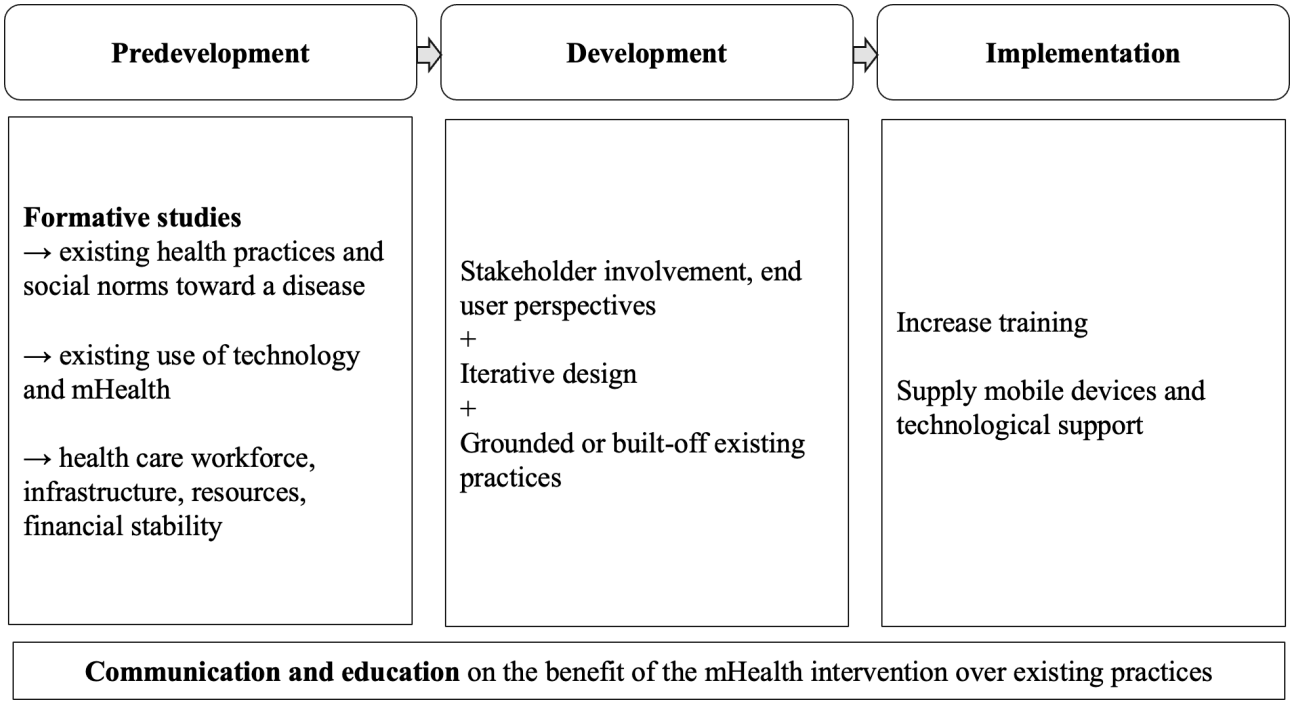
Recommendations

Overview

Insights from this review can help shape health policies and identify key considerations when developing mHealth interventions to improve their efficacy and sustainability in improving health outcomes. A full list of recommendations as reported in each study is noted in Multimedia Appendix 2.

When interpreting findings, it is important to consider the different contexts within which mHealth interventions are implemented, as they must be tailored to the context. Several considerations are important during predevelopmental, developmental, and implementation phases of the mHealth intervention, as seen in Figure 4.

Figure 4. Identified recommendations for successful implementation of mHealth interventions. mHealth: mobile health.



Predevelopmental Considerations

A thorough understanding of existing health and technology practices and social norms toward a disease are crucial to predict the feasibility of an mHealth intervention in a specific context. This understanding of the cultural context and structural factors—such as the broader health care workforce, capacity of existing infrastructure, and resource availability—can be achieved through formative research [42]. Communication and education about the importance of health care practices and mHealth's ability to facilitate health can aid in the acceptance of the mHealth intervention.

Development of the Intervention

There were 4 studies that performed prepilot testing and designed the intervention iteratively to maximize participant engagement with the intervention during implementation [7,8,31,42]. For example, through pilot testing, Mangam et al identified the need to incorporate interactive voice messaging in their SMS intervention, as many users were unable to read or understand the texts [8]. To complement the theoretical foundations of an intervention, studies emphasized the importance of stakeholder involvement in the iterative design process during the development of the initiative [2,8,15,16,19,20,42].

Implementation of the Intervention

As most interventions require a change in knowledge or behavior, it is important to communicate the benefit of the mHealth intervention compared to existing practice, prior to and during implementation. Increased training will increase confident user engagement with the intervention for long-term improvement of health outcomes [10,33,36]. Supplying mobile devices and offering technological support were recommended by studies to mitigate technological barriers [9,10,36].

It is important to tailor mHealth interventions to complement existing health services and face-to-face practices to optimize the desired health outcome. Adeagbo et al commented on how the mHealth intervention alone was insufficient in completely removing the barrier of accessing HIV testing and health care services [41]. Mangam et al discussed how future mHealth-based mobile communication should complement the community's existing familiarity with interpersonal communication, particularly as their SMS notifications of health prevention measures were not met with improvement compared to the non-mHealth, face-to-face status quo [8]. An example of complementing mHealth with non–technology-based communication is seen in the study by Bhattarai et al, who paired SMS text messaging with pamphlets [37].

Conclusion

This review provided comprehensive insight and an analysis of factors influencing the implementation of mHealth initiatives in LMICs. This review underscores the importance of iterative development of the intervention and deep consideration of the structural factors and cultural context before mHealth implementation to ensure scalability and sustainability to improve communicable health outcomes in LMICs.

Data Availability

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Authors' Contributions

JGO was responsible for the formal analysis, investigation, writing of the original draft, and visualization. HMY was responsible for the conceptualization, methodology, and review and editing of the draft. GF and JN were responsible for supervision and conceptualization.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search terms for databases.
[DOCX File, 18 KB - mhealth_v12i1e55189_app1.docx]

Multimedia Appendix 2
Reported gaps and recommendations.
[DOCX File, 19 KB - mhealth_v12i1e55189_app2.docx]

References

1. mHealth: New horizons for health through mobile technologies. Based on the findings of the second global survey on eHealth. Global Observatory for eHealth series - Volume 3. World Health Organization. 2012. URL: https://www.afro.who.int/publications/mhealth-new-horizons-health-through-mobile-technologie [accessed 2024-11-24]
2. Agarwal S, Perry HB, Long LA, Labrique AB. Evidence on feasibility and effective use of mHealth strategies by frontline health workers in developing countries: systematic review. Trop Med Int Health 2015 Aug;20(8):1003-1014. [doi: 10.1111/tmi.12525] [Medline: 25881735]
3. Alghamdi M, Gashgari H, Househ M. A systematic review of mobile health technology use in developing countries. Stud Health Technol Inform 2015;213:223-226. [Medline: 26152999]
4. Hall CS, Fottrell E, Wilkinson S, Byass P. Assessing the impact of mHealth interventions in low- and middle-income countries--what has been shown to work? Glob Health Action 2014;7:25606. [doi: 10.3402/gha.v7.25606] [Medline: 25361730]
5. Osei E, Mashamba-Thompson TP. Mobile health applications for disease screening and treatment support in low- and middle-income countries: a narrative review. Heliyon 2021 Mar;7(3):e06639. [doi: 10.1016/j.heliyon.2021.e06639] [Medline: 33869857]
6. WHO guideline: Recommendations on digital interventions for health system strengthening. World Health Organization. 2019. URL: https://www.who.int/publications/i/item/9789241550505 [accessed 2024-11-24]
7. Hodges J, Waldman AL, Koshkina O, et al. Process evaluation for the adaptation, testing and dissemination of a mobile health platform to support people with HIV and tuberculosis in Irkutsk, Siberia. BMJ Open 2022 Mar 29;12(3):e054867. [doi: 10.1136/bmjopen-2021-054867] [Medline: 35351714]
8. Mangam K, Fiekowsky E, Bagayoko M, et al. Feasibility and effectiveness of mHealth for mobilizing households for indoor residual spraying to prevent malaria: a case study in Mali. Glob Health Sci Pract 2016 Jun 20;4(2):222-237. [doi: 10.9745/GHSP-D-15-00381] [Medline: 27353616]
9. Ellington LE, Najjingo I, Rosenfeld M, et al. Health workers' perspectives of a mobile health tool to improve diagnosis and management of paediatric acute respiratory illnesses in Uganda: a qualitative study. BMJ Open 2021 Jul 19;11(7):e049708. [doi: 10.1136/bmjopen-2021-049708] [Medline: 34281930]
10. Ginsburg AS, Tawiah Agyemang C, Ambler G, et al. mPneumonia, an innovation for diagnosing and treating childhood pneumonia in low-resource settings: a feasibility, usability and acceptability study in Ghana. PLoS One 2016;11(10):e0165201. [doi: 10.1371/journal.pone.0165201] [Medline: 27788179]
11. Global strategy on digital health 2020-2025. World Health Organization. 2021. URL: https://www.who.int/docs/default-source/documents/gs4dhdaa2a9f352b0445bafbc79ca799dce4d.pdf [accessed 2024-11-01]
12. Measuring digital development: facts and figures 2021. : International Telecommunication Union; 2021 URL: https://www.itu.int/en/ITU-D/Statistics/Documents/facts/FactsFigures2021.pdf [accessed 2024-11-24]
13. Roser M, Ritchie H, Spooner F. Burden of disease. 2021. URL: https://ourworldindata.org/burden-of-disease [accessed 2024-11-24]
14. GBD 2019 Viewpoint Collaborators. Five insights from the Global Burden of Disease Study 2019. Lancet 2020 Oct 17;396(10258):1135-1159. [doi: 10.1016/S0140-6736(20)31404-5] [Medline: 33069324]
15. Aranda-Jan CB, Mohutsiwa-Dibe N, Loukanova S. Systematic review on what works, what does not work and why of implementation of mobile health (mHealth) projects in Africa. BMC Public Health 2014 Feb 21;14:188. [doi: 10.1186/1471-2458-14-188] [Medline: 24555733]
16. Chib A, van Velthoven MH, Car J. mHealth adoption in low-resource environments: a review of the use of mobile healthcare in developing countries. J Health Commun 2015;20(1):4-34. [doi: 10.1080/10810730.2013.864735] [Medline: 24673171]

17. Wallis L, Blessing P, Dalwai M, Shin SD. Integrating mHealth at point of care in low- and middle-income settings: the system perspective. Glob Health Action 2017 Jun;10(sup3):1327686. [doi: 10.1080/16549716.2017.1327686] [Medline: 28838302]
18. McCool J, Dobson R, Whittaker R, Paton C. Mobile health (mHealth) in low- and middle-income countries. Annu Rev Public Health 2022 Apr 5;43:525-539. [doi: 10.1146/annurev-publhealth-052620-093850] [Medline: 34648368]
19. Tomlinson M, Rotheram-Borus MJ, Swartz L, Tsai AC. Scaling up mHealth: where is the evidence? PLoS Med 2013;10(2):e1001382. [doi: 10.1371/journal.pmed.1001382] [Medline: 23424286]
20. Walsh JC, Groarke JM. Integrating behavioral science with mobile (mHealth) technology to optimize health behavior change interventions. Eur Psychol 2019 Jan;24(1):38-48. [doi: 10.1027/1016-9040/a000351]
21. Janssen R, Engel N, Pant Pai N, et al. "You're only there on the phone"? A qualitative exploration of community, affect and agential capacity in HIV self-testing using a smartphone app. Sociol Health Illn 2021 Mar;43(3):591-606. [doi: 10.1111/1467-9566.13242] [Medline: 33634889]
22. Kruse C, Betancourt J, Ortiz S, Valdes Luna SM, Bamrah IK, Segovia N. Barriers to the use of mobile health in improving health outcomes in developing countries: systematic review. J Med Internet Res 2019 Oct 9;21(10):e13263. [doi: 10.2196/13263] [Medline: 31593543]
23. Eze E, Gleasure R, Heavin C. Mobile health solutions in developing countries: a stakeholder perspective. Health Syst 2018 Apr 4;9(3):179-201. [doi: 10.1080/20476965.2018.1457134] [Medline: 32939258]
24. LMIC filters. Cochrane Effective Practice and Organisation of Care. 2020. URL: https://epoc.cochrane.org/lmic-filters [accessed 2024-11-01]
25. Flottorp SA, Oxman AD, Krause J, et al. A checklist for identifying determinants of practice: a systematic review and synthesis of frameworks and taxonomies of factors that prevent or enable improvements in healthcare professional practice. Impl Sci 2013 Mar 23;8:35. [doi: 10.1186/1748-5908-8-35] [Medline: 23522377]
26. Wensing M. The Tailored Implementation in Chronic Diseases (TICD) project: introduction and main findings. Impl Sci 2017 Jan 10;12(1):5. [doi: 10.1186/s13012-016-0536-x] [Medline: 28069029]
27. Systems thinking for health systems strengthening. Alliance for Health Policy and Systems Research. 2009. URL: https://iris.who.int/bitstream/handle/10665/44204/9789241563895_eng.pdf?sequence=1 [accessed 2024-11-01]
28. World bank country and lending groups. World Bank. 2022. URL: https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups [accessed 2024-11-01]
29. Ide N, Hardy V, Chirambo G, et al. People welcomed this innovation with two hands: a qualitative report of an mHealth intervention for community case management in Malawi. Ann Glob Health 2019 Apr 25;85(1):61. [doi: 10.5334/aogh.919] [Medline: 31025838]
30. Kaunda-Khangamwa BN, Steinhardt LC, Rowe AK, et al. The effect of mobile phone text message reminders on health workers' adherence to case management guidelines for malaria and other diseases in Malawi: lessons from qualitative data from a cluster-randomized trial. Malar J 2018 Dec 19;17(1):481. [doi: 10.1186/s12936-018-2629-2] [Medline: 30567603]
31. Nhavoto JA, Grönlund Å, Klein GO. Mobile health treatment support intervention for HIV and tuberculosis in Mozambique: perspectives of patients and healthcare workers. PLoS One 2017;12(4):e0176051. [doi: 10.1371/journal.pone.0176051] [Medline: 28419149]
32. Babili A, Nsanzimana S, Rwagasore E, Lester RT. SMS-based digital health intervention in Rwanda's home-based care program for remote management of COVID-19 cases and contacts: a qualitative study of sustainability and scalability. Front Digit Health 2022;4:1071790. [doi: 10.3389/fdgth.2022.1071790] [Medline: 36714610]
33. Haberer JE, Kiwanuka J, Nansera D, Wilson IB, Bangsberg DR. Challenges in using mobile phones for collection of antiretroviral therapy adherence data in a resource-limited setting. AIDS Behav 2010 Dec;14(6):1294-1301. [doi: 10.1007/s10461-010-9720-1] [Medline: 20532605]
34. Twimukye A, Bwanika Naggirinya A, Parkes-Ratanshi R, et al. Acceptability of a mobile phone support tool (Call for Life Uganda) for promoting adherence to antiretroviral therapy among young adults in a randomized controlled trial: exploratory qualitative study. JMIR mHealth uHealth 2021 Jun 14;9(6):e17418. [doi: 10.2196/17418] [Medline: 34121665]
35. Jones COH, Wasunna B, Sudoi R, Githinji S, Snow RW, Zurovac D. "Even if you know everything you can forget": health worker perceptions of mobile phone text-messaging to improve malaria case-management in Kenya. PLoS One 2012;7(6):e38636. [doi: 10.1371/journal.pone.0038636] [Medline: 22719911]
36. Win Han O, Kaung Myat T, Cutts JC, Win H, Kyawt Mon W, May Chan O, et al. Sustainability of a mobile phone application-based data reporting system in Myanmar's malaria elimination program: a qualitative study. BMC Med Inform Decis Mak 2021 Oct 18;21(1):285. [doi: 10.1186/s12911-021-01646-z] [Medline: 34663306]
37. Bhattarai AH, Sanjaya GY, Khadka A, Kumar R, Ahmad RA. The addition of mobile SMS effectively improves dengue prevention practices in community: an implementation study in Nepal. BMC Health Serv Res 2019 Oct 15;19(1):699. [doi: 10.1186/s12913-019-4541-z] [Medline: 31615484]
38. Mohammed S, Siddiqi O, Ali O, et al. User engagement with and attitudes towards an interactive SMS reminder system for patients with tuberculosis. J Telemed Telecare 2012 Oct;18(7):404-408. [doi: 10.1258/jtt.2012.120311] [Medline: 23034935]

39. Venables E, Ndlovu Z, Munyaradzi D, et al. Patient and health-care worker experiences of an HIV viral load intervention using SMS: a qualitative study. PLoS ONE 2019;14(4):e0215236. [doi: 10.1371/journal.pone.0215236] [Medline: 30973925]
40. Straw C, Sanchez-Antelo V, Kohler R, Paolino M, Viswanath K, Arrossi S. Implementation and scaling-up of an effective mHealth intervention to increase adherence to triage of HPV-positive women (ATICA study): perceptions of health decision-makers and health-care providers. BMC Health Serv Res 2023 Jan 18;23(1):47. [doi: 10.1186/s12913-023-09022-5] [Medline: 36653775]
41. Adeagbo O, Kim HY, Tanser F, et al. Acceptability of a tablet-based application to support early HIV testing among men in rural KwaZulu-Natal, South Africa: a mixed method study. AIDS Care 2021 Apr;33(4):494-501. [doi: 10.1080/09540121.2020.1742867] [Medline: 32172596]
42. Chaiyachati KH, Loveday M, Lorenz S, et al. A pilot study of an mHealth application for healthcare workers: poor uptake despite high reported acceptability at a rural South African community-based MDR-TB treatment program. PLoS One 2013;8(5):e64662. [doi: 10.1371/journal.pone.0064662] [Medline: 23724075]
43. Maraba N, Hoffmann CJ, Chihota VN, et al. Using mHealth to improve tuberculosis case identification and treatment initiation in South Africa: results from a pilot study. PLoS ONE 2018;13(7):e0199687. [doi: 10.1371/journal.pone.0199687] [Medline: 29969486]
44. O'Connor Y, Hardy V, Heavin C, Gallagher J. Supporting LIFE: Mobile Health Application for Classifying, Treating and Monitoring Disease Outbreaks of Sick Children in Developing Countries: Springer International Publishing; 2015:366-370. [doi: 10.1007/978-3-319-18714-3_24]
45. Pai NP, Esmail A, Marathe G, et al. 1275. Will an app-optimized HIV self-testing strategy work for South Africans? Results from a large cohort study. Open Forum Infect Dis 2018 Nov 26;5(suppl_1):S388-S388. [doi: 10.1093/ofid/ofy210.1108]
46. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319. [doi: 10.2307/249008]
47. Holden RJ, Karsh BT. The technology acceptance model: its past and its future in health care. J Biomed Inform 2010 Feb;43(1):159-172. [doi: 10.1016/j.jbi.2009.07.002] [Medline: 19615467]
48. Lee SB, Valerius J. mHealth interventions to promote anti-retroviral adherence in HIV: narrative review. JMIR mHealth uHealth 2020 Aug 28;8(8):e14739. [doi: 10.2196/14739] [Medline: 32568720]
49. Sowon K, Maliwichi P, Chigona W. The influence of design and implementation characteristics on the use of maternal mobile health interventions in Kenya: systematic literature review. JMIR mHealth uHealth 2022 Jan 27;10(1):e22093. [doi: 10.2196/22093] [Medline: 35084356]
50. Opoku D, Stephani V, Quentin W. A realist review of mobile phone-based health interventions for non-communicable disease management in sub-Saharan Africa. BMC Med 2017 Feb 6;15(1):24. [doi: 10.1186/s12916-017-0782-z] [Medline: 28162090]
51. Ag Ahmed MA, Gagnon MP, Hamelin-Brabant L, Mbemba GIC, Alami H. A mixed methods systematic review of success factors of mhealth and telehealth for maternal health in Sub-Saharan Africa. Mhealth 2017;3:22. [doi: 10.21037/mhealth.2017.05.04] [Medline: 28736731]
52. Feroz AS, Ali NA, Khoja A, Asad A, Saleem S. Using mobile phones to improve young people sexual and reproductive health in low and middle-income countries: a systematic review to identify barriers, facilitators, and range of mHealth solutions. Reprod Health 2021 Jan 16;18(1):9. [doi: 10.1186/s12978-020-01059-7] [Medline: 33453723]
53. Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. Health Inf Sci Syst 2014;2:3. [doi: 10.1186/2047-2501-2-3] [Medline: 25825667]

Abbreviations

LMIC: low- and middle-income countries
mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
TICD: Tailored Implementation for Chronic Diseases
WHO: World Health Organization

Edited by L Buis; submitted 08.12.23; peer-reviewed by MH Su, X Yu; revised version received 02.07.24; accepted 06.09.24; published 13.12.24.

Please cite as:
Greenall-Ota J, Yapa HM, Fox GJ, Negin J
Qualitative Evaluation of mHealth Implementation for Infectious Disease Care in Low- and Middle-Income Countries: Narrative Review
JMIR Mhealth Uhealth 2024;12:e55189
URL: https://mhealth.jmir.org/2024/1/e55189
doi:10.2196/55189

Original Paper

Effectiveness and Implementation of a Text Messaging mHealth Intervention to Prevent Childhood Obesity in Mexico in the COVID-19 Context: Mixed Methods Study

Ana Lilia Lozada-Tequeanes[1], PhD; Florence L Théodore[1], PhD; Edith Kim-Herrera[1], MSc; Armando García-Guerra[1], MSc; Amado D Quezada-Sánchez[1], MSc; Rocio Alvarado-Casas[1], MPH; Anabelle Bonvecchio[1], PhD

National Institute of Public Health, Instituto Nacional de Salud Pública de México, Cuernavaca, Mexico

Corresponding Author:
Florence L Théodore, PhD
National Institute of Public Health
Instituto Nacional de Salud Pública de México
Avenida Universidad 655
Cuernavaca, 62100
Mexico
Phone: 52 +527773293000
Email: ftheodore@insp.mx

Abstract

Background: Promoting physical activity (PA) and healthy feeding (HF) is crucial to address the alarming increase in obesity rates in developing countries. Leveraging mobile phones for behavior change communication to encourage infant PA and promote HF is particularly significant within the Mexican context.

Objective: This study aims to explore the effectiveness and feasibility of mHealth interventions aimed at promoting PA and HF among primary caregivers (PCs) of Mexican children under the age of 5 years. Additionally, the study aims to disseminate insights gained from intervention implementation amidst the COVID-19 pandemic and assess the potential of behavior change mHealth interventions on a broader population scale.

Methods: NUTRES, an mHealth intervention, underwent an effectiveness-implementation hybrid trial. Over 36 weeks, participants in the intervention group (IG), totaling 230 individuals, received approximately 108 SMS text messages tailored to their children's age. These messages covered topics such as PA and HF and emphasized the significance of proper child nutrition amidst the COVID-19 pandemic. NUTRES participants were recruited from both urban and rural health units across 2 states in Mexico. Given the COVID-19 context, both baseline and follow-up surveys were conducted via mobile or fixed telephone. The evaluation of effectiveness and implementation used a mixed methods approach. Qualitative analysis delved into participants' experiences with NUTRES and various implementation indicators, including acceptance, relevance, and coverage. Grounded theory was used for coding and analysis. Furthermore, difference-in-differences regression models were used to discern disparities between groups (comparison group [CG] versus IG) concerning knowledge and practices pertaining to infant PA and HF.

Results: Of the total 494 PCs enrolled in NUTRES, 334 persisted until the end of the study, accounting for 67.6% (334/494) participation across both groups. A majority of PCs (43/141, 30.5%, always; and 97/141, 68.8%, sometimes) used the SMS text message information. Satisfaction and acceptability toward NUTRES were notably high, reaching 98% (96/98), with respondents expressing that NUTRES was "good," "useful," and "helpful" for enhancing child nutrition. Significant differences after the intervention were observed in PA knowledge, with social interaction favored (CG: 8/135, 5.9% vs IG: 20/137, 14.6%; *P*=.048), as well as in HF practice knowledge. Notably, sweetened beverage consumption, associated with the development of chronic diseases, showed divergence (CG: 92/157, 58.6% vs IG: 110/145, 75.9%; *P*=.003). In the difference-in-differences model, a notable increase of 0.03 in knowledge regarding the benefits of PA was observed (CG: mean 0.13, SD 0.10 vs IG: mean 0.16, SD 0.11; *P*=.02). PCs expressed feeling accompanied and supported, particularly amidst the disruption of routine health care services during the COVID-19 pandemic.

Conclusions: While NUTRES exhibited a restricted impact on targeted knowledge and behaviors, the SMS text messages functioned effectively as both a reminder and a source of new knowledge for PCs of Mexican children under 5 years of age. The key lessons learned were as follows: mHealth intervention strategies can effectively maintain communication with individuals during emergencies, such as the COVID-19 pandemic; methodological and implementation barriers can constrain the effectiveness

of mHealth interventions; and using mixed methods approaches ensures the complementary nature of results. The findings contribute valuable evidence regarding the opportunities and constraints associated with using mobile phones to enhance knowledge and practices concerning PA and HF among PCs of children under 5 years old.

Trial Registration: ClinicalTrials.gov NCT04250896; https://clinicaltrials.gov/ct2/show/NCT04250896

(JMIR Mhealth Uhealth 2024;12:e55509) doi:10.2196/55509

KEYWORDS

effectiveness; feasibility; mHealth; SMS text message; mixed methods; infant obesity; physical activity; healthy feeding; children; COVID-19; Mexico

Introduction

Because of the short and long-term adverse effects of overweight/obesity on health and human capital, the World Health Organization (WHO) emphasizes its prevention early in life as a critical priority [1]. There is mounting evidence indicating that children from low socioeconomic backgrounds exhibit higher rates of overweight/obesity [2,3]. Mexico is particularly alarmed by the escalating prevalence of overweight/obesity among children under 5 years old, with the most recent national estimates projecting a combined rate of approximately 8% in 2022 [4]. Nevertheless, socioeconomically disadvantaged families are often more challenging for the health sector to engage with, and they may be less inclined to participate in programs promoting healthy behaviors [5]. Mobile health (mHealth) technologies and telecommunications present themselves as appealing low-cost interventions capable of reaching vast and remote populations [6].

In low- and middle-income countries [7], resources have been directed toward the development of mHealth interventions, including behavior change communication (BCC) strategies [8]. A 2019 systematic review [9] revealed that the most commonly used mHealth technology was SMS text messaging (60%). Additionally, there is evidence suggesting that eHealth and mHealth interventions are effective in promoting physical activity (PA) and healthy feeding (HF) in developing countries [10].

Hence, leveraging mobile phones for BCC interventions presents a promising opportunity to advocate for HF and PA practices among primary caregivers (PCs) of the Mexican child population. This is especially significant considering the rising and widespread use and accessibility of this technology, which has increased from 71.5% of the total population in 2015 to 79.2% in 2022 [11]. In particular, in emergency contexts where health services are disrupted and face-to-face information dissemination to the population is hindered, as was evident during the COVID-19 pandemic, leveraging innovative mHealth interventions becomes crucial. In this regard, testing and evaluating such interventions in emergency settings can significantly contribute to the existing evidence on mHealth. Moreover, it can serve to stimulate further research endeavors to complement and expand upon the findings of this study.

Hence, this research aimed to assess the effectiveness and implementation of an mHealth intervention, referred to as NUTRES, in promoting PA and HF practices among Mexican PCs of children under 5 years. A secondary objective was to document the lessons learned from implementing and evaluating an mHealth intervention during the COVID-19 pandemic.

Methods

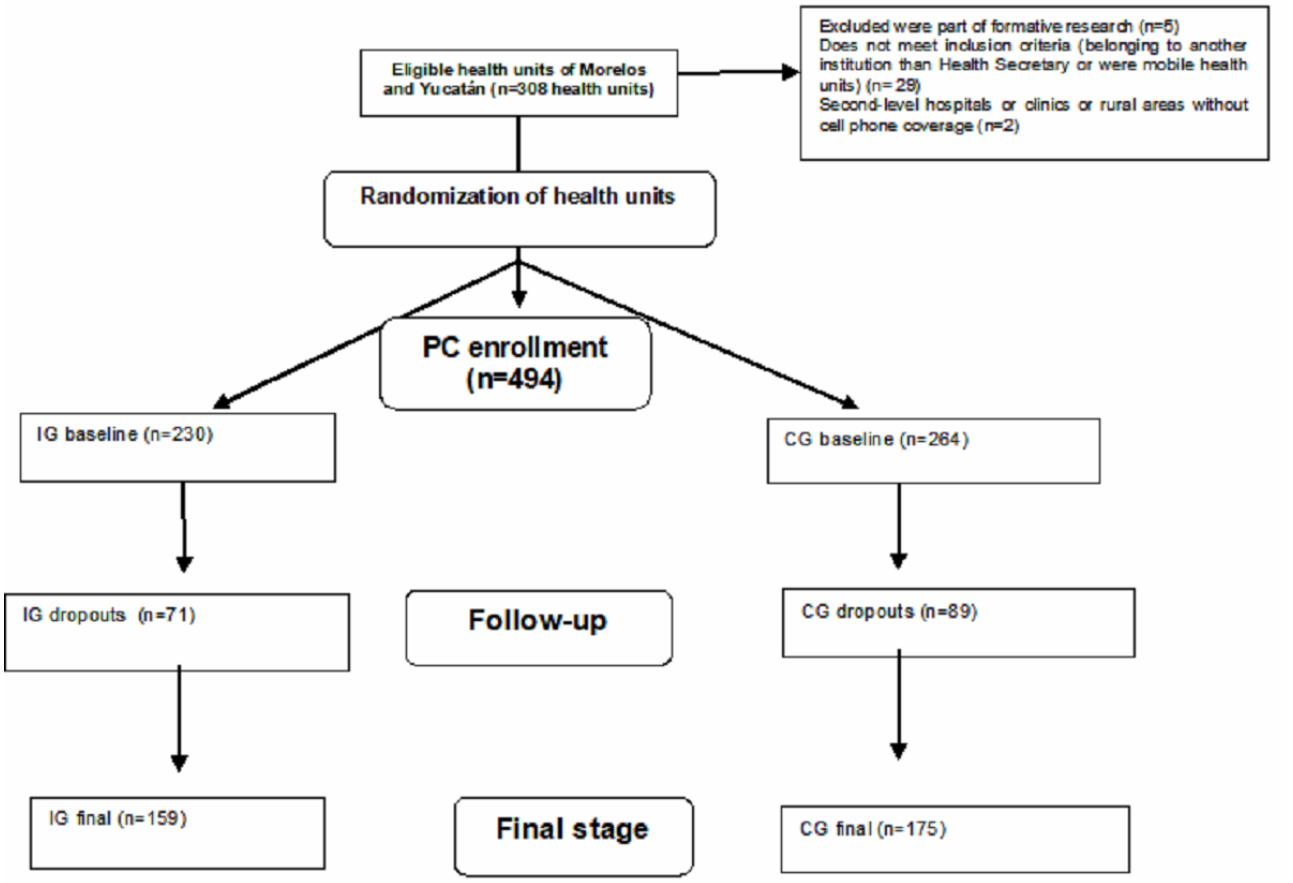
Study Design

This study constitutes an effectiveness-implementation hybrid trial [12], incorporating elements from both effectiveness and implementation research. The trial is registered under the Trial Registration ID NCT04250896. We used a mixed methods approach [13], utilizing a “convergent advanced design” [14]. This design involved the simultaneous use of qualitative and quantitative methods, allowing for the triangulation of results from both methodologies. By combining statistical findings with insights gleaned from individuals’ real-life experiences, we aimed to gain a comprehensive understanding of the effects and implementation process of NUTRES.

Eligibility Criteria and Recruitment of Participants

NUTRES participants were recruited from urban and rural health units situated in 2 states of Mexico: Morelos in the central region and Yucatán in the southern region. Randomization was conducted at the health unit level. Inclusion criteria comprised being a primary health care unit, being located in an area with access to a mobile phone network, having more than 50 registered users under 5 years, and having over 80% of the population with Spanish as their first language. A total of 308 eligible urban and rural health units from the 2 states were included in the study, assigned either to the intervention group (IG) or the comparison group (CG) (Figure 1).

Figure 1. Flow diagram of NUTRES participants. CG: comparison group; IG: intervention group; PC: primary caregiver.



The eligibility criteria for participants in both the IG and the CG were being the PC of an infant under 5 years, having ownership/access to any type of mobile phone, being able to speak and read Spanish, being aged 18 years or older, and being a resident in the coverage area of the participating health unit. Participants were recruited through 3 methods: (1) invitation from primary health providers, (2) snowball sampling, and (3) face-to-face encounters by field workers at health units. The CG had the same eligibility criteria for recruitment, except for the requirement of mobile phone ownership/access.

Sample Size

Quantitative

Sample size calculations indicated that 100 participants per study group would yield 80% power at the .05 significance level to detect a 20%-point difference in the proportion of change for at least one outcome variable (eg, knowledge, attitudes, or practices related to infant PA or HF) between study groups, using a 2-sided test. These calculations accounted for a design effect ranging between 1.5 and 1.9 [15] and factored in a 20% dropout rate [16]. Notably, this sample size estimation did not consider states or areas as strata, thus precluding intergroup comparisons based on geographical regions.

Qualitative

The sample selection was purposive, with the aim of capturing a diverse range of experiences associated with NUTRES. Informants were chosen from both rural and urban areas, encompassing caregivers with children aged 0-23 months and 2-5 years, representing varying levels of interactivity. The use of 2-way SMS text messaging served as an indicator of their engagement with NUTRES (details on interactivity provided below).

NUTRES Intervention

NUTRES is an mHealth BCC strategy designed to prevent childhood overweight/obesity. It achieves this by disseminating SMS text messages to PCs of children under 5 years, as well as to health personnel operating within primary unit services in 2 Mexican states. In this paper, we will primarily concentrate on the outcomes of the SMS text messaging interventions aimed at promoting infant PA [17] and HF [18]. For a comprehensive understanding of the development of NUTRES, readers are referred to previous publications [19,20].

In essence, the NUTRES intervention comprised 1- or 2-way SMS text messages, each containing fewer than 150 characters, disseminated to PCs over 9 months (equivalent to 36 weeks). Participants in the IG received approximately 108 SMS text messages, adjusted based on the child's age. These SMS text messages addressed various aspects, including attitudes and practices concerning infant PA and HF. Additionally, SMS text messages addressing nutrition within the context of the COVID-19 pandemic were incorporated due to the emergency situation. The design of the SMS text message content was informed by extensive formative research [21], which aimed to identify both barriers and facilitators to the adoption of healthy practices within the target population. Every factor that could either facilitate or hinder behavior change was systematically

addressed in line with the Theory of Planned Behavior [22,23]. This was accomplished through the dissemination of truthful information, practical tips, healthy recipes, challenges, and socioemotional support messages to PCs. All SMS text messages were meticulously aligned with the latest national [24] and international guidelines concerning infant PA and HF [17]. See Multimedia Appendix 1 for examples of SMS text messages delivered.

PCs who consented to participate in NUTRES were asked to provide details about the child's name, age, and gender during registration. This information was used to personalize and tailor SMS text messages for each participant accordingly. NUTRES incorporated 2 types of SMS text messages: 1- and 2-way (the latter being sent following a response from PCs to the initial SMS text messages). On average, participants received approximately 3 SMS text messages per week. In addition to informational messages, a weekly positive socioemotional support SMS text message was dispatched to PCs, as research has demonstrated its effectiveness in fostering participation and maintaining interest [25]. These SMS text messages were programmed to be automatically delivered to PCs via the Rapid Pro platform [26], and interactivity, including responses from PCs, was recorded and tracked.

Data Collection and Analysis

Overview

Data collection took place from September 2020 to September 2021. The quantitative component encompassed both the IG and the CG, involving both baseline and final assessments. Meanwhile, the qualitative component solely focused on the IG, conducted after 36 weeks of exposure to NUTRES. Both the qualitative and quantitative teams, consisting of approximately 4 members each, possessed extensive fieldwork experience and proficiency in communication technology. They underwent a week-long virtual training session via Zoom (Zoom Video Communications, Inc.).

Quantitative Data

Because of the COVID-19 context, both baseline and follow-up surveys were conducted via mobile or fixed telephone. The surveys covered a range of topics, including sociodemographic information, health status, and knowledge pertaining to infant PA and HF. Knowledge of recommended practices was assessed by querying respondents about the advantages, disadvantages, or known recommendations regarding PA and HF. Each answer was assigned a score based on its correctness, as outlined in Multimedia Appendix 2.

Furthermore, exposure to NUTRES was gauged by querying PCs regarding their receipt of the SMS text messages, as well as their perception of the usefulness and practicality of the messages. For instance, PCs were asked if they recall a specific message from NUTRES that they had implemented in practice, and their responses were recorded without any prompts or suggestions from the interviewer.

Data collection was conducted using the REDCap (Research Electronic Data Capture; Vanderbilt University) app, with daily verification and backup procedures in place. Following data collection, thorough exploration, cleaning, and recategorization processes were undertaken to prepare for descriptive statistics and basic comparative analyses between study groups. Topics of interest were examined and translated into knowledge and practice indices, with summary variables generated for each question. Practices were categorized into 3 equally sized tertiles for evaluation: the first tertile was termed "limited," the intermediate tertile was termed "moderate," and the tertile with higher values was termed "adequate." Lastly, we used a double-difference approach [27] to evaluate the variance in the change of outcomes between the presence (IG) and absence (CG) of the intervention. A significance level of $P<.05$ was deemed significant. All statistical analyses adhered to an "intention-to-treat" principle [28] and were conducted using the Stata version 14.2 statistical package (StataCorp).

Qualitative Data

Interviews were conducted after the quantitative survey and delved into various topics, including opinions regarding NUTRES; technical issues encountered with receiving, reading, and responding to SMS text messages; overall impressions of the SMS text message content; behaviors encouraged by the messages; and the perceived impact of the intervention on the intention or ability to carry out recommended practices. Additionally, we explored 3 implementation indicators with PCs from the IG:

- "Acceptance" of NUTRES and behaviors promoted by SMS text messages.
- "Pertinency," which examines the relevance of information and behaviors promoted by the SMS text messages in NUTRES.
- "Coverage," which examines the extent to which PCs received the SMS text messages.

The first 2 implementation indicators were proposed by Proctor and colleagues [29], while the last one was by Peters and colleagues [30]. According to these authors, these indicators are crucial determinants of the success, in terms of both implementation and expected outcomes, of an intervention.

The interview guide underwent a pilot phase involving role-play to ensure its smooth flow and enable interviewers to practice its implementation. Telephone interviews typically lasted between 40 and 80 minutes, and they were recorded and transcribed verbatim. Following the principles of grounded theory [31], data analysis was conducted using a coding tree [31] to streamline the coding process. This analysis was performed utilizing NVivo 2020 (QSR International). See Multimedia Appendix 3 for the category tree used for coding.

Ethical Approval and Consent

This study received approval from the Ethics, Research, and Biosafety Committees of the National Institute of Public Health (INSP) of Mexico, with the reference number CI 1547. All participants provided verbal informed consent to take part in the study.

Results

General Data About Design, Participants, and Sample Size

Quantitative

The total number of participants at baseline was 494 PCs, with 230 in the IG and 264 in the CG. Out of these, a total of 334 participants completed the study, comprising 118 from Morelos and 216 from Yucatán (Figure 1) [32]. The primary reasons for the loss to follow-up were the inability to locate the PC for final measurement, accounting for 9.1% (20/220) from the IG and 18.1% (41/227) from the CG, and the unwillingness of some participants to continue with the research, constituting 4.3% (8/187) from the IG.

The baseline characteristics of PCs revealed that the majority were young, with an average age of approximately 28 years. Most PCs were married or cohabiting and had a low level of education, with less than one-quarter having completed primary education. Additionally, at baseline, it was observed that the children of PCs from the CG were on average 3.7 months older than those from the IG (P<.001). Furthermore, PCs from the CG were less likely to report having a paid job compared with PCs from the IG (P<.004; Table 1). Baseline characteristics were found to be similar between PCs who continued in the study and those who were lost to follow-up. This is detailed in Multimedia Appendix 4, which presents the descriptive characteristics of individuals lost to follow-up versus those who completed the study.

Table 1. Sociodemographic characteristics of the primary caregivers (N=494) at baseline and final line, by study group (NUTRES, 2020-21).

Sociodemographic characteristics	Baseline		
	Comparison group (n=264)	Intervention group (n=230)	*P* value
States of Mexico, n (%)			.80
Morelos	90 (34.1)	88 (38.3)	
Yucatán	174 (65.9)	142 (61.7)	
Area, n (%)			.86
Urban	162 (61.4)	148 (64.3)	
Rural	102 (38.6)	82 (35.7)	
Age of primary caregiver, mean (SD)	27.5 (7.25)	27.1 (6.80)	.49
Relationship to the child, n (%)			.49
Mother	259 (98.1)	226 (98.3)	
Grandmother	1 (0.4)	2 (0.9)	
Aunt/sister	3 (1.1)	1 (0.4)	
Father/grandfather	0 (0)	1 (0.4)	
Sex of child, n (%)			.44
Girl	126 (47.7)	104 (45.2)	
Boy	138 (52.3)	125 (54.3)	
Age of child (months), mean (SD)	20.5(16.8)	16.8 (15.7)	.001
<24, n (%)	165 (62.5)	164 (71.3)	.12
24-59, n (%)	99 (37.5)	66 (28.7)	
Marital status (married/free union), n (%)	232 (87.9)	191 (83.0)	.33
Schooling (basic or less), n (%)	10 (3.8)	27 (11.7)	.35
Employment with payment (last week) (yes), n (%)	22 (8.3)	40 (17.4)	.004
Socioeconomic status[a], n (%)			.19
Tertile 1	105 (39.8)	72 (31.3)	
Tertile 2	90 (34.1)	67 (29.1)	
Tertile 3	69 (26.1)	91 (39.6)	
Beneficiary/affiliation social program (yes), n (%)	64 (24.2)	61 (26.5)	.51

[a]Socioeconomic status (tercile 1 represents the lowest welfare conditions).

Qualitative

Twenty-four PCs from the IG were interviewed, with 12 participants from each state. These PCs had an average age of 30.2 years, with 14 participants having completed high school education or higher. The majority of interviewed PCs were homemakers, totaling 16 individuals. Multimedia Appendix 5 provides further details on the main characteristics of PCs interviewed for the NUTRES study conducted between 2020 and 2021. Interestingly, only a few differences were observed in the experiences reported by PCs of children aged 0-23 months and those aged 24-59 months, as well as between rural and urban areas. Consequently, qualitative findings and PC experiences are presented in a general manner.

NUTRES Intervention Implementation

PCs expressed appreciation for the SMS text messages from NUTRES, noting that they valued their brevity and clarity. Informants unanimously agreed that the NUTRES strategy was "good," "useful," and "helpful." PCs highlighted that NUTRES served as a "reminder" of previously acquired knowledge, while also introducing "new" information, particularly regarding the promotion of PA. PCs also remarked that the changing topics in the SMS text messages as their children grew made them feel supported and motivated to implement the recommendations, highlighting the acceptability and pertinence of the intervention (Textbox 1; Multimedia Appendix 6). Additionally, an overwhelming majority of PCs (96/98, 98%), expressed their desire to continue receiving NUTRES SMS text messages. However, some PCs raised concerns about sharing sensitive information, such as personal data, during the registration phase. Furthermore, a portion of PCs, specifically 32/145 (22%), reported encountering barriers to receiving SMS text messages from NUTRES. These barriers were damaged equipment (11/32, 34%) and issues related to connectivity or lack of mobile phone credit (7/32, 22%), indicating challenges in coverage (Table 2). Additionally, PCs indicated that they consistently use (43/141, 30.5%) and interact (27/140, 19.3%) with the 2-way SMS text message feature. Testimonies revealed that low interaction by PCs did not necessarily correspond to a lack of interest or rejection of NUTRES but rather to external circumstances such as a lack of credit or signal to receive SMS text messages, or a lack of awareness about the reply option (Textbox 1).

Textbox 1. Implementation indicators (English quotes) of NUTRES by primary caregivers exposed to NUTRES.

Acceptance

- Well, I don't know, I say how....one year, I don't know...because I still don't feel as prepared for...as for...well...yes, well yes, I do feel prepared for my baby, but I still don't know how many things, so it does help me [#14]
- They make me feel calm [the messages]. (S) how do I explain it? They make me feel calm because they are helping me" (...) and it helped me a lot with the advice is reaching me (...) [#15].
- Mmm, well, it's a program that helps us, helps us complement what we already do at home for the children [#24].

Pertinency

- Since I am a first-timer, I felt that it helped me in the sense that, like right now, because of the Covid, there is no need for you to take, for example, to your health clinic, so there they told you how the processes that touched them...in what month could you give certain foods to your baby [#5].
- Yes, because they were at the age, they were perfectly fine at my son's age. Hey? forever! [#13].

Coverage

- Good. Well, in fact, we had to answer "Yes, No", or if the goal was achieved or not, I tell him. There were times that the same and maybe I did not realize some messages because there was no coverage, they did not enter. Sometimes several days passed, and until I went out to some place that did exist, I was aware of the messages [#23]
- And I told him "it's that they don't reach me", that only when I go to a higher part or I left here, they began to reach me. And I tell him, when I'm away, well, yes, I answer them, yes or...like no or something, well, the things they asked me [#21]
- Here in the town the current fails a lot, and the signal goes away...it's been like [P] four months now that the current goes out constantly here in the town and it's in parts [#6].
- Well, almost all of them because it was at the beginning when I began to receive the messages that they were going to charge me. And then, well, this...well, my recharge ran out very quickly, so to speak, and I preferred to just read them and put them into practice...Sometimes I would send them and say no, that I had to charge I don't know what...things like of money or something like that, and others if they let them answer [#14].
- What happens, that as there was a problem, that the telephones were lost. There came a time when this number that I have, my husband had it, and since we removed the chip, it disappeared [#5].

Table 2. Reception and reading of SMS text messages and the most useful format to put them into practice in primary caregivers participating in NUTRES.

Responses	Values (n=145)[a]
Difficulties receiving NUTRES SMS text messages (n=145), n (%)	32 (22.1)
Reading of NUTRES SMS text messages (n=144), n (%)	
Always	96 (66.7)
Sometimes	45 (31.3)
Use of NUTRES SMS text messages (n=141), n (%)	
Always	43 (30.5)
Sometimes	97 (68.8)
Interactivity of NUTRES SMS text messages (n=140), n (%)	
Always	27 (19.3)
Sometimes	71 (50.7)
PCs[b] reported barriers to receiving NUTRES SMS text messages (n=32), n/N (%)	32/145 (22.1)
Dropped telephone line, n (%)	3 (9.4)
Lack of telephone signal, n (%)	7 (21.9)
Change of phone number, n (%)	1 (3.1)
Lack own mobile phone equipment, n (%)	3 (9.4)
Damaged equipment, n (%)	11 (34.4)
Lack of connectivity or mobile phone credit, n (%)	7 (21.9)
Problem or difficulty was solved (yes), n (%)	11 (34.4)
Reasons for not reading NUTRES SMS text messages (n=3), n/N (%)	3/145 (2.1)
Other (damaged or missing equipment), n (%)	3 (100.0)
Reasons PCs did not implement NUTRES SMS text messages (n=1), n/N (%)	1/145(0.7)
Lack of time, n (%)	1 (100.0)
Reasons PCs did not respond to NUTRES SMS text messages (n=42), n/N (%)	42/145 (29.0)
Lack of time and participants did not respond to SMS text messages within 12 hours, n (%)	7 (16.7)
It is complicated, many issues, n (%)	1 (2.4)
Information is missing, n (%)	2 (4.8)
Other (eg, lack of own equipment, the economic burden for payment of phone line, error when sending responses, read SMS text messages later), n (%)	31 (73.8)
Does not know/no comments, n (%)	1 (2.4)
Perception of NUTRES (satisfied and very satisfied) (n=141), n/N (%)	141/145 (97.2)
Because is practical, n (%)	92 (65.2)
A good resource in times of COVID-19, n (%)	10 (7.1)
Received interesting messages, n (%)	83 (58.9)
Other (eg, because of the interest they show toward the family, important messages for the health of babies), n (%)	7 (5.0)
Reasons for dissatisfaction or some dissatisfaction with NUTRES (n=3), n/N (%)	3/145 (2.1)
Unclear or confusing information, n/N (%)	1/145 (0.7)
Relevance of NUTRES (relevant/appropriate and highly relevant/very appropriate) (n=140), n/N (%)	140/145 (96.6)
It helps them with feeding and PA[c] of their daughters and sons or with counseling of health providers, n (%)	87 (62.1)
You can apply the recommendations in your daily life, n (%)	53 (37.9)
Would you like to continue receiving SMS text messages from NUTRES? (Yes) (n=98), n (%)	96 (98.0)
Recommendations to improve NUTRES from PCs (n=159), n (%)	

Responses	Values (n=145)[a]
Have a balance or credit/signal	28 (17.6)
Have access to cell phone	3 (1.9)
Know what to respond/clear instructions	6 (3.8)
Have time	9 (5.7)
Include images or videos (eg, WhatsApp)	16 (10.1)
Be interesting	5 (3.1)
Have training	8 (5.0)
Other (eg, have a follow-up, include information about diseases, more examples, include foods from the region, make it face-to-face, more interaction, none)	51 (32.1)
Do not know/no answer	63 (39.6)
Most useful PA SMS text message format for PC (n=133), n (%)	
Challenge	61 (45.9)
Socioemotional support	17 (12.8)
Informative	21 (15.8)
Recipes	2 (1.5)
Tips	13 (9.8)
With examples	15 (11.3)
Links for more information	2 (1.5)
Do not know/no answer	4 (3.0)
Most useful HF[d] SMS text message format for PC (n=142), n (%)	
Challenge	25 (17.6)
Socioemotional support	20 (14.1)
Informative	36 (25.4)
Recipes	24 (16.9)
Tips	21 (14.8)
With examples	8 (5.6)
Links for more information	2 (1.4)
Do not know/no answer	5 (3.5)

[a]Totals may vary because of missing data.

[b]PC: primary caregiver.

[c]PA: physical activity.

[d]HF: healthy feeding.

The majority of PCs (92/141, 65.2%) reported that the SMS text messages were practical. Additionally, 57/145 (39.3%) PCs expressed a preference for the informative format over other formats, such as tips or challenges. PCs particularly appreciated the challenges related to PA, with 61/133 (45.9%) expressing appreciation because it involved other family members (Table 2). This aspect was particularly valued, especially in the context of COVID-19.

> *What I liked most about the SMS was that they sent advice, physical activity, not only for Mateo, but for the whole family and everything. And the same advice for the whole family to follow, not just for the baby, for my son* [#3]

PCs recalled messages regarding the importance of avoiding certain unhealthy food items for their children, such as sugar-sweetened beverages or sausages. These messages emphasized that these items were not suitable for children and were associated with the risk of early chronic diseases.

> *I remember that they recommended not to give him boxed juices, but to give him plain water or natural juice, that is, natural fruit juice* [#2]

> *The same, [NUTRES] would send us, for example (...) the sausages and all that, because it was not the right thing to do, because they have fat, if not, that it was better for them to eat chicken, pork, beef* [#2]

NUTRES Intervention Effectiveness

At baseline, PCs from the IG exhibited greater knowledge and awareness regarding the risk of developing anxiety or depression due to lack of PA compared with the CG (4/179, 2.2% vs 13/149, 8.7%, respectively; *P*=.02). Following the intervention, significant differences between the 2 groups were observed, with the IG showing higher rates of favoring social interaction and integration in relation to knowledge of PA (CG: 8/135, 5.9% vs IG: 20/137, 14.6%; *P*=.048). Regarding HF practices, significant differences were observed between the IG and the CG in recognizing that diabetes and chronic diseases can develop as a consequence of sugar-sweetened beverage consumption (CG: 92/157, 58.6% vs IG: 110/145, 75.9%; *P*=.003; Table 3). However, in the difference-in-differences model, there was no significant improvement observed in the practice of infant PA and HF (*P*>.05; Table 4). The only significant increase between the study groups was observed in terms of knowledge about the benefits of PA (CG: mean 0.13, SD 0.10 vs IG: mean 0.16, SD 0.11; *P*>.02; Table 4). Nevertheless, several PCs reported improvements in their perceived control and intention related to some of the behaviors promoted by NUTRES, although detailed data on this aspect are not shown.

> *Well, as far as I know, I used to give my older children juice...soda, and [now] I don't want to give her, I mean, any sweets* [#16]

Table 3. Knowledge and practices about infant physical activity and healthy feeding from primary caregivers: baseline versus final line (NUTRES, 2020-21).[a]

Knowledge and practice	Baseline			Final line		
	Comparison group (n=264)	Intervention group (n=230)	*P* value	Comparison group (n=175), %	Intervention group (n=159), %	*P* value
Physical activity (PA[b]), n/N (%)						
Main benefits of PA identified	177/264 (67.0)	158/230 (68.7)	.43	135/171 (78.9)	137/157 (87.3)	.09
Strengthens muscles and bones	103/177 (58.2)	91/158 (57.6)	.92	76/135 (56.3)	76/137 (55.5)	.89
Strengthens the immune system	45/177 (25.4)	21/158 (13.3)	.17	28/135 (20.7)	19/137 (13.9)	.16
Improves cognitive function, school performance, or both	42/177 (23.7)	43/158 (27.2)	.55	30/135 (22.2)	30/137 (21.9)	.94
Promotes relaxation and well-being and improves sleep patterns	11/177 (6.2)	17/158 (10.8)	.18	15/135 (11.1)	19/137 (13.9)	.56
Helps to prevent chronic diseases	44/177 (24.9)	26/158 (16.5)	.07	25/135 (18.5)	37/137 (27.0)	.09
Helps to improve chronic disease control	35/177 (19.8)	21/158 (13.3)	.32	16/135 (11.9)	19/137 (13.9)	.62
Helps to avoid anxiety/depression	3/177 (1.7)	4/158 (2.5)	.67	3/135 (2.2)	4/137 (2.9)	.70
Promotes socialization and social interaction	12/177 (6.8)	14/158 (8.9)	.47	8/135 (5.9)	20/137 (14.6)	.048
Promotes water consumption (improves hydration)	1/177 (0.6)	__[c]	.34	1/135 (0.7)	3/137 (2.2)	.33
Others (less constipation, promotes growth, gives them energy, they express themselves better)	16/177 (9.0)	16/158 (10.1)	.83	6/135 (4.4)	4/137 (2.9)	.56
Do not know/answer	1/177 (0.6)	—	.36	1/135 (0.7)	2/137 (1.5)	.53
Main consequences of not doing PA	179/264 (67.8)	149/230 (64.8)	.83	132/171 (77.2)	129/157 (82.2)	.38
Decreased flexibility and weak muscles/bones	39/179 (21.8)	41/149 (27.5)	.32	31/132 (23.5)	40/129 (31.0)	.30
Weak immune system	32/179 (17.9)	18/149 (12.1)	.40	25/132 (18.9)	19/129 (14.7)	.38
Poor cognitive function/school performance	9/179 (5.0)	11/149 (7.4)	.48	20/132 (15.2)	12/129 (9.3)	.18
Stress, annoyance, sadness, and insomnia	11/179 (6.1)	14/149 (9.4)	.31	10/132 (7.6)	12/129 (9.3)	.55
Increased risk of chronic disease	131/179 (73.2)	92/149 (61.7)	.17	79/132 (59.8)	85/129 (65.9)	.37
Increased risk of anxiety or depression	4/179 (2.2)	13/149 (8.7)	.02	4/132 (3.0)	9/129 (7.0)	.14
Less socialization and social interaction	9/179 (5.0)	16/149 (10.7)	.11	13/132 (9.8)	18/129 (13.9)	.42
Others (fatigue, poor oxygenation, become lazy, or sedentary)	8/179 (4.5)	10/149 (6.7)	.49	3/132 (2.3)	3/129 (2.3)	.97
Do not know/answer	2/179 (1.1)	1/149 (0.7)	.67	1/132 (0.8)	2/129 (1.5)	.54
PA recommendations in children <5 years	56/264 (21.2)	49/229 (21.4)	.35	48/171 (28.1)	29/157 (18.5)	.14
At least 3 hours per day (180 minutes/day)	2/56 (3.6)	2/49 (4.1)	.63	2/48 (4.2)	2/29 (6.9)	.58
Screen time recommendations for children <5 years	88/264 (33.3)	106/229 (46.3)	.30	90/171 (52.6)	85/157 (54.1)	.91

Knowledge and practice	Baseline			Final line		
	Comparison group (n=264)	Intervention group (n=230)	*P* value	Comparison group (n=175), %	Intervention group (n=159), %	*P* value
<30 minutes/day	27/88 (30.7)	34/106 (32.1)	.59	33/90 (36.7)	33/85 (38.8)	.77
<60 minutes/day	31/88 (35.2)	37/106 (34.9)		35/90 (38.9)	35/85 (41.2)	N/A[d]
<2 hours/day	15/88 (17.0)	13/106 (12.2)		12/90 (13.3)	10/85 (11.8)	N/A
Is avoided	12/88 (13.6)	12/106 (11.3)		7/90 (7.8)	4/85 (4.7)	N/A
It depends on the age	2/88 (2.3)	2/106 (1.9)		1/90 (1.1)	0 (0)	N/A
Other	—	4/106 (3.8)		2/90 (2.2)	2/85 (2.4)	N/A
Do not know/answer	1/88 (1.1)	4/106 (3.8)		0 (0)	1/85 (1.2)	N/A
Healthy feeding						
Identifies that it is important to include vegetables in the dishes' preparation (vegetable consumption)	28/264 (10.6)	26/230 (11.3)	.86	159/175 (90.9)	146/159 (91.8)	.78
Identifies the consumption of natural water as recommended	36/264 (13.6)	28/230 (12.2)	.80	11/175 (6.3)	19/159 (11.9)	.13
Identifies the consumption of sugar-sweetened beverages as not recommended	6/264 (2.3)	5/230 (2.2)	.94	8/175 (4.6)	3/159 (1.9)	.13
Identifies the consumption of ultra-processed products (eg, cupcakes, cookies) as not recommended	—	1/230 (0.4)	.30	3/175 (1.7)	4/159 (2.5)	.63
Consequences of an unhealthy feeding						
Undernourishment	196/264 (74.2)	162/224 (72.3)	.63	145/175 (82.9)	126/159 (79.2)	.41
Overweight or obesity	71/264 (26.9)	64/224 (28.6)	.61	46/175 (26.3)	57/159 (35.8)	.13
Respiratory diseases	21/264 (8.0)	16/224 (7.1)	.76	15/175 (8.6)	9/159 (5.7)	.28
Musculoskeletal or skin problems	2/264 (0.8)	2/224 (0.9)	.88	16/175 (9.1)	3/159 (1.9)	.002
Sleep disturbances, discouragement and tiredness, difficulty practicing a physical activity or any activity, learning difficulties	9/264 (3.4)	7/224 (3.1)	.83	7/175 (4.0)	4/159 (2.5)	.58
Diabetes, hypertension, and hypercholesterolemia	25/264 (9.5)	25/224 (11.2)	.44	17/175 (9.7)	20/159 (12.6)	.48
Candies and soft drinks are identified as prizes for life	0 (0)	1/224 (0.4)	.20	0 (0)	1/159 (0.6)	.32
Harmful habits are established that are for life	0 (0)	1/224 (0.4)	.20	0 (0)	1/159 (0.6)	.32
Others (eg, bulimia and anorexia, gastritis, diarrhea and vomiting, lack of appetite, dehydration, getting sick in general)	14/264 (5.3)	8/224 (3.6)	.23	1/175 (0.6)	2/159 (1.3)	.46
Do not know/answer	14/264 (5.3)	13/224 (5.8)	.79	5/175 (2.9)	8/159 (5.0)	.31
Consequences of sugar-sweetened beverage consumption				157/171 (91.8)	145/157 (92.4)	.67
Undernourishment	14/215 (6.5)	13/185 (7.0)	.90	14/171 (8.2)	10/144 (6.9)	.46
Overweight or obesity	96/215 (44.7)	97/187 (51.9)	.33	75/157 (47.8)	67/145 (46.2)	.79
Respiratory diseases or diarrhea	5/215 (2.3)	3/185 (1.6)	.64	4/157 (2.5)	3/143 (2.1)	.81
Musculoskeletal or skin problems	10/215 (4.7)	11/185 (5.9)	.62	17/157 (10.8)	8/145 (5.5)	.10

Knowledge and practice	Baseline			Final line		
	Comparison group (n=264)	Intervention group (n=230)	*P* value	Comparison group (n=175), %	Intervention group (n=159), %	*P* value
Sleep disturbances, discouragement and tiredness, difficulty practicing a physical activity or any activity, learning difficulties	5/215 (2.3)	2/185 (1.1)	.92	13/157 (8.3)	2/143 (1.4)	.01
Diabetes, hypertension, and hypercholesterolemia	139/215 (64.7)	116/187 (62.0)	.65	92/157 (58.6)	110/145 (75.9)	.003
Candies and soft drinks are identified as prizes for life	0 (0)	1/185 (0.5)	.21	0 (0)	2/145 (1.4)	.18
Harmful habits are established that are for life	2/215 (0.9)	2/185 (1.1)	.90	0 (0)	3/145 (2.1)	.12
Others (eg, kidney problems, addiction, hyperactivity, stomach ache)	21/215 (9.8)	20/185 (10.8)	.84	4/157 (2.5)	5/145 (3.4)	.56
Do not know/answer	1/215 (0.5)	2/185 (1.1)	.47	2/157 (1.3)	0 (0)	.31

[a]Totals may vary because of missing data.

[b]PA: physical activity.

[c]Not available.

[d]N/A: not applicable.

Table 4. Changes in knowledge and practices about infant physical activity and healthy feeding among primary caregivers in NUTRES.

Outcomes of interest	Baseline		Postintervention			*P* value
	Comparison group (n=264)	Intervention group (n=230)	Comparison group (n=175)	Intervention group (n=159)	Estimated change, double difference[a]	
Knowledge						
PA[b,c,d]						
PA concept, mean (SD)	0.26 (0.18)	0.3 (0.18)	0.28 (0.17)	0.27 (0.15)	–0.05	.04
PA benefits, mean (SD)	0.13(0.11)	0.12 (0.11)	0.13 (0.10)	0.16 (0.11)	0.037	.02
PA recommendations, mean (SD)	0.04 (0.10)	0.05 (0.10)	0.05 (0.09)	0.03 (0.06)	–0.024	.07
Screen time recommendations (adequate), n (%)	88 (33.3)	106 (46.1)	90 (51.4)	85 (53.5)	N/A[e]	<.001
Healthy feeding[b], mean (SD)						
Healthy menu	0.18 (0.07)	0.19 (0.07)	0.19 (0.07)	0.21 (0.10)	0.014	.19
Unhealthy feeding	0.25 (0.12)	0.25 (0.14)	0.26 (0.15)	0.29 (0.16)	0.025	.21
Healthy drinks	0.25 (0.13)	0.26 (0.12)	0.25 (0.12)	0.26 (0.11)	0.001	.94
Unhealthy drinks	0.14 (0.07)	0.14 (0.06)	0.14 (0.08)	0.15 (0.09)	0.013	.21
Practices						
PA[b,c], mean (SD)	5.30 (3.28)	4.94 (3.20)	8.65 (3.50)	7.67 (3.50)	–0.758	.08
PC[f] of children between 6 and 23.9 months, mean (SD)	4.81 (3.15)	4.60 (3.03)	9.68 (3.08)	8.28 (3.68)	–1.175	.10
Adequate, n/N (%)	38/102 (37.3)	25/88 (28.4)	25/66 (37.9)	13/61 (21.3)	N/A	N/A
Moderate, n/N (%)	19/102 (18.6)	13/88 (14.8)	24/66 (36.4)	21/61 (34.4)	N/A	N/A
Limited, n/N (%)	45/102 (44.1)	50/88 (56.8)	17/66 (25.8)	27/61 (44.3)	N/A	N/A
PC of children ≥24 months, mean (SD)	7.53 (3.05)	7.84 (3.14)	9.0 (2.22)	8.58 (2.58)	0.705	.30
Adequate, n/N (%)	31/95 (32.6)	22/65 (33.8)	21/63 (33.3)	16/49 (32.7)	N/A	N/A
Moderate, n/N (%)	33/95 (34.7)	20/65 (30.8)	22/63 (34.9)	14/49 (28.6)	N/A	N/A
Limited, n/N (%)	31/95 (32.6)	23/65 (35.4)	20/63 (31.7)	19/49 (38.8)	N/A	N/A
Healthy feeding[g], mean (SD)	10.22 (3.09)	10.72 (2.69)	10.23 (2.32)	10.77 (2.25)	0.029	.96
Adequate, n (%)	31/99 (31.3)	24/66 (36.4)	22/65 (33.8)	16/50 (32.0)	N/A	N/A
Moderate, n (%)	34/99 (34.3)	21/66 (31.8)	18/65 (27.7)	20/50 (40.0)	N/A	N/A
Limited, n (%)	34/99 (34.3)	21/66 (31.8)	25/65 (38.5)	14/50 (28.0)	N/A	N/A

[a]Difference-in-differences models adjusted for the baseline age of the child and maternal employment.

[b]Applied to all participants.

[c]An index of PA and healthy feeding practices was calculated, whose score was divided into tertiles to obtain the categories of "adequate," "moderate," and "limited" practices according to children's age (annex 6).

[d]PA: physical activity.

[e]N/A: not applicable.

[f]PC: primary caregiver.

[g]Applied only to PC of children >24 months.

Discussion

Principal Findings

To our knowledge, this study represents the first report on the feasibility of implementation and effectiveness of an mHealth intervention aimed at preventing childhood overweight/obesity in Mexico within the context of COVID-19. This research offers valuable insights into the implementation process of NUTRES and its impact on expected outcomes. Although NUTRES demonstrated limited impact on target knowledge and behaviors, the SMS text messages served as reminders or sources of new knowledge among PCs of Mexican children under 5 years.

The absence of significant effects observed in this study may be partially attributed to methodological adaptations necessitated by the COVID-19 context, such as recruitment and data collection conducted via telephone. However, it is conceivable that NUTRES could have served to mitigate the impact of the pandemic on the participating population. Given that the COVID-19 pandemic exacerbated food insecurity and contributed to increased consumption of ultra-processed products, along with physical inactivity during lockdowns, the intervention may have provided some degree of protection to the population involved [33]. Furthermore, the COVID-19 pandemic severely disrupted health services, significantly limiting the population's access to essential nutritional counseling [34].

Additionally, the lack of significant effects could be attributed to the high rate of loss to follow-up, which, although not markedly different between the study groups (71/230, 30.9%, loss in the IG and 89/264, 33.7%, loss in the CG), represented a loss to follow-up exceeding 20% (160/494, 32.4%) [35,36]. However, considering a sample size of 150 participants per study group and assuming a correlation of 0.6 between repeated measures within participants, our analysis sample achieved an approximate power of 83% to detect a difference of 0.15 in the change of proportions [37]. Nevertheless, it is plausible that the registration and follow-up process, which involved requesting contact information such as the child's name or age, may have discouraged the participation of PCs due to concerns about privacy and security. To address this issue, other studies have implemented strategies such as providing payments or reminders, resulting in a high retention rate of over 80% [38]. Such approaches could potentially enhance participant engagement and retention in future interventions.

Another important consideration to mention is that, due to the nature of the intervention, access to and use of a mobile phone was a prerequisite for inclusion in the IG. This aspect could introduce potential selection bias [39], especially considering that mobile phone ownership was not a requirement for the CG and mobile phone ownership might be correlated with income level. However, this potential bias was addressed in the analysis by adjusting for paid maternal employment (40/230, 17.4%, in the IG vs 22/264, 8.3%, in the CG), among other baseline characteristics. Adjusting for these baseline characteristics helps mitigate the potential influence of confounding factors on the study outcomes.

Previous studies have not consistently shown significant effects of exposure to BCC mHealth interventions on the intended audience [8]. In this trial, although 96/144 (66.7%) PCs always read NUTRES SMS text messages, only 43/141 (30.5%) always used the information provided. However, based on PC testimonies, NUTRES served as a reliable source of recommendations or reminders for child care, which they incorporated into their practices. Simultaneously, these recommendations empowered them within their immediate family and social environment. Furthermore, while PC narratives expressed some enthusiasm, the limitations of conducting telephone interviews prevented the observation of informants' physical responses, hindering further exploration and discussion of their enthusiasm. Therefore, future mHealth interventions should incorporate features to enhance engagement over time, using various methods such as incentives, push notifications, and other interactive elements. These strategies could help sustain interest and participation in the intervention among the target audience.

The results from both the qualitative and quantitative components of NUTRES are consistent with the lack of effectiveness observed in other studies regarding the impact of mHealth interventions on PC practices related to PA and HF. For example, a study conducted in China found that a weekly SMS text messaging intervention over 12 months had no significant effect on infant feeding practices [40]. It could therefore be speculated that the 9-month duration of exposure in NUTRES may have been insufficient to bring about changes in infant HF practices. However, NUTRES did demonstrate an improvement in PC knowledge regarding some benefits of PA. Indeed, PCs emphasized in interviews that family challenges played a crucial role in guiding and supporting them in implementing the recommendations provided by NUTRES. While the effectiveness of NUTRES could not be conclusively demonstrated, the mixed methods methodology used in this study identified that the obstacles to implementation were primarily external to NUTRES and largely technological in nature. Similarly, other studies conducted in Mexico have documented persistent barriers to mHealth implementation, such as the cost of mobile phones, lack of network infrastructure, and low digital literacy [41]. These barriers to implementing mHealth interventions may diminish the potential impact of smaller-scale studies and hinder their scale-up [41].

Our findings suggest that multicomponent and multichannel interventions have greater potential to promote behavior change compared with single-channel multitheme interventions. While the original design of NUTRES aimed for health providers to reinforce the intervention topics among PCs, this approach was not feasible due to the disruption of health services during the COVID-19 pandemic [42,43]. Nevertheless, an mHealth intervention could serve as a promising channel to maintain contact when face-to-face interaction is not feasible, complementing other remote interventions such as telemedicine [44,45].

One strength of this study is its effectiveness-implementation hybrid trial with an advanced convergent design, which enabled the comparison and interpretation of quantitative and qualitative results. Using a mixed methods approach, we were able to

comprehensively observe the effects of the intervention and implementation process [13]. For example, our study revealed that the low interaction with SMS text messages by PCs did not necessarily indicate a lack of interest in NUTRES. Instead, it was often attributed to external factors such as poor signal reception, economic constraints, and other challenges. However, despite these insights, we were generally unable to demonstrate a significant effect of the NUTRES intervention on the target behaviors. Nevertheless, previous studies [46] investigating the usefulness of SMS text messages, as reported by PCs, suggest that valuable behavioral changes could potentially be achieved in the long term. Additionally, our difference-in-differences models were adjusted for child age and maternal employment, which were variables that differed between the groups at baseline. This adjustment allowed us to estimate the specific effect of the intervention. Furthermore, the double-difference approach accounts for time-invariant observed factors (eg, education) and unobserved variables (eg, motivation to comply) [27], providing a more robust analysis of the intervention's impact.

Lessons Learned

The positive feedback received through both quantitative and qualitative data aligns with findings from another mHealth intervention [47,48]. A systematic review conducted in 2019 [9] identified the top 3 barriers among participants, which included signal coverage, lack of equipment (unaffordable phones), and technology gap (limited knowledge of how to use a phone). In this study, despite the participants' positive perceptions, similar challenges were encountered in terms of interaction with SMS text messages. For instance, some participants reported that SMS text messages were blocked, requiring them to contact their mobile operator to resolve the issue, which they were often unaware of. Additionally, other reasons for not receiving the SMS text messages were dropped lines, being out of coverage area, or changes in phone numbers (15/32, 47%). When participants did not respond to the SMS text messages within 12 hours, they were removed from the screen (7/42, 17%), and a significant portion of participants dropped out of the study (160/494, 32.4%). Overcoming such barriers is crucial to increase exposure to the intervention, thereby enhancing its impact and sustainability [49].

Based on the results presented here, this study provides valuable insights that can serve as a foundation for further research using mHealth interventions. The advantages of mHealth interventions, including broad coverage, cost-effectiveness, and feasibility of implementation in the population, underscore the potential for future studies in this area. It is suggested that future research should focus on addressing the identified challenges to improve the effectiveness, efficiency, and scalability of interventions. Some potential areas for improvement are (1) using WhatsApp (Facebook, Inc.) instead of SMS text messages (or in addition to SMS text messages), as it is a useful and accepted medium among the target population; (2) using informative material to recruit PCs; (3) improving SMS text message reception; (4) clarifying that SMS text message interaction has no cost; and (5) including resubscription for loss/change of mobile phone (Table 2).

Another valuable lesson learned is the importance of engaging key health care providers [50] and identifying informal strategies to recruit participants during emergency situations [33]. Furthermore, we have observed that mHealth interventions during lockdowns may represent the sole means of supporting socioeconomically vulnerable populations, who were often the hardest hit [51].

Conclusions

The findings contribute to the existing evidence on the opportunities and limitations associated with mHealth interventions for BCC aimed at enhancing knowledge and practices related to PA and HF among caregivers of infants. Integrating mobile phone technology into BCC interventions facilitates the promotion of healthy lifestyles, particularly during emergency situations and among vulnerable populations. Moreover, the integration of effectiveness and implementation findings is valuable for optimizing mHealth interventions in particular settings, enhancing their sustainability, and fostering their application in diverse contexts [52].

Acknowledgments

This work received financial support from the United Nations Children's Fund of Mexico. The authors thank Luz Dinorah González-Castell, MSc for her review and suggestions to improve the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Examples of SMS text messages delivered by mobile phone to primary caregivers' of the intervention group of NUTRES, 2020-21.
[DOCX File , 30 KB - mhealth_v12i1e55509_app1.docx]

Multimedia Appendix 2
Index of knowledge and practices on topics of interest, NUTRES.
[DOCX File , 15 KB - mhealth_v12i1e55509_app2.docx]

Multimedia Appendix 3
Category tree for coding testimonies of primary caregivers of NUTRES, 2021.
[DOCX File , 15 KB - mhealth_v12i1e55509_app3.docx]

Multimedia Appendix 4
Sociodemographic characteristics of primary caregivers at baseline and lost to follow-up. NUTRES, 2020-21.
[DOCX File , 20 KB - mhealth_v12i1e55509_app4.docx]

Multimedia Appendix 5
Main characteristics of primary caregivers interviewed. NUTRES, 2020-21.
[DOCX File , 15 KB - mhealth_v12i1e55509_app5.docx]

Multimedia Appendix 6
Implementation indicators of NUTRES by primary caregivers exposed to NUTRES.
[DOCX File , 17 KB - mhealth_v12i1e55509_app6.docx]

References

1. The Commission on Ending Childhood Obesity. Implementation plan: executive summary. WHO. Geneva, Switzerland: World Health Organization; 2017. URL: https://tinyurl.com/55rwmvmz [accessed 2024-03-20]
2. Wells JC, Sawaya AL, Wibaek R, Mwangome M, Poullas MS, Yajnik CS, et al. The double burden of malnutrition: aetiological pathways and consequences for health. Lancet 2020 Jan 04;395(10217):75-88 [FREE Full text] [doi: 10.1016/S0140-6736(19)32472-9] [Medline: 31852605]
3. Wake M, Hardy P, Canterford L, Sawyer M, Carlin JB. Overweight, obesity and girth of Australian preschoolers: prevalence and socio-economic correlates. Int J Obes (Lond) 2007 Jul;31(7):1044-1051. [doi: 10.1038/sj.ijo.0803503] [Medline: 17146451]
4. Romero-Martínez M, Barrientos-Gutiérrez T, Cuevas-Nasu L, Bautista-Arrendondo S, Colchero MA, Gaona-Pineda EB, et al. Metodología de la Encuesta Nacional de Salud y Nutrición 2022 y planeación y diseño de la Ensanut Continua 2020-2024. Salud Publica Mex 2022 Aug 26;64(5, sept-oct):522-529. [doi: 10.21149/14186] [Medline: 36130337]
5. Hillier-Brown FC, Bambra CL, Cairns J, Kasim A, Moore HJ, Summerbell CD. A systematic review of the effectiveness of individual, community and societal level interventions at reducing socioeconomic inequalities in obesity amongst children. BMC Public Health 2014 Aug 11;14:834 [FREE Full text] [doi: 10.1186/1471-2458-14-834] [Medline: 25113624]
6. World Health Organization (WHO) Global Observatory for eHealth. Global diffusion of eHealth: making universal health coverage achievable. Report of the Third Global Survey on eHealth. WHO. Geneva, Switzerland: WHO; 2016. URL: https://iris.who.int/bitstream/handle/10665/252529/9789241511780-eng.pdf?sequence=1 [accessed 2024-03-21]
7. Militello LK, Kelly SA, Melnyk BM. Systematic review of text-messaging interventions to promote healthy behaviors in pediatric and adolescent populations: implications for clinical practice and research. Worldviews Evid Based Nurs 2012 Apr;9(2):66-77. [doi: 10.1111/j.1741-6787.2011.00239.x] [Medline: 22268959]
8. Gurman TA, Rubin SE, Roess AA. Effectiveness of mHealth behavior change communication interventions in developing countries: a systematic review of the literature. J Health Commun 2012;17 Suppl 1:82-104. [doi: 10.1080/10810730.2011.649160] [Medline: 22548603]
9. Kruse C, Betancourt J, Ortiz S, Valdes Luna SM, Bamrah IK, Segovia N. Barriers to the use of mobile health in improving health outcomes in developing countries: systematic review. J Med Internet Res 2019 Oct 09;21(10):e13263 [FREE Full text] [doi: 10.2196/13263] [Medline: 31593543]
10. Müller AM, Alley S, Schoeppe S, Vandelanotte C. Int J Behav Nutr Phys Act 2016 Oct 10;13(1):109 [FREE Full text] [doi: 10.1186/s12966-016-0434-2] [Medline: 27724911]
11. Instituto Nacional de Estadística y Geografía (INEGI), Secretaría de Comunicaciones y Transportes (SCT), Instituto Federal de Telecomunicaciones (IFT). Encuesta Nacional sobre Disponibilidad y Uso de Tecnologías de la Información en los Hogares (ENDUTIH) 2022. INEGI. URL: https://www.inegi.org.mx/programas/dutih/2022/ [accessed 2024-03-21]
12. Curran GM, Bauer M, Mittman B, Pyne JM, Stetler C. Effectiveness-implementation hybrid designs: combining elements of clinical effectiveness and implementation research to enhance public health impact. Med Care 2012 Mar;50(3):217-226 [FREE Full text] [doi: 10.1097/MLR.0b013e3182408812] [Medline: 22310560]
13. Palinkas LA, Mendon SJ, Hamilton AB. Innovations in mixed methods evaluations. Annu Rev Public Health 2019 Apr 01;40(1):423-442 [FREE Full text] [doi: 10.1146/annurev-publhealth-040218-044215] [Medline: 30633710]
14. Klassen AC, Creswell J, Plano Clark VL, Smith KC, Meissner HI. Best practices in mixed methods for quality of life research. Qual Life Res 2012 Apr;21(3):377-380. [doi: 10.1007/s11136-012-0122-x] [Medline: 22311251]
15. Balakrishnan R, Gopichandran V, Chaturvedi S, Chatterjee R, Mahapatra T, Chaudhuri I. Continuum of care services for maternal and child health using mobile technology - a health system strengthening strategy in low and middle income

countries. BMC Med Inform Decis Mak 2016 Jul 07;16:84 [FREE Full text] [doi: 10.1186/s12911-016-0326-z] [Medline: 27387548]
16. Harari N, Rosenthal MS, Bozzi V, Goeschel L, Jayewickreme T, Onyebeke C, et al. Feasibility and acceptability of a text message intervention used as an adjunct tool by WIC breastfeeding peer counsellors: the LATCH pilot. Matern Child Nutr 2018 Jan;14(1):e12488 [FREE Full text] [doi: 10.1111/mcn.12488] [Medline: 28766913]
17. Bull FC, Al-Ansari SS, Biddle S, Borodulin K, Buman MP, Cardon G, et al. World Health Organization 2020 guidelines on physical activity and sedentary behaviour. Br J Sports Med 2020 Dec;54(24):1451-1462 [FREE Full text] [doi: 10.1136/bjsports-2020-102955] [Medline: 33239350]
18. Fondo de las Naciones Unidas para la Infancia. In: , editor. El Sobrepeso en la Niñez: Un Llamado Para la Prevención en América Latina y el Caribe. Panama City, Panama: UNICEF; 2022.
19. Dirección de políticas y programas en nutrición, Centro de Investigación en Nutrición y Salud. Protocolo Para la Prevención del Sobrepeso y Obesidad Infantil a Través de Una Intervención de mHealth. UNICEF. Cuernavaca, México: Instituto Nacional de Salud Pública; 2018. URL: https://www.unicef.org/mexico/estrategia-nutres-mhealth-o-salud-m%C3%B3vil [accessed 2024-03-21]
20. Théodore F, Lozada-Tequeanes A, Alvarado R, Kim-Herrera E, García-Guerra A, Rodríguez-Ramírez S. Exploring facilitators and barriers of children's feeding with 'junk food': insights from Mexican women's experiences. Public Health Nursing 2024 (forthcoming).
21. Théodore FL, Bonvecchio A, Lozada Tequeanes AL, Alvarado R, García-Guerra A, Villanueva Borbolla MA, et al. Challenges around child-feeding practices with 'comida chatarra': a qualitative study to understand the role of sociocultural factors in caregiver feeding decisions. Nutrients 2023 Mar 07;15(6):1317 [FREE Full text] [doi: 10.3390/nu15061317] [Medline: 36986041]
22. Ajzen I. From intentions to actions: a theory of planned behavior. In: Kuhl J, Beckman J, editors. Action Control. Berlin, Heidelberg: Springer; 1985:11-39.
23. Ajzen I. The theory of planned behavior. Organizational Behavior and Human Decision Processes 1991;50(2):179-211 [FREE Full text] [doi: 10.1016/0749-5978(91)90020-t]
24. SSA, INSP, UNICEF. 2023 Guías alimentarias saludables y sostenibles para la población mexicana. Google Drive. 2023. URL: https://drive.google.com/file/d/1fGQoNbTZ0zlfVrYyZ8eT0efQWEfflwae/view [accessed 2024-03-21]
25. Gobierno de México. Prospera digital: plataforma digital para el desarrollo social y la inclusión financiera. Gobierno de Mexico. 2015. URL: https://tinyurl.com/yc4bbsu8 [accessed 2023-03-12]
26. UNICEF. RapidPro. UNICEF. 2018. URL: https://www.unicef.org/innovation/stories/what-exactly-rapidpro [accessed 2023-03-14]
27. Gertler PJ, Martinez S, Premand P, Rawlings LB, Vermeersch CMJ. Impact Evaluation in Practice. Washington, DC: World Bank; 2011.
28. McCoy CE. Understanding the intention-to-treat principle in randomized controlled trials. West J Emerg Med 2017 Oct;18(6):1075-1078 [FREE Full text] [doi: 10.5811/westjem.2017.8.35985] [Medline: 29085540]
29. Proctor E, Silmere H, Raghavan R, Hovmand P, Aarons G, Bunger A, et al. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. Adm Policy Ment Health 2011 Mar;38(2):65-76 [FREE Full text] [doi: 10.1007/s10488-010-0319-7] [Medline: 20957426]
30. Peters DH, Adam T, Alonge O, Agyepong IA, Tran N. Implementation research: what it is and how to do it. BMJ 2013 Nov 20;347:f6753. [doi: 10.1136/bmj.f6753] [Medline: 24259324]
31. Glaser BG, Strauss AL. The Discovery of Grounded Theory. Strategies for Qualitative Research. Chicago, IL: Aldine; 1967.
32. Eldridge SM, Chan CL, Campbell MJ, Bond CM, Hopewell S, Thabane L, PAFS consensus group. CONSORT 2010 statement: extension to randomised pilot and feasibility trials. Pilot Feasibility Stud 2016 Oct 21;2(1):64-29 [FREE Full text] [doi: 10.1186/s40814-016-0105-8] [Medline: 27965879]
33. Hennis A, Coates A, Del Pino S, Ghidinelli M, Gomez Ponce de Leon R, Bolastig E, et al. COVID-19 and inequities in the Americas: lessons learned and implications for essential health services. Rev Panam Salud Publica 2021;45:e130 [FREE Full text] [doi: 10.26633/RPSP.2021.130] [Medline: 34987555]
34. Osendarp S, Akuoku J, Black R, Headey D, Ruel M, Scott N, et al. The COVID-19 crisis will exacerbate maternal and child undernutrition and child mortality in low- and middle-income countries. Nat Food 2021 Jul;2(7):476-484 [FREE Full text] [doi: 10.1038/s43016-021-00319-4] [Medline: 37117686]
35. Akl EA, Briel M, You JJ, Lamontagne F, Gangji A, Cukierman-Yaffe T, et al. LOST to follow-up information in trials (LOST-IT): a protocol on the potential impact. Trials 2009 Jun 11;10:40 [FREE Full text] [doi: 10.1186/1745-6215-10-40] [Medline: 19519891]
36. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. Trials 2010 Mar 24;11:32 [FREE Full text] [doi: 10.1186/1745-6215-11-32] [Medline: 20334632]
37. Fitzmaurice GM, Laird NM, Ware JH. Applied Longitudinal Analysis. Hoboken, NJ: John Wiley and Sons, Inc; 2004:411-413.

38. Laws R, Denney-Wilson E, Taki S, Russell CG, Zheng M, Litterbach E, et al. Key lessons and impact of the growing healthy mHealth program on milk feeding, timing of introduction of solids, and infant growth: quasi-experimental study. JMIR Mhealth Uhealth 2018 Apr 19;6(4):e78-e17 [FREE Full text] [doi: 10.2196/mhealth.9040] [Medline: 29674313]
39. Kleinbaum DG, Kupper LL, Morgenstern H. Epidemiologic research. In: Principles and Quantitative Methods. Belmont, CA: Lifetime Learning Publications; 1982.
40. Jiang H, Li M, Wen LM, Hu Q, Yang D, He G, et al. Effect of short message service on infant feeding practice: findings from a community-based study in Shanghai, China. JAMA Pediatr 2014 May;168(5):471-478. [doi: 10.1001/jamapediatrics.2014.58] [Medline: 24639004]
41. Gaitán-Rossi P, De la Cerda Lobato S, Pérez Navarro AC, Aguilar Esteva A, Vargas García MR, Vilar-Compte M. Fidelity of implementation of Prospera Digital: evaluation of a multi-site intervention aimed at improving maternal health outcomes in Mexico. Curr Dev Nutr 2019 Oct;3(10):nzz107. [doi: 10.1093/cdn/nzz107] [Medline: 31637367]
42. Okely AD, Kariippanon KE, Guan H, Taylor EK, Suesse T, Cross PL, et al. Global effect of COVID-19 pandemic on physical activity, sedentary behaviour and sleep among 3- to 5-year-old children: a longitudinal study of 14 countries. BMC Public Health 2021 May 17;21(1):940-915 [FREE Full text] [doi: 10.1186/s12889-021-10852-3] [Medline: 34001086]
43. Cena H, Fiechtner L, Vincenti A, Magenes VC, De Giuseppe R, Manuelli M, et al. COVID-19 pandemic as risk factors for excessive weight gain in pediatrics: the role of changes in nutrition behavior. A narrative review. Nutrients 2021 Nov 26;13(12):4255 [FREE Full text] [doi: 10.3390/nu13124255] [Medline: 34959805]
44. Lukas H, Xu C, Yu Y, Gao W. Emerging telemedicine tools for remote COVID-19 diagnosis, monitoring, and management. ACS Nano 2020 Dec 22;14(12):16180-16193. [doi: 10.1021/acsnano.0c08494] [Medline: 33314910]
45. Drees-Gross F, Zhang P. Less than 50% of Latin America has fixed broadband. Here are 3 ways to boost the region's digital access. In: World Economic Forum. Geneva, Switzerland: World Economic Forum; 2021.
46. Tamrat T, Kachnowski S. Special delivery: an analysis of mHealth in maternal and newborn health programs and their outcomes around the world. Matern Child Health J 2012 Jul 19;16(5):1092-1101. [doi: 10.1007/s10995-011-0836-3] [Medline: 21688111]
47. Fassnacht D, Ali K, Silva C, Gonçalves S, Machado P. Use of text messaging services to promote health behaviors in children. J Nutr Educ Behav 2015;47(1):75-80 [FREE Full text] [doi: 10.1016/j.jneb.2014.08.006] [Medline: 25282200]
48. Banna J, Campos M, Gibby C, Graulau RE, Meléndez M, Reyes A, et al. Multi-site trial using short mobile messages (SMS) to improve infant weight in low-income minorities: development, implementation, lessons learned and future applications. Contemp Clin Trials 2017 Nov;62:56-60 [FREE Full text] [doi: 10.1016/j.cct.2017.08.011] [Medline: 28827160]
49. Chambers DA, Glasgow RE, Stange KC. The dynamic sustainability framework: addressing the paradox of sustainment amid ongoing change. Implement Sci 2013 Oct 02;8:117 [FREE Full text] [doi: 10.1186/1748-5908-8-117] [Medline: 24088228]
50. Kruk ME, Gage AD, Joseph NT, Danaei G, García-Saisó S, Salomon JA. Mortality due to low-quality health systems in the universal health coverage era: a systematic analysis of amenable deaths in 137 countries. Lancet 2018 Nov 17;392(10160):2203-2212 [FREE Full text] [doi: 10.1016/S0140-6736(18)31668-4] [Medline: 30195398]
51. Abaza H, Marschollek M. mHealth application areas and technology combinations. A comparison of literature from high and low/middle income countries. Methods Inf Med 2017 Aug 08;56(7):e105-e122 [FREE Full text] [doi: 10.3414/ME17-05-0003] [Medline: 28925418]
52. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci 2009 Aug 07;4:50 [FREE Full text] [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]

Abbreviations

BCC: behavior change communication
CG: comparison group
HF: healthy feeding
IG: intervention group
INSP: National Institute of Public Health
mHealth: mobile health
PA: physical activity
PC: primary caregiver
REDCap: Research Electronic Data Capture
WHO: World Health Organization

Edited by L Buis; submitted 18.12.23; peer-reviewed by J Betancourt, R Zhuang; comments to author 22.02.24; revised version received 28.02.24; accepted 29.02.24; published 09.04.24.

Please cite as:
Lozada-Tequeanes AL, Théodore FL, Kim-Herrera E, García-Guerra A, Quezada-Sánchez AD, Alvarado-Casas R, Bonvecchio A Effectiveness and Implementation of a Text Messaging mHealth Intervention to Prevent Childhood Obesity in Mexico in the COVID-19 Context: Mixed Methods Study
JMIR Mhealth Uhealth 2024;12:e55509
URL: https://mhealth.jmir.org/2024/1/e55509
doi:10.2196/55509
PMID:38592753

Original Paper

Perceptions of Wearable Health Tools Post the COVID-19 Emergency in Low-Income Latin Communities: Qualitative Study

Stefany Cruz[1], MS; Claire Lu[1], BS; Mara Ulloa[2], MS; Alexander Redding[3], BS; Josiah Hester[4], BS, PhD; Maia Jacobs[2,5], BS, PhD

[1]Department of Electrical and Computer Engineering, McCormick School of Engineering, Northwestern University, Evanston, IL, United States

[2]Department of Computer Science, McCormick School of Engineering, Northwestern University, Evanston, IL, United States

[3]Department of Computer Science and Engineering, Irwin & Joan Jacobs School of Engineering, University of California, San Diego, La Jolla, CA, United States

[4]Interactive Computing and Computer Science, College of Computing, Georgia Institute of Technology, Atlanta, GA, United States

[5]Department of Preventative Medicine, Feinberg School of Medicine, Northwestern University, Evanston, IL, United States

Corresponding Author:
Stefany Cruz, MS
Department of Electrical and Computer Engineering
McCormick School of Engineering
Northwestern University
633 Clark Street
Evanston, IL, 60208
United States
Phone: 1 8474913741
Email: stefanycruz2024@u.northwestern.edu

Abstract

Background: Mobile health (mHealth) wearable devices are increasingly being adopted by individuals to help manage and monitor physiological signals. However, the current state of wearables does not consider the needs of racially minoritized low–socioeconomic status (SES) communities regarding usability, accessibility, and price. This is a critical issue that necessitates immediate attention and resolution.

Objective: This study's aims were 3-fold, to (1) understand how members of minoritized low-SES communities perceive current mHealth wearable devices, (2) identify the barriers and facilitators toward adoption, and (3) articulate design requirements for future wearable devices to enable equitable access for these communities.

Methods: We performed semistructured interviews with low-SES Hispanic or Latine adults (N=19) from 2 metropolitan cities in the Midwest and West Coast of the United States. Participants were asked questions about how they perceive wearables, what are the current benefits and barriers toward use, and what features they would like to see in future wearable devices. Common themes were identified and analyzed through an exploratory qualitative approach.

Results: Through qualitative analysis, we identified 4 main themes. Participants' perceptions of wearable devices were strongly influenced by their COVID-19 experiences. Hence, the first theme was related to the impact of COVID-19 on the community, and how this resulted in a significant increase in interest in wearables. The second theme highlights the challenges faced in obtaining adequate health resources and how this further motivated participants' interest in health wearables. The third theme focuses on a general distrust in health care infrastructure and systems and how these challenges are motivating a need for wearables. Lastly, participants emphasized the pressing need for community-driven design of wearable technologies.

Conclusions: The findings from this study reveal that participants from underserved communities are showing emerging interest in using health wearables due to the COVID-19 pandemic and health care access issues. Yet, the needs of these individuals have been excluded from the design and development of current devices.

(JMIR Mhealth Uhealth 2024;12:e50826) doi:10.2196/50826

KEYWORDS

mHealth; mobile health; wearable; wearables; Health wearables; COVID-19; digital divide; low-socioeconomic status; socioeconomic; adoption; underserved; poverty; low income; low resource; marginalized; equity; attitude; attitudes; opinion;

opinions; perception; perceptions; perspective; perspectives; acceptance; Spanish; Hispanic; Latinx; Hispanics; interview; interviews

Introduction

Background

As mobile computing has advanced, wearable technologies have become more ubiquitous; however, in their current state, wearables threaten to worsen digital and health inequities and perpetuate structural harm in the health care sector and society as a whole [1,2]. Health disparities persist that limit the positive health outcomes of those from low–socioeconomic status (SES) backgrounds. The inequitable allocation of resources prevents many from low-SES communities from accessing quality health care, and systemic racism continues to pervade the health care system on both an individual provider and more macro level [3-6]. Further, the social determinants of health, which include factors such as economic status and race, are now understood to be the primary drivers of health outcomes, meaning that health disparities can manifest from societal structures even before an individual interacts with the health care system [7,8]. As a result, current systems fail to adequately provide for the health of Latine and low-SES communities, contributing to the preventable higher rates of illness and death among these groups.

Wearable devices are electronic smart devices delivered in a range of form factors including accessories and clothing that can sense biological and environmental factors and perform influential predictive computations. Most recent wearables, such as the Apple Watch (Apple Inc), have opened new avenues for capturing important physiological data such as measuring heart rate, tracking sleeping, and capturing electrocardiogram data. While wearables have demonstrated their potential to improve the health of low-SES communities by helping them increase physical activity through fitness trackers [9], research shows that wearable devices have primarily been designed with the participation of more affluent communities, and low-SES communities have historically been and continue to be excluded from their design [2-6]. The exclusion of low-income communities in technological design can have unintended and harmful consequences [10]. For example, the photoplethysmography sensor, commonly used for measuring heart rate, detecting arrhythmia, and tracking sleep, has been shown to be less accurate or fail to function properly for individuals with darker skin tones [11,12]. Given that poverty rates are over twice as high for Black and Hispanic or Latine or Latinx groups (hereby referred to as Latine) in comparison to White individuals [13,14] inaccurate readings from a photoplethysmography sensor may disproportionately impact low-SES communities. This is particularly troubling, given that our research participants all identified as members of low-income Latine communities. There is a critical need for wearables to overcome existing access and accuracy issues for these marginalized communities and to be developed directly with community members to ensure goal and value alignment.

Objectives

The development and design of current health wearables have predominantly been driven by the experiences of affluent communities. Consequently, there is a need to collaborate with low-SES community members to better understand the potential of wearables for addressing their health needs, goals, and experiences. Through an exploratory qualitative research study, we aimed to understand (1) how members of low-SES communities perceive mobile health (mHealth) wearable devices and (2) identify the barriers and facilitators toward adoption. The results point to requirements and recommendations for designing and developing better mHealth wearable devices more equitably, enabling low-SES community members to understand and self-manage their health and well-being.

Methods

Study Design

This exploratory qualitative study aimed to identify low-income community members' perspectives and needs toward wearable devices. This study was part of a larger project investigating low-SES community members' perspectives of wearable devices and the barriers and facilitators toward adoption [15]. We conducted 19 semistructured interviews with members of low-SES Latine communities from 2 metropolitan cities in the United States. Interviews were conducted separately in 2 different rounds. Our initial interviews, consisting of 8 participants, aimed at understanding low-SES community members' perspectives on wearable technology more broadly to obtain an understanding of community members' needs and expectations. Though questions on the impacts of COVID-19 and its relation to wearable devices were not asked, the topic was brought up by participants so frequently that we felt the topic warranted more in-depth discussion. Therefore, we ran a second round of interviews with 11 additional participants, in which questions were focused on participants' perspectives on how mHealth wearables can support them in their everyday lives. This research study was conducted from December 2021 to March 2022. Due to the COVID-19 Omicron variant, all interviews were conducted in English over a Zoom (Zoom Video Communications, Inc) video call. The results are reported following the Standards for Reporting Qualitative Research by O'Brien (see Multimedia Appendix 1) [16].

Recruitment

We were interested in working with low-SES community members in the United States. Flyers containing information about this study and a link to a Qualtrics (Qualtrics) screening survey were posted on the research team's social media sites, such as Facebook and Instagram (Meta Platforms). Interested participants completed a screening survey that asked for basic demographic information (eg, race, education level, household income, and the number of persons in the household). The inclusion criteria were determined if participants identified as (1) aged older than 18 years, (2) BIPOC (Black, Indigenous, and people of color), and (3) of low-income. Criteria 3 was met if an individual's income level fell at or below the low-income threshold according to their county's Department of Housing and Community Development (in the United States, the

Department of Housing and Community development uses State Income Limits provided by the US Department of Housing and Urban Development [17]). Individuals were excluded if they did not meet all the above criteria. After an eligible participant signed a digital consent form (see Multimedia Appendix 2), an online semistructured interview session was scheduled. We note that all participants came from 2 metropolitan cities in the United States but being from a particular city or location was not part of the eligibility criterion.

Study Procedure

Overview

Interview sessions were conducted by the lead author. All interviews were 45 to 60 minutes in length and were conducted over a Zoom video call. In these interview sessions, we sought to learn about participants' opinions regarding wearable technology for health.

A preliminary round of interviews centered on exploring participants' access to Wi-Fi connectivity, technology, and any resource constraints they may encounter was conducted. Additionally, we delved into their perspectives on wearable technology, their community's perception of wearables, their likes and dislikes about current wearable devices, and the features they would like to see in future wearable devices; interview questions from the preliminary study can be found in Multimedia Appendix 3.

A second round of interviews was initiated to focus on participants' views of mHealth wearable devices, their opinions on the general health and well-being of people in their neighborhood and community, the impact of COVID-19 in their community, and the types of health information considered by participants to be useful for them. The participants were asked if they had an idea what a wearable device was, and if they were unfamiliar, the interviewer described a brief definition and example. Interview questions from the second part of this study can be found in Multimedia Appendix 4. Interviews were halted once saturation was reached. All interview sessions were audio recorded and participants were compensated with a US $40 gift card at the end of the interview.

Data Analysis

Audio recordings of the interviews were transcribed, resulting in a total of 21 hours of interviews. During the preliminary investigation, the authors performed open coding on the transcripts and analyzed the data using a grounded theory approach, following the methods defined by Charmaz and Belgrave [18]. Further, 2 members of the research team performed open coding on the transcripts and identified initial themes. The research team then reviewed the transcripts and collaboratively discussed associated codes to look for consistencies and differences in the data. Through a collaborative process involving group discussions, an iterative refinement of themes was conducted. Among the emerging themes, one stood out as particularly significant: the relationship between health and wearables in the context of COVID-19. This theme was subsequently used to develop the interview guide for the second round of interviews.

In the secondary investigation, the lead author and fourth author once again performed open coding on the transcripts and analyzed the data using a grounded theory approach. Author 3 was added to the research team to help identify and narrow themes. The research team once again reviewed the transcripts and collaboratively discussed associated codes to look for consistencies and differences in the data. Codes were merged into subthemes and then grouped into 4 prominent themes. Consensus was reached by involving the sixth author to determine the final themes.

Research Reflexivity

The research team consisting of the first, third, and fourth authors are part of Latine low-SES communities. The first author, a doctoral student, recruited participants, designed this study, conducted the semistructured interviews, and analyzed the transcripts. The third and fourth authors are doctoral students and helped collect and analyze the data. All other authors contributed to drafting and revising this paper.

Ethical Considerations

This project's study protocol was reviewed and approved by Northwestern University's institutional review board (STU00216152).

Results

Participants

We recruited 19 adults from low-SES communities in 2 metropolitan cities in the United States. In total, 8 participated in the first round of interviews and 11 in the second round. All participants identified as Hispanic or Latinx and had low income. Additionally, 2 participants identified as members of the LGBTQ+ (lesbian, gay, bisexual, transgender, and queer) community. Furthermore, 94.7% (18/19) of the participants were from Los Angeles and 5% (1/19) of the participants were from Chicago. Participants' ages ranged from 18 to 54 (mean 29.7, SD 10.81) years. Table 1 summarizes participant demographics.

Table 1. Sample characteristics (N=19).

Characteristics and variable or category	Participant, n (%)
Sex	
Female	14 (78.9)
Male	5 (21.1)
Age (years)	
18-29	12 (63.2)
30-44	6 (31.5)
45-54	1 (5.26)
≥55	0 (0)
Education	
<High school	0 (0)
High school	3 (10.5)
Two years or some college	8 (42.1)
Bachelors	8 (42.1)
Graduate school	0 (0)
Income range (US $)	
<26,000	7 (36.8)
26,000-50,000	9 (47.4)
50,000-75,000	3 (15.8)
75,000-100,000	0 (0)
<100,000	0 (0)
Owns wearable	
Apple Watch	5 (26.3)
Samsung Galaxy Watch (SAMSUNG)	1 (5.26)
Fitbit (previously; Google)	2 (10.5)

Themes

Overview

Four major themes emerged from our data analysis: (1) how COVID-19 changed community members' interest in wearables for health; (2) barriers to health care resources, seeking alternatives through wearables, (3) distrust in the medical system, motivating the needs for wearables as a potential solution; and (4) community-based technical requirements. All participants said they were aware of wearable technology. We discuss our results in more depth in the following sections.

Theme 1: How COVID-19 Increased Interest in Personal Health Monitoring Through Wearables for Health

The impact of COVID-19 on underrepresented communities was significant and had a clear and direct effect of exacerbating existing health disparities [19]. Even though the COVID-19 public health emergency has ended, the residual effects of the pandemic have left individuals with long-term symptoms.

Participants (n=14) shared that they themselves had been infected with COVID-19 and had long-lasting health problems as a result. Participants (n=7) mentioned that they were still experiencing breathing problems despite having no health issues before being infected. Further, 1 participant elaborated as follows:

> *I got COVID the first time, right when the pandemic started... Afterwards it was hard for me to breathe when walking... The second time around that I got it, which was recently, it hit me a lot harder. Before I got COVID, I was healthy like there was nothing wrong with me. So, it's definitely taking its toll. I was used to walking 5 miles a day. I was walking 20 something miles a week so you know it wasn't normal for me to have breathing problems.*

As a result of experiencing ongoing COVID-19 symptoms, participants shared that for the first time, they are considering how health data, such as monitoring their oxygen levels, could be helpful. This led to conversations about how participants wished they had wearable tools that automatically measured their oxygen levels and allowed them to monitor the data themselves. Participants expressed the following:

> *I guess the one thing that scares me that I never even thought of until I got COVID were like my oxygen*

> *levels. Like, am I at normal levels? Is that an issue that I need to kind of think about, you know?*

Another participant commented:

> *One thing I noticed, especially with COVID right now which is, I think a very big topic. The timing of getting all your vitals measured can actually save somebody's life. So, I think that's a very important thing. Like oxygen levels to be measured.*

Through these interviews, we uncovered overwhelming interest in the types of data that can be provided through health wearables, primarily as a result of the long-term consequences of COVID-19 and resulting interest in engaging in personal health monitoring.

Theme 2: Barriers to Health Care Resources, Seeking Alternative Health Monitoring Through Wearables

Participants frequently discussed a severe lack of health resources and infrastructure in their communities, made worse by COVID-19, and expressed interest in potential alternatives to manage their health care through wearables to compensate for the lack of local health resources. Specifically, participants shared that the hospitals in their neighborhoods were shut down, forcing community members to seek health care in small clinics. A participant explained:

> *Most Hispanics don't have health care. I do not have that great health care myself and I have two jobs ... We lost, we had a hospital down the street, and they went bankrupt. Right now, all we really have is small clinics. So, I'm pretty sure that's all the help that anybody around here can get and it's really busy.*

Many participants (n=11) mentioned that these small clinics were completely inaccessible due to overcrowding.

> *It's overly populated. Even if you make an appointment, you're there all day. Whatever time you go, whatever day you go, it's always crowded, because it's one of the very few [clinics] that accepts Medi-Cal. So low-income communities, they don't have the resources, it's always crowded.*

Lack of health care coverage was another stressor all participants cited in conversation. Participants mentioned they could not be seen or treated and were left to fend for themselves. Individuals who experienced more critical symptoms were not able to receive treatment and passed away.

> *Health coverage was a very big issue, because not everybody was able to afford to, let's say, be able to get seen or get treated. A lot of stuff was very limited, especially to the community. If you got sick, well, your best guess to do was rub VapoRub on yourself cause that's what we only have. … A lot of these people didn't get treated and passed [away].*

Many participants felt the impact of COVID-19 could have been mitigated if the proper health care infrastructures were implemented. However, the idea of having wearables that can potentially measure vital signs and other health parameters that do not require doctor visits came up often in conversations as a practical alternative. Participants shared how they felt wearables could be useful for individuals who live in underserved areas and who do not have access to the proper health care infrastructure.

> *If [the wearable] was easy for people to use then for sure. Just knowing when people who are in areas that need more resources or who need more hospitals. Like knowing what's going on with people without having them come into the hospital to find out, because how often do people really go to the clinic and get their vitals checked, so just maybe having an idea of what is going on with people beforehand, that would be really cool.*

Thus, we found that undersourced and poor health care infrastructure has led participants to express interest in using wearable technologies for health self-management.

Theme 3: Distrust in Health Care Infrastructure and Systems Is Motivating a Need for Wearables

We found that community members' distrust of their health systems increased their interest in health wearables. Participants described experiences that led to their mistrust and fear of the health care system. For example, being turned away from immediate hospital care resulted in a participant's family member's medical conditions deteriorating leading to their death.

> *My uncle, that's another issue, they didn't want to give him a covid test for some reason and he just stayed at home for like 2-3 weeks feeling sick and only [when]he had very large symptoms he [got seen at] the hospital, and he died there two days later.*

The participants expressed that inadequate infrastructure and being denied medical assistance have made digital health care tools, like wearables, a valuable option for health testing and monitoring. The participants desired tools that could enable them to receive timely care and prevent late hospital visits. A participant highlighted that community members' lack of trust in doctors, coupled with high medical expenses, made seeking medical help unfeasible. Nonetheless, they suggested that this distrust in the health system could motivate people to take charge of their health and use wearable devices to conduct essential vital tests, rather than relying solely on medical professionals.

> *I think if you're worried about a wearable device running tests on people, cause let's be honest. Hispanic people don't go to the doctor because they don't believe in the doctor. They think the doctors are gonna kill them and then they're poor, so they can't pay for the doctor. So, like if [a wearable] could do basic [vital] tests that would be great.*

Participants suggested that wearables could offer patients a new avenue to engage in health care, especially in cases where conventional medical services are not accessible or when patients are hesitant to rely on data collected by their physicians. By using wearables instead, users can set and track their own health goals through data, alleviating dependence on the health care system to gain ownership over personal preventative health strategies. Consequently, participants also expressed a desire

to have ownership of their health data and to better understand the significance of this information for their personal health.

> *I think the biggest benefit to wearables will probably be the health aspect, like if you want to be constantly measuring your heart rate or measuring how many steps you walked in a day, I think that's where those devices will be very useful.*

Further, 17 other participants echoed this comment and expressed enthusiasm for having the ability to track health data that is important to them. Additionally, 18 out of the 19 participants expressed a strong interest in adopting wearables to support their health goals. Only 1 participant felt that they did not see the benefits of current wearables, specifically because they felt they could cheat on how their physical activity is being tracked by moving their hand around instead of moving to obtain their steps. Table 2 provides an overview of the most sought-after health parameters for monitoring. Notably, we observed that individuals from low-SES communities demonstrated a heightened interest in wearables and identifying the most critical health information to track.

Table 2. Signals participants (N=19) want to capture through wearable sensors.

Signal category	People, n
Heart rate sensors	18
Physical activity	17
Oxygen levels	13
COVID-19 symptoms	11
Blood pressure	11
Diabetes	10
Vitals	10
Breathing	7
Temperature	5
Stress	4
Glucose	4
Mental health	3
Cholesterol	3
Blood sugar	3
Allergies	1

Theme 4: Community-Based Technical Requirements

Participants shared several contextual considerations on how wearables can be designed to meet the needs of marginalized groups. The most common technical requirements emphasized throughout the interviews were durability, autonomy over data, and affordability.

Durability for Employment Reasons

As mentioned earlier, participants live in communities where most of the population identify as Hispanic or Latinx and they tend to work in more physically demanding environments. Hence, a common preference participants discussed if a wearable device were to be designed to meet the constraints of low-income communities was durability. A participant elaborated on the importance of durability:

> *I do think that it has to be very durable because the purpose is [for] low income communities. They don't have money to replace it. We just don't have comfy jobs. A lot of us work more physically demanding jobs. Some of us are plumbers, some are construction workers, some of us are gardeners. Some of us run a business and like that business involves pots and pans like we're restaurant workers. If [the device] breaks, they're just gonna say oops and throw it away. Or they're gonna like cry about it and be really upset. And they're like how do I pay for this again, you know? If it is more durable that's one of the biggest keys to wearing it.*

Participants demonstrated a significant inclination toward health wearables, specifically those that monitor vital signals such as oxygen levels. However, since such wearables must be worn continuously throughout the day, their durability becomes a critical factor. Therefore, it is imperative to ensure that these wearables are built to last and withstand the rigors of daily wear.

Autonomy: Would Rather Have Control Over Data

Privacy and having control over their own data was another critical component participants discussed. Participants did not feel comfortable having their data shared with big tech companies or a single health system due to the mistrust of both entities:

> *I think that's actually one of the biggest things that stops me from getting these, 'smart stuff.' It can measure your heart rate, has your location and all that. That sounds cool but for me my concern is how do I know if they're not sending it to a server? If it's*

> *being sent to server and I wouldn't feel comfortable with [tech companies] having access to all that data and just knowing where I'm going, what my vitals are that's just for me a bit Orwellian. I think the big thing for me is feeling comfortable and having control [over data].*

Participants shared conflicting viewpoints on this topic. For instance, 1 participant said:

> *People will be skeptical, you know they'll say, 'oh, now they're gonna read my mind or something'. Especially in low-income communities, because unfortunately we don't have the same education. A lot of people rush to conclusions and say ' oh, that's not real' or 'they're just trying to track me'. So I am thinking from that point of view, because I do have relatives that think that way. But at the same time, you always have the batch of people who are like, well, you know, let's try this [technology].*

Overall, these findings highlight diverse opinions regarding privacy and information sharing with big tech companies.

Affordability: Can Health Insurance Cover the Cost of Wearable Devices?

All participants mentioned that affordability was a barrier to adopting wearable devices, in particular wearables that are designed for health purposes.

> *I feel it's really limited because you either get a cheaper wearable technology and it's just not that advanced or not as reliable or you get expensive wearable technology that is honestly out of your price range. We would have to save up for it. And if you don't have access to like insurance then it sucks. My mom doesn't have access to one of those skin sugar sensors they have come out with. because insurance just doesn't cover it… wearable technology it definitely works, it's just expensive.*

Community members often face a challenging trade-off between functionality and affordability. Further efforts are necessary to cultivate innovative strategies that can ensure the reliability and affordability of such devices. Participants highlighted the potential health impacts of wearable devices on their lives, but they also expressed concerns about the way wearables are currently designed. Table 3 summarizes our study's findings and implications.

Table 3. Summary of key findings and implications.

Theme	Summary of participant needs	Recommendations in technical solutions moving forward
Theme 1: how COVID-19 increased interest in personal health monitoring through wearables for health	• Increased interest in tools to monitor health signals through wearables to monitor long-term COVID-19 symptoms.	• Hardware: develop embedded electronic sensors in mobile health and clinical tools that do not perpetuate racial disparities. • Software: algorithms deployed for signal processing and machine learning must not perpetuate bias.
Theme 2: interest in self-checking vital signs in response to health care access barriers	• In light of systemic challenges, participants wanted alternatives to self-manage their health care through wearables when limited resources are available.	• Develop low-cost, low-power sensors that are still capable of measuring physiological signals robustly. • Repurpose and upcycle existing hardware components. • Open-source hardware and software for transparency and reliability. • Increase service life for affordability and repairability.
Theme 3: distrust in health care infrastructure and systems is motivating a need for wearables	• Participants expressed a desire to use health wearables to set and monitor personal health goals, with a focus on preventive health measures, such as increased exercise.	• Develop wearables that robustly measure important physiological signals (as shown in Table 2) and allow individuals to self-manage these health parameters.
Theme 4: community-based technical requirements	• Participants require durability due to physically demanding blue-collar jobs. • Participants held varying opinions on privacy and the sharing of their health data with large tech corporations. • Cost was the most commonly discussed barrier to adoption.	• Wearable devices must be long-lived, durable, and adaptable for individuals who work in occupations that demand physical labor. • Wearables should allow users to have autonomy over how their data are disseminated. • Enhancing the affordability of wearables is crucial to ensure wider accessibility.

Discussion

Principal Findings

Overview

This study examined how members of minoritized low-income communities perceive mHealth wearable devices, what their needs and preferences are for using wearable devices, and understood the important contextual considerations for designing wearables in these communities.

The participants of this study expressed a newfound interest in health wearables for addressing their most critical health needs. Participants expressed a need for wearable devices that enable them to track vital signs and longitudinal symptoms (particularly those related to COVID-19) for setting and achieving preventive health goals, as illustrated in Table 2. Additionally, the participants emphasized that the wearable devices must be durable, reasonably priced, and within their budget or covered by their insurance, particularly state-provided insurance. It is evident from our study that the significant consequences of the COVID-19 pandemic on low-SES communities have led to an increased interest in health engagement and, subsequently, health wearables. However, current wearable devices fail to account for these individuals' lived experiences. Below, we elaborate on this, discussing how mHealth wearable devices can be better designed to improve the health care experience of low-income community members by integrating with medical systems, increasing autonomy, and empowering them to make informed decisions about their care.

Overcoming Technology Limitations

Though participants expressed strong interest in using wearable devices to improve and promote their understanding of their health, existing wearables are not currently built to meet their needs.

According to Table 2, monitoring heart rate, physical activity, and oxygen levels were the top 3 areas of interest for participants. However, current commercial wearables' optical sensors and the signal processing methods used to measure physiological signals are not reliable on darker skin tones [11]. This is a troubling concern because our study's participants identify as Latine or Latinx, which are a minoritized group of color, are members of low-SES communities, and they stand to be disproportionately affected by inaccurate photoplethysmography sensors.

Photoplethysmography sensors are widely used in pulse oximeters to measure blood oxygen saturation levels. However, earlier research has demonstrated that the accuracy of pulse oximeters decreases when used on individuals with darker skin [19-22]. Sjoding et al [22] highlighted the dangers of relying on pulse oximeters, particularly during the COVID-19 pandemic. Patients who identified as Black were three times more likely to be incorrectly classified as having normal oxygen levels in the blood than their White counterparts. Such misclassification can result in a failure to detect hypoxemia (low oxygen levels in the blood), which can lead to severe health consequences. To address this issue, the Food and Drug Administration has cautioned against using pulse oximeters on individuals with dark skin tones, thick skin, poor circulation, and other factors that can impact the precision of the results [23].

Given the significant limitations of this technology, we see a clear need to (1) develop new embedded electronic sensors in mHealth and clinical tools that do not perpetuate racial disparities and (2) in software, we must ensure that the algorithms deployed for signal processing and machine learning do not perpetuate bias.

Regarding the development of new sensors, a couple of avenues of research hold promise. In light of the complications found in photoplethysmography sensors, researchers are now looking at multiwavelength photoplethysmography signals to target a deeper range of measuring blood pressure and other cardiovascular parameters from the skin and show potential for being implemented in future wearables [24]. Single-channel bioimpedance [25] is another method being investigated by researchers and shows promise for potentially being implemented in future wearable devices. The key to making these hardware modifications usable in future wearable devices and minimizing racial bias is to diversify subject testing and include individuals with a broad spectrum of skin tones.

The signal processing, machine learning, and high-level algorithmic approach to prediction from these sensors must be considered at the software level so as not to perpetuate racial bias. Software mitigations for poor sensor resolution and precision are commonplace in critical systems—similar approaches are needed here. Wearable devices have the potential to perpetuate racial bias unless addressed. Many machine learning approaches to things like recidivism prediction for parolees [26], mortgage loans [27], and facial recognition [28] have already proven that without care and attention in their design and training, these software systems perpetuate racial discrimination among racially minoritized individuals. Software and firmware developers of wearables must ensure that the algorithms they deploy to capture data from biomedical sensors do not perpetuate these racial harms. Our study extends this work by establishing wearable devices as an important application for these sensor and software developments. By leveraging more equitable research on both the hardware and software level, we can address the health concerns exacerbated by the COVID-19 pandemic.

Durability and Adaptability for Diverse User Groups

Designing wearables acutely attuned to resource-constrained populations requires ensuring that the tools are both durable and scalable. As mentioned by participants, if wearables are being designed with low-income community members in mind, they must be made more durable as most individuals from these communities work in "more physically demanding jobs."

Taking into consideration their income and employment concerns, we see a need to ensure that wearables built to serve the general public also scale to this already underserved subset of the broader population. This is particularly important since literature has shown that technology that serves broader populations may not scale to subgroups [29,30].

Our findings point to the need to rethink the development of wearable technology. Strategies for improving the durability and affordability of wearable devices are to develop and make use of low-cost low-power sensors that are still capable of robustly sensing physiological signals, reuse and upcycle existing materials such as hardware components, increase service life by allowing wearable devices to be repairable inexpensively, open-sourcing hardware and software would allow wearables to be more affordable because people will not have to pay for licensing fees, and exploring different form-factors of wearable design beyond smartwatches. We argue that devices that can withstand the pressures of the everyday difficulties faced by our study population would better serve them and be a tool by which teams building wearables encode the values and preferences of underserved populations, thus promoting health equity.

Autonomy Over Data

In contrast to previous research where test subjects claimed they would like to share their data with health care providers [31], participants in our study were skeptical about sharing too much information with their health care providers due to a lack of trust in the health care system and fear of being monitored by big technology companies. Prior research discusses that the distrust of technology and the medical system among racialized low-SES communities is not new and increases barriers to adopting technology [32-34]. However, in our study, most participants believed that they should have more control over how their data is used and shared. Thus, we see interesting and important tension in participants' desire to use health technologies but concern about using them in partnership with health care systems due to a history of distrust. This points to important future research on how such tools might be able to foster patient or provider collaboration and communication within these communities where such distrust is prominent.

Serious privacy concerns regarding the data collected by wearable devices have also been reported and highlight the lack of protection of consumers' data [35]. Health insurance and tech companies such as Aetna and Apple have partnered to offer Apple watches at a discounted price; customers have to meet their fitness goals in 24 months or pay the full price if goals are not met [36]. The implications of these partnerships between insurance companies and tech companies raise privacy concerns as the health data collected by wearables allow insurance companies to determine which customers seem profitable and can raise the premium rates or deny insurance for others who are not meeting their fitness goals [37]. Apart from the privacy concerns, the potential consequences of insurance companies using data collected by wearable devices to raise insurance rates or deny health insurance to customers can lead to further health disparities for individuals from low-SES communities who may not have the funds to purchase the device even at a discounted price or have the ability to meet their fitness requirements due to other family and work responsibilities. Members of low-SES communities are already being denied access to health care, wearables should not be used as another tool to perpetuate health disparities among these populations.

Studies on contact tracing [38,39], COVID-19 mobile apps [40], and health informatics adoption among low-income individuals [32,41] have shown that data privacy is important to general users. This is especially true for low-income populations due to their mistrust of the health care system. Previous research in human-computer interaction has demonstrated the feasibility of creating privacy-preserving tools for mobile apps [42,43], software tools [44-46], and Internet of Things devices [47-49] that help users retain control and autonomy over their data.

Despite the significant progress in disseminating privacy tools to the general public, these tools have yet to be extended to the wearable space. Therefore, further research is necessary to develop privacy-focused tools for wearables, such as mobile apps or wearable-based interfaces, that enable individuals to manage and comprehend their data flow. Additionally, these privacy tools must be accessible and user-friendly to promote fair use among underserved communities.

Potential solutions for addressing privacy concerns call for stronger privacy regulations and encouraging tech companies to be more transparent about their data collection and usage. Additionally, individuals must have the autonomy to selectively share data that they feel comfortable disclosing to health care providers and tech companies. This can be facilitated through user-centric mobile apps or wearable interface privacy tools that empower people to understand and manage the flow of their data.

Beyond addressing participants' privacy concerns and data autonomy needs, we see an interesting and important tension between participants' desire to use health technologies and concern about using them in partnership with health care systems due to a history of distrust. While health wearables provide tools that can be empowering and engaging for patients, they are not meant to circumvent health care systems or medical care. We see a critical need for future research on how such tools might be able to foster patient or provider collaboration and communication within these communities where such distrust is prominent.

Affordability

The 18 participants who expressed interest in adopting wearables in their daily lives all mentioned that affordability was the strongest barrier toward adoption. These findings are in line with previous works [1,15,31]. Of participants who already owned a wearable device, 3 mentioned that they were gifted with the device and would not have been able to afford one otherwise. The remaining participants who already owned a wearable reported that they either "*bought it during a Black Friday sale*" or saved up to buy one.

Possible solutions for making wearables more affordable include exploring form factors of wearables that can be redesigned to be more affordable. Many wearables are overdesigned and engineered for sleekness, small size, and fashion, as well as feature sets that may not align with the needs of low-SES communities. In essence, the question is what specific features might be more useful than having a general-purpose platform that does it all (but therefore is much more expensive). Similar to how low-feature phones (text, phone call, and light browsing

ability) are a 10th of the price of a smartphone, wearable devices must also explore this avenue of sacrificing features (like sophisticated GPUs and video streaming accelerators or advanced machine learning processing) for things like optical respiratory rate measurement.

From an academic perspective, as stated previously, including low-SES communities in research studies is important to diversify participant populations to reduce potential racial bias. However, wearable researchers can take a step further and allow participants from low-SES communities to keep any technology used within the research. Giving people technology for a 2-month deployment and then taking it away when they do not have the means to obtain something like that themselves is a serious ethical concern.

Limitations and Future Directions

We acknowledge that our work has sampling limitations. This study was exploratory, and though our intentions were not meant to only target participants from Hispanic or Latine low-SES communities from 2 metropolitan cities, we were not able to sample the perceptions of other racially minoritized BIPOC communities from smaller cities or rural areas. City representation was likely due to the research team members being from these 2 cities. In future work, examining the differences in how members from other racially minoritized groups perceiving wearables could vary would provide valuable insights.

Conclusions

This research study aimed to investigate the needs and perspectives of individuals from low-income communities regarding the adoption and usage of wearable devices. Our findings indicate that there is considerable interest among members of these communities in employing wearables to promote their well-being. Participants expressed frustration with the current health system citing how the lack of health resources and the health effects of the COVID-19 pandemic has led them to seek alternative methods to manage their health. Participants recommended design considerations for the utility of wearable devices that included, durability, sustainability, and accessibility. Additionally, autonomy on how their data is used was important for the majority of participants. Affordability was the primary barrier to the adoptability of wearable devices. Participants believe that if health insurance companies can help pay for partial costs of wearable devices, more people in the community would be more interested in using them. The insights from this study serve as a first step for researchers and technology companies in the domain of wearable technology to develop tools that account for the contextual and cultural perspectives of low-SES communities to help democratize the utility of wearable devices.

Acknowledgments

This research is based upon work supported by the National Science Foundation (CNS-2145584 and CNS-2107400). The first author was also funded by a doctoral PhD fellowship from Microsoft Research to fund her and this work. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation or Microsoft Research.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Standards for reporting qualitative research checklist.
[DOCX File , 18 KB - mhealth_v12i1e50826_app1.docx]

Multimedia Appendix 2
Study consent form.
[PDF File (Adobe PDF File), 89 KB - mhealth_v12i1e50826_app2.pdf]

Multimedia Appendix 3
Interview guidelines and questions round 1.
[DOCX File , 14 KB - mhealth_v12i1e50826_app3.docx]

Multimedia Appendix 4
Interview guidelines and questions round 2.
[DOCX File , 16 KB - mhealth_v12i1e50826_app4.docx]

References

1. Holko M, Litwin TR, Munoz F, Theisz KI, Salgin L, Jenks NP, et al. Wearable fitness tracker use in federally qualified health center patients: strategies to improve the health of all of us using digital health devices. NPJ Digit Med 2022 Apr 25;5(1):53 [FREE Full text] [doi: 10.1038/s41746-022-00593-x] [Medline: 35469045]

2. Yao R, Zhang W, Evans R, Cao G, Rui T, Shen L. Inequities in health care services caused by the adoption of digital health technologies: scoping review. J Med Internet Res 2022;24(3):e34144 [FREE Full text] [doi: 10.2196/34144] [Medline: 35311682]
3. Dubay LC, Lebrun LA. Health, behavior, and health care disparities: disentangling the effects of income and race in the United States. Int J Health Serv 2012;42(4):607-625 [FREE Full text] [doi: 10.2190/HS.42.4.c] [Medline: 23367796]
4. Khanijahani A, Iezadi S, Gholipour K, Azami-Aghdash S, Naghibi D. A systematic review of racial/ethnic and socioeconomic disparities in COVID-19. Int J Equity Health 2021;20(1):248 [FREE Full text] [doi: 10.1186/s12939-021-01582-4] [Medline: 34819081]
5. Kirksey L, Tucker DL, Taylor E, White Solaru KT, Modlin CS. Pandemic superimposed on epidemic: Covid-19 disparities in Black Americans. J Natl Med Assoc 2021;113(1):39-42 [FREE Full text] [doi: 10.1016/j.jnma.2020.07.003] [Medline: 32747313]
6. Kondo N. Socioeconomic disparities and health: impacts and pathways. J Epidemiol 2012;22(1):2-6 [FREE Full text] [doi: 10.2188/jea.je20110116] [Medline: 22156290]
7. Krouse HJ. COVID-19 and the widening gap in health inequity. Otolaryngol Head Neck Surg 2020;163(1):65-66 [FREE Full text] [doi: 10.1177/0194599820926463] [Medline: 32366172]
8. Alcaraz KI, Wiedt TL, Daniels EC, Yabroff KR, Guerra CE, Wender RC. Understanding and addressing social determinants to advance cancer health equity in the United States: a blueprint for practice, research, and policy. CA Cancer J Clin 2020;70(1):31-46 [FREE Full text] [doi: 10.3322/caac.21586] [Medline: 31661164]
9. Levine JA. The application of wearable technologies to improve healthcare in the world's poorest people. TI 2017;08(02):83-95 [FREE Full text] [doi: 10.4236/ti.2017.82007]
10. Glied S, Lleras-Muney A. Technological innovation and inequality in health. Demography 2008;45(3):741-761 [FREE Full text] [doi: 10.1353/dem.0.0017] [Medline: 18939670]
11. Colvonen PJ, DeYoung PN, Bosompra NO, Owens RL. Limiting racial disparities and bias for wearable devices in health science research. Sleep 2020;43(10):zsaa159 [FREE Full text] [doi: 10.1093/sleep/zsaa159] [Medline: 32893865]
12. Shcherbina A, Mattsson CM, Waggott D, Salisbury H, Christle JW, Hastie T, et al. Accuracy in wrist-worn, sensor-based measurements of heart rate and energy expenditure in a diverse cohort. J Pers Med 2017;7(2):3 [FREE Full text] [doi: 10.3390/jpm7020003] [Medline: 28538708]
13. Lin AC, Harris DR. The Colors of Poverty: Why Racial and Ethnic Disparities Persist. New York, New York: Russell Sage Foundation; 2009:43091.
14. Creamer J. Inequalities persist despite decline in poverty for all major race and hispanic origin groups. Census.gov. URL: https://www.census.gov/library/stories/2020/09/poverty-rates-for-blacks-and-hispanics-reached-historic-lows-in-2019.html [accessed 2023-12-18]
15. Cruz S, Redding A, Chau CW, Lu C, Persche J, Hester J, et al. EquityWare: Co-designing wearables with and for low income communities in the U.S. USA: Association for Computing Machinery; 2023 Presented at: CHI '23: CHI Conference on Human Factors in Computing Systems; April 23-28, 2023; Hamburg Germany. [doi: 10.1145/3544548.3580980]
16. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: a synthesis of recommendations. Acad Med 2014;89(9):1245-1251 [FREE Full text] [doi: 10.1097/ACM.0000000000000388] [Medline: 24979285]
17. Office of Policy Development and Research (PD and R). URL: https://www.huduser.gov/portal/datasets/il.html [accessed 2024-04-10]
18. Charmaz K, Belgrave LL. Qualitative interviewing and grounded theory analysis. In: Gubrium JF, editor. The SAGE Handbook of Interview Research: The Complexity of the Craft. Thousand Oaks, Calif: SAGE Publications; 2012:347-365.
19. Bickler PE, Feiner JR, Severinghaus JW. Effects of skin pigmentation on pulse oximeter accuracy at low saturation. Anesthesiology 2005;102(4):715-719 [FREE Full text] [doi: 10.1097/00000542-200504000-00004] [Medline: 15791098]
20. Feiner JR, Severinghaus JW, Bickler PE. Dark skin decreases the accuracy of pulse oximeters at low oxygen saturation: the effects of oximeter probe type and gender. Anesth Analg 2007;105(6 Suppl):S18-S23 [FREE Full text] [doi: 10.1213/01.ane.0000285988.35174.d9] [Medline: 18048893]
21. Moço AV, Stuijk S, de Haan G. Skin inhomogeneity as a source of error in remote PPG-imaging. Biomed Opt Express 2016;7(11):4718-4733 [FREE Full text] [doi: 10.1364/BOE.7.004718] [Medline: 27896011]
22. Sjoding MW, Dickson RP, Iwashyna TJ, Gay SE, Valley TS. Racial bias in pulse oximetry measurement. N Engl J Med 2020;383(25):2477-2478 [FREE Full text] [doi: 10.1056/NEJMc2029240] [Medline: 33326721]
23. Jubran A, Tobin MJ. Reliability of pulse oximetry in titrating supplemental oxygen therapy in ventilator-dependent patients. Chest 1990;97(6):1420-1425 [FREE Full text] [doi: 10.1378/chest.97.6.1420] [Medline: 2347228]
24. Garmin Ltd. Garmin watch optical heart rate accuracy tips. Garmin Customer Support. URL: https://support.garmin.com/en-US/?faq=xQwjQjzUew4BF1GYcusE59 [accessed 2023-07-13]
25. Liu J, Yan BP, Zhang YT, Ding XR, Su P, Zhao N. Multi-Wavelength photoplethysmography enabling continuous blood pressure measurement with compact wearable electronics. IEEE Trans Biomed Eng 2019;66(6):1514-1525 [FREE Full text] [doi: 10.1109/TBME.2018.2874957] [Medline: 30307851]

26. Angwin J, Larson J, Mattu S, Kirchner L. Machine bias. ProPublica. URL: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing [accessed 2023-07-13]
27. Bartlett R, Morse A, Stanton R, Wallace NE. Consumer-Lending Discrimination in the FinTech Era. Cambridge, MA: National Bureau of Economic Research; 2019.
28. Buolamwini J, Gebru T. Gender shades: intersectional accuracy disparities in commercial gender classification. Conf Fairness Account Transparency Proc Mach Learn Res 2018;81:77-91 [FREE Full text]
29. Grimes A, Grinter RE. Designing persuasion: health technology for low-income African American communities. In: De Kort Y, IJsselsteijn W, Midden C, Eggen B, Fogg BJ, editors. Persuasive Technology. Heidelberg: Springer Berlin Heidelberg; 2007:24.
30. Orji R, Mandryk RL. Developing culturally relevant design guidelines for encouraging healthy eating behavior. Int J Hum Comput Stud 2014;72(2):207-223 [FREE Full text] [doi: 10.1016/j.ijhcs.2013.08.012]
31. Lima FV, Kadiyala V, Huang A, Agusala K, Cho D, Freeman AM, et al. At the crossroads! time to start taking smartwatches seriously. Am J Cardiol 2022;179:96-101 [FREE Full text] [doi: 10.1016/j.amjcard.2022.06.020] [Medline: 35842279]
32. Veinot TC, Mitchell H, Ancker JS. Good intentions are not enough: how informatics interventions can worsen inequality. J Am Med Inform Assoc 2018;25(8):1080-1088 [FREE Full text] [doi: 10.1093/jamia/ocy052] [Medline: 29788380]
33. Jimison H, Gorman P, Woods S, Nygren P, Walker M, Norris S, et al. Barriers and drivers of health information technology use for the elderly, chronically ill, and underserved. Evid Rep Technol Assess (Full Rep) 2008(175):1-1422 [FREE Full text] [Medline: 19408968]
34. Armstrong K, Ravenell KL, McMurphy S, Putt M. Racial/ethnic differences in physician distrust in the United States. Am J Public Health 2007;97(7):1283-1289. [doi: 10.2105/AJPH.2005.080762] [Medline: 17538069]
35. Montgomery K, Chester J, Kopp K. Health wearables: ensuring fairness, preventing discrimination, and promoting equity in an emerging internet-of-things environment. J Inf Policy 2018;8:34-77 [FREE Full text] [doi: 10.5325/jinfopoli.8.2018.0034]
36. Farr C. Apple and Aetna are teaming up on a new app to help track and reward healthy behavior. CNBC. 2019. URL: https://www.cnbc.com/2019/01/28/apple-aetna-team-up-on-attain-health-tracking-app.html [accessed 2023-07-02]
37. Raber I, McCarthy CP, Yeh RW. Health insurance and mobile health devices: opportunities and concerns. JAMA 2019;321(18):1767-1768 [FREE Full text] [doi: 10.1001/jama.2019.3353] [Medline: 30973573]
38. Bonnell TJ, Revere D, Baseman J, Hills R, Karras BT. Equity and accessibility of Washington State's COVID-19 digital exposure notification tool (WA Notify): survey and listening sessions among community leaders. JMIR Form Res 2022;6(8):e38193 [FREE Full text] [doi: 10.2196/38193] [Medline: 35787520]
39. Kolasa K, Mazzi F, Leszczuk-Czubkowska E, Zrubka Z, Péntek M. State of the art in adoption of contact tracing apps and recommendations regarding privacy protection and public health: systematic review. JMIR mHealth uHealth 2021;9(6):e23250 [FREE Full text] [doi: 10.2196/23250] [Medline: 34033581]
40. Idris MY, Korin M, Araya F, Chowdhury S, Medina P, Cruz L, et al. Including the public in public eHealth: the need for community participation in the development of state-sponsored COVID-19-related mobile apps. JMIR mHealth uHealth 2022;10(3):e30872 [FREE Full text] [doi: 10.2196/30872] [Medline: 35113793]
41. Niazkhani Z, Toni E, Cheshmekaboodi M, Georgiou A, Pirnejad H. Barriers to patient, provider, and caregiver adoption and use of electronic personal health records in chronic care: a systematic review. BMC Med Inform Decis Mak 2020;20(1):153 [FREE Full text] [doi: 10.1186/s12911-020-01159-1] [Medline: 32641128]
42. Jin H, Liu M, Dodhia K, Li Y, Srivastava G, Fredrikson M, et al. Why are they collecting my data? Proc ACM Interact Mob Wearable Ubiquitous Technol 2018;2(4):1-27 [FREE Full text] [doi: 10.1145/3287051]
43. Chitkara S, Gothoskar N, Harish S, Hong JI, Agarwal Y. Does this app really need my location?: Context-aware privacy management for smartphones. Proc ACM Interact Mob Wearable Ubiquitous Technol 2017;1(3):1-22 [FREE Full text] [doi: 10.1145/3132029]
44. Sheng S, Magnien B, Kumaraguru P, Acquisti A, Cranor L, Hong J, et al. Anti-phishing phil: the design and evaluation of a game that teaches people not to fall for phish. USA: Association for Computing Machinery; 2007 Presented at: SOUPS '07: The third Symposium on Usable Privacy and Security; July 18-20, 2007; Pittsburgh Pennsylvania USA. [doi: 10.1145/1280680.1280692]
45. Li T, Agarwal Y, Hong JI. Coconut: an IDE plugin for developing privacy-friendly apps. Proc ACM Interact Mob Wearable Ubiquitous Technol 2018;2(4):1-35 [FREE Full text] [doi: 10.1145/3287056]
46. Schaub F, Marella A, Kalvani P, Ur B, Pan C, Forney E, et al. Watching them watching me: browser extensions impact on user privacy awareness and concern. : Internet Society; 2016 Presented at: Proceedings 2016 Workshop on Usable Security; February 21, 2016; San Diego, CA. [doi: 10.14722/usec.2016.23017]
47. Huang DY, Apthorpe N, Li F, Acar G, Feamster N. IoT Inspector: crowdsourcing labeled network traffic from smart home devices at scale. Proc ACM Interact Mob Wearable Ubiquitous Technol 2020;4(2):1-21 [FREE Full text] [doi: 10.1145/3397333]
48. Wan Y, Xu K, Wang F, Xue G. IoTAthena: unveiling IoT device activities from network traffic. IEEE Trans Wireless Commun 2022;21(1):651-664 [FREE Full text] [doi: 10.1109/twc.2021.3098608]

49. Cruz S, Danek L, Liu S, Kraemer C, Wang Z, Feamster N, et al. Augmented reality's potential for identifying and mitigating home privacy leaks. ArXiv Preprint posted online on January 27 2023. [doi: 10.14722/usec.2023.238930]

Abbreviations

BIPOC: Black, Indigenous, people of color
LGBTQ+: lesbian, gay, bisexual, transgender, queer
mHealth: mobile health
SES: socioeconomic status

Edited by L Buis; submitted 14.07.23; peer-reviewed by B Green, L Connell Bohlen; comments to author 08.09.23; revised version received 21.12.23; accepted 25.01.24; published 08.05.24.

Please cite as:
Cruz S, Lu C, Ulloa M, Redding A, Hester J, Jacobs M
Perceptions of Wearable Health Tools Post the COVID-19 Emergency in Low-Income Latin Communities: Qualitative Study
JMIR Mhealth Uhealth 2024;12:e50826
URL: https://mhealth.jmir.org/2024/1/e50826
doi:10.2196/50826
PMID:38717816

Original Paper

Effectiveness of a Multifaceted Mobile Health Intervention (Multi-Aid-Package) in Medication Adherence and Treatment Outcomes Among Patients With Hypertension in a Low- to Middle-Income Country: Randomized Controlled Trial

Muhammad Arshed[1,2], MBBS, MPH, PhD; Aidalina Mahmud[1], MBChB, MPH, PhD; Halimatus Sakdiah Minhat[1], MBChB, MPH, DrPH; Poh Ying Lim[1], BSc, MSc, PhD; Rubeena Zakar[3], MBBS, MPS, PhD

[1]Department of Community Health, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, Serdang, Malaysia

[2]University Institute of Public Health, Faculty of Allied Health Sciences, University of Lahore, Punjab, Lahore, Pakistan

[3]Department of Public Health, Institute of Social and Cultural Studies, University of the Punjab, Lahore, Pakistan

Corresponding Author:
Aidalina Mahmud, MBChB, MPH, PhD
Department of Community Health
Faculty of Medicine and Health Sciences
Universiti Putra Malaysia
UPM Serdang
Serdang, 43400
Malaysia
Phone: 60 397692416
Email: aidalina@upm.edu.my

Abstract

Background: The high prevalence of uncontrolled hypertension in Pakistan is predominantly attributed to poor medication adherence. As more than 137 million people in Pakistan use cell phones, a suitable mobile health (mHealth) intervention can be an effective tool to overcome poor medication adherence.

Objective: We sought to determine whether a novel mHealth intervention is useful in enhancing antihypertensive therapy adherence and treatment outcomes among patients with hypertension in a low- to middle-income country.

Methods: A 6-month parallel, single-blinded, superiority randomized controlled trial recruited 439 patients with hypertension with poor adherence to antihypertensive therapy and access to smartphones. An innovative, multifaceted mHealth intervention (Multi-Aid-Package), based on the Health Belief Model and containing reminders (written, audio, visual), infographics, video clips, educational content, and 24/7 individual support, was developed for the intervention group; the control group received standard care. The primary outcome was self-reported medication adherence measured using the Self-Efficacy for Appropriate Medication Adherence Scale (SEAMS) and pill counting; the secondary outcome was systolic blood pressure (SBP) change. Both outcomes were evaluated at baseline and 6 months. Technology acceptance feedback was also assessed at the end of the study. A generalized estimating equation was used to control the covariates associated with the probability of affecting adherence to antihypertensive medication.

Results: Of 439 participants, 423 (96.4%) completed the study. At 6 months post intervention, the median SEAMS score was statistically significantly higher in the intervention group compared to the controls (median 32, IQR 11 vs median 21, IQR 6; U=10,490, *P*<.001). Within the intervention group, there was an increase in the median SEAMS score by 12.5 points between baseline and 6 months (median 19.5, IQR 5 vs median 32, IQR 11; *P*<.001). Results of the pill-counting method showed an increase in adherent patients in the intervention group compared to the controls (83/220, 37.2% vs 2/219, 0.9%; *P*<.001), as well as within the intervention group (difference of n=83, 37.2% of patients, baseline vs 6 months; *P*<.001). There was a statistically significant difference in the SBP of 7 mmHg between the intervention and control groups (*P*<.001) at 6 months, a 4 mmHg reduction (*P*<.001) within the intervention group, and a 3 mmHg increase (*P*=.314) within the controls. Overall, the number of patients with uncontrolled hypertension decreased by 46 in the intervention group (baseline vs 6 months), but the control group remained unchanged. The variables groups (adjusted odds ratio [AOR] 1.714, 95% CI 2.387-3.825), time (AOR 1.837, 95% CI

1.625-2.754), and age (AOR 1.618, 95% CI 0.225-1.699) significantly contributed (P<.001) to medication adherence. Multi-Aid-Package received a 94.8% acceptability score.

Conclusions: The novel Multi-Aid-Package is an effective mHealth intervention for enhancing medication adherence and treatment outcomes among patients with hypertension in a low- to middle-income country.

Trial Registration: ClinicalTrials.gov NCT04577157; https://clinicaltrials.gov/study/NCT04577157

(JMIR Mhealth Uhealth 2024;12:e50248) doi:10.2196/50248

KEYWORDS

mobile health; mHealth; intervention; medication adherence; hypertension; low- to middle-income country; effectiveness; randomized controlled trial; Pakistan; drug adherence; tool; mHealth module; self-efficacy; systolic blood pressure; feedback

Introduction

Hypertension is a significant health challenge worldwide and a leading cause of morbidity and mortality [1]. In the 21st century, hypertension has become a growing health issue worldwide. It is expected to increase from 918 million individuals in 2000 to 1.56 billion in 2025 [1]. Compared to high-income countries, the prevalence of hypertension among adults is more remarkable in low- and middle-income countries (LMICs) [2,3]. Hypertension is responsible for approximately 9.4 million fatalities worldwide, making it a significant cause of death [4]. These deaths are mostly preventable, as lowering the systolic blood pressure (SBP) can lessen fatalities from all causes and cardiovascular disease (CVD) [5].

The risk of mortality due to cardiovascular events and stroke is lowered by twofold for each 20 mmHg drop in the SBP or each 10 mmHg drop in the diastolic blood pressure (DBP) between the ages of 40 and 69 years [6]. The dosage of blood pressure medications administered and adherence to therapy are 2 important aspects that influence blood pressure control in patients receiving treatment with clinically corrected blood pressure levels. Patient compliance is a critical aspect of blood pressure management, and medication is worthless for those who refuse to take it [7]. What is more worrying is the chronic nature of hypertension and the need to be compliant with the medications usually for more than 1 year. It has been noted that 1 year after starting antihypertensive medication, 50% of patients still use it [8,9]. Unfortunately, in general, the percentage of treated patients with hypertension who achieve control levels is only between 20% and 50% [2,10].

In Pakistan, hypertension is a crucial matter of public health, where nearly 19% of the youth and 33% of individuals older than 45 years have hypertension, with the majority of the population with hypertension having poor blood pressure control [11]. Poor medication adherence has been noted to contribute to poor blood pressure control in Pakistan [12]. A recent investigation found that 37.7% of patients fail to take their antihypertensive medications, as directed [12]. This situation is of concern because as mentioned, medication adherence is a proven and cost-effective treatment for hypertension [13], in addition to lifestyle modification and medical risk assessment. Furthermore, medications may lower the risk of stroke and myocardial infarction by 30% and 15%, respectively, among the population with hypertension [14]. Lower levels of adherence are connected to poorer blood pressure control and unfavorable outcomes [15].

For a few years, there has been an upsurge in the use of mobile health (mHealth) apps to improve medication compliance [16,17]. Using mobile technology, such as cell phones, personal digital assistants, patient-monitoring equipment, and other wireless devices, for medical care is referred to as “mHealth” [18]. mHealth is an ideal tool for LMICs due to its low cost and ease of use. For mHealth, all that is required are mobile devices, cellular communication technologies, and an internet connection. According to the Pakistan Telecommunication Authority, more than 137 million Pakistanis use cell phones, corresponding to a cellular density of 77% of the population [19]. However, despite the growing popularity of cell phones in LMICs, mHealth approaches in these countries remain limited.

Furthermore, no specific association between the use of mHealth apps and enhancement of medication adherence in CVD in LMICs has been shown to date [20]. Several studies have suggested that further investigations be conducted to determine whether mHealth can enhance medication adherence in LMICs [20,21] compared to traditional methods [22]. WhatsApp facilitates the collection of real-time data over both time and place. WhatsApp offers a plethora of health-related uses, including optimizing communication and the delivery of health education [23,24]. A survey found that Pakistanis primarily use social media for communication and information exchange in the health sector, with WhatsApp and YouTube being the most widely used social media platforms for health-related topics [25]. Several important observations, particularly those gleaned from the body of the existing literature [25], guide our decision to use WhatsApp in implementing this cutting-edge intervention, since it is an efficient way to provide interventions with respect to cost, time, and dissemination.

Using the Health Belief Model [26] and self-determination theory [27] as a foundation, we developed a novel mHealth intervention module for the population with hypertension in LMICs, particularly Pakistan. The module is called “Multi-Aid-Package.” It is a multifaceted intervention integrated with educational guidelines and a reminder component. The module addresses individual patients’ perspectives and concerns and incites health-related beliefs toward better medication adherence. This trial’s distinctive feature was its all-encompassing, multimodal strategy, which combined various previous interventions [28-30] into a single intervention. Second, animated images and videos were used in place of text in this

study. Therefore, as far as we are aware, this is the first study in Pakistan to develop and assess the efficacy of a comprehensive and multifaceted mHealth intervention.

Consequently, this study sought to use the mHealth-based multifaceted intervention with the aid of WhatsApp to help patients who were not adhering to their medication and to assess the efficiency of Multi-Aid-Package in optimizing adherence to medication and manage the SBP among patients with hypertension in the LMIC context. We hypothesized that this mHealth module intervention would improve medication adherence, lower the SBP, and eventually lower the mortality and morbidity due to hypertension.

Methods

Trial Design

This trial was a parallel, single-blinded, superiority randomized controlled study that lasted for 6 months and had a 2-arm, parallel design. The trial was carried out following the CONSORT (Consolidated Standards of Reporting Trails) Statement 2010 standards [31]. Participants were randomly concurrently allocated 1 of 2 groups (intervention or control) in a 1:1 ratio. The intervention group underwent the Multi-Aid-Package intervention, while the control group received regular treatment (as per the hospitals' routine practice) [32]. Evaluations were carried out at baseline and 6 months after the implementation of the intervention. The trial was registered with Clinical Trials (NCT04577157; registration date October 6, 2020) before the start of recruitment, which began on January 3, 2021.

Sampling Method and Study Setting

A 2-stage random sample procedure was used to carry out sampling. The first stage required selecting a hospital at random from a list of hospitals, and the second stage entailed randomly selecting patients with hypertension from the selected hospital.

The study site selected was a public tertiary care hospital in Punjab's provincial capital, Lahore. Lahore is Pakistan's second-biggest city and has a population of 11,302,285, with a gross domestic product (GDP) of US $84 billion [33].

Study Participants

Study participants were selected from among patients diagnosed with hypertension at the hospital's cardiology and medical outpatient departments. Patient screening involved identifying those who were registered as hypertensive for the past month. The selection process was conducted by specially assigned registrars who screened patients using the Self-Efficacy for Appropriate Medication Scale (SEAMS) [34] and by asking the patients the number of pills they consumed during a specified period [35]. Although these 2 approaches to measure medication adherence are distinct, both adherence measures were used in this trial for inclusion/exclusion considerations. Only those patients who satisfied both criteria were recruited. Due to the selection criteria's restriction to using both approaches, any participant could be classified as nonadherent. These results indicated the patients' medication adherence status. Based on these assessments and other eligibility criteria, 439 participants were selected. The data collected included sociodemographic information, health-related profiles, baseline SEAMS score, and number of pills (representing medication adherence status). All the information was collected through face-to-face interviews conducted by trained research staff. In addition, each participant's baseline SBP reading was also recorded.

Eligibility Criteria

The inclusion criteria were as follows: patients who were at least 18 years old, diagnosed with hypertension within the previous month, prescribed antihypertensive medications, had poor medication adherence (a low SEAMS score ranging from 13-21 and pill-counting rate<80% were coded as nonadherent), had a smartphone with WhatsApp installed, and had the ability to read and send messages using WhatsApp.

The exclusion criteria were as follows: patients who had plans to leave the study area during the study period that would prohibit them from accessing cell signals; had a history of cancer, as they would need medication adjustments over time; would undergo a planned surgery or intervention; had blood pressure>220/120 mmHg (in the hypertensive emergency category); or were pregnant, breastfeeding, or 3 months postpartum.

Sample Size

The sample size was estimated to assess a 1-point difference in SD (SD 2) on the major outcome metric of adherence change when comparing the 2 groups. To evaluate the 2-tailed hypothesis, the α level (type 1 error) was set at .05, with a 95% CI interval, Z=1.96, and strength to obtain a power of 90% [36]. The adherence reference value was as determined by a recent study [17]. After a 30% attrition rate, using Lemeshow et al's [37] formula:

the calculated sample size for this study was 440 participants, equally allocated to the intervention and control groups (n=220, 50%, per group).

Randomization and Concealment

A simple complete randomization method was used [38]. First, a random sequence was generated in Microsoft Excel using the formula =ROUNDUP (RAND ()*440,0). Participants were then split into 1 of 2 groups at random in a 1:1 ratio using their unique identification numbers. Opaque envelopes were used to disseminate information concerning participant allocation.

An independent biostatistician performed all the subsequent randomization steps. In addition, the staff involved in the randomization assessment and intervention delivery were separated, thus ensuring they did not know which patient belonged to which group.

Blinding

The research team, which consisted of the research supervisor and research assistants responsible for data collection, was unaware of the intervention and control groups [39]. Due to the subjective nature of the intervention, participants were aware of their allocation to either the intervention or the control group.

Outcome Measures

The primary outcome was the change in antihypertensive medication adherence at 6 months. This change was measured using the SEAMS questionnaire and self-reported pill counting (number of pills consumed over a certain period divided by pills prescribed for that specific period) [35]. SEAMS is a validated and reliable questionnaire, a 13-item assessment of medication self-efficacy in managing chronic conditions, found appropriate for individuals with limited literacy [34]. SEAMS uses a 3-point answer scale, where 1 denotes a lack of confidence, 2 denotes a moderate level of confidence, and 3 denotes a high level of confidence. A conceivable score is 13-39 points. Greater medication adherence is associated with higher scores, and vice versa. Based on prior studies, a cutoff value of 80% was used to distinguish between adherence status and nonadherence status. Participants were questioned regarding the number of pills they had been prescribed for a certain period, the number of pills they had consumed, and the number of pills they had forgotten to take during that certain period. Adherence rates were then calculated [35]. Patients who scored <80% were categorized as nonadherents, while those who scored ≥80% were classified as adherents [40]. At baseline, all participants were nonadherent; therefore, no further analysis could be conducted.

The secondary outcome was the SBP change at 6 months. This outcome was assessed in the hospital by a nurse who was not aware of the allocation of the study participants. The blood pressure was measured using a calibrated upper-arm mercury sphygmomanometer (MODEL-605P YAMASU). Standard principles were used to measure each participant's blood pressure [41].

Interim Analysis

At baseline, 3 months, and 6 months, both primary and secondary outcomes were evaluated. The 3-month analysis was used as a bridge to evaluate attrition rate patterns and the trends of change in outcomes. It was performed as an interim analysis; therefore, its results were not reported. The 6-month analysis was regarded as final.

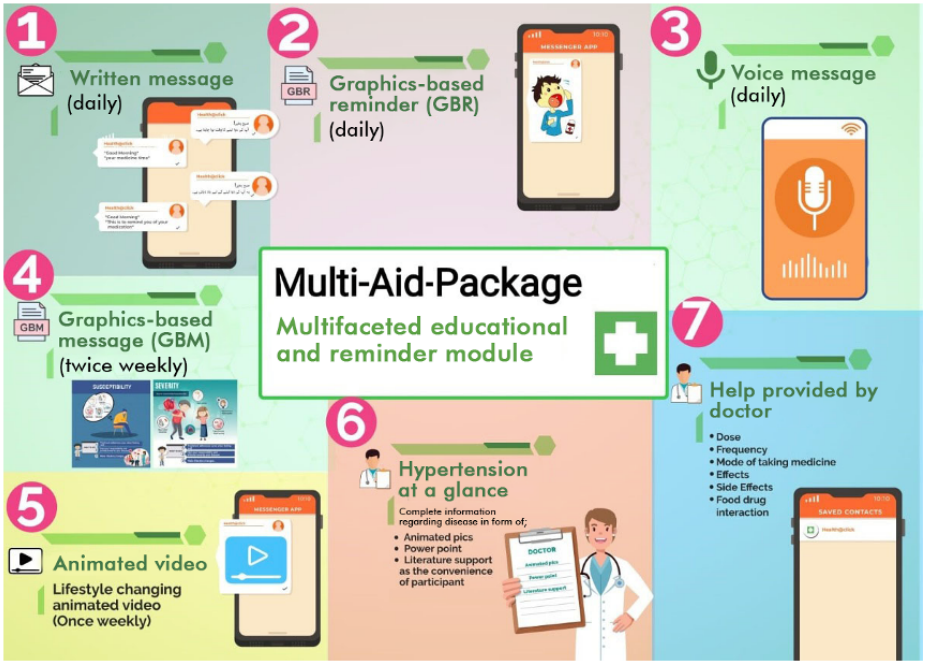
Intervention

The intervention's main objective was to enhance adherence to antihypertensive therapy in the intervention group using Multi-Aid-Package, a novel mHealth module. Multiple procedures were used in the intervention development. The first step was a thorough literature search for the theories and determinants of medication nonadherence. Additionally, hypotheses on patient acceptance of electronic/mobile devices were looked up in the accessible literature. The procedure for consulting with a group of specialists came next. Experts in epidemiology, behavioral intervention, health education, IT, and cardiology specializing in hypertension management used the Health Belief Model, self-determination theory, and relevant clinical standards and recommendations in the development of this module.

The content of this module included 7 items, a multifaceted approach with educational instructions, and reminders. Multi-Aid-Package comprised written and voice reminders, as well as graphics-based reminders (GBRs) and graphics-based messages (GBMs), which were all disseminated daily and weekly to participants in the intervention group via WhatsApp.

Reminder text and voice messages for medication intake were in Urdu, as it was the most commonly used language among the study participants. Examples of such messages are "Good morning, it's time for your medication" and "Good morning, this is a reminder for you to take your pills" (see items 1 and 3 in Figure 1).

Figure 1. Contents of Multi-Aid-Package. AI: artificial intelligence.



GBMs were an animated series of messages developed with the help of a professional team of software developers (see item 4 in Figure 1). Furthermore, an animated video was also created by IT professionals with the help of clinical experts. The resulting animated video was divided into 3 sections: (1) awareness of hypertension, (2) the negative consequences of uncontrolled hypertension, and (3) medical and lifestyle changes for better health (see item 5 in Figure 1). In this module, participants were also provided with the "Hypertension at a Glance" component, a portfolio of instructional and educational tools that shows details of the condition's causes, diagnosis, treatment, complications, and prognosis (see item 6 in Figure 1).

In addition to the components mentioned earlier, the module also contained live support provided by a certified doctor 24 hours a day. The support included information on the medicine's dose, the dose frequency, the administration method, the effects of therapy on the present sickness, adverse effects, and interactions with particular meals (see item 7 in Figure 1). Live support and the "Hypertension at a Glance" portfolio were provided on a demand or need basis to only those participants who encountered problems from day 1 of the intervention, while the remaining components of Multi-Aid-Package were disseminated following the timetable provided. The contents of Multi-Aid-Package are summarized in Figure 1.

Pilot Testing

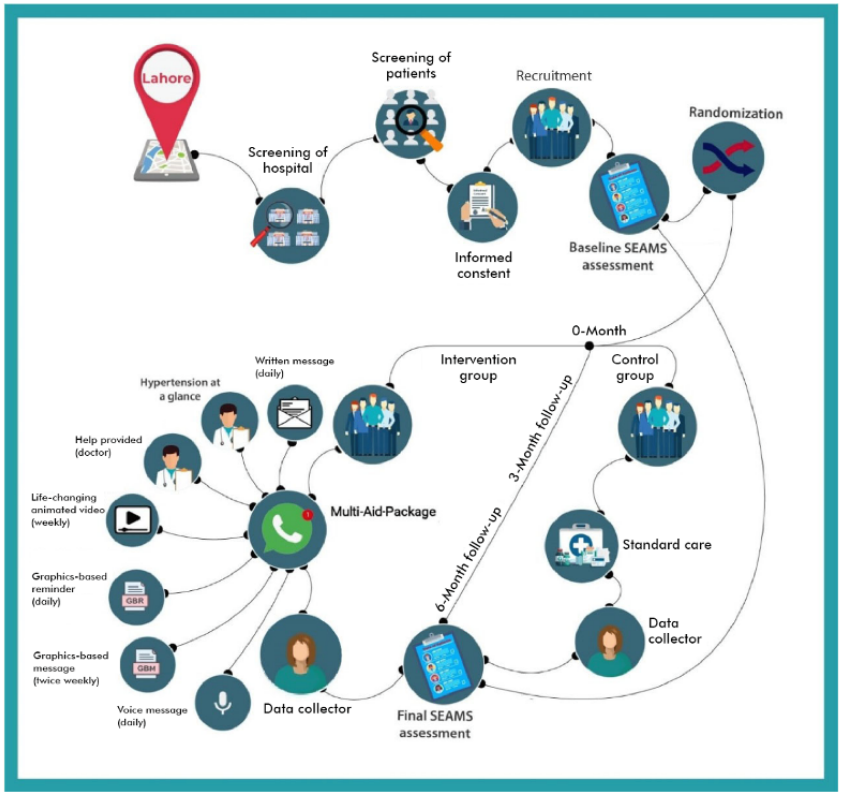
In addition, the Multi-Aid-Package was subject to a pilot test among 44 patients with hypertension to determine if they could understand the module's contents. As per usual hospital practice, only standard care was given to those in the control group. After being pilot tested, the intervention didn't change significantly. Only a few minor issues were seen, such as delivery issues, network issues, being outside of the coverage region, and unsuccessful file and video downloads. During the trial, this application's final version was made available.

Implementation of Multi-Aid-Package

The trial's design and execution adhered to the 2010 CONSORT criteria, which included a rigorous protocol for the provision of interventions to participants. This included the validation of the Multi-Aid-Package intervention in a pilot study, and an orientation and training session on the intervention was provided to the participants. Contact numbers were also provided to the participants in case they experienced any inconvenience. Moreover, a strict protocol was followed to disseminate the various contents of the multifaceted intervention. The implementation of Multi-Aid-Package was coordinated with the assistance of an IT specialist and 2 trained research assistants. In essence, these individuals were tasked to oversee the dissemination of the various contents of Multi-Aid-Package to the intervention group's participants via WhatsApp and to support data collection. WhatsApp was used in this study because it contains a feature on its interface that indicates to the sender whether the receiver has seen the message: a sign on the message sent through this app changes its color to blue when the recipient sees the message. This was the only way to check whether the participants had read the message. Second, the outcome results illustrated whether the participants took their medication, as prescribed. Research staff members were trained and the intervention pretested to ensure quality control. Using a pre-established curriculum for training by experts, 2 days of on-site instruction sessions on hypertension and questionnaire completion and how to respond to typical queries related to medication adherence were provided to all recruited research workers. Finally, the intervention module was rolled out to the selected participants in the trial. Figure 2 shows the overall flow of the study implementation.

No financial or other benefits were provided to the participants, except 6 months of free-of-cost WhatsApp use. There were no other direct benefits provided.

Figure 2. Trial flow. SEAMS: Self-Efficacy for Appropriate Medication Adherence Scale.



Participants' Timeline

The recruitment process was completed from January to May 2021, and a total of 439 participants were included. From June to December 2021, the intervention group received the Multi-Aid-Package intervention. The intervention duration was 6 months. The 6-month time frame of the intervention was chosen based on data from the previous literature on the topic, which included studies conducted over 2 and 3 months [30,42], and the fact that 6 months is a reasonable time limit to observe changes in behavior.

Data Collection

The data collection tool was a questionnaire in Urdu and English languages. The questionnaire was divided into 4 sections: A, B, C, and D.

Section A collected data on sociodemographic and health-related variables. Section B was the validated SEAMS questionnaire [34]. Section C collected self-reported pill-counting activity. Section D was a postintervention survey performed to assess the acceptability of the intervention. A 5-item Likert scale, with each item having 7 options, was used to evaluate participants' perceptions of their intervention experience with regard to usefulness, simplicity of use, and fulfillment of information.

Validity and Reliability of the Study Instrument

The internal consistency of SEAMS was good (Cronbach α=0.89). The test-retest reliability was moderate (Spearman coefficient=0.62, P<.001). The item total correlation coefficients ranged from 0.36 to 0.67, and the mean interitem correlation was 0.32 (range 0.08-0.71) [34].

The SEAMS questionnaire was translated from English to Urdu (SEAMS-U) in Pakistan using the standard "forward-backward" procedure. A convenient sample of 1011 patients with hypertension who were being treated at a tertiary care hospital in Lahore, Pakistan, was used to validate the translated version. The internal consistency of the translated questionnaire was good (Cronbach α=0.897). Cronbach α for part 1 was 0.838 and for part 2 was 0.789 using split-half reliability. The test-retest reliability was moderate (Spearman correlation=0.686, P<.001), and the intraclass correlation coefficient score was 0.814. The entire translation validity and reliability process was performed by our team and during publication.

Data Management and Statistical Analysis

Data management was the responsibility of a study supervisor, a biostatistician, and 2 research assistants. First, the research assistants ensured that no data collection form was incomplete or missing. Next, the research supervisor received all the data in sealed boxes. If there were any missing pieces of information in the data, the participants were contacted via a phone call to finish the form. Lastly, a biostatistician entered, cleaned, and analyzed the data. SPSS version 26.0 (IBM Corp) and RStudio (version 4.0.3; Posit PBC) were used to analyze the data.

The intention-to-treat analysis was used in this study [43]. The Shapiro-Wilk test was performed to determine whether the data were normally distributed. Categorical data were represented using frequencies and percentages, while continuous data were represented using medians (IQRs). The nonparametric Mann-Whitney U test was used on the data for the primary and secondary outcomes between groups, while the Wilcoxon signed rank test was used for within-group differences between baseline and 6 months. For categorical variables, the chi-square test was

used. The significance test was run with a *P* value of <.05. In addition, missing data were reported and treated using the single imputation approach. A generalized estimating equation (GEE) was used to control the covariates associated with the probability of affecting adherence to antihypertensive medication and the covariates that were significantly different between the intervention and control groups.

Adverse Events

No other adverse results were reported, except the primary and secondary outcomes that were reported in relation to the participants and our intervention strategy. All COVID-19 standard operating procedures were strictly followed during recruitment, randomization, and data collection. Furthermore, no issues were reported regarding COVID-19.

Ethical Considerations

The University Putra Malaysia (UPM) Ethical Committee on Human Research approved the research protocol (reference number: JKEUPM-2020-391) and the Institutional Review Board of the Sheikh Zayed Medical Complex Lahore (SZMC/IRB/163/2021). Participation in this trial was discretionary, and informed consent was obtained in writing from each participant before the start of the study. Strict confidentiality and privacy were ensured by assigning each participant an identification number to protect their identity [44]. The confidentiality of participant data was also ensured. This study was designed according to Good Clinical Practice (GCP) [45,46].

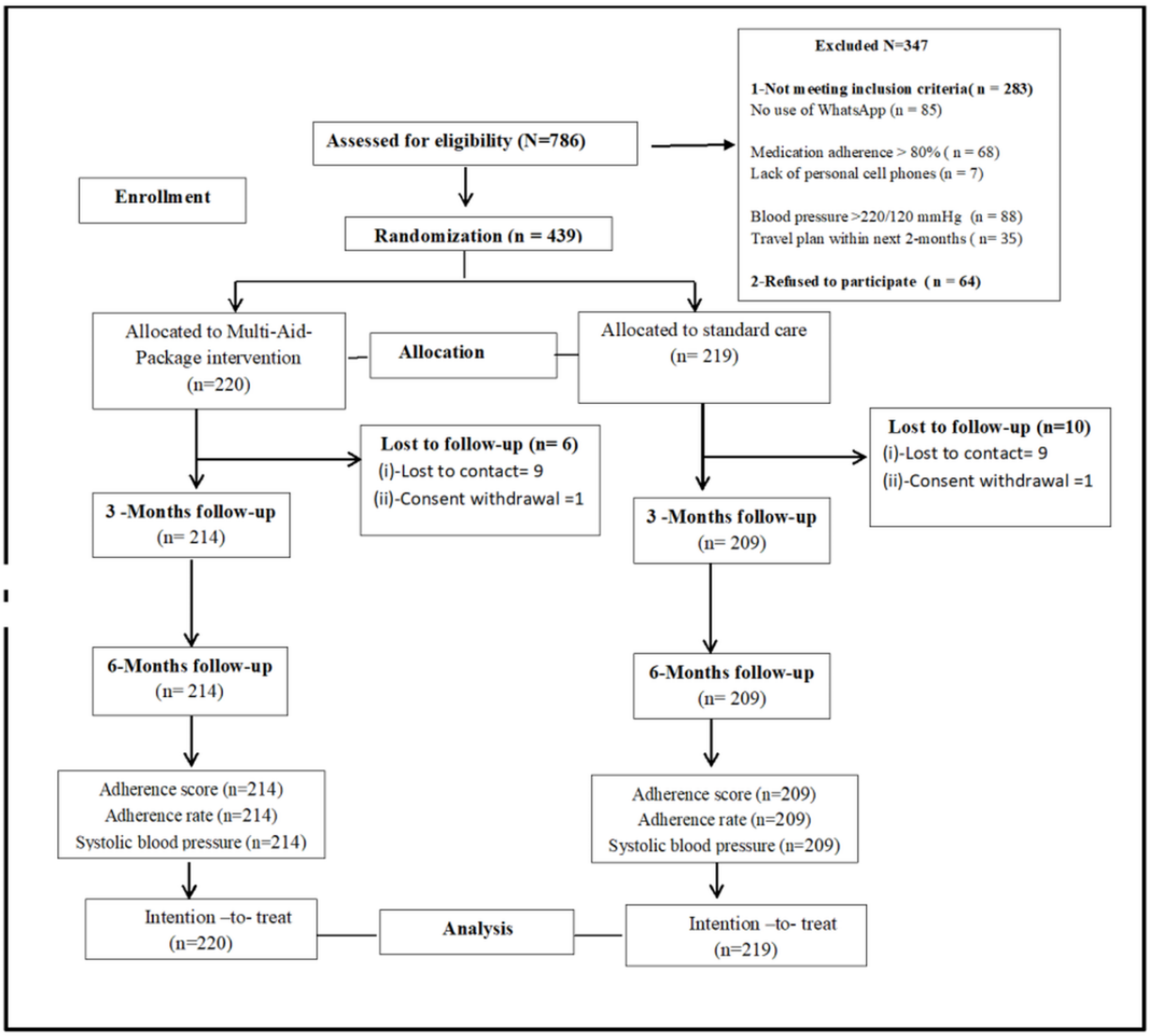
The study was designed to barely cause any risk to both patients and medical personnel. The participants were questioned in a private room, away from onlookers, in order to prevent minor psychological discomforts related to personal issues involving their income and the embarrassment that they might experience when answering questions about their subpar adherence status.

Results

Response Rate

From January to May 2021, a total of 786 participants were initially assessed based on the eligibility criteria. Of these, 347 (44.1%) participants were excluded based on inclusion criteria (n=283, 81.6%) and refusal to participate (n=64, 18.4%) in the trial. The details of the subcategories under “not meeting inclusion criteria” are elaborated in Figure 3. Of them, in June 2021, 439 participants fulfilled the criteria, consented to participate, and were randomly assigned to either the control group, which received standard care (n=219), or the intervention group, which received the Multi-Aid-Package (n=220). The randomization was performed according to the CONSORT flow diagram [31,47] (Figure 3). The total response rate at the end of the intervention was 423 (96.3%), with 209 (95.4%) participants in the control group and 214 (97.2%) participants in the intervention group until the completion of follow-up in December 2021. In the control group, 10 (4.8%) participants were lost to follow-up, while 1 (0.5%) participant withdrew consent. In the intervention group, 6 (2.8%) participants were lost to follow-up. The reason for the failure to follow up was loss of contact. No mortality was reported.

Figure 3. CONSORT flow diagram. CONSORT: Consolidated Standards of Reporting Trails.



Baseline Demographic Characteristics

At baseline, the 2 groups did not differ statistically significantly from each other across most of the variables, except gender, which had substantial differences among both groups (P=.02). In general, the intervention group's baseline characteristics were similar to those of the control group regarding age, ethnicity, marital status, education, family status, employment, and monthly income. Most of the participants were aged between 30 and 49 years old and were male, with a graduate level of education and a high-income status (Table 1).

Table 1. Baseline characteristics of study participants according to their group allocation (N=439).[a]

Characteristics	Intervention group (n=220), n (%)	Control group (n=219), n (%)	*P* value[a]
Age (years)			.30
≤50	91 (41.4)	102 (46.6)	
30-49	116 (52.7)	105 (47.9)	
18-29	13 (5.9)	12 (5.5)	
Gender			.02
Female	91 (41.4)	80 (36.5)	
Male	129 (58.6)	139 (63.5)	
Ethnicity			.54
Urdu	27 (12.3)	17 (7.8)	
Punjabi	143 (65.0)	163 (74.4)	
Suraiki	46 (20.9)	37 (16.9)	
Others	4 (1.8)	2 (0.9)	
Marital status			.76
Married	167 (75.9)	163 (74.4)	
Single	32 (14.5)	25 (11.4)	
Others	21 (9.5)	31 (14.2)	
Education			.59
Primary and secondary	72 (32.7)	60 (27.4)	
Graduate	80 (36.4)	71 (32.4)	
Postgraduate	68 (30.9)	88 (40.2)	
Do you smoke?			.29
No	171 (77.7)	151 (68.9)	
Yes	44 (20.0)	57 (26.0)	
Ex-smoker	5 (2.3)	11 (5.0)	
Family status			.31
Joint family	125 (56.8)	126 (57.5)	
Nuclear family	95 (43.2)	93 (42.5)	
Employment			.83
Yes	192 (87.3)	182 (83.1)	
No	28 (12.7)	37 (16.9)	
Monthly income (PKR[b])			.60
<10,000 (<US $35.96)[c]	1 (0.5)	1 (0.5)	
10,000-25,999 (US $35.96-$93.49)	44 (20.0)	35 (16.0)	
26,000-50,999 (US $93.50-$183.39)	31 (14.1)	30 (13.7)	
51,000-100,000 (US $183.40-$359.60)	66 (30.0)	66 (30.1)	
>100,000 (>US $359.60)	78 (35.5)	87 (39.7)	
Use of reminder alarm			.27
Yes	42 (19.1)	67 (30.6)	
No	178 (80.9)	152 (69.4)	

[a]*P*<.05 was considered statistically significant.

[b]PKR: Pakistani Rupee.

[c]An exchange rate of PKR 1=US $0.0036 was used.

Baseline Health-Related Characteristics

Regarding health-related characteristics, there was no significantly significant difference between the 2 groups. Additionally, the median SEAMS score of the control group was substantially higher than that of the intervention group (*P*=.03). Otherwise, the 2 groups had a balanced distribution of subjects regarding the duration of hypertension, comorbid conditions, the number of medications used daily, dose frequency, SBP, and controlled status of SBP<140 mmHg (Table 2).

Table 2. Comparison of health-related characteristics between intervention and control groups (N=439).[a]

Characteristics	Intervention group (n=220)	Control group (n=219)	*P* value
Duration of hypertension (years), n (%)			.40
<1	22 (10.0)	22 (10.0)	
1-5	79 (35.9)	82 (37.4)	
>5	119 (54.1)	115 (52.5)	
Concomitant disease, n (%)			.34
Yes	141 (64.1)	142 (64.8)	
No	79 (35.9)	77 (35.2)	
Comorbid conditions, n (%)			.57
1	87 (39.5)	88 (40.2)	
>1	133 (60.5)	131 (59.8)	
Daily medication number, n (%)			.47
<5	126 (57.3)	117 (53.4)	
5-9	74 (33.6)	83 (37.9)	
>10	20 (9.1)	19 (8.7)	
Daily dose frequency, n (%)			.91
Once daily	64 (29.1)	55 (25.1)	
Twice daily	104 (47.3)	111 (50.7)	
Thrice daily	52 (23.6)	53 (24.2)	
Controlled SBP[b] <140 mmHg, n (%)			.72
Uncontrolled	217 (98.6)	215 (98.2)	
Controlled	3 (1.36)	4 (1.82)	
Pill counting, n (%)			—[c]
Nonadherent	220 (100.0)	219 (100.0)	
Adherent	0	0	
SBP, median (IQR)	159 (23)	159 (27)	.77
SEAMS[d] adherence score, median (IQR)	19.5 (5)	21.0 (6)	.03

[a]*P*<.05 was considered statistically significant.

[b]SBP: systolic blood pressure.

[c]Not applicable.

[d]SEAMS: Self-Efficacy for Appropriate Medication Adherence Scale.

Effect of the Multi-Aid-Package Intervention on Medication Adherence

The effect of the Multi-Aid-Package intervention on medication adherence on the 2 groups at baseline and 6 months was measured using the median (IQR) SEAMS score. At baseline, the median SEAMS score was 19.5 (IQR 5) for the intervention group and 21 (IQR 6) for the control group. Compared to the intervention group, the control group's median SEAMS score was significantly higher (1.5 points; *P*=.011). At 6-month follow-up, the median SEAMS score was significantly different between the intervention and control groups (*P*<.001).

Regarding the medication adherence status, at baseline, all participants were nonadherent; therefore, no further analysis

could be performed. At 6 months, however, there was an increase in adherent patients between the intervention and control groups (difference n=81, 18.4% of patients; *P*<.001; Table 3).

The effect of the Multi-Aid-Package intervention on medication adherence within groups at baseline and 6 months was also measured using the median (IQR) SEAMS score. At baseline, the median SEAMS score within the intervention group was 19.5 (IQR 5), which increased to 32 (IQR 11) at the 6-month follow-up, with a median difference of 12.5 points. The median SEAMS score statistically significantly changed from baseline to 6 months (*P*<.001), while there was no statistically significant change in the median SEAMS score in the control group from baseline to 6 months (*P*=.29). A total of 83 (37.7%) participants achieved adherent status in the intervention group (*P*<.001), while 2 (0.9%) participants achieved adherent status in the control group (*P*=.78) from baseline to 6 months (refer to Multimedia Appendix 1).

Table 3. Primary outcomes for the intervention and control groups from baseline to 6 months (N=439).

Variable	Intervention group (n=220)	Control group (n=219)	Difference (intervention – control)	Test statistics	*P* value
SEAMS[a] score at baseline, median (IQR)	19.5 (5)	21.0 (6)	–1.5	*U*[b]=20,717.500	.01
SEAMS score at 6 months, median (IQR)	32.0 (11)	21.0 (6)	11	*U*=10,490.000	<.001
Adherence status at 6 months, n (%)	83 (37.72)	2 (.91)	81	95.266[c]	<.001

[a]SEAMS: Self-Efficacy for Appropriate Medication Adherence Scale.

[b]Mann-Whitney *U* test.

[c]Fisher exact test.

Effect of the Multi-Aid-Package Intervention on the SBP

At baseline, there was no difference in the SBP between the 2 groups. At 6 months, however, the median SBP was statistically different between the intervention and control groups (*P*<.001). A binary variable "controlled systolic blood pressure" was computed to evaluate the success of the treatment. The controlled SBP code was "Controlled <140 mmHg and uncontrolled >140 mmHg."

Overall, at 6 months, the number of patients with uncontrolled hypertension decreased by 46 in the intervention group (*P*<.001) but remained unchanged in the control group (*P*=.724; Table 4).

At baseline, the median SBP in the intervention group was 159 (IQR 23) mmHg, which decreased to 155 (IQR 29) mmHg, with a median difference of 4 mmHg. The median SBP significantly changed from baseline to 6 months (*P*<.001), while there was no statistically significant change in the median SBP in the control group from baseline to 6 months (*P*=.31). A total of 49 (22.3%) participants achieved controlled SBP status in the intervention group (*P*<.001), whereas there was no change in the control group's controlled SBP status (*P*=.78) from baseline to 6 months (refer to Multimedia Appendix 1).

Table 4. Secondary outcome change between intervention and control groups from baseline to 6 months (N=439).

Variable	Intervention group (n=220)	Control group (n=219)	Difference (intervention – control)	Test statistics	*P* value
SBP[a] (mmHg) at baseline, median (IQR)	159 (23)	159 (27)	0	*U*[b]=23,768.000	.81
SBP (mmHg) at 6 months, median (IQR)	155 (29)	162 (17)	–7	*U*=18,276.000	<.001[c]
Controlled SBP (mmHg) at baseline, n (%)	3 (1.36)	4 (1.83)	–1	N/A[d]	.72
Controlled SBP (mmHg) at 6 months, n (%)	49 (22.27)	4 (1.83)	45	43.221[e]	<.001[c]

[a]SBP: systolic blood pressure.

[b]Mann-Whitney *U* test.

[c]Significant *P* value.

[d]Not applicable.

[e]Fisher exact test.

Covariates Affecting Medication Adherence

A GEE was used to control the covariates associated with the probability of affecting adherence to antihypertensive medication using pill counting and the covariates significantly different between the intervention and control groups. We used the forward method. A working correlation matrix was gender. A total of 3 factors were found significant: group, time, and age. The group variable significantly contributed to medication adherence. The intervention group had a 1.714 times higher probability of being adherent to antihypertensive medication than the control group (adjusted odds ratio [AOR] 1.714, 95% CI 2.387-3.825; *P*<.001). Time points also contributed significantly to medication adherence. A 6-month postintervention time had a 1.837 times higher probability of showing adherence to antihypertensive medication than baseline

(AOR 1.837, 95% CI 1.625-2.754; *P*<.001). Age also contributed significantly. The 18-29 years of age group was found more likely to be adherent to antihypertensive treatment, with a 1.618 times higher probability than the other 2 age groups (AOR 1.618, 95% CI 0.225-1.699; *P*<.001). Income was a significant predictor for adherence to antihypertensive treatment (Table 5).

Table 5. Effect of the Multi-Aid-Package intervention on medication adherence, with adjusted covariates by the GEE[a] (N=439).

Variable	B[b] (SE)	Wald chi-square (*df*=1)	AOR[c], exp(B) (95% CI)	*P* value
Group				
Intervention	0.672 (0.863)	4.813	1.714 (2.387-3.825)	<.001[d]
Control	Reference	__[e]	—	—
Time point				
6 months	0.748 (0.216)	2.765	1.837 (1.625-2.754)	<.001[d]
Baseline	Reference	—	—	—
Age (years)				
≥50	–4.302 (0.574)	2.953	0.014 (0.002-0.074)	<.001[d]
>18	Reference	—	—	—
Gender				
Female	0.141 (0.287)	0.242	1.152 (0.655-2.025)	.62
Male	Reference	—	—	—
Education				
Primary and secondary	0.583 (0.365)	2.548	1.792 (0.876-3.667)	.11
Graduate	0.207 (0.358)	0.332	1.230 (0.609-2.485)	.56
Postgraduate	Reference	—	—	—
Monthly income (PKR[f])				
26,000-50,999 (US $93.50-$183.39)[g]	–0.441 (0.379)	1.346	0.644 (0.306-1.355)	.25
51,000-100,000 (US $183.40-$359.60)	–0.933 (0.672)	0.936	0.393 (0.157-0.983)	.17
>100,000 (>US $359.60)	Reference	—	—	—
Duration of hypertension (years)				
<1	–0.194 (0.447)	0.189	0.823 (0.343-1.978)	.66
1-5	–0.011 (0.300)	0.001	0.989 (0.549-1.780)	.97
>5	Reference	—	—	—
Concomitant disease				
Yes	–0.197 (0.286)	0.472	0.821 (0.468-1.441)	.49
No	Reference	—	—	—
Comorbid conditions				
1	0.043 (0.399)	0.012	1.044 (0.478-2.282)	.91
>1	Reference	—	—	—
Daily medication number				
<5	0.357 (0.829)	0.186	1.429 (0.281-7.259)	.67
5-9	–0.148 (0.686)	0.046	0.863 (0.225-3.313)	.83
>10	Reference	—	—	—
Daily dose frequency				
1	0.245 (0.680)	0.129	1.277 (0.337-4.848)	.72
2	0.041 (0.566)	0.005	1.042 (0.343-3.163)	.94
3	Reference	—	—	—

[a]GEE: generalized estimating equation.

[b]B: unstandardized β.

[c]AOR: adjusted odds ratio.

[d]$P<.05$ was considered statistically significant.

[e]Not applicable.

[f]PKR: Pakistani Rupee.

[g]An exchange rate of PKR 1=US $0.0036 was used.

Technology Acceptance Feedback

At the end of the study, an intervention acceptance survey for Multi-Aid-Package was performed. A total of 214 (97.3%) participants from the intervention group participated in the survey. The survey consisted of 5 questions on (1) the information provided by Multi-Aid-Package about the disease, disease management, and complications (1 question); (2) how easy the participants found the intervention to use (2 questions); and (3) utility (2 questions). Each question offered 7 possible answers. Ratings ranged from 7 to 35. The minimum score was 7, while the maximum score was 35. Next, the mean score was calculated for the 214 (97.3%) participants. The Multi-Aid-Package intervention received a mean score of 33.21 (SD 4.39) of 35 points (94.8%), with good feedback on its usefulness, simplicity of use, and fulfillment of information for treating hypertension.

Discussion

Principal Findings

The comprehensive and unique multifaceted Multi-Aid-Package comprised 7 potential components, including continuous reminders integrated with education and support components. Multi-Aid-Package was designed for the intervention group and revealed a significant increase in adherence to antihypertensive medication and a substantial reduction in SBP in patients with hypertension. This study showed that using Multi-Aid-Package led to a significant improvement in medication adherence among patients who were nonadherent to their antihypertensive medication at the beginning of the trial.

Similar results from an existing body of literature concur with this trial's findings, where SMS text message interventions revealed positive results in patients with hypertension compared to controls [17,29]. Some trials have also shown significant results in patients with hypertension using advanced cell phone apps [48-50]. Overall, mHealth technology interventions have revealed positive results in patients with CVD [51]. In another trial, an SMS text message intervention demonstrated substantial improvement in adherence to treatment, from 49% to 62.3%, in patients with hypertension [52]. In the previous literature, mHealth interventions have been reported to lower blood pressure and improve medication adherence, with adequate acceptance and feasibility [16,23,53-56]. Patients also significantly benefit over time with self-management and blood pressure control [57-60]. However, a few trials were unable to reveal any significant improvement in medication adherence post intervention: one used a mobile app, while the other studies used a web-based talking intervention to enhance medication adherence in CVD [16,30]. Similarly, another trial using mailing and automated calls [61] and one using video interventions in patients with stroke [62] were unable to reveal any substantial change. mHealth is also paramount in medication adherence in other chronic illnesses, such as tuberculosis, CVD, diabetes, and chronic liver diseases [63-66].

Research Innovation and Clinical Implications

Multi-Aid-Package is a modified version of the preexisting literature on this subject as it contains multiple facets (SMS text messages, apps, interactive messages, and calls) in 1 application compared to only 1 or 2 facets per app in other interventions. Previous trials have used different facets of mHealth, for instance, SMS text messaging interventions used to improve adherence to antihypertensive medication [29,67], interactive voice interventions [68], talking treatment interventions [30], advanced mobile apps [48,50], and mail-outs [61]. Some of these interventions have demonstrated positive results, while others have been unable to reveal any improvement or insignificant improvement in adherence to antihypertensive medication. The unique aspect of this study is that the Multi-Aid-Package intervention combines multiple facets in 1 intervention. Multi-Aid-Package is superior because it contains 7 different parts, such as SMS text messaging, videos, and graphics.

Multi-Aid-Package is a comprehensive and effective tool for enhancing antihypertensive medication adherence and subsequently managing the SBP. Much preexisting literature supports our findings. In most studies, intervention group participants have reported being adherent more likely compared to the controls, where the intervention eventually altered health beliefs concerning medication adherence but was unable to show any significant effect on the SBP. For example, a cell phone app for patients with hypertension improved medication adherence but failed to significantly control the SBP in an intervention arm compared to the control arm [48]. Similarly, another 12-month trial found a minor change in the SBP [29] compared to our study, which revealed better results, even with a shorter duration. mHealth interventions are also influential in changing lifestyles [67]. Although there can be various explanations for no significant or even a minute change in the SBP despite considerable improvement in medication adherence, factual evidence shows that a reasonably long time is needed to see a change in clinical outcomes. Nevertheless, becoming highly adherent to therapy is essential. The literature also emphasizes that to obtain more clinical benefits, patients must strictly adhere to their antihypertensive medications [69]. There is no evidence of the efficiency of interventions in lowering the DBP, although an increased SBP is the primary aim of antihypertensive medication. Furthermore, according to epidemiological research, untreated hypertension, especially the SBP, ought to be the main goal for hypertension treatment [70].

It is crucial to emphasize that using Multi-Aid-Package in the framework of clinical care in a resource-limited setting can alone help support patients in managing their hypertension. Several studies have found that improving medication adherence

has a higher impact on clinical outcomes linked to additional assistance, primarily through connections to health care professionals [71]. Some effective interventions has shown that support via a cell phone can be provided without further blood pressure monitoring [72] and with nonadherent patients not being contacted by any health care providers [73]. In our study, there was no permanent connection between the patient, the health care provider, and the targeted blood pressure monitor.

Conclusively, to the best of our knowledge, this multifaceted approach is the first technology-based intervention in Pakistan to be built comprehensively and uniquely using only WhatsApp, which was cost-, time-, and resource-efficient. In contrast to prior cell phone–based interventions, this multifaceted intervention has a more substantial impact on antihypertensive medication adherence and SBP outcomes.

Future Suggestions

Further exploration by covering multiple cities, a large sample size, and a longer duration of follow-up is required to validate the results of this study. Our findings cannot be generalized to populations with different sociodemographic and medical profiles, so more diverse studies are required to generalize the findings to a wider population. There are recommendations for more sophisticated designs and efficient interventions to enhance medication adherence in CVD [21]. To improve outcomes, we suggest implementing new educational interventions with more effective designs and sophisticated adherence measurement techniques (mobile apps or devices) at a relatively low cost and implementing successful treatments in clinical settings. A recent study compared an innovative mHealth strategy to peer counseling to improve adherence to medication in patients with hypertension [51]. Future trials should consider the type of antihypertensive medication being taken. Finally, high-quality research is needed to explore mixed qualitative evidence with quantitative studies.

Strengths and Limitations

Two methods were used to measure medication adherence to strengthen the method of measurement and to obtain robust findings, supported by a preexisting body of literature on the subject [74]. The SBP was also assessed as a secondary outcome to increase the credibility of adherence to medication in patients with hypertension. An interim analysis was performed to monitor the dropout status, increasing the trial's strength. All the steps in the trial were in line with recent SPIRIT guidelines [75]. The Multi-Aid-Package intervention could help minimize inequity and prevent discrimination among different sociodemographic groups. Technology acceptance feedback was also assessed at the study's end, and the intervention received excellent feedback.

This trial also has a few limitations. First, the trial was a single-center study conducted in only 1 city due to time constraints, limited funds, and response burden. Therefore, extension over the entire province or multiple towns might be impossible. The possible effect size at various sites may vary, which may potentially affect the findings and may constrain external validity. In addition, the study continued for 6 months; consequently, the researchers might not be able to ascertain the effect of adherence to medication on treatment outcomes for a longer duration. The study also did not consider the difference in gender between groups, and this might be a good area for future research. Finally, self-reporting was used to assess medication adherence. Social desirability bias could cause self-report questionnaires to overestimate genuine adherence [76]. However, several technology-based or mHealth methods exist to follow and measure adherence, which could not be used in the study due to the nonavailability of such devices, poor communication of patients, or fear of loss to follow-up. Please refer to the eHealth CONSORT checklist for more information about this Multi-Aid-Package trial [77].

Conclusions

In the context of this study, the multifaceted Multi-Aid-Package is an effective mHealth intervention that increased medication adherence among patients with hypertension and subsequently improves their SBP readings. The findings revealed a statistically significant change in the medication adherence score and pill-counting rates and a reduction in the SBP 6 months after the intervention began. The outcomes also demonstrated that users value Multi-Aid-Package for its applicability, ease of use, and informational content for the management of hypertension. Multi-Aid-Package should be considered as an approach for boosting the adherence of hypertension patients to their medication in Pakistan and other similar LMICs.

Acknowledgments

We acknowledge Professor Dr Sunil Kripalani, Professor of Medicine and Director, Center for Clinical Quality and Implementation Research, Emory University, who graciously allowed us to use the Self-Efficacy for Appropriate Medication Adherence Scale. We are also grateful to the patients, doctors, and nurses who participated in this study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Primary and secondary outcomes for the intervention and control groups.
[PDF File (Adobe PDF File), 119 KB - mhealth_v12i1e50248_app1.pdf]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 17470 KB - mhealth_v12i1e50248_app2.pdf]

References

1. Kearney PM, Whelton M, Reynolds K, Muntner P, Whelton PK, He J. Global burden of hypertension: analysis of worldwide data. Lancet 2005;365(9455):217-223. [doi: 10.1016/S0140-6736(05)17741-1] [Medline: 15652604]
2. Mills KT, Stefanescu A, He J. The global epidemiology of hypertension. Nat Rev Nephrol 2020 Apr;16(4):223-237 [FREE Full text] [doi: 10.1038/s41581-019-0244-2] [Medline: 32024986]
3. NCD Risk Factor Collaboration (NCD-RisC). Worldwide trends in blood pressure from 1975 to 2015: a pooled analysis of 1479 population-based measurement studies with 19·1 million participants. Lancet 2017 Jan 07;389(10064):37-55 [FREE Full text] [doi: 10.1016/S0140-6736(16)31919-5] [Medline: 27863813]
4. Kintscher U. The burden of hypertension. EuroIntervention 2013 May;9(Suppl R):R12-R15 [FREE Full text] [doi: 10.4244/EIJV9SRA3] [Medline: 23732143]
5. Bundy JD, Li C, Stuchlik P, Bu X, Kelly TN, Mills KT, et al. Systolic blood pressure reduction and risk of cardiovascular disease and mortality: a systematic review and network meta-analysis. JAMA Cardiol 2017 Jul 01;2(7):775-781 [FREE Full text] [doi: 10.1001/jamacardio.2017.1421] [Medline: 28564682]
6. Lewington S, Clarke R, Qizilbash N, Peto R, Collins R, Prospective Studies Collaboration. Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. Lancet 2002 Dec 14;360(9349):1903-1913. [doi: 10.1016/s0140-6736(02)11911-8] [Medline: 12493255]
7. Lindenfeld J, Jessup M. 'Drugs don't work in patients who don't take them' (C. Everett Koop, MD, US Surgeon General, 1985). Eur J Heart Fail 2017 Nov;19(11):1412-1413 [FREE Full text] [doi: 10.1002/ejhf.920] [Medline: 28891126]
8. Hill MN, Miller NH, Degeest S, American Society of Hypertension Writing Group, Materson BJ, Black HR, et al. Adherence and persistence with taking medication to control high blood pressure. J Am Soc Hypertens 2011;5(1):56-63. [doi: 10.1016/j.jash.2011.01.001] [Medline: 21320699]
9. Vrijens B, De Geest S, Hughes DA, Przemyslaw K, Demonceau J, Ruppar T, et al. A new taxonomy for describing and defining adherence to medications. Br J Clin Pharmacol 2012 May;73(5):691-705 [FREE Full text] [doi: 10.1111/j.1365-2125.2012.04167.x] [Medline: 22486599]
10. Burnier M, Egan BM. Adherence in hypertension. Circ Res 2019 Mar 29;124(7):1124-1140 [FREE Full text] [doi: 10.1161/CIRCRESAHA.118.313220] [Medline: 30920917]
11. National Institute of Population Studies. Pakistan Demographic and Health Survey 2017-18. The DHS Program. 2019 Jan 1. URL: https://dhsprogram.com/pubs/pdf/FR354/FR354.pdf [accessed 2024-05-14]
12. Mahmood S, Jalal Z, Hadi MA, Orooj H, Shah KU. Non-adherence to prescribed antihypertensives in primary, secondary and tertiary healthcare settings in Islamabad, Pakistan: a cross-sectional study. Patient Prefer Adherence 2020;14:73-85 [FREE Full text] [doi: 10.2147/PPA.S235517] [Medline: 32021119]
13. Modi D, Saha S, Vaghela P, Dave K, Anand A, Desai S, et al. Costing and cost-effectiveness of a mobile health intervention (ImTeCHO) in improving infant mortality in tribal areas of Gujarat, India: cluster randomized controlled trial. JMIR Mhealth Uhealth 2020 Oct 14;8(10):e17066 [FREE Full text] [doi: 10.2196/17066] [Medline: 33052122]
14. Law MR, Morris JK, Wald NJ. Use of blood pressure lowering drugs in the prevention of cardiovascular disease: meta-analysis of 147 randomised trials in the context of expectations from prospective epidemiological studies. BMJ 2009 May 19;338:b1665 [FREE Full text] [doi: 10.1136/bmj.b1665] [Medline: 19454737]
15. DiMatteo MR, Haskard KB, Williams SL. Health beliefs, disease severity, and patient adherence: a meta-analysis. Med Care 2007 Jun;45(6):521-528. [doi: 10.1097/MLR.0b013e318032937e] [Medline: 17515779]
16. Ni Z, Liu C, Wu B, Yang Q, Douglas C, Shaw RJ. An mHealth intervention to improve medication adherence among patients with coronary heart disease in China: development of an intervention. Int J Nurs Sci 2018 Oct 10;5(4):322-330 [FREE Full text] [doi: 10.1016/j.ijnss.2018.09.003] [Medline: 31406843]
17. Zhai P, Hayat K, Ji W, Li Q, Shi L, Atif N, et al. Efficacy of text messaging and personal consultation by pharmacy students among adults with hypertension: randomized controlled trial. J Med Internet Res 2020 May 20;22(5):e16019 [FREE Full text] [doi: 10.2196/16019] [Medline: 32432556]
18. WHO Global Observatory for eHealth. mHealth: New Horizons for Health through Mobile Technologies. Geneva: World Health Organization; 2011.
19. Type approval. Pakistan Telecommunication Authority. URL: https://www.pta.gov.pk/en/type-approval [accessed 2020-04-11]
20. Adler AJ, Martin N, Mariani J, Tajer CD, Owolabi OO, Free C, et al. Mobile phone text messaging to improve medication adherence in secondary prevention of cardiovascular disease. Cochrane Database Syst Rev 2017 Apr 29;4(4):CD011851 [FREE Full text] [doi: 10.1002/14651858.CD011851.pub2] [Medline: 28455948]
21. Arshed M, Mahmud AB, Minhat HS, Ying LP, Umer MF. Effectiveness of mHealth interventions in medication adherence among patients with cardiovascular diseases: a systematic review. Diseases 2023 Mar 01;11(1):41 [FREE Full text] [doi: 10.3390/diseases11010041] [Medline: 36975590]

22. M M, Kadir MM. Use of m-health technology for preventive medicine in Pakistan. Health Care: Curr Rev 2016 Jan;04(04):1000178. [doi: 10.4172/2375-4273.1000178]
23. McBride CM, Morrissey EC, Molloy GJ. Patients' experiences of using smartphone apps to support self-management and improve medication adherence in hypertension: qualitative study. JMIR Mhealth Uhealth 2020 Oct 28;8(10):e17470 [FREE Full text] [doi: 10.2196/17470] [Medline: 33112251]
24. Lima DL, Cordeiro RN, Carvalho GL, Malcher F. The influence of social media in minimally invasive surgery education: how surgeons exchange experience and knowledge in these platforms. J Minim Access Surg 2019;15(3):275-276 [FREE Full text] [doi: 10.4103/jmas.JMAS_270_18] [Medline: 30618437]
25. Ittefaq M, Seo H, Abwao M, Baines A. Social media use for health, cultural characteristics, and demographics: a survey of Pakistani millennials. Digit Health 2022;8:20552076221089454 [FREE Full text] [doi: 10.1177/20552076221089454] [Medline: 35401998]
26. Champion VL. Revised susceptibility, benefits, and barriers scale for mammography screening. Res Nurs Health 1999 Aug;22(4):341-348. [doi: 10.1002/(sici)1098-240x(199908)22:4<341::aid-nur8>3.0.co;2-p] [Medline: 10435551]
27. Deci EL, Ryan RM. The general causality orientations scale: self-determination in personality. J Res Pers 1985 Jun;19(2):109-134. [doi: 10.1016/0092-6566(85)90023-6]
28. Johnston N, Bodegard J, Jerström S, Åkesson J, Brorsson H, Alfredsson J, et al. Effects of interactive patient smartphone support app on drug adherence and lifestyle changes in myocardial infarction patients: a randomized study. Am Heart J 2016 Aug;178:85-94 [FREE Full text] [doi: 10.1016/j.ahj.2016.05.005] [Medline: 27502855]
29. Bobrow K, Farmer AJ, Springer D, Shanyinde M, Yu L, Brennan T, et al. Mobile phone text messages to support treatment adherence in adults with high blood pressure (SMS-Text Adherence Support [StAR]): a single-blind, randomized trial. Circulation 2016 Feb 09;133(6):592-600 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.115.017530] [Medline: 26769742]
30. Kamal AK, Khalid W, Muqeet A, Jamil A, Farhat K, Gillani SRA, et al. Making prescriptions "talk" to stroke and heart attack survivors to improve adherence: results of a randomized clinical trial (The Talking Rx Study). PLoS One 2018;13(12):e0197671 [FREE Full text] [doi: 10.1371/journal.pone.0197671] [Medline: 30571697]
31. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMJ 2010 Mar 23;340:c332 [FREE Full text] [doi: 10.1136/bmj.c332] [Medline: 20332509]
32. Tetzlaff JM, Chan A, Kitchen J, Sampson M, Tricco AC, Moher D. Guidelines for randomized clinical trial protocol content: a systematic review. Syst Rev 2012 Sep 24;1:43 [FREE Full text] [doi: 10.1186/2046-4053-1-43] [Medline: 23006870]
33. Pakistan Bureau of Statistics. Pakistan Bureau of Statistics. URL: http://www.pbs.gov.pk/ [accessed 2020-04-11]
34. Risser J, Jacobson TA, Kripalani S. Development and psychometric evaluation of the Self-efficacy for Appropriate Medication Use Scale (SEAMS) in low-literacy patients with chronic disease. J Nurs Meas 2007 Dec 01;15(3):203-219. [doi: 10.1891/106137407783095757] [Medline: 18232619]
35. Osterberg L, Blaschke T. Adherence to medication. N Engl J Med 2005 Aug 04;353(5):487-497. [doi: 10.1056/nejmra050100]
36. Noordzij M, Tripepi G, Dekker FW, Zoccali C, Tanck MW, Jager KJ. Sample size calculations: basic principles and common pitfalls. Nephrol Dial Transplant 2010 May;25(5):1388-1393. [doi: 10.1093/ndt/gfp732] [Medline: 20067907]
37. Lemeshow S, Hosmer Jr DW, Klar J, Lwanga SK. Adequacy of Sample Size in Health Studies. Hoboken, NJ: John Wiley & Sons; 1990.
38. Lim C, In J. Randomization in clinical studies. Korean J Anesthesiol 2019 Jun;72(3):221-232 [FREE Full text] [doi: 10.4097/kja.19049] [Medline: 30929415]
39. Karanicolas PJ, Farrokhyar F, Bhandari M. Practical tips for surgical research: blinding: who, what, when, why, how? Can J Surg 2010 Oct;53(5):345-348 [FREE Full text] [Medline: 20858381]
40. Matthews JN, Altman DG. Interaction 3: how to examine heterogeneity. BMJ 1996 Oct 05;313(7061):862 [FREE Full text] [doi: 10.1136/bmj.313.7061.862] [Medline: 8870577]
41. Muntner P, Shimbo D, Carey RM, Charleston JB, Gaillard T, Misra S, et al. Measurement of blood pressure in humans: a scientific statement from the American Heart Association. Hypertension 2019 May;73(5):e35-e66 [FREE Full text] [doi: 10.1161/HYP.0000000000000087] [Medline: 30827125]
42. Kamal AK, Shaikh Q, Pasha O, Azam I, Islam M, Memon AA, et al. A randomized controlled behavioral intervention trial to improve medication adherence in adult stroke patients with prescription tailored Short Messaging Service (SMS)-SMS4Stroke study. BMC Neurol 2015 Oct 21;15:212 [FREE Full text] [doi: 10.1186/s12883-015-0471-5] [Medline: 26486857]
43. Del Re AC, Maisel NC, Blodgett JC, Finney JW. Intention-to-treat analyses and missing data approaches in pharmacotherapy trials for alcohol use disorders. BMJ Open 2013 Nov 12;3(11):e003464 [FREE Full text] [doi: 10.1136/bmjopen-2013-003464] [Medline: 24227870]
44. Strobl J, Cave E, Walley T. Data protection legislation: interpretation and barriers to research. BMJ 2000 Oct 07;321(7265):890-892 [FREE Full text] [doi: 10.1136/bmj.321.7265.890] [Medline: 11021874]
45. The Declaration of Helsinki and public health. World Health Organization. URL: https://www.who.int/bulletin/volumes/86/8/08-050955/en/ [accessed 2020-04-11]

46. Rand CS, Sevick MA. Ethics in adherence promotion and monitoring. Control Clin Trials 2000 Oct;21(5 Suppl):241S-247S [FREE Full text] [doi: 10.1016/s0197-2456(00)00085-4] [Medline: 11018582]
47. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. BMJ 2010 Mar 23;340:c869 [FREE Full text] [doi: 10.1136/bmj.c869] [Medline: 20332511]
48. Morawski K, Ghazinouri R, Krumme A, Lauffenburger JC, Lu Z, Durfee E, et al. Association of a smartphone application with medication adherence and blood pressure control: the MedISAFE-BP randomized clinical trial. JAMA Intern Med 2018 Jun 01;178(6):802-809 [FREE Full text] [doi: 10.1001/jamainternmed.2018.0447] [Medline: 29710289]
49. Al-Arkee S, Mason J, Lane DA, Fabritz L, Chua W, Haque MS, et al. Mobile apps to improve medication adherence in cardiovascular disease: systematic review and meta-analysis. J Med Internet Res 2021 May 25;23(5):e24190 [FREE Full text] [doi: 10.2196/24190] [Medline: 34032583]
50. Santo K, Richtering SS, Chalmers J, Thiagalingam A, Chow CK, Redfern J. Mobile phone apps to improve medication adherence: a systematic stepwise process to identify high-quality apps. JMIR Mhealth Uhealth 2016 Dec 02;4(4):e132 [FREE Full text] [doi: 10.2196/mhealth.6742] [Medline: 27913373]
51. Arshed M, Zakar R, Umer MF, Kiran M, Ullah N, Iftikhar G, et al. Efficacy of mHealth and education-led peer counseling for patients with hypertension and coronary artery disease in Pakistan: study protocol for a double-blinded pragmatic randomized-controlled trial with factorial design. Trials 2023 Jul 10;24(1):448 [FREE Full text] [doi: 10.1186/s13063-023-07472-0] [Medline: 37424031]
52. Varleta P, Acevedo M, Akel C, Salinas C, Navarrete C, García A, et al. Mobile phone text messaging improves antihypertensive drug adherence in the community. J Clin Hypertens (Greenwich) 2017 Dec;19(12):1276-1284 [FREE Full text] [doi: 10.1111/jch.13098] [Medline: 28941056]
53. Xu H, Long H. The effect of smartphone app–based interventions for patients with hypertension: systematic review and meta-analysis. JMIR Mhealth Uhealth 2020 Oct 19;8(10):e21759 [FREE Full text] [doi: 10.2196/21759] [Medline: 33074161]
54. Li R, Liang N, Bu F, Hesketh T. The effectiveness of self-management of hypertension in adults using mobile health: systematic review and meta-analysis. JMIR Mhealth Uhealth 2020 Mar 27;8(3):e17776 [FREE Full text] [doi: 10.2196/17776] [Medline: 32217503]
55. Schoenthaler A, Leon M, Butler M, Steinhaeuser K, Wardzinski W. Development and evaluation of a tailored mobile health intervention to improve medication adherence in black patients with uncontrolled hypertension and type 2 diabetes: pilot randomized feasibility trial. JMIR Mhealth Uhealth 2020 Sep 23;8(9):e17135 [FREE Full text] [doi: 10.2196/17135] [Medline: 32965230]
56. Chew S, Lai PSM, Ng CJ. Usability and utility of a mobile app to improve medication adherence among ambulatory care patients in Malaysia: qualitative study. JMIR Mhealth Uhealth 2020 Jan 31;8(1):e15146 [FREE Full text] [doi: 10.2196/15146] [Medline: 32003748]
57. Chen B, Dou Y, Yu X, Ma D. Influence of internet-based health management on control of clinical parameters in patients with hypertension: four-year longitudinal study. J Med Internet Res 2023 Mar 20;25:e42896 [FREE Full text] [doi: 10.2196/42896] [Medline: 36939826]
58. Cao W, Milks MW, Liu X, Gregory ME, Addison D, Zhang P, et al. mHealth interventions for self-management of hypertension: framework and systematic review on engagement, interactivity, and tailoring. JMIR Mhealth Uhealth 2022 Mar 02;10(3):e29415 [FREE Full text] [doi: 10.2196/29415] [Medline: 35234655]
59. Liu F, Song T, Yu P, Deng N, Guan Y, Yang Y, et al. Efficacy of an mHealth app to support patients' self-management of hypertension: randomized controlled trial. J Med Internet Res 2023 Dec 19;25:e43809 [FREE Full text] [doi: 10.2196/43809] [Medline: 38113071]
60. Aovare P, Abdulai K, Laar A, van der Linden EL, Moens N, Richard E, et al. Assessing the effectiveness of mHealth interventions for diabetes and hypertension management in Africa: systematic review and meta-analysis. JMIR Mhealth Uhealth 2023 Aug 29;11:e43742 [FREE Full text] [doi: 10.2196/43742] [Medline: 37646291]
61. Ivers NM, Schwalm J, Bouck Z, McCready T, Taljaard M, Grace SL, et al. Interventions supporting long term adherence and decreasing cardiovascular events after myocardial infarction (ISLAND): pragmatic randomised controlled trial. BMJ 2020 Jun 10;369:m1731 [FREE Full text] [doi: 10.1136/bmj.m1731] [Medline: 32522811]
62. Kamal A, Khoja A, Usmani B, Magsi S, Malani A, Peera Z, et al. Effect of 5-minute movies shown via a mobile phone app on risk factors and mortality after stroke in a low- to middle-income country: randomized controlled trial for the Stroke Caregiver Dyad Education Intervention (Movies4Stroke). JMIR Mhealth Uhealth 2020 Jan 28;8(1):e12113 [FREE Full text] [doi: 10.2196/12113] [Medline: 32012080]
63. Hamine S, Gerth-Guyette E, Faulx D, Green BB, Ginsburg AS. Impact of mHealth chronic disease management on treatment adherence and patient outcomes: a systematic review. J Med Internet Res 2015 Feb 24;17(2):e52 [FREE Full text] [doi: 10.2196/jmir.3951] [Medline: 25803266]
64. Lee S, Rajaguru V, Baek JS, Shin J, Park Y. Digital health interventions to enhance tuberculosis treatment adherence: scoping review. JMIR Mhealth Uhealth 2023 Dec 04;11:e49741 [FREE Full text] [doi: 10.2196/49741] [Medline: 38054471]

65. Zhou TT, Wang R, Gu SJ, Xie LL, Zhao QH, Xiao MZ, et al. Effectiveness of mobile medical apps in ensuring medication safety among patients with chronic diseases: systematic review and meta-analysis. JMIR Mhealth Uhealth 2022 Nov 22;10(11):e39819 [FREE Full text] [doi: 10.2196/39819] [Medline: 36413386]
66. Pouls BPH, Vriezekolk JE, Bekker CL, Linn AJ, van Onzenoort HAW, Vervloet M, et al. Effect of interactive eHealth interventions on improving medication adherence in adults with long-term medication: systematic review. J Med Internet Res 2021 Jan 08;23(1):e18901 [FREE Full text] [doi: 10.2196/18901] [Medline: 33416501]
67. Chow CK, Redfern J, Hillis GS, Thakkar J, Santo K, Hackett ML, et al. Effect of lifestyle-focused text messaging on risk factor modification in patients with coronary heart disease: a randomized clinical trial. JAMA 2015;314(12):1255-1263. [doi: 10.1001/jama.2015.10945] [Medline: 26393848]
68. Vollmer WM, Owen-Smith AA, Tom JO, Laws R, Ditmer DG, Smith DH, et al. Improving adherence to cardiovascular disease medications with information technology. Am J Manag Care 2014 Nov;20(11 Spec No. 17):SP502-SP510 [FREE Full text] [Medline: 25811824]
69. Choudhry NK, Glynn RJ, Avorn J, Lee JL, Brennan TA, Reisman L, et al. Untangling the relationship between medication adherence and post-myocardial infarction outcomes: medication adherence and clinical outcomes. Am Heart J 2014 Jan;167(1):51-58.e5. [doi: 10.1016/j.ahj.2013.09.014] [Medline: 24332142]
70. Strandberg TE, Pitkala K. What is the most important component of blood pressure: systolic, diastolic or pulse pressure? Curr Opin Nephrol Hypertens 2003 May;12(3):293-297. [doi: 10.1097/01.mnh.0000069868.94246.ef] [Medline: 12698068]
71. Margolis KL, Asche SE, Bergdall AR, Dehmer SP, Maciosek MV, Nyboer RA, et al. A successful multifaceted trial to improve hypertension control in primary care: why did it work? J Gen Intern Med 2015 Nov;30(11):1665-1672 [FREE Full text] [doi: 10.1007/s11606-015-3355-x] [Medline: 25952653]
72. Carrasco MP, Salvador CH, Sagredo PG, Márquez-Montes J, González de Mingo MA, Fragua JA, et al. Impact of patient-general practitioner short-messages-based interaction on the control of hypertension in a follow-up service for low-to-medium risk hypertensive patients: a randomized controlled trial. IEEE Trans Inf Technol Biomed 2008 Nov;12(6):780-791. [doi: 10.1109/TITB.2008.926429] [Medline: 19000959]
73. Wald DS, Bestwick JP, Raiman L, Brendell R, Wald NJ. Randomised trial of text messaging on adherence to cardiovascular preventive treatment (INTERACT trial). PLoS One 2014;9(12):e114268 [FREE Full text] [doi: 10.1371/journal.pone.0114268] [Medline: 25479285]
74. Pednekar PP, Ágh T, Malmenäs M, Raval AD, Bennett BM, Borah BJ, et al. Methods for measuring multiple medication adherence: a systematic review-report of the ISPOR Medication Adherence and Persistence Special Interest Group. Value Health 2019 Feb;22(2):139-156 [FREE Full text] [doi: 10.1016/j.jval.2018.08.006] [Medline: 30711058]
75. Chan A, Tetzlaff JM, Gøtzsche PC, Altman DG, Mann H, Berlin JA, et al. SPIRIT 2013 explanation and elaboration: guidance for protocols of clinical trials. BMJ 2013 Jan 08;346:e7586 [FREE Full text] [doi: 10.1136/bmj.e7586] [Medline: 23303884]
76. Shi L, Liu J, Koleva Y, Fonseca V, Kalsekar A, Pawaskar M. Concordance of adherence measurement using self-reported adherence questionnaires and medication monitoring devices. Pharmacoeconomics 2010;28(12):1097-1107. [doi: 10.2165/11537400-000000000-00000] [Medline: 21080735]
77. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]

Abbreviations

AOR: adjusted odds ratio
CONSORT: Consolidated Standards of Reporting Trails
CVD: cardiovascular disease
DBP: diastolic blood pressure
GBM: graphics-based message
GBR: graphics-based reminder
GEE: generalized estimating equation
LMIC: low- and middle-income country
mHealth: mobile health
PKR: Pakistani Rupee
SBP: systolic blood pressure
SEAMS: Self-Efficacy for Appropriate Medication Adherence Scale

Edited by L Buis; submitted 02.07.23; peer-reviewed by M Östbring, HF Hsieh, B Castonguay; comments to author 06.11.23; revised version received 31.01.24; accepted 30.04.24; published 19.06.24.

Please cite as:
Arshed M, Mahmud A, Minhat HS, Lim PY, Zakar R
Effectiveness of a Multifaceted Mobile Health Intervention (Multi-Aid-Package) in Medication Adherence and Treatment Outcomes Among Patients With Hypertension in a Low- to Middle-Income Country: Randomized Controlled Trial
JMIR Mhealth Uhealth 2024;12:e50248
URL: https://mhealth.jmir.org/2024/1/e50248
doi:10.2196/50248
PMID:

Original Paper

Assessment of Heat Exposure and Health Outcomes in Rural Populations of Western Kenya by Using Wearable Devices: Observational Case Study

Ina Matzke[1]; Sophie Huhn[1]; Mara Koch[1]; Martina Anna Maggioni[2,3], PD, PhD; Stephen Munga[4], PhD; Julius Okoth Muma[4]; Collins Ochieng Odhiambo[4]; Daniel Kwaro[4], MD; David Obor[4], MSc; Till Bärnighausen[1,5,6], MD, Prof Dr; Peter Dambach[1], PhD; Sandra Barteit[1], Dr sc hum

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

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

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

[4]Centre for Global Health Research KISUMU, Kenya Medical Research Institute, Kisumu, Kenya

[5]Department of Global Health and Population, Harvard TH Chan School of Public Health, Havard University, Boston, MA, United States

[6]Africa Health Research Institute, KwaZulu-Natal, Somkhele, South Africa

Corresponding Author:
Sandra Barteit, Dr sc hum
Heidelberg Institute of Global Health
Heidelberg University Hospital
Heidelberg University
Im Neuenheimer Feld 130
Heidelberg, 69120
Germany
Phone: 49 62215634030
Email: barteit@uni-heidelberg.de

Abstract

Background: Climate change increasingly impacts health, particularly of rural populations in sub-Saharan Africa due to their limited resources for adaptation. Understanding these impacts remains a challenge, as continuous monitoring of vital signs in such populations is limited. Wearable devices (wearables) present a viable approach to studying these impacts on human health in real time.

Objective: The aim of this study was to assess the feasibility and effectiveness of consumer-grade wearables in measuring the health impacts of weather exposure on physiological responses (including activity, heart rate, body shell temperature, and sleep) of rural populations in western Kenya and to identify the health impacts associated with the weather exposures.

Methods: We conducted an observational case study in western Kenya by utilizing wearables over a 3-week period to continuously monitor various health metrics such as step count, sleep patterns, heart rate, and body shell temperature. Additionally, a local weather station provided detailed data on environmental conditions such as rainfall and heat, with measurements taken every 15 minutes.

Results: Our cohort comprised 83 participants (42 women and 41 men), with an average age of 33 years. We observed a positive correlation between step count and maximum wet bulb globe temperature (estimate 0.06, SE 0.02; *P*=.008). Although there was a negative correlation between minimum nighttime temperatures and heat index with sleep duration, these were not statistically significant. No significant correlations were found in other applied models. A cautionary heat index level was recorded on 194 (95.1%) of 204 days. Heavy rainfall (>20 mm/day) occurred on 16 (7.8%) out of 204 days. Despite 10 (21%) out of 47 devices failing, data completeness was high for sleep and step count (mean 82.6%, SD 21.3% and mean 86.1%, SD 18.9%, respectively), but low for heart rate (mean 7%, SD 14%), with adult women showing significantly higher data completeness for heart rate than men (2-sided *t* test: *P*=.003; Mann-Whitney *U* test: *P*=.001). Body shell temperature data achieved 36.2% (SD 24.5%) completeness.

Conclusions: Our study provides a nuanced understanding of the health impacts of weather exposures in rural Kenya. Our study's application of wearables reveals a significant correlation between physical activity levels and high temperature stress, contrasting with other studies suggesting decreased activity in hotter conditions. This discrepancy invites further investigation

into the unique socioenvironmental dynamics at play, particularly in sub-Saharan African contexts. Moreover, the nonsignificant trends observed in sleep disruption due to heat expose the need for localized climate change mitigation strategies, considering the vital role of sleep in health. These findings emphasize the need for context-specific research to inform policy and practice in regions susceptible to the adverse health effects of climate change.

(JMIR Mhealth Uhealth 2024;12:e54669) doi:10.2196/54669

KEYWORDS

wearables; wearable; tracker; trackers; climate; Africa; environment; environmental; heat; weather; exposure; temperature; rural; fitness trackers; climate change; health; heat; sub-Saharan Africa; Kenya; outcome; outcomes

Introduction

Climate Change and Health

Anthropogenic climate change has led to a mean global temperature increase of approximately 1 °C from preindustrial levels, with projections indicating a continued rise if substantial reductions in greenhouse gas emissions are not achieved; this warming trend poses profound health risks in low-and middle-income countries (LMICs) due to limited resources for environmental adaptation [1,2]. An emerging body of research indicates that wearable devices (wearables)—compact, noninvasive electronic devices capable of continuously monitoring various health metrics—may offer valuable insights into assessing the health impacts of climate change, especially in LMICs, where data on climate change and health are limited [3-5]. Climate change disproportionately affects regions such as sub-Saharan Africa, where increased temperatures exacerbate vulnerabilities, adversely impacting human health and agricultural productivity [2,6]. Kenya, the focus of our study, is increasingly vulnerable to climate change, with forecasts anticipating higher temperatures and more frequent extreme weather events; yet, there remains a lack of preparedness for necessary adaptation measures [1,7].

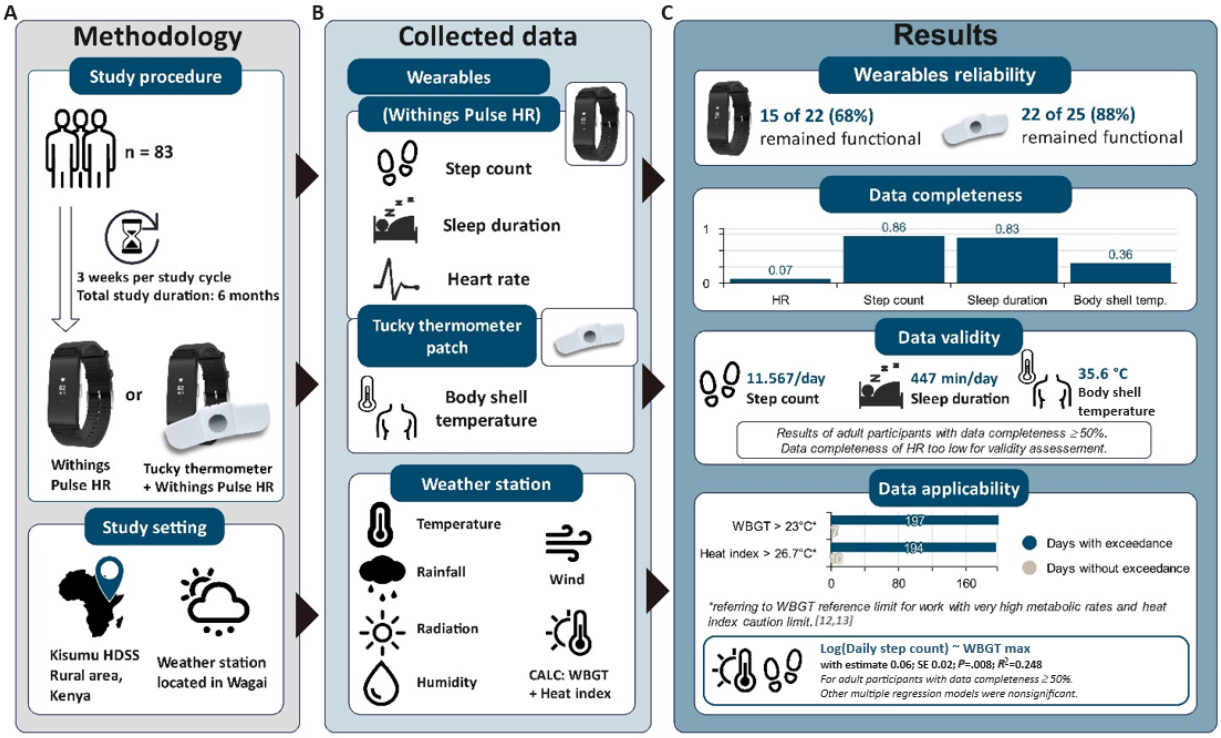
Need for Nuanced Understanding

In LMIC settings, a more nuanced understanding of individual exposure to extreme weather events and the resulting health outcomes is essential for creating tailored interventions and allocating resources efficiently [2]. Wearable devices, given their ability to monitor health metrics continuously and noninvasively, provide valuable insights into the health risks faced by vulnerable communities due to climate change [4,8]. Numerous large-scale studies in high-income settings have explored the use of wearables in health care [9,10], highlighting their potential as early warning systems for outbreaks of flu-like illnesses, among other applications. Although wearables have been utilized in studies within LMICs, notably in India [11], there is a lack in research concerning their use in other LMICs, particularly for assessing the impacts of climate change [3,4].

Objective of This Study

The primary aim of this study was to assess the feasibility and effectiveness of wearable devices in continuously and objectively monitoring the health impacts of weather exposures on individuals, particularly in a rural setting in Siaya, Kenya (Figure 1). We will integrate these technologies into routine data collection methods of the Health and Demographic Surveillance Systems (HDSS) in Siaya, Kenya, aiming to fill the data void in LMICs by providing measured health metrics that can approximate health impacts, thus offering individual-level, objectively measured health responses to weather exposures.

Figure 1. Schematic overview of the observational case study. Reference values are according to Parsons [12] and the National Weather Service [13]. Wearables depiction provided by Withings and e-TakesCare. (A) Methodology involves 83 participants from rural Siaya Health and Demographic Surveillance System, Kenya, equipped with wearables for a 3-week data collection period (total study duration: 9 weeks), coupled with local climate monitoring via a state-of-the-art weather station. (B) Collected data include metrics such as sleep duration, step count, pulse rate, and body temperature, as well as environmental data. (C) Results explore the reliability, completeness, and validity of data, with implications for climate change and health research. HDSS: Health and Demographic Surveillance System; HR: heart rate; WBGT: wet bulb globe temperature.



Methods

Ethics Approval

Ethics approval was granted by the Kenya Medical Research Institute (KEMRI/RES/7/3/1) and the ethics committee of the University Hospital Heidelberg, Germany (S-294/2019). This study is reported in accordance with the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) statement [14].

Study Design and Participants

This study employs an observational case study methodology in Siaya, a rural county situated in western Kenya, conducted from September 2021 to April 2022. For comprehensive details on the study protocol and a related case study from Burkina Faso, please refer to [15,16]. The Siaya county, about 40 km from Kisumu and 1000 meters above the sea level, hosts the Kenya Medical Research Institute–operated Siaya HDSS, covering an area of 700 km² and serving a population of around 260,000 people [16,17]. The HDSS has over 20 years of retrospective health and demographic data since 1990. Our study's sample population was stratified by gender, age, and wearable type, with age categories of 6-16 years, 17-45 years, and >45 years. Participants were assigned to either a group with the Withings Pulse Heart Rate (WPHR) wearable or a group using WPHR and the Tucky thermometer patch. Eligibility criteria included age >6 years, living within 5 km of Wagai health center, not planning to move during the study, and providing informed consent. Recruitment methods varied by age, with random selection for those older than 22 years and snowball sampling for those aged 6-22 years.

Procedures

The study protocol required participants to provide demographic and anthropometric data for wearable calibration, collected during each 3-week study cycle at the Wagai health center. Participants received a compensation of 200 Kenyan shillings (US $1.82) for travel, a smartphone with mobile data for data sync, and a battery pack for charging the wearables. Field personnel conducted weekly visits for data synchronization. Participants wore either the WPHR or both WPHR and a thermometer patch, with WPHR monitoring activity, sleep, and pulse rate, and the patch measuring body shell temperature at night (for detailed information, see Multimedia Appendix 1 [18,19]). The initial inventory consisted of 22 WPHR and 25 thermometer patches. Before reuse, both devices were cleaned and sanitized. This study also used a weather station in Wagai to record various weather parameters. The wet bulb globe temperature (WBGT), indicating heat strain, was calculated using a specific formula incorporating wet bulb temperature, global radiation, relative humidity, and air temperature [20]:

$$WBGT = (0.7*w) + (0.2*[0.009624*y - 0.00404*z + 1.102*x - 2.2776]) + (0.1*x)$$

where w represents wet bulb temperature, y represents global radiation, x represents relative humidity, and z represents air temperature.

Statistical Analysis

Participants were categorized into 4 age groups: school children (6-11 years), adolescents (12-18 years), young adults (19-45 years), and older adults (>45 years). BMI was classified as underweight (<18.5 kg/m^2), normal weight (18.5-24.9 kg/m^2), and overweight (>25 kg/m^2), following World Health Organization guidelines [21]. A descriptive analytical approach was used for demographic details and participant dropouts. BMI for adults was measured at recruitment. Wearable condition and wear were tracked for reliability assessment, and community interviewers' implementation challenges and infrastructure needs were thematically analyzed.

We analyzed 4 variables to ensure data quality: sleep duration, total step count, heart rate, and body shell temperature. The measured pulse rate was assumed to be equivalent to the participant's heart rate. We systematically excluded data that showed significant bias or anomalies such as unusually high or low heart rate readings (>expected maximum heart rate or <30 bpm) [22,23] and body shell temperature measurements indicative of protein denaturation or hypothermia-triggered loss of consciousness [24] (detailed criteria for data analysis can be found in Multimedia Appendix 2). Expected maximum heart rate was derived using the equation 208 – 0.7 × age [23]. Based on literature and the WPHR user guide, sleep measurements less than 3 hours (including naps) and those exceeding 13 hours were excluded as nonvalid. Wake times recorded by the wearable were not considered part of sleep duration [25,26]. Sleep episodes exceeding 3 hours were combined if the last episode began before noon on the following day. Differences across gender, study arm, and BMI for adult participants as well as across age groups were analyzed using Welch 2-sided *t* test and Mann-Whitney *U* test, with a 95% CI for error calculation.

To evaluate data completeness, we assessed the proportion of study duration covered by the measurements of the 4 key variables from wearables: sleep duration, step count, heart rate, and body shell temperature. The criteria for data completeness were based on existing literature (detailed in Multimedia Appendix 2 [15,22-28]). For external data validity, we combined individual measurements into a descriptive summary, including data from participants with at least 50% completeness per variable.

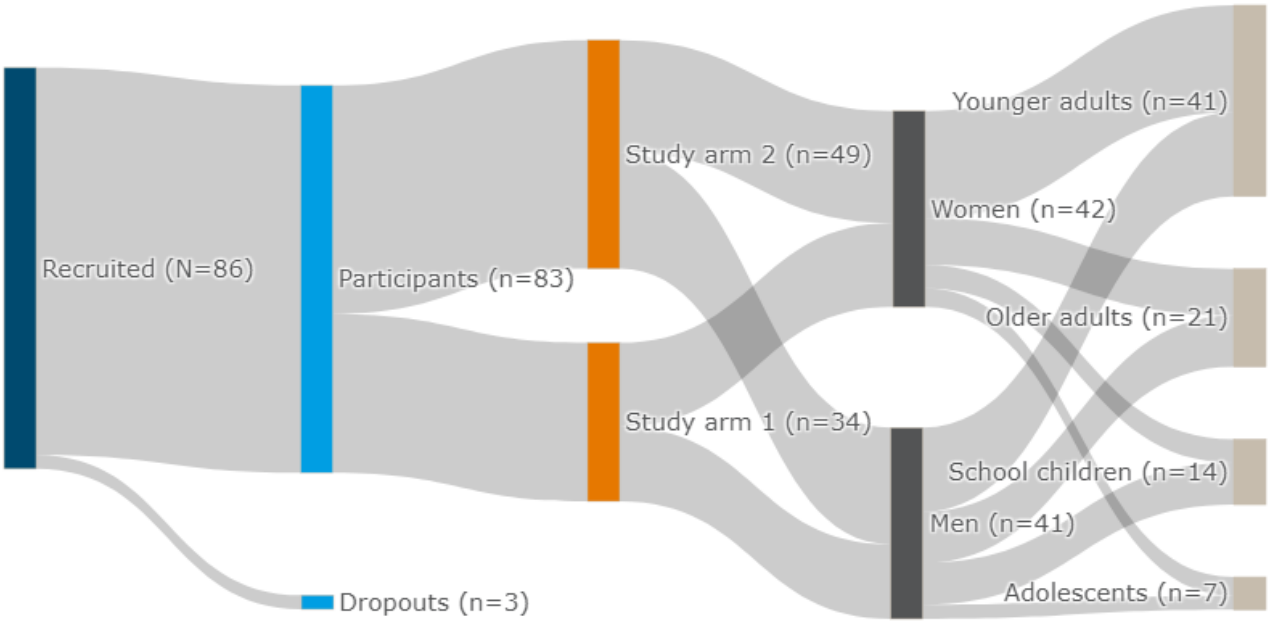
We descriptively analyzed the weather profile during the study period and correlations between weather events and adult study participant's health metrics that had at least 50% data completeness (sleep duration, step count, and body shell temperature). The measurements of WBGT and heat index were categorized using the reference values according to Parsons [12] and the National Weather Service [13]. We employed unadjusted linear regression and multilinear regression analyses [4,8,29-32] considering factors such as maximum daily and minimal nighttime heat and heavy rainfall. The multilinear models further considered gender, age, and BMI as confounders. This study did not account for within-subject trends or the effects of prior day heat stress due to the limitations of the 21-day duration of wearable usage by each participant. Analyses were conducted on R4.1.2 (RStudio version 4.1.2; PBC) with stats3.6.2 for statistics and ggplot2 and car packages for visualization, providing standard error and adjusted R^2 for error quantification and model assessment (see Multimedia Appendix 3 for variable plots and residual plots of the regression models). Numerical values with final digits <5 were rounded down, while numerical values with final digits >4 were rounded up.

Results

Demographics of the Participants

We initially enrolled 86 participants in our study; 3 withdrew their consent, resulting in 83 participants for analysis (Figure 2). In the 2 study arms, 34 (41%) participants wore solely the WPHR wearable, while the remaining 49 (59%) wore both the WPHR and the thermometer patch. The mean age of the participants was 33.3 (SD 19) years (range 6-83 years). A further breakdown of this age distribution showed that our study involved 14 school children (17%, age range 6-11 years), 7 adolescents (8%, age range 12-18 years), 41 young adults (49%, age range 19-45 years), and 21 older adults (25%, age >45 years). Women comprised 51% (42/83) of the all the participants. In the adult demographics, which accounted for 75% (62/83) of the participants, the average BMI calculated was 23.8 (SD 4.7) kg/m^2 (range 16.0-37.4 kg/m^2).

Figure 2. Participant stratification in this study.



Wearables' Reliability

Of the 22 WPHRs and 25 thermometer patches initially deployed, 7 (32%) WPHRs and 3 (12%) thermometer patches malfunctioned, primarily from physical damage. Technical issues, including data synchronization, were frequent initially but reduced over the study's duration. Physical damage was the predominant reason for WPHR failures, causing 5 (23%) devices to malfunction due to broken components or overall failure. One WPHR and its charger (5% of the total) were lost. As for thermometer patches, 2 (8%) had damaged charging ports, 1 (4%) was lost, and 2 (8%) had nonretrievable data, despite being intact.

Data Completeness

Data quality varied across health metrics and participant demographics. Accelerometer metrics such as step count and sleep duration exhibited high completeness, registering 86.1% (SD 18.9%) and 82.6% (SD 21.3%), respectively. Photoplethysmography-based heart rate measurements lagged behind at 7% (SD 14%), while body shell temperature recorded 36.2% (SD 24.5%) completeness. Data completeness was calculated as percentage of study duration covered with measurements in distinct intervals for all participants. Sleep data completeness varied by age: younger adults (19-45 years; n=41) recorded lower data completeness (79%) than older adults (>45 years; n=21; data completeness 90%) (*t* test: *P*=.02; 95% CI –0.20 to –0.02; Mann-Whitney *U* test [MWU] test: *P*=.03; 95% CI –0.14 to 0.00). Further, adult women (n=33) showed less data missingness (87%) than adult men (n=29, 78%) (MWU test: *P*=.03; 95% CI 0.00-0.14). Heart rate data completeness in adult women (12%) was significantly higher than that in men (2%) (*t* test: *P*=.003; 95% CI 0.04-0.17; MWU test: *P*=.001; 95% CI 0.01-0.04). Body shell temperature data completeness for school children (6-11 years; data completeness 47%) was higher than that for adolescents (12-18 years; data completeness 23%) (*t* test: *P*=.03; 95% CI 0.03-0.45; MWU test: *P*=.03; 95% CI 0.05-0.48). After correcting for multiple tests using Holm sequential Bonferroni method [33], only the difference in the completeness of heart rate data between men and women remained statistically significant (*t* test: *P*=.02; MWU: *P*=.004). For detailed data measurements stratified by age, gender, study arm, and BMI, as well as data completeness results corrected for multiple testing, refer to Multimedia Appendix 4 [33].

Environmental Exposure

During the 204 days study, rainfall was recorded on 97 (47.6%) days. Out of these, heavy rain (≥20 mm per day) was recorded on 16 (7.8%) days. The average daily heat index and WBGT were 22.1 (SD 1.1) °C and 20 (SD 0.7) °C, respectively. The heat index caution limit (26.67 °C or 80 °F) was reached on 194 (95.1%) of 204 days, and the extreme caution limit (32.22 °C or 90 °F) exceeded on only 2 (<1%) days. For the WBGT, reference values for outdoor work with very high metabolic rates were met on 197 (96.6%) days and for work with high metabolic rates on 96 (47.1%) days. Figure 3 illustrates the instances where individual reference values for WBGT, as proposed by Parsons [12], and the heat index, as per the standards of the National Weather Service [13], were exceeded (Table 1).

Figure 3. Average daily wet bulb globe temperature in °C (dark blue line) and average daily heat index in °C (orange line) with daily ranges (daily wet bulb globe temperature range in blue ribbon; daily heat index range in orange ribbon). WBGT: wet bulb globe temperature.

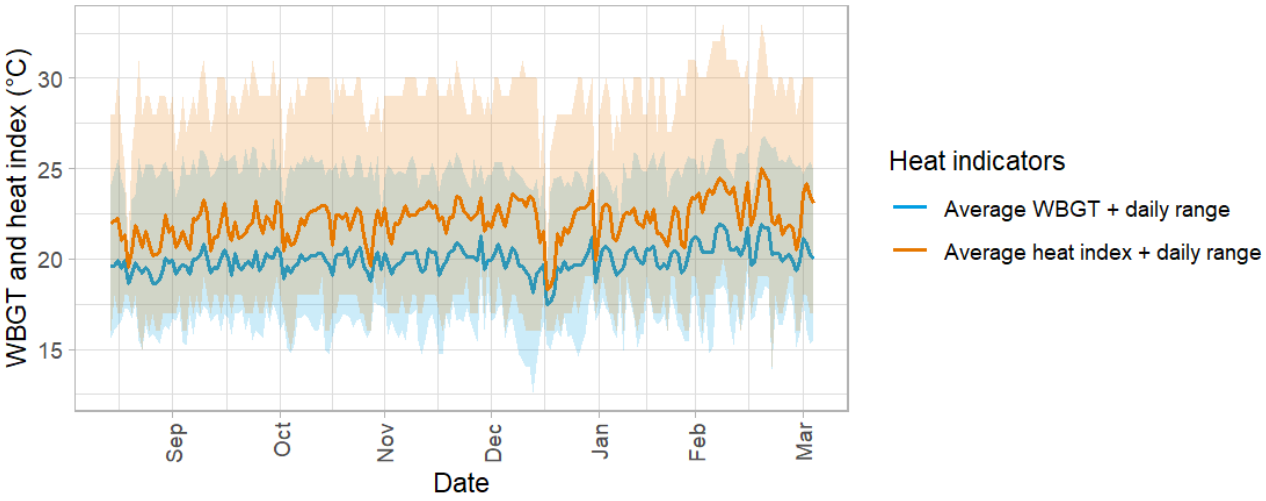


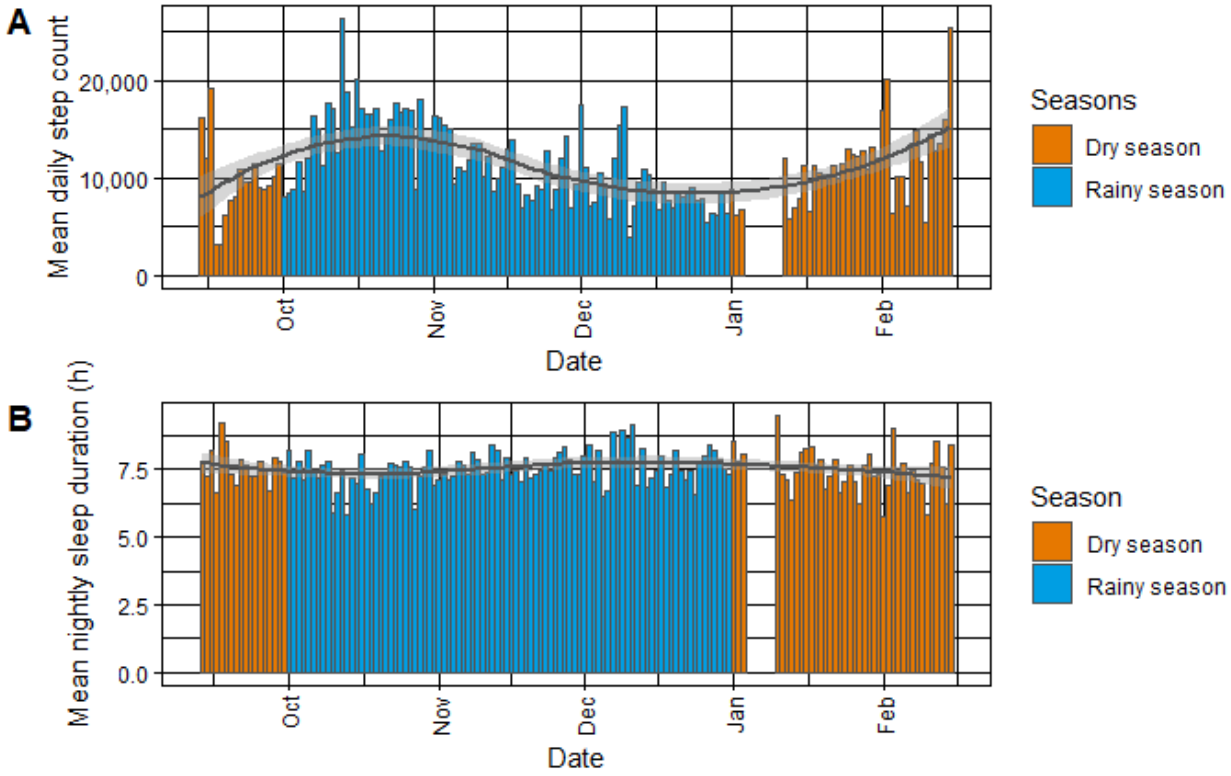
Table 1. Allocation of the respective reference values for wet bulb globe temperature and heat index to the individual risk levels for heat-related diseases depending on the type of work performed and listing of the proportion of days in the observed study period in which these reference values were exceeded. Reference values are according to Parsons [12] and the National Weather Service [13].

Reference	Days with exceedance (n=204), n (%)
Wet bulb globe temperature reference [12]	
Limit for work with a very high metabolic rate (metabolic rate>260 W/m^2)	197 (96.6)
Limit for work with a high metabolic rate (200 W/m^2<metabolic rate<260 W/m^2)	96 (47.1)
Limit for work with a medium metabolic rate (130 W/m^2<metabolic rate<200 W/m^2)	0 (0)
Heat index reference [13]	
Caution limit (26.67 °C/80 °F)	194 (95.1)
Extreme caution limit (32.22 °C/90 °F)	2 (0.1)
Danger limit (39.44 °C/103 °F)	0 (0)

Data Applicability: Heat Exposure and Health Outcomes

The unadjusted regression analysis for daily step count indicated a significant positive correlation with maximal daily WBGT and heat index (WBGT: estimate 974.4, SE 242.3; P<.001; heat index: estimate 317.6, SE 152.4; P=.04) with a low R^2 value (WBGT: R^2=0.014; heat index: R^2=0.003). On days without heavy rainfall, a similar positive yet nonsignificant correlation was observed for daily step count (estimate 1466.2, SE 852.9; P=.09; R^2=0.002). The analysis for daily maximal temperature did not demonstrate any predictive power for step count (R^2=–0.001; estimate 87.54, SE 145.08; P=.55). After a logarithmic transformation of the dependent variable (step count) due to a funnel shape observed in residuals, the confounder-adjusted models revealed a significant positive relationship with the maximum WBGT (estimate 0.06, SE 0.02; P=.008; R^2=0.248) (see Figure 4 for details). Other heat indicators showed positive yet nonsignificant associations with daily step count. Age and BMI emerged as significant predictors of daily step count (P<.001).

Figure 4. (A) Mean daily step count of adult participants per day and season, wherein dry season is highlighted in orange and rainy season is highlighted in blue. (B) Mean daily sleep duration per adult participant per day and season. Trend lines are added using locally weighted scatterplot smoothing; seasons are classified according to Odhiambo et al [17].



For sleep duration in minutes, the linear regression model revealed a negative, however nonsignificant, correlation with minimal nighttime temperature and heat index at small *P* values, but with low predictive power (temperature: estimate –5.61, SE 3.24; *P*=.08, R^2=0.002; heat index: estimate –4.77, SE 2.91; *P*=.10; R^2=0.002). For minimal nighttime WBGT and sleep duration, simple linear regression had no explanatory power (R^2=–0.001, estimate –0.67, SE 2.98; *P*=.82). Similarly, the multiple linear regression analysis indicated that an increase in minimum nighttime temperatures and nightly minimal heat index corresponded with nonsignificant reductions in sleep duration, that is, by 5.56 minutes per °C increase (*P*=.09; SE 3.234; R^2=0.006) and 4.50 minutes per °C increase (*P*=.12; SE 2.905; R^2=0.006), respectively. We did not detect a significant relationship between nightly minimum WBGT and sleep duration (estimate 0.07 minutes; *P*=.98; SE 2.985; R^2=0.003). Age proved to be a significant determinant of sleep duration across all models (*P*=.03).

Unadjusted linear regression analysis of body shell temperature and minimal nighttime heat had no explanatory power giving their negative adjusted R^2 of –0.003 across all heat indicators (temperature: estimate 0.05, SE 0.07; *P*=.51; WBGT: estimate 0.05, SE 0.07; *P*=.51; heat index: estimate 0.04, SE 0.07; *P*=.53). The multiple linear regression models of nightly body shell temperature did not show significant results either, however at relatively high explanatory power, as indicated by their R^2 values (0.168, 0.170, and 0.170, respectively; temperature: estimate 0.03, SE 0.07; *P*=.63; WBGT: estimate 0.05, SE 0.07; *P*=.42; heat index: estimate 0.05, SE 0.06; *P*=.41, respectively). Gender (*P*<.001) and BMI (*P*=.04) were significant predictors in all 3 models.

The additional analysis incorporating the temporal confounder of weekday or weekend, detailed in Multimedia Appendix 5, did not reveal any significant impact in the models. Regression models were not applied to heart rate measurements, as only the data sets of the 3 participants met the required 50% completeness. A summary of all the multiple regression models, including estimates, standard errors, *P* values, and other considered confounders for each health parameter and the associated extreme weather indicators is provided in Table 2.

Table 2. Regression analysis results for health parameters.[a]

Health parameter, variable	Temperature		WBGT[b]		Heat index		Precipitation <20 mm	
	Estimate (SE)	*P* value	Estimate (SE)	*P* value	Estimate (SE)	*P* value	Estimate (SE)	*P* value
Step count								
Intercept	10.47 (0.41)	<.001	9.26 (0.62)	<.001	10.12 (0.45)	<.001	10.79 (0.16)	<.001
Gender (men)	0.06 (0.05)	.25	0.05 (0.05)	.28	0.06 (0.05)	.27	0.06 (0.05)	.25
Age	–0.02 (0.00)	<.001	–0.02 (0.00)	<.001	–0.02 (0.00)	<.001	–0.02 (0.00)	<.001
Maximal daily weather measurement	0.01 (0.01)	.33	0.06 (0.02)	.008	0.03 (0.01)	.08	0.07 (0.08)	.36
BMI	–0.03 (0.01)	<.001	–0.03 (0.01)	<.001	–0.03 (0.01)	<.001	–0.03 (0.01)	<.001
Sleep duration								
Intercept	537.52 (58.98)	<.001	438.98 (52.84)	<.001	520.52 (54.42)	<.001	—[c]	—
Gender (men)	1.42 (6.01)	.81	0.93 (6.01)	.88	1.32 (6.01)	.83	—	—
Age	–0.39 (0.17)	.03	–0.38 (0.18)	.03	–0.38 (0.17)	.03	—	—
Minimal nightly weather measurement	–5.56 (3.23)	.09	0.07 (2.99)	.98	–4.50 (2.91)	.12	—	—
BMI	1.02 (0.61)	.09	1.03 (0.61)	.09	1.00 (0.61)	.10	—	—
Body shell temperature								
Intercept	36.31 (1.34)	<.001	35.98 (1.25)	<.001	36.02 (1.19)	<.001	—	—
Gender (men)	–0.72 (0.12)	<.001	–0.72 (0.12)	<.001	–0.72 (0.12)	<.001	—	—
Minimum nightly weather measurement	0.03 (0.07)	.63	0.05 (0.07)	.42	0.05 (0.06)	.41	—	—
Age	0.00 (0.01)	.41	0.00 (0.01)	.44	0.00 (0.01)	.40	—	—
BMI	–0.03 (0.02)	.04	–0.03 (0.02)	.04	–0.04 (0.02)	.04	—	—

[a]This is a logarithmically transformed model for step count and standard models for sleep duration and body temperature. The table includes estimates, standard errors, and *P* values for each health parameter—sleep duration, step count, and body temperature—along with their associated extreme weather indicators. Health parameters were included of adult participants having at least 50% data completeness for the respective health parameter. Depending on the model, the weather indicators considered were temperature, heat index, wet bulb globe temperature, and precipitation. Additionally, the multiple linear regression models incorporated gender, age, and BMI, as the data validity analysis demonstrated the significance of these confounders.

[b]WBGT: wet bulb globe temperature.

[c]Not available.

Discussion

Principal Results

The findings of our study highlight the advantages and challenges associated with the use of wearable devices for the continuous monitoring of vital signs in rural sub-Saharan populations. We found using wearables a pertinent approach for understanding individual impacts of weather exposures. Our research emphasizes the feasibility and effectiveness of integrating wearable technology into health research, in particular, to understand individual exposures and activity patterns such as daily steps and sleep duration. We identified a correlation between weather exposures and various health metrics. Notably, there is a positive relationship between daily step count and the maximum WBGT as well as a potential negative association between nighttime temperatures and sleep duration. These findings contribute to our understanding of the possible health impacts of climate change, with a particular focus on rural regions in western Kenya.

Comparison With Prior Work

Our study shows environmental exposures such as frequent heavy rainfall and extreme heat, with approximately 20.7% (4044/19,576) of our weather station readings surpassing the heat index caution threshold [13]—a pattern consistent with WBGT findings. We found a strong positive correlation between daily step count and maximum WBGT across both models. Additionally, a positive relationship with the heat index was observed in the unadjusted model. Although many studies indicate a negative link between heat and physical activity [30], the relationship between temperature and activity levels remains inconsistent. This discrepancy necessitates further investigation. For instance, a study [34] conducted in Qatar utilizing wearable pedometers discovered negative correlations between daily step count and both precipitation and temperature. However, that study reported varying associations with WBGT depending on the analytical model employed. Our findings emphasize the need for deeper investigation into this relationship [4]. External factors such as seasonal farming activities illustrated in Figure 4 or specific mitigation practices might significantly influence

this correlation and therefore merit additional investigation [17], especially given the scarcity of objectively generated data on activity levels and physical activity profiles in rural sub-Saharan African populations [35].

In our study, we observed a negative trend between minimum nighttime temperature and sleep duration as well as between the minimal nighttime heat index and sleep duration, but these relationships were not statistically significant in either model. However, this trend aligns with previous research that has linked warmer nights to shorter sleep durations, suggesting that climate change could significantly affect sleep health [4,8,32]. For example, a comprehensive study by Minor et al [8] utilizing sleep data of 47,628 participants across 68 countries found a correlation between shorter sleep duration and warmer nights. In line with these findings, all studies in the scoping review of Koch et al [4] reported a negative correlation between sleep duration and heat. These studies underscore the risk of insufficient sleep, in light of the expected impact of global warming due to climate change on local heat exposure [8]. The review of Caddick et al [36] suggests that optimal ambient temperatures for sleep lie between 17 °C and 28 °C, with 40%-60% humidity, although this may vary based on other factors. Insufficient sleep, whether due to short duration or disruptions, can negatively affect human health, potentially compromising the immune system [37] and increasing cardiovascular risk [38].

Regarding body shell temperature, our study did not find significant associations with average nightly body shell temperature and heat. However, previous studies [4,31] have suggested a possible connection. In our analysis, gender and age emerged as critical factors in all 3 models, emphasizing the importance of demographic and physiological factors in body temperature regulation. The inconclusive results in our study might be attributed to the limited sample size and issues with data integrity. Furthermore, natural thermoregulatory processes such as evaporation and the thermoregulatory behaviors of participants should be considered when evaluating body shell temperature [24].

Our study shows improved data completeness compared to previous research in Burkina Faso [15] and the United States [27], especially in thermometer patch data, likely due to enhanced adhesion using medical tape [15]. However, heart rate data showed lower completeness, which may have been affected by technical factors such as the proximity of the sensors to the skin, the impact of motion, and potential device errors. Similar studies using certain wearable devices reported data loss due to wearable malfunctions such as connection issues [39]. Other factors such as blood vessel thickness, skin thickness, obesity, and age might also affect measurement quality and completeness, potentially explaining the gender-specific differences observed in our data, where factors such as thinner skin in women or older populations or thicker blood vessels in men could enhance photoplethysmography signals [40,41]. Other factors that have been mentioned in the scientific literature are the skin pigmentations of the participants; several studies have noted a correlation between the Fitz-Patrick skin scale and heart rate measurements via photoplethysmography, suggesting reduced heart rate measurement accuracy in individuals with higher Fitz-Patrick scale values [40,42], although some studies [43] did not find this correlation. To mitigate this, wearables are starting to implement enhanced photoplethysmography sensors or infrared measurements [42]. Moreover, gender may present contradictory effects on the data collection [40]. The influence of BMI on signal quality remains uncertain [40], with additional variables such as blood vessel dilation during physical activity also playing a role [40]. Participant adherence challenges were also observed, as some felt uncomfortable wearing the device continuously, especially at night.

Our research aligns with previous studies on average daily activity and sleep duration, particularly regarding the influence of age [15,32,34]. Although body shell temperature readings aligned with physiological norms, indicating their usefulness in identifying individual anomalies [24], they were notably low, suggesting potential nighttime evaporation processes [24]. As with prior research, factors such as gender and BMI affected body shell temperature [4,24], necessitating further research to comprehensively grasp temperature impacts.

Limitations

The internal validity of the wearables used in this study was not evaluated; however, models older than the WPHR used here, some requiring manual sleep initiation, have shown largely satisfactory results. Gruwez et al [44] identified a significant correlation between Withings wearables and a research-grade actigraph for step count during daily activities. However, a review by Fuller et al [45] revealed an underestimation of step count in Withings wearables in most examined studies. Comparisons of Withings wearables to polysomnography revealed no significant differences in sleep duration measurements [39], and various studies have noted consistent correlations with polysomnography-recorded total sleep duration [39,44]. Regulatory constraints restricted the recruitment of younger participants in our study, affecting its generalizability. Furthermore, smartphone usage restrictions in Kenyan boarding schools, especially pertinent to the substantial under-15 years demographic, could pose challenges for future research [17]. In addition, as explained in the methods section, only the data sets of the adult participants were analyzed with regard to weather effects. Our study has other limitations such as uncorrected multiple testing and potential influences of the COVID-19 pandemic on data. In Kenya, the COVID-19 pandemic caused governmental restrictions on public life such as nighttime curfew until October 20, 2021, which affected 22 (26.5%) participants during their study participation—likely influencing the data collected [46]. Research during the 2020 COVID lockdowns reported reduced total sleep duration and increased napping [47], and a South African study observed decreased mobility, particularly on weekends [48]. Future research should consider a wider range of variables, account for the carryover effects of prior day heat strain [49], and address technical challenges such as inaccessible raw data, software issues, high data missingness for heart rate due to factors such as skin type [40], and possible algorithmic biases. Our study does not account for factors such as clothing, air circulation, bedding, and individual differences such as gender, which can influence the relationship between heat and sleep duration [36].

Conclusions

Our study shows the potential of wearable devices to monitor vital signs and assess the impact of environmental exposures on health in rural sub-Saharan settings, with implications for understanding the nuanced effects of climate change. Despite a robust data set, our findings indicate the need for improved wearable technology to ensure data completeness across diverse demographic groups, acknowledging the impact of factors such as age, gender, and BMI. The positive correlation between physical activity levels and high WBGT offers new insights into behavior during extreme weather conditions, while the nonsignificant trends in sleep duration in relation to temperature call for further investigation. These observations are crucial for public health strategies in climate-vulnerable regions, guiding the integration of wearables in longitudinal health monitoring and climate resilience research. Future studies should expand on the relationship between weather and health outcomes, including a broader demographic and addressing technical challenges identified in wearable data collection. This research contributes to a growing body of knowledge that will inform both technological innovation in health monitoring and the development of interventions to mitigate the health impacts of global climate dynamics.

Acknowledgments

We thank the German Research Foundation (Deutsche Forschungsgemeinschaft) for supporting this work as part of a Deutsche Forschungsgemeinschaft–funded research unit (Forschungsgruppe FOR 2936/project: 409670289). We acknowledge the support of Else Kröner-Fresenius-Stiftung from the Heidelberg Graduate School of Global Health. For the publication fee we acknowledge financial support by Heidelberg University and by the Else Kröner-Fresenius-Stiftung within the Heidelberg Graduate School of Global Health. The funders had no influence on the design, data collection and analysis, the decision to publish or the preparation of the manuscript. During the preparation of this work, the author(s) used ChatGPT in order to improve language and readability of this manuscript. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Data Availability

Deidentified wearable data, weather data, and R scripts will be available for scientific purposes with publication on reasonable request to the corresponding author SB (barteit@uni-heidelberg.de).

Authors' Contributions

SB, TB, and SM planned this study and led its implementation. SM, DO, DK, JOM, COO, and IM conducted and supervised the study on site, remotely advised by SB and IM. SH, MK, MAM, SB, and IM assessed the data and performed the data analysis. IM analyzed the data, supervised and verified by SB. All authors critically and freely contributed to the final draft and confirmed the publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Detailed technical information on the wearables used.
[DOCX File , 57 KB - mhealth_v12i1e54669_app1.docx]

Multimedia Appendix 2
Additional information on data analysis.
[DOCX File , 18 KB - mhealth_v12i1e54669_app2.docx]

Multimedia Appendix 3
Added variables and residual plots of regression models.
[DOCX File , 1049 KB - mhealth_v12i1e54669_app3.docx]

Multimedia Appendix 4
Detailed results on data completeness and data validity.
[DOCX File , 131 KB - mhealth_v12i1e54669_app4.docx]

Multimedia Appendix 5
Additional regression models incorporating weekday or weekend.
[DOCX File , 23 KB - mhealth_v12i1e54669_app5.docx]

Multimedia Appendix 6
ChatGPT conversation.
[DOCX File , 53 KB - mhealth_v12i1e54669_app6.docx]

References

1. Green H, Bailey J, Schwarz L, Vanos J, Ebi K, Benmarhnia T. Impact of heat on mortality and morbidity in low and middle income countries: A review of the epidemiological evidence and considerations for future research. Environ Res 2019 Apr;171:80-91. [doi: 10.1016/j.envres.2019.01.010] [Medline: 30660921]
2. IPCC. Climate change 2022: impacts, adaption, and vulnerability. Working group II contribution to the IPCC sixth assessment report. In: Pörtner HO, Roberts DC, Tignor M, Poloczanska ES, Mintenbeck K, Alegría A, et al, editors. IPCC. Cambridge, UK: Cambridge University Press; 2022.
3. Huhn S, Axt M, Gunga HC, Maggioni MA, Munga S, Obor D, et al. The impact of wearable technologies in health research: scoping review. JMIR Mhealth Uhealth 2022 Jan 25;10(1):e34384 [FREE Full text] [doi: 10.2196/34384] [Medline: 35076409]
4. Koch M, Matzke I, Huhn S, Gunga H, Maggioni MA, Munga S, et al. Wearables for measuring health effects of climate change-induced weather extremes: scoping review. JMIR Mhealth Uhealth 2022 Sep 09;10(9):e39532 [FREE Full text] [doi: 10.2196/39532] [Medline: 36083624]
5. Ming DK, Sangkaew S, Chanh HQ, Nhat PTH, Yacoub S, Georgiou P, et al. Continuous physiological monitoring using wearable technology to inform individual management of infectious diseases, public health and outbreak responses. Int J Infect Dis 2020 Jul;96:648-654 [FREE Full text] [doi: 10.1016/j.ijid.2020.05.086] [Medline: 32497806]
6. Ortiz-Bobea A, Ault TR, Carrillo CM, Chambers RG, Lobell DB. Anthropogenic climate change has slowed global agricultural productivity growth. Nat Clim Chang 2021 Apr 01;11(4):306-312. [doi: 10.1038/s41558-021-01000-1]
7. Gleixner S, Lüttringhaus S, Kaufmann J, Smytzek P, Gornott C. Assessment of climatic impact drivers in Kenya. United Nations Office for Disaster Risk Reduction. Bonn and Eschborn, Germany: Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH; 2022. URL: https://www.preventionweb.net/publication/assessment-climatic-impact-drivers-kenya [accessed 2022-09-15]
8. Minor K, Bjerre-Nielsen A, Jonasdottir SS, Lehmann S, Obradovich N. Rising temperatures erode human sleep globally. One Earth 2022 May;5(5):534-549. [doi: 10.1016/j.oneear.2022.04.008]
9. Natarajan A, Pantelopoulos A, Emir-Farinas H, Natarajan P. Heart rate variability with photoplethysmography in 8 million individuals: a cross-sectional study. Lancet Digit Health 2020 Dec;2(12):e650-e657 [FREE Full text] [doi: 10.1016/S2589-7500(20)30246-6] [Medline: 33328029]
10. Radin JM, Wineinger NE, Topol EJ, Steinhubl SR. Harnessing wearable device data to improve state-level real-time surveillance of influenza-like illness in the USA: a population-based study. The Lancet Digital Health 2020 Feb;2(2):e85-e93. [doi: 10.1016/s2589-7500(19)30222-5]
11. Sahu S, Sett M, Kjellstrom T. Heat exposure, cardiovascular stress and work productivity in rice harvesters in India: implications for a climate change future. Ind Health 2013;51(4):424-431 [FREE Full text] [doi: 10.2486/indhealth.2013-0006] [Medline: 23685851]
12. Parsons K. Heat stress standard ISO 7243 and its global application. Ind Health 2006 Jul;44(3):368-379 [FREE Full text] [doi: 10.2486/indhealth.44.368] [Medline: 16922180]
13. What is the heat index? National Weather Service. URL: https://www.weather.gov/ama/heatindex [accessed 2022-09-19]
14. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. PLoS Med 2007 Oct 16;4(10):e296 [FREE Full text] [doi: 10.1371/journal.pmed.0040296] [Medline: 17941714]
15. Huhn S, Matzke I, Koch M, Gunga H, Maggioni M, Sié A, et al. Using wearable devices to generate real-world, individual-level data in rural, low-resource contexts in Burkina Faso, Africa: A case study. Frontiers in Public Health 2022;10:972177 [FREE Full text] [doi: 10.3389/fpubh.2022.972177] [Medline: 36249225]
16. Barteit S, Boudo V, Ouedraogo A, Zabré P, Ouremi L, Sié A, et al. Feasibility, acceptability and validation of wearable devices for climate change and health research in the low-resource contexts of Burkina Faso and Kenya: Study protocol. PLoS One 2021;16(9):e0257170 [FREE Full text] [doi: 10.1371/journal.pone.0257170] [Medline: 34591893]
17. Odhiambo FO, Laserson KF, Sewe M, Hamel MJ, Feikin DR, Adazu K, et al. Profile: the KEMRI/CDC Health and Demographic Surveillance System--Western Kenya. Int J Epidemiol 2012 Aug;41(4):977-987. [doi: 10.1093/ije/dys108] [Medline: 22933646]
18. Pulse HR - User Guide. URL: https://support.withings.com/hc/en-us/articles/360015826813-Pulse-HR-User-Guid [accessed 2022-09-18]
19. e-TakesCare - Client Support. URL: https://www.e-takescare.com/en/assistance-particuliers [accessed 2022-09-18]
20. Carter AW, Zaitchik BF, Gohlke JM, Wang S, Richardson MB. Methods for estimating wet bulb globe temperature from remote and low‐cost data: a comparative study in Central Alabama. GeoHealth 2020 May 21;4(5):e2019GH000231. [doi: 10.1029/2019gh000231]

21. A healthy lifestyle - WHO recommendations. World Health Organization. 2010. URL: https://www.who.int/europe/news-room/fact-sheets/item/a-healthy-lifestyle---who-recommendations [accessed 2022-11-17]
22. Mangrum JM, DiMarco JP. The evaluation and management of bradycardia. N Engl J Med 2000 Mar 09;342(10):703-709. [doi: 10.1056/NEJM200003093421006] [Medline: 10706901]
23. Tanaka H, Monahan KD, Seals DR. Age-predicted maximal heart rate revisited. J Am Coll Cardiol 2001 Jan;37(1):153-156 [FREE Full text] [doi: 10.1016/s0735-1097(00)01054-8] [Medline: 11153730]
24. Campbell I. Body temperature and its regulation. Anaesthesia & Intensive Care Medicine 2008 Jun;9(6):259-263. [doi: 10.1016/j.mpaic.2008.04.009]
25. Jonasdottir SS, Minor K, Lehmann S. Gender differences in nighttime sleep patterns and variability across the adult lifespan: a global-scale wearables study. Sleep 2021 Feb 12;44(2):zsaa169. [doi: 10.1093/sleep/zsaa169] [Medline: 32886772]
26. Pulse HR - I'm having issues with my sleep data. Withings. URL: https://support.withings.com/hc/en-us/articles/360011368074-Pulse-HR-I-m-having-issues-with-my-sleep-data-What-should-I-do- [accessed 2022-03-31]
27. Troiano RP, Berrigan D, Dodd KW, Mâsse LC, Tilert T, McDowell M. Physical activity in the United States measured by accelerometer. Med Sci Sports Exerc 2008 Jan;40(1):181-188. [doi: 10.1249/mss.0b013e31815a51b3] [Medline: 18091006]
28. Sallis JF, Cerin E, Conway TL, Adams MA, Frank LD, Pratt M, et al. Physical activity in relation to urban environments in 14 cities worldwide: a cross-sectional study. Lancet 2016 May 28;387(10034):2207-2217 [FREE Full text] [doi: 10.1016/S0140-6736(15)01284-2] [Medline: 27045735]
29. Davis C, Engelbrecht F, Tadross M, Wolski P, Archer E. Future climate change over Southern Africa. CSIR. 2017 Oct. URL: https://researchspace.csir.co.za/dspace/handle/10204/10085 [accessed 2024-03-12]
30. Jehn M, Gebhardt A, Liebers U, Kiran B, Scherer D, Endlicher W, et al. Heat stress is associated with reduced health status in pulmonary arterial hypertension: a prospective study cohort. Lung 2014 Aug;192(4):619-624. [doi: 10.1007/s00408-014-9587-4] [Medline: 24816966]
31. Okamoto-Mizuno K, Tsuzuki K, Mizuno K, Iwaki T. Effects of partial humid heat exposure during different segments of sleep on human sleep stages and body temperature. Physiol Behav 2005 Jan 17;83(5):759-765. [doi: 10.1016/j.physbeh.2004.09.009] [Medline: 15639161]
32. Rifkin DI, Long MW, Perry MJ. Climate change and sleep: A systematic review of the literature and conceptual framework. Sleep Med Rev 2018 Dec;42:3-9 [FREE Full text] [doi: 10.1016/j.smrv.2018.07.007] [Medline: 30177247]
33. Abdi H. Holm's sequential Bonferroni procedure. In: Salkind N, editor. Encyclopedia of Research Design. Thousand Oaks, CA: SAGE Publications, Inc; 2010.
34. Al-Mohannadi AS, Farooq A, Burnett A, Van Der Walt M, Al-Kuwari MG. Impact of climatic conditions on physical activity: a 2-year cohort study in the arabian gulf region. Journal of Physical Activity and Health 2016 Sep;13(9):929-937. [doi: 10.1123/jpah.2015-0593] [Medline: 27631650]
35. Yiga P, Seghers J, Ogwok P, Matthys C. Determinants of dietary and physical activity behaviours among women of reproductive age in urban sub-Saharan Africa: a systematic review. Br J Nutr 2020 May 28;124(8):761-772. [doi: 10.1017/s0007114520001828]
36. Caddick ZA, Gregory K, Arsintescu L, Flynn-Evans EE. A review of the environmental parameters necessary for an optimal sleep environment. Building and Environment 2018 Mar;132:11-20. [doi: 10.1016/j.buildenv.2018.01.020]
37. Irwin MR. Why sleep is important for health: a psychoneuroimmunology perspective. Annu Rev Psychol 2015 Jan 03;66:143-172 [FREE Full text] [doi: 10.1146/annurev-psych-010213-115205] [Medline: 25061767]
38. Krittanawong C, Tunhasiriwet A, Wang Z, Zhang H, Farrell A, Chirapongsathorn S, et al. Association between short and long sleep durations and cardiovascular outcomes: a systematic review and meta-analysis. Eur Heart J Acute Cardiovasc Care 2019 Dec;8(8):762-770. [doi: 10.1177/2048872617741733] [Medline: 29206050]
39. Mantua J, Gravel N, Spencer RMC. Reliability of sleep measures from four personal health monitoring devices compared to research-based actigraphy and polysomnography. Sensors (Basel) 2016 May 05;16(5):646 [FREE Full text] [doi: 10.3390/s16050646] [Medline: 27164110]
40. Fine J, Branan KL, Rodriguez AJ, Boonya-Ananta T, Ajmal, Ramella-Roman JC, et al. Sources of inaccuracy in photoplethysmography for continuous cardiovascular monitoring. Biosensors (Basel) 2021 Apr 16;11(4):126 [FREE Full text] [doi: 10.3390/bios11040126] [Medline: 33923469]
41. Moço AV, Stuijk S, de Haan G. Skin inhomogeneity as a source of error in remote PPG-imaging. Biomed. Opt. Express 2016 Oct 26;7(11):4718. [doi: 10.1364/boe.7.004718]
42. Colvonen PJ, DeYoung PN, Bosompra NA, Owens RL. Limiting racial disparities and bias for wearable devices in health science research. Sleep 2020 Oct 13;43(10):zsaa159 [FREE Full text] [doi: 10.1093/sleep/zsaa159] [Medline: 32893865]
43. Bent B, Goldstein BA, Kibbe WA, Dunn JP. Investigating sources of inaccuracy in wearable optical heart rate sensors. NPJ Digit Med 2020;3:18 [FREE Full text] [doi: 10.1038/s41746-020-0226-6] [Medline: 32047863]
44. Gruwez A, Libert W, Ameye L, Bruyneel M. Reliability of commercially available sleep and activity trackers with manual switch-to-sleep mode activation in free-living healthy individuals. Int J Med Inform 2017 Jun;102:87-92. [doi: 10.1016/j.ijmedinf.2017.03.008] [Medline: 28495352]

45. Fuller D, Colwell E, Low J, Orychock K, Tobin MA, Simango B, et al. Reliability and validity of commercially available wearable devices for measuring steps, energy expenditure, and heart rate: systematic review. JMIR Mhealth Uhealth 2020 Sep 08;8(9):e18694 [FREE Full text] [doi: 10.2196/18694] [Medline: 32897239]
46. Otieno J. Council of governors say no to lockdown over COVID-19 surge. The Star. 2021 Dec 20. URL: https://www.the-star.co.ke/news/2021-12-20-council-of-governors-say-no-to-lockdown-over-covid-19-surge/ [accessed 2022-09-19]
47. Petrov ME, Pituch KA, Kasraeian K, Jiao N, Mattingly J, Hasanaj K, et al. Impact of the COVID-19 pandemic on change in sleep patterns in an exploratory, cross-sectional online sample of 79 countries. Sleep Health 2021 Aug;7(4):451-458. [doi: 10.1016/j.sleh.2021.05.007] [Medline: 34193394]
48. Borofsky Y, Günther I. Mobility in informal settlements during a public lockdown: A case study in South Africa. PLoS One 2022;17(12):e0277465 [FREE Full text] [doi: 10.1371/journal.pone.0277465] [Medline: 36548350]
49. Garzon-Villalba XP, Mbah A, Wu Y, Hiles M, Moore H, Schwartz SW, et al. Exertional heat illness and acute injury related to ambient wet bulb globe temperature. Am J Ind Med 2016 Dec;59(12):1169-1176. [doi: 10.1002/ajim.22650] [Medline: 27779310]

Abbreviations

HDSS: Health and Demographic Surveillance System
LMICs: low- and middle-income countries
MWU: Mann-Whitney U test
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology
WBGT: wet bulb globe temperature
WPHR: Withings pulse heart rate

Edited by G Eysenbach, L Buis; submitted 18.11.23; peer-reviewed by H Aghayan Golkashani; comments to author 22.12.23; revised version received 16.01.24; accepted 26.01.24; published 04.07.24.

Please cite as:
Matzke I, Huhn S, Koch M, Maggioni MA, Munga S, Muma JO, Odhiambo CO, Kwaro D, Obor D, Bärnighausen T, Dambach P, Barteit S
Assessment of Heat Exposure and Health Outcomes in Rural Populations of Western Kenya by Using Wearable Devices: Observational Case Study
JMIR Mhealth Uhealth 2024;12:e54669
URL: https://mhealth.jmir.org/2024/1/e54669
doi:10.2196/54669
PMID:38963698

Original Paper

Implementation of a Technology-Based Mobile Obstetric Referral Emergency System (MORES): Qualitative Assessment of Health Workers in Rural Liberia

Christopher W Reynolds[1], MD, MS; HaEun Lee[2], PhD; Joseph Sieka[3], MD, MSc; Joseph Perosky[4], MD; Jody R Lori[2], PhD

[1]University of Michigan Medicine, Ann Arbor, MI, United States

[2]University of Michigan School of Nursing, Ann Arbor, MI, United States

[3]A.M. Dogiloti College of Medicine, University of Liberia, Monrovia, Liberia

[4]Spectrum Health, Grand Rapids, MI, United States

Corresponding Author:
Christopher W Reynolds, MD, MS
University of Michigan Medicine
1500 E. Medical Center Dr.
Ann Arbor, MI, 48109
United States
Phone: 1 302 377 9925
Email: chwre@med.umich.edu

Abstract

Background: Maternal mortality remains a persistent challenge in low- and middle-income countries, where evidence-based interventions of obstetric triage and prehospital communication remain sparse. There is limited implementation evidence for technology-based approaches to improve obstetric care in such contexts. Liberia struggles with maternal mortality, particularly in rural areas where deaths are attributable to delays from absent triage and interfacility communication. We implemented a Mobile Obstetric Referral Emergency System (MORES) in rural Bong County to improve prehospital transfer, health worker attentiveness, and patient care for critical obstetric patients. MORES consisted of triage training and a 2-way, templated WhatsApp communication system to reduce delays among patients transferred from rural health facilities (RHF) to hospitals.

Objective: This study aimed to examine MORES implementation outcomes of usability, fidelity, effectiveness, sustainability, and scalability, as well as additional impacts on the wider health system.

Methods: A structured case study design interview was developed by Liberian and US experts in obstetric triage. Participants included 62 frontline obstetric health providers including midwives (38/62, 61%), nurses (20/62, 32%), physicians assistants (3/62, 5%), and physicians (1/62, 2%) from 19 RHFs and 2 district hospitals who had used MORES for 1 year. Individual interviews were conducted on MORES implementation outcomes, transcribed, and analyzed in NVivo (version 12; Lumivero) with a team-based coding methodology. Content analysis with a deductive approach examined implementation outcomes of usability, fidelity, effectiveness, sustainability, and scalability, while an inductive approach categorized the unanticipated impacts of MORES on the wider health system.

Results: Four domains were identified regarding MORES implementation: Usability and Fidelity, Effectiveness, Sustainability and Scalability, and Health System Impact. All participants perceived MORES to have high usability and fidelity, as the triage and messaging system was implemented as intended for critical obstetric patients (62/62, 100%). For effectiveness, MORES accomplished its intended aims by improving prehospital transfer (57/62, 92%), increasing health worker attentiveness (39/62, 63%), and contributing to improved patient care (34/62, 55%). MORES was perceived as sustainable and scalable (62/62, 100%), particularly if technological barriers (21/62, 34%) and staff training (19/62, 31%) were addressed. MORES impacted the wider health system in unanticipated ways including improved coordination and accountability (55/62, 89%), feedback mechanisms for hospitals and RHFs (48/62, 77%), interprofessional teamwork (21/62, 34%), longitudinal follow-up care (20/62, 32%), creating a record of care delays (17/62, 27%), and electronic health record infrastructure (13/62, 21%).

Conclusions: MORES was perceived to have high usability, fidelity, effectiveness, sustainability, and scalability by frontline obstetric providers in rural Liberia. MORES accomplished the intended aims of improving prehospital transfer, increasing health worker attentiveness, and contributing to improved patient care. Additionally, MORES strengthened the health system through

6 domains which impacted individual and system levels. Future studies should quantitatively evaluate delay and morbidity reductions and strategies for scaling MORES.

(JMIR Mhealth Uhealth 2024;12:e58624) doi:10.2196/58624

KEYWORDS

mHealth; mobile triage; referral pathways; Liberia; LMIC; low- income country; obstetric triage; third delay; mobile health; mobile application; digital health; digital intervention; smartphone; middle-income country

Introduction

Despite increased attention, funding, and innovation, maternal mortality continues to disproportionately affect low- and middle-income countries (LMICs) [1]. Liberia in West Africa has one of the highest maternal mortality ratios in the world with 652 deaths per 100,000 births [2]. This inequity is due to many factors including resource limitations, workforce shortages, and an overall unstable and underfunded health system following decades of conflict, the 2014-2016 Ebola epidemic, and COVID-19 pandemic [3]. A major contributing factor impacting Liberia's high maternal mortality is a lack of standardized obstetric triage protocols and prehospital communication systems between health facilities [4]. Such systems allow for efficient patient transfer and have consistently been shown to improve maternal and neonatal health outcomes [5,6]. In Liberia, a lack of standardized triage, referral, and transfer systems leads to care delays and increases morbidity at the prehospital and intra-facility level, also known as the "third delay," which occurs after patients arrive at health facilities [7]. This lack of infrastructure creates a critical blind spot for timely maternal care, especially for patients in rural areas who must travel hours on foot or by personal car to reach a health facility.

Over the last decade, the Liberian Ministry of Health has emphasized the importance of combating maternal and perinatal mortality at community and national levels [8]. Rural health facilities (RHFs) are community clinics established throughout rural areas of Liberia to serve as the first line of care for patients living far from district hospitals [9]. RHF staff, midwives, and nurses are the frontline obstetric health workers in Liberia, responsible for preliminary assessments, treatment and stabilization, and escalation of care by referring patients to district hospitals when necessary [10]. However, care delays persist due to system limitations. Verbal autopsy reports of 35 maternal deaths from Bong County, Liberia revealed 2 major contributors to maternal death: inadequate staff training and ineffective communication between RHFs and hospitals [4]. A promising solution to improve maternal transfer and prehospital communication systems leverages mobile health interventions (mHealth), including messaging platforms to transmit patient information from a referring facility to the hospital prior to arrival [11]. Such systems may be a promising intervention to reduce prehospital and intrafacility delays by allowing health workers to prepare resources for immediate intervention upon patient arrival [12]. Several mHealth approaches have been successful in reducing care delays in LMICs including Ghana, and a recent study by our team showed high desirability for such a program among 130 health workers from RHFs and hospitals in Bong County [13,14].

In response, our team developed and implemented the Mobile Obstetric Referral Emergency System (MORES) among 20 RHFs and 2 district referral hospitals in Bong County, Liberia. MORES has two main components as follows: (1) triage capacity building training among obstetric health workers, and (2) implementation of a mobile, WhatsApp-based communication platform, aimed at reducing prehospital and intrafacility care delays [15]. Specifically, the intended aims of MORES were to reduce care delays through improved prehospital transfer, increased health worker attentiveness, and improved patient care. MORES works in the following ways: when a pregnant patient presents to an RHF in critical condition, she will undergo standardized triage by a community nurse midwife to assess acuity. If the RHF midwife determines that this patient requires higher level or urgent care at the district hospital, the RHF midwife will send a templated WhatsApp message to the district hospital with information including the patient's condition, past medical and obstetric history, any interventions administered, and method of transport to the hospital. Immediately upon receiving this message from the RHF midwife, hospital obstetric nurses begin preparing for that patient's arrival including gathering needed medications, informing physicians, readying the emergency department, and preparing the operating room if urgent surgical intervention is anticipated. If delays emerge during patient transport, the RHF and facility providers can work collaboratively through MORES messaging to overcome these barriers. Upon arrival at the hospital, the referred obstetric patient will undergo an additional standardized triage by the obstetric nurses and receive a green, yellow, or red wristband corresponding to acuity level before admission or initiation of urgent intervention. Implemented in March 2022, MORES demonstrated significant improvements in triage knowledge and was used by more than 50 health workers to send 359 referral messages to transfer patients over the course of a year.

In this qualitative study, we evaluated the implementation outcomes of MORES including usability, fidelity, effectiveness of accomplishing intended aims, sustainability, and scalability. We also evaluated if MORES had additional impacts on the health system more broadly beyond its intended impacts.

Methods

Study Setting

This study took place in Bong County, Liberia, the third most populous and primarily rural county in Liberia [14]. Researchers at the University of Liberia and the University of Michigan

partnered with the Bong County Health Team to design and implement MORES. Further detail of the MORES intervention, including its impact on health outcomes, is published elsewhere [15]. This study was conducted 1 year post implementation. Bong County was selected as the study site following efforts by the Liberia Ministry of Health to focus on improving maternal outcomes in this region given recent increases in maternal deaths [4]. Additionally, we have strong institutional partnerships between researchers and the County Health Team. Prior to MORES, Bong County had limited prehospital infrastructure consisting of 2 ambulances, no medical dispatch, and limited avenues for RHFs to communicate with hospitals. Health workers previously documented all medical information on a referral paper, which the patient was responsible for carrying and presenting to the referral hospital upon arrival. Often taking hours to arrive at said hospital, these papers would frequently be lost or damaged in transit.

Mores Implementation

In March 2022, specialists from Ghana and the University of Michigan with extensive experience in mHealth prehospital obstetric systems conducted capacity-building training among health workers from 20 RHFs and 2 district hospitals. Details regarding the templated communication messages and initial MORES training have been described [15]. Following this training, health workers returned to their respective health facilities to train colleagues in MORES. The 2 district hospitals were provided with smartphones to be used in the emergency department solely for MORES, while RHF workers used their personal cell phones. RHF and hospital workers were supplied cellular scratch cards to reload data at regular intervals throughout the study.

Study Design

To determine health worker perceptions of MORES implementation outcomes, we conducted a qualitative study using content analysis, following 1 year of MORES implementation in Bong County. First, US and Liberian experts in obstetric care and qualitative methods developed a structured interview script. The interview script was reviewed with Liberian health workers to adjust the language to the local context, purge repetitive items, and consolidate questions. The final interview script contained 20 items evaluating 6 major metrics including demographics, MORES purpose, usability, impact on referrals, facility linkages and feedback, and areas of strength and improvement (Multimedia Appendix 1).

Data Collection and Participants

In March 2023, 10 research assistants (RAs) from the University of Liberia traveled to Bong County to conduct interviews. The RAs consisted of both male and female Master of Public Health Students who had completed their course of study and were working on their Master theses and Master students who had completed a qualitative methods course. All were bachelor degree holders in various health-related fields and had extensive prior experience conducting qualitative interviews, which was a criterion for RA selection. During the data collection period, participants knew that RAs were students from the University of Liberia with an interest in obstetric health access and part of the research team who would be studying the implementation of MORES within their health system. A prior relationship with the study team had been established, given that this group was the same that implemented MORES 1 year prior.

In collaboration with the Bong County Health Team, RAs invited all health workers from the 20 RHFs and 2 district hospitals who had used MORES in the past year to participate using purposive sampling. Health workers were eligible if they were 18 years or older, currently employed by a Bong County health facility, and had used MORES at least once. Participants were excluded if they were unable to allocate sufficient time for an uninterrupted interview (10 minutes) or had never used MORES. RAs approached eligible participants face-to-face and invited them to complete the interview. Written consent was collected from eligible health workers who were willing to participate. It was made clear that the interview was anonymous, voluntary, and that their answers or decision of whether to participate would not influence their employment or ability to use MORES in the future. Individual interviews were conducted in a private location at the RHF or hospital according to participant preference. Each interview took approximately 10 minutes, and the interviews were audio recorded and stored in a secure lock box for transport to the University of Liberia. Researchers from Liberia and the University of Michigan met periodically throughout the data collection period to determine data saturation. Though saturation occurred after approximately 30 interviews, the opportunity to participate remained available so that all health workers who wished to share their perceptions were able to.

Data Analysis

Interviews were transcribed in Microsoft Word and immediately uploaded to a secure Dropbox folder only available to study researchers. All identifiable information was omitted during transcription. Two researchers began analysis through iterative immersion in the data, using content analysis with a deductive approach to categorize implementation outcomes of usability, fidelity, effectiveness, sustainability, and scalability to guide codebook development. Upon immersion in the data, it was determined that additional, unanticipated themes of MORES impact on the wider health system were frequently mentioned throughout the interviews. Therefore, a combined deductive-inductive content analysis approach was used to generate codes that categorized study aims while also accounting for the emergence of new themes. Once domains and major themes were identified, the codebook was further refined to identify subthemes under the major themes [16]. Transcripts were imported to NVivo (version 12; Lumivero) and coded using the codebook with a content analysis approach to answer the research question: “What are health workers’ perceptions on the implementation of MORES in light of its intended aims and impact on the wider health system?” [17]. Multiple measures were taken to ensure trustworthiness including (1) triangulation of interviews from multiple health specialists, (2) creation and validation of a consensus codebook, (3) inclusion of Liberian health workers as core members of the analytic team, and (4) member checking with Liberian health workers regarding key results.

Ethical Considerations

This study was approved by the University of Michigan institutional review board (HUM00195449) and the University of Liberia institutional review board (IRB00013730), whose protocols were followed throughout the entire study period. All participants underwent a process of written, informed consent prior to participation and all data were deidentified prior to analysis. This study adheres to COREQ (consolidated criteria for reporting qualitative research) guidelines for reporting in qualitative research [18].

Results

Participant Demographics

Sixty-two health workers from 19 RHFs and 2 district hospitals participated, representing 95% of total facilities taking part in MORES (Figure 1). Health workers from one facility included in the original training did not transmit any messages due to the lack of a smartphone with WhatsApp capability, therefore health workers from 19 facilities were included in our data collection. Participants included registered nurses, midwives, physician assistants, and medical doctors (Table 1). Three participants had additional specialization as neonatal specialists, and 12 occupied leadership roles, including nursing officer-in-charge (n=9), obstetrics supervisor (n=2), and chief medical officer (n=1). Most participants were between 35 and 45 years old, with years of professional experience ranging from 1 to 19 years (median 8 years, IQR 5-10). Across facilities, participants working in hospitals were slightly older and had more work experience than those in RHFs.

Four major domains were used to evaluate the implementation of MORES: (1) usability and fidelity, (2) effectiveness of intended impact, (3) sustainability and scalability, and (4) health system impact. Among these domains were 15 themes and 43 subthemes, outlined in Figure 2.

Figure 1. Rural health facilities and hospitals participating in the Mobile Obstetric Referral Emergency Systems (MORES) in Bong County, Liberia. Hospitals are denoted in yellow (2), with RHFs in red (19).

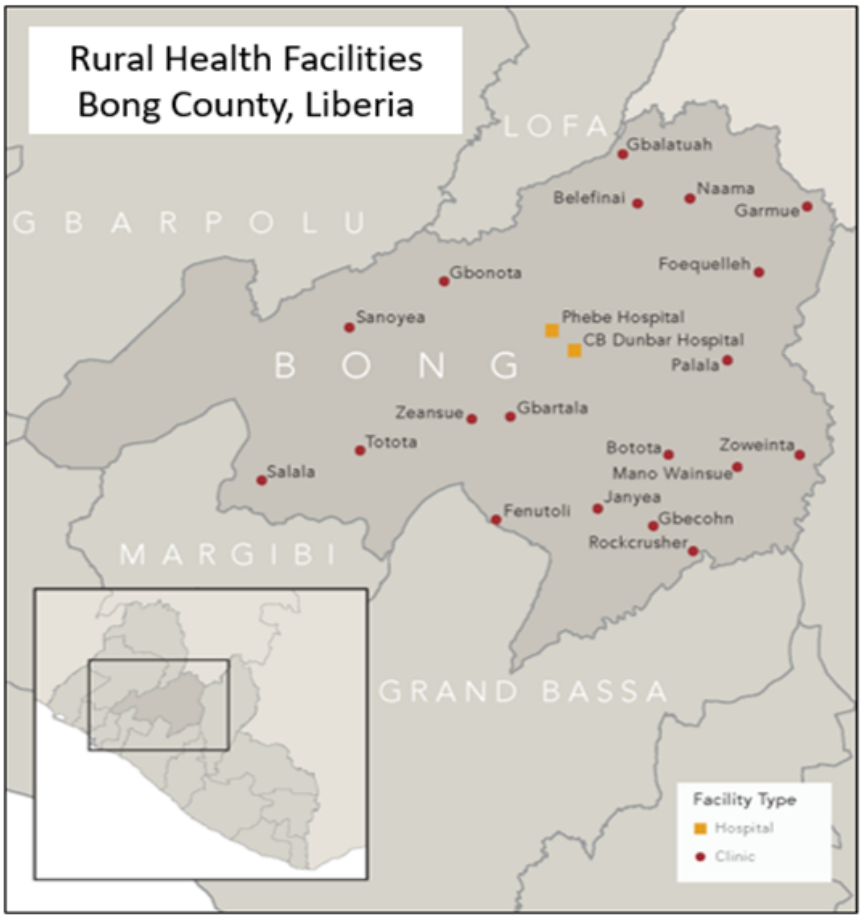
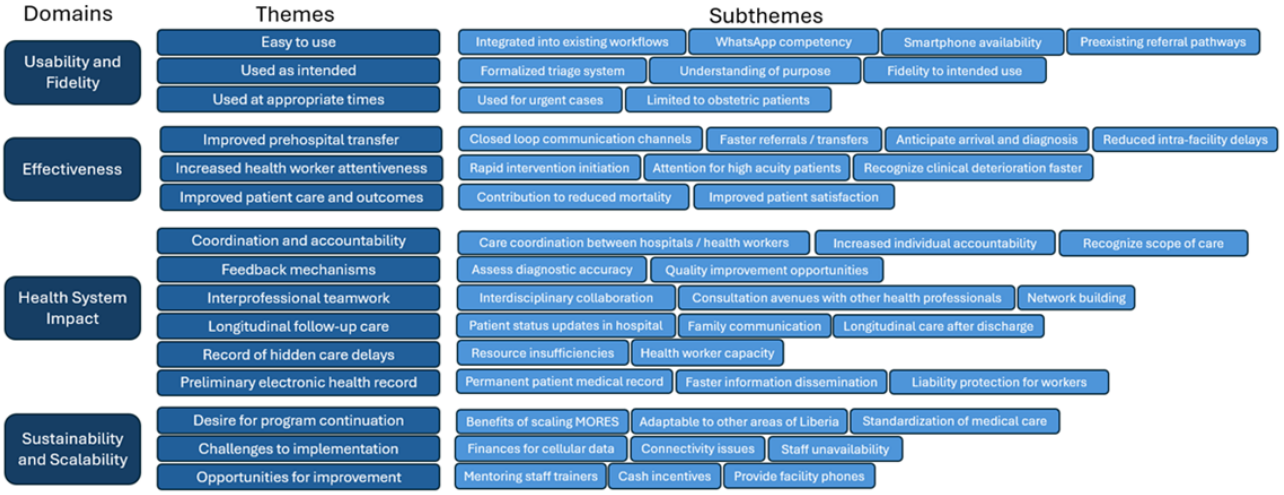


Table 1. Participant demographics for the Mobile Obstetric Referral Emergency System (MORES) evaluation study in Bong County, Liberia.

	Total (N=62)	Rural health facilities (n=33)	Hospital 1 (CB Dunbar) (n=15)	Hospital 2 (Phebe) (n=14)
Age (years), n (%)				
<20	0 (0)	0 (0)	0 (0)	0 (0)
20-29	6 (10)	5 (15)	0 (0)	1 (7)
30-39	30 (48)	19 (58)	7 (47)	4 (29)
40-49	17 (27)	9 (27)	5 (33)	3 (21)
>50	7 (11)	0 (0)	2 (13)	5 (36)
Unknown	2 (3)	0 (0)	1 (7)	1 (7)
Age (years), mean (SD)	39.2 (7.22)	36.7 (4.96)	42.1 (6.91)	42.7 (9.46)
Role, n (%)				
Registered nurse	20 (32)	10 (30)	5 (33)	5 (36)
Midwife	38 (61)	23 (70)	9 (60)	6 (43)
Physician assistant	3 (5)	0 (0)	1 (7)	2 (14)
Medical doctor	1 (2)	0 (0)	0 (0)	1 (7)
Time in practice, n (%)				
<3 years	4 (6)	4 (12)	0 (0)	0 (0)
3-6 years	19 (31)	13 (39)	2 (13)	4 (29)
7-10 years	24 (39)	12 (36)	7 (47)	5 (36)
11-15 years	11 (18)	3 (9)	4 (27)	4 (29)
16-20 years	2 (3)	1 (3)	1 (7)	0 (0)
Unknown	2 (3)	0 (0)	1 (7)	1 (7)
Time in practice (years), mean (SD)	7.8 (3.61)	6.7 (3.60)	9.6 (3.20)	8.5 (3.10)

Figure 2. Organization of domains, themes, and subthemes from MORES interviews. MORES: Mobile Obstetric Referral Emergency System.



MORES Usability and Fidelity

MORES was evaluated for its usability and fidelity, meaning if participants used MORES as it was intended to be implemented. For usability, all 62 participants affirmed the MORES intervention was easy to use and beneficial to the patients they served (100%). They described MORES as easily integrated into existing workflows within their health system, as most had access to smartphones, knew how to use WhatsApp, and used MORES as part of a preexisting RHF-hospital referral system. For fidelity, participants described how they used MORES as it was intended to be implemented, including for making patient referrals, interfacility communication, and triage upon hospital arrival: “We get the patient’s name, age, reasoning of referral, mode of transfer. If referring through an ambulance or commercial car, we can place it in the WhatsApp group chat and send it” [P41].

Upon arrival at the hospital, patients were immediately triaged with a green, yellow, or red wristband according to acuity: "The band is there to help us to separate patients according to the conditions. The red is a critical condition; yellow is for someone that has a bad condition but not very critical; and the green is for normal patients" [P29]. Bands were worn until discharge, with health workers continually monitoring patients for changes in the category and exchanging wristbands appropriately.

All participants expressed that they used MORES for acutely sick patients at appropriate times. For hospitals, this was multiple times per day, given their high volume: "We use it all of the time, when the patient arrives in the facility. Any patient that comes in the labor room or the entry point, the ER" [P15]. At RHFs with smaller catchment areas, rates of emergency presentation were less frequent: "Sometimes in a month, we can get 4-6 [emergencies]. All depends on how the patients will come" [P31].

There was a clear understanding of the purpose of MORES from all providers and facilities, as every participant (62/62, 100%) accurately defined at least 1 goal of the program:

> *[MORES] is a quick emergency guide that will inform you where to refer, because you put the patient's information on the WhatsApp platform and send it to the referral site. They will receive it and be preparing for the patient* [P56]

MORES Effectiveness

MORES was evaluated for effectiveness, meaning whether or not the program accomplished its intended purpose. These goals included improving prehospital transfer, increasing health worker attentiveness, and improving patient care and outcomes. Nearly all participants reported improved prehospital transfer (57/62, 92%). A major factor for improved transfer was that MORES established an avenue for rapid, prehospital communication between RHF and hospital health professionals, which facilitated closed-loop communication:

> *If [RHFs] communicate that they are sending a patient and we reply, 'yes, we're awaiting the patient' then they are aware that we have received the message. Then we have to give them feedback that the patient came and arrived at this time. We send information like the patient came with this kind of condition and we've started management or the patient is improving or the patient is on treatment. Until that patient can be discharged, we still have to give them feedback.* [P1]

Participants believed that this communication resulted in more timely referrals from RHFs to hospitals (53/62, 85%). MORES also allowed health professionals to collaboratively overcome referral and transport facilitation barriers and informed clinicians on patient arrival times and conditions: "Before, people were not there to receive the patient on time. But now, if you send information, they will already be standing at the entrance waiting to receive the patient" [P48]. This led to a reduction in transfer time and intrafacility, or third delays, at hospitals, which was especially important for urgent cases requiring surgical intervention: "It has helped us limit delays from patients being referred. It informs the ER and OB Ward in times of emergency and gives us prior notice about a referred patient that has limited time for surgery" [P21].

Participants reported being able to prepare operating rooms and gather needed materials for urgent intervention before patient arrival, to more quickly evaluate and treat critically ill patients upon arrival: "As a health worker, it's helpful in the sense that before my patient arrives, I've already received their information and setup. The moment they arrive, immediate intervention starts" [P6].

Second, health workers believed that MORES increased their attentiveness toward patients (39/62, 63%). One stated "It's useful because it helps to keep us focused and makes us know that there's an emergency that you need to act right away" [P9]. Another agreed that MORES trained her to recognize clinical deterioration faster: "Most often, if a patient comes, I will already know and be on guard to monitor because any time the patient's condition can change" [P7].

This attentiveness was partially attributed to improved triage capacity following the triage training sessions and wristband implementation (35/62, 56%). One hospital health worker described the case of a woman with peripartum hemorrhage who received timely care because of the triage system. Another affirmed that the triage wristbands helped focus health workers' urgency, with patients with red wristbands being seen first on clinician rounds (Table 2).

Finally, more than half of the participants (34/62, 55%) believed that MORES contributed to improving patient outcomes. They perceived MORES as helping to improve maternal and neonatal mortality, as well as patient satisfaction:

> *It has helped me to save lives that were to be lost, especially in pregnant women cases. If someone has a ruptured ectopic, before the person comes, we already know the person's condition…and the person will receive the care in time.* [P14]

Table 2. Effectiveness domain: themes, subthemes, and representative quotes.

Theme and subtheme	Representative quote
Theme 1: Improved prehospital transfer (57/62, 92%)	
Closed loop communication channels	*It provides information ahead of time to inform a bigger facility that a patient is coming with their diagnosis. It allows for information dissemination.* [P50]
Faster referral and transfer times	*It helps reduce the time if you're referring a patient. You have to give the patient's information, then the health worker, doctor, or midwife will be able to know the condition that is coming and prepare before time to avoid delays.* [P34]
Anticipate patient arrival and diagnosis	*If the RHFs have any critical patients, they will send a message before the patient reaches the hospital, and we will already know the type of patient that is coming and their condition so that we can get prepared for the patient.* [P18]
Reduced intrafacility delays through materials preparation and intervention initiation	*When we know the case coming, we will setup and get prepared. Whatsoever materials that they will need like drugs, the team will get everything set as they arrive, so we just start our care. No more going to look for MgSO4 for an eclamptic patient. You get your tray set so as soon as they enter, you start treatment.* [P1]
Theme 2: Increased health worker attentiveness (39/62, 63%)	
Formalized and improved triage system	*[MORES] improves care quality because when the patients come to the emergency room and they are already triaged, they will place the band on the patients' hands to know the kind of patients so everyone will be on the guard to monitor the patients.* [P7]
Increased attention for high acuity patients	*The moment the doctors enter the emergency room (ER), they go directly to the patients with the red band because they are patients of concern. It provides information on which patient to see first.* [P6]
Recognize patient deterioration or complications faster	*Staff that is on shift will receive the patient and label them for the next person who is coming. As soon the person enters the place and sees the patients, they will know who to care for first; who to start with or who's the patient of concern. Patients with red are our patients of concern. If green, you know it's a normal patient; someone who's in yellow will require attention because they can change to red.* [P1]
Theme 3: Improved patient care and outcomes (34/62, 55%)	
Contribution to reducing mortality	*It helps the pregnant woman to be well taken care of and to help save their lives and that of the babies.* [P29]
Patient satisfaction	*Because information is disseminated and prompt referral is given before patient arrival, it makes patients happy.* [P50]

MORES Sustainability and Scalability

Interview responses reflected an overwhelmingly positive desire for MORES to continue and expand (62/62, 100%): "My recommendation is for the program to continue because it's really helpful to us and reduces our maternal deaths" [P35]. Participants discussed the benefit of scaling MORES to other RHFs, who: "complained on why they are not part of the program," and citing that, "there are some terrible cases that come from the other facilities that are not trained" [P6]. They believed benefits could be realized in standardization of medical care, reduction in error, and ultimately improvement in maternal and neonatal mortality if the program was scaled to other counties in Liberia. Participants directly cited MORES usability, feasibility, and effectiveness as justification for its scaling potential.

Despite successes, there were ongoing challenges to be addressed prior to scale-up, with 89% (n=55) of participants naming at least 1 issue. Challenges included financial support for cellular data (39/62, 63%), poor cellular networks preventing connectivity (21/62, 34%), and staff unavailability due to high turnover and limited motivation (19/62, 31%). RHF health workers were more likely to mention poor cellular networks than hospital staff, who worked in urban and suburban settings. Three participants also explained how conflict could arise between patient care and MORES, particularly when checking messages on a busy service: "Especially at the emergency unit, sometimes we are busy and forget to check the WhatsApp phone. If you have an emergency, you will be busy with that while other clinics are sending messages." [P10].

From these challenges, participants offered suggestions for improvement. For workforce shortages, participants encouraged mentoring more staff to be MORES trainers to onboard new health workers: "If the knowledge of nurses can be increased and incorporating more nurses and increasing the training method, it will help support the program." [P21]. Others suggested increasing staff motivation through cash incentives and providing facility phones to all RHFs with mechanisms for data recharging, rather than relying on personal phone use.

MORES Health System Impact

In addition to evaluating MORES implementation based on intended aims, there were multiple effects of MORES on the wider health system that were not part of the original design. In total, 61 (98%) participants described ways that MORES implementation strengthened aspects of the health system beyond the intended aims of implementation. Participants identified 6 ways in which MORES contributed to broader health systems strengthening (Figure 3). First, participants noted increased coordination and accountability among health workers

and facilities when caring for patients (55/62, 89%). This was perceived through rapidly updated information and the ability to coordinate care between receiving hospitals which previously had limited communication:

> *I appreciate the program because care is on time these days and people are aware. When a patient is coming but C.B Dunbar is crowded and there's no space, as long we get the message, we send it to Phebe for preparations. It helps a lot.* [P1]

Coordination and accountability were also experienced at a personal level, with health workers reporting increased responsibility to their patients, including their scope of care: "[MORES] got us to be on our guard and to know which case is above our limit and needs to be transferred to a bigger facility" [P54]. Workers felt more comfortable acknowledging personal limitations and used the MORES system to consult higher-level colleagues more frequently than prior.

Second, feedback mechanisms for hospitals and RHFs were a key benefit of the 2-way WhatsApp messaging (48/62, 77%). Hospital staff offered constructive feedback to RHF workers on their care and referral process. This feedback included the accuracy of diagnoses and opportunities for quality improvement:

> *Based on the feedback, we will know that we are doing good work or understanding the referral system. Using the paper-based system, getting feedback used to be difficult, but now it comes as soon as the patient arrives* [P37]

Feedback was rapid, with 89% of RHF health workers receiving updates within the first 24 hours (n=55). Such communication also facilitated improved interprofessional teamwork (21/62, 34%). Participants reported reduced work tensions between RHF and facility providers and improved interdisciplinary collaboration through network building: "It is helpful because my community health assistants and midwives are working together. If a patient is not willing to come, they will continue encouraging the patient" [P40]. RHF health workers appreciated that they could consult dozens of other professionals when managing complex patients, establishing consultation avenues for health workers who are often the sole clinicians working in a low-resource facility, and learning from the outcomes of one another: "I use it to refer, but also see if my colleagues are using it, because we all are trying to stop the maternal death rate in Liberia. I use it to see the outcome of my colleagues' referral" [P47].

Fourth, MORES facilitated longitudinal follow-up care for RHFs (20/62, 32%), especially through patient tracking while still in the hospital. These updates helped family members unable to visit loved ones: "It makes us know more about our patients because the hospital will give [updates]. If the patient's relatives don't have a phone, we will be able to inform them on the status" [P35]. Hospitals additionally used MORES to communicate discharge instructions to RHFs for patients requiring follow-up. One RHF midwife elaborated on how MORES empowered her to help a patient continue her hospital-prescribed medication following discharge (Table 3).

Figure 3. Six themes of health systems strengthening which emerged from the implementation of MORES according to participant interviews. MORES: Mobile Obstetric Referral Emergency System; RHF: rural health facility.

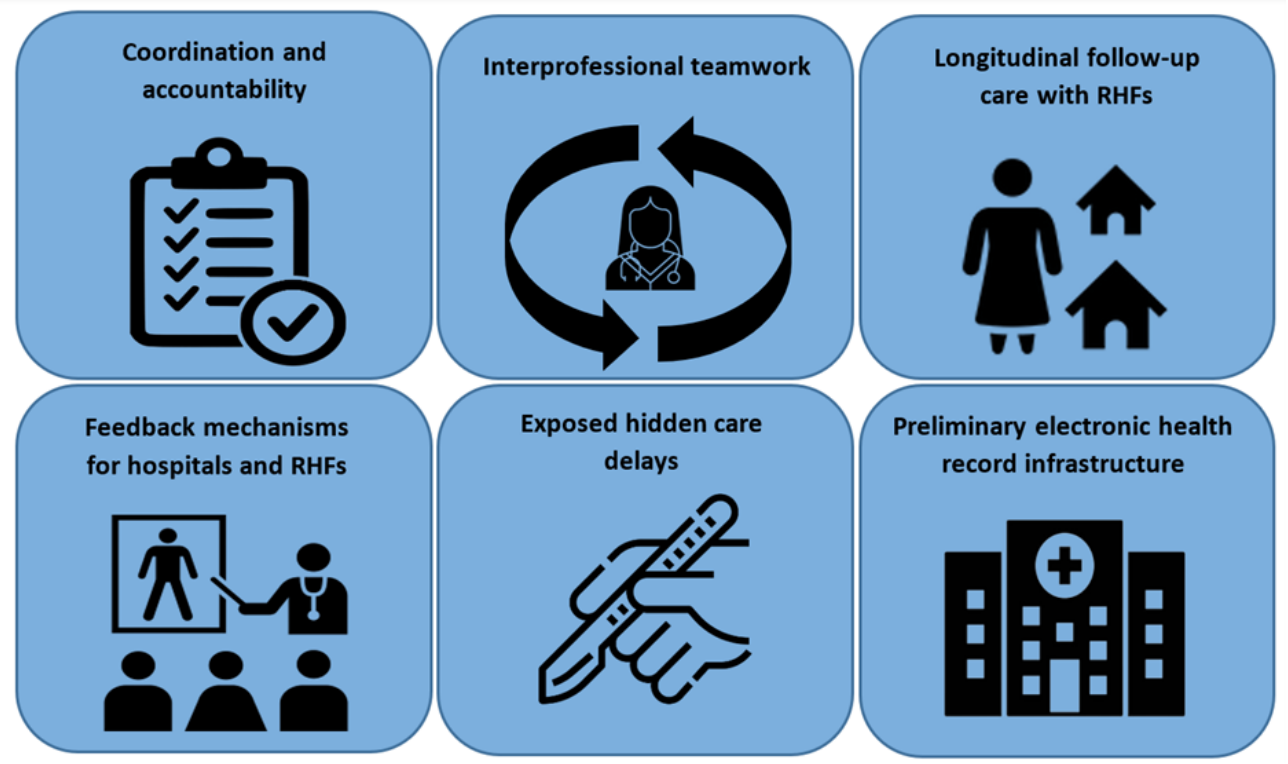


Table 3. Health system impact domain: themes, subthemes, and representative quotes.

Theme and subtheme	Representative quotes
Theme 1: Coordination and accountability (55/62, 89%)	
Care coordination between hospitals and health workers	*When the patient comes as an emergency, if we call and can't get the ambulance, as soon as we send the information the people at Phebe will inform the ambulance team to come for the patient.* [P35]
Increased individual accountability	*Before, there was conflict. Midwives would delay the patients and write referral notes to suit their taste while patients were saying something different. No one was willing to take the blame.* [P24]
Recognize scope of care	*It's very helpful because we don't have to keep cases that we know we can't manage. Once they come to the facility, if the case is in our reach, we can handle it, but if not, we refer them.* [P41]
Theme 2: Feedback mechanisms for hospitals and RHFs (48/62, 77%)	
Identify quality improvement opportunities	*It's helpful. Sometimes we get negative feedback from the patients, like if a patient gives birth and loses her baby.* [P49]
Assess diagnostic accuracy	*You are doing work and get feedback. It can help you in your diagnosis because if the people at the hospital diagnose the same thing, you can be confident of yourself* [P30]
Theme 3: Interprofessional teamwork (21/62, 34%)	
Interdisciplinary collaboration	*Using [MORES], I will be in communication with hospital nurses at CB Dunbar or Phebe to know if the patient is in good condition or if there is something that should make me concerned.* [P38]
Consultation avenues with other health workers	*The program is good; it gives information. Whatever happens in Palala Clinic is being communicated in all facilities and we are aware of all our patients.* [P50]
Network building	*It helps us do plenty things. It makes us to know lot of people and how to really deal with the phone. It also makes us to know every patient information that needs to be communicated.* [P42]
Theme 4: Longitudinal follow-up care with RHFs (20/62, 32%)	
Patient status updates in the hospital	*It makes us help our patients get quick health services at the higher facilities and for us to get information on how they are doing.* [P35]
Communication with family members	*The feedback is very helpful because it confirms and updates on the conditions of our patients that we sent. We can use the feedback to tell the patient's people.* [P46]
Longitudinal care after discharge	*The feedback helps us continue the treatment or medication of a patient that has been referred [once discharged].* [P58]
Theme 5: Exposed hidden care delays (17/62, 27%)	
Resource insufficiencies	*[We realized we] need emergency materials to make work easier. All OB cases are actually emergency cases and should be treated.* [P22]
Health worker capacity	*One difficulty we experience is having misdiagnosis from the clinics. Communication sometimes isn't clear.* [P17]
Theme 6: Preliminary electronic health record infrastructure (13/62, 21%)	
Permanent patient medical record	*[MORES is] helping a lot especially with data [capture]. Patient referral and treatment between facilities is being communicated.* [P54]
Faster information dissemination	*We now receive electronic communication. It makes us more confident in our work and we feel satisfied after getting our findings.* [P50]
Liability protection for workers	*It's very helpful to me because it will save me. If I use WhatsApp to refer my patient, the time I refer will be mentioned. If there's any delay, they will know where it comes from. So, it will protect me.* [P55]

Fifth, MORES created a record of hidden care delays in Bong County by documenting reasons for delays as they arose (17/62, 27%). The most common included resource insufficiencies and worker capacity: "The purpose of [MORES] is to do quick referral and…to know where the delay comes from; whether from the community, the relatives or the midwife or from the county because they are responsible to send the ambulance" [P55]. In such situations, even when MORES worked properly, patients still experienced delays. More than half of respondents listed factors outside of MORES that worsened care delays for patients (35/62, 56%), including a lack of medications, intravenous fluids, and ambulance insufficiency.

Finally, MORES benefited the health system by creating a preliminary electronic health record (EHR) infrastructure (13/62, 21%). No EHR exists in Bong, and all documentation occurs on paper charts. Participants appreciated the ease, accessibility, and speed that MORES provided, believing it sped information dissemination:

> *[MORES] enables me to do my work with ease. The time I take to write a whole lot of papers…wastes the patient's time, and it will also waste time at the referral hospital…With [MORES], the hospital will already get the information before the patient can arrive so they will start to attend immediately.* [P47]

This EHR also started a permanent record of patient conditions, transfers, and treatments administered, so that health teams could review patients requiring special accommodations or difficult cases. Finally, health workers appreciated the protections that having a permanent record of electronic communication offered in terms of liability for malpractice.

Discussion

Principal Findings

From this qualitative study of frontline health workers in rural Liberia, we found that implementation of the MORES intervention was perceived as having high usability, fidelity, and effectiveness for its intended aims. Furthermore, MORES had the unanticipated impact of strengthening multiple components of the health system and demonstrated exciting potential for scaling and sustainability. These findings confirmed the results of our preintervention study which showed high desirability and acceptability of the WhatsApp-based triage, transfer, and referral system among rural Liberian health workers to better care for obstetric patients [14]. Our analysis further revealed health worker perceptions regarding the benefits and health system improvement opportunities through MORES. These in-depth results contribute to the findings from other studies on the impact of MORES, including increased rates of cesarean section and lower odds of nonvigorous symptoms in neonates [15].

Users perceived MORES to be highly usable, as it could be seamlessly integrated into existing workflows and referral pathways. MORES use also demonstrated high fidelity, as it was implemented as intended and at appropriate times to refer critical patients. MORES participants reported the program to be effective across its intended aims: improving prehospital transfer, increasing health worker attentiveness, and improving patient care. The triage training and templated, bidirectional WhatsApp communication components of MORES worked collaboratively to address 2 key gaps in rural care delivery in Liberia: a lack of standardized obstetric triage protocols and a limited prehospital communication system. Increasing capacity and infrastructure to support triage protocols and prehospital emergency systems have been shown to significantly reduce patient morbidity and mortality, particularly in LMICs [19]. This opportunity for impact is especially prevalent for maternal mortality, where the third delay accounts for a massive burden of maternal and neonatal deaths, in some cases as high as 80% [20]. Participants were enthusiastic about the impact of MORES and desired its sustainability and scaling to other health facilities throughout Liberia. Continuing to implement and evaluate programs that increase capacity in obstetric triage, prehospital communication, and facility referral and transfer will be important for improving care and understanding the effects of mHealth technologies in resource-limited environments.

Though not an intended aim of implementing MORES, we found that the program facilitated health systems strengthening across 6 themes: coordination and accountability, feedback mechanisms for hospitals and RHFs, interprofessional teamwork, longitudinal follow-up care with RHFs, a record of hidden care delays, and establishment of a preliminary EHR infrastructure. While not the primary goals of MORES, these effects highlight important benefits across many internal and external domains of health system strengthening. For internal effects, participants reported MORES improved personal accountability, quality improvement opportunities, network building, and reduced workplace tensions and stress from liability. Externally, participants perceived faster information dissemination and facility preparation, consultation mechanisms, and care continuity. These findings highlight how MORES contributed to capacity building and the beneficial effects that can be realized when interventions are implemented to improve systems beyond individual disease states [21]. For example, while MORES was intended for obstetric and peripartum referrals, it exhibited the potential to be adapted to other specialties such as emergency medicine and surgery [22]. MORES provides a pragmatic example of how horizontally implementing interventions across multiple facilities, worker levels, and patient conditions can have multiplicative effects on health systems improvement.

This study adds to the growing literature on mHealth implementation in LMICs to reduce delays and improve care access. One comprehensive review of mHealth interventions for maternal health in LMICs showed that health worker buy-in and motivation, perceived intervention usability, and efficient messaging were key implementation-related mechanisms for success [23]. Our findings on MORES account for all 3, showing high usability and desires for sustainability and scaling, with efficient communication mechanisms that were used appropriately, bidirectionally, and rapidly. A study in South Africa demonstrated high usability for a mobile health platform to aid health workers in the management of pregnant women at risk for preeclampsia [24]. From this program, most user issues resulted from phone features including scroll wheels and touch screens. In MORES, we leveraged the free, WhatsApp communication platform already familiar to Liberian health workers, thereby mitigating user errors from unfamiliarity. Similar to our findings on the impact on interprofessional dynamics, other studies have demonstrated the collaborative effects of a WhatsApp messaging platform among obstetric providers, including as avenues for consultation, network building, and interprofessional communication [25]. Recently in Ghana, a WhatsApp messaging technology was implemented among 13 facilities and 81 health workers for more than 600 patient referrals from rural to district hospitals [26]. Findings from this study mirrored ours, including exposing hidden care delays and creating a preliminary EHR. Specifically, the program identified delay factors, referral characteristics associated with faster transport, and timely updates on patient arrival and outcomes more quickly and permanently compared with paper charting. A mHealth intervention among traditional birth attendants for indigenous women in rural Guatemala demonstrated increased referrals to facility-level care but similar complication rates across intervention and control groups [27]. Such inconclusive outcomes data suggest opportunities for future evaluation of maternal mHealth interventions. Results of a meta-analysis from 2016 revealed that out of 15 articles, only 2 studies showed a low risk of bias and 1 demonstrated a mortality benefit when text messaging women during pregnancy [28]. Future quantitative findings regarding the impact of

MORES on maternal outcomes and referral times could clarify the potential clinical impact of well-implemented and highly acceptable maternal mHealth in LMICs.

Limitations

This study had several limitations. First, our qualitative methods and geographic homogeneity limit the generalizability of results. However, this approach was considered most effective given the study aims which allowed for in-depth explanations of MORES usability, fidelity, and effectiveness. Second, our sample had an unevenly distributed representation of health worker roles, with most participants being midwives and nurses. Increasing recruitment of physicians and other personnel who interact with MORES such as ambulance drivers, administrative staff, and patients being referred within the system could capture more in-depth data beyond what was offered by frontline health workers. Finally, we elected not to utilize a preexisting framework for the conceptualization of our results, given that commonly referenced usability, acceptability, and feasibility metrics for mHealth did not align with the evaluative approach and subsequent immersion in the data. This was apparent, especially for the impact of MORES on health system strengthening. To increase study rigor, we used a validated codebook and 4 separate metrics to ensure accuracy including member checking. However, developing flexible frameworks to better conceptualize usability, effectiveness, and broader benefits of mHealth technologies in LMICs could be highly valuable for future studies.

Conclusions

The MORES intervention was perceived to have high usability, fidelity, and effectiveness for reducing delays and optimizing referrals in rural Liberia. MORES also demonstrated 6 components of health systems strengthening which positively impacted accountability among health workers, feedback mechanisms, interprofessional dynamics, longitudinal follow-up care, and system-level efficiencies through recording delay factors and electronic health record infrastructure. Findings from this study suggest that MORES demonstrates high feasibility for scaling to other facilities and counties in Liberia to improve maternal and neonatal referral pathways. These conclusions can have direct policy and program implications for supporting MORES funding, implementation, and scaling at a national level. More widely, these findings can be highly useful to policymakers, clinicians, and researchers intending to implement and scale technology-based obstetric referral systems in low-resource contexts to accomplish intended aims, anticipate challenges prior to implementation, and enhance impacts on health system strengthening.

Acknowledgments

The authors would like to thank T-Max Zorh, BSc, from Cuttington University and Lovetta B Dixon, BSN, from the University of Liberia for their contributions to data collection and transcription in this study. We would also like to thank the many rural health facilities and hospital workers who have used and continued the MORES program in Bong County. This research received funding support from the Bill and Melinda Gates Foundation (INV-023274) and the USAID PEER Grant (project 9–518). The funders, however, did not contribute to the study design, data collection and analysis, manuscript preparation, or the decision to publish.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Structured interview script for the usability study among front-line obstetric health workers using Mobile Obstetric Referral Emergency System in rural Liberia.
[DOCX File , 15 KB - mhealth_v12i1e58624_app1.docx]

References

1. Lawrence ER, Klein TJ, Beyuo TK. Maternal mortality in low and middle-income countries. Obstet Gynecol Clin North Am 2022;49(4):713-733. [doi: 10.1016/j.ogc.2022.07.001] [Medline: 36328676]
2. WHO, UNICEF, UNFPA, World Bank Group, and UNDESA/Population Division. Trends in maternal mortality 2000 to 2020. Geneva, World Health Organization, 2023. URL: https://data.worldbank.org/indicator/SH.STA.MMRT?locations=LR [accessed 2024-01-31]
3. Nyenswah T. Reflections on leadership and governance from the incident manager of Liberia's ebola response. Health Secur 2017;15(4):445-449. [doi: 10.1089/hs.2017.0040] [Medline: 28806109]
4. Lee H, Perosky J, Horton M, Reynolds C, Nyanplu A, Lori JR. Verbal autopsy analysis of maternal mortality in bong county, Liberia: a retrospective mixed methods study. BMJ Open Qual 2023;12(2):e002147 [FREE Full text] [doi: 10.1136/bmjoq-2022-002147] [Medline: 37076249]
5. Avoka CK, McArthur E, Banke-Thomas A. Interventions to improve obstetric emergency referral decision making, communication and feedback between health facilities in sub-Saharan africa: a systematic review. Trop Med Int Health 2022;27(5):494-509. [doi: 10.1111/tmi.13747] [Medline: 35322914]

6. Naz S, Saleem S, Shamsul Islam Z, Bhamani S, Sheikh L. Obstetric triage improvement process using the donabedian model of quality care: a quality improvement initiative. BMJ Open Qual 2022;11(2):e001483. [doi: 10.1136/bmjoq-2021-001483] [Medline: 35577399]
7. Lori JR, Starke AE. A critical analysis of maternal morbidity and mortality in Liberia, West Africa. Midwifery 2012;28(1):67-72. [doi: 10.1016/j.midw.2010.12.001] [Medline: 21232836]
8. United NPF. Liberia launches ambitious five-year plan to reduce maternal and newborn deaths. United Nations Population Fund. 2011. URL: https://www.unfpa.org/news/liberia-launches-ambitious-five-year-plan-reduce-maternal-and-newborn-deaths [accessed 2024-01-29]
9. Kentoffio K, Kraemer JD, Griffiths T, Kenny A, Panjabi R, Sechler GA, et al. Charting health system reconstruction in post-war Liberia: a comparison of rural vs. remote healthcare utilization. BMC Health Serv Res 2016;16(1):478 [FREE Full text] [doi: 10.1186/s12913-016-1709-7] [Medline: 27604708]
10. Lori JR, Munro ML, Rominski S, Williams G, Dahn BT, Boyd CJ, et al. Maternity waiting homes and traditional midwives in rural Liberia. Int J Gynaecol Obstet 2013;123(2):114-118 [FREE Full text] [doi: 10.1016/j.ijgo.2013.05.024] [Medline: 23992657]
11. Bossman E, Johansen MA, Zanaboni P. mHealth interventions to reduce maternal and child mortality in sub-saharan africa and southern asia: a systematic literature review. Front Glob Womens Health 2022;3:942146 [FREE Full text] [doi: 10.3389/fgwh.2022.942146] [Medline: 36090599]
12. Pagalday-Olivares P, Sjöqvist BA, Adjordor-van de Beek J, Abudey S, Silberberg AR, Buendia R. Exploring the feasibility of eHealth solutions to decrease delays in maternal healthcare in remote communities of ghana. BMC Med Inform Decis Mak 2017;17(1):156 [FREE Full text] [doi: 10.1186/s12911-017-0552-z] [Medline: 29197391]
13. Goodman DM, Srofenyoh EK, Ramaswamy R, Bryce F, Floyd L, Olufolabi A, et al. Addressing the third delay: implementing a novel obstetric triage system in ghana. BMJ Glob Health 2018;3(2):e000623 [FREE Full text] [doi: 10.1136/bmjgh-2017-000623] [Medline: 29707245]
14. Reynolds CW, Horton M, Lee H, Harmon W, Sieka J, Lockhart N, et al. Acceptability of a Whatsapp triage, referral, and transfer system for obstetric patients in rural Liberia. Ann Glob Health 2023;89(1):34 [FREE Full text] [doi: 10.5334/aogh.4030] [Medline: 37273491]
15. Lee H, Dahn B, Sieka J, Nyanplu A, Reynolds CW, Edson C, et al. The use of a mobile obstetric emergency system to improve obstetric referrals in bong county, Liberia: a pre-post study. Int J Gynaecol Obstet 2024;165(1):229-236. [doi: 10.1002/ijgo.15175] [Medline: 37789637]
16. MacQueen KM, McLellan E, Kay K, Milstein B. Codebook development for team-based qualitative analysis. CAM J 1998;10(2):31-36. [doi: 10.1177/1525822x980100020301]
17. Neubauer BE, Witkop CT, Varpio L. How phenomenology can help us learn from the experiences of others. Perspect Med Educ 2019;8(2):90-97 [FREE Full text] [doi: 10.1007/s40037-019-0509-2] [Medline: 30953335]
18. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
19. Bhattarai HK, Bhusal S, Barone-Adesi F, Hubloue I. Prehospital emergency care in low- and middle-income countries: a systematic review. Prehosp Disaster Med 2023;38(4):495-512. [doi: 10.1017/S1049023X23006088] [Medline: 37492946]
20. Knight HE, Self A, Kennedy SH. Why are women dying when they reach hospital on time? a systematic review of the 'third delay'. PLoS One 2013;8(5):e63846. [doi: 10.1371/journal.pone.0063846] [Medline: 23704943]
21. Béhague DP, Storeng KT. Collapsing the vertical-horizontal divide: an ethnographic study of evidence-based policymaking in maternal health. Am J Public Health 2008;98(4):644-649. [doi: 10.2105/AJPH.2007.123117] [Medline: 18309123]
22. Mehmood A, Rowther AA, Kobusingye O, Hyder AA. Assessment of pre-hospital emergency medical services in low-income settings using a health systems approach. Int J Emerg Med 2018;11(1):53 [FREE Full text] [doi: 10.1186/s12245-018-0207-6] [Medline: 31179939]
23. Kabongo EM, Mukumbang FC, Delobelle P, Nicol E. Explaining the impact of mHealth on maternal and child health care in low- and middle-income countries: a realist synthesis. BMC Pregnancy Childbirth 2021;21(1):196 [FREE Full text] [doi: 10.1186/s12884-021-03684-x] [Medline: 33750340]
24. Lim J, Cloete G, Dunsmuir DT, Payne BA, Scheffer C, von Dadelszen P, et al. Usability and feasibility of PIERS on the move: an mHealth app for pre-eclampsia triage. JMIR Mhealth Uhealth 2015;3(2):e37 [FREE Full text] [doi: 10.2196/mhealth.3942] [Medline: 25887292]
25. Carmona S, Alayed N, Al-Ibrahim A, D'Souza R. Realizing the potential of real-time clinical collaboration in maternal-fetal and obstetric medicine through WhatsApp. Obstet Med 2018;11(2):83-89 [FREE Full text] [doi: 10.1177/1753495X18754457] [Medline: 29997691]
26. Owen MD, Ismail HM, Goodman D, Batakji M, Kim SM, Olufolabi A, et al. Use of WhatsApp messaging technology to strengthen obstetric referrals in the Greater Accra Region, Ghana: findings from a feasibility study. PLoS One 2022;17(4):e0266932 [FREE Full text] [doi: 10.1371/journal.pone.0266932]
27. Martinez B, Ixen EC, Hall-Clifford R, Juarez M, Miller AC, Francis A, et al. mHealth intervention to improve the continuum of maternal and perinatal care in rural Guatemala: a pragmatic, randomized controlled feasibility trial. Reprod Health 2018;15(1):120 [FREE Full text] [doi: 10.1186/s12978-018-0554-z] [Medline: 29973229]

28. Lee SH, Nurmatov UB, Nwaru BI, Mukherjee M, Grant L, Pagliari C. Effectiveness of mHealth interventions for maternal, newborn and child health in low- and middle-income countries: systematic review and meta-analysis. J Glob Health 2016;6(1):010401 [FREE Full text] [doi: 10.7189/jogh.06.010401] [Medline: 26649177]

Abbreviations

COREQ: consolidated criteria for reporting qualitative research
EHR: electronic health record
LMIC: low- and middle-income country
mHealth: mobile health
MORES: Mobile Obstetric Referral Emergency System
RA: research assistant
RHF: rural health facility

Edited by L Buis; submitted 20.03.24; peer-reviewed by R Sewpaul, TJRT Moncatar; comments to author 13.05.24; revised version received 08.07.24; accepted 26.08.24; published 13.11.24.

Please cite as:
Reynolds CW, Lee H, Sieka J, Perosky J, Lori JR
Implementation of a Technology-Based Mobile Obstetric Referral Emergency System (MORES): Qualitative Assessment of Health Workers in Rural Liberia
JMIR Mhealth Uhealth 2024;12:e58624
URL: https://mhealth.jmir.org/2024/1/e58624
doi:10.2196/58624
PMID:

Hearing Rehabilitation With a Chat-Based Mobile Auditory Training Program in Experienced Hearing Aid Users: Prospective Randomized Controlled Study

Jae Sang Han[1], MD, PhD; Ji Hyung Lim[1], MD; Yeonji Kim[1], MD; Aynur Aliyeva[1,2], MD; Jae-Hyun Seo[1], MD, PhD; Jaehyuk Lee[3], BSc; Shi Nae Park[1], MD, PhD

[1]Department of Otorhinolaryngology–Head and Neck Surgery, Seoul St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea

[2]Department of Pediatric Otolaryngology, Cincinnati Children's Hospital, Cincinnati, OH, United States

[3]Nara Information Co, Ltd, Seoul, Republic of Korea

Corresponding Author:
Shi Nae Park, MD, PhD
Department of Otorhinolaryngology–Head and Neck Surgery, Seoul St. Mary's Hospital, College of Medicine, The Catholic University of Korea, , Seoul, , Republic of Korea

Abstract

Background: Hearing rehabilitation with auditory training (AT) is necessary to improve speech perception ability in patients with hearing loss. However, face-to-face AT has not been widely implemented due to its high cost and personnel requirements. Therefore, there is a need for the development of a patient-friendly, mobile-based AT program.

Objective: In this study, we evaluated the effectiveness of hearing rehabilitation with our chat-based mobile AT (CMAT) program for speech perception performance among experienced hearing aid (HA) users.

Methods: A total of 42 adult patients with hearing loss who had worn bilateral HAs for more than 3 months were enrolled and randomly allocated to the AT or control group. In the AT group, CMAT was performed for 30 minutes a day for 2 months, while no intervention was provided in the control group. During the study, 2 patients from the AT group and 1 patient from the control group dropped out. At 0-, 1- and 2-month visits, results of hearing tests and speech perception tests, compliance, and questionnaires were prospectively collected and compared in the 2 groups.

Results: The AT group (n=19) showed better improvement in word and sentence perception tests compared to the control group (n=20; *P*=.04 and *P*=.03, respectively), while no significant difference was observed in phoneme and consonant perception tests (both *P*>.05). All participants were able to use CMAT without any difficulties, and 85% (17/20) of the AT group completed required training sessions. There were no changes in time or completion rate between the first and the second month of AT. No significant difference was observed between the 2 groups in questionnaire surveys.

Conclusions: After using the CMAT program, word and sentence perception performance was significantly improved in experienced HA users. In addition, CMAT showed high compliance and adherence over the 2-month study period. Further investigations are needed to validate long-term efficacy in a larger population.

Trial Registration: Clinical Research Information Service (CRiS) KCT0006509; https://cris.nih.go.kr/cris/search/detailSearch.do?seq=22110&search_page=L

(JMIR Mhealth Uhealth 2024;12:e50292) doi:10.2196/50292

KEYWORDS

hearing loss; hearing aids; hearing rehabilitation; auditory training; mobile program

Introduction

According to a World Health Organization report in 2021, it is predicted that a quarter of people will have some degree of hearing loss by 2050 [1]. Hearing loss is known to be related to the risk of vocational problems, depressed mood, and even cognitive impairment in adults [2-4]. Adult aural or hearing rehabilitation, a multifactorial strategy for patients with acquired hearing impairment, includes 4 main components: sensory management, instruction, perceptual training, and counseling [5]. Sensory management involves enhancing auditory function through the use of hearing aids (HAs), cochlear implants, and alternative devices. Additional perceptual training, including auditory training (AT), seems to be required because sensory management alone has limitations in improving speech perceptual ability [5,6].

Although AT is known to improve hearing performance and lower the HA return rate, it has been not sufficiently provided to patients with hearing loss due to its poor time- and cost-effectiveness [7-9]. In our previous randomized controlled trial, we reported significant improvements in consonant perception and subjective satisfaction after 8 weeks of AT consisting of in-hospital face-to-face education and at-home self-training among HA users [10]. However, we experienced the limitations of in-hospital training programs during the COVID-19 pandemic, and this led us to consider a mobile-based AT program.

Smartphone penetration is rapidly increasing worldwide, and South Korea is a leading county for smartphone use, with a penetration rate of around 93% in 2021 [11,12]. Among the older population, around 94% of those older than 60 years and 60% of those older than 70 years own smartphones, and most of them (80%) use the internet with a mobile device [12,13]. The increasing smartphone penetration rate is a global trend, and mobile-based health care programs are expected to be used effectively even in the older population [14]. We developed a new mobile AT program with a chat-based interface to increase compliance with hearing rehabilitation, with attention to the fact that most of the adult population is familiar with message apps on smartphones [15].

Several mobile and web-based AT programs have been introduced in the last decade. However, many mobile-based AT applications are simple gamified training programs for pediatric patients, and most of them have no clinical validation to support their efficacy [16]. Even though a state-of-the-art review showed AT to be effective for improving auditory perceptions [17], it is unclear whether mobile-based training will still be effective. Digital therapeutics are expected to change the future of the health care system; however, clinical validation should first be performed to prove their effectiveness [18]. In addition, the effect of AT on experienced HA users has been inconsistently reported, while the effect of AT on novel HA users is relatively clear. Of the 16 studies included in a state-of-the-art review on AT, only 3 investigated experienced HA users, and only 1 of these studies showed improvements in speech perception [19-21].

Therefore, this prospective randomized controlled study was conducted to investigate the efficacy of our novel chat-based mobile AT (CMAT) program for 3 aspects of hearing rehabilitation in experienced HA users: speech perception ability, subjective satisfaction, and training compliance.

Methods

Ethical Considerations

This prospective study was approved by the ethical committee of Seoul St. Mary's Hospital (KC21EISI0525) and followed the tenets of the Declaration of Helsinki. The patient records and information were anonymized and deidentified before analysis. All participants provided written informed consent prior to commencement of the study and voluntarily participated in this clinical trial.

Participants

A prospective randomized controlled trial with an unblinded study setting was carried out. Bilateral HA users were recruited from the department of otorhinolaryngology–head and neck surgery at a tertiary referral center between September and December 2021. Study eligibility criteria were as follows: (1) age ≥20 years; (2) bilateral moderate to severe sensorineural hearing loss (mean threshold of pure tone audiometry measured at 500, 1000, 2000, and 4000 Hz was 41 to 80 dB hearing level, and the air-bone gap was less than 15 dB hearing loss); and (3) the patient had been using bilateral HAs for more than 3 months, demonstrated sufficient functional gain, and consistently used the HAs for more than 8 hours per day. Patients were excluded if they had fluctuation in hearing loss, brain tumor, or difficulty using the program or coordinating hearing tests.

The randomization was performed using premade random cards with a 1:1 allocation ratio prepared by a contract research organization (Medical Excellence). The cards were then opened by the health care provider after obtaining informed consent. Participants allocated to the AT group (ATG) were provided instructions about the CMAT program by a coauthor (JL) and were encouraged to use it daily for 8 weeks. One CMAT session consisted of 20 questions, took about 5 to 10 minutes, and was repeated 3 times a day. Those assigned to the control group (CG) did not receive the CMAT. All participants were asked to complete audiologic tests, speech perception tests, and questionnaires before enrollment and at 1 and 2 months after the start of the study, respectively.

Assuming that the effect of offline AT and CMAT would be similar, we determined the number of target participants in the same way as in our previous study [10]. For statistical significance at a .05 confidence level with 80% power, the sample size required for the 2 groups was estimated as 18 patients per group. Allowing for a 10% dropout rate, 42 patients were estimated to be required in total.

Chat-Based AT Program

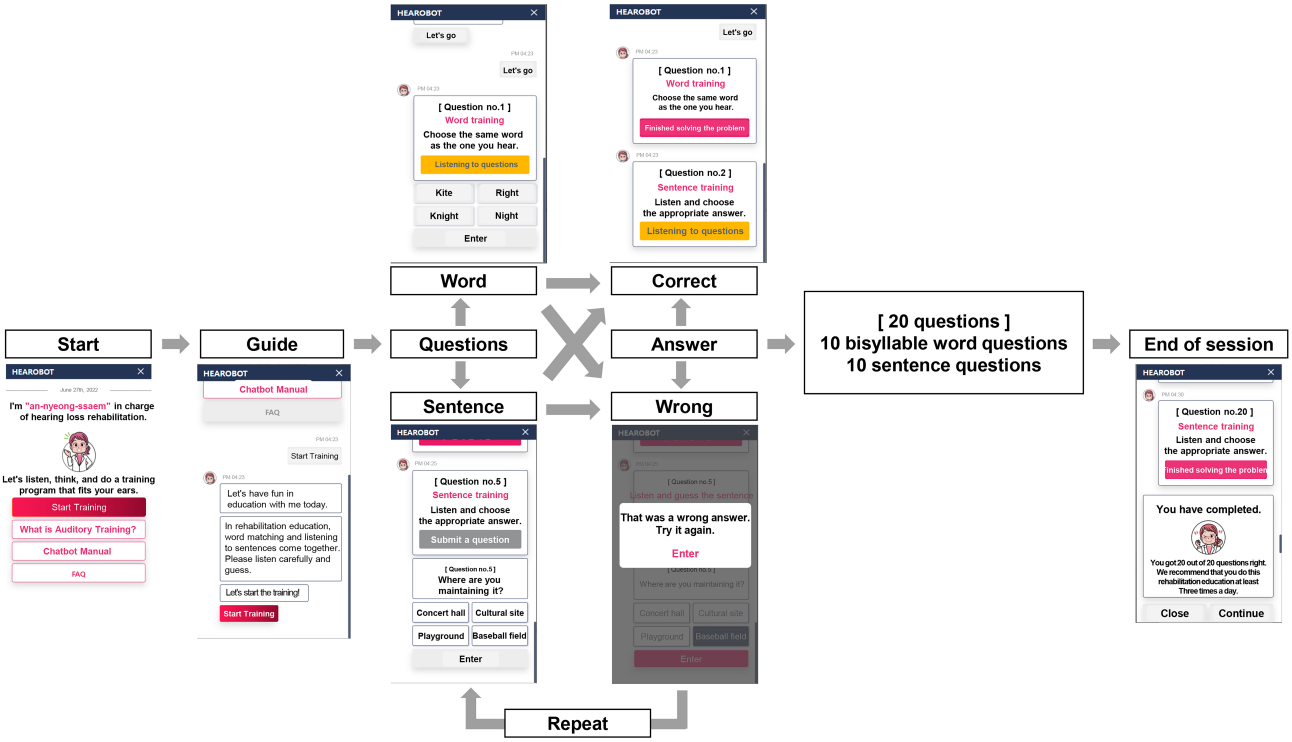
Our CMAT program was developed in the form of a web-based program that can be accessed from computers, tablets, and smartphones. However, all participants were requested to connect with their smartphones to avoid device bias in this clinical trial. When a participant accessed a provided link, the ID and password given during participant registration were required to be entered, and a page was provided to additionally confirm the consent form for clinical trials and the collection of personal information. Afterward, a problem was presented in a chat-based interface. In the chat-based interface, we designed a conversational partner as a character representing a medical professional, intending to give users the feeling of exchanging chat bubbles with a health care provider. Participants could choose 1 of 4 answers, and a message was shown giving feedback after a problem was answered. All participants were instructed to train in a quiet place while wearing their HAs.

The rehabilitation program consisted of 2 parts: word and sentence training. For word training, a total of 1540 two-syllable Korean words that are frequently used in daily life were extracted based on data provided by the National Institute of

the Korean Language. Four words with the same or similar vowels but different consonants were provided, and then one of these words was provided twice. Then, the participant was asked to choose the word that he or she heard. For sentence training, daily phrases consisting of 2- to 12-word phrases were provided, and the frequency of the words used was checked to correct or delete infrequently used words (eg, “We decided to meet at Seoul Station”). A total of 1225 sentences were composed, and a question was made for each sentence. The sentences and related questions were provided once, and 3 words similar to the correct answer were shown with it (eg, “Question: Where would you like to meet?”; “Answers: (1) Sadang Station, (2) Sung-soo Station, (3) Sung-book Station, (4) Seoul Station”).

This CMAT program was presented as if chatting with a virtual character. If a participant answered the question correctly, the next question among the word and sentence training question pools was shown randomly. If a participant chose a wrong answer, the question was repeated up to 3 times. Ten-word and 10-sentence questions were randomly assigned to participants for solving within a single session. AT dosage was planned to be around 30 minutes of training per day, similar to our previous study, as solving 20 questions was expected to take approximately 10 minutes. However, if a participant wanted to solve more questions, they could engage in an additional session. Participants included in the ATG were instructed to repeat this session 3 times a day, spaced out, following a similar approach to our previous study [10]. A schema of the CMAT is shown in Figure 1.

Figure 1. Schematic diagram of chatbot-delivered auditory training program.



The Korean language was used in this program, and words and sentences were synthesized using Google Cloud text-to-speech software. To facilitate scalability for increases in the number of sentences or upgrades in the future, we determined that software-generated voices would be a reasonable choice. The speech was randomly generated using male or female voices provided by the software, and the default settings of the program were used for other suprasegmental characteristics of sentences, including speed, pitch, and volume. The log-in time and the number of questions solved were automatically collected by the program; however, prompts to use the CMAT were not provided.

Outcome Measures

Study Design

All participants completed audiologic tests, speech perception tests, and questionnaires at the 0-, 4-, and 8-week visits, and the results were collected using Excel (Microsoft). Among these measured outcomes, speech perception test results were used as the primary end point, and questionnaire results were used as the secondary end point.

Audiologic Tests

Pure tone audiograms (PTAs) without HAs and sound field threshold audiometry tests with HAs were conducted to exclude changes in hearing level and inappropriate function of HAs. Speech audiometry tests, including the speech reception threshold test and speech discrimination scores, were evaluated both with and without the participants wearing HAs.

Speech Perception Tests

A total of 5 speech perception tests were conducted by a female audiologist in quiet conditions while the patient wore an HA. The Ling Six Sound test was conducted for phoneme detection [22], and the Vowel and Consonant Imitation Test (VCIT) was used to measure phoneme perception [23]. To evaluate word perception ability, monosyllable and bisyllable open-set tests were used [24]. Sentence perception ability was evaluated using

the Korean version of Central Institute for the Deaf (K-CID) test [25].

Questionnaires

Subjective benefits were measured by 3 validated questionnaires. The Korean version of the Hearing Handicap Inventory for the Elderly (K-HHIE) was used to measure situational and emotional handicaps in everyday life due to hearing problems [26]. In addition, the Korean versions of the International Outcome Inventory for Hearing Aids (K-IOI-HA) and Abbreviated Profile of Hearing Aid Benefit (APHAB) tests were conducted to evaluate the subjective satisfaction with HA use [27,28].

Statistical Analysis

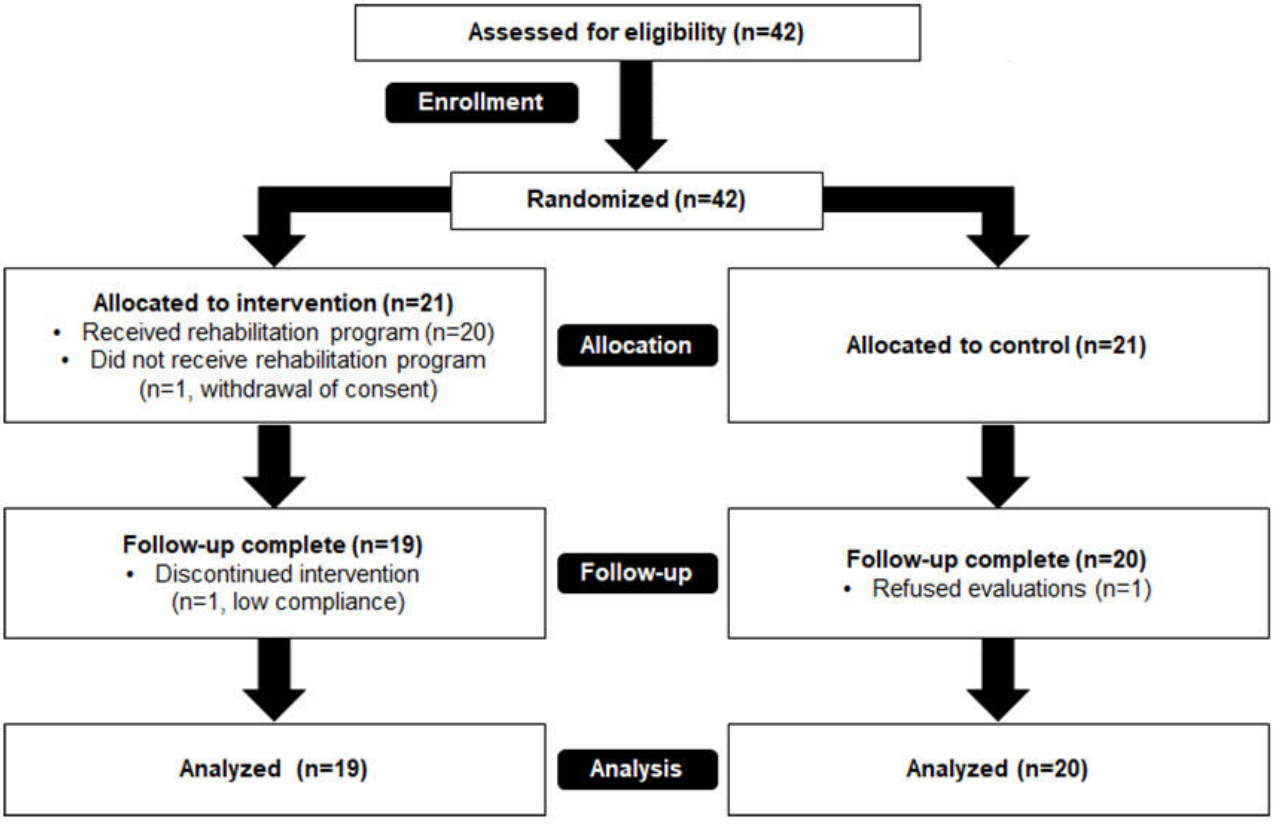
Statistical analysis was conducted using SAS (version 9.4; SAS Institute). The Shapiro-Wilk test was used to examine the normality of the measured variables. Data were expressed as mean, SD, and percentage. *P* values were calculated using the chi-square test or Fisher exact test for categorical variables, the Mann-Whitney test or independent 2-tailed *t* test for pairs of independent variables, and the Wilcoxon rank sum test or paired *t* test for continuous variables. All speech perception tests were collected as percentage data, and as the data did not meet the normality assumption based on the Shapiro-Wilk test, we conducted additional analyses using arcsine transformation. Speech perception tests and questionnaire surveys were analyzed with a linear mixed-effect model and post hoc test. Random effects included the intercept, while fixed effects included group (ATG and CG), time (initial, 1 month, and 2 month), group × time interaction, and the covariate "aided 4 kHz threshold result of the right ear." The inclusion of the aided 4 kHz threshold result of the right ear as a covariate was due to the significant differences between the 2 groups in the initial hearing test despite randomization. Further details regarding this will be provided in the *Results* section. Two covariance structures in the linear mixed model, variance components and compound symmetry, were compared. There were no substantial differences in the results of the fixed effect type 3 test, so we described the variance components estimation. The Akaike information criterion values representing model fit varied depending on the specified model, from 280 to 680. Bonferroni correction was used for multiple comparisons and adjusted *P* values are presented. Correlations between pairs of variables were analyzed with the Pearson correlation test. Differences were considered significant when the *P* value was less than .05.

Results

Clinical Characteristics

A total of 42 participants were initially enrolled after informed consent was obtained. The ATG and CG were each randomly allocated 21 patients, with 1 participant subsequently withdrawn from the ATG due to withdrawal of consent. During the 2-month study follow-up period, 1 participant in the ATG was excluded from the analysis due to low compliance (CMAT completion rate <50%), and 1 participant in the CG was excluded for refusing to perform speech perception tests. Therefore, 19 participants in the ATG and 20 participants in the CG were analyzed. The CONSORT (Consolidated Standards of Reporting Trials) flow diagram is shown in Figure 2. None of the participants in the ATG had any problems using CMAT during the study.

Figure 2. CONSORT (Consolidated Standards of Reporting Trials) flow diagram.



There were no significant differences between the 2 groups in age, sex, duration of hearing loss, type of HA being used, or duration of HA use. The detailed patient demographics are shown in Table 1.

Table . Clinical characteristics of the participants.

		Total (n=41)	Auditory training group (n=20)	Control group (n=21)	*P* value
Age (years), mean (SD)		70.0 (10.9)	70.5 (6.9)	69.5 (13.8)	.41[a]
Sex, n					.90[b]
	Male	16	8	8	
	Female	25	12	13	
Hearing aid type, n					.75[c]
	Complete in the canal	67	33	34	
	Invisible hearing aid	10	4	6	
	Receiver in the canal	5	3	2	
Hearing loss duration (months), mean (SD)		103.0 (86.5)	107.9 (83.3)	98.3 (91.3)	.48[a]
Hearing aid use duration (months), mean (SD)		54.5 (62.1)	56.1 (72.2)	53.0 (50.7)	.80[a]

[a]Mann-Whitney test.

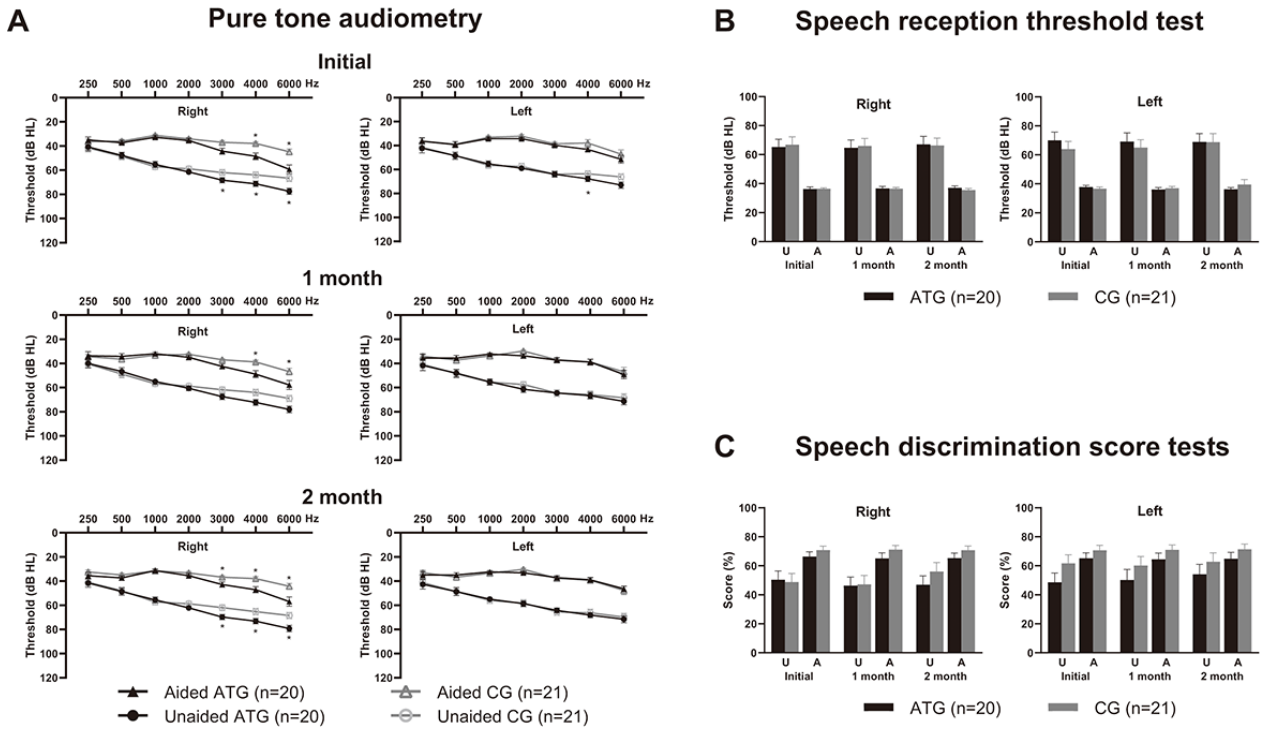
[b]Fisher exact test.

[c]Chi-square test.

Audiologic Evaluations

Although we randomly allocated participants to the 2 groups, there were unintended differences in the baseline hearing tests. In the ATG, the high-frequency thresholds (3000, 4000, and 6000 Hz) of the right ear were significantly higher than in the CG in both unaided and aided conditions (independent *t* test or Mann-Whitney test, all $P<.05$). However, there was no significant difference in the left ear. In speech audiometry, mean values for speech discrimination score were higher in the CG without a statistically significant difference (independent *t* test or Mann-Whitney test, all $P>.05$). No significant changes over time were observed in the repeated audiologic tests under either unaided or aided conditions (linear mixed model, all $P>.05$). These audiologic test results are presented in Figure 3.

Figure 3. Results of audiologic tests. (A) Pure tone audiometry showing significantly higher thresholds in the auditory training group (ATG) at 3000, 4000, and 6000 Hz frequencies in the right ear. (B) Speech reception threshold test showing no significant difference between the ATG and the control group (CG). (C) Speech discrimination scores were higher in the CG without statistical significance. A: aided; dB HL: decibel hearing level; U: unaided. *P<.05; independent t test or Wilcoxon rank sum test.



Speech Perception Tests

With the exception of the vowel imitation test, the initial mean scores for the remaining 5 tests were better in the CG. Although the Ling Six Sound test, monosyllable and bisyllable open-set tests, and the K-CID test showed no significant differences (adjusted P>.05 for all tests; Mann-Whitney test with Bonferroni correction), the initial consonant imitation test exhibited a statistically significant difference (adjusted P=.003; Mann-Whitney test with Bonferroni correction). Therefore, the aided 4 kHz threshold result of the right ear, which exhibited a significant difference in the initial PTA result (P=.002; Mann-Whitney test), was included as a variable in the analysis of the linear mixed model, considering that high-frequency hearing loss is a well-known factor affecting hearing perception [29].

When conducting a linear mixed model that includes the 4 kHz threshold of the right ear as a variable, we observed no significant difference in the initial scores of any speech perception tests (all P>.05).

In the Ling Six Sound test, the VCIT (vowel and consonant) and monosyllable test results showed no significant group differences (all P>.05, linear mixed model). However, in the bisyllable condition, the ATG demonstrated significantly better results compared to the CG (P=.04, linear mixed model). Similarly, in the K-CID test, the ATG showed significantly greater improvement than the CG (P=.03, linear mixed model). These results are visualized in Figure 4, and detailed linear mixed model analyses, including post-hoc P values, are presented in Table 2. Additionally, an arcsine transformation was applied to the percentage values, yielding almost the same results. Detailed values and statistical analyses of both the percentage and arcsine-transformed data are provided in Multimedia Appendix 1.

Figure 4. Results of speech perception tests. The auditory training group (ATG) showed significant improvement in speech perception compared to the control group (CG) in the intergroup analysis of the 2-syllable word recognition test and the K-CID (Korean version of Central Institute for the Deaf) test. ns: not significant; VCIT: Vowel and Consonant Imitation Test. *P<.05; linear mixed model.

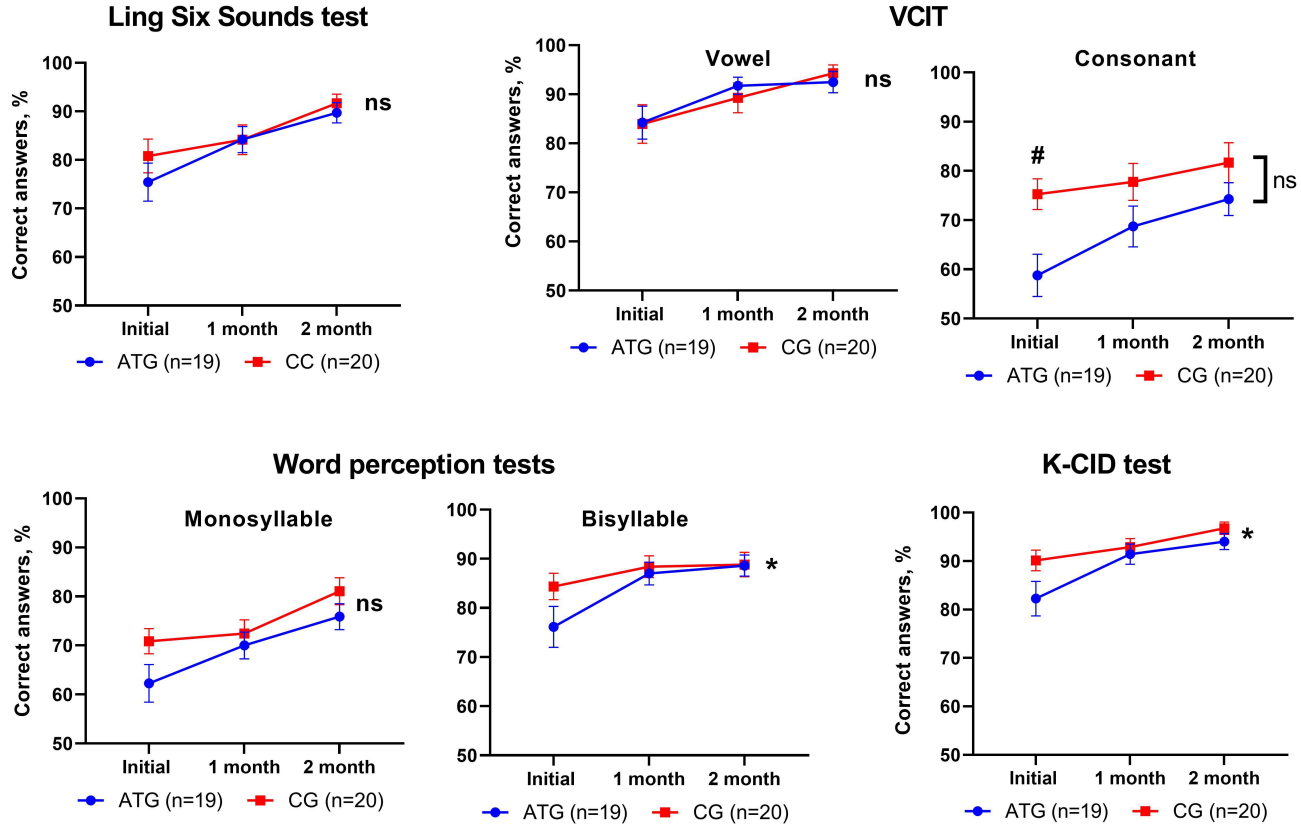


Table . Linear mixed model analysis of speech perception tests.

	Group, post hoc P value			Time, post hoc P value						Group × time, post hoc P value			F (df)	P value[a]
				Auditory training group			Control group							
	Initial	1 mo	2 mo	Initial minus 1 mo	1 mo minus 2 mo	Initial minus 2 mo	Initial minus 1 mo	1 mo minus 2 mo	Initial minus 2 mo	Initial minus 1 mo	1 mo minus 2 mo	Initial minus 2 mo		
Ling Six Sound test	>.99	.79	>.99	.05	<.001	.004	>.99	.11	.008	>.99	>.99	.03	0.64 (2, 73)	.53
VCIT[b] (vowel)	>.99	.74	>.99	.04	>.99	.03	.21	.27	.002	>.99	>.99	.25	0.52 (2, 73)	.60
VCIT (consonant)	.20	>.99	>.99	.01	.28	<.001	>.99	.70	.15	.07	.65	.01	2.24 (2, 73)	.11
Monosyllable test	.92	>.99	>.99	.04	.17	<.001	>.99	.02	.003	.51	.35	.004	1.04 (2, 73)	.36
Bisyllable test	>.99	.85	.47	<.001	>.99	<.001	.28	>.99	.19	.30	<.001	.24	3.37 (2, 73)	.04
K-CID[c]	.83	>.99	>.99	<.001	.35	<.001	.36	.09	.001	.16	<.001	.006	3.81 (2, 73)	.03

[a]Group × time interaction effects compared the initial minus 1 month or initial minus 2 months.

[b]VCIT: Vowel and Consonant Imitation Test.

[c]K-CID: Korean version of Central Institute for the Deaf test (linear mixed model).

Subjective Satisfaction

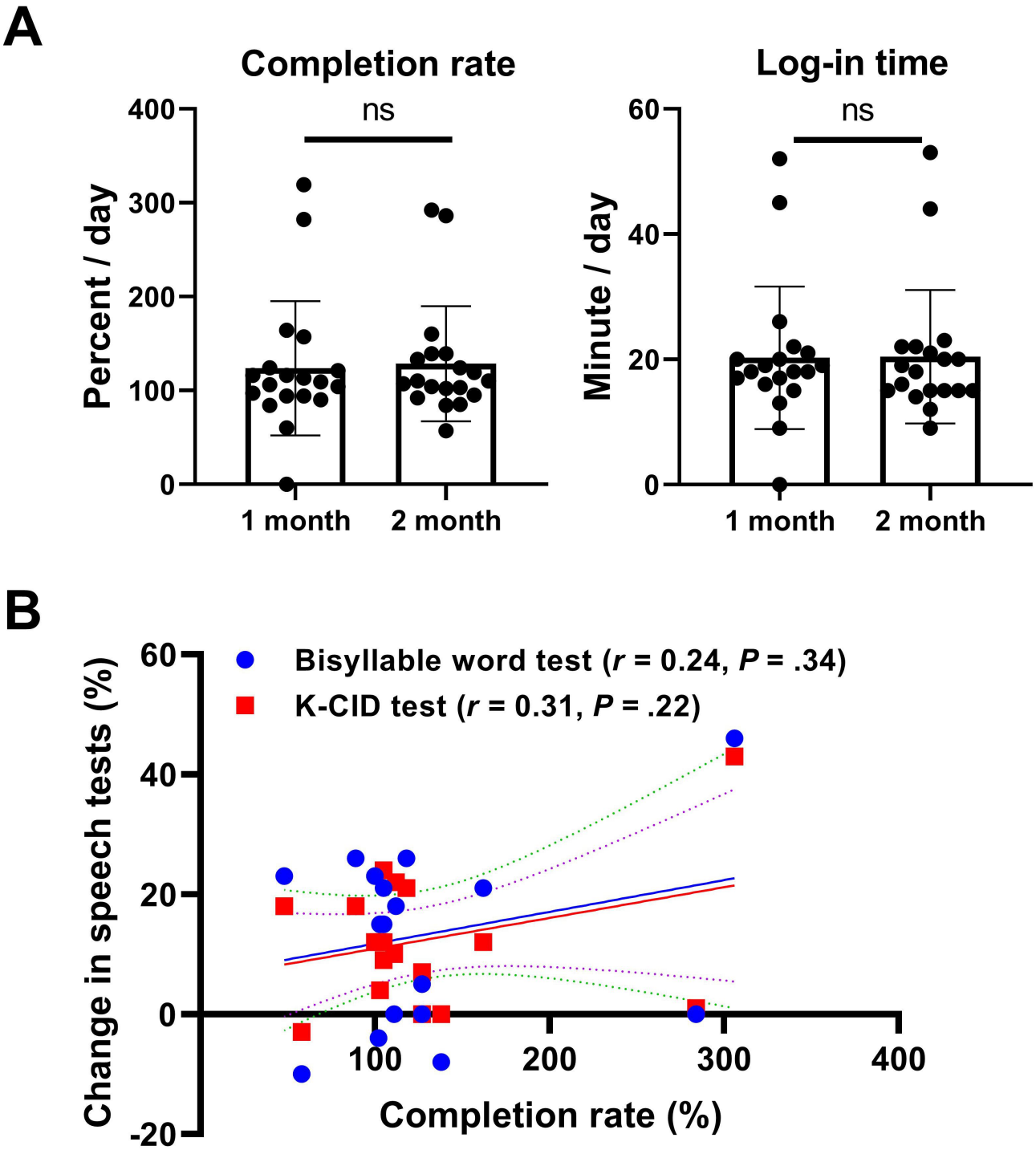
There was no significant difference between the 2 groups in all 3 types of questionnaire surveys (all $P>.05$; linear mixed model). Detailed results are described in Multimedia Appendix 2.

Compliance With CMAT Program

Except for one participant in the ATG who withdrew consent, all participants (n=20) in the intervention group accessed the CMAT program in a mobile environment. One of 20 participants (5%) was excluded from the analysis in this study because the duration of CMAT use was less than 50% of the required AT. The remaining 19 participants (95%) completed at least 50% of the training sessions, and 17 (85%) completed more than the required AT.

Log-in times, completion rates, and scores of all participants were collected by the CMAT program and analyzed. Mean daily application log-in time was 20.3 (SD 11.4) minutes in the first month and 20.4 (SD 10.7) minutes in the second. The completion rate was higher than our recommendation; 123.7% (SD 71.4%) in the first month and 128.5% (SD 61.3%) in the second month. There were no statistically significant differences among these variables ($P=.80$ for completion rate and $P=.50$ for log-in time; Wilcoxon signed rank test) (Figure 5A). The analysis of compliance and speech perception tests revealed no significant correlation (Figure 5B).

Figure 5. Compliance of chatbot-delivered mobile auditory training program and its correlation with speech perception tests. (A) Completion rate (%) and log-in time (min) were similar in the first month and second month (both *P*>.05; Wilcoxon signed rank test; error bars indicate SD). (B) No significant correlations were observed between compliance and speech perception tests (all *P*>.05, Pearson correlation test). K-CID: Korean version of Central Institute for the Deaf.



Discussion

Principal Results

In this study, we developed a new CMAT program in the Korean language, and we confirmed that it was effective in improving speech perception abilities and had high compliance among experienced bilateral HA users.

Our study evaluated speech perception performance at the levels of phoneme, word, and sentence. When the tests were repeated, they all showed improvements in both the ATG and CG. This is a similar finding as our previous report, and it might have been caused by the participants adapting to the test materials [10,30]. Interaction effects implied that compared to the initial time point, at 1 month and 2 months there was a greater performance improvement in the ATG in the bisyllable test and K-CID test. However, no significant findings were observed for the Ling Six Sound test, VCIT, or monosyllable test. The prominent improvement in bisyllable word and sentence perception tests could be explained by the effect of AT having

been achieved only to a limited degree for the trained task. The CMAT program used bisyllable words and sentences, which resulted in marked improvement on the tests assessing the same ability. Several previous studies suggested that the effect of AT appeared only in trained tasks and did not lead to improvements in untrained tasks [21,31,32].

All 3 questionnaires measuring subjective hearing disability and HA satisfaction showed no improvement after 2 months of CMAT, a result that is contrary to our previous study. Our previous study investigated the effectiveness of hospital-based AT and showed significant improvements on the K-IOI-HA and K-HHIE questionnaires in the intervention group [10]. A systematic review suggested that telemedicine has similar feasibility as conventional hearing rehabilitation; however, this review paper was not focused on AT; it focused more on HA fitting, device testing, and counseling [33]. To the best of our knowledge, only 3 comparative studies have been conducted of telemedicine and conventional AT [34-36]. A pilot study comparing the feasibility of tablet-based AT and conventional face-to-face AT in adult cochlear implant recipients reported similar results to our study. Even though tablet-based AT showed better speech improvements in tablet-based training, subjective auditory ability measured by the Oldenburger Inventory-R questionnaire showed partial improvement only in the conventional AT group and not in the teletherapeutic group [36]. The reason why we could not find subjective speech perception improvement in the questionnaire seems to be the lack of interaction between the participants and health care providers, who encourage and reassure patients during each session of conventional AT programs. A relatively short study period might be another reason. A long-term follow-up study or a study with participants who are more exposed to health care providers during CMAT might be needed to support our suggestions in the future.

Of the 20 participants who agreed to CMAT, 17 (85%) completed all or more requested training times, and there was no difference in compliance during the 2-month study. Our completion rate is similar to previous studies [10,37], and it seems that mobile-based AT is easily accessible and that the chat-based interface gives users a more interactive impression [8].

In our study, there was no statistically significant correlation between CMAT compliance and listening performance, which might be due to the high and homogeneous completion rate of an average of 120% of the program in the ATG. However, previous studies suggested that training duration is related to the improvement in listening abilities. A study using the Listening and Communication Enhancement training program in veterans using HAs reported that 84% (42/50) of participants completed the required sessions, and improvement in off-task performance was significantly better in the group of participants who completed all the sessions [38]. Another study examined the use of ReadmyQuips among new HA users and reported a correlation between the words-in-noise test and training time but not between the hearing-in-noise test and training time [39].

Strengths and Limitations

The mobile-based AT program we introduced has several advantages that set it apart from other internet-based AT programs. The first advantage lies in the interface, which features chatting, thus catering to users familiar with mobile messengers and providing a more convenient experience. Simplistic problem-solving might bore users easily, and gamification could come across as juvenile for adult users. Additionally, our interface uses a format in which users chat with a character resembling a medical professional, aiming to give users the impression of receiving remote treatment from experts. The second advantage pertains to the quality and quantity of the provided problems. Other AT programs sometimes present meaningless word lists or scenarios where the sound completely differs from the presented words. However, our CMAT program was designed with the intention of using 1540 frequently used words provided by the National Institute of the Korean Language to offer assistance in users' everyday lives. We aimed to enhance consonant discrimination abilities that are challenging for individuals with hearing impairments by creating problems that categorize cases where the vowels are similar but the consonants differ. Additionally, due to the vast number of questions, each exceeding 1000, participants might not encounter the same question more than once during the 2-month test period.

Nevertheless, CMAT also has its limitations. Recent artificial intelligence (AI)–powered language models like ChatGPT, which enable real-time interactions, were not integrated. The program did not incorporate various user-responsive technologies such as those found in AI-driven tools. Furthermore, additional settings such as background noise or speech speed adjustments were not possible in this version of CMAT. We refrained from including these adjustable settings to maintain the consistency of the training program and avoid introducing bias in the study. However, in the next version, we plan to incorporate features that allow users to adjust the program to their skill level.

From a clinical study perspective, there are several other advantages. Foremost among them is the fact that this is the first well-designed study conducted using a mobile-based AT program using the Korean language. This study also suggests the possibility that AT could be widely applied to the hearing rehabilitation process by showing the effect of AT on users with HA experience, which has been controversial in previous reports [19-21].

However, several limitations that emerged during the course of this study should also be taken into consideration. First, despite the randomization, the high-frequency hearing threshold in the right ear was significantly higher in the ATG than the CG, indicating that the CG score was better in most initial speech perception tests. We considered a subgroup analysis such as participant matching to avoid differences in initial hearing threshold, but we could not conduct such an analysis due to the limitations of a randomized prospective study with a relatively small number of participants. Instead, we tried to minimize unintentional bias by adjusting the 4 kHz threshold of the PTA when analyzing with the linear mixed model; however, there is

a possibility that the effect of CMAT was underestimated due to differences in initial hearing threshold. Further research using a larger number of participants is needed in the future. Second, a 2-month training time and follow-up duration might not be long enough to show the effect of CMAT on all types of hearing performance and subjective satisfaction. Therefore, a future study to evaluate the long-term effects of CMAT with a longer training time is needed. Third, this study was not blinded, and the potential risk of type I error should be considered. Lastly, fitting verification results, such as real-ear measurements, were not included, and data on HA use time during the trial were not included in this analysis. Instead, fitting validation was conducted using sound field tests, and only those who had been using HAs for more than 3 months and for at least 8 hours a day were included in this study. Incorporating more objective fitting verification and HA use time monitoring with data logging would be beneficial in future studies.

Several points for improving this CMAT program can be considered in the future. First and foremost, AI technology can be adapted to enable 2-way interactions between the user and the program. The AI system can generate an almost infinite number of new questions and adjust the question difficulty level based on the user's responses. Additionally, while this study used a chat-based interface for adults, more casual training methods, such as gamified training, could be adopted for pediatric patients with hearing loss.

Conclusions

In this study, we developed a novel mobile-accessible AT program with a chat-based interface and showed that it enhanced speech perception ability in experienced HA users. CMAT seems to increase the accessibility of and compliance to AT, and we expect that it will be used in all stages of the hearing rehabilitation process.

Acknowledgments

This work was supported by the Technology Development Program (S2964607), funded by the Ministry of SMEs and Startups of Korea. The funding organization had no role in the design or conduct of the study; in the collection, analysis, or interpretation of the data; in the decision to submit the article for publication; or in the preparation, review, or approval of the article.

Data Availability

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available because they contain information that could compromise the privacy of the research participants.

Conflicts of Interest

JL is the developer of the software used in this study (Nara Information Co, Ltd).

Multimedia Appendix 1
Linear mixed model analysis of speech perception tests after arcsine transformation.
[DOCX File, 33 KB - mhealth_v12i1e50292_app1.docx]

Multimedia Appendix 2
Detailed results of 3 questionnaire surveys.
[DOCX File, 25 KB - mhealth_v12i1e50292_app2.docx]

Checklist 1
CONSORT eHEALTH checklist (V 1.6.1).
[PDF File, 1032 KB - mhealth_v12i1e50292_app3.pdf]

References

1. WHO: 1 in 4 people projected to have hearing problems by 2050. World Health Organization. 2021 Mar 2. URL: https://www.who.int/news/item/02-03-2021-who-1-in-4-people-projected-to-have-hearing-problems-by-2050 [accessed 2024-01-17]
2. Cosh S, Helmer C, Delcourt C, Robins TG, Tully PJ. Depression in elderly patients with hearing loss: current perspectives. Clin Interv Aging 2019 Aug 14;14:1471-1480. [doi: 10.2147/CIA.S195824] [Medline: 31616138]
3. Lasak JM, Allen P, McVay T, Lewis D. Hearing loss: diagnosis and management. Prim Care 2014 Mar;41(1):19-31. [doi: 10.1016/j.pop.2013.10.003] [Medline: 24439878]
4. Loughrey DG, Kelly ME, Kelley GA, Brennan S, Lawlor BA. Association of age-related hearing loss with cognitive function, cognitive impairment, and dementia: a systematic review and meta-analysis. JAMA Otolaryngol Head Neck Surg 2018 Feb 1;144(2):115-126. [doi: 10.1001/jamaoto.2017.2513] [Medline: 29222544]
5. Boothroyd A. Adult aural rehabilitation: what is it and does it work? Trends Amplif 2007 Jun;11(2):63-71. [doi: 10.1177/1084713807301073] [Medline: 17494873]

6. Ferguson M, Maidment D, Henshaw H, Heffernan E. Evidence-based interventions for adult aural rehabilitation: that was then, this is now. Semin Hear 2019 Feb;40(1):68-84. [doi: 10.1055/s-0038-1676784] [Medline: 30728650]
7. Lawrence BJ, Jayakody DMP, Henshaw H, et al. Auditory and cognitive training for cognition in adults with hearing loss: a systematic review and meta-analysis. Trends Hear 2018;22:2331216518792096. [doi: 10.1177/2331216518792096] [Medline: 30092719]
8. Sweetow RW, Sabes JH. Auditory training and challenges associated with participation and compliance. J Am Acad Audiol 2010 Oct;21(9):586-593. [doi: 10.3766/jaaa.21.9.4] [Medline: 21241646]
9. Martin M. Software-based auditory training program found to reduce hearing aid return rate. Hear J 2007;60(8):32. [doi: 10.1097/01.HJ.0000286505.76344.10]
10. Han JS, Park JM, Kim Y, et al. Effect of hearing rehabilitation therapy program in hearing aid users: a prospective randomized controlled study. Clin Exp Otorhinolaryngol 2022 May;15(2):144-152. [doi: 10.21053/ceo.2021.00948] [Medline: 35255665]
11. Global smartphone penetration rate as share of population from 2016 to 2022. Statista. 2023. URL: https://www.statista.com/statistics/203734/global-smartphone-penetration-per-capita-since-2005/ [accessed 2024-01-17]
12. Smartphone ownership rate. Korea Information Society Development Institute. 2022. URL: https://stat.kisdi.re.kr/statHtml/statHtml.do?orgId=405&tblId=DT_405001_I182&conn_path=I2&language=en [accessed 2022-07-01]
13. Kim K, Park SY, Kang HC. Smartphone proficiency and use, loneliness, and ego integrity: an examination of older adult smartphone users in South Korea. Behav Inf Technol 2021 May 19;40(7):625-634. [doi: 10.1080/0144929X.2020.1713213]
14. Abbaspur-Behbahani S, Monaghesh E, Hajizadeh A, Fehresti S. Application of mobile health to support the elderly during the COVID-19 outbreak: a systematic review. Health Policy Technol 2022 Mar;11(1):100595. [doi: 10.1016/j.hlpt.2022.100595] [Medline: 35018280]
15. Top 3 frequently used applications of smart devices. Korea Information Society Development Institute. 2022. URL: https://stat.kisdi.re.kr/kor/tblInfo/TblInfoList.html?vw_cd=MT_ATITLE&siteGb=SITE002 [accessed 2024-01-04]
16. Olson A, Williams R, Livingston E, Futscher C. Review of auditory training mobile apps for adults with hearing loss. Perspect ASHA SIGs 2018 Jan;3(7):12-23. [doi: 10.1044/persp3.SIG7.12]
17. Stropahl M, Besser J, Launer S. Auditory training supports auditory rehabilitation: a state-of-the-art review. Ear Hear 2020 Aug;41(4):697-704. [doi: 10.1097/AUD.0000000000000806] [Medline: 31613823]
18. Dang A, Arora D, Rane P. Role of digital therapeutics and the changing future of healthcare. J Family Med Prim Care 2020 May;9(5):2207-2213. [doi: 10.4103/jfmpc.jfmpc_105_20] [Medline: 32754475]
19. Barcroft J, Spehar B, Tye-Murray N, Sommers M. Task- and talker-specific gains in auditory training. J Speech Lang Hear Res 2016 Aug 1;59(4):862-870. [doi: 10.1044/2016_JSLHR-H-15-0170] [Medline: 27567015]
20. Olson AD, Preminger JE, Shinn JB. The effect of LACE DVD training in new and experienced hearing aid users. J Am Acad Audiol 2013 Mar;24(3):214-230. [doi: 10.3766/jaaa.24.3.7] [Medline: 23506666]
21. Saunders GH, Smith SL, Chisolm TH, Frederick MT, McArdle RA, Wilson RH. A randomized control trial: supplementing hearing aid use with listening and communication enhancement (LACE) auditory training. Ear Hear 2016 Jul;37(4):381-396. [doi: 10.1097/AUD.0000000000000283] [Medline: 26901263]
22. Glista D, Scollie S, Moodie S, Easwar V. The LING 6(HL) test: typical pediatric performance data and clinical use evaluation. J Am Acad Audiol 2014 Nov;25(10):1008-1021. [doi: 10.3766/jaaa.25.10.9] [Medline: 25514453]
23. Kim SJ, Kim LS, Ahn YM, Lee HO, Rhee KS. Speech perception abilities over time in children with a cochlear implant on vowel and consonant imitation test. Korean J Otorhinolaryngol-Head Neck Surg 1997;40(12):1741-1751 [FREE Full text]
24. Kim LS, Lee MY, Huh MJ, Oh YJ. Open-set speech perception development in children with cochlear implants. Korean J Otorhinolaryngol-Head Neck Surg 2004;47(1):15-21. [doi: 10.3342/kjorl-hns.2010.53.9.534]
25. Jang JH, Chang HK, Park HY, et al. Comparison analysis between Korean Central Institute for the Deaf sentence and Korean hearing in noise test sentence. Korean J Otorhinolaryngol-Head Neck Surg 2012;55(2):85. [doi: 10.3342/kjorl-hns.2012.55.2.85]
26. Park SN, Han GC, Cho YS, et al. Standardization for a Korean version of Hearing Handicap Inventory for the Elderly. Korean J Otorhinolaryngol-Head Neck Surg 2011;54(12):828. [doi: 10.3342/kjorl-hns.2011.54.12.828]
27. Chu H, Cho YS, Park SN, et al. Standardization for a Korean adaptation of the International Outcome Inventory for Hearing Aids: study of validity and reliability. Korean J Otorhinolaryngol-Head Neck Surg 2012;55(1):20. [doi: 10.3342/kjorl-hns.2012.55.1.20]
28. Yun DH, Yoon TH, Lee KS. Subjective satisfaction in hearing aid users by APHAB. Korean J Otorhinolaryngol-Head Neck Surg 2000;43(7):698-702.
29. Hogan CA, Turner CW. High-frequency audibility: benefits for hearing-impaired listeners. J Acoust Soc Am 1998 Jul;104(1):432-441. [doi: 10.1121/1.423247] [Medline: 9670535]
30. Yund EW, Woods DL. Content and procedural learning in repeated sentence tests of speech perception. Ear Hear 2010 Dec;31(6):769-778. [doi: 10.1097/AUD.0b013e3181e68e4a] [Medline: 20562624]
31. Humes LE, Kinney DL, Brown SE, Kiener AL, Quigley TM. The effects of dosage and duration of auditory training for older adults with hearing impairment. J Acoust Soc Am 2014 Sep;136(3):EL224. [doi: 10.1121/1.4890663] [Medline: 25190425]

32. Reis M, McMahon CM, Távora-Vieira D, Humburg P, Boisvert I. Effectiveness of computer-based auditory training for adult cochlear implant users: a randomized crossover study. Trends Hear 2021 Jan;25:23312165211025938. [doi: 10.1177/23312165211025938] [Medline: 34591702]
33. Bush ML, Thompson R, Irungu C, Ayugi J. The role of telemedicine in auditory rehabilitation: a systematic review. Otol Neurotol 2016 Dec;37(10):1466-1474. [doi: 10.1097/MAO.0000000000001236] [Medline: 27755363]
34. Bernstein C, Bakke M, Mazevski A, et al. Benefits of speech tracking training on sentence recognition, tracking rate, and self-assessed communication function in adult cochlear implant users. J Acad Rehabil Audiol 2012 Jan;45:11-39 [FREE Full text]
35. Verma H, N B, Mishra R, Panda NK. Outcome measures following tele-rehabilitation and conventional face to face rehabilitation in paediatric cochlear implant users during COVID-19 pandemic: a pilot study in a tertiary care setup. J Otol 2022 Jan;17(1):31-38. [doi: 10.1016/j.joto.2021.10.002] [Medline: 34745238]
36. Völter C, Stöckmann C, Schirmer C, Dazert S. Tablet-based telerehabilitation versus conventional face-to-face rehabilitation after cochlear implantation: prospective intervention pilot study. JMIR Rehabil Assist Technol 2021 Mar 12;8(1):e20405. [doi: 10.2196/20405] [Medline: 33709934]
37. Henshaw H, Ferguson MA. Efficacy of individual computer-based auditory training for people with hearing loss: a systematic review of the evidence. PLoS One 2013;8(5):e62836. [doi: 10.1371/journal.pone.0062836] [Medline: 23675431]
38. Chisolm TH, Saunders GH, Frederick MT, McArdle RA, Smith SL, Wilson RH. Learning to listen again: the role of compliance in auditory training for adults with hearing loss. Am J Audiol 2013 Dec;22(2):339-342. [doi: 10.1044/1059-0889(2013/12-0081)] [Medline: 24018575]
39. Abrams HB, Bock K, Irey RL. Can a remotely delivered auditory training program improve speech-in-noise understanding? Am J Audiol 2015 Sep;24(3):333-337. [doi: 10.1044/2015_AJA-15-0002] [Medline: 26649542]

Abbreviations

AI : artificial intelligence
APHAB: Abbreviated Profile of Hearing Aid Benefit
AT: auditory training
ATG: auditory training group
CG: control group
CMAT: chat-based mobile auditory training
CONSORT: Consolidated Standards of Reporting Trials
HA: hearing aid
K-CID: Korean version of Central Institute for the Deaf
K-HHIE: Korean version of the Hearing Handicap Inventory for the Elderly
K-IOI-HA: Korean version of International Outcome Inventory for Hearing Aids
PTA: pure tone audiogram
VCIT: Vowel and Consonant Imitation Test

Edited by L Buis; submitted 27.06.23; peer-reviewed by A Olson, E Picou; revised version received 11.12.23; accepted 11.12.23; published 09.02.24.

Please cite as:
Han JS, Lim JH, Kim Y, Aliyeva A, Seo JH, Lee J, Park SN
Hearing Rehabilitation With a Chat-Based Mobile Auditory Training Program in Experienced Hearing Aid Users: Prospective Randomized Controlled Study
JMIR Mhealth Uhealth 2024;12:e50292
URL: https://mhealth.jmir.org/2024/1/e50292
doi:10.2196/50292

Original Paper

Documentation Completeness and Nurses' Perceptions of a Novel Electronic App for Medical Resuscitation in the Emergency Room: Mixed Methods Approach

Kin Cheung[1*], PhD; Chak Sum Yip[2,3*], MSN

[1]School of Nursing, The Hong Kong Polytechnic University, Hong Kong, China (Hong Kong)

[2]Accident and Emergency Department, Tuen Mun Hospital, Hong Kong, China (Hong Kong)

[3]Quality & Safety Office, The Hong Kong Children's Hospital, Hong Kong, China (Hong Kong)

[*]all authors contributed equally

Corresponding Author:
Kin Cheung, PhD
School of Nursing
The Hong Kong Polytechnic University
11 Yuk Choi Road, Hung Hom
Hong Kong
China (Hong Kong)
Phone: 852 2766 6773
Email: kin.cheung@polyu.edu.hk

Abstract

Background: Complete documentation of critical care events in the accident and emergency department (AED) is essential. Due to the fast-paced and complex nature of resuscitation cases, missing data is a common issue during emergency situations.

Objective: This study aimed to evaluate the impact of a tablet-based resuscitation record on documentation completeness during medical resuscitations and nurses' perceptions of the use of the tablet app.

Methods: A mixed methods approach was adopted. To collect quantitative data, randomized retrospective reviews of paper-based resuscitation records before implementation of the tablet (Pre-App Paper; n=176), paper-based resuscitation records after implementation of the tablet (Post-App Paper; n=176), and electronic tablet-based resuscitation records (Post-App Electronic; n=176) using a documentation completeness checklist were conducted. The checklist was validated by 4 experts in the emergency medicine field. The content validity index (CVI) was calculated using the scale CVI (S-CVI). The universal agreement S-CVI was 0.822, and the average S-CVI was 0.939. The checklist consisted of the following 5 domains: basic information, vital signs, procedures, investigations, and medications. To collect qualitative data, nurses' perceptions of the app for electronic resuscitation documentation were obtained using individual interviews. Reporting of the qualitative data was guided by Consolidated Criteria for Reporting Qualitative Studies (COREQ) to enhance rigor.

Results: A significantly higher documentation rate in all 5 domains (ie, basic information, vital signs, procedures, investigations, and medications) was present with Post-App Electronic than with Post-App Paper, but there were no significant differences in the 5 domains between Pre-App Paper and Post-App Paper. The qualitative analysis resulted in main categories of "advantages of tablet-based documentation of resuscitation records," "challenges with tablet-based documentation of resuscitation records," and "areas for improvement of tablet-based resuscitation records."

Conclusions: This study demonstrated that higher documentation completion rates are achieved with electronic tablet-based resuscitation records than with traditional paper records. During the transition period, the nurse documenters faced general problems with resuscitation documentation such as multitasking and unique challenges such as software updates and a need to familiarize themselves with the app's layout. Automation should be considered during future app development to improve documentation and redistribute more time for patient care. Nurses should continue to provide feedback on the app's usability and functionality during app refinement to ensure a successful transition and future development of electronic documentation records.

(JMIR Mhealth Uhealth 2024;12:e46744) doi:10.2196/46744

KEYWORDS

tablet computer; nursing documentation; paper resuscitation record; electronic resuscitation record; medical resuscitation; electronic medical record; documentation; resuscitation; electronic health record; nurses' perception; traditional paper record; nurse

Introduction

Background

The completeness of documentation of critical care events in the accident and emergency department (AED) is essential for (1) the continuity of patient care, (2) medicolegal issues [1], (3) improving accessibility to critical information needed for research [2], and (4) serving as evidence for quality outcome measures [3]. Traditionally, documentation is performed on paper. Due to the fast-paced and complex nature of resuscitation cases, missing data is a common issue during emergency situations. As much as 60% of essential data fields in prehospital paper records can be incomplete [4]. A study in a trauma center also found incompleteness in 18% of the mandatory elements for trauma resuscitation [5].

Over the last 2 decades, there has been a global trend of switching to electronic medical records (EMRs). It was estimated that about 46% of AEDs in the United States used EMRs in 2010 [6]. This percentage is expected to increase in the future. Fully functional EMRs have been shown to improve efficiency in AEDs [7]. Despite the growth in usage, very few studies have explored the impact of EMRs in AED settings. Furthermore, the perceptions toward EMRs are mixed. A study found that nurses and physicians generally had a negative perception toward EMRs in the AED. EMRs are considered to be ineffective, redundant, and prone to error [8]. In contrast, another study found that nurses perceived that their productivity increased and care was better coordinated after implementing EMRs [9].

Among the limited studies in the area, a retrospective review of trauma resuscitations in AED settings showed that EMRs can improve documentation completeness [5,10]. However, there is a lack of studies on medical resuscitations, which are more common than trauma resuscitations [11]. Medical resuscitations are performed on triage category I and II patients with life-threatening conditions such as myocardial infarction, sepsis, and stroke [12]. Medical resuscitations differ from trauma resuscitations in that they do not follow a single protocol. The differences in management protocols can make the process of documenting medical resuscitations different from that of trauma resuscitations.

The EMR system used by the Hospital Authority of Hong Kong is called the Clinical Management System (CMS). It is an integrated platform that allows clinical users to manage the following daily clinical activities [13]: (1) obtain clinical data including consultation notes, laboratory, and imaging results; (2) document clinical activities; and (3) provide clinical decision support.

In AEDs in Hong Kong, medical records are still part paper and part electronic, with the patient's clinical notes being documented on paper. This type of mixed documentation has been shown to hinder effective communication and utilization of information in either record [14].

Recently, there has been a trend of switching to electronic documentation in AEDs. In 2020, 3 of 18 AEDs in Hong Kong had switched to an EMR system called the eAED. It was expected that, by 2023, about two-thirds of AEDs would have switched to the eAED. The eAED is meant to replace paper charts previously used to document a patient's progress [15]. Despite the gradual adoption of EMRs in AEDs in Hong Kong, the use of electronic documentation during medical resuscitations has not occurred owing to the time-critical, fast-paced nature and lack of a suitable application.

However, with advances in computer processing power, a tablet-based system could fill the gap. Documentation efficiency and data precision have improved when a tablet-based app was used, in comparison with a desktop EMR, during a simulation [16]. In Hong Kong, a tablet-based system called "eResus" is being developed by the Hong Kong Hospital Authority for medical and trauma resuscitation documentation. With the implementation of the eAED and eResus, documentation in AEDs would become fully electronic.

Aim and Objectives

This study aimed to evaluate the impact of a tablet-based app on documentation completeness during medical resuscitations. The research questions were the following:

- What are the differences between paper and electronic tablet-based records on the levels of documentation completeness?
- What are the perceptions of emergency room nurses regarding documentation completeness when using eResus?

Hypothesis

This study hypothesized that the completeness of resuscitation documentation using electronic tablet-based records would be higher than that that using paper records.

Methods

Design

To answer research question 1, a randomized retrospective review of paper and electronic resuscitation medical records (N=528) was conducted using a documentation completeness checklist. The study was implemented during the transition from paper to electronic documentation, when only triage category II cases would be documented using the tablet-based app called eResus. Therefore, triage category II records were collected before (from November 2020 to December 2020) and after (from February 2021 to March 2021) implementation of the tablet-based eResus app. Paper records were collected before (Pre-App Paper) and 1 month after (Post-App Paper) implementation of eResus, while electronic records were collected 1 month after (Post-App Electronic) implementation of eResus. We randomly selected 176 records each for the

Pre-App Paper, Post-App Paper, and Post-App Electronic record sets from CMS using a random number generator.

To answer research question 2, emergency nurses' perceptions of the advantages, challenges, and areas for improvement of the electronic app for resuscitation documentation were obtained in individual interviews conducted in mid-April 2021, 3 months after the implementation of eResus. Reporting of the qualitative findings was guided by the Consolidated Criteria for Reporting Qualitative Studies (COREQ) [17], as delineated in the following sections, to enhance rigor.

Ethical Considerations

Ethical clearance (NTWC/REC/20098) from the study hospital and the Human Subjects Ethics Sub-committee of the Hong Kong Polytechnic University (HSEARS20200826001) was obtained before the commencement of the study.

Data

Quantitative Data

A documentation completeness checklist was established based on the literature and a review of the department's current medical resuscitation event documentation. The checklist consisted of 5 essential domains (ie, basic information, vital signs, procedures, investigations, and medications) of medical resuscitation as illustrated in Multimedia Appendix 1. Face and content validity of the checklist were determined by 4 experts in the emergency medicine field [18]. Experts were invited based on the following criteria: (1) worked in an AED and (2) published at least one article related to the accident and emergency field. The expert panel consisted of 1 associate consultant, 1 medical officer, and 2 advanced practice nurses (1 of which was a Fellow in Emergency Nursing). The content validity index (CVI) was calculated using the scale CVI (S-CVI). The S-CVI is calculated based on the number of items in the scale rated by the expert as "quite relevant" or "highly relevant" [19]. The S-CVI was further analyzed by universal agreement (UA) among experts (S-CVI/UA) and the average (S-CVI/Ave). The checklist's S-CVI/UA was 0.822, and the S-CVI/Ave was 0.939.

The medical resuscitation documents were reviewed against the validated checklist by a researcher (CSY), and intrarater reliability was determined to ensure consistency. Intrarater agreement was calculated using the Cohen kappa [19]. We evaluated 5 cases at week 0 and week 2. The agreement between the 2 records was considered acceptable at a κ of 0.884 (95% CI 0.671-1.105; P<.001).

For each resuscitation documentation review, the researcher provided a dichotomous response of "Yes or No" for each item and identified the level of completeness based on the checklist. The patient diagnosis, length of medical resuscitation, initial triage category, and demographics including age and gender were also collected as part of the basic information. However, patient names and identification numbers were not collected. Data were kept anonymous by assigning codes only identifiable to the researcher.

Qualitative Data

Emergency room nurses were guided to discuss their thoughts on the eResus app's features for documentation completeness through individual interviews with an onsite nurse who was one of the researchers (CSY). They understood the aim of the study, and their experience with the app was explored. Participants' demographic data including age, gender, years of experience after graduation, and years of experience in their current specialty were collected for subsequent data analysis.

Sample Size

For the quantitative data, the Chi-square test was used to compare the differences in documentation completeness between the 3 groups. Based on the findings from a level 1 pediatric trauma center in 2015 [10], with an α of .05 and power of 0.80, a minimum sample size of 153 medical records per record set was required. To ensure an adequate sample size, 176 patient records were included in each of the paper and electronic record sets, resulting in a total sample size of 528 records (ie, 176 records each for the Pre-App Paper, Post-App Paper, and Post-App Electronic record sets).

For qualitative data, data saturation is the criterion to determine the sample size. Data are considered saturated when no new theoretical insights are gained from new data [20]. For this study, data saturation was achieved after 10 individual interviews, and 2 more interviews were conducted to confirm the data saturation.

Recruitment

The study was conducted in the AED of 1 hospital in Hong Kong. It is one the major local trauma centers providing 24-hour accident and emergency services and serves more than 190,000 patients per year, with over 300 resuscitation cases per month [21]. The tablet app was scheduled to be implemented in June 2020 but was postponed due to COVID-19. The app was eventually implemented in January 2021.

Quantitative Data Collection Method

In this study, we reviewed 2 types of resuscitation documents, namely paper and tablet-based resuscitation records. Completed resuscitation documents in paper format were attached to the patient's CMS record by optical scanning as per usual practice. These records were stored in the CMS.

Training prior to the implementation of the electronic resuscitation record could lead to bias toward improved completeness of electronic documents [10]. Therefore, paper documentation records were collected before and 1 month after implementation of eResus to address this issue. First, baseline paper resuscitation records were collected prior to implementation of the eResus app (Pre-App Paper). After the implementation of the eResus app with training, there was a washout period of 1 month. After 1 month, the paper (Post-App Paper) and tablet-based (Post-App Electronic) resuscitation records were retrieved for analysis. Both paper and tablet resuscitation records involving trauma team activation or triage category I cases were excluded because the app did not cover these 2 types of cases at this stage.

For the purpose of this study, 3 lists of 3 months of case lists including eligible medical records from the Pre-App Paper, Post-App Paper, and Post-App Electronic record sets were retrospectively generated from the CMS and assigned a serial number. From each group, 176 records were randomly selected using a random number generator.

Qualitative Data Collection Method

In terms of qualitative data collection, purposive sampling was applied. The researcher conducted individual, voice-recorded interviews with each emergency nurse 3 months after the eResus implementation. The nurses were provided an explanation of the study, and written consent was obtained. The inclusion criteria included nurses (1) working in the AED of the hospital, (2) with experience using the eResus app, (3) who spoke Cantonese and were able to read English, and (4) working in their current position for more than 3 months.

Invitation emails were sent to colleagues. Eligible colleagues who replied to the email or expressed interest were invited to be interviewed according to their years of experience. Individual interviews were conducted in a quiet room or via Zoom. Each interview lasted about 1 hour or stopped when the interviewee felt that their viewpoint had been fully expressed. The interview guide is shown in Textbox 1.

Textbox 1. Interview guide.

Opening question:

Can you tell me your experience with using the eResus app until now?

Guiding questions:

1. What are the main advantages and challenges with achieving high documentation completeness when using eResus in the resuscitation room?
2. How do you think eResus can be improved to help you achieve better documentation completeness?

Data Analysis Methods

For the quantitative data analysis, SPSS version 25 (IBM Corp) was used. Descriptive statistics such as means, standard deviations, frequencies, and percentages were used to present the study variables. Normality was tested using the Kolmogorov-Smirnov test, and the data were found to be not normally distributed. The Mann Whitney *U* test was used to compare mean ranks for age and clinical characteristics between the Pre-App Paper and Post-App Paper record sets as well as between the Post-App Paper and Post-App Electronic record sets to ensure the clinical characteristics of the 3 groups were comparable. Subsequently, Chi-square tests were used to compare the differences in proportions, such as the percentage of completeness between the Pre-App Paper and Post-App Paper record sets to determine any historical bias or effect from training and then between the Post-App Paper and Post-App Electronic record sets. Results with a *P* value <.05 were considered significant.

Each resuscitation record was manually reviewed against the study checklist for data element completeness. Each record was reviewed individually. Any incomplete data element was entered as an incomplete domain for the respective domain of the 5 domains, namely basic information, vital signs, procedures, investigations, and medications. For example, for records of the administration of 2 medications that use the same route, the record was treated as 2 separate data entries. If 1 of the data items (such as 1 missing medication name) was incomplete, the medication domain for that case was entered as incomplete. The number of entries for each domain of the resuscitation documentation was analyzed, delineating sections that were recorded at higher or lower frequencies.

For the qualitative data analysis, content analysis was performed [22]. First, the interview was audio-recorded and transcribed verbatim into Chinese. NVivo Pro 12 was used for data analysis. The researcher read through the transcript multiple times to become immersed in the data. Participants' experiences with the eResus app, challenges, and possible solutions were extracted and summarized into meaning units. Third, each meaning unit was condensed and labelled with codes. Fourth, subcategories were identified by comparing the similarities and differences between different codes. Finally, the latent meanings of the subcategories were sorted into themes.

Results

Quantitative Results

Tables 1 and 2 present the characteristics of the 5 domains of resuscitation documentation. Comparisons were made between the Pre-App Paper and Post-App Paper record sets. There were no significant differences in characteristics or documentation completion between the Pre-App Paper and Post-App Paper record sets (Table 3).

Table 1. Comparisons using the Mann-Whitney U test among the 3 resuscitation record sets in patient age; length of resuscitation; and total numbers of vital sign entries, procedures, investigations, and medications in the resuscitation documentation for patients requiring medical resuscitation (N=528).

5 domains	Pre-App Paper record set (n=176)		Post-App Paper record set (n=176)		Difference between Pre-App Paper and Post-App Paper record sets		Post-App Electronic record set (N=176)		Difference between Post-App Paper and Post-App Electronic record sets	
	Range	Mean (SD)	Range	Mean (SD)	*U*	*P* value	Range	Mean (SD)	*U*	*P* value
Basic information										
Age (years)	4-98	61.6 (20.6)	3-101	61.6 (22.0)	15,121	.84	4-100	62.5 (21.1)	15,188	.90
Length of resuscitation (minutes)	3-216	40.6 (30.2)	5-175	41.7 (28.6)	14,673	.39	8-367	40.5 (37.8)	14,214	.18
Total number of vital sign entries	1-58	10.9 (8.0)	2-47	11.1 (7.1)	15,089	.68	2-72	9.9 (7.6)	13,258	.02
Total number of procedures	0-7	1.83 (1.34)	0-8	1.76 (1.26)	15,050	.61	0-10	1.50 (1.12)	13,900	.06
Total number of investigations	0-11	5.72 (1.89)	0-11	5.49 (1.92)	15,091	.67	0-11	5.70 (1.91)	14,741	.42
Total number of medications	0-12	1.65 (2.07)	0-15	1.70 (2.31)	15,085	.66	0-19	1.68 (2.41)	15,288	.83

Table 2. Gender differences among the 3 groups of resuscitation records for patients requiring medical resuscitation (N=528), as assessed using the Chi-square test.

Gender	Pre-App Paper record set (n=176), n (%)	Post-App Paper record set (n=176), n (%)	Post-App Electronic record set (n=176), n (%)	Difference among the groups	
				χ^2 (df)	*P* value
Female	84 (47.7)	84 (47.7)	86 (48.9)	0.06 (2)	.97

Table 3. Differences in completion of the 5 domains of documentation between paper and electronic resuscitation records (N=528).

5 domains	Pre-App Paper record set (n=176), n (%)	Post-App Paper record set (n=176), n (%)	Post-App Electronic record set (n=176), n (%)	Differences among groups	
				χ^2 (df)	*P* value
Basic information	113 (64.2)	105 (59.7)	128 (72.7)	6.86 (2)	<.001
Vital sign	116 (65.9)	108 (61.4)	158 (89.8)	40.97 (2)	<.001
Procedures	123 (69.9)	127 (72.2)	176 (100)	63.50 (2)	<.001
Investigations	101 (57.4)	93 (52.8)	128 (72.7)	16.06 (2)	<.001
Medications	158 (89.8)	163 (92.6)	175 (99.4)	15.24 (2)	<.001

For the post-app comparison, there were no significant differences in the characteristics, except the number of vital sign entries, between the Post-App Paper and Post-App Electronic record sets (Tables 1 and 2). To answer research question 1, there was a significantly higher completion rate for all 5 domains in the Post-App Electronic record set than in Post-App Paper record set (Table 3).

Qualitative Results

Categories

The objective of the qualitative study was to explore nurses' perceptions of the use of eResus for documentation completeness. The main categories identified were "advantages of tablet-based documentation of resuscitation records," "challenges with tablet-based documentation of resuscitation records," and "areas for improvement of tablet-based resuscitation records" (Textbox 2).

Textbox 2. Summary of the categories and subcategories.

Advantages of tablet-based documentation of resuscitation records

- Structural guidance for documentation
- Easy to review and edit documentation
- Comparable mobility to paper and superior to desktop

Challenges with tablet-based documentation of resuscitation records

- System loading speed and stability
- Familiarization with the app layout

Areas for improvement of tablet-based resuscitation records

- Need for speedy documentation and automated documentation

Data saturation was achieved after conducting individual interviews with 12 nurses. The mean age of the participants was 26.9 (SD 2.68) years, and 9 participants were female. Participants' mean length of work experience was 4.46 (SD 2.20) years, with a range of 2.5 years to 8.5 years. Their mean length of work experience in AED was 3.3 (SD 1.87) years, with a range of 1.5 years to 8.5 years.

Advantages of Tablet-Based Documentation of Resuscitation Records

Structural Guidance for Documentation

The electronic app included an extensive database that encompassed the essential aspects of resuscitation documentation. The participants appreciated the app's preset data fields that prompted users to input essential data during documentation.

> *(During documentation of blood glucose,) the interface displayed all the data field such as time, result, performer. You definitely cannot forget to input.* [D168-169]
>
> *(After urinary catheter insertion) I may forget to write urinary output..., but eResus would prompt you if you did not enter.* [H 170-173]
>
> *The application made sure that you have 2 colleagues to countercheck the medication and documented their name before administration.* [C164-165]

The built-in logic set by emergency physicians and nurses provided clinical management support and guidance to users during documentation. Certain data fields were auto filled, saving more time for nursing care.

> *I found it convenient because the application would lead you how to input data in a step-by-step fashion.* [A49]
>
> *After inputting the systolic blood pressure, it would automatically divert you to the diastolic blood pressure.* [C70-71]
>
> *When asystole rhythm was chosen, the data field on blood pressure, pulse etc. would be prohibited from inputting...We no longer have to write "undetectable" over and over again.* [E47-52]

Furthermore, the electronic app reduced the need for verbal order prescriptions and allowed structured electronic prescriptions, which are less prone to error during documentation and administration.

> *Verbal order was prone to miscommunication, distraction, and error in administration.* [C146-147]
>
> *In the past, I would have to remember or write down physician's verbal order..., but now the drug name, dosage, infusion speed etc. would all be on the screen.* [K 41-47]

Easy to Review and Edit Documentation

Medical resuscitation documentation has to be done contemporaneously during resuscitation. Electronic documentation can ensure legibility compared with handwriting, and users were able to review specific aspects of the documentation for completeness using in-app features.

> *Colleagues' handwriting could be illegible; maybe everyone was in a hurry. And colleagues could misspell words, which could affect handover to ward colleagues.* [E22-24]
>
> *Someone may accidently splash alcohol onto the paper chart, and the word would become illegible.* [H209-210]
>
> *The application has a filter function which allows you to choose vital signs, allowing you to review vital sign inputs and trends or procedures, allowing you to review whether you have forgotten to document something.* [K123-125]

Fragmented information was conveyed to the documenter from various sources, in a random sequence. Not all users can correctly recall the exact sequence of medical resuscitation as they document. Electronic documentation allowed the users time to edit the sequence rather than having to rewrite the whole resuscitation event on a new paper resuscitation record.

> *If the handwriting was too ugly and the time sequence is too out of place, such as the medication administration time did not align to the corresponding row, then I would cross out the whole paper chart and rewrite it. But now, eResus can easily amend it.* [I 104-106]

Comparable Mobility to Paper and Superior to Desktop

Patients requiring medical resuscitation would often need to be transferred to another department for investigation or intervention. A tablet app can provide the mobility needed to document in various locations.

> *Let's say the patient has to be escorted to computed tomography (CT). I would take the table to the CT suite (to continue the document). When we returned to the resuscitation bay, I could use the Bluetooth keyboard to continue the document. It's better than desktop.* [L 169-171]
>
> *You can bring it (to CT) like paper...don't even need to bring pen, just use your fingers.* [J 179-180]

Challenges With Tablet-Based Documentation of Resuscitation Records

System Loading Speed and Stability

Participants embraced the transition to electronic documentation. However, participants reported technical challenges due to the internet connection or app coding issues when using the electronic app that could compromise documentation completeness. The fast pace of medical resuscitations and contemporaneous nature of the documentation exacerbated the problem.

> *It has some technical problems...there was a time when it kept crashing and could not input data.* [A58-60]
>
> *Sometimes, switching between different tab pages is rather slow.* [F60]
>
> *The patient was...in asystole, we were conducting chest compression, and administering medication, but the application was still loading.* [I 116-117]

Familiarization With the App Layout

Navigating through the various tab bars, interface, and data fields of the app was different from the paper resuscitation record that presented all the data fields on the same page. Users were required to tab multiple times to access the desired data fields on the tablet, which was more time-consuming. All participants received training prior to using the electronic app in clinical settings. They believed that being familiar with the design of the app takes time and practice:

> *I have to tab this and that before I could input data...if it was handwritten, it would be much quicker.* [H 53-57]
>
> *This application has different tabs and options, which require a bit of thinking...it is like using a phone.* [C33-43]
>
> *When you first encounter the application, you would need to spend time to learn the layout. But after you have become familiar with it, you would find the documentation process very smooth.* [D147-149]

Areas for Improvement of Tablet-Based Resuscitation Records

All participants reported that speed of documentation was an important aspect in resuscitation documentation. During medical resuscitations, the case nurse was required to perform patient care and document contemporaneously. These resuscitation events were highly demanding and required speedy documentation:

> *Sometime, the documentation with eResus could take up lots of time. There were cases when we needed to document lots of medication right at the beginning. The application may not be able to document events in real time.* [A63-64]
>
> *When you have many items pending documentation, you would be naturally prone to incomplete documentation.* [B76-77]

The multitasking nature of the nursing practice posed competing demands between managing patient care and documentation, which required the nurses to compromise. Nurses decreased the frequency of taking vital signs. One participant said:

> *When handling less critical cases...I would take vital signs every 5 minutes (instead of 3) so that I can be more at ease when managing both the patient and documentation.* [D113-117]

Users appreciated the auto retrieval of data from the Hospital Authority's network and the auto fill of relevant fields. Relevant data previously inputted into the app were prepopulated either automatically or after the user's approval:

> *It would auto-capture allergy status from CMS, an electronic health record system used in Hong Kong).* [G 102-103]
>
> *It can retrieve the previous (vital sign) data. Then, I can tab it and paste it onto the data field...such as Glasgow Coma Scale score, etc.* [H 44-46]

Faster documentation speed can improve documentation completeness and overall resuscitation quality. Participants believed that speed and improved care could be gained by automation:

> *If a multimonitor could automatically record vital signs and transfer data into the application, the user would spend less time inputting data and more time looking after the patient...or checking the (resuscitation) record for incompletion.* [D254-262]
>
> *(Automated vital sign recording) would be useful...But some factors could affect the reading's accuracy; it should allow health care workers to verify the readings prior to documenting.* [K 217-220]

Documentation speed can also be gained by flattening the user interface (UI). Participants found that, although organizing data fields into different categories and layers was logical, it made navigating through the layers inevitably slow. A more direct, intuitive UI is needed to improve the speed of data entry:

> *(The commonly used) items should be accessible with one tap.* [G194]

The commonly performed investigations...that has many data fields should be more easily accessible. [L 216-220]

Discussion

Principal Findings

This is the first study, particularly in Asia, to compare the completion rates of documentation between paper and tablet-based resuscitation records in the emergency room. Our results indicate that electronic documentation is promising, with a higher completion rate than with the paper format.

Our study results support the hypothesis that tablet-based documentation of resuscitation records results in a higher documentation completion rate than paper formats in all 5 domains. Previous studies in AED settings have been conducted to compare the completion rates of key data elements for trauma resuscitation records between paper and electronic formats for adult [23] and pediatric [10] trauma cases. Both studies found areas for improvement and degradation in the key data elements. However, it is difficult to directly compare the studies since the outcome measures were different and none of the studies in AED settings used tablet-based devices. Nevertheless, our study results were consistent with those of previous studies that supported that, with electronic resuscitation records, documentation completion rates were higher, particularly of basic information such as case start time and disposal but not for serial vital signs [10,23]. Interestingly, although no difference was found in the completeness of documenting vital signs and interventions in these previous studies, our study showed improvement in the vital signs, procedures, investigations, and medication domains. This may be due to the differences in the inherent design of the EMRs and the use of a tablet-based device instead of desktops, as explained by interview participants. In addition, contrary to the concern of bias for improved documentation completeness caused by training [10], our study showed no statistically significant differences between the Pre-App Paper and Post-App Paper record sets. This indicates that missing data with the paper format could be consistent since the documentation format has not changed (such as using the same paper form).

Our qualitative results further explain the reasons why the tablet-based device could improve documentation completeness. The structural design of the tablet-based resuscitation record provided guidance that contributed to the completeness. This guidance provides support to the documenter via various clinical support features such as structured prescriptions, preset data fields, and preset documentation logic. This structural guidance was developed by consulting local emergency physicians and nurses working in the AED. The guidance mimicked the normal workflow and thus supported the documentation process. Similar results were found in a previous study in which nurses had higher confidence using the EMR when they perceived that their suggestions were used to customize the system [24]. This also implies that the tablet-based device will be considered useful if it is country, institution, and department-specific.

However, similar to other studies, our qualitative findings supported that nurses have to multitask during work, which has been shown to compromise documentation completeness [8]. Furthermore, documentation in an EMR was perceived to be more time-consuming and complex [25]. Our study participants also experienced similar concerns with slow app loading speeds and needing to navigate through various tabbed pages, which increases the complexity of documentation.

Our participants further suggested that future development of the app should include automation features that would spare the documenter from manually inputting individual data into the app. Automating data input can reduce the documenters' need to tab multiple times before finding the desired data field. This would be particularly useful for vital signs, which were the most frequently documented in this study. The UI should be flattened to facilitate input of other common data fields, which supports the concept that data fields that are more frequently recorded should be located in readily accessible spots [26]. With the automation of vital signs and an improved UI, the documenter would be able to spend more time on patient care and document in real time.

Implications for Emergency Room Nurses

Most emergency room nurses believed that the transition from paper to electronic charting can improve the quality of resuscitation documentation and patient safety. This study clearly demonstrates the potential of electronic charting to achieve that. During the transition period, the nurse documenters faced general problems with resuscitation documentation such as multitasking and unique challenges such as software updates and a subsequent need to become familiar with the app's layout. Therefore, systematic, periodic needs assessments of nurse documenters using tablet-based devices, followed by corresponding training, should be conducted. Emergency room nurses should also be actively involved in the development and implementation phases to ensure success in the transition and future development of electronic documentation. Automation functions should be considered during the development of future apps to improve documentation and redistribute more time for patient care.

Limitations

This study has some limitations. It was not able to demonstrate the effect on documentation accuracy, rates of medication errors, the quality of patient care, or the process of clinical decision-making. Furthermore, since the sample was obtained from 1 AED only, the study findings may not be generalizable to other AEDs or other acute ward settings where the staffing and workflow may be different. In addition, this study excluded trauma cases and cardiopulmonary resuscitation cases; thus, its findings cannot be generalized to all resuscitation room situations in the AED.

Conclusions

This study demonstrates that a statistically higher completion rate in 5 domains essential to resuscitation documentation was achieved with a tablet-based device than with the traditional paper resuscitation documentation. Refinement of the device should be ongoing and include consultation with the users. Further studies can expand the scope to involve all medical resuscitation cases across AEDs.

Acknowledgments

We acknowledge all the emergency room nurses who participated in the interviews in this study. This study did not receive any funding.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Documentation completeness checklist.
[PDF File (Adobe PDF File), 343 KB - mhealth_v12i1e46744_app1.pdf]

References

1. Kuah BT. Medico-legal issues in cardiopulmonary resuscitation and defibrillation. Singapore Med J 2011 Aug;52(8):642-643 [FREE Full text] [Medline: 21879227]
2. Karlsson V, Dankiewicz J, Nielsen N, Kern KB, Mooney MR, Riker RR, et al. Association of gender to outcome after out-of-hospital cardiac arrest--a report from the International Cardiac Arrest Registry. Crit Care 2015 Apr 21;19(1):182 [FREE Full text] [doi: 10.1186/s13054-015-0904-y] [Medline: 25895673]
3. ANA's Principles for Nursing Documentation: Guidance for Registered Nurses. American Nurses Association. 2010. URL: http://www.nursingworld.org/~4af4f2/globalassets/docs/ana/ethics/principles-of-nursing-documentation.pdf [accessed 2023-12-16]
4. Bergrath S, Skorning M, Rörtgen D, Beckers SK, Brokmann JC, Mutscher C, et al. Is paper-based documentation in an emergency medical service adequate for retrospective scientific analysis? An evaluation of a physician-run service. Emerg Med J 2011 Apr 15;28(4):320-324. [doi: 10.1136/emj.2009.086538] [Medline: 20713363]
5. D'Huyvetter C, Lang AM, Heimer DM, Cogbill TH. Efficiencies gained by using electronic medical record and reports in trauma documentation. J Trauma Nurs 2014;21(2):68-71. [doi: 10.1097/JTN.0000000000000031] [Medline: 24614296]
6. Geisler BP, Schuur JD, Pallin DJ. Estimates of electronic medical records in U.S. Emergency departments. PLoS One 2010 Feb 17;5(2):e9274 [FREE Full text] [doi: 10.1371/journal.pone.0009274] [Medline: 20174660]
7. Furukawa MF. Electronic medical records and the efficiency of hospital emergency departments. Med Care Res Rev 2011 Feb 16;68(1):75-95. [doi: 10.1177/1077558710372108] [Medline: 20555014]
8. Franco M, Baird J, Brown LL, Overly FL. Electronic medical record in the ED. Pediatr Emer Care 2018;34(5):303-309. [doi: 10.1097/pec.0000000000001441]
9. Macatangay GC, Caritativo MT, Mestarihi H, Dela PG, Torkzi A. Perception of emergency nurses on electronic health record (Cerner) in acute-care setting. Journal of Emergency Medicine, Trauma and Acute Care 2016:107. [doi: 10.5339/jemtac.2016.icepq.107]
10. Coffey C, Wurster LA, Groner J, Hoffman J, Hendren V, Nuss K, et al. A comparison of paper documentation to electronic documentation for trauma resuscitations at a level I pediatric trauma center. J Emerg Nurs 2015 Jan;41(1):52-56. [doi: 10.1016/j.jen.2014.04.010] [Medline: 24996509]
11. Weiss S, Fullerton L, Oglesbee S, Duerden B, Froman P. Does ambulance response time influence patient condition among patients with specific medical and trauma emergencies? Southern Medical Journal 2013;106(3):230-235. [doi: 10.1097/smj.0b013e3182882c70]
12. A&E Triage Guidelines. Hong Kong: Hospital Authority; Aug 2016:3-7.
13. Cheung A. Past, Present and Future Clinical Management System (CMS) for Hospital Authority in Hong Kong - a journey of 20+ years. Hospital Authority Convention. 2016. URL: https://www.ha.org.hk/haconvention/hac2016/proceedings/downloads/IHF1.4.pdf [accessed 2023-12-16]
14. Farley HL, Baumlin KM, Hamedani AG, Cheung DS, Edwards MR, Fuller DC, et al. Quality and safety implications of emergency department information systems. Ann Emerg Med 2013 Oct;62(4):399-407 [FREE Full text] [doi: 10.1016/j.annemergmed.2013.05.019] [Medline: 23796627]
15. Annual Plan 2019-20. Hospital Authority. 2019. URL: https://www.ha.org.hk/haho/ho/ap/AP2019-20_Eng2.pdf [accessed 2023-12-16]
16. Grundgeiger T, Albert M, Reinhardt D, Happel O, Steinisch A, Wurmb T. Real-time tablet-based resuscitation documentation by the team leader: evaluating documentation quality and clinical performance. Scand J Trauma Resusc Emerg Med 2016 Apr 16;24:51 [FREE Full text] [doi: 10.1186/s13049-016-0242-3] [Medline: 27084746]
17. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]

18. Yusoff MSB. ABC of content validation and content validity index calculation. EIMJ 2019 Jun 28;11(2):49-54. [doi: 10.21315/eimj2019.11.2.6]
19. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159. [doi: 10.2307/2529310]
20. Vasileiou K, Barnett J, Thorpe S, Young T. Characterising and justifying sample size sufficiency in interview-based studies: systematic analysis of qualitative health research over a 15-year period. BMC Med Res Methodol 2018 Nov 21;18(1):148 [FREE Full text] [doi: 10.1186/s12874-018-0594-7] [Medline: 30463515]
21. Annual Report 2017-2018. Hospital Authority. 2018. URL: https://www.ha.org.hk/ho/corpcomm/AR201718/PDF/HA_Annual_Report_2017-2018.pdf [accessed 2023-12-16]
22. Graneheim UH, Lundman B. Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. Nurse Educ Today 2004 Feb;24(2):105-112. [doi: 10.1016/j.nedt.2003.10.001] [Medline: 14769454]
23. Bilyeu P, Eastes L. Use of the electronic medical record for trauma resuscitations: how does this impact documentation completeness? J Trauma Nurs 2013;20(3):166-168. [doi: 10.1097/JTN.0b013e3182a17195] [Medline: 24005122]
24. Abu Raddaha AH. Nurses' perceptions about and confidence in using an electronic medical record system. Proceedings of Singapore Healthcare 2017 Sep 21;27(2):110-117. [doi: 10.1177/2010105817732585]
25. Callen J, Paoloni R, Li J, Stewart M, Gibson K, Georgiou A, et al. Perceptions of the effect of information and communication technology on the quality of care delivered in emergency departments: a cross-site qualitative study. Ann Emerg Med 2013 Feb;61(2):131-144. [doi: 10.1016/j.annemergmed.2012.08.032] [Medline: 23083964]
26. Jagannath S, Sarcevic A, Multak N, Myers S. Understanding paper-based documentation practices in medical resuscitations to inform the design of electronic documentation tools. Pediatr Emer Care 2018 Dec 26;37(8):e436-e442. [doi: 10.1097/pec.0000000000001676]

Abbreviations

AED: accident and emergency department
CMS: Clinical Management System
COREQ: Consolidated Criteria for Reporting Qualitative Studies
CT: computed tomography
CVI: content validity index
EMR: electronic medical record
S-CVI: scale content validity index
UA: universal agreement
UI: user interface

Edited by L Buis; submitted 06.03.23; peer-reviewed by KM Kuo, S Kim, J Walsh; comments to author 25.07.23; revised version received 19.09.23; accepted 29.11.23; published 05.01.24.

Please cite as:
Cheung K, Yip CS
Documentation Completeness and Nurses' Perceptions of a Novel Electronic App for Medical Resuscitation in the Emergency Room: Mixed Methods Approach
JMIR Mhealth Uhealth 2024;12:e46744
URL: https://mhealth.jmir.org/2024/1/e46744
doi:10.2196/46744
PMID:38180801

Original Paper

Health Care Professionals' Experiences With a Mobile Self-Care Solution for Low Complex Orthopedic Injuries: Mixed Methods Study

Jelle Spierings[1*], MD; Gijs Willinge[2*], MD; Marike Kokke[1*], MD; Bas Twigt[2*], MD, PhD; Wendela de Lange[3*], BSc; Thijs Geerdink[2*], MD, PhD; Detlef van der Velde[1*], MD, PhD; Sjoerd Repping[4*], MD, PhD; Carel Goslings[2], MD, PhD

[1]Department of Traumasurgery, St Antonius Hospital, Utrecht, Netherlands

[2]Onze Lieve Vrouwe Gasthuis, The Department of Trauma Surgery, Amsterdam, Netherlands

[3]The Healthcare Innovation Center, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, Netherlands

[4]Department of Public Health, Amsterdam UMC, Amsterdam, Netherlands

[*]these authors contributed equally

Corresponding Author:
Jelle Spierings, MD
Department of Traumasurgery
St Antonius Hospital
Soestwetering 1
Utrecht, 3543AZ
Netherlands
Phone: 31 883203000
Email: j.spierings@antoniusziekenhuis.nl

Abstract

Background: To cope with the rising number of patients with trauma in an already constrained Dutch health care system, Direct Discharge (DD) has been introduced in over 25 hospitals in the Netherlands since 2019. With DD, no routine follow-up appointments are scheduled after the emergency department (ED) visit, and patients are supported through information leaflets, a smartphone app, and a telephone helpline. DD reduces secondary health care use, with comparable patient satisfaction and primary health care use. Currently, little is known about the experiences of in-hospital health care professionals with DD.

Objective: The aim of this study was to explore the experiences of health care professionals with the DD protocol to enhance durable adoption and improve the protocol.

Methods: We conducted a mixed methods study parallel to the implementation of DD in 3 hospitals. Data were collected through a preimplementation survey, a postimplementation survey, and semistructured interviews. Quantitative data were reported descriptively, and qualitative data were reported using thematic analysis. Outcomes included the Bowen feasibility parameters: implementation, acceptability, preliminary efficacy, demand, and applicability. Preimplementation expectations were compared with postimplementation experiences. Health care professionals involved in the daily clinical care of patients with low-complex, stable injuries were eligible for this study.

Results: Of the 217 eligible health care professionals, 128 started the primary survey, 37 completed both surveys (response rate of 17%), and 15 participated in semistructured interviews. Health care professionals expressed satisfaction with the DD protocol (median 7.8, IQR 6.8-8.9) on a 10-point scale, with 82% (30/37) of participants noting improved information quality and uniformity and 73% (27/37) of patients perceiving reduced outpatient follow-up and imaging. DD was perceived as safe by 79% (28/37) of participants in its current form, but a feedback system to reassure health care professionals that patients had recovered adequately was suggested to improve DD. The introduction of DD had varying effects on workload and job satisfaction among different occupations. Health care professionals expressed intentions to continue using DD due to increased efficiency, patient empowerment, and self-management.

Conclusions: Health care professionals perceive DD as an acceptable, applicable, safe, and efficacious alternative to traditional treatment. A numerical in-app feedback system (eg, in-app communication tools or recovery scores) could alleviate health care professionals' concerns about adequate recovery and further improve DD protocols. DD can reduce health care use, which is

important in times of constrained resources. Nonetheless, both advantages and disadvantages should be considered while evaluating this type of treatment. In the future, clinicians and policy makers can use these insights to further optimize and implement DD in clinical practice and guidelines.

(JMIR Mhealth Uhealth 2024;12:e51510) doi:10.2196/51510

KEYWORDS

application; direct discharge; eHealth; experience; healthcare professional; mixed method study; orthopaedic surgery; orthopaedic; policy; policymaker; self-care application; self-care; trauma surgery; utilization; virtual fracture clinic

Introduction

The global increase in the number of patients with trauma presents a major challenge to the already strained health care systems [1,2]. To achieve more sustainable health care, digital alternatives to face-to-face outpatient follow-up have been introduced as a supported strategy [3-5]. These alternatives, also known as "eHealth," are defined by Eysenbach [6] as the intersection of medical informatics, public health, and business, referring to health services and information delivered or enhanced through the internet and related technologies. In a broader sense, the term characterizes not only a technical development but also a state of mind; a way of thinking; an attitude; and a commitment to networked, global thinking to improve health care locally, regionally, and worldwide by using information and communication technology.

Based on a British example, a Dutch teaching hospital implemented the Direct Discharge (DD) protocol to maintain the quality of care for patients with trauma in 2019 [7,8]. The DD protocol involves discharging patients from the emergency department (ED) without scheduled follow-up while providing patients with information through a self-care mobile eHealth app. DD significantly reduces secondary health care use (SHU) with similar levels of patient satisfaction and primary health care use (PHU) [9-11]. Based on these results and catalyzed by the COVID-19 pandemic, over 25 Dutch hospitals have implemented this protocol since 2019 [12].

The successful and sustainable adoption of digital health technology is complex and influenced by various factors at organizational, technological, and social levels [13,14]. This complexity is widely recognized in eHealth and eHealth evaluation frameworks [15-17]. The organizational and logistic benefits and patient satisfaction scores following DD have been well described in the literature [18-20]. However, the social aspects of the DD protocol for health care professionals remain underreported [21]. An in-depth exploration is warranted to better understand the adoption of the DD protocol within its social context, including insight into the experiences of stakeholders. The aim of this study was to explore the experiences of health care professionals with the DD protocol parallel to the implementation of this protocol in 3 Dutch hospitals to enhance durable adoption and improve the protocol.

Methods

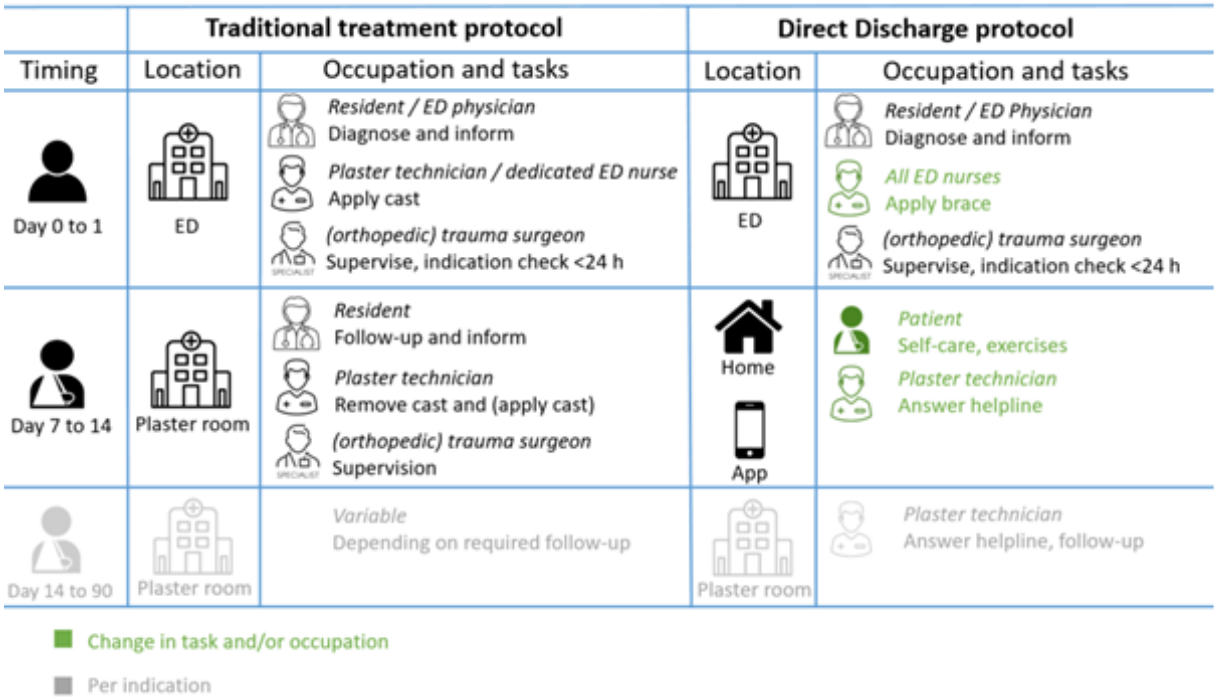
Design

An observational mixed methods study was conducted among health care professionals from August 2021 to June 2022, parallel to the implementation of DD in 3 hospitals. Both quantitative and qualitative data were collected and analyzed separately by a quantitative team (GW and JS) and a qualitative team (WDL and Elke Mathijssen). The Bowen feasibility framework was used to organize both data sources with the following parameters: implementation, acceptability, preliminary efficacy, demand, and applicability [22]. After separate analyses, quantitative and qualitative data were triangulated with the Pillar Integration Process [23]. This study was reported according to the Good Reporting on a Mixed Methods Study (GRAMMS) criteria (Multimedia Appendix 1) [24].

Context

The 3 participating centers were urban, level-2 trauma centers with up to 3 locations per hospital, treating between 1200 and 1800 patients with low-complex traumatic musculoskeletal injuries annually. Each center had a similar size and structure. All centers consisted of 3 locations, with 1 large location focusing on low-to-high complex traumatic injuries and having an ED with more rooms compared with the other locations. The 2 other locations were smaller and had no particular focus on patients with trauma, but they treated low-complex patients with trauma if they sought care at these locations. Per center, all 3 locations have 1 team taking care of all patients. These teams consisted of (orthopedic) trauma surgeons, residents, plaster technicians, ED physicians, and ED nurses. In total, 217 eligible health care professionals were exposed to DD based on data provided by participating hospitals. Changes in tasks per health care professional are described in Figure 1. These changes apply per center, including all 3 hospitals per center. The variance in the number of employees was correlated with the size of the hospital. The 2 centers implemented DD in September 2021 and 1 in March 2022.

Figure 1. Treatment protocols before and after implementation of DD and changes in location, involved stakeholders and tasks. DD: Direct Discharge; ED: emergency department.



Traditional Treatment

Before DD was implemented, patients were treated according to the local trauma protocols. These protocols consisted of immobilization or support with either a cast, sling, bandage, or splint and brief information about the injury at the ED. At least 1 outpatient follow-up appointment was scheduled at the plaster room or in the outpatient clinic within 2 weeks after the injury for review, extensive information, and definitive management planning.

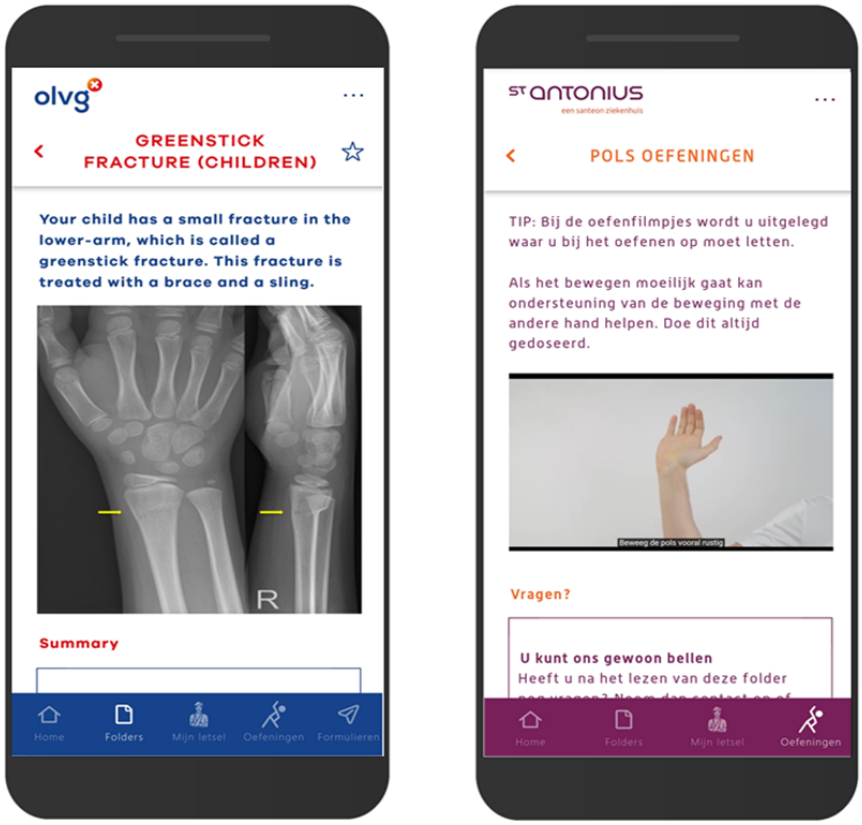
Direct Discharge Protocol

This protocol was derived from the British model of a Virtual Fracture Clinic (VFC) and adapted to the Dutch health care setting in 2019 [25]. In its Dutch adaptation, DD includes 11 treatment protocols for low-complex, stable traumatic orthopedic injuries with additional injury-related criteria (Multimedia Appendix 2) [25]. If patients met the injury-related inclusion criteria in Multimedia Appendix 2 and spoke Dutch or English fluently, they were included; no further predefined restrictions (eg, age or comorbidity) were used. Patients were excluded from the protocol at the ED if they had initial treatment in another hospital; follow-up in another hospital (eg, closer to home); multiple injuries; a reason for follow-up other than the injury (eg, social-care reasons); an eye-, motor-, or verbal-score <15 at presentation; or intoxication. With DD, patients were discharged directly from the ED without routine outpatient follow-up. They received a removable orthosis or a sling (eg, brace instead of a cast) and extensive information at the ED, summarized in a mobile self-care app (the VFC app). Patient eligibility for the protocol was re-evaluated on the next workday (within 24 hours) by a team consisting of an (orthopedic) trauma surgeon and a radiologist. If patients were incorrectly discharged directly based on the injury-related or social inclusion criteria during the second review the next day, then they were contacted by phone and scheduled for a face-to-face appointment. This re-evaluation was a standard procedure in both protocols and was used to check previous decisions of young doctors by a senior group of medical professionals based on the radiographs and electronic patient record.

The VFC App

The VFC app provides self-care assistance through information, videos, and a helpline and can be downloaded for free at the Google Play Store and iOS App Store (Figure 2). Injury-specific leaflets with recovery information, treatment rules, and red flags were included. Furthermore, frequently asked questions, audiovisual exercise, immobilization, and analgesic instructions were included to assist patients. If patients required human contact in addition to the information, a helpline by phone operated by a health care professional was available during working hours. The VFC app aimed to increase self-management and self-care during recovery and to substitute face-to-face follow-up.

Figure 2. English and Dutch in-app screenshots of the Virtual Fracture Clinic app used in the Direct Discharge protocol.



Implementation Tools and Materials

Interested centers contacted 1 of the trauma surgeons or the main email address of the initiating Onze Lieve Vrouwe Gasthuis (OLVG) Hospital, Amsterdam, the Netherlands. The OLVG Hospital created several tools to assist the implementation in other hospitals. The tools were created to streamline the implementation process as hospitals had similar questions during implementation and medical doctors struggled with the implementation of eHealth in their daily practice, partly due to inexperience with implementation and accompanying barriers and facilitators. These tools consisted of an email address that health care professionals could contact, an implementation guide, a digital PowerPoint presentation with an overview of the concept, and an information set with standardized information. This information set included posters, pocket cards, training guides, smart phrases for electronic patient records, standardized discharge letters, and a tool to personalize the layout of the VFC app. Participating centers started implementation preparations 3 months before the actual start of DD as the standard of care. Based on the experiences of the implementation in previous centers, 3 months was considered a sufficient amount of time to implement DD. There were minor differences between hospitals to optimize local fit (eg, the available hours of the staffed helpline).

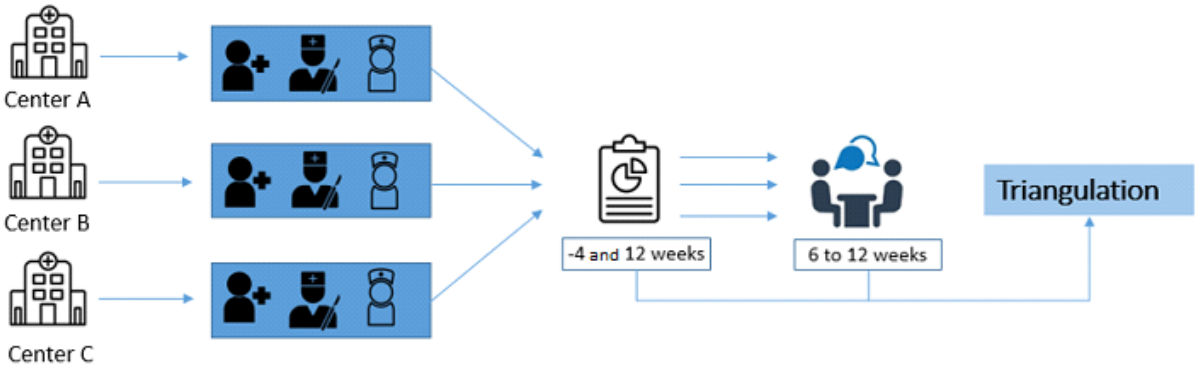
Study Population

Health care professionals involved in the in-hospital treatment of patients with eligible low-complex, stable traumatic orthopedic injuries on a daily basis were included in this study (Figure 1). Health care professionals were excluded from the final analysis if they did not provide the correct contact details or did not complete both surveys.

Sampling and Recruitment

A total of 4 weeks before the implementation of DD, the 217 potentially eligible health care professionals were asked to participate in the study through education moments, e-learnings, and by email. The health care professionals were given a survey distribution link by email to a preimplementation survey in Research Electronic Data Capture (REDCap; Vanderbilt University), a digital survey system [26]. Potentially eligible participants were remembered twice, 1 week after the initial email. Participants were excluded if they did not complete both surveys or if they did not provide any contact details to send the second survey to. Within the survey, consent for an additional semistructured interview was asked. Age, sex, occupation, medical specialty, and hospital were used to select a purposive sample among health care professionals who consented. During sample selection for the interviews, ED nurses and plaster technicians were underrepresented. Through an open call, nonresponders were recruited through email, after which the definitive sample was selected. Eligible health care professionals were contacted to schedule a web-based, semistructured interview. Health care professionals were reminded through email to complete the survey. The second survey was sent 3 months after the implementation of the DD protocol (Figure 3). We aimed to collect completed quantitative data from 100 health care professionals and qualitative data from 15 health care professionals.

Figure 3. Summary of hospitals, procedures, and models used to evaluate the Direct Discharge protocol among health care professionals. All centers consisted of 3 hospitals with similar teams and similar sizes.



Data Collection

Data were collected from surveys and semistructured interviews.

Surveys

A total of 2 surveys, a preimplementation survey and a postimplementation survey (Multimedia Appendix 3), with 46 questions, including close-ended questions, multiple-choice questions, 5-point Likert scales, visual analogue scales, and free-text questions, measured 5 Bowen feasibility parameters. As no golden standard for the evaluation of innovations exists, the surveys and topic list were developed by 4 researchers (JS, GW, BT, and TG) and checked by 2 experts on relevance: a professor in trauma surgery (CG) and an associate professor in process evaluations of health care innovations (Jaap Trappenburg). We pretested the survey with 5 health care professionals to improve clarity. After data collection, the preimplementation expectations were compared with the postimplementation experiences.

Semistructured Interviews

A total of 2 independent researchers from "The Healthcare Innovation Centre," not involved in daily clinical practice or the VFC research team, conducted digital semistructured interviews to minimize social-desirability bias. The web-based interviews were held at least 8-12 weeks after the implementation of DD to warrant an optimal recall. The interviews were guided by a topic list based on literature, including the previously mentioned Bowen feasibility parameters (Multimedia Appendix 4). The research team piloted the topic list for clarity and completeness and modified it during data collection. A pragmatic choice was made to interview 5 different types of health care professionals from each participating center, which was deemed sufficient to get a good understanding of the experiences of health care professionals and to reach saturation. No new themes were identified in the last step of the analysis, indicating that saturation was reached.

Data Analysis

Quantitative Data

Quantitative data were analyzed using SPSS (version 27; IBM Corporation) [27]. Baseline characteristics and outcomes were reported descriptively using numbers and proportions for categorical variables and mean (SD) or median (IQR) for continuous variables as appropriate. The normal distribution of continuous data was assessed with a visual analysis of boxplots. The paired 2-tailed *t* test or Mann-Whitney *U* test was used to determine the statistical significance of parametric variables for normally and nonnormally distributed data, respectively.

Qualitative Data

Qualitative data were analyzed using an inductive approach. Data analysis started after the first 5 interviews. Interviews were audiotaped, summarized, and analyzed using NVivo (version 12; QSR International) [28]. A total of 2 researchers (WDL and Elke Mathijssen) used inductive analysis with methods to ensure reliability and validity [29,30]. The data was independently analyzed by 1 researcher (WDL), and another researcher (Elke Mathijssen) reviewed the analysis. Discrepancies and remarks were discussed until they reached a consensus about the interpretations of the data. Memos were made to track research decisions during analysis. Code saturation was reached when no new categories or themes emerged from the new raw data [31]. We considered 15 interviews sufficient to reach saturation and get a good understanding of the experiences of professionals. Therefore, the number of interviews was limited by a pragmatic choice of available time. The final themes were used to describe the value and feasibility of DD from the perspective of involved health care professionals.

Triangulation

After the separate quantitative and qualitative analyses, the findings were triangulated with a simplified approach of the Pillar Integration Process technique [23]. This approach uses a transparent and rigorous 4-stage technique for integrating and presenting qualitative and quantitative findings in a joint display (Microsoft Excel, 2018; Microsoft Corporation) [32]. A researcher (JS) presented the quantitative findings per study parameter, and another presented the qualitative findings (WDL). Dissimilarities and self-contained themes were objectified. These themes were merged by 1 of the researchers (Elke Mathijssen) into a meaningful narrative (the pillar), which was reviewed by 2 researchers (JS and WDL).

Patient and Public Involvement

Patients or health care professionals were not involved in the design, intervention, research question, or outcome measures of this study.

Ethical Considerations

This study, including the process analysis, was reviewed and approved by the Medical Ethical Committee of Utrecht, Netherlands (W21.261).

Patients provided consent for participation in the research, and could opt out at any time after request by e-mail. The original consent or IRB approval covers secondary analysis without additional consent. Data is de-identified. A data key is stored at the local hospitals in a secured map and coded file. This is only accessible to JS and GW. The accessible data has been de-identified as far as possible (e.g., age in years instead of date of birth). Patients received no compensation to participate in this research.

Results

Demographics

Of the 217 estimated eligible health care professionals, 128 started the primary survey, 42 did not complete the primary survey, and 49 did not complete both surveys (Figure 4). Of the 37 included health care professionals (response rate of 17%), 23 (62%) were female, and the median age was 38 (IQR 32-45) years. Current occupations were medical specialists (14/37, 38%), residents (14/37, 38%), plaster technicians (7/37, 19%), and ED nurses (2/37, 5%; Table 1). The baseline characteristics of health care professionals who solely filled out the primary surveys did not vary statistically significantly in age (P=.98) or sex (P=.28) as compared to those who filled out both surveys. A total of 15 health care professionals, 5 per hospital, participated in the web-based, semistructured interviews, of which 60% (9/15) were female.

Figure 4. Flow diagram of the included health care professionals in the evaluation of the Direct Discharge (DD) protocol.

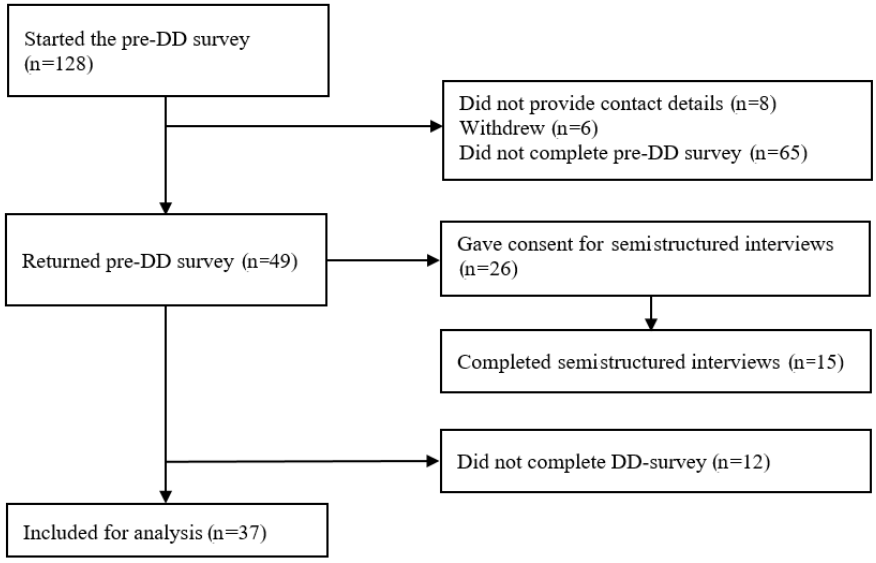


Table 1. Baseline characteristics of health care professionals included in a mixed methods evaluation of Direct Discharge.

Characteristics	Survey (n=37)	Interview (n=15)
Female, n (%)	23 (62)	9 (60)
Age (years), median (IQR)	38 (32-45)	40 (32-44)
Hospital, n (%)		
Hospital A	19 (51)	5 (33)
Hospital B	11 (30)	5 (33)
Hospital C	7 (19)	5 (33)
Current function, n (%)		
Trauma surgeon	5 (13)	3 (20)
Orthopedic surgeon	1 (3)	0 (0)
Emergency physician	8 (22)	2 (13)
Surgery resident in training	8 (22)	3 (20)
Surgery resident not in training	6 (16)	1 (7)
Plaster technician	7 (19)	3 (20)
ED[a] nurse	2 (5)	3 (20)
Total	37 (100)	16 (100)

[a]ED: emergency department.

Implementation

Qualitative data showed that the implementation strategy varied between hospitals and was adjusted to improve the local fit (eg, available hours of the helpline or brace brands). Changes in tasks, immobilization material, and the number of follow-up appointments influenced the implementation experience the most. Some nurses reported that coworkers had difficulties applying the braces, sometimes because their schooling during implementation was suboptimal or delayed, for example, because someone was ill (quote 1 in Table 2). Health care professionals reported that adequate schooling and involvement of the previous responsible health care professional in the implementation process were essential to executing the DD protocol adequately. Mainly because the ED nurse, now responsible for applying the braces, had less experience immobilizing fractures than the plaster technician.

Table 2. Health care professionals' quotes and associated feasibility parameters.

Theme and quote number	Quote	Health care professional
Implementation of Direct Discharge among health care professionals		
1	Some of my colleagues have difficulty with the materials. How does it work and what goes where? It takes a bit longer for some of them to get the hang of it. Having to learn so many new things sometimes causes resistance.	ED[a] nurse
Acceptability of the Direct Discharge among health care professionals		
2	This is a significant improvement for patients and appeals to their autonomy and control, as well as their own influence on the healing process. I believe it is motivating and in line with the current times.	ED nurse
3	I no longer have to do these routine outpatient clinical check-ups. I could only provide limited contributions besides providing information, allowing me to have more peace and tranquility in the clinic. I can use that time for other patients to add more value.	Trauma surgeon
4	Both among doctors and nurses, DD has been widely embraced and well implemented, but for both professions, it requires valuable extra minutes due to additional explanation. Currently, the workload is very high.	Resident
5	But if that care is taken away from us, I do believe that we have a responsibility towards the patients to ensure the proper transition of that care.	Plaster technician
Preliminary efficacy of the Direct Discharge among health care professionals		
6	We now have a tool in our hands to change healthcare without it deteriorating, which convinces people who tended towards over-treatment.	ED physician
7	The quality is not affected, assuming the doctor was already good. It is mainly more efficient. Information provision has improved. It has become more modern. I think DD is not worse, but we're not certain yet.	Trauma surgeon
8	We are sometimes called about 2-3 times per day on the fracture line. I don't think that's a bad score.	Plaster technician
Demand for Direct Discharge among health care professionals		
9	I was already familiar with DD because I was looking for good and reliable information on injuries during my training as an ED nurse. I came across the article published by OLVG and started using the app in my work.	ED nurse
10	Every day, a few patients are treated through the app. I am starting to notice the reduction in daily practice!	Trauma surgeon
11	In my work, it has changed that we see less patients, but new things have also been added. However, we no longer see minor injuries. The easier type of care has decreased a bit.	Plaster technician
Applicability of the Direct Discharge among health care professionals		
12	I see the advantage of a brace instead of a cast, a great improvement. I would also prefer DD myself.	Resident
13	The walking boot is difficult to fit, which poses a risk of misuse. This results in most complications being caused by misuse. What does that do to the recovery process?	ED physician

[a]ED: emergency department.

Acceptability

Overview

Quantitative data showed a median satisfaction with treatment of 7.8 (IQR 6.8-8.9) on a 10-point scale. This finding complemented qualitative data, as most health care professionals were satisfied with DD (quote 2 in Table 2). Health care professionals suggested that in-app patient feedback, such as pain scores or patient-reported outcome measures, could further increase health care professional satisfaction levels and address health care professionals' possible concerns about adequate recovery. Some health care professionals were hesitant about DD due to its novelty, limited education during implementation, changes in work activities, and concerns about the short- and long-term outcomes of certain injuries.

VFC App Acceptability

The introduction of the VFC app at the ED yielded both advantages and disadvantages. Both qualitative and quantitative data showed that introducing the VFC app and treatment information in the ED required more time than the traditional treatment (19/37, 52%; Multimedia Appendix 5). Qualitative data showed that this time decreased once DD was fully implemented but remained longer than traditional treatment. Despite the increased time to inform patients, the perceived reduction in logistics at the outpatient clinic was perceived as more valuable (quotes 3 and 4 in Table 2). Both data sources underline the benefits of the app in terms of uniform, on-demand, and adequate information for patients after discharge. The downsides of the VFC app include less personal health care and the current limited language availability (Dutch and English).

Workload and Job Satisfaction

Qualitative data showed that workload and job satisfaction decreased slightly after the introduction of DD. Quantitative data did not support this, as no statistically significant differences between expectations and experiences were found regarding workload (P=.37) and increased job satisfaction (P=.42). Plaster technicians reported that the introduction of DD has led to losing a "fun" part of their job. Some reported they could provide less service for patients and felt responsible for educating ED nurses who had less experience with the immobilization of fractures compared with them (quote 5 in Table 2).

Preliminary Efficacy

Quality of Care

Quantitative data showed that the quality of care with the DD protocol is perceived as comparable to traditional treatment (25/37, 67%). Health care professionals reported an improvement in the quality of information and uniformity (30/37, 82%; Multimedia Appendix 6). No statistically significant differences were found between expectations and experiences of quality of care (P=.86), quality of information (P=.42), and quantity of information (P=.18). Qualitative data supported these findings, highlighting the benefits of uniform, injury-specific information (quotes 6 and 7 in Table 2). In both the survey and interviews, most (27/37, 73%) health care professionals reported a reduction in outpatient follow-up and injury-related imaging. Qualitative data showed that (orthopedic) trauma surgeons experienced the reduction as beneficial, whereas some plaster technicians experienced the reduction as a disadvantage. Most residents and ED physicians reported that the logistical benefits at the outpatient clinic outweighed the slight increase in time at the ED.

Perceived Safety

Both data sources showed that most (29/37, 79%) health care professionals perceived DD as safe and that sufficient scientific evidence exists to treat patients safely. Health care professionals assumed that patients had fully recovered if they did not contact the hospital again. Nevertheless, they proposed a numerical feedback system in the VFC app to ensure adequate recovery and alleviate concerns about the poor long-term functional outcomes of their patients. The frequency of helpline use was low (≤5 times per week; quote 8 in Table 2). Some residents reported that the introduction of DD decreased their exposure to low-complex traumatic injuries, which might influence their learning curve in the future. Furthermore, the lack of follow-up introduced the tendency for some residents to be more explanatory at the ED. However, some residents stated that during crowding at the ED, they limited their explanation to downloading the VFC app with minimal instructions. The frequency of calling varied per injury, with patients with a greenstick fracture rarely requiring contact. Reasons to call the helpline were similar among the 3 centers. Patient questions were related to a poor recall of the ED visit, suboptimal information provision at the ED, doubts about their recovery, or insufficient reading of the app's content. The daily multidisciplinary radiologic evaluation and helpline were considered effective safety nets.

Demand

Qualitative analysis revealed that some health care professionals had previous knowledge or experience with the DD protocol (quote 9 in Table 2). The COVID-19 pandemic positively influenced their perceptions of digitally assisted care. Some health care professionals reported that it fits the general demand to develop more efficient outpatient follow-up models and that the DD protocol is an example of "tomorrow's health care." Reduced outpatient follow-up, hospital use (eg, treatment rooms or parking lots), and staff were mentioned as benefits. Health care professionals expected the DD protocol to stay and intended to continue using it. Health care professionals reported that the DD protocol could improve patient empowerment and enhance self-management and independence, especially among younger patients. Health care professionals had different experiences integrating the DD protocol into their daily activities. Orthopedic trauma surgeons perceived a decrease in patients at the outpatient clinic following DD implementation and stated that their workload was starting to decrease due to DD (quote 10 in Table 2). Plaster technicians also reported a decrease in patients. However, they perceived this sometimes as a disadvantage as they no longer treated these low-complex injuries, which is considered a loss of a fun part of their job. No reduction in workload was perceived among plaster technicians due to additional tasks and an increase in a patient population with more complex injuries (quote 11 in Table 2).

Residents reported that extensive information has changed from the ED and outpatient clinic to solely the ED. The ED nurses reported applying fewer casts and more orthoses, such as braces or walking boots.

Applicability

Before implementation, some health care professionals expressed concerns about the incorrect use of immobilization materials. However, after implementation, the types of immobilization materials were perceived as adequate, with 74% (27/37) of health care professionals finding the braces easy to use for patients (Multimedia Appendix 7). The braces were less immobilizing than a cast, which was perceived as a benefit (quote 12 in Table 2). The less immobilizing treatment regimens could result in an earlier return to full function. However, some were concerned that the braces could lead to inferior long-term functional outcomes, even though the scientific evidence for these concerns was lacking (quote 13 in Table 2).

Discussion

Principal Findings

Dutch health care professionals considered the DD protocol a safe and satisfactory alternative to traditional treatment, leading to a perceived decrease in SHU. Although providing information with DD required more (explanation) time for residents and physicians at the ED, the logistical benefits (ie, reduced number of follow-up appointments) outweighed the slight increase in time. Before and after implementation, no statistically significant differences were found regarding workload or job satisfaction.

However, qualitative data reported benefits in workload and job satisfaction for trauma surgeons and residents, and disadvantages for some plaster technicians in terms of job satisfaction and workload. Furthermore, health care professionals reported increased quality of information and comparable quality of care. Almost all health care professionals would like to continue using DD after implementation. To improve DD, several new app functionalities were suggested, and the earlier involvement of stakeholders who performed new tasks was suggested.

Comparison With Existing Literature

The study results show that DD is an acceptable alternative to traditional follow-up for health care professionals. Health care professionals reported similar, high satisfaction scores with treatment as previously reported patient satisfaction scores for similar protocols [18,33]. To further improve health care professionals' satisfaction scores, health care professionals in this study suggested a short numerical feedback system to monitor injury recovery remotely (eg, recovery questions after 3 months). In the literature, health care professionals have also mentioned this as an important feature of eHealth developed for patients with musculoskeletal injuries [34,35]. A challenge to incorporating this is that these functions would require a more enhanced app that complies with current laws and regulations for data storage and requires substantial financial investment. In addition, this request might also be a sign of early-stage adoption, where health care professionals feel a bit uncertain about the patients they would normally see for follow-up but who are now out of sight. As time progresses, it seems likely they will feel more comfortable not seeing these patients anymore, as this is then considered standard of care.

Preliminary efficacy was partly in line with the literature [18,20,33]. Most health care professionals considered DD a safe alternative to face-to-face follow-up, leading to a perceived reduction of SHU. This finding is consistent with previous articles, which concluded that DD is safe based on low, comparable complication numbers, with significantly reduced SHU compared to traditional treatment [9,18,33]. The perceived reduction in SHU varied among stakeholders. The reduction was perceived as most beneficial in qualitative data regarding workload for (orthopedic) trauma surgeons and least beneficial for plaster technicians, as they enjoyed this particular part of their workload. This study has been unable to demonstrate statistically significant differences regarding workload and job satisfaction before and after implementation. Even though the introduction of DD at the ED increased valuable treatment time for some stakeholders, which decreased but remained longer after implementation, almost all reported that the benefits of less SHU outweighed this downside. These findings are interesting, as the increase in time has been reported as one of the most important personal barriers to implementing and adopting eHealth but was not reported as an important disadvantage in this study [14,35].

Similar quality of care and increased quality of information after the introduction of DD are in accordance with earlier findings [11,20,36]. Particular advantages of DD were uniform, injury-specific, and on-demand information, consistent with a study pointing out the current heterogeneity in treatments for these injuries and the demand for uniform treatment [37]. The disadvantages of DD were the care being less personal, the loss of care activities for several stakeholders, and the inability to monitor recovery. A previous study described that VFCs might influence learning curves as residents become less experienced with the follow-up and full recovery of these injuries [38]. This was not supported by our results. However, a lack of follow-up has led to a tendency for some (often inexperienced) residents to be more explanatory at the ED to ensure the patient had received all information during the only contact moment. To reassure themselves and identify the small group of patients with persistent complaints, some health care professionals suggested a feedback system to ensure they had recovered adequately. Such a system should not increase workload but only filter patients with remaining questions, pain, or complications. This should be developed with patients and health care professionals and could contain anchor-based questions or patient-reported outcome measures with predefined cut-off values.

Health care professionals report that DD fits the public demand to develop more efficient (outpatient) care, and digital assistance could help deal with constrained resources. This partly aligns with the literature, as attitudes toward the usefulness of eHealth vary [39-41]. Almost all health care professionals would like to continue using DD after its implementation. This finding complements current literature showing the widespread implementation of DD protocols and VFCs in the United Kingdom, Australia, New Zealand, and India [8,33,42,43]. This further strengthens the idea of at least adequate acceptance and the general applicability of this concept in different countries. The applicability of DD varied among stakeholders and was influenced by the increase or loss of tasks after implementation of the protocol and their experience with that particular increase or loss of tasks. Even though evaluation of applicability for health care professionals and health care professional satisfaction with the protocol seems vital to determining feasibility, it has not been reported previously for DD protocols, despite the many centers that have implemented similar protocols. Health care professionals reported that early stakeholder involvement during implementation could be beneficial to cope with the changes in tasks and transfer of knowledge, which aligned with the results of Logishetty [21] reporting the importance of early stakeholder involvement during VFC implementation in a quality improvement approach. The lack of early involvement of stakeholders and limited feasibility among health care professionals are known risks for nonadoption or abandonment [3,44,45].

Strengths and Limitations

This study has several strengths. To date, this is the first study to explore the in-depth views of several stakeholders involved in DD protocols and VFCs. The COVID-19 pandemic has accelerated the implementation of DD in the Netherlands, emphasizing the demand for a shift in patient care, where eHealth alternatives have become the new standard [46]. The views and perceptions of health care professionals about different eHealth services are valuable to further tailor these services to their needs and preferences. A second strength of

this study is the multidisciplinary involvement of the research team during evaluation and analysis. This approach ensured that the data were analyzed from all possible angles. A third strength is a mixed methods approach to evaluating DD because the separate collection of both data sources combined with the triangulation increased the likelihood of our results being a realistic representation of the daily clinical activities of the involved health care professionals. The fourth strength is the use of a validated framework to evaluate DD, which allowed for a structured insight into each feasibility parameter. The fifth strength was the heterogeneity in the sizes and locations of the three study sites. This allowed us to investigate the applicability within different types of hospitals.

This study also has limitations that need to be addressed. The first limitation is the small quantitative sample size and low response rate. Even though the response rate among health care professionals is 10% to 15% lower compared to patient studies, the quantitative sample size remains limited [47]. Nevertheless, for almost all parameters, quantitative and qualitative data were similar, indicating a realistic response from most health care professionals. A third limitation is that ED nurses were not involved in the initial design of this study and did not provide quantitative data. Nevertheless, they have been included in the qualitative data to strengthen the study results.

Implications for Clinicians and Policy Makers and Future Perspectives

DD protocols reduce SHU without negatively influencing PHU, satisfaction, complications, or functional outcomes [17]. By assisting patients through the VFC app, they can receive care at home, potentially reducing health care costs [9]. Health care costs in the Netherlands have increased drastically, particularly due to specialized in-hospital care. In addition, there is a growing shortage of personnel. Digitally assisted solutions are suggested to cope with the rise of these costs and limited personnel, but implementation remains difficult [14]. The technical (eg, compliance and complication numbers) and logistic outcomes of this concept are well explored, but the social and cultural elements of DDs have not been explored. These findings help optimize future implementation strategies for eHealth in orthopedic and trauma surgery by providing preconditions and learning lessons such as early stakeholder involvement. These findings might be generalizable for other short treatment processes in other (surgical) departments with high volumes of relatively low-complex surgical patients (eg, low-complex dermatology or otorhinolaryngology).

As DD is introduced rapidly to cope with constrained financial and human resources and health care professionals expect DD to stay, a thorough evaluation of caregivers and patients is essential to ensuring sustainable adoption. Results of this study have improved our DD protocol and have led to 2 corresponding major points of improvement among all stakeholders and many points of improvement, such as ideas for illustrations, adjusted language (levels), and adjustable font size from 1 or more health care professionals. Future studies should focus on co-designing numerical feedback with patients and health care professionals. Furthermore, a thorough evaluation of patient perspectives should be performed to gain insight into the end users' perspective on this innovation. Particularly to investigate the potential link between different levels of literacy (health literacy, digital literacy, and literacy) and health equity, as suggested in previous studies [48,49].

Conclusion

Health care professionals perceive DD as an acceptable, applicable, safe, and efficacious alternative to traditional treatment. A short numerical feedback system could alleviate concerns about a full recovery and further improve DD protocols. DD can reduce SHU, which is important in times of constrained resources. Nonetheless, both advantages and disadvantages should be considered while evaluating this type of protocol. In the future, clinicians and policy makers can use these insights to further optimize and implement DD and VFC in clinical practice and guidelines.

Conflicts of Interest

The institutions of the authors affiliated with St Antonius Hospital and Onze Lieve Vrouwe Gasthuis (OLVG) Hospital have received funding to support this work through an unrestricted grant from the Dutch organization for health care sciences and health care innovation, ZonMw, the Hague, Netherlands (516012524).

Multimedia Appendix 1
Assessment of Good Reporting on a Mixed Methods Study (GRAMMS) criteria for the study.
[DOCX File , 24 KB - mhealth_v12i1e51510_app1.docx]

Multimedia Appendix 2
Additional criteria and immobilization for low-complex, traumatic orthopedic injuries included in the Direct Discharge Protocol.
[DOCX File , 15 KB - mhealth_v12i1e51510_app2.docx]

Multimedia Appendix 3
Surveys used to evaluate Direct Discharge among health care professionals.
[DOCX File , 140 KB - mhealth_v12i1e51510_app3.docx]

Multimedia Appendix 4

Topic list for health care professionals the Direct Discharge protocol.
[DOCX File , 24 KB - mhealth_v12i1e51510_app4.docx]

Multimedia Appendix 5
5-Point Likert scale regarding applicability of the Direct Discharge among health care professionals.
[DOCX File , 42 KB - mhealth_v12i1e51510_app5.docx]

Multimedia Appendix 6
5-Point Likert scale regarding preliminary efficacy of the Direct Discharge among health care professionals.
[DOCX File , 63 KB - mhealth_v12i1e51510_app6.docx]

Multimedia Appendix 7
5-Point Likert scale regarding applicability of the Direct Discharge among health care professionals.
[DOCX File , 42 KB - mhealth_v12i1e51510_app7.docx]

References

1. Goslings C, Gorzeman M, Offeringa-Klooster M, Berdowski J. Brandbrief 'regionale spoedzorg - de rek is er uit'. Wetenschappelijk Onderzoek Verkeersveiligheids. 2016. URL: https://swov.nl/nl/publicatie/brandbrief-regionale-spoedzorg-de-rek-er-uit [accessed 2023-12-30]
2. GBD 2019 Fracture Collaborators. Global, regional, and national burden of bone fractures in 204 countries and territories, 1990-2019: a systematic analysis from the global burden of disease study 2019. Lancet Healthy Longev 2021;2(9):e580-e592 [FREE Full text] [doi: 10.1016/S2666-7568(21)00172-0] [Medline: 34723233]
3. Van Velthoven MH, Cordon C. Sustainable adoption of digital health innovations: perspectives from a stakeholder workshop. J Med Internet Res 2019;21(3):e11922 [FREE Full text] [doi: 10.2196/11922] [Medline: 30907734]
4. van den Heuvel D. Zorg op afstand dichterbij? The Dutch Council for Health and Society. 2020. URL: https://mijn.bsl.nl/zorg-op-afstand/708328?fulltextView=true&doi=10.1007%2F978-90-313-9233-9_16 [accessed 2023-12-30]
5. Samenvatting Integraal Zorgakkoord. Ministry of Health Welfare and Sport. 2022. URL: https://www.rijksoverheid.nl/documenten/publicaties/2022/09/16/samenvatting-integraal-zorgakkoord [accessed 2023-12-30]
6. Eysenbach G. What is e-health? J Med Internet Res 2001;3(2):E20 [FREE Full text] [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]
7. O' Reilly M, Breathnach O, Conlon B, Kiernan C, Sheehan E. Trauma assessment clinic: virtually a safe and smarter way of managing trauma care in Ireland. Injury 2019;50(4):898-902. [doi: 10.1016/j.injury.2019.03.046] [Medline: 30955873]
8. Brogan K, Bellringer S, Akehurst H, Gee C, Ibrahim N, Cassidy L, et al. Virtual fracture clinic management of fifth metatarsal, including Jones', fractures is safe and cost-effective. Injury 2017;48(4):966-970. [doi: 10.1016/j.injury.2017.02.003] [Medline: 28284470]
9. Geerdink TH, Spierings JF, Schuijt HJ, Eversdijk HAJ, van Dongen JM, Kokke MC, et al. Healthcare utilization and satisfaction with treatment before and after direct discharge from the emergency department of simple stable musculoskeletal injuries in the Netherlands. Eur J Trauma Emerg Surg 2022;48(3):2135-2144 [FREE Full text] [doi: 10.1007/s00068-021-01835-5] [Medline: 34997258]
10. Geerdink TH, Geerdink NJ, van Dongen JM, Haverlag R, Goslings JC, van Veen RN, et al. Cost-effectiveness of direct discharge from the emergency department of patients with simple stable injuries in the Netherlands. Trauma Surg Acute Care Open 2021;6(1):e000763 [FREE Full text] [doi: 10.1136/tsaco-2021-000763] [Medline: 34722930]
11. Little M, Huntley D, Morris J, Jozsa F, Hardman J, Anakwe RE. The virtual fracture clinic improves quality of care for patients with hand and wrist injuries: an assessment of 3709 patients. J Hand Surg Eur Vol 2020;45(7):748-753 [FREE Full text] [doi: 10.1177/1753193420930610] [Medline: 32539577]
12. Virtual fracure care implemented in 25th Dutch hospital. LinkedIn. 2023. URL: https://nl.linkedin.com/in/virtual-fracture-care-0b4174232 [accessed 2023-12-30]
13. Lennon MR, Bouamrane MM, Devlin AM, O'Connor S, O'Donnell C, Chetty U, et al. Readiness for delivering digital health at scale: lessons from a longitudinal qualitative evaluation of a national digital health innovation program in the United Kingdom. J Med Internet Res 2017;19(2):e42 [FREE Full text] [doi: 10.2196/jmir.6900] [Medline: 28209558]
14. Schreiweis B, Pobiruchin M, Strotbaum V, Suleder J, Wiesner M, Bergh B. Barriers and facilitators to the implementation of ehealth services: systematic literature analysis. J Med Internet Res 2019;21(11):e14197 [FREE Full text] [doi: 10.2196/14197] [Medline: 31755869]
15. Odendaal W, Goudge J, Griffiths F, Tomlinson M, Leon N, Daniels K. Healthcare workers' perceptions and experiences on using mHealth technologies to deliver primary healthcare services: a qualitative evidence synthesis. Cochrane Database Syst Rev 2015;2015(11):CD011942 [FREE Full text] [doi: 10.1002/14651858.CD011942] [Medline: 27478408]
16. Dehzad F, Hilhorst C, de Bie C, Claassen E. Adopting health apps, what's hindering doctors and patients? Health 2014;06(16):49627 [FREE Full text] [doi: 10.4236/health.2014.616256]

17. Jacob C, Lindeque J, Klein A, Ivory C, Heuss S, Peter MK. Assessing the quality and impact of eHealth tools: systematic literature review and narrative synthesis. JMIR Hum Factors 2023;10:e45143 [FREE Full text] [doi: 10.2196/45143] [Medline: 36843321]
18. Geerdink TH, Verbist J, van Dongen JM, Haverlag R, van Veen RN, Goslings JC. Direct discharge of patients with simple stable musculoskeletal injuries as an alternative to routine follow-up: a systematic review of the current literature. Eur J Trauma Emerg Surg 2022;48(4):2589-2605 [FREE Full text] [doi: 10.1007/s00068-021-01784-z] [Medline: 34529086]
19. Davey MS, Coveney E, Rowan F, Cassidy JT, Cleary MS. Virtual fracture clinics in orthopaedic surgery - a systematic review of current evidence. Injury 2020;51(12):2757-2762. [doi: 10.1016/j.injury.2020.11.001] [Medline: 33162011]
20. Murphy EP, Fenelon C, Murphy RP, O'Sullivan MD, Pomeroy E, Sheehan E, et al. Are virtual fracture clinics during the COVID-19 pandemic a potential alternative for delivering fracture care? a systematic review. Clin Orthop Relat Res 2020;478(11):2610-2621 [FREE Full text] [doi: 10.1097/CORR.0000000000001388] [Medline: 32657810]
21. Logishetty K. Adopting and sustaining a virtual fracture clinic model in the district hospital setting - a quality improvement approach. BMJ Qual Improv Rep 2017;6(1):u220211.w7861 [FREE Full text] [doi: 10.1136/bmjquality.u220211.w7861] [Medline: 28243440]
22. Bowen DJ, Kreuter M, Spring B, Cofta-Woerpel L, Linnan L, Weiner D, et al. How we design feasibility studies. Am J Prev Med 2009;36(5):452-457 [FREE Full text] [doi: 10.1016/j.amepre.2009.02.002] [Medline: 19362699]
23. Johnson RE, Grove AL, Clarke A. Pillar integration process: a joint display technique to integrate data in mixed methods research. J Mix Methods Res 2017;13(3):301-320 [FREE Full text] [doi: 10.1177/1558689817743108]
24. O'Cathain A, Murphy E, Nicholl J. The quality of mixed methods studies in health services research. J Health Serv Res Policy 2008;13(2):92-98 [FREE Full text] [doi: 10.1258/jhsrp.2007.007074] [Medline: 18416914]
25. Geerdink TH, Haverlag R, van Veen RN, Bouwmeester OVA, Goslings JC. Direct discharge from the ED for patients with simple stable injuries: a Dutch pilot study. Article in Dutch. Ned Tijdschr Geneeskd 2020;164:D4604. [Medline: 32391996]
26. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research Electronic Data Capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
27. IBM SPSS Statistics for Windows, Version 27.0. IBM. 2020. URL: https://www.ibm.com/support/pages/downloading-ibm-spss-statistics-27 [accessed 2023-12-30]
28. QSR International Pty Ltd. NVivo. 2020. URL: https://support.qsrinternational.com/s/ [accessed 2023-12-30]
29. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77-101 [FREE Full text] [doi: 10.1191/1478088706qp063oa]
30. Boeije HR. Analysis in Qualitative Research. 2009. URL: https://www.researchgate.net/publication/50205377_Analysis_in_Qualitative_Research [accessed 2023-12-30]
31. Morse JM, Barrett M, Mayan M, Olson K, Spiers J. Verification strategies for establishing reliability and validity in qualitative research. Int J Qual Methods 2016;1(2):13-22 [FREE Full text] [doi: 10.1177/160940690200100202]
32. Microsoft Excel. Microsoft Corporation. 2018. URL: https://office.microsoft.com/excel [accessed 2023-12-30]
33. Khan SA, Asokan A, Handford C, Logan P, Moores T. How useful are virtual fracture clinics?: a systematic review. Bone Jt Open 2020;1(11):683-690 [FREE Full text] [doi: 10.1302/2633-1462.111.BJO-2020-0107.R1] [Medline: 33263108]
34. Reinecke F, Dittrich F, Dudda M, Stang A, Polan C, Müller R, et al. Acceptance, barriers, and future preferences of mobile health among patients receiving trauma and orthopedic surgical care: paper-based survey in a prospective multicenter study. JMIR Mhealth Uhealth 2021;9(4):e23784 [FREE Full text] [doi: 10.2196/23784] [Medline: 33881401]
35. Houwen T, Vugts MAP, Lansink KWW, Theeuwes HP, Neequaye N, Beerekamp MSH, et al. Developing mHealth to the context and valuation of injured patients and professionals in hospital trauma care: qualitative and quantitative formative evaluations. JMIR Hum Factors 2022;9(2):e35342 [FREE Full text] [doi: 10.2196/35342] [Medline: 35723928]
36. Cavka B, Cross E, Montvida O, Plunkett G, Oppy A, Bucknill A, et al. Retrospective cohort study evaluating the efficacy and safety of an orthopaedic consultant-led virtual fracture clinic in an Australian level 1 trauma centre. ANZ J Surg 2021;91(7-8):1441-1446 [FREE Full text] [doi: 10.1111/ans.16574] [Medline: 33459513]
37. Geerdink TH, Uijterwijk BA, Meijer DT, Sierevelt IN, Mallee WH, van Veen RN, et al. Adoption of direct discharge of simple stable injuries amongst (orthopaedic) trauma surgeons. Injury 2021;52(4):774-779 [FREE Full text] [doi: 10.1016/j.injury.2020.11.026] [Medline: 33276960]
38. McAuliffe O, Lami M, Lami T. The impact of virtual fracture clinics on medical education - a medical student perspective. Med Educ Online 2016;21:30950 [FREE Full text] [doi: 10.3402/meo.v21.30950] [Medline: 26983936]
39. Oakes AH, Patel MS. A nudge towards increased experimentation to more rapidly improve healthcare. BMJ Qual Saf 2020;29(3):179-181 [FREE Full text] [doi: 10.1136/bmjqs-2019-009948] [Medline: 31744847]
40. Granath A, Eriksson K, Wikström L. Healthcare workers' perceptions of how eHealth applications can support self-care for patients undergoing planned major surgery. BMC Health Serv Res 2022;22(1):844 [FREE Full text] [doi: 10.1186/s12913-022-08219-4] [Medline: 35773687]
41. Kerst A, Zielasek J, Gaebel W. Smartphone applications for depression: a systematic literature review and a survey of health care professionals' attitudes towards their use in clinical practice. Eur Arch Psychiatry Clin Neurosci 2020;270(2):139-152 [FREE Full text] [doi: 10.1007/s00406-018-0974-3] [Medline: 30607530]

42. Iyengar KP, Jain VK, Nallakumarasamy A, Jeyaraman M. Virtual fracture clinic model in India: a technological innovation. Indian J Orthop 2023;57(1):1-6 [FREE Full text] [doi: 10.1007/s43465-022-00763-9] [Medline: 36660488]
43. Fuschini C, Bussoletti T, Shaw C, Shazhad M, Qi L, Cleary A. Validation of the primary care pathway model for management of orthopedic injuries: results of a prospective, queensland study. J Prim Care Community Health 2020;11:2150132720967228 [FREE Full text] [doi: 10.1177/2150132720967228] [Medline: 33106089]
44. Tossaint-Schoenmakers R, Versluis A, Chavannes N, Talboom-Kamp E, Kasteleyn M. The challenge of integrating eHealth into health care: systematic literature review of the donabedian model of structure, process, and outcome. J Med Internet Res 2021;23(5):e27180 [FREE Full text] [doi: 10.2196/27180] [Medline: 33970123]
45. De Leeuw JA, Woltjer H, Kool RB. Identification of factors influencing the adoption of health information technology by nurses who are digitally lagging: in-depth interview study. J Med Internet Res 2020;22(8):e15630 [FREE Full text] [doi: 10.2196/15630] [Medline: 32663142]
46. Keuper J, Batenburg R, Verheij R, van Tuyl L. Use of e-health in Dutch general practice during the COVID-19 pandemic. Int J Environ Res Public Health 2021;18(23):12479 [FREE Full text] [doi: 10.3390/ijerph182312479] [Medline: 34886204]
47. Meyer VM, Benjamens S, Moumni ME, Lange JFM, Pol RA. Global overview of response rates in patient and health care professional surveys in surgery: a systematic review. Ann Surg 2022;275(1):e75-e81 [FREE Full text] [doi: 10.1097/SLA.0000000000004078] [Medline: 32649458]
48. Latulippe K, Hamel C, Giroux D. Social health inequalities and eHealth: a literature review with qualitative synthesis of theoretical and empirical studies. J Med Internet Res 2017;19(4):e136 [FREE Full text] [doi: 10.2196/jmir.6731] [Medline: 28450271]
49. Neter E, Brainin E. eHealth literacy: extending the digital divide to the realm of health information. J Med Internet Res 2012;14(1):e19 [FREE Full text] [doi: 10.2196/jmir.1619] [Medline: 22357448]

Abbreviations

DD: Direct Discharge
ED: emergency department
GRAMMS: Good Reporting on a Mixed Methods Study
OLVG: Onze Lieve Vrouwe Gasthuis
PHU: primary health care use
REDCap: Research Electronic Data Capture
SHU: secondary health care use
VFC: Virtual Fracture Clinic

Edited by L Buis; submitted 02.08.23; peer-reviewed by F Weber, D Frost; comments to author 22.09.23; revised version received 24.10.23; accepted 03.11.23; published 02.02.24.

Please cite as:
Spierings J, Willinge G, Kokke M, Twigt B, de Lange W, Geerdink T, van der Velde D, Repping S, Goslings C
Health Care Professionals' Experiences With a Mobile Self-Care Solution for Low Complex Orthopedic Injuries: Mixed Methods Study
JMIR Mhealth Uhealth 2024;12:e51510
URL: https://mhealth.jmir.org/2024/1/e51510
doi:10.2196/51510
PMID:38306162

Original Paper

The Effectiveness of a Cell Phone eHealth App in Changing Knowledge, Stigmatizing Attitudes, and Intention to Seek Help Associated With Obsessive-Compulsive Disorder: Pilot Questionnaire Study

Antonio Chaves[1], PhD; Sandra Arnáez[2], PhD; Gemma García-Soriano[2], PhD

[1]Departamento de Orientación Educativa, IES Cid Campeador, Conselleria d'Educació, Cultura i Esport, Valencia, Spain

[2]Departamento de Personalidad, Evaluación y Tratamientos Psicológicos, Universitat de València, Valencia, Spain

Corresponding Author:
Gemma García-Soriano, PhD
Departamento de Personalidad, Evaluación y Tratamientos Psicológicos
Universitat de València
Avda Blasco Ibáñez, 21
Valencia, 46010
Spain
Phone: 34 963983389
Email: gemma.garcia@uv.es

Abstract

Background: Obsessive-compulsive disorder (OCD) is a disabling disorder associated with high interference in people's lives. However, patients with OCD either do not seek help or delay seeking help. Research suggests that this could be explained by poor mental health literacy about the disorder and the associated stigma.

Objective: This study aims to evaluate the feasibility, acceptability, and preliminary effectiveness of a mental health mobile app, esTOCma, developed to improve knowledge about OCD and its treatment, increase help-seeking intention, and reduce stigmatizing attitudes and social distance associated with OCD.

Methods: We used preintervention, postintervention, and 3-month follow-up assessments in this single-arm pilot intervention. Overall, 90 participants were recruited from the community using the snowball sampling method. We used esTOCma to defeat the "stigma monster" over the course of 10 missions. The participants completed the sociodemographic information and Obsessive-Compulsive Inventory–Revised at preassessment and an acceptability questionnaire at postassessment. All other measures were completed at the preassessment, postassessment, and 3-month follow-up (ie, the Spanish Mental Illness Stigma Attribution Questionnaire–27, the General Help-Seeking Questionnaire, the Social Distance Scale, and the Mental Health Literacy Questionnaire).

Results: Of the 90 participants from the community that were assessed for eligibility, 86% (n=78) were allocated to intervention. Of these 78 participants, 79% (n=62) completed the game and answered the postintervention assessment (completer group). Overall, 69% (43/62) of the participants also completed the 3-month follow-up assessment. The participants completing the study were older (P=.003) and had a higher baseline knowledge of OCD (P=.05). The participants took an average of 13.64 (SD 10.50) days to complete the intervention, including the pre- and postassessments. The participants spent an average of 4.56 (SD 3.33) days completing the 10 missions included in the app. Each mission took a mean of between 2 (SD 3.01) and 9.35 (SD 3.06) minutes. The app was rated as useful or very useful by the vast majority of participants 90% (56/62). Moreover, 90% (56/62) of the participants reported that they had learned or learned a lot, and 98% (61/62) of the participants reported that they would recommend the app to a friend. Repeated measures ANOVA (43/62, 69%) showed that after the intervention participants showed an increased knowledge of mental health and intention to seek help as well as fewer stigmatizing attitudes and less social distance.

Conclusions: Preliminary data show that esTOCma is a feasible and acceptable app, and after completing its 10 missions, there is an increase in the understanding of OCD and help-seeking intention along with a decrease in the social stigma and social distance associated with OCD that lasts for at least 3 months. The results support the potential of technology-based interventions to increase the intention to seek help and reduce the stigma associated with OCD. A larger, community-controlled study is also recommended.

(JMIR Mhealth Uhealth 2024;12:e48027) doi:10.2196/48027

KEYWORDS

obsessive-compulsive disorder; OCD; mental health literacy; stigma; app; help-seeking intention; seek help; mobile phone

Introduction

Background

Obsessive-compulsive disorder (OCD) is a clinically heterogeneous condition characterized by obsessions, compulsions or both that cause clinically significant levels of distress or functional impairment [1,2]. At present, effective treatments exist for OCD [3,4]; however, many people delay seeking treatment [5,6], and this may contribute to its chronic course [7,8]. Research suggests that this delay in seeking treatment by patients with OCD may be explained by social stigma and poor mental health literacy (MHL) about the disorder [5,9,10].

To date, many interventions have proved to be effective in reducing stigma associated with mental disorders, producing knowledge, and achieving attitudinal improvements [11-13]. Data show that contact and education strategies produce small-to-medium short-term reductions in stigmatizing attitudes, and there is limited evidence on long-term effectiveness [14,15]. A few of these interventions have benefited from the advantages of new technologies (eg, video games and electronic contact with patients), showing medium effects on reducing social stigma and suggesting that new technologies are a useful tool to decrease stigmatizing attitudes toward mental disorders, at least in the young population (the mean age of participants ranged between 15.7 and 24 years) [16]. However, none of these studies have examined the benefits of using app-based interventions. Furthermore, most of these interventions have focused on reducing the stigma of mental disorders such as schizophrenia [17-19], depression [20-22] or bipolar disorder [23,24].

In the case of OCD, programs are scarce, and most proposals have studied the impact of the educational mechanism [25-28], showing a reduction in stigmatizing attitudes and an improvement in participants' MHL. Furthermore, a proposal has reported a significant reduction in social stigma and social distance using an indirect contact strategy through a video of a patient with OCD and a family member talking about their experience with the problem [29]. However, none of these interventions are based on innovative technologies, and only 1 of them includes >1 intervention strategy (ie, contact and education) [29]. In this context, as a response to the limited interventions focused on OCD and the need to bridge the gap between the onset of symptoms and seeking help, a gamified mental health app named esTOCma has been developed [30] (refer to the Methods section). In this way, the intervention will take advantage of mobile mental health interventions as low-cost tools that are available 24 hours a day for a large number of people [31,32], in addition to including gaming benefits such as providing immediate feedback, motivating users to achieve goals, and being easy to use [33].

Objective and Hypothesis

The aim of this study is to explore the feasibility, acceptability, and preliminary effectiveness of the beta version of the eHealth mobile app esTOCma. Regarding the effectiveness, and based on the reviewed literature on interventions to reduce stigma and increase knowledge of OCD [25-29] and other mental disorders [11,14,16], we hypothesize that the esTOCma intervention will (1) improve knowledge associated with OCD and its treatment, (2) decrease stigmatizing attitudes and social distance, and (3) increase intention to seek help. Moreover, we hypothesize that changes will be maintained at the 3-month follow-up assessment.

Methods

Study Design

This study was a single-arm pilot intervention with 3 measures at the pretest, posttest, and 3-month follow-up. Data were collected from people residing in Spain.

Participants and Procedure

Data were collected from a convenience sample. Participants were recruited from the general community and university setting by snowball sampling after providing relevant information via face-to-face classes and inviting the participants to share information about the study with their acquaintances. The inclusion criteria for this study were as follows: (1) being aged >18 years, (2) residing in Spain, (3) owning a smartphone, and (4) self-reporting not having an OCD diagnosis. Interested participants were invited to participate in a study consisting of downloading an app (esTOCma) in an Android Package Kit file format, playing with it, and completing a set of questionnaires before and after using the app. The participants performed all the tasks individually at home and at their own pace at the time they deemed most convenient. Furthermore, they chose the rate at which they completed the game, although the app recommends completing 1 mission a day.

The participants signing the informed consent form were given an identification number automatically generated by the app. The data from the questionnaires and the game were matched with the personal player ID. The participants were randomized by the app using a sampling without replacement method to 1 of 6 vignettes describing a person with obsessive-compulsive (OC) symptoms from 1 out of 6 types of content (ie, aggression or harm; sexual; religious, blasphemous, or immoral; contamination or washing; doubts or checking; or superstition, symmetry, or order). The vignettes consisted of descriptions of patients presenting with OC symptoms, with similar severity. Furthermore, interference and impairment in quality of life were described. The descriptions were based on real clinical cases [34]. All of them meet the diagnosis criteria following the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition*, and the labels of obsessions or compulsions were avoided. All the described patients were referred to as A; sex

was not specified, and they were middle aged. Furthermore, all descriptions included between 166 and 175 words.

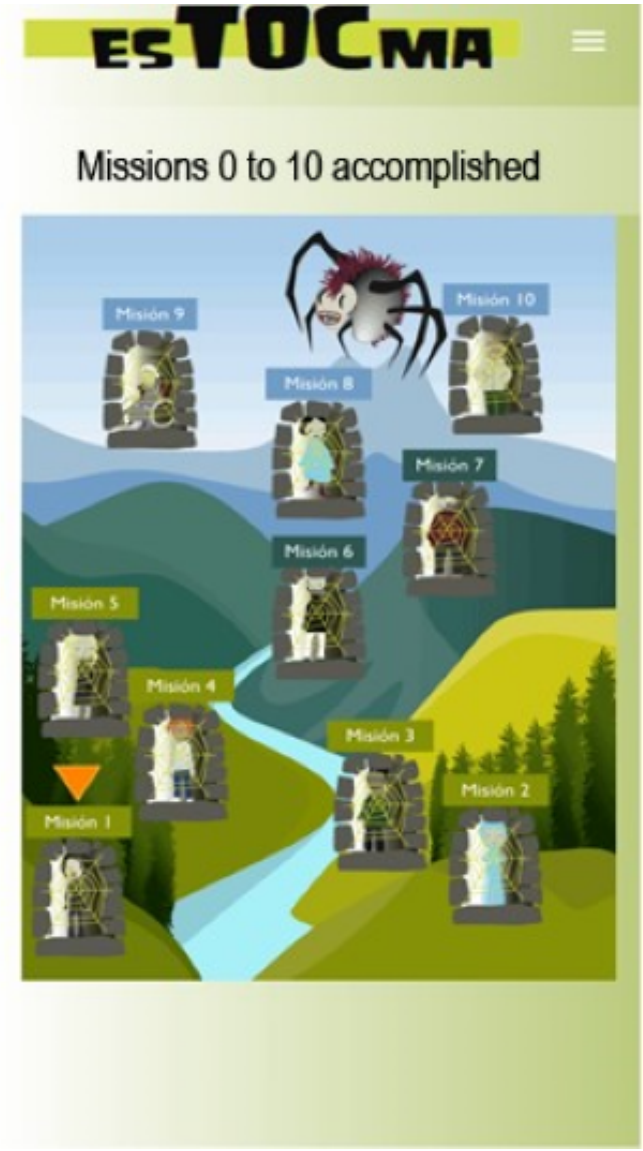
After reading the assigned vignette, the participants were asked by the app to complete the preintervention measures. Most of them (ie, Spanish Mental Illness Stigma Attribution Questionnaire–27 [AQ-27-E)], General Help-Seeking Questionnaire [GHSQ], Social Distance Scale [SDS], and MHL questionnaire) were answered in reference to the assigned vignette. Only after completing the preassessment were the participants able to begin playing. The participants who completed the 10 missions and finished the game were asked through the app to complete the postintervention assessment and the 3-month follow-up assessment. To encourage the participants to complete the game, they were given entries to a prize draw for a voucher to spend on the web.

Intervention Program

The esTOCma beta version is a serious game whose content and videos were developed by Doctor of Philosophy–level clinical psychologists who are experts on OCD, together with a usability expert. A professional designer developed the graphic elements, and a computer engineer developed the app. After testing multiple prototypes, this version was developed (Multimedia Appendix 1).

During the game, participants are asked to fight against the OCD "stigma monster" with their knowledge by accomplishing 10 missions and freeing the 10 characters who are affected by the esTOCma monster, a creature that feeds on false beliefs and misinformation in society [35] (Figure 1). The participants are guided through the game by a woman who describes herself as an OCD expert.

Figure 1. The missions, set along a mountain road.



The game is organized into three different intervention mechanisms: (1) psychoeducation (general information about OCD, OCD heterogeneity, OCD dimensionality, OCD cognitive model, evidence-based treatments, and options for seeking help in OCD), which includes 5 missions, and 1 of them—mission 3—includes a video explaining the OCD cognitive model (1 min); (2) indirect contact (including 6 videos of approximately 2 to 3 minutes of 3 patients diagnosed with OCD who talk about their own experience with OCD: symptom description, interference, how long it took them to seek help, their experience

in disclosing that they had a disorder, and experience with psychotherapy), which is organized into 2 missions; and (3) cognitive restructuring to replace dysfunctional beliefs related to rejection toward people with OCD, variables involved in OCD development, and treatment options and effectiveness, which includes 3 missions.

The missions are organized as follows: (1) the expert describes the objective of the mission and introduces the character to be freed, (2) the expert presents activities associated with the mission (between 6 and 8 activities, mostly consisting of reading a text and answering questions about it), and (3) the freed character appears. During the game, users receive basic internet-based rewards (ie, a key to free the character), together with a message of reinforcement and a video of the character actually being freed from esTOCma. At the end of the game, there is new visual reinforcement through a video in which all the characters are freed from the monster. Moreover, users receive reinforcement through a diploma that certifies them as OCD experts. A further description of the game can be found in the study by Chaves et al [35].

Measures

The participants completed the sociodemographic information and Obsessive-Compulsive Inventory–Revised at preassessment and an acceptability questionnaire at postassessment. All other measures were completed at the preassessment, postassessment, and 3-month follow-up assessment.

Sociodemographic Variables

The sociodemographic variables include gender, age, educational level, level of information and communications technology (ICT) knowledge (from 1 *little or none* to 5 *expert level*), and the question of whether they have an OCD diagnosis.

Acceptability Questionnaire

It consisted of 3 questions developed ad hoc to assess the acceptability of the esTOCma app. The survey included 2 questions related to usefulness (ie, “Did you find the app useful?” [from *very useful* to *not useful at all*] and “Did you learn from the app?” [from *I learned a lot* to *I learned nothing*]) and 1 question related to satisfaction with the app (ie, “Would you recommend this app to a friend?” [from *a lot* to *not at all*]). The questions were multiple-choice questions with 4 alternatives.

AQ-27-E Measures

This measures the social stigma associated with a vignette describing a person showing OC symptoms through 27 items rated on a Likert-type scale ranging from 1 to 9 [36,37]. It includes 9 subscales with 3 items: responsibility, pity, anger, dangerousness, fear, no help, coercion, segregation, and avoidance. Higher scores indicate higher social stigma. In this study, the AQ-27-E scales showed acceptable to excellent internal consistency across subscales (from 0.70 [anger, 3-month follow-up] to 91 [fear, 3-month follow-up]), except for responsibility (ranging from 0.40 [postassessment] to 0.63 [preassessment]) and pity (ranging from 0.36 [preassessment] to 0.52 [3-month follow-up]).

GHSQ Measures

This measures the intention to seek help from 10 different sources with regard to the specific content described in a vignette describing a person showing OC symptoms [38,39]. Participants rated 10 items regarding their help-seeking intentions if they were experiencing from symptoms similar to those described in the vignette on a 7-point Likert-type scale ranging from 1 (*extremely unlikely*) to 7 (*extremely likely*). Scores were calculated by summing up the items and dividing by 10. Higher scores indicate a higher intention to seek help. In this, the GHSQ showed acceptable internal consistency at the different assessment points (Cronbach α ranging from 0.68 [preassessment] to 0.75 [postassessment]).

SDS Measures

This assesses an individual’s willingness to interact with an individual with a mental disorder described in a vignette across 7 different situations on a 4-point Likert scale from 0 (*definitely willing*) to 3 (*definitely unwilling*) [40]. The total score has been calculated by adding the scores and dividing by the 7 items, with higher scores indicating a greater preference for social distance. In this, the SDS showed between good and excellent internal consistency at all assessment points; Cronbach α ranged from 0.86 at postassessment to 0.93 at the 3-month follow-up.

MHL Questionnaire

This is an instrument developed for this study based on previous studies [25,41] and assesses 1 of the components defined by Kutcher et al [42] as MHL but only referring to OCD: the understanding of OCD and its treatment. Part 1 has 4 multiple-choice questions including between 2 and 7 alternative answers, only 1 of which is correct. The questions refer to the assigned vignette (person A) and evaluate the following dimensions: (1) problem recognition (ie, “What happens to A is cause for concern?”; response alternatives: yes and no), (2) OCD identification (ie, “What do you think might be happening to A?” This question includes 7 response alternatives: family problems, adjustment problems, anxiety disorder, generalized anxiety disorder, schizophrenia, OCD, and depression), (3) perception of causality, and (4) effective treatment option. Part 2 has four multiple-choice questions with 3 alternative answers, only 1 of which is correct, and refers to participants’ general knowledge of OCD: (1) identification of OCD as a mental disorder (ie, OCD is [a] a learning disorder, [b] a mental disorder, or [c] a set of manias); (2) definition of obsession; (3) definition of a compulsion; and (4) role played by compulsions and other control strategies in the maintenance of obsessions. A total score has been calculated as the sum of the correct answers, thus ranging from 0 to 8.

Obsessive-Compulsive Inventory–Revised Measures

This is an 18-item self-report questionnaire assessing distress caused by OC symptoms and rated on a 5-point Likert scale ranging from 0 (not at all) to 4 (extremely) [43,44]. A total score was calculated. The total score of the Obsessive-Compulsive Inventory-Revised showed excellent internal consistency at all assessment points in this study (Cronbach α ranging from 0.89 [postassessment] to 0.90 [preassessment]).

Statistical Analysis

Descriptive statistics (eg, means, SD, and frequencies or percentages) were used to analyze demographic data, study variables, and app use patterns. Chi-square and 1-tailed *t* tests were used to test whether the groups (completers vs noncompleters) had preexisting differences. The change in study variables over time was determined by a repeated measures ANOVA. Partial eta–squared was used to report the effect size of the intervention on the dependent measures. A mixed model was used to determine whether the pattern of use of the app affected its effectiveness. The within-participants factor was time, and the between-participants factor was whether the user followed the 1 mission per day recommendation. The statistical significance level was set at *P*=.05. SPSS Statistics (version 26; IBM Corp) was used for statistical analysis.

Ethical Considerations

All procedures described in the study have been approved by the Human Research Ethics Committee of the University of Valencia, Spain (1276901). All study participants provided informed consent before study enrollment.

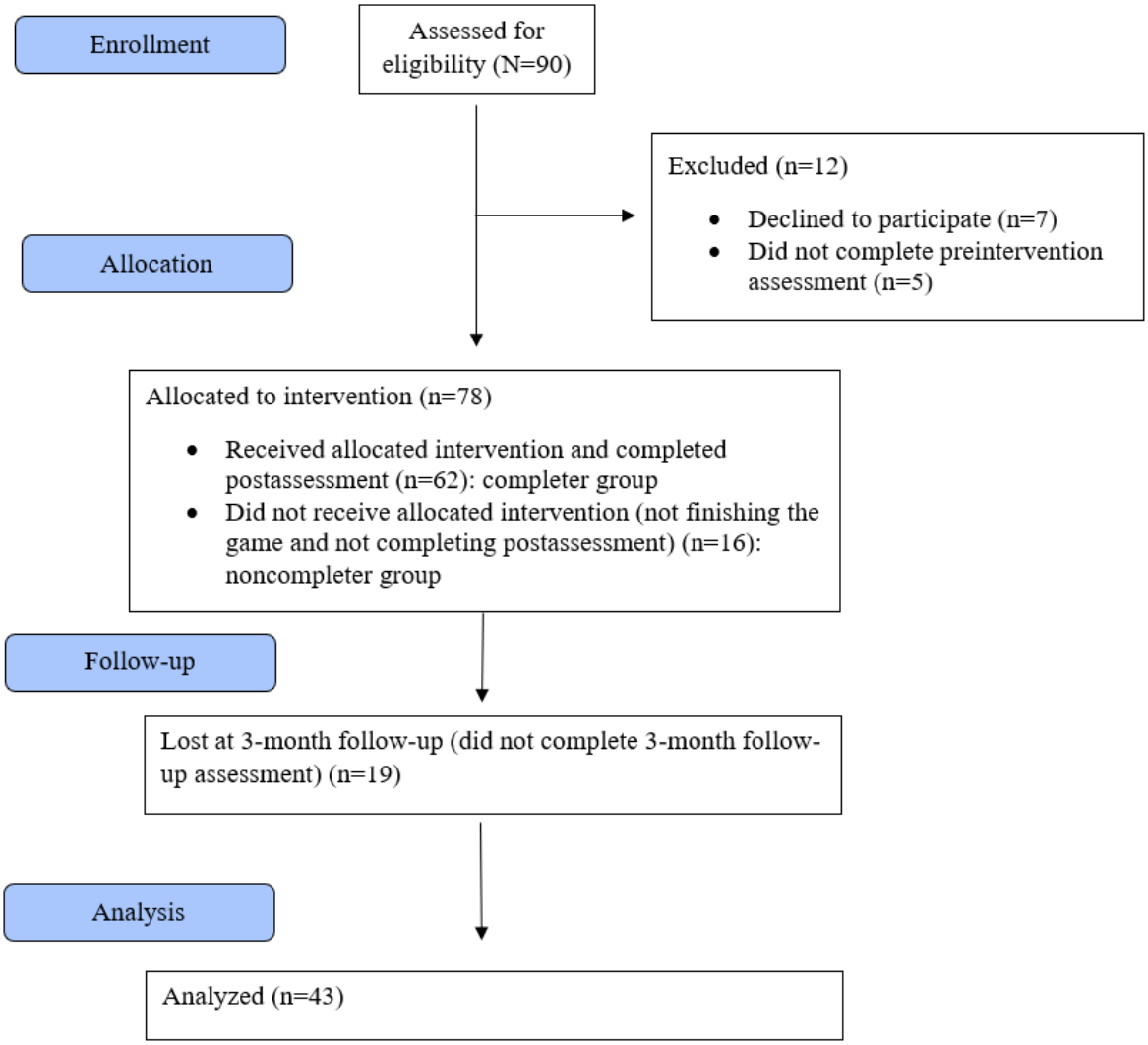
Results

Feasibility

Recruitment

A total of 90 participants were enrolled in the study; they downloaded the app and met the inclusion criteria. Of these 90 participants, 86% (78/90) were allocated to intervention and 13% (12/90) were excluded as they did not provide informed consent or did not complete the preassessment intervention (Figure 2). Of these 78 participants, 20% (16/78) did not complete the game or the postassessment (noncompleter group) and 79% (62/78) completed the game and answered the postintervention assessment (completer group). In the completer group, 69% (43/62) of the participants completed the 3-month follow-up assessment, and the remaining 30% (19/43) of the participants were lost at the 3-month follow-up.

Figure 2. Participants' flowchart.



Participants' Characteristics

The completer group had a mean age of 36.74 years, ranging from 18 to 71 years; were mostly women; and had university studies and moderate ICT knowledge. Differences in sociodemographic characteristics and preintervention assessment (baseline) between the completer group and the noncompleter group were calculated (Table 1).

Table 1. Differences between the completers and noncompleters on demographic and study variables (N=90).

Variable or measure	Completers (n=62)	Noncompleters (n=16)	Chi-square (*df*)	*t* test (*df*)	*P* value
Age (y), mean (SD)	36.74 (14.41)	27.19 (9.26)	10.4 (2)	N/A[a]	.003
Gender (women), n (%)	38 (61)	7 (44)	2.0 (2)	N/A	.35
Education level, n (%)			2.2 (2)	N/A	.31
Primary	3 (5)	0 (0)			
Secondary	11 (18)	1 (6)			
University	48 (77)	15 (94)			
Knowledge of ICT[b], n (%)			2.8 (2)	N/A	.58
Little or none	1 (2)	0 (0)			
Low	5 (8)	0 (0)			
Moderate	35 (57)	10 (63)			
Advanced	18 (29)	4 (25)			
Expert level	3 (5)	2 (13)			
MHL[c] (total[d]), mean (SD)	6.79 (1.50)	5.93 (1.61)	N/A	1.992 (76, 36.104)	.05
AQ-27-E[e] (total[f]), mean (SD)	84.77 (24.81)	77.25 (25.52)	N/A	1.075 (76, 36.104)	.28
SDS[g] (total[h]), mean (SD)	1.00 (0.64)	0.78 (0.70)	N/A	1.207 (76, 36.104)	.23
GSHQ[i] (total[j]), mean (SD)	4.20 (0.92)	4.27 (0.83)	N/A	0.308 (76, 36.104)	.38
OCI-R[k] (total[l]), mean (SD)	18.17 (13.23)	20.06 (10.70)	N/A	0.526 (76, 36.104)	.60

[a]N/A: not applicable.

[b]ICT: information and communications technology.

[c]MHL: Mental Health Literacy Questionnaire.

[d]Total score ranging from 0 to 8.

[e]AQ-27-E: Spanish Mental Illness Stigma Attribution Questionnaire–27.

[f]Total scoring from 27 to 243.

[g]SDS: Social Distance Scale.

[h]Total score ranging from 0 to 3.

[i]GHSQ: General Help-Seeking Questionnaire.

[j]Total scoring from 1 to 7.

[k]OCI-R: Obsessive-Compulsive Inventory–Revised.

[l]Total scoring from 0 to 72.

Statistically significant differences were only observed in age, with the completer group being older, and MHL scores, which were higher in the completer group. In addition, in the completer group, we explored the differences between the participants who completed the 3-month assessment (43/62, 69%) and those who did not (19/62, 31%). The only discernible difference observed was in age (t_{60}=2.167; *P*=.01), with participants who completed the follow-up assessment being older.

Of the 43 participants included in the effectiveness analyses, most were women (n=28, 65%) with a mean age of 39.30 (SD 14.58; range 21-71) years, with university-level education (n=35, 81%), and with ICT knowledge between moderate (n=23, 54%) and advanced (n=15, 35%; only 3 (7%) participants described having between little or none and low ICT knowledge level).

App Use Pattern

An analysis of the app use pattern was conducted with the completer group. First, we explored the number of days that elapsed from the preassessment to the postassessment. The participants took a mean of 13.64 (SD 10.50; range 1-44) days to complete the app (which includes having done the preassessment, the 10 missions and the postassessment), with a mode of 1, that is, the most frequent pattern was conducting the pre- and postassessments and missions of the app in 1 day. Second, we analyzed the number of days the person spent performing the missions. The participants spent between 1 and

10 days performing the missions, with a mean of 4.56 (SD 3.33) days and a mode of 1. In total, 30% (19/62) of the participants completed the 10 missions in 1 day, whereas another 21% (13/62) of the participants spent between 9 and 10 days completing the app, that is, approximately 1 session per day.

Finally, we explored the minutes spent performing each mission, first excluding the participants who stayed on 1 mission for >20 minutes, as we assumed that they left the app open without using it. The data on the participants completing the missions in <20 minutes is presented in Table 2.

Table 2. Time (in min) spent in completing each of the 10 missions for participants who complete the mission in <20 minutes.

	Participants, n (%)[a]	Time (min), mean (SD; range)	Mode (min)
Mission 1	58 (94)	3.93 (1.02; 1-6)	4
Mission 2	54 (87)	4.44 (1.90; 2-10)	4
Mission 3	35 (56)	3.80 (1.37; 1-8)	3
Mission 4	35 (56)	2.06 (1.64; 0-10)	2
Mission 5	35 (56)	5 (3.11; 2-20)	4
Mission 6	56 (96)	8.71 (1.78; 7-20)	8
Mission 7	29 (47)	9.35 (3.06; 7-20)	8
Mission 8	33 (53)	3.85 (1.48; 2-8)	3
Mission 9	58 (94)	2 (3.01; 0-19)	0
Mission 10	56 (90)	2.09 (2.26; 0-9)	0

[a]n (%) of participants completing the mission in ≤20 minutes.

The data are displayed in minutes.

Between the participants who completed missions in ≤20 minutes, and considering the mode, missions included in module 1 (ie, missions 1 to 5) took between 2 and 4 minutes, in module 2 (ie, missions 6 to 7) took 8 minutes, and in module 3 (ie, missions 8 to 10) took between <1 minute and 3 minutes.

Acceptability

After completing the app until the end of the game, most participants (56/62, 90%) perceived the app as useful or very useful. Moreover, 90% (56/62) of the participants indicated that they had learned or learned a lot, and 98% (61/62) of the participants indicated that they would recommend the app to a friend.

Preliminary Effectiveness: Differences Between Pre- and Postintervention Assessments and 3-Month Follow-Up

Repeated measures ANOVA was conducted to examine differences between pre- and postintervention and the 3-month follow-up on knowledge of OCD and its treatments (MHL questionnaire), stigmatizing attitudes (AQ-27-E), social distance (SDS), and intention to seek help (GHSQ) associated with OCD (Table 3). The results showed statistically significant differences (P≤.05) in all variables with medium-to-large effect sizes, except for the second part of the MHL questionnaire and the pity and coercion subscales (AQ-27-E). In general, the results show that using esTOCma until the end of the game results in an increase in MHL and intention to seek treatment (GHSQ) and a decrease in stigmatizing attitudes (AQ-27-E) and social distance desire (SDS). Post hoc pairwise comparisons showed statistically significant differences between preintervention and the other 2 assessment points (postintervention and 3-month follow-up). No significant differences were observed between postintervention and the 3-month follow-up in the variables assessed, except for the MHL total score, in which follow-up scores did not differ from pre- to postintervention.

Table 3. Means (SDs) and repeated measures ANOVA on pre-, post-, and 3-month follow-up intervention scores (n=43).

Variable or measure	Pretreatment, mean (SD)	Posttreatment, mean (SD)	3 month follow-up, mean (SD)	*F* test (*df*)[a]	*P* value	ηp^{2b}
MHL[c]						
Part 1[d]	3.32 (0.80)[e]	3.72 (0.54)[e]	3.62 (0.61)[e]	6.523 (1.566, 65.769)	.005	0.134
Part 2[f]	3.58 (0.69)	3.72 (0.50)	3.77 (0.57)	1.896 (1.513, 63.526)	.16	0.043
Total score	6.90 (1.34)[e]	7.44 (0.93)[e]	7.39 (0.90)[e]	5.754 (1.325, 55.662)	.01	0.120
AQ-27-E[g]						
Responsibility	9.04 (4.07)[e]	6.97 (3.70)[e]	7.30 (4.35)[e]	8.103 (2, 84)	.001	0.162
Pity	17.30 (4.15)	18.62 (4.36)	18.20 (4.68)	1.984 (2, 84)	.14	0.045
Anger	8.23 (4.44)[e]	5.81 (3.42)[e]	5.76 (3.19)[e]	10.554 (1.614, 67.813)	<.001	0.201
Dangerousness	7.11 (4.31)[e]	4.72 (2.65)[e]	5.37 (4.01)[e]	7.386 (2, 84)	<.001	0.150
Fear	6.39 (4.26)[e]	5.00 (3.72)[e]	4.23 (2.42)[e]	6.489 (2, 84)	.002	0.134
No help	8.09 (4.68)[e]	5.69 (3.32)[e]	5.83 (4.05)[e]	9.320 (2, 84)	<.001	0.182
Coercion	12.00 (5.30)	11.18 (5.50)	10.93 (6.47)	0.983 (1.490, 62.601)	.35	0.023
Segregation	5.16 (3.92)[e]	4.00 (1.96)[e]	4.04 (2.22)[e]	4.762 (1.555, 65.297)	.01	0.102
Avoidance	11.88 (5.66)[e]	8.58 (6.35)[e]	7.76 (5.48)[e]	16.938 (2, 84)	<.001	0.287
SDS[h]	0.99 (0.58)[e]	0.64 (0.65)[e]	0.59 (0.63)[e]	10.597 (2, 84)	<.001	0.201
GSHQ[i]	4.33 (0.84)[e]	4.84 (0.86)[e]	4.58 (1.03)[e]	6.818 (2, 84)	.002	0.140

[a]Dfs were Greenhouse-Geisser corrected where appropriate.

[b]ηp^2:partial eta squared for within-subject contrasts (ANOVA).

[c]MHL: Mental Health Literacy Questionnaire, total score ranging from 0 to 8.

[d]Part 1 scoring from 0 to 4.

[e]Significant differences among groups ($P \leq .05$).

[f]Part 2 scoring from 0 to 4.

[g]AQ-27-E: Spanish Mental Illness Stigma Attribution Questionnaire–27; subscales ranging from 3 to 27.

[h]SDS: Social Distance Scale, ranging from 0 to 3.

[i]GHSQ: General Help-Seeking Questionnaire, scoring from 1 to 7.

Finally, we examined whether the pattern of use of the app influenced the effectiveness of the intervention with a 2 (group: individuals who follow the recommendation of 1 mission per day [ie, 9/10, % days] [12/N, %; percentile 75] vs individuals who did it in another way [31/N, %])×3 (time: pre-, post-, and follow-up assessments) repeated measures mixed ANOVA. The results show that there was no significant group×time interaction ($F_{26,16}$=0.838; P=.67). Univariate follow-up analyses also indicated no significant group×time effect for any of the measured variables (P>.05).

Discussion

Principal Findings

This study is the first to investigate the feasibility, acceptability, and effectiveness of a mobile health app designed to enhance several variables related to OCD, which were identified by Kutcher et al [42] as components of MHL: understanding OCD and its treatments, decreasing stigma associated with OCD, and enhancing help-seeking effectiveness. Our findings show that esTOCma was feasible and acceptable and that after using it until game completion, there was a positive change in the variables of interest, which lasted for at least 3 months.

Of the participants allocated to the intervention, approximately 80% (78/90) completed the app, and of the who completed the intervention, approximately 70% (62/78) completed the follow-up assessment. This adherence rate is similar to or higher than that reported in other internet-based studies with self-help interventions, with dropouts being one of the main challenges of interventions with mental health apps [45-47].

The participants who completed the study were older and had higher knowledge of OCD than those who began the study but did not complete it. However, there were no differences in participants' knowledge of new technologies, which suggests that the app is easy to use and does not require a significant amount of knowledge to become involved in its use, although >50% (35/62) of the participants described themselves as having

moderate ICT knowledge. Regarding the pattern of use, although it was recommended that participants complete 1 mission per day and participants were reminded of this recommendation after finishing each mission, the most common pattern of use was to complete the app, including pre- and postevaluations, in 1 day. In fact, only 21% (13/62) of the participants completing the intervention followed the recommendation to perform 1 mission per day. It seems that it is more comfortable for participants to perform more missions per day; otherwise, they forget to complete the mission the following day. In fact, the participants completed their participation in the study within a range of 1 to 44 days.

Regarding the time invested in each mission, although missions were quick to complete and always took <10 minutes, the participants often left missions midway and continued later the same day or days later. Furthermore, certain missions were abandoned in the middle more frequently, suggesting that they could be reformulated to make them more "attractive." This was the case for mission 7, which was interrupted by >50% (31/62) of the participants, as well as for missions 3, 4, 5, and 8. The pattern of use and interruptions does not seem to be associated with the intervention mechanism on which the mission is based but perhaps with the content or the duration of the mission. If we analyze missions 1, 2, 6, 9, and 10, those that were carried out without interruption, we see that missions 1 and 2 are the initial ones and deal with content describing obsessions and compulsions; mission 6 includes the first videos that also describe symptomatology and interference, without additional theoretical content to read; and missions 9 and 10 (cognitive restructuring) are characterized by being very brief and with less theoretical content than missions 1 to 5 (psychoeducation). It seems that the description of symptomatology, as well as more dynamic and shorter missions, result in more attractive missions or at least in missions that capture participants' attention to a greater extent.

In general, the participants seemed satisfied with the app, as approximately 100% (62/62) would recommend it to a friend and >90% (56/62) consider it useful and that they have learned about OCD.

Regarding the effectiveness of the app, the results show an increase in OCD knowledge. Our data match those of the previous interventions that have found increases in general knowledge of OCD after offering written information about the disorder [25,26]. The data are also consistent with other technology interventions that have increased MHL levels on different mental health problems [48-50]. However, the differences were not statistically significant in those questions that asked about OCD in general (eg, the definition of an obsession), that is, not referring to the description of a person showing OC symptoms. This could be due to a ceiling effect, as the scores were already high in the preassessment evaluation. In fact, they were higher among those participants who played with the app until the game was over in comparison with those who did not finish the game. In this sense, our first hypothesis was only partially supported.

Regarding the hypothesized decrease in stigmatizing attitudes, the results support our hypothesis, as they suggest that after completing the 10 missions of the app, there was a decrease in some stereotypes or public attitudes, such as the perception of OC symptoms as dangerous; emotional reactions of anger or fear toward people showing OC symptoms; discriminating behaviors such as the intention of not helping, segregating, or avoiding people showing OC symptoms; or the desire to maintain social distance. Although a video-based intervention decreased social distance desire [29] and 2 interventions centered on reading the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition diagnostic criteria for OCD decreased negative attitudes about violent and sexual thoughts [28,51], other interventions consisting of reading information on OCD (psychoeducation) showed small changes in stigmatizing attitudes [25]. The results suggest that the esTOCma intervention, which includes psychoeducation but also incorporates components such as contact, seems to change stigmatizing attitudes to a greater extent, with medium-to-large effect sizes.

Our data are also consistent with previous research also using new technologies to reduce stigma associated with other mental disorders that have reported a decrease in dangerousness, anger, fear, segregation, and avoidance [15,52,53]; a decrease in stigma as a general measure [53-55]; an increase in the help factor [52]; or a decrease in social distance [52,56]. This is a remarkable result, as traditional antistigma interventions not using innovative technologies often report small-to-medium effect sizes [14,52]. Moreover, the results are based on a community sample with an average age higher than those used in studies reporting interventions using innovative technologies [16], suggesting that these types of interventions could also be useful for older people.

In addition to the changes in stigmatizing attitudes after using esTOCma, there were no significant changes in 2 of the social stigma dimensions measured by the AQ-27-E: pity and coercion. Previous studies have also reported a lack of changes in pity [57]. In fact, it has been suggested that pity should be considered, at least in some contexts, not to be a factor of stigma but rather a reflection of compassion and the capacity to empathize with people with mental health problems [58], and research has shown associations between pity and the tendency to help [37]. Regarding coercion, there is no significant change after the intervention in the belief that people with OCD should receive treatment, even if they refuse it (coercion). This is consistent with previous interventions in the OCD field [25].

Finally, the results also support our third hypothesis, as there was a large effect size improvement in the intention to seek help when experiencing symptoms similar to those represented in the different vignettes. Thus, increasing knowledge of OCD and its treatment could have acted as a help-seeking facilitator [59,60]. To the best of our knowledge, there are no interventions to improve help seeking associated with OCD, and thus, our results are of great relevance, as research shows that early help seeking is associated with a better treatment response, earlier remission of symptoms, and recovery from the disorder [9,61]. The data are consistent with other studies that, through mental health apps [62] and other technology-based interventions, have improved the intention to seek help for other mental health problems [63].

Moreover, our study showed a maintained effect in the study variables at the 3-month follow-up assessment. This is especially relevant as most of the interventions do not include a follow-up assessment [16,50,63,64], and only 1 intervention on OCD has included a follow-up assessment [29].

Limitations and Recommendations

This study has limitations. As a pilot study, the study sample was small and did not include a control group. There was also a considerable dropout rate that could be associated with the fact that participants forgot to complete the app (to do all 10 missions) as well as the large number of assessment questionnaires included. Furthermore, considering that OCD is a heterogeneous disorder and that there are differences in stigma and OCD recognition between different types of content [41,64,65], we decided to randomize participants to 6 vignettes that represent OCD heterogeneity. However, this decision could be considered as a limitation of the design of this study that could affect the effectiveness of the data.

Despite these limitations, our results provide valuable information for a broader implementation of the app. First, as future lines of work, we propose making an effort to motivate participants, given that higher adherence was associated with initial knowledge of OCD or sensitivity toward mental health problems. Second, a potential strategy to attract a younger audience, as adherence was also associated with older age, could involve transforming the app into a serious game, aligning esTOCma with the immersive experience of video games. Actively engaging participants through gamification elements could further contribute to increasing adherence. Looking ahead, evolving esTOCma into a serious game with an adventure game theme holds promise for future developments and for addressing the stigma associated with OCD. Third, pop-ups and emails to remind participants to “play” could be incorporated, as those participants who completed the participation took up to 44 days to finish an app recommended to be conducted in 10 days. Fourth, given that the number of missions per day seems not to influence the effectiveness of the app, the recommendation to play 1 mission per day could be omitted, as this is not the preferred pattern of use of participants. Finally, given that some missions were completed with interruptions by approximately half of the participants, it would be recommended to inform users of how long each mission will last. Furthermore, missions 3 to 5, 7, and 8 should be revised, as although they were quick to complete (between 3 and 8 min), the participants often left them midway and continued later. They could perhaps be shortened or made more dynamic.

Future Research

The results suggest the potential for the use of this intervention app and provide the basis for developing a larger randomized controlled study to validate the use of esTOCma (version 1.0), as has been proposed [35]. Importantly, we will also explore differences in OCD types of content through a controlled study, as research shows that there are differences in OCD recognition and stigma considering the different OCD types of content [41,64,65]. Furthermore, it will be of interest to explore the effectiveness of each intervention mechanism (ie, psychoeducation, indirect contact, and cognitive restructuring) to understand whether there are differences between them. This analysis would allow us to develop a new app that would eliminate those mechanisms that are less effective or improve them. Future studies should also examine the usefulness of this app in OCD cohorts, especially in the first stages of diagnosis, and their families. Through esTOCma, people with OCD could receive (or reinforce) cognitive-behavioral psychoeducation and be assisted in their search for empirical-based treatments. Furthermore, esTOCma could eliminate self-stigma, improving their quality of life. In terms of family involvement, esTOCma could assist them in comprehending and managing OCD symptoms as well as support their relatives with OCD in seeking effective treatment.

Conclusions

Our findings show that esTOCma is a feasible and acceptable app and that after completing its 10 missions, there is an increase in the understanding of OCD and help-seeking intention as well as a decrease in the social stigma and social distance associated with OCD that lasts for at least 3 months. These changes might result in less delay in seeking help and a better treatment response to the problem and prognosis. Moreover, providing mental health knowledge to the community population as to the nature and universality of intrusive thoughts may protect and prevent the general population from developing OCD and, furthermore, reduce the economic and personal costs associated with OCD.

Acknowledgments

Funding for this study (RTI2018-098349-B-I00 and PID2021-124409OB-I00) was provided by the Ministry of Science and Innovation—State Research Agency of Spain and by the European Union European Regional Development Fund: “A way of making Europe.” None of the funders had any role in the design, preparation, review, or approval of the manuscript or in the decision to submit the manuscript for publication.

Authors' Contributions

The work was conceptualized by AC and GG-S. The methodology and research, and the writing of the original draft were carried out by AC, GG-S and SA. Formal analyses were developed by SA. The final manuscript was reviewed and edited by GG-S.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Screenshots of the esTOCma app with examples of the Psychoeducation mechanism.
[PDF File (Adobe PDF File), 960 KB - mhealth_v12i1e48027_app1.pdf]

References

1. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, 5th Edition. Washington, DC: American Psychiatric Association; 2013.
2. McKay D, Abramowitz JS, Calamari JE, Kyrios M, Radomsky A, Sookman D, et al. A critical evaluation of obsessive-compulsive disorder subtypes: symptoms versus mechanisms. Clin Psychol Rev 2004 Jul;24(3):283-313. [doi: 10.1016/j.cpr.2004.04.003] [Medline: 15245833]
3. Abramowitz JS, Blakey SM, Reuman L, Buchholz JL. New directions in the cognitive-behavioral treatment of OCD: theory, research, and practice. Behav Ther 2018 May;49(3):311-322. [doi: 10.1016/j.beth.2017.09.002] [Medline: 29704962]
4. Olatunji BO, Davis ML, Powers MB, Smits JA. Cognitive-behavioral therapy for obsessive-compulsive disorder: a meta-analysis of treatment outcome and moderators. J Psychiatr Res 2013 Jan;47(1):33-41. [doi: 10.1016/j.jpsychires.2012.08.020] [Medline: 22999486]
5. Belloch A, Del Valle G, Morillo C, Carrió C, Cabedo E. To seek advice or not to seek advice about the problem: the help-seeking dilemma for obsessive-compulsive disorder. Soc Psychiatry Psychiatr Epidemiol 2009 Apr 31;44(4):257-264. [doi: 10.1007/s00127-008-0423-0] [Medline: 18668187]
6. Pinto A, Mancebo MC, Eisen JL, Pagano ME, Rasmussen SA. The brown longitudinal obsessive compulsive study: clinical features and symptoms of the sample at intake. J Clin Psychiatry 2006 May 15;67(5):703-711 [FREE Full text] [doi: 10.4088/jcp.v67n0503] [Medline: 16841619]
7. Fullana MA, Mataix-Cols D, Caspi A, Harrington H, Grisham J, Moffitt TE, et al. Obsessions and compulsions in the community: prevalence, interference, help-seeking, developmental stability, and co-occurring psychiatric conditions. Am J Psychiatry 2009 Mar;166(3):329-336 [FREE Full text] [doi: 10.1176/appi.ajp.2008.08071006] [Medline: 19188283]
8. Subramaniam M, Abdin E, Vaingankar JA, Chong SA. Obsessive--compulsive disorder: prevalence, correlates, help-seeking and quality of life in a multiracial Asian population. Soc Psychiatry Psychiatr Epidemiol 2012 Dec 17;47(12):2035-2043. [doi: 10.1007/s00127-012-0507-8] [Medline: 22526825]
9. García-Soriano G, Rufer M, Delsignore A, Weidt S. Factors associated with non-treatment or delayed treatment seeking in OCD sufferers: a review of the literature. Psychiatry Res 2014 Dec 15;220(1-2):1-10. [doi: 10.1016/j.psychres.2014.07.009] [Medline: 25108591]
10. Williams MT, Domanico J, Marques L, Leblanc NJ, Turkheimer E. Barriers to treatment among African Americans with obsessive-compulsive disorder. J Anxiety Disord 2012 May;26(4):555-563 [FREE Full text] [doi: 10.1016/j.janxdis.2012.02.009] [Medline: 22410094]
11. Griffiths KM, Carron-Arthur B, Parsons A, Reid R. Effectiveness of programs for reducing the stigma associated with mental disorders. A meta-analysis of randomized controlled trials. World Psychiatry 2014 Jun 02;13(2):161-175 [FREE Full text] [doi: 10.1002/wps.20129] [Medline: 24890069]
12. Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Cochrane Bias Methods Group, Cochrane Statistical Methods Group. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011 Oct 18;343(oct18 2):d5928 [FREE Full text] [doi: 10.1136/bmj.d5928] [Medline: 22008217]
13. Clement S, Lassman F, Barley E, Evans-Lacko S, Williams P, Yamaguchi S, et al. Mass media interventions for reducing mental health-related stigma. Cochrane Database Syst Rev 2013 Jul 23;2013(7):CD009453 [FREE Full text] [doi: 10.1002/14651858.CD009453.pub2] [Medline: 23881731]
14. Morgan AJ, Reavley NJ, Ross A, Too LS, Jorm AF. Interventions to reduce stigma towards people with severe mental illness: systematic review and meta-analysis. J Psychiatr Res 2018 Aug 16;103(4):120-133 [FREE Full text] [doi: 10.1016/j.jpsychires.2018.05.017] [Medline: 29843003]
15. Maunder RD, White FA. Intergroup contact and mental health stigma: a comparative effectiveness meta-analysis. Clin Psychol Rev 2019 Aug;72:101749. [doi: 10.1016/j.cpr.2019.101749] [Medline: 31254936]
16. Rodríguez-Rivas ME, Cangas AJ, Cariola LA, Varela JJ, Valdebenito S. Innovative technology-based interventions to reduce stigma toward people with mental illness: systematic review and meta-analysis. JMIR Serious Games 2022 May 30;10(2):e35099 [FREE Full text] [doi: 10.2196/35099] [Medline: 35635744]
17. Jorm AF, Kitchener BA, Fischer JA, Cvetkovski S. Mental health first aid training by e-learning: a randomized controlled trial. Aust N Z J Psychiatry 2010 Dec 01;44(12):1072-1081. [doi: 10.3109/00048674.2010.516426] [Medline: 21070103]
18. Corrigan PW, River LP, Lundin RK, Penn DL, Uphoff-Wasowski K, Campion J, et al. Three strategies for changing attributions about severe mental illness. Schizophr Bull 2001 Jan 01;27(2):187-195. [doi: 10.1093/oxfordjournals.schbul.a006865] [Medline: 11354586]
19. Penn DL, Chamberlin C, Mueser KT. The effects of a documentary film about schizophrenia on psychiatric stigma. Schizophr Bull 2003 Jan 01;29(2):383-391. [doi: 10.1093/oxfordjournals.schbul.a007012] [Medline: 14552511]

20. Kiropoulos LA, Griffiths KM, Blashki G. Effects of a multilingual information website intervention on the levels of depression literacy and depression-related stigma in Greek-born and Italian-born immigrants living in Australia: a randomized controlled trial. J Med Internet Res 2011 Apr 19;13(2):e34 [FREE Full text] [doi: 10.2196/jmir.1527] [Medline: 21504872]
21. Gulliver A, Griffiths KM, Christensen H, Mackinnon A, Calear AL, Parsons A, et al. Internet-based interventions to promote mental health help-seeking in elite athletes: an exploratory randomized controlled trial. J Med Internet Res 2012 Jun 29;14(3):e69 [FREE Full text] [doi: 10.2196/jmir.1864] [Medline: 22743352]
22. Farrer L, Christensen H, Griffiths KM, Mackinnon A. Web-based cognitive behavior therapy for depression with and without telephone tracking in a national helpline: secondary outcomes from a randomized controlled trial. J Med Internet Res 2012 Jun 27;14(3):e68 [FREE Full text] [doi: 10.2196/jmir.1859] [Medline: 22738715]
23. Patten SB, Remillard A, Phillips L, Modgill G, Szeto AC, Kassam A, et al. Effectiveness of contact-based education for reducing mental illness-related stigma in pharmacy students. BMC Med Educ 2012 Dec 05;12(1):120 [FREE Full text] [doi: 10.1186/1472-6920-12-120] [Medline: 23216787]
24. Rusch LC, Kanter JW, Angelone AF, Ridley RC. The impact of in our own voice on stigma. Am J Psychiatr Rehabil 2008 Oct 13;11(4):373-389. [doi: 10.1080/15487760802397660]
25. Chaves A, Arnáez S, Roncero M, García-Soriano G. Teachers' knowledge and stigmatizing attitudes associated with obsessive-compulsive disorder: effectiveness of a brief educational intervention. Front Psychiatry 2021 Jun 2;12:677567 [FREE Full text] [doi: 10.3389/fpsyt.2021.677567] [Medline: 34149482]
26. Jassi AD, Kolvenbach S, Heyman I, Macleod T, Rose J, Diamond H. Increasing knowledge about obsessive compulsive disorder and support for parents and schools: evaluation of initiatives. Health Educ J 2016 Jul 28;75(5):600-609. [doi: 10.1177/0017896915608513]
27. White SW, Sukhodolsky DG, Rains AL, Foster D, McGuire JF, Scahill L. J Dev Phys Disabil 2010 Aug 3;23(1):5-14. [doi: 10.1007/s10882-010-9209-x]
28. Snethen C, Warman DM. Effects of psychoeducation on attitudes towards individuals with pedophilic sexual intrusive thoughts. J Obsessive Compuls Relat Disord 2018 Oct;19:92-98. [doi: 10.1016/j.jocrd.2018.10.001]
29. Gürbüz AA, Yorulmaz O, DURNA G. Reducing the social stigma associated with obsessive compulsive disorder: a controlled trial of an intervention program in a Turkish community sample. J Evid Based Psychother 2020 Sep 1;20(2):101-120. [doi: 10.24193/jebp.2020.2.14]
30. Home page. EsTOCma. URL: http://estocma.com/ [accessed 2024-03-05]
31. Miralles I, Granell C, Díaz-Sanahuja L, Van Woensel W, Bretón-López J, Mira A, et al. Smartphone apps for the treatment of mental disorders: systematic review. JMIR Mhealth Uhealth 2020 Apr 02;8(4):e14897 [FREE Full text] [doi: 10.2196/14897] [Medline: 32238332]
32. Lecomte T, Potvin S, Corbière M, Guay S, Samson C, Cloutier B, et al. Mobile apps for mental health issues: meta-review of meta-analyses. JMIR Mhealth Uhealth 2020 May 29;8(5):e17458 [FREE Full text] [doi: 10.2196/17458] [Medline: 32348289]
33. Fitzgerald M, Ratcliffe G. Serious games, gamification, and serious mental illness: a scoping review. Psychiatr Serv 2020 Feb 01;71(2):170-183. [doi: 10.1176/appi.ps.201800567] [Medline: 31640521]
34. Belloch A, Cabedo E, Carrió C. TOC: Obsesiones y Compulsiones: Tratamiento Cognitivo del Trastorno Obsesivo Compulsivo. Madrid, Spain: Alianza Editorial; 2011.
35. Chaves A, Arnáez S, Castilla D, Roncero M, García-Soriano G. Enhancing mental health literacy in obsessive-compulsive disorder and reducing stigma via smartphone: a randomized controlled trial protocol. Internet Interv 2022 Sep;29:100560 [FREE Full text] [doi: 10.1016/j.invent.2022.100560] [Medline: 35874968]
36. Corrigan P, Markowitz FE, Watson A, Rowan D, Kubiak MA. An attribution model of public discrimination towards persons with mental illness. J Health Soc Behav 2003 Jun;44(2):162. [doi: 10.2307/1519806]
37. Muñoz M, Guillén AI, Pérez-Santos E, Corrigan PW. A structural equation modeling study of the Spanish Mental Illness Stigma Attribution Questionnaire (AQ-27-E). Am J Orthopsychiatry 2015 May;85(3):243-249. [doi: 10.1037/ort0000059] [Medline: 25985111]
38. Wilson CJ, Deane FP, Ciarrochi JV, Rickwood D. Measuring help seeking intentions: properties of the general help seeking questionnaire. Can J Couns Psychother 2005;39:15-28. [doi: 10.1037/t42876-000]
39. Pacheco del Castillo LA. Dominican college students' experiences of distress, help-seeking and stigma. Western Michigan University. 2017. URL: https://scholarworks.wmich.edu/cgi/viewcontent.cgi?article=4111&context=dissertations [accessed 2024-03-05]
40. Link BG, Cullen FT, Frank J, Wozniak JF. The social rejection of former mental patients: understanding why labels matter. Am J Sociol 1987 May;92(6):1461-1500. [doi: 10.1086/228672]
41. García-Soriano G, Roncero M. What do Spanish adolescents think about obsessive-compulsive disorder? Mental health literacy and stigma associated with symmetry/order and aggression-related symptoms. Psychiatry Res 2017 Apr;250:193-199. [doi: 10.1016/j.psychres.2017.01.080] [Medline: 28161615]
42. Kutcher S, Bagnell A, Wei Y. Mental health literacy in secondary schools: a Canadian approach. Child Adolesc Psychiatr Clin N Am 2015 Apr;24(2):233-244. [doi: 10.1016/j.chc.2014.11.007] [Medline: 25773321]

43. Foa EB, Huppert JD, Leiberg S, Langner R, Kichic R, Hajcak G, et al. The obsessive-complusive inventory: development and validation of a short version. Psychol Assess 2002;14(4):485-495. [doi: 10.1037//1040-3590.14.4.485]
44. Fullana MA, Tortella-Feliu M, Caseras X, Andión O, Torrubia R, Mataix-Cols D. Psychometric properties of the Spanish version of the Obsessive-Compulsive Inventory--revised in a non-clinical sample. J Anxiety Disord 2005 Jan;19(8):893-903. [doi: 10.1016/j.janxdis.2004.10.004] [Medline: 16243637]
45. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
46. Kelders SM, Van Gemert-Pijnen JE, Werkman A, Nijland N, Seydel ER. Effectiveness of a web-based intervention aimed at healthy dietary and physical activity behavior: a randomized controlled trial about users and usage. J Med Internet Res 2011 Apr 14;13(2):e32 [FREE Full text] [doi: 10.2196/jmir.1624] [Medline: 21493191]
47. Ludden GD, van Rompay TJ, Kelders SM, van Gemert-Pijnen JE. How to increase reach and adherence of web-based interventions: a design research viewpoint. J Med Internet Res 2015 Jul 10;17(7):e172 [FREE Full text] [doi: 10.2196/jmir.4201] [Medline: 26163456]
48. Brijnath B, Protheroe J, Mahtani KR, Antoniades J. Do web-based mental health literacy interventions improve the mental health literacy of adult consumers? Results from a systematic review. J Med Internet Res 2016 Jun 20;18(6):e165 [FREE Full text] [doi: 10.2196/jmir.5463] [Medline: 27323907]
49. Li TM, Chau M, Wong PW, Lai ES, Yip PS. Evaluation of a web-based social network electronic game in enhancing mental health literacy for young people. J Med Internet Res 2013 May 15;15(5):e80 [FREE Full text] [doi: 10.2196/jmir.2316] [Medline: 23676714]
50. Tay JL, Tay YF, Klainin-Yobas P. Effectiveness of information and communication technologies interventions to increase mental health literacy: a systematic review. Early Interv Psychiatry 2018 Dec 13;12(6):1024-1037. [doi: 10.1111/eip.12695] [Medline: 29897166]
51. Warman DM, L. Phalen P, Martin JM. Impact of a brief education about mental illness on stigma of OCD and violent thoughts. J Obsessive Compuls Relat Disord 2015 Apr;5:16-23. [doi: 10.1016/j.jocrd.2015.01.003]
52. Mullor D, Sayans-Jiménez P, Cangas AJ, Navarro N. Effect of a serious game (Stigma-Stop) on reducing stigma among psychology students: a controlled study. Cyberpsychol Behav Soc Netw 2019 Mar;22(3):205-211. [doi: 10.1089/cyber.2018.0172] [Medline: 30855993]
53. Rodríguez-Rivas ME, Cangas AJ, Fuentes-Olavarría D. Controlled study of the impact of a virtual program to reduce stigma among university students toward people with mental disorders. Front Psychiatry 2021 Feb 9;12:632252 [FREE Full text] [doi: 10.3389/fpsyt.2021.632252] [Medline: 33633613]
54. Cangas AJ, Navarro N, Parra JM, Ojeda JJ, Cangas D, Piedra JA, et al. Stigma-stop: a serious game against the stigma toward mental health in educational settings. Front Psychol 2017 Aug 21;8:1385 [FREE Full text] [doi: 10.3389/fpsyg.2017.01385] [Medline: 28878702]
55. Yuen AS, Mak WW. The effects of immersive virtual reality in reducing public stigma of mental illness in the university population of Hong Kong: randomized controlled trial. J Med Internet Res 2021 Jul 14;23(7):e23683 [FREE Full text] [doi: 10.2196/23683] [Medline: 34259636]
56. Ferchaud A, Seibert J, Sellers N, Escobar Salazar N. Reducing mental health stigma through identification with video game avatars with mental illness. Front Psychol 2020 Sep 9;11:2240 [FREE Full text] [doi: 10.3389/fpsyg.2020.02240] [Medline: 33013585]
57. Maunder RD, White FA, Verrelli S. Modern avenues for intergroup contact: using E-contact and intergroup emotions to reduce stereotyping and social distancing against people with schizophrenia. Group Process Intergr Relat 2018 Sep 19;22(7):947-963. [doi: 10.1177/1368430218794873]
58. Saavedra J, Arias-Sánchez S, Corrigan P, López M. Assessing the factorial structure of the mental illness public stigma in Spain. Disabil Rehabil 2021 Sep 06;43(18):2656-2662. [doi: 10.1080/09638288.2019.1710769] [Medline: 31906732]
59. Aguirre Velasco A, Cruz IS, Billings J, Jimenez M, Rowe S. What are the barriers, facilitators and interventions targeting help-seeking behaviours for common mental health problems in adolescents? A systematic review. BMC Psychiatry 2020 Jun 11;20(1):293 [FREE Full text] [doi: 10.1186/s12888-020-02659-0] [Medline: 32527236]
60. Picco L, Abdin E, Pang S, Vaingankar JA, Jeyagurunathan A, Chong SA, et al. Association between recognition and help-seeking preferences and stigma towards people with mental illness. Epidemiol Psychiatr Sci 2016 Dec 08;27(1):84-93. [doi: 10.1017/s2045796016000998]
61. Eisen JL, Pinto A, Mancebo MC, Dyck IR, Orlando ME, Rasmussen SA. A 2-year prospective follow-up study of the course of obsessive-compulsive disorder. J Clin Psychiatry 2010 Aug 15;71(08):1033-1039. [doi: 10.4088/jcp.08m04806blu]
62. Wiljer D, Shi J, Lo B, Sanches M, Hollenberg E, Johnson A, et al. Effects of a mobile and web app (Thought Spot) on mental health help-seeking among college and university students: randomized controlled trial. J Med Internet Res 2020 Oct 30;22(10):e20790 [FREE Full text] [doi: 10.2196/20790] [Medline: 33124984]
63. Johnson JA, Sanghvi P, Mehrotra S. Technology-based interventions to improve help-seeking for mental health concerns: a systematic review. Indian J Psychol Med 2022 Jul 21;44(4):332-340 [FREE Full text] [doi: 10.1177/02537176211034578] [Medline: 35949632]

64. Ponzini GT, Steinman SA. A systematic review of public stigma attributes and obsessive–compulsive disorder symptom subtypes. Stig Health 2022 Feb;7(1):14-26. [doi: 10.1037/sah0000310]
65. Ponzini GT, Signorelli M, Claydon EA, Lilly C, Steinman SA. Stereotypes and OCD-symptom presentations: a mixed-methods evaluation using male-character vignettes. Clinical Psychological Science 2023 Sep 24. [doi: 10.1177/21677026231192893]

Abbreviations

AQ-27-E: Spanish Mental Illness Stigma Attribution Questionnaire–27
GHSQ: General Help-Seeking Questionnaire
ICT: information and communications technology
MHL: mental health literacy
OC: obsessive-compulsive
OCD: obsessive-compulsive disorder
SDS: Social Distance Scale

Edited by L Buis; submitted 09.04.23; peer-reviewed by B Kalema, R McCarty, B Pascual, AM Ros; comments to author 25.08.23; revised version received 17.12.23; accepted 01.02.24; published 29.03.24.

Please cite as:
Chaves A, Arnáez S, García-Soriano G
The Effectiveness of a Cell Phone eHealth App in Changing Knowledge, Stigmatizing Attitudes, and Intention to Seek Help Associated With Obsessive-Compulsive Disorder: Pilot Questionnaire Study
JMIR Mhealth Uhealth 2024;12:e48027
URL: https://mhealth.jmir.org/2024/1/e48027
doi:10.2196/48027
PMID:38551629

Original Paper

Establishing a Consensus-Based Framework for the Use of Wearable Activity Trackers in Health Care: Delphi Study

Kimberley Szeto[1], BClinExPhys; John Arnold[1], PhD; Erin Marie Horsfall[2], BPTh; Madeline Sarro[2], BPTh; Anthony Hewitt[3], BAPSC, MBA; Carol Maher[1], PhD

[1]Alliance for Research in Exercise, Nutrition and Activity, Allied Health and Human Perfomance, University of South Australia, Adelaide, Australia

[2]Allied Health and Human Perfomance, University of South Australia, Adelaide, Australia

[3]Southern Adelaide Local Health Network, South Australia Health, Adelaide, Australia

Corresponding Author:
Kimberley Szeto, BClinExPhys
Alliance for Research in Exercise, Nutrition and Activity, Allied Health and Human Perfomance
University of South Australia
GPO Box 2471
Adelaide, 5001
Australia
Phone: 61 (08) 8302 2283
Email: kimberley.szeto@mymail.unisa.edu.au

Abstract

Background: Physical activity (PA) plays a crucial role in health care, providing benefits in the prevention and management of many noncommunicable diseases. Wearable activity trackers (WATs) provide an opportunity to monitor and promote PA in various health care settings.

Objective: This study aimed to develop a consensus-based framework for the optimal use of WATs in health care.

Methods: A 4-round Delphi survey was conducted, involving a panel (n=58) of health care professionals, health service managers, and researchers. Round 1 used open-response questions to identify overarching themes. Rounds 2 and 3 used 9-point Likert scales to refine participants' opinions and establish consensus on key factors related to WAT use in health care, including metrics, device characteristics, clinical populations and settings, and software considerations. Round 3 also explored barriers and mitigating strategies to WAT use in clinical settings. Insights from Rounds 1-3 informed a draft checklist designed to guide a systematic approach to WAT adoption in health care. In Round 4, participants evaluated the draft checklist's clarity, utility, and appropriateness.

Results: Participation rates for rounds 1 to 4 were 76% (n=44), 74% (n=43), 74% (n=43), and 66% (n=38), respectively. The study found a strong interest in using WATs across diverse clinical populations and settings. Key metrics (step count, minutes of PA, and sedentary time), device characteristics (eg, easy to charge, comfortable, waterproof, simple data access, and easy to navigate and interpret data), and software characteristics (eg, remote and wireless data access, access to multiple patients' data) were identified. Various barriers to WAT adoption were highlighted, including device-related, patient-related, clinician-related, and system-level issues. The findings culminated in a 12-item draft checklist for using WATs in health care, with all 12 items endorsed for their utility, clarity, and appropriateness in Round 4.

Conclusions: This study underscores the potential of WATs in enhancing patient care across a broad spectrum of health care settings. While the benefits of WATs are evident, successful integration requires addressing several challenges, from technological developments to patient education and clinician training. Collaboration between WAT manufacturers, researchers, and health care professionals will be pivotal for implementing WATs in the health care sector.

(JMIR Mhealth Uhealth 2024;12:e55254) doi:10.2196/55254

KEYWORDS

wearable activity tracker; health care; physical activity; sedentary behavior; wearable; wearables; wearable tracker; tracker; wearable technology; support; exercise; prevention; management; monitor; promote; survey; utility

Introduction

Background

Physical activity (PA) is critical in preventing and managing many noncommunicable diseases [1]. For chronic disease populations, PA can mitigate disease progression, improve physical function, and reduce the risk and burden of comorbid health conditions [2]. In the hospital setting, extremely low levels of patient PA are linked to poor outcomes, including functional decline, increased frailty and disability, and increased mortality risk [3-5]. Inactivity is also associated with higher rates of hospitalization [6-8], longer length of hospital stay [9,10], and increased risk of readmission [11,12]. Global guidelines recommend adults perform at least 150 minutes of moderate or 75 minutes of vigorous-intensity PA weekly, muscle strengthening exercise at least 2 days per week, and reduce sedentary behavior (SB) [2]. Yet just 1 in 5 adults globally meet these guidelines [13], resulting in a substantial global burden of physical inactivity, contributing to millions of premature deaths each year [1], and costing an estimated US $53.6 billion annually in health care costs and loss of productivity [14].

The need to address PA in the health care sector is clear. Indeed, organizations such as the World Health Organization [15] and the International Society for Physical Activity [16] have identified health care as a crucial setting for investment and implementation of strategies to promote PA. Health care professionals (HCPs) are strategically positioned to influence a wide range of people either living with or at risk of developing many diseases that could be alleviated or prevented through increased PA. Accordingly, PA has been endorsed as a "vital sign" that should be routinely assessed by HCPs to identify inactive individuals, prompt interventions targeted at increasing PA, and form a baseline for such interventions [17-19].

Wearable activity trackers (WATs) offer substantial potential for measuring and influencing patient PA. They have gained widespread acceptance among researchers, HCPs, and the general population. Consumer-oriented WATs, such as Fitbits, use accelerometers to measure PA, and have seen an uptick in popularity [20]. These consumer-oriented WATs may also incorporate additional sensors, like heart rate and blood oxygen saturation monitors, allowing users to track multiple health parameters via the device interface and associated smartphone applications. Compared with other methods of PA assessment, such as self-report questionnaires, WATs demonstrate superior validity, reliability, and reduced biases [21,22]. Furthermore, there is a substantial body of evidence demonstrating the effectiveness of WAT-based interventions for augmenting PA across diverse populations, with subsequent benefits like improved physical function, body composition, and blood pressure [23-25].

Despite considerable evidence demonstrating the opportunities and value of using WATs in health care settings, their routine use in health care has not yet been achieved. Nonetheless, HCPs and researchers are endeavoring to adopt WATs and are exhibiting optimism about their potential for PA measurement, informing exercise prescription and PA promotion, and fostering patient motivation and self-monitoring [26,27]. While smaller-scale efforts that have been made so far are encouraging and demonstrate interest from a spectrum of stakeholders, barriers to broader and routine use of WATs in health care persist. These barriers range from practical obstacles related to the WAT itself, such as battery life and wear issues, to clinician workload, software and data access limitations, lack of interdisciplinary support, and costs and resource concerns [26,28,29]. This highlights the need for standardized and coordinated approaches to WAT deployment in health care. A more systematic approach could facilitate the comparison and amalgamation of data across similar populations from disparate settings, circumvent recurring issues, and enhance the desired outcomes of using WATs in health care.

Objectives

This study aimed to establish a standardized and consensus-based framework for the quality use of WATs in health care settings. To do this, we sought to identify the most important metrics and device characteristics for health care–focused WAT use, identify suitable clinical populations and settings for WAT deployment, and explore stakeholders' perceptions of the most significant barriers to WAT adoption in health care, alongside potential solutions.

Methods

Study Design

A 4-round web-based Delphi study was conducted to meet the study aims between March 2021 and June 2022. This research adhered to the Conducting and Reporting of Delphi Studies (CREDES; Multimedia Appendix 1 [30]).

Ethical Considerations

This research was conducted in accordance with the Helsinki declaration of 1975, as revised in 2000. The University of South Australia's Human Research Ethics Committee granted ethical approval (approval number: 203069). Given the web-based survey design, opening the survey weblink and completing the survey were considered implied informed consent. Participants could elect to withdraw from this study if requested.

Participants and Recruitment

We recruited participants representing three stakeholder groups with experience or expertise on the topic: (1) HCPs, (2) researchers, (3) health service managers (eg, department officials or heads of services), or any combination thereof. Potential participants were identified through multiple channels: HCPs from a preceding study [26], professional associations and networks, participant referrals (ie, invitees were encouraged to share with relevant contacts), by searching for health service managers and officials from health networks (eg, government health departments), and international researchers with relevant expertise. Potential participants were emailed study information and invited to participate. A follow-up email was sent after 1 week to nonrespondents. We aimed to recruit 50 participants based on recommendations for Delphi studies with multiple stakeholder groups [31]. We anticipated similar representation from HCPs and researchers, with fewer health service managers (approximately 15% of the total sample) due to the limited

number of these positions. All participants who agreed were invited to complete all survey rounds, regardless of earlier rounds' completion.

Delphi Surveys

Four survey rounds were distributed via email (SurveyMonkey). Surveys were piloted by 5 people who were not participants (2 HCPs, 1 researcher, and 2 both HCP and researcher) to ensure clarity, timeliness, and appropriate topic exploration, with amendments made if required. Surveys took 15-25 minutes to complete, and participants had 3 weeks to complete each survey. Reminders were sent after 7 and 14 days. Results from each round were provided to participants within subsequent rounds, with summaries incorporated within the surveys for context.

Round 1

The Round 1 survey (Multimedia Appendix 2) aimed to generate items and identify broad themes on the topic. Round 1 collected data on participants' professional experience with WATs, and used 12 open-ended questions to explore perspectives on the following: clinical value of WATs, suitable clinical populations and settings, important metrics, important device characteristics, and barriers and enablers to use in health care settings. At the outset of the survey, we defined "wearable activity trackers" as "a wearable device for tracking activity-related metrics such as steps, sleep, energy expenditure (eg, calories), and in some cases, activity minutes." Open-response questions were analyzed thematically, categorized, and converted to items for the subsequent round.

Round 2

The Round 2 survey (Multimedia Appendix 3) used findings from Round 1 to refine opinions and establish consensus on 64 items across 4 sections and introduced a fifth section on software, based on its emergence as a significant topic in Round 1. The first 4 sections considered key metrics, essential WAT characteristics, and usefulness for various populations and settings. Participants rated their agreement with items on a 9-point Likert scale (eg, "it is critically important for the wearable to measure 'x'": 1=strongly disagree to 9=strongly agree), and were encouraged to provide comments on open-response questions in each section. One multiple-choice question determined participants' preferred wear site for WATs. The fifth section on software comprised multiple-choice, item rating, and open-response questions which were tailored for participants with direct WAT experience in health care.

Round 3

The Round 3 survey (Multimedia Appendix 4) comprised 3 sections and was designed using insights from earlier rounds. It sought to refine opinions and reach consensus on new items, rerate items that did not meet consensus in Round 2, and further explore barriers and potential strategies identified in Round 1. The first section required participants to rate 28 items on the same topics explored in Round 1 (6 rerating, 22 new). The second section explored software, requiring participants to rate 11 new items, and included an optional open-response question for additional comments. The third section explored barriers, and required participants to select 1 multiple-choice response for what they considered to be the most important barrier to using WATs in health care across 3 categories (patient, clinician and interdisciplinary team, and health care–system), and included open-response questions to provide suggestions for strategies to address the barriers. Barriers were explored in Round 3, given the length and item volume of the Round 2 survey.

Round 4

Between Rounds 3 and 4, three authors (KS, CM, and JA) drafted a 12-item checklist containing key elements for clinicians and service planners to consider when developing procedures for using WATs in health care settings (Multimedia Appendix 5 [26-28,32-47]). Results from Rounds 1-3 and the Consolidated Framework for Implementation Research (CFIR) [48] informed the constituents of the draft checklist. The CFIR provides a structure for approaching and evaluating intervention implementation and recognizes that to be effective, new innovations need to be adapted to "fit" the needs of a setting while retaining their "core" components. The draft checklist included "core" elements to consider when developing procedures for using WATs in different settings while allowing the user to adapt details based on their specific circumstances. Each item included prompts and explanations. During Round 4, participants evaluated the checklist items' clarity and usefulness and provided overall feedback.

Data Analysis

Response rates, participant characteristics, and responses on Likert scales and multiple-choice questions were analyzed using descriptive statistics (frequency of responses and percentages). Results were presented for each stakeholder type ("health system" and "research") and for the entire sample. Qualitative data from open-response questions were analyzed thematically and organized into emergent categories and themes, or converted to items for rating in subsequent rounds.

Consensus Agreement

Consensus was determined based on responses on 9-point Likert scales. Participants rated their agreement with statements regarding either the importance, appropriateness, or usefulness of items. Ratings were categorized as "not important/appropriate/useful" (1-3), "neutral" (4-6), and "critically important/appropriate/useful" (7-9) based on the 9-point GRADE methodology [49]. Items were scored as either:

- "Critically important/appropriate/useful": ≥75% of participants rating ≥7, and ≤15% of participants rating ≤3.
- "Somewhat important/appropriate/useful": 50%-74% rated an item ≥7
- "Not important" when <50% rated an item ≥7.

Item scores were grouped based on responses from different stakeholder groups: "healthcare" (HCPs and health-service officials) and "research" (researchers). Scores were summarized by group and for the entire sample. To establish consensus, we considered item scores from the 2 groups, and scores from the entire sample. Scores from each group were considered, as we were interested in different perspectives that the types of stakeholders may have, and if this influenced the overall rating of items.

Items reached consensus based on the following criteria:

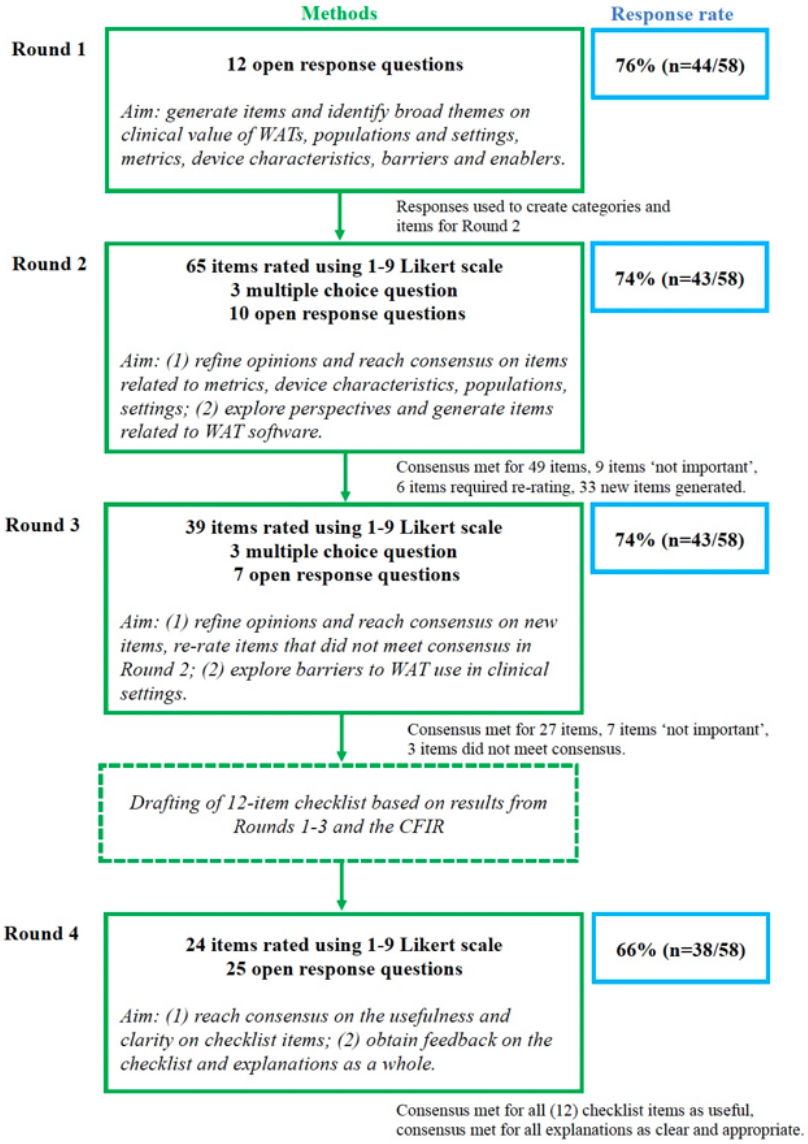
- "Critically important/appropriate/useful": when this was the score for both groups or for one group only and the overall score was "critically important/appropriate/useful."
- "Somewhat important/appropriate/useful" when this was the score for both groups or for one group and the overall rating was "somewhat important/appropriate/useful."
- "Consensus not met": the score from one group was either "critically important" or "somewhat important," and the score from the other group was "not important."
- "Not important" (omitted from exploration in subsequent rounds): when the rating from both groups was "not important."

Results

Participants

Of 82 potential participants who were directly invited to the study, 38 (46%) individuals expressed interest in participating. An additional 20 individuals contacted the research team after learning about the study via word of mouth, yielding a total sample of 58. Response rates for rounds 1 to 4 were 76% (n=44), 74% (n=43), 74% (n=43), and 66% (n=38) respectively, with 52 (90%) participants responding to at least 1 round, and 28 (48%) participants completing all 4 rounds. An overall summary of the structure of the Delphi rounds, item consensus, and response rates is shown in Figure 1.

Figure 1. Overview of Delphi process. WAT: wearable activity tracker; CFIR: Consolidated Framework for Implementation Research.



Round 1

Of the 44 participants who completed Round 1, 8 participants were health service managers, 24 participants were HCPs, and 17 participants were researchers (Table 1). All participants considered WATs to be useful in health care settings, predominantly for the purposes of monitoring patient activities (ie, PA, SB, and sleep), intervening in patient PA, and monitoring physiological parameters (ie, heart rate). Participants provided a wide variety of responses for suitable populations and settings (eg, people with metabolic conditions, rehabilitation, inpatient, and outpatient). Various PA and physiological metrics were suggested, with 26 discrete metrics identified. The most frequently reported metrics were daily step

count (n=27), daily PA minutes (n=20), and heart rate (n=19). A total of 28 device characteristics were reported as important for clinical settings, which were categorized under battery and charging (eg, long battery life), wear (eg, waterproof), device interface (eg, real-time feedback), data (eg, easy data access), and other (eg, affordable cost). Qualitative analysis suggested that the selection of metrics and characteristics would be influenced by the purpose of use, the population and setting, and patient factors such as goals and level of mobility. Numerous barriers and enablers to using WATs in health care settings were reported, which were grouped into 4 categories: device-related (eg, battery: short battery life), patient-related (eg, patient suitability: cognition and mobility), clinician-/multidisciplinary team-related (eg, time constraints and competing demands: setting up devices) and system-level (eg, lack of procedures and systems: distributing and managing devices). Enablers were generally the inverse of barriers (for instance, if short battery life was a barrier, long battery life was an enabler). A full summary of results from Round 1 is available in Multimedia Appendix 2.

Table 1. Demographics of the cohort (n=44).

Characteristics	Round 1 responses, n (%)
Type of participant	
Health care professional	24 (55)
Health service managers	8 (18)
Researcher	17 (39)
Country	
Australia	41 (94)
United States	1 (2)
United Kingdom	1 (2)
Spain	1 (2)
Years of experience in field	
Up to 5	3 (7)
>5-10	6 (14)
>10-20	14 (32)
>20	21 (47)
Clinical field (n=33)	
Physiotherapy	23 (70)
Exercise physiology	2 (5)
Other allied health	2 (5)
Medical doctor	3 (9)
Nurse	1 (3)
Other	2 (5)

Round 2

Round 2 comprised five sections. The first 4 sections required item rating to indicate agreement with the importance of different metrics and characteristics of WATs for clinical use, and to indicate agreement with the appropriateness of using WATs with different clinical populations and settings.

The consensus was reached for 26/64 items as being “critically important” (metrics: 2/14, device characteristics: 8/19), or “highly appropriate” (populations: 6/10, settings and purposes: 10/21), and 23/64 items as being “somewhat important” (metrics: 2/14, device characteristics: 8/19), or “somewhat appropriate” (populations: 3/10, settings and purposes: 10/21; Tables 2 and 3). Nine items reached an outcome of “not important.” Six items did not reach consensus and were rerated in the following round. Interestingly, despite step count being the most frequently reported useful metric in Round 1, 74% of participants overall rated it as “critically important” in Round 2, leading to an overall score of “somewhat important” (<75% of participants rating ≥7). Given its frequent mention and relevance in research [23,50], the team decided this item warranted rerating in the next round. Most participants (72%) considered the wrist to be the most appropriate wear site, with 18 participants providing open-response comments on the wear site. Comments mostly related to wear site affecting the validity and reliability of data, as well as patient engagement and preferences, with themes identified that wear site should consider both the purpose of using WATs (ie, is patient engagement the priority), and the patient context (ie, walking speed; Multimedia Appendix 3).

Table 2. Device-related items that reached consensus in Rounds 2 and 3.

Item	Percentage agreement (≥7 on 9-point Likert scale)	
	Round 2 (n=43)	Round 3 (n=43)
Metrics		
Critically important (≥75% rated ≥7) ≥7		
Daily minutes of PA[a]	95	N/A[b]
Daily minutes of SB[c]	81	N/A
Daily step count	74	86
Somewhat important (50%-74% rated ≥7)		
Intensity of PA (heart rate zones)	56	N/A
Details of PA intensities (eg, light/moderate/vigorous)	__[d]	74
All daily activities (physical activity, sedentary time, sleep)	58	N/A
Sit to stand transitions	—	58
Position of SB (eg, sitting/standing/lying down)	—	56
Heart rate	44[e]	58[e]
On versus off body time (wear time)	—	60
Characteristics		
Critically important (≥75% rated ≥7)		
Easy to charge	95	N/A
Comfortable to wear day and night	93	N/A
Water resistant	81	N/A
Easy to clean and disinfect	81	N/A
Simple data syncing process	98	N/A
Simple data download or export process	91	N/A
Store at least 5 days of data without syncing or downloading	88	N/A
Easy to navigate and set up	93	N/A
Provide data that is easy to interpret	—	88
Wrist-worn is most appropriate for patient acceptability and compliance	MCQ[f]	77
Adaptable wear site for different populations and individuals (eg, those with walking frames)	—	84
Somewhat important (50%-74% rated ≥7)		
Battery that lasts at least 2 days on a single charge	74	N/A
Battery that lasts at least 5 days on a single charge	63	N/A
Quick to charge (eg, reach full charge in 1 hour)	70	N/A
Measure steps in slow ambulators	72	N/A
Ability to set personalized goals	63	N/A
Simple data interface (eg, interpret key data at a glance)	74	N/A
Real-time feedback is provided on the interface	67	N/A
Data can be accessed remotely by clinician	56	N/A
Aesthetically pleasing	51[e]	50
Ability to set specific reminders	—	56
Ability to select between different metrics for viewing	—	70
Ability to wear at different bodily locations	49[e]	63

Item	Percentage agreement (≥7 on 9-point Likert scale)	
	Round 2 (n=43)	Round 3 (n=43)
Wear site may need to be adapted for different purposes (eg, assessment vs intervention)	—	56

[a]PA: physical activity.

[b]N/A: not applicable (agreement met in prior round).

[c]SB: sedentary behavior.

[d]not rated (identified from this round).

[e]Disagreement between groups.

[f]MCQ: multiple choice question.

Table 3. Wearable activity tracker usefulness items that reached consensus in Rounds 2 and 3.

Item	Percentage agreement (≥7 on 9-point Likert scale)	
	Round 2 (n=43)	Round 3 (n=43)
Patient populations		
Highly appropriate (≥75% rated ≥7)		
Cardiovascular	95	N/A[a]
Pulmonary	88	N/A
Metabolic (eg, obesity, bariatrics, and diabetes)	98	N/A
Mixed rehabilitation	86	N/A
Chronic pain	79	N/A
Older adults (geriatrics and aged care)	88	N/A
Somewhat appropriate (50%-74% rated ≥7)		
Orthopedic	74	N/A
Neurological	74	N/A
Pediatric	65	N/A
Oncology	—[b]	70
Mental Health	—	65
Disability sector	—	72
Measuring activity		
Highly appropriate (≥75% rated ≥7)		
During home-based rehabilitation	95	N/A
Following a hospital admission (after discharge)	84	N/A
Outpatient settings	88	N/A
Community-based settings (inside or outside the home)	95	N/A
Residential aged-care (home-based or live-in facilities)	81	N/A
Somewhat appropriate (50%-74% rated ≥7)		
Prior to an elective hospital admission	60	N/A
During an inpatient hospital admission	70	N/A
Measuring physiological parameter		
Highly appropriate (≥75% rated≥7)		
During home-based rehabilitation	76	N/A
Somewhat appropriate (50%-74% rated ≥7)		
During an inpatient hospital admission	56	N/A
Following a hospital admission (after discharge)	70	N/A
Outpatient settings	65	N/A
Community-based settings (inside or outside the home)	72	N/A
Residential aged-care (home-based or live-in facilities)	65	N/A
Prior to an elective hospital admission	47[c]	65
Intervening on activity		
Highly appropriate (≥75% rated ≥7)		
During home-based rehabilitation	86	N/A
Outpatient settings	77	N/A
Community-based settings (inside or outside the home)	86	N/A

Item	Percentage agreement (≥7 on 9-point Likert scale)	
	Round 2 (n=43)	Round 3 (n=43)
Following a hospital admission (after discharge)	77	N/A
Somewhat appropriate (50%-74% rated ≥7)		
Prior to an elective hospital admission	63	N/A
During an inpatient hospital admission	58	N/A
Residential aged-care (home-based or live-in facilities)	72	N/A

[a]N/A: not applicable (agreement met in prior round).

[b]Not rated (identified from this round).

[c]Disagreement between groups.

Section 5 focused on the software aspects of using WATs in health care settings. Participants with direct experience were specifically asked to rate their satisfaction with the existing WAT software they had used in health care settings. Of the 32 respondents to this item, most (59%) rated their software experience as neutral or dissatisfied. Twenty-nine participants described the software they had used and how they thought software could be improved to better meet the needs of health care settings. The majority reported using proprietary software linked to the device or reading outputs from the device interface, and some either invested in third-party software or developed their own. Those who described difficulties generally reported that software was difficult to use, time-consuming, and expensive (eg, logging in and out of separate accounts to view data on consumer WATs, or needing to purchase specialized software to download data). Participants provided numerous suggestions regarding the requirements of WAT software to meet the needs of health care settings, which were converted into items for consensus rating in Round 3. A full summary of results for Round 2 is available in Multimedia Appendix 3.

Round 3

Round 3 comprised 3 sections. The first section covered metrics, device characteristics, populations, and settings and purposes, and focused on rating new items and rerating items that did not meet consensus in Round 2. The second section focused on software, and involved rating items generated from Round 2 to identify "ideal" software features. The third section focused on barriers to WAT use in health care.

Consensus was reached for 11/39 items as being "critically important" (metrics: 1/13, device characteristics: 3/11, software: 7/11), and 16/39 items as being "somewhat important" (metrics: 3/13, device characteristics: 4/11, software: 4/11) or "somewhat appropriate" (patient populations: 3/3, settings and purposes: 1/1; Tables 2-4). Four items did not meet consensus (metrics: 2/13, device characteristics: 2/11). Seven items reached an outcome of "not important."

Table 4. Ideal software items that reached consensus in Round 3 (n=43).

Item	Percentage agreement (≥7 on 9-point Likert scale)
Critically important (≥75% rated ≥7)	
Present relevant data at varying levels of simplicity	93
Centralized access of patient data	84
Access to multiple patients' data simultaneously	79
Wireless data upload or export	93
Batch data downloads	84
Access to raw data sets	77
Ability to conduct separate or additional analyses to those provided by the device	79
Somewhat important (50%-74% rated ≥7)	
Capacity to input self-report data (eg, RPE[a], pain, and fatigue VAS[b])	53
Provide useful and relevant data on the wearable interface (eg, without further analysis needed)	70
Capacity to download and access data sets instantly	72
Automatically present data in different formats (eg, graphs, tables, summary scores without having to edit or analyze data sets)	65

[a]RPE: rate of perceived exertion.

[b]VAS: visual analog scale.

The barriers section of the survey asked participants to identify the most significant obstacles to WAT use in health care across 3 categories: patient-related, clinician– and interdisciplinary team–related, and health care system. Results were presented by group (health care and research participants) and overall. No single patient-related barrier dominated, but there were nuances between health care and research participants (Multimedia Appendix 4). Most health care participants (37%) considered "clinical unsuitability" to be the most important, and most research participants (37%) considered "patient reluctance" to be the most important. Participants from both groups considered "time constraints and competing demands" to be the most important clinician-related barrier (overall: 51%, health system: 53%, research: 50%), and "lack of funding and resources" to be the most important health care system–related barrier (overall: 58%, health system: 63%, research: 54%). Participants provided strategies to address these barriers, with the top 5 strategy categories, in order of frequency, being: clinician information and support (19 mentions); standardized approach (17 mentions), broader involvement of teams or families and caregivers (15 mentions), top-down support (14 mentions), and patient education and support and improved funding and resourcing (13 mentions each). A full summary of results for Round 3 is available in Multimedia Appendix 4.

Round 4

A draft 12-item user checklist was developed based on insights from Rounds 1-3 and guided by the CFIR. This checklist was designed to assist HCPs and health service managers in planning to implement WATs across various health care settings. During Round 4, participants were presented with the draft checklist with the goal of gauging consensus on the usefulness of individual items and assessing the clarity and appropriateness of explanations. There was clear support for individual items and explanations, with every item on the checklist reaching consensus as being "very useful," and all explanations reached consensus as being "very clear and appropriate" (Multimedia Appendix 5 [26-28,32-47]). Fourteen participants provided comments on the checklist as a whole, which was predominantly positive feedback. A few comments related to the presentation of the final product, shortcomings, and the clarity or amount of information. Twenty-three participants provided feedback on individual items and explanations, which was mostly general comments related to the content of items (eg, suggestions to provide more detail and citations regarding the accuracy and validity of devices in different contexts). Other common responses included suggestions for additional information, suggestions to condense information, specific suggestions on wording or formatting, and feedback on the clarity of wording. All comments for individual items and the expanded checklist including elaboration statements for each of the 12 items are provided in Multimedia Appendix 5 [26-28,32-47].

Discussion

Principal Results

This Delphi study brought together HCPs, health service managers, and researchers to gather expert opinions regarding the application of WATs in health care settings. Findings revealed that participants believed that WATs offer utility for monitoring and improving patient PA in a wide variety of clinical populations and health care settings. The study identified various metrics and device characteristics as critical, with the specific choice of metrics being contingent on the clinical context and patient factors. Software was acknowledged as a vital element, with most participants being unsatisfied with the currently available software. Key barriers to WAT adoption in health care were recognized in relation to patients, clinicians, and health care system categories, with time constraints and resource limitations being major obstacles. To overcome these, participants suggested strategies such as clinician support, standardized approaches, and improved funding. The study culminated in the development of a 12-item checklist to assist HCPs and service planners in successfully integrating WATs within various health care settings.

Potential for Widespread Adoption of WATs in Health Care

Our findings highlight widespread interest in using WATs in a very wide range of clinical populations, settings, and purposes. Various clinical groups were identified as suitable for WAT applications, across different health care contexts. This multifaceted application of WATs is corroborated by existing literature, with previous studies using WATs with post-surgical patients and acutely ill patients in hospitals for daily step count monitoring to predict hospital readmission and length of stay [10,11], as well as using WATs in various community-dwelling chronic disease populations [24,51,52] and various hospitalized populations [25] for promoting patient PA. Together, these findings highlight the versatility of using WATs in health care settings and their promising role in advancing patient care.

Toward More Systematic Integration of WATs in Health Care

This study identified a desire for more standardization in WAT use in health care, yet with such a range of different potential applications for WATs, a universal approach seems unfeasible. Nevertheless, there was consensus on essential components for diverse WAT applications. Out of 45 important elements identified, 21 were classified as "critically important," encompassing 3 metrics, 11 device characteristics, and 7 software characteristics. Furthermore, an additional 24 elements were classified as "somewhat important." Recognizing these pivotal components can steer HCPs and potential users to select devices and software that meets their clinical needs. These elements were incorporated into the user-oriented Wearable Activity Tracker Checklist for Health care (WATCH) developed from this study, which is presented in the companion article [53]. Given that the barriers this study identified were similar to those described in international studies [28,29], it seems likely that the core elements we identified will be relevant for applications in different locations.

Strengths and Limitations

This study is the first to use an expert consensus approach to advance the use of WATs in health care settings. It adhered to rigorous Delphi methodology, including defining consensus, having clear criteria for accepting, rerating, and omitting items,

defining the number of rounds to be performed a priori and piloting surveys before sending them to participants [54]. The average response rate of 73% was another strength of this study, given the large number of items and larger size of the panel, both of which have been associated with lower response rates for Delphi studies [55].

This study has several limitations that warrant consideration. First, a limitation of the Delphi methodology is that it uses a limited sample of participants to represent various stakeholder groups of interest. It is possible that the participants in this Delphi study may not represent the broader population of HCPs, health service managers, and researchers with expertise and experience in the application of WATs in health care. While participants in this Delphi study represented various professions and had diverse experiences with WATs in different populations and settings, the sample was predominantly Australian. This may limit the generalizability of our findings to other contexts, as digital technology and data landscapes vary significantly across countries. These variations can influence the accessibility, usability, and integration of wearable technologies and associated software in health care practices. For instance, differences in regulatory environments, data privacy laws, and technological infrastructure can impact the deployment and effectiveness of WATs. Therefore, HCPs from different jurisdictions may hold different attitudes and perspectives. A further limitation of the Delphi methodology is that, as a consensus-driven approach, it can potentially overlook important opinions or knowledge if they are not widely held by the sample [31]. In this study, some metrics were considered more important than others for clinical use. However, the combination of multiple metrics (and metrics considered “less important”) should not be overlooked in gaining other important insights from WAT data. For example, heart rate data can be used to determine if devices are being worn by patients. This Delphi study did not involve patients, who are also end users of WATs, thereby lacking their perspectives. Patients will be an important stakeholder group in future research as the integration of WATs in health care becomes more widespread. Finally, as WAT technology continues to advance rapidly, the considerations and factors explored in this study may require updating.

Implications

Our study identified various barriers to the use of WATs in health care that will need to be addressed to support widespread implementation. First, patient-related barriers included clinical unsuitability for certain conditions, and patient reluctance, often rooted in apprehension or lack of understanding. Addressing this will require comprehensive patient education and selecting WATs suitable for specific patient needs. Additionally, clinician-related barriers like time constraints necessitate streamlining of WAT integration processes. Incorporating training programs for HCPs is likely to enhance their competency in efficiently using WATs. At the system level, financial constraints and lack of resources appear to be significant obstacles. Strategically advocating for funding, possibly through demonstrating the long-term cost benefits of WATs in patient care, may be instrumental to addressing this. Lastly, interoperability and data integration into electronic health records will be pivotal in the future for clinicians to effectively access and interpret data. The study also reveals a substantial gap between the ideal and actual capabilities of currently available WAT devices and software in health care settings. Current options fall short in areas deemed critical by participants, particularly in software features, wearability, and data accuracy. This highlights an urgent call for the innovation of WAT technology and software tailored specifically for health care applications. Such innovations may include the integration of behavior change and gamification techniques (eg, goal-setting, feedback loops, rewards, and nudges) into devices to foster use and engagement, potentially leading to increased effectiveness of PA promotion efforts. A collaborative approach between WAT manufacturers, researchers, and HCPs will be imperative to develop solutions that not only address the practical necessities of health care settings but also rigorously uphold data privacy standards required for handling patient information.

Conclusions

This Delphi study offers valuable insights into the prospects and challenges of integrating WATs in health care settings for PA promotion. The collective perspectives of HCPs, health service managers, and researchers underscore the broad potential of WATs across diverse clinical scenarios. Yet, for WATs to be fully effective in patient care, several hurdles must be addressed, ranging from patient education to technological advancements specific to the health care sector. The evolving nature of WAT technology will necessitate continuous collaboration and re-evaluation. As the health care landscape seeks more personalized and data-driven approaches, the integration of WATs presents a promising avenue to enhance PA promotion and patient outcomes, optimize clinical processes, and elevate the overall quality of health care delivery.

Acknowledgments

This project received no specific funding. KS was supported by the Australian Government Research Training Program domestic stipend as a postgraduate student at the University of South Australia. CM is funded by a Medical Research Future Fund Investigator Grant (GNT1193862). The authors thank all of the participants who were involved in this study for sharing their time, expertise, and experiences. The authors also thank those who shared this study among their networks, helping us to identify participants who were involved with this study.

Authors' Contributions

KS, JA, and CM conceived and designed the study. All authors contributed to recruitment and the development of the study protocol and survey instruments. KS, EMH, MS, JA, and CM carried out data analysis. KS, JA, and CM contributed to the checklist draft. KS drafted the full manuscript and all authors reviewed the drafts and approved the final submission.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CREDES checklist.
[DOCX File , 17 KB - mhealth_v12i1e55254_app1.docx]

Multimedia Appendix 2
Round 1 survey and results.
[PDF File (Adobe PDF File), 366 KB - mhealth_v12i1e55254_app2.pdf]

Multimedia Appendix 3
Round 2 survey and results.
[PDF File (Adobe PDF File), 464 KB - mhealth_v12i1e55254_app3.pdf]

Multimedia Appendix 4
Round 3 survey and results.
[PDF File (Adobe PDF File), 359 KB - mhealth_v12i1e55254_app4.pdf]

Multimedia Appendix 5
Round 4 survey and results.
[PDF File (Adobe PDF File), 508 KB - mhealth_v12i1e55254_app5.pdf]

References

1. Katzmarzyk PT, Friedenreich C, Shiroma EJ, Lee I. Physical inactivity and non-communicable disease burden in low-income, middle-income and high-income countries. Br J Sports Med 2022;56(2):101-106. [doi: 10.1136/bjsports-2020-103640] [Medline: 33782046]
2. Bull FC, Al-Ansari SS, Biddle S, Borodulin K, Buman MP, Cardon G, et al. World health organization 2020 guidelines on physical activity and sedentary behaviour. Br J Sports Med 2020;54(24):1451-1462 [FREE Full text] [doi: 10.1136/bjsports-2020-102955] [Medline: 33239350]
3. Gill TM, Gahbauer EA, Han L, Allore HG. The relationship between intervening hospitalizations and transitions between frailty states. J Gerontol A Biol Sci Med Sci 2011;66(11):1238-1243 [FREE Full text] [doi: 10.1093/gerona/glr142] [Medline: 21852286]
4. Brown CJ, Friedkin RJ, Inouye SK. Prevalence and outcomes of low mobility in hospitalized older patients. J Am Geriatr Soc 2004;52(8):1263-1270. [doi: 10.1111/j.1532-5415.2004.52354.x] [Medline: 15271112]
5. Gill TM, Allore HG, Holford TR, Guo Z. Hospitalization, restricted activity, and the development of disability among older persons. JAMA 2004;292(17):2115-2124. [doi: 10.1001/jama.292.17.2115] [Medline: 15523072]
6. Esteban C, Arostegui I, Aburto M, Moraza J, Quintana JM, Aizpiri S, et al. Influence of changes in physical activity on frequency of hospitalization in chronic obstructive pulmonary disease. Respirology 2014;19(3):330-338. [doi: 10.1111/resp.12239] [Medline: 24483954]
7. Lu Y, Ballew SH, Kwak L, Selvin E, Kalbaugh CA, Schrack JA, et al. Physical activity and subsequent risk of hospitalization with peripheral artery disease and critical limb ischemia in the ARIC study. J Am Heart Assoc 2019;8(21):e013534 [FREE Full text] [doi: 10.1161/JAHA.119.013534] [Medline: 31642360]
8. Koo P, Gjelsvik A, Choudhary G, Wu WC, Wang W, McCool FD, et al. Prospective association of physical activity and heart failure hospitalizations among black adults with normal ejection fraction: the jackson heart study. J Am Heart Assoc 2017;6(9):e006107 [FREE Full text] [doi: 10.1161/JAHA.117.006107] [Medline: 28882818]
9. Fisher SR, Kuo YF, Graham JE, Ottenbacher KJ, Ostir GV. Early ambulation and length of stay in older adults hospitalized for acute illness. Arch Intern Med 2010;170(21):1942-1943 [FREE Full text] [doi: 10.1001/archinternmed.2010.422] [Medline: 21098357]
10. Daskivich TJ, Houman J, Lopez M, Luu M, Fleshner P, Zaghiyan K, et al. Association of wearable activity monitors with assessment of daily ambulation and length of stay among patients undergoing major surgery. JAMA Netw Open 2019;2(2):e187673 [FREE Full text] [doi: 10.1001/jamanetworkopen.2018.7673] [Medline: 30707226]

11. Fisher SR, Graham JE, Ottenbacher KJ, Deer R, Ostir GV. Inpatient walking activity to predict readmission in older adults. Arch Phys Med Rehabil 2016;97(9 Suppl):S226-S231 [FREE Full text] [doi: 10.1016/j.apmr.2015.09.029] [Medline: 27264549]
12. Low CA, Bovbjerg DH, Ahrendt S, Choudry MH, Holtzman M, Jones HL, et al. Fitbit step counts during inpatient recovery from cancer surgery as a predictor of readmission. Ann Behav Med 2018;52(1):88-92 [FREE Full text] [doi: 10.1093/abm/kax022] [Medline: 29538623]
13. Garcia-Hermoso A, López-Gil JF, Ramírez-Vélez R, Alonso-Martínez AM, Izquierdo M, Ezzatvar Y. Adherence to aerobic and muscle-strengthening activities guidelines: a systematic review and meta-analysis of 3.3 million participants across 32 countries. Br J Sports Med 2023;57(4):225-229. [doi: 10.1136/bjsports-2022-106189] [Medline: 36418149]
14. Ding D, Lawson KD, Kolbe-Alexander TL, Finkelstein EA, Katzmarzyk PT, van Mechelen W, Lancet Physical Activity Series 2 Executive Committee. The economic burden of physical inactivity: a global analysis of major non-communicable diseases. Lancet 2016;388(10051):1311-1324. [doi: 10.1016/S0140-6736(16)30383-X] [Medline: 27475266]
15. Organization WH. Global Action Plan on Physical Activity 2018-2030: More Active People for a Healthier World. USA: World Health Organization; 2019.
16. Milton K, Cavill N, Chalkley A, Foster C, Gomersall S, Hagstromer M, et al. Eight investments that work for physical activity. J Phys Act Health 2021;18(6):625-630 [FREE Full text] [doi: 10.1123/jpah.2021-0112] [Medline: 33984836]
17. Cowan RE. Exercise is medicine initiative: physical activity as a vital sign and prescription in adult rehabilitation practice. Arch Phys Med Rehabil 2016;97(9 Suppl):S232-S237. [doi: 10.1016/j.apmr.2016.01.040] [Medline: 27470321]
18. Sallis R. Developing healthcare systems to support exercise: exercise as the fifth vital sign. Br J Sports Med 2011;45(6):473-474. [doi: 10.1136/bjsm.2010.083469] [Medline: 21292925]
19. Lobelo F, Rohm Young D, Sallis R, Garber MD, Billinger SA, Duperly J, et al. Routine assessment and promotion of physical activity in healthcare settings: a scientific statement from the American heart association. Circulation 2018;137(18):e495-e522 [FREE Full text] [doi: 10.1161/CIR.0000000000000559] [Medline: 29618598]
20. Forecast wearables unit shipments worldwide from 2014-2024. Statista. 2021. URL: https://tinyurl.com/36kkr829 [accessed 2024-07-26]
21. Prince SA, Adamo KB, Hamel M, Hardt J, Connor Gorber S, Tremblay M. A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. Int J Behav Nutr Phys Act 2008;5:56 [FREE Full text] [doi: 10.1186/1479-5868-5-56] [Medline: 18990237]
22. Prince SA, Cardilli L, Reed JL, Saunders TJ, Kite C, Douillette K, et al. A comparison of self-reported and device measured sedentary behaviour in adults: a systematic review and meta-analysis. Int J Behav Nutr Phys Act 2020;17(1):31 [FREE Full text] [doi: 10.1186/s12966-020-00938-3] [Medline: 32131845]
23. Ferguson T, Olds T, Curtis R, Blake H, Crozier AJ, Dankiw K, et al. Effectiveness of wearable activity trackers to increase physical activity and improve health: a systematic review of systematic reviews and meta-analyses. Lancet Digit Health 2022;4(8):e615-e626 [FREE Full text] [doi: 10.1016/S2589-7500(22)00111-X] [Medline: 35868813]
24. Franssen WMA, Franssen GHLM, Spaas J, Solmi F, Eijnde BO. Can consumer wearable activity tracker-based interventions improve physical activity and cardiometabolic health in patients with chronic diseases? a systematic review and meta-analysis of randomised controlled trials. Int J Behav Nutr Phys Act 2020;17(1):57 [FREE Full text] [doi: 10.1186/s12966-020-00955-2] [Medline: 32393357]
25. Szeto K, Arnold J, Singh B, Gower B, Simpson CEM, Maher C. Interventions using wearable activity trackers to improve patient physical activity and other outcomes in adults who are hospitalized: a systematic review and meta-analysis. JAMA Netw Open 2023;6(6):e2318478 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.18478] [Medline: 37318806]
26. Maher C, Szeto K, Arnold J. The use of accelerometer-based wearable activity monitors in clinical settings: current practice, barriers, enablers, and future opportunities. BMC Health Serv Res 2021;21(1):1064 [FREE Full text] [doi: 10.1186/s12913-021-07096-7] [Medline: 34625076]
27. Wu M, Luo J. Wearable technology applications in healthcare: a literature review. Online J Nurs Inform 2019;23(3) [FREE Full text]
28. Keogh A, Taraldsen K, Caulfield B, Vereijken B. It's not about the capture, it's about what we can learn": a qualitative study of experts' opinions and experiences regarding the use of wearable sensors to measure gait and physical activity. J Neuroeng Rehabil 2021;18(1):78 [FREE Full text] [doi: 10.1186/s12984-021-00874-8] [Medline: 33975600]
29. Braakhuis HEM, Bussmann JBJ, Ribbers GM, Berger MAM. Wearable activity monitoring in day-to-day stroke care: a promising tool but not widely used. Sensors (Basel) 2021;21(12):4066 [FREE Full text] [doi: 10.3390/s21124066] [Medline: 34204824]
30. Jünger S, Payne SA, Brine J, Radbruch L, Brearley SG. Guidance on conducting and REporting DElphi studies (CREDES) in palliative care: recommendations based on a methodological systematic review. Palliat Med 2017;31(8):684-706. [doi: 10.1177/0269216317690685] [Medline: 28190381]
31. Hsu CC, Sandford BA. The delphi technique: making sense of consensus. Pract Assess Res Eval 2007;12(1):10 [FREE Full text] [doi: 10.7275/pdz9-th90]

32. Duncan M, Murawski B, Short CE, Rebar AL, Schoeppe S, Alley S, et al. Activity trackers implement different behavior change techniques for activity, sleep, and sedentary behaviors. Interact J Med Res 2017 Aug 14;6(2):e13 [FREE Full text] [doi: 10.2196/ijmr.6685] [Medline: 28807889]
33. Wu H, Gal R, van Sleeuwen NC, Brombacher AC, IJsselsteijn WA, May AM, et al. Breast cancer survivors' experiences with an activity tracker integrated into a supervised exercise program: qualitative study. JMIR Mhealth Uhealth 2019 Feb 21;7(2):e10820 [FREE Full text] [doi: 10.2196/10820] [Medline: 30789349]
34. Ummels D, Beekman E, Moser A, Braun SM, Beurskens AJ. Patients' experiences with commercially available activity trackers embedded in physiotherapy treatment: a qualitative study. Disabil Rehabil 2020 Nov;42(23):3284-3292. [doi: 10.1080/09638288.2019.1590470] [Medline: 30973026]
35. Farina N, Sherlock G, Thomas S, Lowry RG, Banerjee S. Acceptability and feasibility of wearing activity monitors in community-dwelling older adults with dementia. Int J Geriatr Psychiatry 2019 Apr;34(4):617-624. [doi: 10.1002/gps.5064] [Medline: 30701592]
36. Straiton N, Alharbi M, Bauman A, Neubeck L, Gullick J, Bhindi R, et al. The validity and reliability of consumer-grade activity trackers in older, community-dwelling adults: a systematic review. Maturitas 2018 Jun;112:85-93 [FREE Full text] [doi: 10.1016/j.maturitas.2018.03.016] [Medline: 29704922]
37. Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. Int J Behav Nutr Phys Act 2015 Dec 18;12:159 [FREE Full text] [doi: 10.1186/s12966-015-0314-1] [Medline: 26684758]
38. Simpson LA, Eng JJ, Klassen TD, Lim SB, Louie DR, Parappilly B, et al. Capturing step counts at slow walking speeds in older adults: comparison of ankle and waist placement of measuring device. J Rehabil Med 2015 Oct 05;47(9):830-835 [FREE Full text] [doi: 10.2340/16501977-1993] [Medline: 26181670]
39. Gaz DV, Rieck TM, Peterson NW, Ferguson JA, Schroeder DR, Dunfee HA, et al. Determining the validity and accuracy of multiple activity-tracking devices in controlled and free-walking conditions. Am J Health Promot 2018 Nov;32(8):1671-1678. [doi: 10.1177/0890117118763273] [Medline: 29558811]
40. Farmer C, van den Berg ME, Vuu S, Barr CJ. A study of the accuracy of the Fitbit Zip in measuring steps both indoors and outdoors in a mixed rehabilitation population. Clin Rehabil 2022 Jan;36(1):125-132. [doi: 10.1177/02692155211035293] [Medline: 34313149]
41. Ferguson T, Rowlands AV, Olds T, Maher C. The validity of consumer-level, activity monitors in healthy adults worn in free-living conditions: a cross-sectional study. Int J Behav Nutr Phys Act 2015 Mar 27;12:42 [FREE Full text] [doi: 10.1186/s12966-015-0201-9] [Medline: 25890168]
42. Henriksen A, Sand AS, Deraas T, Grimsgaard S, Hartvigsen G, Hopstock L. Succeeding with prolonged usage of consumer-based activity trackers in clinical studies: a mixed methods approach. BMC Public Health 2020 Aug 27;20(1):1300 [FREE Full text] [doi: 10.1186/s12889-020-09406-w] [Medline: 32854671]
43. Hardcastle SJ, Galliott M, Lynch BM, Nguyen NH, Cohen PA, Mohan GR, et al. Acceptability and utility of, and preference for wearable activity trackers amongst non-metropolitan cancer survivors. PLoS One 2018;13(12):e0210039 [FREE Full text] [doi: 10.1371/journal.pone.0210039] [Medline: 30596781]
44. Smuck M, Odonkor CA, Wilt JK, Schmidt N, Swiernik M. The emerging clinical role of wearables: factors for successful implementation in healthcare. NPJ Digit Med 2021 Mar 10;4(1):45 [FREE Full text] [doi: 10.1038/s41746-021-00418-3] [Medline: 33692479]
45. Colley R, Connor Gorber S, Tremblay MS. Quality control and data reduction procedures for accelerometry-derived measures of physical activity. Health Rep 2010 Mar;21(1):63-69 [FREE Full text] [Medline: 20426228]
46. Campos C, DePaul VG, Knorr S, Wong JS, Mansfield A, Patterson KK. Validity of the ActiGraph activity monitor for individuals who walk slowly post-stroke. Top Stroke Rehabil 2018 May;25(4):295-304. [doi: 10.1080/10749357.2018.1446487] [Medline: 29557275]
47. van der Weegen S, Verwey R, Spreeuwenberg M, Tange H, van der Weijden T, de Witte L. It's LiFe! mobile and web-based monitoring and feedback tool embedded in primary care increases physical activity: a cluster randomized controlled trial. J Med Internet Res 2015 Jul 24;17(7):e184 [FREE Full text] [doi: 10.2196/jmir.4579] [Medline: 26209025]
48. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci 2009;4:50 [FREE Full text] [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]
49. Guyatt GH, Oxman AD, Kunz R, Atkins D, Brozek J, Vist G, et al. GRADE guidelines: 2. framing the question and deciding on important outcomes. J Clin Epidemiol 2011;64(4):395-400. [doi: 10.1016/j.jclinepi.2010.09.012] [Medline: 21194891]
50. Chan A, Chan D, Lee H, Ng CC, Yeo AHL. Reporting adherence, validity and physical activity measures of wearable activity trackers in medical research: a systematic review. Int J Med Inform 2022;160:104696 [FREE Full text] [doi: 10.1016/j.ijmedinf.2022.104696] [Medline: 35121356]
51. Hodkinson A, Kontopantelis E, Adeniji C, van Marwijk H, McMillian B, Bower P, et al. Interventions using wearable physical activity trackers among adults with cardiometabolic conditions: a systematic review and meta-analysis. JAMA Netw Open 2021;4(7):e2116382 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.16382] [Medline: 34283229]

52. Braakhuis HEM, Berger MAM, Bussmann JBJ. Effectiveness of healthcare interventions using objective feedback on physical activity: a systematic review and meta-analysis. J Rehabil Med 2019;51(3):151-159 [FREE Full text] [doi: 10.2340/16501977-2522] [Medline: 30843082]
53. Szeto K, Arnold J, Maher C. The wearable activity tracker checklist for healthcare (WATCH): a 12-point guide for the implementation of wearable activity trackers in healthcare. Int J Behav Nutr Phys Act 2024;21(1):30 [FREE Full text] [doi: 10.1186/s12966-024-01567-w] [Medline: 38481238]
54. Diamond IR, Grant RC, Feldman BM, Pencharz PB, Ling SC, Moore AM, et al. Defining consensus: a systematic review recommends methodologic criteria for reporting of delphi studies. J Clin Epidemiol 2014;67(4):401-409. [doi: 10.1016/j.jclinepi.2013.12.002] [Medline: 24581294]
55. Gargon E, Crew R, Burnside G, Williamson PR. Higher number of items associated with significantly lower response rates in COS Delphi surveys. J Clin Epidemiol 2019;108:110-120 [FREE Full text] [doi: 10.1016/j.jclinepi.2018.12.010] [Medline: 30557677]

Abbreviations

CFIR: Consolidated Framework for Implementation Research
CREDES: Conducting and Reporting of Delphi Studies
HCP: health care professional
PA: physical activity
SB: sedentary behavior
WAT: wearable activity tracker
WATCH: Wearable Activity Tracker Checklist for Health Care

Edited by L Buis; submitted 06.12.23; peer-reviewed by A Mazeas, H Liu; comments to author 23.04.24; revised version received 01.05.24; accepted 18.06.24; published 23.08.24.

Please cite as:
Szeto K, Arnold J, Horsfall EM, Sarro M, Hewitt A, Maher C
Establishing a Consensus-Based Framework for the Use of Wearable Activity Trackers in Health Care: Delphi Study
JMIR Mhealth Uhealth 2024;12:e55254
URL: https://mhealth.jmir.org/2024/1/e55254
doi:10.2196/55254
PMID:39178034

Original Paper

Reliable Contactless Monitoring of Heart Rate, Breathing Rate, and Breathing Disturbance During Sleep in Aging: Digital Health Technology Evaluation Study

Kiran K G Ravindran[1,2], PhD; Ciro della Monica[1,2], PhD; Giuseppe Atzori[1,2], MSc, RPGST; Damion Lambert[1,2], MSc; Hana Hassanin[2,3,4], MBBS, Dr med; Victoria Revell[1,2], PhD; Derk-Jan Dijk[1,2], PhD

[1]Surrey Sleep Research Centre, Guildford, United Kingdom

[2]UK Dementia Research Institute, Care Research and Technology Centre at Imperial College, London, United Kingdom, and the University of Surrey, Guildford, London, United Kingdom

[3]Surrey Clinical Research Facility, University of Surrey, Guildford, United Kingdom

[4]NIHR Royal Surrey Clinical Research Facility, Guildford, United Kingdom

Corresponding Author:
Kiran K G Ravindran, PhD
Surrey Sleep Research Centre
University of Surrey
Guildford, GU2 7XP
United Kingdom
Phone: 44 01483 68 3709
Email: k.guruswamyravindran@surrey.ac.uk

Abstract

Background: Longitudinal monitoring of vital signs provides a method for identifying changes to general health in an individual, particularly in older adults. The nocturnal sleep period provides a convenient opportunity to assess vital signs. Contactless technologies that can be embedded into the bedroom environment are unintrusive and burdenless and have the potential to enable seamless monitoring of vital signs. To realize this potential, these technologies need to be evaluated against gold standard measures and in relevant populations.

Objective: We aimed to evaluate the accuracy of heart rate and breathing rate measurements of 3 contactless technologies (2 undermattress trackers, Withings Sleep Analyzer [WSA] and Emfit QS [Emfit]; and a bedside radar, Somnofy) in a sleep laboratory environment and assess their potential to capture vital signs in a real-world setting.

Methods: Data were collected from 35 community-dwelling older adults aged between 65 and 83 (mean 70.8, SD 4.9) years (men: n=21, 60%) during a 1-night clinical polysomnography (PSG) test in a sleep laboratory, preceded by 7 to 14 days of data collection at home. Several of the participants (20/35, 57%) had health conditions, including type 2 diabetes, hypertension, obesity, and arthritis, and 49% (17) had moderate to severe sleep apnea, while 29% (n=10) had periodic leg movement disorder. The undermattress trackers provided estimates of both heart rate and breathing rate, while the bedside radar provided only the breathing rate. The accuracy of the heart rate and breathing rate estimated by the devices was compared with PSG electrocardiogram-derived heart rate (beats per minute) and respiratory inductance plethysmography thorax-derived breathing rate (cycles per minute), respectively. We also evaluated breathing disturbance indexes of snoring and the apnea-hypopnea index, available from the WSA.

Results: All 3 contactless technologies provided acceptable accuracy in estimating heart rate (mean absolute error <2.12 beats per minute and mean absolute percentage error <5%) and breathing rate (mean absolute error ≤1.6 cycles per minute and mean absolute percentage error <12%) at 1-minute resolution. All 3 contactless technologies were able to capture changes in heart rate and breathing rate across the sleep period. The WSA snoring and breathing disturbance estimates were also accurate compared with PSG estimates (WSA snore: r^2=0.76; P<.001; WSA apnea-hypopnea index: r^2=0.59; P<.001).

Conclusions: Contactless technologies offer an unintrusive alternative to conventional wearable technologies for reliable monitoring of heart rate, breathing rate, and sleep apnea in community-dwelling older adults at scale. They enable the assessment of night-to-night variation in these vital signs, which may allow the identification of acute changes in health, and longitudinal monitoring, which may provide insight into health trajectories.

International Registered Report Identifier (IRRID): RR2-10.3390/clockssleep6010010

(JMIR Mhealth Uhealth 2024;12:e53643) doi:10.2196/53643

KEYWORDS

Withings Sleep Analyzer; Emfit; Somnofy; contactless technologies; vital signs; evaluation; apnea-hypopnea index; wearables; nearables

Introduction

Background

Vital signs measured in clinical practice include heart rate, breathing rate, blood pressure, and body temperature. These serve as objective measurements of normal physiological functions and play a fundamental role in the assessment of health [1,2]. With aging, there is an increased incidence of functional limitations and chronic conditions, including hypertension, coronary heart disease, stroke, type 2 diabetes, and sleep apnea. More than 65% of adults aged ≥65 years report multimorbidity. The presence of uncontrolled comorbidities reduces the quality of life; leads to loss of independence; and increases the incidence of falls, hospitalization, and mortality [3-6].

Standardized, continuous vital signs monitoring systems, when implemented for at-home monitoring and care of older adults, including people living with dementia, can serve as an important tool for early identification of changes in health, improve care in older people, and reduce the burden on the health care system [7-10]. Commercially available wearable devices (wearables) and contactless technologies (nearables) are increasingly used for home monitoring and have the potential to enable remote health monitoring and promote independent living [11-18]. These technologies offer secure digital infrastructure that allows reliable and seamless transfer of collected data to cloud servers and can facilitate long-term remote monitoring opportunities for health care.

Wearables are widely used for continuous, at-home monitoring of heart rate, and some have been evaluated in-laboratory settings, predominantly in younger age groups [17-25]. Although several wearables have been shown to be acceptable for older adults, lower technology adoption rate; user comfort trade-off; and burden of maintenance (eg, removal during some daily activities such as showers, periodic recharging, and mobile app use) may make them unsuitable for long-term use in people with cognitive impairment due to their associated behavioral and psychological symptoms [23,26].

Contactless technologies can be embedded in the living environment such as under the bed mattress (undermattress devices or bed sensors) or on the bedside table (eg, bedside radars) and allow contactless monitoring when the user is in bed [27]. They are powered by the mains and securely stream the collected data wirelessly. They use several contactless sensing modalities to measure a composite signal (ballistographic signal) containing movements resulting from breathing and cardiac activity to extract vital signs (heart rate and breathing rate) information. The bedside radars use the Doppler radar technique, while the undermattress devices use several technologies, such as electromechanical films and pneumatic sensors [27,28]. Due to their inconspicuous nature and low maintenance, they do not pose any of the burdens imposed by wearables and are an ideal tool for continuous monitoring of vital signs, behavioral information, and sleep in community-dwelling older adult populations with long-term conditions and in people living with dementia [13,14,29,30].

Objective

To realize the potential of contactless technologies for monitoring vital signs such as heart rate and breathing rate in the community, the validity of their measurements needs to be evaluated in relevant populations. While the validity of the heart rate and breathing rate estimates collected from a few contactless technologies has been evaluated in younger populations, to the best of our knowledge, there are no vital signs evaluation studies in older adults (aged >65 years) although these devices have been implemented in longitudinal studies [13,29,31-33]. Here, we evaluated the validity of heart rate and breathing rate measurements collected from 3 contactless technologies (a bedside radar and 2 undermattress devices) against polysomnography (PSG) electrocardiogram (ECG)-derived heart rate and respiratory inductance plethysmography thorax (RIP thorax)-derived breathing rate in a laboratory setting. Throughout this paper, we have used the term vital signs to denote heart rate or breathing rate and vice versa. The evaluation addresses aspects of overnight average estimates; the ability to capture overnight trends; variability in heart rate and breathing rate; and accuracy at different sleep stages and time resolutions (60-, 10-, and 1-minute intervals) of estimates. We also discuss the data collection reliability in a home environment and summary estimates of breathing disturbance from the devices. To enhance the relevance of this study, we applied liberal inclusion and exclusion criteria for the participant selection, such that several participants had comorbidities that were representative of the general older population.

Methods

Cohort Characteristics

The study data were collected at home for a period of 7 to 14 days, followed by an overnight laboratory session (with full PSG) at the Surrey Sleep Research Centre in 2 cohorts (cohort 1: n=18, from January to March 2020; cohort 2: n=17, from June to November 2021). The participant group consisted of 35 individuals (men: n=21, 60%) aged between 65 and 83 (mean 70.8 SD 4.9) years. The participants were identified and recruited through the Surrey Clinical Research Facility. All participants attended an in-person screening visit, during which a range of assessments were completed. These included medical history (self-reported), the Pittsburgh Sleep Quality Index, the International Consultation on Incontinence

Questionnaire-Urinary Incontinence, intermediate activities of daily living questionnaire, and the Epworth Sleepiness Scale. The vital signs (heart rate, breathing rate, blood pressure, and temperature) were also collected and reviewed to determine inclusion in the study. To ensure the ecological validity of the collected data in this population, participants with stable comorbidities such as hypertension, type 2 diabetes, arthritis, and so on, were included in the study if their comorbidity and concomitant medications were stable and were not considered to pose a safety risk. Eligible participants had to be able to independently perform activities of daily life, as assessed by the intermediate activities of daily living questionnaire, and comply with study procedures. Participants were provided with detailed information about the study and provided written informed consent before any study procedures were performed. A detailed description of the inclusion and exclusion criteria can be found in our previous publications, in which we evaluated the ability of these technologies to estimate sleep timing in the home environment and sleep stages as defined by PSG [34,35].

Ethical Considerations

The study received a favorable opinion from the University of Surrey Ethics Committee (UEC-2019-065-FHMS) and was conducted in line with the Declaration of Helsinki and the principles of Good Clinical Practice. Potential participants were given detailed information about the study protocol, and they provided written informed consent before any study procedures were performed. Complete details of the study, along with the data management, privacy, and confidentiality measures, are discussed in the protocol [36].

Study Protocol

The undermattress devices (Withings Sleep Analyzer [WSA]; Withings] and Emfit QS; Emfit Ltd) were deployed both in the laboratory and at home, while the bedside radar (Somnofy; VitalThings) was only used in the laboratory. The participants also used an actigraphy device (Actiwatch Spectrum [AWS]; Philips Respironics) and maintained a consensus sleep diary at home [37]. During the home deployment period, the contactless technologies did not require any manual intervention or maintenance by the participants, and the data was transmitted automatically via Wi-Fi. To ensure the anonymity of the data, no personally identifiable participant information was added to the device applications, and a portable Wi-Fi router was used for data transfer. The contactless devices remained active and collected data continuously for the entire period of the study. The participants wore AWS on their nondominant hands throughout the study period, removing them briefly in scenarios that could lead to the device becoming wet. AWS data have been reported elsewhere [34].

After the home data collection period, 1 overnight full clinical PSG recording was conducted, which included an extended time in bed of 10 hours. On average, the participants slept 385.97 (SD 65.67) minutes in the sleep laboratory as measured via PSG and 405.96 (SD 84.00) minutes as measured via consensus sleep diary at home [34,35]. The WSA was used in both cohort 1 and cohort 2, while the Somnofy and Emfit were deployed only in cohort 2. The data from the contactless technologies were collected simultaneously along with PSG in the laboratory and with AWS and sleep diary at home. Empatica E4 (Empatica Srl), a wrist-worn device that collects activity and photoplethysmography, was also deployed during the laboratory session, and our evaluation of this device is reported elsewhere [26]. All devices and PSG data were collected for the entire 10-hour in-bed period in the laboratory. A detailed description of the study protocol is given in the protocol [36].

The Reference Vital Signs Data

During the in-laboratory session, PSG data were collected using the SomnoHD system (SOMNOmedics GmbH). The collected data included electroencephalography (256 Hz; F3-M2, C3-M2, O1-M2, F4-M1, C4-M1, and O2-M1); ECG (modified lead II subclavicular electrode placement; 256 Hz); RIP thorax and abdomen (128 Hz); photoplethysmography (128 Hz); electromyography (256 Hz, both submental and limb); and electrooculography (256 Hz; E2-M1 and E1-M2). In addition, data on the snore sensor (256 Hz) and airflow via nasal cannula and flow thermistor (128 Hz) were also collected. Sleep was scored at 30-second intervals in the Domino software environment as per American Academy of Sleep Medicine (AASM) guidelines by 2 independent scorers (a registered polysomnographic technologist and a trained scorer), and a consensus hypnogram was generated [38]. The sleep hypnogram contains 5 stages: rapid eye movement (REM), stage 1 of non-REM sleep (N1), stage 2 of non-REM sleep (N2), stage 3 of non-REM sleep (N3), and wake. The apnea-hypopnea index (AHI; number of apnea/hypopnea events per hour) and period limb movement index (PLMI; number of period limb movement events per hour) were determined by the Registered Polysomnographic Technologist using scoring rules recommended by AASM. An apnea was scored when there was a ≥90% drop in airflow lasting for at least 10 seconds, while a hypopnea was scored using the 3% drop in oxygen saturation and an arousal in the electroencephalogram criteria. The severity of apnea was determined using the following thresholds as per AASM guidelines: AHI score of <5 (normal), 5 to <15 (mild apnea), 15 to <30 (moderate apnea), and ≥30 (severe). A periodic limb movement event was scored when at least 4 consecutive limb movements occurred, each separated from the preceding limb movement by at least 5 seconds but not more than 90 seconds apart. PLMI of >15 seconds is used as the cutoff for the presence of periodic limb movement disorder [39,40]. In addition, participants with cardiac arrhythmia were identified using the arrhythmia index generated by the Domino software and verified by visual inspection of the record.

The PSG data were exported as standard EDF+ files along with recording markers and the consensus hypnogram. The ECG from the PSG was used for the extraction of the heart rate reference data, while the RIP thorax was used as the breathing rate reference data. For one of the participants in whom RIP thorax was unavailable, RIP abdomen was used to create the breathing rate reference data. MATLAB 2021b was used for all data analyses reported. The RR intervals used for the computation of the heart rate were derived from the ECG using the PhysioNet cardiovascular signal toolbox and a well-evaluated beat detection toolbox [41,42]. This beat-to-beat information was used to estimate reference heart rate (beats per minute [bpm]) at 30-second intervals, which is the same as the

PSG hypnogram. For extracting the breathing rate (cycles per minute [cpm]) from the RIP thorax signal at 30-second intervals, the RRest package was used [43,44].

Contactless Technologies: Data Overview

The WSA and Somnofy data (json format) were downloaded using the respective manufacturer's application programming interface, while the Emfit data (CSV format) were downloaded from the manufacturer's web interface. All the compared contactless technologies (WSA, Emfit, and Somnofy) provided breathing rate data, while only the undermattress devices (WSA and Emfit) provided heart rate data.

The devices provided vital signs (heart rate and breathing rate) data and sleep hypnograms at different resolutions (WSA: 60 seconds; Emfit and Somnofy: 30 seconds). The WSA and Somnofy heart rate and breathing rate estimations were available at 60 and 30-second resolutions (same as the respective device hypnogram resolution), while the Emfit estimated heart rate at a 4-second resolution. These 4-second estimates were averaged to generate estimates at 30-second intervals to match the Emfit sleep label intervals. To allow data analysis relative to local time, daylight savings correction was applied to the Coordinated Universal Time timeseries generated by the devices. The sleep hypnograms generated by the devices contain 4 stages: deep sleep (DS=N3), light sleep (LS=N2/N1), REM, and wake.

Apart from heart rate and breathing rate, Emfit generates continuous heart rate variability and activity measures, while Somnofy provides the estimates of movement and environmental variables such as ambient light, sound, temperature, pressure, humidity, and indoor air quality, which were out of scope for this evaluation and are not discussed here.

Both at home and in the laboratory, all devices were connected to the same network, and the devices used the manufacturer's time synchronization protocol such as network time protocol to timestamp the data. Although this ensured that the devices were synchronized to local time, we performed another synchronization step to allow an accurate comparison of the data between the devices. The device vital signs measures were aligned to the PSG reference vital signs estimates via cross-correlation between the device and PSG vital signs and hypnograms, and the lag (within a 5-minute window) that provided the best alignment of both the vital signs data and hypnograms was then applied. The WSA data were converted from 60- to 30-second intervals by upsampling. Epochs in the PSG and device hypnograms that were scored as artifacts or no presence were excluded from the assessment.

Vital Signs Assessment

Overview

The evaluation of the epoch-by-epoch heart rate and breathing rate data collected by the contactless technologies was performed against reference estimates derived from PSG ECG and RIP thorax. The accuracy and reliability of the heart rate and breathing rate estimates were performed at different levels of time resolution to determine use cases in which the contactless technologies can be used. These include accuracy assessment of overnight average estimates; ability to capture overnight trends; variability in vital signs; and accuracy in different sleep stages and at different time resolutions (60-, 10-, and 1-minute intervals) of estimates. All laboratory data analyses were performed over the total recording period of the PSG. At all temporal resolutions of comparisons, only complete or valid pairs of estimates were used.

Performance Measures

To assess the accuracy of the vital signs estimates (heart and breathing rates), mean absolute error (MAE) and mean absolute percentage error (MAPE) were used as the primary metrics. MAE and MAPE are used to measure the error in the estimate between the device and the PSG reference vital signs. Bland-Altman metrics such as minimum detectable change (MDC), bias, and limits of agreement (LoA) were also computed to provide an overview of the agreement of the estimates and to allow comparison with the evaluations reported in the literature [45,46]. All the measures are reported with 95% CIs. MDC is the smallest change in the estimate that can be detected by the device that exceeds the measurement error. It is equal to half the agreement width [47],

$$MDC = (LoA_{\text{Upper bound}} - LoA_{\text{Lower bound}}) / 2 \quad \textbf{(1)}$$

We used intraclass correlation (ICC) with 2-way random effects to measure the reliability of measurement and standardized absolute difference, a directionless Cohen *d*, described by Guruswamy Ravindran et al [35] and Haghayegh et al [47], for measuring the dispersion in the bias. All ICC values (range 0-1) were estimated with α of .05; ranges used for interpretation were as follows: ICC<0.5 (poor), ICC=0.5 to 0.75 (moderate), and ICC>0.75 (good reliability). Apart from the above metrics, the coefficient of determination (r^2, a measure of how close the measured estimates are to the reference, computed using simple linear regression) was also used for the concordance analysis. For estimating the significance of differences between vital signs during different sleep stages, devices, and time courses, we used ANOVA followed by linear mixed effects models with the different groups (devices, sleep stages, and time) as fixed effects (with interactions) and participant as a random effect. MATLAB 2021b was used for all statistical analyses.

Acceptable Agreement for Heart Rate

The satisfactory level of agreement between the PSG reference heart rate and the device-determined heart rate was set to an error of 10% or +5 or –5 bpm as recommended by the Association for the Advancement of Medical Instrumentation [42,48].

Acceptable Agreement for Breathing Rate

For breathing rate, to the best of our knowledge, no device-specific satisfactory level of agreement is discussed in the literature. Hence, the agreement in breathing rate estimates between human observers is used to set the acceptable level for our evaluation. We set the permissible level of error in breathing rate estimation to be +4 or –4 cpm, as reported by Lim et al [49].

Breathing Disturbance Estimates

Apart from epoch-by-epoch heart rate and breathing rate, WSA also provides a "snoring" signal, which is a binary variable

depicting snore presence detected by the device. In addition, the WSA also provides the summary estimates of snoring duration (WSA snore), breathing disturbance intensity (WSA breathing disorder index [BDI]), and AHI (WSA AHI). The Emfit and Somnofy devices do not generate breathing disturbance measures. Only complete or valid pairs of estimates were used to analyze the summary measures. For WSA BDI and AHI estimates, we explored the relationship between them and the concordance of these WSA estimates to the PSG AHI in the laboratory.

For the WSA snore estimates, we explored the concordance of the all-night snore duration estimated by the PSG snore sensor (ie, PSG snore microphone placed on the side of the neck) and the WSA. The PSG snore sensor data were scored by the Somnomedics Domino software using 30 dB as the snore amplitude threshold and a minimum snore duration of 300 ms. We further explored the differences in the snore intensity as measured by the PSG snore sensor for the epochs determined to contain snore events by the WSA, followed by an exploration of the distribution of the snore events during the different sleep stages.

Results

Characteristics of the Study Population

Of the 35 participants, more than half of the participants (n=20, 57%) reported comorbidities, including type 2 diabetes, obesity, arthritis, and hypertension, with concomitant medication. In this study population, the mean heart rate was 62.2 (SD 8.9) bpm (men: n=21, 60%; mean 60.6, SD 9.3 bpm; women: n=14; 64.4, SD 7.8 bpm), and mean breathing rate was 14.7, SD 2.9 cpm (men: 14.6, SD 2.9 cpm; women: 14.7, SD 3.0 cpm), as assessed from the overnight laboratory PSG. The average BMI of the participants was 27.0 (SD 4.8) kg/m^2, with 17% (n=6, BMI >30) being obese. The mean systolic and diastolic blood pressures measured during screening were 148.7 (SD 16.2) mmHg and 87.0 (SD 9.6) mmHg, respectively. During the clinical PSG, it emerged that 95% (n=33) of the participants in the study had some degree of apnea. Of the participants with apnea, 8 (23%) had severe (AHI>30), 9 (26%) had moderate (AHI=15 to <30), and 16 (46%) had mild apnea (AHI=5 to <15). A total of 10 (29%) participants had PLMI>15, which is similar to the prevalence of periodic limb movement syndrome in community-dwelling older adults [50,51]. Some form of cardiac arrhythmia was found in 51% (n=18) of the participants, with 29% (n=10) of them also having severe or moderate apnea. A detailed description of the population characteristics can be found in the study by Ravindran et al [34].

Overview of Vital Signs Data

Example Case

An example of vital sign data collected by WSA for 14 days at home followed by 1 day in the laboratory is shown in Figure 1. The participant had moderate apnea with an AHI of 24.1 events/hour, as determined by the clinical PSG during the laboratory visit (day 0). The raster plot (Figure 1A) shows heart rate, sleep-wake stage, and breathing rate as estimated by the WSA. The vital sign data were available when the participant was in bed at night and during daytime in-bed periods, which were also reported as naps by the participant. The days –14 to –1 depict the data collected at home, while day 0 depicts data collected in the sleep laboratory. The darker, purple-colored regions denote sleep, while the lighter regions denote wake as identified by the WSA. The gray areas represent out-of-bed periods (ie, periods during which the device did not record data). The first day of data from the WSA at home were lost. This participant had irregular nocturnal bed timing, with an average time in bed during the nocturnal period of 10 hours 8 minutes at home. Mean nocturnal heart rate and breathing rate varied across nights. The WSA AHI showed night-to-night variability (Figure 1B), and the WSA AHI during the laboratory visit was 25 events/hour, which was close to the PSG AHI value. The heart rate showed a trend across the nocturnal sleep period, with a higher heart rate at the beginning of the nocturnal sleep period and a lower heart rate just before the end of the sleep period.

Figure 2 shows the contactless technology–derived vital sign data collected alongside PSG reference data during the laboratory visit for the participant depicted in Figure 1. The PSG heart rate and breathing rate both show changes as the hypnogram shows transitions between different stages of sleep, with higher variability during wake and REM sleep and lower variability during non-REM sleep. The differences in the data resolution (WSA: 60 seconds; Emfit and Somnofy: 30 seconds) between the contactless technologies can be seen from the plots. The WSA vital signs estimates provided by the device are rounded to the nearest integers, leading to more discretized vital sign data. The trends in the heart rate and breathing rate data recorded by the PSG and the contactless technologies are more similar during the sleep periods compared with the wake periods. The Somnofy had a number of missing estimates of breathing rate during many of the wake epochs determined by the device, and these missing vital sign epochs also coincided with periods of higher activity, as detected by a wrist-worn activity device (Empatica E4). Snoring, as detected by the WSA, followed a pattern that closely followed the snoring signal detected by PSG, with some disagreement in the detection of the snoring event (Figure 2).

Figure 1. Vital signs from a male participant aged 65 to 70 years collected at home and in the laboratory. (A) Raster plot showing the heart rate (bpm), and breathing rate (cpm) as detected by the Withings Sleep Analyzer along with device-detected sleep (or) wake period and sleep diary information. (B) Estimates of mean heart rate, mean breathing rate, and apnea-hypopnea index (AHI; depicted as circles adjacent to the time courses) were made during the night as determined by the device.

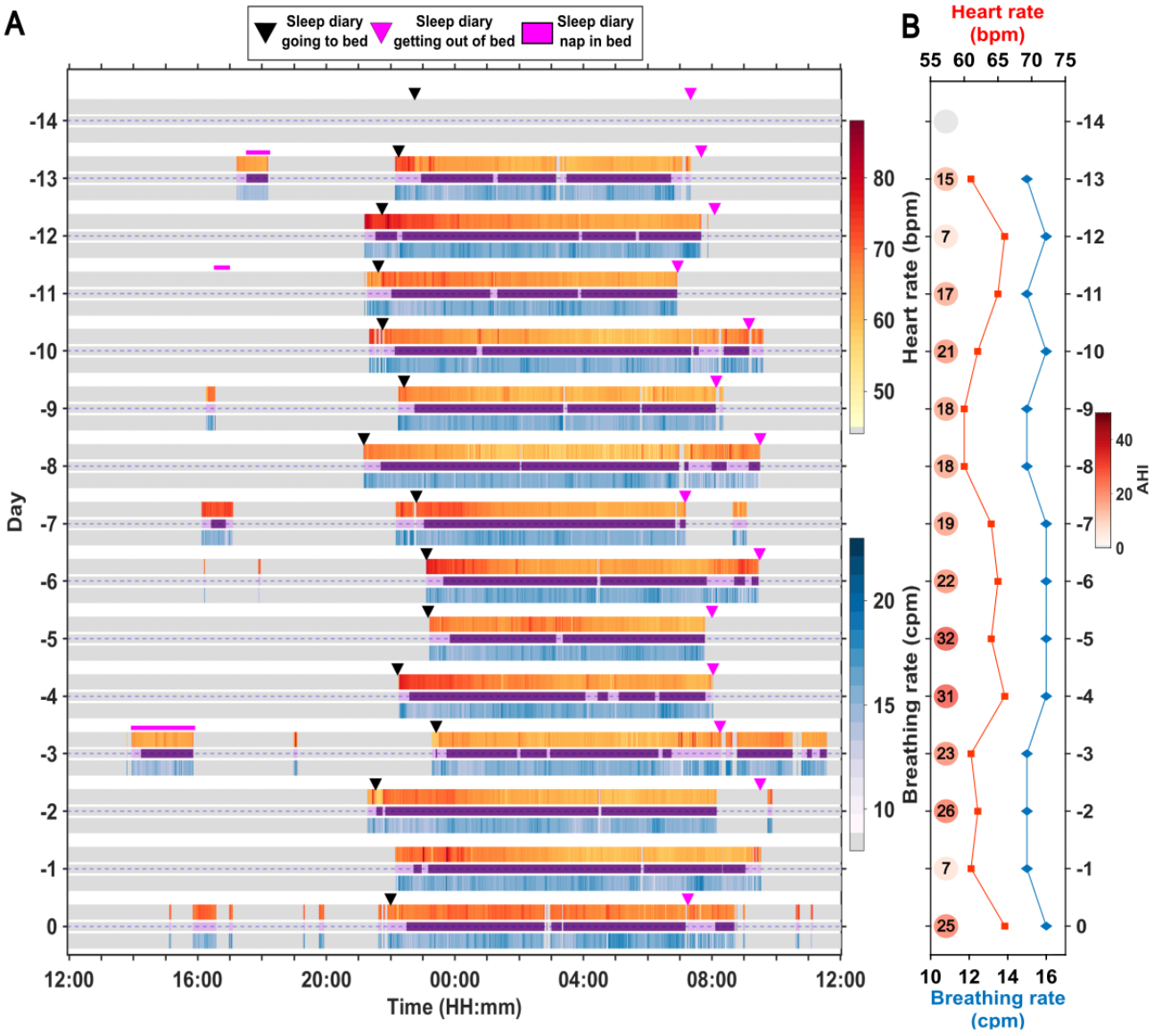
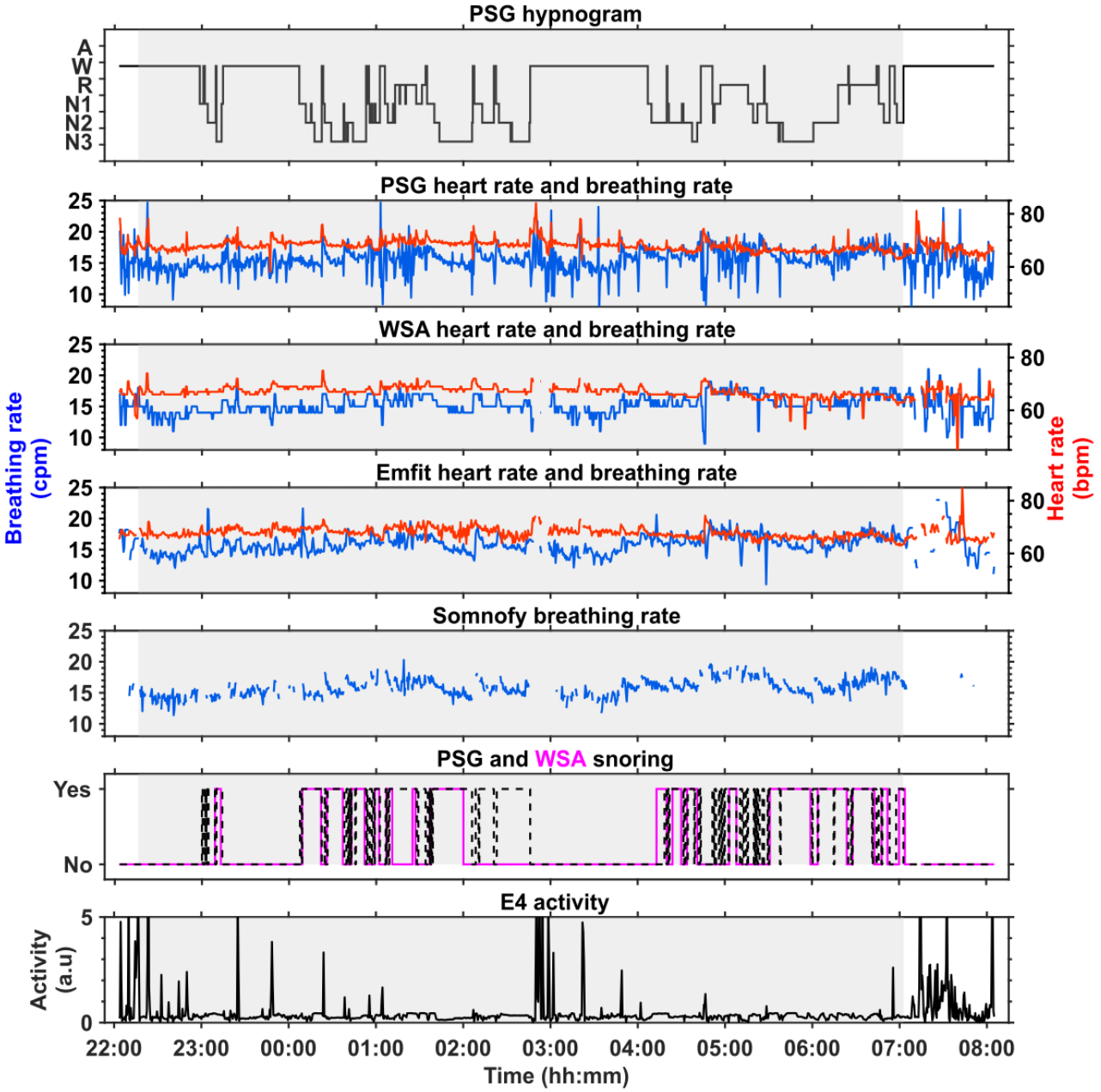


Figure 2. Vital sign data simultaneously collected from 3 contactless technologies and polysomnography (PSG) in the laboratory for the male participant described in Figure 1. The PSG consensus hypnogram (top) is depicted at the top followed by the heart rate (red), and breathing rate (blue) from PSG and contactless devices, and activity data in arbitrary units are from the Empatica E4 device. The gray regions in all the plots correspond to the lights-off period. A: artifact; bpm: beats per minute; cpm: cycles per minute; N1: stage 1 of nonrapid eye movement (non-REM) sleep; N2: stage 2 of non-REM sleep; N3: stage 3 of non-REM sleep; R: REM sleep; W: wake; WSA: Withings Sleep Analyzer.



Summary of Collected Data

In the laboratory study, all 35 PSGs (ground truth or reference data, cohort 1: n=18, 51%; cohort 2: n=17, 49%) were available. The total number of nights of data collected in the laboratory for each of the 3 contactless technologies were 35 for WSA, 16 for Emfit, and 17 for Somnofy. One night of data was lost due to a device malfunction for Emfit. At home, a total of 401 days of data were collected across the 35 participants, with 321 days of data available for WSA (cohort 1: n=10, 56%; cohort 2: n=17, 100%) and 228 days of data available for Emfit (cohort 2: n=17, 100%). At home, for WSA, portions of data from 8 participants were lost in cohort 1 due to deployment errors and Wi-Fi dropouts, with a data loss of 3.3% (11/332 days lost). For Emfit, the data loss was 4.2% (10/238 days lost).

In the laboratory, the range of the heart rate estimated (minimum to maximum) by the contactless technologies was 40 to 90 bpm for WSA and 40 to 135 bpm for Emfit, whereas for breathing rate, the ranges were 8 to 35 cpm for WSA, 6 to 30 cpm for Emfit, and 6 to 30 cpm for Somnofy. The undermattress device–generated vital sign (both heart rate and breathing rate) data for 100% of the in-bed periods. Somnofy bedside radar, by contrast, generated breathing rate data less continuously, resulting in data unavailability at 32.21% (54.41/169.14 hours) of the in-bed period (in-laboratory). Most of these missed breathing rate epochs were found to be in the Somnofy-predicted

wake state (the total percentage of breathing rate epochs unavailable per label was follows: wake=34.35/54.41 hours, 62.39%, REM=6.96/54.41 hours, 12.73%; LS=12.19/54.41 hours, 23.1%, and DS=0.91/54.41 hours, 1.78%).

All-Night Vital Signs

The concordance between the nightly average heart rate and breathing rate estimates of the contactless technologies against PSG are shown in Figure 3 and Table 1. The WSA (MAPE 3.28%; ICC=0.87) had a lower level of agreement compared with Emfit (MAPE 1.83%; ICC=0.96), and this was due to an outlier participant with severe arrhythmia (see the outlier in Figure 3). When the outlier was removed WSA (WSA*: MAPE 1.87%; ICC=1.0) had an agreement with the PSG that was similar to that of Emfit. The MDC was higher for Emfit (3.25 bpm) compared with WSA* (1.17 bpm), which can be seen from the higher dispersion in the Emfit estimates (Figure 2A). For the breathing rate, Somnofy (MAPE 4.64%; ICC=0.82) had a high agreement, followed by WSA (MAPE 6.29%; ICC=0.78) and Emfit (MAPE 5.46%; ICC=0.76). The MDC follows the agreement results, with Somnofy (1.98 cpm) having a somewhat lower value compared with WSA (2.08 cpm) and Emfit (2.21 cpm).

Figure 3. Association between vital signs estimated by 3 contactless devices and estimates from polysomnography (PSG) averaged across the night while sleeping in the sleep laboratory. (A) Heart rate (beats per minute [bpm]); (B) breathing rate (cycles per minute [cpm]). The error bars represent the SD of the estimate within participants. The PSG heart rate and breathing rate are derived from electrocardiogram (ECG) and respiratory inductance plethysmography thorax (RIP thorax), respectively. The number of participants (ie, nights) available for the devices are Withings Sleep Analyzer (WSA; n=34), Emfit (n=16), and Somnofy (n=17).

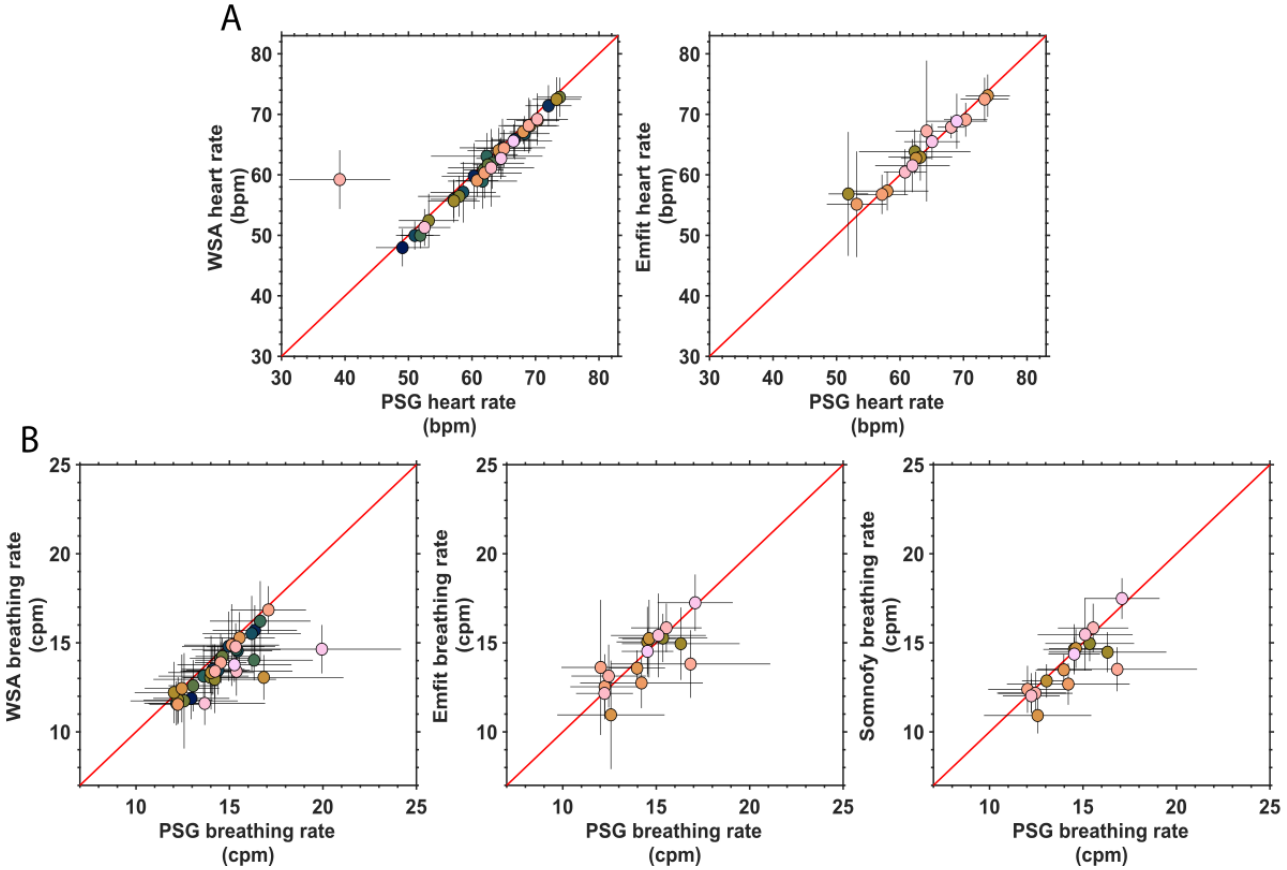


Table 1. All-night average of vital signs and their agreement metrics[a].

Vital signs	Device, mean (SD)	PSG[b], mean (SD)	Bias[c] (SD; 95% CI)	LoA[d] (lower bound), (95% CI)	LoA (upper bound), (95% CI)	MDC[e]	MAE[f] (95% CI)	SAD[g] (95% CI)	MAPE[h] (95% CI)	ICC[i] (95% CI)
Heart rate										
WSA[j]	61.66 (6.51)	62.16 (7.52)	–0.5 (3.63; 1.75 to 0.74)	–7.61 (–9.76 to –5.46)	6.61 (4.46 to 8.75)	7.11	1.69 (2.80 to 0.58)	0.24 (–0.09 to 0.58)	3.28 (0.40 to 6.16)	0.87 (0.75 to 0.93)
WSA*[k]	61.73 (6.6)	62.84 (6.46)	–1.11 (0.6; –1.31 to –0.9)	–2.28 (–2.64 to –1.92)	0.07 (–0.29 to 0.43)	1.17	1.15 (1.33 to 0.97)	0.18 (–0.16 to 0.52)	1.87 (1.56 to 2.17)	1 (0.99 to 1)
Emfit	63.85 (5.66)	63.42 (6.47)	0.44 (1.66; –0.45 to 1.32)	–2.81 (–4.35 to –1.27)	3.68 (2.14 to 5.22)	3.25	1.08 (1.78 to 0.39)	0.18 (–0.33 to 0.7)	1.83 (0.52 to 3.13)	0.96 (0.90 to 0.99)
Breathing rate										
WSA	13.66 (1.38)	14.61 (1.69)	–0.95 (1.06; –1.32 to –0.59)	–3.04 (–3.66 to –2.41)	1.13 (0.5 to 1.75)	2.08	0.97 (1.33 to 0.6)	0.63 (0.3 to 0.97)	6.29 (4.31 to 8.27)	0.76 (0.58 to 0.87)
Emfit	14.13 (1.6)	14.35 (1.67)	–0.22 (1.13; –0.82 to 0.38)	–2.43 (–3.48 to –1.38)	1.99 (0.94 to 3.03)	2.21	0.78 (1.22 to 0.34)	0.49 (–0.02 to 1.01)	5.46 (2.55 to 8.36)	0.76 (0.44 to 0.91)
Somnofy	13.77 (1.68)	14.28 (1.65)	–0.51 (1.01; –1.02 to 0.01)	–2.48 (–3.38 to –1.58)	1.47 (0.57 to 2.37)	1.98	0.69 (1.14 to 0.24)	0.43 (–0.07 to 0.93)	4.64 (1.79 to 7.5)	0.82 (0.56 to 0.93)

[a]Metrics of agreement for overall heart rate of the devices against the electrocardiogram (ECG) estimates (included in the PSG). The number of participants contributing to each of these devices was as follows: WSA (n=35), WSA* (n=34), Emfit (n=16), and Somnofy (n=17).

[b]PSG: polysomnography.

[c]Bias is the difference in measurement between the device and PSG ECG.

[d]LoA: limits of agreement.

[e]MDC: minimum detectable changes; smallest detectable change independent of measurement error (half of Bland-Altman agreement width).

[f]MAE: mean absolute error.

[g]SAD: standardized absolute difference; directionless version of Cohen *d*.

[h]MAPE: mean absolute percentage error.

[i]ICC: intraclass correlation coefficient with 2-way random effects; measure of measurement reliability.

[j]WSA: Withings Sleep Analyzer.

[k]WSA* depicts the outlier-removed WSA data.

Vitals Signs During Different Vigilance States

We investigated the agreement of the vital signs estimated by the contactless technologies to the PSG reference during the different vigilance states of the consensus PSG hypnogram. The distribution of the vital signs for the different vigilance states is provided in Figure S1 in Multimedia Appendix 1. The heart rate and breathing rate estimates of both the PSG and contactless devices were not normally distributed (via the Kolmogorov-Smirnov test). The difference between the mean heart rate estimated by the devices and PSG was <2.5 bpm across all sleep stages. The difference between the mean breathing rate estimated by the devices and the PSG was <1.5 cpm across all sleep stages. Overall, the concordance between the estimates of both heart and breathing rate provided by the contactless technologies and PSG was good across all vigilance states (Figure S2 and Tables S1 and S2 in Multimedia Appendix 1).

Time Course of Vital Signs During Sleep

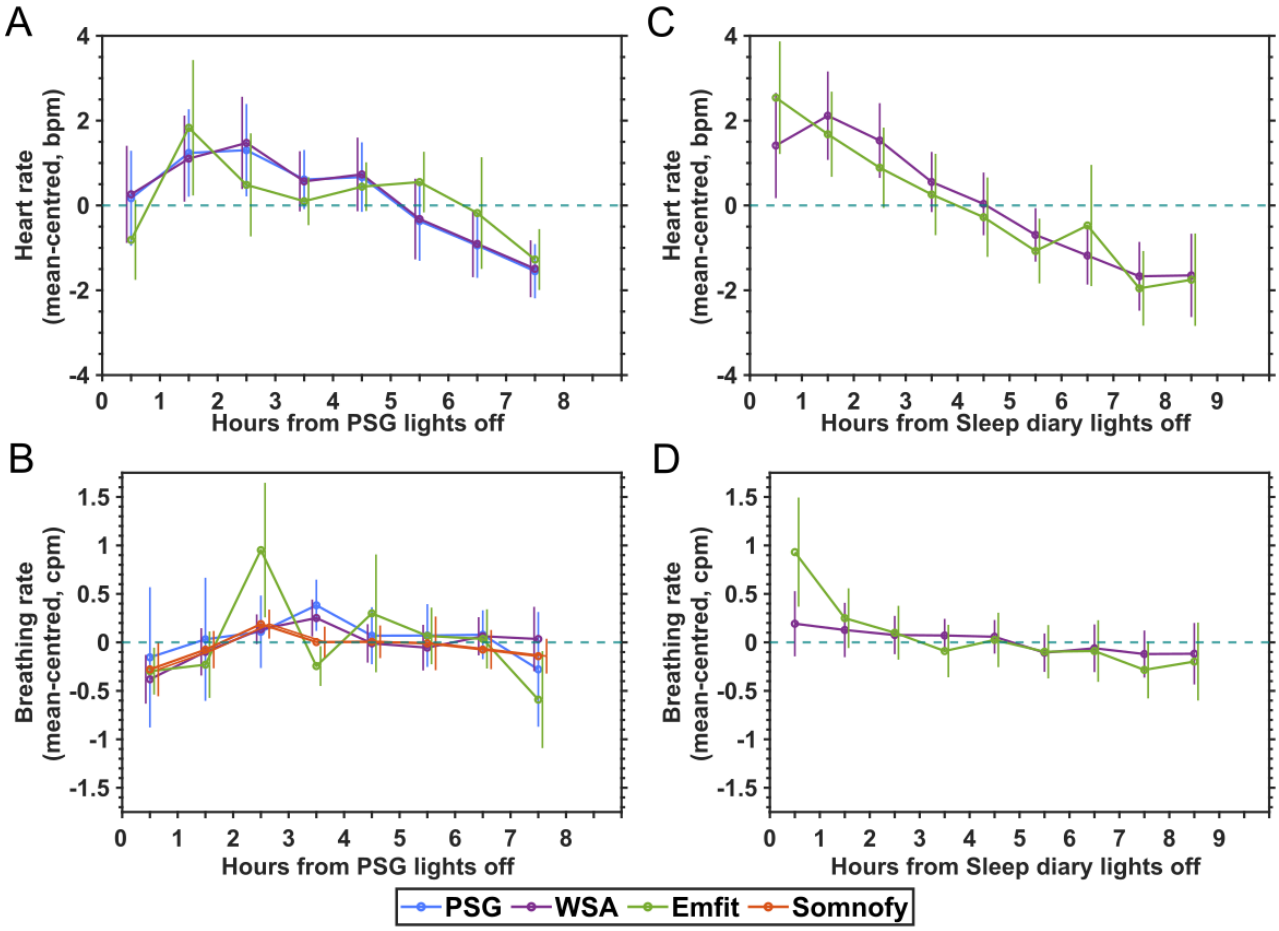
Figure 4A shows the time course of the heart rate and breathing rate estimated by the contactless devices and PSG over the PSG lights-off period in the laboratory. Figure 4B shows these time courses for the sleep diary–defined lights-off periods recorded at home (Figure 4B). The vital sign data were mean centered and averaged over the epochs during which participants were asleep as detected by the PSG hypnogram for the laboratory data and as detected by the sleep scoring algorithms of the devices for the at-home data. The data are plotted per hour starting from the onset of the lights-off period. The error bars represent the SD of the estimate and are shifted along the x-axis to improve visibility.

Both in the laboratory and at home, the heart rate started close to or above the mean, gradually decreased over the night, and reached the nightly minimum in the second part of the sleep period before a slight increase at the end of the sleep period. The similarity between the contactless technology and PSG heart rate hourly time series in the laboratory was determined using MAPE. The WSA (MAPE 14.42%) closely follows the PSG, while the Emfit (MAPE 144.34%) is less similar.

The breathing rate trends at home were similar to that of the heart rate, starting above the overnight mean and gradually decreasing to a lower value closer to the end of the night. In the laboratory, the breathing rate hourly estimates were fluctuating and did not show any clear trend. Both the WSA (MAPE 129.95%) and Somnofy (MAPE 130.01%) had the highest

similarity with the PSG, while the Emfit had the lowest similarity (MAPE 295.76%).

Figure 4. Time course of vital signs during sleep in the laboratory (A and B) and at home (C and D). The number of participants available for each of the devices (measured in nights) in the laboratory (polysomnography [PSG]: n=35, 35 nights; Withings Sleep Analyzer [WSA]: n=34, 34 nights; Emfit: n=16, 16 nights; and Somnofy: n=17, 17 nights) and at home (WSA: n=27, 295 nights; Emfit: n=17, 213 nights). bpm: beats per minute; cpm: cycles per minute.



Effect of Temporal Resolution on the Vital Signs Accuracy

To examine the effect of the length of the time period over which the vital signs are computed, we averaged the heart rate and breathing rate estimates over 60 minutes, 10 minutes, and 1 minute and computed the agreement with the corresponding PSG reference estimates (Tables 2 and 3). We examined the cumulative distribution function (CDF) of the MAE at these resolutions to better characterize the estimation error. The CDFs are depicted in Figure 5, while more detailed scatter plots and associated agreement measures are provided in Figure S3 and Tables S2 and S3 in Multimedia Appendix 1.

For all devices and both heart and breathing rates, the CDFs become steeper with increasing duration of the time window over which these variables are computed. For both heart rate and breathing rate, the agreement (measured by ICC) with the PSG reference estimates increased with decreasing temporal resolution (Figure S3 in Multimedia Appendix 1). On closer inspection of the CDFs, we find that for the heart rate estimates, the error at the 90th percentile is lower for WSA than for Emfit for the 1-minute and 60-minute estimations (overall error<4 bpm). For the Emfit, the median (50th percentile) error of the heart rate estimates became smaller with increasing duration of the time window, but the WSA median error was always close to 1. For the breathing rate, at all 3 resolutions, 50% of the estimates had an error of <1 cpm. When we inspected the 90th percentile error of breathing rate estimates at 1 minute, we found that Somnofy had the lowest error (2.56 cpm), followed by WSA (3.46 cpm) and Emfit (4.26 cpm). This trend was seen for lower resolutions as well. At the 50th percentile, we see the effect of discrete breathing rate output from WSA on the 50th percentile error, where the error of the other 2 devices falls below 0.5, while the WSA error does not. A detailed discussion of the effects of temporal resolution on the agreement between the device and PSG vital sign estimate is provided in Multimedia Appendix 1.

Table 2. Effect of temporal resolution on reliability of estimates of heart rate[a].

Heart rate resolution	Device, mean (SD)	PSG[b], mean (SD)	Bias[c] (SD; 95% CI)	LoA[d] (lower bound), (95% CI)	LoA (upper bound), (95% CI)	MDC[e]	MAE[f] (95% CI)	SAD[g] (95% CI)	MAPE[h] (95% CI)	ICC[i] (95% CI)
60 minutes										
WSA[j]	61.76 (7.0)	62.09 (8.03)	−0.33 (3.84; −0.78 to 0.12)	−7.86 (−8.63 to −7.1)	7.2 (6.44 to 7.97)	7.53	1.75 (2.15 to 1.35)	0.23 (0.12 to 0.35)	3.46 (2.41 to 4.51)	0.87 (0.84 to 0.9)
Emfit	64.04 (6.52)	63.58 (6.78)	0.46 (2.85; −0.04 to 0.96)	−5.12 (−5.98 to −4.26)	6.04 (5.18 to 6.9)	5.58	1.16 (1.63 to 0.69)	0.18 (0 to 0.35)	1.95 (1.09 to 2.8)	0.91 (0.87 to 0.93)
10 minutes										
WSA	61.71 (7.18)	62.07 (8.21)	−0.36 (3.94; −0.54 to −0.18)	−8.07 (−8.38 to −7.76)	7.36 (7.05 to 7.66)	7.71	1.83 (1.99 to 1.67)	0.24 (0.19 to 0.28)	3.58 (3.15 to 4)	0.87 (0.86 to 0.88)
Emfit	64.03 (7.04)	63.57 (7.04)	0.47 (4.01; 0.19 to 0.74)	−7.39 (−7.86 to −6.91)	8.32 (7.84 to 8.79)	7.85	1.4 (1.66 to 1.14)	0.2 (0.13 to 0.27)	2.38 (1.9 to 2.86)	0.84 (0.82 to 0.86)
1 minute										
WSA	61.71 (7.44)	62.08 (8.59)	−0.37 (4.54; −0.43 to −0.3)	−9.27 (−9.38 to −9.16)	8.54 (8.42 to 8.65)	8.9	2.08 (2.14 to 2.02)	0.26 (0.24 to 0.27)	4 (3.85 to 4.14)	0.84 (0.84 to 0.84)
Emfit	64.02 (7.54)	63.56 (7.44)	0.46 (5.24; 0.35 to 0.58)	−9.8 (−10 to −9.61)	10.73 (10.54 to 10.92)	10.27	2.12 (2.22 to 2.02)	0.28 (0.26 to 0.3)	3.52 (3.33 to 3.7)	0.76 (0.75 to 0.76)

[a]Metrics of agreement for overall heart rate estimates from the devices against PSG electrocardiogram (ECG) estimates (beats per minute) at various temporal resolutions.

[b]PSG: polysomnography.

[c]Bias is the difference in measurement between the device and PSG ECG (device–PSG ECG).

[d]LoA: limits of agreement.

[e]MDC: minimum detectable changes; smallest detectable change independent of measurement error (half of Bland-Altman agreement width).

[f]MAE: mean absolute error.

[g]SAD: standardized absolute difference; a directionless version of Cohen *d*.

[h]MAPE: mean absolute percentage error.

[i]ICC: intraclass correlation coefficient; with two-way random effects (measures of measurement reliability).

[j]WSA: Withings sleep analyzer.

Table 3. Effect of temporal resolution on the reliability of estimates of breathing rate[a].

Breathing rate resolution	Device, mean (SD)	PSG[b], mean (SD)	Bias[c] (SD; 95% CI)	LoA[d] (lower bound), (95% CI)	LoA (upper bound), (95% CI)	MDC[e]	MAE[f] (95% CI)	SAD[g] (95% CI)	MAPE[h] (95% CI)	ICC[i] (95% CI)
60 minutes										
WSA[j]	13.67 (1.51)	14.58 (1.87)	−0.91 (1.28; −1.06 to −0.76)	−3.42 (−3.67 to −3.17)	1.6 (1.34 to 1.85)	2.51	1 (1.14 to 0.86)	0.59 (0.47 to 0.71)	6.48 (5.73 to 7.22)	0.72 (0.65 to 0.77)
Emfit	14.18 (1.86)	14.33 (1.85)	−0.14 (1.66; −0.44 to 0.15)	−3.39 (−3.9 to −2.89)	3.11 (2.6 to 3.62)	3.25	0.92 (1.16 to 0.67)	0.5 (0.32 to 0.67)	6.35 (4.68 to 8.02)	0.6 (0.47 to 0.7)
Somnofy	13.89 (1.65)	14.31 (1.8)	−0.42 (1.19; −0.65 to −0.19)	−2.76 (−3.16 to −2.36)	1.92 (1.52 to 2.31)	2.34	0.55 (0.77 to 0.33)	0.32 (0.13 to 0.51)	3.47 (2.27 to 4.67)	0.76 (0.67 to 0.83)
10 minutes										
WSA	13.67 (1.67)	14.6 (2.2)	−0.93 (1.67; −1.01 to −0.86)	−4.21 (−4.34 to −4.08)	2.34 (2.21, 2.47)	3.28	1.14 (1.21 to 1.07)	0.59 (0.54 to 0.63)	7.29 (6.94 to 7.65)	0.63 (0.61 to 0.66)
Emfit	14.16 (2.2)	14.35 (2.17)	−0.18 (2.1; −0.33 to −0.04)	−4.31 (−4.56 to −4.06)	3.94 (3.69 to 4.19)	4.12	1.09 (1.22 to 0.96)	0.5 (0.43 to 0.57)	7.42 (6.6 to 8.25)	0.54 (0.48 to 0.58)
Somnofy	13.87 (1.77)	14.35 (2.05)	−0.48 (1.49; −0.59 to −0.36)	−3.41 (−3.6 to −3.21)	2.45 (2.26 to 2.65)	2.93	0.68 (0.79 to 0.57)	0.36 (0.28 to 0.43)	4.2 (3.64 to 4.76)	0.69 (0.65 to 0.73)
1 minute										
WSA	13.67 (2.04)	14.61 (2.61)	−0.93 (2.32; −0.97 to −0.9)	−5.48 (−5.54 to −5.42)	3.61 (3.56 to 3.67)	4.55	1.51 (1.54 to 1.48)	0.64 (0.63 to 0.66)	9.85 (9.68 to 10.02)	0.51 (0.5 to 0.52)
Emfit	14.17 (2.5)	14.35 (2.57)	−0.18 (2.73; −0.24 to −0.12)	−5.54 (−5.65 to −5.44)	5.18 (5.07 to 5.28)	5.36	1.6 (1.65 to 1.56)	0.63 (0.61 to 0.65)	11.07 (10.74 to 11.4)	0.42 (0.4 to 0.44)
Somnofy	13.87 (1.91)	14.35 (2.43)	−0.47 (1.99; −0.52 to −0.43)	−4.38 (−4.46 to −4.3)	3.43 (3.35 to 3.51)	3.9	1.06 (1.1 to 1.01)	0.48 (0.46 to 0.51)	6.76 (6.54 to 6.99)	0.58 (0.57 to 0.6)

[a]Metrics of agreement for overall heart rate estimates from the devices against PSG electrocardiogram (ECG) estimates (beats per minute) at various temporal resolutions.

[b]PSG: polysomnography.

[c]Bias is the difference in measurement between the device and PSG ECG (device–PSG ECG).

[d]LoA: limits of agreement.

[e]MDC: minimum detectable changes; smallest detectable change independent of measurement error (half of Bland-Altman agreement width).

[f]MAE: mean absolute error.

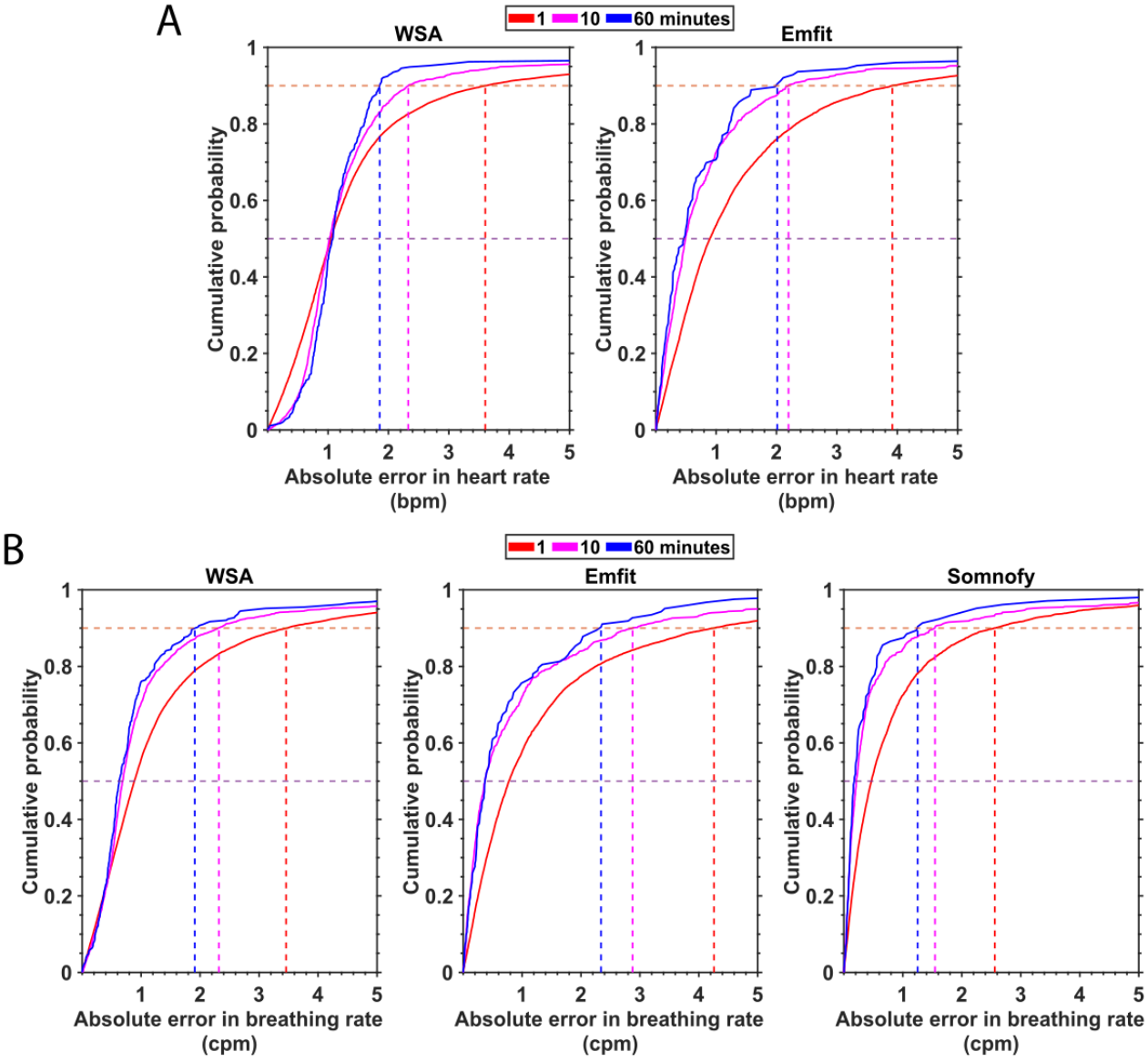
[g]SAD: standardized absolute difference; a directionless version of Cohen *d*.

[h]MAPE: mean absolute percentage error.

[i]ICC: intraclass correlation coefficient; with two-way random effects (measures of measurement reliability).

[j]WSA: Withings sleep analyzer.

Figure 5. Effect of time window over which vital signs are estimated on device measurement error; (A) heart rate (beats per minute [bpm]) and (B) breathing rate (cycles per minute [cpm]). The cumulative density function of the absolute error is represented for each of the devices for the window lengths 1, 10, and 60 minutes. The median (50th percentile) and the 90th percentile are represented by horizontal lines. PSG: polysomnography; WSA: Withings Sleep Analyzer.



Estimating Breathing Disturbances During Sleep

WSA Snore

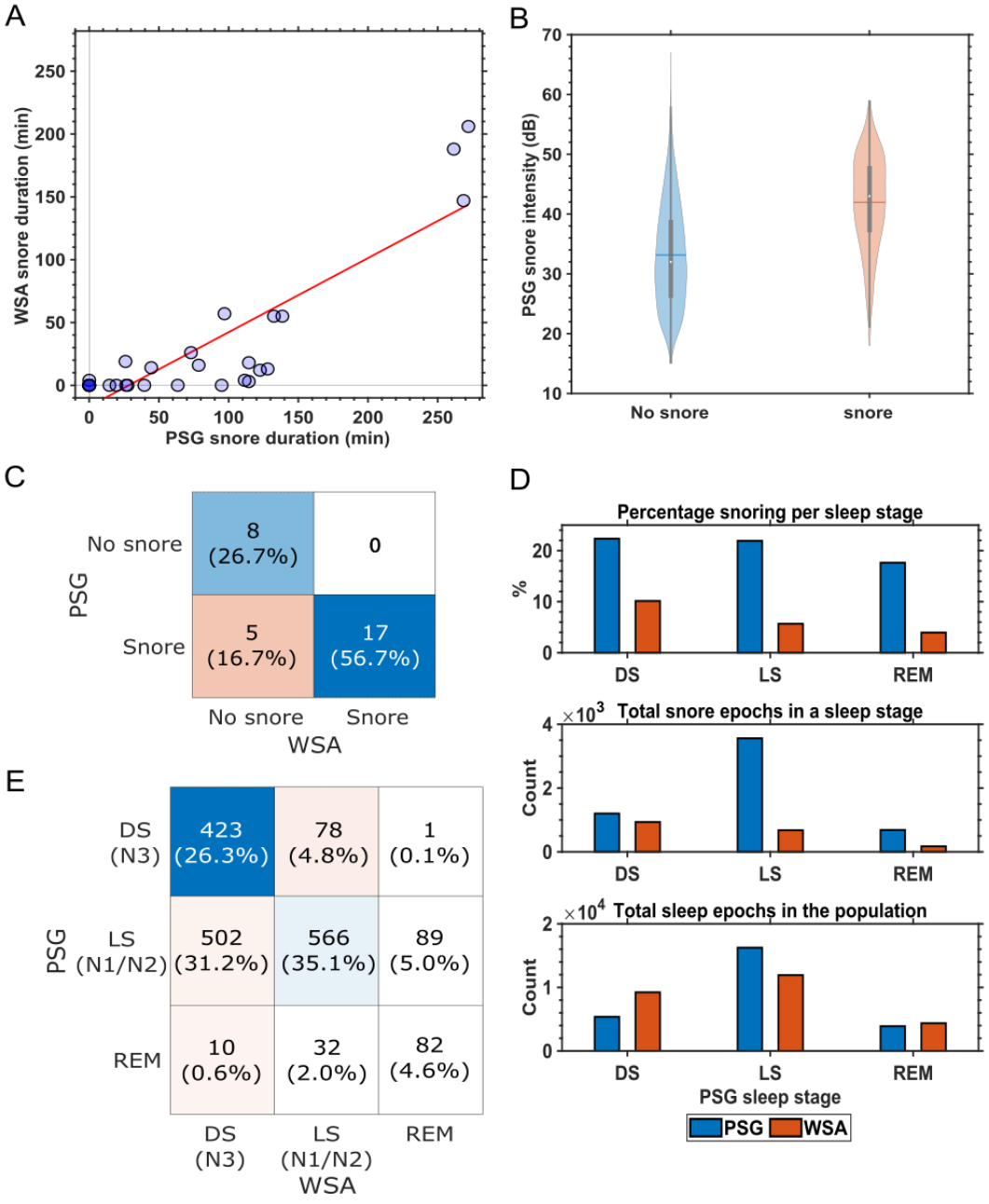
The results of the WSA snore analysis and an example of the overnight time course of the snore data are depicted Figure 6, as well as in Figure S6 in Multimedia Appendix 1. Out of the 35 participants, snore data from both PSG and WSA were available for 30 (86%) participants (PSG snore sensor data were not available for 2 participants, and WSA snore data were unavailable for 3 participants). Of the remaining 30 participants, it was determined from PSG snore sensor data that 8 (27%) participants did not have any form of snoring, whereas 22 (73%) participants snored. The WSA incorrectly determined that 5 (23%) of these 22 participants had no snoring, leading to a moderate performance with a Matthews correlation coefficient of 0.69 (Figure 6C).

PSG determined that the 22 snorers had a nightly snoring duration ranging between 10 and 270 minutes (Figure 6A). The concordance between the snore duration estimates from the WSA and PSG was high (r^2=0.76; P<.001; n=30). When the snoring intensity determined by the PSG snore sensor was grouped based on the WSA snore labels (Figure 6B), we found that, on average, the intensity of the PSG-detected snore events that were not detected by WSA was lower than the intensity of the snore events detected by both PSG and WSA, but there was a considerable overlap between the distributions.

Overall, the snore events were underestimated by the WSA compared with PSG. The distribution of the snoring events as determined by the PSG snore sensor showed that snoring was present across all sleep states, with >20% of all non-REM epochs having snore events, while for REM epochs, this was 17.48% (Figure 6D). By contrast, when the distribution of snore events automatically identified by WSA was analyzed, the

WSA-determined snore events were high during LS and DS and low during REM (Figure 6E). The corresponding confusion matrix between the PSG consensus sleep stage and Withings sleep stage prediction during WSA-identified snore events is depicted in Figure 6E, which shows that the WSA does not score wake when snore events are detected. WSA also scored more snore epochs as DS, followed by LS and REM sleep.

Figure 6. Withings Sleep Analyzer (WSA) snore analysis. (A) Concordance of polysomnography (PSG)-assessed snore duration and WSA snore (n=30). The linear fit is depicted by the red line. (B) Snore intensity as detected by the PSG snore sensor during WSA predicted snore and no snore. (C) Confusion matrix for participants identified as snorers or nonsnorers (n=30). (D) Distribution of snore events across PSG-derived sleep stages for both PSG and WSA. (E) Confusion matrix of epoch-to-epoch (EBE) concordance between PSG and WSA during WSA snore events (total WSA snore epochs: 1611). bpm: beats per minute; cpm: cycles per minute; DS: deep sleep; LS: light sleep; REM: rapid eye movement.



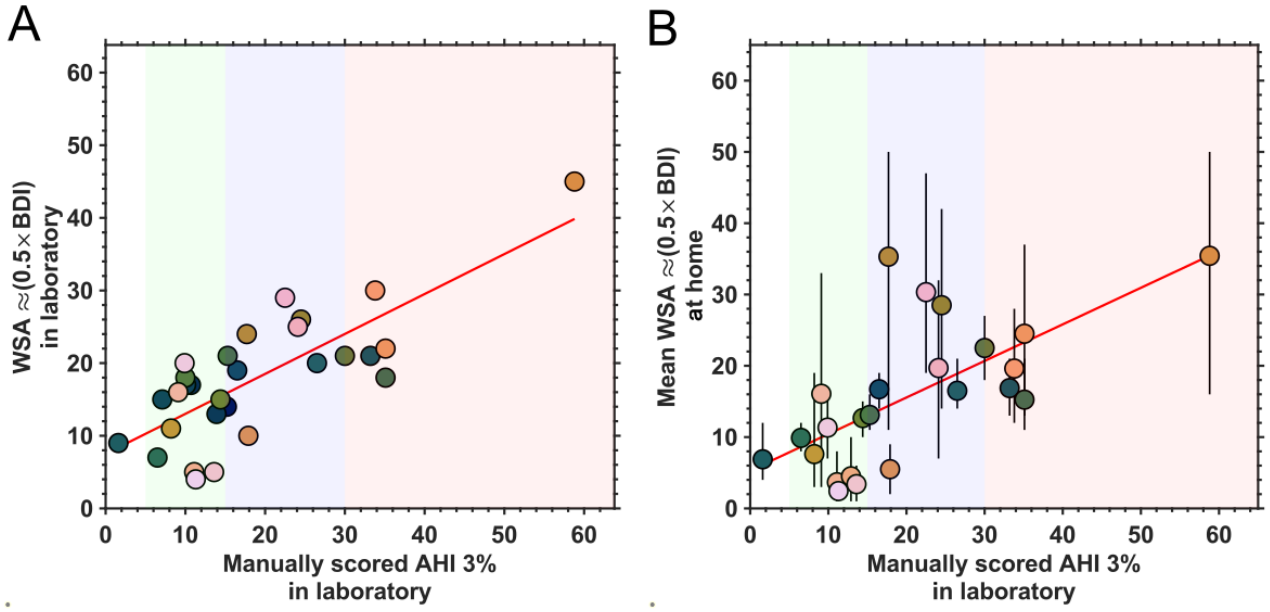
WSA AHI

The WSA summary measures relevant to breathing disorders are the BDI and AHI. WSA BDI data were available for 29 participants in the laboratory and for 24 participants at home (a total of 222 nights available). Both WSA AHI and BDI were available for only 64 nights across 7 participants at home. Upon inspection of potential correlations between the WSA AHI and WSA BDI, we found that the WSA BDI was double the value of WSA AHI (WSA BDI=(2.006×WSA AHI)–0.133; r^2=.999; P<.001; n=64; Figure S7 in Multimedia Appendix 1). On the basis of this inference, we used half BDI as the proxy for WSA AHI for the remainder of the analysis.

We found a high correlation between the WSA AHI and PSG AHI in the laboratory (r^2=0.59; P<.001; n=29; Figure 7A). The MAE was 6.49 (95% CI 4.89-8.10) events per hour. We further explored the relationship between the PSG AHI in the laboratory

and the mean WSA AHI at home (Figure 7B). We found that there was a moderate level of correlation between the 2 (r^2=0.44; P<.001; n=24). There was also a high level of agreement between the WSA AHI 1 night before the laboratory and both in-laboratory WSA AHI (r^2=0.74; P<.001; n=15) and PSG AHI (r^2=0.59; P<.001; n=16). We have depicted the WSA AHI (half WSA BDI) for the 24 participants' data available at home as a heat map showing the night-to-night variability and missing values in Figure S8 in Multimedia Appendix 1.

Figure 7. Relationship between polysomnography (PSG)-based apnea-hypopnea index (AHI) and Withings Sleep Analyzer (WSA) breathing disorder index (BDI). (A) In laboratory (n=35) and (B) at home (n=29). The linear fit is depicted by the red line. AHI reference ranges are as follows: 0 to 4 (no apnea), 5 to 14 (mild; shown in green), 15 to 29 (moderate; shown in blue), and ≥30 (severe; shown in red). For the WSA AHI estimates at home, each data point depicts the mean per participant, and the vertical bars depict the minimum and maximum values.



Discussion

Principal Findings

In this study, we provide an evaluation of 3 contactless technologies for monitoring heart rate, breathing rate, and breathing disturbances during sleep in older men and women. Overall, the contactless technologies provided heart rate (WSA and Emfit) and breathing rate estimates (all 3 devices) with acceptable agreement compared with standard reference estimates from PSG ECG and RIP thorax. We also found that these devices can be used for detecting respiratory events, including apnea and snoring, in this population of older men and women with stable comorbidities.

We were able to successfully collect data at home with limited (<5%) data loss. The data loss was primarily due to Wi-Fi dropouts, where the device spontaneously lost connection to the Wi-Fi network. Overall, this demonstrates the ability of these contactless devices to reliably collect continuous vital sign data remotely in the community with little oversight and maintenance.

The devices also captured the time course of vital signs during sleep, in good agreement with the PSG, with relatively small differences in performance between the devices. The heart rate estimate range of both WSA and Emfit was narrower than the Association for the Advancement of Medical Instrumentation–recommended minimum allowable range of 30 to 200 bpm, with WSA having a more limited range compared with Emfit [48]. WSA performed somewhat better than Emfit at capturing heart rate trends, while Somnofy performed the best in terms of breathing rate, followed by WSA and Emfit. The decline of heart rate during the sleep period is in accordance with previous studies, although the upswing in heart rate at the end of the sleep period, which has been observed in younger participants [52], was not very clear in the data collected by any of the methods in this study. This may be because, here, we studied older participants with comorbidities including sleep apnea.

The contactless technologies provided estimates of heart rate and breathing, which, when averaged across the night, were in very good agreement with the PSG estimates. Outliers in both the breathing rate and heart rate agreement plots were found to originate from participants with severe breathing disturbances or significant abnormal cardiac rhythm. The overall agreement between contactless technology–derived estimates and the estimates derived from PSG becomes poorer when the time period over which the estimate is computed becomes shorter. This is not surprising, but it puts limitations on the use cases in which these devices can be applied. Improvement in the estimates from WSA with reducing temporal resolution was limited by the discretized or rounded output of vital signs from the WSA, with 50% of the estimates having a minimum error of 1 bpm.

At 1-minute resolution, the undermattress devices had an acceptable accuracy with an MAE of <2.12 bpm and an MAPE of <5% for heart rate estimates, which is lower than the errors reported in the literature for wearable technologies during daily activities and for many contactless technologies during sleep

[24,27]. The breathing rate estimates at 1-minute resolution were acceptable across all 3 devices with an MAE of ≤ 1.6-cpm and an MAPE of >12% and were comparable to other contactless technologies in previous evaluations in young participants [27]. The accuracy of both the vital signs estimates was higher during sleep than in the wake period primarily due to reduced body movements. Somnofy, the best-performing device in estimating breathing rate, was also the best-performing device in our evaluation of sleep stage classification performance [35].

Although in our evaluation of older participants, all the compared contactless devices provided acceptable performance, this performance was poorer than what was previously reported in studies (Emfit and Somnofy) in a younger population. In the evaluation conducted by Ranta et al [32] in a population of 34 participants with a median age of 32 years, the Emfit had an MAE of 1.34 bpm for heart rate and an MAE of 0.59 cpm for breathing rate. In contrast, in the evaluation conducted by Toften et al [31], Somnofy had an MAE of 0.18 cpm in a population of 37 participants with a mean age of 32.6 years. Although the vital sign estimates from WSA have been used in large-scale studies, there is no existing evaluation of the WSA-estimated vital signs in the literature to the best of our knowledge [53,54].

The higher accuracy of Somnofy in estimating breathing rate compared with the undermattress devices can be attributed to the device not estimating breathing rate when the signal quality is affected by body movements. Although this leads to some loss of data, this also leads to better accuracy, highlighting the need for a signal quality index associated with the device-generated vital signs estimates.

Finally, our evaluation revealed that the WSA snore and BDI estimates were accurate, and the performance of the WSA AHI in terms of MAE in our study (MAE 6.49 events per hour; n=29) was better than the results reported by Edouard et al [55] (MAE 9.5 events per hour; N=118; mean age 49.3 years). The variability of WSA AHI across nights, as seen in Figure S8 in Multimedia Appendix 1, highlights the ability of contactless monitoring devices like WSA to capture fluctuations in obstructive sleep apnea and their potential to play a crucial role in understanding changes in daytime function, comorbid conditions, and personalized management.

To the best of our knowledge, WSA snore has not been previously evaluated in other studies. The breathing disturbance detection has been performed using Emfit raw ballistography data in the literature, but these algorithms are not open source or available directly from the manufacturer and hence not used in our analysis [56,57]. The availability of these breathing disturbance estimates, along with the acceptable agreement of the vital sign measures generated by the contactless technologies, demonstrates their immediate potential usefulness in population-wide deployment for home monitoring and care [58].

Limitations

One of the limitations of the work is that the data synchronization of the PSG reference data and the device data is based on the best alignment of the vital signs data and hypnogram, which is not ideal. Second, the algorithms used by the different devices in deriving the ballistography signal and heart rate and breathing rate information are hidden due to their proprietary nature, and hence the interpretability of several observations made in this study such as the bounded nature of the output vital signs, outliers, and unavailability of data is limited. Due to a lack of detailed documentation on the snore-detection approach used by WSA, the minimum intensity of the snoring that is detected as a snoring event by WSA is unclear. Finally, out of the compared contactless devices, the Somnofy radar does not provide heart rate estimates, which is a limitation.

Conclusions

With their ability to reliably collect heart rate, breathing rate, and breathing disturbance data longitudinally and at scale, contactless technologies have the potential to be a powerful tool for unintrusive remote vital signs monitoring in community-dwelling older adults and people living with dementia. Applications range from early detection of abnormalities and deterioration of health to monitoring the impact of interventions to improve health (eg, treatment of sleep apnea). This is particularly valuable for patients using prescribed medications for long-term conditions, as it may indicate the need for dosage adjustments or even discontinuation (eg, bradycardia in patients on cholinesterase inhibitors). Together, these applications could improve overall home care. They also allow for investigation into the night-to-night variation in sleep and vital signs and how this variation is associated with health outcomes and daytime function. Such approaches have already shown that night-to-night variation in sleep apnea is associated with uncontrolled hypertension [58] and that night-to-night variation in sleep continuity is associated with day-to-day variation in symptoms in people living with Alzheimer disease [59].

Acknowledgments

This work was supported by the UK Dementia Research Institute (award number UKDRI-7005) through UK DRI Ltd, principally funded by the UK Medical Research Council, and additional funding partner includes Alzheimer's Society. The authors thank the staff members of the Clinical Research Facility for their help with data collection and the members of Care Research & Technology for helpful discussions.

Data Availability

The data sets generated during and analyzed during this study are available from the coauthor CdM on reasonable request.

Authors' Contributions

KKGR conducted the data exploration and analysis and prepared the manuscript. DJD conceived the study and contributed to the writing of the manuscript. CdM, GA, HH, and VR contributed to the design of the study and were responsible for participant recruitment and screening, and study conduct. The devices were setup and data sets were downloaded and curated by DL, KKGR, GA, CdM, and VR. All the authors contributed to the data collection and finalization of the manuscript.

Conflicts of Interest

The authors declare no competing nonfinancial interests but the following competing financial interests: the Withings and Emfit devices used in this study were purchased from the manufacturers without any price reductions. The Somnofy devices used in the study were provided by Vital Things, Norway, at no cost. The manufacturers of these devices were not involved in the design and conduct of this study, the analysis and interpretation of the data, or the preparation of the manuscript.

Multimedia Appendix 1
Supplementary tables and figures.
[DOCX File , 6332 KB - mhealth_v12i1e53643_app1.docx]

References

1. Chester JG, Rudolph JL. Vital signs in older patients: age-related changes. J Am Med Dir Assoc 2011 Jul;12(5):337-343 [FREE Full text] [doi: 10.1016/j.jamda.2010.04.009] [Medline: 21450180]
2. Spruijt B, Vergouwe Y, Nijman RG, Thompson M, Oostenbrink R. Vital signs should be maintained as continuous variables when predicting bacterial infections in febrile children. J Clin Epidemiol 2013 May;66(4):453-457 [FREE Full text] [doi: 10.1016/j.jclinepi.2012.09.014] [Medline: 23306061]
3. Barnett K, Mercer SW, Norbury M, Watt G, Wyke S, Guthrie B. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. Lancet 2012 Jul 07;380(9836):37-43 [FREE Full text] [doi: 10.1016/S0140-6736(12)60240-2] [Medline: 22579043]
4. Divo MJ, Martinez CH, Mannino DM. Ageing and the epidemiology of multimorbidity. Eur Respir J 2014 Oct 19;44(4):1055-1068 [FREE Full text] [doi: 10.1183/09031936.00059814] [Medline: 25142482]
5. Freid VM, Bernstein AB, Bush MA. Multiple chronic conditions among adults aged 45 and over: trends over the past 10 years. NCHS Data Brief 2012 Jul(100):1-8 [FREE Full text] [Medline: 23101759]
6. Browne J, Edwards DA, Rhodes KM, Brimicombe DJ, Payne RA. Association of comorbidity and health service usage among patients with dementia in the UK: a population-based study. BMJ Open 2017 Mar 09;7(3):e012546 [FREE Full text] [doi: 10.1136/bmjopen-2016-012546] [Medline: 28279992]
7. Yamamoto S, Kataoka Y. Delta HR/BT as a possible predictor of severity or mortality, not a predictor of bacterial infection. Postgrad Med 2017 Oct 28;129(7):777-778. [doi: 10.1080/00325481.2017.1358583] [Medline: 28730913]
8. Wrede C, Braakman-Jansen A, van Gemert-Pijnen L. How to create value with unobtrusive monitoring technology in home-based dementia care: a multimethod study among key stakeholders. BMC Geriatr 2022 Dec 30;22(1):921 [FREE Full text] [doi: 10.1186/s12877-022-03550-1] [Medline: 36451119]
9. Lee JP, Freeman G, Cheng M, Brown L, De la Hoz Siegler H, Conly J. Clinical relevance of home monitoring of vital signs and blood glucose levels: a narrative review. Int J Technol Assess Health Care 2019 Jul 26;35(4):334-339. [doi: 10.1017/s0266462319000527]
10. Tiersen F, Batey P, Harrison MJ, Naar L, Serban AI, Daniels SJ, et al. Smart home sensing and monitoring in households with dementia: user-centered design approach. JMIR Aging 2021 Aug 11;4(3):e27047 [FREE Full text] [doi: 10.2196/27047] [Medline: 34383672]
11. Liu L, Stroulia E, Nikolaidis I, Miguel-Cruz A, Rios Rincon A. Smart homes and home health monitoring technologies for older adults: a systematic review. Int J Med Inform 2016 Jul;91:44-59. [doi: 10.1016/j.ijmedinf.2016.04.007] [Medline: 27185508]
12. Valero M, Clemente J, Li F, Song W. Health and sleep nursing assistant for real-time, contactless, and non-invasive monitoring. Pervasive Mob Comput 2021 Aug;75:101422. [doi: 10.1016/j.pmcj.2021.101422]
13. Schütz N, Saner H, Botros A, Pais B, Santschi V, Buluschek P, et al. Contactless sleep monitoring for early detection of health deteriorations in community-dwelling older adults: exploratory study. JMIR Mhealth Uhealth 2021 Jul 11;9(6):e24666 [FREE Full text] [doi: 10.2196/24666] [Medline: 34114966]
14. Hashizaki M, Nakajima H, Kume K. Monitoring of weekly sleep pattern variations at home with a contactless biomotion sensor. Sensors (Basel) 2015 Aug 03;15(8):18950-18964 [FREE Full text] [doi: 10.3390/s150818950] [Medline: 26247948]
15. Seelye A, Leese MI, Dorociak K, Bouranis N, Mattek N, Sharma N, et al. Feasibility of in-home sensor monitoring to detect mild cognitive impairment in aging military veterans: prospective observational study. JMIR Form Res 2020 Jul 08;4(6):e16371 [FREE Full text] [doi: 10.2196/16371] [Medline: 32310138]

16. Fonseka RD, Natarajan P, Maharaj MM, Rooke K, Mobbs RJ. Two-year continuous data capture using a wearable sensor to remotely monitor the surgical spine patient: a case report. J Spine Surg 2022 Mar;8(1):170-179 [FREE Full text] [doi: 10.21037/jss-21-89] [Medline: 35441110]
17. Hilty M, Oldrati P, Barrios L, Müller T, Blumer C, Foege M, et al. Continuous monitoring with wearables in multiple sclerosis reveals an association of cardiac autonomic dysfunction with disease severity. Mult Scler J Exp Transl Clin 2022 Jun 01;8(2):20552173221103436 [FREE Full text] [doi: 10.1177/20552173221103436] [Medline: 35677598]
18. Siddi S, Bailon R, Giné-Vázquez I, Matcham F, Lamers F, Kontaxis S, PAB Members, et al. The usability of daytime and night-time heart rate dynamics as digital biomarkers of depression severity. Psychol Med 2023 Jul;53(8):3249-3260. [doi: 10.1017/S0033291723001034] [Medline: 37184076]
19. Leenen JP, Leerentveld C, van Dijk JD, van Westreenen HL, Schoonhoven L, Patijn GA. Current evidence for continuous vital signs monitoring by wearable wireless devices in hospitalized adults: systematic review. J Med Internet Res 2020 Jul 17;22(6):e18636 [FREE Full text] [doi: 10.2196/18636] [Medline: 32469323]
20. Weenk M, Bredie SJ, Koeneman M, Hesselink G, van Goor H, van de Belt TH. Continuous monitoring of vital signs in the general ward using wearable devices: randomized controlled trial. J Med Internet Res 2020 Jul 10;22(6):e15471 [FREE Full text] [doi: 10.2196/15471] [Medline: 32519972]
21. Henriksen A, Svartdal F, Grimsgaard S, Hartvigsen G, Hopstock LA. Polar vantage and Oura physical activity and sleep trackers: validation and comparison study. JMIR Form Res 2022 May 27;6(5):e27248 [FREE Full text] [doi: 10.2196/27248] [Medline: 35622397]
22. Nelson BW, Allen NB. Accuracy of consumer wearable heart rate measurement during an ecologically valid 24-hour period: intraindividual validation study. JMIR Mhealth Uhealth 2019 Mar 11;7(3):e10828 [FREE Full text] [doi: 10.2196/10828] [Medline: 30855232]
23. Fuller D, Colwell E, Low J, Orychock K, Tobin MA, Simango B, et al. Reliability and validity of commercially available wearable devices for measuring steps, energy expenditure, and heart rate: systematic review. JMIR Mhealth Uhealth 2020 Oct 08;8(9):e18694 [FREE Full text] [doi: 10.2196/18694] [Medline: 32897239]
24. Benedetti D, Olcese U, Frumento P, Bazzani A, Bruno S, d'Ascanio P, et al. Heart rate detection by Fitbit ChargeHR : a validation study versus portable polysomnography. J Sleep Res 2021 Dec;30(6):e13346 [FREE Full text] [doi: 10.1111/jsr.13346] [Medline: 33837981]
25. Haveman ME, van Rossum MC, Vaseur RM, van der Riet C, Schuurmann RC, Hermens HJ, et al. Continuous monitoring of vital signs with wearable sensors during daily life activities: validation study. JMIR Form Res 2022 Jan 07;6(1):e30863 [FREE Full text] [doi: 10.2196/30863] [Medline: 34994703]
26. Ravindran KK, della Monica C, Atzori G, Lambert D, Revell V, Dijk DJ. Evaluating the Empatica E4 derived heart rate and heart rate variability measures in older men and women. In: Proceedings of the 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society. 2022 Presented at: EMBC 2022; July 11-15, 2022; Glasgow, UK. [doi: 10.1109/embc48229.2022.9871559]
27. Boiko A, Martínez Madrid N, Seepold R. Contactless technologies, sensors, and systems for cardiac and respiratory measurement during sleep: a systematic review. Sensors (Basel) 2023 May 24;23(11):5038 [FREE Full text] [doi: 10.3390/s23115038] [Medline: 37299762]
28. Saner H, Knobel SE, Schuetz N, Nef T. Contact-free sensor signals as a new digital biomarker for cardiovascular disease: chances and challenges. Eur Heart J Digit Health 2020 Dec;1(1):30-39 [FREE Full text] [doi: 10.1093/ehjdh/ztaa006] [Medline: 36713967]
29. Saner H, Schütz N, Botros A, Urwyler P, Buluschek P, du Pasquier G, et al. Potential of ambient sensor systems for early detection of health problems in older adults. Front Cardiovasc Med 2020 Jul 15;7:110 [FREE Full text] [doi: 10.3389/fcvm.2020.00110] [Medline: 32760739]
30. Zhang G, Vahia IV, Liu Y, Yang Y, May R, Cray HV, et al. Contactless in-home monitoring of the long-term respiratory and behavioral phenotypes in older adults with COVID-19: a case series. Front Psychiatry 2021 Oct 27;12:754169 [FREE Full text] [doi: 10.3389/fpsyt.2021.754169] [Medline: 34777058]
31. Toften S, Kjellstadli JT, Thu OK, Ellingsen OJ. Noncontact longitudinal respiratory rate measurements in healthy adults using radar-based sleep monitor (Somnofy): validation study. JMIR Biomed Eng 2022 Aug 12;7(2):e36618 [FREE Full text] [doi: 10.2196/36618] [Medline: 38875674]
32. Ranta J, Aittokoski T, Tenhunen M, Alasaukko-oja M. EMFIT QS heart rate and respiration rate validation. Biomed Phys Eng Express 2019 Jan 15;5:025016. [doi: 10.1088/2057-1976/aafbc8]
33. Mishica C, Kyröläinen H, Hynynen E, Nummela A, Holmberg HC, Linnamo V. Evaluation of nocturnal vs. morning measures of heart rate indices in young athletes. PLoS One 2022 Jan 5;17(1):e0262333 [FREE Full text] [doi: 10.1371/journal.pone.0262333] [Medline: 34986202]
34. Ravindran KK, Della Monica C, Atzori G, Lambert D, Hassanin H, Revell V, et al. Contactless and longitudinal monitoring of nocturnal sleep and daytime naps in older men and women: a digital health technology evaluation study. Sleep 2023 Oct 11;46(10):zsad194 [FREE Full text] [doi: 10.1093/sleep/zsad194] [Medline: 37471049]
35. G Ravindran KK, Della Monica C, Atzori G, Lambert D, Hassanin H, Revell V, et al. Three contactless sleep technologies compared with actigraphy and polysomnography in a heterogeneous group of older men and women in a model of mild

sleep disturbance: sleep laboratory study. JMIR Mhealth Uhealth 2023 Oct 25;11:e46338 [FREE Full text] [doi: 10.2196/46338] [Medline: 37878360]
36. Della Monica C, Ravindran KK, Atzori G, Lambert DJ, Rodriguez T, Mahvash-Mohammadi S, The UK Dementia Research Institute Care Research Amp Technology Research Group, et al. A protocol for evaluating digital technology for monitoring sleep and circadian rhythms in older people and people living with dementia in the community. Clocks Sleep 2024 Mar 29;6(1):129-155 [FREE Full text] [doi: 10.3390/clockssleep6010010] [Medline: 38534798]
37. Carney CE, Buysse DJ, Ancoli-Israel S, Edinger JD, Krystal AD, Lichstein KL, et al. The consensus sleep diary: standardizing prospective sleep self-monitoring. Sleep 2012 Mar 01;35(2):287-302 [FREE Full text] [doi: 10.5665/sleep.1642] [Medline: 22294820]
38. Berry RB, Brooks R, Gamaldo CE, Harding SM, Marcus CL, Vaughn BV. The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications, Version 2.0. Darien, IL: American Academy of Sleep Medicine; 2018.
39. Montplaisir J, Allen RP, Arthur W, Ferini-Strambi L. Restless legs syndrome and periodic limb movements during sleep. In: Principles and Practice of Sleep Medicine. Amsterdam, The Netherlands: Elsevier; 2010:1026-1037.
40. International Classification of Sleep Disorders – Third Edition (ICSD-3). Darien, IL: American Academy of Sleep Medicine; 2014.
41. Vest AN, Da Poian G, Li Q, Liu C, Nemati S, Shah AJ, et al. An open source benchmarked toolbox for cardiovascular waveform and interval analysis. Physiol Meas 2018 Oct 11;39(10):105004 [FREE Full text] [doi: 10.1088/1361-6579/aae021] [Medline: 30199376]
42. Charlton PH, Kotzen K, Mejía-Mejía E, Aston PJ, Budidha K, Mant J, et al. Detecting beats in the photoplethysmogram: benchmarking open-source algorithms. Physiol Meas 2022 Aug 19;43(8):085007 [FREE Full text] [doi: 10.1088/1361-6579/ac826d] [Medline: 35853440]
43. Charlton PH, Bonnici T, Tarassenko L, Clifton DA, Beale R, Watkinson PJ. An assessment of algorithms to estimate respiratory rate from the electrocardiogram and photoplethysmogram. Physiol Meas 2016 May;37(4):610-626 [FREE Full text] [doi: 10.1088/0967-3334/37/4/610] [Medline: 27027672]
44. Charlton PH, Bonnici T, Tarassenko L, Alastruey J, Clifton DA, Beale R, et al. Extraction of respiratory signals from the electrocardiogram and photoplethysmogram: technical and physiological determinants. Physiol Meas 2017 May 03;38(5):669-690. [doi: 10.1088/1361-6579/aa670e] [Medline: 28296645]
45. Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res 1999 Jul;8(2):135-160. [doi: 10.1177/096228029900800204] [Medline: 10501650]
46. Menghini L, Cellini N, Goldstone A, Baker FC, de Zambotti M. A standardized framework for testing the performance of sleep-tracking technology: step-by-step guidelines and open-source code. Sleep 2021 Mar 12;44(2):zsaa170 [FREE Full text] [doi: 10.1093/sleep/zsaa170] [Medline: 32882005]
47. Haghayegh S, Kang HA, Khoshnevis S, Smolensky MH, Diller KR. A comprehensive guideline for Bland-Altman and intra class correlation calculations to properly compare two methods of measurement and interpret findings. Physiol Meas 2020 Jul 15;41(5):055012. [doi: 10.1088/1361-6579/ab86d6] [Medline: 32252039]
48. Physical activity monitoring for heart rate - real world analysis (ANSI/CTA-2065.1). Consumer Technology Association. URL: https://shop.cta.tech/collections/standards/products/physical-activity-monitoring-for-heart-rate [accessed 2024-07-31]
49. Lim WS, Carty SM, Macfarlane JT, Anthony RE, Christian J, Dakin KS, et al. Respiratory rate measurement in adults--how reliable is it? Respir Med 2002 Jan;96(1):31-33 [FREE Full text] [doi: 10.1053/rmed.2001.1203] [Medline: 11863207]
50. Gehrman P, Stepnowsky C, Cohen-Zion M, Marler M, Kripke DF, Ancoli-Israel S. Long-term follow-up of periodic limb movements in sleep in older adults. Sleep 2002 May 01;25(3):340-343. [doi: 10.1093/sleep/25.3.340] [Medline: 12003165]
51. Ancoli-Israel S, Kripke DF, Klauber MR, Mason WJ, Fell R, Kaplan O. Periodic limb movements in sleep in community-dwelling elderly. Sleep 1991 Dec;14(6):496-500. [doi: 10.1093/sleep/14.6.496] [Medline: 1798881]
52. Vandewalle G, Middleton B, Rajaratnam SM, Stone BM, Thorleifsdottir B, Arendt J, et al. Robust circadian rhythm in heart rate and its variability: influence of exogenous melatonin and photoperiod. J Sleep Res 2007 Jul 31;16(2):148-155. [doi: 10.1111/j.1365-2869.2007.00581.x] [Medline: 17542944]
53. Soreq E, Kolanko M, Guruswamy Ravindran KK, Monica CD, Revell V, Lai H, et al. 042 Longitudinal assessment of sleep/wake behaviour in dementia patients living at home. J Neurol Neurosurg Psychiatry 2022 Aug 12;93(9):e2.237. [doi: 10.1136/jnnp-2022-abn2.86]
54. Soreq E, Kolanko MA, Monica CD, Ravindran KK, Revell VL, de Villèle P, et al. Monitoring abnormal nocturnal behaviour in the homes of patients living with dementia. Alzheimers Dement 2022 Dec 20;18(S2):e067936. [doi: 10.1002/alz.067936]
55. Edouard P, Campo D, Bartet P, Yang RY, Bruyneel M, Roisman G, et al. Validation of the Withings Sleep Analyzer, an under-the-mattress device for the detection of moderate-severe sleep apnea syndrome. J Clin Sleep Med 2021 Jul 01;17(6):1217-1227 [FREE Full text] [doi: 10.5664/jcsm.9168] [Medline: 33590821]
56. Perez-Macias JM, Tenhunen M, Varri A, Himanen SL, Viik J. Detection of snores using source separation on an Emfit signal. IEEE J Biomed Health Inform 2018 Jul;22(4):1157-1167. [doi: 10.1109/jbhi.2017.2757530]

57. Nokelainen P, Perez-Macias JM, Himanen SL, Hakala A, Tenhunen M. Methods for detecting abnormal ventilation in children - the case study of 13-years old Pitt-Hopkins girl. Child Neurol Open 2023 Feb 20;10:2329048X231151361 [FREE Full text] [doi: 10.1177/2329048X231151361] [Medline: 36844470]
58. Lechat B, Loffler KA, Reynolds AC, Naik G, Vakulin A, Jennings G, et al. High night-to-night variability in sleep apnea severity is associated with uncontrolled hypertension. NPJ Digit Med 2023 Mar 30;6(1):57 [FREE Full text] [doi: 10.1038/s41746-023-00801-2] [Medline: 36991115]
59. Balouch S, Dijk DA, Rusted J, Skene SS, Tabet N, Dijk DJ. Night-to-night variation in sleep associates with day-to-day variation in vigilance, cognition, memory, and behavioral problems in Alzheimer's disease. Alzheimers Dement (Amst) 2022 May 16;14(1):e12303 [FREE Full text] [doi: 10.1002/dad2.12303] [Medline: 35603140]

Abbreviations

AASM: American Academy of Sleep Medicine
AHI: apnea-hypopnea index
AWS: Actiwatch Spectrum
BDI: breathing disorder index
bpm: beats per minute
CDF: cumulative distribution function
cpm: cycles per minute
ECG: electrocardiogram
ICC: intraclass correlation
LoA: limits of agreement
MAE: mean absolute error
MAPE: mean absolute percentage error
MDC: minimum detectable change
PLMI: period limb movement index
PSG: polysomnography
REM: rapid eye movement
RIP thorax: respiratory inductance plethysmography thorax
WSA: Withings Sleep Analyzer

Edited by L Buis; submitted 13.10.23; peer-reviewed by M Chee, J Starling-Smith, M Shekh; comments to author 25.03.24; revised version received 13.05.24; accepted 25.06.24; published 27.08.24.

Please cite as:
G Ravindran KK, della Monica C, Atzori G, Lambert D, Hassanin H, Revell V, Dijk DJ
Reliable Contactless Monitoring of Heart Rate, Breathing Rate, and Breathing Disturbance During Sleep in Aging: Digital Health Technology Evaluation Study
JMIR Mhealth Uhealth 2024;12:e53643
URL: https://mhealth.jmir.org/2024/1/e53643
doi:10.2196/53643
PMID:39190477

Original Paper

A Remote Patient Monitoring System With Feedback Mechanisms Using a Smartwatch: Concept, Implementation, and Evaluation Based on the activeDCM Randomized Controlled Trial

Reto Wettstein[1]; Farbod Sedaghat-Hamedani[2,3,4]; Oliver Heinze[5]; Ali Amr[2,3,4]; Christoph Reich[2,3,4]; Theresa Betz[2,4,6]; Elham Kayvanpour[2,3,4]; Angela Merzweiler[1]; Christopher Büsch[7]; Isabell Mohr[2,4]; Birgit Friedmann-Bette[6]; Norbert Frey[2,3,4]; Martin Dugas[1]; Benjamin Meder[2,3,4]

[1]Institute of Medical Informatics, Heidelberg University Hospital, Heidelberg, Germany

[2]Institute for Cardiomyopathies Heidelberg, Heidelberg University Hospital, Heidelberg, Germany

[3]German Centre for Cardiovascular Research, Heidelberg-Mannheim, Germany

[4]Department of Internal Medicine III, Heidelberg University Hospital, Heidelberg, Germany

[5]RheinMain University of Applied Sciences, Wiesbaden, Germany

[6]Department of Sports Medicine, Medical Clinic, Heidelberg University Hospital, Heidelberg, Germany

[7]Institute of Medical Biometry, Heidelberg University, Heidelberg, Germany

Corresponding Author:
Reto Wettstein
Institute of Medical Informatics
Heidelberg University Hospital
Im Neuenheimer Feld 130.3
Heidelberg, 69120
Germany
Phone: 49 62215636617
Email: reto.wettstein@med.uni-heidelberg.de

Abstract

Background: Technological advances allow for recording and sharing health-related data in a patient-centric way using smartphones and wearables. Secure sharing of such patient-generated data with physicians would enable close management of individual health trajectories, monitoring of risk factors, and asynchronous feedback. However, most remote patient monitoring (RPM) systems currently available are not fully integrated into hospital IT systems or lack a patient-centric design.

Objective: The objective of this study was to conceptualize and implement a user-friendly, reusable, interoperable, and secure RPM system incorporating asynchronous feedback mechanisms using a broadly available consumer wearable (Apple Watch). In addition, this study sought to evaluate factors influencing patient acceptance of such systems.

Methods: The RPM system requirements were established through focus group sessions. Subsequently, a system concept was designed and implemented using an iterative approach ensuring technical feasibility from the beginning. To assess clinical feasibility, the system was used as part of the activeDCM prospective randomized interventional study focusing on dilated cardiomyopathy. Each patient used the system for at least 12 months. The System Usability Scale was used to measure usability from a subjective patient perspective. In addition, an evaluation was conducted on the objective wearable interaction frequency as well as the completeness of transmitted data classified into sensor-based health data (SHD) and patient-reported outcome measures (PROMs). Descriptive statistics using box plots and bootstrapped multiple linear regression with 95% CIs were used for evaluation analyzing the influence of age, sex, device experience, and intervention group membership.

Results: The RPM system comprised 4 interoperable components: patient devices, a data server, a data viewer, and a notification service. The system was evaluated with 95 consecutive patients with dilated cardiomyopathy (28/95, 29% female; mean age 50, SD 12 y) who completed the activeDCM study protocol. The system's app achieved a mean System Usability Scale score of 78 (SD 17), which was most influenced by device experience. In total, 87% (83/95) of the patients could integrate the use of the app well or very well into their daily routine, and 71% (67/95) saw a benefit of the RPM system for management of their health condition. On average, patients interacted with the wearable on 61% (SD 26%) of days enrolled in the study. SHD were available

on average for 78% (SD 23%) of days, and PROM data were available on 64% (SD 27%) of weeks enrolled in the study. Wearable interaction frequency, SHD, and PROM completeness were most influenced by intervention group membership.

Conclusions: Our results mark a first step toward integrating RPM systems based on a consumer wearable device for primary patient input into standardized clinical workflows. They can serve as a blueprint for creating a user-friendly, reusable, interoperable, and secure RPM system that can be integrated into patients' daily routines.

(JMIR Mhealth Uhealth 2024;12:e58441) doi:10.2196/58441

KEYWORDS

wearable; consumer device; mobile phone; mobile health; telemedicine; remote patient monitoring; usability; Health Level 7 Fast Healthcare Interoperability Resources; HL7 FHIR; cardiology; heart failure; dilated cardiomyopathy

Introduction

Background

With constant advancements in digitalization, wearable devices such as smartwatches, fitness trackers, or chest straps are becoming an integral part of everyday life. These devices facilitate continuous recording and monitoring of patient-generated data, such as sensor-based health data (SHD) and electronic patient-reported outcome measures (PROMs). If these data were efficiently shared with physicians, they could gain a more comprehensive view of a patient's lifestyle and longitudinal insights into a patient's health trajectory [1]. Although most wearables are designed for the consumer market and are not primarily intended as medical devices, initial clinical trials have shown that the quality of patient-generated data from wearable consumer devices is sufficiently high to answer medical questions [2-4]. In addition, selected wearable consumer devices have been approved for various diagnostic purposes by authorities such as the US Food and Drug Administration. Especially for physical activity and cardiovascular monitoring, the development of wearables with accurate sensors is well advanced (eg, for step count as well as heart rate and electrocardiogram [ECG] measurements).

Patients show willingness to share their self-generated data with physicians, anticipating that the integration of wearables into their health care journey positively enhances their health [5]. Consequently, this holds potential for both patients and physicians to engage in collaborative health management. This can include monitoring the patient's health status; managing risk factors; and facilitating asynchronous feedback to support effective self-management, especially in the case of chronic conditions [6]. In addition, it has the potential of reducing the frequency of needed physician consultations; improving quality of life; and, ultimately, reducing long-term treatment costs [7-10].

To exploit this potential, suitable IT systems following a patient-centric design must be available. In the literature, these systems are often referred to as *telehealth*, *telemedicine*, *eHealth*, *mHealth*, or *remote patient monitoring* (RPM) systems; the terminology is used interchangeably and inconsistently [11,12]. RPM systems must be able to process and transfer large amounts of data in an automated manner and should facilitate the integration of wearable devices into a patient's everyday life as well as into standardized primary care and clinical research workflows. Methods for realizing such integrations are still immature due to numerous challenges in areas including patient digital literacy, data overload, interoperability, data privacy, data protection, and information security [13-16]. Overall, high usability and, consequently, high acceptance of RPM systems among patients during their treatment process is needed a priori. In addition, wearables are mostly used by healthy individuals as lifestyle devices [17] or for early disease detection in younger adults, such as in the Apple Heart Study [18] or the Fitbit Heart Study [19].

Objectives

In its position papers on wearable-based detection of arrhythmias [20] and eCardiology [21], the German Cardiac Society and other international organizations emphasize the benefits of wearables and RPM systems for primary care and clinical research, including the treatment of heart failure. This chronic disease has a high worldwide prevalence, with an estimate of up to 64 million individuals affected. In high-income countries, it is assumed that 1% to 2% of the adult population has heart failure [22]. These patients usually have multiple contact with health care providers each year; the rehospitalization rates can approach 30% within 90 days of discharge [23]. RPM systems have been shown to detect early health deterioration of patients with chronic heart failure, triggering therapeutic interventions that could reduce rehospitalizations. Overall, approximately 30% to 40% of patients with heart failure have nonischemic cardiomyopathy, such as dilated cardiomyopathy (DCM) [24]. Patients with DCM face physical and quality of life limitations and are often young. In the past, they have been discouraged from physical activity [25]. However, exercise has, in principle, been shown to positively impact morbidity, quality of life, and patients' psychological state, which in turn enhances their physical well-being [26-28]. Sports-related complications, such as ventricular arrhythmias, can counteract the positive effects. Therefore, the activeDCM study [29] investigated the impact of an individualized exercise program in patients with DCM. In this paper, we introduce the RPM system used in activeDCM, detailing its conceptualization, implementation, and evaluation. The objective of this study was to create a user-friendly, reusable, interoperable, and secure RPM system incorporating asynchronous feedback mechanisms using a broadly available consumer wearable device for primary patient input. In addition, this study sought to evaluate factors influencing patient acceptance of the developed RPM system.

Methods

Study Design

activeDCM was designed as a prospective, randomized, interventional case-control study and is described by Sedaghat-Hamedani et al [29]. The primary outcome measure of activeDCM was defined as the change in maximum oxygen uptake, whereas the secondary outcome measures focused on changes in quality of life and behavioral lifestyle. The inclusion criteria for patients were age between 18 and 75 years and diagnosis of nonischemic DCM (left ventricular ejection fraction of ≤50% and New York Heart Association Classification of I-III). The exclusion criteria were acute myocarditis or Takotsubo syndrome; known or suspected ischemic heart disease; known syndromic DCM; history of syncope, cardiac arrest, sustained ventricular tachycardia, or cardiac decompensation within the previous 3 months; contraindications for exercise testing or training; and pregnancy or breastfeeding in women. Patients were identified during initial presentations or routine examinations at the outpatient center for cardiomyopathies at Heidelberg University Hospital and were screened against the inclusion and exclusion criteria by a physician familiar with the study. Participation in the study lasted at least 12 months for each patient. On the day of enrollment, patients were randomly assigned to one of three study arms: (1) intervention group with an individualized exercise program and feedback messages (IG+), (2) intervention group with an individualized exercise program but without feedback messages (IG–), and (3) control group (CG).

Requirements

Using the focus group methodology [30], the following requirements for a patient-centric RPM system were systematically delineated through collaborative sessions involving cardiologists (n=3), mobile health experts (n=2), and potential users (n=2) affiliated with Heidelberg University Hospital, Germany.

The functional requirements were as follows:

1. Patients should use a wearable as primary input device. They should be equipped with a wearable by the study team and not use their own device.
2. Patients should be able to record and transmit patient-generated data (ie, the SHD steps, active burned energy, heart rate, and ECG as well as the PROMs of a study-specific, weekly 7-part questionnaire and an optional exercise diary) in near real time to the study center.
3. Physicians should be able to send personalized feedback messages to patients with motivational content about the individualized exercise program.
4. Physicians should be able to review the substantial volume of transmitted patient-generated data in a streamlined format.

The nonfunctional requirements were as follows:

1. The RPM system should prioritize usability, thereby minimizing barriers to integrating the wearable into patients' daily routines as many patients are older and a prestudy showed little penetration of wearable devices in this cohort (data not shown).
2. The RPM system should be based on medical IT standards for interoperability.
3. The RPM system should meet the regulatory requirements for data privacy and data protection as well as provide state-of-the-art information security.
4. The RPM system should be usable beyond the activeDCM study.

Concept and Implementation

On the basis of the identified requirements, a concept of an RPM system with asynchronous feedback mechanisms was designed and subsequently implemented using an iterative approach. In each iteration, a new module of patient-generated data or for feedback messages was added to the system. Throughout the implementation process, a test environment was available, allowing for the immediate testing of each newly added module. This facilitated short feedback cycles with test participants and physicians from the activeDCM study and ensured the technical feasibility of a patient-centric RPM system.

The final concept, as well as the implementation, consisted of 4 main components: patient devices, data server, data viewer, and notification service (see the Results section). The patient devices used in this study included an iPhone (SE generation 1 or newer) and an Apple Watch (Series 4 or newer, generously provided by Apple Inc). A study-specific wearable and smartphone app was implemented using the Swift programming language (version 5.1; Apple Inc) [31] for the operating systems of both devices (ie, iOS version 14 and watchOS version 7). The extraction of the SHD (steps, active burned energy, heart rate, and ECG) was based on the Core Motion [32] and HealthKit [33] frameworks for iOS and watchOS development. The PROMs (study-specific, weekly questionnaire and optional exercise diary) were recorded by means of a self-designed user interface using Apple's UIKit [34] framework. The transfer of data from the Apple Watch to the iPhone and vice versa was based on the iOS and watchOS Watch Connectivity protocol [35]. The data storage on the iPhone was based on the Realm database (version 5.5; MongoDB Inc) [36].

The implementation of the data server as well as the transmission of the patient-generated data was based on the standardized application programming interface (API) and the standardized data model of Health Level 7 (HL7) Fast Healthcare Interoperability Resources (FHIR) version R4 [37], as well as the Logical Observation Identifiers Names and Codes (LOINC) [38] and the Unified Medical Language System (UMLS) [39] (see the Data Model section) in the Java programming language (version 11; Oracle Corporation) [40] using the HAPI library (version 5.2) [41]. The database technology used was PostgreSQL (version 12; PostgreSQL Global Development Group) [42].

The physician data viewer component was implemented as a mobile-first TypeScript (version 4.1; Microsoft Corp) [43] web application using the vue.js framework (version 2) [44] and the chart.js library (version 2) [45]. The notification service was provided by the device manufacturer (ie, Apple Inc). Local on-device and remote push notifications for the patient devices

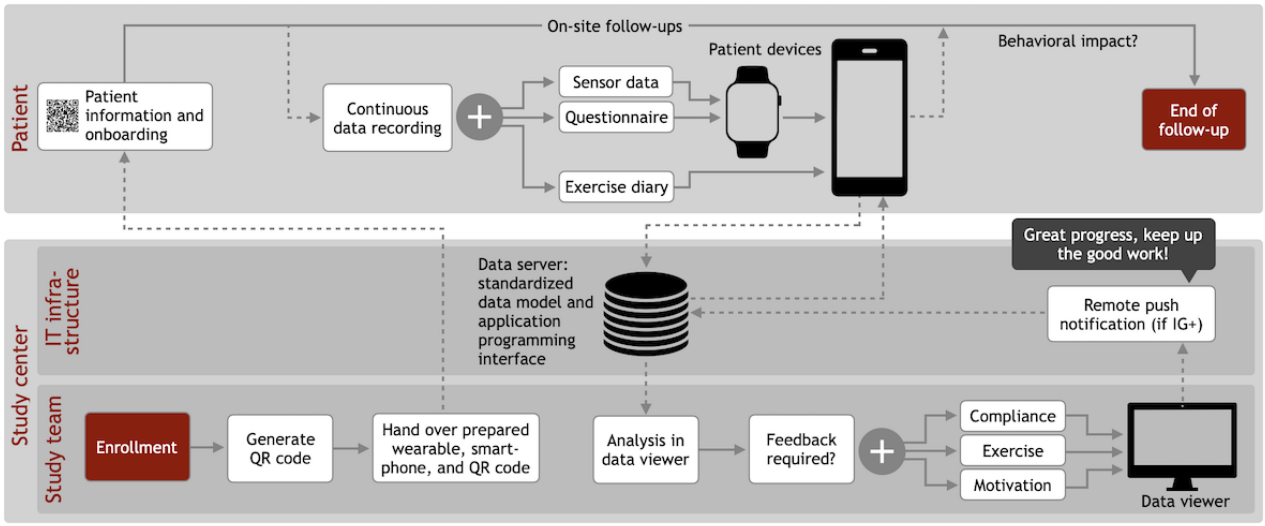
were implemented using the User Notifications [46] framework for iOS and watchOS development.

The aforementioned version numbers correspond to the ones at the beginning of the implementation process and were regularly updated during the use of the RPM system if new applicable versions were released or APIs changed. These updates did not change any functionality or user interfaces of any component.

Procedure

To ensure both technical and clinical feasibility, the RPM system was used in the activeDCM study following a procedure that was developed in addition to the requirements as part of the focus group sessions. The procedure is illustrated in Figure 1. On the day of enrollment, patients were randomized to 1 of the 3 study arms and provided with a configured device bundle consisting of a consumer wearable (Apple Watch Series 4 or newer) and a smartphone (iPhone SE generation 1 or newer) on which the corresponding app of the RPM system was preinstalled. Patients were not provided with a mobile data volume contract for the devices; instead, they were required to set up a Wi-Fi connection at their home. After a thorough patient educational session regarding the study and the RPM system covering instructions on use of the wearable and smartphone, including the RPM-specific app and how to launch a workout session with resultant increased frequency of activity measures, patients signed the written informed consent form. After that, they were onboarded to the app by scanning a patient-specific QR code using the smartphone's camera. From this point forward, the required SHD were continuously recorded by the wearable. The synchronization of the recorded data with the study center was confirmed by patients through a short daily interaction with the wearable. In addition, patients were asked to answer 1 randomly selected question daily from the study-specific, weekly 7-part PROM questionnaire [47], requiring a second daily interaction. To remind patients of these 2 mandatory interactions with the wearable, they were alerted daily at 11 AM and 5 PM with local on-device push notifications sent directly by the wearable app. An interaction with the smartphone was not mandatory but could take place to keep an optional PROM exercise diary, which was synchronized with the study center as well. Each patient was part of the activeDCM study over a period of at least 12 months. Patients were allowed to withdraw at any time from the study. During participation in the study, patient-generated data were collected and transmitted from the wearable via the smartphone to the study center, which in turn provided feedback messages to the patients in the corresponding intervention group (IG+). Feedback messages containing motivational content about the individualized exercise program were sent by a physician using the study center's data viewer approximately every 7 to 10 days using remote push notifications.

Figure 1. Use of the remote patient monitoring (RPM) system in the activeDCM study. After onboarding, patients are equipped with devices whose RPM system app is activated by scanning a patient-specific QR code. Patient-generated data are then continuously recorded and transmitted to the study center until the end of follow-up. Physicians analyze the data and send feedback messages to the patients' devices if appropriate. IG+: intervention group with feedback messages.



Evaluation

The evaluation of the RPM system was conducted in 2 parts. First, to assess patients' subjective perspectives on the RPM system, a 2-part evaluation questionnaire was used to address the following two end points: (1) the first part of the evaluation questionnaire asked about experience (no: without previous device experience [Exp–], yes: with previous device experience [Exp+]) using Apple devices (iPhone or Apple Watch) before participating in the activeDCM study and used a German version [48] of the standardized System Usability Scale (SUS) [49] to assess user-friendliness, and (2) the second part contained 5 self-designed questions to assess patients' attitudes toward the use of an RPM system for disease management.

Both parts of the evaluation questionnaire—the SUS and self-designed questions—used a 5-point ordinal Likert scale with each item ranging from *strongly disagree* to *strongly agree*. An SUS value of >68 was considered user-friendly [50]. The evaluation questionnaire was administered to each patient at completion of the activeDCM study protocol after at least 12 months and can be found in Multimedia Appendix 1.

Second, to assess patients' objective use of the wearable and the interaction with the study-specific wearable and smartphone app, three additional end points were evaluated: (3) to assess *wearable interaction frequency*, the number of days with at least one wearable interaction (either SHD transmission or answering a PROM questionnaire item) was compared with the total number of days that a patient was enrolled in the study; (4) *SHD completeness* was assessed by comparing the days with available SHD on the data server with the total number of days that a patient was enrolled in the study; and (5) finally, to assess *PROM completeness*, the number of completed and available PROM questionnaires on the data server was compared to the total number of PROM questionnaires expected to be completed during the study period (1 per week).

The statistical analysis of all 5 end points was conducted using descriptive statistics incorporating box plots. Continuous variables were described using mean and SD, and categorial variables were described using absolute and relative frequencies. Assessment of differences in the patient demographics of the evaluation groups was performed using the Kruskal-Wallis test [51] for continuous variables. For categorical variables, the Pearson chi-square test [52] was used.

As the number of days for which each patient participated in the activeDCM study varied, we calculated relative values for end points 3, 4, and 5 for each patient. We then described these relative values using mean and SD and used them for further statistical analysis. For comprehensiveness, we reported the ratio of mean absolute values in days and weeks across all patients for end points 3, 4, and 5 as well. However, the mean of the relative values offers a more detailed understanding of the differences between patients and produces more accurate results than the ratio of mean absolute values. It is important to note that these 2 measures are not mathematically equivalent.

Furthermore, bias-corrected and accelerated bootstrapped multiple linear regression [53] was performed to investigate evaluation group differences in the SUS, patient wearable interaction frequency, and SHD or PROM completeness. Thus, the outcome (dependent) variables were the SUS score, patient wearable interaction frequency, and SHD or PROM completeness after activeDCM study protocol completion. As independent variables, age in years at activeDCM study protocol completion, sex, device experience (no: Exp–, yes: Exp+), and study arm membership (CG, IG–, IG+) were chosen before the analysis. The analysis involved 10,000 bootstrap resampling distributions applying a CI of 95%. CIs not including 0 were considered as significant (P<.05). The analysis was conducted using Python (version 3.10; Python Software Foundation) [54] using the packages *scipy.stats.kruskal* (version 1.11.1) [55], *scipy.stats.chi2_contingency* (version 1.11.1) [56], *scipy.stats.bootstrap* (version 1.11.1) [57], and *statsmodels.regression.linear_model.OLS* (version 0.14.0) [58].

Due to the exploratory characteristics of this study, the analysis is intended for hypothesis generation only; thus, P values were not adjusted for multiplicity, and P<.05 was regarded as significant.

Ethical Considerations

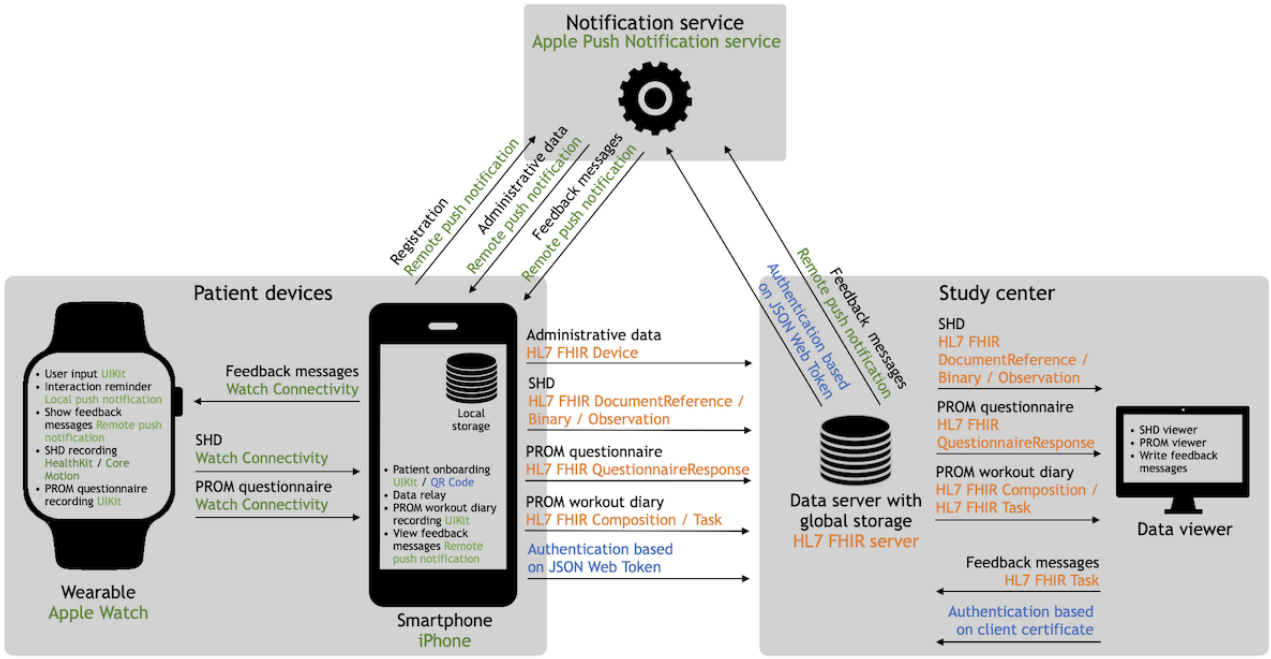
The activeDCM study, as well as its RPM system concept, its implementation, and its evaluation method, received ethics approval from Heidelberg University's research ethics committee (reference numbers S-740/2018 and S-740/2021). The study is registered on ClinicalTrials.gov (NCT04359238). It adheres to the principles of the Declaration of Helsinki and good clinical practice guidelines. Participation in the activeDCM study was voluntary and patients received no compensation. An Apple Watch (Series 4 or newer) and an iPhone (SE generation 1 or newer) were provided during participation in the activeDCM study. After study completion these devices had to be returned to the study team. A written informed consent form was signed by each patient participating in the study. In addition, the RPM system's data protection concept and its security mechanisms were audited by the data protection officer of Heidelberg University Hospital. No deficiencies were identified regarding data privacy, data protection, and information security.

Results

Concept

On the basis of the identified requirements, a concept of a patient-centric RPM system with asynchronous feedback mechanisms was designed. It consisted of 4 interoperable components: patient devices (ie, a smartphone and a connected consumer wearable device, both having a corresponding RPM system app installed), a data server, a data viewer, and a notification service. A graphical representation of the concept and its implementation can be found in Figure 2 summarizing the key features and functionalities of the individual components. In this concept, the smartphone was used for patient onboarding and as a data relay with local storage for data transmission to the data server, which was located at the study center. The wearable was intended to be the patient's primary input device. It was used to record SHD as well as collect PROMs to transmit these patient-generated data to the data server, relayed via the smartphone. In addition, the wearable was supposed to remind patients about necessary interactions and display personalized feedback messages, which were received from the notification service and forwarded by the smartphone. The data server was responsible for global storage of patient-generated data and feedback messages in a standard-compliant, interoperable manner. In addition, it made patient-generated data available to the data viewer. The data viewer, operated at the study center, allowed physicians to analyze patient-generated data and send personalized feedback messages to the data server. These feedback messages were stored on the data server and transmitted to the notification service. The notification service was responsible for identifying the smartphone that would receive the forwarded feedback messages. For the notification service to uniquely identify a smartphone, the latter must undergo an initial registration process during patient onboarding. At the end of this process, administrative data were generated and forwarded from the smartphone to the data server. The data server used the administrative data to identify the smartphone when forwarding feedback messages to the notification service.

Figure 2. Architecture of the remote patient monitoring system with asynchronous feedback mechanisms describing the 4 main components and their interaction. Black writing shows the initial concept, and colored writing shows concepts realized subsequently as part of the implementation. The services, frameworks, and protocols of the Apple ecosystem are shown in green; the standardized data model based on Health Level Seven (HL7) Fast Healthcare Interoperability Resources (FHIR) are shown in orange; and the authentication mechanisms are shown in blue. PROM: patient-reported outcome measure; SHD: sensor-based health data.



Implementation

Overview

The described concept was subsequently implemented for the activeDCM study. For patient devices, an iPhone (SE generation 1 or newer) was used as the smartphone, and an Apple Watch (Series 4 or newer) was used as the wearable for which a study-specific app was implemented. Selected screenshots of the Apple Watch app can be found in Figure 3. The onboarding of a patient to the RPM system was based on scanning a patient-specific QR code on the iPhone's app. After that, SHD recording started immediately on the Apple Watch. Security mechanisms permitting data access and extraction from HealthKit exclusively if the device was unlocked forced the extraction of SHD on the Apple Watch app, which must be unlocked during interaction. Each type of SHD and PROM was implemented in its own module so that each module could be activated and deactivated independently of one another. To remind patients to answer the daily question of the PROM questionnaire and transfer patient-generated data to the study center, local on-device notifications were implemented directly as part of the Apple Watch app.

Figure 3. Screenshots of the Apple Watch app. (A) and (B) show the daily notification to remind patients about answering 1 question of the 7-part weekly patient-reported outcome measure (PROM) questionnaire. (C) and (D) show the user interface to enter 1 of 2 different answer types for the PROM questionnaire—(C) allows for entering a star rating between 1 and 4, and (D) allows for entering a Boolean answer using yes or no. (E) shows the user interface during the transmission of the patient-generated data to the study center.

Transmission of patient-generated data from the Apple Watch to the iPhone was executed using background processes triggered by interactions with the Apple Watch on-device notifications. Transmission errors between Apple Watch and iPhone were displayed to the patient in the user interface, cached, and retried on the next transmission. Once the patient-generated data had been transmitted to the iPhone, they were forwarded again through background processes to the data

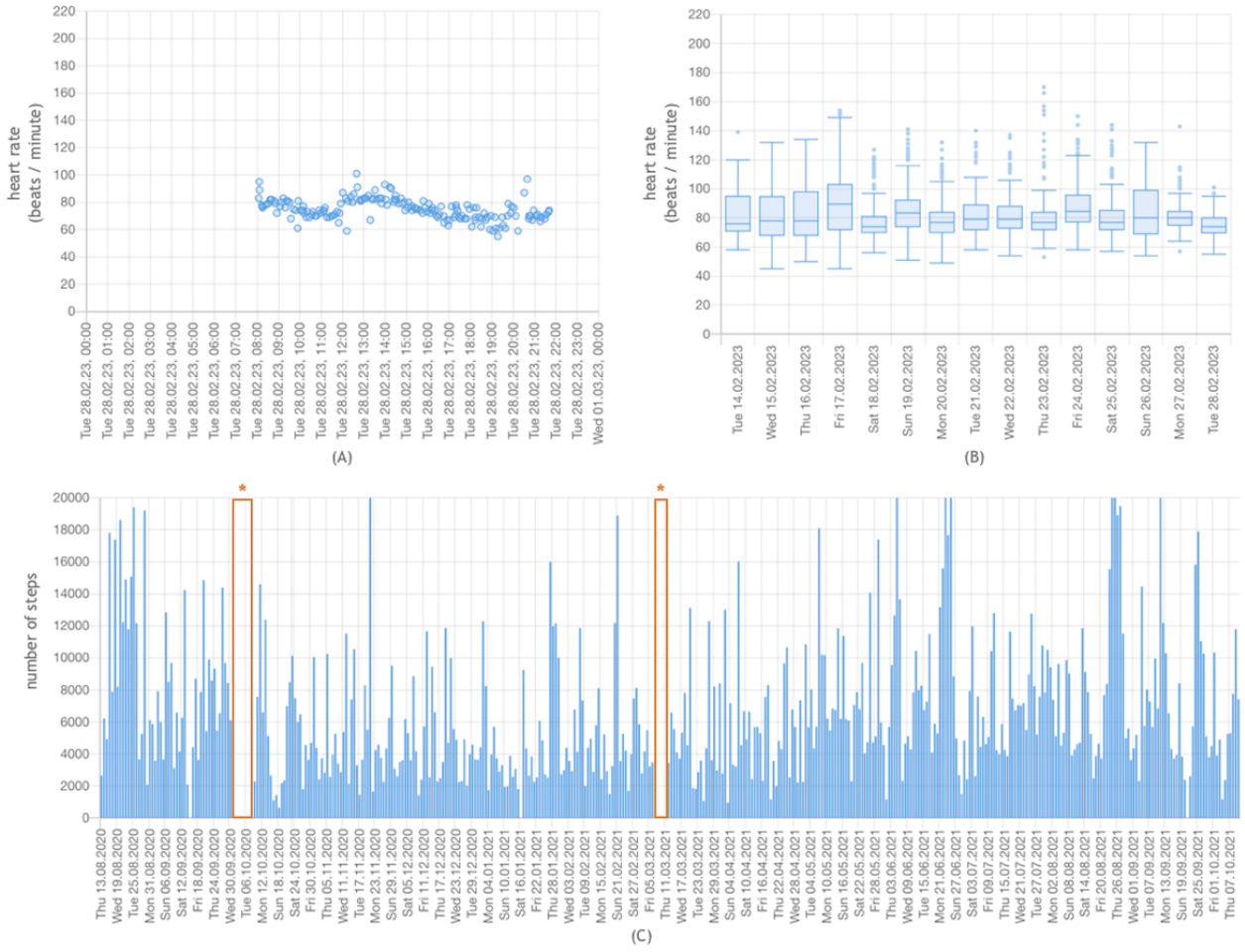
server without any user interaction. This meant that no patient interaction with the iPhone was necessary. It could remain plugged into a power supply socket for the duration of the study and did not have to be carried around. Data transmissions used the standardized HL7 FHIR API provided by the data server. Each data transmission from the iPhone to the data server was recorded with an audit message, which could be viewed by a patient on the user interface of the iPhone app. If a transmission error occurred, the patient-generated data were temporarily stored in an encrypted local database, and the transmission was retried during the next scheduled data transfer to the data server.

The data server provided global storage of patient-generated data and feedback messages using standardized HL7 FHIR R4 resources as a data model. In addition, it executed algorithms periodically to check the date and time of the last data transmission of each patient. If >5 days had passed since a patient's last data transmission, an automated feedback message was sent to the patient's devices with a request to synchronize data.

The data viewer retrieved patient-generated data via the HL7 FHIR API and displayed them through a self-designed user interface. Data transmitted by a patient could be viewed at several levels of granularity (ie, data for a single day, data for 2 weeks, data for 1 month, or all recorded and transmitted data since study enrollment). Depending on the data type, they were presented using different visualization formats: (1) bar plots illustrated various granularities of active burned energy and step data, (2) scatter plots were used for daily heart rate views, and (3) box plots were used for 2 weekly and all-data heart rate views.

If all the data of a patient since study enrollment were displayed, the visualization was zoomable so that a closer look at areas of interest was possible. Selected screenshots of the data viewer user interface can be found in Figure 4. A screencast showing the functionality of the data viewer (with a German user interface) can be found in Multimedia Appendix 2.

Figure 4. Screenshots of the data viewer application for 1 patient; (A) shows the recorded and transmitted heart rate data of a single day as a scatter plot; (B) shows the recorded and transmitted heart rate data over 2 weeks as daily box plots; (C) shows the recorded and transmitted step counts for all days included in the study as a zoomable bar plot, where asterisks indicate days without any recordings or transmissions, for example, because the wearable was not worn.



To receive remote push notifications from the study center, the patient's iPhone registered itself with Apple Push Notification service (APNs) directly after onboarding. APNs provided a device-specific identifier, which was forwarded by the iPhone app to the data server of the study center as administrative data. On the basis of the device identifier, the data server could forward a feedback message received from the data viewer to the APNs server after the message was transformed from the

standardized HL7 FHIR format to the APNs-required proprietary format. The device identifier then enabled the APNs server to forward the feedback message as a remote push notification to the patient's devices. Each feedback message was recorded, in addition to an audit message on the iPhone, and could be viewed by a patient in the app's user interface.

Data Model

The data model was based on standardized HL7 FHIR R4 resources to establish syntactic interoperability. A patient was identified by a *Patient* resource containing the patient's study pseudonym. The *Device* resource stored the device-specific identifier provided by APNs, which was needed to send and retrieve feedback messages via remote push notifications and linked to the *Patient* resource. Feedback messages were modeled using the *Task* resource, which contained information such as title, message, category, and recipient, which was a reference to the *Patient* resource. The SHD steps, active burned energy, and heart rate were modeled using the *Observation* resource. For steps and active burned energy, the recorded data were summarized in 1-hour intervals. For heart rate, 25 consecutive measurements were listed and transferred in 1 *Observation* resource to reduce network calls. The *DocumentReference* resource was used to store metadata related to ECG data, such as recording date, classification, and average heart rate. In addition, it referenced a *Binary* resource to transport the actual ECG recording (512 Hz; 30 seconds) as millivolt values in a CSV file. The PROM questionnaire was based on the *Questionnaire* resource, which contained the questions to be answered weekly, and the *QuestionnaireResponse* resource, which stored a patient's completed questionnaire. The exercise diary was based on the *Composition* resource, which stored the date of a completed workout and was linked to several *Task* resources containing information about the executed strength and endurance exercises during the workout.

To establish semantic interoperability with other clinical information systems (eg, from primary care), the SHD were annotated using LOINC codes, and the PROM questionnaire was annotated using postcoordinated UMLS codes. For semantic annotation of the PROM exercise diary, we used a self-defined code system based on the exercise program. HL7 FHIR profiles, concrete resource examples, and validation information for the data model can be found in Multimedia Appendix 3.

Security, Authentication, and Authorization

For general security, all messages exchanged between the components of the RPM system were encrypted using Transport Layer Security [59] using a minimum version of 1.2. The Realm database used to cache patient-generated data on the iPhone was encrypted using the Advanced Encryption Standard-256+Secure Hash Algorithm 2 algorithm and could not be analyzed if the device was lost. The encryption key of the database was stored in the secure Keychain [60] of the iPhone and could only be accessed through the iPhone app.

Authentication and authorization to the RPM system distinguished between the roles of patient and study center staff using distinct technical approaches for each. For the patient role, the QR code scanned during onboarding contained a signed JSON Web Token (JWT) [61] valid for study enrollment, storing authentication information and the patient's pseudonym in the iPhone's secure Keychain. This JWT was used when transmitting patient-generated data, ensuring authentication and data provenance. The role of study center staff, authenticated using X.509 client certificates [62], could retrieve patient-generated data and stored push notifications on the data server. Revoking compromised certificates or JWTs was managed via configuration options on the data server. The server's authentication to APNs used a JWT as well that was signed and encrypted by the study center data server using a private key from Apple that was revocable on the Apple developer website if compromised.

Evaluation

The evaluation of the RPM system was carried out with patients who completed the activeDCM study protocol between October 2021 and February 2024 (N=110). A total of 13.6% (15/110) of the patients were excluded from the evaluation because of missing answers in the evaluation questionnaire on experience or SUS answers. The evaluated cohort consisted of 95 patients (n=28, 29% female) with a mean age of 50 (SD 12) years at the end of study participation. Of these 95 patients, 39 (41%) had previous experience using Apple devices (Exp+, n=30, 77% using iPhone only and n=9, 23% using iPhone and Apple Watch), and 56 (59%) had no experience (Exp–) at the time of enrollment in the study. A total of 26% (25/95) of the patients were randomized into the intervention group with an individualized exercise program and feedback messages (IG+), 36% (34/95) were randomized into the intervention group with an individualized exercise program but without feedback messages (IG–), and 38% (36/95) were randomized into the CG study arm. The mean enrollment time of these patients in the activeDCM study was 396 (SD 39) days, which corresponds to 56 (SD 5) weeks. The baseline characteristics of the evaluated patient cohort are summarized in Table 1.

Table 1. Baseline characteristics of patients by device experience and study arm membership in the activeDCM study (N=95).

Variable	Total sample	Exp–[a] (n=56)	Exp+[b] (n=39)	CG[c] (n=36)	IG–[d] (n=34)	IG+[e] (n=25)
Sex, n (%)[f]						
Male	67 (71)	40 (71)	27 (69)	27 (75)	24 (71)	16 (64)
Female	28 (29)	16 (29)	12 (31)	9 (25)	10 (29)	9 (36)
Age (y)[g], mean (SD; range)	50 (12; 23-66)	51 (10; 23-66)	49 (13; 24-66)	50 (11; 24-65)	48 (13; 23-66)	53 (11; 24-66)
Time in study (d)[h], mean (SD; range)	396 (39; 315-625)	398 (44; 315-625)	394 (30; 330-456)	402 (50; 315-625)	395 (28; 336-448)	389 (31; 330-484)

[a]Exp–: without previous device experience.

[b]Exp+: with previous device experience.

[c]CG: control group.

[d]IG–: intervention group without feedback messages.

[e]IG+: intervention group with feedback messages.

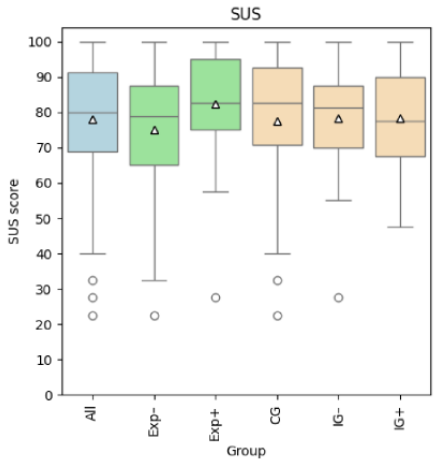
[f]Device experience: Pearson $\chi^2_1 P>.99$; study arm: Pearson $\chi^2_2 P=.65$.

[g]Device experience: Kruskal-Wallis test $P=.58$; study arm: Kruskal-Wallis test $P=.29$.

[h]Device experience: Kruskal-Wallis test $P=.95$; study arm: Kruskal-Wallis test $P=.53$.

The results of the patients' subjective perspective on the RPM system based on the evaluation questionnaire were divided into 2 end points. The results of the first end point, analyzing the usability of the wearable and smartphone app based on the SUS, are summarized using box plots in Figure 5. The app achieved a mean SUS score of 78 (SD 17). A total of 25% (24/95) of the patients reported an SUS score of <68 (insufficient usability), of whom 75% (18/24) were aged >50 years and 75% (18/24) had no previous experience using Apple devices (Exp–). Patients with previous experience using Apple (Exp+) devices rated usability more highly, with a mean SUS score of 82 (SD 15), compared with patients without any experience (Exp–), who reported a mean SUS score of 75 (SD 18). There were no major differences among patients in the study arms. Patients in the intervention group with feedback messages (IG+) reported a mean SUS score of 78 (SD 15), patients in the intervention group without feedback messages (IG–) reported a mean score of 78 (SD 15), and patients in the CG reported a mean score of 77 (SD 20).

Figure 5. Results of the wearable and smartphone app usability analysis using the System Usability Scale (SUS) displayed as box plots. All data combined are shown in blue, experience using Apple devices is shown in green, and study arm membership is shown in orange. Triangles show mean values. CG: control group; Exp–: without previous device experience; Exp+: with previous device experience; IG–: intervention group without feedback messages; IG+: intervention group with feedback messages.



As shown in Table 2, the bias-corrected and accelerated bootstrapped multiple linear regression analysis of the SUS scores resulted in a significant effect for those with previous experience (Exp+) with a 95% CI of 0.25-13.47 and an estimate of 6.75. This means that patients with Apple device experience were estimated to rate the usability of the wearable and smartphone app 6.75 points higher on the SUS than patients without Apple device experience. This is similar to the measured differences in SUS scores between those without and with previous experience (7-point difference). Age approached significance with a 95% CI of −0.46 to 0.14. The estimate of −0.19 showed that a 40-year age difference between 2 patients was estimated to result in a 7.60-point worse SUS score for the wearable and smartphone app from the older patient.

Table 2. Bias-corrected and accelerated bootstrapped multiple linear regression analysis of the System Usability Scale (SUS) scores, the wearable interaction frequency (percentage of days enrolled in the study), the sensor-based health data (SHD; percentage of days enrolled in the study), and patient-reported outcome measure (PROM) completeness (percentage of weeks enrolled in the study).

Variable	Estimate, mean (95% CI)
SUS score	
Sex (male)	−3.61 (−10.35 to 4.54)
Age	−0.19 (−0.46 to 0.14)
Exp+[a]	*6.75 (0.25 to 13.47)* [b]
Group	
IG+[c]	0.95 (−7.80 to 10.10)
IG–[d]	−0.37 (−7.81 to 8.00)
Wearable interaction frequency (%)	
Sex (male)	−7.06 (−17.7 to 3.93)
Age	0.39 (−0.08 to 0.84)
Exp+	−1.38 (−10.81 to 8.65)
Group	
IG+	6.08 (−6.54 to 19.51)
IG–	*13.53 (1.02 to 25.4)* [b]
SHD completeness (%)	
Sex (male)	−8.22 (−16.39 to 0.32)
Age	*0.45 (0.06 to 0.95)* [b]
Exp+	5.36 (−2.44 to 13.88)
Group	
IG+	*12.18 (1.65 to 22.43)* [b]
IG–	*11.73 (1.17 to 21.86)* [b]
PROM completeness (%)	
Sex (male)	−5.49 (−17.14 to 5.71)
Age	0.42 (−0.04 to 0.92)
Exp+	−1.43 (−12.03 to 9.31)
Group	
IG+	7.17 (−7.39 to 20.9)
IG–	12.22 (−0.37 to 24.29)

[a]Exp+: with previous device experience.

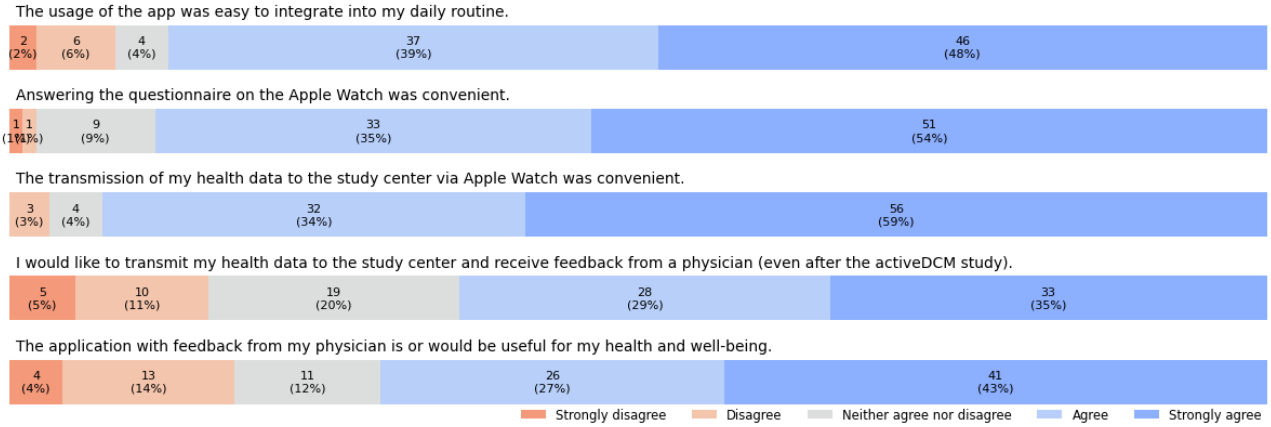
[b]CIs not including 0 were considered as significant (P<.05).

[c]IG+: intervention group with feedback messages.

[d]IG–: intervention group without feedback messages.

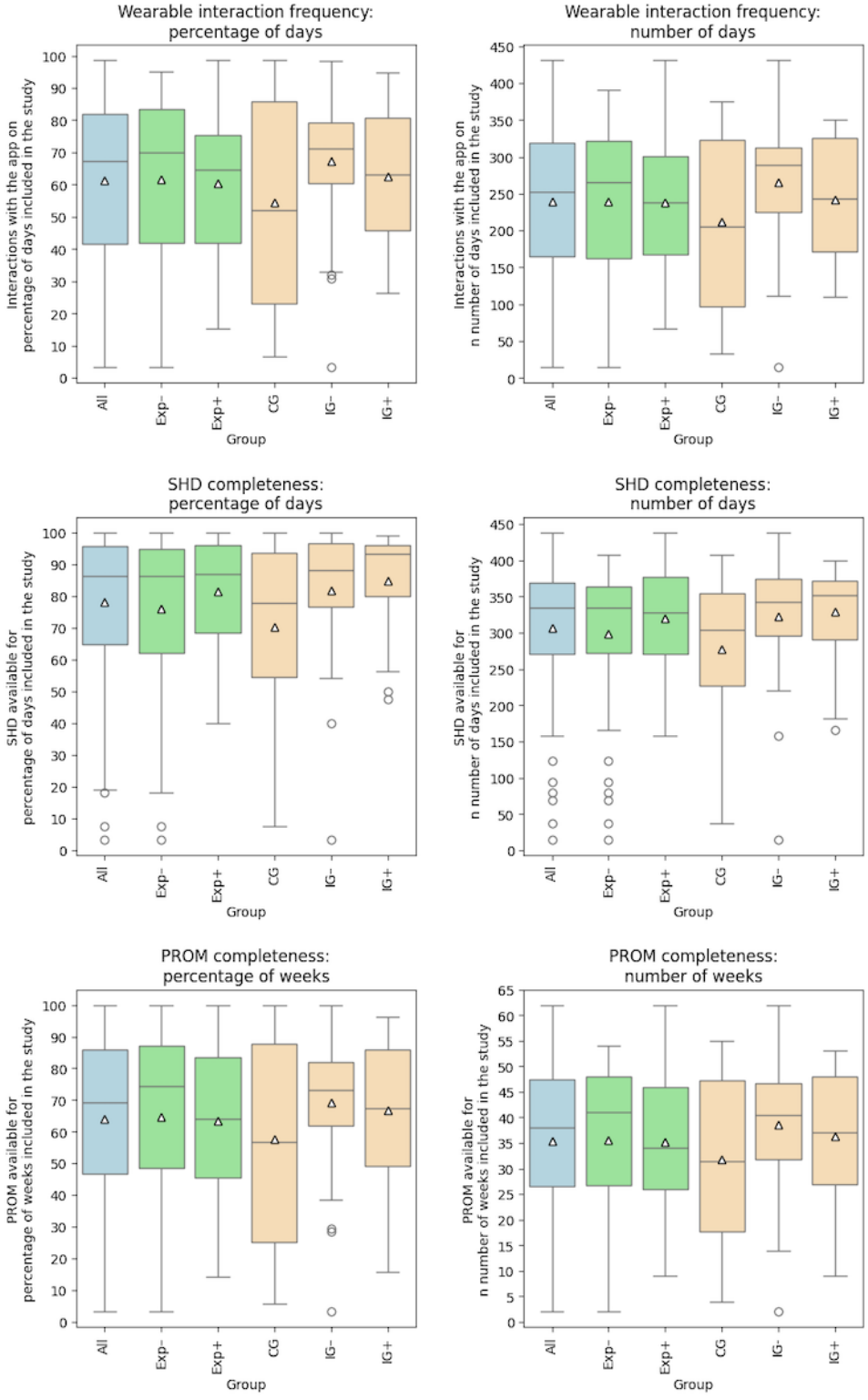
The results of the second evaluation questionnaire end point, regarding the 5 self-designed questions about the patients' attitudes toward using the RPM system as well as their participation in the study, are shown in Figure 6. A total of 87% (83/95) of the patients (strongly) agreed that the use of the app was easy to integrate into their daily routine. Similarly, 88% (84/95) and 93% (88/95) of the patients considered answering the PROM questionnaire and transmitting their SHD to the study center via Apple Watch as convenient, respectively. In total, 64% (61/95) of the patients would like to submit their health data to the study center and receive feedback from a physician (even after the activeDCM study), and the app with feedback from their physician was or would be useful for 71% (67/95) of the patients regarding their health and well-being.

Figure 6. Results of the 5 self-designed questions regarding the patients' attitudes toward using the remote patient monitoring system as well as their participation in the activeDCM study. Responses to the statements were provided by patients using a 5-point ordinal Likert scale ranging from strongly disagree in orange to strongly agree in blue.



In Figure 7, the results of patients' objective use of the wearable device are summarized using box plots, showing wearable interaction frequency and SHD as well as PROM completeness for end points 3, 4, and 5. On average, patients interacted with the wearable app on 61% (SD 26%) of the days that they were enrolled in the study, which corresponds to 239 (SD 99) of 396 (SD 39) days. On average, SHD were available for 78% (SD 23%) of the days that patients were enrolled in the study, which corresponds to 307 (SD 87) of 396 (SD 39) days. In terms of PROM questionnaires answered, data were available on average for 64% (SD 27%) of the weeks that the patients were enrolled in the study, which corresponds to 35 (SD 15) of 56 (SD 5) weeks. There were no major differences between those without and with previous experience (wearable interaction frequency: mean 62%, SD 27% of days [240, SD 103 of 398, SD 44 days] vs mean 60%, SD 23% of days [238, SD 91 of 394, SD 30 days], respectively; SHD completeness: mean 76%, SD 25% of days [299, SD 97 of 398, SD 44 days] vs mean 81%, SD 17% of days [319, SD 69 of 394, SD 30 days], respectively; PROM completeness: mean 65%, SD 28% of weeks [36, SD 15 of 56, SD 6 weeks] vs mean 63%, SD 24% of weeks [35, SD 14 of 56, SD 4 weeks], respectively). Only the dispersion of the data was slightly narrower for those with previous experience (Exp+). Regarding the 3 different study arms, the interaction with the wearable was more frequent and patient-generated data completeness was higher in both the IG– (wearable interaction frequency: mean 67%, SD 20% of days [266, SD 84 of 395, SD 28 days]; SHD completeness: mean 82%, SD 20% of days [322, SD 80 of 395, SD 28 days]; PROM completeness: mean 69%, SD 21% of weeks [39, SD 12 of 56, SD 4 weeks]) and IG+ (wearable interaction frequency: mean 62%, SD 20% of days [242, SD 78 of 389, SD 31 days]; SHD completeness: mean 85%, SD 16% of days [330, SD 63 of 389, SD 31 days]; PROM completeness: mean 67%, SD 23% of weeks [36, SD 12 of 55, SD 4 weeks]) than in the CG (wearable interaction frequency: mean 54%, SD 31% of days [212, SD 115 of 402, SD 50 days]; SHD completeness: mean 70%, SD 26% of days [277, SD 99 of 402, SD 50 days]; PROM completeness: mean 58%, SD 32% of weeks [32, SD 17 of 57, SD 7 weeks]). In both the IG+ and IG–, the dispersion of the data was smaller than in the CG.

Figure 7. Results of patients' wearable interaction frequency as well as completeness of recorded and transmitted patient-generated data classified into sensor-based health data (SHD) and electronic patient-reported outcome measures (PROMs) of the wearable and smartphone app as box plots. All data combined are shown in blue, experience using Apple devices is shown in green, and study arm membership is shown in orange. Triangles show mean values. CG: control group; Exp–: without previous device experience; Exp+: with previous device experience; IG–: intervention group without feedback messages; IG+: intervention group with feedback messages.



The data volume recorded and transferred per patient to the data server varied based on health status, physical activity levels, and workout duration, which influenced both recorded data point density and the number of ECG recordings. On average, the data volume amounted to approximately 300 MB per patient and year in the HL7 FHIR database on the data server.

The bias-corrected and accelerated bootstrapped multiple linear regression analysis (Table 2) resulted in a significant influence of IG– membership on the wearable interaction frequency and SHD completeness end points and approached significance for the PROM completeness end point (wearable interaction frequency estimate: 13.53%, 95% CI 1.02%-25.4%; SHD

completeness estimate: 11.73%, 95% CI 1.17%-21.86%; PROM completeness estimate: 12.22%, 95% CI –0.37% to 24.29%). This means that, on average, a patient in the IG– was expected to interact with the wearable app on 13.35% more days enrolled in study than a patient in the CG and that both SHD and PROM data were expected to be available on 11.73% more days and 12.22% more weeks that a patient was enrolled in the study, respectively.

Discussion

Principal Findings

In this paper, we presented a concept, implementation, and evaluation of a reusable RPM system following a patient-centric approach for RPM without direct physician-patient contact using asynchronous feedback mechanisms. The system was based on patient-generated data such as SHD and PROMs. It leveraged a consumer wearable device (smartwatch) instead of a smartphone as the primary input device. The implementation of the concept and its subsequent use as part of the activeDCM randomized controlled clinical trial demonstrated that the identified challenges and requirements of a sophisticated precision digital health trial could be addressed. The implementation relied on wearable devices and smartphones from Apple Inc and their notification service, but the concept should be versatile enough to be used with devices based on other operating systems, such as the Android-based Wear OS for the Samsung Galaxy Watch or the Google Pixel Watch.

Decisions during RPM system conceptualization and implementation aimed to minimize barriers to use, enhancing usability and ensuring high patient acceptance as well as compliance. The study team established a device preparation and distribution scheme providing patients with ready-to-use devices, thereby reducing the configuration efforts to scanning an onboarding QR code. Patient education sessions focused not only on the medical aspects of the study but also on digital literacy and provided written and video-based material about the wearable and smartphone app, requiring additional effort from patient educational sessions. Opting for a wearable as the primary input device likely played a pivotal role in seamlessly integrating the RPM system into the patients' daily routines. Its proximity to the patient and its intuitive interaction, aligned with the chosen device, operating system, and study-specific app, contributed to a successful integration and reduced the interaction time to a few seconds per day. To enhance the wearable and smartphone app usability, we implemented only the essential functionalities in a straightforward, simplistic manner, trying to reduce the interactions for these functionalities to a minimum. This was significantly shaped by the official human interface guidelines provided by Apple [63,64]. Storing authentication information during onboarding further streamlined data transmissions, eliminating the need for patients to remember usernames and passwords.

The Apple Watch sensors capture heart rate measurements in an interval of seconds during an activity session [65], and active burned energy and step data can be extracted in hourly intervals. Given this vast amount of data points, the physician-centered data viewer prioritized simplicity, presenting summarized information. In total, 3 visualization granularity levels allowed physicians to analyze patient-generated data, identifying short-, medium-, and long-term trends, and compare data from multiple patients effortlessly. Direct remote push notifications from the data viewer interface streamlined physician support, eliminating the need to switch between various communication channels, as observed in other wearable-based studies [66,67].

The HL7 FHIR standard enhanced medical data exchange by addressing health care challenges beyond the capabilities of previous standards, such as HL7 version 2, HL7 version 3, and Clinical Document Architecture [68]. Therefore, HL7 FHIR's standardized API and data models used in combination with the LOINC and UMLS medical terminologies ensured syntactic and semantic interoperability with other clinical information systems. However, complete integration with other clinical information systems requires these systems to provide standardized interfaces as well. Without these interfaces, integration becomes more complex, necessitating adaptations and mappings. In Germany, legislative efforts are promoting the adoption of HL7 FHIR interfaces for clinical information systems [69].

The wearable and smartphone app, data server, and data viewer implementation were not bound to a specific medical domain and could be used independently beyond the activeDCM study as long as the counterpart supported the HL7 FHIR standard as well. The components of the wearable and smartphone app and the data viewer were built iteratively in a modular system and, thus, can be adapted for further studies and other diseases such as diabetes or chronic obstructive pulmonary disease without major effort. If required, new modules could be implemented, integrating additional sensors and data types. Similarly, not required data types could be removed easily.

By equipping patients with dedicated devices, it was possible to ensure that data transmissions only took place in a pseudonymized form. During the educational session, patients were briefed on the RPM system's privacy and security mechanisms, emphasizing the storage of only the pseudonym on devices after handover. This ensured that no data were stored in a cloud solution provided by the manufacturer of the devices. As such, the smartphone and wearable were setup using Apple Configurator 2, and the app was deployed using the over-the-air method without the need to sign in using an Apple ID. Furthermore, by separating patient identification from device identification, feedback messages could be sent via the manufacturer's notification service without having to know the patient's identity or pseudonym. The connection of the pseudonym to the device could only be established on the data server.

The SUS evaluation, resulting in a mean score of 78 (SD 17), indicated that the wearable and smartphone app was considered user-friendly by patients [50]. The significant difference between patients with and without experience using Apple devices (estimate: 6.75, 95% CI 0.25-13.47) suggests that patients are generally more comfortable using devices they are familiar with. However, to allow use of wearables from different manufacturers in the RPM system, achieving comparable sensor accuracies for the same SHD recordings would be essential, but

the literature shows that there are measurable differences [70,71], posing challenges for physicians in comparing patient data and providing feedback. Despite the additional effort to explain the devices and study-specific app, patient age had an influence on the SUS score approaching significance (estimate: −0.19, 95% CI –0.46 to 0.14). Insufficient usability (SUS score of <68; 24/95, 25% of the patients) was mainly reported by patients aged >50 years (18/24, 75%) and with no previous Apple device experience (18/24, 75%). This suggests that the challenges faced by these individuals may not have been due to problems with the usability of the wearable and smartphone app but rather with the general use of the devices. These findings are consistent with results reported by Sonderegger et al [72] and illustrate the importance of digital literacy and the supplementary material provided for patients during the onboarding process.

On average, patients engaged with the wearable more than every other day (61%, SD 26% of days, 239, SD 99 of 396, SD 39 days). For SHD recording and transmission, as few as 1 interaction per week would suffice for complying with the activeDCM study protocol. Complete datasets would still be transferred as Apple Watch sensors store SHD for up to 10 days. The data server algorithm that monitored transmission frequency and sent automatic reminders likely played a crucial role in obtaining complete datasets. The measured user-friendliness was also reflected in the completeness of SHD data, which was available on average for 3 out of 4 days (78%, SD 23% of days, 307, SD 87 of 396, SD 39 days). This aligns with the results by Werhahn et al [73], where heart rate measurements and step counts based on a worn Apple Watch were recorded on 84.8% and 83.5%, respectively, of days that patients were included in their study. However, the monitoring period lasted only 2 months, which is significantly shorter than that in activeDCM. The average completeness of PROM questionnaires showed similar results, which were sufficient to see relevant health changes in patients and corresponds to a transmission of a complete PROM questionnaires more than every second week (mean 64%, SD 27% of weeks, 35, SD 15 of 56, SD 5 weeks).

Gaps in data, whether SHD or PROMs, could be attributed to various factors, such as illness, hospitalization, vacation, or lapses in wearing the Apple Watch. Technical issues, poor internet connection, or oversight in configuring the patient device's Wi-Fi connection at home may also have resulted in missing patient-generated data.

However, home-based Wi-Fi connection did not restrict patient recruitment, the daily use of the wearable and smartphone app, or recording of patient-generated data. Data were still collected even without a current Wi-Fi connection. The Wi-Fi connection was only necessary for data transmission to the study center, thereby transmitting all recorded data since the last transmission. A mobile data volume contract is possible for data transfer, enabling transfers outside the patient's home. However, as the devices and the wearable and smartphone app can temporarily store patient-generated data until they can be transferred to the data server, data completeness would remain unchanged.

There was a notable difference in the evaluation results of objective RPM use between the CG, IG–, and IG+, showing significance for wearable interaction frequency (estimate: 13.53%, 95% CI 1.02%-25.4%) and SHD completeness (estimate: 11.73%, 95% CI 1.17%-21.86%), as well as approaching significance for PROM completeness (estimate: 12.22%, 95% CI −0.37%-24.29%) in the IG– in comparison with the CG. This suggests the presence of a Hawthorne effect [74] for increased RPM system use as anticipated given the expected behavioral impact of the activeDCM study in the intervention groups. Within each randomized controlled trial, a potential bias exists in the case of a Hawthorne effect that could have increased compliance with the treatment and the use of the wearables in this study. To study the use of the RPM system in real-world settings, another design of an observational trial would be needed.

The inferior performance of the IG+ (only significant SHD completeness with an estimate of 12.18%, 95% CI 1.65%-22.43% compared with the CG) in comparison with the IG– may be attributed to more patients with no Apple device experience and more patients aged >50 years in the IG+ compared with the IG–.

Positive responses from patients to self-designed questions 1 to 3 regarding the wearable as a primary input device and its integration into daily routines aligned with the usability and data completeness results, affirming our design decisions. In addition, responses to questions 4 and 5 indicated a strong willingness among patients to share patient-generated data, acknowledging the benefits of an RPM system with feedback mechanisms. This reflects a positive attitude from patients toward using technology for chronic disease management, which is consistent with the evaluations by Turner et al [75] and Rising et al [5].

While sensors for cardiovascular monitoring and smartphone apps for PROM recording are advanced, RPM systems are not commonly used in cardiology research or standardized clinical workflows, highlighting the need for more guidance on the design and implementation of such systems. In a systematic review, Kinast et al [76] found a limited number of studies (between January 1, 2001, and March 31, 2021) implementing wearables and noninvasive sensors, as in the activeDCM study, for the management of cardiovascular diseases. In contrast, activeDCM featured continuous recording and transmission of patient-generated data for at least 12 months—more than twice the duration of most studies in the review. This extended time frame offers initial evidence supporting the feasibility of an RPM system for the longitudinal monitoring of patients with chronic diseases. A systematic review by Vegesna et al [11] examining RPM systems for various other chronic diseases, including cardiac, respiratory, neurological, and metabolic conditions, revealed similar distinctions. Only a few studies for each disease type could be identified, shorter monitoring time frames were mainly applied, and the use of wearables as primary device was rather modest (computerized systems, smartphones, websites, biosensors, or a combination of these were often used). Finally, many studies in both aforementioned reviews predominantly emphasized clinical outcomes, with limited attention given to providing in-depth insights into the technical design and implementation of the RPM system as well as its usability, which in turn impacts patient adherence and

completeness of patient-generated data—all factors influencing the sustained long-term success of RPM systems [77].

Limitations

This study has several potential limitations. First, the evaluated cohort consisted of twice as many men than women, which is typical in cardiovascular trials and especially in DCM trials as men are more often affected than women and the activeDCM study design enrolled independently of gender selection. The RPM system can be used in its current implementation state only with iOS devices and requires a Wi-Fi or cellular data connection of the smartphone for data transfers.

Furthermore, the RPM system was only used and evaluated in a research context. We hypothesize that the experience and evaluation results obtained from this research can facilitate the transition of the RPM system into primary care. We anticipate that this transition will maintain high levels of interaction frequency and completeness of patient-generated data, as observed in the research setting. Nevertheless, additional investigation is required to ensure the successful integration and adoption of the RPM system into standardized clinical workflows. Potential challenges, especially in managing a considerably larger number of patients and their devices, may necessitate adjustments to current physician-patient interaction workflows as established for the activeDCM study. Who would bear the costs for the RPM-based treatment approach would also have to be clarified. One possibility for a reimbursement scheme to explore in Germany would be to certify the RPM system as a digital health application [78,79]. As a result, the wearable and smartphone app and the monitoring process could be prescribed by physicians and, thus, paid for by health insurance companies. In addition, cardiologists would need training in interpreting patient-generated data. The calculation of scores and the implementation of clinical decision support tools would streamline the analysis of patient-generated data, thereby saving time and increasing the scalability of RPM-based clinical workflows.

The RPM system module to answer PROM questionnaires has limited response options, using *yes* or *no*, or a scale from 1 to 4. Future efforts should explore methods for adapting more complex PROM questionnaires with diverse response options for use on wearables, maintaining their role as primary input devices. An example could be the Kansas City Cardiomyopathy Questionnaire–Short Version [80], an important questionnaire in cardiology that may require adaptation for wearable use.

Finally, the RPM system's reliance on manually triggered data transmission by patients, coupled with the consequent time delay in analysis by study physicians, makes it unsuitable for monitoring acute medical emergencies such as myocardial infarctions. To enable real-time monitoring, algorithms would need to be developed for direct analysis of patient-generated data on wearables. Nevertheless, the observational data collected as part of the activeDCM study could serve as a foundation for the future development of such algorithms.

Conclusions

As life expectancy increases and the older adult population experiences more chronic diseases, there is a growing need for ready-to-use RPM systems. In addition, physicians and researchers require more guidance on designing and implementing such systems for both primary care settings and research purposes [77]. This paper presents a blueprint of an RPM system concept, its subsequent implementation, and its procedure for use as part of the activeDCM study. The selected patient-centric approach and the evaluation results show the feasibility of creating a user-friendly, reusable, interoperable, and secure RPM system including an asynchronous feedback mechanism using a consumer wearable without a designated medical purpose as the primary input device for patient-generated data such as SHD and PROMs. The main factors influencing usability and patient acceptance of the RPM system were device experience, age, and intervention group membership. Our research provides a first step toward integrating RPM systems into standardized clinical workflows and could help other researchers in tailoring RPM systems for their studies. Further research is needed to translate such systems into primary care settings. If this translation is successful, the RPM systems have the potential to change the traditional patient-physician interaction in the future [81].

Acknowledgments

This project was kindly supported by the German Cardiac Society eCardiology program, the German Heart Foundation, and Informatics for Life (Klaus Tschira Foundation). The authors acknowledge Apple Inc for generously providing Apple Watch devices as part of a research agreement with the Institute for Cardiomyopathies Heidelberg, Heidelberg University Hospital, Heidelberg, Germany. The authors extend their sincere gratitude to the study nurses and team assistants for their support in using the remote patient monitoring system and providing technical assistance to their patients, with special recognition to Anja Rippert.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Paper-based evaluation questionnaire in German.
[DOCX File , 86 KB - mhealth_v12i1e58441_app1.docx]

Multimedia Appendix 2

Screencast showing the functionality of the data viewer with the German user interface.
[MP4 File (MP4 Video), 29427 KB - mhealth_v12i1e58441_app2.mp4]

Multimedia Appendix 3
Health Level 7 Fast Healthcare Interoperability Resources profiles and concrete resource examples of the data model.
[ZIP File (Zip Archive), 180822 KB - mhealth_v12i1e58441_app3.zip]

Multimedia Appendix 4
CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) checklist (v1.6.1).
[PDF File (Adobe PDF File), 1108 KB - mhealth_v12i1e58441_app4.pdf]

References

1. Izmailova ES, Wagner JA, Perakslis ED. Wearable devices in clinical trials: hype and hypothesis. Clin Pharmacol Ther 2018 Jul 02;104(1):42-52 [FREE Full text] [doi: 10.1002/cpt.966] [Medline: 29205294]
2. Ringeval M, Wagner G, Denford J, Paré G, Kitsiou S. Fitbit-based interventions for healthy lifestyle outcomes: systematic review and meta-analysis. J Med Internet Res 2020 Oct 12;22(10):e23954 [FREE Full text] [doi: 10.2196/23954] [Medline: 33044175]
3. Hershman SG, Bot BM, Shcherbina A, Doerr M, Moayedi Y, Pavlovic A, et al. Physical activity, sleep and cardiovascular health data for 50,000 individuals from the MyHeart counts study. Sci Data 2019 Apr 11;6(1):24 [FREE Full text] [doi: 10.1038/s41597-019-0016-7] [Medline: 30975992]
4. Riffenburg KM, Spartano NL. Physical activity and weight maintenance: the utility of wearable devices and mobile health technology in research and clinical settings. Curr Opin Endocrinol Diabetes Obes 2018 Oct;25(5):310-314. [doi: 10.1097/MED.0000000000000433] [Medline: 30063553]
5. Rising CJ, Gaysynsky A, Blake KD, Jensen RE, Oh A. Willingness to share data from wearable health and activity trackers: analysis of the 2019 health information national trends survey data. JMIR Mhealth Uhealth 2021 Dec 13;9(12):e29190 [FREE Full text] [doi: 10.2196/29190] [Medline: 34898448]
6. Sharon T. Self-tracking for health and the quantified self: re-articulating autonomy, solidarity, and authenticity in an age of personalized healthcare. Philos Technol 2016 Apr 18;30(1):93-121. [doi: 10.1007/S13347-016-0215-5]
7. Lee J, Kim D, Ryoo HY, Shin BS. Sustainable wearables: wearable technology for enhancing the quality of human life. Sustainability 2016 May 11;8(5):466. [doi: 10.3390/SU8050466]
8. Phillips SM, Cadmus-Bertram L, Rosenberg D, Buman MP, Lynch BM. Wearable technology and physical activity in chronic disease: opportunities and challenges. Am J Prev Med 2018 Jan;54(1):144-150. [doi: 10.1016/j.amepre.2017.08.015] [Medline: 29122356]
9. Steinhubl SR, Muse ED, Topol EJ. The emerging field of mobile health. Sci Transl Med 2015 Apr 15;7(283):283rv3 [FREE Full text] [doi: 10.1126/scitranslmed.aaa3487] [Medline: 25877894]
10. Amin T, Mobbs RJ, Mostafa N, Sy LW, Choy WJ. Wearable devices for patient monitoring in the early postoperative period: a literature review. Mhealth 2021;7:50 [FREE Full text] [doi: 10.21037/mhealth-20-131] [Medline: 34345627]
11. Vegesna A, Tran M, Angelaccio M, Arcona S. Remote patient monitoring via non-invasive digital technologies: a systematic review. Telemed J E Health 2017 Jan;23(1):3-17 [FREE Full text] [doi: 10.1089/tmj.2016.0051] [Medline: 27116181]
12. Atreja A, Otobo E, Ramireddy K, Deorocki A. Remote patient monitoring in IBD: current state and future directions. Curr Gastroenterol Rep 2018 Mar 07;20(2):6. [doi: 10.1007/s11894-018-0611-3] [Medline: 29516186]
13. Vercruyssen A, Schirmer W, Geerts N, Mortelmans D. How “basic” is basic digital literacy for older adults? Insights from digital skills instructors. Front Educ 2023 Sep 13;8:5. [doi: 10.3389/feduc.2023.1231701]
14. Smuck M, Odonkor CA, Wilt JK, Schmidt N, Swiernik MA. The emerging clinical role of wearables: factors for successful implementation in healthcare. NPJ Digit Med 2021 Mar 10;4(1):45 [FREE Full text] [doi: 10.1038/s41746-021-00418-3] [Medline: 33692479]
15. Dinh-Le C, Chuang R, Chokshi S, Mann D. Wearable health technology and electronic health record integration: scoping review and future directions. JMIR Mhealth Uhealth 2019 Sep 11;7(9):e12861 [FREE Full text] [doi: 10.2196/12861] [Medline: 31512582]
16. Dehling T, Gao F, Schneider S, Sunyaev A. Exploring the far side of mobile health: information security and privacy of mobile health apps on iOS and Android. JMIR Mhealth Uhealth 2015 Jan 19;3(1):e8 [FREE Full text] [doi: 10.2196/mhealth.3672] [Medline: 25599627]
17. Wiesner M, Zowalla R, Suleder J, Westers M, Pobiruchin M. Technology adoption, motivational aspects, and privacy concerns of wearables in the German Running Community: field study. JMIR Mhealth Uhealth 2018 Dec 14;6(12):e201 [FREE Full text] [doi: 10.2196/mhealth.9623] [Medline: 30552085]
18. Perez MV, Mahaffey KW, Hedlin H, Rumsfeld JS, Garcia A, Ferris T, et al. Large-scale assessment of a smartwatch to identify atrial fibrillation. N Engl J Med 2019 Nov 14;381(20):1909-1917. [doi: 10.1056/nejmoa1901183]

19. Lubitz SA, Faranesh AZ, Atlas SJ, McManus DD, Singer DE, Pagoto S, et al. Rationale and design of a large population study to validate software for the assessment of atrial fibrillation from data acquired by a consumer tracker or smartwatch: the Fitbit heart study. Am Heart J 2021 Aug;238:16-26. [doi: 10.1016/j.ahj.2021.04.003] [Medline: 33865810]
20. Veltmann C, Ehrlich JR, Gassner UM, Meder B, Möckel M, Radke P, et al. Wearable-basierte Detektion von Arrhythmien. Kardiologe 2021 Jul 07;15(4):341-353. [doi: 10.1007/s12181-021-00488-3]
21. Meder B, Duncker D, Helms TM, Leistner DM, Goss F, Perings C, et al. eCardiology: ein strukturierter Ansatz zur Förderung der digitalen Transformation in der Kardiologie. Kardiologie 2023 Jan 16;17(1):12-26. [doi: 10.1007/S12181-022-00584-Y]
22. Groenewegen A, Rutten FH, Mosterd A, Hoes AW. Epidemiology of heart failure. Eur J Heart Fail 2020 Aug;22(8):1342-1356 [FREE Full text] [doi: 10.1002/ejhf.1858] [Medline: 32483830]
23. Gheorghiade M, Vaduganathan M, Fonarow GC, Bonow RO. Rehospitalization for heart failure: problems and perspectives. J Am Coll Cardiol 2013 Jan 29;61(4):391-403 [FREE Full text] [doi: 10.1016/j.jacc.2012.09.038] [Medline: 23219302]
24. Haas J, Frese KS, Peil B, Kloos W, Keller A, Nietsch R, et al. Atlas of the clinical genetics of human dilated cardiomyopathy. Eur Heart J 2015 May 07;36(18):1123-135a. [doi: 10.1093/eurheartj/ehu301] [Medline: 25163546]
25. Alpert C, Day SM, Saberi S. Sports and exercise in athletes with hypertrophic cardiomyopathy. Clin Sports Med 2015 Jul;34(3):489-505. [doi: 10.1016/j.csm.2015.03.005] [Medline: 26100424]
26. Chung CJ, Schulze PC. Exercise as a nonpharmacologic intervention in patients with heart failure. Phys Sportsmed 2011 Nov 13;39(4):37-43 [FREE Full text] [doi: 10.3810/psm.2011.11.1937] [Medline: 22293766]
27. Arbelo E, Protonotarios A, Gimeno JR, Arbustini E, Barriales-Villa R, Basso C, ESC Scientific Document Group. 2023 ESC guidelines for the management of cardiomyopathies. Eur Heart J 2023 Oct 01;44(37):3503-3626. [doi: 10.1093/eurheartj/ehad194] [Medline: 37622657]
28. Bock JO, Hajek A, König HH. The Longitudinal Association between psychological factors and health care use. Health Serv Res 2018 Apr 15;53(2):1065-1091 [FREE Full text] [doi: 10.1111/1475-6773.12679] [Medline: 28294328]
29. Sedaghat-Hamedani F, Amr A, Betz T, Kayvanpour E, Reich C, Wettstein R, et al. Personalized care in dilated cardiomyopathy: rationale and study design of the activeDCM trial. ESC Heart Fail (Forthcoming) 2024 Jul 15 [FREE Full text] [doi: 10.1002/ehf2.14970] [Medline: 39010317]
30. Kitzinger J. Qualitative research. Introducing focus groups. BMJ 1995 Jul 29;311(7000):299-302 [FREE Full text] [doi: 10.1136/bmj.311.7000.299] [Medline: 7633241]
31. The Swift programming language - documentation. Swift. URL: https://docs.swift.org/swift-book/documentation/the-swift-programming-language/ [accessed 2024-07-08]
32. Core motion - Apple developer documentation. Apple Inc. URL: https://developer.apple.com/documentation/coremotion [accessed 2024-07-08]
33. HealthKit - Apple developer documentation. Apple Inc. URL: https://developer.apple.com/documentation/healthkit [accessed 2024-07-08]
34. UIKit - Apple developer documentation. Apple Inc. URL: https://developer.apple.com/documentation/uikit [accessed 2024-07-08]
35. Watch connectivity - Apple developer documentation. Apple Inc. URL: https://developer.apple.com/documentation/watchconnectivity [accessed 2024-07-08]
36. Welcome to the Atlas Device SDK docs. MongoDB. URL: https://realm.io/ [accessed 2024-07-08]
37. FHIR v 4.0.1. Health Level Seven. URL: https://www.hl7.org/fhir/R4 [accessed 2024-07-08]
38. Forrey AW, McDonald CJ, DeMoor G, Huff SM, Leavelle D, Leland D, et al. Logical observation identifier names and codes (LOINC) database: a public use set of codes and names for electronic reporting of clinical laboratory test results. Clin Chem 1996 Jan;42(1):81-90. [Medline: 8565239]
39. Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. Nucleic Acids Res 2004 Jan 1;32(Database issue):D267-D270 [FREE Full text] [doi: 10.1093/nar/gkh061] [Medline: 14681409]
40. JDK 11 documentation. Oracle. URL: https://docs.oracle.com/en/java/javase/11 [accessed 2024-07-08]
41. HAPI FHIR - the open source FHIR API for Java. Smile CDR. URL: https://hapifhir.io [accessed 2024-07-08]
42. PostgreSQL: the world's most advanced open source database. The PostgreSQL Global Development. URL: https://www.postgresql.org/ [accessed 2024-07-08]
43. Bierman G, Abadi M, Torgersen M. Understanding TypeScript. In: Proceedings of the 28th European Conference on Object-Oriented Programming. 2014 Presented at: ECOOP '14; July 28-August 1, 2014; Uppsala, Sweden p. 257-281 URL: https://link.springer.com/chapter/10.1007/978-3-662-44202-9_11 [doi: 10.1007/978-3-662-44202-9_11]
44. The progressive JavaScript framework. Vue.js. URL: https://vuejs.org/guide/introduction.html [accessed 2024-07-08]
45. Home page. Chart.js. URL: https://www.chartjs.org/docs/latest/ [accessed 2024-07-08]
46. User notifications - Apple developer documentation. Apple Inc. URL: https://developer.apple.com/documentation/usernotifications [accessed 2024-07-08]
47. Meder B. ActiveDCM wearable app. Institut für Cardiomyopathien Heidelberg, Innere Medizin III, Universitätsklinikum Heidelberg. URL: https://medical-data-models.org/45696 [accessed 2024-07-08]
48. System usability scale. SAP Community. 2016. URL: https://blogs.sap.com/2016/02/01/system-usability-scale-jetzt-auch-auf-deutsch/ [accessed 2024-07-08]

49. Brooke J. SUS: a 'quick and dirty' usability scale. In: Jordan PW, Thomas B, McClelland IL, Weerdmeester B, editors. Usability Evaluation in Industry. Boca Raton, FL: Taylor & Francis Group; 1996:189-194.
50. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: adding an adjective rating scale. J Usability Stud 2009;4(3):114-123 [FREE Full text]
51. Kruskal WH, Wallis WA. Use of ranks in one-criterion variance analysis. J Am Stat Assoc 1952 Dec;47(260):583-621. [doi: 10.1080/01621459.1952.10483441]
52. Pearson K. X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. London Edinburgh Philos Mag J Sci 2009 Apr 21;50(302):157-175 [FREE Full text] [doi: 10.1080/14786440009463897]
53. Efron B. Better bootstrap confidence intervals. J Am Stat Asso 1987 Mar;82(397):171-185. [doi: 10.1080/01621459.1987.10478410]
54. Python language reference, version 3.10. Python Software Foundation. URL: https://www.python.org [accessed 2024-07-08]
55. scipy.stats.kruskal, version 1.11.1. 2023. The SciPy community. URL: https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.kruskal.html [accessed 2024-07-08]
56. scipy.stats.chi2_contingency, version 1.11.1. The SciPy community. URL: https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.chi2_contingency.html [accessed 2024-07-08]
57. scipy.stats.bootstrap, version 1.11.1. The SciPy community. URL: https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.bootstrap.html [accessed 2024-07-08]
58. Perktold J, Seabold S, Taylor J. statsmodels.regression.linear_model.OLS, version 0.14.0. statsmodels-developers. URL: https://www.statsmodels.org/dev/generated/statsmodels.regression.linear_model.OLS.html [accessed 2024-07-08]
59. Dierks T, Rescorla E. The Transport Layer Security (TLS) protocol version 1.2. Internet Engineering Task Force (IETF). 2008. URL: https://www.rfc-editor.org/rfc/pdfrfc/rfc5246.txt.pdf [accessed 2024-04-29]
60. Keychain services - Apple developer documentation. Apple Inc. URL: https://developer.apple.com/documentation/security/keychain_services [accessed 2024-07-08]
61. Jones M, Bradley J, Sakimura N. JSON Web Token (JWT). Internet Engineering Task Force (IETF). 2015. URL: https://www.rfc-editor.org/rfc/pdfrfc/rfc7519.txt.pdf [accessed 2024-04-29]
62. Cooper D, Santesson S, Farrell S, Boeyen S, Housley R, Polk W. Internet X.509 public key infrastructure certificate and certificate revocation list (CRL) profile. Internet Engineering Task Force (IETF). 2008. URL: https://datatracker.ietf.org/doc/html/rfc5280 [accessed 2024-04-29]
63. Designing for iOS. Apple Developer. URL: https://developer.apple.com/design/human-interface-guidelines/designing-for-ios [accessed 2024-07-08]
64. Designing for watchOS. Apple Developer. URL: https://developer.apple.com/design/human-interface-guidelines/designing-for-watchos [accessed 2024-07-08]
65. Monitor your heart rate with Apple Watch - Apple Support. Apple Inc. URL: https://support.apple.com/en-us/HT204666# [accessed 2024-07-08]
66. Hickey KT, Hauser NR, Valente LE, Riga TC, Frulla AP, Masterson Creber R, et al. A single-center randomized, controlled trial investigating the efficacy of a mHealth ECG technology intervention to improve the detection of atrial fibrillation: the iHEART study protocol. BMC Cardiovasc Disord 2016 Jul 16;16(1):152 [FREE Full text] [doi: 10.1186/s12872-016-0327-y] [Medline: 27422639]
67. Ong MK, Romano PS, Edgington S, Aronow HU, Auerbach AD, Black JT, Better Effectiveness After Transition–Heart Failure (BEAT-HF) Research Group. Effectiveness of remote patient monitoring after discharge of hospitalized patients with heart failure: the better effectiveness after transition -- heart failure (BEAT-HF) randomized clinical trial. JAMA Intern Med 2016 Mar;176(3):310-318 [FREE Full text] [doi: 10.1001/jamainternmed.2015.7712] [Medline: 26857383]
68. Ayaz M, Pasha MF, Alzahrani MY, Budiarto R, Stiawan D. The Fast Health Interoperability Resources (FHIR) standard: systematic literature review of implementations, applications, challenges and opportunities. JMIR Med Inform 2021 Jul 30;9(7):e21929 [FREE Full text] [doi: 10.2196/21929] [Medline: 34328424]
69. ISiK (Informationstechnische Systeme in Krankenhäusern). Gematik. URL: https://www.gematik.de/anwendungen/isik [accessed 2024-07-08]
70. Fuller D, Colwell E, Low J, Orychock K, Tobin MA, Simango B, et al. Reliability and validity of commercially available wearable devices for measuring steps, energy expenditure, and heart rate: systematic review. JMIR Mhealth Uhealth 2020 Sep 08;8(9):e18694 [FREE Full text] [doi: 10.2196/18694] [Medline: 32897239]
71. Case MA, Burwick HA, Volpp KG, Patel MS. Accuracy of smartphone applications and wearable devices for tracking physical activity data. JAMA 2015 Feb 10;313(6):625-626. [doi: 10.1001/jama.2014.17841] [Medline: 25668268]
72. Sonderegger A, Schmutz S, Sauer J. The influence of age in usability testing. Appl Ergon 2016 Jan;52:291-300. [doi: 10.1016/j.apergo.2015.06.012] [Medline: 26360221]
73. Werhahn SM, Dathe H, Rottmann T, Franke T, Vahdat D, Hasenfuß G, et al. Designing meaningful outcome parameters using mobile technology: a new mobile application for telemonitoring of patients with heart failure. ESC Heart Fail 2019 Jun 13;6(3):516-525 [FREE Full text] [doi: 10.1002/ehf2.12425] [Medline: 30868756]

74. Adair JG. The Hawthorne effect: a reconsideration of the methodological artifact. J Appl Psychol 1984 May;69(2):334-345. [doi: 10.1037/0021-9010.69.2.334]
75. Turner K, Jo A, Wei G, Tabriz AA, Clary A, Jim HS. Sharing patient-generated data with healthcare providers: findings from a 2019 national survey. J Am Med Inform Assoc 2021 Mar 15;28(2):371-376. [doi: 10.1093/jamia/ocaa272] [Medline: 33180896]
76. Kinast B, Lutz M, Schreiweis B. Telemonitoring of real-world health data in cardiology: a systematic review. Int J Environ Res Public Health 2021 Aug 27;18(17):9070 [FREE Full text] [doi: 10.3390/ijerph18179070] [Medline: 34501659]
77. Thomas EE, Taylor ML, Banbury A, Snoswell CL, Haydon HM, Gallegos Rejas VM, et al. Factors influencing the effectiveness of remote patient monitoring interventions: a realist review. BMJ Open 2021 Aug 25;11(8):e051844 [FREE Full text] [doi: 10.1136/bmjopen-2021-051844] [Medline: 34433611]
78. Das Fast-Track-Verfahren für digitale Gesundheitsanwendungen (DiGA) nach § 139e SGB V. Bundesinstitut für Arzneimittel und Medizinprodukte. 2020. URL: https://www.bfarm.de/SharedDocs/Downloads/DE/Medizinprodukte/diga_leitfaden.pdf [accessed 2024-07-08]
79. DiGa-Verzeichnis. Bundesinstitut für Arzneimittel und Medizinprodukte. 2021. URL: https://diga.bfarm.de/de [accessed 2024-07-08]
80. Spertus JA, Jones PG. Development and validation of a short version of the Kansas City cardiomyopathy questionnaire. Circ Cardiovasc Qual Outcomes 2015 Sep;8(5):469-476. [doi: 10.1161/circoutcomes.115.001958]
81. Williams GJ, Al-Baraikan A, Rademakers FE, Ciravegna F, van de Vosse FN, Lawrie A, et al. Wearable technology and the cardiovascular system: the future of patient assessment. Lancet Digit Health 2023 Jul;5(7):e467-e476 [FREE Full text] [doi: 10.1016/S2589-7500(23)00087-0] [Medline: 37391266]

Abbreviations

API: application programming interface
APNs: Apple Push Notification service
CG: control group
DCM: dilated cardiomyopathy
ECG: electrocardiogram
Exp–: without previous device experience
Exp+: with previous device experience
FHIR: Fast Healthcare Interoperability Resources
HL7: Health Level 7
IG–: intervention group without feedback messages
IG+: intervention group with feedback messages
JWT: JSON Web Token
LOINC: Logical Observation Identifiers Names and Codes
PROM: patient-reported outcome measure
RPM: remote patient monitoring
SHD: sensor-based health data
SUS: System Usability Scale
UMLS: Unified Medical Language System

Edited by L Buis; submitted 15.03.24; peer-reviewed by R Pryss, SH Park; comments to author 02.07.24; revised version received 22.07.24; accepted 13.09.24; published 22.11.24.

Please cite as:
Wettstein R, Sedaghat-Hamedani F, Heinze O, Amr A, Reich C, Betz T, Kayvanpour E, Merzweiler A, Büsch C, Mohr I, Friedmann-Bette B, Frey N, Dugas M, Meder B
A Remote Patient Monitoring System With Feedback Mechanisms Using a Smartwatch: Concept, Implementation, and Evaluation Based on the activeDCM Randomized Controlled Trial
JMIR Mhealth Uhealth 2024;12:e58441
URL: https://mhealth.jmir.org/2024/1/e58441
doi:10.2196/58441
PMID:

The Impact of User Engagement With Exposure Components on Posttraumatic Stress Symptoms in an mHealth Mobile App: Secondary Analysis of a Randomized Controlled Trial

C Adrian Davis[1], MA; Madeleine Miller[2], BA; Carmen P McLean[1,3], PhD

[1]National Center for Posttraumatic Stress Disorder, Dissemination and Training Division, VA Palo Alto Healthcare System, Menlo Park, CA, United States

[2]Health Psychology and Clinical Science, The Graduate Center, The City University of New York, New York, NY, United States

[3]Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, CA, United States

Corresponding Author:
C Adrian Davis, MA
National Center for Posttraumatic Stress Disorder, Dissemination and Training Division, VA Palo Alto Healthcare System, , Menlo Park, CA, , United States

Abstract

Background: Mobile mental health apps are a cost-effective option for managing mental health problems, such as posttraumatic stress disorder (PTSD). The efficacy of mobile health (mHealth) apps depends on engagement with the app, but few studies have examined how users engage with different features of mHealth apps for PTSD.

Objective: This study aims to examine the relationship between app engagement indices and PTSD symptom reduction using data from an unblinded pilot randomized controlled trial of "Renew" (Vertical Design), an exposure-based app for PTSD with and without coaching support. Because exposure is an effective approach for treating PTSD, we expected that engagement with exposure activities would be positively related to symptom reduction, over and above overall app usage.

Methods: Participants were veterans (N=69) with clinically significant PTSD symptoms who were recruited online using Facebook advertisements and invited to use the Renew app as often as they wanted over a 6-week period. Participants completed screening and assessments online but provided informed consent, toured the app, and completed feedback interviews via telephone. We assessed users' self-reported PTSD symptoms before and after a 6-week intervention period and collected app usage data using a research-instrumented dashboard. To examine overall app engagement, we used data on the total time spent in the app, the number of log-in days, and the number of points that the user gained in the app. To examine engagement with exposure components, we used data on total time spent completing exposure activities (both in vivo and imaginal), the number of in vivo exposure activities completed, and the number of characters written in response to imaginal exposure prompts. We used hierarchical regression analyses to test the effect of engagement indices on change in PTSD symptoms.

Results: Usage varied widely. Participants spent an average of 166.09 (SD 156.52) minutes using Renew, over an average of 14.7 (SD 10.71) mean log-in days. Engagement with the exposure components of the app was positively associated with PTSD symptom reduction ($F_{6,62}$=2.31; P=.04). Moreover, this relationship remained significant when controlling for overall engagement with the app ($\Delta F_{3,62}$=4.42; P=.007). The number of characters written during imaginal exposure (β=.37; P=.009) and the amount of time spent completing exposure activities (β=.36; P=.03) were significant contributors to the model.

Conclusions: To our knowledge, this is the first study to show a relationship between symptom improvement and engagement with the active therapeutic components of an mHealth app (ie, exposure) for PTSD. This relationship held when controlling for overall app use, which suggests that it was engagement with exposure, specifically, that was associated with symptom change. Future work to identify ways of promoting greater engagement with self-guided exposure may help improve the effectiveness of mHealth apps for PTSD.

Trial Registration: ClinicalTrials.gov NCT04155736; https://clinicaltrials.gov/ct2/show/NCT04155736

(JMIR Mhealth Uhealth 2024;12:e49393) doi:10.2196/49393

KEYWORDS

posttraumatic stress disorder; PTSD; mHealth apps; user engagement; mHealth interventions; digital interventions; veterans

Introduction

Overview

Mobile mental health apps have continued to proliferate in recent years [1]. As of last year, more than 80 mobile mental health (mHealth) apps aimed at managing posttraumatic stress disorder (PTSD)—a significant public health concern with an estimated 7% lifetime prevalence in the United States of America—are now available for download [2,3]. Research on mHealth apps for PTSD has also increased and suggests that these tools are generally feasible and acceptable to users [4]. However, a recent review found that evidence for the efficacy of PTSD mHealth apps was not consistently superior to inactive control conditions [5].

The efficacy of an mHealth app for PTSD will depend upon how the user engages with the tool. Only one study has examined the relationship between engagement with a PTSD self-management app and symptom change. This study found that the number of log-ins to the mHealth app PTSD Coach (Vertical Design) was not related to PTSD symptom reduction [6]. This contrasts with findings from depression mHealth apps showing that number of log-ins [7] and number of activities per log-in [8] are linked with change in depressive symptoms. Additional research on the relationship between app engagement and PTSD symptom change is clearly needed and could inform user guidance on how to benefit from using PTSD mHealth self-management apps.

Objectives

The current analysis aims to address this gap by exploring the relationship between user engagement indices and symptom reduction in users of a PTSD mHealth self-management app called "Renew" [9,10]. Based on principles of exposure therapy, Renew uses gamified elements and activities designed to help users target their PTSD symptoms using in vivo and written exposure. Given the evidence supporting exposure as an efficacious therapeutic approach for treating PTSD [11], we hypothesized that engagement with the exposure activities in Renew would be uniquely related to symptom reduction, over and above general app use.

Methods

Participants

Potential participants were recruited through Facebook (Meta) advertisements because we were interested in examining a nationally representative sample of veterans. As of the Veterans Health Administration's last publicly released report, more than half of veterans do not receive care through the Department of Veterans Affairs (VA) [12], so a sample recruited strictly from among VA patients would not necessarily represent veterans as a whole. Potential participants completed an online screening measure through REDCap (Research Electronic Data Capture; Vanderbilt University) [13] to assess veteran status, trauma exposure [14], and PTSD symptoms. Eligible participants were veterans with clinically significant PTSD symptoms, defined by a score of ≥31 on the Posttraumatic Stress Disorder Checklist for *DSM-5* (PCL-5) [14]; access to an Android phone (Renew was only developed for Android OS); and sufficient smartphone literacy and internet connection to successfully download the app with telephone guidance from study staff. There were no exclusion criteria.

Ethical Considerations

All participants gave their informed consent for inclusion before they participated in this trial. The protocol was approved by the Stanford University Institutional Review Board (52829).

Renew

Renew is a research-instrumented mobile app designed to reduce symptoms of PTSD using in vivo exposure and imaginal exposure, developed by Vertical Design. Other components include psychoeducation, behavioral activation through self-care, and assessment. Users gain points for completing activities in the app. Renew was designed to be used both with and without human coaching support, provided through a one-way messaging function. Users downloaded Renew from the Google Play Store and accessed the intervention using an invitation code provided by the study team during app orientation. Readers of this article who possess an Android phone may email the corresponding author to receive a Renew invitation code if they wish to access the intervention. Before this trial, we tested the feasibility and acceptability of Renew in a small pilot study, after which the content of Renew was updated according to user feedback and frozen throughout the duration of the current trial. Miller et al [10] and McLean et al [9] give more details about Renew, including the details of the feasibility and acceptability pilot study.

Measures

Renew Usage Data

Participant usage data were collected through a secure research dashboard. User access to the app was restricted to only those users who were provided with an individualized code. Once linked with a given user's app, this code allowed researchers to track the user's engagement with the Renew app through the secure research dashboard, including time-stamped button presses, time spent on each screen, number of characters typed into free text boxes, log-ins, and dates of use.

Consistent with previous mHealth app studies that have used multiple objective engagement criteria [15], we examined several indices of overall app engagement: total time spent in Renew, number of log-in days, and number of points gained in Renew. To specifically capture engagement with exposure, we examined time spent completing exposure activities (both imaginal and in vivo), the number of in vivo exposure activities completed, and the number of characters written during imaginal exposure. Number of characters was chosen as a proxy for the degree to which users engaged with their worst traumatic memory.

Posttraumatic Stress Symptoms

The PCL-5 [15] is a 20-item self-report measure of PTSD symptoms as defined by the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* (*DSM-5*) [16]. The PCL-5 has demonstrated strong psychometric properties [17,18] and has been validated for online use [19]. The Cronbach α in the

current sample was 0.84. The PCL-5 was administered to each participant at 3 time points. For participants immediately assigned to one of the active use conditions, the PCL-5 was completed through REDCap before app use as part of the study screening procedures, after the 6-week active use period, and after a 6-week follow-up period. Participants assigned to the delayed use condition completed the PCL-5 before the delayed use period, at the end of the delayed use period before receiving a code to access Renew, and after their 6-week active use period.

Procedures

The Stanford University Institutional Review Board approved the study. Informed consent was collected via telephone call, in which the study staff read a script that included information about Renew and the 3 study conditions to which they could be randomized. Study staff then allowed participants to ask questions about the study before providing consent. Eligible participants who consented to the study were randomly assigned to either 6 weeks of (1) Renew without study staff support, (2) Renew with a study support person as a support team member, or (3) delayed use. Neither participants nor study staff were blinded to conditions. Participants in the delayed use group were invited to use the app for 6 weeks after a 6-week wait period. Immediately before beginning the 6-week active use period, participants received a 30-minute app orientation by phone. During the orientation, a study staff member walked participants through each section of the app, described the features of Renew, and offered participants the chance to ask questions. Following the orientation, participants were then asked to use the app as often as they wanted for the next 6 weeks. Participants could choose to set reminders to use the app. Participants completed postuse measures through the REDCap online survey platform [13] at the end of the app-use period. Usage data were collected using a secure research dashboard. All participants who were assigned to the delayed use condition were given an access code for Renew at the end of the 6-week delayed use period. Data about their app usage was collected during a 6-week follow-up period, in addition to pre- and postuse symptom measures. For the current analysis, data were pooled from participants assigned to all 3 conditions, because there were no significant (P=.31) differences in PTSD symptom change during the active use periods across groups [9]. The majority of data collection took place online or through the online research dashboard that recorded Renew usage data, however, participants did complete postuse feedback interviews with study staff via telephone. Of the 93 participants who enrolled in the parent pilot trial [9], 69 participants had completed both pre- and postuse symptom measures and were included in this study.

Data Analysis

Analyses were conducted using SPSS (version 26; IBM Corp) [20]. Descriptive and correlational analyses were conducted for all engagement variables. A hierarchical linear regression analysis was conducted with the overall app engagement variables entered in step 1 and exposure engagement variables entered in step 2. The dependent variable was change on the PCL-5. All app use data in the analysis came from the participants' active use period (ie, for participants initially assigned to delayed use, usage data was taken from the active use period that followed the delayed use period).

Results

Participant demographics are reported in Table 1. Participants were predominantly middle-aged, women, and White. More than half of participants in this subsample were army veterans (n=36, 52.17%), and the average PCL-5 score was well above the cutoff for clinically significant PTSD symptoms.

Means and SDs for user engagement variables are presented in Table 2.

Table . Participant demographic characteristics (N=69).

Variable	Value
Age (years), mean (SD)	45.28 (9.05)
Gender [a] , n (%)	
Men	22 (33.33)
Women	47 (66.67)
Race or ethnicity [b] , n (%)	
White	47 (68.12%)
Black or African American	13 (18.84)
Asian or Asian American	2 (2.89)
American Indian or Alaska Native	4 (5.8)
Middle Eastern or North African	1 (14.49)
Other	2 (2.89)
Military branch, n (%)	
Army	36 (52.17)
Navy	13 (18.84)
Air force	15 (21.74)
Marines	5 (7.25)
Baseline PCL-5[c] score, mean (SD)	55.94 (9.87)

[a]No participants identified as transgender, nonbinary, or another gender identity.

[b]Participants could select more than one race or ethnicity. Because of this, percentages sum to >100%.

[c]PCL-5: Posttraumatic Stress Disorder Checklist for *DSM-5*.

Table . Descriptive statistics for app engagement indices.

Overall engagement variables	Value, mean (SD)
Log-in days	14.7 (10.71)
Minutes in Renew	166.09 (156.52)
Number of points gained	350.14 (451.42)
Exposure engagement variables	
Number of in vivo exposure activities completed	1.84 (4.22)
Minutes spent completing exposure activities	78.19 (135.35)
Number of characters written during imaginal exposure activities	5116 (8522.43)

Over the 6-week active use period, participants logged into Renew about one-third of the days. Users spent a little over 2.5 hours total engaging with Renew, not including their initial 30-minute orientation. Users spent over an hour engaging in exposure activities, averaging more than one complete in vivo exposure activity. In response to exposure prompts, participants wrote over 5000 characters on average, which roughly equals 750 - 1000 words.

Correlations between app usage variables are presented in Table 3. Given the overlap in the engagement variables, there was high intercorrelation between the general and exposure-specific usage variables. Relationships between all variables were significant. The variable “minutes spent completing exposure activities” reflects minutes spent practicing both imaginal and in vivo exposure. Time spent on imaginal and in vivo exposure were combined in order to examine the impact of overall engagement with exposure in addition to the individual effects of engagement with imaginal and in vivo alone.

Table . Correlations (Pearson *r*) among usage variables.

Variables	Log-in days	Minutes in Renew	Number of points gained	Number of in vivo exposure activities completed	Minutes spent completing exposure activities	Number of characters written during imaginal exposure
Overall usage variables						
Log-in days						
r	1	0.74	0.81	0.47	0.41	0.45
P value	—[a]	<.001	<.001	<.001	<.001	<.001
Minutes in Renew						
r	0.74	1	0.79	0.64	0.65	0.54
P value	<.001	—	<.001	<.001	<.001	<.001
Number of points gained						
r	0.81	0.79	1	0.58	0.41	0.46
P value	<.001	<.001	—	<.001	<.001	<.001
Exposure engagement variables						
Number of in vivo exposure activities completed						
r	0.47	0.64	0.58	1	0.58	0.33
P value	<.001	<.001	<.001	—	<.001	.01
Minutes spent completing exposure activities						
r	0.41	0.65	0.41	0.58	1	0.38
P value	<.001	<.001	<.001	<.001	—	0
Number of characters written during imaginal exposure						
r	0.45	0.54	0.46	0.33	0.38	1
P value	<.001	<.001	<.001	.01	0	—

[a]Not applicable.

The results of the regression analysis are presented in Table 4. The initial model with overall engagement predicting PTSD symptom change was not significant and accounted for 1% of the variance observed in the model. However, the model with the exposure engagement variables added was significant and accounted for an additional 18% of variance in the model. Time spent completing exposure activities (*P*=.03) and the number of characters written during imaginal exposure (*P*=.009) were significant contributors to the final model.

Table . Hierarchical linear regression model (N=69).

Model steps and variables	β[a]	R	R^2	ΔR^2	*F* test (*df*)	*P* value[b]	Δ*F* test (*df*)	*P* value[c]
Step 1: overall engagement	—[d]	0.1	0.01	—	0.18 (3,65)	0.9	—	—
Log-in days	0.01	—	—	—	—	0.97	—	—
Time in Renew	0.01	—	—	—	—	0.64	—	—
Number of points gained	−0.02	—	—	—	—	0.53	—	—
Step 2: exposure engagement	—	0.44	0.19	0.18	2.31 (6,62)	0.04	4.42 (3,62)	0.01
Number of in vivo exposure activities completed	−0.03	—	—	—	—	0.86	—	—
Time spent completing exposure activities	0.36	—	—	—	—	0.03	—	—
Number of characters written during imaginal exposure	0.37	—	—	—	—	0.01	—	—

[a]β: Standardized beta weight.

[b]Significance for standardized beta weights of predictors and steps of the model (*F* test statistic).

[c]Significance for change in *F* test statistic (Δ*F* test) at second step.

[d]Not applicable.

Discussion

Overview

We examined the relationship between engagement with imaginal and in vivo exposure and PTSD symptom reduction in an mHealth app, Renew. Few randomized controlled trials have examined mobile app-delivered interventions for PTSD [21], and even fewer studies have evaluated the impact of usage factors on symptom change in digital PTSD interventions [22,23]. As hypothesized, we found that engagement with exposure was associated with greater PTSD symptom reduction over and above overall use of the app in general. Our finding that overall app use was unrelated to PTSD symptom change is consistent with previous research on PTSD Coach [6] but inconsistent with meta-analytic findings showing that engagement with mHealth apps in general is associated with symptom change [24]. To our knowledge, this is the first study to demonstrate a relationship between engagement with the active therapeutic components of an mHealth app and improvement in PTSD symptoms.

Our findings are consistent with the limited research on the relationship between app usage and symptom change in self-management apps for depression and anxiety, which suggests that higher rates of engagement with active intervention components are associated with symptom change. Specifically, Zhang et al [25] found that participants' depressive symptom reduction was associated with greater engagement in self-tracking and moderate engagement with learning and goal setting, but anxiety symptom reduction was unrelated to engagement with specific app components. One other study, Nardi et al [26], examined data from a clinical trial of a mindfulness app and also found evidence that engagement with specific components, such as meditation, was linked to improvement in generalized anxiety symptom reduction. To our knowledge, there are no studies of PTSD apps that have examined how engagement with different intervention components relates to symptom change. Even among apps for more common mental health concerns, robust reporting on usage metrics remains scant [27]. Our analysis of exposure-specific use metrics in this study contributes to a profound gap in the literature on digital mental health interventions. Our finding that exposure may be helpful for reducing PTSD symptoms through a self-guided mHealth app is consistent with the broader literature on efficacious PTSD interventions [28]. Future work should examine digital interventions that include active cognitive behavioral therapy (CBT) components such as exposure, which are associated with symptom change. Indeed, in a large meta-analysis of 176 randomized controlled trials, effect sizes were larger for self-management apps for anxiety and depression that included CBT components compared to those that did not [29].

The number of characters entered during imaginal exposure, which we considered a proxy for the degree to which

participants engaged with and recounted their worst traumatic memory, and the time spent in any exposure activity, both significantly contributed to the model predicting PTSD symptom change, whereas the number of exposure activities did not. This may suggest that the depth of engagement and time spent approaching trauma-related memories and situations is important for therapeutic change in the context of mHealth apps for PTSD. This pattern of findings maps to the literature on in-person psychotherapy, where most exposure therapies emphasize the importance of imaginal exposure [30].

Limitations

Several study limitations should be noted. Data for this study were drawn from a pilot trial, and replication of the current findings is needed. Participants were veterans with clinically significant symptoms of PTSD; findings may not generalize to other populations diagnosed with PTSD. Data were cross-sectional, which prevents causal interpretations. In addition, many of the usage variables were highly intercorrelated and there is no established gold standard approach for operationalizing app engagement. These limitations notwithstanding, this study provides preliminary evidence of a unique relationship between engagement with exposure and PTSD symptom change among veterans with clinically significant symptoms of PTSD who used a self-guided mHealth app (Renew) for a 6-week period.

Acknowledgments

The authors would like to thank Vertical Inc and Yuzu Labs Inc for their assistance with app development and participant recruitment, respectively. The views expressed herein are solely those of the authors and do not reflect an endorsement by or the official policy or position of the Department of Veterans Affairs or the US government. This material is the result of work supported with resources and the use of facilities at the VA Palo Alto Health Care System. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflicts of Interest

None declared.

References

1. Livingston NA, Shingleton R, Heilman ME, Brief D. Self-help smartphone applications for alcohol use, PTSD, anxiety, and depression: addressing the new research-practice gap. J Technol Behav Sci 2019 Jun;4(2):139-151. [doi: 10.1007/s41347-019-00099-6]
2. Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. Arch Gen Psychiatry 2005 Jun;62(6):593-602. [doi: 10.1001/archpsyc.62.6.593] [Medline: 15939837]
3. Ting A, McLachlan C. Dr. Smartphone, can you support my trauma? an informatics analysis study of app store apps for trauma- and stressor-related disorders. PeerJ 2023 May;11:e15366. [doi: 10.7717/peerj.15366] [Medline: 37187512]
4. Kulakli A, Shubina I. Scientific publication patterns of mobile technologies and apps for posttraumatic stress disorder treatment: bibliometric co-word analysis. JMIR Mhealth Uhealth 2020 Nov 26;8(11):e19391. [doi: 10.2196/19391] [Medline: 33242019]
5. Wickersham A, Petrides PM, Williamson V, Leightley D. Efficacy of mobile application interventions for the treatment of post-traumatic stress disorder: a systematic review. Digit Health 2019 Apr;5:2055207619842986. [doi: 10.1177/2055207619842986] [Medline: 31019722]
6. Kuhn E, Kanuri N, Hoffman JE, Garvert DW, Ruzek JI, Taylor CB. A randomized controlled trial of a smartphone app for posttraumatic stress disorder symptoms. J Consult Clin Psychol 2017 Mar;85(3):267-273. [doi: 10.1037/ccp0000163] [Medline: 28221061]
7. Van Gemert-Pijnen JE, Kelders SM, Bohlmeijer ET. Understanding the usage of content in a mental health intervention for depression: an analysis of log data. J Med Internet Res 2014 Jan 31;16(1):e27. [doi: 10.2196/jmir.2991] [Medline: 24486914]
8. Donkin L, Hickie IB, Christensen H, et al. Rethinking the dose-response relationship between usage and outcome in an online intervention for depression: randomized controlled trial. J Med Internet Res 2013 Oct 17;15(10):e231. [doi: 10.2196/jmir.2771] [Medline: 24135213]
9. McLean C, Davis CA, Miller M, Ruzek J, Neri E. The effects of an exposure-based mobile app on symptoms of posttraumatic stress disorder in veterans: pilot randomized controlled trial. JMIR Mhealth Uhealth 2022 Nov 4;10(11):e38951. [doi: 10.2196/38951] [Medline: 36331540]
10. Miller ML, Davis AC, McLean CP. Development and pilot testing of a trauma-focused cognitive-behavioral self-management mobile app for post-traumatic stress symptoms. J Technol Behav Sci 2022 Mar;7(1):63-72. [doi: 10.1007/s41347-020-00188-x]
11. McLean CP, Levy HC, Miller ML, Tolin DF. Exposure therapy for PTSD: a meta-analysis. Clin Psychol Rev 2022 Feb;91:102115. [doi: 10.1016/j.cpr.2021.102115] [Medline: 34954460]

12. National Center for Veterans Analysis and Statistics. VA utilization profile FY 2016. United States Department of Veterans Affairs. 2017 Nov. URL: https://www.va.gov/vetdata/docs/Quickfacts/VA_Utilization_Profile.pdf [accessed 2024-06-07]
13. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap) – a metadata-driven methodology and workflow process for providing translational research Informatics support. J Biomed Inform 2009 Apr;42(2):377-381. [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
14. Weathers FW, Litz BT, Keane TM, Palmieri PA, Marx BP, Schnurr PP. The Posttraumatic Stress Disorder Checklist for DSM-5 (PCL-5). PTSD: National Center for PTSD. 2013. URL: https://www.ptsd.va.gov/professional/assessment/adult-sr/ptsd-checklist.asp [accessed 2024-06-07]
15. Ng MM, Firth J, Minen M, Torous J. User engagement in mental health apps: a review of measurement, reporting, and validity. Psychiatr Serv 2019 Jul 1;70(7):538-544. [doi: 10.1176/appi.ps.201800519] [Medline: 30914003]
16. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition: American Psychiatric Publishing, Inc; 2013. [doi: 10.1176/appi.books.9780890425596]
17. Bovin MJ, Marx BP, Weathers FW, et al. Psychometric properties of the PTSD Checklist for Diagnostic and Statistical Manual of Mental Disorders–Fifth Edition (PCL-5) in veterans. Psychol Assess 2016 Nov;28(11):1379-1391. [doi: 10.1037/pas0000254] [Medline: 26653052]
18. Wortmann JH, Jordan AH, Weathers FW, et al. Psychometric analysis of the PTSD Checklist-5 (PCL-5) among treatment-seeking military service members. Psychol Assess 2016 Nov;28(11):1392-1403. [doi: 10.1037/pas0000260] [Medline: 26751087]
19. Taylor LS, Caloudas SG, Haney LC, Teng E. Asynchronous assessment with the PCL-5: practice considerations and recommendations. Psychol Serv 2023 Nov 30. [doi: 10.1037/ser0000824] [Medline: 38032650]
20. IBM Corp. IBM SPSS Statistics for Windows, version 26.0. : IBM Corp; 2019.
21. Kuhn E, Owen JE. Advances in PTSD treatment delivery: the role of digital technology in PTSD treatment. Curr Treat Options Psych 2020 Jun;7(2):88-102. [doi: 10.1007/s40501-020-00207-x]
22. Rodriguez-Paras C, Tippey K, Brown E, et al. Posttraumatic stress disorder and mobile health: app investigation and scoping literature review. JMIR Mhealth Uhealth 2017 Oct 26;5(10):e156. [doi: 10.2196/mhealth.7318] [Medline: 29074470]
23. Steubl L, Sachser C, Baumeister H, Domhardt M. Mechanisms of change in internet-and mobile-based interventions for PTSD: a systematic review and meta-analysis. Eur J Psychotraumatol 2021 Feb 19;12(1):1879551. [doi: 10.1080/20008198.2021.1879551] [Medline: 34025920]
24. Gan DZQ, McGillivray L, Han J, Christensen H, Torok M. Effect of engagement with digital interventions on mental health outcomes: a systematic review and meta-analysis. Front Digit Health 2021 Nov;3:764079. [doi: 10.3389/fdgth.2021.764079] [Medline: 34806079]
25. Zhang R, Nicholas J, Knapp AA, et al. Clinically meaningful use of mental health apps and its effects on depression: mixed methods study. J Med Internet Res 2019 Dec 20;21(12):e15644. [doi: 10.2196/15644] [Medline: 31859682]
26. Nardi W, Roy A, Dunsiger S, Brewer J. Analyzing the impact of mobile app engagement on mental health outcomes: secondary analysis of the unwinding anxiety program. J Med Internet Res 2022 Aug 15;24(8):e33696. [doi: 10.2196/33696] [Medline: 35969440]
27. Lipschitz JM, Van Boxtel R, Torous J, et al. Digital mental health interventions for depression: scoping review of user engagement. J Med Internet Res 2022 Oct 14;24(10):e39204. [doi: 10.2196/39204] [Medline: 36240001]
28. Galovski TE, McLean CP, Davis CA, Wachen JS. Psychosocial treatment for adults with PTSD. In: Friedman MJ, Schnurr PP, Keane TM, editors. Handbook of PTSD: Science and Practice: Guilford Press; 2021.
29. Linardon J, Torous J, Firth J, Cuijpers P, Messer M, Fuller-Tyszkiewicz M. Current evidence on the efficacy of mental health smartphone apps for symptoms of depression and anxiety. a meta‐analysis of 176 randomized controlled trials. World Psychiatry 2024 Feb;23(1):139-149. [doi: 10.1002/wps.21183] [Medline: 38214614]
30. Foa EB, Hembree EA, Rothbaum BO, Rauch SAM. Prolonged Exposure Therapy for PTSD: Emotional Processing of Traumatic Experiences—Therapist Guide: Oxford University Press; 2019. [doi: 10.1093/med-psych/9780190926939.001.0001]

Abbreviations

***DSM-5*:** *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition*
mHealth: mobile health
PCL-5: Posttraumatic Stress Disorder Checklist for *DSM-5*
PTSD: posttraumatic stress disorder
RedCap: Research Electronic Data Capture
VA: Department of Veterans Affairs

Edited by L Buis; submitted 26.05.23; peer-reviewed by M Gasmi, R Sripada, Y Pachankis; revised version received 02.04.24; accepted 22.04.24; published 18.07.24.

Please cite as:
Davis CA, Miller M, McLean CP
The Impact of User Engagement With Exposure Components on Posttraumatic Stress Symptoms in an mHealth Mobile App: Secondary Analysis of a Randomized Controlled Trial
JMIR Mhealth Uhealth 2024;12:e49393
URL: https://mhealth.jmir.org/2024/1/e49393
doi:10.2196/49393

Efficacy of a Web-Based Home Blood Pressure Monitoring Program in Improving Predialysis Blood Pressure Control Among Patients Undergoing Hemodialysis: Randomized Controlled Trial

Tingting Chen[1*], MSN, RN; Wenbo Zhao[2*], MD, PhD; Qianqian Pei[3], MSN, RN; Yanru Chen[2], BSN, RN; Jinmei Yin[4], BSN, RN; Min Zhang[5], MSc; Cheng Wang[4], PhD, MD; Jing Zheng[1], PhD, RN

[1]School of Nursing, Guangdong Pharmaceutical University, Guangzhou, China

[2]Division of Nephrology, The Third Affiliated Hospital of Sun Yat-sen University, Guangzhou, China

[3]School of Tourism Management, Three Gorges Tourism Polytechnic College, Yichang, China

[4]Division of Nephrology, The Fifth Affiliated Hospital Sun Yat-sen University, Zhuhai, China

[5]School of Public Health, Guangdong Pharmaceutical University, Guangzhou, China

[*]these authors contributed equally

Corresponding Author:
Jing Zheng, PhD, RN
School of Nursing, Guangdong Pharmaceutical University, , Guangzhou, , China

Abstract

Background: Hypertension is highly prevalent among patients undergoing hemodialysis, with a significant proportion experiencing poorly controlled blood pressure (BP). Digital BP management in this population has been underused.

Objective: This study aimed to explore the efficacy of a web-based home BP monitoring (HBPM) program in improving predialysis BP control and enhancing knowledge, perception, and adherence to HBPM among patients with hypertension undergoing hemodialysis.

Methods: A multicenter, open-label, randomized controlled trial was conducted at 2 hemodialysis units. Patients were randomly allocated in a 1:1 ratio to either the web-based HBPM program as the intervention group or to usual care as the control group over a 6-month period. The primary outcomes were the predialysis BP control rate, defined as less than 140/90 mm Hg, and the predialysis systolic and diastolic BP, assessed from baseline to the 6-month follow-up. Secondary outcomes included patient knowledge, perception, and adherence to HBPM, evaluated using the HBPM Knowledge Questionnaire, HBPM Perception Scale, and HBPM Adherence Scale, respectively. A generalized estimating equations analysis was used to analyze the primary outcomes in the intention-to-treat analysis.

Results: Of the 165 patients enrolled in the program (n=84, 50.9% in the web-based HBPM group and n=81, 49.1% in the control group), 145 (87.9%) completed the follow-up assessment. During the follow-up period, 11 instances of hypotension occurred in 9 patients in the web-based HBPM group, compared to 15 instances in 14 patients in the control group. The predialysis BP control rate increased from 30% (25/84) to 48% (40/84) in the web-based HBPM group after the 6-month intervention, whereas in the control group, it decreased from 37% (30/81) to 25% (20/81; χ^2_2=16.82, *P*<.001; odds ratio 5.11, 95% CI 2.14-12.23, *P*<.001). The web-based HBPM group demonstrated a significant reduction after the 6-month intervention in the predialysis systolic BP (t_{163}=2.46, *P*=.02; β=−6.09, 95 % CI −10.94 to −1.24, *P*=.01) and the predialysis diastolic BP (t_{163}=3.20, *P*=.002; β=−4.93, 95% CI −7.93 to −1.93, *P*=.001). Scores on the HBPM Knowledge Questionnaire (t_{163}=−9.18, *P*<.001), HBPM Perception Scale (t_{163}=−10.65, *P*<.001), and HBPM Adherence Scale (t_{163}=−8.04, *P*<.001) were significantly higher after 6 months of intervention.

Conclusions: The implementation of a web-based HBPM program can enhance predialysis BP control and the knowledge, perception, and adherence to HBPM among patients undergoing hemodialysis. This web-based HBPM program should be promoted in appropriate clinical settings.

Trial Registration: China Clinical Trial Registration Center ChiCTR2100051535; https://www.chictr.org.cn/showproj.html?proj=133286

(JMIR Mhealth Uhealth 2024;12:e53355) doi:10.2196/53355

KEYWORDS

hemodialysis; hypertension; home blood pressure monitoring; eHealth; randomized controlled trial

Introduction

Background

Hemodialysis has become the predominant renal replacement therapy for individuals with end-stage renal disease, the final stage of chronic kidney disease. The prevalence of patients receiving hemodialysis has surged alongside the increasing incidence of end-stage renal disease cases. Globally, the number of patients requiring renal replacement therapy is expected to reach 5.4 million by 2030, with nearly 89% of them undergoing hemodialysis [1]. According to the Chinese National Renal Data System, more than 1 million patients underwent dialysis in 2022 [2].

Hypertension is nearly ubiquitous among patients receiving hemodialysis, who often have suboptimal blood pressure (BP) control. The epidemiology of hypertension in this population varies across studies, depending on the definition of hypertension and the method used to measure BP. The overall prevalence of hypertension in patients undergoing hemodialysis ranged from 80% to 90% [3-5]. The European Registry of Cardiovascular and Renal Medicine of the European Renal Association—European Dialysis and Transplant Association showed that hypertension, defined as having ambulatory BP ≥130/80 mm Hg at 48 hours or current use of medications to reduce BP, was prevalent in 84.3% of patients undergoing hemodialysis [3]. Another study based on the criteria of predialysis BP ≥140/90 mm Hg or current use of antihypertensives indicated that 86.2% of patients undergoing hemodialysis have hypertension [4]. Additionally, the China Dialysis Outcomes and Practice Patterns Study reported that 87.3% of patients undergoing hemodialysis were complicated by hypertension [5]. Although 86.8% of these patients received pharmacological treatment, fewer than 30% had their BP well controlled [3].

Increased BP in patients undergoing hemodialysis has been found to have a linear association with adverse cardiovascular events and mortality. Recent studies have identified a U-shaped or J-shaped correlation between BP before or after dialysis and the risk of all-cause mortality in patients undergoing hemodialysis. This phenomenon, termed the “reverse epidemiology” of hypertension, is predominantly observed within dialysis units [6,7]. Elevated BP levels measured outside the dialysis unit demonstrated better prognostic capacity for adverse cardiovascular events and mortality compared to measurements taken within the dialysis unit [8-10]. Therefore, it is imperative to develop a more effective BP management model for patients undergoing hemodialysis, especially BP measurements taken outside the dialysis unit, to optimize the reduction in mortality risk.

BP monitoring plays a crucial role in the management of BP, representing a fundamental prerequisite therein, and is a crucial tool for the identification, diagnosis, and prognostic evaluation of hypertension. For patients undergoing hemodialysis, BP measurements are taken at the hemodialysis unit and outside the hemodialysis unit, and the latter consists of home BP monitoring (HBPM) and ambulatory BP monitoring [11]. Despite the conventional clinical application of BP measurements taken within dialysis units for diagnosing hypertension, the limitations of this approach persist [11]. In contrast, HBPM offers improved reproducibility [12], superior diagnostic accuracy [13], a more intimate nexus with target organ damage [14], and more prognostic information [15] and serves as a guide for long-term antihypertensive regimens [15]. Compared to ambulatory BP monitoring, HBPM has advantages in terms of lower cost, simpler operation, and greater patient acceptance [11]. Furthermore, the existing literature highlighted that positive behavior with HBPM can promote patient adherence to obtaining or reporting home BP, achieving better management of BP [16].

Under the proposal of national information health care policies [17,18], the maturation of information technology, and the amalgamation of network information, medical health services have engendered novel perceptions for the management of BP. Within the realm of existing digital interventions for BP, a variety of modalities such as BP telemonitoring, smartphone app-based tracking, phone calls, websites, emails, and SMS text messages have been used, demonstrating notable efficacious results [19]. Compared to conventional BP management paradigms, digital BP management is more convenient and flexible, with greater patient acceptance [19]. However, digital BP management has been applied predominantly in the primary hypertension population, rather than in patients undergoing hemodialysis [19]. It is unclear whether the efficacy of digital BP management modalities has the potential to influence predialysis BP control in patients with hypertension undergoing hemodialysis.

Objective

To bridge this knowledge gap, our goal was to develop a customized digital BP management program for patients with hypertension undergoing hemodialysis and to evaluate its long-term feasibility, efficacy, and safety in these patients. A 6-month, prospective, multicenter, randomized controlled trial was designed, with the aim of a comprehensive investigation of the efficacy of a digital BP management program in improving predialysis BP control rates, along with an evaluation of its implications for cognition, perception, and adherence to HBPM among patients with hypertension undergoing hemodialysis.

Methods

Study Design

This study was a multicenter, open-label, randomized controlled trial comparing the web-based HBPM program against usual care for patients with hypertension undergoing hemodialysis, which was performed at the dialysis centers of 2 tertiary hospitals in Guangdong Province, China, from August 2022 to February 2023. The design of the study was in accordance with the specifications outlined in the CONSORT-eHEALTH checklist (Checklist 1).

Ethical Considerations

The study protocol was registered with the Chinese Clinical Trial Registry (ChiCTR2100051535) and was approved by the ethics committees of Zhongshan Third Affiliated Hospital of

Sun Yat-sen University ([2019] 02-520-01) and Zhongshan Fifth Affiliated Hospital of Sun Yat-sen University ([2022] K141-1). The authors adhered to the tenets of the Declaration of Helsinki and its amendments. Written informed consent was obtained from each participant. All data collected from participants were recorded in an anonymized format. Participants in both groups were paid equally, such as a sphygmomanometer or an equivalent gift. These details were conveyed to the participants at the time of signing the informed consent form.

Participants

The inclusion criteria were (1) aged between 18 to 80 years; (2) mean predialysis systolic BP (SBP) of ≥140 mm Hg, mean diastolic BP (DBP) of ≥90 mm Hg over 3 consecutive sessions, the use of antihypertensive medication [20], or any combination of the above; (3) stage 5 and estimated glomerular filtration rate less than 15 mL/(min×1.73 m^2); (4) dialysis initiation period exceeding 3 months; (5) competencies in communication, perception, and learning; and (6) proficiency in smartphone operation and willingness to participate in the study. Exclusion criteria included patients who (1) transferred to intensive care units for any reason; (2) experienced complicating infection or bleeding; (3) died during the research period; (4) had mental or physical disabilities impairing their ability to respond to inquiries; (5) were unable to see or hear or were unable to measure BP due to upper-limb disability; or (6) demonstrated an inability to engage in self-care, as indicated by a score below 60 on the Activities of Daily Living scale. Participants were enrolled by researchers (YC and JY).

A simple randomization procedure was performed by researchers (JZ and MZ) using a random number table generated from the list of patients undergoing hemodialysis, at a 1:1 randomization ratio. Since patient dialysis management was conducted by different groups, each group of 8 - 10 patients had a relatively fixed dialysis time, dialysis area, and group of attending nurses. To minimize bias, the allocation of participants into study groups was conducted using a block randomization procedure. The random allocation sequence was generated according to the dialysis group through a publicly available web-based tool [21].

Intervention

Intervention Group

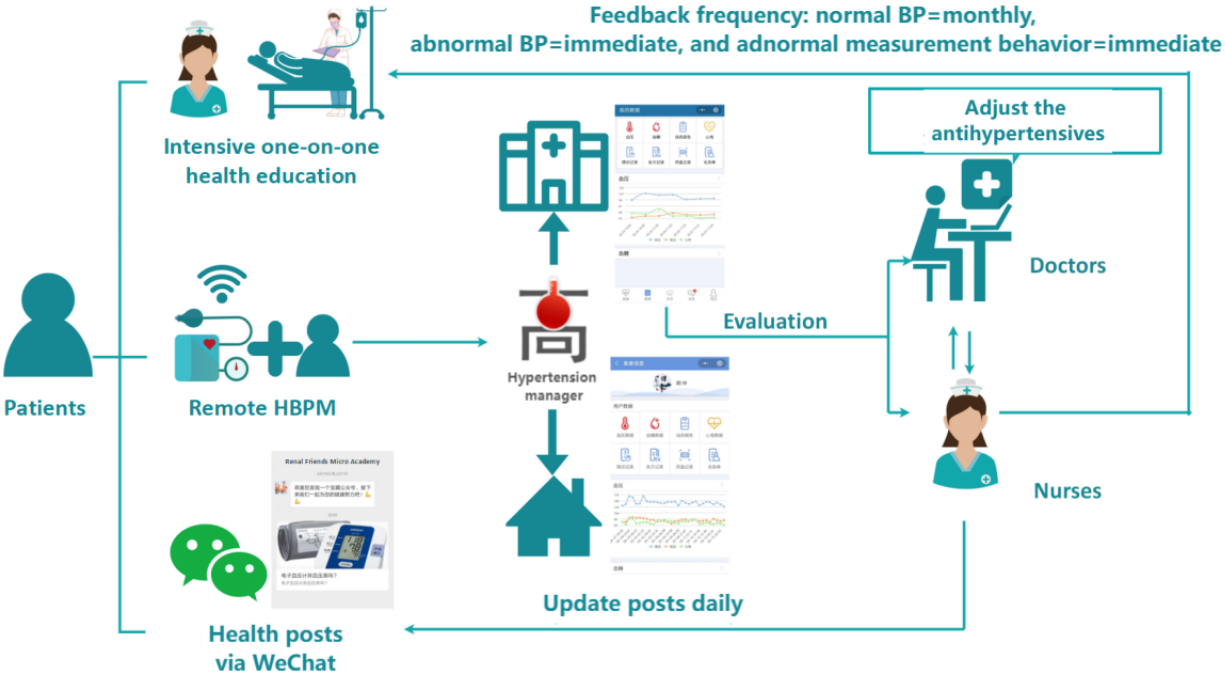
Development of the Web-Based HBPM Program

Concomitant with the management of the disease according to the *Chinese Standard Operating Procedures for Blood Purification* [22], the patients in the intervention group received complementary management of BP in the form of the web-based HBPM program. The theoretical framework of this program was developed according to the Health Promotion Model [23]. The program protocol was methodically formulated by the research group by systematically searching, screening, evaluating, and synthesizing the literature and based on clinical experience. Subsequently, the final version of the web-based HBPM program was validated using a peer-reviewed expert consensus approach.

Web-Based HBPM Program Protocol

The web-based HBPM program was characterized by a tripartite framework that encompassed intensive one-on-one health education; remote HBPM; and health posts disseminated through the WeChat platform (Figure 1). The program was implemented by health care professionals, including doctors and nurses in the hemodialysis centers, who received training in hypertension management conducted by the research group prior to the intervention.

Figure 1. The framework of the web-based HBPM program. BP: blood pressure; HBPM: home blood pressure monitoring.



Intensive Individualized Health Education

Intensive, one-on-one health education was a component of the web-based HBPM program. At the beginning of the intervention, the nurses in charge of the patients in the web-based HBPM group implemented dedicated, intensive, and individualized health education according to the *Nurse Education Manual*, which was formulated through a review of the literature. First, nurse leaders provided comprehensive instruction to participants that included the proper techniques for normative BP measurement, including the required measurement frequency, time, and recording of BP readings. The instruction was provided in half an hour. Subsequently, nurse leaders connected the patient's sphygmomanometers to the WeChat platform to facilitate remote HBPM. After that, nurse leaders provided monthly one-on-one education for 6 months. Nurse leaders offered feedback according to patients' home BP readings from the previous month; laboratory test results; and assessments of patient knowledge, perceptions, and adherence to HBPM. They provided personalized coaching to patients monthly, using positive encouragement to improve patients' confidence in monitoring their BP at home appropriately. Furthermore, if patients developed conditions (abnormal BP, abnormal measurement behavior, and relative interdialytic weight gain ≥5%), immediate education was required. Two researchers (TC and WZ) were responsible for providing web-based counseling and responding to patients at any time.

Remote HBPM

The core component of the web-based HBPM program was remote HBPM. To establish a digital conduit for the transmission and recording of BP data, internet-connected sphygmomanometers were used. These devices automatically uploaded BP measurement data to the cloud for monitoring purposes. The cloud was divided into 2 portals: the health care professional portal and the patient portal. Both health care professionals and patients used this cloud on the internet to collaboratively monitor patients' home BP.

After registering the sphygmomanometers, which automatically transmitted BP measurement data to the patient's WeChat account, nurse leaders guided patients to check BP reports on the web. The patients were then instructed to monitor their home BP with these sphygmomanometers following the nurse's instructions. The requirements for HBPM behavior were as follows: for patients whose BP met the standards (home BP <135/85 mm Hg [24]), it was recommended that they monitor their BP 1-2 days per week; for patients with unstable or substandard BP, it was recommended that they measure their BP 5-7 days per week. During these monitoring sessions, measurements were performed both in the morning and in the evening, following a rest period of 1-2 minutes.

Nurse leaders tracked both home BP and dialysis BP using the health care professional portal via the cloud and hemodialysis machine. Nurse leaders observed and evaluated the measurements of BP monthly through the health care professional portal and hemodialysis machine. They observed abnormal BP instances, including uncontrolled home BP (defined as an average home BP of ≥135/85 mm Hg [24]), interdialytic hypotension (defined as a reduction in SBP of 20 mm Hg or more or a decrease in mean arterial pressure by 10 mm Hg, accompanied by symptoms on nondialysis days), intradialytic hypotension (a decrease in BP accompanied by symptoms in dialysis procedure [25]), and uncontrolled predialysis BP (an average predialysis BP of ≥140/90 mm Hg [22]). They provided feedback to physicians about abnormal BP and discussed the adjustments to achieve BP management. Meanwhile, nurse leaders evaluated patient compliance with HBPM through the health care professional portal.

The physicians used the health care professional portal on the cloud and hemodialysis machine to monitor the patients. Initially, they adopted comprehensive nonpharmacological strategies, such as adjusting patient's dry weight and controlling fluid and sodium intake. If these measures did not normalize the patient's BP (predialysis BP <140/90 mm Hg), the patient's antihypertensive regimen would be adjusted. Additionally, adjustments to ultrafiltration and dry weight would be made, and instructions on managing interdialytic hypotension would be conducted according to the *China Standard Operating Procedures for Blood Purification* [22] to ensure safety at home. Throughout the continuum of BP management, health care professionals remained concerned about the effect of BP reduction, patient feelings, and adverse events through BP management and set goals for the next stage.

Health Posts via WeChat

Health posts via WeChat served as an integral component of the web-based HBPM program. Before the intervention, nurse leaders assisted patients in subscribing to a public WeChat account dedicated to self-care knowledge and skills related to BP management for patients undergoing hemodialysis, created by the researchers in this study. A researcher updated the health posts on this public WeChat account and distributed them to patients daily, while also monitoring their engagement and encouraging active participation. The WeChat health posts covered 35 topics consisting of 196 posts, including discussions on HBPM, fluid management, and calcium-phosphorus management. The content was regularly updated to ensure relevance and timeliness.

Control Group

Participants assigned to the control group received routine disease management based on the *China Standard Operating Procedures for Blood Purification* [22]. Patients were advised to measure their home BP, record the results in a notebook, and return it to the health care professional voluntarily 1 or several months later. If the home BP of the patients did not reach the standard BP, the physicians would adjust treatment through nonpharmacological or pharmacological methods to control the BP of the patients according to *Chinese Standard Operating Procedures for Blood Purification* [22]. Patients were provided with education from nurses on BP self-monitoring and BP management at the initiation of antihypertensive therapy.

Outcomes and Instruments

The primary outcomes were the predialysis BP control rate, with targets set at a predialysis BP of less than 140/90 mm Hg [22], and the predialysis SBP and DBP. Secondary outcomes

included patient knowledge, perception, and adherence to HBPM.

The Predialysis BP and HBPM Instrument

Predialysis BP was obtained before dialysis treatment using the automated BP monitor attached to the hemodialysis machine (Dialog+) in the dialysis unit. Predialysis BP was measured using standardized in-office methods [11]. At baseline and at each follow-up (after 1, 3, and 6 months of the study period), the BP readings were collected and averaged from the latest week, measured by the automated BP monitor attached to the hemodialysis machine. The outcome assessors (QP and CW) for BP were blinded to the intervention assignment.

Home BP was measured by an automatic electronic sphygmomanometer (A666G; Kangkang Shengshi Information Technology Co., Ltd [26]) in the web-based HBPM group.

HBPM Knowledge Questionnaire

Cognition of hypertension prevention and BP monitoring was measured using the HBPM Knowledge Questionnaire (HBPMKQ), which was developed according to the guidelines [27,28] by the research group. It consists of 20 multiple-choice questions with a potential total score ranging from 0 to 60, with lower scores indicating weaker knowledge. The difficulty coefficient and discrimination indices of the questionnaire were moderate and well defined [29].

HBPM Perception Scale

The perception of HBPM was assessed using the HBPM Perception Scale (HBPMPS) [29]. This scale is a 5-point Likert scale comprising 27 items across 5 dimensions: perceived benefit of HBPM, perceived barriers to HBPM, perceived self-efficacy of HBPM, situational influence, commitment to an HBPM plan. The total score ranges from 27 to 135, with lower scores indicating a worse attitude toward HBPM.

HBPM Adherence Scale

Adherence to HBPM was measured using the HBPM Adherence Scale (HBPMAS) [30]. The HBPMAS uses a 5-point Likert scale and consists of 8 items, with lower scores indicating poorer adherence to HBPM.

Data Collection

A trained investigator collected the data by interviewing patients. Sociodemographic and clinical characteristics were collected by a trained investigator using electronic questionnaires and by retrieving clinical data from the charts at the beginning of the study. The predialysis BP values between the 2 groups were recorded by a trained investigator and measured during the same period at baseline (T0) and at the 1-, 3-, and 6-month follow-ups (T1, T2, and T3, respectively). Electronic questionnaires for the HBPMKQ, HBPMPS, and HBPMAS were completed by both groups at the beginning of the study and at the 1-, 3-, and 6-month follow-ups.

Statistical Analysis

To test the differences between the 2 groups, α was set at .05 and the test power ($1 - \beta$) was set at .90. A previous study reported a predialysis BP control rate of 28.2% (defined as <140/90 mm Hg) [3]. Assuming a 25% improvement after the intervention, the sample size was calculated to be 124 cases according to the 2 independent sample rates. Considering a dropout rate of 20%, at least 155 patients must be enrolled.

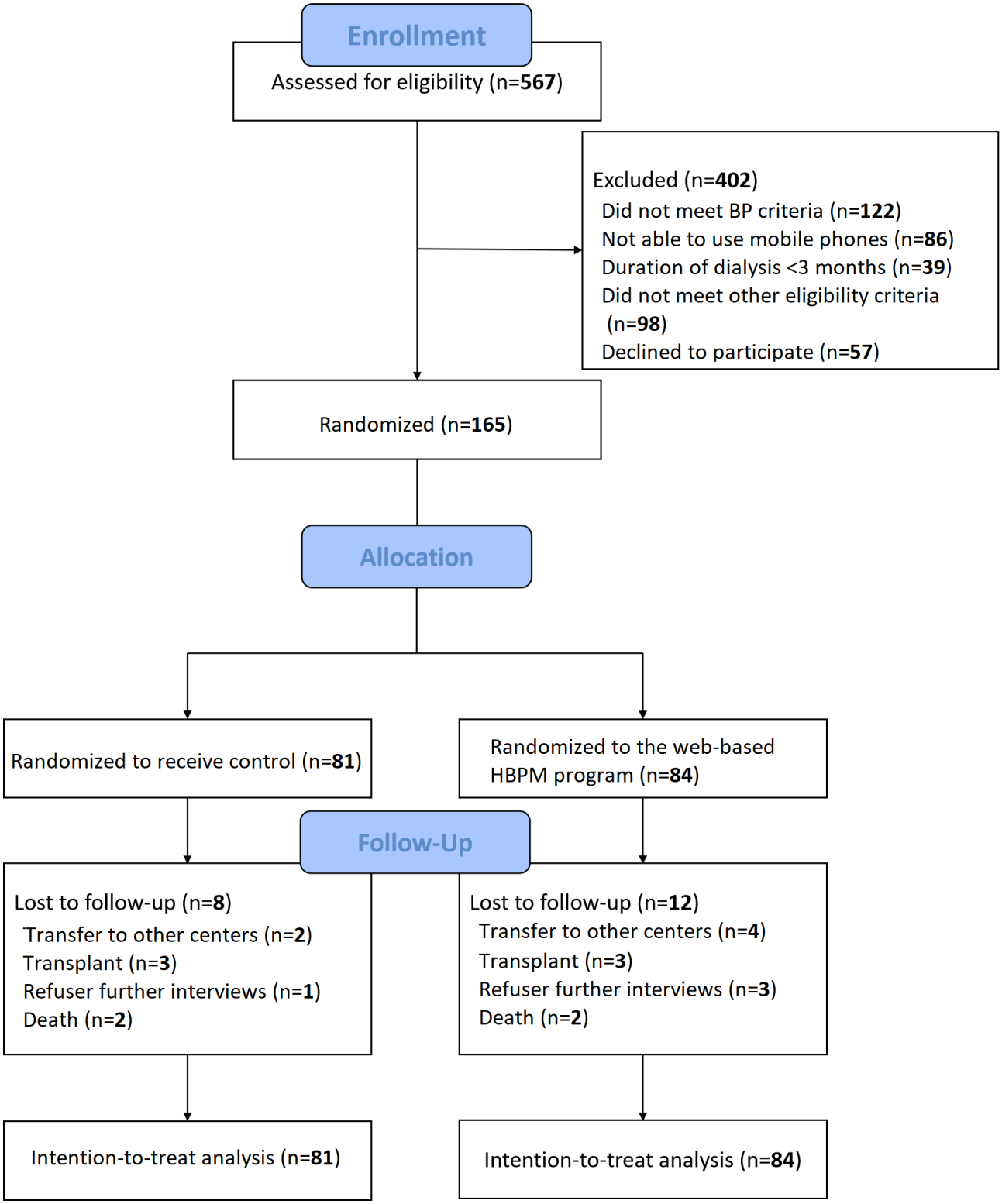
All analyses used intention-to-treat principles, and missing data were input using the last-observation-carried-forward approach. Sociodemographic and clinical characteristics and study outcomes were described as mean (SD) or n (%). The comparison of sociodemographic and clinical variables and study outcomes between the 2 groups was conducted using independent-sample (2-tailed) *t* tests and χ^2 tests. The primary outcomes were analyzed using generalized estimating equations. Model 1 was adjusted for group and time. Model 2 also controlled for hospital, age, sex, education, employment status, and marital status. Model 3 additionally controlled for smoking, BMI, the number of antihypertensive agents, duration of dialysis, interdialytic weight gain, urea clearance index, and weekly dialysis frequency. All analyses were processed in SPSS software (version 25.0; IBM Corp).

Results

Enrollment and Participant Allocation

During the recruitment periods, a total of 567 patients undergoing hemodialysis were assessed for eligibility. Among them, 165 enrolled patients were randomly assigned to either the web-based HBPM program group (n=84, 50.9%) or the usual care group (n=81, 49.1%; Figure 2). Dialysis unit A comprised 81 (49.1%) patients, whereas dialysis unit B comprised 84 (50.9%) patients. No statistically significant differences were observed in the demographic and clinical characteristics of the participants between the 2 dialysis units (all *P*>.05). A total of 20 participants prematurely discontinued the study, primarily due to transferring to other dialysis centers, receiving kidney transplantation, declining further interviews, and death. Throughout the follow-up period, the web-based HBPM program group reported 11 instances of hypotension in 9 patients, including 10 instances of intradialytic hypotension and 1 instance of interdialytic hypotension. The control group reported 15 instances of hypotension in 14 patients, including 11 instances of intradialytic hypotension and 4 instances of interdialytic hypotension. No hypotension-related adverse events, such as cardiovascular events (eg, acute myocardial infarction or ischemic stroke) or falls, were observed in either group.

Figure 2. Flow diagram of the web-based HBPM program. BP: blood pressure; HBPM: home blood pressure monitoring.



Demographic and Clinical Characteristics of the Participants

Of the 165 patients, the mean age was 53.7 (14.0) years and 66.1% (n=109) were male. The mean duration of hemodialysis was 57.8 (45.9) months. Glomerulonephritis (n=70, 42.4%) was the main etiology of renal disease. A total of 84.8% (n=140) of patients received treatment for their hypertension, 43.6% (61/140) were treated with 1 antihypertensive agent, and 56.4% (79/140) were given a combination of antihypertensives. Among those who received a combination of antihypertensives, the combination of calcium channel blockers+β-blockers+angiotensin II receptor antagonists was the most common (25/79, 32%), followed by calcium channel blockers+β-blockers (22/79, 28%). Except for education, other characteristics were similar between the 2 groups (Table 1).

Table . Demographic and clinical characteristics of participants (n=165).

Demographic and clinical characteristics		Web-based HBPM[a] group (n=84)	Control group (n=81)	*t* test or chi-square (*df*)	*P* value
Demographic information					
Age (years), mean (SD)		53.15 (14.65)	54.17 (13.36)	.47 (163)	.64
Sex, n (%)				.03 (1)	.87
	Male	56 (67)	53 (65)		
	Female	28 (33)	28 (35)		
Education, n (%)				7.98 (3)	.046
	Elementary school	8 (10)	13 (16)		
	Junior middle school	21 (25)	32 (40)		
	High school	26 (31)	14 (17)		
	College or above	29 (34)	22 (27)		
Employment status, n (%)				2.00 (2)	.37
	Employed	27 (32)	20 (25)		
	Unemployed	32 (38)	29 (36)		
	Retired	25 (30)	32 (40)		
Marital status, n (%)				.25 (1)	.62
	Married or cohabiting	66 (79)	61 (75)		
	Single, divorced, or widowed	18 (21)	20 (25)		
Current smoking, n (%)				.26 (1)	.61
	Yes	15 (18)	17 (21)		
	No	69 (82)	64 (79)		
Medical history					
Cause of the renal disease, n (%)				1.02 (5)	.96
	Glomerulonephritis	34 (40)	36 (44)		
	Diabetic nephropathy	20 (24)	21 (26)		
	Hypertensive nephropathy	10 (12)	9 (11)		
	Polycystic renal disease	5 (6)	3 (4)		
	Obstructive nephropathy	3 (4)	3 (4)		
	Other or unknown	12 (14)	9 (11)		
Comorbidities, n (%)					
	Diabetes mellitus	23 (27)	26 (32)	.44 (1)	.51
	Heart failure	8 (10)	9 (11)	.03 (1)	.86
	Peripheral vascular disease	10 (12)	14 (17)	.96 (1)	.33
	Cerebrovascular disease	7 (8)	6 (7)	.05 (1)	.83
	Asthma or COPD[b]	2 (2)	1 (1)	.30 (1)	.58
	Cancer	4 (5)	4 (5)	.003 (1)	.96

Demographic and clinical characteristics		Web-based HBPM[a] group (n=84)	Control group (n=81)	*t* test or chi-square (*df*)	*P* value
BMI (kg/m^2), mean (SD)		21.08 (3.46)	22.55 (3.36)	.88 (163)	.38
Family history of hypertension, n (%)				.84 (1)	.36
	Yes	37 (44)	30 (37)		
	None	47 (56)	51 (63)		
Number of antihypertensive agents used, n (%)				4.65 (3)	.20
	0	8 (10)	17 (21)		
	1	35 (42)	26 (32)		
	2	21 (25)	20 (25)		
	≥3	20 (24)	18 (22)		
Duration of dialysis (month), mean (SD)		54.62 (47.91)	61.07 (43.78)	−.90 (163)	.37
IDWG/d[c], mean (SD)		.83 (.38)	.87 (.44)	.72 (163)	.47
Weekly dialysis frequency, n (%)				.03 (1)	.86
	2	9 (11)	8 (10)		
	3	75 (89)	73 (90)		
Kt/V[d], mean (SD)		1.36 (0.27)	1.40 (0.33)	.90 (163)	.37

[a]HBPM: home blood pressure monitoring.

[b]COPD: chronic obstructive pulmonary disease.

[c]IDWG/d: interdialytic weight gain.

[d]Kt/V: urea clearance index, calculated as the ratio of the urea clearance by the dialyzer and the volume for a particular dialysis duration.

Predialysis BP Control Rate, BP, Knowledge of Patients, Perception, and Adherence to the HBPM

Descriptive statistics were used to describe the control rate of predialysis BP; predialysis BP; and HBPMKQ, HBPMPS, and HBPMAS scores at baseline and at the 1-, 3-, and 6-month follow-ups (Table 2). Preliminary calculations suggested that the web-based HBPM program improved patient' predialysis BP control rate; predialysis BP; and the patients' knowledge, perception, and adherence to the HBPM after the 6-month intervention.

Table . Descriptive statistics for outcomes from baseline to the 6-month postintervention follow-up.

Variables and time point	Web-based HBPM[a] group (n=84)	Δ (T3 – T0)	Control group (n=81)	Δ (T3 – T0)	*t* test or chi-square (*df*)	*P* value
Predialysis BP[b] control, n (%)		15 (18)		−10 (−12)	16.82 (2)	<.001
T0[c]	25 (30)		30 (37)			
T1[d]	32 (38)		22 (27)			
T2[e]	30 (36)		21 (26)			
T3[f]	40 (48)		20 (25)			
Predialysis SBP[g] (mm Hg), mean (SD)		−3.93 (13.16)		2.16 (18.32)	2.46 (163)	.02
T0	146.75 (15.97)		144.05 (17.70)			
T1	146.94 (16.93)		145.69 (17.80)			
T2	145.85 (15.10)		146.33 (15.83)			
T3	142.82 (15.45)		146.21 (16.56)			
Predialysis DBP[h] (mm Hg), mean (SD)		−2.16 (10.40)		2.77 (9.38)	3.20 (163)	.002
T0	83.71 (12.19)		82.02 (11.31)			
T1	84.01 (11.68)		83.38 (12.89)			
T2	82.87 (11.96)		83.86 (12.10)			
T3	81.55 (12.20)		84.79 (12.80)			
HBPMKQ[i] score, mean (SD)		16.5 (9.96)		2.07 (10.24)	−9.18 (163)	<.001
T0	40.36 (9.46)		41 (7.79)			
T1	48.71 (6.22)		42.67 (7.75)			
T2	53.71 (4.94)		44.33 (7.75)			
T3	56.86 (5.72)		43.07 (8.63)			
HBPMPS[j] score, mean (SD)		18.32 (15.76)		−4.19 (11.04)	−10.65 (163)	<.001
T0	95.7 (9.44)		91.09 (7.26)			
T1	102.12 (9.26)		96.8 (10.82)			
T2	103.19 (7.57)		95.72 (9.61)			
T3	114.02 (10.58)		86.9 (9.32)			
HBPMAS[k] score, mean (SD)		6.44 (5.97)		−2.43 (8.08)	−8.04 (163)	<.001
T0	31.18 (4.70)		29.86 (6.06)			
T1	36.46 (3.65)		30.02 (6.82)			
T2	35.06 (3.38)		28.38 (7.41)			
T3	37.62 (3.65)		27.43 (7.51)			

[a]HBPM: home blood pressure monitoring.

[b]BP: blood pressure.

[c]T0: at the baseline of the study.

[d]T1: 1-month follow-up.

[e]T2: 3-month follow-up.

[f]T3: 6-month follow-up.

[g]SBP: systolic blood pressure.

[h]DBP: diastolic blood pressure.

[i]HBPMKQ: Home Blood Pressure Monitoring Knowledge Questionnaire.

[j]HBPMPS: Home Blood Pressure Monitoring Perception Scale.

[k]HBPMAS: Home Blood Pressure Monitoring Adherence Scale.

Effect of the Web-Based HBPM Program on Predialysis BP

The results of the generalized estimating equation analysis demonstrated the importance of the duration of follow-up and intergroup interaction across various models (Table 3 and Multimedia Appendix 1). In model 1, controlling for groups and time effects, a significant difference was found between the web-based HBPM group and the control group regarding the predialysis BP control rate at T1 (odds ratio [OR] 2.29, 95% CI 1.16 - 4.53; *P*=.02), T2 (OR 2.20, 95% CI 1.19 - 4.08; *P*=.01), and T3 (OR 3.85, 95% CI 1.89 - 7.83; *P*<.001). The OR of the predialysis BP control rate increased slightly after an additional adjustment for the demographic data of the patients in model 2, while after further adjustment for the clinical characteristics of the patients in model 3, the odds markedly increased. The 2 groups were affected differently during the follow-up period. The 3 models indicated a significant reduction in both predialysis SBP and DBP after 6 months of intervention (Multimedia Appendix 2).

Table . Analysis of the general estimating equations for the effect of the intervention on predialysis BP[a] control rate (n=165). Model 1 controlled for group and time; model 2 also controlled for hospital, age, sex, education, employment status, and marital status; and model 3 additionally controlled for smoking, BMI, the number of antihypertensive agents, duration of dialysis, interdialytic weight gain, urea clearance index, and frequency of weekly dialysis.

Outcomes		Model 1, OR[b] (95% CI)	*P* value	Model 2, OR (95% CI)	*P* value	Model 3, OR (95% CI)	*P* value
Predialysis BP control rate							
Group[c]		0.72 (0.38-1.38)	.32	1.55 (0.78-3.07)	.21	0.70 (0.33-1.48)	.35
Time [d]							
	1 month (T1)	0.63 (0.40-1.01)	.06	1.64 (0.99-2.71)	.06	0.57 (0.32-1.02)	.049
	3 months (T2)	0.60 (0.37-0.97)	.04	1.75 (1.04-2.95)	.04	0.53 (0.29-0.96)	.04
	6 months (T3)	0.56 (0.31-0.99)	.048	1.876 (1.00-3.51)	.049	0.49 (0.24-1.00)	.049
Group [c] **×time** [d]							
	1 month	2.29 (1.16-4.53)	.02	2.43 (1.17-5.06)	.02	2.74 (1.19-6.32)	.02
	3 months	2.20 (1.19-4.08)	.01	2.33 (1.20-4.51)	.01	2.62 (1.23-5.59)	.01
	6 months	3.85 (1.89-7.83)	<.001	4.22 (1.96-9.09)	<.001	5.11 (2.14-12.23)	<.001

[a]BP: blood pressure.

[b]OR: odds ratio.

[c]Reference: baseline.

[d]Reference: control group.

Discussion

Principal Findings

To our knowledge, this is the first study to develop a digital BP management program specifically tailored for patients with hypertension undergoing hemodialysis. Additionally, it explores the application of a web-based HBPM program to manage BP outside of the hemodialysis unit in this patient population. Unlike previous studies that have focused on the management of BP in patients undergoing hemodialysis by adopting conventional HBPM [12,15,16,31]; modifications to hemodialysis regimens [32-34]; patient education and counseling intervention [35,36]; reduction in dry weight [37,38]; dietary sodium restriction [37]; physical exercise [39]; continuity of care through web-based education, telephone visits, and outpatient visits [40]; and cognitive behavior therapy [41], our study innovatively used remote HBPM. Sheppard et al [42] highlighted the categorization of HBPM interventions into 4 intensity levels for managing hypertension. They emphasized that a multicomponent approach integrating goal setting, health education, telemonitoring, and prompt feedback from health care professionals could achieve better BP control. The web-based HBPM program is a high-intensity, level-4 intervention for people with hypertension undergoing hemodialysis, which was complemented by 2 educational approaches: health posts via WeChat and intensive personalized health education. In this study, a total of 165 patients with hypertension undergoing hemodialysis were initially included, with 145 patients ultimately completing the study. The results

unequivocally demonstrated the beneficial impact of the web-based HBPM program in patients undergoing hemodialysis. The program led to enhancements in predialysis BP control rates; predialysis BP; and patient knowledge, perception, and adherence to HBPM, which were particularly evident at the 6-month follow-up assessment. These findings support the viability, efficacy, and security of implementing the web-based HBPM program in patients undergoing maintenance hemodialysis.

In this study, preliminary analysis suggested a notable improvement in the predialysis BP control rate among patients who participated in the web-based HBPM program. Specifically, the proportion of the patients achieving predialysis BP control in the web-based HBPM group increased from 30% (25/84) to 48% (40/84), whereas in the control group, it decreased from 37% (30/81) to 25% (20/81; *P*<.001). Similar results were also observed in patients with essential hypertension [43], revealing an 11.6% improvement in the BP control rate after a 12-week HBPM intervention. The initial predialysis BP control rate among all participants in our study was 33.3% (55/165), which is consistent with the study by Sarafidis et al [3], which reported a rate of 28.2% among European patients undergoing hemodialysis. The implementation of a web-based HBPM program resulted in a remarkable elevation in the predialysis BP control rate. The elevated rate was assumed to be a consequence of improved adherence to treatment, resulting in better control of predialysis BP. The OR of the control rate for BP before dialysis increased when further controlling for clinical characteristics, compared to controlling for demographic characteristics alone. The multivariate adjustment models demonstrated that the predialysis BP control rate in the web-based HBPM group was 2.74 times that of the control group at the 1-month intervention time point, 2.62 times at the 3-month intervention time point, and 5.11 times at the 6-month follow-up. This indicated that the predialysis BP control rate was also affected by clinical characteristics, especially dry weight, ultrafiltration, dialysis adequacy, and weekly dialysis frequency. The intricate interaction between BP and volume management is well recognized in patients undergoing hemodialysis [44]. By controlling for clinical characteristics, the web-based HBPM program has been found to more effectively promote the predialysis BP control rate.

The results showed a significant reduction in predialysis SBP and DBP in the web-based HBPM group compared to the control group at the 6-month follow-up (SBP: β=−6.09, 95% CI −10.94 to −1.24; DBP: β=−4.93, 95% CI −7.93 to −1.93; Multimedia Appendix 2), despite no significant reduction in BP at the 1-month and 3-month follow-ups (*P*>.05). These findings aligned with a previous study, which found that guiding antihypertensive therapy based on BP measurements at home and before dialysis exhibited a significant 6 - to 10–mm Hg change in BP compared to those subjected only to predialysis BP measurements [15]. Our findings contribute substantially to the literature on the utility of HBPM in reducing BP among patients undergoing hemodialysis. Despite these promising findings, the development of HBPM to treat patients with hypertension undergoing hemodialysis is limited [11,45]. Adoption rates among these patients are low [11], with only 2% adhering to standardized BP measurement protocols and 61.1% measuring their BP independently at home [45]. Therefore, it is imperative to advocate for the widespread implementation of standardized HBPM in patients undergoing hemodialysis to improve hypertension management. Information technology under favorable conditions could promote efficient management of BP among patients.

In this study, to systematically investigate knowledge, perception, and adherence to HBPM, improvements in secondary outcomes were observed after a 6-month intervention (Δ_{HBPMKQ}=14.4, SD 8.3; Δ_{HBPMPS}=22.5, SD 9.4; Δ_{HBPMAS}=8.9, SD 6.0), which was consistent with primary hypertension research [46]. In a study by Sun et al [46], a remote interactive approach “Internet+” was used to promote the management of BP, leading to a 50.25% increase in the proportion of timely measurements and a 45.18% increase in high-degree disease awareness after the intervention. In the short term, the web-based HBPM program used health education to improve the patient’s knowledge and improved perception of the benefits, barriers, and self-efficacy associated with HBPM. This, in turn, facilitated the adoption of health-orientated behaviors by patients. In essence, the knowledge, perception, and behavior of the patients about HBPM exhibited a rapid increase and eventually stabilized over a period of time. However, translating these enhanced perceptions and behaviors into favorable health outcomes presented a certain temporal delay, indicating an existing delay. Thus, the introduction of the clinically relevant health management model should consider the overall participation of individuals, as well as the impact on cognition, perception, and behavior of the corresponding health management, to promote the improvement of outcomes.

Limitations

In general, there were several limitations in this trial. First, due to the nature of the health behavior intervention, participant blinding was not feasible. Similarly, health care professionals could not be blinded as the same individuals enrolled participants and conducted follow-ups. Second, due to resource constraints, the study relied on a relatively short follow-up period. Lastly, given the modest sample size, the web-based HBPM program appeared to have improved BP control. However, its impact on health outcomes, such as cardiovascular events or the enhancement of dialysis services, remained uncertain. This underscores the necessity for further research with larger sample sizes across multiple centers to verify the web-based HBPM program’s clinical applicability and reliability in future studies.

Conclusions

The web-based HBPM program introduced in this study produced significant improvements in the predialysis BP control rate and the knowledge, perception, and adherence to HBPM by patients with hypertension undergoing hemodialysis. BP was only significantly decreased at the 6-month follow-up. The web-based HBPM program exhibited benefits in BP control, potentially stemming from improved treatment adherence and optimized prescriptions. This study advocates for the implementation of the web-based HBPM program within the

framework of a hemodialysis unit, as aligned with the prevailing circumstances.

Acknowledgments

We are grateful to all the experts who provided valuable feedback and suggestions for this program, including those from the Division of Rehabilitation at Sun Yat-Sen Memorial Hospital of Sun Yat-sen University; the Division of Cardiology, Nephrology, and Psychology at the Third Affiliated Hospital of Sun Yat-Sen University; as well as the Division of Nutrition at the Fifth Affiliated Hospital of Sun Yat-Sen University. We also thank all health care providers from both The Third Affiliated Hospital of Sun Yat-sen University and The Fifth Affiliated Hospital of Sun Yat-sen University for their collaboration and support in carrying out the project. Finally, we express our appreciation to the patients who participated in this study at the Blood Purification Centres of both hospitals. This research was supported by the Natural Science Foundation of Guangdong Province, China (grant 2018A030313514).

Authors' Contributions

TC, WZ, CW, and JZ designed the whole study. TC and WZ managed the intervention. YC and JY provided the intervention program. QP contributed to data collection. MZ conducted the statistical analysis. TC and WZ prepared the manuscript. JZ and CW supervised the study process. All authors reviewed and edited the manuscript. TC and WZ are joint first authors. CW is the co-corresponding author, with an email address of wangch2@mail.sysu.edu.cn.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Forest plots of the predialysis blood pressure control rate in the interaction of intergroup and research time based on the 3 tested models.
[DOCX File, 106 KB - mhealth_v12i1e53355_app1.docx]

Multimedia Appendix 2
Analysis of the general estimating equations for the effect of the intervention on predialysis blood pressure (n=165).
[DOCX File, 19 KB - mhealth_v12i1e53355_app2.docx]

Checklist 1
CONSORT-eHEALTH checklist.
[PDF File, 91 KB - mhealth_v12i1e53355_app3.pdf]

References

1. Himmelfarb J, Vanholder R, Mehrotra R, Tonelli M. The current and future landscape of dialysis. Nat Rev Nephrol 2020 Oct;16(10):573-585. [doi: 10.1038/s41581-020-0315-4] [Medline: 32733095]
2. Tan R, Kang Y, Dou P, Wang Y, Yang Y. Dietary Guidelines for Adults with Chronic Kidney Disease (2024 Edition): Tongfang Knowledge Network (Beijing) Technology Co., Ltd; 2024:357-362. [doi: 10.19813/j.cnki.weishengyanjiu.2024.03.003]
3. Sarafidis PA, Mallamaci F, Loutradis C, et al. Prevalence and control of hypertension by 48-h ambulatory blood pressure monitoring in haemodialysis patients: a study by the European Cardiovascular and Renal Medicine (EURECA-m) working group of the ERA-EDTA. Nephrol Dial Transplant 2019 Sep 1;34(9):1542-1548. [doi: 10.1093/ndt/gfy147] [Medline: 30007295]
4. Tsikliras N, Georgianos PI, Vaios V, et al. Prevalence and control of hypertension among patients on haemodialysis. Eur J Clin Invest 2020 Oct;50(10):e13292. [doi: 10.1111/eci.13292] [Medline: 32463486]
5. Zhao X, Niu Q, Gan L, et al. Baseline data report of the China Dialysis Outcomes and Practice Patterns Study (DOPPS). Sci Rep 2021 Jan 13;11(1):873. [doi: 10.1038/s41598-020-79531-4] [Medline: 33441625]
6. Georgianos PI, Agarwal R. Blood pressure and mortality in long-term hemodialysis-time to move forward. Am J Hypertens 2017 Mar 1;30(3):211-222. [doi: 10.1093/ajh/hpw114] [Medline: 27661097]
7. Robinson BM, Tong L, Zhang J, et al. Blood pressure levels and mortality risk among hemodialysis patients in the Dialysis Outcomes and Practice Patterns Study. Kidney Int 2012 Sep;82(5):570-580. [doi: 10.1038/ki.2012.136] [Medline: 22718187]
8. Alborzi P, Patel N, Agarwal R. Home blood pressures are of greater prognostic value than hemodialysis unit recordings. Clin J Am Soc Nephrol 2007 Nov;2(6):1228-1234. [doi: 10.2215/CJN.02250507] [Medline: 17942773]

9. Bansal N, McCulloch CE, Rahman M, et al. Blood pressure and risk of all-cause mortality in advanced chronic kidney disease and hemodialysis: the Chronic Renal Insufficiency Cohort study. Hypertension 2015 Jan;65(1):93-100. [doi: 10.1161/HYPERTENSIONAHA.114.04334] [Medline: 25287404]
10. Bansal N, McCulloch CE, Lin F, et al. Blood pressure and risk of cardiovascular events in patients on chronic hemodialysis: the CRIC study (Chronic Renal Insufficiency Cohort). Hypertension 2017 Aug;70(2):435-443. [doi: 10.1161/HYPERTENSIONAHA.117.09091] [Medline: 28674037]
11. Georgianos PI, Vaios V, Sgouropoulou V, Eleftheriadis T, Tsalikakis DG, Liakopoulos V. Hypertension in dialysis patients: diagnostic approaches and evaluation of epidemiology. Diagnostics (Basel) 2022 Nov 26;12(12):2961. [doi: 10.3390/diagnostics12122961] [Medline: 36552968]
12. Agarwal R, Satyan S, Alborzi P, et al. Home blood pressure measurements for managing hypertension in hemodialysis patients. Am J Nephrol 2009;30(2):126-134. [doi: 10.1159/000206698] [Medline: 19246891]
13. Vaios V, Georgianos PI, Vareta G, et al. Clinic and home blood pressure monitoring for the detection of ambulatory hypertension among patients on peritoneal dialysis. Hypertension 2019 Oct;74(4):998-1004. [doi: 10.1161/HYPERTENSIONAHA.119.13443] [Medline: 31401878]
14. Agarwal R, Brim NJ, Mahenthiran J, Andersen MJ, Saha C. Out-of-hemodialysis-unit blood pressure is a superior determinant of left ventricular hypertrophy. Hypertension 2006 Jan;47(1):62-68. [doi: 10.1161/01.HYP.0000196279.29758.f4] [Medline: 16344376]
15. da Silva GV, de Barros S, Abensur H, Ortega KC, Mion D Jr, Cochrane Renal Group Prospective Trial Register: CRG060800146. Home blood pressure monitoring in blood pressure control among haemodialysis patients: an open randomized clinical trial. Nephrol Dial Transplant 2009 Dec;24(12):3805-3811. [doi: 10.1093/ndt/gfp332] [Medline: 19586971]
16. Bansal N, Glidden DV, Mehrotra R, et al. Treating home versus predialysis blood pressure among in-center hemodialysis patients: a pilot randomized trial. Am J Kidney Dis 2021 Jan;77(1):12-22. [doi: 10.1053/j.ajkd.2020.06.014] [Medline: 32800842]
17. General Office of the State Council. Opinions of the General Office of the State Council on promoting the development of “internet+medical health” [Article in Chinese]. Government of China. 2018 Apr 28. URL: http://www.gov.cn/zhengce/content/2018-04/28/content_5286645.htm [accessed 2024-07-30]
18. General Office of the Ministry of Industry and Information Technology. Notice of the General Office of the National Health Commission on further strengthening the capacity building of telemedicine network [Article in Chinese]. Government of China. 2020 Oct 22. URL: http://www.gov.cn/zhengce/zhengceku/2020-11/04/content_5557432.htm [accessed 2024-07-30]
19. Cavero-Redondo I, Saz-Lara A, Sequí-Dominguez I, et al. Comparative effect of eHealth interventions on hypertension management-related outcomes: a network meta-analysis. Int J Nurs Stud 2021 Dec;124:104085. [doi: 10.1016/j.ijnurstu.2021.104085] [Medline: 34601205]
20. K/DOQI Workgroup. K/DOQI clinical practice guidelines for cardiovascular disease in dialysis patients. Am J Kidney Dis 2005 Apr;45(4 Suppl 3):S1-S153. [doi: 10.1053/j.ajkd.2005.01.019] [Medline: 15806502]
21. Random.org. URL: https://www.random.org/sequences/ [accessed 2024-07-30]
22. General Office of the State Council. Notice of the General Office of the National Health Commission on printing and distributing the standard operating procedures for blood purification (version 2021) [Article in Chinese]. Government of China. 2021 Nov 8. URL: http://www.nhc.gov.cn/yzygj/s7659/202111/6e25b8260b214c55886d6f0512c1e53f.shtml [accessed 2024-07-30]
23. Pender NJ, Murdaugh CL, Parsons MA. Health Promotion in Nursing Practice: Prentice Hall; 2006.
24. Williams B, Mancia G, Spiering W, et al. 2018 ESC/ESH guidelines for the management of arterial hypertension: the Task Force for the Management of Arterial Hypertension of the European Society of Cardiology and the European Society of Hypertension: the Task Force for the Management of Arterial Hypertension of the European Society of Cardiology and the European Society of Hypertension. J Hypertens 2018 Oct;36(10):1953-2041. [doi: 10.1097/HJH.0000000000001940] [Medline: 30234752]
25. Reeves PB, Mc Causland FR. Mechanisms, clinical implications, and treatment of intradialytic hypotension. Clin J Am Soc Nephrol 2018 Aug 7;13(8):1297-1303. [doi: 10.2215/CJN.12141017] [Medline: 29483138]
26. Kangkang Shengski Information Technology. URL: https://saas.kang.cn/saas/ [accessed 2024-07-30]
27. Writing Committee of the Chinese Hypertension Committee, Chinese Medical Doctors Association. Home blood pressure monitoring: a consensus document. Chin J Front Med Sci 2012;4(4):43-47. [doi: 10.3969/j.issn.1674-7372.2012.04.012]
28. Writing Group of 2018 Chinese Guidelines for the Management of Hypertension, Chinese Hypertension League, Chinese Society of Cardiology. 2018 Chinese guidelines for the management of hypertension. Chin J Cardiovasc Med 2019;24(1):24-56. [doi: 10.3969/j.issn.1007-5410.2019.01.002]
29. Li Y, Li M, Liu H, et al. Knowledge and compliance status of home blood pressure monitoring and its correlation in patients with hemodialysis combined hypertension [Article in Chinese]. Modern Clinical Nursing 2022;21(4):25-30. [doi: 10.3969/j.issn.1671-8283.2022.04.004]
30. Wang Y, Li K, Li H, et al. Development, reliability, and validity of the Home Blood Pressure Monitoring Adherence Scale for patients with chronic kidney disease. PPA 2020 Oct 12;14:1863-1872. [doi: 10.2147/PPA.S264287] [Medline: 33116431]

31. Miskulin DC, Jiang H, Gul A, et al. Comparison of dialysis unit and home blood pressures: an observational cohort study. Am J Kidney Dis 2021 Nov;78(5):640-648. [doi: 10.1053/j.ajkd.2021.04.013] [Medline: 34144104]
32. Kotanko P, Garg AX, Depner T, et al. Effects of frequent hemodialysis on blood pressure: results from the randomized Frequent Hemodialysis Network trials. Hemodial Int 2015 Jul;19(3):386-401. [doi: 10.1111/hdi.12255] [Medline: 25560227]
33. Ebrahimi H, Safavi M, Saeidi MH, Emamian MH. Effects of sodium concentration and dialysate temperature changes on blood pressure in hemodialysis patients: a randomized, triple-blind crossover clinical trial. Ther Apher Dial 2017 Apr;21(2):117-125. [doi: 10.1111/1744-9987.12506] [Medline: 28185407]
34. Rocco MV, Lockridge RSJ, Beck GJ, et al. The effects of frequent nocturnal home hemodialysis: the Frequent Hemodialysis Network Nocturnal Trial. Kidney Int 2011 Nov;80(10):1080-1091. [doi: 10.1038/ki.2011.213] [Medline: 21775973]
35. Kauric-Klein Z, Peters RM, Yarandi HN. Self-efficacy and blood pressure self-care behaviors in patients on chronic hemodialysis. West J Nurs Res 2017 Jul;39(7):886-905. [doi: 10.1177/0193945916661322] [Medline: 27456461]
36. Moattari M, Ebrahimi M, Sharifi N, Rouzbeh J. The effect of empowerment on the self-efficacy, quality of life and clinical and laboratory indicators of patients treated with hemodialysis: a randomized controlled trial. Health Qual Life Outcomes 2012 Sep 20;10:115. [doi: 10.1186/1477-7525-10-115] [Medline: 22992449]
37. Rodrigues Telini LS, de Carvalho Beduschi G, Caramori JCT, Castro JH, Martin LC, Barretti P. Effect of dietary sodium restriction on body water, blood pressure, and inflammation in hemodialysis patients: a prospective randomized controlled study. Int Urol Nephrol 2014 Jan;46(1):91-97. [doi: 10.1007/s11255-013-0382-6] [Medline: 23340794]
38. Chen Z, Liu J, Sun F, Shen Y, Ma L, Zhou Y. Effect of improving volume overload on home blood pressure variability in hemodialysis patients. Ther Apher Dial 2020 Apr;24(2):163-168. [doi: 10.1111/1744-9987.13427] [Medline: 31381247]
39. Molsted S, Eidemak I, Sorensen HT, Kristensen JH. Five months of physical exercise in hemodialysis patients: effects on aerobic capacity, physical function and self-rated health. Nephron Clin Pract 2004;96(3):c76-c81. [doi: 10.1159/000076744] [Medline: 15056989]
40. Lai B, Shen L, Ye S, et al. Influence of continuity of care on self-management ability and quality of life in outpatient maintenance hemodialysis patients. Ther Apher Dial 2022 Dec;26(6):1166-1173. [doi: 10.1111/1744-9987.13800] [Medline: 35043556]
41. Valsaraj BP, Bhat SM, Prabhu R, Kamath A. Follow-up study on the effect of cognitive behaviour therapy on haemodialysis adherence: a randomised controlled trial. Sultan Qaboos Univ Med J 2021 Feb;21(1):e58-e65. [doi: 10.18295/squmj.2021.21.01.008] [Medline: 33777424]
42. Sheppard JP, Tucker KL, Davison WJ, et al. Self-monitoring of blood pressure in patients with hypertension-related multi-morbidity: systematic review and individual patient data meta-analysis. Am J Hypertens 2020 Mar 13;33(3):243-251. [doi: 10.1093/ajh/hpz182] [Medline: 31730171]
43. Zhang D, Huang QF, Li Y, Wang JG. A randomized controlled trial on home blood pressure monitoring and quality of care in stage 2 and 3 hypertension. Hypertens Res 2021 May;44(5):533-540. [doi: 10.1038/s41440-020-00602-0] [Medline: 33442031]
44. Flythe JE, Chang TI, Gallagher MP, et al. Blood pressure and volume management in dialysis: conclusions from a Kidney Disease: Improving Global Outcomes (KDIGO) controversies conference. Kidney Int 2020 May;97(5):861-876. [doi: 10.1016/j.kint.2020.01.046] [Medline: 32278617]
45. Zheng J, Huang K, Zhang X, Xie Y, Xu M, You T. The current status and affecting factors of blood pressure monitoring at home in maintenance hemodialysis patients with hypertension. J Pract Med 2021;37(16):2147-2151. [doi: 10.3969/j.issn.1006-5725.2021.16.023]
46. Sun GH, Shen MZ, Xu WH, et al. Application of remote “Internet+” interactive mode in the management of patients with hypertension during normalized epidemic prevention and control of COVID-19 [Article in Chinese]. Zhonghua Xin Xue Guan Bing Za Zhi 2021 Nov 24;49(11):1089-1093. [doi: 10.3760/cma.j.cn112148-20210615-00509] [Medline: 34775718]

Abbreviations

BP: blood pressure
DBP: diastolic blood pressure
HBPM: home blood pressure monitoring
HBPMAS: Home Blood Pressure Monitoring Adherence Scale
HBPMKQ: Home Blood Pressure Monitoring Knowledge Questionnaire
HBPMPS: Home Blood Pressure Monitoring Perception Scale
OR: odds ratio
SBP: systolic blood pressure

Edited by L Buis, M Sobolev; submitted 04.10.23; peer-reviewed by A Logan, J Edwards; revised version received 12.02.24; accepted 10.06.24; published 09.08.24.

Please cite as:
Chen T, Zhao W, Pei Q, Chen Y, Yin J, Zhang M, Wang C, Zheng J
Efficacy of a Web-Based Home Blood Pressure Monitoring Program in Improving Predialysis Blood Pressure Control Among Patients Undergoing Hemodialysis: Randomized Controlled Trial
JMIR Mhealth Uhealth 2024;12:e53355
URL: https://mhealth.jmir.org/2024/1/e53355
doi:10.2196/53355

Cardiac Health Assessment Using a Wearable Device Before and After Transcatheter Aortic Valve Implantation: Prospective Study

Rob Eerdekens[1], MSc, MD; Jo Zelis[1], MD, PhD; Herman ter Horst[2], PhD; Caia Crooijmans[3], MD; Marcel van 't Veer[1,4], MSc, PhD; Danielle Keulards[1], MD, PhD; Marcus Kelm[5], MD, PhD; Gareth Archer[6], MD; Titus Kuehne[5], MD; Guus Brueren[1], MD, PhD; Inge Wijnbergen[1], MD, PhD; Nils Johnson[7,8], MS, MD; Pim Tonino[1], MD, PhD

[1]Department of Cardiology, Catharina Hospital Eindhoven, Eindhoven, Netherlands

[2]Philips Research Department, Eindhoven, Netherlands

[3]Department of Cardiology, RadboudUMC, Nijmegen, Netherlands

[4]Department of Biomedical Engineering, Eindhoven University of Technology, Eindhoven, Netherlands

[5]Deutsches Herzzentrum der Charité, Institute of Computer-assisted Cardiovascular Medicine, Berlin, Germany

[6]Department of Cardiology, Sheffield Teaching Hospital, Sheffield, United Kingdom

[7]Weatherhead PET Imaging Center for Preventing and Reversing Atherosclerosis, Houston, TX, United States

[8]Division of Cardiology, Department of Medicine, McGovern Medical School at UTHealth, Houston, TX, United States

Corresponding Author:
Rob Eerdekens, MSc, MD
Department of Cardiology, Catharina Hospital Eindhoven, , Eindhoven, , Netherlands

Abstract

Background: Due to aging of the population, the prevalence of aortic valve stenosis will increase drastically in upcoming years. Consequently, transcatheter aortic valve implantation (TAVI) procedures will also expand worldwide. Optimal selection of patients who benefit with improved symptoms and prognoses is key, since TAVI is not without its risks. Currently, we are not able to adequately predict functional outcomes after TAVI. Quality of life measurement tools and traditional functional assessment tests do not always agree and can depend on factors unrelated to heart disease. Activity tracking using wearable devices might provide a more comprehensive assessment.

Objective: This study aimed to identify objective parameters (eg, change in heart rate) associated with improvement after TAVI for severe aortic stenosis from a wearable device.

Methods: In total, 100 patients undergoing routine TAVI wore a Philips Health Watch device for 1 week before and after the procedure. Watch data were analyzed offline—before TAVI for 97 patients and after TAVI for 75 patients.

Results: Parameters such as the total number of steps and activity time did not change, in contrast to improvements in the 6-minute walking test (6MWT) and physical limitation domain of the transformed WHOQOL-BREF questionnaire.

Conclusions: These findings, in an older TAVI population, show that watch-based parameters, such as the number of steps, do not change after TAVI, unlike traditional 6MWT and QoL assessments. Basic wearable device parameters might be less appropriate for measuring treatment effects from TAVI.

(JMIR Mhealth Uhealth 2024;12:e53964) doi:10.2196/53964

KEYWORDS

aortic valve stenosis; health watch; quality of life; heart; cardiology; cardiac; aortic; valve; stenosis; watch; smartwatch; wearables; 6MWT; walking; test; QoL; WHOQOL-BREF; 6-minute walking test

Introduction

As transcatheter aortic valve implantation (TAVI) for severe aortic stenosis is increasingly used for older adults, including a high percentage of patients with substantial comorbidity, improvement in quality of life (QoL) is as important as extending life expectancy [1,2]. Not all TAVI patients benefit from improved physical activity, as assessed by a 6-minute walking test (6MWT) or the QoL questionnaire [3,4], nor does physical activity change the incidence of aortic stenosis [5]. Nevertheless, improvement in, for example, the baseline 6MWT distance in TAVI studies can be a marker for better survival [6,7]. However, these tests could be influenced by other factors and comorbidities such as peripheral vascular disease for the 6MWT or depression for the QoL questionnaire. Another concern with such tools is that they merely provide a snapshot of a patient's life and might change under different circumstances. Consequently, an unbiased and longer-term tool

to anticipate the benefit from TAVI would allow physicians and patients to personalize treatment and expectations.

In recent years, digital health has begun to transform medicine [8]. Smart phones and health watches, in particular, have found their way into the clinic [9]. These devices can detect atrial fibrillation [10], predict the wearer's 5-year risk of dying [11], and aid in primary prevention to reduce the risk of atherosclerotic cardiovascular disease [12]. The wearable device used in this study, the Philips Health Watch [13], continuously measures physical parameters such as heart rate (HR), number of steps, and amount of physical activity. Combining parameters from the health watch might facilitate a more physiological and comprehensive assessment of functional status before and after TAVI. After intervention for aortic stenosis, patient symptoms often improve (as measured using a QoL questionnaire), but the question arises whether they objectively become more active as measured using a wearable tracker (cq, do patients become more active after TAVI or do they exhibit the same daily routine as that prior to TAVI?). Despite advancements in digital health, current controlled clinical assessments often rely on controlled tests such as the 6MWT and a QoL questionnaire, which may not fully capture the nuances of patients' daily lives.

In this study, we evaluated the change in parameters collected using the Philips Health Watch ("daily activity parameters" such as walking distance) among patients before and after TAVI in comparison to standard clinical and research tests ("controlled environment tests" such as the 6MWT and QoL questionnaire). We hypothesized that after a TAVI procedure, physiological parameters such as step count, total activity time, and daily total energy expenditure (TEE) would increase, whereas respiration rate and HR would decrease.

Methods

This prospective exploratory study sought to identify parameters from the Philips Health Watch (DL8791, Philips) that changed after successful TAVI, and their relationship with standard clinical and research tests (including the 6MWT and QoL questionnaire).

Ethical Considerations

The study complied with the tenets of the Declaration of Helsinki and local regulations. All participants provided written informed consent, and this study was approved by an independent medical ethics committee (MEC-U approval ID: W16.141).

Study Population

Between July 2017 and September 2018, a total of 100 consecutive patients (aged ≥18 years) with severe aortic valve stenosis undergoing a clinically indicated TAVI after the Heart Team's decision were included. Exclusion criteria were immobility and not being able to wear an electronic health watch. All patients were recruited at the Catharina Hospital in Eindhoven, the Netherlands.

Study Protocol

Before TAVI, all patients underwent transthoracic echocardiography, computed tomography for valve sizing and access site evaluation, laboratory testing, and clinical assessment per local protocol. Patients were screened at the outpatient clinic, and eligible and consenting patients received the Philips Health Watch. The watch was placed around the patient's wrist after configuration with patient-specific parameters (height, weight, resting HR, and birth year). It was locked on the time screen, thereby blinding patients from all activity parameters, and worn for a week before being returned for data extraction. TAVI took place within 3-6 months of the baseline assessment. Three months after the TAVI procedure, patients visited the outpatient clinic for follow-up and again wore the Philips Health Watch for 1 week. At baseline and follow-up, a 6MWT and questionnaire (transformed WHOQOL-BREF) were administered [14].

Analysis of the Health Watch Data

The Philips Health Watch is a wrist-worn, photoplethysmography-based, HR and activity monitor (Figure 1). Once per minute, it measures parameters such as HR, respiration rate, step count, and TEE (ie, the number of calories needed to carry out physiological functions such as breathing and physical activity, but excluding the energy required for digesting food) as described previously by Hendrikx et al [13]. Parameters are measured at a 1-Hz sampling rate and stored on the device as 1-minute average values. Data can be extracted via Bluetooth by means of an iPod, using a proprietary iOS application from Philips, and sent to a Philips Research server for use in analyses.

A full report including primary data from the watch and derived parameters consists of a summary averaged over 1 day (Table 1), distributions of HR and respiration rate (Figure 2), and log plot of the HR and TEE (Figure 3). TEE is divided into subcategories of the Metabolic Equivalent Task (MET) scale. As the older TAVI population of this cohort seemed rather inactive, a subdivision of the MET scale was designed: basal activity corresponded to a MET score of 1.5 to 2, light activity from 2 to 3, moderate activity from 3 to 6, and high activity from 6 upwards (we used standard thresholds for the last 2 categories) [15].

Each red dot in Figure 3 represents a particular measurement: the 1-minute average HR and corresponding energy expenditure level. The fitted line quantifies the cardiac energy expenditure slope (CEES): as the HR increases, more energy is needed to maintain the resulting hemodynamic state. When the slope is less steep, more energy is needed to maintain an HR of, for example, 60 beats per minute. Conversely, when the slope is steeper, less energy is needed to maintain the same hemodynamic state. The steepness of the CEES potentially serves as an indicator for the energy efficiency of the cardiovascular system.

The report and concept of CEES were proposed and made available as data derived from the raw data from the health watch by HtH from Philips Research Eindhoven and used in clinical data analysis at Catharina Hospital in Eindhoven.

Figure 1. The Philips Health Watch. The Philips Health Watch is a wrist-worn, photoplethysmography-based, heart rate and activity monitor. It measures parameters such as heart rate, respiratory rate, step count, total energy expenditure, and activity time. Measurements use a 1-Hz sampling rate and are stored on the device as 1-minute average values. Data can be extracted in the hospital via Bluetooth using a proprietary iOS application and sent through Wi-Fi to the Philips Research server for analysis.

Table . Watch data before and after transcatheter aortic valve implantation (TAVI; overall and good responders cohorts).

Parameter	Before TAVI (n=97)	After TAVI (n=75)	*P* value	Good responders before TAVI (n=43)	*P* value[a]
Resting heart rate (1/minute), mean (SD)	62.5 (8.9)	62.3 (8.2)	.88	62.6 (10.3)	.90
Respiratory rate at rest (1/minute), mean (SD)	16.5 (1.9)	16.2 (1.8)	.54	16.5 (2.0)	.70
Heart rate (1/minute), mean (SD)	69.9 (8.3)	69.5 (7.3)	.72	70.2 (9.6)	.95
Heart rate during sleep (1/minute), mean (SD)	63.9 (8.9)	63.3 (8.5)	.57	64.3 (10.0)	.61
Respiratory rate during sleep (1/minute), mean (SD)	16.1 (2.1)	15.9 (2.1)	.92	16.1 (2.3)	.67
Daily percentage of HR[b] observations <60: bradycardia, median (IQR)	10.1 (1.2-34.5)	14.7 (2.4-35.4)	.92	13.3 (0.8-3.0)	.46
Daily percentage of HR observations >100: tachycardia, median (IQR)	1.3 (0.3-2.9)	1.3 (0.5-2.2)	.96	1.0 (0.3-1.0)	.86
Daily total number of steps, median (IQR)	3586 (2607-4946)	4341 (2093-6083)	.36	3633 (2763-5135)	.18
Daily cumulative active energy expenditure (kcal), mean (SD)	718.5 (206.5)	722.6 (226.6)	.98	733.0 (221.4)	.15
Daily cumulative total energy expenditure (kcal), mean (SD)	2313.3 (401.3)	2296.1 (436.6)	.72	2310.7 (428.1)	.19
Slope of log(HR/TEE[c])	0.27 (0.21-0.34)[d]	0.29 (0.23-0.36)[d]	.26	0.26 (0.1)[e]	.04
Daily sleep time (hours)	7.9 (1.8)[e]	10.1 (15.7)[e]	.23	8.2 (7.0-8.7)[d]	.28
Daily basal activity time (minutes), median (IQR)	209 (173-253)	198 (167-262)	.43	209 (173-261)	.94
Daily light activity time (minutes), mean (SD)	183.1 (83.8)	190.3 (100.3)	.63	195.2 (88.8)	.63
Daily moderate activity time (minutes)	48.8 (60.7)[e]	58.3 (51.3)[e]	.10	20.2 (8.8-66.3)[d]	.01
Daily high activity time (minutes), median (IQR)	0.0 (0.0-0.0)	0.0 (0.0-1.5)	.18	0.0 (0.0-0.0)	.19
Daily total active (minutes), mean (SD)	101.9 (91.4)	102.3 (71.9)	.98	108.3 (100.3)	.41

[a]*P* values apply to the comparison between pre- and post-TAVI values in the good responders cohort. See Table S3 in Multimedia Appendix 1 for the full comparison.

[b]HR: heart rate.

[c]TEE: total energy expenditure.

[d]Median (IQR) values.

[e]Mean (SD) values.

Figure 2. Distributions of heart and respiratory rate (overall percentage of heart rate observations <60: 57.9%; overall percentage of heart rate observations >100: 0.3%). This page of the output report from the health watch depicts density plots of the heart and respiration rates during the day and during sleep. It also shows the frequency of resting heart rate and the distribution of activity levels for heart rate observations >100 beats per minute.

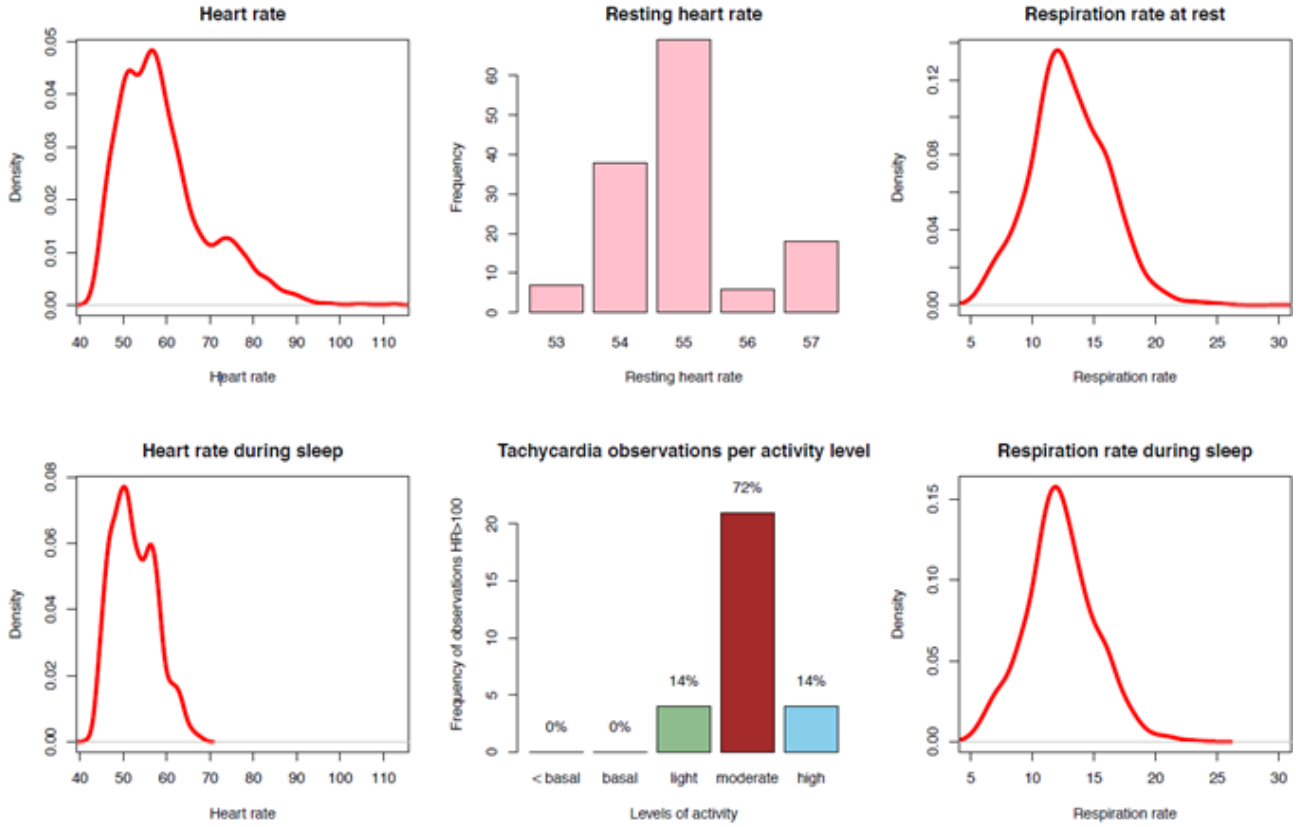
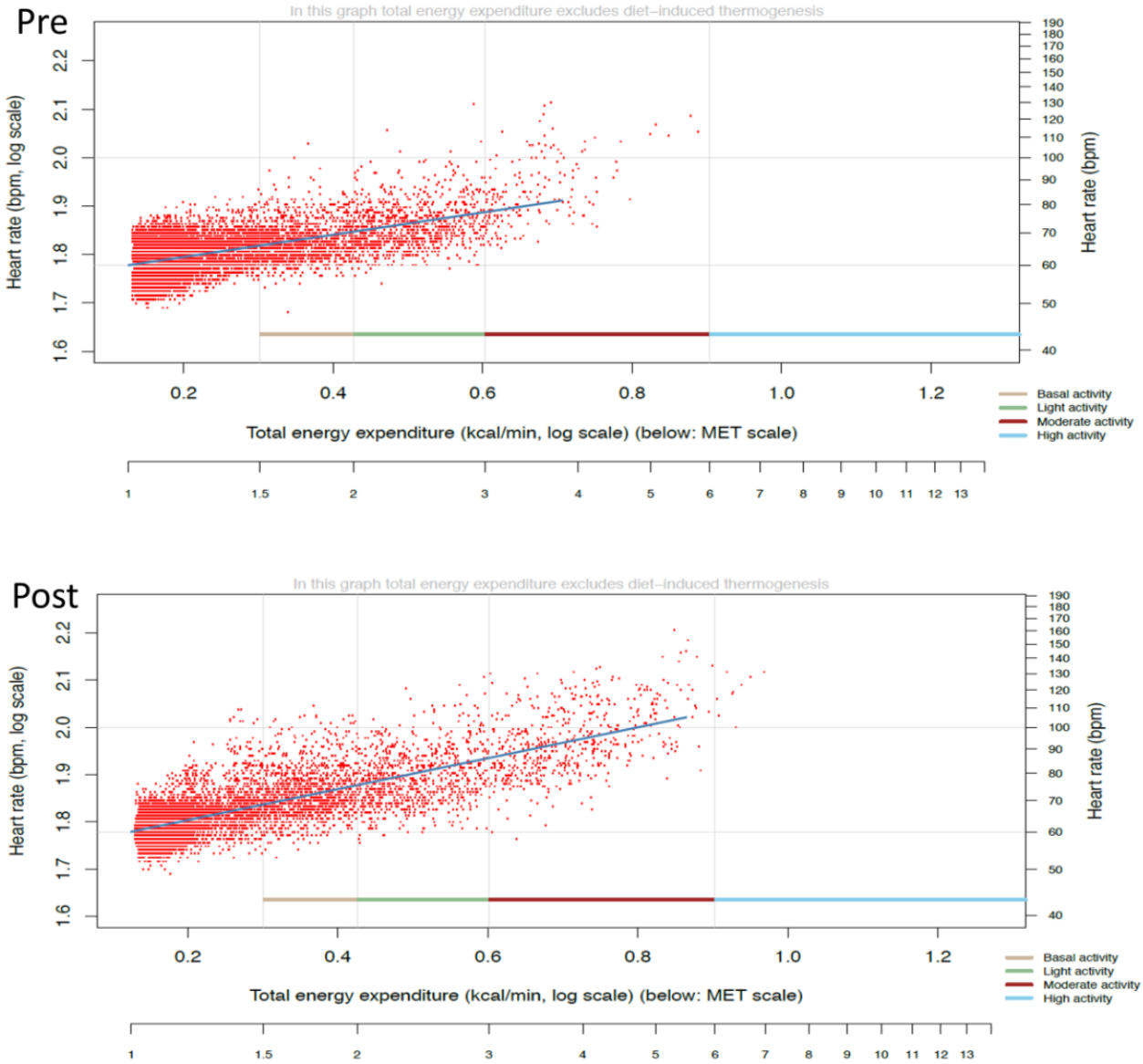


Figure 3. Log plot of heart rate versus total energy expenditure. Another page of the report plots each heart rate and corresponding total energy expenditure on a log plot, divided into subcategories of the Metabolic Equivalent Task (MET) scale. The fitted line relates to the energy efficiency of the cardiovascular system. bpm: beats per minute.



Statistical Analysis

Analyses were performed using SPSS (version 29.0; IBM Corp). The data are displayed as mean and SD values unless stated otherwise. Dichotomous variables are displayed as percentages (%) and absolute numbers (n). Applicable tests were 2-tailed, and $P<.05$ was considered statistically significant. Student *t* tests were used to compare normally distributed variables. Categorical variables were analyzed using the chi-square test, Fisher exact test, or McNemar-Bowker test, whichever was appropriate. As this was an exploratory study, no sample size was prespecified. Analyses were performed for the overall population ("overall cohort"), men versus women ("gender cohort"), those older than and those younger than 81 years ("81 y cohort"), those older than and those younger than 85 years ("85 y cohort"), and a cohort that had an increase of >40 m on the 6MWT after TAVI ("good responders cohort"). The 81-year cutoff is based on the median age, the 85-year cutoff was arbitrary, and the good responders cutoff is based on data from Tuttle et al [16].

Results

Demographic Characteristics

A total of 100 participants were enrolled. Their demographics and medications (before and after TAVI) are displayed in Table 2 and Table 3, respectively. After TAVI, 11 patients died and 11 patients were lost to follow-up. Data extraction for 3 patients failed before TAVI. Complete watch data were thus obtained for 97 patients before TAVI and 75 patients post TAVI. The population consisted of more men (n=57, 57%) than women. Demographic characteristics were representative of a clinical TAVI population with a median age of 81.0 years, NYHA (New York Heart Association) class II or higher in 92% (n=92) of patients, hypertension in 66% (n=66) of patients, and dyslipidemia in two-thirds of patients (n=67, 67%). All patients

fulfilled the criteria for severe aortic stenosis. Procedural data can be found in the supplemental material in Multimedia Appendix 1.

Table . Demographic characteristics at baseline.

Characteristics			Overall cohort (N=100)	Pre-TAVI[a] good responders cohort (n=43)	*P* value[b]
Age (years), median (IQR)			81.0 (76.0-84.0)	81.0 (74.0-84.0)	N/A[c]
Male participants, n (%)			57 (57)	22 (51)	N/A
Risk factors, n (%)					
	Active smoking		8 (8)	5 (12)	.19
	Hypertension		66 (66)	31 (72)	.15
	Dyslipidemia		67 (67)	28 (65)	.36
	Diabetes mellitus		23 (23)	6 (14)	.13
Cardiac history, n (%)					
	Prior myocardial infarction		31 (31)	14 (33)	.86
	Prior PCI[d]		41 (41)	21 (49)	.33
	Prior CABG[e]		22 (22)	12 (28)	.55
Cardiovascular disease, n (%)					
	Cerebral vascular disease		18 (18)	6 (14)	.37
	Peripheral vascular disease		16 (16)	5 (12)	.74
	COPD[f]		27 (27)	13 (30)	.42
	Permanent pacemaker		9 (9)	4 (9)	>.99
Laboratory values					
	Hemoglobin (mmol/L), mean (SD)		8.0 (0.9)	7.9 (0.7)	.42
	hs-cTnT[g] (ng/L) level, median (IQR)		21.0 (14.0-37.8)	23.5 (15.0-34.7)	.63
	NT-proBNP[h] (pmol/L), median (IQR)		1484 (835-3178)	1250 (835-2763)	.79
	Creatinine (mg/dL), median (IQR)		97.0 (77.0-119.0)	101.0 (81.0-118.0)	.93
Echocardiographic parameters					
	Left ventricular ejection fraction (%), median (IQR)		56 (46-63)	56 (42-64)	.62
	AV[i] maximum velocity (cm/second), median (IQR)		424 (381-467)	412 (386-464)	.96
	AV mean pressure gradient (mm Hg), mean (SD)		45 (14)	42 (13)	.97
	AVA[j] (cm^2), mean (SD)		0.8 (0.2)	0.7 (0.1)	.70
Symptoms, n (%)					
	NYHA[k] heart failure class				.15
		I	8 (8)	4 (9)	
		II	26 (26)	6 (14)	
		III	57 (57)	21 (63)	
		IV	6 (6)	5 (12)	
	CCS[l] angina grade≥III		21 (21)	9 (21)	.41
	Syncope		9 (9)	3 (7)	.42

[a]TAVI: transcatheter aortic valve implantation.

[b]*P* values correspond to the comparison between the good responder cohort versus non–good responders before TAVI.

[c]N/A: not applicable.

[d]PCI: percutaneous coronary intervention.

[e]CABG: coronary artery bypass grafting.

[f]COPD: chronic obstructive pulmonary disease.
[g]hs-cTnT: high-sensitivity cardiac troponin T.
[h]NT-proBNP: N-terminal B-type natriuretic peptide.
[i]AV: aortic valve.
[j]AVA: aortic valve area.
[k]NYHA: New York Heart Association.
[l]CCS: Canadian Cardiovascular Society.

Table . Medications.

	Before TAVI[a], n (%)	Post TAVI, n (%)	*P* value
Aspirin	45 (56)	45 (56)	.21
Antiplatelet	28 (35)	72 (90)	<.001[b]
Anticoagulation	38 (47)	34 (43)	.77
Beta-blocker	56 (70)	38 (48)	.26
RAAS[c]	58 (73)	40 (50)	.21
Potassium sparing diuretic	12 (15)	8 (10)	>.99
Statin	61 (76)	41 (51)	.07
Calcium channel blocker	23 (29)	15 (19)	>.99
Nitrates	22 (28)	14 (18)	.15
Alpha blocker	10 (13)	7 (9)	>.99
Antiarrhythmic	5 (6)	5 (6)	>.99
Insulin	12 (15)	7 (9)	.13
Oral diabetic	23 (29)	17 (21)	.63
Loop diuretic	38 (48)	30 (38)	.75

[a]TAVI: transcatheter aortic valve implantation.
[b]*P*<.05.
[c]RAAS: renin-angiotensin-aldosterone system.

Health Watch Parameters

Changes in all health watch parameters before and after TAVI are displayed in Table 1 and Table S3 in Multimedia Appendix 1 for the good responders cohort. Notably, in the total cohort, no parameter changed significantly. For the female cohort, there was a small increase in the light-to-moderate activity time (206.6 minutes before vs 207.3 minutes after TAVI, *P*=.03). For responders younger than 81 years in the good responders cohort, there was an increase in daily moderate activity: 14.2 minutes versus 39.3 minutes (*P*=.02) and 20.2 minutes versus 71.5 minutes (*P*=.01), respectively. A slight decrease was seen in the respiratory rate for responders older than 81 years (16.1 vs 15.1 per minute, *P*=.04). Heart rate, total number of steps, and daily total active minutes did not change after TAVI compared to those before TAVI for the overall group or for the different subgroups. There was no decrease in HR after TAVI despite a trend toward decreased use of beta-blockers (70% before TAVI vs 48% post TAVI, *P*=.26). Univariate analysis of health watch data could not identify a predictor for the good responders cohort (a >40 m improvement during the 6MWT). The results are displayed in Table S2 in Multimedia Appendix 1.

Energy Efficiency of the Cardiovascular System

The CEES—the slope of the fitted line in Figure 3 between HR and TEE on a log scale—serves as an indicator for the energy efficiency of the cardiovascular system. The CEES did not change significantly before versus after TAVI for the overall cohort (*P*=.26) but increased significantly for the good responders cohort (*P*=.04).

6MWT and the Transformed WHOQOL-BREF Questionnaire

The distance on the 6MWT increased after TAVI compared to that before TAVI (342.8 vs 289.7 m, respectively; *P*<.001) both for the total cohort as well as all subgroups. An improvement in the physical limitation score (domain 1 of the questionnaire) could be seen in the overall cohort (from 54.5 to 61.4, *P*=.005); on subgroup analyses, male participants (55.4 vs 62.0, *P*=.03), those younger than 81 years (51.8 vs 61.1, *P*=.001), and those younger than 85 years (54.4 vs 60.8, *P*=.01) showed a similar improvement. Results from the other domains (psychological, level of independence, social relationships, and overall) did not change for the total cohort. However, we detected an improvement in the psychological domain score among participants older than 81 years (67.6 vs 71.7, *P*=.03) and in the overall score among participants younger than 81 years (248.4

vs 267.1, *P*=.009). The parameters in the 6MWT and QoL questionnaire are summarized in Table 4 and Table 5, respectively.

Table . Parameters in the 6-minute walking test (6MWT) before and after TAVI[a].

Parameter	Before TAVI (N=100)	Post TAVI (n=76)	*P* value
Heart rate before the 6MWT (beats per minute), median (IQR)	69 (62 to 79)	70 (61 to 75)	.66
SpO_2[b] before the 6MWT (%), median (IQR)	97 (96 to 98)	98 (97 to 99)	.005
Distance (m), mean (SD)	289.7 (123.0)	342.8 (121.6)	<.001
Heart rate post the 6MWT (beats per minute), median (IQR)	88 (77 to 103)	89 (79 to 98)	.88
SpO_2 after the 6MWT (%), median (IQR)	97 (95 to 98)	97 (94 to 98)	.17
Beats Above Baseline Index, median (IQR)	19 (10 to 28)	20 (12 to 29)	.99
SpO_2 difference (%), IQR	–1 to 1	N/A[c] to 3	.09
Walking speed (m/second), mean (SD)	0.8 (0.3)	1.0 (0.3)	<.001

[a]TAVI: transcatheter aortic valve implantation.

[b]SpO_2: peripheral pulse oximeter saturation.

[c]N/A: not applicable.

Table . Quality of life questionnaire before versus post TAVI[a].

	Before TAVI (N=100)	Post TAVI (n=76)	*P* value
Domain I - physical, mean (SD)	54.5 (18.7)	61.4 (19.7)	.005
Domain II - psychological, mean (SD)	68.1 (14.7)	67.4 (14.4)	.64
Domain III - level of independence, median (IQR)	69 (53-75)	69 (56-75)	.88
Domain IV - social relationships, mean (SD)	71.4 (15.2)	73.5 (15.8)	.31
Overall score, mean (SD)	258.8 (53.6)	268.7 (58.7)	.16

[a]TAVI: transcatheter aortic valve implantation.

Discussion

Principal Findings

To our knowledge, this is the first study in which extensive 1-week physiological data before and after TAVI were assessed using a sophisticated wearable sensor, the Philips Health Watch. Watch parameters such as activity time and step count did not increase after TAVI for the overall cohort, in contrast with significant improvements in distance on the 6MWT (53 m; an 18% increase from baseline) and physical limitation score from the QoL questionnaire (7 points; a 13% increase from baseline). The increase in distance on the 6MWT mirrors the results obtained in a previous randomized trial: 254 m at baseline, 288 m at 30 days, and 297 m 1 year after TAVI [16].

One explanation for our findings is that daily activity does not actually increase in a relatively older population after TAVI, simply because they do not have to or do not want to increase it (ie, lack of necessity or motivation). In such cases, step count and HR are not expected to increase. However, distance on the 6MWT increased significantly after 1 year in a cohort randomized to TAVI compared to no change in those randomized to medical therapy [17], arguing against a Hawthorne effect. A second explanation for our findings is that QoL questionnaires and the 6MWT [4] generate abundant motivation in the hospital setting, which does not correspond with daily life. In this case, the findings on standard tests (including the 6MWT and QoL questionnaires) would appear to have improved, whereas parameters on the health watch would remain unchanged (eg, the daily total number of steps). Thus, both tests may be accurate given their circumstances. A third explanation is that the health watch parameters contain bias or imprecision, which their paired comparison, even in a reasonably sized cohorts such as our own. This seems to be an interesting topic for further research.

Novel metrics such as the CEES may be valuable parameters in the future or in other settings for evaluating cardiac energy efficiency. The steepness of the slope objectively quantifies the relationship between HR and TEE. A less steep slope corresponds with a lower CEES value. In such cases, more energy is needed to maintain HR than when the slope is steeper (and CEES value higher). In this study, there was a significant increase in the CEES for the good responders cohort, which was accompanied by a significant increase in moderate activity time. There was no significant improvement in the CEES or moderate daily activity time in patients who had no or moderate improvement in distance on the 6MWT (<40 m) after TAVI. This novel metric could be used in future research as a tool to identify patient improvement after a TAVI intervention, independent of subjective variables. More research on this metric is warranted.

How to best identify patients whose symptoms benefit from treatment with TAVI remains an important and unanswered clinical question. A post hoc analysis of a randomized trial of TAVI compared to a surgical aortic valve replacement demonstrated that 36% of patients had no change in 6MWT after outcomes after 30 days and 12 months, and 23%-28% of patients demonstrated no improvement in their QoL questionnaire scores (albeit using a different tool than the one used in our study) [16]. When considering an intervention, both procedural risks and economic costs should be balanced against the potential improvement in QoL. Since the patients who commonly qualify for TAVI treatment are relatively older and frailer, an increase in physical performance can be equally or even more important than an increase in life expectancy. With increasing health care costs, the benefit of an intervention should be clear and personalized [18], and for TAVI, this cost-benefit ratio has been disputed [19]. Unfortunately, the overall findings from this study cannot identify patients using the health watch who would be expected to have an above-average response.

The impact of health watches and other sensor technologies on cardiologic care warrants more research. The Apple Heart study [10] found an irregular cardiac rhythm in only 0.52% of over 400,000 people followed up for 8 months, of whom, only 21% completed further testing, and of them, a 34% minority were ultimately diagnosed with atrial fibrillation, accounting for a paltry yield of 153 people, or much lesser than 0.1% of the total. Findings such as these show the inevitable trade-offs between mass testing and the pretest probability of an actionable diagnosis.

Limitations

Not all patients had available follow-up data, including 11 patients who died. However, follow-up and health watch data were complete for >80% of patients. The second health watch measurement was performed 3 months after the procedure as that timing fit best with the local follow-up protocol. It might be speculated that not all patients have completely recovered at 3 months already. However, most studies comparing TAVI to surgical aortic valve replacement have reported good functional improvement in the TAVI cohort at 30 days and at 6 months [20] compared to the surgical aortic valve replacement group, with no further improvement in the TAVI cohort in 1 year. This implies that most patients are already at full capacity at 3 months' follow-up.

Conclusions

This is the first study that compares pre- and post-TAVI outcomes with extensive, 1-week functional assessment with a sophisticated wearable sensor, the Philips Health Watch. Data from the health watch did not register an increase in activity time, total step count, or other parameters after TAVI, whereas the traditional 6MWT and QoL assessment outcomes improved. Watch-based parameters such as these might be less appropriate for measuring treatment effects in the TAVI population. However, our findings relating to the good responder subpopulation suggest that using data such as the CEES parameter derived from wearable device data, might be useful to objectively identify patient improvement after TAVI intervention. This seems to be interesting for further research.

Conflicts of Interest

As this is part of the EURValve project, a European project, we received funding from Horizon 2020. RE, JZ, CC, MvV, DK, MK, GA, TK, GB, and IW report no additional support or industry relationships. HtH is from Philips Research, which provided the health watches used in this study. PT reports research grants from Biosensors, ZonMW (10070012010001), and opSens. NJ received internal funding from the Weatherhead PET Center for Preventing and Reversing Atherosclerosis; outside of this study, NJ and PT have a pending patent on diagnostic methods for quantifying aortic stenosis and TAVI physiology.

Multimedia Appendix 1
Procedural data.
[DOCX File, 1380 KB - mhealth_v12i1e53964_app1.docx]

References

1. Tsevat J, Dawson NV, Wu AW. Health values of hospitalized patients 80 years or older. JAMA 1998 Feb 4;279(5):371. [doi: 10.1001/jama.279.5.371]
2. Kim CA, Rasania SP, Afilalo J, Popma JJ, Lipsitz LA, Kim DH. Functional status and quality of life after transcatheter aortic valve replacement: a systematic review. Ann Intern Med 2014 Feb 18;160(4):243-254. [doi: 10.7326/M13-1316] [Medline: 24727842]

3. Astin F, Horrocks J, McLenachan J, Blackman DJ, Stephenson J, Closs SJ. The impact of transcatheter aortic valve implantation on quality of life: a mixed methods study. Heart Lung 2017;46(6):432-438. [doi: 10.1016/j.hrtlng.2017.08.005] [Medline: 28985898]
4. Hiltrop N, Belmans A, Claes M, et al. Functional performance and quality of life in high-risk comorbid patients undergoing transcatheter aortic valve implantation for symptomatic aortic valve stenosis. Eur Heart J Qual Care Clin Outcomes 2016 Jul 1;2(3):184-192. [doi: 10.1093/ehjqcco/qcw001] [Medline: 29474610]
5. Sarajlic P, Wolk A, Bäck M, Larsson SC. Physical activity does not reduce aortic valve stenosis incidence. Circ J 2018 Aug 24;82(9):2372-2374. [doi: 10.1253/circj.CJ-18-0598] [Medline: 29998916]
6. Green P, Cohen DJ, Généreux P, et al. Relation between six-minute walk test performance and outcomes after transcatheter aortic valve implantation (from the PARTNER trial). Am J Cardiol 2013 Sep 1;112(5):700-706. [doi: 10.1016/j.amjcard.2013.04.046] [Medline: 23725996]
7. Mok M, Nombela-Franco L, Urena M, et al. Prognostic value of exercise capacity as evaluated by the 6-minute walk test in patients undergoing transcatheter aortic valve implantation. J Am Coll Cardiol 2013 Feb 26;61(8):897-898. [doi: 10.1016/j.jacc.2012.10.050] [Medline: 23428221]
8. Sharma A, Harrington RA, McClellan MB, et al. Using digital health technology to better generate evidence and deliver evidence-based care. J Am Coll Cardiol 2018 Jun 12;71(23):2680-2690. [doi: 10.1016/j.jacc.2018.03.523] [Medline: 29880129]
9. Dörr M, Nohturfft V, Brasier N, et al. The WATCH AF trial: smartwatches for detection of atrial fibrillation. JACC Clin Electrophysiol 2019 Feb;5(2):199-208. [doi: 10.1016/j.jacep.2018.10.006] [Medline: 30784691]
10. Perez MV, Mahaffey KW, Hedlin H, et al. Large-scale assessment of a smartwatch to identify atrial fibrillation. N Engl J Med 2019 Nov 14;381(20):1909-1917. [doi: 10.1056/NEJMoa1901183] [Medline: 31722151]
11. Zhou H, Zhu R, Ung A, Schatz B. Population analysis of mortality risk: predictive models from passive monitors using motion sensors for 100,000 UK Biobank participants. PLOS Digit Health 2022 Oct;1(10):e0000045. [doi: 10.1371/journal.pdig.0000045] [Medline: 36812566]
12. Tekkeşin A, Hayıroğlu M, Çinier G, et al. Lifestyle intervention using mobile technology and smart devices in patients with high cardiovascular risk: a pragmatic randomised clinical trial. Atherosclerosis 2021 Feb;319:21-27. [doi: 10.1016/j.atherosclerosis.2020.12.020] [Medline: 33465658]
13. Hendrikx J, Ruijs LS, Cox LG, Lemmens PM, Schuijers EG, Goris AH. Clinical evaluation of the measurement performance of the Philips health watch: a within-person comparative study. JMIR Mhealth Uhealth 2017 Feb 2;5(2):e10. [doi: 10.2196/mhealth.6893] [Medline: 28153815]
14. Development of the World Health Organization WHOQOL-BREF quality of life assessment. The WHOQOL Group. Psychol Med 1998 May;28(3):551-558. [doi: 10.1017/S0033291798006667]
15. Jetté M, Sidney K, Blümchen G. Metabolic equivalents (METS) in exercise testing, exercise prescription, and evaluation of functional capacity. Clin Cardiol 1990 Aug;13(8):555-565. [doi: 10.1002/clc.4960130809] [Medline: 2204507]
16. Tuttle MK, Kiaii B, Van Mieghem NM, et al. Functional status after transcatheter and surgical aortic valve replacement: 2-year analysis from the SURTAVI trial. JACC Cardiovasc Interv 2022 Apr 11;15(7):728-738. [doi: 10.1016/j.jcin.2022.01.284] [Medline: 35393106]
17. Kapadia SR, Leon MB, Makkar RR, et al. 5-Year outcomes of transcatheter aortic valve replacement compared with standard treatment for patients with inoperable aortic stenosis (PARTNER 1): a randomised controlled trial. Lancet 2015 Jun 20;385(9986):2485-2491. [doi: 10.1016/S0140-6736(15)60290-2] [Medline: 25788231]
18. Wammes JJG, van der Wees PJ, Tanke MAC, Westert GP, Jeurissen PPT. Systematic review of high-cost patients' characteristics and healthcare utilisation. BMJ Open 2018 Sep 8;8(9):e023113. [doi: 10.1136/bmjopen-2018-023113] [Medline: 30196269]
19. Neyt M, Van Brabandt H, Devriese S, Van De Sande S. A cost-utility analysis of transcatheter aortic valve implantation in Belgium: focusing on a well-defined and identifiable population. BMJ Open 2012;2(3):e001032. [doi: 10.1136/bmjopen-2012-001032] [Medline: 22561354]
20. Smith CR, Leon MB, Mack MJ, et al. Transcatheter versus surgical aortic-valve replacement in high-risk patients. N Engl J Med 2011 Jun 9;364(23):2187-2198. [doi: 10.1056/NEJMoa1103510] [Medline: 21639811]

Abbreviations

6MWT: 6-minute walking test
CEES: cardiac energy expenditure slope
HR: heart rate
MET: Metabolic Equivalent Task
NYHA: New York Heart Association
QoL: quality of life
TAVI: transcatheter aortic valve implantation
TEE: total energy expenditure

Edited by M Dorsch; submitted 25.10.23; peer-reviewed by M Hayıroğlu, R Rahman; revised version received 26.04.24; accepted 30.04.24; published 03.06.24.

Please cite as:
Eerdekens R, Zelis J, ter Horst H, Crooijmans C, van 't Veer M, Keulards D, Kelm M, Archer G, Kuehne T, Brueren G, Wijnbergen I, Johnson N, Tonino P
Cardiac Health Assessment Using a Wearable Device Before and After Transcatheter Aortic Valve Implantation: Prospective Study
JMIR Mhealth Uhealth 2024;12:e53964
URL: https://mhealth.jmir.org/2024/1/e53964
doi:10.2196/53964

Original Paper

Evidence of How Physicians and Their Patients Adopt mHealth Apps in Germany: Exploratory Qualitative Study

Tanja Schroeder[1,2], BA, MA; Maximilian Haug[2], BEng, MSc; Andrew Georgiou[1], BA, DipArts, MSc, PhD; Karla Seaman[1], BPharm, MSCE, PhD; Heiko Gewald[2], BSc, MBA, EMBSc, PhD

[1]Centre for Health Systems and Safety Research, Australian Institute of Health Innovation, Macquarie University, Sydney, Australia

[2]Institute for Digital Innovation, Faculty of Information Management, Neu-Ulm University of Applied Sciences, Neu-Ulm, Germany

Corresponding Author:
Tanja Schroeder, BA, MA
Centre for Health Systems and Safety Research
Australian Institute of Health Innovation
Macquarie University
Level 6, 75 Talavera Road
Sydney, 2109
Australia
Phone: 61 2 9850 ext 6281
Email: tanja.schroeder@mq.edu.au

Abstract

Background: The enactment of the “Act to Improve Healthcare Provision through Digitalisation and Innovation ” (Digital Healthcare Act; *Digitale-Versorgung-Gesetz* [DVG]) in Germany has introduced a paradigm shift in medical practice, allowing physicians to prescribe mobile health (mHealth) apps alongside traditional medications. This transformation imposes a dual responsibility on physicians to acquaint themselves with qualifying apps and align them with patient diagnoses, while requiring patients to adhere to the prescribed app use, similar to pharmaceutical adherence. This transition, particularly challenging for older generations who are less skilled with technology, underscores a significant evolution in Germany’s medical landscape.

Objective: This study aims to investigate physicians’ responses to this novel treatment option, their strategies for adapting to this form of prescription, and the willingness of patients to adhere to prescribed mHealth apps.

Methods: Using an exploratory qualitative study design, we conducted semistructured interviews with 28 physicians and 30 potential patients aged 50 years and older from August 2020 to June 2021.

Results: The findings reveal several factors influencing the adoption of mHealth apps, prompting a nuanced understanding of adoption research. Notably, both physicians and patients demonstrated a lack of information regarding mHealth apps and their positive health impacts, contributing to a deficiency in trust. Physicians’ self-perceived digital competence and their evaluation of patients’ digital proficiency emerge as pivotal factors influencing the prescription of mHealth apps.

Conclusions: Our study provides comprehensive insights into the prescription process and the fundamental factors shaping the adoption of mHealth apps in Germany. The identified information gaps on both the physicians’ and patients’ sides contribute to a trust deficit and hindered digital competence. This research advances the understanding of adoption dynamics regarding digital health technologies and highlights crucial considerations for the successful integration of digital health apps into medical practice.

(JMIR Mhealth Uhealth 2024;12:e48345) doi:10.2196/48345

KEYWORDS

mobile health apps; DiGA; adoption; prescription; mHealth; aging and individual differences

Introduction

Mobile devices have enormous potential to enhance the way patients receive medical care and health education [1]. Mobile health (mHealth) is a dynamic and expanding area of health care with short innovation cycles [1,2]. mHealth is closely related to telemedicine and eHealth. The fundamental difference lies in how the related services are delivered to the patient. According to the World Health Organization (WHO), mHealth expands the spectrum of functionalities to include a mobile component. Thus, mHealth is delivered through any mobile

device, from simple mobile phones, smartphones, and tablets to wearable devices used in health care settings [2].

To accommodate the growing number of mHealth apps (ie, software products for smartphones designed to support good health), Germany passed the "Act to Improve Healthcare Provision through Digitalisation and Innovation" (Digital Healthcare Act; *Digitale-Versorgung-Gesetz* [DVG]) in December 2019 and the Digital Health Applications Ordinance (*Digitale-Gesundheitsanwendungen-Verordnung* [DiGAV]) in October 2020 [3,4]. This ordinance enables physicians to prescribe mHealth apps (in German referred to as *Digitale Gesundheitsanwendungen* [DiGAs; digital health applications]) to their patients in the same way as any other medicine. A DiGA supports the recognition, monitoring, treatment, and alleviation of diseases, injuries, or disabilities [5]. To qualify as a DiGA, an mHealth app has to successfully pass the evaluation of the Federal Institute for Drugs and Medical Devices (*Bundesinstitut für Arzneimittel und Medizinprodukte*; BfArM). Only then will an app be included in the list of reimbursable digital health apps, the so-called DiGA directory. In September 2023, this directory comprised 48 admitted DiGAs [6]. DiGAs cover a wide array of the medical spectra. Popular examples are "Endo-App" (to treat endometriosis), "Kaia COPD" (for chronic obstructive pulmonary diseases), or "sinCephalea" (for the treatment of migraine).

The introduction of the "Act to Improve Healthcare Provision through Digitalisation and Innovation" represents a significant and innovative change to the German health care system [7]. The new legal framework elevates DiGAs to the ranks of medical devices [5]. Therefore, if physicians want to prescribe a DiGA, they are required to thoroughly inform themselves about which apps can help their patients, how they are to be used, and how a positive influence is expected to manifest. Patients are required and need to be able to adhere to the prescription; for example, they have to use the DiGA as stated by the physician. Thus, the physician needs to assess (implicitly or explicitly) whether the patient is likely to use the DiGA as prescribed. This does not only include adherence to, for example, training intervals (for orthopedic DiGAs) but also the general technology savviness of the patient, for example, whether the patient is able to download the app, install it on the smartphone, maintain updates, and so on. This question is specifically challenging when it comes to older users, who are often regarded as being not technology savvy.

Consequently, this innovation raises several questions for technology adoption research. Typically, adoption research concerns the individual user's decision whether to use technology, either mandatory [8,9] or voluntary [10]. Now, a concerned third party (the physician) decides on behalf of the user whether the patient is expected to be willing and able to use an app on their smartphone voluntarily. Therefore, the physician's assessment now includes not only the medical relevance of the DiGA but also whether the user might be able and willing to use it as prescribed. Although the latter assessment seems easy for younger people, the case is much more difficult for older patients. Potential doctor misperception raises concerns about a possible digital divide and ageism by doctors. A recent report by the WHO and the United Nations (UN) raises awareness of this issue and urges action to combat ageism, as it negatively impacts well-being and can lead to premature death and higher health care costs [11].

In the health care context, the resulting research questions are specifically relevant, as the influence of age on the adoption of mHealth apps has not yet received sufficient attention in the scientific discourse [12]. Therefore, this research addresses the complementary research questions:

1. What factors enable or hinder physicians to prescribe DiGAs?
2. What factors enable or hinder older users' adoption of DiGAs?

Methods

Study Design

We developed an exploratory qualitative study design to answer the research questions and gain insights through semistructured interviews with representatives of the 2 relevant stakeholder groups: physicians and patients. The study adhered to the standards for reporting on qualitative research [13].

Study Setting

The data for our research were gathered in Germany.

Participants

The first qualitative study was conducted from August 2020 to June 2021. We interviewed 28 physicians (demographics are provided in Table 1) to assess how DiGAs can improve the health of their patients. Physicians were chosen as research objects because of their unique role as prescribers of DiGAs. The interviewees had different backgrounds in terms of IT affinity and previous experience with mHealth apps in general.

Table 1. Demographics of the physicians.

Code	Specialty	Sex	Age (years)
E1	General practitioner and cardiologist	Male	63
E2	Dermatologist	Male	67
E3	Dermatologist	Male	45
E4	Urologist	Male	48
E5	General practitioner	Female	63
E6	Dermatologist	Female	54
E7	Dermatologist	Male	41
E8	Internist and gerontologist	Male	38
E9	Orthopedist	Male	53
E10	General practitioner	Male	64
E11	Geriatric therapist	Female	58
E12	Urologist	Male	38
E13	General practitioner and pain therapist	Male	59
E14	Assistant doctor cardiology	Male	35
E15	Urologist	Male	42
E16	General practitioner	Male	68
E17	Dermatologist	Female	36
E18	General practitioner	Female	38
E19	General practitioner	Female	59
E20	General practitioner	Female	41
E21	General practitioner	Female	45
E22	General practitioner	Male	45
E23	Neurologist	Male	44
E24	General practitioner	Male	67
E25	Neurologist	Male	65
E26	Neurologist	Male	42
E27	Molecular neurologist	Female	45
E28	General practitioner	Female	37

In the second study, we interviewed patients (ie, the potential users of a DiGA). To reflect the specific issues of aging patients, we chose interview partners over 50 years of age. Evidence shows that from this age onward, chronic diseases increase significantly [14]. Therefore, this age group is likely to represent a large part of the target group for the prescription of DiGAs. Furthermore, studies suggest that there are still age disparities in attitudes toward technology and that the aging population is often less comfortable using technology [15].

We conducted 30 interviews to determine the factors that influence potential patients to adopt DiGAs or not. Demographics of the interviewees are given in Table 2.

Table 2. Demographics of the patients.

Code	Sex	Age (years)
P1	Male	68
P2	Female	60
P3	Female	57
P4	Female	76
P5	Female	56
P6	Female	65
P7	Male	69
P8	Female	64
P9	Male	68
P10	Female	66
P11	Male	65
P12	Male	67
P13	Female	57
P14	Female	72
P15	Female	67
P16	Male	53
P17	Female	61
P18	Female	69
P19	Female	61
P20	Male	63
P21	Male	65
P22	Female	68
P23	Female	67
P24	Female	61
P25	Female	67
P26	Female	59
P27	Male	64
P28	Female	69
P29	Female	64
P30	Female	68

Recruitment

We recruited physicians by telephone from a community-based physician network in Germany and via social media. For the interviews with patients, we promoted our study to doctors and approached medical centers and clinics. To identify suitable interview partners (ie, *potential* patients), we presented the research project to local sports, communication, and civic groups; promoted the study on various social media platforms; and spoke to the local press. In this way, we motivated suitable candidates to contact us. The participation of all study participants, both physicians and patients, was voluntary.

Data Collection

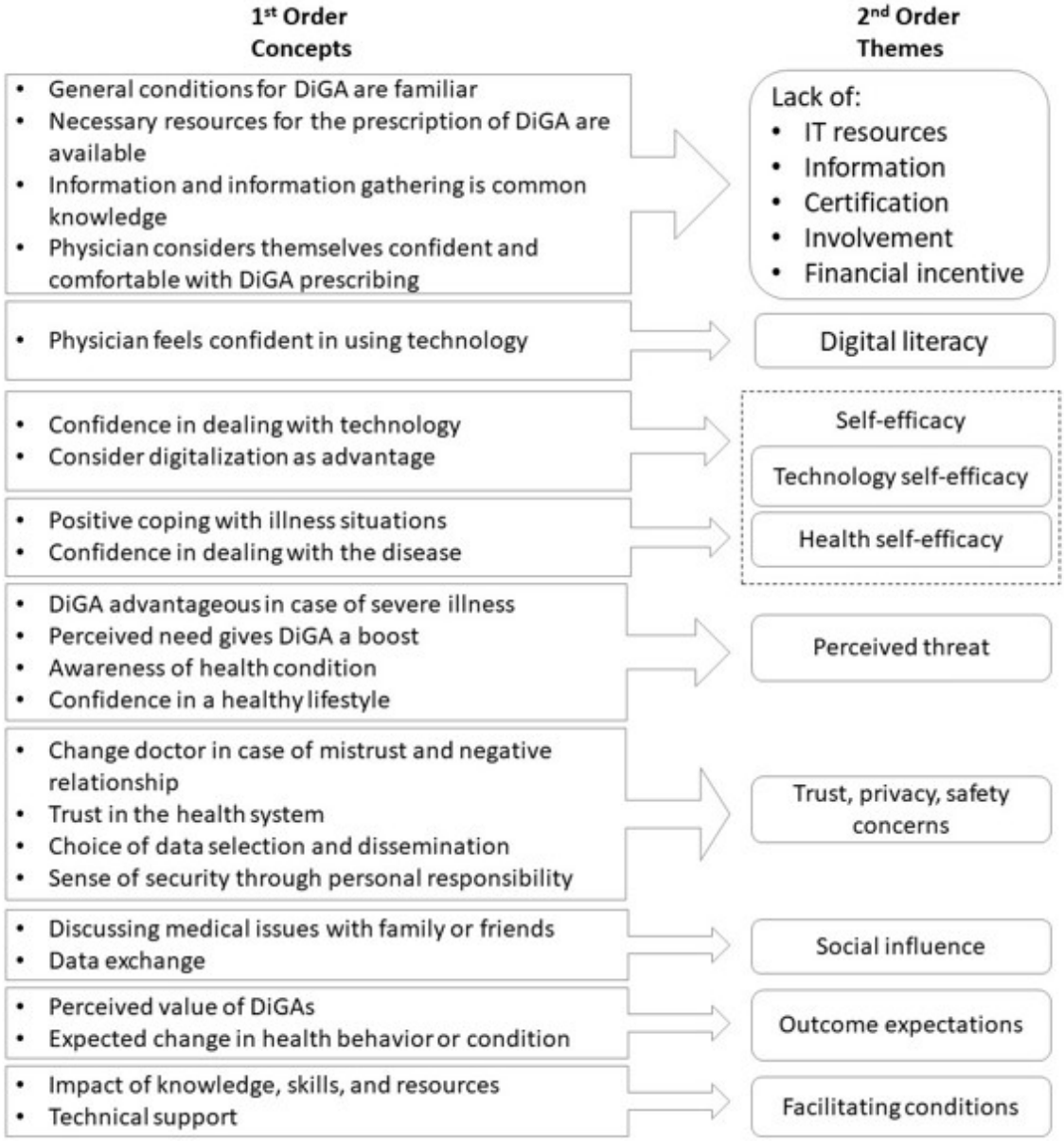
All interviews were semistructured and led by a list of questions and general topics that the interviewers were supposed to address. The semistructured interview guidelines are provided in Multimedia Appendix 1. As a structure and topic basis for creating the questions, the constructs from well-known technology adoption research models (unified theory of acceptance and use of technology [UTAUT] [9]) and models for the analysis of health behavior (health belief model [HBM] [16]) were considered. The questions were primarily open and allowed the interviewees to explore their experiences and views. Supported by a systematic and comprehensive interview process, the interviewers had high degrees of freedom to conduct the interview in order to gain deeper insights. The questions were adjusted correspondingly for the following interviews to gain

more profound knowledge for each interview. Interviews lasted around 20-45 minutes and were conducted face-to-face or over the phone by 1 researcher (TS). The interviews were conducted in German and were recorded, transcribed, and translated into English for further analysis [17].

Data Analysis

For the coding process, we used the NVivo10 software (QSR International). The research was conducted using an interpretative phenomenological analysis (IPA) [18]. IPA aims to examine the world perspective of the participants and, if possible, take an "insider perspective" [18] of the phenomenon under investigation. At the same time, IPA recognizes that research is a dynamic process. In parallel with the data collection, we scanned and coded the data from the first round of analysis. One researcher (TS) conducted a thematic analysis to identify patterns and themes. After analyzing the first 3 interview transcripts, 2 researchers (MH and TS) developed a coding framework using an inductive approach that allowed for identification of predominant themes. All emerging themes were cross-checked and discussed within the whole research team and developed iteratively to ensure definition and reliability. In the process, commonalities and differences between the respective perspectives were identified. This led to 9 different key themes concerning benefits and barriers, influencing factors on the intention to adopt a DiGA, and the outcome expectation from different perspectives as the main areas. Figure 1 lists the corresponding coding scheme according to Gioia et al [19]. The aim was to understand the specific properties of these areas and the influence of these factors in the context of the introduction of DiGAs.

Figure 1. Coding scheme using the methodology by Giola et al. DiGA: Digitale Gesundheitsanwendungen (digital health application).



Ethical Considerations

This study did not require ethical approval according to the guideline of the applicable Ethics Committee of the Bavarian Universities [20], as no risks or harm was brought forward to the participants. All participants received an information and consent form explaining the requirements for participation. This included the opportunity to have the form explained to them if needed. All interviewees gave written or verbal consent before the interview started. Data collection, storage, and analysis were conducted in adherence to European Union General Data Protection Regulation (EU-GDPR). None of the participants were compensated.

Results

Overview

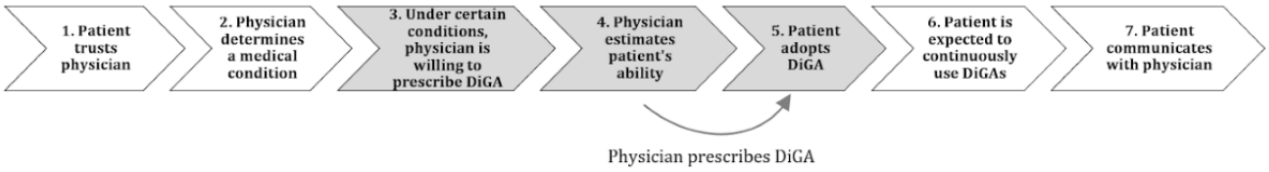
Our data demonstrate that only 4 of the interviewed physicians already prescribed DiGAs and none of the patients had used a DiGA before (Table S1 in Multimedia Appendix 1). Nevertheless, most physicians have experienced mHealth apps themselves. These were mostly used for medical information in their professional routine or privately for personal fitness and nutrition goals.

On the patient side, 20 participants had already had experience with mHealth apps or were actively using them (Table S2 in Multimedia Appendix 1). In this context, only fitness and nutrition apps were mentioned. These apps have a preventive character to help people to consciously lead a healthier and more active life. Most of the apps mentioned were those used in conjunction with smartwatches to measure activity, such as pedometers.

The DiGA Adoption Process of Physicians and Patients

The analysis indicates that for DiGAs in particular, additional steps are added to the traditional prescribing process. Because of the special role of the doctor as a gatekeeper, the doctor is the first to decide whether the patient is suitable for a DiGA. First, the familiar steps such as the patient's trust in the doctor and the treating doctor's determination of the patient's medical condition represent the keystone of the process. Following these steps, when prescribing a DiGA, the next steps are the doctor's consent to prescribe a DiGA under certain conditions and the assessment of the patient's ability to use a DiGA. Thereafter, the doctor prescribes a DiGA, and the patient is given the opportunity to adopt it. The process ends with the expectation that the patient will continuously use the prescribed DiGA and report the results to the treating physician. This process is visualized in Figure 2, with the doctor and the patient being influenced by different factors that determine whether the DiGA is prescribed (doctor's perspective) and accepted (patient's perspective).

Figure 2. The adoption process for DiGAs by the physician and patient. DiGA: Digitale Gesundheitsanwendungen (digital health application).



Factors Enabling or Hindering Physicians to Prescribe DiGAs

Overview

Our results indicated that DiGAs are not yet widely known and used at this early stage by the physicians we interviewed. They indicated that the level of information is still insufficient and that very few DiGAs are prescribed compared with the prescription of drugs. DiGAs represent a completely new and innovative approach, so entry challenges are not considered unusual. Nevertheless, these challenges are initially barriers to prescribing DiGAs according to our participants. We were able to identify the following factors that influence physicians to prescribe a DiGA or not.

Lack of IT Resources

The lack of IT resources was mentioned by the interviewed doctors as a barrier to prescribing DiGAs within the framework of the DVG. Numerous efforts have been made toward digitalization. These include networking with various players in the German health system, such as doctors, hospitals, pharmacies, and health insurance companies. So far, only limited resources are available for the implementation of these plans. The doctors interviewed describe that, for example, the provision of services for an electronic health record (EHR) could help to digitalize various processes and information. They see the benefit of a DiGA for their work as low as long as the patient files and the exchange between other actors in the health care system are not fully digitized. The introduction of an EHR in Germany has failed so far because of technical challenges.

Lack of Information

Another reason often mentioned in connection with rejection was a lack of information. Physicians often do not know which DiGAs are available and where they could get the necessary information. As a result, physicians are often reluctant and skeptical about prescribing DiGAs ("No, I have not been educated on what DiGAs are available, how to prescribe, and how to tell if a DiGA is effective. I would not know where to look for this information" [E2, E3, E8, E12, E16, E20, E21, E25, and E28]; "I have little time to search for all the information I need to prescribe a DiGA in good conscience" [E1, E4, E8, E15, E16, E20, and E21]). They tend to be negative out of concern for malpractice and the resulting liability risk. Likewise, the physicians described the requirements for evidence of a DiGA (requirements for security, functionality, data protection, information security and quality, and positive effects on care) in the assessment procedure of a DiGA as insufficient. Along with this concern is the fact that for provisionally included DiGAs, evidence-based studies on the benefit of the

app are not yet available, and the apps are therefore only provisionally included in the directory. A comprehensive explanation of the BfArM authorization process of a DiGA can counteract these problems. However, participants reported receiving little information from insurers, DiGA manufacturers, or the BfArM. Physician respondents stated that proactive communication from DiGA providers was limited and possible involvement in the development process was unknown ("These apps have often been developed without the support of physicians and are now being 'fast-tracked' to market—that is not exactly building trust" [E25]). This statement reflects a high level of mistrust. In contrast, one participant stated that he worked with an app development company to develop an app in his field and described that this approach was extremely helpful to reduce the information deficit, mistrust, and gain more confidence. This shows that involvement is possible, but that an exchange of information is necessary.

Lack of Certification

In medical circles, a lack of certification has been widely mentioned. Half of all interviewed physicians agree that DiGAs, similar to other medical products (eg, drugs and medical aids such as wheelchairs or bandages), should be provided with a known official certification to ensure more trust in the product ("I think a well-known and recognized certification in medicine could help to build confidence in DiGAs" [E2, E13, E15, E17, E26, and E28]). Certification would make it easy to recognize that the DiGAs are safe as a medical device and are also medically or technically suitable within the scope of the intended purpose stated by the manufacturer. A certification equivalent to that for medicines could be conceivable in this context.

Lack of Involvement

Another aspect that our interviewees criticized was the lack of involvement after the prescription of a DiGA. They described that they are little involved in the procedure after prescribing a DiGA and in many cases currently only take on the role of a "prescriber" ("I would like to be involved in the whole process from the manufacturing to the evaluation of the data with the patient" [E5]). This shows that the participating physicians apparently make a great distinction between a traditional medicine in the form of medication administration and the use of digital components. After prescribing a traditional medicine, the physician only gets feedback when the patient comes back after some time and tells them how the medicine works. But with digital options such as DiGAs, the requirement is now higher: participants demanded an adjustment of the involvement in the postprescription process. A preview of future digitization plans shows that the involvement of physicians in the digital feedback process will be considered.

Lack of Financial Incentives

Participants further stated that counseling for a DiGA is much more time-consuming than for medicines, but the monetary incentive is not there. As a result, we identified a lack of financial incentive. Financial pressure weighs on the physicians in this regard, which is not compensated for by health insurance companies. As a result, the incentive for prescribing (€2 [US $2.16] per prescription) and treatment support (eg, successful monitoring) is currently considered too low ("2 euros for prescribing or 7 euros for treatment support are in no way an incentive to prescribe a DiGA" [E2]).

Finally, the study also found differences in physicians' skills, knowledge, and attitudes toward digital technology. We define this factor as digital literacy, which has 2 specific characteristics. Some of the surveyed physicians stated that they have the impression that a physician is either digitally interested and very open-minded or completely ignorant of new innovations, so that even educational conversations might fail ("Either you find it good as a physician and have dealt with it once or recommended it to your patients. Or you ignore it at first" [E6]). Along with this result, the physicians interviewed expressed the fear that they might lack knowledge, for example, when it comes to patients' technical questions or that they would have to become a kind of "technical support" in the event of problems in this area or in the evaluation of DiGA analyses. Others, however, considered the introduction of DiGAs to be an advantage for their work and less of a hindrance or a problem.

Factors Enabling or Hindering Older Users' Adoption of DiGAs

Overview

As indicated above, a process of influencing factors leads to the adoption of DiGAs by patients. Thus, if the physician determines a medical condition, is willing to prescribe a DiGA, and considers the patient's ability to use a DiGA to be positive, the physician will eventually prescribe a DiGA so that the patient will have the opportunity to adopt it.

Lack of Information

Similar to the physicians, we also identified a lack of information on the patient side. The DiGA concept is rather rarely known by the patients, and experiences were only described in 3 of the 30 interviewed participants. Nevertheless, from the patient perspective, DiGAs are recognized as an innovative and profitable treatment option that can be carried out independently of the time and place of the doctor's visit ("I want to have the flexibility to do my therapy when it suits me" [P1, P6, P9, P12, P18, P26, P29, and P30]). Patients described DiGAs as a helpful "bridge" and a refresher or repetition of therapy content, especially for patients waiting for an appointment with a specialist (eg, psychotherapist). Some patients pointed to a long history of illness and low chances of success of conventional therapies and considered DiGAs as another treatment option. The interviews thus confirmed that the population's willingness for DiGAs is high. However, many patients lack further information ("My physician or my health insurance company haven't informed me about it yet—how should I know?" [P2, P3, P12, P16, P19, P22, and P29]).

The factors influencing the adoption of DiGAs are discussed in more detail below. In this context, we are oriented toward the most well-known models in adoption research: the UTAUT [9], a model from information systems that measures the acceptance of a technology by users to gain access to individual user behavior, and the HBM [16], a theoretical model from health psychology that analyzes and predicts health-related behavior.

Self-Efficacy

Self-efficacy refers to a sense of control over one's environment and behavior. Participants considered self-efficacy important but also taken for granted. A distinction was made between technology self-efficacy about the DiGA and health self-efficacy with a health aspect. We define technology self-efficacy as the patient's belief in their competence to use the DiGA. Health self-efficacy here describes the patient being confident in managing their health. Participants expressed confidence in their technology self-efficacy. For all participants, everyday use of smartphones and various apps was normal and regular. Some indicated that they still prefer paper calendars or dislike apps for health prevention but saw digital documentation and treatment support as an advantage in case of possible illness ("I'm more used to pen and paper, but if you get used to it, it certainly has its advantages" [P1, P4, P12, P14, P21, and P29]). In terms of health self-efficacy, participants were mostly self-confident. No participant stated that they did not want to deal with illnesses or would rather not know how healthy or ill their body was. Most participants seemed to have a high level of health literacy and wanted to actively deal with their health ("Yes, I would rather be the person who would then say, I would like to know this to know my enemy and then fight against it" [P19]). Participants indicated that a personal sense of control facilitates health behavior change ("Then an app like that would be great because you can calm down a bit more and have more security for yourself" [P10]). Participants confirmed that they focus on the opportunities rather than the obstacles ("Yes, I usually don't go right away, but after a short scare I always face all the problems, so I'm more for problem solving rather than suppression" [P3]).

Perceived Threat

Perceived threat was mentioned as a crucial central factor for the use of a DiGA in the context of an impending or existing chronic disease. According to the HBM, perceived threat stems from beliefs about perceived susceptibility to disease and the perceived severity of disease consequences [16]. Susceptibility here refers to a person's risk of contracting a disease. Severity refers not only to the medical consequences but also to the potential impact of an illness on a person's daily life, family life, and social relationships. Participants indicated that a threat has a strong influence on health behavior. They explained that if they were seriously ill, they would use any means to support treatment, both digital and analogue ("So, I would say if I got sick now, I would be very interested in a DiGA already" [P29]; "Yes, when I get sick, I try everything possible to get better, whether digital or not" [P3]). Nevertheless, participants indicated that they generally felt very body and health conscious ("I eat healthy and exercise as much as I can, that's part of my everyday life" [P9, P11, P15, P16, P21, P23, P26, and P29]). Only 1 respondent indicated that they felt an increased fear of possible illness. All participants indicated that they were primarily concerned with their health and well-being and that preventive measures were a natural part of their lives ("I prefer to focus positively on my health instead of worrying daily about illness and negative health" [P23, P29, and P31]).

Trust in the Physician, Attitudes Toward Privacy, and Safety Concerns

Partly different from classical adoption research, trust in the physician, attitudes toward privacy, and safety concerns were mentioned as key characteristics. Due to the medical field, the existing models require expansion. Trust does not describe a direct influencing factor but rather a precondition. For the participants, a deep trust relationship with their physician was crucial to consult a physician in case of a health problem and to receive good medical treatment. If this precondition is not given, an exchange about digital treatment options does not take place ("If I feel that I cannot trust him, then I would change" [P1, P7, P10, P13, and P22]; "The most important thing is the trust relationship. If I don't trust my physician, I change physicians" [P27]). Subordinated are the characteristics of privacy and security. All participants considered it important to be able to set privacy settings themselves. The vast majority said that they did not particularly care about the content of privacy settings, but wanted to decide for themselves who could access which health data ("I would like to know who knows what about me" [P6, P7, P8, P15, and P26]; "There should be settings options. I don't want to transmit everyday occurrences" [P9, P10, P11, P13, and P27]). Based on the accreditation of a DiGA as a medical device and the assessment by the BfArM, they have no concerns about the safety of DiGAs and feel confident in using DiGAs ("I trust that our federal system is highly secure" [P11]). On the basis of these correlations, we categorized the 3 concepts together. We rank the trust factor as the most important, as the privacy and security factors can be mitigated by a high level of trust.

Social Influence

The concept of social influence reflects the effect of environmental factors, for example, the opinion of friends and family, and is a significant factor in traditional adoption research [21]. However, in this study, we could not find any relevant results. Here, the medical context seems to have an important role. Most participants indicated that they did not discuss their health behaviors, personal diagnosis, or treatment plans with their social circle ("I don't want to share all health data, including the fact that I use a health app, with other people" [P9]). We suspect that social influence may be a factor that is difficult to capture due to the sensitive nature of the data.

Outcome Expectations

Outcome expectations are defined as the expected consequences of a certain health behavior, which can be negative or positive [22], that is, what does the patient expect from using a DiGA. Participants stated that they would be able to reduce the frequency of visits to the physician, thus saving travel and time; that an existing illness would be better monitored by the physician; that they would feel safe and well cared for as a result; and that they would receive health-promoting treatment in the form of the DiGA. However, these expectations were tied to the continued use of features of the DiGA. Participants agreed that a DiGA must be simple to use, regardless of age. Many participants expect the app to provide more detailed information about the disease, symptoms, medication, and contraindications. However, this information needs to be understandable,

meaningful, and informative for every patient, regardless of age, education level, and professional background ("It should be easy for me, and I should be able to understand and comprehend it. The ease of use." [P1, P5, P16, P19, and P21]).

Facilitating Conditions

Eventually, these expectations also cross over to the facilitating conditions, which includes the impact of the patient's knowledge, skills, and resources. Here, the participants considered technical support to be particularly important. This was not because they felt unsure about using the technology but rather because it was a new type of health intervention. Some participants considered the physician to be an appropriate point of contact when difficulties arose in a few cases. Other participants stated that if they had difficulties or questions, they would ask their family and friends for help, as they do with other technical matters ("I think it would be good to have a number that I can call and that can help me" [P2]; "Well, I trust my son, he knows me anyway. And he would also know the diseases I have" [P18]). In summary, this study distinguishes between technical infrastructure and health support, with the health aspect (eg, knowledge and understanding of the diseases and their treatment) being more important here.

Post Hoc Analysis: Physician Assessment of the Patient's Ability to Use a DiGA

During our interviews with the physicians, we discovered a phenomenon that we did not initially anticipate and that has not yet been described in the relevant technology adoption literature: the pre–user adoption decision of another instance, whether the user will be able and willing to voluntary use a technology. Typically, in technology adoption research analyses, this means whether a user is willing to use a technology in an either voluntary or mandatory environment. In any case, the decision remains with the user.

Now we see a new phenomenon: the assessment of one instance (physician), whether a subsequent instance (patient) would be able to use the DiGA. Only if the assessment is positive would the physician offer the DiGA to the patient, which will then trigger the traditional adoption questions and corresponding behavior of the user, as described in the well-researched technology adoption models such as the technology acceptance model, UTAUT, and HBM [9,16,23].

There are no guidelines under which circumstances a patient should be assumed to be able to use a DiGA. Thus, each physician needs to do this assessment individually. If they conclude that the patient will likely not be able to install, maintain, and use the app as foreseen by its developers, there is no point in prescribing the DiGA. As there are no objective guidelines, the assessment is either done explicitly, by asking the patient, or implicitly, by assuming what the patient is capable of.

Interestingly, the physicians interviewed were very consistent regarding the assessment of digital literacy of their patients. They indicated that prioritization of certain patient groups is facilitated by anticipating the digital literacy of their patients. Unfortunately, this often leads to a negative bias toward older users; the physicians described the typical DiGA user as a young and tech-savvy patient ("I would not consider my older patients for the use of DiGAs" [E1, E2, E4, E7, E8, E11, E12, E15, E16, E17, E20, E21, E22, and E28]; "There are certainly exceptions, but most of my older patients are totally overwhelmed with a tablet or a smartphone, because the interest would not even be there" [E11]). Consistently, physicians expressed that they would not even consider an older patient as a DiGA user.

As mentioned before, these findings arose from the data and were not anticipated before. Thus, both issues, the second-order technology adoption process and the (possible) systematic disadvantage of older users, need deeper investigation in further research.

Discussion

Overview

In this research, we identified the salient factors that were either beneficial or hindering the adoption of DiGAs from the physician and the patient perspective. Furthermore, the results of our study suggest that the adoption process for a DiGA does not only depend on patient behavior but also on the physician's behavior.

Most informants have a positive attitude toward the digitalization in general. Nevertheless, physicians' demands on DiGAs are high, and their perceptions can be affected by a lack of facilitating conditions, trust, and digital competence. Certain influencing factors for the adoption of DiGAs by patients are consistent with the literature on established adoption research [9,24-28].

Principal Implications

Our study contributes to the field by investigating factors influencing the adoption of DiGAs to inform future research and guide strategies and efforts for this user group. DiGAs represent a wide range of assistive apps that aim to support disease behaviors, manage various health conditions, and maintain the well-being of those with chronic diseases. There are very few empirical studies addressing the factors influencing users' adoption of DiGAs [7,29,30]; hence, there is limited knowledge and guidance from the existing literature.

First, it is important to demonstrate that existing technology acceptance models reach their limits when used in the context of DiGAs. In contrast to Davis [23] and Venkatesh et al [9], our interviews with physicians and potential patients led to the assumption that, in addition to usefulness and ease of use, there are more constructs that play a significant role. We found that technology and health aspects such as technology- and health-related self-efficacy, trust, and a trustful doctor-patient relationship play a major role in the intention to use DiGAs. So far, these aspects have rarely been brought together. A study by Uncovska et al [30] confirms this finding. However, we note that there are few studies on the adoption process of DiGA. Other studies regarding mHealth app adoption have highlighted that health consciousness of individuals is a factor that directly influences both the intention to use mHealth apps and the actual use behavior [31,32]. Public trust in the health care system [33] and a strong doctor-patient relationship can empower patients to contribute to treatment decision-making [34].

Second, we found that our interviewees on the patient side distinguished between technology-related and health-related self-efficacy. The consideration of a health component is not integrated into traditional technology adoption research. To date, it has not seemed necessary to consider the health domain in adopting general technologies. However, this is an important difference in the adoption of DiGAs. With DiGAs, the focus is on the health aspect for both the physician and the patient. Patients who may have low self-efficacy with technology do not simultaneously have to have low self-efficacy with their health. Older people, in particular, may have very sophisticated health self-efficacy while lacking technology self-efficacy [35-38]. Distinguishing these forms of self-efficacy provides a more detailed explanation of adoption behavior, increasing our understanding in the context of DiGAs. Personal beliefs, such as outcome expectations and self-efficacy expectations, are among the most critical variables in terms of intention formation and bridging the gap between intention and behavior, according to existing literature [39-41]. Nevertheless, a division into different areas of self-efficacy has not yet been made in information systems research but also in research on health adoption. As technology and health self-efficacy positively impact the adoption of DiGAs, we believe that it is important to consider both factors.

Third, related to the previous aspect, is the construct of perceived threat. Previous research shows that people are concerned about adopting technology in different areas, such as privacy, effort, or performance [7,42,43]. However, when using DiGAs, health is firmly in focus from a medical perspective. Therefore, the perceived threat of diseases and the need to use a DiGA strongly influenced the adoption of a DiGA. In our interviews, this construct was strongly emphasized, and we suspect moderation effects on other constructs, such as technology and health self-efficacy. A high perceived threat can increase the influence of the perceived health self-efficacy on adopting a DiGA because a threat can be better assessed by someone with high health self-efficacy and is, therefore, more likely to act. As a result, it is also possible that DiGAs will have a higher use rate, especially for hazardous diseases. A recent study by Pourhaji et al [44] investigated the perceived threat and stress response to the COVID-19 pandemic and found that the Iranian population's health behavior was influenced by the perceived severity and susceptibility of the infection, which meant that preventive interventions were more likely to be accepted. Further studies related to COVID-19 found that risk severity also tends to increase with age, but the perception of susceptibility to contracting COVID-19 decreases [45-47]. Thus, risk perception does not seem to increase with age, but vulnerability and severity show opposite patterns [48]. The HBM postulates that individual beliefs about risk can be influenced by various factors such as sociodemographic and sociopsychological variables as well as knowledge, experience, and awareness [16]. However, patient awareness can also become an important issue, as these patients may not perceive a threat and, therefore, not adopt a DiGA.

Fourth, in addition to the patient, this study involves another important stakeholder, the physician. This stakeholder is not considered in the technology adoption models as they do not provide a specific gatekeeper for the technology or basically consider different stakeholders. But in the case of DiGA adoption by the patient, the first step requires the physician's adoption of a DiGA. This observation has also been noted in previous studies [7,49-52]. Subsequently, the physician's positive assessment of the patient's competence to perform a certain behavior is one of the essential conditions for a patient to consistently perform a health intervention. This result relates to previous research without reference to DiGAs as well as with reference to DiGAs [7,53]. Similarly, this view can be developed in a negative direction when doctors decide that the patient is not capable of adopting and using a DiGA, which could be justified by digital ageism. Ageism is a societal bias conceptualized as (1) prejudicial attitudes toward older adults, (2) discriminatory practices toward older adults, or (3) institutionalized policies and social practices that promote these attitudes [54]. Ball et al [55] show that both the development and use of technology have excluded older adults, resulting in a "physical-digital divide," which exists when a group feels excluded because they are unable to engage with the technologies used around them. Some studies suggest that ageism is widespread in the health care system [56-58]. For example, Walter et al [59] showed that physicians promote less preventive care for older patients. Chu et al [60] emphasized that the exclusion of older people from technology development leads to a broader cycle of inequity and ageist social attitudes, widening the digital divide. In contrast, we noticed that the trust factor impacts a patient's health behavior, which is in line with Wildenbos et al [61]. Beyond this, we also found that the physician's trust in the DiGA is equally important for their prescription of a DiGA. A physician needs a strong relationship of trust with the patient to convince the patient of the treatment methods. By motivating their patients to adopt a DiGA and use it to support their therapy, physicians focus on their social influence on the patient [7,62]. We provide justification for the incorporation of the physician as an important influence on adoption behavior in this context. After all, DiGAs live and die with physicians' willingness to prescribe to their patients and influence them to understand the technology's necessity. We argue for an adoption model that does not only incorporate a human-technology interaction but also a human-human-technology interaction.

Implication for Practice

The results of our study demonstrated that there is insufficient information available and published for both physicians and patients, as well as a lack of comprehensive technical support. Some statements (eg, the involvement of medical professionals in the development process of a DiGA, DiGA list unknown, and lack of evidence-based sources) confirm the knowledge deficit. It is important for health policy makers and public authorities such as the Ministry of Health, the Medical Association, and insurers to address these issues. Extensive information and source references are needed to take into account the needs of physicians and to enable DiGAs to get started more effectively. Including the consideration of conflicting goals in technology development from the beginning seems necessary. Codevelopment can improve app use and effectiveness in the long term by using a user-centered design

to develop DiGAs that are effective in chronic disease self-management [63,64]. Likewise, we demonstrated that physicians distinguish clearly between digital and traditional treatment options and have significantly higher expectations of digital resources. Despite a previously excluded position of the physician in the development and introduction process of medicines, the physician now anticipates more involvement in digital developments. In order to counteract the feeling of exclusion and disconnection from the supposedly nonmedically focused DiGA development, appropriate education seems to be required. One way to achieve this is through promotion within public networks, local authorities, and medical associations.

Insufficient participation in the follow-up to a prescription was also a concern for medical practitioners. In this context, it is important to fully communicate to physicians the many opportunities DiGAs offer. DiGAs have higher potential than other treatment interventions such as medicines to maintain a meaningful exchange of information and stay in touch with the patient even after prescription. We found that physicians feel a loss of control when the DiGA is prescribed and then used by the patient. A high degree of self-management is demanded of the patient, leaving the physician feeling incapable of action. Digital monitoring with the help of a DiGA also results in an advantage after the prescription compared with the conventional prescription of medical devices and medicines. Furthermore, negative attitudes and lack of digital competence among physicians are major barriers to physician prescription of DiGAs and, thus, patient adoption. At the same time, the lack of facilitating conditions and the high demands regarding the introduction of various digital changes (eg, EHR) exert excessive pressure on physicians. In this case, further information and education of physicians would be useful. In addition, a trial period of the DiGA can demonstrate digital connectivity to the physician.

Limitations and Future Research

Because of the exploratory character of the research design, our findings naturally lack generalizability and should be regarded as a first starting point on the investigation of a new phenomenon. As DiGAs are a new option for physicians, several interview partners have not really experienced them yet. We found that some participants, especially those who had not yet heard of DiGAs, found it difficult to properly understand the use and benefits of DiGAs.

Another issue corresponding to the novelty of the phenomenon is that informants' perceptions change quickly. Therefore, our findings reflect the perceptions of interview partners in the early phase of introductions of DiGAs into the market. It is likely that some of the issues raised will not be present in a couple of years, when DiGAs are more common to the market and perceived as natural to prescribe as all other medicines today.

This poses interesting questions for further research. It would be interesting to conduct longitudinal studies to gain a better understanding about the diffusion of such innovations in the medical space from the legislative setting into the physicians' toolkit and finally to the patients. This could generate valuable insights for future management of digital innovations in the medical area. In close conjunction to these questions, a cross-national comparison could generate advice for policy to smoothen the introduction phases of digital medical innovations in new other countries.

Finally, the study of the second-order adoption mechanisms—highlighted earlier in the document—could lead to interesting theoretical insights and valuable advice for practitioners to enhance the prescription and adoption of DiGAs and comparable digital innovations.

Conclusions

DiGAs provide an opportunity to support people with severe (often chronic) diseases, to live independently with greater confidence and understanding of their condition, better symptom management, and ultimately enhanced quality of life. Our study provides deep insights into the needs and circumstantial evidence that enables a better understanding of the perspectives and preferences for adopting DiGAs by physicians and potential patients. We found that there is a considerable lack of information on both physicians' and patients' sides, resulting in poor trust and digital competence. Furthermore, we identified several factors influencing the adoption of DiGAs, which led to a new understanding of adoption research concerning digital health technologies.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview guidelines and descriptive results.
[DOCX File , 42 KB - mhealth_v12i1e48345_app1.docx]

References

1. Albrecht UV. Chances and Risks of Mobile Health Apps (CHARISMHA). Hannover Medical School. 2016. URL: https://www.bundesgesundheitsministerium.de/fileadmin/Dateien/3_Downloads/A/App-Studie/charismha_abr_v.01.1e-20160606.pdf [accessed 2023-12-07]
2. WHO Global Observatory for eHealth. mHealth: New Horizons for Health through Mobile Technologies: Based on the Findings of the Second Global Survey on eHealth. Geneva, Switzerland: World Health Organization; 2011.

3. Ältere Menschen und Digitalisierung: Erkenntnisse und Empfehlungen des Achten Altersberichts. Bundesministerin für Familie, Senioren, Frauen und Jugend. URL: https://www.bmfsfj.de/resource/blob/159704/3dab099fb5eb39d9fba72f6810676387/achter-altersbericht-aeltere-menschen-und-digitalisierung-data.pdf [accessed 2023-12-07]
4. Driving the digital transformation of Germany's healthcare system for the good of patients. Federal Ministry of Health. 2020. URL: https://www.bundesgesundheitsministerium.de/en/digital-healthcare-act [accessed 2023-12-07]
5. The fast-track process for digital health applications (DiGA) according to Section 139e SGB V. A guide for manufacturers, service providers and users. Federal Institute for Drugs and Medical Devices. 2020. URL: https://www.bfarm.de/SharedDocs/Downloads/EN/MedicalDevices/DiGA_Guide.html [accessed 2023-12-07]
6. DiGA Directory 2023. Federal Institute for Drugs and Medical Devices. URL: https://diga.bfarm.de/de/verzeichnis [accessed 2023-04-09]
7. Dahlhausen F, Zinner M, Bieske L, Ehlers JP, Boehme P, Fehring L. Physicians' attitudes toward prescribable mHealth apps and implications for adoption in Germany: mixed methods study. JMIR Mhealth Uhealth 2021;9(11):e33012 [FREE Full text] [doi: 10.2196/33012] [Medline: 34817385]
8. Venkatesh V, Davis FD. A theoretical extension of the technology acceptance model: four longitudinal field studies. Manage Sci 2000;46(2):186-204. [doi: 10.1287/mnsc.46.2.186.11926]
9. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Quarterly 2003;27(3):425-478. [doi: 10.2307/30036540]
10. Venkatesh V, Thong JYL, Xu X. Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. MIS Quarterly 2012;36(1):157-178. [doi: 10.2307/41410412]
11. Global report on ageism. World Health Organization. Geneva; 2021. URL: https://www.who.int/teams/social-determinants-of-health/demographic-change-and-healthy-ageing/combatting-ageism/global-report-on-ageism [accessed 2023-12-07]
12. Taherdoost H. A review of technology acceptance and adoption models and theories. Procedia Manuf 2018;22:960-967 [FREE Full text] [doi: 10.1016/j.promfg.2018.03.137]
13. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: a synthesis of recommendations. Acad Med 2014;89(9):1245-1251 [FREE Full text] [doi: 10.1097/ACM.0000000000000388] [Medline: 24979285]
14. Huang CQ, Dong BR, Lu ZC, Yue JR, Liu QX. Chronic diseases and risk for depression in old age: a meta-analysis of published literature. Ageing Res Rev 2010;9(2):131-141. [doi: 10.1016/j.arr.2009.05.005] [Medline: 19524072]
15. Lee CC, Czaja SJ, Moxley JH, Sharit J, Boot WR, Charness N, et al. Attitudes toward computers across adulthood from 1994 to 2013. Gerontologist 2019;59(1):22-33 [FREE Full text] [doi: 10.1093/geront/gny081] [Medline: 29982458]
16. Hochbaum G, Rosenstock I, Kegels S. Health belief model. United States Public Health Service. 1952. URL: https://www.usphs.gov/ [accessed 2023-12-07]
17. Flick U. An Introduction to Qualitative Research, 6th Edition. Los Angeles: Sage Publications Limited; 2018.
18. Conrad P. The experience of illness: recent and new directions. In: Roth JA, Conrad P, editors. Research in the Sociology of Health Care. Vol 6: The Experience and Management of Chronic Illness. Greenwich, CT: JAI Press; 1987:1-31.
19. Gioia DA, Corley KG, Hamilton AL. Seeking qualitative rigor in inductive research. Organ Res Methods 2012;16(1):15-31. [doi: 10.1177/1094428112452151]
20. Swoboda W, Schmieder M, Bulitta C, Gaisser S, Hofmann GR, Kremer-Rücker P, et al. Die gemeinsame Ethik-Kommission der Hochschulen Bayerns—GEHBa. Article in German. mdi—Forum der Medizin-Dokumentation und Medizin-Informatik 2021(3):80-83 [FREE Full text]
21. Chang YZ, Yu CW, Chao CM, Lin FC. Influences on medical app adoption by patients: the unified theory of acceptance and use of technology model and the moderating effects of technology readiness. Soc Sci J 2020:1-14. [doi: 10.1080/03623319.2020.1848338]
22. Reesor L, Vaughan EM, Hernandez DC, Johnston CA. Addressing outcomes expectancies in behavior change. Am J Lifestyle Med 2017;11(6):430-432 [FREE Full text] [doi: 10.1177/1559827617722504] [Medline: 29302249]
23. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly 1989;13(3):319-340. [doi: 10.2307/249008]
24. Ajzen I. The directive influence of attitudes on behavior. In: Gollwitzer PM, Bargh JA, editors. The Psychology of Action: Linking Cognition and Motivation to Behavior. New York: Guilford Press; 1996:385-403.
25. Deng Z, Hong Z, Ren C, Zhang W, Xiang F. What predicts patients' adoption intention toward mHealth services in China: empirical study. JMIR Mhealth Uhealth 2018;6(8):e172 [FREE Full text] [doi: 10.2196/mhealth.9316] [Medline: 30158101]
26. Glanz K, Rimer BK, Viswanath K. Health Behavior and Health Education: Theory, Research, and Practice, 4th Edition. San Francisco, CA: Jossey-Bass; 2008.
27. Guo X, Zhang X, Sun Y. The privacy–personalization paradox in mHealth services acceptance of different age groups. Electron Commer Res Appl 2016;16:55-65. [doi: 10.1016/j.elerap.2015.11.001]
28. Li J, Talaei-Khoei A, Seale H, Ray P, Macintyre CR. Health care provider adoption of eHealth: systematic literature review. Interact J Med Res 2013;2(1):e7 [FREE Full text] [doi: 10.2196/ijmr.2468] [Medline: 23608679]

29. Wangler J, Jansky M. The use of health apps in primary care-results from a survey amongst general practitioners in Germany. Wien Med Wochenschr 2021;171(7-8):148-156 [FREE Full text] [doi: 10.1007/s10354-021-00814-0] [Medline: 33570692]
30. Uncovska M, Freitag B, Meister S, Fehring L. Patient acceptance of prescribed and fully reimbursed mHealth apps in Germany: an UTAUT2-based online survey study. J Med Syst 2023;47(1):14 [FREE Full text] [doi: 10.1007/s10916-023-01910-x] [Medline: 36705853]
31. Alam MMD, Alam MZ, Rahman SA, Taghizadeh SK. Factors influencing mHealth adoption and its impact on mental well-being during COVID-19 pandemic: a SEM-ANN approach. J Biomed Inform 2021;116:103722 [FREE Full text] [doi: 10.1016/j.jbi.2021.103722] [Medline: 33705856]
32. Cho J, Park D, Lee HE. Cognitive factors of using health apps: systematic analysis of relationships among health consciousness, health information orientation, eHealth literacy, and health app use efficacy. J Med Internet Res 2014;16(5):e125 [FREE Full text] [doi: 10.2196/jmir.3283] [Medline: 24824062]
33. Gille F, Smith S, Mays N. What is public trust in the healthcare system? a new conceptual framework developed from qualitative data in England. Soc Theory Health 2020;19(1):1-20. [doi: 10.1057/s41285-020-00129-x]
34. Luk CY. The impact of digital health on traditional healthcare systems and doctor-patient relationships: the case study of Singapore. In: Manoharan A, McQuiston J, editors. Innovative Perspectives on Public Administration in the Digital Age. Hershey, Pennsylvania: IGI Global; 2018:143-167.
35. Cajita MI, Hodgson NA, Budhathoki C, Han HR. Intention to use mHealth in older adults with heart failure. J Cardiovasc Nurs 2017;32(6):E1-E7 [FREE Full text] [doi: 10.1097/JCN.0000000000000401] [Medline: 28248747]
36. Kuerbis A, Mulliken A, Muench F, Moore AA, Gardner D. Older adults and mobile technology: factors that enhance and inhibit utilization in the context of behavioral health. Ment Health Addict Res 2017;2(2):1-11 [FREE Full text] [doi: 10.15761/MHAR.1000136]
37. Luszczynska A, Sutton S. Physical activity after cardiac rehabilitation: evidence that different types of self-efficacy are important in maintainers and relapsers. Rehabilitation Psychology 2006;51(4):314-321. [doi: 10.1037/0090-5550.51.4.314]
38. Schwarzer R, Renner B. Social-cognitive predictors of health behavior: action self-efficacy and coping self-efficacy. Health Psychol 2000;19(5):487-495. [doi: 10.1037/0278-6133.19.5.487]
39. Rudy BJ, Murphy DA, Harris DR, Muenz L, Ellen J, Adolescent Trials Network for HIV/AIDS Interventions. Patient-related risks for nonadherence to antiretroviral therapy among HIV-infected youth in the United States: a study of prevalence and interactions. AIDS Patient Care STDS 2009;23(3):185-194 [FREE Full text] [doi: 10.1089/apc.2008.0162] [Medline: 19866536]
40. Sirur R, Richardson J, Wishart L, Hanna S. The role of theory in increasing adherence to prescribed practice. Physiother Can 2009;61(2):68-77 [FREE Full text] [doi: 10.3138/physio.61.2.68] [Medline: 20190989]
41. Wong N. Investigating the effects of cancer risk and efficacy perceptions on cancer prevention adherence and intentions. Health Commun 2009;24(2):95-105 [FREE Full text] [doi: 10.1080/10410230802676474] [Medline: 19280453]
42. Schroeder T, Haug M, Gewald H. Data privacy concerns using mHealth apps and smart speakers: comparative interview study among mature adults. JMIR Form Res 2022;6(6):e28025 [FREE Full text] [doi: 10.2196/28025] [Medline: 35699993]
43. Zhou L, Bao J, Watzlaf V, Parmanto B. Barriers to and facilitators of the use of mobile health apps from a security perspective: mixed-methods study. JMIR Mhealth Uhealth 2019;7(4):e11223 [FREE Full text] [doi: 10.2196/11223] [Medline: 30990458]
44. Pourhaji F, Tehrani H, Talebi M, Peyman N. Perceived threat and stress responses in the face of COVID-19 based on health belief model. J Health Lit 2022;7(1):17-25.
45. de Bruin WB. Age differences in COVID-19 risk perceptions and mental health: evidence from a national U.S. survey conducted in march 2020. J Gerontol B Psychol Sci Soc Sci 2021;76(2):e24-e29 [FREE Full text] [doi: 10.1093/geronb/gbaa074] [Medline: 32470120]
46. Kivi M, Hansson I, Bjälkebring P. Up and about: older adults' well-being during the COVID-19 pandemic in a Swedish longitudinal study. J Gerontol B Psychol Sci Soc Sci 2021;76(2):e4-e9 [FREE Full text] [doi: 10.1093/geronb/gbaa084] [Medline: 32599622]
47. Pasion R, Paiva TO, Fernandes C, Barbosa F. The AGE effect on protective behaviors during the COVID-19 outbreak: sociodemographic, perceptions and psychological accounts. Front Psychol 2020;11:561785 [FREE Full text] [doi: 10.3389/fpsyg.2020.561785] [Medline: 33178069]
48. Rosi A, van Vugt FT, Lecce S, Ceccato I, Vallarino M, Rapisarda F, et al. Risk perception in a real-world situation (COVID-19): how it changes from 18 to 87 years old. Front Psychol 2021;12:646558 [FREE Full text] [doi: 10.3389/fpsyg.2021.646558] [Medline: 33737899]
49. Byambasuren O, Beller E, Hoffmann T, Glasziou P. Barriers to and facilitators of the prescription of mHealth apps in Australian general practice: qualitative study. JMIR Mhealth Uhealth 2020;8(7):e17447 [FREE Full text] [doi: 10.2196/17447] [Medline: 32729839]
50. Sarradon-Eck A, Bouchez T, Auroy L, Schuers M, Darmon D. Attitudes of general practitioners toward prescription of mobile health apps: qualitative study. JMIR Mhealth Uhealth 2021;9(3):e21795 [FREE Full text] [doi: 10.2196/21795] [Medline: 33661123]
51. Schroeder T, Haug M, Gewald H. The difference between motivation and volition matters!-a qualitative study on mobile health application adoption. 2021 Presented at: European Conference on Information Systems 2021 Proceedings Research

Paper; 2021; Marrakesh, Morocco p. 1-18 URL: https://web.archive.org/web/20220802072524id_/https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1096&context=ecis2021_rp
52. Schroeder T, Seaman K, Nguyen A, Gewald H, Georgiou A. Enablers and inhibitors to the adoption of mHealth apps by patients—a qualitative analysis of German doctors' perspectives. Patient Educ Couns 2023;114:107865 [FREE Full text] [doi: 10.1016/j.pec.2023.107865] [Medline: 37356116]
53. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. Psychol Rev 1977;84(2):191-215. [doi: 10.1037//0033-295x.84.2.191] [Medline: 847061]
54. Rosales A, Fernández-Ardèvol M. Ageism in the era of digital platforms. Convergence (Lond) 2020;26(5-6):1074-1087 [FREE Full text] [doi: 10.1177/1354856520930905] [Medline: 33239961]
55. Ball C, Francis J, Huang KT, Kadylak T, Cotten SR, Rikard RV. The physical-digital divide: exploring the social gap between digital natives and physical natives. J Appl Gerontol 2019;38(8):1167-1184. [doi: 10.1177/0733464817732518] [Medline: 29165009]
56. Meisner BA. Physicians' attitudes toward aging, the aged, and the provision of geriatric care: a systematic narrative review. Crit Public Health 2012;22(1):61-72. [doi: 10.1080/09581596.2010.539592]
57. Pritchard KI. Have we been guilty of ageism in the primary treatment of breast cancer? Br J Cancer 2007;96(7):1011-1012 [FREE Full text] [doi: 10.1038/sj.bjc.6603697] [Medline: 17406347]
58. Silliman RA. When cancer in older adults is undermanaged: the breast cancer story. J Am Geriatr Soc 2009;57(Suppl 2):S259-S261 [FREE Full text] [doi: 10.1111/j.1532-5415.2009.02506.x] [Medline: 20122024]
59. Walter U, Flick U, Neuber A, Fischer C, Hussein RJ, Schwartz FW. Putting prevention into practice: qualitative study of factors that inhibit and promote preventive care by general practitioners, with a focus on elderly patients. BMC Fam Pract 2010;11:68 [FREE Full text] [doi: 10.1186/1471-2296-11-68] [Medline: 20854654]
60. Chu CH, Nyrup R, Leslie K, Shi J, Bianchi A, Lyn A, et al. Digital ageism: challenges and opportunities in artificial intelligence for older adults. Gerontologist 2022;62(7):947-955 [FREE Full text] [doi: 10.1093/geront/gnab167] [Medline: 35048111]
61. Wildenbos GA, Peute L, Jaspers M. Aging barriers influencing mobile health usability for older adults: a literature based framework (MOLD-US). Int J Med Inform 2018;114:66-75. [doi: 10.1016/j.ijmedinf.2018.03.012] [Medline: 29673606]
62. Schroeder T, Dodds L, Georgiou A, Gewald H, Siette J. Older adults and new technology: mapping review of the factors associated with older adults' intention to adopt digital technologies. JMIR Aging 2023;6:e44564 [FREE Full text] [doi: 10.2196/44564] [Medline: 37191976]
63. Matthew-Maich N, Harris L, Ploeg J, Markle-Reid M, Valaitis R, Ibrahim S, et al. Designing, implementing, and evaluating mobile health technologies for managing chronic conditions in older adults: a scoping review. JMIR Mhealth Uhealth 2016;4(2):e29 [FREE Full text] [doi: 10.2196/mhealth.5127] [Medline: 27282195]
64. Schnall R, Rojas M, Bakken S, Brown W, Carballo-Dieguez A, Carry M, et al. A user-centered model for designing consumer mobile health (mHealth) applications (apps). J Biomed Inform 2016;60:243-251 [FREE Full text] [doi: 10.1016/j.jbi.2016.02.002] [Medline: 26903153]

Abbreviations

BfArM: Bundesinstitut für Arzneimittel und Medizinprodukte (Federal Institute for Drugs and Medical Devices)
DiGA: Digitale Gesundheitsanwendungen (digital health applications)
DiGAV: Digitale-Gesundheitsanwendungen-Verordnung (Digital Health Applications Ordinance)
DVG: Digitale-Versorgung-Gesetz (Digital Healthcare Act)
EHR: electronic health record
EU-GDPR: European Union General Data Protection Regulation
HBM: health belief model
IPA: interpretative phenomenological analysis
mHealth: mobile health
UN: United Nations
UTAUT: unified theory of acceptance and use of technology
WHO: World Health Organization

Edited by L Buis; submitted 20.04.23; peer-reviewed by J Claggett, R Azevedo; comments to author 17.06.23; revised version received 30.09.23; accepted 27.11.23; published 17.01.24.

Please cite as:
Schroeder T, Haug M, Georgiou A, Seaman K, Gewald H
Evidence of How Physicians and Their Patients Adopt mHealth Apps in Germany: Exploratory Qualitative Study
JMIR Mhealth Uhealth 2024;12:e48345
URL: https://mhealth.jmir.org/2024/1/e48345
doi:10.2196/48345
PMID:38231550

Original Paper

Engagement With a Remote Symptom-Tracking Platform Among Participants With Major Depressive Disorder: Randomized Controlled Trial

Katie M White[1], PhD; Ewan Carr[2], PhD; Daniel Leightley[1], PhD; Faith Matcham[1,3], PhD; Pauline Conde[2], BSc; Yatharth Ranjan[2], BSc; Sara Simblett[4], PhD; Erin Dawe-Lane[4], MSc; Laura Williams[5], BA, MSc; Claire Henderson[6,7], PhD; Matthew Hotopf[1,7], PhD

[1]Department of Psychological Medicine, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom

[2]Department of Biostatistics and Health Informatics, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom

[3]School of Psychology, University of Sussex, Falmer, United Kingdom

[4]Department of Psychology, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom

[5]NIHR MindTech MedTech Co-operative, Institute of Mental Health and Clinical Neurosciences, University of Nottingham, Nottingham, United Kingdom

[6]Health Services & Population Research Department, King’s College London, London, United Kingdom

[7]South London and Maudsley NHS Foundation Trust, London, United Kingdom

Corresponding Author:
Katie M White, PhD
Department of Psychological Medicine
Institute of Psychiatry, Psychology and Neuroscience
King's College London
16 De Crespigny Park
London, SE5 8AF
United Kingdom
Phone: 44 020 7848 0002
Email: katie.white@kcl.ac.uk

Abstract

Background: Multiparametric remote measurement technologies (RMTs), which comprise smartphones and wearable devices, have the potential to revolutionize understanding of the etiology and trajectory of major depressive disorder (MDD). Engagement with RMTs in MDD research is of the utmost importance for the validity of predictive analytical methods and long-term use and can be conceptualized as both objective engagement (data availability) and subjective engagement (system usability and experiential factors). Positioning the design of user interfaces within the theoretical framework of the Behavior Change Wheel can help maximize effectiveness. In-app components containing information from credible sources, visual feedback, and access to support provide an opportunity to promote engagement with RMTs while minimizing team resources. Randomized controlled trials are the gold standard in quantifying the effects of in-app components on engagement with RMTs in patients with MDD.

Objective: This study aims to evaluate whether a multiparametric RMT system with theoretically informed notifications, visual progress tracking, and access to research team contact details could promote engagement with remote symptom tracking over and above the system as usual. We hypothesized that participants using the adapted app (intervention group) would have higher engagement in symptom monitoring, as measured by objective and subjective engagement.

Methods: A 2-arm, parallel-group randomized controlled trial (participant-blinded) with 1:1 randomization was conducted with 100 participants with MDD over 12 weeks. Participants in both arms used the RADAR-base system, comprising a smartphone app for weekly symptom assessments and a wearable Fitbit device for continuous passive tracking. Participants in the intervention arm (n=50, 50%) also had access to additional in-app components. The primary outcome was objective engagement, measured as the percentage of weekly questionnaires completed during follow-up. The secondary outcomes measured subjective engagement (system engagement, system usability, and emotional self-awareness).

Results: The levels of completion of the Patient Health Questionnaire-8 (PHQ-8) were similar between the control (67/97, 69%) and intervention (66/97, 68%) arms (P value for the difference between the arms=.83, 95% CI −9.32 to 11.65). The intervention

group participants reported slightly higher user engagement (1.93, 95% CI −1.91 to 5.78), emotional self-awareness (1.13, 95% CI −2.93 to 5.19), and system usability (2.29, 95% CI −5.93 to 10.52) scores than the control group participants at follow-up; however, all CIs were wide and included 0. Process evaluation suggested that participants saw the in-app components as helpful in increasing task completion.

Conclusions: The adapted system did not increase objective or subjective engagement in remote symptom tracking in our research cohort. This study provides an important foundation for understanding engagement with RMTs for research and the methodologies by which this work can be replicated in both community and clinical settings.

Trial Registration: ClinicalTrials.gov NCT04972474; https://clinicaltrials.gov/ct2/show/NCT04972474

International Registered Report Identifier (IRRID): RR2-10.2196/32653

(JMIR Mhealth Uhealth 2024;12:e44214) doi:10.2196/44214

KEYWORDS

remote measurement; technology; engagement; app; depression; smartphones; wearable devices; engagement; symptom tracking; self-awareness; community; mobile phone

Introduction

Background

Multiparametric remote measurement technologies (RMTs), which comprise smartphone apps and wearable devices, have the potential to revolutionize the clinical care of people with chronic, episodic health conditions [1]. Major depressive disorder (MDD) is one such condition, characterized by the relapse and remission of low mood and anhedonia over time [2]. Continuously measured longitudinal RMT data on the symptoms of MDD (mood variability, activity, cognition, and sleep) can capture a less biased picture of clinical state than retrospective self-report data [3]. Research using multiparametric sources might identify signals that could potentially predict future depressive episodes [4]. Such data could be ultimately implemented in patient self-management and shared decision-making in clinical practice [5].

It is important to understand how users engage with RMTs for depression symptom tracking. A recent systematic review found that engagement with RMTs can be measured objectively, for example, as the number of app-based symptom-tracking assessments completed, and subjectively, for example, as the perceived usability of and experience of using the RMT system [6]. Higher levels of objective engagement result in increased data availability, which, in turn, increases the validity of the machine learning approaches used for relapse prediction [7]. Objective engagement can also be used as an indicator of real-world uptake [8,9]. Further evidence suggests that increased satisfaction with mobile health apps is positively associated with the intention to continually use the tools [10]. Therefore, understanding engagement with RMTs is key to realizing their potential for relapse prediction.

Previous studies have reported inconsistent levels of engagement with RMT systems. Data completion, based on the total data expected, ranges from 42% to 82% for app-based symptom reporting and from 50% to 75% for device wear time [11]. The largest, multisite study of multiparametric RMTs for tracking depression to date, Remote Assessment of Disease and Relapse–Major Depressive Disorder (RADAR-MDD) [12], tracked 623 participants for 2 years using a smartphone app for mood tracking and a wrist-worn wearable for continuous passive data collection. The study has recently reported data availability metrics; 55.4% (345/623) of the sample completed >50% of the self-reported mood questionnaires expected to be completed, and 70.1% (437/623) had wearable heart rate data for >50% of the study days. Qualitative analyses from RADAR-MDD have revealed that the presence of a physical research team providing technological support and planned task reminders was a fundamental facilitator of long-term engagement in the study [13]. To ensure the scalability and real-world implementation of RMT systems, it is important to investigate methods that maximize engagement with RMTs while minimizing the human resources needed.

Focusing on the user interface (UI) of RMT systems is the logical first step for promoting engagement. Positioning the design of system UI within a theoretical framework of behavior change could help maximize effectiveness [14]. The Behavior Change Wheel [15] posits that researchers should begin by identifying a target behavior before considering the barriers to and facilitators of this behavior in terms of capability, opportunity, and motivation (the capability, opportunity, motivation, and behavior [COM-B] model). In the case of RMTs, the target behavior can be defined as objective engagement with symptom monitoring tasks. A series of published studies have evaluated both perceived [11,16] and experienced [13] barriers to RMT use in MDD research. Factors such as the knowledge of the utility of the research (capability), motivation linked to mood (motivation), and confirmation of logged data (opportunity) have been suggested to be prominent. The Behavior Change Wheel further provides a series of “intervention functions” best suited to address these factors, each with its own related behavior change techniques. With regard to RMTs, these have been suggested to be the provision of information from credible sources, visual feedback on behavior, and access to support.

The app design literature provides several options for incorporating behavior change techniques into RMT system design. First, following the Fogg behavioral model [17], push notifications can provide a trigger to perform a behavior, such as completing a monitoring task. Notifications can include tailored content, such as insights into the benefits of self-monitoring, which serves to simultaneously motivate the

user to respond to the notification and engage them in future tasks [18]. Second, visual incentives, such as graphs, can be embedded into the app to reflect on patterns in user progress and spark intrinsic motivation to complete future tasks [19]. Visualization can also help users manage uncertainty by attending to information about themselves [20]. A combination of qualitative and single-arm evaluation studies supports the perceived value of data visualization [21,22] and progress viewing [19] in encouraging symptom-tracking completion. Provision of contact details directly within an app can allow the user to directly and immediately access support, if required.

Without a control group, it is difficult to quantify the effect of in-app components on engagement [23]. A randomized controlled trial (RCT) of a substance abuse tracking app [20] suggested that users were 5% more likely to self-report on a day if they received a prior notification with an inspirational quote, although these results were not statistically significant. Conversely, users were 2% less likely to self-report following the provision of personalized visual data summaries; however, this main effect was significantly moderated by the prior day task completion such that those who had not completed the previous task were 36% more likely to self-report after receiving data visualization [20]. Users receiving prompts with tailored health messages, such as those highlighting the beneficial effects of symptom monitoring, were 4% more likely to engage in self-monitoring via another app for mental well-being [24]. It is important to replicate this work with multiparametric symptom monitoring systems, as it is currently unclear which combination of in-app features best promotes engagement with these technologies.

This Study

This study aimed to evaluate whether in-app components could promote engagement with a multiparametric RMT system for symptom tracking in depression. We conducted a 2-arm RCT to compare the system as usual with an adapted system that contained informative notifications, a visual progress report, and access to the research team contact details as a substitute for planned research team contact. We measured engagement as both objective and subjective concepts. This study had the following four specific objectives: (1) to describe data availability in an RCT of a multiparametric RMT system for tracking depression, (2) to test whether in-app components increase the rates of objective data completion, (3) to explore how in-app components influence the subjective experience of using the app, and (4) to understand how the components of the system are used by participants via process evaluation measures.

For objectives 2 and 3, we hypothesized that participants using the adapted app (intervention group) would have higher engagement in symptom monitoring, as measured by both objective engagement (completion of mood questionnaires) and subjective engagement (usability, utility, and emotional self-awareness).

Methods

Ethical Considerations

This study was approved by the Psychiatry, Nursing, and Midwifery Research Ethics Subcommittee at King's College London (reference number: RESCM-20/21-21083) and registered as a clinical trial (reference number: NCT04972474). A trial protocol has been previously published [25].

Trial Design

This was a single-center, 2-arm, parallel-group RCT (participant-blinded) with 1:1 randomization conducted in London, United Kingdom. We compared a remote symptom-tracking system (RADAR-base [26]; the control arm) with a system that contained additional in-app components (the intervention arm). Both the control and intervention arms were delivered via the RADAR-base system [26] using a smartphone app and a wearable Fitbit Charge (Fitbit Inc) device. Participants in the intervention arm had additional access to (1) theoretically informed notifications, (2) progress visualization, and (3) research team contact details through the study app. All participants were asked to use the system for 12 weeks.

Data were collected at baseline (0 weeks) and follow-up (12 weeks after randomization). Participants in both arms were sent 3 symptom-tracking tasks each week via the app; Fitbit data were collected continuously.

Participants

All participants were recruited from the RADAR-MDD study between April and May 2021. The inclusion criteria were as follows: (1) previous participation in the RADAR-MDD study at the London site (which required experiencing at least 1 episode of MDD in the 2 years before enrollment), (2) consent to be contacted, (3) willingness and ability to continue to use an Android (Google LLC) smartphone (provided for use by RADAR-MDD; see the study by Matcham et al [3] for the full study protocol), and (4) willingness and ability to complete a remote enrollment session owing to the COVID-19 pandemic. Participants were excluded if they were diagnosed with one of the following comorbid psychiatric disorders: bipolar disorder, schizophrenia, psychosis, schizoaffective disorder, or dementia.

Potential participants were invited to take part (up to 3 invitations were sent per participant, as per ethical considerations) and subsequently checked for eligibility, both via email. If eligible, contact details were entered into the REDCap (Research Electronic Data Capture [27]) system, which emailed an automated link to the informed consent form and baseline questionnaires. After participants provided consent and completed the baseline questionnaires, they were sent a link to book an enrollment session (via email, phone call, or video call).

On the day of the enrollment session, the principal investigator (KMW) initiated the REDCap randomization module and generated unique QR codes to link the study devices to the RADAR-base management portal. Each participant was sent a personalized set of instructions for downloading and logging

into the system using the QR codes at the chosen enrollment time, accompanied by a phone or video call as requested.

Participants were purposefully not contacted by the research team during the follow-up period, aside from sending 1 check-in email at the 6-week time point. However, participants were able to initiate contact with the team if they had any queries during follow-up. The research team did not make withdrawals based on "lost to follow up," given the fundamental aims of the study; however, participants were aware that they could withdraw at any point.

Suicidal ideation was assessed at baseline and follow-up using the Inventory of Depressive Symptomatology–Self-Report [28] item "thoughts of death or suicide." Participants who reported suicidal ideation and intent at either time point were contacted via phone call by the principal investigator, advised to contact their treating physician, and emailed a list of signposting resources.

At the 12-week end point, participants were directed to debrief information that explained the aims of the study and provided instructions for logging out of the system.

RADAR-Base

The RADAR-base system is an open-source platform that supports data collection via remote devices [3,26]. It requires users to download and log into an Android smartphone app in addition to wearing and syncing a wearable device. All participants were asked to complete the following three validated symptom-tracking tasks per week via the study app: (1) Patient Health Questionnaire-8 (PHQ-8 [29]); (2) Rosenberg Self-Esteem Scale (RSES [30]); and (3) a speech task, during which the user records themselves reading aloud a short paragraph (Multimedia Appendix 1). All tasks became available on the same day each week, 1 hour apart, beginning at the point of enrollment. All tasks had to be completed within 24 hours.

Interventions

Control Arm

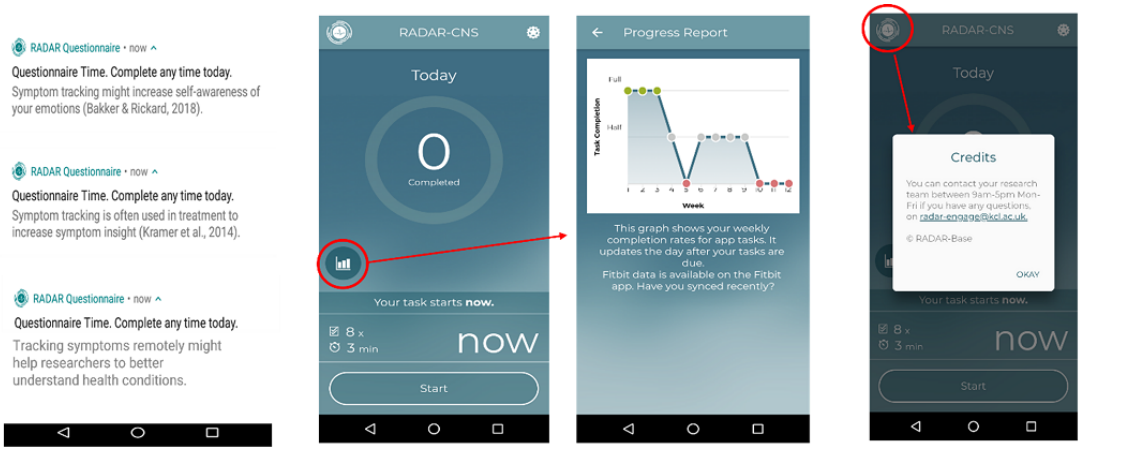
Participants were sent 3 tasks per week via the RADAR-base study app, as outlined in the previous section. For each task, they received a notification on the day that the task was due that read, "Questionnaire Time. Won't usually take longer than 3 minutes." They were unable to view any data other than those available on the Fitbit app.

Intervention Arm

The design of the additional in-app components was grounded in behavioral theory and user research on the barriers to and facilitators of RMT use in patients with MDD [11,13,16]. The COM-B [31] framework of behavior change highlighted education, incentivization, and enablement as the most suitable forms of intervention function. Findings from research with users of the RADAR-base system allowed for the translation of these functions into tangible components tailored specifically to the needs and preferences of the target cohort [32]. It was decided that an engaging app should include notifications with information on symptom tracking from a credible source, behavioral feedback via progress visualization, and instant access to researcher contact details (see the study by White et al [25] and Multimedia Appendix 2 for a detailed overview of this process).

Participants in the intervention arm received notifications and tasks at the same time as those in the control arm but with the following additional content (Figure 1):

Figure 1. Screenshots of the in-app components included in the intervention arm.



(A) Theoretically informed notifications alternating between 3 phrases

(B) Progress visualization, which is accessible through an icon on the home page, displaying full, half, or non-completion of weekly tasks for each of the 12 weeks

(C) Research team contact details, which are accessible through an icon on the home page

1. Theoretically informed notifications: the notifications included additional sentences that described the potential benefits of symptom monitoring for emotional self-awareness, clinical practice, and research. Participants were also reminded that they could complete the task "any time today."

2. Progress visualization: participants were provided with a graph in the app that tracked the completion of the tasks. This graph could be viewed at any time from the main app home page.
3. Researcher contact details: the main app home page included a phone number, an email address, and contact hours of the research team for the reporting of technical issues or requests for support.

Measures

A summary of measures and data collection time points is presented in Table 1. The measures were identical between the intervention and control arms.

Table 1. A summary of measures and data collection points across the 12-week follow-up period.

Measures	Baseline	End point	Weekly	Continuously
REDCap[a] survey				
Consent	✓			
Contact information	✓			
Study devices	✓			
Sociodemographics	✓			
Social environment	✓			
Medical history	✓			
LIDAS[b]	✓			
IDS-SR[c]	✓	✓		
The World Health Organization CIDI-SF[d]	✓	✓		
GAD-7[e]	✓	✓		
WSAS[f]	✓	✓		
BIPQ[g]	✓	✓		
Life events	✓	✓		
CSRI[h]	✓	✓		
UES[i]	✓	✓		
ESQ[j]	✓	✓		
MAUQ[k]		✓		
Active app measures				
PHQ-8[l]			✓	
RSES[m]			✓	
Speech task			✓	
Fitbit				
Heart rate, step count, and GPS				✓
Process evaluation				
App use metrics				✓
Qualitative interviews		✓		

[a]REDCap: Research Electronic Data Capture.
[b]LIDAS: Lifetime Depression Assessment Self-Report.
[c]IDS-SR: Inventory of Depressive Symptomatology–Self-Report.
[d]CIDI-SF: Composite International Diagnostic Interview-Short Form.
[e]GAD-7: Generalized Anxiety Disorder-7.
[f]WSAS: Work and Social Adjustment Scale.
[g]BIPQ: Brief Illness Perception Questionnaire.
[h]CSRI: Client Service Receipt Inventory.
[i]UES: User Engagement Scale.
[j]ESQ: Emotional Self-Awareness Questionnaire.
[k]MAUQ: mHealth App Usability Questionnaire.
[l]PHQ-8: Patient Health Questionnaire-8.
[m]RSES: Rosenberg Self-Esteem Scale.

Questionnaires

After registration for the study, participants completed web-based baseline questionnaires via REDCap, providing information on sociodemographics and physical and mental health history, including the presence of depression, recent life events, and service use. The principal investigator also manually extracted data pertaining to previous participation in the RADAR-MDD study, including participation length and technology use. At the 12-week time point, participants repeated these questionnaires.

Outcome Measures

The primary outcome was objective engagement with the system, measured as the number and percentage of weekly PHQ-8 questionnaires completed during follow-up (compared with the total of 12 questionnaires that were sent). Completion of 1 PHQ-8 questionnaire was defined as the completion of all 8 questions.

There were four secondary outcomes, three of which measured subjective engagement with the system:

1. User engagement: this was measured using the User Engagement Scale (UES) [33] adapted to mobile health use [34], a 30-item questionnaire measuring focused attention, perceived usability, esthetic appeal, and reward. All items are scored on a Likert scale ranging from 1 (“strongly disagree”) to 5 (“strongly agree”). Total scores are calculated by summing the scores for each item in each of the 4 subscales and dividing the resultant value by the number of items in each subscale. An overall engagement score can be calculated using the average of each subscale. A higher average score indicates higher user engagement. The UES has been widely adopted and shows good reliability and construct validity [35].
2. Emotional self-awareness: this was measured using the Emotional Self-Awareness Questionnaire (ESQ) [36], a 33-item scale measuring recognition, contextualization, and decision-making in relation to self-emotion. All items are scored on a 5-point Likert scale ranging from 0 (“never”) to 4 (“a lot”). The total score is calculated as a continuous variable that ranges from 0 to 132, with a higher score reflecting higher emotional self-awareness. The ESQ has a reliability of 0.92 and shows significant positive correlations with the Emotional Intelligence Test [36].
3. System usability: this was measured using the mHealth App Usability Questionnaire (MAUQ) for stand-alone apps [37]. The MAUQ is an 18-item scale that measures the immediate and long-term self-reported usability of an app, including its ease of use and utility for self-management (overall Cronbach α=.914). All items are scored on a 7-point Likert scale ranging from 1 (“disagree”) to 7 (“agree”). The app usability score is calculated as the sum of scores across the items for each participant, ranging from 18 to 126, with a higher score reflecting higher reported usability of the app.
4. Overall adherence to the RADAR-base system: a participant was considered to have adhered to the system if they (1) responded to at least 50% of the 3 weekly tasks and (2) >2 heart rate data points were recorded by the Fitbit device on at least 50% of the days during the 12-week intervention period. This measure was chosen to align with previous data availability reporting [12] and other studies [38].

Process evaluation measures were collected to evaluate the use of the in-app components. Quantitative measures covered app engagement, in-app interactions, and notification engagement. A total of 20 participants, split evenly across the 2 arms, were also invited at the study endpoint to qualitatively discuss their experiences with the components through a 1:1 interview (Multimedia Appendix 3).

Sample Size

Power calculations were performed based on data availability from the RADAR-MDD study [12]. To detect a difference of 25% completion of PHQ-8 tasks between the control and intervention arms, with 80% power and 95% CIs, 132 participants were required (66 per arm). We decided on 25% as the minimum difference that would be practically useful for analyses.

Randomization and Blinding

Participants were randomly allocated in a 1:1 ratio to either the control or intervention arm using simple randomization via the REDCap randomization module.

The principal investigator was unblinded to allocation to ensure that remote enrollments had been carried out successfully and had access to incoming data throughout the study. The trial data manager (DL) was blinded to arm allocation, as this information was stored elsewhere. Participants had previously used the RADAR-base system and, therefore, could not be fully blinded to arm allocation. However, the explicit aims and arm assignments of the study were not revealed until study debrief.

Statistical Methods

Sociodemographic and clinical variables at baseline were described by arm using appropriate summary statistics (counts and percentages for categorical variables and mean and SD or median and IQR for continuous variables). We reported data availability for all outcomes. Data availability for each study app task (PHQ-8, RSES, and speech task) was summarized as the median (IQR) number of weekly tasks completed. Fitbit wear time was summarized as mean (SD) days with >2 heart rate data points. The overall completion of all 4 data sources was also reported, calculated as a percentage of the total expected count (n=12) for the study app tasks and the total expected days of wear time (n=84) for the Fitbit.

The primary outcome, objective engagement, was analyzed using 2-sample 2-tailed *t* tests, which tested the difference in the mean percentage of PHQ-8 completion over 12 weeks between the study arms.

Three secondary outcomes (UES, ESQ, and MAUQ) were analyzed using separate linear regression models. Each model included the follow-up score as the dependent variable and arm allocation (0=control; 1=intervention) as the only covariate. Models for outcomes measured at baseline and follow-up (UES and ESQ) additionally included the baseline values of the outcomes. Differences in the combined adherence to the system (0=<50% total data completion; 1=>50% total data completion)

were tested using Pearson chi-square test. The threshold for statistical significance in all the analyses was *P*=.05.

All outcomes were analyzed under the intention-to-treat principle using R (version 4.1; R Foundation for Statistical Computing) [39]. All data, including those from withdrawn participants, were included in the analyses.

Supplementary Analysis

A supplementary analysis of all outcomes was conducted to estimate complier average causal effect (CACE) [40]. We defined compliers as participants in the intervention group who viewed the progress report page at least once in the 12 weeks. The CACE analyses were performed using a 2-stage least squares regression with arm allocation as the instrumental variable.

Process Evaluation

We examined quantitative app use measures from data retrieved from Google Analytics (Google LLC). These were reported under the following three categories: (1) app engagement (user-initiated app opening and active weeks), (2) in-app interactions (questionnaire initiation, progress report viewing, viewed progress report >1 time, and progress report viewing duration), and (3) notification engagement (notifications received, notifications opened, and the percentage of notifications opened out of notifications received). The number of active weeks was calculated as the number of weeks the participant was active out of the total 12 weeks, with at least 3 screen view or user engagement metrics recorded per participant. Each indicator was summarized by arm as mean (SD) and median (IQR). The qualitative experiences of the study were thematically analyzed and reported as a brief narrative synthesis. These data will be reported in more detail elsewhere.

Results

Recruitment

A total of 347 individuals were contacted between April and May 2021. Of them, 114 (32.9%) agreed to participate, and 100 (28.8%) completed an enrollment session and were enrolled in the study. Enrollment sessions took place via email (89/100, 89%), video call (9/100, 9%), or phone call (2/100, 2%). Figure 2 details the participation rate and reasons for nonparticipation.

Sample Characteristics

All (100/100, 100%) participants completed the baseline outcome assessment, and 87 (87%) participants completed the 12-week follow-up assessment. Among the total 100 participants, 1 (1%) participant in the intervention group withdrew from the study before the 12-week point, citing technological issues with the study apps as the main reason for withdrawal. The follow-up period was from April to September 2021.

Baseline characteristics were similar between the 2 groups (Table 2). The groups contained an equal number of participants (n=50). The mean age of the sample was 53.3 (SD 14.3) years, and 76 (76%) of the 100 participants were female. Most participants reported mild (36/100, 36%) or moderate (29/100, 29%) symptoms of depression at enrollment, as measured by the Inventory of Depressive Symptomatology–Self-Report. Overall, 12 (12%) participants reported suicidal ideation at baseline. Among the 100 participants, 59 (59%) “strongly agreed” that they were confident in using the smartphone they were using for the study, and 51 (51%) “strongly agreed” that they were confident in using the Fitbit device.

Figure 2. Participant flowchart following CONSORT (Consolidated Standards of Reporting Trials) guidelines.

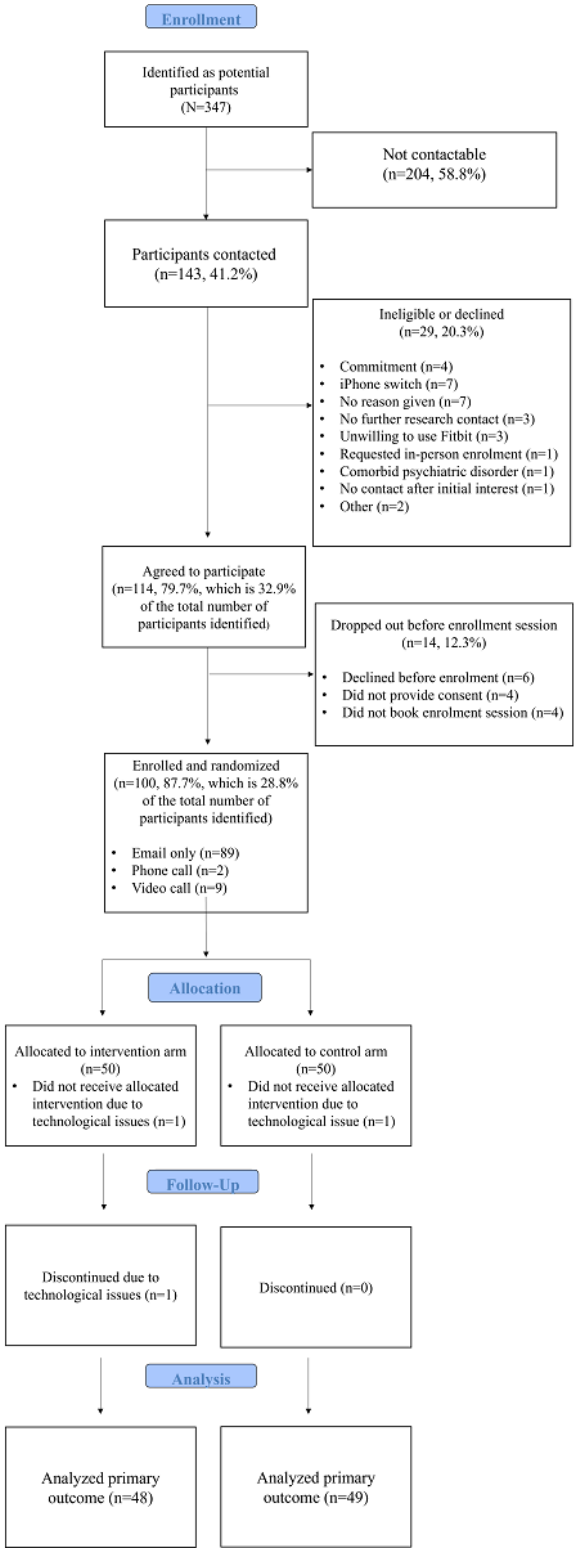


Table 2. Baseline characteristics and demographics of the study sample.

	Intervention (n=50)	Control (n=50)	Overall (N=100)
Age (y), mean (SD)	55.3 (12.7)	51.2 (15.7)	53.3 (14.3)
Gender, n (%)			
Women	40 (80)	36 (72)	76 (76)
Men	10 (20)	14 (28)	24 (24)
Ethnicity, n (%)			
Black or mixed ethnicity	3 (6)	3 (6)	6 (6)
White British	40 (80)	41 (82)	81 (81)
White other	5 (10)	4 (8)	9 (9)
Other	2 (4)	2 (4)	4 (4)
Total time in education (y), mean (SD)	20.2 (3.34)	20.5 (3.71)	20.4 (3.51)
Benefit receipt, n (%)			
Yes	24 (48)	23 (46)	47 (47)
No	26 (52)	27 (54)	53 (53)
Income (£; US $), n (%)			
<15,000 (US $18,828.67)	9 (18)	12 (24)	21 (21)
15,000-24,000 (US $18,828.67-$30,125.88)	8 (16)	9 (18)	17 (17)
24,000-40,000 (US $30,125.88-$50,209.8)	15 (30)	10 (20)	25 (25)
40,000-55,000 (US $50,209.8-$69,038.47)	11 (22)	7 (14)	18 (18)
>55,000 (US $69,038.47)	7 (14)	12 (24)	19 (19)
Employment status, n (%)			
Employed	24 (48)	25 (50)	49 (49)
Sick leave	0 (0)	3 (6)	3 (3)
Retired	20 (40)	14 (28)	34 (34)
Unemployed	2 (4)	4 (8)	6 (6)
Other	4 (8)	4 (8)	8 (8)
Current depression (continuous), mean (SD)[a]	24.8 (13.7)	26.5 (13.3)	25.7 (13.5)
Current depression (categorical), n (%) [a]			
None	10 (20)	8 (16)	18 (18)
Mild	20 (40)	16 (32)	36 (36)
Moderate	12 (24)	17 (34)	29 (29)
Severe	5 (10)	7 (14)	12 (12)
Very severe	3 (6)	2 (4)	5 (5)
Suicidal ideation, n (%)			
Yes	3 (6)	9 (18)	12 (12)
No	47 (94)	41 (82)	88 (88)
Current anxiety (continuous), mean (SD)[b]	6.34 (4.62)	7.10 (5.21)	6.72 (4.92)
Current anxiety (categories), n (%)[b]			
None	22 (44)	17 (34)	39 (39)
Mild	16 (32)	20 (40)	36 (36)
Moderate	9 (18)	6 (12)	15 (15)
Severe	3 (6)	7 (14)	10 (10)

	Intervention (n=50)	Control (n=50)	Overall (N=100)
Medical comorbidity, n (%)			
Yes	25 (50)	34 (68)	59 (59)
No	25 (50)	16 (32)	41 (41)
Functional disability, n (%)			
No impairment	14 (28)	17 (34)	31 (31)
Some impairment	17 (34)	17 (34)	34 (34)
Significant impairment	19 (38)	16 (32)	35 (35)
Life events in the past year, mean (SD)	0.680 (1.04)	0.920 (1.07)	0.800 (1.05)
Confidence in smartphone use, n (%)			
Strongly agree	29 (58)	30 (60)	59 (59)
Agree	15 (30)	16 (32)	31 (31)
Neither agree nor disagree	5 (10)	2 (4)	7 (7)
Disagree	0 (0)	2 (4)	2 (2)
Strongly disagree	1 (2)	0 (0)	1 (1)
Confidence in Fitbit (Fitbit Inc) use, n (%)			
Strongly agree	27 (54)	24 (48)	51 (51)
Agree	16 (32)	22 (44)	38 (38)
Neither agree nor disagree	6 (12)	3 (6)	9 (9)
Disagree	0 (0)	0 (0)	0 (0)
Strongly disagree	0 (0)	0 (0)	0 (0)
Not using Fitbit	1 (2)	1 (2)	2 (2)
Existing RADAR-MDD[c] status, n (%)			
Finished 2 years	0 (0)	2 (4)	2 (2)
Continuing past 2 years	13 (26)	17 (34)	30 (30)
Not reached 2 years	36 (72)	31 (62)	67 (67)
Withdrawn	1 (2)	0 (0)	1 (1)
Existing phone status, n (%)			
Existing Android (Google LLC)	30 (60)	27 (54)	57 (57)
Switching from iPhone (Apple Inc)	13 (26)	14 (28)	27 (27)
Switching from nonsmartphone	2 (4)	4 (8)	6 (6)
Upgrading existing Android	5 (10)	5 (10)	10 (10)

[a]Measured using the Inventory of Depressive Symptomatology–Self-Report. The maximum score possible is 84. The scores are categorized as follows: none=0-13, mild=14-25, moderate=26-38, severe=39-48, and very severe=49-84.

[b]Measured using the Generalized Anxiety Disorder-7 item scale. The maximum score possible is 21. The scores are categorized as follows: none=0-5, mild=6-10, moderate=11-15, and severe=16-21.

[c]RADAR-MDD: Remote Assessment of Disease and Relapse–Major Depressive Disorder.

Data Availability

Of the 100 participants, 97 (97%) provided any data via the study app, and 93 (93%) had any recorded Fitbit data. A total of 2 (2%) participants were unable to use either the study app or Fitbit immediately following enrollment, 1 (1%) received no notifications from the study app during follow-up, and 4 (4%) were unable to sync the Fitbit with their smartphone. Moreover, 1 (1%) participant opted out of using a Fitbit for the study at enrollment. As data from the sources were unavailable owing to technical limitations, rather than nonengagement, these participants were excluded from the respective analyses (3/100, 3% for primary analysis and 8/100, 8% for the secondary combined adherence analysis).

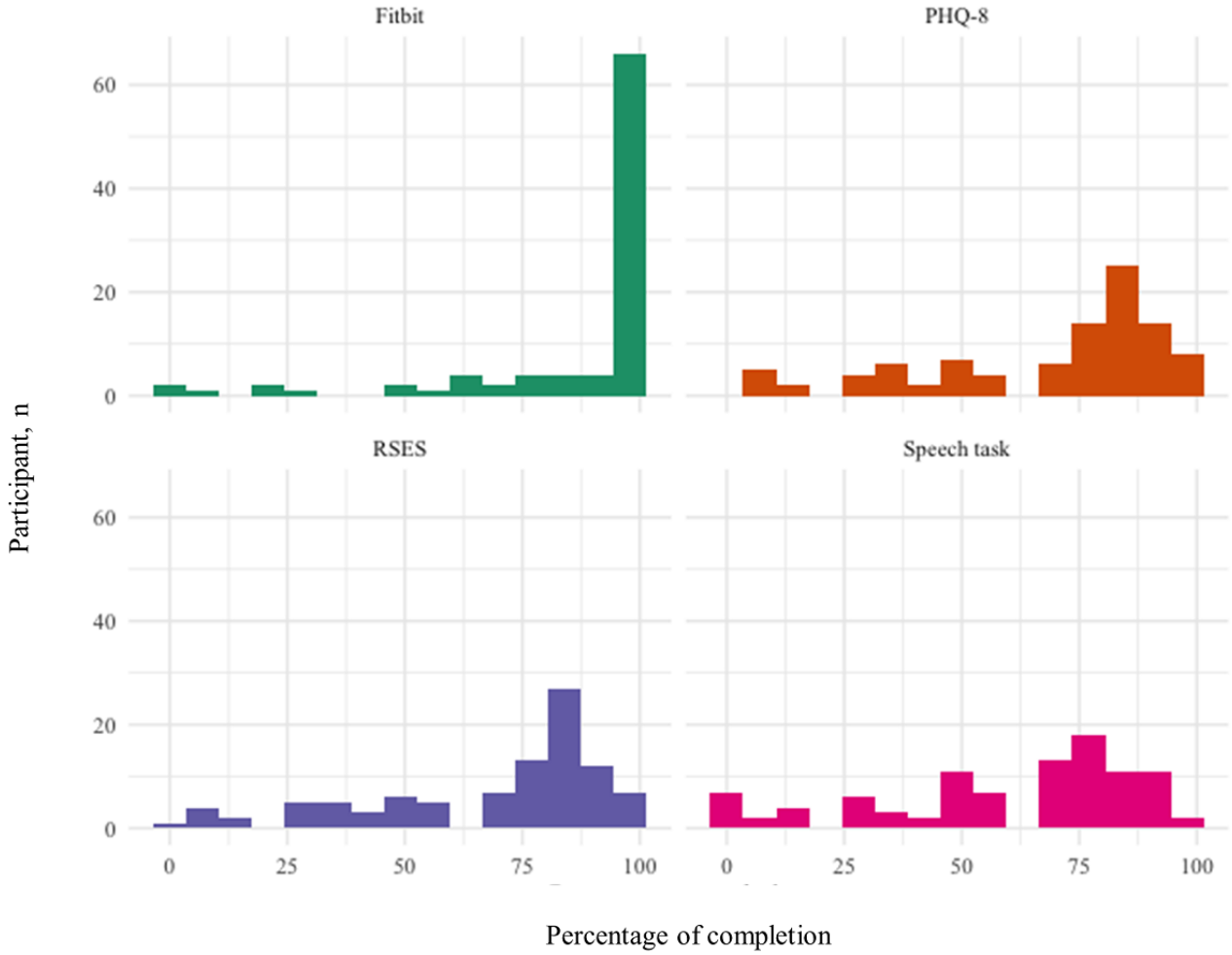
Each app task had a maximum count of 12 (1 per week). Overall, participants completed a median of 9 (IQR 6-10) PHQ-8 tasks, 9 (IQR 6-10) RSES tasks, and 8 (IQR 6-9) speech tasks. Among the 100 participants, 2 (2%) completed all available tasks, and 7 (7%) completed all available PHQ-8 tasks. A total of 35 (35%)

participants completed all 3 tasks at each point when they completed a PHQ-8 task.

The participants provided sufficient Fitbit data (at least 2 recordings per day) on a mean average of 74 (SD 19.7) days per participant during the 12-week (84-day) intervention period.

Figure 3 shows the percentage of completion for all 4 data sources across the sample.

Figure 3. Percentage of completion for the Fitbit and the 3 active tasks (Patient Health Questionnaire-8 [PHQ-8], Rosenberg Self-Esteem Scale [RSES], and speech task).



Primary and Secondary Outcomes

The primary analysis included 97 participants. The levels of completion of the PHQ-8 task were similar between the control (67/97, 69%) and intervention (66/97, 68%) arms (*P* value for the difference between the arms=.83, 95% CI −9.32 to 11.65).

For the secondary outcomes (Table 3), we found that those in the intervention group reported slightly higher UES (1.93, 95% CI −1.91 to 5.78), ESQ (1.13, 95% CI −2.93 to 5.19), and MAUQ (2.29, 95% CI −5.93 to 10.52) scores than those in the control group at follow-up. However, all CIs were wide and included 0.

Table 3. Linear regression model coefficients for each of the 3 secondary outcomes.

Subjective engagement outcome	Treatment effect (95% CI)	Participant (N=100), n (%)[a]
UES[b]	1.93 (−1.91 to 5.78)	89 (89)
ESQ[c]	1.13 (−2.93 to 5.19)	89 (89)
MAUQ[d,e]	2.29 (−5.93 to 10.52)	87 (87)

[a]For end point measures only.

[b]UES: User Engagement Scale.

[c]ESQ: Emotional Self-Awareness Questionnaire.

[d]MAUQ: mHealth App Usability Questionnaire.

[e]Only includes end point measure.

The combined adherence secondary analysis included 92 participants. The proportion of participants adhering to the system was similar between both arms (control=32/48, 67%; intervention=35/44, 80%; *P* value for difference between arms=.98).

For the CACE analyses, of the 48 participants in the intervention group, 29 (60%) met the complier definition of viewing the progress report at least once during the intervention period (Table 4). Table 5 presents the CACE treatment estimates for the primary and secondary outcomes. The between-arm difference in PHQ-8 completion was −1.92 (95% CI −19.93 to 15.54; *P*=.83), showing no evidence of a statistical difference. The treatment effect estimates for the UES, ESQ, and MAUQ were larger than the intention-to-treat estimates in favor of the intervention, but the effect sizes remained small.

Table 4. Number of participants and mean percentage of completion among the control group, intervention group compliers, and intervention group noncompliers (n=97).

Study arm	Participant, n (%)	Completion (%), mean (SD)
Intervention group compliers[a]	29 (30)	75.3 (23.9)
Intervention group noncompliers[b]	19 (20)	57.0 (28.0)
Control group	49 (50)	69.2 (25.1)

[a]Viewed the progress report module at least once during the intervention period.

[b]Did not view the progress report module during the intervention period.

Table 5. Intention-to-treat (ITT) and complier average causal effect (CACE) treatment estimates for the primary and secondary outcomes.

Outcome	Participant, n (%)	ITT		CACE[a]	
		Treatment effect (95% CI)	*P* value	Treatment effect (95% CI)	*P* value
PHQ-8[b]	97 (100)	−1.16 (−11.65 to 9.32)	.83	−1.92 (−19.39 to 15.54)	.83
UES[c]	89 (92)	1.93 (−1.91 to 5.78)	.32	3.49 (−3.75 to 10.73)	.34
ESQ[d]	89 (92)	1.13 (−2.93 to 5.19)	.58	2.03 (−5.32 to 9.38)	.58
MAUQ[e]	87 (90)	2.29 (−5.93 to 10.52)	.58	4.21 (−10.87 to 19.28)	.58

[a]Complier average causal effect estimates of intervention group compliers, defined as those who viewed the progress report module at least once during the intervention period.

[b]PHQ-8: Patient Health Questionnaire-8.

[c]UES: User Engagement Scale.

[d]ESQ: Emotional Self-Awareness Questionnaire.

[e]MAUQ: mHealth App Usability Questionnaire.

Process Evaluation

Table 6 presents the quantitative process evaluation measures collected throughout the study. Over the entire study period, participants in the intervention arm opened the app a mean average of 21.2 (SD 13.5) times, whereas the participants in the control group opened the app a mean average of 19.0 (SD 9.10) times. In total, 60% (35/58) of the participants who were able to view the progress report viewed it multiple times throughout the study, viewing for a mean average of 14.7 (SD 10.9) seconds per time. Participants in both groups received a similar number of notifications, although those in the control group opened a higher mean percentage (39.9, SD 25.9) of the notifications.

Table 6. Process evaluation use statistics by arm over the 12-week follow-up period.

	Intervention (n=48)	Control (n=49)
App engagement, mean (SD)		
User-initiated app opening	21.2 (13.5)	19.0 (9.10)
Active weeks[a]	8.96 (3.14)	8.88 (2.60)
In-app interactions		
Questionnaire initiation, mean (SD)	23.6 (10.3)	25.0 (10.5)
Progress report viewing		
Values, mean (SD)	3.60 (7.64)	N/A[b]
Values, median (IQR)	1.00 (0-2.25)	N/A
Viewed progress report >1 time , n (%)		
No	19 (40)	49 (100)
Yes	29 (60)	N/A
Progress report viewing duration (seconds), mean (SD)	14.7 (10.9)	N/A
Notification engagement		
Notifications received, mean (SD)	22.0 (13.3)	22.6 (9.82)
Notifications opened, mean (SD)	6.58 (6.45)	8.69 (6.36)
Percentage of notifications opened, mean (SD)[c]	34.3 (31.8)	39.9 (25.9)
None received, n (%)	2 (4)	1 (2)

[a]Calculated as the number of weeks over the 12-week period with at least 3 screen view or user engagement metrics recorded, as per Google Analytics (Google LLC) data.

[b]N/A: not applicable; participants in the control arm were unable to view the progress report.

[c]Percentage of notifications opened based on the total notifications received for each participant.

Participants generally liked the new in-app components. They felt that the progress report could motivate task completion by providing clarity that previous tasks had been successfully logged:

> *[It allowed me] certainly to feel more engaged and understand...how it contributes, but also to gain an overview of my own input into it, so not just I enter the data and it disappears.* [P99]

Most participants found the notifications somewhat informative:

> *The pop up things with little quotes about "doing this helps you"...yeah I liked those, I thought that was really good. You're doing it for a purpose.* [P29]

However, many participants were unsure whether they had seen all the notifications that their phone had received. Some highlighted the potentially demotivating effects of the progress report, depending on previous completion:

> *I think it depends what mood you're in...if I hadn't completed everything and I wasn't in a good mind space I could be thinking "ooh I've failed.*" [P40]

Whereas most participants agreed that the components might motivate others, the impact of the components on participants' own task completion was more nuanced. Instead, participating in the research study seemed to be the strongest motivation for task completion:

> *Because I had committed to do the study it meant that I said I am going to do it so I can't be half-hearted about it...I want to do the best I could because it was for somebody else's use.*" [P99]

Many participants discussed the beneficial effects of taking part in symptom monitoring generally, such as increased awareness of their depression and communication with others. Several additional in-app components were suggested, including a direct communication channel between the app and research team.

Harms and Protocol Violations

No adverse or serious adverse events were reported. Among the 100 participants, 1 (1%) withdrew owing to technological issues.

Discussion

Principal Findings

This study conducted the first, fully remote RCT of the RADAR-base symptom-tracking system to test the effect of additional in-app components, based on behavioral change theory, on objective and subjective engagement. Overall, objective engagement was high across the sample. We found that the participants who received the adapted system (incorporating theoretically informed notifications, real-time progress reports, and researcher contact details) did not show higher levels of engagement with the system than the participants who received the system as usual. Although

subjective engagement (emotional self-awareness, system utility, and usability) was slightly higher in those who received the adapted app, the difference was small and did not reach statistical significance.

Implications and Links With Previous Work

Previous research, both usability studies [19,21,22] and RCTs [20,24], has suggested that providing notifications and progress visualization can prompt objective engagement in remote symptom monitoring. We propose several explanations as to why our results did not reflect past findings.

First, our findings may reflect the sample used. Participants were recruited from a previous study that used the RADAR-base system. This meant that they had prior experience of and interest in symptom monitoring. Previous work has also highlighted the impact of the academic setting on engagement through altruistic motivations [41]. It is possible that our results reflect a ceiling effect, whereby participants in both groups were motivated to participate in the research and complete symptom monitoring regardless of the changes to the app. This is particularly apparent given that 2 of the in-app components were designed to reflect individual achievement and benefits, aspects that might not have been as relevant in this research context.

Second, the combination of in-app components used in the adapted system might not have been sufficiently tailored to the user. The development of the app was grounded in both behavioral theory [31] and user involvement [21], which suggested that viewing real-time progress and being reminded of the proposed benefits of symptom monitoring might combat the barriers to engagement. However, although these components are proposed to encourage future tracking behavior, in practice, it is unclear how they interact with the *motivation* section of the COM-B model, in this case, the low motivation linked to low mood in our cohort of people with depression. Previous work has focused on symptom tracking for substance abuse [20] or general population [18,24] cohorts, both of which might react to incentivization in different ways from those with depression. Our qualitative discussions indeed suggested that the impact of viewing data progress might be affected by individual mood and motivational fluctuations. The addition of other virtual incentives, such as gamification [42], might have been more effective in promoting engagement with the tasks here, alongside the ability to personalize which components are seen and when they are seen.

Moreover, our components were static in that they were accessible to all the participants in the intervention group at the same time and frequency. Previous work has suggested that several factors can significantly moderate the relationship between in-app components and engagement. For example, Nahum-Shani and colleagues [20] found that receiving data insights only increased the likelihood of future self-reporting in those who were not frequent users of the app, suggesting that visualizing progress is not incentivizing (or is even perhaps "irritating") for those who are actively engaged in the task from the beginning. Several studies have found a link between notification timing and engagement [20,24], although attempts at sensor-driven notification sending based on location have so far been unsuccessful in improving data availability [43]. Taken together, this suggests that future work is needed to understand the process of interacting with in-app components in this cohort.

Third, with regard to subjective engagement, the measures used in this study might not have reflected the experience of self-monitoring in the most nuanced manner. We used previous findings [6,41] to inform our operationalization of subjective engagement with RMTs as usability (UES), emotional self-awareness (ESQ), and utility (MAUQ). Our qualitative evaluation suggested that participants generally saw the in-app components as helpful in increasing task completion, which, in turn, might have promoted the feelings of emotional self-awareness they gained from monitoring their symptoms. We also saw that participants who viewed the progress report did so for around 15 seconds at a time, often repeatedly, which suggests a sustained interest in viewing progress. Although we did not see significant differences in either objective or subjective engagement, we did see slight treatment effects for all 3 subjective measures, which were higher still when adjusting for those who viewed the progress report. It is possible that different measures might have revealed a more significant change. For example, the UES is a tool designed primarily for digital health interventions and measures concepts such as focused attention, which are not as relevant to RMTs [6]. Measures tapping into other aspects of the experience of symptom monitoring, such as being seen as an individual [44] or the provision of a safety net [45], might have provided a more detailed understanding of the interaction among the in-app components, objective engagement, and subjective engagement in the study; however, to our knowledge, these have yet to be developed.

Strengths and Limitations

To our knowledge, this was the first study to attempt to quantify the effect of in-app components on objective and subjective engagement with a multiparametric symptom-tracking system for depression. We used an established system that was previously used to conduct the largest, longitudinal study on RMT in MDD to date [12] and demonstrated the successful transference of the system to a remote RCT design. Methodologically, this study laid the foundation for future work to measure both objective and subjective engagements with symptom-tracking devices. We used an adapted system with in-app components, which allowed for an active control group (the system as usual) and embedded data collection to reduce confounding factors associated with the delivery medium [46]. In reference to our first aim, we have shown good data availability in the first fully remote trial of the system, with 87% (87/100) of the participants completing follow-up data collection, a median of 75% completion of symptom-tracking tasks, and a mean of 74 of 84 days of wear time data without planned researcher contact.

There are several limitations to this study. First, as mentioned, the sample was previously engaged in remote symptom tracking and driven by research altruism. This allowed for the recruitment of a large sample from an established group, obtaining results quickly and efficiently. However, it is unclear how far these results might generalize to community cohorts using symptom tracking in their daily life. Second, the study was conducted

during the COVID-19 lockdown periods in the United Kingdom. A combination of increased free time and interest in health tracking could have resulted in increased engagement rates. Third, despite the large sample size, the study did not reach the intended number of participants needed to achieve the optimum statistical power. Fourth, although the app design was grounded in previous research, working within the confines of an established system gave way to certain design constraints. Some additional facilitators that arose from the COM-B analysis, such as the in-app reporting of technological malfunctions, could not be included or assessed in terms of their impact on engagement.

Avenues for Future Work

Future work should use these findings as a basis for further RCTs quantifying the effects of RMT system design on objective and subjective engagement with remote symptom tracking. Context-specific, dynamic tailoring of notifications and data insights could be key here. Although in-app components reduce the need for human resources, the impact of external factors should not be dismissed. Our system amendments did not promote engagement over and above the system as usual; future work could seek to understand how incentives such as research team support could interact with in-app components to increase engagement, such as the use of supportive chatbots [23]. Of major importance is replicating this work with different cohorts. Using the adapted system with non–help-seeking participants or those with lower technological literacy might affect the impact of the components that we tested. For example, the impact of the theoretically informed notifications might be greater in those who are less aware of the proposed benefits of symptom monitoring. Similarly, engagement with the app is likely to vary if the app is implemented in clinical practice; progress tracking and notification content might be more impactful for those who use the system for their own direct benefit. This work could also seek to complement the RCT design with additional analysis manuscripts for increased insight into the impact of UI features. For example, this could include correlational analyses of in-app component use with the measures of objective and subjective engagement or exploring whether baseline demographics are predictive of engagement in such trials. Another area for exploration is the measurement of the subjective experience of remote symptom tracking. The development of a suitable instrument that encapsulates experiential engagement would propel the understanding of the promotion of engagement across the field.

Conclusions

This study found that a combination of informative notifications, progress visualization, and research team contact details did not increase engagement in remote symptom tracking in our research cohort. However, the system provided good data availability, and the process evaluation measures suggested that participants saw benefits in using the adapted system. We have provided the methodology and scope for future exploration in this area, as well as opportunities to replicate this work in both community and clinical cohorts to further the promotion of engagement in remote health symptom tracking for both data collection and clinical management.

Acknowledgments

This paper represents independent research partly funded by the National Institute for Health and Care Research (NIHR) Maudsley Biomedical Research Centre at South London and Maudsley National Health Service (NHS) Foundation Trust and King's College London. The views expressed are those of the authors and not necessarily those of the NHS, NIHR, or Department of Health and Social Care.

The Remote Assessment of Disease and Relapse–Central Nervous System project received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement number 115902. This joint undertaking receives support from the European Union's Horizon 2020 research and innovation program and European Federation of Pharmaceutical Industries and Associations (EFPIA). This communication reflects the views of the RADAR-CNS consortium, and neither Innovative Medicines Initiative nor the European Union and EFPIA are liable for any use that may be made of the information contained herein. The funding body was not involved in the design of the study or the collection, analysis, or interpretation of the data. This research was reviewed by a team with experience of mental health problems and their carers who were specially trained to advise on research proposals and documentation through the Feasibility and Acceptability Support Team for Researchers: a free, confidential service in England provided by the NIHR Maudsley Biomedical Research Centre via King's College London and South London and Maudsley NHS Foundation Trust. Finally, the authors would like to thank all members of the RADAR-CNS patient advisory board who have experience of living with or supporting those who are living with depression, epilepsy, or multiple sclerosis.

Authors' Contributions

KMW designed, coordinated, and conducted the study as well as led data processing, data coding, data analysis, and the writing of this manuscript. EC contributed to the design of the study and advised on data analysis. DL contributed to the design of the study and aided with data extraction. FM contributed to the design of the study. PC contributed to in-app component development and data extraction. YR contributed to app development and data extraction. SS contributed to service user involvement research and in-app component development. ED-L contributed to service user involvement research and in-app component development. LW contributed to the coding and thematic analysis of the interview data. CH contributed to the development and design of the study, reviewed data analysis, and supervised the first author. MH contributed to the development and design of the study, reviewed data analysis, and supervised the first author.

Conflicts of Interest

MH is the principal investigator of the RADAR-CNS consortium, a private-public precompetitive consortium with research funding from Janssen, Union Chimique Belge, Merck Sharp & Dohme, Biogen, and Lundbeck. All other authors declare no other conflicts of interest.

Multimedia Appendix 1
Speech task paragraph.
[DOCX File , 10 KB - mhealth_v12i1e44214_app1.docx]

Multimedia Appendix 2
Development of the adapted system based on the capability, opportunity, motivation, and behavior framework and service user research.
[DOCX File , 243 KB - mhealth_v12i1e44214_app2.docx]

Multimedia Appendix 3
Semistructured interview schedules.
[DOCX File , 21 KB - mhealth_v12i1e44214_app3.docx]

Multimedia Appendix 4
CONSORT checklist.
[PDF File (Adobe PDF File), 65 KB - mhealth_v12i1e44214_app4.pdf]

References

1. Liao Y, Thompson C, Peterson S, Mandrola J, Beg MS. The future of wearable technologies and remote monitoring in health care. Am Soc Clin Oncol Educ Book 2019 Jan;39:115-121 [FREE Full text] [doi: 10.1200/EDBK_238919] [Medline: 31099626]
2. Verduijn J, Verhoeven JE, Milaneschi Y, Schoevers RA, van Hemert AM, Beekman AT, et al. Reconsidering the prognosis of major depressive disorder across diagnostic boundaries: full recovery is the exception rather than the rule. BMC Med 2017 Dec 12;15(1):215 [FREE Full text] [doi: 10.1186/s12916-017-0972-8] [Medline: 29228943]
3. Matcham F, Barattieri di San Pietro C, Bulgari V, de Girolamo G, Dobson R, Eriksson H, RADAR-CNS consortium. Remote assessment of disease and relapse in major depressive disorder (RADAR-MDD): a multi-centre prospective cohort study protocol. BMC Psychiatry 2019 Feb 18;19(1):72 [FREE Full text] [doi: 10.1186/s12888-019-2049-z] [Medline: 30777041]
4. Naslund JA, Marsch LA, McHugo GJ, Bartels SJ. Emerging mHealth and eHealth interventions for serious mental illness: a review of the literature. J Ment Health 2015 May 28;24(5):321-332 [FREE Full text] [doi: 10.3109/09638237.2015.1019054] [Medline: 26017625]
5. Godkin FE, Turner E, Demnati Y, Vert A, Roberts A, Swartz RH, et al. Feasibility of a continuous, multi-sensor remote health monitoring approach in persons living with neurodegenerative disease. J Neurol 2022 May;269(5):2673-2686 [FREE Full text] [doi: 10.1007/s00415-021-10831-z] [Medline: 34705114]
6. White KM, Williamson C, Bergou N, Oetzmann C, de Angel V, Matcham F, et al. A systematic review of engagement reporting in remote measurement studies for health symptom tracking. NPJ Digit Med 2022 Jun 29;5(1):82 [FREE Full text] [doi: 10.1038/s41746-022-00624-7] [Medline: 35768544]
7. Mohr DC, Zhang M, Schueller SM. Personal sensing: understanding mental health using ubiquitous sensors and machine learning. Annu Rev Clin Psychol 2017 May 08;13(1):23-47 [FREE Full text] [doi: 10.1146/annurev-clinpsy-032816-044949] [Medline: 28375728]
8. Baumel A, Muench F, Edan S, Kane JM. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. J Med Internet Res 2019 Sep 25;21(9):e14567 [FREE Full text] [doi: 10.2196/14567] [Medline: 31573916]
9. Leightley D, Williamson C, Rona RJ, Carr E, Shearer J, Davis JP, et al. Evaluating the efficacy of the drinks: ration mobile app to reduce alcohol consumption in a help-seeking military veteran population: randomized controlled trial. JMIR Mhealth Uhealth 2022 Jun 20;10(6):e38991 [FREE Full text] [doi: 10.2196/38991] [Medline: 35724966]
10. Milward J, Khadjesari Z, Fincham-Campbell S, Deluca P, Watson R, Drummond C. User preferences for content, features, and style for an app to reduce harmful drinking in young adults: analysis of user feedback in app stores and focus group interviews. JMIR Mhealth Uhealth 2016 May 24;4(2):e47 [FREE Full text] [doi: 10.2196/mhealth.5242] [Medline: 27220371]
11. Simblett S, Greer B, Matcham F, Curtis H, Polhemus A, Ferrão J, et al. Barriers to and facilitators of engagement with remote measurement technology for managing health: systematic review and content analysis of findings. J Med Internet Res 2018 Jul 12;20(7):e10480 [FREE Full text] [doi: 10.2196/10480] [Medline: 30001997]

12. Matcham F, Leightley D, Siddi S, Lamers F, White KM, Annas P, RADAR-CNS consortium. Remote Assessment of Disease and Relapse in Major Depressive Disorder (RADAR-MDD): recruitment, retention, and data availability in a longitudinal remote measurement study. BMC Psychiatry 2022 Feb 21;22(1):136 [FREE Full text] [doi: 10.1186/s12888-022-03753-1] [Medline: 35189842]
13. White KM, Dawe-Lane E, Siddi S, Lamers F, Simblett S, Riquelme Alacid G, et al. Understanding the subjective experience of long-term remote measurement technology use for symptom tracking in people with depression: multisite longitudinal qualitative analysis. JMIR Hum Factors 2023 Jan 26;10:e39479 [FREE Full text] [doi: 10.2196/39479] [Medline: 36701179]
14. Truelove S, Vanderloo LM, Tucker P, Di Sebastiano KM, Faulkner G. The use of the behaviour change wheel in the development of ParticipACTION's physical activity app. Prev Med Rep 2020 Dec;20:101224 [FREE Full text] [doi: 10.1016/j.pmedr.2020.101224] [Medline: 33134041]
15. Michie S, Atkins L, West R. The Behaviour Change Wheel: A Guide to Designing Interventions. Surrey, UK: Silverback Publishing; 2014.
16. Simblett S, Matcham F, Siddi S, Bulgari V, Barattieri di San Pietro C, Hortas López J, RADAR-CNS Consortium. Barriers to and facilitators of engagement with mHealth technology for remote measurement and management of depression: qualitative analysis. JMIR Mhealth Uhealth 2019 Jan 30;7(1):e11325 [FREE Full text] [doi: 10.2196/11325] [Medline: 30698535]
17. Fogg BJ. Persuasive technology: using computers to change what we think and do. Ubiquity 2002 Dec;2002(December). [doi: 10.1145/764008.763957]
18. Bidargaddi N, Pituch T, Maaieh H, Short C, Strecher V. Predicting which type of push notification content motivates users to engage in a self-monitoring app. Prev Med Rep 2018 Sep;11:267-273 [FREE Full text] [doi: 10.1016/j.pmedr.2018.07.004] [Medline: 30109172]
19. Rabbi M, Philyaw-Kotov M, Li J, Li K, Rothman B, Giragosian L, et al. Translating behavioral theory into technological interventions: case study of an mHealth app to increase self-reporting of substance-use related data. arXiv. Preprint posted online March 30, 2020 2023 [FREE Full text]
20. Nahum-Shani I, Rabbi M, Yap J, Philyaw-Kotov ML, Klasnja P, Bonar EE, et al. Translating strategies for promoting engagement in mobile health: a proof-of-concept microrandomized trial. Health Psychol 2021 Dec;40(12):974-987 [FREE Full text] [doi: 10.1037/hea0001101] [Medline: 34735165]
21. Polhemus A, Novak J, Majid S, Simblett S, Morris D, Bruce S, et al. Data visualization for chronic neurological and mental health condition self-management: systematic review of user perspectives. JMIR Ment Health 2022 Apr 28;9(4):e25249 [FREE Full text] [doi: 10.2196/25249] [Medline: 35482368]
22. Li I, Dey AK, Forlizzi J. Understanding my data, myself: supporting self-reflection with ubicomp technologies. In: Proceedings of the 13th international conference on Ubiquitous computing. 2011 Sep Presented at: Ubicomp '11: The 2011 ACM Conference on Ubiquitous Computing; September 17-21, 2011; Beijing, China. [doi: 10.1145/2030112.2030166]
23. Perski O, Crane D, Beard E, Brown J. Does the addition of a supportive chatbot promote user engagement with a smoking cessation app? An experimental study. Digit Health 2019 Sep 30;5:2055207619880676 [FREE Full text] [doi: 10.1177/2055207619880676] [Medline: 31620306]
24. Bidargaddi N, Almirall D, Murphy S, Nahum-Shani I, Kovalcik M, Pituch T, et al. To prompt or not to prompt? A microrandomized trial of time-varying push notifications to increase proximal engagement with a mobile health app. JMIR Mhealth Uhealth 2018 Nov 29;6(11):e10123 [FREE Full text] [doi: 10.2196/10123] [Medline: 30497999]
25. White KM, Matcham F, Leightley D, Carr E, Conde P, Dawe-Lane E, et al. Exploring the effects of in-app components on engagement with a symptom-tracking platform among participants with major depressive disorder (RADAR-Engage): protocol for a 2-armed randomized controlled trial. JMIR Res Protoc 2021 Dec 21;10(12):e32653 [FREE Full text] [doi: 10.2196/32653] [Medline: 34932005]
26. Ranjan Y, Rashid Z, Stewart C, Conde P, Begale M, Verbeeck D, et al. RADAR-Base: open source mobile health platform for collecting, monitoring, and analyzing data using sensors, wearables, and mobile devices. JMIR Mhealth Uhealth 2019 Aug 01;7(8):e11734 [FREE Full text] [doi: 10.2196/11734] [Medline: 31373275]
27. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
28. Rush AJ, Gullion CM, Basco MR, Jarrett RB, Trivedi MH. The Inventory of Depressive Symptomatology (IDS): psychometric properties. Psychol Med 1996 May 09;26(3):477-486. [doi: 10.1017/s0033291700035558] [Medline: 8733206]
29. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
30. Rosenberg M. Society and the Adolescent Self-Image. Princeton, NJ: Princeton University Press; 1965.
31. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011 Apr 23;6(1):42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]

32. Simblett S, Dawe-Lane E, Gilpin G, Morris D, White K, Erturk S, RADAR-CNS Consortium. Data visualisation preferences in remote measurement technology for individuals living with depression, epilepsy, and multiple sclerosis: a qualitative study. JMIR Preprints. Preprint posted online November 21, 2022 2023 [FREE Full text] [doi: 10.2196/preprints.43954]
33. O'Brien HL, Toms EG. The development and evaluation of a survey to measure user engagement. J Am Soc Inf Sci Technol 2009 Oct 19;61(1):50-69. [doi: 10.1002/asi.21229]
34. Holdener M, Gut A, Angerer A. Applicability of the user engagement scale to mobile health: a survey-based quantitative study. JMIR Mhealth Uhealth 2020 Jan 03;8(1):e13244 [FREE Full text] [doi: 10.2196/13244] [Medline: 31899454]
35. O'Brien HL, Cairns P, Hall M. A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form. Int J Hum Comput Stud 2018 Apr;112:28-39. [doi: 10.1016/j.ijhcs.2018.01.004]
36. Killian KD. Development and validation of the Emotional Self-Awareness Questionnaire: a measure of emotional intelligence. J Marital Fam Ther 2012 Jul;38(3):502-514. [doi: 10.1111/j.1752-0606.2011.00233.x] [Medline: 22804468]
37. Zhou L, Bao J, Setiawan IM, Saptono A, Parmanto B. The mHealth App Usability Questionnaire (MAUQ): development and validation study. JMIR Mhealth Uhealth 2019 Apr 11;7(4):e11500 [FREE Full text] [doi: 10.2196/11500] [Medline: 30973342]
38. Van Til K, McInnis MG, Cochran A. A comparative study of engagement in mobile and wearable health monitoring for bipolar disorder. Bipolar Disord 2020 Mar 25;22(2):182-190 [FREE Full text] [doi: 10.1111/bdi.12849] [Medline: 31610074]
39. RStudio: integrated development environment for R. RStudio Team. Boston, MA: RStudio; 2015. URL: http://www.rstudio.com/ [accessed 2023-02-01]
40. Peugh JL, Strotman D, McGrady M, Rausch J, Kashikar-Zuck S. Beyond intent to treat (ITT): a complier average causal effect (CACE) estimation primer. J Sch Psychol 2017 Feb;60:7-24. [doi: 10.1016/j.jsp.2015.12.006] [Medline: 28164801]
41. White KM, Dawe-Lane E, Siddi S, Lamers F, Simblett S, Riquelme Alacid G, et al. Understanding the subjective experience of long-term remote measurement technology use for symptom tracking in people with depression: multisite longitudinal qualitative analysis. JMIR Hum Factors 2023 Jan 26;10:e39479 [FREE Full text] [doi: 10.2196/39479] [Medline: 36701179]
42. Comello ML, Qian X, Deal AM, Ribisl KM, Linnan LA, Tate DF. Impact of game-inspired infographics on user engagement and information processing in an eHealth program. J Med Internet Res 2016 Sep 22;18(9):e237 [FREE Full text] [doi: 10.2196/jmir.5976] [Medline: 27658469]
43. Morrison LG, Hargood C, Pejovic V, Geraghty AW, Lloyd S, Goodman N, et al. The effect of timing and frequency of push notifications on usage of a smartphone-based stress management intervention: an exploratory trial. PLoS One 2017 Jan 3;12(1):e0169162 [FREE Full text] [doi: 10.1371/journal.pone.0169162] [Medline: 28046034]
44. Gustavell T, Sundberg K, Langius-Eklöf A. Using an interactive app for symptom reporting and management following pancreatic cancer surgery to facilitate person-centered care: descriptive study. JMIR Mhealth Uhealth 2020 Jun 17;8(6):e17855 [FREE Full text] [doi: 10.2196/17855] [Medline: 32554375]
45. Faurholt-Jepsen M, Vinberg M, Frost M, Debel S, Margrethe Christensen E, Bardram JE, et al. Behavioral activities collected through smartphones and the association with illness activity in bipolar disorder. Int J Methods Psychiatr Res 2016 Dec 01;25(4):309-323 [FREE Full text] [doi: 10.1002/mpr.1502] [Medline: 27038019]
46. Murray E, Hekler EB, Andersson G, Collins LM, Doherty A, Hollis C, et al. Evaluating digital health interventions: key questions and approaches. Am J Prev Med 2016 Nov;51(5):843-851 [FREE Full text] [doi: 10.1016/j.amepre.2016.06.008] [Medline: 27745684]

Abbreviations

CACE: complier average causal effect
COM-B: capability, opportunity, motivation, and behavior
ESQ: Emotional Self-Awareness Questionnaire
MAUQ: mHealth App Usability Questionnaire
MDD: major depressive disorder
PHQ-8: Patient Health Questionnaire-8
RADAR-MDD: Remote Assessment of Disease and Relapse–Major Depressive Disorder
RCT: randomized controlled trial
REDCap: Research Electronic Data Capture
RMT: remote measurement technology
RSES: Rosenberg Self-Esteem Scale
UES: User Engagement Scale
UI: user interface

Edited by L Buis; submitted 30.11.22; peer-reviewed by VRD Herbuela, J Meyer, A Daros; comments to author 07.03.23; revised version received 21.05.23; accepted 09.06.23; published 19.01.24.

Please cite as:
White KM, Carr E, Leightley D, Matcham F, Conde P, Ranjan Y, Simblett S, Dawe-Lane E, Williams L, Henderson C, Hotopf M Engagement With a Remote Symptom-Tracking Platform Among Participants With Major Depressive Disorder: Randomized Controlled Trial
JMIR Mhealth Uhealth 2024;12:e44214
URL: https://mhealth.jmir.org/2024/1/e44214
doi:10.2196/44214
PMID:38241070

Original Paper

Effectiveness of a Nurse-Led Mobile-Based Health Coaching Program for Patients With Prostate Cancer at High Risk of Metabolic Syndrome: Randomized Waitlist Controlled Trial

Kyoungjin Lee[1,2], PhD; Jeongok Park[3], PhD; Eui Geum Oh[3], PhD; JuHee Lee[3], PhD; Chang Park[4], PhD; Young Deuk Choi[5], PhD

[1]College of Nursing and Brain Korea 21 FOUR Project, Yonsei University, Seoul, Republic of Korea

[2]College of Nursing, Kyungbok University, Namyangju, Republic of Korea

[3]Mo-Im Kim Nursing Research Institute, College of Nursing, Yonsei University, Seoul, Republic of Korea

[4]Department of Population Health Nursing Science, College of Nursing, University of Illinois at Chicago, Chicago, IL, United States

[5]College of Medicine, Yonsei University, Seoul, Republic of Korea

Corresponding Author:
Jeongok Park, PhD
Mo-Im Kim Nursing Research Institute, College of Nursing
Yonsei University
Seodaemun-gu, Seoul, Korea
Seoul, 03722
Republic of Korea
Phone: 82 2 2228 3390
Fax: 82 2 2227 8303
Email: JOPARK02@yuhs.ac

Abstract

Background: Androgen deprivation therapy (ADT), a standard treatment for prostate cancer (PC), causes many physical side effects. In particular, it causes metabolic changes such as fasting glucose abnormalities or accumulation of body fat, and its continuation can lead to metabolic syndrome (MetS), which is closely related to diabetes and cardiovascular disease. Therefore, it is important to maintain and practice a healthy lifestyle in patients with PC.

Objective: This study aims to evaluate the effectiveness of a nurse-led mobile-based program that aims to promote a healthy lifestyle in patients with PC undergoing ADT with MetS risk factors.

Methods: This was a single-blind, randomized, waitlist control interventional study. A total of 48 patients were randomly assigned to the experimental and waitlist control groups at the urology cancer clinic of a tertiary general hospital in South Korea. The inclusion criteria were patients who had undergone ADT for >6 months, had at least 1 of the 5 MetS components in the abnormal range, and could access a mobile-based education program. The experimental group attended a 4-week mobile-based program on exercise and diet that included counseling and encouragement to maintain a healthy lifestyle, whereas the control group was placed on a waitlist and received usual care during the follow-up period, followed by the intervention. The primary outcome was a change in the lifestyle score. The secondary outcomes were changes in 5 MetS components, body composition, and health-related quality of life. The outcomes were measured at 6 weeks and 12 weeks after the initiation of the intervention. Each participant was assigned to each group in a sequential order of enrollment in a 4×4 permuted block design randomization table generated in the SAS (SAS Institute) statistical program. A linear mixed model was used for statistical analysis.

Results: A total of 24 participants were randomly assigned to each group; however, 2 participants in the experimental group dropped out for personal reasons before starting the intervention. Finally, 46 participants were included in the intention-to-treat analysis. The experimental group showed more positive changes in the healthy lifestyle score (β=29.23; P≤.001), level of each MetS component (fasting blood sugar: β=−12.0; P=.05 and abdominal circumference: β=−2.49; P=.049), body composition (body weight: β=−1.52; P<.001 and BMI: β=−0.55; P<.001), and the urinary irritative and obstructive domain of health-related quality of life (β=14.63; P<.001) over time than the waitlist control group.

Conclusions: Lifestyle changes through nurse-led education can improve level of each MetS components, body composition, and ADT side effects. Nurses can induce positive changes in patients' lifestyles and improve the self-management of patients starting ADT through this program.

Trial Registration: Clinical Research Information Service KCT0006560; http://tinyurl.com/yhvj4vwh

(JMIR Mhealth Uhealth 2024;12:e47102) doi:10.2196/47102

KEYWORDS

nurses; prostate neoplasms; healthy lifestyle; metabolic syndrome; exercise; diet; mobile phone

Introduction

Adverse Effects of Androgen Deprivation Therapy in Prostate Cancer

Prostate cancer (PC) is a commonly occurring cancer in men worldwide, and it is the fourth most common cancer among adult men in South Korea [1]. The incidence of PC in South Korea has increased rapidly in recent years, from 2.2% in 2000 to 14.3% in 2018 [1], owing to changes in dietary patterns and the development of new diagnostic technologies. PC is affected by androgenic activity in the body [2]. Therefore, the goal of treatment is to either remove androgens using castration or neutralize the effects of androgens. Androgen deprivation therapy (ADT), which disrupts the mechanisms that create androgens, is a standard treatment for PC [3]. Initially, ADT was used to treat patients with metastatic PC or older persons with cancer with limited treatment options. However, in recent years, there has been a gradual expansion in the use of ADT to include treating patients in the early stages of PC [4,5].

Repeated ADT can lead to climacteric symptoms in male individuals [6], including sexual dysfunction, hot flashes [7], gynecomastia [8], depression, fatigue, changes in sleep patterns [9], loss of muscle strength [10,11], osteoporosis [12], metabolic syndrome (MetS) [13], and cardiovascular disease [14,15]. These issues can reduce the quality of life of patients and may even be life threatening [16,17]. Specifically, evidence suggests that ADT leads to metabolic changes [13,18]. MetS, also called insulin resistance syndrome, is the clustering of several risk factors associated with obesity. It is closely related to diabetes [14,15,19] and has also been identified as a major cause of cardiovascular disease owing to its association with dyslipidemia, diabetes, and hypertension [14,15]. PC and MetS have a close relationship: patients with PC who received ADT were found to be 2.5 times more likely to have MetS than those who did not receive ADT [18]. A cross-sectional analysis using Korean National Health Insurance Service data found that the prevalence rate of MetS in patients with PC was 40.1%, which is higher than the 34.5% prevalence rate in older Korean male adults [20]. In addition, as age increases, MetS has been found to increase by 8 times in patients with PC [21]. Given that most patients with PC are older adults, ADT-induced metabolic problems coupled with aging can lead to deterioration of health, resulting in cardiovascular disease or death [22]. Despite the necessity of severe side effects management for ADT, >50% of patients with PC are unaware of these problems [23], and most health care providers tend to focus on cancer treatment. Moreover, no protocols have been developed to manage ADT-induced MetS in patients with PC [24,25].

Importance of Improving Self-Management Skill for ADT-Induced MetS

ADT-induced MetS causes metabolic changes that are different from those caused by classic MetS including changes in body composition [26]. Androgens are a group of male sex hormones, and androgen level reduction leads to an increase in body fat percentage and a decrease in fat-free mass, such as muscle loss and decreased bone density. Although classic MetS is characterized by an increase in visceral fat, there is insufficient evidence to determine whether or not this also occurs in patients with PC [27,28]. Implementing a classic MetS management program is not suitable for patients with PC because ADT-induced MetS has different characteristics than classic MetS, and each patient experiences different side effects, including physical and emotional problems. For example, 80% of patients with PC experience hot flashes and 12% to 14% experience breast tenderness and depression [29].

MetS is a representative chronic lifestyle disease that requires individual self-management. The World Health Organization emphasizes that lifestyle is the most critical factor affecting health status [30]. The current health care system appears to focus its attention and investments toward the discovery and treatment of the causes of disease rather than lifestyle. However, lifestyle modification is the most effective way to reduce the incidence of cardiovascular complications caused by MetS [31]. In general, lifestyle modifications, including exercise and a nutritional diet, are known to result in lower occurrence rates of MetS and a lower risk of cardiovascular disease [31]. Such lifestyle modifications require improving self-management skills. Self-management is a lifelong task that requires engagement in activities that promote good health [32]. Improving self-management allows patients to maintain active lives, leading to a better quality of life [32].

Although the optimal duration of ADT remains undefined, patients with PC typically receive ADT for 2 to 3 years [33]. Furthermore, as metabolic changes begin at least 3 months after ADT initiation, it is very important for patients with PCs to improve their self-management skills from the beginning of ADT. In addition, health care providers should be aware of ADT-related side effects from an early stage and mediate metabolic changes of PC through education. Therefore, newly developed health coaching programs for patients with PC should be configured differently from previous iterations [34-36]. In other words, the focus should be on improving self-management skills to encourage lifestyle changes that take into consideration the side effects of ADT from the beginning of ADT administration. However, effective self-management can be

challenging to maintain. In addition, the COVID-19 pandemic has diminished physical activity and nutritional quality worldwide [37]. The World Health Organization has emphasized the importance of exercise and maintaining a healthy lifestyle during the pandemic [38]. Web-based education has relatively few limitations in terms of time and space, and it has the advantage of being able to feature various types of media and teaching and learning materials compared with offline education programs [39].

Aims

This study aimed to evaluate the effectiveness of a nurse-led mobile-based health coaching program that promotes healthy lifestyle changes, normal range of MetS components, and health-related quality of life (HRQoL) for patients with PC receiving ADT. The primary aim was to identify the changes in (1) healthy lifestyle through this program, and the secondary aims were to identify the changes in (1) the levels of each MetS component, including blood pressure, fasting blood sugar (FBS), high-density lipoprotein (HDL) cholesterol, triglyceride, and abdominal circumference (AC); (2) body composition, such as body weight, BMI, skeletal muscle mass, fat mass, and fat percentage; and (3) quality of life.

Methods

Participants

The study population consisted of men who were diagnosed with PC at the urology cancer center of a single tertiary general hospital in South Korea. The inclusion criteria were (1) patients with PC who had been receiving ADT for >6 months at the time of enrollment in this study, (2) those with an abnormal range of at least 1 of the 5 MetS components, (3) those who were fully aware of the object and contents of the study and voluntarily participated, (4) those who understood spoken and written Korean and who could communicate without cognitive impairment, and (5) those who had a smartphone and were able to access the mobile-based education program. The exclusion criteria were (1) patients diagnosed with and treated for cancers other than PC; (2) those who had undergone surgery or chemotherapy for <3 months before the start of the study; (3) those who answered “yes” to ≥1 question of the Physical Activity Readiness Questionnaires; (4) those with cardiovascular diseases such as unstable angina pectoris, uncontrolled blood pressure, myocardial infarction, or comorbidities such as a musculoskeletal or nervous system disease; (5) those who were participating in other research programs; (6) those who had a change in medication to control their blood pressure, cholesterol, or blood sugar <3 months before the start of the study; and (7) those who had difficulty with typical daily activities.

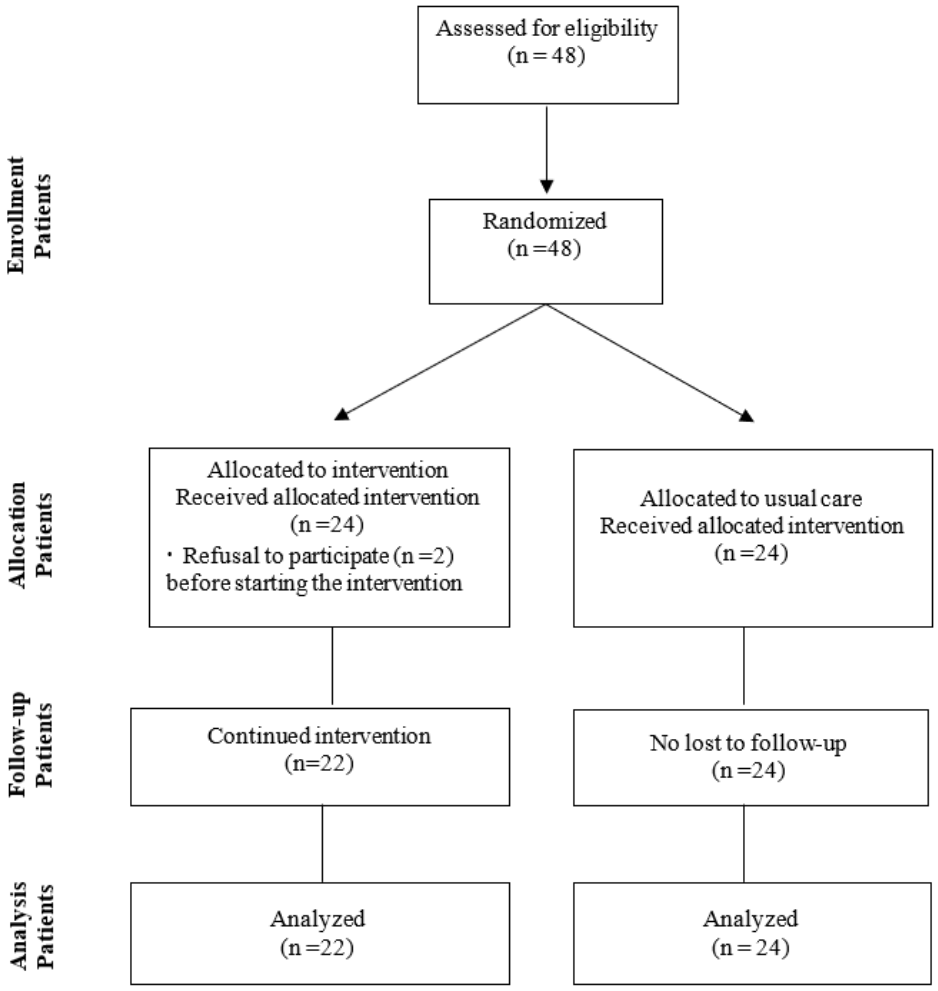
Ethical Considerations

This study was approved by the Institutional Review Board of the Severance Hospital, Yonsei University Health System (4-2020-0889). We informed all participants of the purpose of this study, process, methods, voluntary participation terms, and the possible risks and benefits of participation. Furthermore, we gave the participants 1 week to consider their participation in the study before deciding whether to sign the informed consent form. All study participants were compensated with KRW 50,000 (approximately US $43) as a gift at both the beginning and end of the study.

Enrollment Process

To recruit participants, we posted recruitment announcements on the bulletin board of a urology cancer center. On the day of outpatient visits, we met face-to-face with patients who indicated an interest in participating in the study. We provided potential participants with the opportune time to ask questions. Then, we fully informed the potential participants about all facets of the study and invited them to voluntarily sign the informed consent form. Finally, we registered the participants after screening for the inclusion and exclusion criteria (Figure 1). The allocation process for this study was conducted jointly with a research assistant to ensure that all enrollments were transparent.

Figure 1. Modified CONSORT (Consolidated Standards of Reporting Trials) flow diagram for individual randomized controlled trials.



Using *F* test in the G*Power program version 3.1 (Heinrich Heine University) [40], we calculated the sample size to identify the variance difference of repeated data between the 2 groups. The minimum number of participants was calculated to be 41 based on an effect size of 0.51, which was derived from the study by Bourke et al [41], with a significance level of .05, a power of 0.08, and 3 repeat measures. Considering a projected dropout rate of 15% [41], the total sample size was set at 48. Therefore, 24 participants were registered in each group. We assigned the participants randomly into experimental and waitlist control groups using a pregenerated sequence in SAS 9.4 (SAS Institute) based on the 2×2 permuted randomized block method. Using this method, 4 people are grouped as 1 block, and the 1 block is then divided into 2 groups. Then, we assigned the groups at a 1:1 ratio. The participants did not know which group they were assigned to. They received the intervention individually, which minimized the risk of contamination between groups.

Research Design and Setting

This was a single-blind, single-center randomized waitlist controlled trial. It used a pre- and posttest design and a 2×2 permuted randomization block method. We prepared an allocation table for each group before participants were enrolled. Each participant was assigned to an experimental or a waitlist control group in the order of enrollment using a 2×2 randomization table generated in the SAS program (SAS Institute). The protocol for this study was registered with the Clinical Research Information Service (registration no. KCT0006560). We recruited the participants from a tertiary general hospital located in Seoul, South Korea.

Intervention

To improve the positive health behaviors of patients with PC receiving ADT, a nurse-led mobile-based health coaching program based on the analyze, design, develop, implement, and evaluate model and the information-motivation-behavioral (IMB) skills model was developed. The analyze, design, develop, implement, and evaluate model, known as an instructional design model, is a representative framework used by teaching and learning methods. The IMB model has been widely used as a theoretical basis for developing interventions that aim to encourage patients with chronic diseases to maintain positive health behaviors. According to the IMB model, acquiring sufficient information, enhancing motivation, acquiring the skills required to perform a behavior, and improving self-efficacy are the factors that lead to long-term behavioral changes that improve subjective and objective health outcomes [42]. The nurse-led mobile-based health coaching program used an individualized approach to ensure that the IMB

skills addressed in the program were best suited to each patient to improve their specific lifestyle patterns. In this program, self-management information included personalized recommendations for diet and exercise and strategies to manage both individual ADT-induced side effects and common side effects from other prescribed drugs. Motivation strategies included health contracting, individual goal setting, encouragement and continuous counseling, and rewarding. In addition, we encouraged the patients' main caregivers to become involved in their patients' behavioral self-management. Behavioral self-management skills included sharing vicarious experiences (eg, sharing success stories), self-monitoring through exercises, and maintaining a nutrition diary.

The program consisted of a 4-week intensive program conducted via one-on-one Zoom (Zoom Video Communication, Inc) meetings, followed by an 8-week maintenance program conducted via individual contact through telephone calls and KakaoTalk (Kakao Corporation), which is a popular Korean SMS text messaging app (Figure 2). Participants were provided with relevant information and a to-do list regarding exercise and diet once a week for 4 weeks in an intensive program. We provided an educational package that included a booklet containing PowerPoint (Microsoft Corporation) slides, exercise and nutritional diaries, exercise video files, TheraBand resistance bands (THERABAND), and a pedometer. We focused on overcoming barriers to healthy diet and exercise. For example, we recommended alternative exercises for participants with knee pain, which included a higher proportion of movements performed in a sitting position. Furthermore, the participants received feedback to address problems at the end of every web-based meeting during the intensive program and via SMS text messages and telephone calls during the maintenance program. When participants felt uncertain about an exercise motion, they recorded a video of themselves performing the exercise and sent it to us. We then provided feedback describing the correct motion, using the video as a reference. Regarding diet, we focused on changing unhealthy eating habits, controlling participants' daily calorie intake, encouraging a low-fat and low-carbohydrate diet, and ensuring adequate protein intake. Participants were educated on the proportions of food from each food group they required per day according to their individual daily recommended calories and the diabetic exchange diet. Feedback and questions related to the diet program were addressed via KakaoTalk. We recommended implementing the following strategies for maintaining healthy behaviors and enhancing self-efficacy once a week or more during the maintenance period: vicarious experiences (eg, sharing success stories), emotional support, encouragement, keeping an exercise and nutrition diary, and goal reminders. To minimize the expected bias that can occur in an intervention study, we conducted the intervention according to the study protocol and used a checklist to ensure consistency between the experimental and waitlist control groups. The control group was placed on a waitlist and only received usual care during the intervention period. After completing data collection, we provided them with the same mobile-based program and educational materials as the experimental group (time point 3; T3). A group of experts evaluated this program to confirm its content and construct validity.

Figure 2. Educational material: (A) education on information and management of metabolic syndrome via the internet using Zoom (Zoom Video Communications, Inc), (B) personalized diet coaching, and (C) personalized exercise coaching: providing feedback describing the correct motion.

Data Collection

Data collection happened at 3 different time points: baseline (time point 1; T1), 6 weeks after the beginning of the intervention (time point 2; T2), and 12 weeks after the beginning of the intervention (T3). We collected T1 and T3 data on a face-to-face basis on the day the patient visited the urology cancer center. Variables that required physical contact, such as AC and grip strength, were measured together using a web-based survey at T1 and T3 but not at T2. At T2, we collected data collection on variables, including lifestyle score and HRQoL, on a non–face-to-face basis using a web-based survey. Clinical data that were obtained included treatment type, ADT exposure time, cancer stage, biochemical data, and Gleason score from electronic medical records (EMRs) at T1.

Outcomes

Lifestyle Score

Lifestyle score was the primary outcome of this study. Lifestyle changes were evaluated using the lifestyle evaluation tool by Kang [43] to assess the health behaviors of patients with MetS. A higher score indicates better self-management behaviors. This assessment consists of 36 items across 6 domains: physical activity and weight control, dietary habits, drinking and smoking, sleep and rest, stress, and drug and health management [43].

We recorded responses on a 4-point scale with the options *not at all*, *sometimes*, *often*, and *always*. Total scores could range from 36 to 144 points, with a higher score indicating a healthier lifestyle. The Cronbach α of the original study was .92 (physical activity and weight control=.87, dietary habits=.87, drinking and smoking=.87, sleep and rest=.86, stress=.74, and drug and health management=.70). In this study, Cronbach α was .82 (physical activity and weight control=.90, dietary habits=.86, drinking and smoking=.41, sleep and rest=.67, stress=.63, and drug and health management=.54).

MetS Components

The components of MetS consist of FBS, AC, blood pressure (both systolic and diastolic), fasting triglyceride level, and fasting HDL cholesterol. We instructed the participants not to have a meal and not to take antihypertensive drugs on the test day. Then, at the urology cancer center, we first measured AC using a tape at the umbilicus between the highest point of the iliac crest and the lower edge of the 12th rib with an error range of 0.5 cm while the participants fasted. Second, we measured the blood pressure twice on the participant's nondominant arm using a TM-2657P device (A&D Company Limited) after they had rested for at least 10 minutes. The average systolic blood pressure (SBP) and diastolic blood pressure (DBP) were calculated.

Biochemical Data

Biochemical data included the levels of the following MetS components: fasting triglyceride, fasting HDL cholesterol, and FBS. These variables were measured using blood samples collected while participants had fasted for >6 hours, and the data were obtained from EMRs.

Body Composition

Body composition data were obtained using a body composition analyzer (InBody H20B [Biospace]). We instructed the participants to stand upright and hold the handle attached to the measurement device, which put them into contact with 8 electrodes (2 each on both hands and both feet). Body composition measurements included height (cm), body weight (kg), body fat mass (kg), body fat percentage (%), skeletal muscle mass (kg), and BMI (kg/m^2).

HRQoL Measurement Tool

We measured HRQoL using the Korean version of the 26-item Expanded Prostate Cancer Index Composite (EPIC-26). EPIC-26 is a short-form version of the original expanded PC index composite (EPIC) tool, which contains 50 items. The EPIC tool was developed to understand treatment-related symptoms with a higher degree of sensitivity than previous diagnostic tools and the impact of PC treatment on patients' lives [44]. Higher scores indicated a better HRQoL, with possible scores ranging from 0 to 100. EPIC-26 consists of 5 symptom domains: urinary incontinence, urinary irritation and obstruction, sexual, bowel, and hormonal. There is no Korean version of EPIC-26, but there is a Korean version of the original 50-item EPIC tool. Therefore, the 26 items from EPIC-26 were extracted from the Korean version of the original 50-item EPIC tool [45], and the survey was conducted using this tool. Permission to use both EPIC-26 and the original 50-item EPIC was granted by the original author.

The Cronbach α of EPIC-26 ranged from .70 to .90 in all domains except for the hormonal domain (Cronbach α=.62). The Cronbach α of the Korean version of EPIC was .83 [45]. The Cronbach α of the tool used in this study was .63 (urinary incontinence=.88, urinary irritation and obstruction=.64, sexual=.84, bowel=.13, and vitality or hormonal=.46).

Clinical Data

Disease-related patient information, treatment type, ADT exposure time, cancer stage, and Gleason score were obtained from EMRs.

Data Analysis

Demographic data, disease-related characteristics, and main outcome variables were analyzed using mean, SD, frequency, and percentage. An independent 1-tailed *t* test and a chi-square test were performed to identify differences between the groups. The equality of variance was assessed before using the pooled variance estimator for the *t* test. Fisher exact test was performed as appropriate. According to the International Conference on Harmonization E9 guideline [46], which provides guidance on statistical principles for clinical trials, a modified intention-to-treat analysis was conducted. As 2 participants withdrew from the study for personal reasons before initiating the intervention, there was a lack of data that could evaluate the effect of the main outcomes. We judged that there would be no difference in the intention-to-treat analysis owing to the low dropout rate (2/48, 4%) and high compliance rate (22/22, 81%). Hypothesis testing was conducted using a 1-tailed test and the PROC MIXED procedure in SAS (SAS Institute). SE estimates

were obtained as a result of the PROC MIXED procedure using the *empirical option* to adjust for skewed data from potentially different covariance structures. This method is based on the sandwich estimation approach [47]. It improves variance and covariance with robust and consistent estimates, irrespective of the covariance structure in the actual data. As a follow-up analysis, we determined statistically significant time points within the groups by calculating the difference in the least square means from the baseline at each time point.

Results

Overview

The data for this study were collected across 7 months, from March 24 to September 15, 2021. We took approximately 30 minutes per participant to complete data collection. A total of 48 participants were recruited. Two participants in the experimental group declined to participate in the program before starting the intervention, citing personal reasons. A total of 46 participants were finally included in the analysis, with an attrition rate of 4% (2/48).

Results of General and Disease-Related Characteristics

The general and disease-related characteristics are presented in Table 1. The mean age of the participants was 68.83 (SD 7.09) years. Most participants (41/46, 89%) lived with their spouses or families. Approximately 46% (21/46) of the participants were unemployed, and 35% (16/46) had jobs requiring relatively less physical activity, such as office workers, taxi drivers, and service workers. Most participants (27/46, 59%) were exsmokers, and the mean smoking duration was 16.13 (SD 21.38) pack-years.

Table 1. Homogeneity tests in general characteristics, disease-related characteristics, and main outcome variables between groups (N=46).

Variables	Total	Experimental group (n=22)	Control group (n=22)	χ^2 (*df*)	*t* (*df*)	*P* value
General characteristics						
Age (y), mean (SD)	68.83 (7.09)	67.59 (6.59)	69.96 (7.48)	N/A[a]	−1.13 (44)	.26
Monthly income (US $), mean (SD)	3679.77 (4603.72)	4813.0 (6188.7)	2640.9 (2035.5)	N/A	1.57 (44)	.13
Religion, n (%)						.99
No	21 (46)	10 (22)	11 (24)	0.0 (1)	N/A	
Yes	25 (54)	12 (26)	13 (28)	N/A	N/A	
Education, n (%)						.40
Less than or equal to middle school	27 (59)	11 (24)	16 (35)	0.7 (1)	N/A	
Greater than or equal to college	19 (41)	11 (24)	8 (17)	N/A	N/A	
Living, n (%)						.77[b]
Alone	5 (11)	3 (7)	2 (4)	0.5 (2)	N/A	
With spouse only	21 (46)	9 (20)	12 (26)	N/A	N/A	
With their family	20 (43)	10 (22)	10 (22)	N/A	N/A	
Job-related physical activities, n (%)						.93[b]
Unemployed	21 (46)	10 (22)	11 (24)	0.3 (2)	N/A	
Less active	16 (35)	7 (15)	9 (35)	N/A	N/A	
Highly active	9 (20)	5 (11)	4 (9)	N/A	N/A	
Smoking history						
1 pack-year, mean (SD)	16.13 (21.38)	15.78 (21.58)	16.44 (21.65)	N/A	−0.1 (44)	.92
Nonsmoker, n (%)	16 (35)	6 (13)	10 (22)	1.6 (2)	N/A	.45[b]
Exsmoker, n (%)	27 (59)	15 (33)	12 (26)	N/A	N/A	N/A
Current smoker, n (%)	3 (7)	1 (2)	2 (4)	N/A	N/A	N/A
Disease characteristics						
Number of comorbidities, mean (SD)	1.61 (1.04)	1.64 (1.05)	1.58 (2.03)	N/A	0.17 (44)	.87
0, n (%)	6 (13)	3 (7)	3 (7)	0.8 (3)	N/A	.87[b]
1, n (%)	17 (37)	7 (15)	10 (22)	N/A	N/A	N/A
2, n (%)	14 (30)	8 (17)	6 (13)	N/A	N/A	N/A
≥3, n (%)	9 (20)	4 (9)	5 (11)	N/A	N/A	N/A
Treatment type, n (%)						.10
Operation	16 (35)	5 (11)	11 (24)	2.7 (1)	N/A	
Operation+radiation	30 (65)	17 (37)	13 (28)	N/A	N/A	
ADT[c] type, n (%)						
Antiandrogen	30 (65)	14 (30)	16 (35)	0.0 (1)	N/A	.99
Antiandrogen+LHRH[d]	16 (35)	8 (17)	8 (17)	N/A	N/A	N/A
ADT duration (month), mean (SD)	40.63 (24.71)	40.64 (22.61)	40.63 (26.98)	N/A	0.0 (44)	.99
PSA[e], mean (SD)	0.07 (0.19)	0.09 (0.27)	0.05 (0.08)	N/A	0.59 (44)	.56
Gleason score, mean (SD)	7.80 (0.98)	8.00 (1.07)	7.63 (0.88)	N/A	1.31 (44)	.20
Main outcomes						

Variables	Total	Experimental group (n=22)	Control group (n=22)	χ^2 (*df*)	*t* (*df*)	*P* value
Lifestyle score[f], mean (SD)						
Total score	99.07 (13.56)	100.50 (12.53)	97.71 (14.57)	N/A	0.70 (44)	.48
Exercise and weight loss	19.67 (5.88)	20.23 (5.45)	19.17 (6.32)	N/A	0.61 (44)	.55
Diet	42.17 (6.72)	42.77 (6.05)	41.63 (7.38)	N/A	0.57 (44)	.57
Alcohol and smoking	9.54 (3.40)	10.09 (3.25)	9.04 (3.53)	N/A	1.05 (44)	.30
Stress management	9.82 (1.79)	9.82 (1.79)	9.33 (2.41)	N/A	0.77 (44)	.45
Sleep and rest	6.45 (1.60)	6.45 (1.60)	6.71 (1.76)	N/A	−0.51 (44)	.61
Medication adherence and physical examination	11.52 (2.04)	11.18 (1.84)	11.83 (2.20)	N/A	−1.08 (44)	.28
MetS[g], n (%)						
No	21 (46)	10 (22)	11 (24)	0.0 (1)	N/A	.98
Yes	25 (54)	12 (26)	13 (28)	N/A	N/A	N/A
MetS component, mean (SD)						
SBP[h] (mm/Hg)	136.42 (13.01)	139.7 (11.44)	133.4 (13.83)	N/A	1.69 (44)	.10
DBP[i] (mm/Hg)	83.51 (7.46)	85.18 (6.11)	81.98 (8.35)	N/A	1.47 (44)	.15
AC[j] (cm)	95.73 (5.97)	94.30 (5.29)	97.05 (6.36)	N/A	−1.59 (44)	.12
FBS[k] (mg/dl)	111.17 (22.22)	112.6 (27.64)	109.8 (16.27)	N/A	0.41 (44)	.68
HDL[l] (mg/dl)	52.89 (13.61)	55.50 (13.32)	50.50 (13.71)	N/A	1.25 (44)	.22
Triglyceride (mg/dl)	114.46 (49.05)	116.5 (44.45)	112.5 (53.82)	N/A	0.27 (44)	.79
Body composition, mean (SD)						
Body weight (kg)	73.12 (5.99)	72.53 (4.54)	73.66 (7.13)	N/A	−0.64 (44)	.52
Body fat mass (kg)	22.18 (4.59)	21.75 (3.51)	22.57 (5.44)	N/A	−0.61 (44)	.54
Body fat percent (%)	30.38 (5.31)	30.50 (4.74)	30.28 (5.88)	N/A	0.14 (44)	.89
Skeletal muscle mass (kg)	28.22 (3.08)	28.03 (2.96)	28.39 (3.23)	N/A	−0.39 (44)	.70
BMI (kg/m^2)	25.93 (2.04)	25.65 (1.70)	26.19 (2.32)	N/A	−0.89 (44)	.38
HRQoL[m] domains, mean (SD)						
Urinary incontinence	67.61 (25.99)	69.17 (26.56)	66.17 (25.93)	N/A	0.39 (44)	.70
Urinary irritation and obstruction	88.72 (10.59)	87.78 (10.11)	89.58 (11.16)	N/A	−0.57 (44)	.57
Bowel problem	95.11 (9.31)	95.83 (8.33)	94.44 (10.26)	N/A	0.50 (44)	.62
Sexual problem	19.28 (19.54)	21.89 (19.18)	16.89 (19.97)	N/A	0.86 (44)	.39

Variables	Total	Experimental group (n=22)	Control group (n=22)	χ^2 (*df*)	*t* (*df*)	*P* value
Hormonal problem	85.00 (13.86)	85.00 (78.83)	85.00 (78.29)	N/A	0.00 (44)	.99

[a]N/A: not applicable.

[b]Fisher exact test.

[c]ADT: androgen deprivation therapy.

[d]LHRH: luteinizing hormone–releasing hormone.

[e]PSA: prostate-specific antigen.

[f]Lifestyle score measured by lifestyle evaluation; scores range from 36 to 144.

[g]MetS: metabolic syndrome.

[h]SBP: systolic blood pressure.

[i]DBP: diastolic blood pressure.

[j]AC: abdominal circumference.

[k]FBS: fasting blood sugar.

[l]HDL: high-density lipoprotein.

[m]HRQoL: health-related quality of life. Scores range from 0 to 100, higher scores represent better HRQoL.

The average number of comorbidities was 1.61 (SD 1.04), and the most common comorbid diseases were cardiovascular diseases such as hypertension or dyslipidemia (31/46, 67%). There was no difference in comorbidities between the experimental and waitlist control groups (*P*=.87). Of the 65 participants, 30 (65%) received antiandrogen monotherapy, and the rest received antiandrogen therapy with luteinizing hormone–releasing hormone. The average duration of ADT was 40.63 (SD 24.71) months, the mean prostate-specific antigen level was 0.07 (SD 0.19) ng/ml, and the mean Gleason score, which determines the prognosis and pathological status of cancer, was 7.80 (SD 0.98). The experimental and waitlist control groups showed no statistical differences between general and disease-related characteristics. The mean healthy lifestyle score was 99.07 (SD 13.56). The MetS components with mean values in an abnormal range were SBP (mean 136.42, SD 13.01 mm Hg), AC (mean 95.73, SD 5.97 cm), and FBS (mean 111.17, SD 22.22 mg/dL). There were no statistical differences in MetS components between the groups. The mean body weight and BMI were approximately 73 (SD 5.99) kg and 26 (SD 2.04) kg/m^2, respectively. The participants responded that they had problems in the urinary incontinence (mean 67.61, SD 25.99) and sexual (mean 19.28, SD 19.54) domains of HRQoL. The experimental and waitlist control groups were homogeneous in terms of their main outcome variables.

Results of Primary Outcome Variables Between Groups Over Time

Table 2 and Figure 3 show the results of the lifestyle score variables between the groups over time. The study found that there were no group differences in lifestyle scores at baseline (Table 1). However, over time, the experimental group's lifestyle scores consistently increased (T1=100.55, T2=125.82, and T3=130.27), whereas the waitlist control group's lifestyle scores showed no consistent increase (T1=97.71, T2=95.92, and T3=98.21). Over time, the lifestyle scores of the experimental group significantly increased (β=29.23; *P*≤.001) compared with the waitlist control group. The experimental group developed a healthier lifestyle with time, and remarkable changes were observed during the intervention period.

Table 2. Group comparison in lifestyle score, MetS[a] components, body composition, and HRQoL[b] domain parameters (N=46).

Outcome and group	T1[c], mean (SD)	T2[d], mean (SD)	T3[e], mean (SD)	Difference of changes between groups over time[f]		
				Estimate (SE; 95% CI)	*F* test (*df*)	*P* value
Lifestyle score						
Experimental	100.55 (12.53)	125.82 (11.73)	130.27 (10.15)	29.23 (3.50; 22.26 to 36.19)	38.49 (88)	≤.001
Control	97.71 (14.57)	95.92 (18.64)	98.21 (15.07)	29.23 (3.50; 22.26 to 36.19)	38.49 (88)	≤.001
MetS components						
SBP[g]						
Experimental	139.73 (11.44)	N/A[h]	126.25 (14.08)	−5.64 (3.72; −13.12 to 1.85)	2.30 (44)	.14
Control	133.38 (13.83)	N/A	125.54 (12.58)	−5.64 (3.72; −13.12 to 1.85)	2.30 (44)	.14
DBP[i]						
Experimental	85.18 (6.11)	N/A	79.36 (9.35)	−0.94 (2.21; −5.40 to 3.52)	0.18 (44)	.67
Control	81.98 (8.35)	N/A	77.10 (7.36)	−0.94 (2.21; −5.40 to 3.52)	0.18 (44)	.67
FBS[j]						
Experimental	112.64 (27.64)	N/A	102.59 (12.42)	−12.0 (6.02; −24.14 to 0.13)	3.98 (44)	.05
Control	109.83 (16.27)	N/A	111.79 (16.88)	−12.0 (6.02; −24.14 to 0.13)	3.98 (44)	.05
HDL[k] cholesterol						
Experimental	55.50 (13.33)	N/A	55.32 (12.46)	−0.27 (1.48; −3.25 to 2.72)	0.03 (44)	.86
Control	50.50 (13.71)	N/A	50.58 (12.34)	−0.27 (1.48; −3.25 to 2.72)	0.03 (44)	.86
Triglyceride						
Experimental	116.55 (44.45)	N/A	107.73 (43.19)	−11.65 (12.04; −35.91 to 12.61)	0.94 (44)	.34
Control	112.54 (53.82)	N/A	115.38 (62.71)	−11.65 (12.04; −35.91 to 12.61)	0.94 (44)	.34
AC[l]						
Experimental	94.30 (5.29)	N/A	91.02 (3.94)	−2.49 (1.23; −4.98 to −0.01)	4.09 (44)	.049
Control	97.05 (6.36)	N/A	96.27 (6.83)	−2.49 (1.23; −4.98 to −0.01)	4.09 (44)	.049
Body composition						
Body weight (kg)						
Experimental	72.53 (4.54)	N/A	70.04 (4.53)	−1.52 (0.46; −2.45 to −0.58)	10.71 (44)	<.001
Control	73.66 (7.13)	N/A	72.68 (7.04)	−1.52 (0.46; −2.45 to −0.58)	10.71 (44)	<.001
BMI (kg/m^2)						
Experimental	25.65 (1.70)	N/A	24.76 (1.49)	−0.55 (0.17; −0.88 to −0.21)	10.54 (44)	<.001
Control	26.19 (2.32)	N/A	25.84 (2.27)	−0.55 (0.17; −0.88 to −0.21)	10.54 (44)	<.001
Skeletal muscle mass (kg)						
Experimental	28.03 (2.96)	N/A	27.84 (4.19)	0.53 (0.83; −1.15 to 2.20)	0.40 (44)	.53
Control	28.39 (3.23)	N/A	27.67 (2.84)	0.53 (0.83; −1.15 to 2.20)	0.40 (44)	.53
Fat mass (kg)						
Experimental	21.75 (3.51)	N/A	19.76 (5.01)	−1.93 (1.35; −4.65 to 0.78)	2.06 (44)	.16
Control	22.57 (5.44)	N/A	22.52 (4.66)	−1.93 (1.35; −4.65 to 0.78)	2.06 (44)	.16
Fat percentage						
Experimental	30.50 (4.74)	N/A	28.32 (7.27)	−2.75 (1.90; −6.58 to 1.09)	2.08 (44)	.16
Control	30.28 (5.88)	N/A	30.84 (4.61)	−2.75 (1.90; −6.58 to 1.09)	2.08 (44)	.16
HRQoL domains						

Outcome and group	T1[c], mean (SD)	T2[d], mean (SD)	T3[e], mean (SD)	Difference of changes between groups over time[f]		
				Estimate (SE; 95% CI)	*F* test (*df*)	*P* value
Urinary incontinence						
Experimental	69.17 (26.56)	81.19 (18.87)	83.38 (18.67)	7.70 (6.55; −5.31 to 20.72)	0.70 (88)	.50
Control	66.18 (25.93)	72.29 (23.00)	72.68 (29.53)	7.70 (6.55; −5.31 to 20.72)	0.70 (88)	.50
Urinary irritation and obstruction						
Experimental	87.78 (10.11)	94.60 (7.54)	97.73 (4.93)	14.63 (4.05; 6.57 to 22.69)	7.01 (88)	<.001
Control	89.58 (11.16)	89.32 (11.43)	84.90 (18.14)	14.63 (4.05; 6.57 to 22.69)	7.01 (88)	<.001
Bowel						
Experimental	95.83 (8.33)	95.64 (7.21)	98.48 (4.37)	2.30 (2.98; −3.61 to 8.22)	0.34 (88)	.71
Control	94.44 (10.26)	91.67 (14.33)	94.79 (10.15)	2.30 (2.98; −3.61 to 8.22)	0.34 (88)	.71
Sexual						
Experimental	21.89 (19.18)	18.86 (16.02)	25.49 (16.71)	1.52 (4.50; −7.43 to 10.47)	0.08 (88)	.92
Control	16.89 (19.97)	12.15 (12.18)	18.97 (15.03)	1.52 (4.50; −7.43 to 10.47)	0.08 (88)	.92
Hormonal						
Experimental	85.00 (11.65)	89.32 (12.94)	92.73 (7.03)	6.27 (3.48; −0.64 to 13.18)	1.85 (88)	.16
Control	85.00 (15.88)	86.88 (12.58)	86.46 (13.87)	6.27 (3.48; −0.64 to 13.18)	1.85 (88)	.16

[a]MetS: metabolic syndrome.

[b]HRQoL: health-related quality of life.

[c]T1: time point 1.

[d]T2: time point 2.

[e]T3: time point 3.

[f]Reference: interaction between control group and T1.

[g]SBP: systolic blood pressure.

[h]N/A: not applicable.

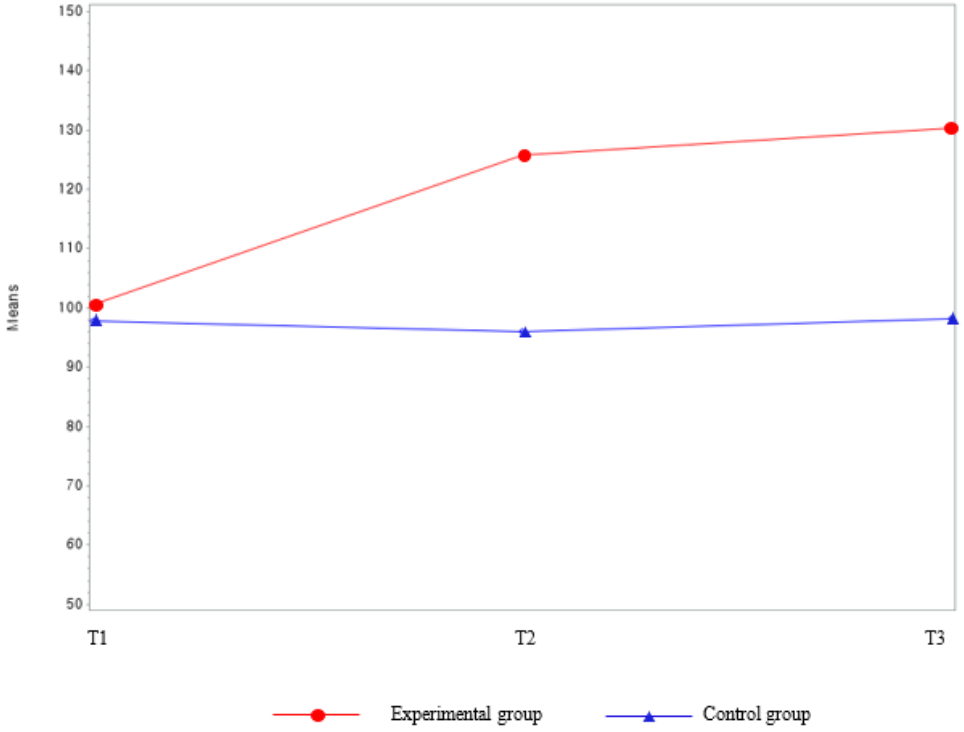
[i]DBP: diastolic blood pressure.

[j]FBS: fasting blood sugar.

[k]HDL cholesterol: high-density lipoprotein cholesterol.

[l]AC: abdominal circumference.

Figure 3. Group comparison among the main outcome variables (Lifestyle score) over time. T1: time point 1; T2: time point 2; T3: time point 3.



Results of Secondary Outcome Variables Between Groups Over Time

The difference in the prevalence of MetS before and after the intervention between the experimental group and the waitlist control group was not statistically significant (χ^2_1=1.1; *P*=.31 at T3 [not presented in the tables]). Among the MetS components, the parameters for FBS (β=−12.0; F_{44}=3.98; *P*=.05) and AC (β=−2.49; F_{44}=4.09; *P*=.049) showed significant interactions between group and time (Table 2 and Figures 4 and 5). Regarding body composition, the mean body weight and BMI in the experimental group decreased significantly by 1.52 kg and 0.55 kg/m^2, respectively (*P*<.001) compared with the baseline values. Group, time, and group and time interactions were also statistically significant between these variables (*P*<.001). Over time, the mean body weight (*P*<.001) and BMI (*P*<.001) decreased more in the experimental group than in the waitlist control group (Table 2 and Figures 6 and 7). Regarding HRQoL domains (Table 2 and Figure 8), a more significant improvement was observed in the experimental group than in the waitlist control group for the urinary irritative and obstructive domain. The mean changes in the urinary irritative and obstructive domain of HRQoL were statistically significant at each time point from the baseline, whereas the mean changes in the control group for this domain were not statistically significant. Group and time interactions were not significant, except in the urinary irritative and obstructive domain of HRQoL in the experimental group (β=14.63; $F_{8,8}$=7.01; *P*<.001).

Figure 4. Group comparison among the main outcome variables (fasting blood sugar) over time. T1: time point 1; T3: time point 3.

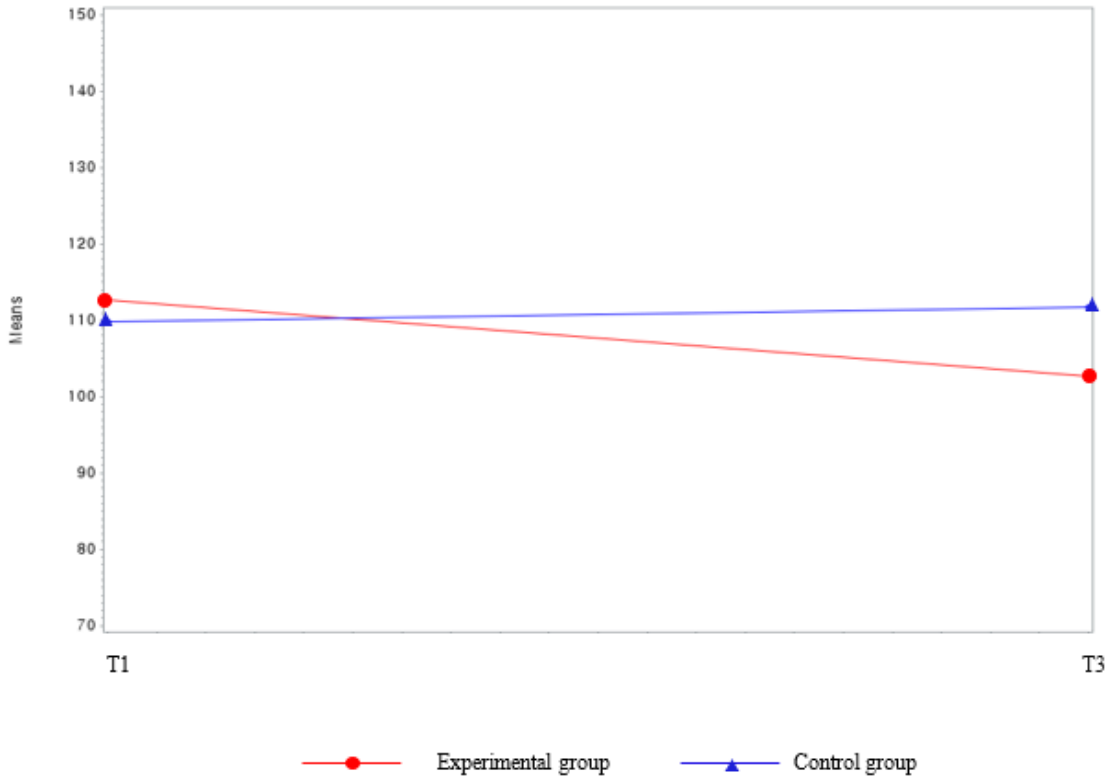


Figure 5. Group comparison among the main outcome variables (abdominal circumference) over time. T1: time point 1; T3: time point 3.

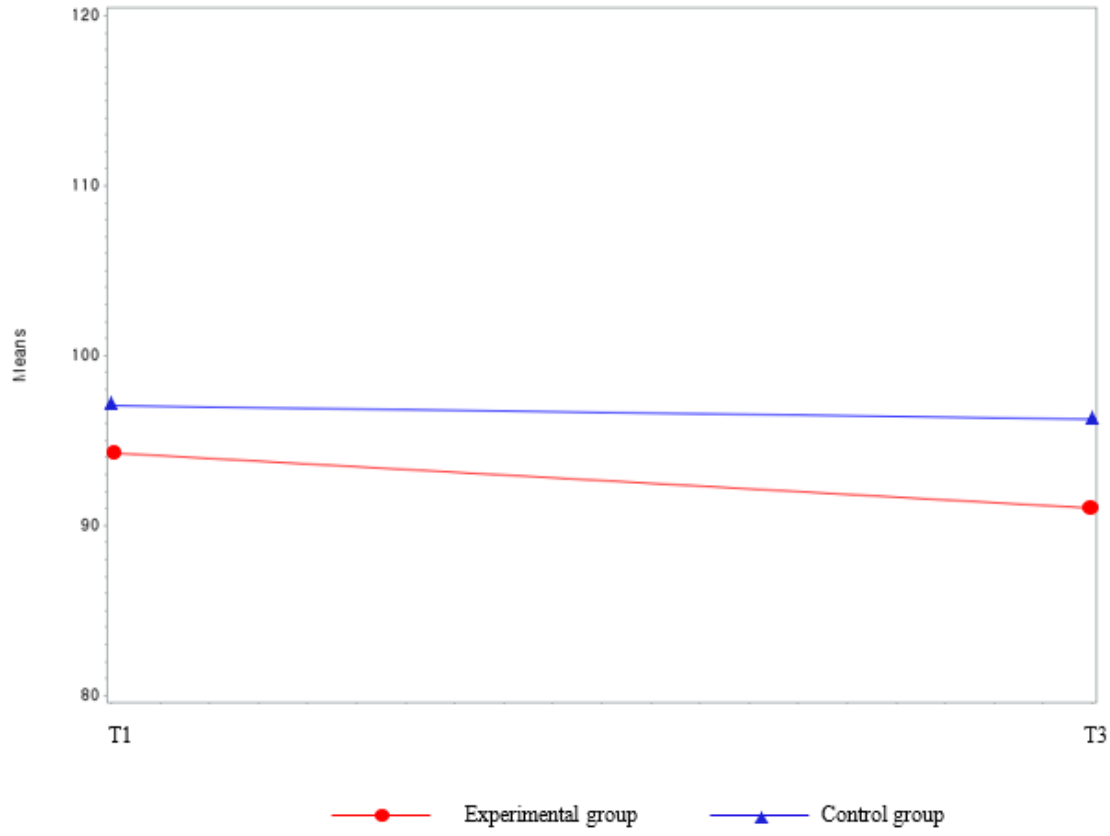


Figure 6. Group comparison among the main outcome variables (body weight) over time. T1: time point 1; T3: time point 3.

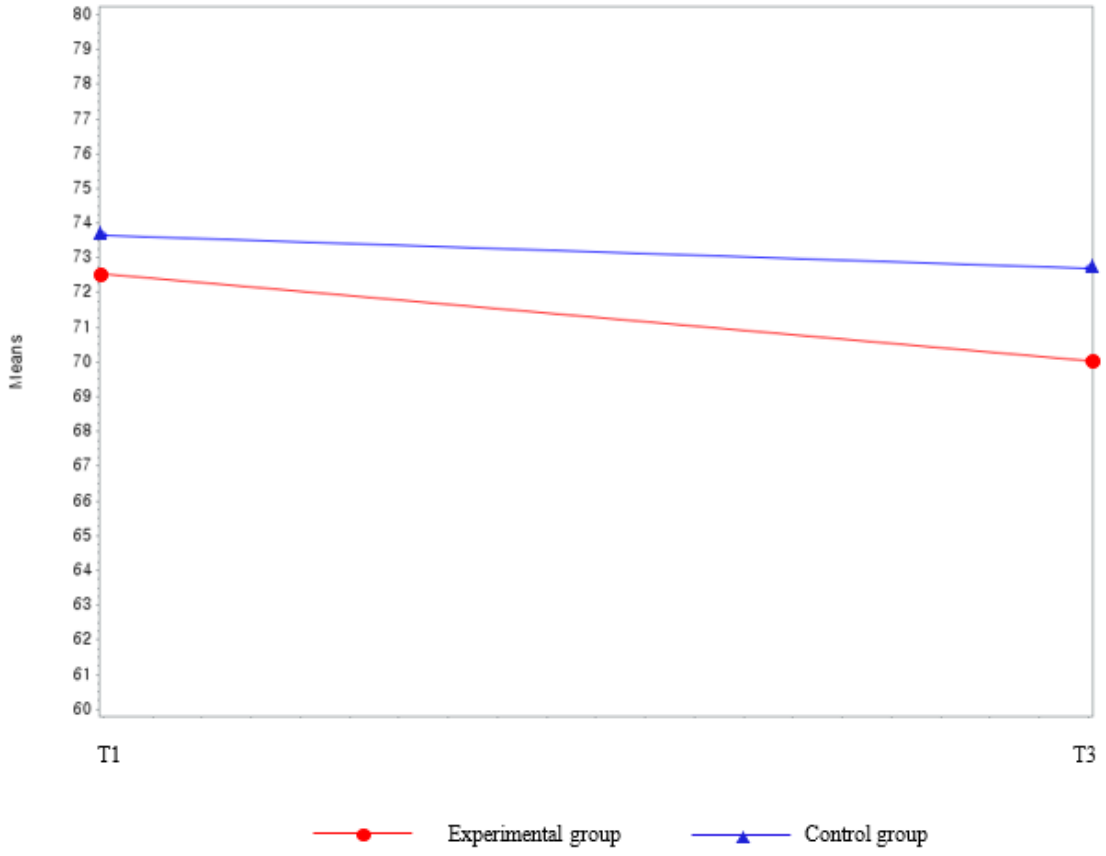


Figure 7. Group comparison among the main outcome variables (BMI) over time. T1: time point 1; T3: time point 3.

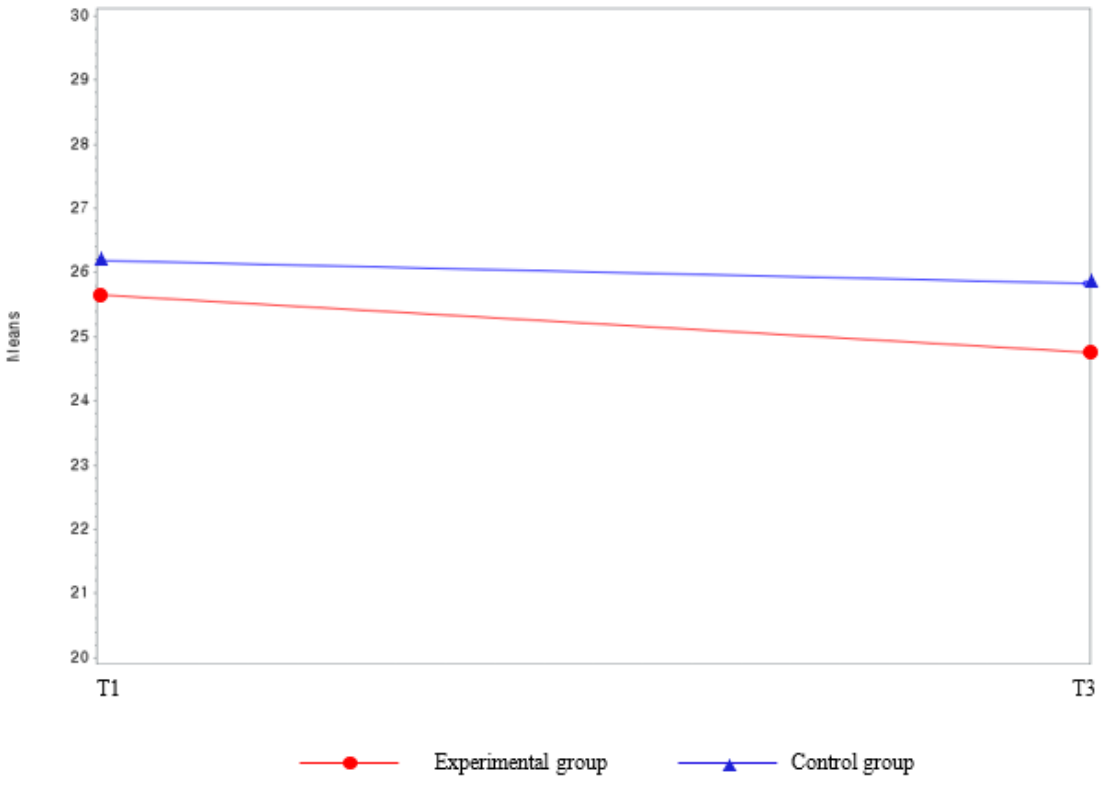
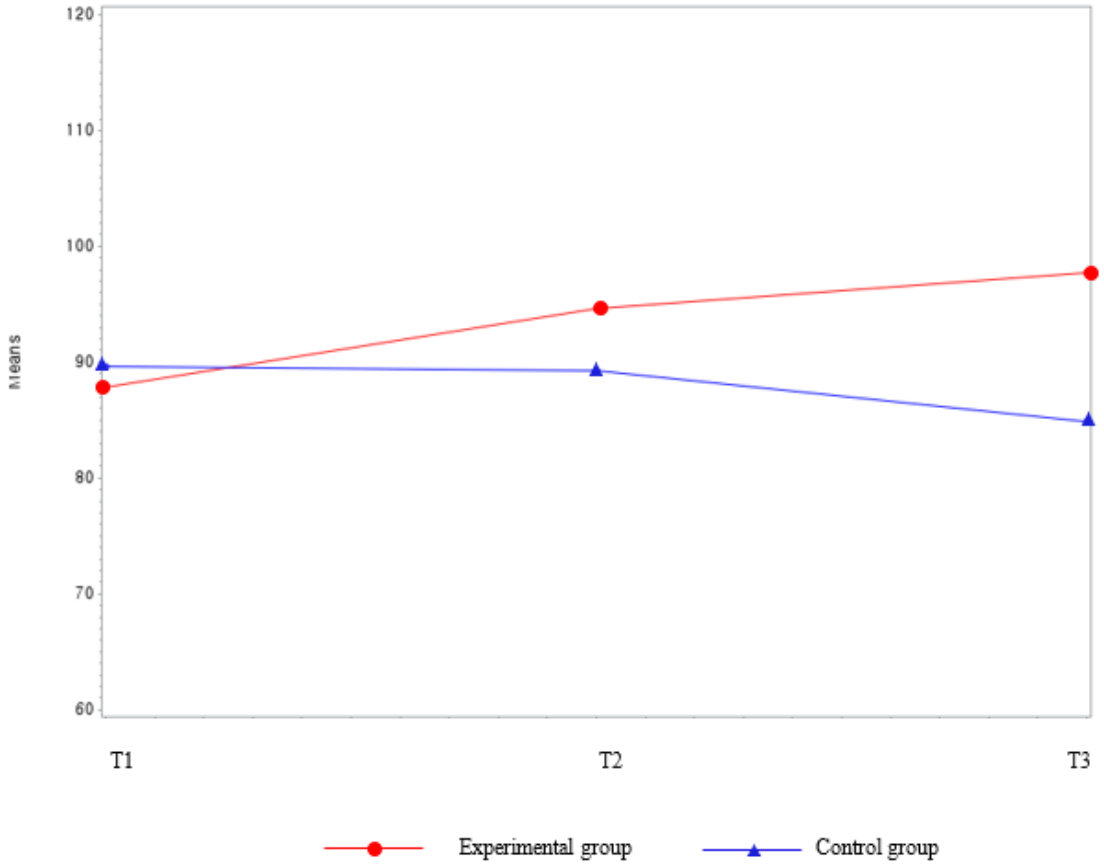


Figure 8. Group comparison among the main outcome variables (urinary irritation and obstruction) over time. T1: time point 1; T2; time point 2; T3: time point 3.



Discussion

Principal Findings

This study was conducted to determine the effectiveness of a nurse-led mobile-based education program for patients with PC who are at risk of MetS. This study showed that a nurse-led mobile-based health coaching program promoted a healthy lifestyle in patients with PC receiving ADT, which ultimately improved the levels of some MetS components (eg, reduction in FBS and AC), body composition (eg, reduction in weight and BMI), and HRQoL (eg, the urinary irritative and obstructive domain).

In this study, with a 3-month intervention, the results indicated that each variable required a different amount of time to show changes. In a previous study, a period of at least 3 to 8 weeks was required to confirm significant changes in weight, BMI, and the levels of each MetS component following lifestyle intervention programs [48-50]. Specifically, AC, FBS, body weight, and BMI have been found to decrease significantly over short periods [49,51]. However, in studies with intervention periods of ≥6 months, significant results were confirmed for MetS components including HDL cholesterol, SBP, and DBP [52,53]. Even in weight control programs that included strict diet control and exercise intervention guidelines, MetS components, including HDL cholesterol, did not change over a period of 8 weeks and only showed a significant change in both men and women 10 months after the end of the intervention [54].

The study by Dawson et al [55] observed decreases in body weight and AC but not in HDL cholesterol level among the MetS components in patients with PC. Focht and colleagues [56] also reported decreased bodyweight in patients with PC. These studies included lifestyle interventions lasting 3 months for patients with PC who had received ADT. The mean duration of ADT was 14 months in the study by Focht et al [56] and 30 months in the study by Dawson et al [55]. Specifically, Focht et al [56] confirmed that the body composition of body weight and body fat decreased significantly at 3 months compared with the control group. Furthermore, Dawson et al [55] conducted a program that emphasized exercise and protein supplementation for 3 months in patients with PC receiving ADT. As a result, of the MetS components, the AC of participants decreased most significantly in the experimental group than in the waitlist control group.

Reductions in body weight, AC, and FBS levels were closely related. When beginning to lose weight, the body temporarily lowers its metabolic function to maintain homeostasis and first metabolizes glucose, which is a basic energy source. When stored glycogen is broken down, the insulin mechanism is activated for additional energy consumption [57]. Weight loss causes a decrease in FBS level along with the action of insulin, which reactivates FBS, stored in the form of excess fat in the liver or abdomen. This fat is continuously used to generate energy, consequently, the fat accumulated in the liver or abdomen is consumed, leading to a reduced volume [58]. In this study, although no significant decrease in body fat was observed, a decrease in AC was observed. Similarly, in previous

studies, decreases in body weight [56], AC [55], and FBS levels [49] were confirmed 3 months after the intervention.

In this study, there were no significant changes in body composition related to body fat and skeletal muscle mass or in MetS components related to blood pressure (SBP and DBP) and lipids (HDL cholesterol and triglyceride). Insulin resistance develops over a long period, which increases the risk of obesity, diabetes, and MetS. In addition, it affects the lipid ratio, leading to an increase in low-density lipoprotein cholesterol and a decrease in HDL cholesterol [58], and causes inflammatory changes, resulting in changes in blood pressure owing to an increase in the residual amount of sodium in the blood [59]. Moreover, in a study that examined prediabetic patients diagnosed with impaired fasting glucose levels over the course of a 10-year follow-up period, 37% of the patients developed diabetes [60]. This finding indicates that the disease mechanism does not change over a short period but rather progresses slowly. Therefore, a long-term follow-up study is required to more accurately confirm the effectiveness of a lifestyle intervention program for MetS.

Among the HRQoL domains, an effect was observed only in the urinary irritative and obstructive domain, which may be caused by the education on how to manage the side effects of ADT treatment in the third and fourth weeks of the intervention. Patients with PC who have undergone multiple surgeries or frequent radiation therapy complained of side effects [61], such as urinary irritation and frequency. During the intervention in this study, information about appropriate water intake and pelvic floor muscle exercises to relieve urinary irritative and obstructive symptoms such as urinary irritation and frequency were included in an educational brochure [57], and appropriate water intake and pelvic floor muscle exercise were recommended for the participants depending on the presence or absence of symptoms. Thus, participants with these symptoms may have experienced relief through this intervention. Of the remaining HRQoL domains, the urinary incontinence, bowel, sexual, and hormonal domains did not show statistically significant changes. HRQoL was assessed using a questionnaire on treatment side effects and symptoms during the preceding 4 weeks. Most participants in this study were receiving long-term treatment after surgery, with an average of 40 (SD 24.71) months of ADT treatment. The bowel domain contains questions about whether the participants had diarrhea and bloody stools, which are acute side effects of participants who have undergone PC resection surgery. This domain had a higher mean score than the other HRQoL domains, with an average of ≥95 beginning at the baseline; therefore, these questions might not be relevant for participants of this study who were not in the acute stage after surgery. In addition, most questions in the hormonal domain examined the side effects that appear toward the start of ADT such as breast tenderness, bloody stools, and weight loss. Therefore, the insignificant changes in this domain might be because the participants in this study had already been taking medication for these side effects. At 19 points, the sexual domain had the lowest mean score of all the HRQoL domains. Sexual function might be limited in recovery as a result of a short-term health coaching program [62,63]. Although a previous study found that exercise can improve sexual function [64], sexual function requires interaction with a partner [65]. To restore sexual function, it is necessary to combine both psychological intervention and drug treatment [66].

This study found lifestyle changes to be critical in reducing the risk of MetS and that improved exercise and nutritional regimens should be implemented consistently for at least 3 months. Furthermore, a patient-centered, individualized approach that considers the side effects of ADT is needed to increase adherence.

Limitations

This study had some limitations. First, the criteria for participant selection included patients with at least 1 MetS component. To more thoroughly examine the effectiveness of this program on MetS management, stricter inclusion criteria should have been applied, such as only including those who were being treated with luteinizing hormone–releasing hormone, which is a type of ADT that causes the occurrence of more MetS components MetS components; those with ≥3 MetS components; those with a sedentary lifestyle; or those with a specific time after radical prostatectomy. Owing to data unavailability, this study only collected the Gleason score but not the information on TNM classification as the patient's pathological data. It is important to establish accurate patient pathological criteria; hence, future studies should incorporate both the TNM classification and the Gleason score with accurate criteria. Second, the 3-month intervention period planned in this study was limited to evaluating the continuation of self-management and improvement in MetS risk factors. It is necessary to extend the intervention period to at least 6 months to further evaluate the persistence of self-management. As older men with PC receiving ADT are at high risk of osteoporosis, measuring bone health–related indicators in addition to changes in body composition is highly recommended for future studies. Third, it is difficult to generalize the results of this study because the participants were recruited from a single hospital and not from a multicenter. Fourth, obtaining low Cronbach α values for some items in the lifestyle score may indicate a mismatch between the measurement and the participant's lifestyle. This is because the lifestyle score was created for the general adult population; however, the participants in this study were older adults who had PC, had undergone surgery, and were taking hormone-suppressing medications.

Conclusions

A nurse-led mobile-based health coaching program was developed to promote a healthy lifestyle among patients with PC receiving ADT. The research findings confirmed that lifestyle changes through nurse-led education can improve the components of MetS, body composition, and ADT's side effects. Therefore, by participating in this program, nurses were capable of creating changes in patients' lifestyles and improving the self-management of patients who were beginning ADT for the first time. In addition, this study can provide a basis for the development of other mobile-based education programs and tools.

Acknowledgments

This work was supported by the Korea Medical Device Development Fund grant funded by the Korean Government (the Ministry of Science and Information and Communications Technology; the Ministry of Trade, Industry and Energy; the Ministry of Health and Welfare; and the Ministry of Food and Drug Safety; project number: 1711174508, KMDF_PR_20200901_0077). This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2020R1A6A3A13060728). This work was supported by the NRF of Korea grant funded by the Korean Government (MSIT; Ministry of Science and Information Communications Technology number 2019R1F1A1062769 and number 2022R1A2C1092084). KL received a scholarship from Brain Korea 21 FOUR Project funded by National Research Foundation NRF of Korea, Yonsei University College of Nursing.

Authors' Contributions

KL contributed to data curation, formal analysis, investigation, methodology, project administration, software work, statistical analysis, visualization, original draft, writing, review and editing, funding acquisition. JP contributed to funding acquisition, investigation, methodology, project administration, visualization, review and editing, and supervision. All authors contributed to the conceptualization and have read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1218 KB - mhealth_v12i1e47102_app1.pdf]

References

1. Ministry of Health and Welfare. Cancer registration statistics. Statistics Korea. South Korea; 2019 Dec. URL: http://meta.narastat.kr/ [accessed 2021-09-18]
2. Huggins C, Hodges CV. Studies on prostatic cancer: I. The effect of castration, of estrogen and of androgen injection on serum phosphatases in metastatic carcinoma of the prostate. 1941. J Urol 2002 Jul;168(1):9-12. [doi: 10.1016/s0022-5347(05)64820-3] [Medline: 12050481]
3. Liede A, Hallett DC, Hope K, Graham A, Arellano J, Shahinian VB. International survey of androgen deprivation therapy (ADT) for non-metastatic prostate cancer in 19 countries. ESMO Open 2016 Mar 18;1(2):e000040 [FREE Full text] [doi: 10.1136/esmoopen-2016-000040] [Medline: 27843596]
4. Koo KC, Chung BH. Epidemiology and treatment patterns of urologic cancers in Korea. Korean J Urol Oncol 2015 Sep 27;13(2):51-57 [FREE Full text]
5. Park I, Kang BW, Koh SJ, Kim KH, Kim SH, Kim CK, et al. Korean treatment guidelines for metastatic prostate cancer developed by the Korean association for clinical oncology. Korean J Med 2017;92(2):124-141 [FREE Full text]
6. Pivonello R, Menafra D, Riccio E, Garifalos F, Mazzella M, de Angelis C, et al. Metabolic disorders and male hypogonadotropic hypogonadism. Front Endocrinol (Lausanne) 2019 Jul 25;10:345 [FREE Full text] [doi: 10.3389/fendo.2019.00345] [Medline: 31402895]
7. Gonzalez BD, Small BJ, Cases MG, Williams NL, Fishman MN, Jacobsen PB, et al. Sleep disturbance in men receiving androgen deprivation therapy for prostate cancer: the role of hot flashes and nocturia. Cancer 2018 Feb 01;124(3):499-506 [FREE Full text] [doi: 10.1002/cncr.31024] [Medline: 29072790]
8. Eckman A, Dobs A. Drug-induced gynecomastia. Expert Opin Drug Saf 2008 Nov 04;7(6):691-702. [doi: 10.1517/14740330802442382] [Medline: 18983216]
9. Dirksen SR, Epstein DR, Hoyt MA. Insomnia, depression, and distress among outpatients with prostate cancer. Appl Nurs Res 2009 Aug;22(3):154-158. [doi: 10.1016/j.apnr.2007.09.001] [Medline: 19616162]
10. Freedland SJ, Terris MK, Platz EA, Presti JC. Body mass index as a predictor of prostate cancer: development versus detection on biopsy. Urology 2005 Jul;66(1):108-113. [doi: 10.1016/j.urology.2005.01.060] [Medline: 15992911]
11. Park C, Sung GT, Kim SY, Kim SJ, Hong S, Kim J, et al. Metabolic change after androgen deprivation therapy in Korean men with prostate cancer. Korean J Urol Oncol 2010 Apr;8(2):65-70. [doi: 10.1016/s1569-9056(10)60142-x]
12. Taylor LG, Canfield SE, Du XL. Review of major adverse effects of androgen-deprivation therapy in men with prostate cancer. Cancer 2009 Jun 01;115(11):2388-2399 [FREE Full text] [doi: 10.1002/cncr.24283] [Medline: 19399748]
13. Chi JT, Lin PH, Chen E, Tolstikov V, Oyekunle T, Gao F, et al. Systemic metabolic responses of prostate cancer patients treated with androgen deprivation therapy (ADT) with and without low carb diet. J Urol 2019 Apr;201(Supplement 4):e407 [FREE Full text] [doi: 10.1097/01.ju.0000555715.74466.11]

14. Alexander CM, Landsman PB, Teutsch SM, Haffner SM, Third National Health and Nutrition Examination Survey (NHANES III), National Cholesterol Education Program (NCEP). NCEP-defined metabolic syndrome, diabetes, and prevalence of coronary heart disease among NHANES III participants age 50 years and older. Diabetes 2003 May;52(5):1210-1214. [doi: 10.2337/diabetes.52.5.1210] [Medline: 12716754]
15. Wilson PW, D'Agostino RB, Parise H, Sullivan L, Meigs JB. Metabolic syndrome as a precursor of cardiovascular disease and type 2 diabetes mellitus. Circulation 2005 Nov 15;112(20):3066-3072. [doi: 10.1161/CIRCULATIONAHA.105.539528] [Medline: 16275870]
16. Dacal K, Sereika SM, Greenspan SL. Quality of life in prostate cancer patients taking androgen deprivation therapy. J Am Geriatr Soc 2006 Jan;54(1):85-90. [doi: 10.1111/j.1532-5415.2005.00567.x] [Medline: 16420202]
17. Shin B, Kang TW. Quality of life in prostate cancer patient undergoing androgen deprivation therapy. Korean J Urol Oncol 2018 Apr;16(1):46 [FREE Full text] [doi: 10.22465/kjuo.2018.16.1.46]
18. Braga-Basaria M, Dobs AS, Muller DC, Carducci MA, John M, Egan J, et al. Metabolic syndrome in men with prostate cancer undergoing long-term androgen-deprivation therapy. J Clin Oncol 2006 Aug 20;24(24):3979-3983. [doi: 10.1200/JCO.2006.05.9741] [Medline: 16921050]
19. Lee K, Lee H, Park J, Jee SH. Sex differences in factors associated with metabolic syndrome among Korean adults without diabetes mellitus: results from the Korea National Health and Nutrition Examination Survey from 2010 to 2013. Fam Pract 2019 Mar 20;36(2):140-146. [doi: 10.1093/fampra/cmy053] [Medline: 29878157]
20. Yi Y, An J. Sex differences in risk factors for metabolic syndrome in the Korean population. Int J Environ Res Public Health 2020 Dec 18;17(24):9513 [FREE Full text] [doi: 10.3390/ijerph17249513] [Medline: 33353082]
21. Yoo S, Oh S, Park J, Cho SY, Cho MC, Son H, et al. Effects of metabolic syndrome on the prevalence of prostate cancer: historical cohort study using the national health insurance service database. J Cancer Res Clin Oncol 2019 Mar 17;145(3):775-780. [doi: 10.1007/s00432-019-02842-1] [Medline: 30656408]
22. Lu-Yao G, Stukel TA, Yao SL. Changing patterns in competing causes of death in men with prostate cancer: a population based study. J Urol 2004 Jun;171(6 Pt 1):2285-2290. [doi: 10.1097/01.ju.0000127740.96006.1a] [Medline: 15126804]
23. Walker LM, Tran S, Wassersug RJ, Thomas B, Robinson JW. Patients and partners lack knowledge of androgen deprivation therapy side effects. Urol Oncol 2013 Oct;31(7):1098-1105. [doi: 10.1016/j.urolonc.2011.12.015] [Medline: 22285008]
24. Damji AN, Bies K, Alibhai SM, Jones JM. Bone health management in men undergoing ADT: examining enablers and barriers to care. Osteoporos Int 2015 Mar 20;26(3):951-959. [doi: 10.1007/s00198-014-2997-6] [Medline: 25526712]
25. Wibowo E, Wassersug RJ, Robinson JW, Matthew A, McLeod D, Walker LM. How are patients with prostate cancer managing androgen deprivation therapy side effects? Clin Genitourin Cancer 2019 Jun;17(3):e408-e419. [doi: 10.1016/j.clgc.2018.12.006] [Medline: 30745202]
26. Vermeulen A, Goemaere S, Kaufman JM. Testosterone, body composition and aging. J Endocrinol Invest 1999;22(5 Suppl):110-116. [Medline: 10442580]
27. Cheung AS, Hoermann R, Dupuis P, Joon DL, Zajac JD, Grossmann M. Relationships between insulin resistance and frailty with body composition and testosterone in men undergoing androgen deprivation therapy for prostate cancer. Eur J Endocrinol 2016 Sep;175(3):229-237. [doi: 10.1530/EJE-16-0200] [Medline: 27340081]
28. Saylor PJ, Smith MR. Metabolic complications of androgen deprivation therapy for prostate cancer. J Urol 2009 May;181(5):1998-2008 [FREE Full text] [doi: 10.1016/j.juro.2009.01.047] [Medline: 19286225]
29. Walker LM, Tran S, Robinson JW. Luteinizing hormone--releasing hormone agonists: a quick reference for prevalence rates of potential adverse effects. Clin Genitourin Cancer 2013 Dec;11(4):375-384. [doi: 10.1016/j.clgc.2013.05.004] [Medline: 23891497]
30. Ziglio E, Currie C, Barnekow V. The WHO cross-national study of health behavior in school-aged children from 35 countries: findings from 2001-2002. J Sch Health 2004 Aug;74(6):204-206. [doi: 10.1111/j.1746-1561.2004.tb07933.x] [Medline: 15468523]
31. Can lifestyle modifications using therapeutic lifestyle changes (TLC) reduce weight and the risk for chronic disease. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/nutrition/downloads/r2p_life_change.pdf [accessed 2021-12-07]
32. Lorig KR, Holman H. Self-management education: history, definition, outcomes, and mechanisms. Ann Behav Med 2003 Aug;26(1):1-7. [doi: 10.1207/S15324796ABM2601_01] [Medline: 12867348]
33. Nabid A, Carrier N, Martin AG, Bahary JP, Lemaire C, Vass S, et al. Duration of androgen deprivation therapy in high-risk prostate cancer: a randomized phase III trial. Eur Urol 2018 Oct;74(4):432-441. [doi: 10.1016/j.eururo.2018.06.018] [Medline: 29980331]
34. Keogh JW, Patel A, MacLeod RD, Masters J. Perceived barriers and facilitators to physical activity in men with prostate cancer: possible influence of androgen deprivation therapy. Eur J Cancer Care (Engl) 2014 Mar;23(2):263-273. [doi: 10.1111/ecc.12141] [Medline: 24134506]
35. Lee CE, Kilgour A, Lau YK. Efficacy of walking exercise in promoting cognitive-psychosocial functions in men with prostate cancer receiving androgen deprivation therapy. BMC Cancer 2012 Jul 30;12(1):324 [FREE Full text] [doi: 10.1186/1471-2407-12-324] [Medline: 22846379]

36. Sharpley CF, Birsika V, Denham JW. Factors associated with feelings of loss of masculinity in men with prostate cancer in the RADAR trial. Psychooncology 2014 May 28;23(5):524-530. [doi: 10.1002/pon.3448] [Medline: 24829952]
37. Lim S, Lim H, Després JP. Collateral damage of the COVID-19 pandemic on nutritional quality and physical activity: perspective from South Korea. Obesity (Silver Spring) 2020 Oct 20;28(10):1788-1790 [FREE Full text] [doi: 10.1002/oby.22935] [Medline: 32543113]
38. Responding to com- munity spread of COVID-19: interim guidance. World Health Organization. 2020 Mar 07. URL: https://apps.who.int/iris/handle/10665/331421 [accessed 2021-12-09]
39. Piskurich GM. Online learning: e-learning. Fast, cheap, and good. Perf Improv 2006 Aug 17;45(1):18-24 [FREE Full text] [doi: 10.1002/pfi.2006.4930450105]
40. Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007 May;39(2):175-191. [doi: 10.3758/bf03193146] [Medline: 17695343]
41. Bourke L, Gilbert S, Hooper R, Steed LA, Joshi M, Catto JW, et al. Lifestyle changes for improving disease-specific quality of life in sedentary men on long-term androgen-deprivation therapy for advanced prostate cancer: a randomised controlled trial. Eur Urol 2014 May;65(5):865-872. [doi: 10.1016/j.eururo.2013.09.040] [Medline: 24119318]
42. Fisher JD, Fisher WA. Changing AIDS-risk behavior. Psychol Bull 1992 May;111(3):455-474. [doi: 10.1037/0033-2909.111.3.455] [Medline: 1594721]
43. Kang SW. The validity and reliability of a lifestyle evaluation tool for patients with metabolic syndrome. J Korean Acad Fundam Nurs 2010;17(4):487-497 [FREE Full text]
44. Wei JT, Dunn RL, Litwin MS, Sandler HM, Sanda MG. Development and validation of the expanded prostate cancer index composite (EPIC) for comprehensive assessment of health-related quality of life in men with prostate cancer. Urology 2000 Dec 20;56(6):899-905. [doi: 10.1016/s0090-4295(00)00858-x] [Medline: 11113727]
45. Chung KJ, Kim JJ, Lim SH, Kim TH, Han DH, Lee SW. Development and validation of the Korean version of expanded prostate cancer index composite: questionnaire assessing health-related quality of life after prostate cancer treatment. Korean J Urol 2010 Sep;51(9):601-612 [FREE Full text] [doi: 10.4111/kju.2010.51.9.601] [Medline: 20856644]
46. ICH E9 (R1) addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials Available: Accessed Dec 5 2022. Committee for Medicinal Products for Human Use. 2020 Feb. URL: http://tinyurl.com/4cb6avwm [accessed 2024-01-04]
47. Diggle P, Heagerty P, Liang KY, Zeger S. Analysis of Longitudinal Data. Oxford, UK: Oxford University Press; 2013.
48. Bassi N, Karagodin I, Wang S, Vassallo P, Priyanath A, Massaro E, et al. Lifestyle modification for metabolic syndrome: a systematic review. Am J Med 2014 Dec;127(12):1242.e1-1242.10. [doi: 10.1016/j.amjmed.2014.06.035] [Medline: 25004456]
49. Courteix D, Valente-dos-Santos J, Ferry B, Lac G, Lesourd B, Chapier R, et al. Multilevel approach of a 1-year program of dietary and exercise interventions on bone mineral content and density in metabolic syndrome--the RESOLVE randomized controlled trial. PLoS One 2015 Sep 16;10(9):e0136491 [FREE Full text] [doi: 10.1371/journal.pone.0136491] [Medline: 26376093]
50. McCarthy CG, Canale RE, Alleman Jr R, Reed JP, Bloomer RJ. Biochemical and anthropometric effects of a weight loss dietary supplement in healthy men and women. Nutr Metab Insights 2012;5:13-22 [FREE Full text] [doi: 10.4137/NMI.S8566] [Medline: 23882144]
51. Dutheil F, Lesourd B, Courteix D, Chapier R, Doré E, Lac G. Blood lipids and adipokines concentrations during a 6-month nutritional and physical activity intervention for metabolic syndrome treatment. Lipids Health Dis 2010 Dec 31;9(1):148. [doi: 10.1186/1476-511x-9-148]
52. Bo S, Ciccone G, Baldi C, Benini L, Dusio F, Forastiere G, et al. Effectiveness of a lifestyle intervention on metabolic syndrome. A randomized controlled trial. J Gen Intern Med 2007 Dec 6;22(12):1695-1703 [FREE Full text] [doi: 10.1007/s11606-007-0399-6] [Medline: 17922167]
53. Guzmán M, Zbella E, Alvarez SS, Nguyen JL, Imperial E, Troncale FJ, et al. Effect of an intensive lifestyle intervention on the prevalence of metabolic syndrome and its components among overweight and obese adults. J Public Health (Oxf) 2020 Nov 23;42(4):828-838 [FREE Full text] [doi: 10.1093/pubmed/fdz170] [Medline: 31840755]
54. Kukkonen-Harjula KT, Borg PT, Nenonen AM, Fogelholm MG. Effects of a weight maintenance program with or without exercise on the metabolic syndrome: a randomized trial in obese men. Prev Med 2005 Sep;41(3-4):784-790. [doi: 10.1016/j.ypmed.2005.07.008] [Medline: 16125218]
55. Dawson JK, Dorff TB, Todd Schroeder E, Lane CJ, Gross ME, Dieli-Conwright CM. Impact of resistance training on body composition and metabolic syndrome variables during androgen deprivation therapy for prostate cancer: a pilot randomized controlled trial. BMC Cancer 2018 Apr 03;18(1):368 [FREE Full text] [doi: 10.1186/s12885-018-4306-9] [Medline: 29614993]
56. Focht BC, Lucas AR, Grainger E, Simpson C, Fairman CM, Thomas-Ahner JM, et al. Effects of a group-mediated exercise and dietary intervention in the treatment of prostate cancer patients undergoing androgen deprivation therapy: results from the IDEA-P trial. Ann Behav Med 2018 Apr 19;52(5):412-428 [FREE Full text] [doi: 10.1093/abm/kax002] [Medline: 29684136]

57. Mul JD, Stanford KI, Hirshman MF, Goodyear LJ. Exercise and regulation of carbohydrate metabolism. Prog Mol Biol Transl Sci 2015;135:17-37 [FREE Full text] [doi: 10.1016/bs.pmbts.2015.07.020] [Medline: 26477909]
58. Bray GA, Bouchard C. Handbook of Obesity: Epidemiology, Etiology, and Physiopathology. Boca Raton, FL: CRC Press; 2014.
59. Kim MK, Park JH. Metabolic syndrome. J Korean Med Assoc 2012;55(10):1005 [FREE Full text] [doi: 10.5124/jkma.2012.55.10.1005]
60. Kim KA, Kim SY, Kim JH, Lee SM, Baek JH. Research on Ways to Promote Self-Health Management to Strengthen Follow-Up Care for Examinees in the Pre-Disease Stage : Focusing on Blood Pressure and Blood Sugar Management. Wonju, South Korea: National Health Insurance Corporation Health Insurance Policy Research Institute; 2017.
61. Chae J, Kim Y. Factors influencing the quality of life of prostate cancer patients. Asian Oncol Nurs 2017;17(4):237-245. [doi: 10.5388/aon.2017.17.4.237]
62. Cormie P, Newton RU, Taaffe DR, Spry N, Joseph D, Akhlil Hamid M, et al. Exercise maintains sexual activity in men undergoing androgen suppression for prostate cancer: a randomized controlled trial. Prostate Cancer Prostatic Dis 2013 Jun 15;16(2):170-175. [doi: 10.1038/pcan.2012.52] [Medline: 23318529]
63. Galvão DA, Taaffe DR, Chambers SK, Fairman CM, Spry N, Joseph D, et al. Exercise intervention and sexual function in advanced prostate cancer: a randomised controlled trial. BMJ Support Palliat Care 2022 Mar 16;12(1):29-32 [FREE Full text] [doi: 10.1136/bmjspcare-2020-002706] [Medline: 33328218]
64. Belardinelli R, Lacalaprice F, Faccenda E, Purcaro A, Perna G. Effects of short-term moderate exercise training on sexual function in male patients with chronic stable heart failure. Int J Cardiol 2005 May 11;101(1):83-90. [doi: 10.1016/j.ijcard.2004.05.020] [Medline: 15860388]
65. Lee DM, Nazroo J, O'Connor DB, Blake M, Pendleton N. Sexual health and well-being among older men and women in England: findings from the English longitudinal study of ageing. Arch Sex Behav 2016 Jan 27;45(1):133-144 [FREE Full text] [doi: 10.1007/s10508-014-0465-1] [Medline: 25624001]
66. Mollaioli D, Ciocca G, Limoncin E, Di Sante S, Gravina GL, Carosa E, et al. Lifestyles and sexuality in men and women: the gender perspective in sexual medicine. Reprod Biol Endocrinol 2020 Feb 17;18(1):10 [FREE Full text] [doi: 10.1186/s12958-019-0557-9] [Medline: 32066450]

Abbreviations

AC: abdominal circumference
ADT: androgen deprivation therapy
DBP: diastolic blood pressure
EMR: electronic medical record
EPIC: expanded prostate cancer index composite
EPIC-26: 26-item Expanded Prostate Cancer Index Composite
FBS: fasting blood sugar
HDL: high-density lipoprotein
HRQoL: health-related quality of life
IMB: information-motivation-behavioral
MetS: metabolic syndrome
PC: prostate cancer
SBP: systolic blood pressure
T1: time point 1
T2: time point 2
T3: time point 3

Edited by L Buis; submitted 22.03.23; peer-reviewed by Y Kim, SH Kim; comments to author 13.09.23; revised version received 29.10.23; accepted 20.12.23; published 01.02.24.

Please cite as:
Lee K, Park J, Oh EG, Lee J, Park C, Choi YD
Effectiveness of a Nurse-Led Mobile-Based Health Coaching Program for Patients With Prostate Cancer at High Risk of Metabolic Syndrome: Randomized Waitlist Controlled Trial
JMIR Mhealth Uhealth 2024;12:e47102
URL: https://mhealth.jmir.org/2024/1/e47102
doi:10.2196/47102
PMID:38300697

Original Paper

Patient Engagement With and Perspectives on a Mobile Health Home Spirometry Intervention: Mixed Methods Study

Andrew W Liu[1], BA; William Brown, III[1,2,3,4], MA, DrPH, PhD; Ndubuisi E Madu[1], BA; Ali R Maiorano[1], BA; Olivia Bigazzi[1], MS; Eli Medina[1], MBA; Christopher Sorric[1], MS; Steven R Hays[2], MD; Anobel Y Odisho[1,3,5], MD, MPH

[1]Center for Digital Health Innovation, University of California, San Francisco, CA, United States

[2]Department of Medicine, University of California, San Francisco, CA, United States

[3]Department of Epidemiology and Biostatistics, University of California, San Francisco, CA, United States

[4]Bakar Computational Health Sciences Institute, University of California, San Francisco, CA, United States

[5]Department of Urology, University of California, San Francisco, CA, United States

Corresponding Author:
Anobel Y Odisho, MD, MPH
Center for Digital Health Innovation
University of California
1700 Owens St 541
San Francisco, CA, 94158
United States
Phone: 1 415 353 7171
Email: anobel.odisho@ucsf.edu

Abstract

Background: Patient engagement attrition in mobile health (mHealth) remote patient monitoring (RPM) programs decreases program benefits. Systemic disparities lead to inequities in RPM adoption and use. There is an urgent need to understand patients' experiences with RPM in the real world, especially for patients who have stopped using the programs, as addressing issues faced by patients can increase the value of mHealth for patients and subsequently decrease attrition.

Objective: This study sought to understand patient engagement and experiences in an RPM mHealth intervention in lung transplant recipients.

Methods: Between May 4, 2020, and November 1, 2022, a total of 601 lung transplant recipients were enrolled in an mHealth RPM intervention to monitor lung function. The predictors of patient engagement were evaluated using multivariable logistic and linear regression. Semistructured interviews were conducted with 6 of 39 patients who had engaged in the first month but stopped using the program, and common themes were identified.

Results: Patients who underwent transplant more than 1 year before enrollment in the program had 84% lower odds of engaging (odds ratio [OR] 0.16, 95% CI 0.07-0.35), 82% lower odds of submitting pulmonary function measurements (OR 0.18, 95% CI 0.09-0.33), and 78% lower odds of completing symptom checklists (OR 0.22, 95% CI 0.10-0.43). Patients whose primary language was not English had 78% lower odds of engaging compared to English speakers (OR 0.22, 95% CI 0.07-0.67). Interviews revealed 4 prominent themes: challenges with devices, communication breakdowns, a desire for more personal interactions and specific feedback with the care team about their results, understanding the purpose of the chat, and understanding how their data are used.

Conclusions: Care delivery and patient experiences with RPM in lung transplant mHealth can be improved and made more equitable by tailoring outreach and enhancements toward non-English speakers and patients with a longer time between transplant and enrollment. Attention to designing programs to provide personalization through supplementary provider contact, education, and information transparency may decrease attrition rates.

(JMIR Mhealth Uhealth 2024;12:e51236) doi:10.2196/51236

KEYWORDS

mobile health; mHealth; remote patient monitoring; interview; interviews; dropout; attrition; eHealth; methods; telemedicine; statistics; numerical data; patient-centered care; spirometry; lung transplant; lung; transplant; transplants; transplantation; organ; organs; engagement; monitor; monitoring; pulmonary; respiratory; lungs; experience; experiences; device; devices; complication; complications

Introduction

Many large health systems have turned to remote patient monitoring (RPM) programs to improve population health outcomes, consistently engage with patients, and deliver care more efficiently at scale. RPM uses technology such as mobile devices (mHealth), wearables, or digital devices to communicate health information between patients and providers [1]. A 2022 systemic review found 268 published studies involving the use of RPM, reporting outcomes such as reduced health system costs, decreased hospitalizations, and improved patient quality of life [2,3]. The diversity of RPM clinical use cases includes dementia, diabetes, inflammatory bowel disease, sleep disorders, COVID-19, and cardiovascular disease, among many others [4-7].

However, RPM mHealth programs are most valuable when they maintain consistent patient engagement, allowing for adequate remote monitoring and increased data quality. High patient drop-off in RPM programs is common and seemingly inevitable upon implementation [8]. In a meta-analysis of 17 studies, 43% of patients stopped using the intervention before study completion [9]. In the same meta-analysis, observational real-world studies had even higher (49%) attrition compared to randomized controlled trials. Individual disengagement in RPM programs may arise from a wide range of factors, such as technological difficulties, lack of face-to-face encounters, or irrelevant content [10]. In addition, systemic disparities exist in terms of digital access and literacy, especially among vulnerable populations, leading to inequities in RPM adoption and use [10-13]. There is an urgent need to understand the patient experiences of RPM in real-world studies, especially for patients who have dropped off in engagement, as these patients may face untraditional and uncommunicated needs and challenges.

At the University of California, San Francisco (UCSF), a real-world RPM program for lung transplant recipients was launched in 2020 and is currently ongoing [14]. Patients used Bluetooth-enabled home spirometers to monitor pulmonary function and reported results and outcomes using a web-based chat interface. With the goal of engaging a higher proportion of patients and improving RPM compliance, we used a mixed methods approach to quantitatively identify the clinical and demographic predictors of patient engagement and conducted qualitative semistructured interviews with patients who had stopped using the home spirometry program after the first month to understand the nuances behind patient drop-off.

Methods

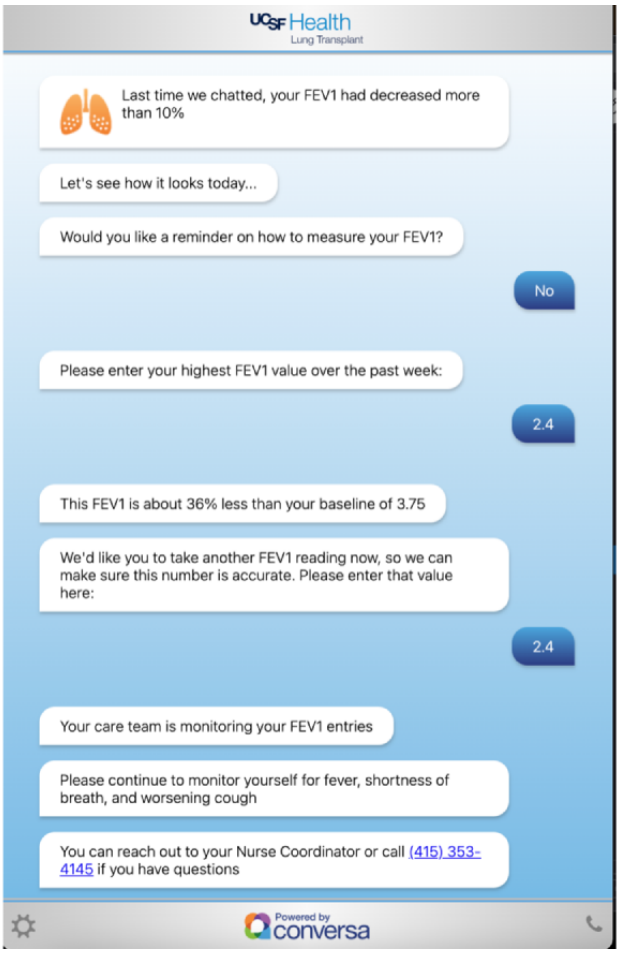
Home Spirometry Program

All patients who have had a lung transplant at UCSF are enrolled in an ongoing real-world RPM mHealth intervention that was implemented as a change in routine care [14]. The intervention is composed of an automated, English-only, chatbot-based symptom monitoring experience powered by a third-party vendor (Conversa Health, Inc), paired with a Bluetooth-enabled handheld home spirometer (Spirobank Smart or SmartOne, MIR Medical International Research) that allows patients to record their forced expiratory volume in the first second (FEV_1) to assess and track their pulmonary function (Figure 1). The program was launched in May 2020, with all patients transplanted after May 1, 2020, automatically enrolled and given the chance to opt out. Additionally, all UCSF patients who had ever had a lung transplant were offered an opportunity to self-enroll (opt in) to the program at the time. In August 2020, all patients who had previously received a lung transplant and who had not self-enrolled were automatically bulk enrolled in the program, allowing them to opt out. Patients transplanted after May 2020 were onboarded in person during their posttransplant care, while patients transplanted before May 2020 were onboarded either digitally through mailed guides or during a routine outpatient follow-up visit.

Through the automated chat, patients can complete individual modules, in which they can report symptoms, manually input new FEV_1 measurements recorded from their Bluetooth spirometer, and receive educational content embedded into chat modules (Figure 2). The goal of the program is early detection of acute and chronic allograft dysfunction and infections. Any abnormal drops in pulmonary function (>10% from baseline for each patient) or concerning patient symptoms generated an immediate alert to an electronic health record shared in-basket that was monitored by the lung transplant care team. Patients were also provided with the clinic contact information and instructed to reach out for additional advice. The transplant team managed clinical findings at their discretion. Additionally, patients are expected to engage with routine automated chat prompts as part of their posttransplant care indefinitely to provide the most up-to-date spirometry and symptom data to their providers, who review with patients during outpatient follow-up visits. Initially, all patients were on a weekly chat reminder cadence, with a reversion to a daily reminder cadence if their condition deteriorated (as defined by a 10% drop in FEV_1 or the reporting of concerning symptoms). In May 2021, the chat reminder cadence was changed to allow patients 1-year posttransplant with stable conditions to opt into a monthly cadence, with the possibility to return to a weekly or daily cadence if their condition deteriorated, and the eventual regraduation to weekly or monthly once they began to recover. Patients could self-initiate chat sessions at will.

Figure 1. Home spirometry device and onboarding.

Figure 2. Screenshot of patient chat experience. FEV1: forced expiratory volume in the first second.



Predictors of Patient Engagement

The 3 primary outcomes were patient engagement with the intervention, including (1) module engagement, defined as completing at least 1 module of any chat, (2) spirometry engagement, defined as patients who submitted at least 1 home spirometry FEV_1 value, or (3) symptom checklist engagement, defined as patients who responded to a symptom-reporting checklist at least once. Unengaged patients were defined as those who did not (1) complete any modules, (2) submit FEV_1 values, or (3) complete symptom checklists. As submitting a FEV_1 or a symptom checklist inherently resulted in the completion of a module, all patients classified as engaged by definitions 2 and 3 were also classified as engaged by definition 1, with definition 1 serving as the broadest outcome. It was possible for patients to have only engaged in educational modules without FEV_1 or symptom submission, resulting in only meeting the module engagement definition without meeting the other 2. Additionally, as secondary outcomes, the number of times patients completed the chat, submitted FEV_1 values, and symptom checklists were also calculated for each patient within the first year after enrollment. Since patients were asked to complete the chat at regularly scheduled repeated intervals, patients who died during their first year of enrollment were excluded from this secondary analysis.

Patient demographic and scheduling data were extracted from the electronic health record. This included patient age, transplant date, sex, race or ethnicity, address, insurance payor, primary language, and marital status. The primary diagnosis resulting in a need for transplantation was classified into 5 categories: restrictive disease, obstructive disease, cystic fibrosis or bronchiectasis, pulmonary hypertension, and other disease. Rural or urban status was assigned at the zip code level using the rural-urban community area codes classification [15]. Area Deprivation Index (ADI) national percentiles, based on a patient's US census block group location, were used as a proxy measure of socioeconomic status [16].

Differences in the patient cohort conditioned on engagement status based on whether or not patients had completed any modules were compared using the chi-square test for categorical features and the 2-sample *t* test (2-tailed) for continuous features. Multivariate logistic and linear regression models were created to assess the predictors of engagement defined by chat module completion and longitudinal engagement, respectively. All analyses were performed using R (version 3.5.1; R Foundation for Statistical Computing), and a P value <.05 was considered significant.

Qualitative Interviews

A convenience sample of patients (n=6) who initially engaged with the mHealth program in the first month but subsequently stopped using it (n=39) were recruited to participate in 40-minute semistructured qualitative interviews. The interviews were led by 1 of 2 user experience designers (NEM and ARM) and conducted through Zoom videoconferencing (Zoom Technologies Inc). Patients were invited to include their significant others or caretakers in the call to help provide additional context and reassurance. Patients were compensated with a US $25 gift card. Interviews were audio-recorded and transcribed. NEM, ARM, AWL, and OB then collaboratively synthesized results to determine concepts and themes by pairing quotes and notes to relevant themes. Three individuals who did not participate in the interviews (AWL, WB III, and AYO) reviewed the selected themes with accompanying quotes and selected quotes for presentation.

The qualitative interview opened with topics concerning the patient's background, including transplant experience, current condition, the frequency of home spirometry, and chat use. Patients were prompted with screenshots of the chat experience and asked about their perceptions, program expectations, reasons for opting out and nonuse, and any challenges they experienced. Patients were also asked about whether solving the challenges they brought up would cause them to reengage at a higher rate. Finally, patients were also given the opportunity to ask questions and provide comments on anything not already touched upon.

Ethical Considerations

The retrospective quantitative and prospective qualitative portions of this study, including patient compensation, were separately reviewed and approved by the University of California, San Francisco institutional review board (22-35950 and 22-35948).

Results

Overview

Between May 4, 2020, and November 1, 2022, a total of 626 patients were enrolled in the chat and 25 patients opted out, resulting in a total of 601 patients included in multivariate logistic regression analysis. Additionally, 33 patients died within their first year of enrollment, resulting in 568 patients included in multivariate linear regression modeling.

In total, 479 (79.7%) patients completed ≥1 module, 433 (72%) patients submitted ≥1 FEV_1 value, and 438 (72.9%) patients submitted ≥1 symptom assessment. The engaged or unengaged cohorts when compared by module completion status were composed of patients of similar sex, race or ethnicity, rural or urban zip code status, marital status, primary language, insurance type, and diagnosis. Engaged patients were marginally younger (66.1 vs 65.2 years; P<.01), had more recent transplants (1.6 years since transplant for engaged patients vs 5.7 years for nonengaged patients; P<.01), and lived in areas with lower ADI national percentiles (14 vs 16; P<.01; Table 1).

Table 1. Patient demographics by module completion (engagement) status (n=601). Patients were considered engaged if they completed any chat modules, including the forced expiratory volume in the first second (FEV1)and symptom submission modules (definition 1).

	Not engaged participants (n=122)	Engaged participants (n=479)	*P* value
Patient age (years), median (IQR)	66.1 (57.7-72.8)	65.2 (55.8-71.2)	<.01
Tim since transplant (years), median (IQR)	5.7 (2.7-9.3)	1.6 (0.2-5.1)	<.01
Transplant date ≥1 year of enrollment date, n (%)	109 (89.7)	279 (58.2)	<.01
Submitted FEV_1 data within 1 year, n (%)	—[a]	433 (90.4)	
Submitted symptom data within 1 year, n (%)	—	437 (91.4)	
FEV_1 submissions in the first year, median (IQR)	—	26 (8-52)	
Symptom submissions in the first year, median (IQR)	—	24 (7-50)	
Total number of modules completed in the first year, median (IQR)	0 (0-0)	23 (9-40)	
Male, n (%)	76 (62.3)	260 (57.6)	.30
Ethnicity, n (%)			.43
Asian, Native Hawaiian, or other Pacific Islander	6 (5.3)	39 (8.5)	
Black or African American	7 (6.1)	37 (8)	
Hispanic or Latino	31 (27.2)	94 (20.4)	
White	62 (54.4)	252 (54.7)	
Other or unknown	8 (7)	39 (8.5)	
Primary language, n (%)			.13
English	99 (86.8)	424 (92)	
Other	15 (13.2)	37 (8)	
Urban, n (%)	114 (95)	439 (95.6)	.96
ADI[b] national percentile, median (IQR)	16 (8-28)	14 (5-30)	<.01
Marital status, n (%)			.92
Married or partnered	79 (69.3)	321 (69.6)	
Single, separated, or other	35 (30.7)	140 (30.4)	
Insurance, n (%)			.09
Commercial	20 (18.2)	127 (28)	
Medicare	73 (66.4)	280 (61.8)	
Medicaid	13 (11.8)	39 (8.6)	
Other	4 (3.6)	7 (1.5)	
Diagnosis, n (%)			.77
Restrictive disease	77 (72)	291 (72.4)	
Obstructive disease	11 (13.1)	31 (12.4)	
Cystic fibrosis	14 (10.3)	50 (7.7)	
Pulmonary hypertension	3 (2.8)	21 (5.2)	
Other disease	2 (1.9)	9 (2.2)	

[a]Not applicable.

[b]ADI: Area Deprivation Index.

Predictors of Engagement

In a multivariate logistic regression model to identify predictors of engagement as defined by completion of a module (definition 1), patients who were enrolled ≥1 year from their transplant had 84% lower odds of engaging compared to those with more recent transplants (odds ratio [OR] 0.16, 95% CI 0.07-0.35; *P*<.01) when demographic factors (race or ethnicity, age, zip code status, insurance type, marital status, diagnosis, and socioeconomic status) were held constant. Similarly, patients

with ≥1 year between transplant and enrollment had 82% lower odds of submitting pulmonary function measurements (OR 0.18, 95% CI 0.09-0.33; *P*<.01) and 78% lower odds of completing symptom checklists (OR 0.22, 95% CI 0.10-0.43; *P*<.01). Patients whose primary language was not English had 78% lower odds of engaging compared to primarily English speakers (OR 0.22, 95% CI 0.07-0.67; *P*<.01). Patient age, race or ethnicity, marital status, insurance type, sex, rural or urban zip code status, ADI national percentile ranking, and diagnosis were not found to have significant associations with engagement in the multivariate logistic regression model (Table 2).

In multivariate linear regression evaluating the number of modules completed in the first year of enrollment, single patient completed 8.69 fewer modules (95% CI –13.75 to –3.64; *P*<.01), patients with non-English primary languages completed 11.89 fewer modules (95% CI –20.62 to –3.17; *P*<.01), and patients with longer duration since transplant completed 15.22 fewer modules (95% CI –19.69 to –10.75; *P*<.01; Table 3). Multivariate analysis conducted on the number of FEV_1 and symptom submissions within the first year found that longer duration between transplant and enrollment and having single or separated or other marital status remained predictors for longitudinal engagement (Table S1 in Multimedia Appendix 1).

Table 2. Predictors of the degree of engagement.[a]

Predictors	Number of FEV_1 submissions		Number of symptom submissions	
	Estimate (95% CI)	*P* value	Estimate (95% CI)	*P* value
Age (years)	0.35 (–0.13 to 0.82)	.16	0.21 (–0.20 to 0.62)	.32
Race or ethnicity (vs White)				
Asian, Native Hawaiian, or other Pacific Islander	0.90 (–13.53 to 7.62)	.92	–2.94 (–17.31 to 11.44)	.69
Black or African American	–8.32 (–0.13 to 0.82)	.23	–8.42 (–20.83 to 3.98)	.18
Hispanic or Latino	–2.96 (–21.82 to 5.19)	.58	–3.49 (–13.16 to 6.18)	.48
Other	5.91 (–6.73 to 18.56)	.36	5.60 (–6.23 to 17.42)	.35
Marital status (vs married or partnered)				
Single or separated or other	–13.86 (–22.13 to –5.59)	<.01	–11.59 (–19.23 to –3.95)	<.01
Insurance (vs commercial)				
Medicare	2.43 (–7.63 to 12.48)	.64	5.40 (–3.62 to 14.42)	.24
Medicaid	1.08 (–13.77 to 15.92)	.89	3.42 (–10.38 to 17.22)	.63
Other	–9.10 (–38.03 to 19.82)	.54	–11.27 (–34.86 to 12.31)	.35
Sex (vs male)				
Female	0.42 (–7.09 to 7.94)	.91	0.07 (–6.82 to 6.96)	.98
Primary language (vs English)				
Non-English	–11.38 (–25.65 to 2.90)	.12	–10.51 (–23.95 to 2.93)	.12
Transplant date ≥1 year of enrollment date	–16.42 (–23.75 to –9.10)	<.01	–14.71 (–21.53 to –7.88)	.01
Rural	–1.55 (–19.04 to 15.94)	.86	0.88 (–15.77 to 17.54)	.92
ADI[b] national percentile	–0.07 (–0.26 to 0.13)	.49	–0.09 (–0.27 to 0.09)	.32
Diagnosis (vs restrictive disease)				
Cystic fibrosis	–3.15 (–19.65 to 13.36)	.71	–7.07 (–21.38 to 7.24)	.33
Obstructive disease	0.39 (–10.56 to 11.34)	.94	2.09 (–8.18 to 12.36)	.69
Pulmonary hypertension	–9.78 (–26.75 to 7.19)	.26	–8.49 (–24.10 to 7.12)	.29
Other disease	8.97 (–16.77 to 34.70)	.49	9.64 (–14.92 to 34.20)	.44

[a]Results from a multivariate linear regression model (n=568).

[b]ADI: Area Deprivation Index.

Table 3. Predictors of module completion engagement. Results from a multivariate logistic regression model (n=601). Patients were considered engaged if they completed any chat modules, including the forced expiratory volume in 1 second and symptom submission modules (definition 1).

Predictors	OR[a] (95% CI)	*P* value
Age (years)	0.99 (0.94-1.03)	.78
Race or ethnicity (vs White)		
Asian, Native Hawaiian, or other Pacific Islander	1.94 (0.51-10.05)	.37
Black or African American	1.19 (0.38-4.63)	.78
Hispanic or Latino	1.20 (0.50-3.10)	.69
Other	1.20 (0.44-3.88)	.74
Marital status (vs married or partnered)		
Single or separated or other	0.95 (0.48-1.94)	.89
Insurance (vs commercial)		
Medicare	0.78 (0.32-1.82)	.58
Medicaid	0.38 (0.12-1.26)	.11
Other	0.20 (0.03-1.28)	.08
Sex (vs male)		
Female	1.25 (0.68-2.34)	.48
Primary language (vs English)		
Non-English	0.22 (0.07-0.67)	<.01
Transplant date ≥1 year of enrollment date	0.16 (0.07-0.35)	<.01
Rural	0.53 (0.13-2.62)	.38
ADI[b] national percentile	1.00 (0.98-1.02)	.95
Diagnosis (vs restrictive disease)		
Cystic fibrosis	1.04 (0.28-4.43)	.96
Obstructive disease	0.85 (0.35-2.23)	.74
Pulmonary hypertension	0.75 (0.19-3.92)	.71
Other disease	1.09 (0.15-22.20)	.94

[a]OR: odds ratio.

[b]ADI: Area Deprivation Index.

Qualitative Interviews

Overview

Interviews were conducted with 2 male and 4 female patients, aged 35-70 years, and White and Latinx. Patients ranged from 1 to 18 years posttransplantation. One patient had his primary caretaker present as his proxy (Table 4). From the interviews, 4 key thematic concepts emerged: challenges with devices, communication breakdowns, desire for more personal interactions and specific feedback with the care team about data and concerns, and understanding the purpose of the care chat.

Table 4. Interviewed patient demographics (n=6). Patients were chosen based on having previously engaged in the first month but then subsequently having had no engagement afterward.

Patient	Age (years); sex	Time since transplant and final date of analysis[a] (years)	Brief background
P1 and PR1	69; male	1.07	Latinx, primary language Spanish, and primary caretaker (PR1) present as proxy
P2	57; male	2.4	White, primary language English, and lives with a partner
P3	36; female	2.17	White and lives alone
P4	30; female	11.18	White and lives alone
P5	44; female	17.84	White and live with partner
P6	70; female	7.31	White and primary caretaker of the spouse

[a]November 1, 2022.

Patients Had Challenges With Devices

Patients reported a diverse set of difficulties associated with the use of their spirometer and mobile devices at home. Challenges encompassed all facets of the experience, such as difficulties with installing the spirometer app on their mobile device and connecting their spirometers through Bluetooth.

> *My machine will not connect to my phone or anybody else's phone. I tried unloading it, reloading it, even took it to AT&T because I'm not very tech savvy.* [Participant 3]

Low technological literacy was a barrier to engagement, and further product development is needed to make it easier to troubleshoot difficulties. One patient reported their preference to carry around an oximeter instead of using their spirometer due to a lack of confidence in their spirometer from experiencing device issues.

Upon use, patients also reported having issues with properly using their spirometer device. Changes in their position or actions led to inconsistent FEV_1 measurements, which were perceived as unreliable in their eyes. For patients, receiving inconsistent FEV_1 readings decreased their confidence, often triggering them to call their care team for guidance.

> *I have a hard time because I have to hold my phone, and I got to blow into it.... if I'm holding the phone down, now I'm having my neck down, it restricts some of my air.* [Participant 2]

Relatedly, patients reported frustration from using their home spirometer.

> *... it deflates me emotionally because it's always low.... I'm not hitting the 3. I'm hitting like, 1.9. So medically, I'm okay. But to me, it keeps on failing.* [Participant 2]

Communication Breakdowns

The transition for patients from using the automated care chat to reaching out to their care team via phone call or secure portal message led to communication breakdowns. Phone calls would occur back and forth between patients and different clinic staff before they were able to reach the right individual and receive relevant answers.

> *Sometimes if you have a question, I guess you have to call your nurse coordinator. And at times you have to wait for them to call you back.... So, I'm at work, so sometimes I miss their phone calls. Sometimes we play phone tag...* [PR1]

Patients Desire More Personal Interactions and Specific Feedback With the Care Team About Data and Concerns

Patients wanted additional opportunities to stay directly connected to their care team. They preferred to communicate with a member of their care team as opposed to entering their FEV_1 or symptom measurements into the chat. “The chat doesn't really offer that opportunity to talk to someone in real life” [PR1].

The automated aspect of the care chat also resulted in a perceived technological barrier between patients and their care team. Patients were keenly aware of the automation and felt it did not provide enough incentive to stay engaged.

> *Lung transplant is a partnership and a lot of hands and a lot of faith and a lot of luck. And a lot of things came together to make a successful surgery.… You’re part of me now, you know. So yeah, any interaction is great. I’d rather talk to you. It's [the Care Chat] not very personal, very impersonal for the drastic surgery that we've had.* [Participant 4]

To patients, the automated chat responses did not provide enough contextual feedback after a FEV_1 measurement was properly submitted to bring patients peace of mind. This resulted in concerns and a lack of confidence in at-home FEV_1 measurements.

> *I don't know what I'm doing in the sense that I'm not getting any feedback from the other end. So,... thinking like, well, what am I doing ?]* [Participant 2]

Patients also wanted to know how their data were being used by the care team to assess their condition and wanted some acknowledgment that someone had reviewed their latest FEV_1 measurement. The lack of feedback from their care team about their inputted values caused patients to want to disengage. There was a strong desire for additional patient-provider interaction around the FEV_1 submission so that patients knew that their data were not going to be ignored.

> *I don't even know if the team ever looks at my what I submit via the care chat, because I don't ever get any feedback from them. So, it's kind of like, I think they see it, but I really don't know. And in that sense, it's like, well, do they really care? Maybe they do. Maybe they don't? This instantly makes me go, oh, well, if Dr. Hays is looking at these numbers, then I definitely want to provide them.* [Participant 5]

Patients Understand the Purpose of the Care Chat Experience

Patients strongly understood that their use of the care chat allowed providers to monitor their progress posttransplant. This resulted in patients being responsive to chat prompts. Furthermore, patients understood the clinical purpose of specifically collecting their FEV_1 measurement.

> *I think the purpose is to monitor our lung function from a distance and for the doctors to be able to get more frequent measurements of our FEV1 without having us come in to do spirometry, or even go into our local hospital. It's kind of a way to keep tabs on us from home in a way that's safe for us and easier for them.* [Participant 5]

At its core, home spirometry helps capture a more comprehensive view of a patient's well-being through consistent remote clinical data measurements and symptom reporting. This more comprehensive view not only allows their care team to be better informed but also respond faster to concerning changes in condition.

> *I think it's pretty great. I found it really helpful. And it's been nice for me to have like a reason to do my FEV1 and have a record of that. I like having more of a stand-up baseline that has more frequent measurements because I've always in clinic other than when I've been sick, it's been really stable, but it's nice to know at home like, oh yeah, this is really kind of where [my condition] lives.* [Participant 5]

Discussion

Principal Findings

We sought to understand engagement in a cohort of 601 patients who underwent lung transplant engaged in a real-world home spirometry mHealth intervention. Engagement and continued use of mHealth tools are critical to effective remote care. However, consistent, quality engagement in mHealth is a difficult challenge and often requires the tailoring of interventions to specific subpopulations [17,18]. We describe several novel predictors of engagement in this intervention and report major themes resulting from qualitative interviews. During our interviews, we found that patients who underwent lung transplant and who had engaged in mHealth RPM spirometry have a strong desire for improved connection to their care team and that when they feel disconnected, are unsure if their data are being received or reviewed, have trouble escalating their concerns to their care team, or experience device difficulties, they are more likely to stop using mobile health (mHealth) tools and home monitoring devices. As the first study (to our knowledge) to focus on in-depth lung transplant patient experiences with drop-off and RPM, this study brings to light patient perspectives that future interventions can learn from for program design specifics.

Multivariate modeling found that having a recent lung transplant (within 1 year of a patient's enrollment date) was consistently found to have a positive association with engagement with the chatbot. One possibility of this is that patients with longer times since transplants might be more confident managing their care without the use of an mHealth app. However, as the potential for chronic allograft dysfunction remains high even 1-year posttransplant, it remains important to engage and reengage patients to reduce the risk of missed allograft dysfunction. Future outreach efforts toward long-term transplant patients to show them the benefits of participation will be required. Additionally, the factors serving as proxies for socioeconomic status (urban or rural status, ADI percentile, and health insurance type) were not found to be associated with engagement in the mHealth program while controlled for other factors during multivariate modeling. This may be due to the fact the spirometers and the chat were provided to patients at no charge. Other factors such as race or ethnicity and age were also not consistently found to be predictive under our 3 definitions of engagement. Finally, our multivariate analysis centered on examining binary engagement with the chat found that having a non-English primary language was associated with lower engagement. Previous literature has shown that patients with a non-English primary language are significantly less likely to engage in mHealth and telemedicine [18]. Our finding has led us to plan to translate the program into Spanish, which is our second largest language cohort (6% of patients), as a next step toward equity.

Comparison With Previous Studies

Our study is not only the first to evaluate predictors of engagement of a mHealth RPM intervention in lung transplant recipients but also one of the first to examine RPM in a practice-wide, real-world setting, as opposed to a randomized controlled trial. Overall, the literature on factors influencing engagement in mHealth is mixed. Previous studies analyzing RPM programs have found age, income, and shorter time in program to be predictive of engagement [19]. Other studies have reported that drops in engagement are not associated with race or ethnicity, disease status, or geography [20]. Our analysis adds to the literature on RPM engagement: reporting that attrition over time remains a large complicated and multifaceted barrier and that trends in engagement are likely program and condition specific rather than being strongly influenced by socioeconomic factors. There is a need to continue to understand how social and clinical factors affect engagement in mHealth interventions to better engage vulnerable populations and not exacerbate existing disparities in care access. In addition, future work can be focused on comparative analyses between differences in engagement rates and outcome measurements (eg, emergency department visits, hospital readmissions).

In addition, interviews also revealed that some patients find the automated chatbot impersonal and crave additional interactions with their care teams. Program design can take this into account by creating defined instances and scenarios where providers

should reach out in case of stress or confusion. Patients desire human contact during their posttransplant care; however, it remains a challenge to balance the need to remotely monitor a large cohort of patients in an automated fashion without overwhelming clinical teams with notifications and also providing a meaningful experience that patients value. Future developments can focus on determining chat design and flow improvements that will offer patients more individualization and education, fine-tuning automated triage mechanisms to identify the right time to have patient-provider interactions, and streamlining patient-provider communication channels to keep patients informed and engaged. Content creation can also be initiated with patient reassurance in mind, such as the development of educational videos featuring their providers or trusted sources. Finally, many patients struggled with proper device setup, spirometer technological difficulties, and confidently using the device to obtain consistent results, leading to confusion and a lack of trust in program quality. These results are consistent with findings showing that low technological literacy and confusion remain major barriers toward even more widespread adoption of RPM and mHealth [10,21,22]. Spirometry device satisfaction levels have been found to greatly influence patient engagement levels, as well as constant contact and follow-up by physicians to increase patient satisfaction [23,24]. In response, program enhancements were built to alleviate the cognitive burden on patients by streamlining onboarding and reducing the number of unnecessary chats. Future developments can aim to help patients build confidence and trust with home spirometry, by adding staff for more detailed onboarding, focusing on education for patients who have technological issues or concerns during the onboarding period, and chat design incorporating more person-to-person real-time feedback by providers.

Limitations

Our study has several limitations. Our multivariate analysis did not account for health and digital literacy levels as these were not systematically assessed as part of routine clinical care and both have been shown to play major roles in a patient's abilities to access health care. While our analysis factored in disease diagnosis type, there are likely additional clinical factors that may affect patient engagement, as transplant recipients with worse pulmonary function may be more apt to consistently engage to monitor their health. Future work can examine the efficacy of home spirometry monitoring in detecting adverse events when controlling for patients' clinical conditions. Next, our study is a single-institution study of patients who have received lung transplant, and therefore the overall generalizability is reduced due to the specific nature of the patient's condition, progression, and experiences with the program specifics such as Bluetooth spirometry. Interviewed patients were recruited only from those who initially dropped out after 1 month, and therefore the themes noted may only be representative of a dropped off in the engagement cohort. More work is required to elucidate the perspectives and reasons for continual engagement from patients whose engagements were consistent, steadily decreased over time, or dropped off at a later point in the program. In the future, we plan to interview patients who had excellent adherence to understand what factors are promoting adherence in these cases. Furthermore, the small sample size of our interview cohort increases the possibility of bias during the process of thematic analysis, particularly because we only interviewed 1 non-English speaker (using an interpreter, with the patient's proxy present). Analyses from numerous implemented mHealth programs have shown that non-English speakers and patients of minority race or ethnicity are significantly less likely to engage in telemedicine [13,25,26]. Patient motivations for drop-off are nuanced and are likely strongly associated with sociocultural factors that can only be uncovered in larger, more diverse studies.

Conclusions

An mHealth intervention consisting of home spirometry paired with an automated care chat results in high engagement rates in patients who have received a lung transplant, particularly in those with more recent transplants. Interviews conducted on patients who have dropped off in engagement revealed program challenges and areas where mHealth care delivery can be improved to reduce engagement attrition, including addressing technological barriers and improving patient confidence.

Acknowledgments

WB was supported by the Agency for Healthcare Research and Quality (K12HS026383) and the National Center for Advancing Translational Sciences (KL2TR001870) during the writing of this research.

Authors' Contributions

Conceptualization, project administration, data curation, formal analysis, investigation, writing—original draft, review, and editing were done by AWL. Project administration, data curation, formal analysis, investigation, writing—review, and editing were done by WB. Data curation, methodology, and project administration were done by NEM. Methodology and project administration were performed by ARM. Conceptualization and funding acquisition was done by OB. Funding acquisition and project administration were performed by EM. Data curation and formal analysis were performed by CS. Funding acquisition, resources, conceptualization, methodology, investigation, writing—review and editing, and supervision were done by SRH. Funding acquisition, conceptualization, formal analysis, methodology, investigation, writing—review and editing, and supervision were done by AYO.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Predictors of the degree of engagement. Results from a multivariate logistic regression model.
[DOCX File , 16 KB - mhealth_v12i1e51236_app1.docx]

References

1. Farias FAC, Dagostini CM, de Assunção Bicca Y, Falavigna VF, Falavigna A. Remote patient monitoring: a systematic review. Telemed J E Health 2020;26(5):576-583 [FREE Full text] [doi: 10.1089/tmj.2019.0066] [Medline: 31314689]
2. Vegesna A, Tran M, Angelaccio M, Arcona S. Remote patient monitoring via non-invasive digital technologies: a systematic review. Telemed J E Health 2017;23(1):3-17 [FREE Full text] [doi: 10.1089/tmj.2016.0051] [Medline: 27116181]
3. Hayes CJ, Dawson L, McCoy H, Hernandez M, Andersen J, Ali MM, et al. Utilization of remote patient monitoring within the United States health care system: a scoping review. Telemed J E Health 2023;29(3):384-394. [doi: 10.1089/tmj.2022.0111] [Medline: 35819861]
4. Gaugler JE, Zmora R, Mitchell LL, Finlay JM, Peterson CM, McCarron H, et al. Six-month effectiveness of remote activity monitoring for persons living with dementia and their family caregivers: an experimental mixed methods study. Gerontologist 2019;59(1):78-89 [FREE Full text] [doi: 10.1093/geront/gny078] [Medline: 29982413]
5. Su D, Michaud TL, Estabrooks P, Schwab RJ, Eiland LA, Hansen G, et al. Diabetes management through remote patient monitoring: the importance of patient activation and engagement with the technology. Telemed J E Health 2019;25(10):952-959. [doi: 10.1089/tmj.2018.0205] [Medline: 30372366]
6. Chugh R, Liu A, Bigazzi O, Maiorano A, Medina E, Pierce L, et al. A digital health platform for the clinical care of inflammatory bowel disease patients. Gastroenterology 2022;162(7):S-257-S-258 [FREE Full text] [doi: 10.1016/s0016-5085(22)60612-7]
7. Kayyali HA, Weimer S, Frederick C, Martin C, Basa D, Juguilon JA, et al. Remotely attended home monitoring of sleep disorders. Telemed J E Health 2008;14(4):371-374. [doi: 10.1089/tmj.2007.0058] [Medline: 18570567]
8. Eysenbach G. The law of attrition. J Med Internet Res 2005;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
9. Meyerowitz-Katz G, Ravi S, Arnolda L, Feng X, Maberly G, Astell-Burt T. Rates of attrition and dropout in app-based interventions for chronic disease: systematic review and meta-analysis. J Med Internet Res 2020;22(9):e20283 [FREE Full text] [doi: 10.2196/20283] [Medline: 32990635]
10. Lie SS, Karlsen B, Oord ER, Graue M, Oftedal B. Dropout from an eHealth intervention for adults with type 2 diabetes: a qualitative study. J Med Internet Res 2017;19(5):e187 [FREE Full text] [doi: 10.2196/jmir.7479] [Medline: 28559223]
11. Ratcliff CL, Krakow M, Greenberg-Worisek A, Hesse BW. Digital health engagement in the US population: insights from the 2018 health information national trends survey. Am J Public Health 2021;111(7):1348-1351. [doi: 10.2105/AJPH.2021.306282] [Medline: 34014759]
12. Yao R, Zhang W, Evans R, Cao G, Rui T, Shen L. Inequities in health care services caused by the adoption of digital health technologies: scoping review. J Med Internet Res 2022;24(3):e34144 [FREE Full text] [doi: 10.2196/34144] [Medline: 35311682]
13. Schenker RB, Laguna MC, Odisho AY, Okumura MJ, Burnett H. Are we reaching everyone? A cross-sectional study of telehealth inequity in the COVID-19 pandemic in an urban academic pediatric primary care clinic. Clin Pediatr (Phila) 2022;61(1):26-33 [FREE Full text] [doi: 10.1177/00099228211045809] [Medline: 34514898]
14. Odisho AY, Liu AW, Maiorano AR, Bigazzi MOA, Medina E, Leard LE, et al. Design and implementation of a digital health home spirometry intervention for remote monitoring of lung transplant function. J Heart Lung Transplant 2023;42(6):828-837. [doi: 10.1016/j.healun.2023.01.010] [Medline: 37031033]
15. Rural-urban commuting area codes. Economic Research Service: U.S. Department of Agriculture. URL: https://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes [accessed 2021-07-19]
16. Kind AJH, Buckingham WR. Making neighborhood-disadvantage metrics accessible-the neighborhood atlas. N Engl J Med 2018;378(26):2456-2458 [FREE Full text] [doi: 10.1056/NEJMp1802313] [Medline: 29949490]
17. Chien I, Enrique A, Palacios J, Regan T, Keegan D, Carter D, et al. A machine learning approach to understanding patterns of engagement with internet-delivered mental health interventions. JAMA Netw Open 2020;3(7):e2010791 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.10791] [Medline: 32678450]
18. Slater H, Stinson JN, Jordan JE, Chua J, Low B, Lalloo C, et al. Evaluation of digital technologies tailored to support young people's self-management of musculoskeletal pain: mixed methods study. J Med Internet Res 2020;22(6):e18315 [FREE Full text] [doi: 10.2196/18315] [Medline: 32442143]
19. Kirkland E, Schumann SO, Schreiner A, Heincelman M, Zhang J, Marsden J, et al. Patient demographics and clinic type are associated with patient engagement within a remote monitoring program. Telemed J E Health 2021;27(8):843-850 [FREE Full text] [doi: 10.1089/tmj.2020.0535] [Medline: 34115942]
20. Pratap A, Neto EC, Snyder P, Stepnowsky C, Elhadad N, Grant D, et al. Indicators of retention in remote digital health studies: a cross-study evaluation of 100,000 participants. NPJ Digit Med 2020;3:21 [FREE Full text] [doi: 10.1038/s41746-020-0224-8] [Medline: 32128451]

21. Steinhubl SR, Muse ED, Topol EJ. The emerging field of mobile health. Sci Transl Med 2015;7(283):283rv3 [FREE Full text] [doi: 10.1126/scitranslmed.aaa3487] [Medline: 25877894]
22. Li J, Wilson LS. Telehealth trends and the challenge for infrastructure. Telemed J E Health 2013;19(10):772-779. [doi: 10.1089/tmj.2012.0324] [Medline: 23980936]
23. Fadaizadeh L, Najafizadeh K, Shajareh E, Shafaghi S, Hosseini M, Heydari G. Home spirometry: assessment of patient compliance and satisfaction and its impact on early diagnosis of pulmonary symptoms in post-lung transplantation patients. J Telemed Telecare 2016;22(2):127-131 [FREE Full text] [doi: 10.1177/1357633X15587435] [Medline: 26026187]
24. Fadaizadeh L, Najafizadeh K, Shafaghi S, Hosseini MS, Ghoroghi A. Using home spirometry for follow up of lung transplant recipients: "a pilot study". Tanaffos 2013;12(1):64-69 [FREE Full text] [Medline: 25191451]
25. Ganeshan S, Pierce L, Mourad M, Judson TJ, Kohli MD, Odisho AY, et al. Impact of patient portal-based self-scheduling of diagnostic imaging studies on health disparities. J Am Med Inform Assoc 2022;29(12):2096-2100 [FREE Full text] [doi: 10.1093/jamia/ocac152] [Medline: 36063414]
26. Sinha S, Garriga M, Naik N, McSteen BW, Odisho AY, Lin A, et al. Disparities in electronic health record patient portal enrollment among oncology patients. JAMA Oncol 2021;7(6):935-937 [FREE Full text] [doi: 10.1001/jamaoncol.2021.0540] [Medline: 33830178]

Abbreviations

ADI: Area Deprivation Index
FEV1: forced expiratory volume in the first second
mHealth: mobile health
OR: odds ratio
RPM: remote patient monitoring
UCSF: University of California, San Francisco

Edited by L Buis; submitted 26.07.23; peer-reviewed by J Edwards, C Peterson; comments to author 08.09.23; revised version received 27.10.23; accepted 12.12.23; published 20.03.24.

Please cite as:
Liu AW, Brown, III W, Madu NE, Maiorano AR, Bigazzi O, Medina E, Sorric C, Hays SR, Odisho AY
Patient Engagement With and Perspectives on a Mobile Health Home Spirometry Intervention: Mixed Methods Study
JMIR Mhealth Uhealth 2024;12:e51236
URL: https://mhealth.jmir.org/2024/1/e51236
doi:10.2196/51236
PMID:38506896

Original Paper

Mining User Reviews From Hypertension Management Mobile Health Apps to Explore Factors Influencing User Satisfaction and Their Asymmetry: Comparative Study

Yunfan He[1], MSc; Wei Zhu[2,3], PhD; Tong Wang[4,5,6], PhD; Han Chen[7], MD, PhD; Junyi Xin[8], MS; Yongcheng Liu[9], MS; Jianbo Lei[10,11,12], MD, PhD; Jun Liang[1,13,14,15], MS

[1]Center for Health Policy Studies, School of Public Health, Zhejiang University, Hangzhou, China
[2]Department of Cardiology, Second Affiliated Hospital, College of Medicine, Zhejiang University, Hangzhou, China
[3]Cardiovascular Key Laboratory of Zhejiang Province, Hangzhou, China
[4]School of Health and Life Sciences, University of Health and Rehabilitation Sciences, Qingdao, China
[5]School of Basic Medical Sciences, Shandong University, Jinan, China
[6]Qingdao Hospital, University of Health and Rehabilitation Sciences (Qingdao Municipal Hospital), Qingdao, China
[7]Department of Cardiology, Second Affiliated Hospital, College of Medicine, Zhejiang University, Hangzhou, China
[8]School of Information Engineering, Hangzhou Medical College, Hangzhou, China
[9]Netease Group, Hangzhou, China
[10]Clinical Research Center, Affiliated Hospital of Southwest Medical University, Luzhou, China
[11]The First Affiliated Hospital, Hainan Medical University, Haikou, China
[12]Center for Medical Informatics, Health Science Center, Peking University, Beijing, China
[13]Department of AI and IT, Second Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou, China
[14]Key Laboratory of Cancer Prevention and Intervention,, China National Ministry of Education, School of Medicine, Zhejiang University, Hangzhou, China
[15]School of Medical Technology and Information Engineering, Zhejiang Chinese Medical University, Hangzhou, China

Corresponding Author:
Jun Liang, MS
Department of AI and IT
Second Affiliated Hospital, School of Medicine
Zhejiang University
88 Jiefang Road, Shangcheng District
Hangzhou, 310000
China
Phone: 86 571 87783942
Fax: 86 571 87783526
Email: junl@zju.edu.cn

Abstract

Background: Hypertension significantly impacts the well-being and health of individuals globally. Hypertension management apps (HMAs) have been shown to assist patients in controlling blood pressure (BP), with their efficacy validated in clinical trials. However, the utilization of HMAs continues to be suboptimal. Presently, there is a dearth of real-world research based on big data and exploratory mining that compares Chinese and American HMAs.

Objective: This study aims to systematically gather HMAs and their user reviews from both China and the United States. Subsequently, using data mining techniques, the study aims to compare the user experience, satisfaction levels, influencing factors, and asymmetry between Chinese and American users of HMAs. In addition, the study seeks to assess the disparities in satisfaction and its determinants while delving into the asymmetry of these factors.

Methods: The study sourced HMAs and user reviews from 10 prominent Chinese and American app stores globally. Using the latent Dirichlet allocation (LDA) topic model, the research identified various topics within user reviews. Subsequently, the Tobit model was used to investigate the impact and distinctions of each topic on user satisfaction. The Wald test was applied to analyze differences in effects across various factors.

Results: We examined a total of 261 HMAs along with their associated user reviews, amounting to 116,686 reviews in total. In terms of quantity and overall satisfaction levels, Chinese HMAs (n=91) and corresponding reviews (n=16,561) were notably fewer compared with their American counterparts (n=220 HMAs and n=100,125 reviews). The overall satisfaction rate among HMA users was 75.22% (87,773/116,686), with Chinese HMAs demonstrating a higher satisfaction rate (13,866/16,561, 83.73%) compared with that for American HMAs (73,907/100,125, 73.81%). Chinese users primarily focus on reliability (2165/16,561, 13.07%) and measurement accuracy (2091/16,561, 12.63%) when considering HMAs, whereas American users prioritize BP tracking (17,285/100,125, 17.26%) and data synchronization (12,837/100,125, 12.82%). Seven factors (easy to use: $P<.001$; measurement accuracy: $P<.001$; compatibility: $P<.001$; cost: $P<.001$; heart rate detection function: $P=.02$; blood pressure tracking function: $P<.001$; and interface design: $P=.01$) significantly influenced the positive deviation (PD) of Chinese HMA user satisfaction, while 8 factors (easy to use: $P<.001$; reliability: $P<.001$; measurement accuracy: $P<.001$; compatibility: $P<.001$; cost: $P<.001$; interface design: $P<.001$; real-time: $P<.001$; and data privacy: $P=.001$) affected the negative deviation (ND). Notably, BP tracking had the greatest effect on PD ($\beta=.354$, $P<.001$), while cost had the most significant impact on ND ($\beta=3.703$, $P<.001$). All 12 factors (easy to use: $P<.001$; blood pressure tracking function: $P<.001$; data synchronization: $P<.001$; blood pressure management effect: $P<.001$; heart rate detection function: $P<.001$; data sharing: $P<.001$; reliability: $P<.001$; compatibility: $P<.001$; interface design: $P<.001$; advertisement distribution: $P<.001$; measurement accuracy: $P<.001$; and cost: $P<.001$) significantly influenced the PD and ND of American HMA user satisfaction. Notably, BP tracking had the greatest effect on PD ($\beta=0.312$, $P<.001$), while data synchronization had the most significant impact on ND ($\beta=2.662$, $P<.001$). In addition, the influencing factors of PD and ND in user satisfaction of HMA in China and the United States are different.

Conclusions: User satisfaction factors varied significantly between different countries, showing considerable asymmetry. For Chinese HMA users, ease of use and interface design emerged as motivational factors, while factors such as cost, measurement accuracy, and compatibility primarily contributed to user dissatisfaction. For American HMA users, motivational factors were ease of use, BP tracking, BP management effect, interface design, measurement accuracy, and cost. Moreover, users expect features such as data sharing, synchronization, software reliability, compatibility, heart rate detection, and nonintrusive advertisement distribution. Tailored experience plans should be devised for different user groups in various countries to address these diverse preferences and requirements.

(JMIR Mhealth Uhealth 2024;12:e55199) doi:10.2196/55199

KEYWORDS

hypertension management; mobile health; topic modeling; satisfaction; 2-factor model; comparative study

Introduction

The global prevalence of hypertension is on the rise. Hypertension management apps (HMAs) serve as convenient tools for effectively managing blood pressure (BP). These apps enhance users' awareness of self-management, dietary and exercise habits, and medication adherence through features such as BP tracking, dietary guidance, exercise monitoring, educational resources, and medication reminders. The ultimate aim is to achieve effective BP control. Their effectiveness has been demonstrated in experimental settings. Globally, the number of patients with hypertension surged from 648 million in 1990 to 1.278 billion in 2019, marking a prevalence of 33% [1]. HMAs represent digital health tools with the potential for effectively controlling BP [2-9]. Their usability [10,11] and effectiveness [12] have been demonstrated in randomized controlled trials. In a systematic review and meta-analysis of 18 randomized controlled trials, Han et al [13] found that HMAs could significantly reduce BP and improve BP control rates. Moreover, within the framework of the current value-based medical policy, a hospital's treatment efficacy relies not only on its in-hospital care but also on postdischarge patient attention and management. As a collaborative tool for out-of-hospital medical services, HMAs can assist patients in sustaining the effects of in-hospital treatments, thereby reducing the rate of hospital readmissions. McManus et al [14] demonstrated that compared with traditional nursing interventions, apps yield superior BP control effects within a year and incur lower incremental costs. However, in practical settings, many HMAs have been developed without adhering to evidence-based medicine [15] and lacked clinical certification before marketing [16], potentially posing adverse effects on hypertension management [17,18].

Although HMAs show significant BP management effects in controlled environments, their real-world outcomes are often unsatisfactory. Both user engagement and retention rates are low, with 62% of mobile health (mHealth) apps having fewer than 1000 monthly active users [19] and only 6.6% of patients with hypertension continuing to use HMAs [20]. These situations directly correlate with user satisfaction [21-24]. Therefore, enhancing user satisfaction can significantly improve the effectiveness of the app [25]. According to the Food and Drug Administration (FDA), in real-world scenarios, health information technology (HIT) is integrated into a complex sociotechnical system, and its actual impact is influenced by 4 primary factors aside from the product itself: people (whether they are involved or not), technologies (including HIT hardware and software), processes (the workflow of health care delivery), and organization (the procedure for HIT installation and configuration). Among these factors, external environmental elements (such as the policies and cultures of different countries) play a significant role [26]. The actual impacts of HMAs can vary significantly among user groups in different countries. Therefore, it is essential to explore and compare the satisfaction levels and influencing factors of user groups across various

countries. This comparative analysis can enable targeted efforts to enhance the practical effectiveness of HMAs.

Traditional studies on user satisfaction of HMAs primarily consist of qualitative analyses [27,28]. However, these studies often lack breadth and depth, characterized by limitations such as small sample sizes, susceptibility to adverse observer effects and recall bias, and limited generalizability of conclusions. Furthermore, these studies often overlook the impact of external environmental factors and do not compare or analyze the satisfaction levels of user groups across different countries. For instance, Li et al [29] conducted semistructured interviews with 13 English-speaking patients with hypertension to explore their surface-level satisfaction with HMAs and analyze factors contributing to dissatisfaction. Kang and Park [30] developed an English HMA grounded on clinical guidelines for hypertension management. They used the modified Morisky Scale to assess perceived usefulness and satisfaction with this app among 38 patients diagnosed with hypertension. Although qualitative analyses allow for a deep exploration of individual attitudes, they have limitations such as a restricted number of apps that can be analyzed [31,32], a small sample size confined to a specific region [33], high research costs, and potential biases in conclusions [34].

Research on user satisfaction with HMAs lacks both quantitative studies driven by large volumes of user-generated content and in-depth exploration of the factors influencing satisfaction, particularly in terms of asymmetry. In addition, there is a notable absence of comparative analyses of user groups across different countries. Plante et al [35] manually annotated and summarized English user reviews of an HMA (Instant BP) from an online app store, discovering that users expressed greater satisfaction with apps yielding lower measurement results. Similarly, Wang et al [36] and Nuo et al [37] conducted quantitative analyses of user reviews for weight and sleep management apps to investigate user satisfaction and influencing factors. However, none of these studies conducted an in-depth evaluation of variations in satisfaction and influencing factors among user groups across different countries. The renowned Herzberg 2-factor theory [38] in management suggests that the factors influencing user satisfaction exhibit asymmetry and can be categorized into motivational factors that enhance satisfaction and hygienic factors that mitigate dissatisfaction. Using large-sample data and incorporating the asymmetry of factors impacting app user satisfaction, while also considering variations among user groups in different countries, can effectively mitigate errors in constructing explanatory models. This approach enhances the predictability and generalizability of the model [39].

Therefore, this study adopts the 2-factor model and uses unsupervised clustering algorithms to quantitatively analyze user reviews of HMAs from major Chinese and American app stores globally. By considering the macro usage environment, the study aims to extract and compare the primary opinions of Chinese and American users, assess differences in satisfaction and its influencing factors, and explore the asymmetry within these factors.

Methods

Informed Consent and Study Approval Statement

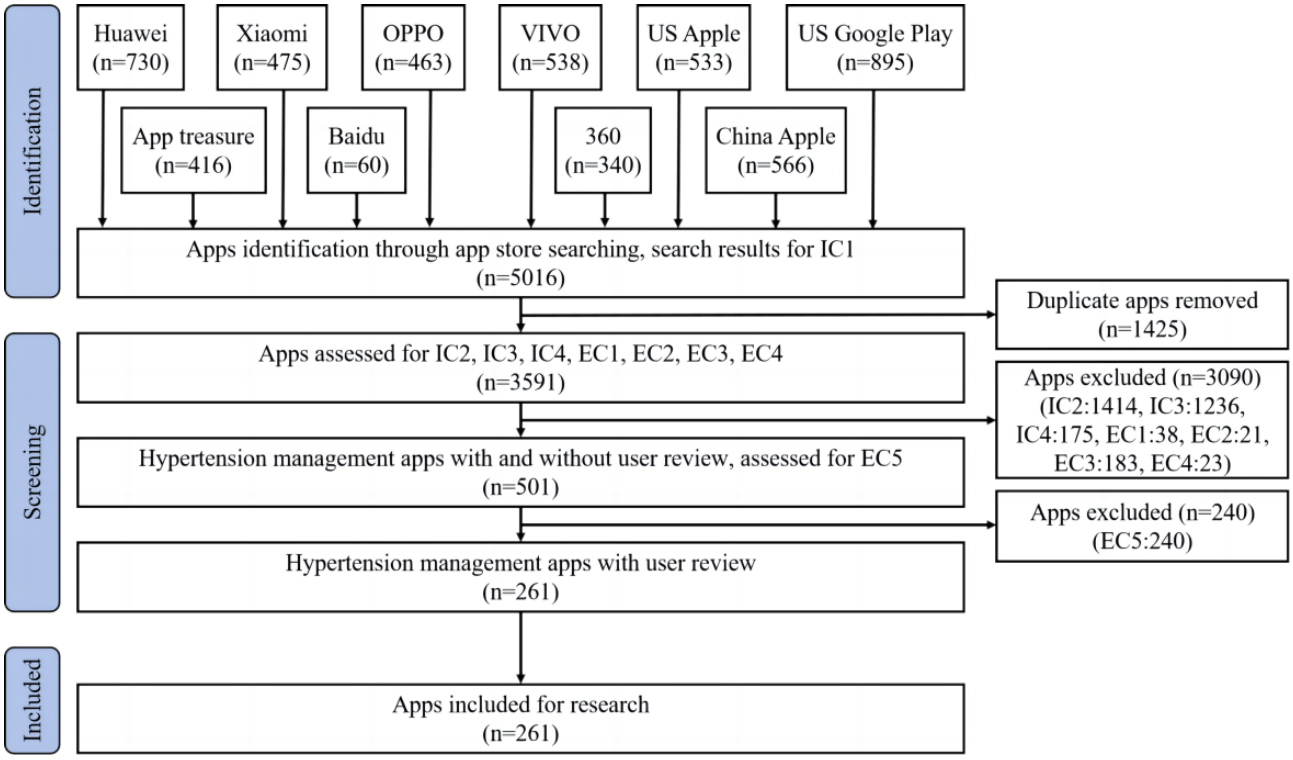
All data used in this study were sourced from publicly accessible internet mobile app stores, encompassing app information and user reviews. Hence, this study does not entail any medical ethics concerns.

Data Collection

HMA Identification

In this comparative study, to conduct a comprehensive review of the primary HMAs, we consulted previous studies [36,37] and identified 10 widely used Chinese and American app stores from the 2 major mobile phone platforms (iOS [Apple Inc.] and Android [Google LLC]). These include 8 Chinese platforms (China Apple, Huawei, Xiaomi, OPPO, VIVO, Baidu, 360, and Application Treasure) and 2 American platforms (US Apple App Store and US Google Play Store). App stores in different countries can only access user data from their respective regions. Therefore, in April 2023, we conducted searches across the aforementioned 10 Chinese and American app stores using keywords such as “hypertension,” “high blood pressure,” “blood pressure management,” and “blood pressure recording.” We retrieved a total of 5016 apps, out of which 3591 remained after deduplication (Table S1 in Multimedia Appendix 1). Following guidelines from previous studies [40], we formulated detailed inclusion and exclusion criteria (Table S2 in Multimedia Appendix 1), which were independently screened by 2 researchers (YFH and J Liang). Both researchers underwent standardized training before the screening process, resulting in high consistency in their screening results (κ=0.84). Any discrepancies between the researchers were resolved through arbitration by another cardiovascular clinical expert (WZ). The specific screening process, following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 guidelines [41], is depicted in Figure 1.

Figure 1. Flowchart of the hypertension management app screening process; IC: inclusion criteria; EC: exclusion criteria.



User Review Collection

All 10 Chinese and American app platforms offer user ratings and review features, enabling users to provide quantitative ratings and qualitative descriptions of their satisfaction with apps. These ratings typically range from 1 (very poor) to 5 (very good). We used Python scripts (Python Foundation) and the Qimai mobile app data analysis platform [42] to extract all quantitative ratings and qualitative user reviews of the included HMAs. As of April 23, 2023, we collected a total of 295,927 quantitative ratings and qualitative user reviews, comprising 250,193 reviews from American users and 45,734 reviews from Chinese users. The proportion of Chinese reviews and downloads (16,561/24,204,832, 0.07%) is similar to that of American reviews and downloads (100,125/148,869,181, 0.07%).

User Review Preprocessing

The presence of false and meaningless user reviews in the app store data significantly impacted the topic mining of user reviews and interfered with the assessment of user satisfaction with HMAs. Hence, we conducted preprocessing on the user reviews with the following steps:

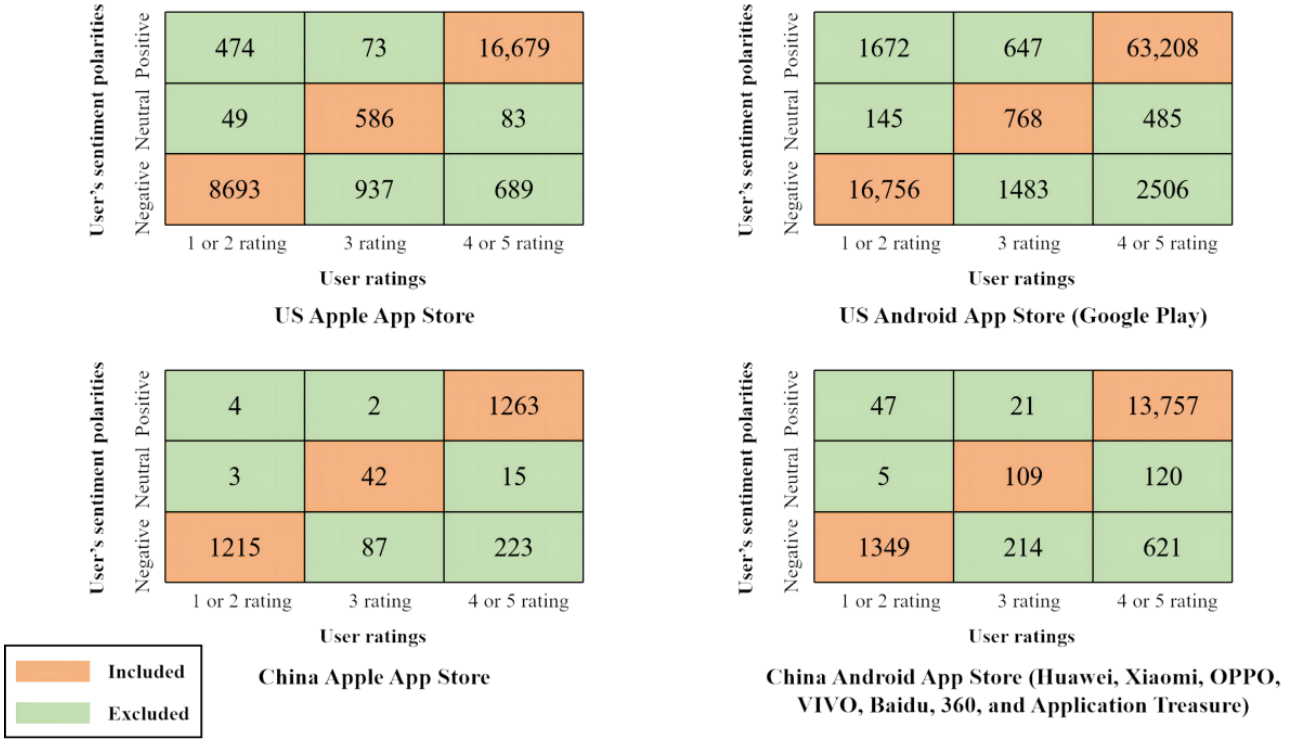
Removing data containing only ratings without accompanying user reviews.

Excluding user reviews posted on bot accounts, using the tweetbotornot package in R [43]. This resulted in the exclusion of 53,431 American and 16,940 Chinese reviews.

Eliminating duplicate reviews, blank values, non-Chinese or English reviews, garbled characters, and reviews deemed meaningless. This process led to the exclusion of 80,829 American and 9697 Chinese reviews.

Following this, we used the sentiment knowledge–enhanced pretraining algorithm [44] to calculate the emotional polarity of each user review, categorized into negative, neutral, and positive sentiments. Furthermore, contradictory data showing inconsistencies between user ratings and reviews were eliminated [45], resulting in the exclusion of 9243 American and 1362 Chinese reviews. Specifically, we removed data with user ratings of 1 or 2 points that lacked negative emotional polarities in user reviews, data with user ratings of 3 points that lacked neutral emotional polarities in user reviews, and data with user ratings of 4 or 5 points that missed positive emotional polarities in user reviews (Figure 2). Finally, we uniformly labeled different data with the same concept (Table S3 in Multimedia Appendix 1). Following the initial data preprocessing steps, a total of 124,425 user reviews were included in the LDA model for topic modeling, comprising 106,690 American reviews and 17,735 Chinese reviews.

Figure 2. Matching between user's sentiment polarities (rows) and user ratings (columns).



Data Analysis

Overview

We used natural language processing technology and the LDA topic model to extract the main topics from both Chinese and American user reviews. Subsequently, following the 2-factor model, we constructed a Tobit model to analyze the correlation between different topics and user satisfaction. Finally, we used the Wald test to analyze the differences in the impact of each topic on user satisfaction.

LDA Topic Modeling

LDA is a widely used probability-based topic modeling algorithm [46]. It is known for its ease of operation, high efficiency, and positive impact on topic clustering and prediction accuracy [47]. As LDA is primarily a language model rather than a classification prediction model, perplexity serves as a common and effective indicator for evaluating the quality of the language model, rather than sensitivity. To efficiently and accurately extract the primary opinions and topics from Chinese and American user reviews, we used LDA, which is a 3-level hierarchical Bayesian model. LDA calculates the distribution probabilities of words and topics, enabling the clustering of latent semantic structures within user reviews to summarize the main topics.

For the LDA model input, we used the word segmentation set and specified the number of Chinese and American review topics, along with manually induced topics. Consequently, we executed the LDA topic modeling process as follows: Initially, we used the Jieba [48] and NLTK packages [49] in Python to segment the Chinese and American reviews, respectively. We then used stop-word lists compiled from Baidu, Harbin Institute of Technology, Sichuan University Machine Intelligence Laboratory, and standard Chinese and English stop words [50]. These stop-word lists were applied to delete stop words, including numbers, punctuation marks, emoticons, and blank values, from both Chinese and English word segmentations. Subsequently, any blank reviews were removed after the stop words had been eliminated. In addition, we conducted morphological restoration of English word segmentation. Subsequently, based on the perplexity curve and actual clustering effect [51], we determined that 11 topics were optimal for Chinese reviews and 12 topics for American reviews (Figures S1–S4 in Multimedia Appendix 1). Furthermore, we used the Gensim package [52] in Python to construct the LDA topic models for both Chinese and American reviews. Two researchers (YFH and J Liang) independently summarized and named the topics for each keyword set in the Chinese and American review topic models, respectively. Any discrepancies were resolved through arbitration by a third expert (WZ). Finally, the topics for each user review were determined based on their distribution probabilities across all topics generated by the LDA model [53].

Statistical Analysis

To delve deeper into the factors influencing user satisfaction with HMAs, in line with the 2-factor theory [38], we introduced 2 variables: positive deviation (PD) and negative deviation (ND) [54]. These variables are defined by the disparity between the user's individual rating of HMAs and the overall rating displayed in the app store. They serve to indicate the discrepancy between the user's personal rating and the average rating. PD and ND are mutually exclusive if a user's deviation is positive and the ND value is 0, and vice versa. To simplify the calculation, we applied absolute value processing to ND. Consequently, the higher the ND value, the greater the degree of dissatisfaction. Both PD and ND values ranged between 0 and 4, calculated based on the difference between the user rating

range and the comprehensive rating range of the app store. We opted for the Tobit model [55], known for its effectiveness in handling limited dependent variables, to assess the factors influencing user satisfaction. The probability distribution of each LDA topic model for each user review served as the independent variable in our analysis. The PD and ND of each user review were used as dependent variables, and the specific model was defined in Multimedia Appendix 1. Finally, we used the Wald test [54] to perform a difference test on the absolute values of the PD and ND model parameters, aiming to identify the asymmetry of the factors influencing user satisfaction. Given that the Tobit model uses the maximum likelihood method to fit the best parameters, conventional methods cannot be applied for the robustness analysis of the model results. In addition, the exogenous variable in this study, namely, the distribution probability of each topic in user reviews, encompasses all factors affecting user satisfaction, thereby mitigating endogeneity concerns. Model establishment and data analysis were conducted using Stata 16.0 (StataCorp) [56], with a significance level set at a 2-sided $P<.05$ for the difference test.

Results

Chinese and American HMA User Review Topics

We initially retrieved 5016 related apps from both Chinese and American app stores, yielding a total of 295,927 original reviews in both languages. Following screening and data preprocessing, we identified 261 HMAs with user reviews. Among these, 41 (15.7%) were exclusively available in Chinese, 170 (65.1%) were exclusively available in American, and 50 (19.2%) were available in both app stores. Ultimately, 116,686 user reviews were included in the analysis, with 100,125 (85.81%) in English and 16,561 (14.19%) in Chinese. Among these reviews, 87,773 (including 73,907 American reviews and 13,866 Chinese reviews) were rated 4 stars and above, resulting in an overall satisfaction rate of 75.22%: 73,907/100,125 (73.81%) for American apps and 13,866/16,561 (83.73%) for Chinese apps.

The results of LDA modeling revealed differences between Chinese and American reviews. In Chinese reviews, the number of reviews related to software reliability (2165/16,561, 13.07%) and measurement accuracy (2091/16,561, 12.63%) substantially exceeded other topics. Conversely, specific hypertension management functions such as BP tracking (17,285/100,125, 17.26%) and data synchronization (12,837/100,125, 12.82%) substantially outnumbered other topics in American reviews.

The LDA modeling revealed 11 topics in the Chinese reviews, with 5 (ease of use: $P<.001$ in both PD and ND; interface design: $P=.01$ in PD and $P<.001$ in ND; measurement accuracy: $P<.001$ in both PD and ND; compatibility: $P<.001$ in both PD and ND; and cost: $P<.001$ in both PD and ND) of them being significant (including 2 motivational factors and 3 hygienic factors). By contrast, the American reviews yielded 12 topics, all of which were significant (easy to use: $P<.001$ in both PD and ND; blood pressure tracking function: $P<.001$ in both PD and ND; data synchronization: $P<.001$ in both PD and ND; blood pressure management effect: $P<.001$ in both PD and ND; heart rate detection function: $P<.001$ in both PD and ND; data sharing: $P<.001$ in both PD and ND; reliability: $P<.001$ in both PD and ND; compatibility: $P<.001$ in both PD and ND; interface design: $P<.001$ in both PD and ND; advertisement distribution: $P<.001$ in both PD and ND; measurement accuracy: $P<.001$ in both PD and ND; cost: $P<.001$ in both PD and ND), comprising 6 motivational factors and 6 hygienic factors (Tables 1 and 2; Tables S4–S7 in Multimedia Appendix 1).

Table 1. Topics and keywords of Chinese reviews formulated by latent Dirichlet allocation topic modeling (N=16,561).

Topics	Keywords	Reviews, n (%)
Topic 1: Easy to use	simple, convenient, easy to use, practical, rapid, recommended, easy to operate	6863 (41.44)
Topic 2: Reliability	healthy, quality, comprehensive, supportive, stable, professional, normal	2165 (13.07)
Topic 3: Measurement accuracy	measurement, accurate, collect, accuracy, test, data, inaccurate	2091 (12.63)
Topic 4: Attitude (positive)	very good, like, recommend, every day, awesome, useful, look forward to	1850 (11.17)
Topic 5: Compatibility	version, download, try, blood pressure monitor, Apple, connection, platform	878 (5.30)
Topic 6: Cost	fee, subscription, free, payment, upgrade, refund, paid	653 (3.94)
Topic 7: Heart rate detection function	heart rate, detection, value, body, monitor, watch, indicator	597 (3.60)
Topic 8: Blood pressure tracking function	blood pressure, function, record, tool, data, share, form	502 (3.03)
Topic 9: Interface design	updated, special, interface, clear, design, components, good-looking	471 (2.84)
Topic 10: Real-time	trial, view, status, anytime, anywhere, patient, daily	299 (1.81)
Topic 11: Data privacy	account, personal, information, security, management, privacy, licensing	192 (1.16)

Table 2. Topics and keywords of American reviews formulated by latent Dirichlet allocation topic modeling (N=100,125).

Topics	Keywords	Reviews, n (%)
Topic 1: Easy to use	easy, love, simple, handy, utility, recommend, worth	34,443 (34.40)
Topic 2: Blood pressure tracking function	blood, pressure, monitor, track, check, record, measure	17,285 (17.26)
Topic 3: Data synchronization	time, phone, update, synchronization, connect, omron, data	12,837 (12.82)
Topic 4: Blood pressure management effect	good, help, care, body, hypertension, maintain, condition	11,881 (11.87)
Topic 5: Heart rate detection function	heart, rate, pulse, check, measure, test, monitor	7065 (7.06)
Topic 6: Data sharing	data, email, export, send, share, access, require	4465 (4.46)
Topic 7: Reliability	quality, bad, screen, complete, fake, uninstalling, bug	3793 (3.79)
Topic 8: Compatibility	version, iphone, upgrade, fine, android, fail, reinstall	2586 (2.58)
Topic 9: Interface design	user, wonderful, friendly, interface, experience, unit, type	2390 (2.39)
Topic 10: Advertisement distribution	advertisement, download, watch, garbage, click, poor, difficult	1599 (1.60)
Topic 11: Measurement accuracy	accurate, cuff, feel, offer, result, manual, actual	1199 (1.20)
Topic 12: Cost	cost, fee, free, money, afford, pay, count	582 (0.58)

Factors Affecting Chinese and American HMA User Satisfaction

We computed the variance inflation factor of the PD and ND models for both Chinese and American reviews to assess multicollinearity among the independent variables. The regression variance inflation factors of all independent variables in the 4 models were found to be <5, indicating the absence of multicollinearity-related issues [22] (Tables S8-S11 in Multimedia Appendix 1). Furthermore, in Chinese reviews, the topic “attitude (positive)” pertains solely to users’ positive attitudes and does not encompass opinions regarding the functions and utility of apps. Therefore, it was not included in the Tobit regression model.

In Table 3, model 1 displays the PD model results for Chinese reviews. With the exception of reliability (P=.56), real-time (P=.21), and data privacy (P=.52), the other 7 topics (easy to use: P<.001; measurement accuracy: P<.001; compatibility: P<.001; cost: P<.001; heart rate detection function: P=.02; blood pressure tracking function: P<.001; and interface design: P=.01) discussed by users significantly influenced the PD of user satisfaction. Among them, the factors with the most substantial positive and negative effects were the BP tracking function (β=.354, P<.001) and cost (β=–.232, P<.001), respectively. Model 2 presents the ND model results for Chinese reviews. Except for the heart rate detection (P=.64) and BP tracking (P=.14) functions, the other 8 topics (easy to use: P<.001; reliability: P<.001; measurement accuracy: P<.001; compatibility: P<.001; cost: P<.001; interface design: P<.001; real-time: P<.001; and data privacy: P=.001) significantly impacted user satisfaction NDs. The factors with the most substantial positive and negative effects were cost (β=3.703, P<.001) and interface design (β=–1.619, P<.001), respectively. In Table 4, model 3 presents the PD model results for American reviews. All 12 topics (easy to use: P<.001; blood pressure tracking function: P<.001; data synchronization: P<.001; blood pressure management effect: P<.001; heart rate detection function: P<.001; data sharing: P<.001; reliability: P<.001; compatibility: P<.001; interface design: P<.001; advertisement distribution: P<.001; measurement accuracy: P<.001; and cost: P<.001) included in the model significantly influenced the user satisfaction PD. Among them, the factors with the most significant positive and negative effects were the BP tracking function (β=.312, P<.001) and data synchronization (β=–.593, P<.001), respectively. Model 4 displays the ND model results for American reviews. All 12 topics (easy to use: P<.001; blood pressure tracking function: P<.001; data synchronization: P<.001; blood pressure management effect: P<.001; heart rate detection function: P<.001; data sharing: P<.001; reliability: P<.001; compatibility: P<.001; interface design: P<.001; advertisement distribution: P<.001; measurement accuracy: P<.001; and cost: P<.001) significantly influenced the user satisfaction ND. Data synchronization (β=2.662, P<.001) had the greatest positive effect, while the BP management effect (β=–2.035, P<.001) had the most substantial negative effect.

Table 3. Determinant factors for rating deviations (Chinese reviews[a]).

Influencing factor	Model 1[b]			Model 2[c]		
	β (95% CI)	SE	*P* value	β (95% CI)	SE	*P* value
Topic 1: Easy to use	0.080 (0.043 to 0.117)	0.019	<.001	–1.575 (–1.792 to –1.358)	0.111	<.001
Topic 2: Reliability	0.013 (–0.030 to 0.055)	0.022	.56	–0.692 (–0.937 to –0.448)	0.125	<.001
Topic 3: Measurement accuracy	–0.181 (–0.224 to –0.138)	0.022	<.001	2.789 (2.564 to 3.014)	0.115	<.001
Topic 5: Compatibility	–0.219 (–0.275 to –0.163)	0.029	<.001	3.170 (2.906 to 3.434)	0.135	<.001
Topic 6: Cost	–0.232 (–0.293 to –0.171)	0.031	<.001	3.703 (3.424 to 3.981)	0.142	<.001
Topic 7: Heart rate detection function	–0.073 (–0.132 to –0.014)	0.030	.02	0.077 (–0.250 to 0.403)	0.167	.64
Topic 8: Blood pressure tracking function	0.354 (0.298 to 0.411)	0.029	<.001	–0.239 (–0.556 to 0.078)	0.162	.14
Topic 9: Interface design	0.082 (0.019 to 0.144)	0.032	.01	–1.619 (–1.934 to –1.304)	0.161	<.001
Topic 10: Real-time	–0.047 (–0.121 to 0.026)	0.038	.21	1.459 (1.091 to 1.826)	0.188	<.001
Topic 11: Data privacy	–0.028 (–0.113 to 0.057)	0.043	.52	0.761 (0.315 to 1.208)	0.228	.001

[a]In Chinese reviews, the topic labeled "topic attitude (positive)" pertains solely to users' positive attitudes and does not encompass opinions regarding the functions and utility of apps. Therefore, it was not included in the Tobit regression model

[b]Positive deviations: The maximum likelihood estimate of model 1 was –16521.383.

[c]Negative deviations: The maximum likelihood estimate of model 2 was –12843.104.

Table 4. Determinant factors for rating deviations (American reviews).

Influencing factor	Model 3[a]			Model 4[b]		
	β (95% CI)	SE	*P* value	β (95% CI)	SE	*P* value
Topic 1: Easy to use	0.068 (0.058 to 0.078)	0.005	<.001	–0.946 (–0.995 to –0.898)	0.025	<.001
Topic 2: Blood pressure tracking function	0.312 (0.301 to 0.323)	0.006	<.001	–1.139 (–1.195 to –1.083)	0.029	<.001
Topic 3: Data synchronization	–0.593 (–0.606 to –0.581)	0.006	<.001	2.662 (2.612 to 2.711)	0.025	<.001
Topic 4: Blood pressure management effect	0.247 (0.235 to 0.260)	0.006	<.001	–2.035 (–2.100 to –1.970)	0.033	<.001
Topic 5: Heart rate detection function	–0.201 (–0.214 to –0.188)	0.007	<.001	0.807 (0.750 to 0.863)	0.029	<.001
Topic 6: Data sharing	–0.150 (–0.165 to –0.135)	0.008	<.001	0.707 (0.642 to 0.771)	0.033	<.001
Topic 7: Reliability	–0.269 (–0.283 to –0.255)	0.007	<.001	1.356 (1.295 to 1.417)	0.031	<.001
Topic 8: Compatibility	–0.233 (–0.252 to –0.214)	0.010	<.001	1.467 (1.394 to 1.540)	0.037	<.001
Topic 9: Interface design	0.155 (0.138 to 0.171)	0.008	<.001	–1.235 (–1.324 to –1.146)	0.046	<.001
Topic 10: Advertisement distribution	–0.577 (–0.599 to –0.555)	0.011	<.001	2.644 (2.565 to 2.723)	0.040	<.001
Topic 11: Measurement accuracy	0.039 (0.019 to 0.058)	0.010	<.001	–0.597 (–0.694 to –0.499)	0.050	<.001
Topic 12: Cost	0.217 (0.192 to 0.243)	0.013	<.001	–0.664 (–0.920 to –0.408)	0.130	<.001

[a]Positive deviations: The maximum likelihood estimate of model 3 was –81779.181.

[b]Negative deviations: The maximum likelihood estimate of model 4 was –102855.05.

Motivational and Hygienic Factor Asymmetry

The effect asymmetry results are presented in Tables S12 and S13 in Multimedia Appendix 1. In models 1 and 2, with the exception of the heart rate detection and BP tracking functions, the remaining 8 factors exhibited significant differences in asymmetric effect (easy to use: *P*<.001; reliability: *P*<.001; measurement accuracy: *P*<.001; compatibility: *P*<.001; cost: *P*<.001; interface design: *P*<.001; real-time: *P*<.001; and data privacy: *P*=.001). In models 3 and 4, all 12 influencing factors demonstrated significant differences (easy to use: *P*<.001; blood pressure tracking function: *P*<.001; data synchronization: *P*<.001; blood pressure management effect: *P*<.001; heart rate detection function: *P*<.001; data sharing: *P*<.001; reliability: *P*<.001; compatibility: *P*<.001; interface design: *P*<.001; advertisement distribution: *P*<.001; measurement accuracy: *P*<.001; and cost: *P*<.001). Consequently, all factors exhibited a significant asymmetric effect on user satisfaction with Chinese (easy to use: *P*<.001; reliability: *P*<.001; measurement accuracy: *P*<.001; compatibility: *P*<.001; cost: *P*<.001; interface design:

P<.001; real-time: *P*<.001; and data privacy: *P*=.001) and American HMAs (easy to use: *P*<.001; blood pressure tracking function: *P*<.001; data synchronization: *P*<.001; blood pressure management effect: *P*<.001; heart rate detection function: *P*<.001; data sharing: *P*<.001; reliability: *P*<.001; compatibility: *P*<.001; interface design: *P*<.001; advertisement distribution: *P*<.001; measurement accuracy: *P*<.001; and cost: *P*=.001).

Regarding the influencing factors of user satisfaction with Chinese HMAs, both ease of use (*P*<.001 for both) and interface design (*P*=.01 and *P*<.001, respectively) had significant positive or negative effects on the PD or ND models, respectively. In addition, the effects on the 2 models were significantly different (*P*<.001). Similarly, measurement accuracy (*P*<.001 for both), compatibility (*P*<.001 for both), and cost (*P*<.001 for both) had a significant negative or positive impact on the PD or ND model, with a significant difference in the effect of the 2 models (*P*<.001).

Regarding the factors influencing user satisfaction with American HMAs, ease of use (*P*<.001 for both), BP tracking function (*P*<.001 for both), BP management effect (*P*<.001 for both), interface design (*P*<.001 for both), measurement accuracy (*P*<.001 for both), and cost (*P*<.001 for both) all exhibited significant positive or negative effects on the PD or ND model. Moreover, there was a significant difference in the effect on the 2 models (*P*<.001). Data synchronization (*P*<.001 for both), heart rate detection function (*P*<.001 for both), data sharing (*P*<.001 for both), reliability (*P*<.001 for both), compatibility (*P*<.001 for both), and advertisement distribution (*P*<.001 for both) each had a significant negative or positive impact on the PD or ND model. Furthermore, the impact on the 2 models was significantly different (*P*<.001).

Discussion

Principal Findings

To the best of our knowledge, this study represents the first quantitative analysis of user satisfaction, influencing factors, and asymmetry of HMA-related factors based on user reviews. By encompassing a large number of apps and user samples, our research achieved high credibility at a low cost, rendering the findings highly generalizable. Furthermore, leveraging the sociotechnical model widely used in HIT evaluation, we conducted a comparative analysis of Chinese and American HMAs, elucidating differences in user satisfaction and influencing factors between Chinese and American user groups. Consequently, we offer targeted improvement suggestions based on our findings. Although the numbers of Chinese HMAs and reviews were lower than those of American apps, user satisfaction was higher with Chinese HMAs. Furthermore, the main factors influencing user satisfaction and dissatisfaction with Chinese HMAs were the BP tracking function and cost, respectively. Conversely, the main factors affecting user satisfaction and dissatisfaction with American HMAs were the BP tracking function and data synchronization, respectively. Regarding the asymmetry of influencing factors, all factors exhibited significantly different effects on user satisfaction and dissatisfaction. Moreover, there were notable disparities in the motivational and hygienic factors between Chinese and American HMA users.

Differences in the Use of Chinese and American HMAs

Globally, users are distributed unevenly, with considerably lower usage of Chinese HMAs compared with American HMAs. As HMAs represent a typical form of digital health software, their usage status can be effectively analyzed using sociotechnical models [25], which are widely used in HIT. External environmental factors, such as medical policies, diagnosis and treatment methods, and payment methods in different countries, serve as major determinants of HIT. These factors may explain the significant disparities in the use of Chinese and American HMA. Regarding medical policies and treatment methods, Chinese patients tend to rely more on hospital-based doctors, while American patients often prefer active health management guided by family doctors. In China, a 3-level accreditation system for general hospitals has been implemented, and diagnosis and treatment modalities primarily revolve around hospital visits. This has resulted in the accumulation of many patients in a small number of tertiary hospitals with high-quality medical resources for offline diagnosis and treatment [57]. Consequently, personal active health management is rarely implemented. However, hierarchical diagnosis and treatment systems and family doctor consultation models have been primarily implemented in American countries [58]. In the American context, family doctors, who constitute around 80% of doctors, are responsible for 80%-90% of primary diagnosis and treatment services [59]. These services primarily entail disease prevention through active health management. Contrastingly, China relies predominantly on traditional hospitals and drug payments. As a result, Chinese hospitals primarily prioritize in-hospital drug efficacy and often overlook out-of-hospital patient management. In addition, limited software copyright protection in China restricts profits for HMA developers from software downloads. Consequently, the substantial economic costs have impeded the development of HMAs in China. However, American countries primarily implement medical value–based payment policies. Medical insurance payers prioritize patient rehospitalization rates, imposing fines on medical institutions for frequent patient rehospitalizations. This has prompted American medical institutions to use digital health apps to manage patients after discharge and monitor their health status in real-time, aiming to reduce rehospitalization rates. Furthermore, in the United States, HMAs can be prescribed to patients by family doctors or specialists through digital therapy prescriptions. Moreover, the software is granted patent rights, and app developers can cover app development–related costs through software downloads or paid functions, thereby promoting the continuous upgrading and iteration of HMAs.

Satisfaction and Focus of Chinese and American MHA Users

The overall user satisfaction with HMAs was generally poor, with Chinese HMAs exhibiting higher user satisfaction compared with American HMAs. The functions, software design, technical stability, and cost of HMAs were common areas of concern for both Chinese and American users, although

their specific focuses differed. The disparity in user satisfaction between Chinese and American HMAs was consistent with the topic *clustering results of user reviews*, with only Chinese HMA user reviews containing topics reflecting positive user attitudes. Furthermore, the overall user satisfaction obtained in this study, using big data generated by real-world users, was lower than that reported by previous small-sample surveys [60] (87,773/116,686, 75.22%, vs 93.5%). Nevertheless, HMA user engagement and intention to continue use were low [19,20], indicating that our study results are more reliable than those of previous small-sample questionnaire surveys.

Regarding qualitative user descriptions, both Chinese and American users expressed concerns about hypertension management functions, software design, technical stability, and costs. Specifically, most users were particularly concerned about hypertension management functions such as BP tracking and heart rate detection, including BP and heart rate measurements, BP recording, and the visual display of BP change trends. Software design, encompassing usability and interface design, was also a primary concern among users. Usability emerged as the topic of greatest concern among both Chinese (6863/16,561, 41.44%) and American (34,443/100,125, 34.40%) users. Attributes such as simple operating procedures, convenient usage environments, excellent interface design, and clear information display were highlighted as factors contributing to improved user satisfaction. Discussions on technical stability primarily revolved around software reliability, including issues such as software crashes and the inability to open software, as well as system compatibility problems such as software version mismatches and difficulties connecting via Bluetooth. In addition, concerns were raised about measurement accuracy. Finally, users also expressed concerns about software costs, including whether the software is free, its affordability, and the possibility of refunds.

However, there was a significant difference in the focus of HMAs between Chinese and American user groups. Chinese users primarily paid attention to technological stability, such as software reliability and measurement accuracy. By contrast, American users were more concerned about specific hypertension management functions. Moreover, Chinese and American users exhibited different concerns regarding software functions and design. Chinese users showed more interest in the real-time software monitoring function, while American users focused more on data synchronization, data sharing, and BP management effects. Furthermore, Chinese users mentioned software design and personalization, whereas American users were more inclined to allow advertisements that did not disrupt normal app use.

Motivation and Hygiene Factors of Chinese and American HMAs

Easy-to-use features and interface design, including simple software operation and convenience, were common motivational factors for both Chinese and American HMA users. Compatibility issues, such as mismatched software versions and the inability to connect via Bluetooth, were common hygienic factors. However, other motivational and hygienic factors differed considerably between the 2 user groups. Tobit model analysis and the Wald test revealed influencing factors with significantly different effects on user satisfaction PD and ND, indicating asymmetric impacts of related factors on Chinese and American HMA user satisfaction. This suggests the presence of both motivational and hygienic factors in influencing user satisfaction.

Simple operating procedures and convenient HMA usage effectively improve user satisfaction. However, other motivational factors had significantly different effects on user satisfaction. In addition to the motivational factors for Chinese users, American users prioritized factors such as the BP tracking function, BP management effect, measurement accuracy, and cost-effectiveness. The presence of these influencing factors in user reviews increased the degree of user satisfaction, while their absence had the opposite effect. Hence, these factors were identified as the main contributors to user satisfaction, although user dissatisfaction was less associated with them. Therefore, according to the 2-factor theory, the aforementioned influencing factors were considered motivational factors for HMA user satisfaction. Enhancing these factors can lead to increased user satisfaction and, consequently, strengthen users' intention to continue using the HMA.

Addressing issues such as software version mismatches and the inability to connect via Bluetooth can effectively mitigate user dissatisfaction. However, other hygienic factors had varying effects on user satisfaction. In addition to measurement accuracy and cost, hygienic factors for Chinese users encompassed aspects such as data sharing (ie, data uploadable to family doctors). By contrast, hygienic factors for American users included reliability, data synchronization, advertisements that did not affect normal use, and heart rate detection functions. The presence of the aforementioned influencing factors in user reviews increased the degree of user dissatisfaction, and vice versa. Hence, while users feel dissatisfied when these factors do not meet their expectations, their satisfaction is not significantly affected. Therefore, according to the 2-factor theory, the aforementioned influencing factors were considered hygienic factors for HMA user satisfaction. Improving these factors can help mitigate user dissatisfaction, thereby enhancing user participation.

Suggestions for Improving User Engagement and Continued Use of HMA

Given the considerable differences in motivational and hygienic factors for HMA use between Chinese and Americans, software developers should tailor improvements to hygienic factors based on the specific needs of each user group. This approach can effectively reduce user dissatisfaction and increase user participation. Furthermore, efforts should be directed toward enhancing motivational factors to improve user satisfaction and foster continued use intention among both Chinese and American users. Considering that hygiene factors have a direct impact on the utilization of HMA, it is advisable for developers to prioritize improving these factors to enhance user participation. Once hygiene factors are satisfactorily addressed, developers can then focus on enhancing motivational factors to increase users' willingness to continue using the HMA. By improving corresponding motivational and hygienic factors tailored to different user groups, developers can facilitate the

adoption and utilization of HMA, thereby aiding patients in effectively controlling BP. For Chinese HMAs, developers should prioritize enhancing the accuracy of indicator measurements by improving the sensitivity of measurement sensors and optimizing software measurement algorithms. This targeted improvement can effectively mitigate user abandonment of Chinese HMAs. In addition, HMA developers should promptly address technical issues related to software compatibility, such as mismatches between phone systems and software versions, as well as Bluetooth connectivity issues. Moreover, given that Chinese users are more price-sensitive, high software charges have resulted in the loss of a significant number of users. Therefore, developers should consider implementing pricing strategies that align with the preferences and financial capabilities of Chinese users to mitigate user attrition. Developers should consider reducing the number of charging items, shifting away from the traditional model of charging for software downloads, and exploring new business models to optimize profits. For instance, HMAs could be bundled with BP monitors or other relevant health devices to provide added value to users. Furthermore, government intervention is crucial to support the growth of digital medical companies and provide funding for the development of mHealth solutions, especially those focused on chronic disease management. Policy support can encourage entrepreneurship in the digital health sector and foster innovation in HMA development. Finally, HMA developers can enhance user satisfaction by focusing on improving app usability and interface design. This could involve streamlining the operation process, providing clear operation guides, and enhancing the overall user-friendliness of the software. By prioritizing these aspects, developers can create a more intuitive and enjoyable user experience, ultimately increasing user satisfaction and engagement with the HMA.

For American HMAs, developers should promptly address software compatibility issues to enhance HMA usage. This includes resolving mismatches between wearable devices and software, as well as fixing Bluetooth connection failures. In addition, HMA developers should consider incorporating health data sharing and export functions to fulfill the fundamental requirements of American users. Furthermore, developers should optimize HMA reliability and promptly address technical issues, such as software crashes and start-up failures. In addition, providing data synchronization functionality for wearable device monitoring software is essential to ensure real-time BP tracking. Furthermore, developers should consider developing additional heart rate detection functionality to meet users' needs for tracking heart rate indicators. In addition, HMA developers should reconsider the placement and frequency of advertisements to avoid disrupting routine HMA use. Finally, in addition to enhancing the satisfaction of American HMA users, developers can improve the effectiveness of BP management within the software, enhance indicator measurement accuracy, optimize interface design, and establish a reasonable charging model. For instance, they could refine BP measurement algorithms and validate measurement results against those obtained from a BP monitor. They can incorporate additional BP management features such as exercise and diet management, BP warnings, and visual displays of BP trends. Enhancing the interface aesthetics and integrating HMAs into medical insurance schemes are also crucial measures to consider.

Advantages Compared With Previous Research

This study marks a significant advancement by exploring user satisfaction and its influencing factors on HMAs based on real-world user reviews. Unlike previous research that primarily relied on on-site surveys and qualitative analysis, often gathering subjective comments through interviews or questionnaires, this study leverages real user feedback obtained from online reviews. By using advanced computational methods such as natural language processing and topic modeling, it provides a more comprehensive and data-driven analysis of user satisfaction factors. For instance, Breil et al [61] discovered, through questionnaire surveys, that both patients and doctors generally accept HMAs, with expected performance being a crucial determining factor. Similarly, Kang et al [30] developed an HMA and assessed user satisfaction using a scale. However, our study adopts a big data–driven approach to quantitatively explore the influencing factors of user satisfaction in HMAs. Using the LDA model, we mined user perspectives from user-generated content and established a connection between user perspectives and satisfaction through the Tobit model, rendering the results more objective and reliable. For practical applications, previous studies have primarily focused on user groups in specific regions, overlooking the exploration of user satisfaction asymmetry. For instance, Hui et al [62] evaluated the functional availability and user satisfaction of HMAs in the United Kingdom, highlighting the need for further improvement. Meanwhile, Melin et al [63] developed an app user satisfaction evaluation scale and assessed app user satisfaction using linear weighting methods. This study compared user satisfaction and influencing factors of HMAs in China and the United States, applying the 2-factor theory to analyze user satisfaction asymmetry. This provided deeper insights into the attributes of influencing factors and offered more accurate improvement suggestions.

Limitations

This study had several limitations. First, due to variations in medical policies, diagnoses, and treatment models between China and the United States, Chinese individuals use HMAs less frequently, resulting in a proportionately lower number of Chinese reviews and downloads (16,561/24,204,832, 0.07%) compared with American ones (100,125/148,869,181, 0.07%). Therefore, the number of collected Chinese HMA user reviews was significantly lower than that for the American apps. However, for the representativeness of user reviews, we did not conduct special sampling but obtained all user reviews. These user reviews were deemed sufficient to reflect user satisfaction and the influencing factors. Second, considering that some users used the HMA but did not leave reviews, there may have been bias in population selection, and we were unable to explore the satisfaction of such users. However, given that 116,686 reviews were included, and the process of user reviews is random, with users of various opinions possibly not leaving reviews, we believe our results are representative and provide useful data for discovering factors and attributes associated with HMA user satisfaction. Third, through user reviews, this study identified

the factors influencing user satisfaction and explored their asymmetry. However, an in-depth analysis of the impact paths could be further explored based on these influencing factors. Finally, through data screening and aggregation, we observed that the number of Chinese HMA user reviews was substantially lower than that of American HMA user reviews, and the factors affecting user satisfaction in China and the United States were considerably different. Although we provided a preliminary analysis of the reasons for these differences through a sociotechnical model, a more in-depth analysis is required in the future. In addition, as users are more concerned about the therapeutic effect of HMAs, it would be meaningful to pair clinical efficacy data with satisfaction in future research to further explore the relationship between user satisfaction and specific clinical efficacy.

Conclusions

Our study reveals that only 87,773/116,686 (75.22%) users are satisfied with HMA use. We also found that the factors influencing Chinese and American HMA user satisfaction exhibit asymmetry. Furthermore, because of differences in user groups and macro usage environments, the motivational and hygienic factors for users in China and the United States are significantly different. Thus, to enhance user participation, developers of HMAs should devise personalized and comprehensive strategies that address issues related to hygienic factors as a priority. Furthermore, efforts should be made to enhance motivational factors to encourage sustained HMA usage.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (grant 81871455), the Zhejiang Provincial Natural Science Foundation of China (grant LY22H180001), the Municipal Natural Science Foundation of Beijing (grant 7222306), the National TCM Innovation Team and Talent Support Projects (grant ZYYCXTD-C-202210), the Key Research and Development Program of Zhejiang Province (grant 2022C03111), the Hainan Province Science and Technology special fund (grant ZDYF2022SHFZ292), the Hainan Province Clinical Medical Center (grant QWYH2022341), and The 13th Five Year Education Reform Project in Zhejiang Province (grant JG20190551).

Data Availability

We used crawler technology and Qimai app data official website to obtain all hypertension management apps and user reviews in 8 Chinese app stores (China Apple App Store, Huawei App Store, Xiaomi App Store, OPPO App Store, VIVO App Store, Baidu App store, 360 App store, and Application Treasure App Store) and 2 American app stores (US Apple App Store and US Google Play Store). The source code for our analysis is publicly available on GitHub [64].

Authors' Contributions

J Liang conceptualized and designed the study, performed data screening, and made significant revisions to the manuscript. YFH drafted the manuscript; summarized the literature review; and performed the data acquisition, screening, analysis, and interpretation. J Lei supervised the data filtering and analysis process. J Liang and WZ supervised the review method and data interpretation and provided valuable suggestions for improvement. HC, JX, TW, and YCL critically revised the manuscript for important intellectual content.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search terms and results; inclusion and exclusion criteria; annotation and examples of reviews (in both Chinese and English). We also present the results of multicollinearity analysis.
[DOCX File , 199 KB - mhealth_v12i1e55199_app1.docx]

References

1. NCD Risk Factor Collaboration (NCD-RisC). Worldwide trends in hypertension prevalence and progress in treatment and control from 1990 to 2019: a pooled analysis of 1201 population-representative studies with 104 million participants. Lancet 2021 Sep 11;398(10304):957-980 [FREE Full text] [doi: 10.1016/S0140-6736(21)01330-1] [Medline: 34450083]
2. Falah F, Sajadi SA, Pishgooie AH. Effect of a mobile-based educational app on blood pressure of patients with hypertension. BMJ Mil Health 2023 Apr 10;169(2):e001577. [doi: 10.1136/bmjmilitary-2020-001577] [Medline: 33303692]
3. Alnooh G, Alessa T, Hawley M, de Witte L. The use of dietary approaches to stop hypertension (DASH) mobile apps for supporting a healthy diet and controlling hypertension in adults: systematic review. JMIR Cardio 2022 Nov 02;6(2):e35876 [FREE Full text] [doi: 10.2196/35876] [Medline: 36322108]

4. Patterson K, Davey R, Keegan R, Kunstler B, Woodward A, Freene N. Behaviour change techniques in cardiovascular disease smartphone apps to improve physical activity and sedentary behaviour: Systematic review and meta-regression. Int J Behav Nutr Phys Act 2022 Jul 07;19(1):81. [doi: 10.1186/s12966-022-01319-8] [Medline: 35799263]
5. Mertens A, Brandl C, Miron-Shatz T, Schlick C, Neumann T, Kribben A, et al. A mobile application improves therapy-adherence rates in elderly patients undergoing rehabilitation: a crossover design study comparing documentation via iPad with paper-based control. Medicine (Baltimore) 2016 Sep;95(36):e4446 [FREE Full text] [doi: 10.1097/MD.0000000000004446] [Medline: 27603339]
6. Alzahrani S, Bin Muammar MF, Bin Muammar AF, Alolah A, Almutawa M. The adoption and acceptance of mHealth interventions for self-management of hypertension among adult patients: a systematic review. Cureus 2022 Nov;14(11):e31584 [FREE Full text] [doi: 10.7759/cureus.31584] [Medline: 36540533]
7. Kassavou A, Wang M, Mirzaei V, Shpendi S, Hasan R. The association between smartphone app-based self-monitoring of hypertension-related behaviors and reductions in high blood pressure: systematic review and meta-analysis. JMIR Mhealth Uhealth 2022 Jul 12;10(7):e34767 [FREE Full text] [doi: 10.2196/34767] [Medline: 35819830]
8. Green BB, Cook AJ, Ralston JD, Fishman PA, Catz SL, Carlson J, et al. Effectiveness of home blood pressure monitoring, web communication, and pharmacist care on hypertension control: a randomized controlled trial. JAMA 2008 Jun 25;299(24):2857-2867 [FREE Full text] [doi: 10.1001/jama.299.24.2857] [Medline: 18577730]
9. Yuting Z, Xiaodong T, Qun W. Effectiveness of a mHealth intervention on hypertension control in a low-resource rural setting: a randomized clinical trial. Front Public Health 2023 Mar 1;11:1049396 [FREE Full text] [doi: 10.3389/fpubh.2023.1049396] [Medline: 36935728]
10. Bengtsson U, Kjellgren K, Hallberg I, Lindwall M, Taft C. Improved blood pressure control using an interactive mobile phone support system. J Clin Hypertens (Greenwich) 2016 Feb;18(2):101-108 [FREE Full text] [doi: 10.1111/jch.12682] [Medline: 26456490]
11. Razai MS, Trinder B, Perry A, Cauldwell M, Reid F, Oakeshott P. Encouraging brisk walking with the free Active10 app in postnatal women who had a hypertensive pregnancy: "Just Walk It" feasibility study. PLoS One 2023 Feb 21;18(2):e0282066 [FREE Full text] [doi: 10.1371/journal.pone.0282066] [Medline: 36809278]
12. Tian M, Ajay VS, Dunzhu D, Hameed SS, Li X, Liu Z, et al. A cluster-randomized, controlled trial of a simplified multifaceted management program for individuals at high cardiovascular risk (SimCard Trial) in rural Tibet, China, and Haryana, India. Circulation 2015 Sep 01;132(9):815-824 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.115.015373] [Medline: 26187183]
13. Han H, Guo W, Lu Y, Wang M. Effect of mobile applications on blood pressure control and their development in China: a systematic review and meta-analysis. Public Health 2020 Aug;185:356-363. [doi: 10.1016/j.puhe.2020.05.024] [Medline: 32738577]
14. McManus RJ, Little P, Stuart B, Morton K, Raftery J, Kelly J, HOME BP investigators. Home and online management and evaluation of blood pressure (HOME BP) using a digital intervention in poorly controlled hypertension: randomised controlled trial. BMJ 2021 Jan 19;372:m4858 [FREE Full text] [doi: 10.1136/bmj.m4858] [Medline: 33468518]
15. Picone DS, Campbell NRC, Schutte AE, Olsen MH, Ordunez P, Whelton PK, et al. Validation status of blood pressure measuring devices sold globally. JAMA 2022 Feb 15;327(7):680-681 [FREE Full text] [doi: 10.1001/jama.2021.24464] [Medline: 35166811]
16. Sharman J, Ordunez P, Brady T, Parati G, Stergiou G, Whelton P, et al. The urgency to regulate validation of automated blood pressure measuring devices: a policy statement and call to action from the world hypertension league. J Hum Hypertens 2023 Feb;37(2):155-159 [FREE Full text] [doi: 10.1038/s41371-022-00747-0] [Medline: 36476777]
17. Whelton PK, Picone DS, Padwal R, Campbell NRC, Drawz P, Rakotz MK, et al. Global proliferation and clinical consequences of non-validated automated BP devices. J Hum Hypertens 2023 Feb 12;37(2):115-119. [doi: 10.1038/s41371-022-00667-z] [Medline: 35279699]
18. Carey RM, Moran AE, Whelton PK. Treatment of hypertension: a review. JAMA 2022 Nov 08;328(18):1849-1861. [doi: 10.1001/jama.2022.19590] [Medline: 36346411]
19. Márquez Contreras E, Márquez Rivero S, Rodríguez García E, López-García-Ramos L, Carlos Pastoriza Vilas J, Baldonedo Suárez A, Compliance Group of Spanish Society of Hypertension (SEH-LELHA). Specific hypertension smartphone application to improve medication adherence in hypertension: a cluster-randomized trial. Curr Med Res Opin 2019 Jan 05;35(1):167-173. [doi: 10.1080/03007995.2018.1549026] [Medline: 30431384]
20. Robbins R, Krebs P, Jagannathan R, Jean-Louis G, Duncan DT. Health app use among US mobile phone users: analysis of trends by chronic disease status. JMIR Mhealth Uhealth 2017 Dec 19;5(12):e197 [FREE Full text] [doi: 10.2196/mhealth.7832] [Medline: 29258981]
21. Kaium M, Bao Y, Alam M, Hoque M. Understanding continuance usage intention of mHealth in a developing country. IJPHM 2020 Mar 23;14(2):251-272. [doi: 10.1108/ijphm-06-2019-0041]
22. Wang T, Fan L, Zheng X, Wang W, Liang J, An K, et al. The impact of gamification-induced users' feelings on the continued use of mhealth apps: a structural equation model with the self-determination theory approach. J Med Internet Res 2021 Aug 12;23(8):e24546 [FREE Full text] [doi: 10.2196/24546] [Medline: 34387550]

23. Cho J. The impact of post-adoption beliefs on the continued use of health apps. Int J Med Inform 2016 Mar;87:75-83. [doi: 10.1016/j.ijmedinf.2015.12.016] [Medline: 26806714]
24. Wang T, Wang W, Liang J, Nuo M, Wen Q, Wei W, et al. Identifying major impact factors affecting the continuance intention of mHealth: a systematic review and multi-subgroup meta-analysis. NPJ Digit Med 2022 Sep 15;5(1):145 [FREE Full text] [doi: 10.1038/s41746-022-00692-9] [Medline: 36109594]
25. Zhou C, Mou H, Xu W, Li Z, Liu X, Shi L, et al. Study on factors inducing workplace violence in Chinese hospitals based on the broken window theory: a cross-sectional study. BMJ Open 2017 Jul 28;7(7):e016290 [FREE Full text] [doi: 10.1136/bmjopen-2017-016290] [Medline: 28756386]
26. FDASIA health IT report--proposed strategy and recommendations for a risk-based framework. FDA. URL: https://www.fda.gov/about-fda/cdrh-reports/fdasia-health-it-report [accessed 2023-09-23]
27. Brandt LR, Hidalgo L, Diez-Canseco F, Araya R, Mohr DC, Menezes PR, et al. Addressing depression comorbid with diabetes or hypertension in resource-poor settings: a qualitative study about user perception of a nurse-supported smartphone app in Peru. JMIR Ment Health 2019 Jun 18;6(6):e11701 [FREE Full text] [doi: 10.2196/11701] [Medline: 31215511]
28. Guo M, Lyu L. A scale to measure the perceived quality of mHealth by elderly patients with hypertension in China. BMC Health Serv Res 2023 Apr 10;23(1):351 [FREE Full text] [doi: 10.1186/s12913-023-09357-z] [Medline: 37038140]
29. Li B, Zhang X, Tang G. Reasons for hypertension patients' refusal to use mobile health platform: a qualitative study. J Nurs Sci 2019;34:37-40. [doi: 10.3870/j.issn.1001-4152.2019.05.037]
30. Kang H, Park HA. A mobile app for hypertension management based on clinical practice guidelines: development and deployment. JMIR Mhealth Uhealth 2016 Feb 02;4(1):e12 [FREE Full text] [doi: 10.2196/mhealth.4966] [Medline: 26839283]
31. Willems SJ, Coppieters MW, Pronk Y, Diks MJF, van der Heijden KWAP, Rooker S, et al. A clinical journey mobile health app for perioperative patients: cross-sectional study. JMIR Hum Factors 2021 Feb 08;8(1):e20694 [FREE Full text] [doi: 10.2196/20694] [Medline: 33555262]
32. Lyzwinski LN, Caffery LJ, Bambling M, Edirippulige S. Consumer perspectives on mHealth for weight loss: a review of qualitative studies. J Telemed Telecare 2017 Feb 09;24(4):290-302. [doi: 10.1177/1357633x17692722]
33. Varnfield M, Redd C, Stoney RM, Higgins L, Scolari N, Warwick R, et al. M♡THer, an mHealth system to support women with gestational diabetes mellitus: feasibility and acceptability study. Diabetes Technol Ther 2021 May 01;23(5):358-366 [FREE Full text] [doi: 10.1089/dia.2020.0509] [Medline: 33210954]
34. Alnasser AA, Alkhalifa AS, Sathiaseelan A, Marais D. What overweight women want from a weight loss app: a qualitative study on Arabic women. JMIR Mhealth Uhealth 2015 May 20;3(2):e41 [FREE Full text] [doi: 10.2196/mhealth.4409] [Medline: 25993907]
35. Plante T, O'Kelly A, Macfarlane Z, Urrea B, Appel L, Miller Iii ER, et al. Trends in user ratings and reviews of a popular yet inaccurate blood pressure-measuring smartphone app. J Am Med Inform Assoc 2018 Aug 01;25(8):1074-1079 [FREE Full text] [doi: 10.1093/jamia/ocy060] [Medline: 29878236]
36. Wang T, Zheng X, Liang J, An K, He Y, Nuo M, et al. Use of machine learning to mine user-generated content from mobile health apps for weight loss to assess factors correlated with user satisfaction. JAMA Netw Open 2022 May 02;5(5):e2215014 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.15014] [Medline: 35639374]
37. Nuo M, Zheng S, Wen Q, Fang H, Wang T, Liang J, et al. Mining the influencing factors and their asymmetrical effects of mHealth sleep app user satisfaction from real-world user-generated reviews: content analysis and topic modeling. J Med Internet Res 2023 Jan 31;25:e42856 [FREE Full text] [doi: 10.2196/42856] [Medline: 36719730]
38. Herzberg F. Motivation to Work. London, UK: Routledge; 2017.
39. Wiertz C, de Ruyter K, Keen C, Streukens S. Cooperating for service excellence in multichannel service systems. Journal of Business Research 2004 Apr;57(4):424-436. [doi: 10.1016/s0148-2963(02)00276-x]
40. Liang J, He X, Jia Y, Zhu W, Lei J. Chinese mobile health apps for hypertension management: a systematic evaluation of usefulness. J Healthc Eng 2018;2018:7328274 [FREE Full text] [doi: 10.1155/2018/7328274] [Medline: 29744027]
41. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
42. Qimai Data (formerly ASO100). Professional mobile product business analysis platform. Qimai Data (formerly ASO100). URL: https://www.qimai.cn/ [accessed 2023-09-17]
43. Tweetbotornot. GitHub. URL: https://github.com/mkearney/tweetbotornot [accessed 2023-08-17]
44. Tian H, Gao C, Xiao X, Liu H, He B, Wu H, et al. SKEP: Sentiment knowledge enhanced pre-training for sentiment analysis. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: Association for Computational Linguistics; 2020 Presented at: The 58th Annual Meeting of the Association for Computational Linguistics; July 5-10, 2020; Seattle, WA p. 4067-4076 URL: https://aclanthology.org/2020.acl-main.374.pdf [doi: 10.48550/arXiv.2005.05635]
45. Hadji B, Degoulet P. Information system end-user satisfaction and continuance intention: a unified modeling approach. J Biomed Inform 2016 Jun;61:185-193 [FREE Full text] [doi: 10.1016/j.jbi.2016.03.021] [Medline: 27033175]

46. Min K, Song S, Min J. Topic modeling of social networking service data on occupational accidents in Korea: latent Dirichlet allocation analysis. J Med Internet Res 2020 Aug 13;22(8):e19222 [FREE Full text] [doi: 10.2196/19222] [Medline: 32663156]
47. Jelodar H, Wang Y, Yuan C, Feng X, Jiang X, Li Y, et al. Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey. Multimed Tools Appl 2018 Nov 28;78(11):15169-15211. [doi: 10.1007/s11042-018-6894-4]
48. Junyi S. Jieba: Chinese words segmentation utilities. GitHub. URL: https://github.com/fxsjy/jieba33 [accessed 2023-09-03]
49. Natural language toolkit. NLTK Project. URL: https://www.nltk.org./ [accessed 2023-09-03]
50. The most comprehensive NLP Chinese and English stop list for the entire site. CSDN. URL: https://blog.csdn.net/qq_52181283/article/details/125013666 [accessed 2023-12-18]
51. Huang L, Ma J, Chen C. Topic detection from microblogs using T-LDA and perplexity. New York, NY: IEEE; 2017 Presented at: 24th Asia-Pacific Software Engineering Conference Workshops (APSECW); December 4-8, 2017; Nanjing, China p. 71-77. [doi: 10.1109/APSECW.2017.11]
52. Rehurek R, Sojka P. Gensim Python framework for vector space modelling. NLP Centre, Faculty of Informatics, Masaryk University. 2011. URL: https://code.activestate.com/pypm/gensim/ [accessed 2023-12-10]
53. Yuan F, Li M, Liu R, Zhai W, Qi B. Social media for enhanced understanding of disaster resilience during Hurricane Florence. International Journal of Information Management 2021 Apr;57:102289. [doi: 10.1016/j.ijinfomgt.2020.102289]
54. Park S, Lee J, Nicolau JL. Understanding the dynamics of the quality of airline service attributes: satisfiers and dissatisfiers. Tour Manag 2020 Dec;81:104163 [FREE Full text] [doi: 10.1016/j.tourman.2020.104163] [Medline: 32518436]
55. Wang X, Parameswaran S, Bagul D, Kishore R. Can online social support be detrimental in stigmatized chronic diseases? A quadratic model of the effects of informational and emotional support on self-care behavior of HIV patients. J Am Med Inform Assoc 2018 Aug 01;25(8):931-944. [doi: 10.1093/jamia/ocy012] [Medline: 29684190]
56. Stata is statistical software for data science. Stata. URL: https://www.stata.com [accessed 2023-09-03]
57. Liang J, He Y, Fan L, Nuo M, Shen D, Xu J, et al. A preliminary study on the abnormal deaths and work burden of Chinese physicians: a mixed method analysis and implications for smart hospital management. Front Public Health 2021;9:803089 [FREE Full text] [doi: 10.3389/fpubh.2021.803089] [Medline: 35059382]
58. Oxtoby K. A true family doctor: the general practitioner. BMJ 2022 May 03;377:o1054. [doi: 10.1136/bmj.o1054] [Medline: 35504652]
59. Mathew R. The difference between a doctor and a family doctor. BMJ 2022 Feb 01;376:o240. [doi: 10.1136/bmj.o240] [Medline: 35105555]
60. Yao L, Shang D, Zhao H, Liu X, Liu Y, Jiang Y. The effect and satisfaction of mobile network in the hypertension management of community dwelling older adults. Chinese General Practice 2024;27(1):85-90. [doi: 10.12114/j.issn.1007-9572.2023.0379]
61. Breil B, Salewski C, Apolinário-Hagen J. Comparing the acceptance of mobile hypertension apps for disease management among patients versus clinical use among physicians: cross-sectional survey. JMIR Cardio 2022 Jan 06;6(1):e31617 [FREE Full text] [doi: 10.2196/31617] [Medline: 34989683]
62. Hui C, Creamer E, Pinnock H, McKinstry B. Apps to support self-management for people with hypertension: content analysis. JMIR Mhealth Uhealth 2019 Jun 03;7(6):e13257 [FREE Full text] [doi: 10.2196/13257] [Medline: 31162124]
63. Melin J, Bonn SE, Pendrill L, Trolle Lagerros Y. A questionnaire for assessing user satisfaction with mobile health apps: development using Rasch measurement theory. JMIR Mhealth Uhealth 2020 May 26;8(5):e15909 [FREE Full text] [doi: 10.2196/15909] [Medline: 32452817]
64. User satisfaction of high blood pressure management apps. GitHub. URL: https://github.com/YunfanHe123/User-satisfaction-of-high-blood-pressure-management-Apps.git [accessed 2024-03-23]

Abbreviations

BP: blood pressure
FDA: Food and Drug Administration
HIT: health information technology
HMA: hypertension management app
LDA: latent Dirichlet allocation
mHealth: mobile health
ND: negative deviation
PD: positive deviation
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by L Buis; submitted 06.12.23; peer-reviewed by S Cao, Z Guo; comments to author 17.12.23; revised version received 19.12.23; accepted 14.03.24; published 28.03.24.

Please cite as:
He Y, Zhu W, Wang T, Chen H, Xin J, Liu Y, Lei J, Liang J
Mining User Reviews From Hypertension Management Mobile Health Apps to Explore Factors Influencing User Satisfaction and Their Asymmetry: Comparative Study
JMIR Mhealth Uhealth 2024;12:e55199
URL: https://mhealth.jmir.org/2024/1/e55199
doi:10.2196/55199
PMID:38547475

Original Paper

Effectiveness of an Interactive mHealth App (EVITE) in Improving Lifestyle After a Coronary Event: Randomized Controlled Trial

María Ángeles Bernal-Jiménez[1,2,3], PhD; German Calle[2,4], MD; Alejandro Gutiérrez Barrios[2,4], PhD; Livia Luciana Gheorghe[2,4], PhD; Celia Cruz-Cobo[1,2,3], RNC; Nuria Trujillo-Garrido[1,2,3], PhD; Amelia Rodríguez-Martín[1,5], PhD; Josep A Tur[6,7], PhD; Rafael Vázquez-García[2,4], PhD; María José Santi-Cano[1,2,3], PhD

[1]Faculty of Nursing and Physiotherapy, University of Cádiz, Cádiz, Spain

[2]Institute of Biomedical Research and Innovation of Cádiz, Cádiz, Spain

[3]Research Group on Nutrition: Molecular, Pathophysiological and Social Issues, University of Cádiz, Cádiz, Spain

[4]Cardiology Unit, Puerta del Mar Hospital, Cádiz, Spain

[5]Biomedicine, Biotechnology and Public Health Department, University of Cádiz, Cádiz, Spain

[6]Research Group on Community Nutrition & Oxidative Stress, University of the Balearic Islands, Palma de Mallorca, Spain

[7]Network Biomedical Research Center "Pathophysiology of Obesity and Nutrition", Carlos III Health Institute, Madrid, Spain

Corresponding Author:
María José Santi-Cano, PhD
Faculty of Nursing and Physiotherapy, University of Cádiz
Avda. Ana de Viya, 52
Cádiz, 11009
Spain
Phone: 34 956019042
Fax: 34 956015685
Email: mariajose.santi@uca.es

Abstract

Background: Coronary heart disease is one of the leading causes of mortality worldwide. Secondary prevention is essential, as it reduces the risk of further coronary events. Mobile health (mHealth) technology could become a useful tool to improve lifestyles.

Objective: This study aimed to evaluate the effect of an mHealth intervention on people with coronary heart disease who received percutaneous coronary intervention. Improvements in lifestyle regarding diet, physical activity, and smoking; level of knowledge of a healthy lifestyle and the control of cardiovascular risk factors (CVRFs); and therapeutic adherence and quality of life were analyzed.

Methods: This was a randomized controlled trial with a parallel group design assigned 1:1 to either an intervention involving a smartphone app (mHealth group) or to standard health care (control group). The app was used for setting aims, the self-monitoring of lifestyle and CVRFs using measurements and records, educating people with access to information on their screens about healthy lifestyles and adhering to treatment, and giving motivation through feedback about achievements and aspects to improve. Both groups were assessed after 9 months. The primary outcome variables were adherence to the Mediterranean diet, frequency of food consumed, patient-reported physical activity, smoking, knowledge of healthy lifestyles and the control of CVRFs, adherence to treatment, quality of life, well-being, and satisfaction.

Results: The study analyzed 128 patients, 67 in the mHealth group and 61 in the control group; most were male (92/128, 71.9%), with a mean age of 59.49 (SD 8.97) years. Significant improvements were observed in the mHealth group compared with the control group regarding adherence to the Mediterranean diet (mean 11.83, SD 1.74 points vs mean 10.14, SD 2.02 points; $P<.001$), frequency of food consumption, patient-reported physical activity (mean 619.14, SD 318.21 min/week vs mean 471.70, SD 261.43 min/week; $P=.007$), giving up smoking (25/67, 75% vs 11/61, 42%; $P=.01$), level of knowledge of healthy lifestyles and the control of CVRFs (mean 118.70, SD 2.65 points vs mean 111.25, SD 9.05 points; $P<.001$), and the physical component of the quality of life 12-item Short Form survey (SF-12; mean 45.80, SD 10.79 points vs mean 41.40, SD 10.78 points; $P=.02$). Overall satisfaction was higher in the mHealth group (mean 48.22, SD 3.89 vs mean 46.00, SD 4.82 points; $P=.002$) and app satisfaction and usability were high (mean 44.38, SD 6.18 out of 50 points and mean 95.22, SD 7.37 out of 100).

Conclusions: The EVITE app was effective in improving the lifestyle of patients in terms of adherence to the Mediterranean diet, frequency of healthy food consumption, physical activity, giving up smoking, knowledge of healthy lifestyles and controlling CVRFs, quality of life, and overall satisfaction. The app satisfaction and usability were excellent.

Trial Registration: Clinicaltrials.gov NCT04118504; https://clinicaltrials.gov/study/NCT04118504

(JMIR Mhealth Uhealth 2024;12:e48756) doi:10.2196/48756

KEYWORDS

coronary artery disease; healthy lifestyle; mHealth; mobile health; percutaneous coronary intervention; randomized controlled trial; secondary prevention; therapeutic adherence

Introduction

Ischemic heart disease, or coronary artery disease (CAD), remains one of the leading causes of mortality worldwide, accounting for 16% of all deaths [1,2]. Coronary revascularization using percutaneous coronary intervention (PCI) is the main treatment option in patients with CAD as it restores coronary blood flow and has excellent clinical outcomes [3]. This procedure, however, does not prevent either the formation of further obstructions in the coronary arteries treated or the disease from progressing with the appearance of new lesions in other parts of the coronary tree. Therefore, secondary prevention strategies aimed at achieving a healthy lifestyle and the control of cardiovascular risk factors (CVRFs) are essential [4].

Secondary prevention and cardiac rehabilitation programs can improve lifestyle, therapeutic adherence, and the control of CVRFs. This facilitates a decrease in the risk of a further cardiovascular event that requires another PCI or even surgery, leading to a reduction in health care costs [5,6]. However, a large number of patients do not make these recommendations a part of their everyday lifestyle or do not take their medication correctly after experiencing a cardiovascular event, which stops them from achieving the treatment goals established [4,7]. The data from the EUROASPIRE IV survey conducted in Europe demonstrated that only 51% of patients who had experienced a coronary event were advised to participate in a cardiac rehabilitation program, and of those, only 41% took part [7]. This suggests that the implementation of attendance-based secondary prevention programs after acute coronary syndrome has stagnated somewhat.

Moreover, during the COVID-19 pandemic, secondary prevention and cardiac rehabilitation programs were affected and even suspended, thereby seriously limiting patient access to available health services [8,9]. Thus, it is necessary to implement innovative secondary prevention strategies that improve adherence to a healthy lifestyle and the long-term treatment prescribed.

The use of information and communication technology may facilitate new communication pathways between patients and health professionals and allow patients more autonomy in the control of their disease while providing professionals with more information about their patients' health status. These days, over 57% of homes have access to the internet, and over 75% of the world's population has a mobile phone with internet access [10,11]. These data highlight that these technologies and smartphones should form part of the communication between patients and health care professionals, as they can be valuable tools for controlling health and lifestyle. The beneficial effects of mobile health (mHealth) systems have been observed in primary prevention patients at high risk of cardiovascular disease [12-15]. A recent systematic review with 2250 participants that analyzed the results obtained after implementing cardiac telerehabilitation programs concluded that these strategies are a way to improve the use and reach of cardiac rehabilitation programs [16]. Likewise, recent meta-analyses of clinical trials performed in patients with CAD concluded that secondary prevention programs involving telehealth interventions may be useful either in isolation or combined with on-site cardiac rehabilitation and that they are related to a decrease in adverse events, lifestyle improvements, and better control of CVRFs [17,18]. These studies reflect the need for further similar research to evaluate the effectiveness of telehealth interventions in this context.

This clinical trial aimed to evaluate the effect of an mHealth intervention on people with heart disease who had undergone PCI in terms of the following variables: improvements in lifestyle, namely diet, physical activity (PA), and smoking; level of knowledge of a healthy lifestyle and the control of CVRFs; and therapeutic adherence and quality of life.

Methods

Study Design

A statistician-blinded randomized controlled clinical trial with a parallel group design was conducted in patients with CAD who underwent a PCI with stent implantation in the cardiology unit of a public specialty reference hospital in the province of Cádiz, Spain, between November 2019 and June 2022.

Study Sample

A total of 134 patients participated, 67 in the intervention group and 67 in the control group. The initial sample size proposed was 240 participants, but the increase in hospital occupancy due to the COVID-19 pandemic reduced the number of patients scheduled for coronary intervention with a nonurgent pathology. However, this sample size is sufficient to detect an effect size (Cohen *d*) of 0.5 on adherence to the Mediterranean diet (mean 8.7, SD 2 points) [19], time spent doing PA (mean 416.8, SD 374 min/week) [20], level of knowledge of CVRFs (mean 104.7, SD 9.7) [21], and a difference in the prevalence of stopping smoking of over 20% [22], with a 95% CI and statistical power of 90%.

Eligibility Criteria

The study included patients with CAD aged between 18 and 75 years who owned a smartphone with an internet connection during the whole study period.

Those excluded from the study were people with severe heart failure, congenital, structural, or serious rheumatic heart disease, chronic kidney or liver disease, physical disability or dementia, and those who were already using a health control app or whose mobile phone was incompatible with the app.

Recruitment, Randomization, and Masking

Overview

The patients meeting the inclusion criteria were randomized and assigned to either the mHealth group or the control group receiving standard health care using a computerized random number generator 1:1. The randomization sequence was generated by staff at the study's coordination center at the university who were not involved in recruiting the participants. The coordinators did not know the randomization group. The researchers analyzing the results were blinded to the allocation of participants. Blinding the participants was not possible because of the nature of the intervention.

Eligible patients and their caregivers were approached during their hospital admission, after the PCI, and before discharge. All the patients were informed of the characteristics of the project and were invited to volunteer to participate and sign the informed consent form. The patients who agreed to participate underwent an initial assessment and were given a follow-up appointment 9 months later. The patients were given an email address and telephone number through which they could resolve any queries or doubts during the study period. Before their hospital discharge, all the patients were encouraged to lead healthy lifestyles. Moreover, they were given written information about CVRFs, lifestyle objectives, examples of healthy diets, and recommendations about the daily portions of each food group.

Control Group

The patients in the standard health care group received advice about medication and lifestyle, the Mediterranean diet, PA, stopping smoking, and therapeutic adherence. They were also provided with written recommendations.

mHealth Group

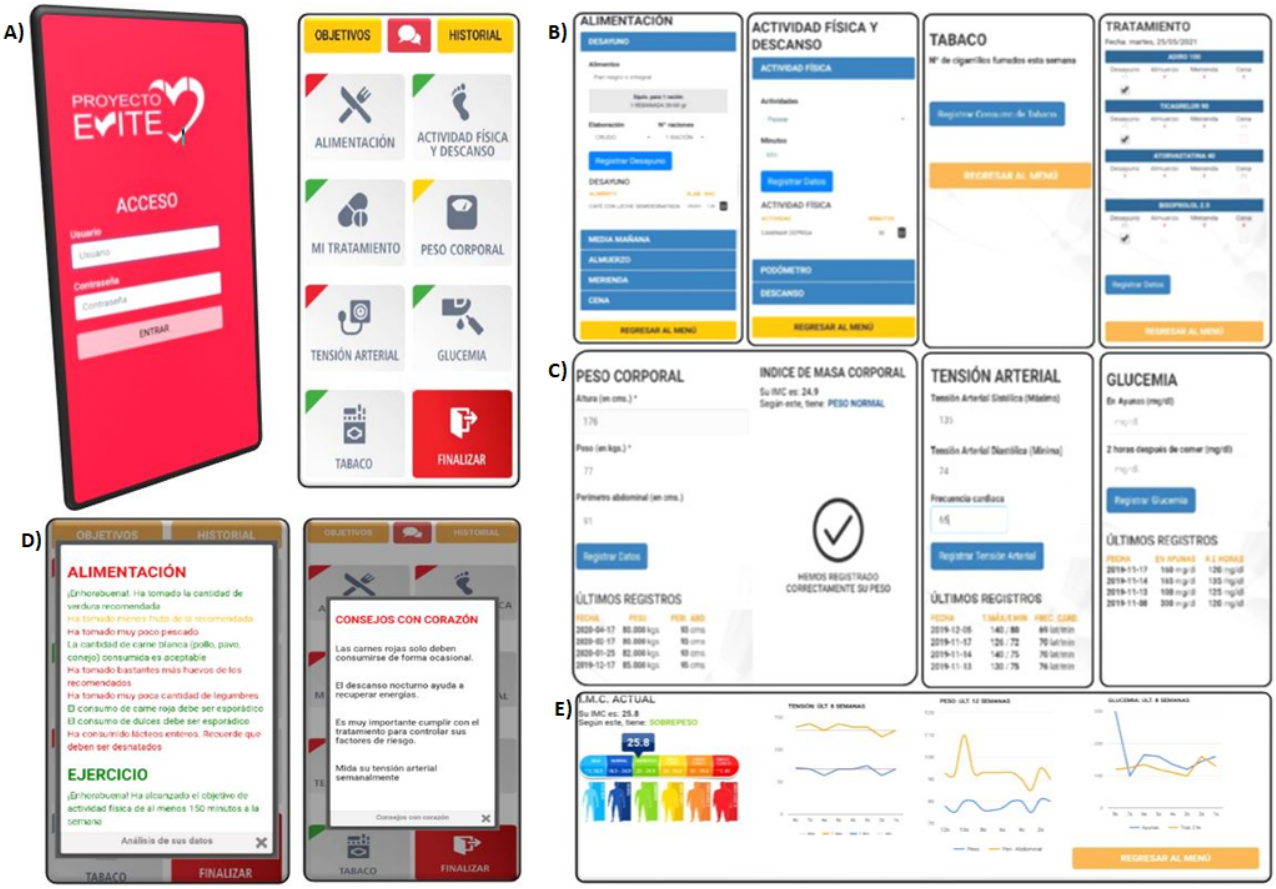
The mHealth intervention began during their hospital stay, after the coronary event. All the patients in this group had an app installed on their mobile phone or tablet and completed a brief online tutorial to learn how to use it. The participants were advised to use the app for a minimum of 15 minutes daily during the 36-week follow-up period. The app established aims to achieve self-control of PA, food consumption, blood pressure, smoking, and therapeutic adherence. The app is based on the phases of change theory (attention, retention, memory, action, and motivation) [23] and on making the process pleasing [24,25]. The user's attention was caught through warnings and bright, attractive colors on the user interface; retention was encouraged by reminders, repetition, and graphs; instructions, advice, and feedback prompted action; and motivation to change was boosted by internal comparisons (progress graphs), setting goals, self-monitoring, and feedback.

A nurse resolved any queries the participants had through the messaging function integrated into the app. The technical data about the website, app, and its contents and components were described previously in greater detail [26]. Briefly, the app provides (1) education about following a healthy lifestyle and about the recommended therapeutic goals in the clinical practice guidelines regarding food, PA, body weight, blood pressure, blood sugar, stopping smoking, and adhering to treatment. The app provided the participants with information on their screens to help them plan a healthier lifestyle and adhere to their treatment; (2) self-monitoring, using records, and a self-checking function to allow patients to monitor each behavioral goal about nutrition, PA, tobacco consumption, blood pressure, body weight, capillary blood glucose in patients with diabetes mellitus and treatment adherence; and (3) motivation to improve and maintain lifestyle habits through automatic reminders about healthy habits generated randomly in a pop-up screen and through personalized messages about reaching goals related with improving their lifestyle and treatment adherence, and recommendations about aspects to be improved (Figure 1).

The personalized messages are produced in response to the information recorded by the patient over the previous 7 days. The patients received weekly feedback through pop-ups that appeared when opening the app. The messages appeared as short sentences in green (when the set goals had been reached), yellow (if they had partially been reached), and red (when the goal was pending), depending on the degree of compliance and control of the goals. As a reminder of achievements and aspects to improve, the top left corner of each section of the app's main screen appeared in the color corresponding to the goals reached during the previous week. The participants could follow their evolution and progress through the graphics generated from the information they had recorded over the previous 8 weeks (Figure 1).

Through its different components (website, messages, emails, and calls), participants were encouraged to (1) follow a healthy eating pattern based on the Mediterranean diet; (2) perform PA of duration and intensity in line with the recommendations of their cardiologist; (3) stop smoking; (4) monitor their blood pressure; and (5) adhere to their treatment by associating taking medication with daily activities, establishing set times for taking it, and with support from a relative, etc. We hypothesize that information is the first step to following a healthy lifestyle. The self-monitoring and recording in the app improves the patients' awareness of their lifestyle behavior, and motivation promotes the initiation and continuation of changes in behavior over time.

Figure 1. (A) Start page app of the EVITE project. (B) Diet, physical activity, smoking cessation, and treatment in the top row (from left to right). (C) Body weight, BMI, blood pressure, and capillary blood glucose modules in the middle row (from left to right). (D) Weekly feedback traffic light and messages, automatic text message reminders. (E) Record: graphics of evolution and progress of BMI, blood pressure, body weight and waist circumference (WC), and capillary blood glucose in the bottom row (from left to right).



Follow-Up

All the participants in both the mHealth group and the control group were evaluated after 36 weeks. Their digital medical records were reviewed to check and record clinical data of medical visits, recent analyses (the same data from the baseline plasma analysis), and adverse events that occurred over that period. Due to the COVID-19 pandemic, the questionnaires were administered by telephone.

Outcome Variables

Overview

The primary outcome variables were adherence to the Mediterranean diet, frequency of consumption of foods, PA performed, smoking (giving up and nicotine dependence test), level of knowledge of a healthy lifestyle, and the control of CVRFs, therapeutic adherence, quality of life, well-being, and satisfaction.

The secondary outcome variables were BMI, WC, systolic blood pressure (SBP), diastolic blood pressure (DBP), heart rate, glycated hemoglobin (HbA_{1c}), total cholesterol, high-density lipoprotein (HDL) cholesterol, low-density lipoprotein (LDL) cholesterol, triglycerides, emotional status, major adverse cardiovascular events (MACEs), and other complications.

While the patients were hospitalized and before their discharge, measurements were taken of their weight and height to calculate their BMI, WC, SBP, and DBP following standard clinical methods. Moreover, records were made of sociodemographic variables (age, sex, and educational level), history of CVRFs (diabetes mellitus, arterial hypertension, hyperlipidemia, obesity, and tobacco consumption), and history of cardiovascular disease (infarction, angina, and stroke). Furthermore, the patient's digital medical records were used to obtain data about left ventricular ejection fraction [27] and plasma analysis (total cholesterol, LDL cholesterol, HDL cholesterol, baseline glycemia, HbA_{1c}, creatinine, and C-reactive protein).

Primary Outcomes

Smoking

Nicotine dependence was assessed using the Fagerström test [28], with scores below 4 points considered low dependence, scores between 4 and 7 considered moderate dependence, and scores from 8 to 10 considered points high dependence. Stopping smoking was self-reported.

Adherence to the Mediterranean Diet and Frequency of Food Consumption

To assess food intake, the following surveys were used: (1) the Mediterranean Adherence Score (maximum score of 14 points; scores below 9 are classified as low adherence, and scores above

9 are classified as high adherence) [19,29]; and (2) the food frequency questionnaire [30].

Physical Activity

Moreover, the time spent performing PA was assessed using the Minnesota Physical Activity Survey (min/week) [31,32], which includes detailed instructions and a list of defined activities. Participants are asked to mark those performed during the past year, month, and week (the activities are expressed in the number of days per week and minutes per day). Total energy expenditure from leisure time PA can be calculated with this questionnaire depending on the type performed: light (≤4 metabolic equivalents [METs]), moderate (4.5-5.5 METs), or heavy (≥6 METs). For this research, only the information referring to the past week was used as the min/week of light PA. The reason was that the participants only performed light PA after PCI. Furthermore, this allowed the comparison of results with other studies that expressed PA in min/week and with the recommendations of the cardiovascular disease prevention guidelines [4]. The PA in the household was omitted. Therapeutic adherence was assessed using the 4-item Morisky Green and Levine Medication Adherence Questionnaire (MGL MAQ). This was considered good if the 4 questions were answered correctly (NO/YES/NO/NO) [33,34].

Level of Knowledge of CVRFs and a Healthy Lifestyle

The level of knowledge of CVRFs and a healthy lifestyle was examined using a validated scale [21]. The scale consists of 24 items with 5 response options and a maximum score of 120 points. Each item is scored from 1 to 5 points, with the highest score corresponding to the most correct answer. The patient's level of knowledge was considered high when they responded correctly to over 75% of the items (90 points).

Quality of Life and Well-Being

The physical and mental dimensions of quality of life were assessed by the SF-12 survey [35], in which the highest scores imply better health-related quality of life. The possible answers for each item are presented as a Likert-type scale, with the number of options varying from 3 to 6 points, depending on the item. To calculate each dimension, the items were recoded, aggregated, and transformed into a score ranging from 0 (worse health status) to 100 points (better health status) [36,37]. Well-being was assessed using the World Health Organization-5 Well-Being Index questionnaire, on which the lowest score is 0 points and the highest 100, a higher score indicating greater well-being [38].

Patient Overall Satisfaction

Overall satisfaction with the health care received was assessed using a specific questionnaire developed by the research team. The questionnaire consists of 10 questions with 5 response options scored from 1 to 5, from the lowest to the highest level of satisfaction. The maximum score is 50 points, with higher scores corresponding to more satisfaction.

Secondary Outcomes

BMI, WC, SBP, DBP, Heart Rate, HbA1c, and Lipids

BMI, WC, SBP, DBP, heart rate, HbA_{1c}, total cholesterol, HDL cholesterol, LDL cholesterol, and triglycerides were obtained following the standard methods mentioned above.

Anxiety and Depression

Levels of anxiety and depression were analyzed using the questionnaire developed by Goldberg et al [39]. This questionnaire consists of 2 subscales; a score over 4 on the anxiety subscale is considered a diagnosis of anxiety, and a score over 2 on the depression subscale is considered a diagnosis of this disorder. The maximum score on each subscale is 9 points, with high scores reflecting a more serious problem [40].

MACEs and Other Complications

At the end of the follow-up period, each patient's digital medical records were analyzed for MACEs (cardiovascular death, acute myocardial infarction, stent thrombosis, new revascularization of target injury, stroke, and other possible complications that can occur, such as bleeding, angina, and the revascularization of another vessel) that had occurred.

Patient App Satisfaction and Usability

Satisfaction with the app itself was assessed using a specific questionnaire developed by the research team. The questionnaire consists of 10 questions with 5 response options scored from 1 to 5, from the lowest to the highest level of satisfaction. The maximum score is 50 points, with higher scores corresponding to more satisfaction.

The app's usability was assessed in the intervention group using the System Usability Scale (SUS) questionnaire, which evaluates user acceptance, with scores ranging from 0 to 100 points [41]. A higher score is an indication of better usability. The scores are classified as excellent for >80.3, good for 68-80.3, poor for 51-67, and very poor for <51.

Statistical Analysis

A descriptive statistical analysis was performed (mean, median, SD, 95% CI, interquartile interval, frequencies, and percentages). The primary quantitative results from the 2 groups—mHealth and control—were compared using the Student 2 tailed *t* test (variables with normal distribution) and the Mann-Whitney *U* test (variables not distributed normally). The χ^2 or Fisher test was used for the comparison of proportions. A 2-tailed *P* value of <.05 was considered statistically significant. The data analysis was performed using SPSS (version 24.0; IBM Corp) for Windows.

Ethical Considerations

The Biomedical Research Ethics Committee of Andalusia approved the study (003_ene19_PI-EVITE-18). The patients signed a written informed consent form. The app included data encryption mechanisms and guaranteed safety measures in accordance with current European Data Protection Regulations. This study was registered on ClinicalTrials.gov (NCT04118504).

Results

Overview

Figure 2 shows a flowchart of the process for recruiting the participants to the clinical trial. During the recruitment period, 1327 patients underwent a PCI and were evaluated for inclusion in the study. After applying the inclusion and exclusion criteria, 1153 were excluded, and 40 refused to participate. The 134 remaining patients were randomized: 67 to the control group and 67 to the mHealth group. There were 6 dropouts in the control group during the follow-up. No participants dropped out of the intervention group. In the end, 128 were analyzed: 61 from the control group and 67 from the mHealth group.

Table 1 shows the baseline characteristics of the participants. Most were male (92/128, 71.9%) with a mean age of 59.49 (SD 8.97) years. In general, both groups were homogeneous.

Figure 2. Flow diagram of the study.

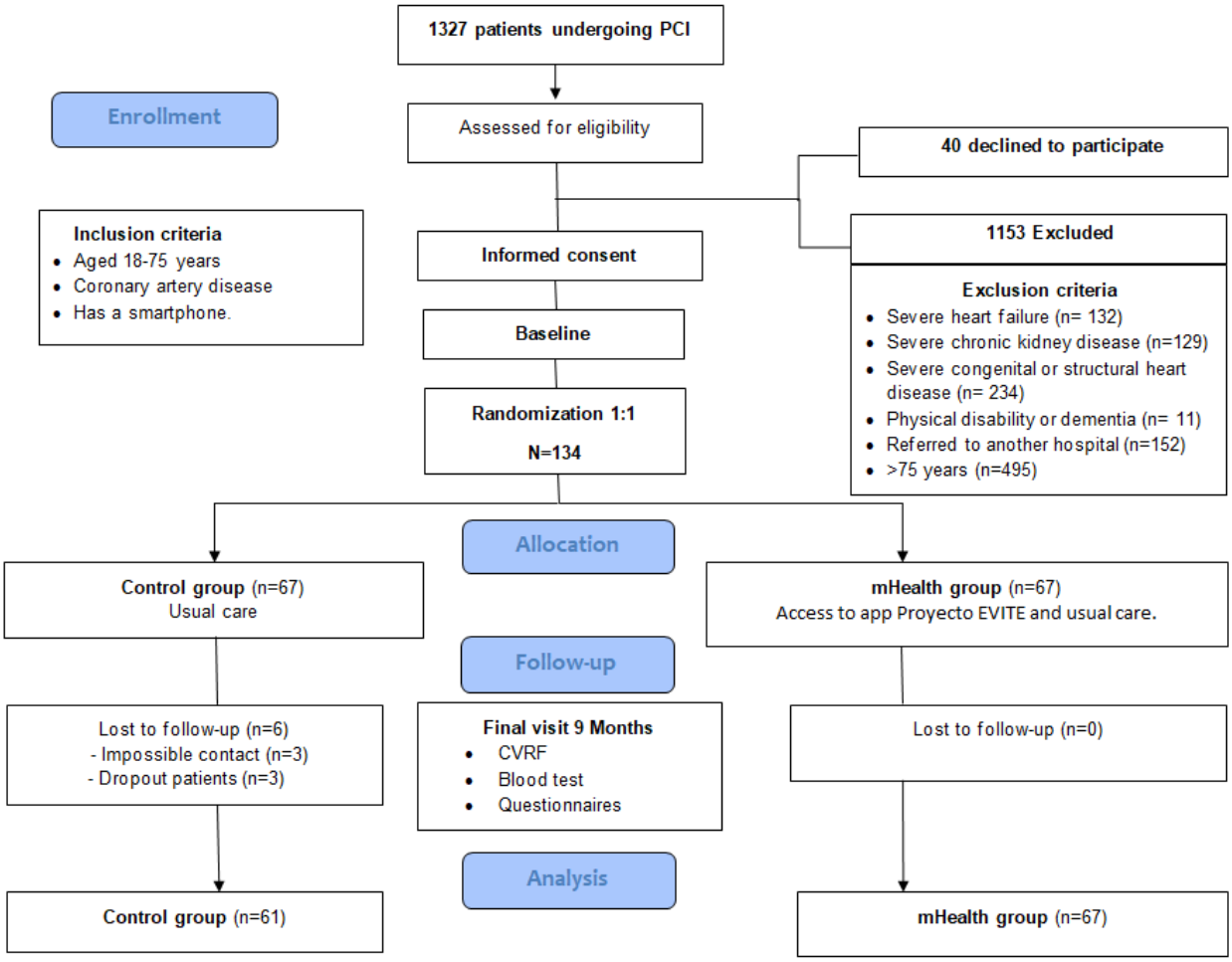


Table 1. Patient baseline characteristics.

	Total (n=128)	mHealth (n=67)	Control (n=61)	*P* value
Male, n (%)	92 (71.9)	53 (79)	39 (64)	.05
Age (years)				.01
Mean (SD)	59.49 (8.97)	57.70 (8.167)	61.46 (9.476)	
95% CI	57.92-61.06	55.71-59.69	59.03-63.89	
Educational level, n (%)				.17
No studies	1 (1.3)	0 (0)	1 (2)	
Primary	21 (26.9)	6 (17)	15 (36)	
Middle school	37 (47.7)	19 (53)	18 (43)	
High school	19 (24.4)	11 (31)	8 (19)	
BMI (kg/m^2)				.55
Mean (SD)	29.34 (4.99)	29.09 (5.02)	29.61 (4.99)	
95% CI	28.47-30.21	27.86-30.32	28.34-30.89	
Waist circumference (cm)				.52
Mean (SD)	104.15 (13.78)	104.91 (10.75)	103.31 (16.55)	
95% CI	101.69-106.60	102.24-107.57	98.99-107.62	
Overweight, n (%)	50 (39.1)	28 (42)	22 (36)	.29
Obesity, n (%)	54 (42.2)	25 (37)	29 (48)	.29
Hypertension, n (%)	74 (57.8)	38 (57)	36 (59)	.79
Diabetes mellitus, n (%)	39 (30.5)	13 (19)	26 (43)	.004
Dyslipidemia, n (%)	66 (51.6)	31 (46)	35 (57)	.20
Smoking, n (%)	59 (46.1)	33 (49)	26 (43)	.45
Former smoker, n (%)	35 (27.3)	21 (31)	14 (23)	.28
Morbidities[a], n (%)	8 (6.3)	3 (5)	5 (8)	.42
LVEF[b] (%)				.78
Mean (SD)	58.27 (9.08)	58.04 (8.71)	58.56 (9.62)	
95% CI	56.40-60.14	55.61-60.46	55.52-61.60	
Reason for catheterization, n (%)				.04
Stable angina	32 (25.0)	11 (16)	21 (34)	
Unstable angina	19 (14.8)	8 (12)	11 (18)	
NSTEMI[c]	29 (22.7)	19 (28)	10 (16)	
STEMI[d]	48 (37.5)	29 (43)	19 (31)	
Number of implanted stents				.63
Mean (SD)	2.34 (1.65)	2.40 (1.89)	2.26 (1.36)	
95% CI	2.05-2.6	1.94-2.86	1.91-2.61	
Discharge treatment, n (%)				
Anticoagulants	6 (4.7)	3 (5)	3 (5)	.93
Antiplatelet	126 (98.4)	66 (100)	60 (97)	.14
Antihypertensives	124 (96.9)	65 (99)	59 (95)	.28
Nitrates	14 (10.9)	4 (6)	10 (16)	.06
Insulin	7 (5.5)	3 (5)	4 (7)	.63
Oral antidiabetics	41 (32.0)	16 (24)	25 (40)	.05

	Total (n=128)	mHealth (n=67)	Control (n=61)	*P* value
Statins	120 (93.8)	64 (97)	56 (30)	.12

[a]Chronic obstructive pulmonary disease, kidney disease, and obstructive sleep apnea syndrome.

[b]LVEF: left ventricular ejection fraction.

[c]NSTEMI: non–ST-segment elevation myocardial infarction.

[d]STEMI: ST-segment elevation acute myocardial infarction.

Primary Outcomes

Overview

The outcome variables at the end of the 9-month follow-up period are shown in Table 2.

Table 2. Primary outcome variables at baseline and 9 months.

	Total (n=128)	mHealth (n=67)	Control (n=61)	*P* value
Mediterranean diet				
Mediterranean adherence score				
Baseline				.57[a]
Mean (SD)	7.38 (2.28)	7.24 (2.21)	7.52 (2.37)	
95% CI	6.98-7.77	6.70-7.78	6.92-8.13	
9 months				<.001[a]
Mean (SD)	11.03 (2.05)	11.83 (1.74)	10.14 (2.02)	
95% CI	10.66-11.41	11.39-12.27	9.60-10.68	
Good adherence, n (%)				
Baseline	36 (28.1)	18 (27)	18 (30)	.74
9 months	99 (83.2)	57 (91)	42 (75)	.02
Food consumption, n (%)				
Red meat ≤1/week				
Baseline	54 (42.2)	32 (48)	22 (36)	.05
9 months	104 (86.7)	60 (95)	44 (77)	.001
Vegetables ≥2/day				
Baseline	36 (28.1)	15 (22)	21 (34)	.54
9 months	76 (62.8)	48 (75)	28 (49)	.03
Fruits ≥2/day				
Baseline	48 (37.5)	23 (34)	25 (41)	.81
9 months	97 (80.1)	58 (91)	39 (68)	.02
Whole grains ≥1/day				
Baseline	9 (7.1)	7 (11)	2 (3)	.48
9 months	61 (50.4)	41 (64)	20 (35)	.003
Industrial pastry <2/week				
Baseline	57 (44.6)	29 (43)	28 (46)	.47
9 months	112 (92.6)	60 (94)	52 (91)	.003
Physical activity (patient-reported)				
Minnesota Physical Activity Survey (min/week)				
Baseline				.51
Mean (SD)	368.97 (304.06)	385.76 (286.53)	350.19 (323.99)	
95% CI	315.14-422.80	315.32-456.20	265.75-434.62	
9 months				.007
Mean (SD)	549.76 (300.87)	619.14 (318.21)	471.70 (261.43)	
95% CI	495.14-604.37	539.00-699.28	401.68-541.71	
Smoking status, n(%)				
Smoker				
9 months	23 (18.0)	8 (12)	15 (25)	.07
Smoking cessation	36 (61.0)	25 (76)	11 (42)	.01
Nicotine dependence (Fagerström test; score)				
Baseline				.58

	Total (n=128)	mHealth (n=67)	Control (n=61)	*P* value
Mean (SD)	4.91 (2.57)	4.74 (2.45)	5.10 (2.73)	
95% CI	4.28-5.54	3.90-5.59	4.09-6.10	
9 months				.02[a]
Mean (SD)	1.56 (2.61)	0.93 (2.29)	2.27 (2.82)	
95% CI	0.86-2.27	0.06-1.81	1.14-3.40	
Cardiovascular risk factors				
CVRF[b] knowledge (score)				
Baseline				.003
Mean (SD)	104.98 (9.83)	107.42 (8.90)	102.31 (10.17)	
95% CI	103.26-106.70	105.25-109.70	99.71-104.92	
9 months				<.001[a]
Mean (SD)	115.19 (7.47)	118.70 (2.65)	111.25 (9.05)	
95% CI	113.84-116.55	118.03-119.37	108.83-113.67	
Medication adherence, n (%)				
MGL MAQ[c]				
Baseline	53 (50.0)	26 (47)	27 (53)	.56
Final	98 (81.7)	54 (84)	44 (79)	.4
Quality of life				
SF-12[d] PCS[e] (score)				
Baseline				.20
Mean (SD)	39.55 (12.24)	40.87 (11.98)	38.10 (12.46)	
95% CI	37.39-41.71	37.93-43.82	34.87-41.32	
9 months				.03[a]
Mean (SD)	43.73 (10.96)	45.80 (10.79)	41.40 (10.78)	
95% CI	41.74-45.72	43.08-48.51	38.52-44.29	
SF-12 MCS[f] (score)				
Baseline				.02
Mean (SD)	48.61 (10.21)	50.51 (9.00)	46.52 (11.09)	
95% CI	46.81-50.41	48.30-52.72	43.65-49.38	
9 months				.57[a]
Mean (SD)	52.01 (9.54)	52.40 (9.27)	51.58 (9.89)	
95% CI	50.28-53.75	50.06-54.73	48.93-54.23	
Well-being				
WHO-5[g] (score)				
Baseline				.24
Mean (SD)	60.88 (19.49)	62.81 (18.95)	58.73 (20.01)	
95% CI	57.46-64.30	58.18-67.43	53.56-63.90	
9 months				.35
Mean (SD)	71.04 (12.31)	72.03 (12.71)	69.93 (11.86)	
95% CI	68.81-73.28	68.83-75.23	66.75-73.11	
Satisfaction				

	Total (n=128)	mHealth (n=67)	Control (n=61)	*P* value
9 months				.002[a]
Mean (SD)	47.18 (4.46)	48.22 (3.86)	46.00 (4.82)	
95% CI	46.37-47.99	47.25-49.20	44.71-47.29	

[a]Mann-Whitney *U* test (median, IQR): Mediterranean diet adherence at baseline (median 7.00, IQR 6.00-9.00); Mediterranean diet adherence at 9 months (median 12.00, IQR 9.00-13.00); nicotine dependence at 9 months (median 0.00, IQR 0.00-3.00); cardiovascular risk factor knowledge at 9 months (median 119.00, IQR 112.00-120.00); 12-item Short Form survey physical component score at 9 months (median 43.04, IQR 34.13-55.19); satisfaction (median 50, IQR 46.00-50.00).

[b]CVRF: cardiovascular risk factor.

[c]MGL MAQ: Morisky Green and Levine Medication Adherence Questionnaire.

[d]SF-12: 12-item Short Form survey.

[e]PCS: physical component score.

[f]MCS: mental component score.

[g]WHO-5: World Health Organization-5 Well-Being Index.

Smoking

Regarding stopping smoking, more participants did so in the mHealth group compared with those receiving standard health care (25/67, 75% vs 11/61, 42%; *P*=.01). The level of nicotine dependence also fell more in the mHealth group than in the control group (mean 0.93, SD 2.29 points vs mean 2.27, SD 2.82 points; *P*=.02; Table 2).

Adherence to the Mediterranean Diet and Frequency of Food Consumption

The score for adherence to the Mediterranean diet was significantly higher among the patients from the mHealth group (mean 11.83, SD 1.74 points) compared with the control group (mean 10.14, SD 2.02 points; *P*<.001) with an effect size (Cohen *d*) of 0.8. Moreover, the prevalence of adherence to the Mediterranean diet (score >9 points) was higher in the mHealth group than in the control group (57/67, 90% vs 42/61, 75%; *P*=.02). Regarding the frequency of food consumption, a significant reduction was observed in the consumption of red meat (≤1/week: 60/67, 95% vs 44/61, 77%; *P*=.001), and industrial pastries (<2/week: 60/67, 93% vs 52/61, 91%; *P*=.003) among the participants of the mHealth group compared with the control group. Furthermore, the patients in the mHealth group significantly increased their intake of vegetables (≥2/day: 48/67, 75% vs 28/61, 49%; *P*=.03), fruit (≥2/day: 58/67, 90% vs 39/61, 68%; *P*=.02), and whole-meal cereals (≥1/day: 41/67, 64% vs 20/61, 35%; *P*=.003).

Physical Activity

Likewise, the patients in the mHealth group spent significantly more time doing PA (patient-reported) every week than those in the control group (mean 619.14, SD 318.21 min/week vs mean 471.70, SD 261.43 min/week; *P*=.007), with an effect size (Cohen *d*) of 0.6.

Knowledge of CVRFs and a Healthy Lifestyle

The patients in the mHealth group presented a significantly higher level of knowledge about a healthy lifestyle and the control of CVRFs than those in the control group (mean 118.70, SD 2.65 points vs mean 111.25, SD 9.05 points; *P*<.001). Therapeutic adherence improved to a similar extent in both groups at the end of the follow-up period, with no significant differences being found between the groups.

Quality of Life and Well-Being

Regarding the quality of life, the physical component was significantly better among the patients in the intervention group (mean 45.80, SD 10.79 points vs mean 41.40, SD 10.78 points; *P*=.02), while both groups presented similar values for the mental component. As for the well-being index, the patients in the mHealth group presented slightly better scores, but they did not reach statistical significance.

Patient Overall Satisfaction

Regarding overall satisfaction with the health care received, the patients in the mHealth group scored significantly higher than those receiving standard health care, with respective average scores of 48.22 (SD 3.89) points and 46.00 (SD 4.82) points (*P*=.002).

Secondary Outcomes

BMI, WC, SBP, DBP, Heart Rate, HbA1c, and Lipids

Regarding the secondary outcome variables (Multimedia Appendix 1), at the end of the follow-up period, lower values were obtained for BMI, WC, SBP, DBP, heart rate, total cholesterol, LDL cholesterol, HDL cholesterol, triglycerides, and HbA_{1c} in both the mHealth and control groups, with no significant differences being found.

Anxiety and Depression

Regarding emotional aspects, the levels of both anxiety and depression were lower at the end of the study than at the beginning, but no significant differences were observed between the groups after the follow-up.

MACE and Other Complications

Major adverse events (3/128, 2.4% of all the participants) and complications such as bleeding, angina, and coronary interventions on another vessel (9/128, 7.2% of all the participants) were scarce in both groups during the 9-month follow-up, with no significant differences being found between the intervention and control groups.

Patient App Satisfaction and Usability

The participants in the mHealth group gave high scores for both satisfaction with the app (mean 44.38, SD 6.176 points) and its usability (mean 95.22, SD 7.369 points).

Discussion

Main Results

This study of people with CAD after a PCI analyzed the effect of an educational intervention based on the use of an mHealth app on the following factors: changes in lifestyle in terms of diet, PA, and tobacco use; level of knowledge of a healthy lifestyle; and the control of CVRFs, therapeutic adherence, and quality of life. The results obtained show that the intervention led to significant improvements in lifestyle in terms of adherence to the Mediterranean diet, PA, and stopping smoking, and that the patients increased their knowledge about a healthy lifestyle and controlling CVRFs. The patients' quality of life improved, and they reported a high level of satisfaction with the health care they received.

Comparison With Previous Work

The Mediterranean diet is considered a cornerstone of a healthy lifestyle to prevent cardiovascular diseases [4,42,43]. In this study, adherence to the Mediterranean diet improved in both groups at the end of the study, but this improvement was more noticeable in the patients in the mHealth group. This group of patients also increased the frequency with which they consumed healthy foods, such as fruit, vegetables, and whole-meal cereals, and decreased the frequency with which they consumed unhealthy foods, such as red meat and industrial pastries, to a greater extent than those in the control group (Table 2). On the contrary, the clinical trial conducted by Choi et al [29] on the effectiveness of dietary advice provided to patients with heart conditions through a smartphone found that the mHealth intervention did not increase adherence to the Mediterranean diet. However, other authors, such as Widmer et al [44], reported a more favorable score for the consumption of healthy food among patients undergoing cardiac rehabilitation who used an mHealth app. Given our results, providing dietary guidance through a mobile app may be a practical way of achieving healthy changes in people's diets, with the added advantage of using fewer resources than face-to-face visits.

As our results show, the time spent performing PA (patient-reported) was significantly higher in the mHealth group than in the control group. These results are in line with other clinical trials that observed improvements in the PA performed associated with an mHealth intervention [45-47]. However, Johnston et al [48] Did not find significant increases in the PA performed by patients using a health app for smartphones. Nevertheless, a recent meta-analysis has highlighted that mHealth apps are effective at encouraging patients to increase the amount of PA they do after a coronary event [18]. In the secondary prevention of CAD, one of the essential lifestyle changes that patients have to make is to perform regular PA, such as walking for at least 30 minutes every day [4,43]. For this reason, our results show that the mobile app was able to encourage the participants to start and continue doing more PA. However, our study did not have data available from an accelerometer as the PA was self-reported.

This study found an important decrease in the proportion of smokers and nicotine dependence measured by the Fagerström test at the end of the follow-up period in both the mHealth and control groups. However, the decrease in this addictive behavior was significantly greater among the participants in the mHealth group. Moreover, 75% (25/67) of the smokers in the mHealth group stopped smoking, while in the group receiving standard health care, only 42% (11/61) did so, even though at the beginning of the study the prevalence of smokers in the 2 groups was similar. Several meta-analyses have studied the effect of telehealth interventions on the prevalence of smoking, obtaining mixed results. While some [18,49] did not find statistically significant improvements in tobacco use when these technologies were used, another meta-analysis did observe a beneficial effect on stopping smoking among patients with CAD through the use of mHealth strategies [50]. Our positive results about stopping smoking with the use of the digital intervention could be due to several reasons, such as the support perceived by the patients and the interaction with the app that kept them motivated to change.

Our results show a significant improvement in the level of knowledge of a healthy lifestyle and the control of CVRFs among the patients using the health app compared to the control group. Few clinical trials involving mHealth interventions analyze the effectiveness of these interventions in improving the patient's knowledge. Among these, Dorje et al [51] reported an increase in the level of knowledge among patients using a telehealth app. Furthermore, the results of the studies included in EUROASPIRE II highlighted that in patients who had experienced a coronary event, a correlation existed between knowledge of CAD and changes in lifestyle and therapeutic adherence [52]. Knowledge of the risk factors for the disease is an essential requirement for patients to decide to adopt behaviors in line with a healthy lifestyle. However, people also need to be motivated to incorporate such behavior into their daily lives [53-55]. Innovative mHealth technology could help to achieve both objectives by increasing the patients' knowledge and motivation.

Therapeutic adherence improved to a similar extent at the end of the follow-up period in both groups, mHealth (54/67, 84%) and standard health care (44/61, 78%). Although data from previous studies indicate that adherence to treatment decreases progressively after a coronary event, reaching around 50% a year after experiencing a myocardial infarction [56-58], our results do not seem to support these data since the therapeutic adherence in both groups was greater than the figures reported by other authors. In addition, recent meta-analyses [18,59] have observed improvements in adherence to treatment in patients with cardiovascular disease with the use of mHealth apps with medication reminders or alarms on mobile devices.

Our results show that the patients in the mHealth group benefited from significant improvements in the physical component of quality of life compared with the participants receiving standard health care. These results are in partial agreement with a recent meta-analysis [18] but not with another meta-analysis [49] that

did not observe significant increases in the quality of life of these patients. The favorable effects of the mHealth intervention highlighted by our results may partly be related to the remote health care option provided by this intervention, whereby patients could resolve queries without having to leave their homes. Patients were given support, thus enhancing their self-sufficiency and the perception they have of their health. Improvements in quality of life may, in turn, motivate patients to comply with secondary prevention recommendations.

The overall satisfaction with the health care received was significantly higher in the patients from the mHealth group than in the control group, and satisfaction with the mHealth app was also high (mean 44.38, SD 6.176 points out of 50). The average score for the usability of the app was 95.22 (SD 7.36) out of 100. These results agree with those obtained by Johnston et al [48]. Our positive results could be attributed to the fact that the patients perceived the app to be a support tool that provided them with information about their disease as well as the motivation to adopt a healthy lifestyle. Telehealth interventions have the potential to overcome barriers and could represent a ground-breaking alternative for providing cardiac rehabilitation to patients for whom face-to-face health care is more complicated. Factors to be taken into consideration when choosing this option could be the patient's risk profile, preferences, and accessibility to health services.

The EVITE interactive tool provided relevant information about the disease, increased the patient's knowledge of their health status, promoted the self-reporting of important CVRF variables, and supplied the patients with visual feedback about changes, motivating them to improve their health. The patients felt more connected with the health professionals, obviating the need for unnecessary visits for trivial reasons. Most of the patients in the mHealth group stated that they would like to continue using the interactive support tool and that they would recommend the tool to other patients.

Our results showed no significant differences between the 2 groups with regard to BMI and WC at the end of the study, although recent meta-analyses show controversial results [18,60].

Lipid variables, blood pressure, and HbA_{1c} did not show significant differences between the 2 groups. These results agree with recent meta-analyses [18,61], which could be due to the intensive prescription of secondary prevention medication among the participants.

Patients have been found to experience the symptoms of anxiety and depression straight after undergoing a PCI, and these may persist for months or even years [62,63]. However, our results show a decrease in the level of anxiety and depression during the follow-up period in both the mHealth group and the control group (Multimedia Appendix 1). These results are in line with previous studies [18], although not with another meta-analysis [60].

Major adverse events were very rare in both groups, as were hospital readmissions. As a result of the important improvements achieved in recent years in the treatment of patients with CAD, the prevalence of short-term adverse events is generally low in different trials.

Clinically Meaningful Improvements in Primary Outcomes

The improvements observed in the mHealth group are clinically relevant, according to scientific evidence [4]. One year after myocardial infarction, patients who continue to smoke have worse cardiovascular outcomes and death than nonsmokers, and the long-term outcomes among those who quit smoking are comparable with those of nonsmokers [64]. We observed a 33% higher prevalence of smoking cessation in the mHealth group than in the control group.

Adults are recommended to perform at least 150-300 minutes a week of moderate-intensity PA, and additional benefits are gained with even more [4,65]. In our study, cardiologists advised the patients to walk for at least 1 hour every day, resulting in 420 minutes per week of PA. While all the participants in our study complied with the PA recommendations, those in the mHealth group performed more, so they would be expected to benefit from this because they are more active than the control group, with an important effect size (Cohen *d*) of 0.6. It should be noted that this study was performed in a geographical area in which the weather is generally good, which encourages people to go outside for a walk.

Adherence to the Mediterranean diet improves survival rates and reduces the risk of cardiovascular disease in individuals with CVD [66]. Moreover, greater adherence to this diet is associated with a 10% reduction in cardiovascular events or mortality [67]. In our study, the prevalence of good adherence to the Mediterranean diet (>9 points) was 15% higher among participants in the mHealth group. A meta-analysis reported a 4% lower risk of cardiovascular mortality for each additional serving of fruit and vegetables per day [68]. Likewise, a high intake of processed meat and unprocessed red meat is associated with an increased risk of atherosclerotic cardiovascular disease of 7% and 3%, respectively [69]. In our study, the prevalence of red meat consumption of ≤1 time per week was 18% higher in the mHealth group.

Knowledge of cardiovascular disease helps patients control risk factors and adopt behaviors that promote cardiovascular health [53,54]. For this reason, it could be considered that the more favorable results in the level of knowledge observed in our patients in the mHealth group could have a clinical impact.

Improving quality of life is one of the most important objectives to be achieved with CAD [70]. Thus, assessing quality of life allows the impact of the disease and its treatment on daily life to be determined. In our study, the score in the physical dimension of quality of life was statistically higher in the mHealth group, a finding that may be related to the capacity of the EVITE app to provide support to patients.

Limitations and Strengths

Our study has a few limitations. First, the participants included in the mHealth group agreed to participate voluntarily and gave their written consent, meaning that they may have been more motivated to adhere to secondary prevention. A second

limitation could be that to participate in the mHealth programs, the patients had to have a mobile phone or tablet with an internet connection, which could suggest that the participants were younger. However, 99% of the European population currently has a mobile telephone, and 86% of homes have an internet connection [71], although our results could also be due to the fact that participants in the mHealth group had a higher socioeconomic status. Moreover, women were less represented in our study, as only 29% (36/128 of the participants belonged to this population group. The third limitation is the limited sample size and the duration of the follow-up. The barrier to recruiting more participants was that the hospital was occupied by patients with COVID-19. A larger sample size and a longer follow-up could enhance the knowledge gleaned about the efficacy of the intervention. The fourth limitation is that most of the primary outcomes were self-reported and the study was open-label, but we assumed the patients responded honestly since they filled out the questionnaire without indicating their names. The questionnaires only contained an identification number so that the statisticians could record information in the database. This was explained to the patients so that they could respond freely. The fifth limitation is detection and performance bias since the blinding of the participants was not possible because of the nature of the intervention. The coordinators and the researchers analyzing the results were blinded to the allocation of participants, however. Finally, this study was conducted at a single public specialty reference hospital, and it would be necessary to test whether the favorable effects observed are generalizable to other health care settings. Additionally, in our study, we did not carry out an economic evaluation that could provide useful information to clarify whether the implementation of this mHealth intervention in the health care of these patients is cost-effective.

This study also presents some strengths, such as the fact that the sample is representative of patients who experience a coronary event in different localities, as the study was conducted on patients treated in the cardiology unit of a reference hospital that receives patients from both rural and urban areas. In addition, all the instruments and tools used in the clinical trial were validated internationally, which allows for generalization and for the study to be compared with others. Another strength was that the trial was conducted using a smartphone app, so the intervention involved the latest, up-to-date technology. In addition, the study included exclusively patients with CAD, so the sample of patients was more homogeneous than in studies that included patients with cardiovascular disease. The inclusion of different kinds of behavioral, metabolic, and psychosocial variables provides a broad view of the results obtained with mHealth technology. Finally, most previous trials were conducted with apps that encouraged behavioral changes related to performing PA, weight loss, or therapeutic adherence but did not comprehensively tackle all the CVRFs as our app does.

Conclusions

Using the EVITE app with patients who have experienced a coronary event is an effective way to improve their lifestyle in terms of adherence to the Mediterranean diet, frequency of healthy food consumption, PA, stopping smoking, knowledge of a healthy lifestyle, and controlling CVRFs, quality of life, and satisfaction with the health care received.

More studies are required to examine the impact of smartphone interventions on people who have undergone a coronary event, with long-term follow-ups that analyze mortality and cardiac-cause hospitalization, because these are important yardsticks of the success of secondary prevention strategies that make it possible to establish the clinical importance of the findings. Cost analyses are also required to promote the generalized use of these tools, their implementation, and their feasibility.

Acknowledgments

We would like to thank the Andalusian Ministry of Health and Families, Spain, for the funding (public funding: PI-0067-2018). JAT is funded by CIBEROBN–Instituto de Salud Carlos III (CB12/03/30038) and co-funded by the European Regional Development Fund.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Secondary outcome variables at baseline and 9 months.
[DOC File , 21 KB - mhealth_v12i1e48756_app1.doc]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1166 KB - mhealth_v12i1e48756_app2.pdf]

References

1. The top 10 causes of death. World Health Organization. 2020. URL: https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death [accessed 2022-09-01]

2. Cardiovascular diseases (CVDs). World Health Organization. URL: https://www.who.int/en/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds) [accessed 2022-09-08]
3. Neumann FJ, Sousa-Uva M, Ahlsson A, Alfonso F, Banning AP, Benedetto U, et al. 2018 ESC/EACTS guidelines on myocardial revascularization. Eur Heart J 2019;40(2):87-165 [FREE Full text] [doi: 10.1093/eurheartj/ehy394] [Medline: 30165437]
4. Visseren FLJ, Mach F, Smulders YM, Carballo D, Koskinas KC, Bäck M, et al. 2021 ESC guidelines on cardiovascular disease prevention in clinical practice: developed by the task force for cardiovascular disease prevention in clinical practice with representatives of the European Society of Cardiology and 12 medical societies with the special contribution of the European Association of Preventive Cardiology (EAPC). Eur Heart J 2021;42(34):3227-3337 [FREE Full text] [doi: 10.1093/eurheartj/ehab484] [Medline: 34458905]
5. Piepoli MF, Abreu A, Albus C, Ambrosetti M, Brotons C, Catapano AL, et al. Update on cardiovascular prevention in clinical practice: a position paper of the European Association of Preventive Cardiology of the European Society of Cardiology. Eur J Prev Cardiol 2020;27(2):181-205 [FREE Full text] [doi: 10.1177/2047487319893035] [Medline: 31826679]
6. Marquina C, Zomer E, Vargas-Torres S, Zoungas S, Ofori-Asenso R, Liew D, et al. Novel treatment strategies for secondary prevention of cardiovascular disease: a systematic review of cost-effectiveness. Pharmacoeconomics 2020;38(10):1095-1113. [doi: 10.1007/s40273-020-00936-0] [Medline: 32583316]
7. Kotseva K, Wood D, De Bacquer D, EUROASPIRE investigators. Determinants of participation and risk factor control according to attendance in cardiac rehabilitation programmes in coronary patients in Europe: EUROASPIRE IV survey. Eur J Prev Cardiol 2018;25(12):1242-1251 [FREE Full text] [doi: 10.1177/2047487318781359] [Medline: 29873511]
8. Nicholls SJ, Nelson M, Astley C, Briffa T, Brown A, Clark R, et al. Optimising secondary prevention and cardiac rehabilitation for atherosclerotic cardiovascular disease during the COVID-19 pandemic: a position statement from the Cardiac Society of Australia and New Zealand (CSANZ). Heart Lung Circ 2020;29(7):e99-e104 [FREE Full text] [doi: 10.1016/j.hlc.2020.04.007] [Medline: 32473781]
9. Indraratna P, Biswas U, Yu J, Schreier G, Ooi SY, Lovell NH, et al. Trials and tribulations: mHealth clinical trials in the COVID-19 pandemic. Yearb Med Inform 2021;30(1):272-279 [FREE Full text] [doi: 10.1055/s-0041-1726487] [Medline: 33882601]
10. Measuring the information society report 2014. International Telecommunication Union. URL: https://www.itu.int/en/ITU-D/Statistics/Documents/publications/mis2014/MIS2014_without_Annex_4.Pdf [accessed 2021-03-06]
11. Digital trends in Europe 2021: ICT trends and developments in Europe, 2017-2020. International Telecommunication Union. URL: https://www.itu.int/dms_pub/itu-d/opb/ind/D-IND-DIG_TRENDS_EUR.01-2021-PDF-E.pdf [accessed 2021-03-06]
12. Tekkeşin A, Hayıroğlu M, Çinier G, Özdemir YS, İnan D, Yüksel G, et al. Lifestyle intervention using mobile technology and smart devices in patients with high cardiovascular risk: a pragmatic randomised clinical trial. Atherosclerosis 2021;319:21-27. [doi: 10.1016/j.atherosclerosis.2020.12.020] [Medline: 33465658]
13. Hayıroğlu M, Çınar T, Çinier G, Karakaya A, Yıldırım M, Güney BC, et al. The effect of 1-year mean step count on the change in the atherosclerotic cardiovascular disease risk calculation in patients with high cardiovascular risk: a sub-study of the LIGHT randomized clinical trial. Kardiol Pol 2021;79(10):1140-1142 [FREE Full text] [doi: 10.33963/KP.a2021.0108] [Medline: 34506630]
14. Hayiroğlu M, Çinier G, Pay L, Yüksel G, Durak F, Palice A, et al. The relation between average 1-year home blood pressure and the change in pro-BNP and left ventricle mass index. Blood Press Monit 2022;27(5):327-333. [doi: 10.1097/MBP.0000000000000611] [Medline: 35866500]
15. Hayıroğlu M, Çınar T, Cilli Hayıroğlu S, Şaylık F, Uzun M, Tekkeşin A. The role of smart devices and mobile application on the change in peak VO2 in patients with high cardiovascular risk: a sub-study of the LIGHT randomised clinical trial. Acta Cardiol 2023;78(9):1000-1005. [doi: 10.1080/00015385.2023.2223005] [Medline: 37318090]
16. Subedi N, Rawstorn JC, Gao L, Koorts H, Maddison R. Implementation of telerehabilitation interventions for the self-management of cardiovascular disease: systematic review. JMIR Mhealth Uhealth 2020;8(11):e17957 [FREE Full text] [doi: 10.2196/17957] [Medline: 33245286]
17. Jin K, Khonsari S, Gallagher R, Gallagher P, Clark AM, Freedman B, et al. Telehealth interventions for the secondary prevention of coronary heart disease: a systematic review and meta-analysis. Eur J Cardiovasc Nurs 2019;18(4):260-271 [FREE Full text] [doi: 10.1177/1474515119826510] [Medline: 30667278]
18. Cruz-Cobo C, Bernal-Jiménez M, Vázquez-García R, Santi-Cano MJ. Effectiveness of mHealth interventions in the control of lifestyle and cardiovascular risk factors in patients after a coronary event: systematic review and meta-analysis. JMIR Mhealth Uhealth 2022;10(12):e39593 [FREE Full text] [doi: 10.2196/39593] [Medline: 36459396]
19. Martínez-González MA, García-Arellano A, Toledo E, Salas-Salvadó J, Buil-Cosiales P, Corella D, et al. A 14-item Mediterranean diet assessment tool and obesity indexes among high-risk subjects: the PREDIMED trial. PLoS One 2012;7(8):e43134 [FREE Full text] [doi: 10.1371/journal.pone.0043134] [Medline: 22905215]
20. Maddison R, Pfaeffli L, Whittaker R, Stewart R, Kerr A, Jiang Y, et al. A mobile phone intervention increases physical activity in people with cardiovascular disease: results from the HEART randomized controlled trial. Eur J Prev Cardiol 2015;22(6):701-709 [FREE Full text] [doi: 10.1177/2047487314535076] [Medline: 24817694]

21. Bernal-Jiménez M, Calle-Pérez G, Gutiérrez-Barrios A, Gheorghe L, Solano-Mulero AM, Trujillo-Garrido N, et al. Design and validation of a scale of knowledge of cardiovascular risk factors and lifestyle after coronary event. J Clin Med 2022;11(10):2773 [FREE Full text] [doi: 10.3390/jcm11102773] [Medline: 35628900]
22. Chow CK, Redfern J, Hillis GS, Thakkar J, Santo K, Hackett ML, et al. Effect of lifestyle-focused text messaging on risk factor modification in patients with coronary heart disease: a randomized clinical trial. JAMA 2015;314(12):1255-1263 [FREE Full text] [doi: 10.1001/jama.2015.10945] [Medline: 26393848]
23. Prochaska JO, DiClemente CC, Norcross JC. In search of how people change: applications to addictive behaviors. Am Psychol 1992;47(9):1102-1114. [doi: 10.1037//0003-066x.47.9.1102] [Medline: 1329589]
24. Robinson TN. Stealth interventions for obesity prevention and control: motivating behavior change. In: Dubé L, Bechara A, Dagher A, Drewnowski A, Lebel J, James P, et al, editors. Obesity Prevention: The Role of Brain and Society on Individual Behavior. New York: Academic Press; 2010:319-327.
25. Bandura A. Social Foundations of Thought and Action: A Social Cognitive Theory. Englewood Cliffs, NJ: Prentice-Hall; 1986.
26. Bernal-Jiménez M, Calle-Pérez G, Gutiérrez-Barrios A, Gheorghe L, Solano-Mulero AM, Rodríguez-Martín A, et al. Lifestyle and treatment adherence intervention after a coronary event based on an interactive web application (EVITE): randomized controlled clinical trial protocol. Nutrients 2021;13(6):1818 [FREE Full text] [doi: 10.3390/nu13061818] [Medline: 34071782]
27. Kosaraju A, Goyal A, Grigorova Y, Makaryus AN. Left ventricular ejection fraction. In: StatPearls. Treasure Island, FL: StatPearls Publishing; 2023.
28. Fagerström KO. Measuring degree of physical dependence to tobacco smoking with reference to individualization of treatment. Addict Behav 1978;3(3-4):235-241. [doi: 10.1016/0306-4603(78)90024-2] [Medline: 735910]
29. Choi BG, Dhawan T, Metzger K, Marshall L, Akbar A, Jain T, et al. Image-based mobile system for dietary management in an american cardiology population: pilot randomized controlled trial to assess the efficacy of dietary coaching delivered via a smartphone app versus traditional counseling. JMIR Mhealth Uhealth 2019;7(4):e10755 [FREE Full text] [doi: 10.2196/10755] [Medline: 31012860]
30. Fernández-Ballart JD, Piñol JL, Zazpe I, Corella D, Carrasco P, Toledo E, et al. Relative validity of a semi-quantitative food-frequency questionnaire in an elderly Mediterranean population of Spain. Br J Nutr 2010;103(12):1808-1816 [FREE Full text] [doi: 10.1017/S0007114509993837] [Medline: 20102675]
31. Elosua R, Marrugat J, Molina L, Pons S, Pujol E. Validation of the Minnesota leisure time physical activity questionnaire in Spanish men. Am J Epidemiol 1994;139(12):1197-1209. [doi: 10.1093/oxfordjournals.aje.a116966] [Medline: 8209878]
32. Elosua R, Garcia M, Aguilar A, Molina L, Covas MI, Marrugat J. Validation of the Minnesota leisure time physical activity questionnaire in Spanish women. Med Sci Sports Exerc 2000;32(8):1431-1437 [FREE Full text] [doi: 10.1097/00005768-200008000-00011] [Medline: 10949009]
33. Morisky DE, Green LW, Levine DM. Concurrent and predictive validity of a self-reported measure of medication adherence. Med Care 1986;24(1):67-74. [doi: 10.1097/00005650-198601000-00007] [Medline: 3945130]
34. Pagès-Puigdemont N, Valverde-Merino MI. Métodos para medir la adherencia terapeútica. Ars Pharm 2018;59(3):163-172. [doi: 10.30827/ars.v59i3.7387]
35. Ware J, Kosinski M, Keller SD. A 12-item short-form health survey: construction of scales and preliminary tests of reliability and validity. Med Care 1996;34(3):220-233. [doi: 10.1097/00005650-199603000-00003] [Medline: 8628042]
36. Dempster M, Donnelly M. Measuring the health related quality of life of people with ischaemic heart disease. Heart 2000;83(6):641-644 [FREE Full text] [doi: 10.1136/heart.83.6.641] [Medline: 10814620]
37. Schmidt S, Vilagut G, Garin O, Cunillera O, Tresserras R, Brugulat P, et al. Normas de referencia para el cuestionario de salud SF-12 versión 2 basadas en población general de Cataluña. Med Clin 2012;139(14):613-625. [doi: 10.1016/j.medcli.2011.10.024] [Medline: 22244683]
38. WHO-5 Well-Being Index. World Health Organization. URL: https://ogg.osu.edu/media/documents/MB%20Stream/who5.pdf [accessed 2022-09-08]
39. Goldberg D, Bridges K, Duncan-Jones P, Grayson D. Detecting anxiety and depression in general medical settings. BMJ 1988;297(6653):897-899 [FREE Full text] [doi: 10.1136/bmj.297.6653.897] [Medline: 3140969]
40. Duch Campodarbe FR, Ruiz de Porras Rosselló L, Gimeno Ruiz de Porras D, Allué Torra B, Palou Vall I. Psicometría de la ansiedad, la depresión y el alcoholismo en atención primaria. Semergen 1999;25(3):209-225.
41. King AC, Hekler EB, Grieco LA, Winter SJ, Sheats JL, Buman MP, et al. Harnessing different motivational frames via mobile phones to promote daily physical activity and reduce sedentary behavior in aging adults. PLoS One 2013;8(4):e62613 [FREE Full text] [doi: 10.1371/journal.pone.0062613] [Medline: 23638127]
42. Estruch R, Ros E, Salas-Salvadó J, Covas MI, Corella D, Arós F, et al. Primary prevention of cardiovascular disease with a mediterranean diet supplemented with extra-virgin olive oil or nuts. N Engl J Med 2018;378(25):e34 [FREE Full text] [doi: 10.1056/NEJMoa1800389] [Medline: 29897866]
43. Salas-Salvadó J, Díaz-López A, Ruiz-Canela M, Basora J, Fitó M, Corella D, et al. Effect of a lifestyle intervention program with energy-restricted mediterranean diet and exercise on weight loss and cardiovascular risk factors: one-year results of

the PREDIMED-Plus trial. Diabetes Care 2019;42(5):777-788 [FREE Full text] [doi: 10.2337/dc18-0836] [Medline: 30389673]
44. Widmer RJ, Allison TG, Lennon R, Lopez-Jimenez F, Lerman LO, Lerman A. Digital health intervention during cardiac rehabilitation: a randomized controlled trial. Am Heart J 2017;188:65-72. [doi: 10.1016/j.ahj.2017.02.016] [Medline: 28577682]
45. Duscha BD, Piner LW, Patel MP, Craig KP, Brady M, McGarrah RW, et al. Effects of a 12-week mHealth program on peak VO2 and physical activity patterns after completing cardiac rehabilitation: a randomized controlled trial. Am Heart J 2018;199:105-114. [doi: 10.1016/j.ahj.2018.02.001] [Medline: 29754647]
46. Lunde P, Bye A, Bergland A, Grimsmo J, Jarstad E, Nilsson BB. Long-term follow-up with a smartphone application improves exercise capacity post cardiac rehabilitation: a randomized controlled trial. Eur J Prev Cardiol 2020;27(16):1782-1792 [FREE Full text] [doi: 10.1177/2047487320905717] [Medline: 32106713]
47. Su JJ, Yu DSF. Effects of a nurse-led eHealth cardiac rehabilitation programme on health outcomes of patients with coronary heart disease: a randomised controlled trial. Int J Nurs Stud 2021;122:104040. [doi: 10.1016/j.ijnurstu.2021.104040] [Medline: 34333211]
48. Johnston N, Bodegard J, Jerström S, Åkesson J, Brorsson H, Alfredsson J, et al. Effects of interactive patient smartphone support app on drug adherence and lifestyle changes in myocardial infarction patients: a randomized study. Am Heart J 2016;178:85-94 [FREE Full text] [doi: 10.1016/j.ahj.2016.05.005] [Medline: 27502855]
49. Huang K, Liu W, He D, Huang B, Xiao D, Peng Y, et al. Telehealth interventions versus center-based cardiac rehabilitation of coronary artery disease: a systematic review and meta-analysis. Eur J Prev Cardiol 2015;22(8):959-971 [FREE Full text] [doi: 10.1177/2047487314561168] [Medline: 25488550]
50. Turan Kavradim S, Özer Z, Boz İ. Effectiveness of telehealth interventions as a part of secondary prevention in coronary artery disease: a systematic review and meta-analysis. Scand J Caring Sci 2020;34(3):585-603. [doi: 10.1111/scs.12785] [Medline: 31747080]
51. Dorje T, Zhao G, Tso K, Wang J, Chen Y, Tsokey L, et al. Smartphone and social media-based cardiac rehabilitation and secondary prevention in China (SMART-CR/SP): a parallel-group, single-blind, randomised controlled trial. Lancet Digit Health 2019;1(7):e363-e374 [FREE Full text] [doi: 10.1016/S2589-7500(19)30151-7] [Medline: 33323210]
52. Alm-Roijer C, Stagmo M, Udén G, Erhardt LR. Better knowledge improves adherence to lifestyle changes and medication in patients with coronary heart disease. Eur J Cardiovasc Nurs 2004;3(4):321-330 [FREE Full text] [doi: 10.1016/j.ejcnurse.2004.05.002] [Medline: 15572021]
53. Liu Q, Huang YJ, Zhao L, Wang W, Liu S, He GP, et al. Association between knowledge and risk for cardiovascular disease among older adults: a cross-sectional study in China. Int J Nurs Sci 2020;7(2):184-190 [FREE Full text] [doi: 10.1016/j.ijnss.2020.03.008] [Medline: 32685615]
54. Rajkumar E, Romate J. Behavioural risk factors, hypertension knowledge, and hypertension in rural India. Int J Hypertens 2020;2020:8108202 [FREE Full text] [doi: 10.1155/2020/8108202] [Medline: 32231796]
55. Mifsud JL, Galea J, Garside J, Stephenson J, Astin F. Motivational interviewing to support modifiable risk factor change in individuals at increased risk of cardiovascular disease: a systematic review and meta-analysis. PLoS One 2020;15(11):e0241193 [FREE Full text] [doi: 10.1371/journal.pone.0241193] [Medline: 33175849]
56. Pietrzykowski Ł, Michalski P, Kosobucka A, Kasprzak M, Fabiszak T, Stolarek W, et al. Medication adherence and its determinants in patients after myocardial infarction. Sci Rep 2020;10(1):12028 [FREE Full text] [doi: 10.1038/s41598-020-68915-1] [Medline: 32694522]
57. Piepoli MF, Hoes AW, Agewall S, Albus C, Brotons C, Catapano AL, et al. 2016 European guidelines on cardiovascular disease prevention in clinical practice: the sixth joint task force of the European Society of Cardiology and other societies on cardiovascular disease prevention in clinical practice (constituted by representatives of 10 societies and by invited experts) developed with the special contribution of the European Association for Cardiovascular Prevention & Rehabilitation (EACPR). Eur Heart J 2016;37(29):2315-2381 [FREE Full text] [doi: 10.1093/eurheartj/ehw106] [Medline: 27222591]
58. De Smedt D, Kotseva K, De Bacquer D, Wood D, De Backer G, Dallongeville J, et al. Cost-effectiveness of optimizing prevention in patients with coronary heart disease: the EUROASPIRE III health economics project. Eur Heart J 2012;33(22):2865-2872 [FREE Full text] [doi: 10.1093/eurheartj/ehs210] [Medline: 22843446]
59. Al-Arkee S, Mason J, Lane DA, Fabritz L, Chua W, Haque MS, et al. Mobile apps to improve medication adherence in cardiovascular disease: systematic review and meta-analysis. J Med Internet Res 2021;23(5):e24190 [FREE Full text] [doi: 10.2196/24190] [Medline: 34032583]
60. Xu Y, Ye H, Zhu Y, Du S, Xu G, Wang Q. The efficacy of mobile health in alleviating risk factors related to the occurrence and development of coronary heart disease: a systematic review and meta-analysis. Clin Cardiol 2021;44(5):609-619 [FREE Full text] [doi: 10.1002/clc.23596] [Medline: 33724494]
61. Akinosun AS, Polson R, Diaz-Skeete Y, De Kock JH, Carragher L, Leslie S, et al. Digital technology interventions for risk factor modification in patients with cardiovascular disease: systematic review and meta-analysis. JMIR Mhealth Uhealth 2021;9(3):e21061 [FREE Full text] [doi: 10.2196/21061] [Medline: 33656444]

62. Richards SH, Anderson L, Jenkinson CE, Whalley B, Rees K, Davies P, et al. Psychological interventions for coronary heart disease: Cochrane systematic review and meta-analysis. Eur J Prev Cardiol 2018;25(3):247-259 [FREE Full text] [doi: 10.1177/2047487317739978] [Medline: 29212370]
63. Liu Y, Zhao Y, Tian J, Tong T, Gao R, Liu Y. The association of depression following percutanous coronary intervention with adverse cardiovascular events: protocol for a systematic review and meta-analysis. Medicine (Baltimore) 2019;98(2):e13952 [FREE Full text] [doi: 10.1097/MD.0000000000013952] [Medline: 30633173]
64. Aker A, Saliba W, Schnaider S, Eitan A, Jaffe R, Zafrir B. The impact of smoking status 1 year after ST-segment elevation myocardial infarction on cardiovascular events and mortality in patients aged ≤60 years. Am J Cardiol 2022;175:52-57. [doi: 10.1016/j.amjcard.2022.04.019] [Medline: 35613953]
65. Kraus WE, Powell KE, Haskell WL, Janz KF, Campbell WW, Jakicic JM, et al. Physical activity, all-cause and cardiovascular mortality, and cardiovascular disease. Med Sci Sports Exerc 2019;51(6):1270-1281 [FREE Full text] [doi: 10.1249/MSS.0000000000001939] [Medline: 31095084]
66. Laffond A, Rivera-Picón C, Rodríguez-Muñoz PM, Juárez-Vela R, Ruiz de Viñaspre-Hernández R, Navas-Echazarreta N, et al. Mediterranean diet for primary and secondary prevention of cardiovascular disease and mortality: an updated systematic review. Nutrients 2023;15(15):3356 [FREE Full text] [doi: 10.3390/nu15153356] [Medline: 37571293]
67. Sofi F, Abbate R, Gensini GF, Casini A. Accruing evidence on benefits of adherence to the Mediterranean diet on health: an updated systematic review and meta-analysis. Am J Clin Nutr 2010;92(5):1189-1196 [FREE Full text] [doi: 10.3945/ajcn.2010.29673] [Medline: 20810976]
68. Wang X, Ouyang Y, Liu J, Zhu M, Zhao G, Bao W, et al. Fruit and vegetable consumption and mortality from all causes, cardiovascular disease, and cancer: systematic review and dose-response meta-analysis of prospective cohort studies. BMJ 2014;349:g4490 [FREE Full text] [doi: 10.1136/bmj.g4490] [Medline: 25073782]
69. Zhong VW, Van Horn L, Greenland P, Carnethon MR, Ning H, Wilkins JT, et al. Associations of processed meat, unprocessed red meat, poultry, or fish intake with incident cardiovascular disease and all-cause mortality. JAMA Intern Med 2020;180(4):503-512 [FREE Full text] [doi: 10.1001/jamainternmed.2019.6969] [Medline: 32011623]
70. Nowbar AN, Howard JP, Finegold JA, Asaria P, Francis DP. 2014 global geographic analysis of mortality from ischaemic heart disease by country, age and income: statistics from World Health Organisation and United Nations. Int J Cardiol 2014;174(2):293-298 [FREE Full text] [doi: 10.1016/j.ijcard.2014.04.096] [Medline: 24794549]
71. Statistics and indicators. International Telecommunication Union. 2021. URL: https://www.itu.int/pub/D-IND [accessed 2022-01-16]

Abbreviations

CAD: coronary artery disease
CVRF: cardiovascular risk factor
DBP: diastolic blood pressure
HbA1c: glycated hemoglobin
HDL: high-density lipoprotein
LDL: low-density lipoprotein
MACE: major adverse cardiovascular event
MET: metabolic equivalent
MGL MAQ: Morisky Green and Levine Medication Adherence Questionnaire
mHealth: mobile health
PA: physical activity
PCI: percutaneous coronary intervention
SBP: systolic blood pressure
SF-12: 12-item Short Form survey
SUS: System Usability Scale
WC: waist circumference

Edited by M Dorsch; submitted 06.05.23; peer-reviewed by M Hayıroğlu; comments to author 17.10.23; revised version received 27.10.23; accepted 17.01.24; published 22.04.24.

Please cite as:
Bernal-Jiménez MÁ, Calle G, Gutiérrez Barrios A, Gheorghe LL, Cruz-Cobo C, Trujillo-Garrido N, Rodríguez-Martín A, Tur JA, Vázquez-García R, Santi-Cano MJ
Effectiveness of an Interactive mHealth App (EVITE) in Improving Lifestyle After a Coronary Event: Randomized Controlled Trial
JMIR Mhealth Uhealth 2024;12:e48756
URL: https://mhealth.jmir.org/2024/1/e48756
doi:10.2196/48756
PMID:38648103

Original Paper

Use of Mobile Technologies to Streamline Pretriage Patient Flow in the Emergency Department: Observational Usability Study

Panzhang Wang[1*], MS; Lei Yu[1*], MS; Tao Li[1*], BS; Liang Zhou[1], PhD; Xin Ma[2], MD, PhD

[1]Department of Medical Informatics, Shanghai Sixth People's Hospital, Shanghai, China

[2]Department of Orthopedics, Shanghai Sixth People's Hospital, Shanghai, China

[*]these authors contributed equally

Corresponding Author:
Xin Ma, MD, PhD
Department of Orthopedics
Shanghai Sixth People's Hospital
600 Yishan Rd
Shanghai, 200233
China
Phone: 86 38297230
Fax: 86 38297230
Email: maxinsix@sina.com

Abstract

Background: In emergency departments (EDs), triage nurses are under tremendous daily pressure to rapidly assess the acuity level of patients and log the collected information into computers. With self-service technologies, patients could complete data entry on their own, allowing nurses to focus on higher-order tasks. Kiosks are a popular working example of such self-service technologies; however, placing a sufficient number of unwieldy and fixed machines demands a spatial change in the greeting area and affects pretriage flow. Mobile technologies could offer a solution to these issues.

Objective: The aim of this study was to investigate the use of mobile technologies to improve pretriage flow in EDs.

Methods: The proposed stack of mobile technologies includes patient-carried smartphones and QR technology. The web address of the self-registration app is encoded into a QR code, which was posted directly outside the walk-in entrance to be seen by every ambulatory arrival. Registration is initiated immediately after patients or their proxies scan the code using their smartphones. Patients could complete data entry at any site on the way to the triage area. Upon completion, the result is saved locally on smartphones. At the triage area, the result is automatically decoded by a portable code reader and then loaded into the triage computer. This system was implemented in three busy metropolitan EDs in Shanghai, China. Both kiosks and smartphones were evaluated randomly while being used to direct pretriage patient flow. Data were collected during a 20-day period in each center. Timeliness and usability of medical students simulating ED arrivals were assessed with the After-Scenario Questionnaire. Usability was assessed by triage nurses with the Net Promoter Score (NPS). Observations made during system implementation were subject to qualitative thematic analysis.

Results: Overall, 5928 of 8575 patients performed self-registration on kiosks, and 7330 of 8532 patients checked in on their smartphones. Referring effort was significantly reduced (43.7% vs 8.8%; P<.001) and mean pretriage waiting times were significantly reduced (4.4, SD 1.7 vs 2.9, SD 1.0 minutes; P<.001) with the use of smartphones compared to kiosks. There was a significant difference in mean usability scores for “ease of task completion” (4.4, SD 1.5 vs 6.7, SD 0.7; P<.001), “satisfaction with completion time” (4.5, SD 1.4 vs 6.8, SD 0.6; P<.001), and “satisfaction with support” (4.9, SD 1.9 vs 6.6, SD 1.2; P<.001). Triage nurses provided a higher NPS after implementation of mobile self-registration compared to the use of kiosks (13.3% vs 93.3%; P<.001). A modified queueing model was identified and qualitative findings were grouped by sequential steps.

Conclusions: This study suggests patient-carried smartphones as a useful tool for ED self-registration. With increased usability and a tailored queueing model, the proposed system is expected to minimize pretriage waiting for patients in the ED.

(JMIR Mhealth Uhealth 2024;12:e54642) doi:10.2196/54642

KEYWORDS

overcrowding; overcrowded; crowding; smartphone; queueing; pretriage; self-service; triage; emergency; urgent; ambulatory; mHealth; mobile health; workflow; health care management; hospital

Introduction

For outpatients, registration at the hospital is completed online via appointments initiated by the patients themselves. However, for patients arriving at the emergency department (ED), registration is normally completed by triage nurses who communicate with patients and then interact with computers for data entry. Such nurse-led registration often takes time and slows down the triage process [1,2].

When a patient presenting at the ED is seen, others have to wait [2]. During peak hours, pretriage waiting tends to be prolonged and unpredictable [2-4], representing one of the main factors contributing to ED overcrowding [5,6]. An extended waiting time in the ED has been associated with increased morbidity and mortality rates [7,8].

Recently, based on the experience in other industries [9], there has been research interest in the use of self-service technologies (SSTs) such as kiosks [10] in the ED setting. For example, Sinha et al [11] demonstrated that the registration process can be easily completed by patients or their proxies using kiosks. However, unlike unmanned retail stores [12], there is no evidence that the use of kiosks can completely replace human-based triage (ie, triage nurses). Without proper control, full self-triage could lead to disparities. For instance, some younger and more computer-literate patients would easily jump ahead of the older, frail, less educated, and less computer-facile patients.

Using ED kiosks requires patients, nurses, and machines to work together. For example, self-awareness and path-finding are required for new arrivals. Otherwise, a referring service is required from triage staff [13], which directs the patient flow to a specific location to complete self-registration. Kiosks could break down at any time, which interrupts the work of the triage staff to handle technical failures [13]. In addition, triage staff need to monitor the kiosk area because long lines at the kiosk could delay patient identification [13].

The coworking context has an impact on the overall triage efficiency, and kiosks work well in offline EDs, but there are still some barriers to overcome for their use as an effective SST. Outside the ED setting, more and more online medical encounters are based on mobile SSTs (ie, smartphones) via social apps [14-18]. For example, smartphones could improve data quality by reducing registration errors [19]. In China, WeChat is the most popular social app, which has more than 1000 million monthly active users [20]. WeChat is not only a popular app but can also serve as a platform to host other web-based apps [21,22]. These apps are built into WeChat and do not need a separate installation on patient smartphones. The apps can even be opened directly when the owner scans an eligible QR code. Beyond medical encounters, a mobile SST is also widely used in the context of mobile payment [23,24]. During COVID-19, the Chinese government used QR technologies to engage its citizens to fight the pandemic [25,26].

For security reasons, patient smartphones cannot be connected to ED triage servers, which limits the exploration of mobile SSTs in EDs. In a previous study, we used a tailored security architecture to send patient-reported data from patients' smartphones to hospital networks via an offline QR code reader [27]. Subsequently, Song et al [28] used a similar approach to transfer medical data to hospital information systems. In this study, we expanded on this work to explore a self-registration tool based on the stack of QR technologies and WeChat. To measure the impact on pretriage patient flow, we compared the use of mobile self-registration with the use of kiosks in a real-world setting. Our goal was to investigate the use of mobile technologies to improve pretriage flow in EDs.

Methods

Rationale

In March 2020, kiosks were introduced at a busy metropolitan ED in Shanghai with an annual census of 306,000 patients. In the crowded greeting area, the ED set up four kiosks to direct nonambulance arrivals for self-registration. With increasing uptake, lines to access the kiosk were frequently observed, prompting development of an alternative method.

Coworking Flow

This study was inspired by the concept of industry 5.0 [29]. With the increasing use of SSTs, the roles of nurses, patients, and machines were redefined in a coworking context (Figure 1A). Some repeated tasks shifted from nurses to tools and patients. Electronic forms were used to replace paper-based forms. The tool guides patients to complete some paperwork that was previously conducted by triage staff.

A patient needs to travel from the front end to the back end of the system, whether via kiosks (Figure 1B) or via smartphones (Figure 1C). The workflow was broken down into five different steps (Table 1). For kiosks, patients swipe their health cards on the front end for registration and present a paper document with a barcode to a triage nurse. After scanning the code with a hand-held device, the nurse quickly finds the record and then updates it with an acuity level based on the "quick-lookup" result [30-32].

For smartphone users, WeChat is used to scan a QR code at the front end. After filling out the form, the result is encoded into a QR code on patient smartphones using a structured format (Figure 1C). Patients show the digital proof to a triage nurse at the back end. The nurse scans the QR code and automatically creates a record after extracting the structured content. The nurse then uses the "quick-lookup" tool to assign an acuity level for patient triage.

Figure 1. Coworking context for self-registration at the emergency department. (A) The roles in the coworking context. (B) The kiosk-led pretriage flow. (C) The smartphone-led pretriage flow.

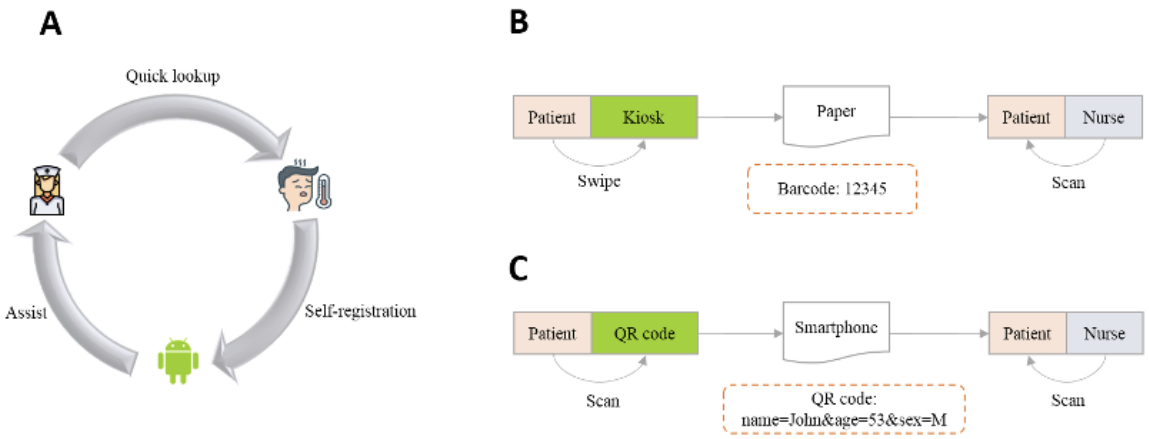


Table 1. Comparison of the workflow in self-registration at the emergency department using a kiosk and a smartphone.

Step	Kiosk	Smartphone
Reach	With or without a referral, a patient walks a certain distance to the kiosk area and finds an available machine for registration	With or without a referral, a patient scans a QR code posted along the route using WeChat. A web-based registration page is then displayed and they need to find a site nearby for registration
Initiate	A patient navigates to the homepage to find the launching icon and swipes a card for log-in or manually inputs the information when the card is not on hand or card-swiping fails	The patient enters the associated phone number to start
Input	On the public touchscreen	On the patient's smartphone
Print	The result is printed into a paper document	The result is printed into a QR code on the smartphone
Submit	The patient brings the printed paper result to the registration desk. The triage nurse uses a personal digital assistant equipped with a scanner to read the barcode.	The patient brings the smartphone to the registration desk. The triage nurse uses a personal digital assistant equipped with a scanner to read the QR code. A short confirmation message is sent to the patient via their registered contact number

Self-Registration Content

For registration, patients have to enter some personal information and the presenting reason for the ED visit (Table 2). With autocomplete-based suggestions for presenting problems [19], the system helps patients enter their chief complaints more efficiently and accurately. For kiosks, swiping a card is useful for saving time by avoiding having to input repeated data. However, when the card is not on hand, for security reasons, the patient has to enter all of the information manually. For kiosk users, the contact number is optional, thereby relying on triage nurses for patient validation. For smartphone users, the contact number is mandatory. When the content is scanned into the ED system by the triage staff, a confirmation message is sent to the patient according to the contact number. For experimental purposes, patients were required to fill out a field to indicate whether self-registration followed a referring service from clinical staff.

Technically, it is possible to use self-triage algorithms to assign an acuity level for a given patient condition [10]. However, this might lead to queue-jumping when the evaluation process is controlled by the patients themselves. In this study, the system was set up so that the evaluation result was not immediately available to patients after self-registration, but rather appeared as a suggestion to the triage staff on the hand-held screen. The triage nurse had the right to modify the result after human-based evaluation. The whole process was designed to mitigate paperwork burdens for triage staff.

Table 2. Content of the registration form.

Element code	Element description
Phone	Contact telephone number
Name	Patient name
Sex	Patient sex
Age	Patient age
Race	Patient race
Complaints	A list of standardized codes for chief complaints, separated by commas
Reason	Patient narrative text to explain the reason for presenting to the emergency department
Referred	Using the self-registration tool after being referred by an assistant

Security Architecture

For kiosks, the platform and the collected data were completely hosted in the hospital private network. Attacks from the internet were handled by the hospital firewall.

For smartphones, the front-end registration page was completely disconnected from the back end. Technically, this was designed as a static webpage hosted by the WeChat platform, which rendered a web form to allow patients to complete the registration online. However, the platform handled data submission in a different manner. Instead of transmitting the collected data over the internet, the data were saved to the local storage on the patient's smartphone, being displayed as a QR result. The QR result could then be read via the scanner held by the triage nurse. Thus, the data flow was executed from online to offline in a smart and secure manner. Since no third-party servers are involved, data security and privacy can be well protected. Hosting the static webpage on a mature platform with WeChat had benefits for both availability and scalability.

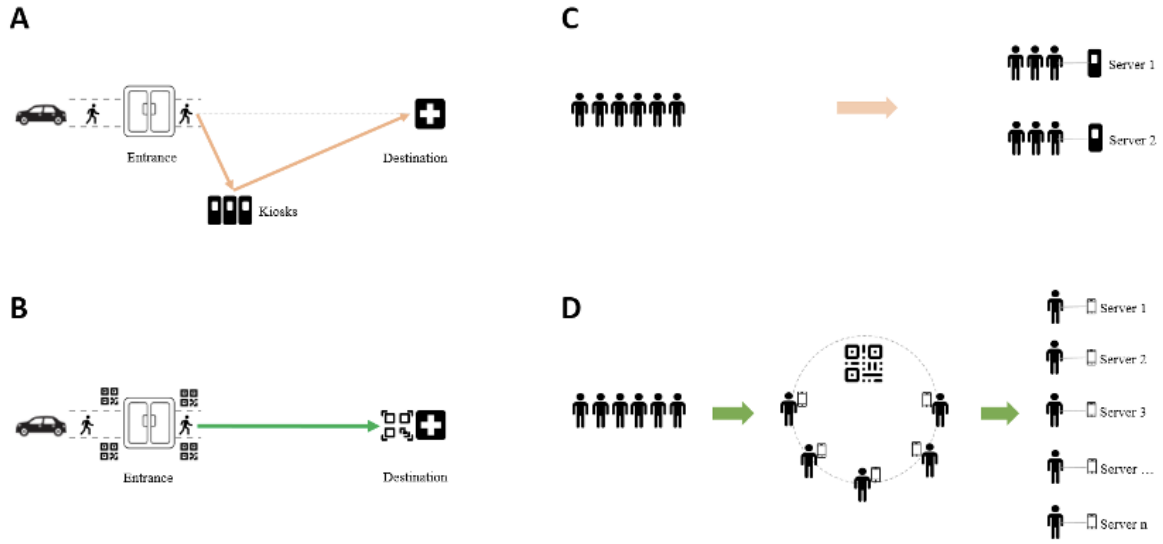
Multilingual Support

As a public machine, it is difficult for a kiosk to determine the language preference of an upcoming user. Therefore, a button for language switching was installed on the public screen. By contrast, for smartphone users, the system can automatically detect the language preference based on the location setting.

Signage

Based on best practice [13,33], the kiosk area was placed inside the ED (Figure 2A), visible from the registration desk. Roadmaps were provided. For smartphones, the signage was printed out as QR symbols, which were posted in four different places, inside and outside the entrance (Figure 2B). The size of the signage was appropriately determined so that it could be easily scanned within a distance (Figure 2D). These QR codes encoded a link to the self-service app.

Figure 2. Signage and queueing. (A) Signage and patient flow in the kiosk group. (B) Signage and patient flow in the smartphone group. (C) Queues in the kiosk group. (D) Queues in the smartphone group.



Study Design

The evaluation was performed during peak hours from 5 PM to 9 PM in three EDs. In each center, an assistant was recruited to work near the entrance to make sure patients traveled to the service quickly by offering a referral service. Patients were told they could bypass self-registration and go directly to the registration desk. Both patients and their proxies were allowed to use the self-registration tools. Data were collected during a 20-day period in each center. Random days were chosen for the use of kiosks and smartphones. At a given time, only one digital tool was evaluated to direct the patient flow. For example, when kiosks were evaluated, the signs with QR codes were removed from the entrance.

In each center, assistants (N=4) were recruited to work at the self-service area to assist patients for self-registration. All of the assistants had a background in qualitative research, and they were trained on how to handle technical issues related to self-registration that could result in triage interruptions [34]. Their observations were documented at the end of each day. Two authors (TL and LY) with expertise in informatics, human factors, quality and safety, and qualitative research served as coders and reviewers. These authors worked with the assistants for coding. Once coding was complete, thematic analysis [35] was used to group codes. Discrepancies were discussed by both the coders and the assistants and reconciled by consensus prior to final analysis.

To evaluate timeliness and usability, medical students (N=210) were recruited to simulate ED arrivals. They received no training and were asked to arrive randomly. Upon arrival, the timestamp (T_{arrive}) was recorded. When they received an acuity result from triage staff, the timestamp, or T_{end}, was recorded. The interval (T_{end}–T_{arrive}) was used to denote pretriage waiting times. For usability evaluation, the After-Scenario Questionnaire (ASQ) [36] was used, which is a standard questionnaire with three questions: “ease of task completion,” “satisfaction with completion time,” and “satisfaction with support.” The ASQ has been validated and is commonly used in studies related to mobile health [37]; this tool is particularly suitable in scenarios where a user might finish a task despite not successfully completing the task. Scale options were set according to a range of 1 to 7, with higher scores representing a higher degree of satisfaction. Two additional assistants were recruited to work in certain posttriage areas to administer paper-based ASQs. An unanswered question was treated in the same manner as a “not applicable” (NA) question. Since the study was designed to evaluate self-services, most participants might complete a session without any help, leaving the question of “satisfaction with support” unanswered. However, since “satisfaction with support” was designed to measure the level of self-service received, these “NA” answers were treated as the highest score in this study. For “ease of task completion” and “satisfaction with completion time,” records with “NA” were treated as invalid feedback.

During the evaluation, triage nurses (N=30) were recruited to work on a shift in each group and two nurses were assigned to the triage unit on each shift. The Net Promoter Score (NPS) [38] was used to solicit the nurses’ opinions in using the tool to handle patient flow, including questions such as “On a scale from 0 to 10, how likely are you to recommend this tool to ED managers to deal with pretriage patient flow?” Participants scoring 9 or 10 were classified as “promoters,” those providing a rating of 7 or 8 were classified as “passives,” and those providing a rating from 0 to 6 were classified as “detractors.” The final NPS was calculated by subtracting the percentage of detractors from the percentage of promoters (excluding the passives). The result could range between –100% and +100%, and the performance was marked as “good” for scores above 50% [39].

Statistical Analysis

Data analysis was performed using SPSS software (version 27.0). The χ^2 test was used to analyze proportion data. All analyses were based on a two-sided P value, with $P<.05$ considered statistically significant.

Ethical Considerations

According to regulations of the Shanghai Ethics Committee for Clinical Research [40], the requirement for ethical approval was exempt since the hospital only exports deidentified data for study purposes. Participants were informed about the study orally. Individual and privacy-related data were not used in this study.

Results

Primary Results

The recruitment flow of real patients is displayed in Figure 3 and the participant characteristics are summarized in Table 3. Within the study time frame, 5928 of 8575 patients (69.13%) performed self-registration on kiosks and 7330 of 8532 patients (85.91%) checked in on their smartphones, representing a significant difference ($P<.001$). There was also a statistically significant difference in the percentage of patients referred to kiosks and smartphone self-registration (Table 3).

Timeliness and usability were evaluated using simulated arrivals (N=210) in each group. Compared to kiosks, pretriage waiting times were significantly reduced in the smartphone group, whereas the smartphone group scored significantly higher in the usability items “ease of task completion,” “satisfaction with completion time,” and “satisfaction with support” (Table 3).

The NPS was calculated based on feedback obtained from triage nurses (N=30) in each group. According to their scores, there were 14 promoters, 6 passives, and 10 detractors in the kiosk group, whereas there were 28 promoters, 2 passives, and 0 detractors in the smartphone group. The use of smartphones significantly improved the final NPS compared to the use of kiosks (Table 3).

Figure 3. Recruitment flowchart. SST: self-service technology.

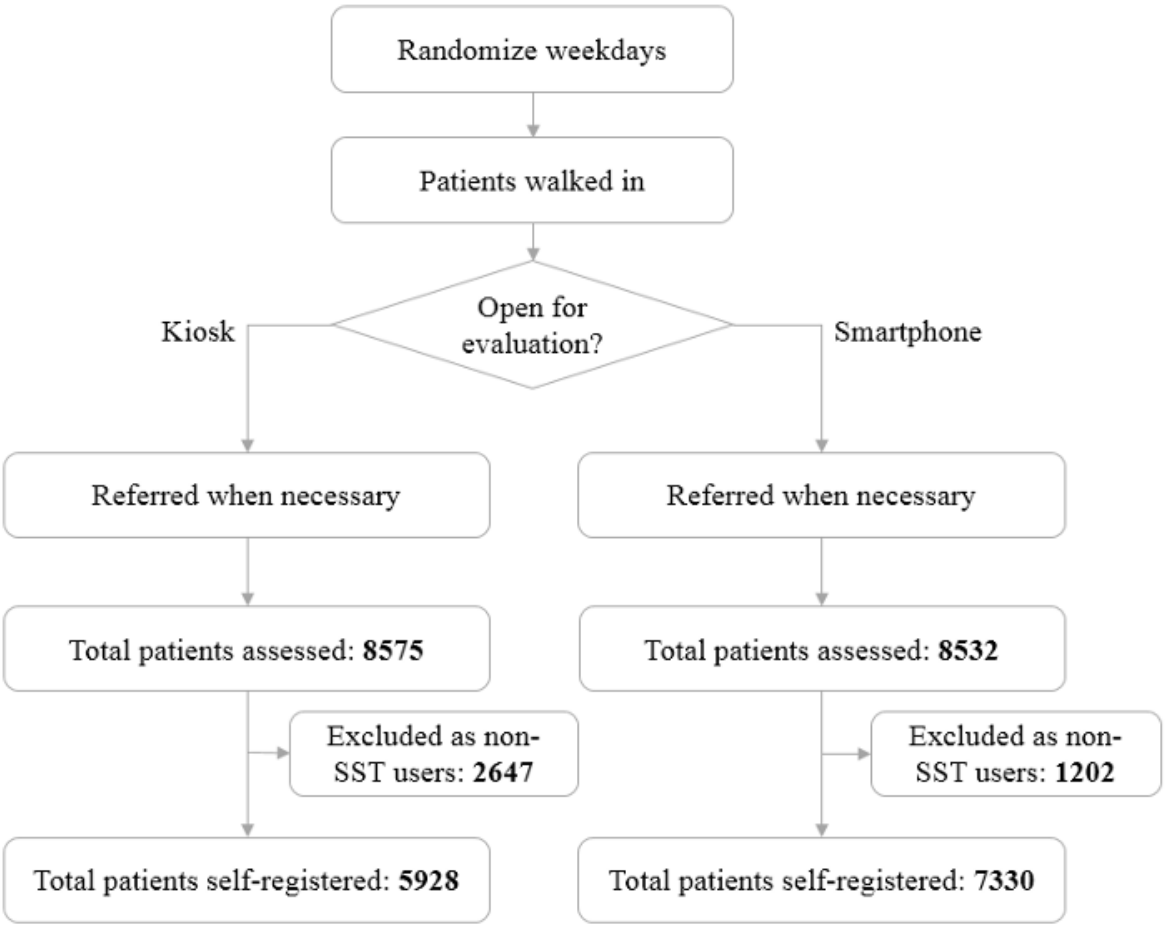


Table 3. Comparison of baseline characteristics across groups.

Characteristics	Kiosk	Smartphone	*P* value
Real patients			
Total patients assessed, n	8575	8532	—[a]
Self-registrations, n (%)	5928 (69.1)	7330 (85.9)	<.001
Age (years), mean (SD)	42.1 (21.1)	43.1 (20.6)	.91
Male sex, n (%)	2727 (46.0)	3291 (44.9)	.20
Sessions completed with referring, n (%)	2591 (43.7)	645 (8.8)	.001
Simulated patients			
Participants, n	210	210	—
Age (years), mean (SD)	22.5 (3.2)	22.6 (2.9)	.75
Male sex, n (%)	106 (50.5)	101 (48.1)	.63
Pretriage waiting time (minutes), mean (SD)	4.4 (1.7)	2.9 (1.0)	.001
Ease of task completion, mean (SD)	4.4 (1.5)	6.7 (0.7)	.001
Satisfaction with completion time, mean (SD)	4.5 (1.4)	6.8 (0.6)	.001
Satisfaction with support, mean (SD)	4.9 (1.9)	6.6 (1.2)	.001
Triage nurses			
Participants, n	30	30	—
Age (years), mean (SD)	36.8 (9.3)	36.0 (8.8)	.73
Male sex, n (%)	3 (10)	4 (13.3)	.69
Promoters, n	14	28	—
Passives, n	6	2	—
Detractors, n	10	0	—
NPS[b], %	13.3	93.3	.001

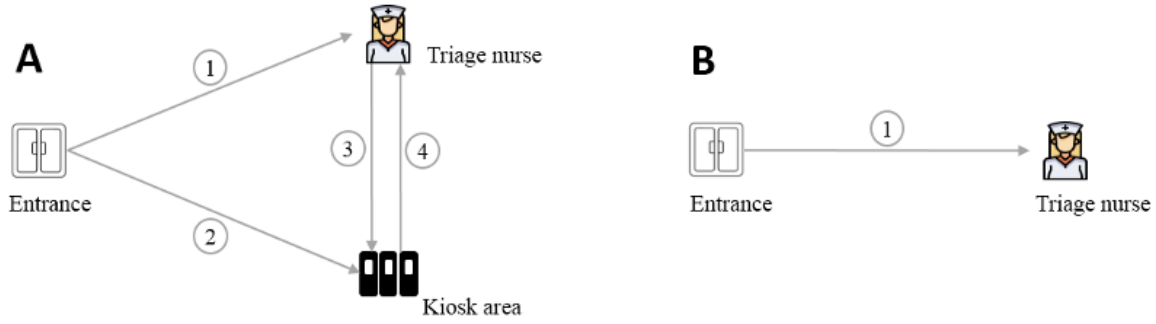
[a]Not applicable.

[b]NPS: Net Promoter Score.

Patient Flow

While kiosk users competed with each other while waiting in lines (Figure 2C) for check-in, smartphone users encircled (Figure 2D) a posted QR symbol to start using the tool. After scanning, smartphone users quickly dispersed around the ED to use their devices privately; thus, no queues were formed in the smartphone group. As illustrated in Figure 4, there were 4 paths in the kiosk group: (1) patients skipped self-registration and walked directly to triage staff, (2) patients walked to the kiosk area for self-registration, (3) patients left the kiosk area to see triage staff, and (4) patients switched to the self-registration line after arriving at the traditional registration desk. By comparison, there was only 1 path in the smartphone group: all patients followed the same path for self-registration and triage.

Figure 4. Paths in the pretriage flow. (A) Patient flow in the kiosk group. (B) Patient flow in the smartphone group.



Reach

Compared to the kiosk area, the posted QR symbols appeared to be easily seen by arrivals and then resulted in followers. Assistant B4 described that "It is good for patients to find and follow, especially posted outside." By contrast, in relation to kiosks, the assistant indicated that "the roadmap seems not working that well."

Although patients waited to be triaged in no particular order, the crowded situation in front of the registration desk could be seen from a distance. In the kiosk area, patients had to wait in order of their arrival, and the lines were also easily seen from the triage area. The two endpoints attracted patients and sometimes resulted in extra traffic of switching. Assistant A2 commented: "It is fine but I have a concern for patient safety when these two lines compete for patients." This concern was not found in the smartphone group.

Patient proxies were allowed to perform the actual data entry. However, in the kiosk group, the proxy often stepped into the line for the kiosk with the patient, resulting in a more crowded area. Although this is helpful for conversation-based data entry, it definitely increased the density of the self-service area. Indeed, assistant B2 advocated that "queue discipline is a must for kiosks." For smartphones, the proxy simply went ahead with scanning the QR code and then returned to the patient in an open area for collaborative data entry. The assistant then concluded, "the queue disappeared and it's privacy-friendly."

A kiosk might break down at any time. When this occurs, the clinical staff should be notified in a timely manner so that an "under repair" sign can be set up. However, this was not often the case in practice; instead, as described by assistant C3, "some patients just switched to another line or simply quit." In the case of smartphones, most interruptions were related to a low battery or bad signal. A smartphone failure would not cause a delay for other patients.

Initiate

Kiosks can be used not only for self-registration but also for other purposes [33], and each case requires its own launching icon on the homepage. Patients faced burdens with respect to page navigation and language switching. For example, the previous session might not end as expected, which would prevent appropriately resetting the screen to the homepage.

Card swiping was considered to be convenient but was also associated with intermittent failures such as an inability to read the card or swiping on the wrong side. Manual input was provided in case of any failure. However, this required some cognition effort, as described by assistant A1: "some patients spent too much time on trying." Patients who forgot their cards could also use the manual input option; however, this caused some issues, as described by assistant B3: "some patients blocked the line and stood too much time in front of a kiosk for searching."

By comparison, it was more straightforward for smartphone users to get started. After scanning, the page showed up right away without requiring any additional information to be input such as a password.

Input

During data entry, a kiosk failure was so disruptive that the entire line was impacted. Assistant C2 logged the following: "the entire line was moved to another." No failures were logged for smartphones in this step.

Print

After data entry was complete, the record was materialized as a proof. Namely, kiosks generated a paper document and smartphones used a QR code. With this proof, patients could confirm with the triage nurse that self-registration had been completed. However, patients had to spend some time to tear the paper document carefully from the printer. In some cases, the workflow was paused due to jammed paper, partially torn paper, or the printer being out of paper or ink. These delays were disruptive, as reported by assistant C1: "it really takes time to recover and most patients cannot wait." By contrast, printing QR codes on smartphones was instant and no failures were reported.

Submit

As mentioned above, patients needed to tear off the paper document from the printer before leaving the kiosk area. However, there were some cases in which more than one document was left in the printer. Consequently, some patients took the wrong record for submission, as logged by assistant A3: "Today, a record was marked with a wrong acuity level. It turned out that the patient took a wrong document." Such an out-of-sequence issue was never found in the smartphone group.

In the kiosk group, some printouts were occasionally found to be left on the floor. No such data privacy issues were observed in the smartphone group.

Discussion

Principal Findings

ED overcrowding is an unresolved issue worldwide [5,41], which threatens patient safety and public health [42]. Therefore, it is an urgent need to address the long pretriage waiting time, as it is associated with an increased level of morbidity and mortality [7,8]. Patient registration becomes a bottleneck when data entry is completed by triage nurses. With SSTs, triage nurses can focus on higher-order tasks [10,43] and enable rapid assessment [44] in EDs.

Unlike outpatients, registration is not a prerequisite for ED patients. This study thus contributes to improving ED self-registration in two ways. The first is related to the actual device provider. Instead of using in-house devices for registration, patients can use their own smartphones with advantages of flexibility and scalability. The second contribution is related to improved digital interaction. "Scan-to-process" should be easy for nurses, as this is already a ubiquitous technique in medical settings (eg, barcode-based medication administration) [45]. During COVID-19, QR codes were widely used for contact tracing [46,47]. Therefore, visitors are now familiar with the rule of "scan to enter" before stepping into public buildings. Accordingly, we applied this concept for patients to adapt to the new workflow of ED check-in.

EDs are space-constrained areas [48] and thus patient flow occurs in an often crowded setting. In addition to the registration desk, the use of a self-registration kiosk introduces another point of convergence. Kiosks cannot be positioned too close together in the case of a sudden increase in arrivals. Theoretically, one flow will start at the ED entrance, pass through to the kiosks, and finally meet up with triage nurses, while the other flow is directly from the entrance to the triage area. In reality, there may be a third path in which patients might return to the kiosk area for registration from the line formed at the registration desk to see a human after balancing out the potential waiting times. Passing through kiosks is helpful for reducing triage burden, but the increased movement within the ED and midway congestion might be detrimental to patient safety [49], especially for patients with potential critical illnesses. Smartphones barely have an impact on the incoming flow. Outside the entrance, scanning takes only a few seconds, and then patients can walk into the ED and complete registration while waiting to be triaged in front of nurses. In this way, experienced nurses could have a better chance of quickly scanning the patients to determine who should be treated first, even if the registration is not complete.

Although there was not a substantial difference in terms of completion time, using smartphones for self-registration could completely eliminate the prominent waiting issue caused by limited resources in EDs, as it enables an infinite-server queueing model [50]. Thus, serviceability and scalability could be refined from a different lens.

Usability is a key factor for users' continuance intention. Some patients who have experienced kiosk malfunctions or delayed help may not use the kiosk area the next time they have to visit the ED. This can explain the low number of participants in the kiosk group. In China, WeChat and QR technologies are frequently used [20]; therefore, patients will be familiar with a QR-based interaction when approaching the ED entrance. Smartphone users also encounter few failures during the process, which might enhance their willingness to continue using the mobile check-in service.

Indeed, the triage nurses gave the smartphones a higher usability score than the kiosks. There could be several reasons for this difference. First, kiosks generate more interruptions than smartphones, which is supported by the ASQ results and observations. Triage staff have to pause their work to handle any kiosk-related interruptions. Second, the overall observation could have an impact, as arrivals are directed in a flow for self-registration and triage. Third, with respect to the signage, the kiosk area is more challenging as a prerequisite for registration compared to the ED entrance, resulting in high referring effort. Finally, smartphones could cause less delay for patient identification than kiosks.

Smartphones increase hospitality in the pretriage flow [51]. First, compared to public screens, smartphones honor patient privacy. Second, the use of personal devices ensures hygiene and cleanliness, which are more difficult to maintain when using shared kiosk touchscreens. Third, social distancing is important in the ED, which could pose a challenge when lining up at kiosks to compete for timely identification. Finally, smartphones are more user-friendly in multilingual scenarios.

Although smartphones have numerous benefits, a computer-literate arrival can still choose to use a kiosk when a smartphone is not at hand. In addition, those who are not comfortable with SSTs can still register at the human-based line. Thus, providing numerous options in EDs can help to improve triage efficiency, especially during peak hours.

Limitations

Our study has limitations. First, this study was performed in three metropolitan ED centers in Shanghai; hence, our findings do not represent other disparate geographical areas with different ED volumes. For example, scanning QR codes might not be a common behavior adopted by patients in other countries. Second, due to the limited time frame of observation, some issues might not have been disclosed. Third, the smartphone-based tool was developed for the purpose of this evaluation and more features should be added in the future, such as integrating the tool with mobile sensors and using algorithms to predict triage results for decision-making. The lack of such features may have affected the usability results but could also serve as a basis to measure future enhancements.

Conclusions

EDs are overcrowded with long waiting times for registration and triage. Patient registration is managed in a single thread when completed by a triage team. Therefore, using SSTs would ease the burden on triage nurses and allow them to focus on higher-order tasks. Compared to kiosks, smartphones seem to be more convenient and suitable as a pretriage SST. However, it is recommended to offer multiple options of self-registration services in EDs.

Acknowledgments

The authors would like to thank the respondents for sharing their experiences and thoughts. This study is funded by Shanghai Sixth People's Hospital under project number DY2019015.

Data Availability

The data sets used or analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

XM and PW designed the study. XM had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. LZ performed the evaluation and assembled the data. PW, LY, and TL wrote the manuscript.

Conflicts of Interest

None declared.

References

1. Bijani M, Rakhshan M, Fararouei M, Torabizadeh C. Development and psychometric assessment of the triage nurses' professional capability questionnaire in the emergency department. BMC Nurs 2020 Sep 01;19(1):82 [FREE Full text] [doi: 10.1186/s12912-020-00476-0] [Medline: 32884445]
2. Sedgman R, Aldridge E, Miller J, Fleming D, Buntine P. Pre-triage wait times for non-ambulance arrivals in the emergency department: a retrospective video audit. Australas Emerg Care 2022 Jun;25(2):126-131. [doi: 10.1016/j.auec.2021.11.002] [Medline: 34824047]
3. Houston C, Sanchez L, Fischer C, Volz K, Wolfe R. Waiting for triage: unmeasured time in patient flow. West J Emerg Med 2015 Jan 01;16(1):39-42 [FREE Full text] [doi: 10.5811/westjem.2014.11.22824] [Medline: 25671006]
4. Yuzeng S, Hui LL. Improving the wait time to triage at the emergency department. BMJ Open Qual 2020 Feb;9(1):e000708 [FREE Full text] [doi: 10.1136/bmjoq-2019-000708] [Medline: 32019749]
5. Pearce S, Marr E, Shannon T, Marchand T, Lang E. Overcrowding in emergency departments: an overview of reviews describing global solutions and their outcomes. Intern Emerg Med 2024 Mar 02;19(2):483-491. [doi: 10.1007/s11739-023-03477-4] [Medline: 38041766]
6. Fekonja Z, Kmetec S, Fekonja U, Mlinar Reljić N, Pajnkihar M, Strnad M. Factors contributing to patient safety during triage process in the emergency department: a systematic review. J Clin Nurs 2023 Sep;32(17-18):5461-5477. [doi: 10.1111/jocn.16622] [Medline: 36653922]
7. Berg LM, Ehrenberg A, Florin J, Östergren J, Discacciati A, Göransson KE. Associations between crowding and ten-day mortality among patients allocated lower triage acuity levels without need of acute hospital care on departure from the emergency department. Ann Emerg Med 2019 Sep;74(3):345-356. [doi: 10.1016/j.annemergmed.2019.04.012] [Medline: 31229391]
8. Pearce S, Marchand T, Shannon T, Ganshorn H, Lang E. Emergency department crowding: an overview of reviews describing measures causes, and harms. Intern Emerg Med 2023 Jun 01;18(4):1137-1158 [FREE Full text] [doi: 10.1007/s11739-023-03239-2] [Medline: 36854999]
9. Mullangi S, Ibrahim AM, Chopra V. Toward patient-centered hospital design: what can airports teach us? Ann Intern Med 2017 Jul 04;167(1):48-49. [doi: 10.7326/M17-0484] [Medline: 28554191]
10. Joseph MJ, Summerscales M, Yogesan S, Bell A, Genevieve M, Kanagasingam Y. The use of kiosks to improve triage efficiency in the emergency department. NPJ Digit Med 2023 Feb 03;6(1):19. [doi: 10.1038/s41746-023-00758-2] [Medline: 36737642]
11. Sinha M, Khor K, Amresh A, Drachman D, Frechette A. The use of a kiosk-model bilingual self-triage system in the pediatric emergency department. Pediatr Emerg Care 2014 Jan;30(1):63-68. [doi: 10.1097/PEC.0000000000000037] [Medline: 24378865]
12. Jeon J, Kang S, Lee H, Kim J. A retail object classification method using multiple cameras for vision-based unmanned kiosks. IEEE Sensors J 2022 Nov 15;22(22):22200-22209. [doi: 10.1109/jsen.2022.3210699]
13. Ackerman SL, Tebb K, Stein JC, Frazee BW, Hendey GW, Schmidt LA, et al. Benefit or burden? A sociotechnical analysis of diagnostic computer kiosks in four California hospital emergency departments. Soc Sci Med 2012 Dec;75(12):2378-2385. [doi: 10.1016/j.socscimed.2012.09.013] [Medline: 23063214]
14. Zhao M, Song S, Zhao Y. Health information seeking on social apps among older adults living with chronic conditions. Proc Assoc Inf Sci Technol 2021 Oct 13;58(1):878-880. [doi: 10.1002/pra2.595]
15. Zhang J, Wu Q, Wang X, Zhang Y, Helena van Velthoven M. High response rate and low cost of WeChat-based data collection of infant and young child feeding data in rural Qinghai Province, China. J Glob Health 2022 Oct 21;12:11011 [FREE Full text] [doi: 10.7189/jogh.12.11011] [Medline: 36269216]
16. Zhang W, Zhang J, Ma R, Chen H, Huang J, Xu B. Patient expectations and awareness of information regarding adverse reactions in drug labelling in China. Int J Clin Pharm 2021 Feb 20;43(1):174-180. [doi: 10.1007/s11096-020-01127-9] [Medline: 32815111]
17. Zhang X, Wen D, Liang J, Lei J. How the public uses social media wechat to obtain health information in china: a survey study. BMC Med Inform Decis Mak 2017 Jul 05;17(Suppl 2):66 [FREE Full text] [doi: 10.1186/s12911-017-0470-0] [Medline: 28699549]
18. Chen X, Zhou X, Li H, Li J, Jiang H. The value of WeChat application in chronic diseases management in China. Comput Methods Programs Biomed 2020 Nov;196:105710 [FREE Full text] [doi: 10.1016/j.cmpb.2020.105710] [Medline: 32858284]

19. Xie W, Cao X, Dong H, Liu Y. The use of smartphone-based triage to reduce the rate of outpatient error registration: cross-sectional study. JMIR Mhealth Uhealth 2019 Nov 11;7(11):e15313 [FREE Full text] [doi: 10.2196/15313] [Medline: 31710300]
20. Monthly active users of the leading apps in China. Statista. URL: https://www.statista.com/statistics/1032630/china-leading-apps-by-monthly-active-users/ [accessed 2024-05-24]
21. Cheng A, Ren G, Hong T, Nam K, Koo C. An exploratory analysis of travel-related WeChat Mini Program usage: affordance theory perspective. In: Pesonen J, Neidhardt J, editors. Information and Communication Technologies in Tourism. Cham: Springer; 2018:333-334.
22. Duan Y, Li X, Guo L, Liang W, Shang B, Lippke S. A WeChat Mini Program-based intervention for physical activity, fruit and vegetable consumption among Chinese cardiovascular patients in home-based rehabilitation: a study protocol. Front Public Health 2022 Mar 22;10:739100 [FREE Full text] [doi: 10.3389/fpubh.2022.739100] [Medline: 35392478]
23. QR code payment. WeChat. URL: https://pay.weixin.qq.com/wechatpay_h5/pages/product/qrcode.shtml [accessed 2024-05-24]
24. Yan L, Tan G, Loh X, Hew J, Ooi K. QR code and mobile payment: the disruptive forces in retail. J Retail Consum Serv 2021 Jan;58:102300. [doi: 10.1016/j.jretconser.2020.102300]
25. Lei T, Deng J, Xie Y, Tian Y. Using health QR code to fight against COVID-19: Chinese government's new way of collaborative governance with enterprises. Adv Econ Bus Manag Res 2020;153(Proceedings of the 2020 International Conference on Management, Economy and Law (ICMEL 2020)):194-198. [doi: 10.2991/aebmr.k.201111.034]
26. Radhi AA. Use QR code and IoT mobile devices to prevent the spread of an epidemic (COVID-19). Int J Interact Mob Technol 2022 Jun 21;16(12):127-136. [doi: 10.3991/ijim.v16i12.32125]
27. Wang P, Li T, Yu L, Zhou L, Yan T. Towards an effective framework for integrating patient-reported outcomes in electronic health records. Digit Health 2022 Jul 13;8:20552076221112152 [FREE Full text] [doi: 10.1177/20552076221112152] [Medline: 35860613]
28. Song C, Kakuta Y, Negoro K, Moroi R, Masamune A, Sasaki E, et al. Collection of patient-generated health data with a mobile application and transfer to hospital information system via QR codes. Comput Methods Programs Biomed Update 2023;3:100099. [doi: 10.1016/j.cmpbup.2023.100099]
29. Özdemir V, Hekim N. Birth of Industry 5.0: making sense of big data with artificial intelligence, "The Internet of Things" and next-generation technology policy. Omics 2018 Jan;22(1):65-76. [doi: 10.1089/omi.2017.0194] [Medline: 29293405]
30. Betz M, Stempien J, Wilde A, Bryce R. A comparison of a formal triage scoring system and a quick-look triage approach. Eur J Emerg Med 2016 Jun;23(3):185-189. [doi: 10.1097/MEJ.0000000000000239] [Medline: 25622182]
31. Abolhassani M, Jacobsen KL. Purification of an acid-stable bovine leukocyte interferon. Vet Immunol Immunopathol 1991 Aug;29(1-2):171-181. [doi: 10.1016/0165-2427(91)90062-h] [Medline: 1949578]
32. Lin Y, Chen K, Wang J, Lai P. Simple triage and rapid treatment protocol for emergency department mass casualty incident victim triage. Am J Emerg Med 2022 Mar;53:99-103. [doi: 10.1016/j.ajem.2021.12.037] [Medline: 35007873]
33. Maramba ID, Jones R, Austin D, Edwards K, Meinert E, Chatterjee A. The role of health kiosks: scoping review. JMIR Med Inform 2022 Mar 29;10(3):e26511 [FREE Full text] [doi: 10.2196/26511] [Medline: 35348457]
34. Johnson K, Gillespie G, Vance K. Effects of interruptions on triage process in emergency department: a prospective, observational study. J Nurs Care Qual 2018;33(4):375-381 [FREE Full text] [doi: 10.1097/NCQ.0000000000000314] [Medline: 29319593]
35. Vaismoradi M, Turunen H, Bondas T. Content analysis and thematic analysis: implications for conducting a qualitative descriptive study. Nurs Health Sci 2013 Sep 11;15(3):398-405. [doi: 10.1111/nhs.12048] [Medline: 23480423]
36. Lewis JR. Psychometric evaluation of an after-scenario questionnaire for computer usability studies. SIGCHI Bull 1991 Jan;23(1):78-81. [doi: 10.1145/122672.122692]
37. Hajesmaeel-Gohari S, Khordastan F, Fatehi F, Samzadeh H, Bahaadinbeigy K. The most used questionnaires for evaluating satisfaction, usability, acceptance, and quality outcomes of mobile health. BMC Med Inform Decis Mak 2022 Jan 27;22(1):22 [FREE Full text] [doi: 10.1186/s12911-022-01764-2] [Medline: 35081953]
38. Reichheld FF. The one number you need to grow. Harv Bus Rev 2003 Dec;81(12):46-54. [Medline: 14712543]
39. Hamilton DF, Lane JV, Gaston P, Patton JT, Macdonald DJ, Simpson AHRW, et al. Assessing treatment outcomes using a single question: the net promoter score. Bone Joint J 2014 May;96-B(5):622-628 [FREE Full text] [doi: 10.1302/0301-620X.96B5.32434] [Medline: 24788496]
40. Ethical regulations. Shanghai Ethics Committee for Clinical Research. URL: http://www.seccr.cn/resource.php [accessed 2024-06-03]
41. Butun A, Kafdag EE, Gunduz H, Zincir V, Batibay M, Ciftci K, et al. Emergency department overcrowding: causes and solutions. Emerg Crit Care Med 2023 Feb 8;3(4):171-176. [doi: 10.1097/EC9.0000000000000078] [Medline: 2211145]
42. Kelen G, Wolfe R, D'Onofrio G, Mills A, Diercks D, Stern S. Emergency department crowding: the canary in the health care system. NEJM Catalyst 2021;2(5):1-4. [doi: 10.1056/CAT.21.0217]
43. Mahmood A, Wyant DK, Kedia S, Ahn S, Powell MP, Jiang Y, et al. Self-check-in kiosks utilization and their association with wait times in emergency departments in the United States. J Emerg Med 2020 May;58(5):829-840. [doi: 10.1016/j.jemermed.2019.11.019] [Medline: 31924466]

44. Anderson JS, Burke RC, Augusto KD, Beagan BM, Rodrigues-Belong ML, Frazer LS, et al. The effect of a rapid assessment zone on emergency department operations and throughput. Ann Emerg Med 2020 Feb;75(2):236-245. [doi: 10.1016/j.annemergmed.2019.07.047] [Medline: 31668573]
45. Grailey K, Hussain R, Wylleman E, Ezzat A, Huf S, Franklin BD. Understanding the facilitators and barriers to barcode medication administration by nursing staff using behavioural science frameworks. A mixed methods study. BMC Nurs 2023 Oct 12;22(1):378 [FREE Full text] [doi: 10.1186/s12912-023-01382-x] [Medline: 37828520]
46. Nakamoto I, Wang S, Guo Y, Zhuang W. A QR code-based contact tracing framework for sustainable containment of COVID-19: evaluation of an approach to assist the return to normal activity. JMIR Mhealth Uhealth 2020 Sep 07;8(9):e22321 [FREE Full text] [doi: 10.2196/22321] [Medline: 32841151]
47. Chambers T, Anglemyer A, Chen A, Baker M. An evaluation of the COVID-19 self-service digital contact tracing system in New Zealand. Health Policy 2024 Apr 17;144:105073 [FREE Full text] [doi: 10.1016/j.healthpol.2024.105073] [Medline: 38657315]
48. Hance K, Lyttle M, Humphreys S. 045 We're going to need a bigger boat! Evening opening of mothballed outpatient areas to reduce crowding in a children's emergency department. Emerg Med J 2019 Nov 21;36(12):805-806. [doi: 10.1136/emermed-2019-rcem.45]
49. Pryce A, Unwin M, Kinsman L, McCann D. Delayed flow is a risk to patient safety: a mixed method analysis of emergency department patient flow. Int Emerg Nurs 2021 Jan;54:100956. [doi: 10.1016/j.ienj.2020.100956] [Medline: 33360361]
50. Worthington D, Utley M, Suen D. Infinite-server queueing models of demand in healthcare: a review of applications and ideas for further work. J Operational Res Soc 2019 May 31;71(8):1145-1160. [doi: 10.1080/01605682.2019.1609878]
51. Pillai S, Haldorai K, Seo W, Kim W. COVID-19 and hospitality 5.0: Redefining hospitality operations. Int J Hosp Manag 2021 Apr 30;94:102869 [FREE Full text] [doi: 10.1016/j.ijhm.2021.102869] [Medline: 34785847]

Abbreviations

ASQ: After-Scenario Questionnaire
ED: emergency department
NA: not applicable
NPS: Net Promoter Score
SST: self-service technology

Edited by L Buis; submitted 16.11.23; peer-reviewed by E Benjamin, C Eccher; comments to author 15.12.23; revised version received 02.01.24; accepted 22.05.24; published 07.06.24.

Please cite as:
Wang P, Yu L, Li T, Zhou L, Ma X
Use of Mobile Technologies to Streamline Pretriage Patient Flow in the Emergency Department: Observational Usability Study
JMIR Mhealth Uhealth 2024;12:e54642
URL: https://mhealth.jmir.org/2024/1/e54642
doi:10.2196/54642
PMID:38848554

Original Paper

Implementing Systematic Patient-Reported Measures for Chronic Conditions Through the Naveta Value-Based Telemedicine Initiative: Observational Retrospective Multicenter Study

Gabriel Mercadal-Orfila[1,2], PharmD; Salvador Herrera-Pérez[3,4], PhD; Núria Piqué[5,6], PhD; Francesc Mateu-Amengual[7], PhD; Pedro Ventayol-Bosch[8], PhD; María Antonia Maestre-Fullana[9], BPH; Joaquín Ignacio Serrano-López de las Hazas[10], BPH; Francisco Fernández-Cortés[11], BPH; Francesc Barceló-Sansó[12], BPH; Santiago Rios[13], PhD

[1]Pharmacy Department, Hospital Mateu Orfila, Mahó, Spain

[2]Department of Biochemistry and Molecular Biology, Universitat de les Illes Balears, Mallorca, Spain

[3]Facultad de Ciencias de la Salud, Universidad Internacional de Valencia, Valencia, Spain

[4]CIBER Epidemiología y Salud Pública (CIBERESP), Barcelona, Spain

[5]Microbiology Section, Department of Biology, Healthcare and Environment, Faculty of Pharmacy and Food Sciences, Universitat de Barcelona, Barcelona, Spain

[6]Research Institute of Nutrition and Food Safety (INSA-UB), Universitat de Barcelona, Barcelona, Spain

[7]Healthcare Industry Solutions at MongoDB - Digital Health & Innovation, Barcelona, Spain

[8]Pharmacy Department, Hospital Universitari Son Espases, Palma de Mallorca, Spain

[9]Pharmacy Department, Hospital de Manacor, Manacor, Spain

[10]Pharmacy Department, Hospital Universitari Son Llàtzer, Palma de Mallorca, Spain

[11]Pharmacy Department, Hospital Comarcal d'Inca, Inca, Spain

[12]Pharmacy Department, Hospital Can Misses, Eivissa, Spain

[13]Departament de Genètica, Microbiologia i Estadística, Facultat de Biologia, Universitat de Barcelona, Barcelona, Spain

Corresponding Author:
Núria Piqué, PhD
Microbiology Section, Department of Biology, Healthcare and Environment
Faculty of Pharmacy and Food Sciences
Universitat de Barcelona
Diagonal Sud, Facultat de Farmàcia, Edifici A, Av Joan XXIII, 27-31
Barcelona, 08028
Spain
Phone: 34 934024496
Fax: 34 934024498
Email: npique@gmail.com

Abstract

Background: Patient-reported outcome and experience measures can play a critical role in providing patient-centered and value-based health care to a growing population of patients who are chronically ill. Value-based telemedicine platforms such as the Naveta initiative may facilitate the effective integration of these tools into health care systems.

Objective: This study aims to evaluate the response rate to electronic patient-reported outcome measures (ePROMs) and electronic patient-reported experience measures (ePREMs) among patients participating in the Naveta telemedicine initiative and its correlations with sociodemographic and clinical characteristics, as well as the evolution of the response rates over time.

Methods: Between January 1, 2021, and June 30, 2023, a total of 53,364 ePREMs and ePROMs for 20 chronic conditions were administered through the Naveta-Phemium platform. Descriptive statistics were used to summarize continuous and categorical variables. Differences in response rates within each sociodemographic variable were analyzed using logistic regression models, with significance assessed via chi-square and post hoc Tukey tests. Two-way ANOVA was used to examine the interaction between time interval and disease type on response rate evolution.

Results: A total of 3372 patients with severe chronic diseases from 64 public hospitals in Spain participated in the Naveta health questionnaire project. The overall response rate to ePROMs and ePREMs during the first 2.5 years of the Naveta initiative was

46.12% (24,704/53,364), with a baseline rate of 53.33% (7198/13,496). Several sociodemographic factors correlated with lower response rates, including male gender, older age, lower education level, frequent alcohol use, being a student, and not being physically active. There were also significant variations in response rates among different types of chronic conditions ($P<.001$), with the highest rates being for respiratory (433/606, 71.5%), oncologic (200/319, 62.7%), digestive (2247/3601, 62.4%), and rheumatic diseases (7506/12,982, 57.82%) and the lowest being for HIV infection (7473/22,695, 32.93%). During the first 6 months of follow-up, the response rates decreased in all disease types, except in the case of the group of patients with oncologic disease, among whom the response rate increased up to 100% (6/6). Subsequently, the overall response rate approached baseline levels.

Conclusions: Recognizing the influence of sociodemographic factors on response rates is critical to identifying barriers to participation in telemonitoring programs and ensuring inclusiveness in patient-centered health care practices. The observed decline in response rates at follow-up may be due to survey fatigue, highlighting the need for strategies to mitigate this effect. In addition, the variation in response rates across chronic conditions emphasizes the importance of tailoring telemonitoring approaches to specific patient populations.

(JMIR Mhealth Uhealth 2024;12:e56196) doi:10.2196/56196

KEYWORDS

chronic conditions; eHealth; value-based care; patient-reported outcome measures; patient-reported experience measures; questionnaires; response rate; telemedicine platform

Introduction

Background

Health care systems and institutions worldwide are undergoing transformation to meet the demands of an aging population, implement new health care standards, and integrate advancements in digital technology [1-4]. Increased life expectancy and improvements in medicine have led to a growing number of people living with chronic conditions [5-7]. This growth represents a major challenge for health care systems, which must provide efficient care for patients with chronic conditions while wisely allocating limited resources [3,8]. To address these complex challenges, governments, health organizations, and health care providers are increasingly recognizing the need to transition toward more personalized and patient-centered care models, aligned with the concept of value-based health care (VBHC) [9-12]. In this new paradigm, value is defined as health outcomes that matter to patients in comparison to the cost of achieving these outcomes [9].

To carry out these changes, health care services and systems need tools to guide the transition [11,13]. Measuring, reporting, and comparing outcomes is essential to assess the value of care provided to patients [9,14,15]. Patient-reported measures are the fundamental tools used to guide health systems and providers in implementing person-centered health care and achieving outcomes that matter to patients [16,17]. Patient-reported measures are administered in the form of questionnaires and include patient-reported outcome measures (PROMs) and patient-reported experience measures (PREMs). PROMs can assess a variety of outcomes, such as physical performance, social functioning, psychological well-being, symptom severity, disability, and impairment, from the perspective of the patient, whereas PREMs focus on the patient's experience of care [18]. These measures can be used to support diagnosis, monitor treatment and patient progress, improve communication between patient and health care professionals, and facilitate shared decision-making [17,19]. Despite their potential value and some examples of successful health system–level PREM and PROM programs, such as the National Health Service PROMs program in the United Kingdom and the Danish patient-reported outcome (PRO) system, their routine use in health systems is not widespread in other countries [11,13,16,18,20-22].

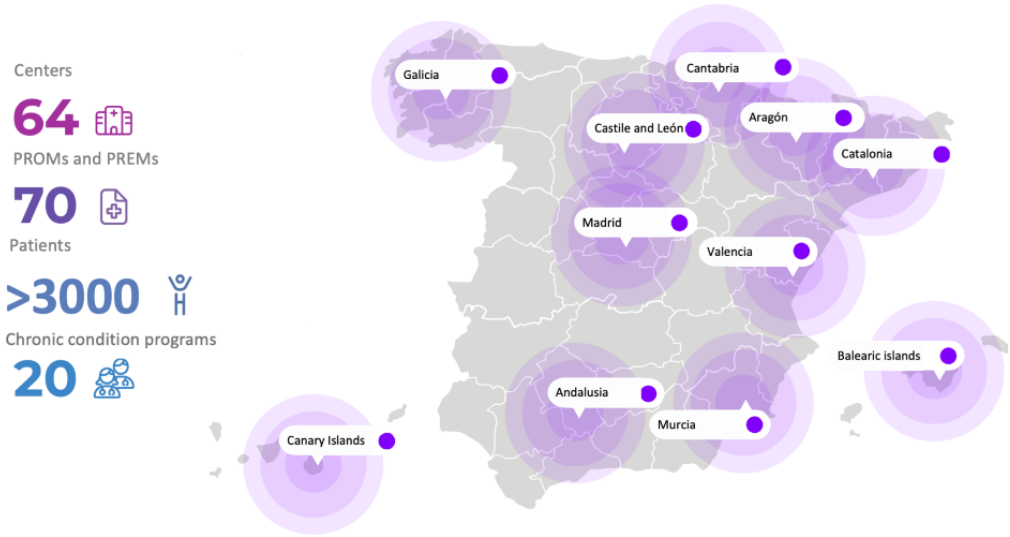
The Naveta Value-Based Telemedicine Initiative

Digital transformation and the integration of eHealth tools and value-based IT platforms in health care settings are critical for achieving the systematic delivery of standardized PREMs and PROMs in clinical practice [11,23,24]. The COVID-19 pandemic has sparked a revolution in digital health technologies that can pave the way for this transformation [4,25]. In this context, the Naveta initiative [26] has emerged as a scientific community that aims to promote VBHC for patients with serious chronic conditions by implementing processes and strategies related to the electronic delivery of PROMs and PREMs and the use of telemedicine tools.

Naveta was created by the Association of Pharmacists of Outpatient Departments of the Balearic Islands (FARUPEIB) with the participation of BiblioPRO, a web-based library of PRO questionnaires in Spanish [27,28]. The technological infrastructure that supports Naveta activities is provided by Phemium [29], a software platform specialized in telemedicine projects. The scientific committee of Naveta is currently composed of physicians, hospital pharmacists, nurses, psychologists, members of BiblioPRO, and experts in eHealth and innovation. The committee's responsibilities include selecting PROM standard sets, determining the frequency and timing of questionnaire administration, evaluating and improving the efficiency of the platform and standard sets, promoting VBHC, mentoring and advising health care professionals and researchers, and fostering new technological developments within the Naveta ecosystem. To date, specific and generic PROM standard sets have been selected for 20 chronic diseases. These measures can assess, among other outcomes, health status, disease activity, impact on quality of life, treatment satisfaction, and productivity from the patient's perspective.

One of the goals of the Naveta initiative is to provide state-of-the-art PROM and PREM tools that enable health care managers and professionals to obtain accurate and valuable information directly from patients to be incorporated into the decision-making process. Another goal of the initiative is to build a body of knowledge on the appropriate use of PROMs, considering both disease type and sociodemographic data, in line with initiatives such as Patient-Reported Outcomes Tools: Engaging Users and Stakeholders (PROTEUS) and the International Consortium for Health Outcomes Measurement (ICHOM) [30,31]. Recognized recently by different Spanish associations as an outstanding innovation project, Naveta is one of the most widespread value-based telemedicine initiatives in Spain [32] and has been widely implemented at hospitals and institutions in different Spanish regions (Figure 1). Furthermore, several observational studies on patients with various chronic diseases are currently being conducted using the Naveta VBHC approach.

Figure 1. Distribution of Spanish hospitals participating in the Naveta initiative. PREM: patient-reported experience measure; PROM: patient-reported outcome measure.



Goal of the Study

Here, we present our experience in implementing the Naveta value-based telemedicine initiative at different hospitals in Spain, which involved the administration of 53,634 electronic PROMs (ePROMs) and electronic PREMs (ePREMs) to 3372 patients with severe chronic conditions over the last 2.5 years.

The primary objective of this study was to assess the response rate to electronic questionnaires among patients who were chronically ill registered in Naveta and to examine its correlation with various sociodemographic characteristics. In addition, this study aimed to track the evolution of response rates over time, specifically among patients with different types of chronic diseases. The purpose of this analysis was to identify potential biases in the collection of information that might have prevented the comprehensive monitoring of the maximum number of patients.

Methods

Study Design and Sample Selection

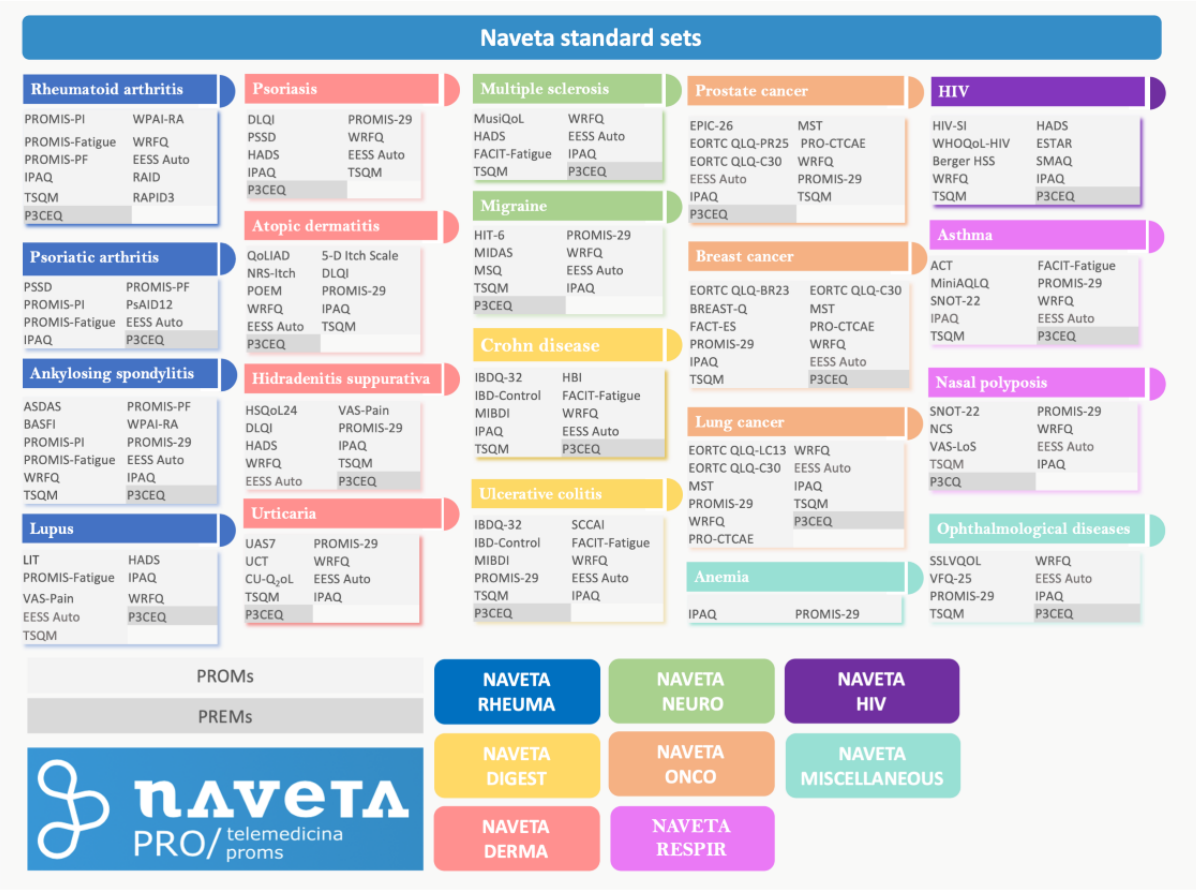
This study was carried out in the context of routine clinical practice in which patients were receiving dual care (on-site plus telematic care with the Naveta platform). We conducted an observational retrospective multicenter study involving patients with severe chronic diseases. These patients were regularly receiving hospital-only medicines at pharmacy outpatient clinics of 64 public hospitals in different regions of Spain (Figure 1). Patients were enrolled at the time of starting or changing their prescribed outpatient medication and registered for ≥1 chronic condition programs based on their diagnosis. The study covered the period from January 1, 2021, to June 30, 2023.

All patients met the following criteria: aged ≥18 years, diagnosed with ≥1 chronic diseases (psoriasis, psoriatic arthritis, rheumatoid arthritis, asthma, cancer, ulcerative colitis, atopic dermatitis, Crohn disease, multiple sclerosis, ankylosing spondylitis, hidradenitis suppurativa, urticaria, and HIV infection, among others), and agreed to participate in a telemedicine project linked to health questionnaires through the Naveta-Phemium platform.

Selection of Patient-Reported Measures

The questionnaires included in each disease standard set, selected by Naveta's scientific committee, are specified in Figure 2. The color code in the figure indicates the type of medical condition: rheumatic diseases, gastrointestinal diseases, neurological diseases, skin diseases, oncologic diseases, respiratory diseases, HIV infection, and other diseases.

Figure 2. Naveta's standard sets of patient-reported measures for 20 chronic conditions. ACT: Asthma Control Test; Ankyl: ankylosing; ARDs: Autoimmune Rheumatoid Diseases; ASDAS: Ankylosing Spondylitis Disease Activity Score; BASFI: Bath Ankylosing Spondylitis Functional Index; Berger HSS: Berger HIV Stigma Scale; CU-Q2oL: Chronic Urticaria Quality of Life Questionnaire; DLQI: Dermatology Life Quality Index; EESS auto: self-reported side effects; EORTC: European Organization for Research and Treatment of Cancer; EORTC QLQ-BR23: EORTC Quality of Life Questionnaire for Breast Cancer; EORTC QLQ-C30: EORTC Core Quality of Life questionnaire; EORTC QLQ-LC13: EORTC Quality of Life Questionnaire module for Lung Cancer patients; EORTC QLQ-PR25: EORTC Quality Of Life Questionnaire Module For Prostate Cancer 25; EPIC-26: Expanded Prostate Cancer Index Composite-26; FACIT-Fatigue: Functional Assessment of Chronic Illness Therapy – Fatigue Scale; FACT-ES: Functional Assessment of Cancer Therapy for patients with Endocrine Symptoms; HADS: Hospital Anxiety and Depression Scale; HBI: Harvey-Bradshaw Index; Hidra.: Hidradenitis; HIT-6: Headache Impact Test -6; HSQoL24: Hidradenitis Suppurativa Quality of Life 24; HIV SI: HIV Symptom Index; IBD Control: Inflammatory Bowel Disease Control Questionnaire; IBDQ32: Inflammatory Bowel Disease Questionnaire; IPAQ: International Physical Activity Questionnaire; LIT: Lupus Impact Tracker; MIBDI: Manitoba Inflammatory Bowel Disease Index; MIDAS: Migraine Disability Assessment; mini AQLQ: Mini Asthma Quality of Life Questionnaire; MSQ: Migraine-Specific Quality-of-Life Questionnaire; MST: Malnutrition screening tool; MusiQoL: multiple sclerosis international quality of life; NCI's PRO-CTCAE: National Cancer Institute's Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events; NCS: Nasal Congestion Score; NRS: Numeric Rating Scale; Ophtalmol: Ophthalmological; PROMIS: Patient Reported Outcome Information System; PROMIS-PF: PROMIS Physical Function; PROMIS-PI: PROMIS Pain Intensity; PsAID12: 12-item Psoriatic Arthritis Impact of Disease; PSSD: Psoriasis Symptoms and Signs Diary; P3CEQ; Person-Centred Coordinated Care Experience Questionnaire; QoLIAD: Quality of Life Index for Atopic Dermatitis; RAID: Rheumatoid Arthritis Impact of Disease; RAPID3: Routine assessment of patient index data 3 ; SCCAI: Simple Clinical Colitis Activity Index; SLVQOL: Spanish low vision quality of life questionnaire; SMAQ: Simplified Medication Adherence Questionnaire; SNOT-22: Sino-nasal outcome test 22; TSQM: Treatment Satisfaction Questionnaire for Medication; UAS7: Urticaria activity Score 7; UCT: Urticaria Control Test; VAS: Visual Analogue Scale; VAS-LoS: VAS for loss of smell; VFG-25: Visual Functioning Questionnaire 25; WHOQoL–HIV: World Health Organization's Quality of Life HIV instrument; WPAI-RA: Work Productivity in Rheumatoid Arthritis; WRFQ: Work Role Functioning Questionnaire.



The scientific committee that selects Naveta's standard sets of PROMs for each pathology is composed of physicians, hospital pharmacists, psychologists, members of BiblioPRO, and eHealth experts.

Their decisions are informed by the study of literature reviews and recommendations from authoritative sources such as the Patient-Reported Outcomes Measurement Information System (PROMIS) [33], the ICHOM initiative [34], and the Patient-Reported Indicator Surveys initiative of the Organisation for Economic Co-operation and Development [35]. The questionnaires used are the validated Spanish versions and meet licensing requirements. The platform mainly included PROMS for a variety of chronic conditions (67), and only a few PREMs were administered, including the Person-Centered Coordinated Care Experience Questionnaire (P3CEQ), which is designed to assess person-centered coordinated care from the perspective of people with long-term conditions [36,37]. Other PREMs included the Self-Injection Assessment Questionnaire (SIAQ)

[38,39] and a Likert scale survey to assess satisfaction with the dual follow-up (on-site and telematic care; administered annually).

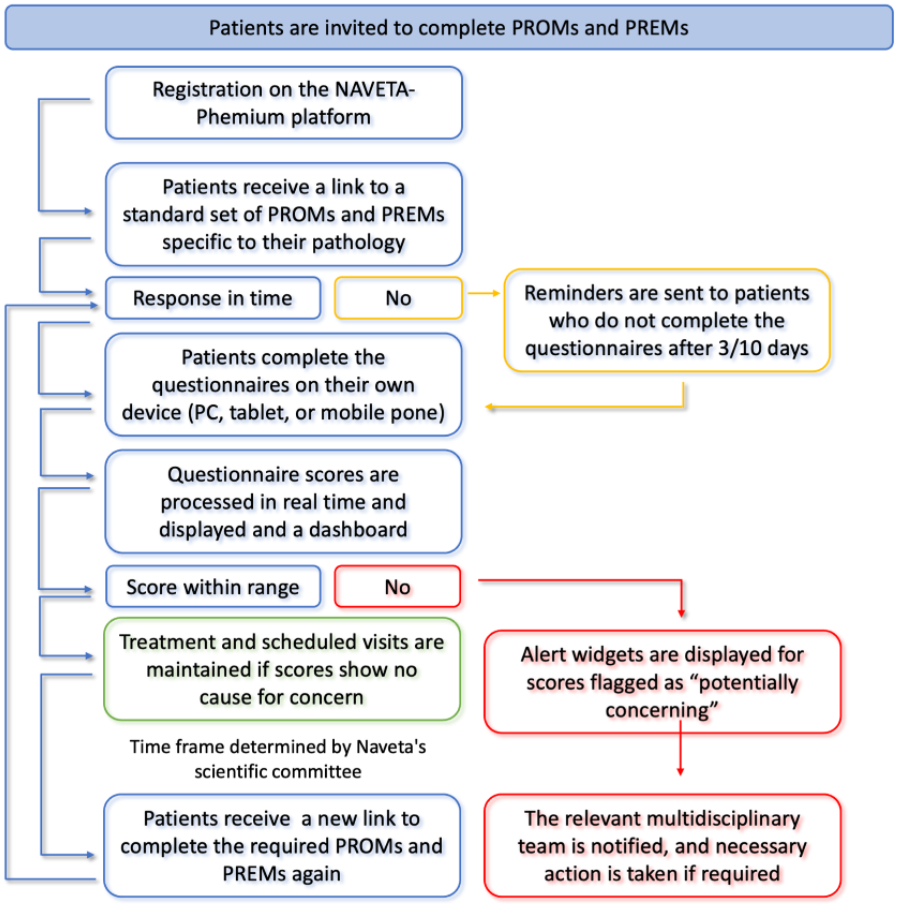
Data Collection Procedures

Patients who provided informed consent (refer to the Ethical Considerations subsection for details) received a link to access the electronic questionnaires, which they could complete on their own device (PC, tablet, or mobile phone). Each patient was provided with a standard set of PROMs tailored to their specific chronic disease. Participants were also asked to complete a sociodemographic questionnaire that included questions about their level of education, employment status, smoking habit and alcohol consumption, BMI, and physical activity. The BMI parameter was categorized according to the World Health Organization classification: underweight (<18.5 kg/m^2), healthy weight (18.5-24.9 kg/m^2), overweight (25.0-29.9 kg/m^2), and obese (≥30 kg/m^2) [40]. After 1 year of follow-up with the Naveta telemedicine program, patients were administered a 10-point Likert scale questionnaire to assess their satisfaction, and this survey was repeated annually.

The process of delivering and collecting PROMs and PREMs through the Naveta-Phemium platform is illustrated in Figure 3. The system automatically generates and sends reminders (via SMS text messaging and email) to patients who do not complete the questionnaires 7 and 15 days after receiving the link. It also generates alert widgets for potentially concerning scores and enables users to specify the score categories that should be flagged for each PROM (eg, scores indicating moderate to severe impairment). These alerts, along with other data from the patient's medical record, are used to make clinical decisions about medication changes, specialist referrals, lifestyle recommendations, and more. The patients were requested to complete the relevant standard set of questionnaires at the start of each pharmacological treatment or at each medication change (baseline collection). The periodicity of the follow-up questionnaires was determined by the Naveta scientific committee according to the specific chronic conditions and the outcome measured.

The team at each center was responsible for seeking permission to use the questionnaires through the Naveta initiative.

Figure 3. Process of delivering and collecting patient-reported measures through the Naveta-Phemium platform. PREM: patient-reported experience measure; PROM: patient-reported outcome measure.



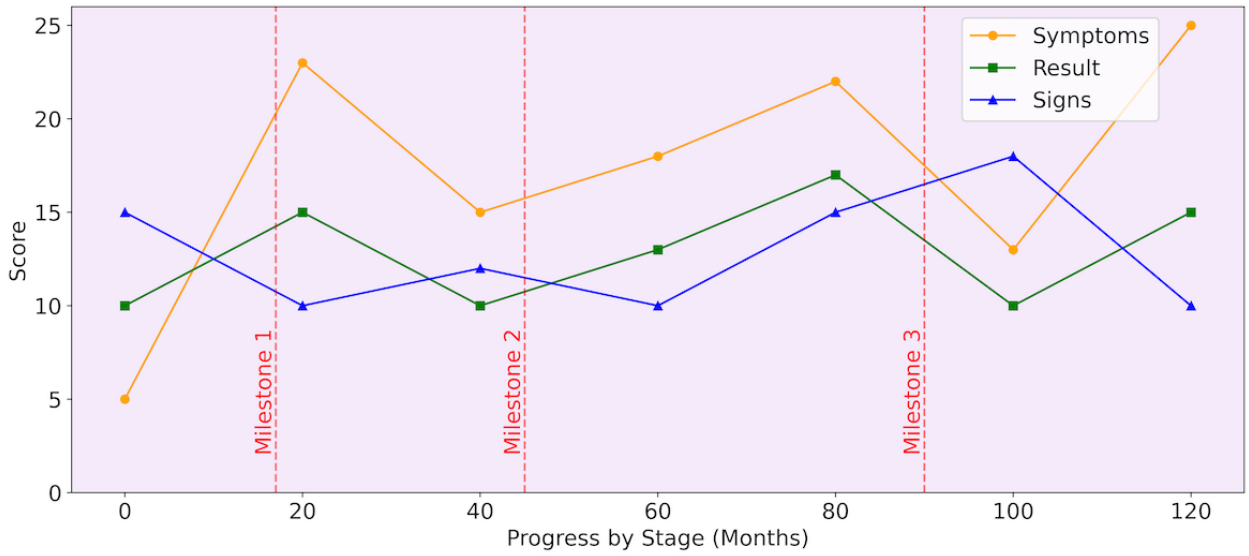
Technology Infrastructure

The Naveta initiative is supported by Phemium [29], a technological multiplatform specialized in the design and implementation of telemedicine projects. This platform has >2000 active professional users in different countries (Spain, Poland, the Czech Republic, Germany, Chile, Peru, Mexico, and Nigeria), with >2 million health care interactions per year globally. It complies with the requirements of a high standard

of data protection, level 3, according to the European General Data Protection Regulation and Spanish Organic Law 3/December 5, 2018, regarding the protection of personal data and the guarantee of digital rights.

The platform facilitates effective communication between health care professionals and patients. To systematize the delivery and interpretation of PROMs and PREMs in health care practice, the system performs the parameterization and customization of protocols. Each PROM is labeled for ≥1 specific conditions or as a generic questionnaire. The system interprets patient responses according to the score intervals of each questionnaire. Responses that do not meet the conditions required for a particular questionnaire are automatically discarded. The results of the questionnaires are processed in real time and presented graphically on a dashboard (Figure 4). This allows health care professionals (nurses, hospital pharmacists, and physicians) to immediately monitor patient outcomes and compare their progress with that of other patients and to monitor the evolution of an individual patient in a clinical context. The system also allows global analysis within groups for each pathology and compares individual and global results. In addition, patients can access their results on their dashboard.

Figure 4. Psoriasis Symptoms and Signs Diary: an example of a patient-reported outcome measure dashboard. This dashboard monitors 3 distinct metrics—symptoms, results, and signs—over a 120-month time frame. Notably, it features milestone markers at roughly 17, 45, and 90 months, with dashed vertical lines denoting significant events in the management of the condition, such as weight loss, changes in alcohol consumption, or alterations in treatment, among potential examples.



Statistical Analysis

Descriptive statistics were used to summarize continuous variables as means and SDs and categorical variables as frequencies and percentages. For results analysis, chronic conditions were categorized based on their respective disease type (Figure 2). To assess the response rates over time, the number of completed questionnaires at distinct intervals was calculated. Differences in response rates among categories within each sociodemographic variable were analyzed using logistic regression models. Their significance was evaluated with chi-square and post hoc Tukey tests (using the *glht* function from the *multcomp* R package). A 2-way ANOVA was used to study the interaction effect between time interval and disease type on response rate evolution. All statistical procedures were performed using R software (version 4.2.2; R Foundation for Statistical Computing) [40], and a significance level of $P \leq .05$ was set for all tests.

Ethical Considerations

Ethics approval was obtained from the research ethics committee of the Balearic Islands (Comitè d'ètica de la investigació de les Illes Balears; IB 5117/23 EOm). This study was conducted in accordance with the principles of the Declaration of Helsinki and its subsequent revisions. The informed consent form and other patient information were reviewed and approved by the clinical research ethics committee and the independent ethics committee of Balearic Islands (Palma de Mallorca, Spain).

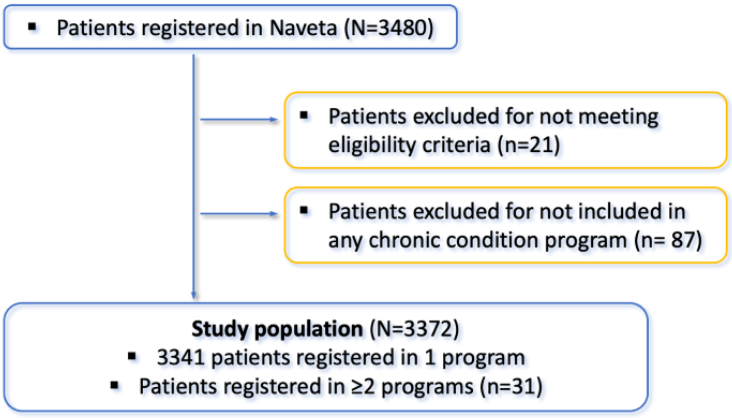
Before enrollment in the Naveta project, all patients received detailed information and provided written informed consent.

Results

Descriptive Analysis of the Sample

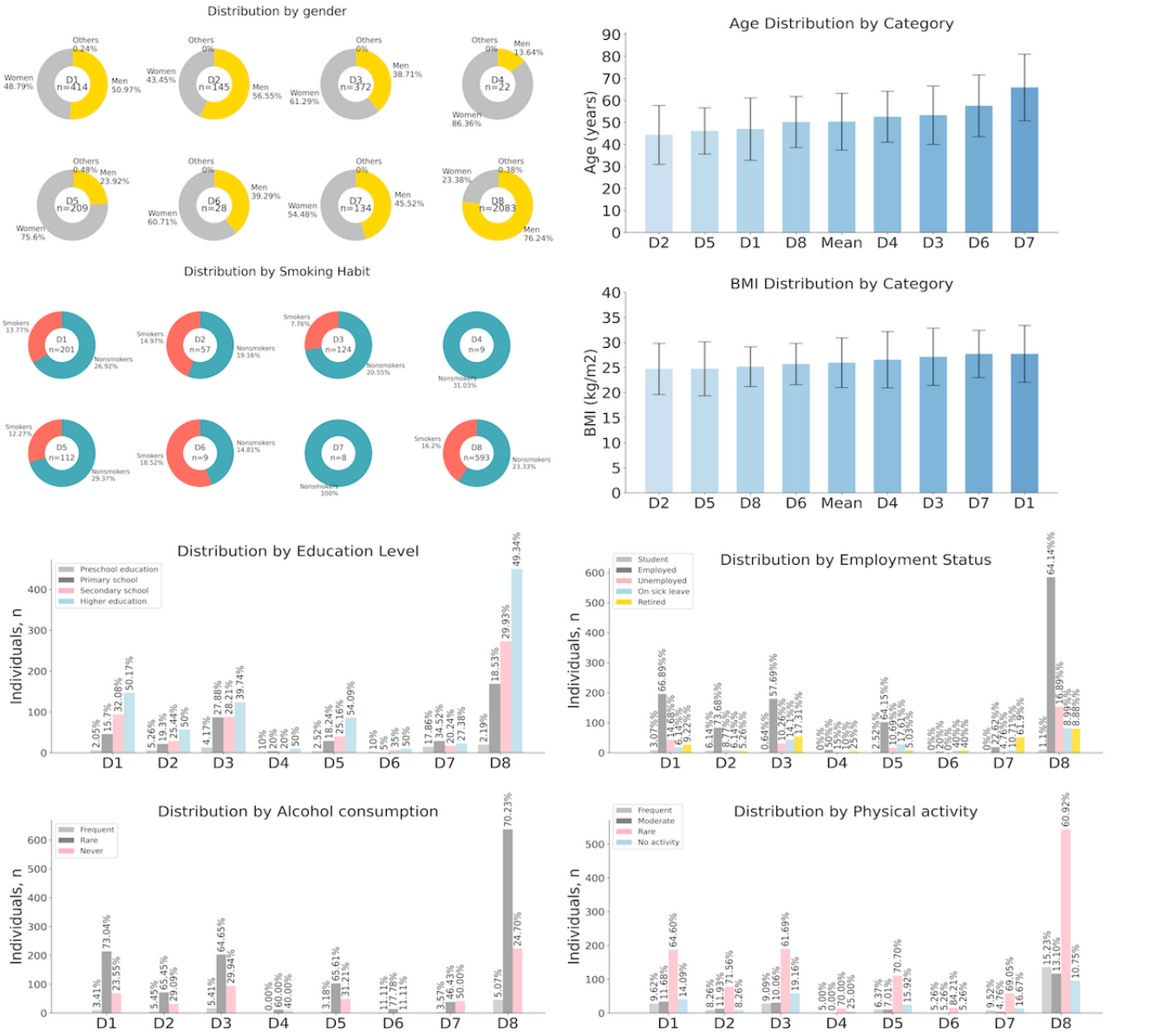
In total, 3372 patients diagnosed with at least 1 chronic condition were registered on the Naveta-Phemium platform: 2128 (63.11%) men, 1234 (36.6%) women, and 10 (0.3%) people who self-identified as *other gender*. Of the 3372 patients, 3341 (99.08%) were registered in only 1 chronic condition program, while 31 (0.92%) had multiple diagnoses and were registered in ≥2 chronic condition programs (n=28, 90% were registered in 2 programs; n=2, 6% in 3 programs; and n=1, 3% in 4 programs; Figure 5). The mean age of the whole cohort was 50.36 (SD 12.91) years.

Figure 5. Flowchart for participant recruitment.



The HIV disease program had the highest number of registered patients (2083/3372, 61.77%), whereas the breast cancer (17/3372, 0.5%), prostate cancer (9/3372, 0.27%) and lung cancer (2/3372, 0.06%) programs had the lowest number of registered patients (Table S1 in Multimedia Appendix 1). Within the disease type categories, the largest group was Naveta HIV (2083/3372, 61.77%), followed by Naveta Derma (414/3372, 12.28%), Naveta Rheuma (372/3372, 11.03%), Naveta Miscellaneous (134/3372, 3.97%), Naveta Neuro (209/3372, 6.2%), Naveta Digest (145/3372, 4.3%), Naveta Onco (28/3372, 0.83%), and finally Naveta Respir (22/3372, 0.65%). Figure 6 and Table S2 in Multimedia Appendix 1 present the main sociodemographic characteristics of the study participants by type of disease.

Figure 6. Sociodemographic characteristics of study participants by disease type (a comprehensive analysis of demographic distribution and health habits across different disease categories). The charts illustrate the distribution—among individuals with dermatological, digestive, respiratory, neurological, oncologic, HIV-related, and miscellaneous conditions—by (A-D) gender, age, BMI, and smoking habits; and (E-H) education level, employment status, alcohol consumption, and physical activity. D1: dermatological diseases; D2: digestive diseases; D3: rheumatological diseases; D4: respiratory diseases; D5: neurological diseases; D6: oncological diseases; D7: miscellaneous diseases; D8; HIV.



Descriptive Analysis of the Questionnaires

A total of 53,364 questionnaires were sent out for all disease categories, and 24,704 (46.12%) were completed. The response rates (percentage of completed questionnaires out of the total number of questionnaires sent) at specific time intervals were as follows: 53.33% (7198/13,496) at baseline, 41.5% (9131/22,000) from baseline to 6 months, 43.7% (4474/10,239) from 6 months to 1 year, and 51.13% (3901/7629) from 1 year to 2.5 years (Figure 7; Table S3 in Multimedia Appendix 1).

The analysis revealed a significant interaction effect between time interval and disease type on response rate evolution over time (P<.001). In all disease categories, the response rate decreased after 6 months, except in the category of oncologic diseases, where the response rate increased up to 100% (6/6). In the following intervals, the evolution of the response rate did not show a homogeneous trend in the different disease categories. It should be noted that the number of questionnaires sent to patients for completion differed based on the type of disease (Figure 8).

Figure 7. Evolution of response rate by disease program.

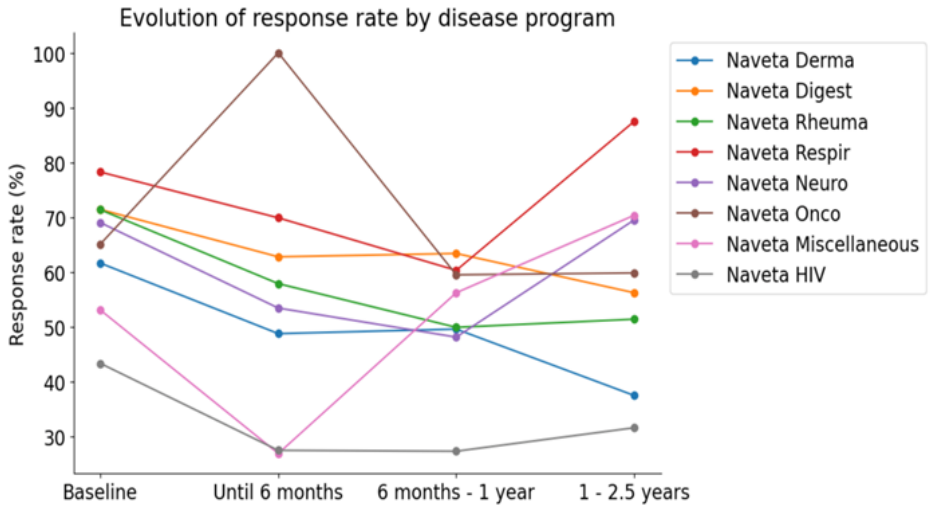
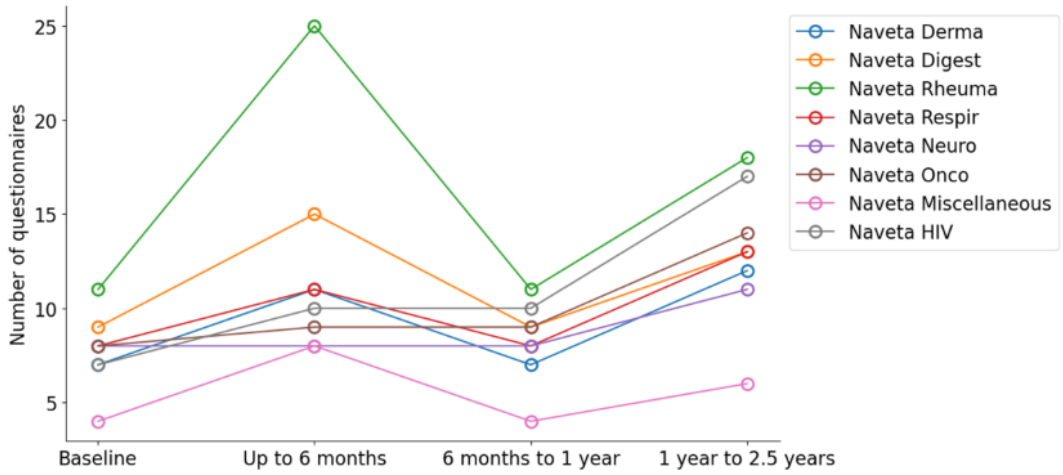
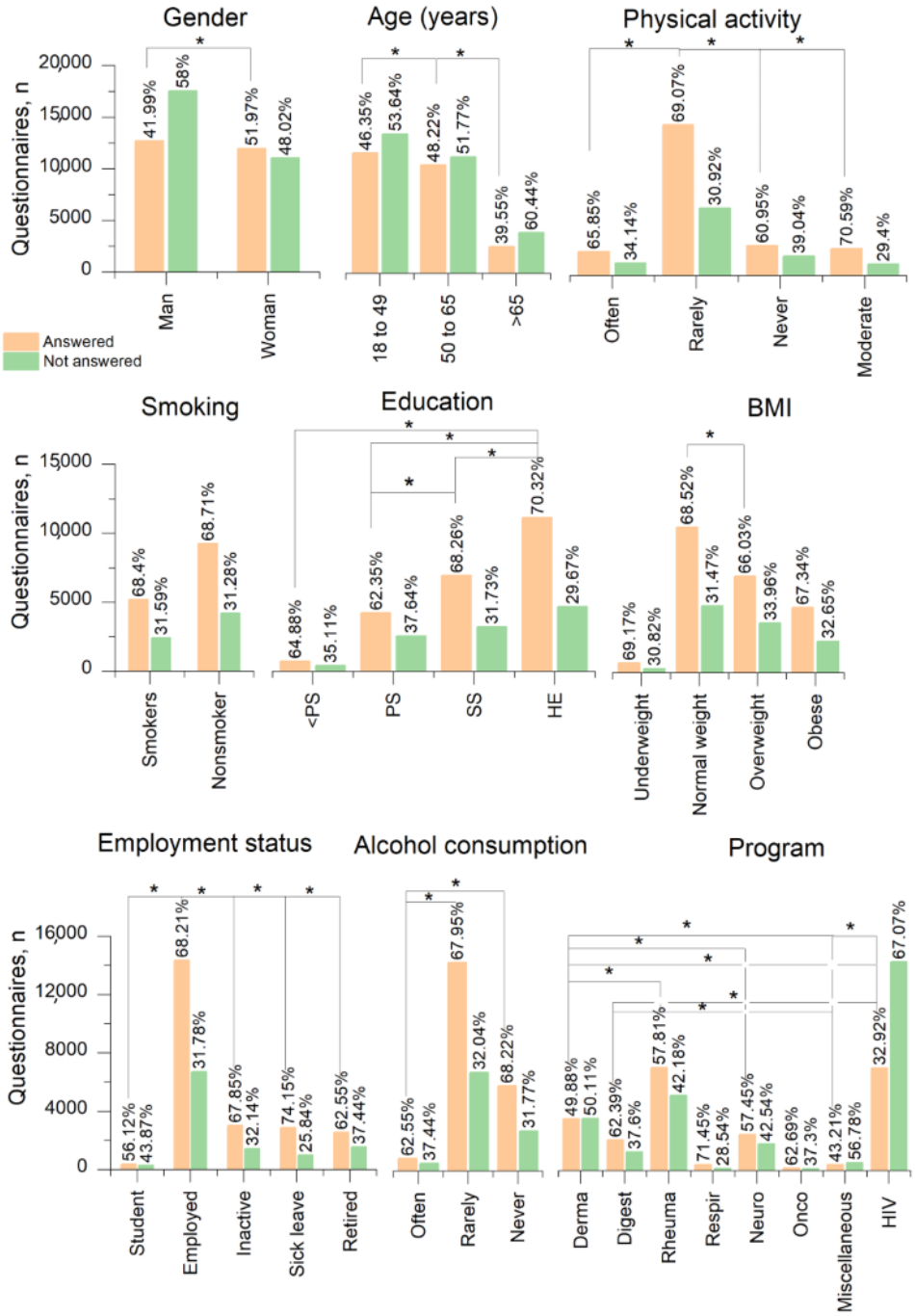


Figure 8. Number of questionnaires sent out by disease program.



Significant differences were found for gender, age, education level, employment status, alcohol consumption, and physical activity. Women were more likely to complete questionnaires than men (P<.001), and patients aged >65 years had a lower response rate than younger patients (P<.001). Higher education correlated with a higher frequency of completed questionnaires compared to the categories of preschool education (P<.001), primary school education (P<.001), and secondary school education (P=.002). Regarding employment status, patients on sick leave had the highest response rate (P<.001), and students had the lowest response rate (P<.001) compared to the other categories. Finally, when comparing the categories of alcohol consumption and physical activity, a lower frequency of completed questionnaires was found for frequent alcohol consumption (P<.001) and no physical activity (P<.001; Figure 9).

Figure 9. Frequencies of answered and unanswered questionnaires compared across different sociodemographic factors. *At post hoc level, the difference in the means is significant, *P*<.05. HE: higher education; PS: primary school; SS: secondary school.



For disease type, significant differences among the categories were observed (Figure 9; Table S4 in Multimedia Appendix 1; refer to Table S5 in Multimedia Appendix 1 for frequencies of answered and unanswered questionnaires by disease program). The highest response rates were found for Naveta Respir, Naveta Digest, Naveta Rheuma, and Naveta Onco; and the lowest response rates were associated with the patients who had tested positive for HIV infection and those included in Naveta Miscellaneous.

In terms of satisfaction with the telemonitoring program via Naveta, 87.7% (381/434) of the patients rated their experience with scores between 7 and 10 (n=197, 45.4% provided a rating of 9 or 10; and n=184, 42.3% provided a rating of 7 or 8). The analysis in the different clinical sets did not show significant differences among them in the level of satisfaction (*P*=.63). A third of the most satisfied patients (61/197, 30.9%; rating of 9 or 10) belonged to the group categorized as Naveta Rheuma.

Discussion

Principal Findings

The Naveta initiative aims to promote the principles of VBHC and improve the quality of care for people with chronic diseases through the systematic integration of ePROMs and ePREMs

into clinical practice settings and a continuous learning process based on result analysis. The routine collection of PROMs and PREMs enables health care providers to assess treatment effectiveness, identify unrecognized issues, and positively impact patient management and satisfaction [41-43]. High response rates to PROMs have been associated with better health outcomes and are important to ensure that the information collected is representative [21,44]. The overall response rate to ePROMs and ePREMs during the first 2.5 years of Naveta was 46.12% (24,704/53,364), with a baseline rate of 53.33% (7198/13,496). In countries with a long tradition of systematic PROM collection, such as the United Kingdom, Denmark, and the United States, response rates might range from 50% to 80% or higher [45-49]. However, comparing response rates among studies is challenging due to the lack of a standardized calculation formula [44,48]. Understanding the variables that influence these rates is fundamental to developing effective strategies to improve patient engagement. We consider that with the publication of similar studies and with the implementation of these systems in more regions, comparable data will be available, and cutoff points could be established. The association of response rate with common clinical surrogate markers would also be necessary.

These data could also be useful in future studies to assess the adequacy of PROMs and PREMs for different profiles of patients and their usefulness in clinical practice and future integration in health services databases.

Our study identified several sociodemographic factors that seemed to negatively impact response rates, including male gender, older age, lower education level, frequent alcohol consumption, being a student, and no physical activity. Women in our cohort exhibited higher response rates than men, a pattern already observed in other studies, although not consistently [44,50]. Further investigation is warranted to confirm this observation and to determine why men may be less motivated to answer health questionnaires. Older age was also significantly associated with lower response rates compared to younger age, and this is a common pattern in previous studies [44,50,51]. Possible reasons for this association include lower levels of digital confidence and skills as well as language and literacy skills, visual impairment, and health problems [50,51]. The higher response rates observed among individuals with higher levels of education in our study align with those in other studies examining barriers to completing ePROMs [52,53].

Lower education levels are often associated with lower health literacy and digital skills [52]. To the best of our knowledge, limited research has been conducted on the impact of employment status, physical activity, and alcohol consumption on PREM and PROM response rates [54,55].

The potential for bias in response rates associated with certain sociodemographic characteristics should be given due consideration and thoroughly examined to implement appropriate compensatory measures. To make the use of PROMs more accessible to a wider range of people, particularly to older and less educated patients, it could be necessary to offer some assistance (from a nurse, data manager, family members, or caregiver). This assistance could help, for example, to complete electronic questionnaires and promote a better understanding of the associated benefits [19,53].

It should be noted that, according to the study by Eriksen et al [21], PROMS equip patients with an improved understanding of their conditions, treatment, and health and increase their awareness of psychosocial issues and symptoms by encouraging disease-related reflections. However, in some cases, particularly patients with some complex conditions, the anxiety stemming from reporting symptoms or being reminded of their deteriorating health status can be a reason for not participating in such health-related survey [56].

Patients' increased access to, and accumulation of, knowledge through digital platforms allow mobilizing and empowering them and might affect their expectations and their political engagement related to the quality of health care [57].

Naveta's scientific committee has so far selected 20 standard sets for different chronic diseases, comprising 70 different PROMs and PREMs. Significant variations in response rates have been observed among different types of chronic conditions, with the highest rates being for respiratory, digestive, rheumatic, and oncologic diseases and the lowest rates being for HIV infection. Studies comparing response rates to ePROMs across different chronic diseases, especially when using condition-specific standard sets of questionnaires, are scarce. In a study describing the implementation of the WestChronic PRO system in Denmark, which included 22 PRO projects for 18 chronic diseases, the highest response rates were observed in patients with epilepsy and prostate cancer, whereas the lowest rates were observed in patients with stroke [58]. A review of response rates available in clinical quality registries and databases that collect PROMs for all types of health conditions found that patients with chronic conditions and cancer have the highest response rates at baseline [48]. The differences observed in response rates in our study may be related to the nature of the specific conditions. However, it would be necessary to determine the role of disease-specific PROMs and other factors such as the severity of symptoms and the presence of comorbidities [59,60]. Further research is necessary to fully understand these differences and to implement strategies to improve response rates across pathologies.

During the first 6 months of follow-up, response rates declined across all disease types, except in the case of the group of patients with oncologic disease, where a notable increase was observed. Subsequently, the overall response rate increased and approached baseline levels, and the response rate pattern in the group of patients with oncologic disease became comparable to the response rate patterns in the other groups. It should be noted that the number of questionnaires that patients were requested to complete varied over time in each group. This initial increase in response rates in patients with oncologic disease has not been described in other studies and should be investigated in a larger cohort to elucidate potential underlying causes [49,61,62]. One possible explanation is the support provided by the cancer functional units, which include multidisciplinary professionals such as oncologists, surgeons, pharmacists, nurses, psychologists, nutritionists, and physiotherapists; nurses, in particular, play an important role

in motivating patients to respond to PROMs. However, as side effects become more pronounced or accumulate over time, patient motivation may diminish.

The overall decline in response rates at follow-up is a common tendency in numerous studies, with only a few exceptions [44,48,49,58,60,63,64]. This decline has been attributed to survey fatigue, which is associated with factors such as response burden, questionnaire length, perceived item irrelevance, older age, disease severity, and the presence of comorbidities [59,64,65]. When selecting standard sets and the frequency of questionnaire administration for each pathology, it is crucial to balance the need to obtain comprehensive information with the potential burden on patients if they have to complete numerous questionnaires. To address survey fatigue and enhance response rates, Naveta's scientific committee is currently exploring new strategies inspired by the PROMIS Computer Adaptive Tests and the studies that have investigated patient burden [59,65-67]. These strategies include reducing the number of PROMs in standard sets; labeling some of them as optional; using shorter questionnaires; and reducing, delaying, or spreading over time the administration of questionnaires.

Unlike the traditional use of PROMs in clinical practice as a supplement to the patient's follow-up, PROM-based follow-up with AmbuFlex and Naveta platforms represents a new model of service delivery where PROMs are used as the basis for outpatient follow-up [68].

Implementing the routine collection of PROMs and PREMs in clinical settings involves integrating technological systems for electronic questionnaire administration [69]. The advantages of ePROMs and ePREMs over paper questionnaires include reduced data entry errors, shorter completion times, real-time results, improved symptom management, the prioritization of clinical visits, and cost reduction [19,70]. In addition, automated alerts enable health care professionals to promptly address serious complications, reducing symptom burden and the risk of hospital readmissions [19]. Challenges associated with electronic questionnaires include privacy concerns, a higher initial investment in infrastructure, and potential bias due to patients' lack of digital literacy [19].

Positive feedback from patients regarding PROM monitoring via the Naveta system was demonstrated by the results of the satisfaction questionnaire. Aligned to our commitment to improve the health and well-being of patients with chronic diseases and guided by the principles of VBHC and right care [8,9], our initiative aims to ensure that digitalization goes hand in hand with humanization. Examples of this approach include health care professionals personally contacting patients in response to flagged alerts identified through questionnaire results or establishing communication channels for consultations. Recently, to help bridge the digital gap, data managers have been introduced to contact patients who do not complete questionnaires. Their role involves understanding the reasons for noncompletion, clarifying doubts, emphasizing the benefits of completing PROMs, and offering assistance to respond if desired. These actions reinforce our commitment to a patient-centered model that embraces technology while reaffirming the human connection.

Moreover, we believe that it is important to involve patients in the design of health technology. Effective patient engagement can profoundly change how patient-centered research is conceptualized and conducted, resulting in better patient-centered care, management, and measurement [71]. In our case, the feedback obtained in the satisfaction surveys is taken into consideration for making improvements to the Naveta platform. While, currently, the perspective of patients is indirectly obtained through contacts with patients' associations, we plan in the future to include patient representatives in the meetings of the scientific committee.

Strengths and Limitations

This study provides empirical evidence on the implementation of a telemonitoring system for patients who are chronically ill through electronic questionnaires at different hospitals and in different regions in Spain. Our comprehensive examination of multiple sociodemographic factors that influence the response rate to ePROMs and ePREMs provides valuables insights into existing literature. This study is one of the few that has compared response rates to ePROMs across different types of chronic diseases when specific standard sets of questionnaires were administered for each disease. Our findings contribute to the current understanding of the strategies needed to implement the systematic collection of ePROMs and ePREMs for patients with chronic conditions in clinical practice according to the principles of value-based care. We present a feasible model of telemedicine, which allows the remote clinical follow-up of a high number of patients with chronic conditions, helped by real-time patient-reported measures, that could substantially change the basis of current clinical management and facilitate shared decision-making, with substantial improvements in comparison with traditional on-site care.

However, 2 limitations should be noted. First, the sample size in certain disease categories (eg, oncologic diseases) was relatively small due to the recent introduction and limited adoption of the Naveta initiative in certain settings. The low number of patients in certain disease groups and the high number of missing values in certain sociodemographic categories reduced the statistical power of the study and did not allow us to perform more specific statistical analysis. Second, the retrospective design of the observational study meant that certain variables, such as the total number of questionnaires and variations in the frequency of questionnaire administration, were not systematically controlled in the analysis of the effect of different diseases on response rates and their evolution.

Conclusions

In conclusion, the overall response rate to the questionnaires during the first 2.5 years of the Naveta initiative was 46.12% (24,704/53,364), with several sociodemographic factors associated with lower response rates. There was a general trend of decreasing rates in the first 6 months, possibly indicating survey fatigue. Recognizing the impact of sociodemographic and clinical factors on response rates may help identify barriers for certain patient groups to participate in telemonitoring programs, suggesting the need to tailor telemonitoring approaches according to patient populations.

Acknowledgments

The authors want to thank Pilar Alba for her editorial assistance (financed by the Association of Pharmacists of Outpatient Departments of the Balearic Islands [FARUPEIB]) and INSA-UB Maria de Maeztu Unit of Excellence funded by MICIN/AEI/FEDER, UE (CEX2021-001234-M).

SR and NP are co-corresponding authors on this article.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Patients' clinical characteristics and response pattern.
[DOCX File , 55 KB - mhealth_v12i1e56196_app1.docx]

References

1. Harper S. The capacity of social security and health care institutions to adapt to an ageing world. Int Soc Secur Rev 2010 Sep 27;63(3-4):177-196. [doi: 10.1111/j.1468-246X.2010.01374.x]
2. Tuckson RV, Edmunds M, Hodgkins ML. Telehealth. N Engl J Med 2017 Oct 19;377(16):1585-1592. [doi: 10.1056/NEJMsr1503323] [Medline: 29045204]
3. Defining value in "valuebased healthcare". Publications Office of the European Union. 2019. URL: https://health.ec.europa.eu/system/files/2019-11/024_defining-value-vbhc_en_0.pdf [accessed 2023-12-06]
4. D'Anza B, Pronovost PJ. Digital health: unlocking value in a post-pandemic world. Popul Health Manag 2022 Feb;25(1):11-22. [doi: 10.1089/pop.2021.0031] [Medline: 34042532]
5. Brennan P, Perola M, van Ommen GJ, Riboli E, European Cohort Consortium. Chronic disease research in Europe and the need for integrated population cohorts. Eur J Epidemiol 2017 Sep;32(9):741-749. [doi: 10.1007/s10654-017-0315-2] [Medline: 28986739]
6. Souza DL, Oliveras-Fabregas A, Minobes-Molina E, de Camargo Cancela M, Galbany-Estragués P, Jerez-Roig J. Trends of multimorbidity in 15 European countries: a population-based study in community-dwelling adults aged 50 and over. BMC Public Health 2021 Jan 07;21(1):76 [FREE Full text] [doi: 10.1186/s12889-020-10084-x] [Medline: 33413239]
7. Reynolds J, Gadsby E, Rijken M, Stoop A, Espallargues M, Lloyd HM, et al. Measuring older peoples' experiences of person-centred coordinated care: experience and methodological reflections from applying a patient reported experience measure in SUSTAIN. Int J Integr Care 2021 Jul 13;21(3):3 [FREE Full text] [doi: 10.5334/ijic.5504] [Medline: 34305488]
8. Kleinert S, Horton R. From universal health coverage to right care for health. Lancet 2017 Jul 08;390(10090):101-102. [doi: 10.1016/S0140-6736(16)32588-0] [Medline: 28077231]
9. Porter ME. What is value in health care? N Engl J Med 2010 Dec 23;363(26):2477-2481. [doi: 10.1056/NEJMp1011024] [Medline: 21142528]
10. Walsh AP, Harrington D, Hines P. Are hospital managers ready for value-based healthcare? A review of the management competence literature. Int J Organ Anal 2019 Jul 24;28(1):49-65. [doi: 10.1108/IJOA-01-2019-1639]
11. Mjåset C, Ikram U, Nagra NS, Feeley TW. Value-based health care in four different health care systems. NEJM Catal Innov Care Deliv 2020 Nov 10:1-23. [doi: 10.1056/CAT.20.0530]
12. Teisberg E, Wallace S, O'Hara S. Defining and implementing value-based health care: a strategic framework. Acad Med 2020 May;95(5):682-685 [FREE Full text] [doi: 10.1097/ACM.0000000000003122] [Medline: 31833857]
13. Implementing value-based health care in Europe. EIT Health. 2020. URL: https://eithealth.eu/wp-content/uploads/2020/05/Implementing-Value-Based-Healthcare-In-Europe_web-4.pdf [accessed 2023-12-06]
14. Sanmartín-Fenollera P, Mangues-Bafalluy I, Talens-Bolos A, Ibarra-Barrueta O, Villamañán-Bueno E, Monte-Boquet E, et al. Telepharmacy scorecard: activity and quality indicators for the pharmaceutical care in a hospital pharmacy service. Farm Hosp 2022 Oct 30;46(7):92-105. [Medline: 36520564]
15. Morillo-Verdugo R, Collado-Borell R, Arrondo-Velasco A, Domínguez-Cantero M, Fernández-Polo A, González-Corominas E, et al. Implementation of pharmaceutical care through telepharmacy: a guide for professionals and patients. Farm Hosp 2022 Oct 30;46(7):115-122. [Medline: 36520566]
16. De Rosis S, Cerasuolo D, Nuti S. Using patient-reported measures to drive change in healthcare: the experience of the digital, continuous and systematic PREMs observatory in Italy. BMC Health Serv Res 2020 Apr 16;20(1):315 [FREE Full text] [doi: 10.1186/s12913-020-05099-4] [Medline: 32299440]
17. Bull C, Callander EJ. Current PROM and PREM use in health system performance measurement: still a way to go. Patient Exp J 2022 Apr 27;9(1):12-18. [doi: 10.35680/2372-0247.1664]

18. Bull C, Teede H, Watson D, Callander EJ. Selecting and implementing patient-reported outcome and experience measures to assess health system performance. JAMA Health Forum 2022 Apr 01;3(4):e220326 [FREE Full text] [doi: 10.1001/jamahealthforum.2022.0326] [Medline: 36218960]
19. Meirte J, Hellemans N, Anthonissen M, Denteneer L, Maertens K, Moortgat P, et al. Benefits and disadvantages of electronic patient-reported outcome measures: systematic review. JMIR Perioper Med 2020 Apr 03;3(1):e15588 [FREE Full text] [doi: 10.2196/15588] [Medline: 33393920]
20. Prodinger B, Taylor P. Improving quality of care through patient-reported outcome measures (PROMs): expert interviews using the NHS PROMs Programme and the Swedish quality registers for knee and hip arthroplasty as examples. BMC Health Serv Res 2018 Feb 07;18(1):87 [FREE Full text] [doi: 10.1186/s12913-018-2898-z] [Medline: 29415714]
21. Eriksen J, Bygholm A, Bertelsen P. The association between patient-reported outcomes (PROs) and patient participation in chronic care: a scoping review. Patient Educ Couns 2022 Jul;105(7):1852-1864 [FREE Full text] [doi: 10.1016/j.pec.2022.01.008] [Medline: 35090802]
22. Hjollund NH. Fifteen years' use of patient-reported outcome measures at the group and patient levels: trend analysis. J Med Internet Res 2019 Sep 30;21(9):e15856 [FREE Full text] [doi: 10.2196/15856] [Medline: 31573943]
23. Bauer G. Delivering value-based care with e-health services. J Healthc Manag 2018;63(4):251-260. [doi: 10.1097/JHM-D-18-00077] [Medline: 29985252]
24. Rosalia RA, Wahba K, Milevska-Kostova N. How digital transformation can help achieve value-based healthcare: balkans as a case in point. Lancet Reg Health Eur 2021 Apr 14;4:100100 [FREE Full text] [doi: 10.1016/j.lanepe.2021.100100] [Medline: 34557815]
25. El-Sherif DM, Abouzid M, Elzarif MT, Ahmed AA, Albakri A, Alshehri MM. Telehealth and artificial intelligence insights into healthcare during the COVID-19 pandemic. Healthcare (Basel) 2022 Feb 18;10(2):385 [FREE Full text] [doi: 10.3390/healthcare10020385] [Medline: 35206998]
26. Naveta Health homepage. Naveta Health. URL: https://navetahealth.com/ [accessed 2023-12-06]
27. Bibliopro homepage. Bibliopro. URL: https://www.bibliopro.org/ [accessed 2023-12-06]
28. Pardo Y, Oriol Zerbe C, Garin O, Vilagut G, García-Forero C, Martí-Pastor M, et al. PRM206 - the periodic systematic review for bibliopro: a decade of Spanish patient-reported outcomes. Value Health 2018 Oct;21(Supplement 3):S391. [doi: 10.1016/j.jval.2018.09.2324]
29. Phemium homepage. Phemium. URL: https://www.phemium.com/ [accessed 2023-12-06]
30. Burns DJ, Arora J, Okunade O, Beltrame JF, Bernardez-Pereira S, Crespo-Leiro MG, et al. International consortium for health outcomes measurement (ICHOM): standardized patient-centered outcomes measurement set for heart failure patients. JACC Heart Fail 2020 Mar;8(3):212-222 [FREE Full text] [doi: 10.1016/j.jchf.2019.09.007] [Medline: 31838032]
31. The PROTEUS-practice guide. PROTEUS. URL: https://theproteusconsortium.org/proteus-practice/proteus-practice-guide/ [accessed 2023-12-06]
32. El proyecto naveta de telefarmacia asociada a los cuestionarios de salud reportados por los pacientes gana la 2ª edición del foro de innovación de la sefh. Sociedad Española de Farmacia Hospitalaria. 2021 Oct 27. URL: https://www.sefh.es/sefh-sala-prensa/noticia/1601 [accessed 2023-12-08]
33. What is PROMIS? PROMIS Health Organization. URL: https://www.promishealth.org/57461-2/ [accessed 2023-12-06]
34. Kelley TA. International Consortium for Health Outcomes Measurement (ICHOM). Trials 2015 Nov 24;16(S3):O4. [doi: 10.1186/1745-6215-16-S3-O4]
35. Patient-Reported Indicator Surveys (PaRIS). Organization for Economic Co-operation and Development. URL: https://www.oecd.org/health/paris/ [accessed 2023-12-06]
36. Sugavanam T, Fosh B, Close J, Byng R, Horrell J, Lloyd H. Codesigning a measure of person-centred coordinated care to capture the experience of the patient: the development of the P3CEQ. J Patient Exp 2018 Sep;5(3):201-211 [FREE Full text] [doi: 10.1177/2374373517748642] [Medline: 30214927]
37. Lloyd H, Fosh B, Whalley B, Byng R, Close J. Validation of the person-centred coordinated care experience questionnaire (P3CEQ). Int J Qual Health Care 2019 Aug 01;31(7):506-512 [FREE Full text] [doi: 10.1093/intqhc/mzy212] [Medline: 30508089]
38. Keininger D, Coteur G. Assessment of self-injection experience in patients with rheumatoid arthritis: psychometric validation of the Self-Injection Assessment Questionnaire (SIAQ). Health Qual Life Outcomes 2011 Jan 13;9:2 [FREE Full text] [doi: 10.1186/1477-7525-9-2] [Medline: 21232106]
39. Pompilus F, Ciesluk A, Strzok S, Ciaravino V, Harris K, Szegvari B, et al. Development and psychometric evaluation of the assessment of self-injection questionnaire: an adaptation of the self-injection assessment questionnaire. Health Qual Life Outcomes 2020 Nov 04;18(1):355 [FREE Full text] [doi: 10.1186/s12955-020-01606-7] [Medline: 33148261]
40. The R project for statistical computing. The Comprehensive R Archive Network. URL: https://www.r-project.org/ [accessed 2023-12-06]
41. Dawson J, Doll H, Fitzpatrick R, Jenkinson C, Carr AJ. The routine use of patient reported outcome measures in healthcare settings. BMJ 2010 Jan 18;340:c186. [doi: 10.1136/bmj.c186] [Medline: 20083546]

42. Chen J, Ou L, Hollis SJ. A systematic review of the impact of routine collection of patient reported outcome measures on patients, providers and health organisations in an oncologic setting. BMC Health Serv Res 2013 Jun 11;13:211 [FREE Full text] [doi: 10.1186/1472-6963-13-211] [Medline: 23758898]
43. Rijken M, Lette M, Baan CA, de Bruin SR. Assigning a prominent role to "the patient experience" in assessing the quality of integrated care for populations with multiple chronic conditions. Int J Integr Care 2019 Sep 26;19(3):19 [FREE Full text] [doi: 10.5334/ijic.4656] [Medline: 31592248]
44. Wang K, Eftang CN, Jakobsen RB, Årøen A. Review of response rates over time in registry-based studies using patient-reported outcome measures. BMJ Open 2020 Aug 06;10(8):e030808 [FREE Full text] [doi: 10.1136/bmjopen-2019-030808] [Medline: 32764078]
45. Schougaard LM, Larsen LP, Jessen A, Sidenius P, Dorflinger L, de Thurah A, et al. AmbuFlex: tele-patient-reported outcomes (telePRO) as the basis for follow-up in chronic and malignant diseases. Qual Life Res 2016 Mar;25(3):525-534 [FREE Full text] [doi: 10.1007/s11136-015-1207-0] [Medline: 26790427]
46. Kjær AS, Rasmussen TA, Hjollund NH, Rodkjaer LO, Storgaard M. Patient-reported outcomes in daily clinical practise in HIV outpatient care. Int J Infect Dis 2018 Apr;69:108-114 [FREE Full text] [doi: 10.1016/j.ijid.2018.02.015] [Medline: 29476900]
47. Sisodia RC, Dankers C, Orav J, Joseph B, Meyers P, Wright P, et al. Factors associated with increased collection of patient-reported outcomes within a large health care system. JAMA Netw Open 2020 Apr 01;3(4):e202764 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.2764] [Medline: 32286657]
48. Ruseckaite R, Mudunna C, Caruso M, Ahern S. Response rates in clinical quality registries and databases that collect patient reported outcome measures: a scoping review. Health Qual Life Outcomes 2023 Jul 11;21(1):71 [FREE Full text] [doi: 10.1186/s12955-023-02155-5] [Medline: 37434146]
49. Unni E, Coles T, Lavallee DC, Freel J, Roberts N, Absolom K. Patient adherence to patient-reported outcome measure (PROM) completion in clinical care: current understanding and future recommendations. Qual Life Res 2024 Jan;33(1):281-290 [FREE Full text] [doi: 10.1007/s11136-023-03505-y] [Medline: 37695476]
50. Ho A, Purdie C, Tirosh O, Tran P. Improving the response rate of patient-reported outcome measures in an Australian tertiary metropolitan hospital. Patient Relat Outcome Meas 2019 Jul 10;10:217-226 [FREE Full text] [doi: 10.2147/PROM.S162476] [Medline: 31372076]
51. Walker RC, Tong A, Howard K, Palmer SC. Patient expectations and experiences of remote monitoring for chronic diseases: systematic review and thematic synthesis of qualitative studies. Int J Med Inform 2019 Apr;124:78-85 [FREE Full text] [doi: 10.1016/j.ijmedinf.2019.01.013] [Medline: 30784430]
52. Neve OM, van Benthem PP, Stiggelbout AM, Hensen EF. Response rate of patient reported outcomes: the delivery method matters. BMC Med Res Methodol 2021 Oct 22;21(1):220 [FREE Full text] [doi: 10.1186/s12874-021-01419-2] [Medline: 34686129]
53. Hughes J, Flood T. Patients' experiences of engaging with electronic Patient Reported Outcome Measures (PROMs) after the completion of radiation therapy for breast cancer: a pilot service evaluation. J Med Radiat Sci 2023 Dec;70(4):424-435 [FREE Full text] [doi: 10.1002/jmrs.711] [Medline: 37550951]
54. Cotter EJ, Hannon CP, Locker P, Davey A, Wang KC, Verma NN, et al. Male sex, decreased activity level, and higher BMI associated with lower completion of patient-reported outcome measures following ACL reconstruction. Orthop J Sports Med 2018 Mar 07;6(3):2325967118758602 [FREE Full text] [doi: 10.1177/2325967118758602] [Medline: 29536023]
55. Ross LA, O'Rourke SC, Toland G, MacDonald DJ, Clement ND, Scott CE. Loss to patient-reported outcome measure follow-up after hip arthroplasty and knee arthroplasty : patient satisfaction, associations with non-response, and maximizing returns. Bone Jt Open 2022 Apr;3(4):275-283 [FREE Full text] [doi: 10.1302/2633-1462.34.BJO-2022-0013.R1] [Medline: 35357243]
56. Nielsen AS, Kidholm K, Kayser L. Patients' reasons for non-use of digital patient-reported outcome concepts: a scoping review. Health Informatics J 2020 Dec 30;26(4):2811-2833 [FREE Full text] [doi: 10.1177/1460458220942649] [Medline: 32731773]
57. Eriksen J, Bygholm A, Bertelsen P. The purpose of patient-reported outcome (PRO) post its digitalization and integration into clinical practice: an interdisciplinary redefinition resembling PROs theoretical and practical evolvement. Appl Sci 2020 Oct 26;10(21):7507 [FREE Full text] [doi: 10.3390/app10217507]
58. Hjollund NH, Larsen LP, Biering K, Johnsen SP, Riiskjær E, Schougaard LM. Use of patient-reported outcome (PRO) measures at group and patient levels: experiences from the generic integrated PRO system, WestChronic. Interact J Med Res 2014 Feb 11;3(1):e5 [FREE Full text] [doi: 10.2196/ijmr.2885] [Medline: 24518281]
59. Aiyegbusi OL, Nair D, Peipert JD, Schick-Makaroff K, Mucsi I. A narrative review of current evidence supporting the implementation of electronic patient-reported outcome measures in the management of chronic diseases. Ther Adv Chronic Dis 2021 May 24;12:20406223211015958 [FREE Full text] [doi: 10.1177/20406223211015958] [Medline: 34104376]
60. Wiegel J, Seppen B, van der Leeden M, van der Esch M, de Vries R, Bos W. Adherence to telemonitoring by electronic patient-reported outcome measures in patients with chronic diseases: a systematic review. Int J Environ Res Public Health 2021 Sep 27;18(19):10161 [FREE Full text] [doi: 10.3390/ijerph181910161] [Medline: 34639463]

61. Graupner C, Kimman ML, Mul S, Slok AH, Claessens D, Kleijnen J, et al. Patient outcomes, patient experiences and process indicators associated with the routine use of patient-reported outcome measures (PROMs) in cancer care: a systematic review. Support Care Cancer 2021 Feb;29(2):573-593 [FREE Full text] [doi: 10.1007/s00520-020-05695-4] [Medline: 32875373]
62. Lehmann J, de Ligt KM, Tipelius S, Giesinger JM, Sztankay M, Voigt S, et al. Adherence to patient-reported symptom monitoring and subsequent clinical interventions for patients with multiple myeloma in outpatient care: longitudinal observational study. J Med Internet Res 2023 Aug 22;25:e46017 [FREE Full text] [doi: 10.2196/46017] [Medline: 37606979]
63. Peters M, Crocker H, Jenkinson C, Doll H, Fitzpatrick R. The routine collection of patient-reported outcome measures (PROMs) for long-term conditions in primary care: a cohort survey. BMJ Open 2014 Feb 21;4(2):e003968 [FREE Full text] [doi: 10.1136/bmjopen-2013-003968] [Medline: 24561495]
64. Gamper EM, Nerich V, Sztankay M, Martini C, Giesinger JM, Scarpa L, et al. Evaluation of noncompletion bias and long-term adherence in a 10-year patient-reported outcome monitoring program in clinical routine. Value Health 2017 Apr;20(4):610-617 [FREE Full text] [doi: 10.1016/j.jval.2017.01.009] [Medline: 28408003]
65. Aiyegbusi OL, Roydhouse J, Rivera SC, Kamudoni P, Schache P, Wilson R, et al. Key considerations to reduce or address respondent burden in patient-reported outcome (PRO) data collection. Nat Commun 2022 Oct 12;13(1):6026 [FREE Full text] [doi: 10.1038/s41467-022-33826-4] [Medline: 36224187]
66. Segawa E, Schalet B, Cella D. A comparison of computer adaptive tests (CATs) and short forms in terms of accuracy and number of items administrated using PROMIS profile. Qual Life Res 2020 Jan;29(1):213-221. [doi: 10.1007/s11136-019-02312-8] [Medline: 31595451]
67. van Muilekom MM, Luijten MA, van Oers HA, Terwee CB, van Litsenburg RR, Roorda LD, et al. From statistics to clinics: the visual feedback of PROMIS CATs. J Patient Rep Outcomes 2021 Jul 10;5(1):55 [FREE Full text] [doi: 10.1186/s41687-021-00324-y] [Medline: 34245390]
68. Mejdahl CT, Schougaard LM, Hjollund NH, Riiskjær E, Thorne S, Lomborg K. PRO-based follow-up as a means of self-management support - an interpretive description of the patient perspective. J Patient Rep Outcomes 2017 Sep 05;2:38 [FREE Full text] [doi: 10.1186/s41687-018-0067-0] [Medline: 30238083]
69. Stover AM, Haverman L, van Oers HA, Greenhalgh J, Potter CM, ISOQOL PROMs/PREMs in Clinical Practice Implementation Science Work Group. Using an implementation science approach to implement and evaluate patient-reported outcome measures (PROM) initiatives in routine care settings. Qual Life Res 2021 Nov;30(11):3015-3033 [FREE Full text] [doi: 10.1007/s11136-020-02564-9] [Medline: 32651805]
70. Glenwright BG, Simmich J, Cottrell M, O'Leary SP, Sullivan C, Pole JD, et al. Facilitators and barriers to implementing electronic patient-reported outcome and experience measures in a health care setting: a systematic review. J Patient Rep Outcomes 2023 Feb 14;7(1):13 [FREE Full text] [doi: 10.1186/s41687-023-00554-2] [Medline: 36786914]
71. Haywood K, Lyddiatt A, Brace-McDonnell SJ, Staniszewska S, Salek S. Establishing the values for patient engagement (PE) in health-related quality of life (HRQoL) research: an international, multiple-stakeholder perspective. Qual Life Res 2017 Jun;26(6):1393-1404 [FREE Full text] [doi: 10.1007/s11136-016-1465-5] [Medline: 27933429]

Abbreviations

ePREM: electronic patient-reported experience measure
ePROM: electronic patient-reported outcome measure
FARUPEIB: Association of Pharmacists of Outpatient Departments of the Balearic Islands
ICHOM: International Consortium for Health Outcomes Measurement
P3CEQ: Person-Centered Coordinated Care Experience Questionnaire
PREM: patient-reported experience measure
PRO: patient-reported outcome
PROM: patient-reported outcome measure
PROMIS: Patient-Reported Outcomes Measurement Information System
PROTEUS: Patient-Reported Outcomes Tools: Engaging Users and Stakeholders
SIAQ: Self-Injection Assessment Questionnaire
VBHC: value-based health care

Edited by L Buis; submitted 09.01.24; peer-reviewed by L Margusino, E Monte, J Eriksen; comments to author 29.02.24; revised version received 12.03.24; accepted 27.03.24; published 28.06.24.

Please cite as:
Mercadal-Orfila G, Herrera-Pérez S, Piqué N, Mateu-Amengual F, Ventayol-Bosch P, Maestre-Fullana MA, Serrano-López de las Hazas JI, Fernández-Cortés F, Barceló-Sansó F, Rios S
Implementing Systematic Patient-Reported Measures for Chronic Conditions Through the Naveta Value-Based Telemedicine Initiative: Observational Retrospective Multicenter Study
JMIR Mhealth Uhealth 2024;12:e56196
URL: https://mhealth.jmir.org/2024/1/e56196
doi:10.2196/56196
PMID:38545697

Original Paper

The Effect of an mHealth Self-Monitoring Intervention (MI-BP) on Blood Pressure Among Black Individuals With Uncontrolled Hypertension: Randomized Controlled Trial

Lorraine R Buis[1], PhD; Junhan Kim[2], MS; Ananda Sen[1,3], PhD; Dongru Chen[1], MS; Katee Dawood[4], MBA; Reema Kadri[1], MLiS; Rachelle Muladore[4], BS; Melissa Plegue[5], MA; Caroline R Richardson[1,6], MD; Zora Djuric[1], PhD; Candace McNaughton[7], MD; David Hutton[8], PhD; Lionel P Robert[2], PhD; Sun Young Park[2], PhD; Phillip Levy[4], MPH, MD

[1]Department of Family Medicine, University of Michigan, Ann Arbor, MI, United States
[2]School of Information, University of Michigan, Ann Arbor, MI, United States
[3]Department of Biostatistics, University of Michigan, Ann Arbor, MI, United States
[4]Department of Emergency Medicine and Integrative Biosciences Center, Wayne State University, Detroit, MI, United States
[5]Department of Pediatrics, University of Michigan, Ann Arbor, MI, United States
[6]Department of Family Medicine, Brown University, Providence, RI, United States
[7]Department of Medicine, Sunnybrook Research Institute, University of Toronto, Toronto, ON, Canada
[8]School of Public Health, University of Michigan, Ann Arbor, MI, United States

Corresponding Author:
Lorraine R Buis, PhD
Department of Family Medicine
University of Michigan
1018 Fuller St.
Ann Arbor, MI, 48104
United States
Phone: 1 734 998 7120
Fax: 1 734 998 7345
Email: buisl@umich.edu

Related Article:

This is a corrected version. See correction statement: https://mhealth.jmir.org/2024/1/e64632

Abstract

Background: Hypertension is one of the most important cardiovascular disease risk factors and affects >100 million American adults. Hypertension-related health inequities are abundant in Black communities as Black individuals are more likely to use the emergency department (ED) for chronic disease–related ambulatory care, which is strongly linked to lower blood pressure (BP) control, diminished awareness of hypertension, and adverse cardiovascular events. To reduce hypertension-related health disparities, we developed MI-BP, a culturally tailored multibehavior mobile health intervention that targeted behaviors of BP self-monitoring, physical activity, sodium intake, and medication adherence in Black individuals with uncontrolled hypertension recruited from ED and community-based settings.

Objective: We sought to determine the effect of MI-BP on BP as well as secondary outcomes of physical activity, sodium intake, medication adherence, and BP control compared to enhanced usual care control at 1-year follow-up.

Methods: We conducted a 1-year, 2-group randomized controlled trial of the MI-BP intervention compared to an enhanced usual care control group where participants aged 25 to 70 years received a BP cuff and hypertension-related educational materials. Participants were recruited from EDs and other community-based settings in Detroit, Michigan, where they were screened for initial eligibility and enrolled. Baseline data collection and randomization occurred approximately 2 and 4 weeks after enrollment to ensure that participants had uncontrolled hypertension and were willing to take part. Data collection visits occurred at 13, 26,

39, and 52 weeks. Outcomes of interest included BP (primary outcome) and physical activity, sodium intake, medication adherence, and BP control (secondary outcomes).

Results: We obtained consent from and enrolled 869 participants in this study yet ultimately randomized 162 (18.6%) participants. At 1 year, compared to the baseline, both groups showed significant decreases in systolic BP (MI-BP group: 22.5 mm Hg decrease in average systolic BP and $P<.001$; control group: 24.1 mm Hg decrease and $P<.001$) adjusted for age and sex, with no significant differences between the groups (time-by-arm interaction: $P=.99$). Similar patterns where improvements were noted in both groups yet no differences were found between the groups were observed for diastolic BP, physical activity, sodium intake, medication adherence, and BP control. Large dropout rates were observed in both groups (approximately 60%).

Conclusions: Overall, participants randomized to both the enhanced usual care control and MI-BP conditions experienced significant improvements in BP and other outcomes; however, differences between groups were not detected, speaking to the general benefit of proactive outreach and engagement focused on cardiometabolic risk reduction in urban-dwelling, low-socioeconomic-status Black populations. High dropout rates were found and are likely to be expected when working with similar populations. Future work is needed to better understand engagement with mobile health interventions, particularly in this population.

Trial Registration: ClinicalTrials.gov NCT02955537; https://clinicaltrials.gov/study/NCT02955537

International Registered Report Identifier (IRRID): RR2-10.2196/12601

(JMIR Mhealth Uhealth 2024;12:e57863) doi:10.2196/57863

KEYWORDS

blood pressure; hypertension; mobile health; mHealth; mobile phone; smartphone

Introduction

Background

Hypertension affects >100 million American adults, which is nearly half of individuals aged ≥20 years [1]. Hypertension is also one of the most important cardiovascular disease risk factors, and when uncontrolled, it can cause adverse health outcomes such as myocardial infarction, stroke, heart failure, and chronic kidney disease [2-6]. Despite the importance of maintaining adequate blood pressure (BP) control, the American Heart Association has reported that only approximately 21.6% of those with hypertension have their BP controlled within age-adjusted criteria. Furthermore, 38.8% are unaware of their condition [1]. Hypertension-related health inequities are abundant in Black communities. Compared to White individuals, Black individuals have a greater prevalence of hypertension, hypertension-associated disease severity, and younger age of onset, making uncontrolled hypertension a significant problem in this population [7]. Moreover, Black individuals are more likely to use the emergency department (ED) for chronic disease–related ambulatory care, which is strongly linked to lower BP control, diminished awareness of hypertension, and adverse cardiovascular events [8-10].

Although uncontrolled hypertension is linked to a host of adverse outcomes, BP can typically be well controlled through lifestyle and behavior changes. Recommendations for managing hypertension have been consistent for decades and center on positive health behaviors such as maintaining a healthy weight, reducing daily sodium intake, increasing physical activity, and adhering to prescribed antihypertensive therapies [11]; however, engaging in these behaviors is difficult for many individuals, and this is especially true in Black individuals, who are less likely than White individuals to report adherence to preventive behaviors [1]. With the increasing national conversation focused on health inequities and social determinants of health [12], population-specific interventions are needed to intervene in communities where the burden of hypertension is disproportionately high.

Objectives

Mobile health (mHealth) for chronic disease self-management is increasing in use, and based on this, as well as on the high penetration of smartphone ownership among Black individuals (currently approximately 83%) [13], we sought to develop and test an mHealth intervention (MI-BP). The MI-BP intervention was developed with the intention of educating and supporting self-monitoring of multiple health behaviors to reduce BP among Black individuals with uncontrolled hypertension recruited from urban EDs and community-based settings. Our goal was to determine the effect of MI-BP on the primary outcome of BP and secondary outcomes of physical activity, sodium intake, medication adherence, and BP control compared to enhanced usual care in a 1-year randomized controlled trial (RCT). We hypothesized that (1) mean systolic BP (SBP) would be significantly lower in the MI-BP arm than in the control group after 1 year (hypothesis 1) and (2) measures of physical activity, sodium intake, medication adherence, and BP control would be significantly better in the MI-BP arm than in the control arm after 1 year (hypothesis 2).

Methods

Overview

This study was a 1-year, 2-group RCT of the MI-BP intervention compared to enhanced usual care. The study was overseen by our Data and Safety Monitoring Board (DSMB). Details on the full study protocol have been previously published [14], but a summary of those procedures follows.

Clinical Setting and Recruitment

All recruitment occurred in Detroit, Michigan, and was primarily conducted at the Detroit Medical Center in the EDs of Detroit Receiving Hospital and Sinai-Grace Hospital. Potentially eligible participants were screened by trained volunteers or by study staff members. Once a potentially eligible participant was identified according to clinical criteria, a research staff member spoke with the treating physician to determine whether they were a good candidate for participation. If so, individuals were informed of the study, screened further, and then consented and enrolled if they were interested and met the eligibility criteria. Additional recruitment occurred at community events where BP screening was conducted, such as mobile health unit visits, health fairs, and other health-related community events. Procedures for these potential participants were the same except for not checking with treating physicians to determine whether our staff should proceed with screening and enrollment.

Eligibility Screening and Consent

Inclusion Criteria

To be eligible to participate in this trial, individuals were required to be Black, between the ages of 25 and 70 years, previously diagnosed with hypertension, have a smartphone compatible with the MI-BP intervention, and have uncontrolled BP (SBP >135 mm Hg) at triage and on repeat measurement using a BpTRU BPM-200 device (Smiths Medical PM, Inc) or Omron HEM 907XL IntelliSense (Omron Healthcare, Inc) at least 1 hour after triage vitals were taken.

Exclusion Criteria

Individuals were excluded from this trial if they were pregnant; had serious existing medical conditions that may make BP control difficult or necessitate frequent hospitalization (ie, previous diagnosis of resistant hypertension, steroid-dependent asthma or emphysema, cirrhosis or hepatic failure, stage-C or stage-D chronic heart failure, stage-IV or stage-V chronic kidney disease, and terminal cancer or ongoing active chemotherapy or radiation therapy); had a history of other serious medical conditions (eg, stroke, dementia, myocardial infarction, or known coronary artery disease); or had a history of alcohol or drug abuse as determined using the Cut down, Annoyed, Guilty, and Eye-opener Adapted to Include Drugs questionnaire (excluded if score was ≥2).

Study Procedures

Baseline Data Collection Visit

After consent and enrollment, participants were scheduled for a return visit 1 to 2 weeks later for baseline data collection at a nearby university building. Transportation to all study visits via taxi or ride-sharing service was offered to anyone requiring transportation assistance. At the baseline visit, a secondary BP screening was conducted to ensure that we were only retaining participants with persistent uncontrolled hypertension in the study. At this time, participants who had an SBP of <130 mm Hg were deemed ineligible and excluded from the study. Next, baseline data were collected. To control response fatigue in the baseline data collection survey, we created 6 different permutations, each with a different order of instruments, which were also balanced within blocks. At this time, participants were also given a prescription for antihypertensive therapy. If needed, referrals to primary care were made by study physicians. In the event that a participant was already taking antihypertensive medications prescribed through a preexisting relationship with a primary care provider (PCP), we contacted their PCP to inform them of our algorithm-based approach to antihypertensive therapy and coordinated with them when medication adjustments were indicated.

Medication Titration and Randomization Visit

At 2 weeks after the baseline visit, participants were assessed for medication titration, the process of adjusting antihypertensive medication dosages to ensure appropriate and optimal treatment. At this time, participants were randomized into 1 of the 2 study arms in the trial. In total, it took approximately 4 weeks for an enrolled participant to be randomized into the study. This month-long de facto washout period was designed to ensure that we were truly reaching individuals with uncontrolled hypertension and who were not just temporarily presenting with elevated BP in the ED. Moreover, our previous experiences conducting work in this setting demonstrated high levels of attrition between ED recruitment and initial follow-up. Delaying randomization also helped ensure identification of individuals who did not intend to fully participate at the outset, increasing the likelihood of randomized participant retention. Trial randomization was stratified by sex in blocks of equal size. Study staff responsible for arm allocation were blinded to block size to prevent contamination. After randomization, all study materials, including any equipment, were distributed to the participants according to the treatment arm. A second titration visit was conducted 6 weeks after randomization, and the need for titration was assessed at each subsequent follow-up visit.

Quarterly Follow-Up Visits

Data collection assessments were conducted at weeks 0, 13, 26, 39, and 52 using a consistent set of study measures. In addition to survey measures, patients were instructed to bring their hypertension medications with them so pill counts could be conducted. As electronic health record data were not available to our study team, all medication data, including prescribed medication names and doses, were self-reported or captured from pill bottles. We also monitored for any potentially harmful renal or metabolic issues at baseline and weeks 26 and 52 and adjusted medications accordingly. To measure sodium intake (a secondary outcome measure of interest), at weeks 0, 26, and 52, participants were given supplies to collect 24-hour urine for sodium measurement. Study staff collected these specimens directly from the participants at their home to improve adherence. All medication titration and study follow-up visits were free; however, participants were responsible for the cost of medications, PCP visits, or copays, as applicable.

Impact of COVID-19 on Study Procedures

In March 2020, the MI-BP trial was closed to new enrollments and in-person data collection due to the COVID-19 pandemic. This necessitated protocol changes in the following weeks and months in an effort to maximize data collection from participants who were enrolled in the study before the pandemic. To

summarize these changes, we pivoted to remote data collection for follow-up assessments via phone or videoconference. This meant that home-monitored BP measurements using study-issued cuffs served as the final outcome measures for participants completing their trial participation between March 2020 and April 2021. In addition, as all in-person participant interaction had been suspended, all laboratory measures were discontinued during the COVID-19 pandemic, and survey-based assessments were conducted verbally by phone or videoconference out of concern for literacy levels among participants. We also removed several instruments from interim follow-up assessments in weeks 13 and 39. Finally, anthropometric assessments, including weight, height, and waist circumference measurements, were self-reported by participants using their own home scales and tape measures. Given the increased reliance on home-based, self-reported data, the chance of missing data from follow-up assessments was greater.

Trial Arms

Participants in this trial were randomized equally to 1 of the 2 treatment arms, which included an enhanced usual care control arm and the MI-BP intervention arm.

Enhanced Usual Care (Control Arm)

Participants in the enhanced usual care group were given a prescription for antihypertensive medications, printed educational materials on hypertension, and a BP monitor for daily use. Participants assigned to the enhanced usual care control group received no further intervention; however, they were asked to take part in all study-specific follow-up visits. The decision to provide home BP cuffs to control participants, above and beyond true usual care, was made to reflect the fact that home BP monitoring is widely accepted as a guideline-based standard of care for individuals being treated for hypertension [6], making it appropriate to include in the usual care arm. We acknowledge that this active control represents a departure from true usual care; however, it does represent an ideal usual care scenario based on current hypertension management guidelines.

MI-BP (Treatment Arm)

Participants randomized to receive the MI-BP intervention were given a prescription for antihypertensive medication, a Bluetooth-enabled pedometer (Fitbit Zip), a BP cuff, and access to the MI-BP mobile app. Participants were asked to use the MI-BP mobile app and related peripheral devices for 12 months.

MI-BP Intervention

Overview

MI-BP is a comprehensive, multicomponent intervention that targets multiple behaviors for managing hypertension via smartphone app, including BP self-monitoring, physical activity tracking, sodium intake tracking, goal setting, educational and motivational messaging, and medication adherence reminders. The MI-BP app was developed by Vibrent Health, a digital health company. Vibrent Health designed the app, study staff web-based portal, and server platforms necessary to support this trial. The MI-BP app was previously described in detail but is summarized in this section [14].

BP Monitoring

To support BP self-monitoring, participants who could use a standard BP cuff (suitable for an arm circumference between 23 and 45 cm) were provided with a Bluetooth-enabled BP cuff (A&D UA-651BLE) that could sync to the MI-BP app. The MI-BP app showed different visualizations of BP over time, including both graph and log form. In the event that a participant required a larger cuff size (between 42 and 60 cm), we provided an extra-large arm monitor (A&D LifeSource UA-789), which was not Bluetooth enabled and required manual data entry. Participants were instructed to measure and sync (or manually enter) their BP to the MI-BP app at home using a commonly accepted home BP-monitoring protocol for a minimum of 3 days per week; however, daily self-monitoring and syncing were encouraged. If participants self-monitored an SBP reading of >180 or <100 mm Hg or a diastolic BP (DBP) reading of >110 mm Hg, they were instructed by the study staff at baseline, as well as by automated notifications within the app at the time of the elevated reading, to check their BP again. If it was still elevated after 3 days, participants were instructed to call the study staff. Participants were also instructed to report to the ED and follow up afterward with a call to the research staff if they were experiencing symptoms of dizziness, chest pain, severe headache, vision changes, or numbness or weakness in the face or extremities.

Physical Activity Monitoring and Tracking

To support physical activity self-monitoring, participants were provided with a Fitbit Zip pedometer that could sync to the MI-BP app, which showed different visualizations (graph and log form) of physical activity data over time. Participants were instructed to wear their Fitbit daily and sync the device at least once per week.

Sodium Intake Monitoring and Tracking

To support sodium intake monitoring, the MI-BP app used a logging approach that encouraged participants to identify their intake of high-sodium foods using a checklist-type log available within the MI-BP app. The checklist comprised 7 categories with 3 to 8 items per category and represented the most common types of high-sodium foods that contribute to high-sodium diets. Although we encouraged users to track their intake of high-sodium foods daily, users were instructed to engage at a minimum in highly focused, 3-day consecutive bouts of logging that were prompted within the MI-BP app.

Goal Setting

Participants received weekly step count goals that were displayed in the MI-BP app and were also delivered via push notifications. On the basis of previous work from our team [15-19], step count goals were gradually incremented and were based on an average of 7 consecutive days of data, during which at least 5 of the days needed to be considered valid. A valid day was defined as >200 steps per day. As we gradually incremented weekly goals, calculated goals never exceeded 600 additional steps over the previous goal. This gradual increment in weekly step count goals was made in an effort to reduce potential adverse events (AEs).

Goal setting for sodium intake was also conducted every 2 to 4 weeks after an intensive 1-week baseline self-monitoring period that was used to calculate the initial goal for each participant. Sodium intake goals were displayed within the app and were also sent via push notifications. Participants were instructed to log their intake of high-sodium foods for a 3-day period approximately 2 weeks after receiving their initial sodium intake goal that limited the number of high-sodium foods to be consumed. When sodium goals were met during a logging period, a new lower sodium intake goal was issued, and participants were asked to log their intake of high-sodium foods 4 weeks later. If the goal was not met, participants were asked to try again in 2 weeks. Additional details on our sodium logging and goal-setting protocols have been published previously [14].

Messaging

MI-BP provided users with 4 different types of messages, which were sent via push notifications and in-app messaging. These included educational messaging focused on hypertension, physical activity, sodium intake, and tips for behavior change and overcoming barriers to behavior change; motivational messaging; tailored messaging, including tips for overcoming specific self-reported barriers to behavior change and daily medication reminders as well as tailored feedback responsive to whether participants were meeting their set goals; and customizable daily medication reminders. In addition to the customizable daily medication reminders, MI-BP sent approximately 7 messages per week. Message content, frequency, and timing were varied and tailored wherever possible to maximize user engagement.

Measures

We collected a variety of measures throughout this study. Full details of our study measures have been published previously [14]; however, those discussed in this paper are described in this section. Data were collected at baseline; medication titration visits at weeks 2 and 8; and planned follow-up assessments at weeks 13, 26, 39, and 52. Although most measures were collected at all time points, some were collected less frequently due to participant burden and cost of administration. The primary outcome measure of BP was collected at the clinic (or at home with study issued BP cuffs during the COVID-19 pandemic) and was assessed at every study visit by a trained study staff member using a BpTRU BPM-200 or Omron HEM 907XL IntelliSense BP-monitoring device. Secondary outcome measures included the following: physical activity as measured using the International Physical Activity Questionnaire–Short Form (IPAQ-SF) [20]; sodium intake as measured using the Block Sodium Screener (BSS) [21] as well as a 24-hour urine sodium test; self-reported medication adherence using the Adherence to Refills and Medication Scale (ARMS-14) [22]; and self-efficacy for changing targeted behaviors, including physical activity via the Self-Efficacy for Exercise Behaviors (SEEB) scale [23] and medication adherence via the Medication Adherence Self-Efficacy Scale (MASES) [24], as well as diet using an investigator-developed 11-item instrument assessing confidence in reducing sodium consumption, avoiding high-fat foods, avoiding sugar-sweetened beverages, and improving vegetable and legume intake. Additional measures included hypertension knowledge measured by the Hypertension Evaluation of Lifestyle and Management (HELM) scale [25]; health literacy measured by the Rapid Assessment of Adult Literacy in Medicine–Short Form (REALM-SF) [26]; patient activation measured by the Patient Activation Measure (PAM) [27], and health-related quality of life measured by the Short Form–12 (SF-12) [28]. In addition to instruments assessing physical activity, sodium intake, and clinic-measured BP, we analyzed related study data from MI-BP treatment arm participants collected in the app.

Statistical Analysis

Sample Size

As stated in our study protocol [14], our sample size was initially developed with 2 co–primary outcomes: SBP measured continuously and SBP control (defined dichotomously as either above or below the SBP target of 130 mm Hg), which is a more conservative measure. After experiencing sustained challenges with recruitment, the more conservative dichotomous BP control measure was dropped as a co–primary outcome. This necessitated a recalculation in sample size based solely on continuously measured SBP [29]. Due to a lack of similar studies available at the time, we estimated a drop of 10 and 17 mm Hg points in SBP in the usual care and MI-BP arms, respectively, at the end of the trial, based on estimates derived from our own previous work [30]. A constant between-subject SD of 10 mm Hg was assumed, along with an intrasubject correlation of 0.5 [31]. With 121 participants per arm, these estimates would allow us to detect a group-by-time interaction with power of >95% at a 5% level of significance. Allowing for 20% attrition, we sought to recruit 152 participants per study arm for a total of 304 participants.

Analysis Plan

Descriptive statistics were calculated for demographic variables and baseline measures. To ensure balance across study arms, these measures were compared using 2-tailed *t* tests, Wilcoxon rank sum tests, or chi-square tests, as appropriate. Linear mixed models were used to investigate differences in change in the outcomes between study arms, with time, study arm, and their interaction as primary covariates. For all outcomes, time was entered as a categorical covariate. The models were further adjusted for age and sex. For the BP outcomes (SBP and DBP), a second set of models was explored, where time was entered as a linear term to capture the rate of change in the outcomes. All models included a random intercept to account for intrasubject correlation. Square root transformations were used for the IPAQ-SF and BSS, and a log transformation was applied to the ARMS-14 before running the linear mixed models to better meet model assumptions.

To investigate whether dropout was associated with any covariates, a time-to-dropout analysis was carried out using a Cox regression model. As this study was partially conducted during the COVID-19 pandemic, we wanted to consider the impact of COVID-19 on dropout. To that end, we defined a new variable, *COVID group*, for each individual at each time point. If the event time (eg, week 13) of individual assessments was before 3 PM on March 16, 2020, we considered these

records as *Before COVID-19*. Otherwise, we considered the records as *During COVID-19*. Thus, the COVID-19 group was modeled as a time-dependent covariate in the Cox model. All statistical analyses were carried out in SAS (version 9.4; SAS Institute).

Ethical Considerations

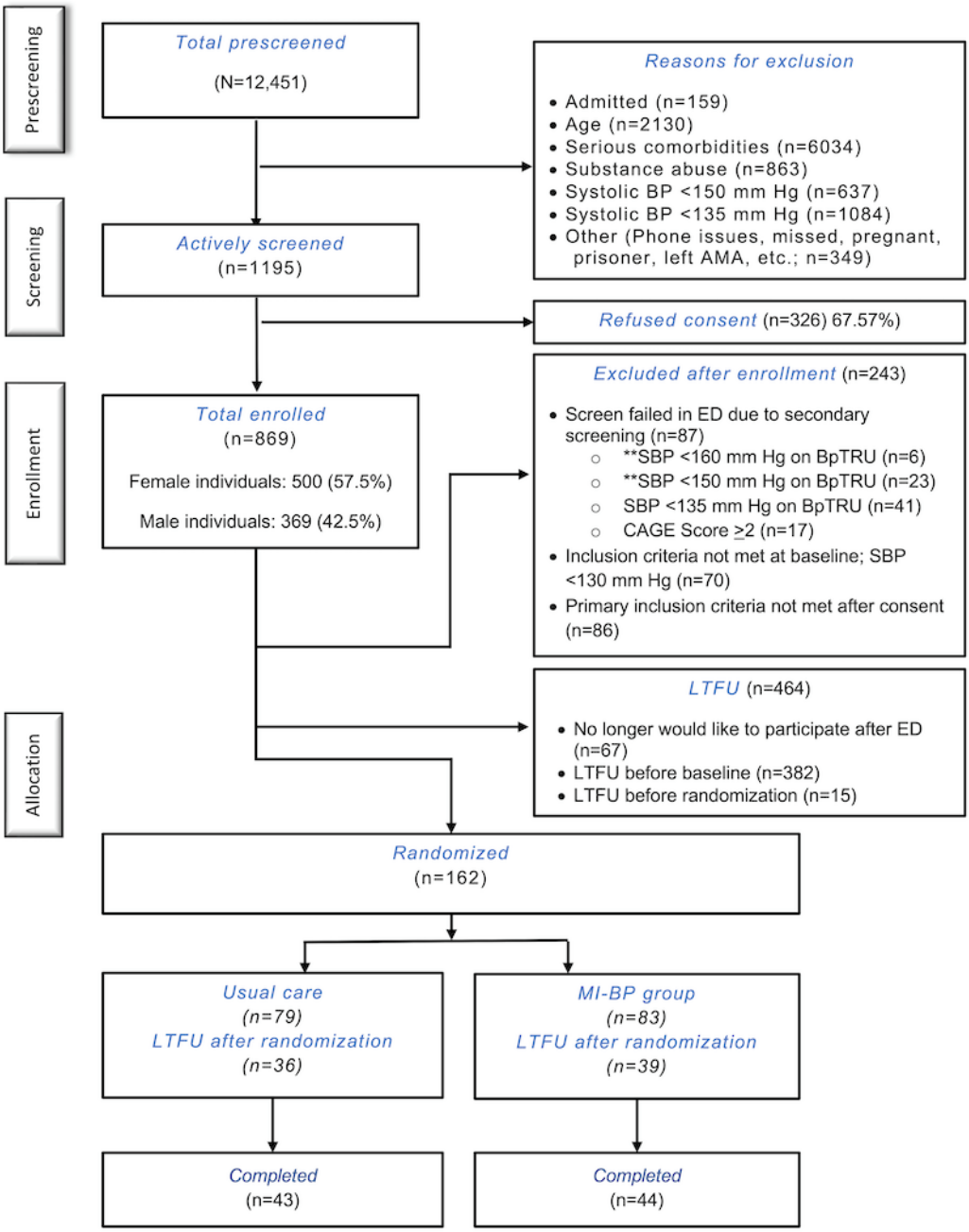
The methods for this study were approved by the institutional review boards (IRBs) at both Wayne State University (WSU; IRB 040416M1F) and the University of Michigan (HUM00114202). All participants provided written informed consent before enrollment. Participants received financial incentives to take part in this study, with each visit individually incentivized, and could earn up to US $275 over the course of 1 year. To protect the privacy and confidentiality of participants, all data are reported in the aggregate and no identifiable information is presented here.

Results

Trial Recruitment

In total, we prescreened 12,451 individuals, predominantly in ED settings (n=12,089, 97.09% in EDs; n=169, 1.36% from community events; and n=193, 1.55% from mobile health units), of whom 1195 (9.6%) were preliminarily eligible for participation. Of those 1195 participants, 869 (72.72%) were consented and enrolled in this study. Most enrolled participants were excluded after enrollment (before randomization) for failing the secondary screening in the ED, not meeting the inclusion criteria at baseline, or not meeting the primary inclusion criteria after consent (243/869, 28%); lost to follow-up before randomization (416/869, 47.9%); or scheduled for baseline or randomization visits that were halted due to the COVID-19 pandemic (48/869, 5.5%). Figure 1 shows the CONSORT (Consolidated Standards of Reporting Trials) diagram of participant flow through the trial. Due to the COVID-19 pandemic, in March 2020, ED recruitment for the trial was suspended precluding new enrollments from the ED. While we were able to eventually transition to community-based recruitment using mobile health units under an IRB-approved protocol amendment within 9 months of this, screening was severely reduced, and no new randomizations occurred. These considerations, combined with the challenge of keeping participants engaged using remote follow-up, prompted a decision by the study team, made in conjunction with our DSMB, to end recruitment for this study in January 2022. Ultimately, we randomized 162 participants to the MI-BP trial.

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) diagram showing participant recruitment and retention. **Reflects secondary blood pressure (BP) criteria in the previous protocol. AMA: against medical advice; CAGE: Cut down, Annoyed, Guilty, and Eye-opener; ED: emergency department; LTFU: lost to follow-up; mHealth: mobile health; SBP: systolic blood pressure.



Participant Characteristics

The 162 participants randomized to this trial (n=79, 48.8% to usual care and n=83, 51.2% to the MI-BP intervention) were predominantly female (n=97, 59.9%) and were, on average, aged 48.3 (SD 9.3; range 29-68) years. As race was an inclusion criterion, 100% (162/162) of our participants were Black individuals. Participants were characterized by being single (86/162, 53.1%) and employed (97/162, 59.9%) and having a high school education or lower (90/162, 55.6%) and an average household income of <US $25,000 (73/162, 45.1%). Table 1 shows participant characteristics as well as summary baseline measures stratified by study arm.

Table 1. Participant characteristics and baseline measures by study arm (N=162).

Participant characteristics and baseline measures	Overall	Intervention (n=83)	Control (n=79)	*P* value[a]
Site, n (%)				.45
COM[b]	17 (10.5)	10 (12)	7 (8.9)	
DRH[c]	83 (51.2)	45 (54.2)	38 (48.1)	
SGH[d]	62 (38.3)	28 (33.7)	34 (43)	
Sex, n (%)				.82
Female	97 (59.9)	49 (59)	48 (60.8)	
Male	65 (40.1)	34 (41)	31 (39.2)	
Age (y), mean (SD)	48.33 (9.28)	48.47 (8.97)	48.18 (9.64)	.84
Height (cm), mean (SD)	170.92 (10.16)	170.82 (10.85)	171.03 (9.46)	.95
Weight (kg), mean (SD)	101.80 (26.48)	100.95 (26.49)	102.67 (26.60)	.82
Waist circumference (mm), mean (SD)	1125.66 (203.53)	1102.4 (215.93)	1150.69 (188.83)	.29
Heart rate (beats/min), mean (SD)	78.09 (13.73)	78.59 (13.23)	77.56 (14.30)	.63
Marital status, n (%)				.81
Single or never been married	86 (53.1)	42 (50.6)	44 (55.7)	
Married or cohabitating	33 (20.4)	18 (21.7)	15 (19)	
Divorced, widowed, or separated	43 (26.5)	23 (27.7)	20 (25.3)	
Education, n (%)				.86
Lower than college	90 (55.6)	47 (56.6)	43 (54.4)	
Some college	40 (24.7)	19 (22.9)	21 (26.6)	
A college or technical degree	32 (19.8)	17 (20.5)	15 (19)	
Insurance, n (%)				.81
Private health insurance	48 (29.6)	24 (28.9)	24 (30.4)	
Medicare	19 (11.7)	10 (11.4)	9 (12.1)	
Medicaid	67 (45.7)	37 (44.6)	30 (38.0)	
No insurance	27 (16.7)	12 (14.5)	15 (19)	
Unknown or refused to answer	1 (0.6)	0 (0)	1 (1.3)	
Employment status, n (%)				.33
Currently employed	97 (59.9)	53 (63.9)	44 (55.7)	
Others	64 (39.5)	30 (36.1)	34 (43)	
Unknown or refused to answer	1 (0.6)	0 (0)	1 (1.3)	
Annual household income (before taxes; US $), n (%)				.42
<10,000	45 (27.8)	24 (28.9)	21 (26.6)	
10,000-24,999	28 (17.3)	16 (19.3)	12 (15.2)	
25,000-49,999	36 (22.2)	21 (25.3)	15 (19)	
≥50,000	14 (8.6)	11 (13.3)	3 (3.8)	
Unknown, refused to answer, or missing	39 (24.1)	11 (13.3)	28 (35.4)	
Health literacy (measured using the REALM-SF[e]), n (%)				.43
0 (third grade and below)	4 (2.5)	1 (1.2)	3 (3.8)	
1-3 (fourth-sixth grade)	13 (8)	5 (6)	8 (10.1)	
4-6 (seventh-eighth grade)	49 (30.2)	24 (28.9)	25 (31.7)	
7 (high school)	95 (58.6)	53 (63.9)	42 (53.2)	

Participant characteristics and baseline measures	Overall	Intervention (n=83)	Control (n=79)	*P* value[a]
Missing	1 (0.6)	0 (0)	1 (1.3)	
Hypertension knowledge (measured using the HELM[f]), mean (SD)	8.38 (2.37)	8.62 (2.27)	8.11 (2.47)	.19
Patient activation (measured using the PAM[g]), mean (SD)	64.00 (12.67)	64.50 (13.50)	63.48 (11.81)	.60
Health-related quality of life–Physical (measured by SF-12[h] PCS[i]), mean (SD)	42.05 (11.88)	42.23 (11.26)	41.86 (12.57)	.98
Health-related quality of life–Mental (measured by SF-12 MCS[j]), mean (SD)	48.85 (11.75)	50.18 (11.07)	47.46 (12.34)	.18

[a]Descriptive statistics were calculated for demographic variables and baseline measures and compared across the study arms using 2-tailed *t* tests, Wilcoxon rank sum tests, or chi-square tests, as appropriate.

[b]COM: community-based recruitment.

[c]DRH: Detroit Receiving Hospital.

[d]SGH: Sinai-Grace Hospital.

[e]REALM-SF: The Rapid Estimate of Adult Literacy in Medicine–Short Form.

[f]HELM: Hypertension Evaluation of Lifestyle and Management Knowledge Scale.

[g]PAM: Patient Activation Measure.

[h]SF-12: Short Form–12 Health Survey.

[i]PCS: Physical Component Summary.

[j]MCS: Mental Component Summary.

Effect of MI-BP on Outcomes

BP Outcome

For our primary outcome of SBP, sex- and age-adjusted average baseline SBP was comparable between the groups (MI-BP group mean 153.92 mm Hg, SD 2.10; enhanced usual care group mean 153.96 mm Hg, SD 2.15; *P*=.99). Both groups saw a mostly steady and similar decline in SBP over the 12-month intervention (unadjusted means shown in Figure 2A). Table 2 shows a model-based assessment of pairwise differences in adjusted mean SBP for each study arm. At week 52, compared to baseline, the MI-BP group exhibited a 22.5 mm Hg decrease (SE 3.35 mm Hg; *P*<.001), and the control group exhibited a 24.12 mm Hg decrease (SE 3.25 mm Hg; *P*<.001) in average estimated SBP adjusted for age and sex. The average declines were not significantly different between the groups (time-by-arm interaction: *P*=.99). A regression model with a linear time term estimated the rate of decline in the MI-BP group to be 0.3 mm Hg per week (SE 0.05 mm Hg/wk; *P*<.001) and 0.34 mm Hg per week (SE 0.05 mm Hg/wk; *P*<.001) in the control group adjusted for age and sex. Again, these rates were not significantly different between the groups (interaction: *P*=.60; Table S1 in Multimedia Appendix 1). At week 52, overall 58% (19/33) of participants in the intervention group had controlled BP (secondary outcome, defined as SBP <130 mm Hg), whereas 53% (18/34) of participants in the nonintervention group achieved BP control. This difference was not statistically significant (*P*=.89).

Figure 2. Unadjusted mean trajectories (with 95% confidence interval) for (A) systolic blood pressure (SBP) and (B) diastolic blood pressure (DBP) by study arm.

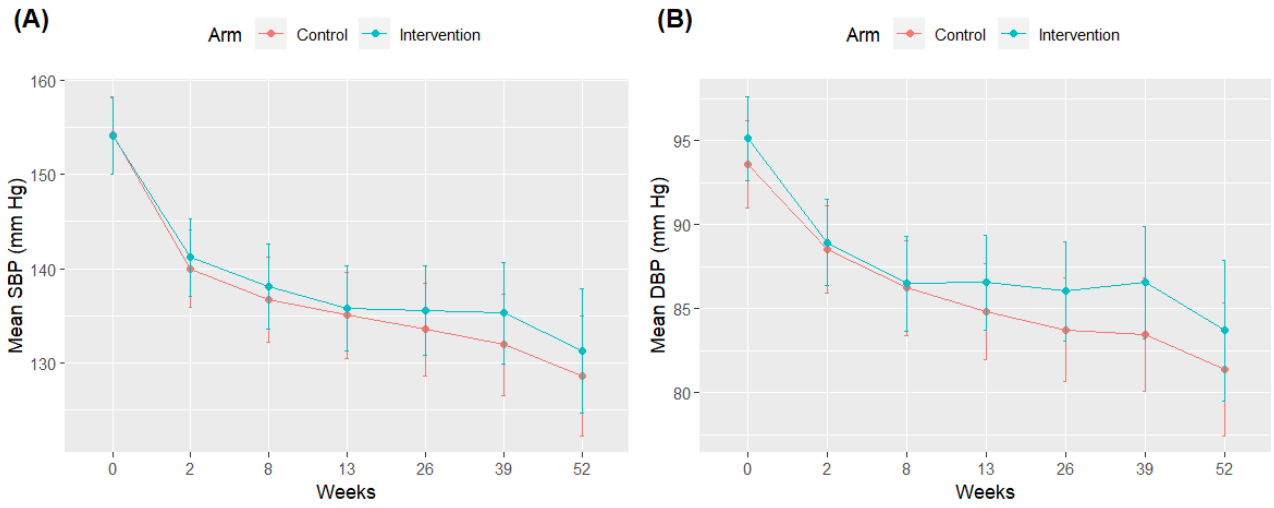


Table 2. Estimated pairwise mean differences and SEs across time for systolic blood pressure by study arm.

Comparison	Intervention		Control	
	Mean difference (mm Hg, time 2 minus time 1) (SE)	*P* value	Mean difference (mm Hg, time 2 minus time 1) (SE)	*P* value
0 vs 2 weeks	–12.8394 (2.4156)	<.001	–14.1392 (2.4466)	<.001
0 vs 8 weeks	–15.3359 (2.5677)	<.001	–16.7350 (2.5790)	<.001
0 vs 13 weeks	–17.9052 (2.5566)	<.001	–18.6283 (2.6044)	<.001
0 vs 26 weeks	–17.4752 (2.6346)	<.001	–19.2175 (2.7335)	<.001
0 vs 39 weeks	–17.3283 (2.8752)	<.001	–21.2452 (2.8972)	<.001
0 vs 52 weeks	–22.5007 (3.3471)	<.001	–24.1173 (3.2502)	<.001
2 vs 8 weeks	–2.4966 (2.5852)	.33	–2.5958 (2.5790)	.31
2 vs 13 weeks	–5.0658 (2.5742)	.05	–4.4891 (2.6044)	.09
2 vs 26 weeks	–4.6358 (2.6520)	.08	–5.0782 (2.7335)	.06
2 vs 39 weeks	–4.4890 (2.8882)	.12	–7.1060 (2.8972)	.01
2 vs 52 weeks	–9.6613 (3.3583)	.004	–9.9781 (3.2502)	.002
8 vs 13 weeks	–2.5692 (2.6855)	.34	–1.8934 (2.6886)	.48
8 vs 26 weeks	–2.1393 (2.7702)	.44	–2.4825 (2.8147)	.38
8 vs 39 weeks	–1.9924 (2.9898)	.51	–4.5102 (2.9664)	.13
8 vs 52 weeks	–7.1647 (3.4428)	.04	–7.3823 (3.3121)	.03
13 vs 26 weeks	0.4300 (2.7428)	.88	–0.5891 (2.8296)	.84
13 vs 39 weeks	0.5768 (2.9774)	.85	–2.6169 (2.9850)	.38
13 vs 52 weeks	–4.5955 (3.4361)	.18	–5.4890 (3.3253)	.10
26 vs 39 weeks	0.1469 (3.0223)	.96	–2.0278 (3.0707)	.51
26 vs 52 weeks	–5.0255 (3.4733)	.15	–4.8999 (3.3983)	.15
39 vs 52 weeks	–5.1723 (3.6255)	.15	–2.8721 (3.5034)	.41

DBP exhibited a very similar pattern to that of SBP (Figure 2B), with the MI-BP and control arms experiencing significant reductions in DBP from baseline to 52 weeks with a 10.20 mm Hg (SE 1.82 mm Hg; *P*<.001) and 11.44 mm Hg (SE 1.75 mm Hg; *P*<.001) estimated average decrease, respectively (Table 3). However, no significant differences were found between the groups (time-by-arm interaction: *P*=.79). Model-based rates of decline were observed in the MI-BP (estimate=0.13 mm Hg/wk; SE 0.03 mm Hg/wk; *P*<.001) and control (estimate=0.17 mm Hg/wk; SE 0.03 mm Hg/wk; *P*<.001; Table S1 in Multimedia Appendix 1) groups. Again, none of these changes were statistically significantly different between the groups (*P*=.21).

Table 3. Estimated pairwise mean differences and SEs across time for diastolic blood pressure by study arm.

Comparison	Intervention		Control	
	Mean difference (mm Hg, time 2–time 1) (SE)	*P* value	Mean difference (mm Hg, time 2–time 1) (SE)	*P* value
0 vs 2 weeks	–6.2249 (1.2940)	<.001	–5.0759 (1.3099)	.001
0 vs 8 weeks	–7.9787 (1.3772)	<.001	–6.8315 (1.3829)	<.001
0 vs 13 weeks	–8.0509 (1.3713)	<.001	–8.1532 (1.3967)	<.001
0 vs 26 weeks	–8.2251 (1.4136)	<.001	–8.8694 (1.4665)	<.001
0 vs 39 weeks	–7.5165 (1.5436)	<.001	–9.5521 (1.5551)	<.001
0 vs 52 weeks	–10.1987 (1.8209)	<.001	–11.4393 (1.7453)	<.001
2 vs 8 weeks	–1.7538 (1.3861)	.21	–1.7555 (1.3829)	.20
2 vs 13 weeks	–1.8261 (1.3803)	.19	–3.0772 (1.3967)	.03
2 vs 26 weeks	–2.0003 (1.4225)	.16	–3.7935 (1.4665)	.01
2 vs 39 weeks	–1.2916 (1.5500)	.41	–4.4762 (1.5551)	.004
2 vs 52 weeks	–3.9738 (1.8263)	.03	–6.3634 (1.7453)	<.001
8 vs 13 weeks	–0.0723 (1.4389)	.96	–1.3217 (1.4403)	.36
8 vs 26 weeks	–0.2465 (1.4854)	.87	–2.0380 (1.5087)	.18
8 vs 39 weeks	0.4622 (1.6035)	.77	–2.7206 (1.5904)	.09
8 vs 52 weeks	–2.2200 (1.8703)	.24	–4.6078 (1.7768)	.01
13 vs 26 weeks	–0.1742 (1.4698)	.91	–0.7163 (1.5164)	.64
13 vs 39 weeks	0.5345 (1.5967)	.74	–1.3989 (1.6003)	.38
13 vs 52 weeks	–2.1477 (1.8669)	.25	–3.2861 (1.7837)	.07
26 vs 39 weeks	0.7087 (1.6201)	.66	–0.6827 (1.6455)	.68
26 vs 52 weeks	–1.9735 (1.8858)	.30	–2.5699 (1.8219)	.16
39 vs 52 weeks	–2.6822 (1.9658)	.17	–1.8872 (1.8775)	.32

Physical Activity

Slight improvements in physical activity over the course of the trial were found for both the MI-BP and enhanced usual care groups as measured using the iPAQ-SF, although the improvements were not statistically significant in general. In the MI-BP group, the age- and sex-adjusted average IPAQ-SF score (after square root transformation) increased by 10.25 metabolic equivalent of task (MET) minutes per week (SE 6.37 MET min/wk; *P*=.11) at 52 weeks, whereas the increase was 10.57 MET minutes per week in the control group (SE 5.87 MET min/wk; *P*=.07; Table S2 in Multimedia Appendix 1). Both groups exhibited fluctuations in the change pattern, where the up-and-down behavior was more prominent in the MI-BP arm (Figure S1 in Multimedia Appendix 2). However, there were no significant differences in the change pattern across the groups (time-by-group interaction: *P*=.93).

Sodium Intake

Figure S2 in Multimedia Appendix 2 shows that mean sodium intake measured using the BSS declined fairly steadily in the MI-BP arm, whereas in the enhanced usual care group, there was a fluctuating pattern in the mean trajectories. However, both arms experienced significant improvements when comparing the baseline with the 52-week values. In the MI-BP group, the average decrease in the adjusted (square root–transformed) BSS score was 0.36 (SE 0.19; *P*=.06), whereas the average decrease in the control arm was 0.60 (SE 0.18; *P*=.001; Table S3 in Multimedia Appendix 1). No significant time-by-group interaction was observed (*P*=.19). We successfully obtained 24-hour urine sodium samples from 136 participants at baseline and 32 participants at 52-week follow-up. In contrast to BSS results, no differential improvements in sodium concentration in 24-hour urine sodium samples were found across arms (time-by-group interaction, *P*=.56).

Medication Adherence

Both treatment groups experienced significant improvements in medication adherence as measured using the ARMS-14 over 1 year (Figure S3 in Multimedia Appendix 2). The average estimated decrease at 52 weeks compared to the baseline in the log-transformed ARMS-14 score in the MI-BP group was 0.20 (SE 0.04; *P*<.001), whereas the corresponding decrease in the control arm was 0.15 (SE 0.04; *P*<.001; Table S4 in Multimedia Appendix 1). No significant difference in the pattern of change was observed (time-by-group interaction: *P*=.30).

Self-Efficacy

Self-efficacies were measured in 4 different ways related to attitude and habits of exercise, medication adherence, and eating habits. Figures S4A and S4B in Multimedia Appendix 2 show

that the trajectories of exercise self-efficacy as measured using 2 subscales of the Self-Efficacy for Exercise Behaviors scale (Sticking to It and Making Time for Exercise) were similar in both study arms. However, in neither arm did the average scores change significantly. In the MI-BP arm, the average estimated decrease in the Sticking to It subscale score at week 52 from the baseline was 0.11 (SE 0.20; *P*=.59), whereas the corresponding decrease in the control arm was 0.01 (SE 0.18; *P*=.94). For the Making Time for Exercise subscale score, the adjusted mean decrease was 0.09 (SE 0.22; *P*=.69) in the MI-BP arm and 0.11 (SE 0.20; *P*=.58) in the control arm (Tables S5 and S6 in Multimedia Appendix 1). As with other outcomes, no significant time-by-group interactions were found (*P*=.88 and *P*=.94 for Sticking to It and Making Time for Exercise, respectively).

For medication adherence self-efficacy (MASES), statistically significant improvements were observed in both arms as the trajectories seemed to follow similar patterns (Figure S4C Multimedia Appendix 2). The increase in the estimated average MASES score in the MI-BP arm was 0.38 (SE 0.16; *P*=.02), and the corresponding increase in the control arm was 0.37 (SE 0.15; *P*=.02; Table S7 in Multimedia Appendix 1). No significant time-by-arm interaction was observed, suggesting no differences in the pattern of change between the groups (overall time-by-group interaction: *P*=.47).

Self-efficacy for eating behaviors showed worse values compared to the baseline (estimated mean 43.95) in the intervention group at week 52 as the estimated mean decreased from a baseline value of 43.95 to 40.04 at week 52 (estimated mean reduction 3.91; SE 1.47; *P*=.008). In contrast, in the control arm, the score increased slightly from baseline (mean 41.64) to week 52 (mean 43.69), although this improvement was not statistically significant (estimated mean improvement 2.05; SE 1.4; *P*=.14; Table S8 in Multimedia Appendix 1; Figure S4D in Multimedia Appendix 2). This was the only outcome for which a significant time-by-group interaction was observed (*P*=.04), albeit in the unintended direction.

Trial Retention

Over the course of the 12-month RCT, we saw steady rates of participant dropout, and only 67 participants (n=33, 49% in the intervention group and n=34, 51% in the control group) remained at the end of the 1-year study (retention rate=67/162, 41.4%). The greatest dropout rates were observed early in the trial between the week 2 and week 8 visits, followed by the periods later in the trial between the week 26 and week 39 visits and between the week 39 and week 52 visits.

Dropout Analysis and Effect of COVID-19

Steady dropout was observed in both treatment arms over the study period, amounting to 60% (50/83) and 57% (45/79) dropout in the MI-BP and control arms, respectively, at the end of study. These dropout patterns were very similar in both study arms (Figure 3). In the time-to-dropout analysis, marital status and COVID-19 group turned out to be statistically significant at the 5% level, with the post–COVID-19 phase showing a strong propensity (Table 4) for dropout (hazard ratio=2.12; SE 0.23; *P*=.001).

Figure 3. Dropout history over the course of the trial by study arm.

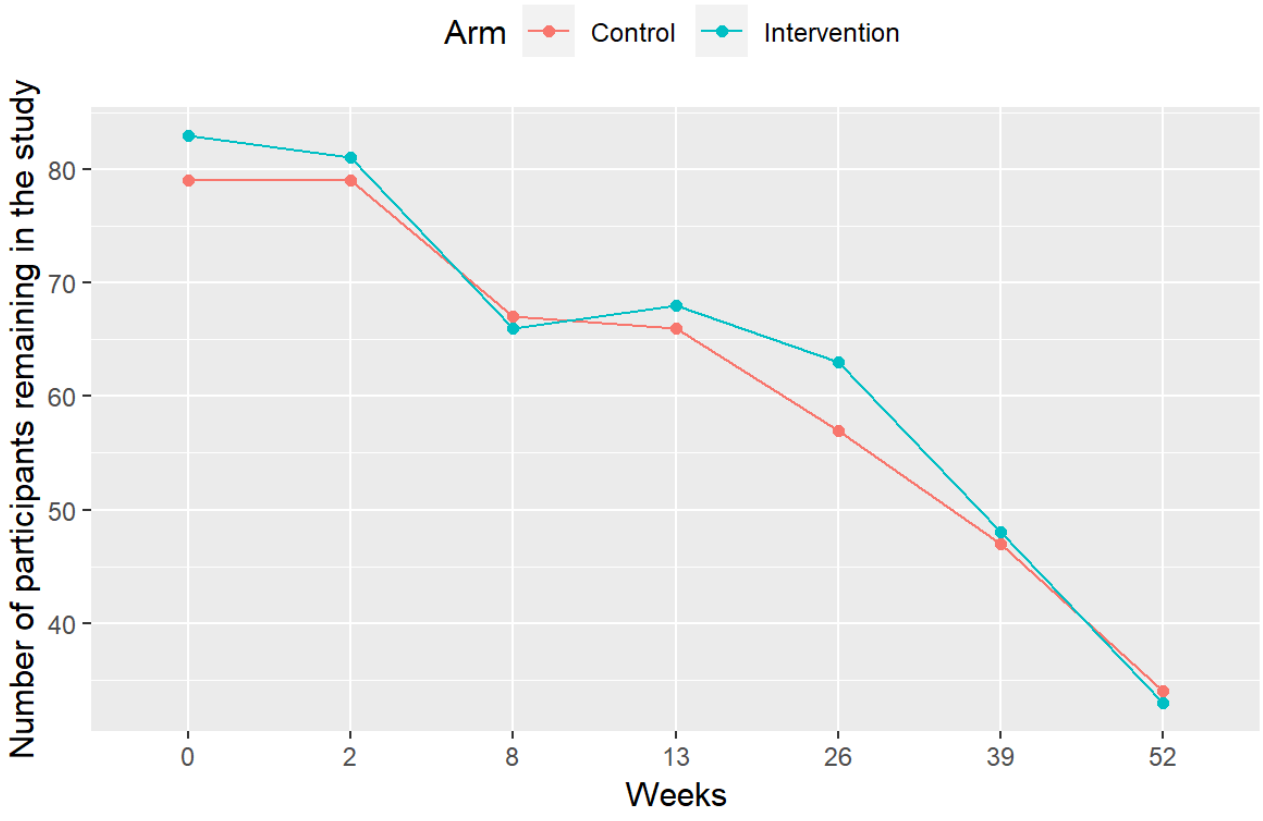


Table 4. Time-to-dropout results based on Cox regression model.

Variable	Hazard ratio (95% CI)	*P* value
COVID-19 group (reference: before COVID-19)	2.120 (1.353-3.322)	.001
Age	0.983 (0.959-1.008)	.18
Marital status (reference: divorced, widowed, or separated)		
Married or cohabitating	1.998 (1.049-3.802)	.04
Single or never been married	1.820 (1.021-3.247)	.04
Sex (reference: male)	0.912 (0.604-1.379)	.66

AE Reporting

AEs for this trial were all determined to be cardiovascular in nature. A total of 15 AEs of SBP>180 were reported. All were determined to be unrelated to the study, and all were reported to both the WSU IRB and the study DSMB. In addition, during the course of this trial, 3 serious AEs (SAEs) were reported by research participants. All 3 SAEs were determined to be unexpected and unrelated to the MI-BP intervention, and all 3 patients recovered with treatment. SAEs reported in this trial included 2 instances of non–ST-elevation myocardial infarction and 1 instance of cerebral visual impairment. All SAEs were reported to both the WSU IRB and the study DSMB.

Discussion

Principal Findings

Results from this study suggest that, compared to our enhanced usual care control group, the MI-BP intervention did not have any significant effects on participants in this study, including the primary outcome measure of BP or the secondary outcome measures of physical activity, sodium intake, medication adherence, and BP control. Even though the trial was underpowered to detect differences due to stopping recruitment early because of the COVID-19 pandemic, trends to suggest that MI-BP had an effect on these outcomes compared to enhanced usual care were not evident. However, it is important to note that, despite a high overall dropout rate, participants in both groups experienced significant reductions in SBP, DBP, and sodium intake, as well as significant increases in physical activity and medication adherence, from baseline to 1 year. This speaks to the general benefit of proactive outreach and engagement focused on cardiometabolic risk reduction in urban-dwelling, low-socioeconomic-status Black populations.

Our findings are similar to those of the recent work by Pletcher et al [32], who found no benefit in terms of SBP reduction for BP self-monitoring using a connected smartphone app compared to standard BP self-monitoring over 6 months. As in our trial, Pletcher et al [32] found that in a sample of 2101 patients with uncontrolled BP, at 6 months, both the intervention and control arms experienced comparable and significant decreases in SBP (–10.8, SD 18 mm Hg vs –10.6, SD 18 mm Hg in the enhanced vs standard group, respectively) with no significant differences between the groups. While our findings and those of Pletcher et al [32] stand in contrast to those of other work suggesting benefit of mHealth apps in reducing BP [14], it should be noted that the evidence base for app-supported self-monitoring of BP is often plagued by short duration and follow-up, small sample sizes, and inconsistent comparison groups, which undermines the quality of research in this area. Moreover, it is important to remember that our control group was not assigned to usual care alone; rather, we used an active control condition where control group participants received a home BP monitor in addition to antihypertensive medication and standard educational materials. This may have led to greater reduction in BP than may have been experienced with standard usual care alone.

The high dropout rate (95/162, 58.6%) among enrolled and randomized participants warrants further mention. While this dropout rate is higher than those in other studies that involved similar digital interventions for monitoring and controlling BP or other study populations [33-35], high dropout rates in studies that deploy digital health interventions are quite common in the mHealth domain and range upward of 80% attrition, with approximately 49% attrition in observational studies and 40% attrition in RCTs. In his seminal piece, Eysenbach [36] described the law of attrition for eHealth interventions, which constitutes the phenomenon of participants dropping out of a research trial before completion or stopping their use of the trial intervention before the study is over. This phenomenon has been described time and time again in the digital health literature and has been specifically evident for mHealth interventions focused on physical activity [37,38], diet [39], and medication adherence (all targeted behaviors in the MI-BP intervention) [40]. Recent studies have demonstrated that a higher dropout rate is even more common for digital health intervention studies involving Black participants [41]. For instance, Jonassaint et al [42] suggested that it may be important to develop a digital intervention system that is culturally tailored to historically marginalized groups. However, our intervention was culturally tailored to the Black community, suggesting room for other explanations for the high dropout rate. We did include a wash-out period to screen out individuals who did not truly have uncontrolled hypertension and identify those who were not fully vested in trial participation, but neither of these is a culture-specific approach. Michaud et al [43] suggested that a high-incentive program may be effective in decreasing the attrition rate for digital health interventions for increasing physical activity in Black women. Although we incorporated a distributed incentive system, which rewarded participants for each visit completion, this was not sufficient to prevent the high dropout rates that echoed those in similar digital intervention studies. It should be mentioned that, particularly when working with historically excluded and underresourced communities, the notion of providing higher incentives to encourage

participation is a hotly contested topic as some believe that it may be considered coercive; however, study incentives are meant to acknowledge participant burden, such as loss of time, and differentially incentivizing study participants based on level of advantage introduces its own host of ethical conundrums.

Our findings, coupled with the high dropout rates found in similar studies, suggest that these types of mHealth behavior change interventions may not be a complete solution that can promote behavior change and improve health outcomes in this population. Rather, mHealth may have the greatest potential as part of a suite of approaches available to health care professionals and patients. It is clear that mHealth solutions are here to stay, but the goal of future research should be to identify the use cases and implementation strategies and factors that contribute to optimization. Moreover, the expectation of high dropout rates for mHealth interventions, especially when working with challenging populations with considerable barriers caused by social determinants of health, should be assumed and addressed at the outset. This is important as underpowered samples compromise the quality of research studies and the evidence base; however, review panels for different funding mechanisms often look unfavorably on proposed research studies that anticipate very high dropout rates. This may cause researchers to intentionally underestimate attrition, which may compromise the research study as a whole.

Limitations

As with all research, this study was not without limitations. Perhaps the largest limitation was the undeniable negative impact of the COVID-19 pandemic, which caused us to stop study recruitment early and likely had significant effects on trial participation for those who were already randomized to the study. Combined with a high dropout rate, early study termination led to an underpowered study, which may have caused us to be unable to find significant differences between the groups. That said, data trends across our participants suggest that both groups experienced significant improvement in both primary and secondary outcomes, and if there really was a benefit of the MI-BP intervention, it was likely much smaller than our initially projected effect size. Given the absence of an indication of differential improvements, future efforts would be better served simply focusing on scalable outreach and engagement programs that facilitate better care of Black patients with uncontrolled hypertension.

We should also note that, although we consented and enrolled 869 individuals in the ED and other community-based settings, 464 (53.4%) were considered lost to follow-up before randomization. Most of this loss to follow-up occurred before the baseline visit (Figure 1). Although we cannot say for certain why there was such tremendous loss to follow-up before randomization, our de facto wash-out period helped remove individuals who were not fully committed to participation in the trial. Given the large loss to follow-up before randomization, we most certainly had some degree of selection bias in our randomized sample. In addition to our loss to follow-up before randomization, as discussed, we experienced high dropout rates in our trial arms, which serves as a potential threat to the validity and generalizability of our findings. We also note that our quarterly follow-up assessments may have helped some participants stay engaged with the intervention and trial, masking even further attrition that may have been experienced without frequent contact. As suggested previously, very high rates of attrition are not uncommon in mHealth studies, and our trial was no exception. More research on how to best engage participants in these types of interventions and trials is desperately needed as the mechanisms that drive engagement are poorly understood. Moreover, our work was met with additional challenges as we are positioned in a community where social determinants of health play an enormous role in the daily lives of our participants. These additional challenges are sure to have contributed to our high attrition rates. While attrition may serve as a threat to the validity of our findings, it is important to remember that this is often the reality of working with populations of individuals who experience health inequity. Rather than shying away from work in this area due to methodological and statistical concerns, we must continue to conduct research with populations of individuals with great need to address inequities.

Conclusions

Although we did not find increased improvement in outcomes for participants using MI-BP compared to enhanced usual care, we are encouraged to see that proactive outreach in a sample of Black individuals with uncontrolled hypertension recruited from EDs and other community-based settings had a significant effect in improving hypertension-related outcomes regardless of treatment group. Given that mHealth approaches for chronic disease self-management are becoming more commonplace, continued work is needed to understand how to better engage and retain users as well as how to better position these types of interventions within a suite of treatment options available to patients. Moreover, participant dropout in mHealth interventions remains high in many studies, including ours, and this phenomenon must be further explored before optimal mHealth use can be achieved.

Acknowledgments

This study was funded by the National Heart, Lung, and Blood Institute at the National Institutes of Health (1R01HL127215-01A1). The authors of this study wish to thank Drs John Flack, MD; Richard Sinert, DO; and Philip Boonstra, PhD, for their service on the MI-BP trial Data Safety and Monitoring Board. Their support of this trial and their guidance were unparalleled and much appreciated. The authors also wish to thank Dr Paula Einhorn, the program officer at the National Heart, Lung, and Blood Institute, for her support. Additional thanks go to Leah Huang and Alexandria Degner for editing assistance as well as Lilly Pritula and Judy Connelly for administrative and grant management support.

Conflicts of Interest

LRB is an Editor in Chief for JMIR mHealth and uHealth.

Multimedia Appendix 1
Supplementary tables for primary and secondary outcomes.
[DOCX File , 35 KB - mhealth_v12i1e57863_app1.docx]

Multimedia Appendix 2
Supplementary figures for secondary outcome trajectories over time.
[DOCX File , 522 KB - mhealth_v12i1e57863_app2.docx]

Multimedia Appendix 3
CONSORT eHEALTH checklist (V 1.6.2).
[PDF File (Adobe PDF File), 303 KB - mhealth_v12i1e57863_app3.pdf]

References

1. Tsao CW, Aday AW, Almarzooq ZI, Alonso A, Beaton AZ, Bittencourt MS, et al. Heart Disease and Stroke Statistics-2022 Update: A Report From the American Heart Association. Circulation 2022 Feb 22;145(8):e153-e639 [FREE Full text] [doi: 10.1161/CIR.0000000000001052] [Medline: 35078371]
2. Feigin VL, Norrving B, Mensah GA. Global burden of stroke. Circ Res 2017 Feb 03;120(3):439-448. [doi: 10.1161/circresaha.116.308413]
3. Muntner P, Carey RM, Gidding S, Jones DW, Taler SJ, Wright JT, et al. Potential US population impact of the 2017 ACC/AHA high blood pressure guideline. Circulation 2018 Jan 09;137(2):109-118. [doi: 10.1161/circulationaha.117.032582]
4. O'Donnell MJ, Xavier D, Liu L, Zhang H, Chin SL, Rao-Melacini P, INTERSTROKE investigators. Risk factors for ischaemic and intracerebral haemorrhagic stroke in 22 countries (the INTERSTROKE study): a case-control study. Lancet 2010 Jul 10;376(9735):112-123. [doi: 10.1016/S0140-6736(10)60834-3] [Medline: 20561675]
5. Valensi P, Lorgis L, Cottin Y. Prevalence, incidence, predictive factors and prognosis of silent myocardial infarction: a review of the literature. Arch Cardiovasc Dis 2011 Mar;104(3):178-188 [FREE Full text] [doi: 10.1016/j.acvd.2010.11.013] [Medline: 21497307]
6. Whelton PK, Carey RM, Aronow WS, Casey DE, Collins KJ, Dennison Himmelfarb C, et al. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA guideline for the prevention, detection, evaluation, and management of high blood pressure in adults: a report of the American College of Cardiology/American Heart Association Task Force on clinical practice guidelines. J Am Coll Cardiol 2018 May 15;71(19):e127-e248 [FREE Full text] [doi: 10.1016/j.jacc.2017.11.006] [Medline: 29146535]
7. Aggarwal R, Chiu N, Wadhera RK, Moran AE, Raber I, Shen C, et al. Racial/ethnic disparities in hypertension prevalence, awareness, treatment, and control in the United States, 2013 to 2018. Hypertension 2021 Dec;78(6):1719-1726 [FREE Full text] [doi: 10.1161/HYPERTENSIONAHA.121.17570] [Medline: 34365809]
8. Mullins PM, Levy PD, Mazer-Amirshahi M, Pines JM. National trends in U.S. emergency department visits for chief complaint of hypertension (2006-15). Am J Emerg Med 2020 Aug;38(8):1652-1657. [doi: 10.1016/j.ajem.2019.10.004] [Medline: 31848039]
9. Ndumele CD, Baer HJ, Shaykevich S, Lipsitz SR, Hicks LS. Cardiovascular disease and risk in primary care settings in the United States. Am J Cardiol 2012 Feb 15;109(4):521-526 [FREE Full text] [doi: 10.1016/j.amjcard.2011.09.047] [Medline: 22112741]
10. Uppal TS, Chehal PK, Fernandes G, Haw JS, Shah M, Turbow S, et al. Trends and variations in emergency department use associated with diabetes in the US by sociodemographic factors, 2008-2017. JAMA Netw Open 2022 May 02;5(5):e2213867 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.13867] [Medline: 35612855]
11. Valenzuela PL, Carrera-Bastos P, Gálvez BG, Ruiz-Hurtado G, Ordovas JM, Ruilope LM, et al. Lifestyle interventions for the prevention and treatment of hypertension. Nat Rev Cardiol 2021 Apr 09;18(4):251-275. [doi: 10.1038/s41569-020-00437-9] [Medline: 33037326]
12. Skolarus LE, Sharrief A, Gardener H, Jenkins C, Boden-Albala B. Considerations in addressing social determinants of health to reduce racial/ethnic disparities in stroke outcomes in the United States. Stroke 2020 Nov;51(11):3433-3439 [FREE Full text] [doi: 10.1161/STROKEAHA.120.030426] [Medline: 33104471]
13. Mobile fact sheet: mobile phone ownership over time. Pew Research Center. 2021. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2023-10-03]
14. Buis LR, Dawood K, Kadri R, Dawood R, Richardson CR, Djuric Z, et al. Improving blood pressure among African Americans with hypertension using a mobile health approach (the MI-BP App): protocol for a randomized controlled trial. JMIR Res Protoc 2019 Jan 25;8(1):e12601 [FREE Full text] [doi: 10.2196/12601] [Medline: 30681965]

15. Krein SL, Kadri R, Hughes M, Kerr EA, Piette JD, Holleman R, et al. Pedometer-based internet-mediated intervention for adults with chronic low back pain: randomized controlled trial. J Med Internet Res 2013 Aug 19;15(8):e181 [FREE Full text] [doi: 10.2196/jmir.2605] [Medline: 23969029]
16. Martinez CH, Moy ML, Nguyen HQ, Cohen M, Kadri R, Roman P, et al. Taking healthy steps: rationale, design and baseline characteristics of a randomized trial of a pedometer-based internet-mediated walking program in veterans with chronic obstructive pulmonary disease. BMC Pulm Med 2014 Feb 03;14(1):12 [FREE Full text] [doi: 10.1186/1471-2466-14-12] [Medline: 24491137]
17. Moy ML, Weston NA, Wilson EJ, Hess ML, Richardson CR. A pilot study of an Internet walking program and pedometer in COPD. Respir Med 2012 Sep;106(9):1342-1350 [FREE Full text] [doi: 10.1016/j.rmed.2012.06.013] [Medline: 22795984]
18. Richardson CR, Buis LR, Janney AW, Goodrich DE, Sen A, Hess ML, et al. An online community improves adherence in an internet-mediated walking program. Part 1: results of a randomized controlled trial. J Med Internet Res 2010 Dec 17;12(4):e71 [FREE Full text] [doi: 10.2196/jmir.1338] [Medline: 21169160]
19. Richardson CR, Mehari KS, McIntyre LG, Janney AW, Fortlage LA, Sen A, et al. A randomized trial comparing structured and lifestyle goals in an internet-mediated walking program for people with type 2 diabetes. Int J Behav Nutr Phys Act 2007 Nov 16;4(1):59 [FREE Full text] [doi: 10.1186/1479-5868-4-59] [Medline: 18021411]
20. Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Ainsworth BE, et al. International physical activity questionnaire: 12-country reliability and validity. Med Sci Sports Exerc 2003 Aug;35(8):1381-1395. [doi: 10.1249/01.MSS.0000078924.61453.FB] [Medline: 12900694]
21. Norris JB, Block G, Block T. Development of a brief screening tool for sodium intake. In: Proceedings of the 140th American Public Health Association Conference on Annual Meeting and Exposition. 2012 Presented at: APHA '12; October 27-31, 2012; San Francisco, CA URL: https://apha.confex.com/apha/140am/webprogram/Paper270258.html
22. Kripalani S, Risser J, Gatti ME, Jacobson TA. Development and evaluation of the adherence to refills and medications scale (ARMS) among low-literacy patients with chronic disease. Value Health 2009 Jan;12(1):118-123 [FREE Full text] [doi: 10.1111/j.1524-4733.2008.00400.x] [Medline: 19911444]
23. Sallis J, Pinski R, Grossman R, Patterson T, Nader P. The development of self-efficacy scales for healthrelated diet and exercise behaviors. Health Educ Res 1988;3(3):283-292. [doi: 10.1093/her/3.3.283]
24. Ogedegbe G, Mancuso CA, Allegrante JP, Charlson ME. Development and evaluation of a medication adherence self-efficacy scale in hypertensive African-American patients. J Clin Epidemiol 2003 Jun;56(6):520-529. [doi: 10.1016/s0895-4356(03)00053-2] [Medline: 12873646]
25. Schapira MM, Fletcher KE, Hayes A, Eastwood D, Patterson L, Ertl K, et al. The development and validation of the hypertension evaluation of lifestyle and management knowledge scale. J Clin Hypertens (Greenwich) 2012 Jul;14(7):461-466 [FREE Full text] [doi: 10.1111/j.1751-7176.2012.00619.x] [Medline: 22747619]
26. Arozullah AM, Yarnold PR, Bennett CL, Soltysik RC, Wolf MS, Ferreira RM, et al. Development and validation of a short-form, rapid estimate of adult literacy in medicine. Med Care 2007 Nov;45(11):1026-1033. [doi: 10.1097/MLR.0b013e3180616c1b] [Medline: 18049342]
27. Hibbard JH, Mahoney ER, Stockard J, Tusler M. Development and testing of a short form of the patient activation measure. Health Serv Res 2005 Dec;40(6 Pt 1):1918-1930 [FREE Full text] [doi: 10.1111/j.1475-6773.2005.00438.x] [Medline: 16336556]
28. Ware Jr J, Kosinski M, Keller SD. A 12-item short-form health survey: construction of scales and preliminary tests of reliability and validity. Med Care 1996 Mar;34(3):220-233. [doi: 10.1097/00005650-199603000-00003] [Medline: 8628042]
29. Appel LJ, Champagne CM, Harsha DW, Cooper LS, Obarzanek E, Elmer PJ, Writing Group of the PREMIER Collaborative Research Group. Effects of comprehensive lifestyle modification on blood pressure control: main results of the PREMIER clinical trial. JAMA 2003;289(16):2083-2093. [doi: 10.1001/jama.289.16.2083] [Medline: 12709466]
30. Buis L, Hirzel L, Dawood RM, Dawood KL, Nichols LP, Artinian NT, et al. Text messaging to improve hypertension medication adherence in African Americans from primary care and emergency department settings: results from two randomized feasibility studies. JMIR Mhealth Uhealth 2017 Feb 01;5(2):e9 [FREE Full text] [doi: 10.2196/mhealth.6630] [Medline: 28148474]
31. Appel LJ, Champagne CM, Harsha DW, Cooper LS, Obarzanek E, Elmer PJ, Writing Group of the PREMIER Collaborative Research Group. Effects of comprehensive lifestyle modification on blood pressure control: main results of the PREMIER clinical trial. JAMA 2003;289(16):2083-2093. [doi: 10.1001/jama.289.16.2083] [Medline: 12709466]
32. Pletcher MJ, Fontil V, Modrow MF, Carton T, Chamberlain AM, Todd J, et al. Effectiveness of standard vs enhanced self-measurement of blood pressure paired with a connected smartphone application: a randomized clinical trial. JAMA Intern Med 2022 Oct 01;182(10):1025-1034 [FREE Full text] [doi: 10.1001/jamainternmed.2022.3355] [Medline: 35969408]
33. Kassavou A, Mirzaei V, Shpendi S, Brimicombe J, Chauhan J, Bhattacharya D, et al. The feasibility of the PAM intervention to support treatment-adherence in people with hypertension in primary care: a randomised clinical controlled trial. Sci Rep 2021 Apr 26;11(1):8897 [FREE Full text] [doi: 10.1038/s41598-021-88170-2] [Medline: 33903656]
34. McManus RJ, Little P, Stuart B, Morton K, Raftery J, Kelly J, HOME BP investigators. Home and online management and evaluation of blood pressure (HOME BP) using a digital intervention in poorly controlled hypertension: randomised controlled trial. BMJ 2021 Jan 19;372:m4858 [FREE Full text] [doi: 10.1136/bmj.m4858] [Medline: 33468518]

35. Skolarus LE, Cowdery J, Dome M, Bailey S, Baek J, Byrd JB, et al. Reach out churches: a community-based participatory research pilot trial to assess the feasibility of a mobile health technology intervention to reduce blood pressure among African Americans. Health Promot Pract 2018 Jul 05;19(4):495-505 [FREE Full text] [doi: 10.1177/1524839917710893] [Medline: 28583024]
36. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
37. Buckingham SA, Williams AJ, Morrissey K, Price L, Harrison J. Mobile health interventions to promote physical activity and reduce sedentary behaviour in the workplace: a systematic review. Digit Health 2019 Mar 27;5:2055207619839883 [FREE Full text] [doi: 10.1177/2055207619839883] [Medline: 30944728]
38. Fukuoka Y, Gay C, Haskell W, Arai S, Vittinghoff E. Identifying factors associated with dropout during prerandomization run-in period from an mHealth physical activity education study: the mPED trial. JMIR Mhealth Uhealth 2015 Apr 13;3(2):e34 [FREE Full text] [doi: 10.2196/mhealth.3928] [Medline: 25872754]
39. Villinger K, Wahl DR, Boeing H, Schupp HT, Renner B. The effectiveness of app-based mobile interventions on nutrition behaviours and nutrition-related health outcomes: a systematic review and meta-analysis. Obes Rev 2019 Oct 28;20(10):1465-1484 [FREE Full text] [doi: 10.1111/obr.12903] [Medline: 31353783]
40. Alvarado MM, Kum H, Gonzalez Coronado K, Foster MJ, Ortega P, Lawley MA. Barriers to remote health interventions for type 2 diabetes: a systematic review and proposed classification scheme. J Med Internet Res 2017 Feb 13;19(2):e28 [FREE Full text] [doi: 10.2196/jmir.6382] [Medline: 28193598]
41. Idris MY, Mubasher M, Alema-Mensah E, Awad C, Vordzorgbe K, Ofili E, et al. The law of non-usage attrition in a technology-based behavioral intervention for black adults with poor cardiovascular health. PLOS Digit Health 2022 Oct 25;1(10):e0000119 [FREE Full text] [doi: 10.1371/journal.pdig.0000119] [Medline: 36812567]
42. Jonassaint CR, Gibbs P, Belnap BH, Karp JF, Abebe KK, Rollman BL. Engagement and outcomes for a computerised cognitive-behavioural therapy intervention for anxiety and depression in African Americans. BJPsych Open 2017 Jan;3(1):1-5 [FREE Full text] [doi: 10.1192/bjpo.bp.116.003657] [Medline: 28058109]
43. Michaud TL, Estabrooks PA, You W, McGuire TJ, Almeida F, Karls K, et al. Sustaining the reach of a scalable weight loss intervention through financial incentives- a pragmatic, feasibility, online randomized trial protocol. Contemp Clin Trials 2020 Nov;98:106142 [FREE Full text] [doi: 10.1016/j.cct.2020.106142] [Medline: 32920241]

Abbreviations

AE: adverse event
ARMS-14: Adherence to Refills and Medication Scale
BP: blood pressure
BSS: Block Sodium Screener
CONSORT: Consolidated Standards of Reporting Trials
DBP: diastolic blood pressure
DSMB: Data and Safety Monitoring Board
ED: emergency department
HELM: Hypertension Evaluation of Lifestyle and Management
IPAQ-SF: International Physical Activity Questionnaire–Short Form
IRB: institutional review board
MASES: Medication Adherence Self-Efficacy Scale
MET: metabolic equivalent of task
mHealth: mobile health
PAM: Patient Activation Measure
PCP: primary care provider
RCT: randomized controlled trial
REALM-SF: Rapid Assessment of Adult Literacy in Medicine–Short Form
SAE: serious adverse event
SBP: systolic blood pressure
SF-12: Short Form–12
WSU: Wayne State University

Edited by T de Azevedo Cardoso; submitted 28.02.24; peer-reviewed by Y Kong, RJ Katz; comments to author 28.03.24; revised version received 12.04.24; accepted 30.04.24; published 28.06.24.

Please cite as:
Buis LR, Kim J, Sen A, Chen D, Dawood K, Kadri R, Muladore R, Plegue M, Richardson CR, Djuric Z, McNaughton C, Hutton D, Robert LP, Park SY, Levy P
The Effect of an mHealth Self-Monitoring Intervention (MI-BP) on Blood Pressure Among Black Individuals With Uncontrolled Hypertension: Randomized Controlled Trial
JMIR Mhealth Uhealth 2024;12:e57863
URL: https://mhealth.jmir.org/2024/1/e57863
doi:10.2196/57863
PMID:

Original Paper

Personalized Smartphone-Enabled Assessment of Blood Pressure and Its Treatment During the SARS-CoV-2 COVID-19 Pandemic in Patients From the CURE-19 Study: Longitudinal Observational Study

Leanne Richardson[1], PhD; Nihal Noori[1], MSc; Jack Fantham[1], MSc; Gregor Timlin[1], MA; James Siddle[1], MSc; Thomas Godec[1], PhD; Mike Taylor[1], PhD; Charles Baum[2], MD

[1]Closed Loop Medicine, Cambridge, United Kingdom

[2]Encore Health, Chicago, IL, United States

Corresponding Author:
Charles Baum, MD
Encore Health
1901 S Calumet Ave
Chicago, IL, 60616
United States
Phone: 44 20 3023 0644
Email: baum73@gmail.com

Abstract

Background: The use of digital interventions by patients for remote monitoring and management of health and disease is increasing. This observational study examined the feasibility, use, and safety of a digital smartphone app for routine monitoring of blood pressure (BP), medication, and symptoms of COVID-19 during the COVID-19 pandemic.

Objective: The objective of this study was to deploy and test electronic data recording using a smartphone app developed for routine monitoring of BP in patients with primary hypertension. We tested the app for ease of data entry in BP management and tracking symptoms of new-onset COVID-19 to determine if participants found this app approach useful and sustainable.

Methods: This remote, decentralized, 12-week, prospective, observational study was conducted in a community setting within the United States. Participants were approached and recruited from affiliated sites where they were enrolled in an ongoing remote decentralized study (CURE-19) of participants experiencing the COVID-19 pandemic. Potential participants were asked to complete a digital screener to determine eligibility and given informed consent forms to read and consent to using the Curebase digital platform. Following enrollment, participants downloaded the digital app to their smartphones for all data collection. Participants recorded daily BP, associated medication use, and emergent symptoms associated with SARS-CoV-2 infection. In addition, usability (adherence, acceptability, and user experience) was assessed using standard survey questions. Adverse events were collected based on participant self-report. Compliance and engagement were determined from user data entry rates. Feasibility and participant feedback were assessed upon study completion using the User Experience Questionnaire.

Results: Of the 389 participants who enrolled in and completed the study, 380 (98%) participants downloaded and entered BP routines in week 1. App engagement remained high; 239 (62.9%) of the 380 participants remained in the study for the full 12-week observation period, and 201 (84.1%) of the 239 participants entered full BP routines into the digital app 80% or more of the time. The smartphone app scored an overall positive evaluation as assessed by the User Experience Questionnaire and was benchmarked as “excellent” for domains of perspicuity, efficiency, and dependability and “above average” for domains of attractiveness and stimulation. Highly adherent participants with hypertension demonstrated well-controlled BP, with no significant changes in average systolic or diastolic BP between week 1 and week 12 (all $P>.05$). Participants were able to record BP medications and symptoms of SARS-CoV-2 infection. No adverse events attributable to the use of the smartphone app were reported during the observational period.

Conclusions: The high retention, engagement and acceptability and positive feedback in this study demonstrates that routine monitoring of BP and medications using a smartphone app is feasible for patients with hypertension in a community setting. Remote monitoring of BP and data collection could be coupled with hypertensive medication in a combination product (drug+digital) for precision management of hypertension.

(JMIR Mhealth Uhealth 2024;12:e53430) doi:10.2196/53430

KEYWORDS

digital diary; hypertension; blood pressure; remote monitoring; smartphone app; mobile phone; app; monitoring; COVID-19; SARS-CoV-2; digital intervention; management; observational study; deployment; feasibility; use; safety; medication; symptoms; community; systolic; diastolic; utilization

Introduction

Hypertension is the leading preventable cause of premature death worldwide [1]. Prevalence of hypertension in the United States is 44% to 49% with an estimated 116 million having the disease [2]. Physician inertia (inadequate up-titration of treatment, especially from monotherapy) and poor patient adherence to treatment (especially when based on multiple pills) are now recognized as major factors contributing to poor blood pressure (BP) control [1,3,4].

During the COVID-19 pandemic, social distancing was a priority measure to limit infection in the general population. This measure alongside the redeployment of clinicians to frontline medicine restricted the medical management of hypertension. In turn, this led to the desire to move as much medical management as possible to arms-length support. To meet this need, health care providers rapidly moved to remote monitoring and patient engagement tools where possible. This trend gained momentum, and as a result, over the past 3 years, a 38-fold increase in the use of remote communication technology to deliver health care services (telehealth) in the United States has been observed [5]. This, alongside a strong patient preference to use telehealth services for ordering medications, clinical visit preparation, and receiving educational material and test results, suggests that widespread adoption of telehealth services by health care providers, insurance companies, and patients will continue [6]. Digitization and automation of these services not only can increase patient access to health services but can also build efficiencies within the health care system.

These recent changes have occurred against a backdrop of longer-term technical developments, allowing the delivery of personalized drug treatment to an individual to get much closer to being feasible. This; the use of electronic diaries; and eventually, as more data are collected over time, decision support by mobile apps, could make that a reality. Our team (Closed Loop Medicine Ltd [CLM]) has expertise in the design and development of electronic apps to develop patient treatment solutions that allow combinations of drug and nondrug therapies. One such solution is a smartphone app created by CLM and coproduced by patients through patient engagement and user testing. This prototype app allows patients' BP diary entries, collection of hypertensive adverse drug reactions, and other patient-reported data to be routed to a study database for analysis and surveillance.

When the full scale of the COVID-19 pandemic became apparent, we responded by adapting the prototype app to collect COVID-19 symptoms and hypertension drug regimen information, using the existing patient experience and data collection technology. This corresponded to our early design thinking for a general-purpose platform. The app was made available to study participants through Google and Apple app stores. This allowed deployment of the app into the community setting within the United States in an observational study facilitated by our study partner Curebase. Additional data collection and integration components and a financial reconciliation program were developed to facilitate the study's operational aspects.

The information gained from this study regarding BP measurements, type of drug taken and symptom data, and the learnings and experience of delivering this study will inform future product development plans and future iterations of the hypertension platform to best meet the needs of health care providers and patients. Furthermore, the information gained will provide evidence that digitization of health care is feasible and a valid means of remote management of hypertension within the existing health care system.

Therefore, the observational study aimed to rapidly modify, deploy, and test a digital smartphone app for routine monitoring of BP, BP-associated medications, and symptoms of SARS-CoV-2 infection in the community setting during the COVID-19 pandemic. This testing would provide valuable information for further development of the app before use in future interventional clinical trials. Alongside this, we examined core interactions and use of the app by patients with hypertension, assessed by participant engagement, compliance, and a User Experience Questionnaire (UEQ) [7]. From a clinical point of view, we assessed BP control in a community setting within the United States and whether BP drug use or level of BP control affected the risk of SAR-CoV-2 infection. We also assessed the safety of participants using the app. In addition, this study established a well-defined cohort of patients ready for recruitment into an interventional trial, should we, or other research groups, need this on account of new insights into COVID-19 and the requirement for rapid study execution.

Methods

Personalized Electronic Record Development and Deployment

A mobile app was created for use in this observational study setting. The app was developed as a data collection interface and modified to additionally collect COVID-19 symptoms as well as current antihypertensive medications, based on a curated set of hypertension drugs and doses that were identified as potentially impacting COVID-19 symptoms. These modifications, along with the existing prototype app experience and an incentives scheme rewarding patient adherence, were used to promote patient engagement in the data collection process. Table 1 lists the standard behavior change techniques [8] and the corresponding participant engagement features used to encourage data entry in the study. Some of the corresponding

user interfaces used in the study are shown in Multimedia Appendix 1.

As shown in the system diagram (Multimedia Appendix 2), the mobile app was supported by several shared backend services hosted in Amazon Web Services. These were extended to support storage of the new data and to allow the data to be shared with CLM's study partner Curebase. A financial reconciliation service was used, where payment eligibility was calculated on the basis of study participation level, then passed on to Curebase in support of a regularly scheduled payment process.

New Google and Apple app store entries were created and configured so that patients in the United States were able to install and run the app. Patient study enrollment was managed using a previously established approach based on a unique patient identifier and an enrollment code shared with the patient, which was then linked to their mobile device. A new set of patient identifiers and enrollment codes was generated and shared with Curebase, which integrated these into their patient training and onboarding process.

Table 1. Participant engagement features.

Behavior changes technique taxonomy [8]	Participant engagement features
Self-monitoring of behavior: Establish a method for the person to monitor and record their behaviors as part of a behavior change strategy.	• Blood pressure monitoring: Enable participants to enter their blood pressure readings into the app. • COVID-19 symptom monitoring: Enable participants to enter any COVID-19 symptoms they experience. • Hypertension medication monitoring: Enable participants to maintain a list of their current hypertensive medications.
Demonstration of the behavior: Provide an observable sample of the performance of the behavior, directly in person or indirectly, for example, through film, pictures, for the person to aspire to or imitate.	• Blood pressure measurement demonstration: Educate participants on the correct way to take consistent blood pressure readings.
Instruction on how to perform a behavior: Advise or agree on how to perform the behavior.	• COVID-19 symptom information: Provide participants with clear information about COVID-19 symptoms so they can correctly identify any symptoms they are experiencing.
Habit formation: Prompt rehearsal and repetition of the behavior in the same context repeatedly so that the context elicits the behavior.	• Morning and evening routine management: Enable consistent routine formation by providing a morning and evening data entry ritual.
Prompts or cues: Introduce or define environmental or social stimulus with the purpose of prompting or cueing the behavior. The prompt or cue would normally occur at the time or place of performance.	• Smartphone reminders: Enable participants to set routine reminders to enter data into the app.
Feedback on outcome of behavior: Monitor and provide feedback on the outcome of performance of the behavior.	• Data history: Provide participants with clear diagrams of previous blood pressure and COVID-19 symptom entries.
Incentive (outcome): Inform that a reward will be delivered if and only if there has been effort or progress in achieving the behavioral outcome.	• Participant incentives: Participants who downloaded the app successfully were given US $10 and then an additional US $10 each week for 12 weeks if all study activities were completed. Participants could continue using the app after the 12 weeks but were not compensated for completion of activities.

Study Design

This remote, decentralized, 12-week, prospective, observational study was conducted in the community with no participant visits to investigational sites during the COVID-19 pandemic. The study consisted of an observational period of 12 weeks, with an optional 12-week extension. Participants recorded daily BP (morning and evening) and associated medication use, as well as any emergent symptoms indicative of SARS-CoV-2 infection.

Participants

Participants were aged 18 years and older with hypertension, defined as receiving prescription drug treatment (minimum of 1 drug) to reduce BP. To take part, participants had to be willing to provide informed consent; be able to speak, read, and comprehend English; and possess a suitable home BP monitoring device (as a BP monitor would not be provided to participants) and a suitable smartphone (to support iOS versions 10.0 [Apple Inc] and newer or Android versions 5.0 [Google] and newer; minimum storage space required to install the digital app) that they could independently use. Main exclusions included any known or suspected COVID-19 symptoms at enrollment, comorbidities incompatible with study participation (ie, that would inhibit completion of daily smartphone app entries), and limited or no understanding of spoken or written English.

Participants were recruited to the study through internet-based recruitment to a larger study called CURE-19 being undertaken by Curebase (contract research organization [CRO]). Study personnel at clinical sites affiliated with the CRO approached and recruited participants from their patient population. Those who were interested in taking part of the study were then screened for eligibility, and if deemed eligible, they continued to the enrollment phase. Digital recruitment methods including

recruitment campaigns through social media and internet advertisement were also used. Potential participants who encountered recruitment information through social media or on the internet expressed interest by clicking on the content, which led them to a study landing page website. To assess eligibility, a digital screener containing relevant questions was used, asking potential participants questions to determine their eligibility. All participants who submitted screeners were asked to provide their email address in order to create an account. Screening responses were automatically analyzed and an outcome for the patient was generated, either eligible or ineligible. If eligible, patients were given informed consent forms to read and were consented using the CRO's digital platform, either remotely on their own personal device or in the presence of study personnel. Electronic signatures were captured and stored electronically in the Curebase platform.

Study Procedures

Following consent, participants were given instructions through the CRO's remote onboarding system to download the digital smartphone app from the app store on their personal smartphone and complete the baseline data entry in the app. In some instances, remote onboarding also included telephone calls as well as emails sent to the participants. Participants were then asked to enter routine daily recordings of BP (morning and evening), BP medication use, and emergent symptoms of SAR-CoV-2 infection for 12 weeks. Following the initial 12-week observational period, participants could continue to an additional optional 12-week extension period. Demographic and adverse event information was collected from participants by self-report and electronic medical records by an electronic case report form. Participants received internet-based surveys to complete in order to collect self-reported data. At the completion of 12 weeks or 24 weeks, the UEQ was sent to the participants as a Google Form to allow participants to give their evaluation of the smartphone app they used during the study.

Outcomes

The main objective of this study was to deploy and test a personalized electronic record in the form of a digital smartphone app to participants with primary hypertension for routine monitoring of BP in a community setting within the United States. We used this app to assess daily BP control within this population. We also assessed interactions with the app over the 12-week observational period, including compliance to BP monitoring routines (percentage of patients entering a BP diary entry on 80% or more occasions) and patient feedback and satisfaction using the smartphone app as assessed by the UEQ. Other outcomes of interest included assessment of antihypertensive medication use and emergent symptoms of SARS-CoV-2 infection including cough, fever, and shortness of breath. The safety of participants in relation to the use of the app was assessed through self-reporting of adverse events (AEs) and serious adverse events (SAEs). Exploratory outcomes included analysis of associations of BP antihypertensive drug use as well as BP level with the incidence of SARS-CoV-2 infection and the severity of COVID-19 outcomes, to determine if antihypertensive drug use or the level of BP control can affect severity of COVID-19 illness or pneumonia.

Statistical Analysis

This study's exploratory nature made finding a specific sample size impractical. Given this study's low risk, there was no cap on enrollment and the intent was to enroll as many participants as possible within the recruitment time period. As this was an observational study, analyses included basic descriptive statistics and statistical tests to assess differences between groups. Analysis groups were defined based on compliance with BP data capture. We planned to analyze patient outcomes, but due to the lack of disease outcomes observed in the study, this was not done. Adherence to BP monitoring routines was assessed by calculating the percentage of patients completing a full BP diary routine entry on 80% or more occasions (a full routine consisted of 3 BP measurements and routines were twice daily—morning and evening). Complete adherence to the digital diary was defined in terms of full routines consisting of morning BP in triplicate with medications summary, or an evening BP triplicate along with a symptoms report.

Ethical Considerations

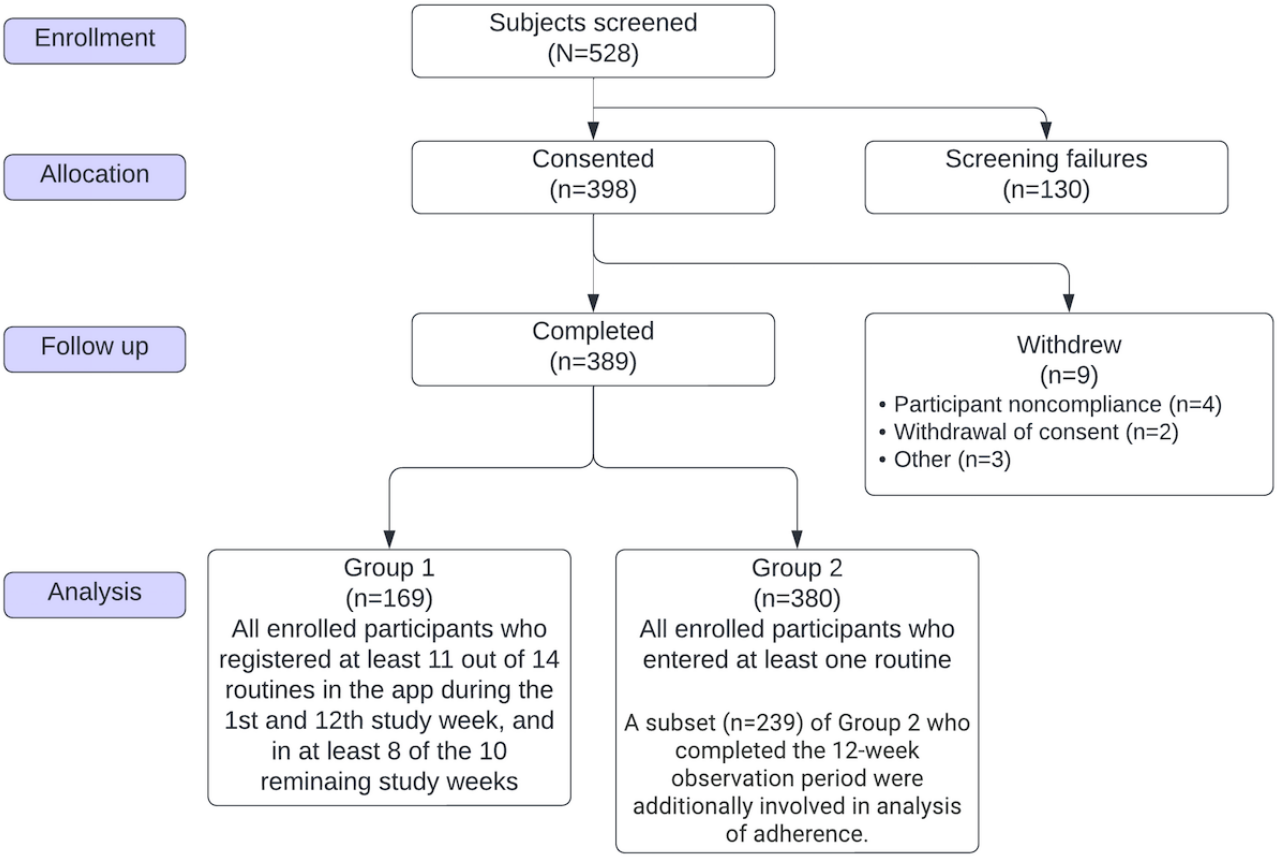
This study was reviewed and approved by IntegReview Institutional Review Board (later became Advarra), which provided a favorable opinion to conduct this study in the United States. Participants who were deemed eligible and after submitting their screeners and signed informed consent forms using electronic signature had their forms stored in the Curebase platform. Participants were able to withdraw from the study at any time. They were able to withdraw consent as well as have all archived clinical data discarded. All information was confidential and data were pseudonymized. Institutional review board–approved small pro rata payments were offered to participants as compensation for participation based on the amount of time in the study and the number of data entry routines completed (US $10 for successfully downloading the app and an additional US $10 for each week of participation up to 12 weeks total).

Results

Overview of Study, Procedures, and Patient Population

A total of 528 participants were screened for entry into the study (Figure 1). From August 2020 to November 2020, a total of 398 patients with hypertension were enrolled, with 389 (98%) participants completing the study. The last participant completed follow-up in February 2021, with 9 participants being lost to follow up. A summary of participant disposition is shown in Figure 1. Participants were analyzed in 2 main groups, with group 1 consisting of 169 participants who were highly compliant at data entry into the smartphone app (participants who registered at least 11 of 14 routines in the app during the 1st and 12th study week, and in at least 8 of the 10 remaining study weeks). At least 80% of routine entries was considered to be the benchmark in this study based on Haynes and colleagues' [9] definition of sufficient adherence of individuals taking antihypertensive medication being ≥80%.

Figure 1. Details of enrollment into the trial.



The other main group, group 2, consisted of 380 participants (overall study population), which included all participants who had entered 1 or more routines into the app. There were 9 participants who did not enter any routines. A subset of group 2 was defined for the purpose of adherence consisting of 239 (62.9%) participants who remained in the study until the end of the 12-week observation period.

A summary of demographic data of all participants is presented in Table 2. The average age of the population was 58.1 (SD 9.9) years in group 1 and 54.1 (SD 12.1) years in group 2. More female participants were present in group 1 (109/169, 64.5%) and group 2 (260/380, 68.4%) than male participants in the study. The average BMI of participants was 33.9 (SD 10.2) kg/m² for group 1 and 34.9 (SD 9.6) kg/m² for group 2. A mix of races and ethnicities were represented, with the majority of participants being White (128/169, 75.7% in group 1 and 283/380, 74.5% in group 2).

Table 2. Patient demographics.

Demographics	Cohort	
	Group 1 population (n=169)[a]	Group 2 population (n=380)[b]
Age (years)		
Participants, n	169	379
Mean (SD)	58.1 (9.9)	54.1 (12.1)
Median (range)	58.0 (34.0-80.0)	55.0 (22.0-87.0)
Sex, n (%)		
Female	109 (64.5)	260 (68.4)
Male	60 (35.5)	118 (31.1)
Missing	0 (0)	2 (0.5)
Race and ethnicity, n (%)		
White	128 (75.7)	283 (74.5)
Black	27 (16)	56 (14.7)
Other	12 (7.1)	28 (7.4)
2 or more[c]	2 (1.2)	11 (2.9)
Missing	0 (0)	2 (0.5)
Height (cm)		
Participants, n	168	377
Mean (SD)	168.6 (10.3)	168.2 (10.4)
Median (range)	167.6 (149.9-198.1)	167.6 (144.8-198.1)
Weight (kg)		
Participants, n	169	377
Mean (SD)	96.4 (30.0)	98.7 (28.5)
Median (range)	90.9 (43.6-239.5)	94.5 (43.6-239.5)
BMI (kg/m^2)		
Participants, n	168	376
Mean (SD)	33.9 (10.2)	34.9 (9.6)
Median (range)	31.8 (18.3-87.9)	33.1 (18.3-90.3)

[a]Includes all participants who registered at least 11 of 14 routines in the app during the 1st and 12th study week, and in at least 8 of the 10 remaining study weeks.

[b]Includes all enrolled participants who completed the study and entered any routines (9 enrolled participants did not enter any routines).

[c]Participants who identified as 2 or more races and ethnicities.

Adherence and Participant Satisfaction

Central to examining the core participant interactions with the app was the adherence of participants to record their daily routines (Table 3). Out of all 380 enrolled participants (group 2), a total of 248 (65.3%) participants recorded full BP routines 80% or more over the 12-week period and 322 (84.7%) participants recorded any data on 80% or more of the days over the 12-week period. Out of the 239 participants who remained in the study for the full 12-week observation period, 201 (84.1%) participants recorded full BP routines 80% or more over the 12-week period, and 227 (95%) participants of patients recorded any data on 80% or more of the days over the 12-week period. The number of study participants decreased from 380 at week 1 to 239 (62.9%) at week 12. Participants also had the option to continue for an additional 12 weeks resulting in an observational period of up to 23 weeks. The adherence to entry of routines for up to an additional 12 weeks was assessed (Multimedia Appendix 3), which looked at the total number of participants listed for each week, including only those participants who entered routines for the week. Of the 239 participants who completed the 12-week study period, 234 progressed to the optional 12-week extension period and started a 13th week. Furthermore, 155 participants progressed to week 14 and 99 participants continued to week 15. At this point, the number of participants dropped to 57 at week 16 and less than 10% of enrollees (30/380, 7.9% of participants) continued to enter data after week 17. However, it should be noted that the length of participation in the study was also dependent on the remaining time for recruitment and overall study completion;

thus, some participants did not have the opportunity to participate in the extension period or could only participate in a time limited portion.

Participant feedback and satisfaction using the digital diary to record BP routines was assessed using the UEQ. The UEQ uses 26 questions, which are presented as a scale from 1 to 7 with a positive term and a negative term on each end of the scale (eg, annoying=1 to enjoyable=7). The 26 questions are grouped into 6 domains, which are attractiveness, perspicuity, efficiency, dependability, simulation, and novelty. A mean score for each domain is presented, whereby mean values for the domains can range from –3 to +3. Mean values between –0.8 and +0.8 are considered neutral, mean values above 0.8 are considered a positive evaluation, and mean values less than –0.8 are considered a negative evaluation. However, it should be noted that question number 26 from the UEQ that asks if users found the app to be "conservative or innovative" was missing from the dataset. Therefore, there are not any data present for this question, and it does not contribute toward the mean scores for the domain it belonged to (novelty).

Table 3. Participant compliance with remote data collection (number of routines[a] recorded).

Analysis group	Full routine[a]	Full BP[b] routine[c]	Any entry within a day
Group 2 [d] (n=380)			
Over 80% adherence, n (%)	241 (63.4)	248 (65.3)	322 (84.7)
Median adherence (%), median (IQR)	87.1 (69.3-97)	88 (70.3-97)	100 (92.5-100)
Participants who completed 12-week observation period (subset of group 2 [d], n=239)			
Over 80% adherence, n (%)	196 (82)	201 (84.1)	227 (95)
Median adherence (%), median (IQR)	92.8 (83.9-97.6)	94 (84.4-98.2)	100 (98.8-100)

[a]A full routine consists of morning BP in triplicate along with medication summary, or evening BP in triplicate along with a symptom report.

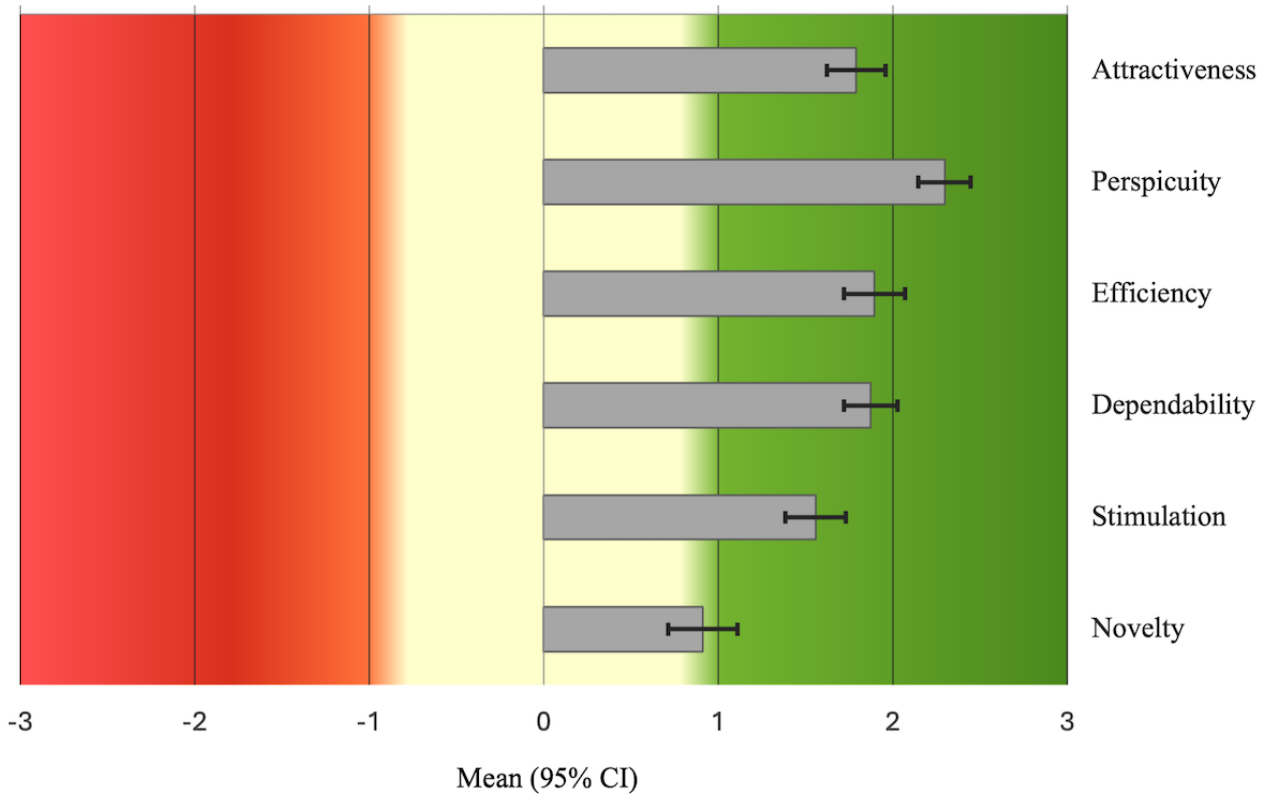
[b]BP: blood pressure.

[c]A full BP routine consists of morning blood pressure in triplicate or evening blood pressure in triplicate.

[d]Includes all enrolled participants who completed the study and entered any routines (9 enrolled participants did not enter any routines).

Almost all single-scored items of the UEQ averaged a positive evaluation (>0.8; only inventive to conventional scored lower [mean score 0.5]), with the majority scoring a very good evaluation (1.5 or above). Figure 2 shows the analysis of the 6 domains representing groups of single items within the UEQ. These results revealed very good to excellent evaluations for attractiveness (mean 1.79, SD 1.29), perspicuity (mean 2.30, SD 1.03), efficiency (mean 1.90, SD 1.43), dependability (mean 1.87, SD 1.09), and stimulation (mean 1.56, SD 1.41). The score for novelty was slightly lower (mean 0.91, SD 1.83) but was still a positive evaluation. The UEQ can be used to benchmark products against other apps, web pages, software, and social networks, with data coming from 21,175 persons involved in 468 studies regarding different products [7]. The digital app in this study was benchmarked as "excellent" (top 10% best results) for perspicuity, efficiency, and dependability and as "good" (10% of results are better and 75% of results are worse) for attractiveness and stimulation. Novelty scored an "above average" UEQ benchmark (25% of results are better and 50% of results are worse).

Figure 2. User Experience Questionnaire scores (all enrolled participants).



BP Control

Changes in self-reported average systolic BP (SBP) and diastolic BP (DBP) during the observational period for group 1 and subgroups within this population are shown in Table 4. No significant changes in average SBP (mean –0.9, SD 10.7 mm Hg; *P*=.30) or DBP (mean –1.0, SD 7.9 mm Hg; *P*=.10) were detected between week 1 and week 12. Subgroup analysis by age (≥65 y and <65 y) and sex also showed no significant changes in BP during the 12-week period (all *P*>.05). Average levels for SBP (week 1: mean 129.5, SD 13.4 mm Hg; week 12: 128.6, SD 13.6 mm Hg) and DBP (week 1: 79.2, SD 9.9 mm Hg; week 12: 78.2, SD 10.8 mm Hg) were below UK and US hypertensive diagnostic thresholds.

Comparison of BP levels measured at week 1 and week 12 by race for group 1 showed no significant changes in SBP or DBP for Black participants or in SBP for White participants (all *P*>.05). A statistically significant change in average DBP between baseline and 12 weeks was observed for White participants (mean –1.6, SD 7.6 mm Hg; *P*=.02). No significant changes in SBP or DBP for participant groupings based on current antihypertensive medication (angiotensin-converting enzyme inhibitor [ACEi] or angiotensin receptor blocker [ARB], β-blockers, or calcium channel blockers [CCB]) between week 1 and week 12 were observed (all *P*>.05), except for a minor decrease in DBP (mean –2.1, SD 6.6 mm Hg; *P*=.01) observed in participants currently taking β-blockers.

Subgroupings of group 1 derived from hypertensive diagnostic thresholds (135/85 mm Hg and 140/90 mm Hg) revealed significant reductions at week 12 in comparison with week 1 in SBP (BP≥135/85 mm Hg group: mean –6.8, SD 10.3 mm Hg, *P*<.001; BP≥140/90 mm Hg group: –9.0, SD 10.8 mm Hg, *P*<.001) and DBP (BP≥135/85 mm Hg group: mean –3.8, SD 10.0 mm Hg, *P*=.02; BP≥140/90 mm Hg group: mean –5.4, SD 9.2 mm Hg, *P*=.008) for participants entering the study with BP equal to or higher that these thresholds. No changes in SBP or DBP over the 12-week period were observed in participants with average BP levels lower than these thresholds, except for a minor increase (mean 2, SD 9.8 mm Hg; *P*=.03) in SBP in participants joining the study with BP<135/85 mm Hg.

Table 4. Change in blood pressure by subgroup (group 1[a]).

Subgroup	Systolic BP[b]					Diastolic BP				
	Participants, n	Week 1, mean (SD)	Week 12, mean (SD)	Week 1-12 change, mean (SD)	*P* value	Participants, n	Week 1, mean (SD)	Week 12, mean (SD)	Week 1-12 change, mean (SD)	*P* value
Overall population	169	129.5 (13.4)	128.6 (13.6)	–0.9 (10.7)	.30	169	79.2 (9.9)	78.2 (10.8)	–1.0 (7.9)	.10
Age group (years)										
≥65	51	126.5 (15.2)	126.4 (14.8)	–0.0 (9.7)	.99	51	74.6 (9.3)	73.4 (8.7)	–1.2 (5.8)	.16
<65	118	130.8 (12.3)	129.5 (12.9)	–1.2 (11.2)	.24	118	81.2 (9.5)	80.2 (10.9)	–1.0 (8.7)	.23
Sex										
Male	60	132.1 (13.9)	131.1 (15.0)	–1.0 (11.6)	.52	60	79.6 (9.7)	79.1 (11.2)	–0.5 (7.6)	.64
Female	109	128.0 (12.9)	127.2 (12.6)	–0.8 (10.3)	.42	109	79.0 (10.0)	77.7 (10.5)	–1.3 (8.1)	.09
Race										
White	128	129.7 (13.7)	128.5 (14.1)	–1.1 (11.3)	.27	128	78.7 (9.7)	77.1 (9.9)	–1.6 (7.6)	.02
Black	27	128.0 (11.4)	128.1 (10.7)	0.1 (8.7)	.94	27	81.2 (9.8)	81.3 (10.4)	0.1 (7.8)	.96
HBP[c] (United States)										
BP ≥140/90 mm Hg at entry	33	150.2 (7.1)	141.2 (12.7)	–9.0 (10.8)	<.001	24	95.3 (7.5)	89.9 (9.3)	–5.4 (9.2)	.008
BP <140/90 mm Hg at entry	136	124.4 (9.0)	125.5 (11.9)	1.1 (9.8)	.19	145	76.5 (7.4)	76.3 (9.7)	–0.3 (7.5)	.65
HBP (United Kingdom)										
BP ≥135/85 mm Hg at entry	55	145.0 (8.5)	138.3 (11.8)	–6.8 (10.3)	<.001	40	92.2 (7.0)	88.5 (10.0)	–3.8 (10.0)	.02
BP <135/85 mm Hg at entry	114	121.9 (7.6)	123.9 (11.8)	2.0 (9.8)	.03	129	75.2 (6.6)	75.0 (8.8)	–0.2 (7.0)	.79
Drug Class										
ACE[d] inhibitors or AR[e] blockers	110	128.9 (13.0)	128.9 (13.5)	0.0 (10.7)	.99	110	8.9 (10.0)	78.4 (11.8)	–0.5 (8.6)	.56
β-blockers	70	131.5 (15.4)	128.9 (15.4)	–2.5 (11.1)	.06	70	80.0 (11.0)	77.9 (10.0)	–2.1 (6.6)	.01
Calcium channel blockers	55	129.7 (12.0)	128.3 (11.1)	–1.4 (11.1)	.35	55	79.1 (9.0)	79.1 (10.8)	0.1 (7.1)	.93

[a]Includes all participants who registered at least 11 of 14 routines in the app during the 1st and 12th study week, and in at least 8 of the 10 remaining study weeks.

[b]BP: blood pressure.

[c]HBP: home blood pressure.

[d]ACE: angiotensin-converting enzyme.

[e]AR: angiotensin receptor.

Antihypertensive Medication Use

Antihypertensive medication use was examined for participants analyzed in group 1 (which includes all participants who registered at least 11 of 14 routines in the app during the 1st and 12th study week, and in at least 8 of the 10 remaining study weeks). No statistically significant differences in average SBP and DBP were seen for participants grouped by their antihypertensive medications (ACEi or ARB, β-blockers, and CCBs; all $P>.05$; Figures 3 and 4). Over the 12-week observational period, a significant but minor reduction in DBP (mean –2.1, SD 6.6 mm Hg; $P=.01$) in those participants

currently taking β-blockers was observed (Table 3). ACEi or ARB were the most common antihypertensive medication taken (n=116), with smaller numbers of participants taking β-blockers (n=70), CCBs (n=55), or thiazide-type diuretics (n=6; Figure 3).

Further categorization of these participants by current antihypertensive medication revealed no statistically significant differences in average SBP and DBP (Figures 5 and 6). The most commonly taken antihypertensive medications in this population included lisinopril (ACEi; n=55), amlodipine (CCB; n=55), losartan (ARB; n=49), and metoprolol (β-blocker; n=38).

Figure 3. Average systolic blood pressure for each drug combination group (group 1 population). ACEi: angiotensin-converting enzyme inhibitor; ARB: angiotensin receptor blocker; CCB: calcium channel blocker.

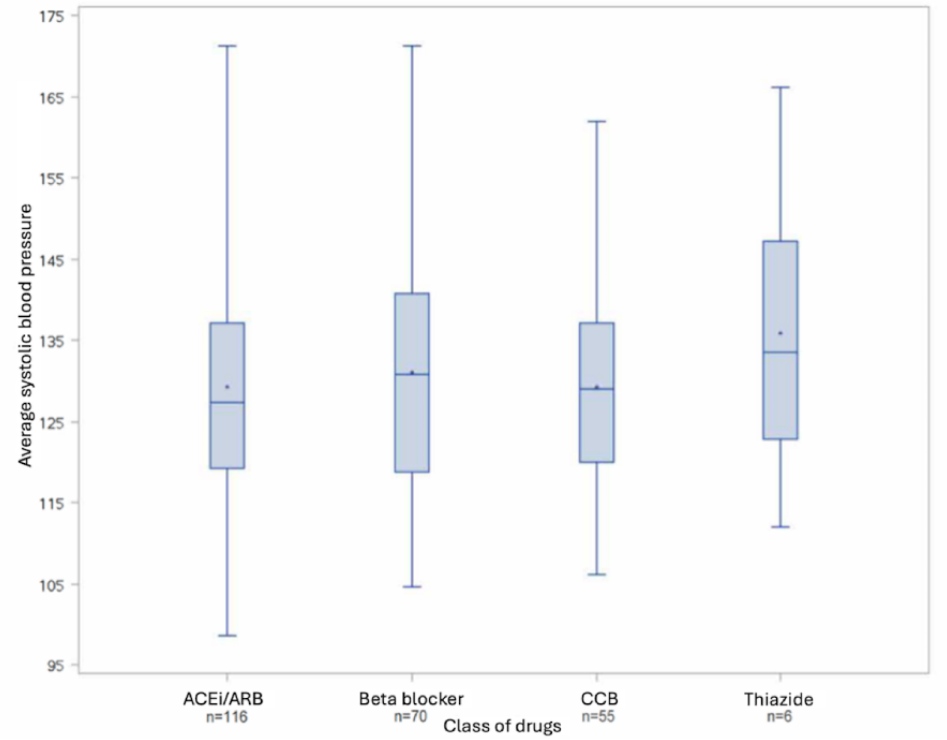


Figure 4. Average diastolic blood pressure for each drug combination group (group 1 population). ACEi: angiotensin-converting enzyme inhibitor; ARB: angiotensin receptor blocker; CCB: calcium channel blocker.

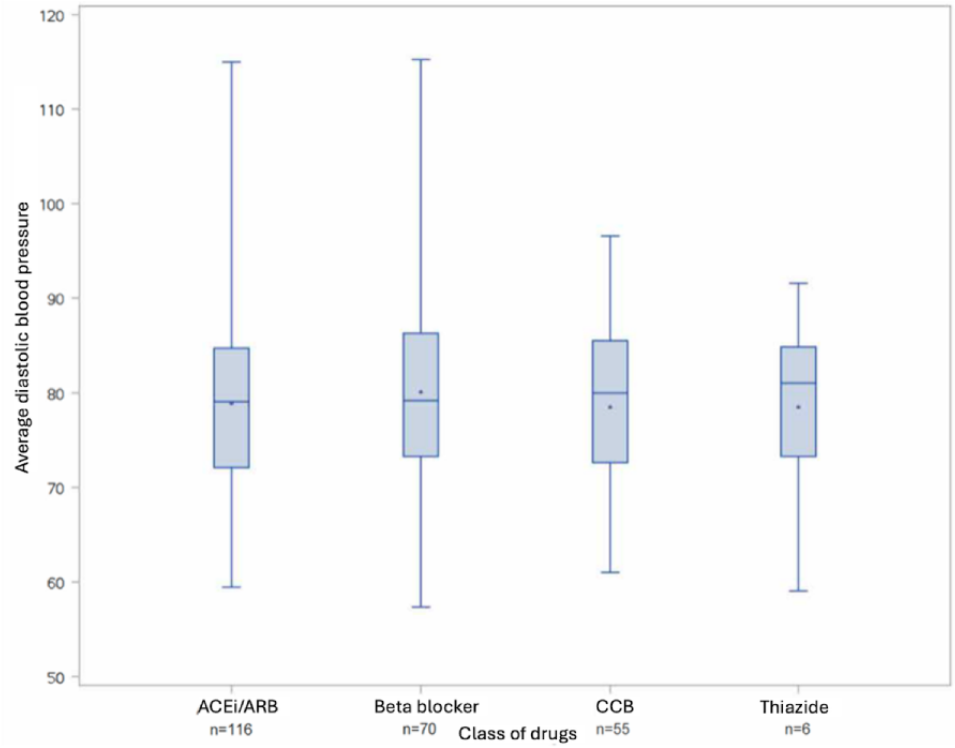


Figure 5. Average systolic blood pressure for each drug (group 1 population). ACEi: angiotensin-converting enzyme inhibitor; ARB: angiotensin receptor blocker; CCB: calcium channel blocker; freq: frequency.

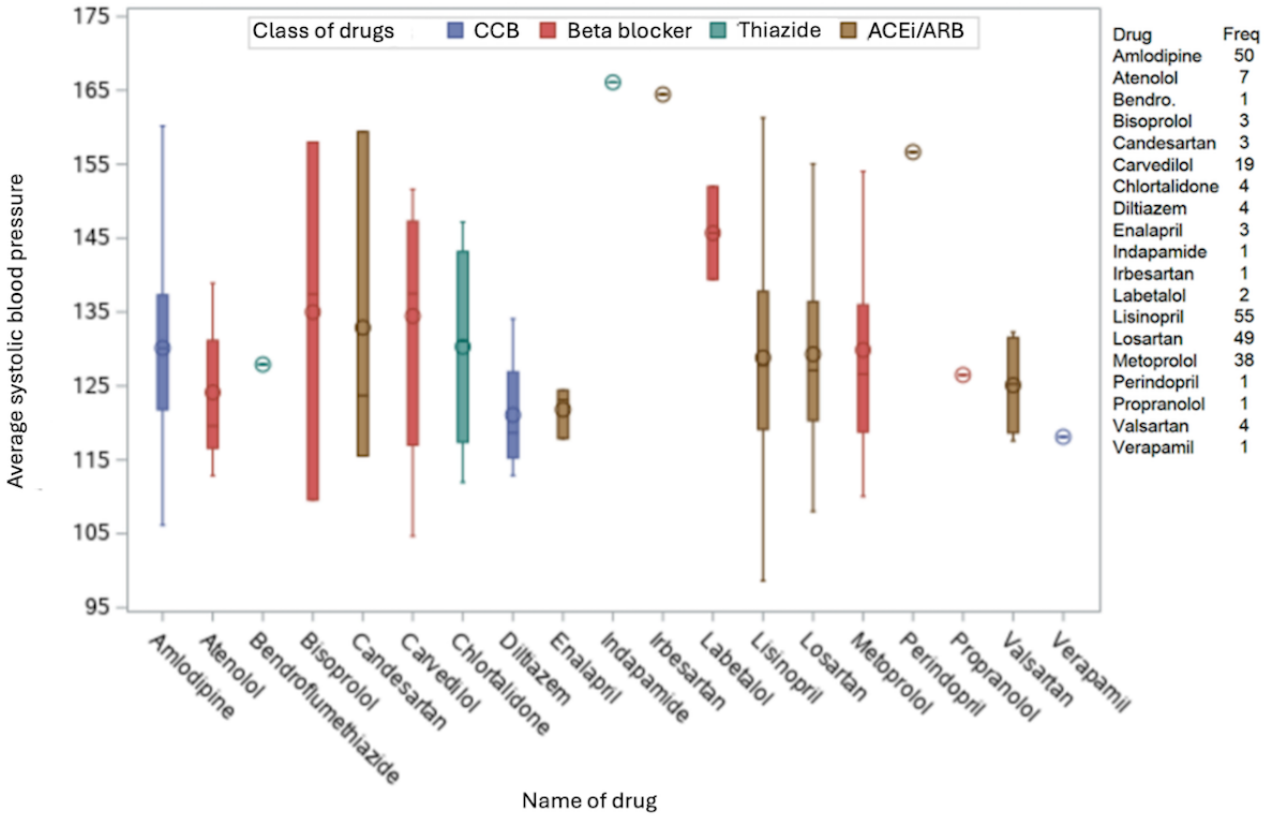
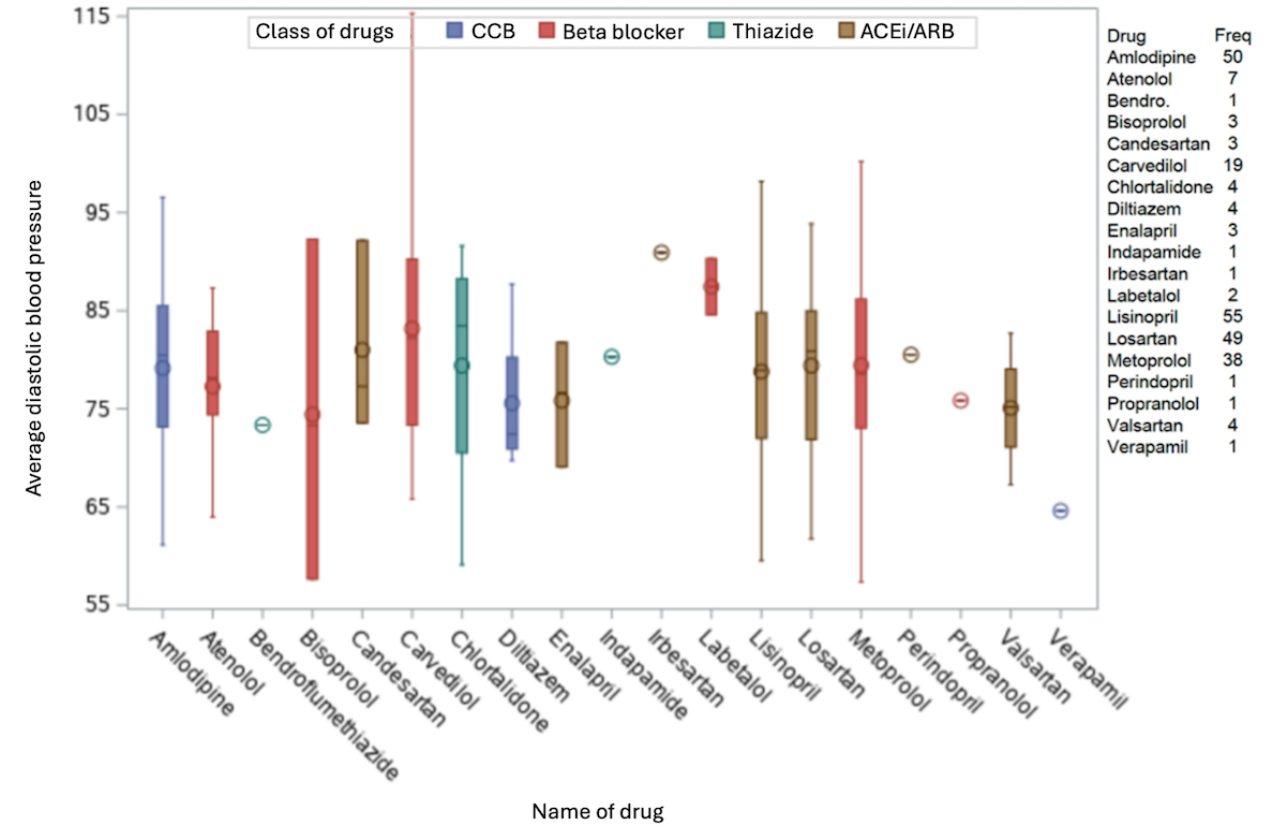


Figure 6. Average diastolic blood pressure for each drug (group 1 population). ACEi: angiotensin-converting enzyme inhibitor; ARB: angiotensin receptor blocker; CCB: calcium channel blocker; freq: frequency.



COVID-19 Symptoms

An analysis of emergent symptoms of SARS-CoV-2 infection, including but not limited to fatigue, high temperature, persistent dry cough, chest pain, and shortness of breath, in all enrolled participants was performed. This revealed very low levels of symptoms related to infection with COVID-19, with only 3 participants experiencing these symptoms. Due to insufficient numbers of participants experiencing COVID-19 symptoms, associations between BP medication and BP level with SARS-CoV-2 infection and COVID-19 outcomes could not be assessed.

Adverse Events

All enrolled participants (n=398) were included in the safety population. There were no AEs attributable to the smartphone app, SAEs, or deaths recorded in this study.

Discussion

Principal Findings

Our study successfully deployed a smartphone app used for monitoring daily BP, medication, and side effects in a community setting within the United States, and almost all participants (380/389, 98%) downloaded and entered BP routines in week 1. A high proportion of participants remained in the study for the full 12-week observation period and entered full BP routines into the digital app 80% or more of the time. Participants reported their experience of using the smartphone app using the UEQ, giving an overall positive evaluation, whereby the app was benchmarked “excellent” and “above average” for almost all domains. Highly adherent participants with hypertension demonstrated well-controlled BP, with no significant changes in average SBP or DBP between week 1 and week 12. Participants were able to record BP medications and symptoms of SARS-CoV-2 infection. No AEs attributable to use of the smartphone app were reported during the observational period.

The study population had a broad age range, between 24 and 87 (mean 54.1) years. There were more female than male participants enrolled (260/380, 68.4%), and analysis of race and ethnicity showed the population consisted of 74.5% (283/380) White participants, 14.7% (56/380) Black participants, 7.4% (28/380) participants who reported “other” for their race and ethnicity, and 2.9% (11/380) participants who reported their race and ethnicity as “2 or more.” Any participant who did not report as being White, Black, or having 2 or more races or ethnicities were required to enter their ethnicity as “other.”

According to the 2020 US decennial census, out of 331,449,281 people, there were 204,277,273 (61.6%) White (alone) individuals; 41,104,200 (12.4%) Black individuals; and 33,848,942 (10.2%) individuals who reported “2 or more” for their race and ethnicity [10]. In comparison with the reported demographics of the Unites States, we observed a higher percentage of White and Black participants enrolled into the study. However, a recent report describing racial and ethnic demographics of adults with hypertension in the United States reported 79,910,050 individuals with hypertension, of which 66.4% (51,03,185) were White and 14.4% (11,061,686) were Black [11]. The percentage of Black participants in our study is representative of the proportion of Black patients with hypertension in the United States; however, we observed a higher proportion of White participants than the proportion of White patients with hypertension in the United States. Our study observed 28 (7.4%) out of 380 participants who reported their race and ethnicity as “other,” which included any individuals did not identify as Black, White, or having “2 or more” races and ethnicities; so, we did not focus on other racial and ethnic groups specifically. The average BMI of the study population was 34.9 kg/m^2. These results indicate that, on average, the population was clinically obese, which are consistent with levels of obesity reported for US-based adults with hypertension [11].

CLM successfully modified an existing app and deployed it for use in the study. The patient experience and data collection components that had been developed as part of an early platform concept proved straightforward to adapt and extend for this study, building confidence that the goal of rapidly developing an adaptable multitherapeutic technology for patient data collection and treatment is viable. During pandemic conditions, the ability to quickly respond with digital patient support and assessment tools is critical [12]. Symptom collection proved straightforward by reusing user interface components from the early platform that allowed patients to rate their symptoms using a visual analog scale (VAS). The symptom component was designed to present a configurable and customizable list of symptoms to the patient, which allows the data to be collected, stored, and presented. A similar approach was taken with the hypertension drug list. The patient experience component required only simple adaptation as it was directly suitable for adoption in this study. The changes were simple, such as updating the app name or contact information presented to patients. This provided reassuring evidence that a common set of user interaction components, supported by an underlying platform, can be viable in a digital health care setting involving mobile devices.

In this study, we used the smartphone app to record daily BP measurements in patients with hypertension. Telemonitoring is another method that is used for routine BP monitoring in hypertension [13]. Industry standards and other clinical studies have shown that 39% to 72% of patients report their BP 80% or more of the time using this method [14-17]. In our study, 248 (65.3%) out of 380 participants (group 2) showed high compliance for entering daily BP routines into the CLM digital diary (entering full BP routines on 80% or more of days). Compliance to entering daily BP routines was even higher in the 239 who completed the full 12 weeks, with 201 (84.1%) entering full BP routines over 80% of the time. While overall adherence levels may have been affected by the daily requirement to enter 2 routines (morning and evening) within a limited time window, recording morning and evening routines each day reflects best practice. Nonetheless, these results demonstrate that our test smartphone app meets industry norms for adherence to routine BP self-monitoring. It should be noted that the digital app development deployed in this study was not designed for high engagement and still achieved these industry standards. Future work will focus on refinement and

incorporation of engagement enhancing features as part of product development.

Participant satisfaction of the CLM smartphone app for remote BP monitoring was assessed using the validated UEQ. The app scored an overall positive evaluation. Analysis of UEQ domains consisting of several questionnaire items, revealed that the app scored an excellent evaluation for perspicuity and very good evaluations for attractiveness, dependability, efficiency, and stimulation. A lower but still positive evaluation was given for novelty. Data from another hypertension study, PERSONAL-COVIDBP, in the United Kingdom has shown similar outcomes for user experience [18,19]. Overall, these results suggest very high levels of participant satisfaction of this smartphone app for daily monitoring of BP.

The UEQ can be used to benchmark products against other apps, web pages, software, and social networks, with data coming from 21,175 persons involved in 468 studies regarding different products [7]. Here, we show that our smartphone app for monitoring BP benchmarked "excellent" for domains of perspicuity, efficiency, and dependability and "good" for domains of attractiveness and stimulation. Novelty was benchmarked at "above average." Again, these results indicate that despite the app being a prototype, a very high level of participant satisfaction was observed. Further app development work at CLM will include a particular focus on novelty, attractiveness, and stimulation.

Clinical Outcomes

Through remote BP monitoring, we demonstrated that in a population of people with hypertension, self-reported BP in study completers was stable over a 3-month period with no significant changes in average BP (129.5/79.2 mm Hg at baseline, 128.6/78.2 mm Hg at week 12). These levels are below the established hypertensive diagnostic thresholds (United Kingdom: 135/85 mm Hg [20] and United States: 130/80 mm Hg [21]). These thresholds represent treatment targets for BP and as such, demonstrate that BP levels in this study population on average are being effectively managed. Although we do acknowledge that these results are based on self-reported data and that direct measurement might have resulted in higher BP data.

Although subgroup analysis showed that there was no difference in our BP outcomes across age and sex, significant reductions in SBP and DBP were observed for those participants who entered this study with baseline BP levels higher than the established hypertensive diagnostic thresholds for the United States and United Kingdom (Table 4). As this was an observational study with no drug intervention, these findings may have in part been a result of remote BP monitoring having a positive impact on BP control, when used in conjunction with standard of care antihypertensive medications. However, it is possible that a lowering of BP during the observational period could be due to other factors such as the Hawthorne Effect or regression to the mean, and in some cases, perhaps patients were at the early stages of their antihypertensive treatment. Nonetheless, our results are in line with the demonstrated effectiveness of BP self-monitoring, combined with other interventions, to lower BP in patients with high BP levels [22-24]. Self-monitoring of BP is thought to be advantageous as it can provide better estimation of underlying BP, increase adherence to medications, and reduce the need for clinic monitoring. In the future, remote BP monitoring would facilitate drug optimization through the development of more sophisticated electronic mobile apps.

According to American College of Cardiology and American Heart Association treatment guidelines for high BP, the first-line antihypertensives used to treat hypertension include ACEi, ARBs, CCBs, and thiazide-type diuretics [21]. After the Joint National Committee guidelines 8 in 2014, β-blockers have been considered secondary agents. Accordingly, this population of hypertensive patients based in the United States were seen taking 1 or more of these antihypertensive medications. Analysis of BP levels with respect to participant-reported antihypertensive drugs revealed no difference in levels of SBP and DBP reported for the participants taking thiazide-type diuretics, CCBs, ACEi, or ARBs. Similar results were seen for analyses of single antihypertensive drugs. We did however observe a minor decrease in DBP (mean 2.1, SD 6.6 mm Hg; P=.01) in those participants currently taking β-blockers at week 12.

Strengths and Limitations

The strengths of this study were the rapid set up of a decentralized study during the COVID-19 pandemic, which involved the rapid modification and deployment of a smartphone app for monitoring BP, BP medication use, and symptoms of SARS-CoV-2 infection. Another strength of this research includes the large sample size and the accurate representation of Black participants in our study against the proportion of Black patients with hypertension in the United States. The low dropout rates also eliminated the need to correct for missing data. However, a limitation is that the study did not specifically look at Hispanic or Latino as well as Asian patients who have been reported under the "other" category of race and ethnicity; therefore, analysis specifically on these groups could not be performed.

Another limitation of this study was that data entry, adherence, and participant engagement rates may have been influenced by (1) participants being incentivized for entry of data into the smartphone app and (2) lockdown social distancing measures, which meant that patients with hypertension were at home without other activities occupying their time. Thus, study routines would have been less interrupted, possibly resulting in enhanced habit formation. Incentives were kept to a minimum, with the monetary value low to temper this limitation. As part of the UEQ, question-26 data were missing, which asked whether users considered the app to be "conservative or innovative." As these data are missing and could have affected the mean score for the novelty domain in a positive or negative manner, it is a further limitation of the study.

Low incidence of symptoms of COVID-19 in the hypertensive study population was another limitation to this study. This meant that we were unable to examine associations between BP levels or BP antihypertensive drug use and the incidence of SARS-CoV-2 infection and severity of COVID-19 outcomes.

In our study, we evaluated changes in BP in a cohort compliant with BP data capture so we could accurately assess their BP levels. It is possible that those who were not as compliant with BP data capture may be different in some ways to this adherent population; however, as it would be less reliable to analyze BP in those with low BP data capture, it would be hard to make that assessment.

Conclusions

We successfully deployed and tested a personalized electronic record in the form of a smartphone app in participants with primary hypertension for routine monitoring of BP in a community setting within the United States. The patient population engaged with the app and provided highly positive feedback. However, novelty features and additional engagement features beyond data monitoring should be considered to bolster patient adherence. This study also demonstrated no safety implications regarding the use of the smartphone app. These are important components for the provision of individualized and targeted treatment for hypertension.

Acknowledgments

This study was funded and sponsored by Closed Loop Medicine Ltd. Curebase carried out the observational study in the United States.

Conflicts of Interest

LR, NN, JF, GT, JS, TG, and MT are employees of Closed Loop Medicine Ltd.

Multimedia Appendix 1
Interface of the digital smartphone app.
[DOCX File , 6529 KB - mhealth_v12i1e53430_app1.docx]

Multimedia Appendix 2
System diagram.
[PNG File , 601 KB - mhealth_v12i1e53430_app2.png]

Multimedia Appendix 3
Length of study participation during weeks 1 to 23 (all enrolled participants).
[PNG File , 29 KB - mhealth_v12i1e53430_app3.png]

References

1. Mills KT, Bundy JD, Kelly TN, Reed JE, Kearney PM, Reynolds K, et al. Global disparities of hypertension prevalence and control: a systematic analysis of population-based studies from 90 countries. Circulation 2016 Aug 09;134(6):441-450 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.115.018912] [Medline: 27502908]
2. Carey RM, Moran AE, Whelton PK. Treatment of hypertension: a review. JAMA 2022 Nov 08;328(18):1849-1861. [doi: 10.1001/jama.2022.19590] [Medline: 36346411]
3. Abegaz TM, Shehab A, Gebreyohannes EA, Bhagavathula AS, Elnour AA. Nonadherence to antihypertensive drugs: a systematic review and meta-analysis. Medicine (Baltimore) 2017 Jan 01;96(4):e5641 [FREE Full text] [doi: 10.1097/MD.0000000000005641] [Medline: 28121920]
4. Lemstra M, Alsabbagh MW. Proportion and risk indicators of nonadherence to antihypertensive therapy: a meta-analysis. Patient Prefer Adherence 2014 Feb 13;8:211-218 [FREE Full text] [doi: 10.2147/PPA.S55382] [Medline: 24611002]
5. Bestsenny O, Gilbert G, Harris A, Rost J. Telehealth: a quarter-trillion-dollar post-COVID-19 reality? McKinsey & Company. 2021 Jul 09. URL: https://www.mckinsey.com/industries/healthcare-systems-and-services/our-insights/telehealth-a-quarter-trillion-dollar-post-covid-19-reality [accessed 2023-09-01]
6. Ebbert JO, Ramar P, Tulledge-Scheitel SM, Njeru JW, Rosedahl JK, Roellinger D, et al. Patient preferences for telehealth services in a large multispecialty practice. J Telemed Telecare 2023;29(4):298-303. [doi: 10.1177/1357633X20980302] [Medline: 33461397]
7. Laugwitz B, Held T, Schrepp M. Construction evaluation of a user experience questionnaire. 2008 Presented at: 4th Symposium of the Workgroup Human-Computer Interaction and Usability Engineering of the Austrian Computer Society, USAB 2008; November 20-21, 2008; Graz, Austria p. 63-76. [doi: 10.1007/978-3-540-89350-9_6]
8. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95 [FREE Full text] [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]

9. Haynes RB, Taylor DW, Sackett DL, Gibson ES, Bernholz CD, Mukherjee J. Can simple clinical measurements detect patient noncompliance? Hypertension 1980;2(6):757-764. [doi: 10.1161/01.hyp.2.6.757] [Medline: 7007235]
10. P8 Race. United States Census Bureau. 2020. URL: https://data.census.gov/table/DECENNIALDHC2020.P8?g=010XX00US [accessed 2024-08-22]
11. Aggarwal R, Chiu N, Wadhera RK, Moran AE, Raber I, Shen C, et al. Racial/ethnic disparities in hypertension prevalence, awareness, treatment, and control in the United States, 2013 to 2018. Hypertension 2021 Aug 09;78(6):1719-1726 [FREE Full text] [doi: 10.1161/HYPERTENSIONAHA.121.17570] [Medline: 34365809]
12. Peek N, Sujan M, Scott P. Digital health and care in pandemic times: impact of COVID-19. BMJ Health Care Inform 2020;27(1):e100166 [FREE Full text] [doi: 10.1136/bmjhci-2020-100166] [Medline: 32565418]
13. Citoni B, Figliuzzi I, Presta V, Volpe M, Tocci G. Home blood pressure and telemedicine: a modern approach for managing hypertension during and after COVID-19 pandemic. High Blood Press Cardiovasc Prev 2021 Dec 02;29(1):1-14 [FREE Full text] [doi: 10.1007/s40292-021-00492-4] [Medline: 34855154]
14. Kerby TJ, Asche SE, Maciosek MV, O'Connor PJ, Sperl-Hillen JM, Margolis KL. Adherence to blood pressure telemonitoring in a cluster-randomized clinical trial. J Clin Hypertens (Greenwich, Conn.) 2012 Jul 26;14(10):668-674 [FREE Full text] [doi: 10.1111/j.1751-7176.2012.00685.x] [Medline: 23031143]
15. Milot J, Birnbaum L, Larochelle P, Wistaff R, Laskine M, Van Nguyen P, et al. Unreliability of home blood pressure measurement and the effect of a patient-oriented intervention. Can J Cardiol 2015 May;31(5):658-663. [doi: 10.1016/j.cjca.2015.03.006] [Medline: 25936491]
16. Park S, Kum H, Morrisey MA, Zheng Q, Lawley MA. Adherence to telemonitoring therapy for Medicaid patients with hypertension: case study. J Med Internet Res 2021 Sep 6;23(9):e29018 [FREE Full text] [doi: 10.2196/29018] [Medline: 34486977]
17. Warner BE, Velardo C, Salvi D, Lafferty K, Crosbie S, Herrington WG, et al. Feasibility of telemonitoring blood pressure in patients with kidney disease (Oxford Heart and Renal Protection Study-1): observational study. JMIR Cardio 2018 Dec 21;2(2):e11332 [FREE Full text] [doi: 10.2196/11332] [Medline: 30596204]
18. Personalised electronic record supported optimisation when alone for patients with hypertension (PERSONALCovdBP). ClinicalTrials.gov. 2020. URL: https://clinicaltrials.gov/study/NCT04559074 [accessed 2024-10-19]
19. Collier DJ, Taylor M, Godec T, Shiel J, James R, Chowdury Y, et al. Personalized antihypertensive treatment optimization with smartphone-enabled remote precision dosing of amlodipine during the COVID-19 pandemic (PERSONAL-CovidBP Trial). J Am Heart Assoc 2024 Feb 20;13(4):e030749 [FREE Full text] [doi: 10.1161/JAHA.123.030749] [Medline: 38323513]
20. Hypertension in adults: diagnosis and management. NICE. 2019. URL: https://www.nice.org.uk/guidance/ng136/chapter/Recommendations#diagnosing-hypertension [accessed 2024-10-19]
21. Whelton PK, Carey RM, Aronow WS, Casey DE, Collins KJ, Himmelfarb CD, et al. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/pcna guideline for the prevention, detection, evaluation, and management of high blood pressure in adults: executive summary: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. Hypertension 2018;71(6):e13-e115 [FREE Full text] [doi: 10.1161/HYP.0000000000000065] [Medline: 29133356]
22. Tucker KL, Sheppard JP, Stevens R, Bosworth HB, Bove A, Bray EP, et al. Self-monitoring of blood pressure in hypertension: a systematic review and individual patient data meta-analysis. PLoS Med 2017 Sep 19;14(9):e1002389 [FREE Full text] [doi: 10.1371/journal.pmed.1002389] [Medline: 28926573]
23. McManus RJ, Little P, Stuart B, Morton K, Raftery J, Kelly J, et al. Home and online management and evaluation of blood pressure (HOME BP) using a digital intervention in poorly controlled hypertension: randomised controlled trial. BMJ 2021 Jan 19;372:m4858 [FREE Full text] [doi: 10.1136/bmj.m4858] [Medline: 33468518]
24. Park S, Shin J, Park J, Choi W. An updated meta-analysis of remote blood pressure monitoring in urban-dwelling patients with hypertension. Int J Environ Res Public Health 2021 Oct 9;18(20):10583 [FREE Full text] [doi: 10.3390/ijerph182010583] [Medline: 34682329]

Abbreviations

ACEi: angiotensin-converting enzyme inhibitor
AE: adverse event
ARB: angiotensin receptor blocker
BP: blood pressure
CCB: calcium channel blocker
CLM: Closed Loop Medicine
CRO: contract research organization
DBP: diastolic blood pressure
SAE: Serious Adverse Event
SBP: systolic blood pressure

UEQ: User Experience Questionnaire
VAS: visual analog scale

Edited by L Buis; submitted 10.10.23; peer-reviewed by R Marshall, X Wang; comments to author 11.04.24; revised version received 05.06.24; accepted 07.10.24; published 03.12.24.

Please cite as:
Richardson L, Noori N, Fantham J, Timlin G, Siddle J, Godec T, Taylor M, Baum C
Personalized Smartphone-Enabled Assessment of Blood Pressure and Its Treatment During the SARS-CoV-2 COVID-19 Pandemic in Patients From the CURE-19 Study: Longitudinal Observational Study
JMIR Mhealth Uhealth 2024;12:e53430
URL: https://mhealth.jmir.org/2024/1/e53430
doi:10.2196/53430
PMID:

Smartphone Screen Time Characteristics in People With Suicidal Thoughts: Retrospective Observational Data Analysis Study

Marta Karas[1], PhD; Debbie Huang[1], PhD; Zachary Clement[1], MS; Alexander J Millner[2], PhD; Evan M Kleiman[3], PhD; Kate H Bentley[2,4,5,6], PhD; Kelly L Zuromski[2,7], PhD; Rebecca G Fortgang[2,4,5,6], PhD; Dylan DeMarco[2], BA; Adam Haim[8], PhD; Abigail Donovan[4,5], MD; Ralph J Buonopane[5,7], PhD; Suzanne A Bird[4,5], MD; Jordan W Smoller[4,5,6,9,10], MD; Matthew K Nock[2,5,7], PhD; Jukka-Pekka Onnela[1], PhD

[1]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Harvard University, Boston, MA, United States

[10]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Boston, MA, United States

[2]Department of Psychology, Harvard University, Cambridge, MA, United States

[3]Department of Psychology, Rutgers University, Piscataway, NJ, United States

[4]Department of Psychiatry, Massachusetts General Hospital, Boston, MA, United States

[5]Department of Psychiatry, Harvard Medical School, Boston, MA, United States

[6]Center for Precision Psychiatry, Massachusetts General Hospital, Boston, MA, United States

[7]Franciscan Children's, Mental Health Research, Brighton, MA, United States

[8]National Institute of Mental Health, Bethesda, MD, United States

[9]Psychiatric and Neurodevelopmental Genetics Unit, Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, United States

Corresponding Author:
Marta Karas, PhD
Department of Biostatistics, Harvard T.H. Chan School of Public Health, Harvard University, , Boston, MA, , United States

Abstract

Background: Smartphone-based monitoring in natural settings provides opportunities to monitor mental health behaviors, including suicidal thoughts and behaviors. To date, most suicidal thoughts and behaviors research using smartphones has primarily relied on collecting so-called “active” data, requiring participants to engage by completing surveys. Data collected passively from smartphone sensors and logs may offer an objectively measured representation of an individual’s behavior, including smartphone screen time.

Objective: This study aims to present methods for identifying screen-on bouts and deriving screen time characteristics from passively collected smartphone state logs and to estimate daily smartphone screen time in people with suicidal thinking, providing a more reliable alternative to traditional self-report.

Methods: Participants (N=126; median age 22, IQR 16-33 years) installed the Beiwe app (Harvard University) on their smartphones, which passively collected phone state logs for up to 6 months after discharge from an inpatient psychiatric unit (adolescents) or emergency department visit (adults). We derived daily screen time measures from these logs, including screen-on time, screen-on bout duration, screen-off bout duration, and screen-on bout count. We estimated the mean of these measures across age subgroups (adults and adolescents), phone operating systems (Android and iOS), and monitoring stages after the discharge (first 4 weeks vs subsequent weeks). We evaluated the sensitivity of daily screen time measures to changes in the parameters of the screen-on bout identification method. Additionally, we estimated the impact of a daylight time change on minute-level screen time using function-on-scalar generalized linear mixed-effects regression.

Results: The median monitoring period was 169 (IQR 42 - 169) days. For adolescents and adults, mean daily screen-on time was 254.6 (95% CI 231.4-277.7) and 271.0 (95% CI 252.2-289.8) minutes, mean daily screen-on bout duration was 4.233 (95% CI 3.565-4.902) and 4.998 (95% CI 4.455-5.541) minutes, mean daily screen-off bout duration was 25.90 (95% CI 20.09-31.71) and 26.90 (95% CI 22.18-31.66) minutes, and mean daily screen-on bout count (natural logarithm transformed) was 4.192 (95% CI 4.041-4.343) and 4.090 (95% CI 3.968-4.213), respectively; there were no significant differences between smartphone operating systems (all *P* values were >.05). The daily measures were not significantly different for the first 4 weeks compared to the fifth week onward (all *P* values were >.05), except average screen-on bout in adults (*P* value = .018). Our sensitivity analysis indicated that in the screen-on bout identification method, the cap on an individual screen-on bout duration has a substantial effect on the resulting daily screen time measures. We observed time windows with a statistically significant effect of daylight time change on screen-on time (based on 95% joint confidence intervals bands), plausibly attributable to sleep time adjustments related to clock changes.

Conclusions: Passively collected phone logs offer an alternative to self-report measures for studying smartphone screen time characteristics in people with suicidal thinking. Our work demonstrates the feasibility of this approach, opening doors for further research on the associations between daily screen time, mental health, and other factors.

(JMIR Mhealth Uhealth 2024;12:e57439) doi:10.2196/57439

KEYWORDS

smartphone; mobile apps; mobile health; screen time; suicidal thoughts and behavior; suicidal; app; observational data; data analysis study; monitor; survey; psychiatric; screen; mental health; feasibility; suicidal ideation; mobile phone

Introduction

The widespread use of smartphones has created new opportunities for capturing social, behavioral, and cognitive phenotypes in free-living settings [1]. The ability to collect data from individuals in their natural environments, rather than in controlled laboratory settings, allows for a more complete ascertainment of an individual's behavior. In addition, the use of smartphones allows for near-continuous data collection, enabling a more detailed and dynamic view of an individual's behavior over time. This approach is well suited to study mental health disorders that are often characterized by a fluctuating and recurrent course [2].

Suicide is a complex mental health problem that is characterized as an "act of intentionally ending one's own life" [3]. According to the National Institute of Mental Health, in 2021, more than 48,000 people died by suicide in the United States [4]. Suicide is often triggered by one or a combination of factors, including mental illness, substance use disorders, traumatic life events, and social isolation [3].

To date, in suicidal thoughts and behaviors (STB) research, smartphone-based monitoring has been primarily used to collect "active" data, which requires a participant to actively engage by contributing data entries via surveys. For example, smartphones can be used to collect ecological momentary assessment (EMA) data, which are frequent surveys on an individual's thoughts, feelings, and behaviors in their natural environment [5,6]. EMA deployed through a smartphone app has been successfully used to understand potential risk factors for STB and characterize dynamics in suicidal ideation over time [7-13]. Importantly, the ability to deploy active data collection via smartphones and access EMA outcomes in real time has given rise to ethical and safety concerns, including when and how to intervene if a participant's responses indicate an elevated risk during the course of the study [14].

The use of smartphone "passive" data, that is, data collected from smartphone sensors and logs without any active engagement from a participant, is less common in STB research. Existing studies have mostly relied on proprietary mobile apps and analytic methods [15]. Using smartphone log data to obtain detailed characteristics of an individual's screen time is a potentially compelling application in STB studies due to the growing research focus on the association between screen time and mental health. Indeed, existing studies have demonstrated a link between increased levels of anxiety and depression and high smartphone screen time, measured either objectively or via a self-report [16,17]. In addition, objectively measured excessive screen time on mobile devices has been linked with a decreased quantity of sleep, which can further exacerbate mental health issues [18]. Interestingly, Rozgonjuk et al [19] reported that in a sample of 101 undergraduate university students, depression and anxiety symptom severity negatively correlated with the frequency of phone screen unlocking but were not related to total screen time; this finding suggests that not only screen-on total duration time but also screen-on use patterns are potentially relevant.

Data missingness is a common issue in digital health studies [20,21]. For smartphone passive data, Kiang et al [22] estimated missingness due to sensor noncollection rates to be 19% for accelerometer data and 27% for GPS data across 6 digital phenotyping studies. Sensor noncollection has been attributed to users' behavior (eg, forgetting to charge a phone, disabling the GPS, and uninstalling the study app) and technological factors (the phone operating system restricting data collection). If the expected data volume or temporal coverage is fixed by study design (as was the case in these studies), one can identify smartphone sensor noncollection by comparing the expected versus observed data volume or temporal coverage. Determining missing data in the context of estimating smartphone screen time from phone state logs is more complicated, and without identifying the missingness, results may be biased. To the best of our knowledge, to date, no systematic approach has been proposed to determine the missingness status of smartphone state logs.

In this paper, we present methods for identifying and characterizing smartphone screen-on bouts, which we define as periods of consecutive smartphone use, in smartphone state logs. These methods were applied to a sample of adolescents (n=50) and adults (n=76) with suicidal thinking. Data were collected for up to 6 months from participants' personal Android and iOS phones using the Beiwe smartphone app (Harvard University) [23]. Our methodological contributions are as follows. First, we present a method for preprocessing raw smartphone state logs into discrete screen-on bouts, capturing the duration and timing of each screen interaction, and handling missingness due to sensor non-collection; additionally, we performed multiple sensitivity analyses for preprocessing of raw logs. Second, we extracted several day-level metrics from the preprocessed bouts, including total screen time volume and fragmentation metrics that quantify the temporal distribution of screen use. Third, we compared screen time measures from early versus later stages of monitoring (which starts after the presentation of suicidal thinking) to identify use pattern shifts potentially associated with recent suicidal thoughts. Finally, we demonstrated how functional data analysis regression techniques

can be used to analyze complex minute-level screen time outcomes. We estimated a time-varying effect of a daylight saving time (DST) change on the outcome using the natural experiment this event creates. The code for all preprocessing and analysis steps is openly available on GitHub [24].

Methods

Ethical Considerations

This study was conducted ethically and in accordance with relevant guidelines and regulations. Institutional review board approval from Harvard University (IRB18-1749) was obtained before the study began. All data collection and maintenance followed university, hospital, state, and national policies and regulations. Participants gave informed consent before taking part in any study procedures. All identifying information, such as names, initials, or hospital numbers, has been omitted from the data used in this analysis to protect participant privacy.

Study Design and Population

The study recruited 2 samples of participants: adults from Massachusetts General Hospital's Acute Psychiatry Service and children or adolescents (and their parents or guardians) from Franciscan Children's psychiatric inpatient unit. Adult participants were eligible if they were 18 years or older, with other criteria including owning a smartphone and presenting to the emergency department with suicidal thoughts. Children and adolescents aged 12 - 19 years were recruited from Franciscan Children's, with eligibility criteria including smartphone ownership, parental consent for minors, child or adolescent assent, and presenting problems that included suicidal thoughts. The recruitment process involved initial identification and informed consent, and it was conducted via both an in-person manner and a remote manner (but fully remote during heightened periods of the COVID-19 pandemic). The recruitment process was designed to promote diversity among participants, without restrictions based on diagnosis or clinical history.

Phone State Logs

Phone state logs were collected with the Beiwe smartphone app, the front end of the open-source Beiwe high-throughput digital phenotyping platform [23]. The Beiwe platform consists of smartphone apps for Android and iOS, a web-based platform for study administration, and a cloud-based back-end system for data storage and processing. Participants were assigned a Beiwe participant ID and had the Beiwe app installed during enrollment, which defined a study start date (Multimedia Appendix 1). The phone state logs were recorded with millisecond-level timestamps. iOS and Android phones report state logs differently (Table S1 in Multimedia Appendix 1). For Android, the logs capture "screen turned on" and "screen turned off" events; for iOS, the logs capture "locked" and unlocked" events. For iOS, the log also includes an event for each 1% change (positive or negative) in battery charge level.

Screen Time Estimation

We define a "screen-on bout" for a smartphone as a period of consecutive screen use and a "screen-off bout" as a period of consecutive screen nonuse. Each screen-on bout is followed by a screen-off bout, and vice versa. To estimate the timing and duration of screen-on bouts for iOS, we used the time intervals between consecutive "unlocked" and "locked" event timestamps, while for Android, we used the time intervals between consecutive "screen turned on" and "screen turned off" event timestamps. During preprocessing, we imputed missing logs, removed bouts attributed to notification arrivals (for Android only), and capped screen-on bouts that lasted longer than 30 minutes (which corresponds to approximately the 97th percentile). A detailed description of our methods is provided in Multimedia Appendix 1. Our preprocessing steps closely resemble those previously presented by Kristensen et al [25]. For comparison, we also estimated timing and duration of screen-on bouts using three comparator approaches: (1) imputing missing logs and capping screen-on bout duration at 6 hours (instead of 30 minutes); (2) not imputing missing logs, that is, only considering consecutive pairs of matching events ("unlocked" and "locked" for iOS and "screen turned on" and "screen turned off" for Android), and not capping screen-on bout duration (ie, minimal preprocessing); and (3) not imputing missing logs with screen-on bout duration capped at 30 minutes.

Daily Measures of Screen Time

To obtain daily measures of smartphone screen time, we used the estimated screen-on bouts and calculated 4 metrics: total screen-on time, average screen-on bout duration, average screen-off bout duration, and screen-on bout count. We defined a day as the period from midnight to midnight in the Coordinated Universal Time (UTC) time zone. The smartphone state logs were originally recorded in UTC, and we assumed the time zone choice would have no impact on the statistical analyses. The total screen-on time and screen-on bout count metrics capture the volume of screen time, while the average bout duration metrics reflect screen-on time accumulation patterns.

Missing Data Labeling

A heuristic was developed to identify periods of missing phone state logs for iOS based on changes in battery charge level log events. Our approach assumes that the battery charge level is in one of four states: (1) decreasing, (2) increasing (when the phone is being charged), (3) constant at 0% (when the phone battery is fully depleted), and (4) constant at 100% (when the phone battery is fully charged and the charger is plugged in). First, we defined a minute of the monitoring time as "valid" based on either of two criteria: (1) phone's battery level was changing at least 1% per hour (decreasing or increasing) or (2) the phone had been recently charged to 100% and not depleted battery for a period of at most 12 hours (likely having a charger plugged in). Minutes that did not meet these criteria were labeled as invalid. A "valid day" was defined as one with at least 1080 valid minutes (18 hours). Unlike for iOS, we were unable to define "valid minutes" for Android due to the absence of phone battery level change logs. Therefore, we only defined "valid days" for Android data as those consisting of at least 8 distinct hours that contained at least 1 screen-on bout. Only valid days were included in the statistical analyses.

Statistical Data Analysis

The data analysis sample consisted of participants who provided at least 28 valid days of data. A day was considered valid if it belonged to a 28-day period with at least 14 valid days. We computed the number of days in the study observation period and the number of valid days of phone state logs for each participant and characterized these measures using the median and range, separately for adolescents, adults, and the combined sample.

To visualize daily measures over time, we aggregated the measures (median, 25th, 75th, 10th, and 90th percentile) across participants for each relative day since the discharge, separately for adolescents and adults. We used a 7-day moving average to smooth the sample statistics and improve readability.

To quantify population-level daily measures of screen time, we considered 4 different linear mixed-effects models (LMMs). Each model had a participant-specific daily measure as the outcome. The LMMs had different fixed-effect coefficients: model 1 had a fixed effect for an age group (adolescents, adults); model 2 had a fixed effect for an age group, operating system (Android, iOS), and their interaction; model 3 had a fixed effect for an age group, monitoring period (one of: the first four weeks after discharge period, from the fifth week up to 6 months period), and their interaction; and model 4 had a fixed effect for an age group, study monitoring period (one of: the first four weeks after discharge period, from the fifth week up to 6 months period), and their interaction. Each LMM included a participant-specific random intercept. Additionally, models 3 and 4 also had a subject-specific random slope for a study period. We fitted each of these 4 different LMMs separately for each of the 4 daily measures: total screen-on time, average screen-on bout duration, average screen-off bout duration, and screen-on bout count (natural logarithm-transformed), yielding a total of 4×4=16 model fits. Using these models, least squares means of daily measures were estimated across age groups, operating systems, and study periods of interest. We also quantified the contrasts of interest: the difference between age groups from model 1, the difference between operating systems across age groups from model 2, and the difference between (differently defined) study periods across age groups from models 3 and 4. The statistical significance of the contrasts of interest was evaluated using a significance level of α=.05.

We calculated screen-on time for each minute of each participant day using the timing and duration of screen-on bouts. The function-on-scalar generalized LMM (FoS-GLMM) was used to estimate the time-varying effect of daylight saving time (DST) on the probability of screen-on time at the minute level. We chose to study the effect of DST changes because they create a natural experiment that allows us to assess the feasibility of our method. Specifically, these changes provide a sudden and exogenous shift in the timing of daylight hours, which can impact individuals' daily routines and phone use patterns. By examining the minute-level screen time outcome around these time changes, we explore whether our method captures changes in phone use.

In our FoS-GLMM, the outcome was a participant- and day-specific functional observation recorded on a minute-level discrete grid, with a value of 1 if any screen-on time was recorded for that minute and 0 otherwise. To quantify the effect of DST, we included a time-varying fixed-effect indicator for the change effect (1 if the functional observation was from a time period after the DST change and 0 otherwise) and a time-varying participant-specific intercept. The model was estimated separately at the start and end of DST (March and November of 2019 - 2022, respectively) using data from ±14 days from the time change, and we performed a sensitivity analysis using data from ±7-day and ±28-day windows. We estimated FoS-GLMMs using the fast inference approach for longitudinal functional models proposed by Cui et al [26], and we made inferences about the time-varying fixed effect based on 95% joint confidence bands.

Results

Sample Characteristics

In total, 297 participants had the Beiwe app installed during enrollment. Of those, 247 participants had at least 1 valid day of data. The final analysis sample consisted of 126 participants (Table 1). The median age of the combined sample (adolescents and adults) was 22 (IQR 12-69) years, with adolescents having a median age of 15 (IQR 12-18) years and adults having a median age of 30 (IQR 18-69) years. The sample was predominantly female (n=81, 64.3%) and self-identified as White (n=101, 80.2%). Most participants (n=75, 59.5%) used smartphones with iOS, but the proportion of iOS users differed between the 2 subsamples, with 84% (42/50) in adolescents and 43.4% (33/76) in adults. The study observation period had a median duration of 169 (IQR 42-169) days, and participants had a median of 94 (IQR 29-167) days of valid smartphone state logs.

Table . Baseline demographic and observation period metadata characteristics of the analysis sample.[a]

Characteristic (statistic)	Statistic value		
	Adolescents (n=50)	Adults (n=76)	Combined (N=126)
Age (years), median (IQR)	15 (12-18)	30 (18-69)	22 (12-69)
Sex, n (%)			
Female (cisgender, she or her pron[b] or unspecified)	35 (70)	40 (52.6)	75 (59.5)
Female (transgender)	0 (0)	3 (3.9)	3 (2.4)
Female (cisgender, other than she or her pron)	2 (4)	1 (1.3)	3 (2.4)
Male (cisgender, he or him pron or unspecified)	11 (22)	31 (40.8)	42 (33.3)
Male (transgender)	1 (2)	1 (1.3)	2 (1.6)
Male (cisgender, other than he or him pron)	1 (2)	0 (0)	1 (0.8)
Self-identified race, n (%)			
Asian	3 (6)	2 (2.6)	5 (4)
Black or African American	1 (2)	6 (7.9)	7 (5.6)
White	41 (82)	60 (78.9)	101 (80.2)
More than 1 race	1 (2)	2 (2.6)	3 (2.4)
Other	0 (0)	3 (3.9)	3 (2.4)
Unavailable	4 (8)	3 (3.9)	7 (5.6)
Smartphone operating system, n (%)			
iOS	42 (84)	33 (43.4)	75 (59.5)
Android	8 (16)	43 (56.6)	51 (40.5)
Observation period metadata, median (IQR)			
Days in observation period	169 (51-169)	169 (42-169)	169 (42-169)
Smartphone state logs (valid data days)	103 (34-158)	86 (29-167)	94 (29-167)

[a]The number of valid data days of smartphone state logs was determined differently for iOS and Android operating systems due to the differences in the type of smartphone state logs recorded.

[b]pron: gender pronouns use.

Daily Measures of Screen Time

Table 2 summarizes the estimated population-level mean daily smartphone screen time. Adults had a mean total screen-on time of 271.0 minutes (4.52 hours), while adolescents had a mean 254.6 minutes (4.24 hours). Adults showed slightly higher average screen-on bout duration (mean 4.998 vs 4.233 minutes), slightly higher average screen-off bout duration (mean 26.90 vs 25.90 minutes), and slightly lower average screen-on bout count (natural logarithm transformed; mean 4.090 vs 4.192) compared to adolescents. None of these differences were statistically significant (Table S3 in Multimedia Appendix 1, rows 1 - 4). Mean daily smartphone screen time measures across different phone operating systems and age groups (adults and adolescents) are shown in Table S2 in Multimedia Appendix 1 (rows 9 - 24). There were no statistically significant differences in any of the 4 screen time measures across phone operating systems within any age group (Table S3 in Multimedia Appendix 1, rows 5 - 12).

Table . Mean daily screen time measures separately for age groups (adolescents and adults).

Daily measure and age group		Daily measure mean (95% CI)
Total screen-on time (minutes)		
	Adolescent	254.6 (231.4-277.7)
	Adult	271.0 (252.2-289.8)
Average screen-on bout (minutes)		
	Adolescent	4.233 (3.565-4.902)
	Adult	4.998 (4.455-5.541)
Average screen-off bout (minutes)		
	Adolescent	25.90 (20.09-31.71)
	Adult	26.90 (22.18-31.62)
Log [a] (screen-on bout count)		
	Adolescent	4.192 (4.041-4.343)
	Adult	4.090 (3.968-4.213)

[a]log(): natural logarithm transformation.

Figure 1 displays heat maps of minute-level smartphone screen time for 4 participants. Screen time per minute is shown across minutes of a day for each day of the participant's observation period. The figure highlights various between- and within-participant screen time patterns. Participant ID 48 (plot A) has a relatively low average total screen-on time (mean 166.7 minutes across the monitoring days) and short average screen-on bouts (mean 0.9 minutes), whereas participant ID 92 (plot B) has a relatively high average total screen-on time (mean 351.0 minutes) and a moderate average screen-on bout duration (mean 4.3 minutes). Participant ID 20 and ID 102 (plots C andD) have similar total screen-on times (mean 253.6 and 266.7 minutes, respectively), but different accumulation patterns, as measured with average screen-on bout durations (mean 2.5 and 8.8 minutes, respectively) and average screen-off bout durations (mean 12.2 and 42.8 minutes, respectively). Participant ID 48 experiences a decline in daily total screen-on time around their 50th day relative to the study start. Participant ID 92 exhibits a pattern of waking up at the same time across many of the 5-day long periods of time (for 15th-80th relative days: waking up at 10:30 AM UTC, ie, 6:30 AM in their current Eastern Daylight Time [EDT] time zone; and for 80th-180th relative days: waking up at 11:30 AM UTC, ie, 6:30 AM in their current EST time zone). Participant ID 102 appears to have very irregular sleep and wake-up time and has screen-on time bouts scattered across the Eastern Time nighttime on multiple days.

Figure 2 displays the median, 25th-75th percentile bounds, and 10th-90th percentile bounds of the screen time daily measures aggregated across participants for each day relative to the study start, separately for adolescents and adults. Although some day-to-day variability is evident, we did not observe any specific patterns during the early versus later monitoring period, which is consistent with the results of our LMMs. Our analysis did not identify any statistically significant differences in population-level daily measures between the first 4 weeks and week 5 and onward nor between the first week and week 5 and onward after the discharge, except for 2 cases: in the adult subsample, the average screen-on bout duration was significantly lower during the first 4 weeks than week 5 and onward; and in the adolescent subsample, the average screen-off bout duration was significantly higher during the first 4 weeks than week 5 and onward. Please refer to Table S2 in Multimedia Appendix 1 for daily measure means across different treatment periods by age group (rows 25 - 56) and to Table S3 in Multimedia Appendix 1 for difference between treatment periods by age group (rows 13 - 28).

In addition to our proposed approach, we estimated the timing and duration of screen-on bouts using 3 comparator methods. Figure S1 illustratesin Multimedia Appendix 1 illustrates the differences in daily measures of total screen-on time and average screen-on bout duration among participants. Comparators 1 and 2 produced substantially higher estimates for daily total screen-on time, with mean differences of −106.1 (range −502.6 to −1.3) minutes and −133.5 (range −666.4 to −2.4) minutes, respectively. These methods used either a generous cap of 6 hours on an individual screen-on bout duration (comparator 1) or no cap at all (comparator 2). In contrast, comparator 3 (no imputation, capping screen-on bout duration at 30 minutes) yielded results similar to our proposed approach, with an average difference of −0.2 (range −11.8 to 7.3) minutes.

Figure 1. Characteristics of minute-level smartphone screen time for 4 participants. Screen time per minute is shown (color-coded; expressed in seconds) across minutes of the day (y-axis; UTC time) for each day of participant's monitoring period relatively to the study start (x-axis). Days labeled as invalid data days are shadowed in gray. Plot A: ID 48 (iOS, adolescent) with the following means of daily measures: total screen-on time=166.7 minutes, average screen-on bout=0.9 minutes, and average screen-off bout=8.4 minutes. Plot B: ID 92 (iOS, adolescent): total screen-on time=351.0 minutes, average screen-on bout=4.3 minutes, and average screen-off bout=13.8 minutes. Plot C: ID 20 (iOS, adolescent): total screen-on time=253.6 minutes, average screen-on bout=2.5 minutes, and average screen-off bout=12.2 minutes. Plot D: ID 102 (iOS, adolescent): total screen-on time=266.7 minutes, average screen-on bout=8.8 minutes, and average screen-off bout=43.8 minutes. UTC: Coordinated Universal Time.

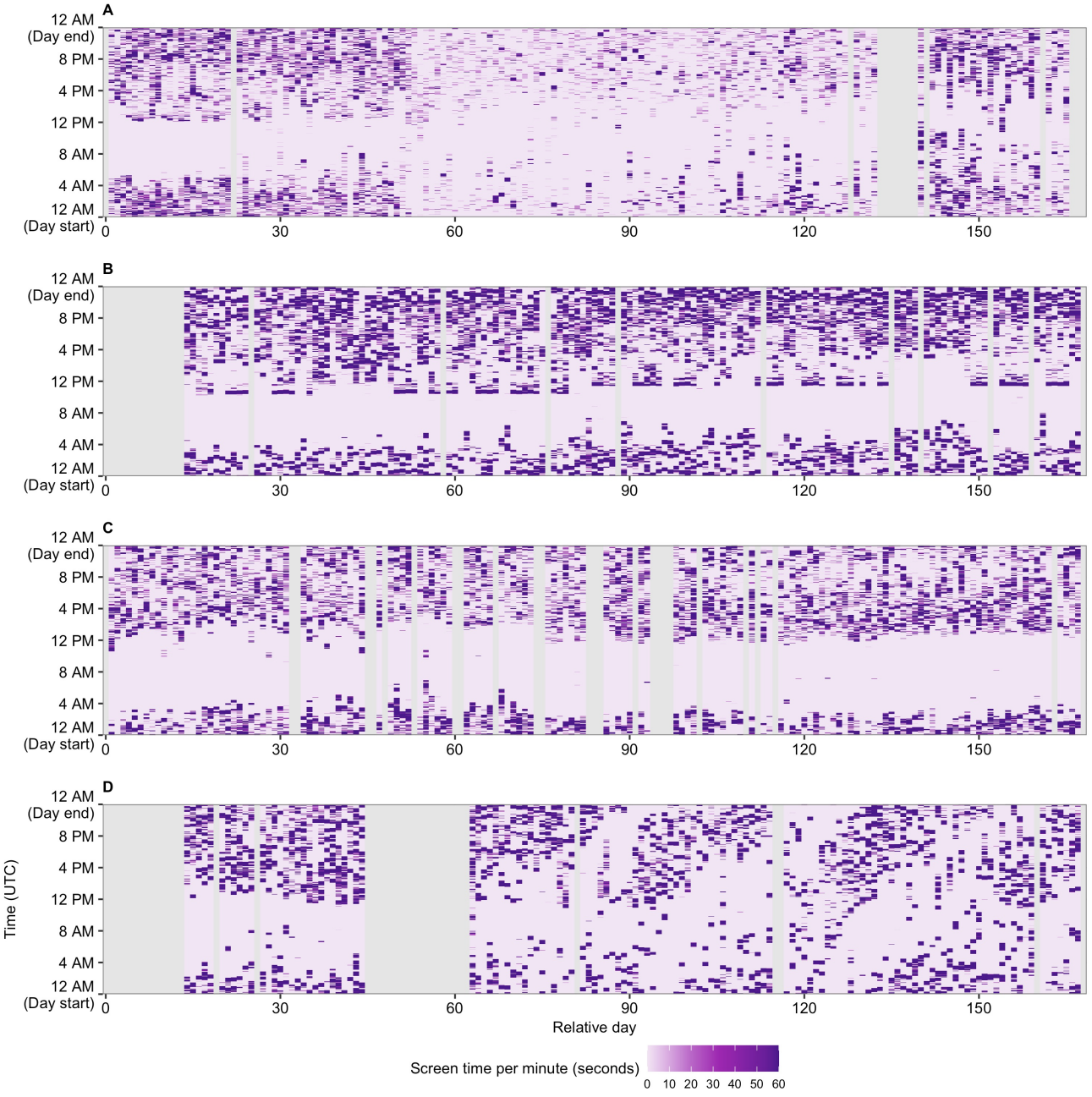
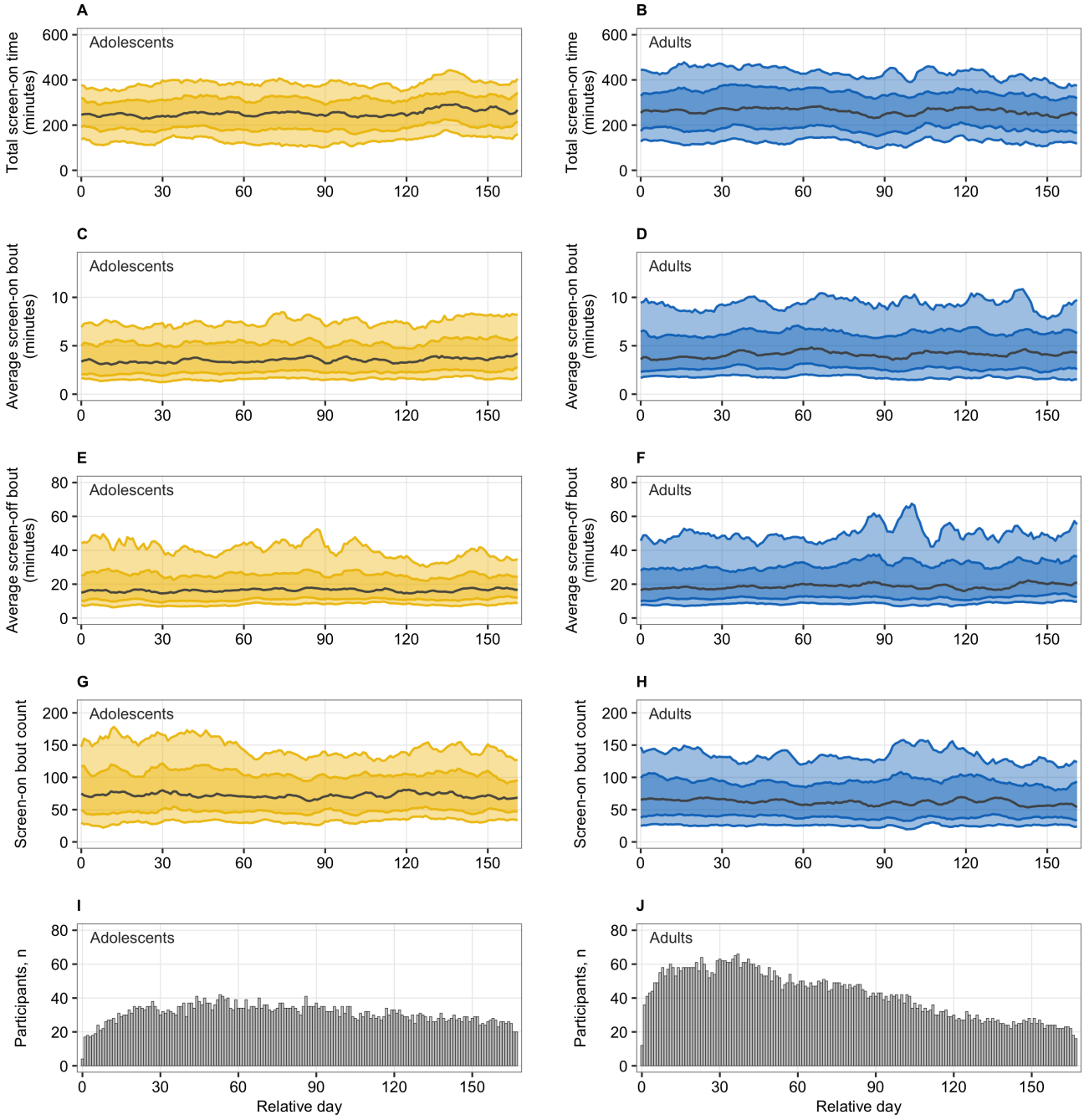


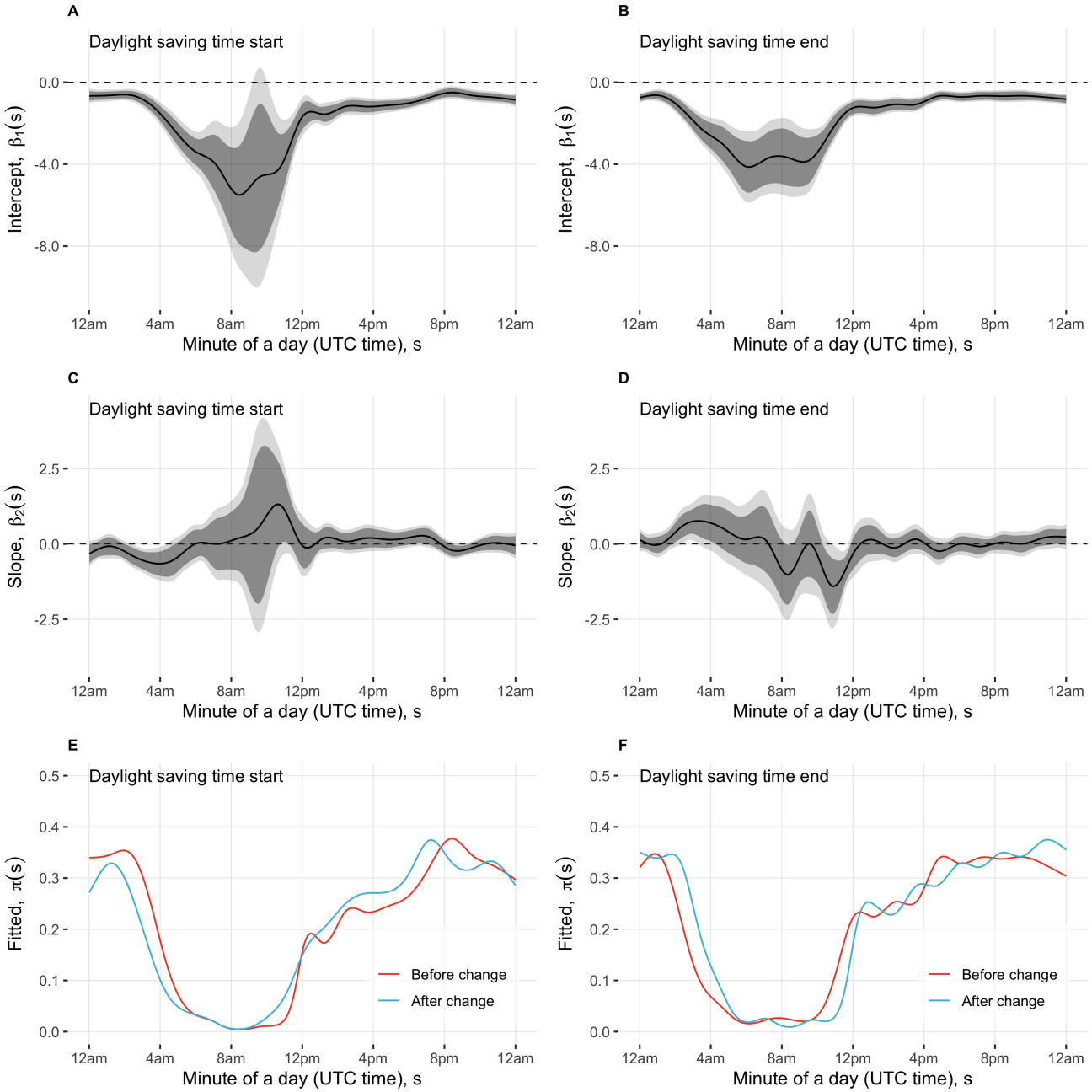
Figure 2. Characteristics of daily screen time measures across time relative from study start. Plots A-H show sample median (black line), 25th-75th percentile bounds (darker color ribbon), and 10th-90th percentile bounds (lighter color ribbon) of day-level measures: total screen-on time (in minutes; plots A and B), average screen-on bout duration (in minutes; plots C and D), average screen-off bout duration (in minutes; plots E and F), and screen-on bout count (plots G and H). Plots I and J show the number of participants contributing a valid day of phone state logs data on a given day relative from the study start.



Time-Varying Effect of a Daylight Time Change on Minute-Level Screen Time

We used FoS-GLMM to estimate a time-varying effect of a daylight time change on the probability of screen-on time at the minute level. Figure 3 shows estimates of a functional intercept and a functional slope for the fixed-effect covariate (1=after and 0=before DST change; plots A-D) as well as fitted probabilities of screen time (plots E and F) in a given minute across the functional domain of a day.

Figure 3. Estimates of functional intercept and slope for daylight saving time (plots A-D) and fitted probabilities of screen time (plots E and F) in a given minute across the minute-level functional domain of a day. Logistic function-on-scalar mixed-effects regression was used to estimate the time-varying effect of the covariate (1=after and 0=before daylight saving time change) on the functional outcome (1=any screen time and 0=no screen time). In rows 1 - 2 (plots A-D), functional coefficient estimates (solid black line), 95% point-wise CIs (dark gray shaded area), and 95% joint CIs (light gray shaded area) are presented. In row 3 (plots E and F), fitted probabilities for before and after the change are shown, with vertical dashed lines indicating the time window with a statistically significant effect based on joint CIs. Separate models were fitted for the start (column 1; plots A, C, and E) and end (column 2; plots B, D, and F) of daylight saving time, each using a ±14-day window of data relative to the change night. UTC: Coordinated Universal Time.



With the start of the DST model, we observed statistically significant negative effects (based on 95% joint confidence bands) on the probability of screen-on time after the time change during the period from 2:24 AM UTC to 4:19 AM UTC (equivalent to 9:24 PM EST to 11:19 PM EST before the time change). In this timeframe, the functional coefficient estimate had an average value −0.57 (range −0.65 to −0.38). This estimate suggests that the odds of phone use after the time change were 0.56 times lower compared to the odds before the time change. The average probabilities of phone use in a given minute were 0.26 and 0.17 for before and after the change in that period, respectively, as depicted by the blue and red lines in Figure 3E, respectively (located between the dashed vertical lines).

With the end of the DST model, we observed statistically significant positive effects (based on 95% joint confidence bands) on the probability of screen-on time after the time change from 2:22 AM UTC to 3:19 AM UTC (10:24 PM EDT to 11:19 PM EDT before the change) and statistically significant negative effects (based on 95% joint confidence bands) during the time window from 10:52 AM UTC to 11:44 AM UTC (6:52 AM EDT to 7:44 AM EDT before the change). In the time window of positive effect significance, the functional coefficient estimate

had an average value 0.71 (range 0.58-0.77), corresponding to the odds of phone use after the time change being 2.03 times the odds before the change, and average probabilities of phone use in a given minute 0.15 and 0.27 for before and after the change, respectively (Figure 3E). In the time window of negative effect significance, the functional coefficient estimate had an average value −1.13 (range −1.40 to −0.66), corresponding to odds of phone use after the time change being 0.32 times the odds before the change, and average probabilities of phone use in a given minute 0.15 and 0.06 for before and after the change, respectively (Figure 3E).

Based on sensitivity analysis results (Figure S2 in Multimedia Appendix 1), the inference was very similar for cases of using ±14 days (main analysis) and ±28 days, whereas for the ±7-day model, the 95% joint confidence bands do overlap with 0 value for the functional coefficient in the DST end model.

Taken together, these results suggest that the effects on screen-on time during these time windows are plausibly attributable to sleep time adjustments related to clocks going 1 hour forward (“we sleep less”) or 1 hour backward (“we sleep more”).

Discussion

Screen Time Data in STB Studies

Digital phenotyping entails the collection and analysis of various types of data from personal digital devices in naturalistic settings; it has applications both within and outside mental health. The focus of our paper is the development of a simple method to enable investigators to use moment-by-moment screen time metrics when studying adolescents and adults with STB. We emphasize that in our approach, the data are collected in naturalistic settings and arise as a byproduct of how participants use their phones, which makes the approach scalable.

In the digital phenotyping literature in mental health, there is surprisingly little existing research on the intersection of STB and screen time. For example, a recent literature review only briefly mentions STB and does not discuss screen time [27]. Another paper focusing on stress, anxiety, and mild depression carried out a systematic review of 40 studies [28]; 7 studies with “student participants” and 2 studies with “adult participants” used screen time data (screen on or off, phone lock or unlock, and similar metrics), but none of the reviewed studies dealt with STB. Finally, a recent narrative review of digital phenotyping for differential diagnosis of major depressive disorder reviewed 74 papers dealing with “digital tools”; 4 of the papers focused on STB, but none were reported to discuss screen time [29].

Principal Findings

We investigated the smartphone screen time characteristics in 126 adolescents and adults with suicidal thinking. Passively collected smartphone state logs data for a median of 169 (IQR 42 - 169) days provided objective measurements of screen time. Our analysis showed that study participants spent an average of 254.6 and 271.0 minutes per day on their smartphones for adolescents and adults, respectively. The means of participants’ average screen-on bout duration were 4.2 and 5.0 minutes, average screen-off bout duration were 25.9 and 26.9 minutes, and screen-on bout count (natural logarithm transformed) were 4.2 and 4.1 for adolescents and adults, respectively.

The daily measures remained relatively constant across the monitoring period. The near constancy of the measures at the daily level could be perceived as a favorable result. First, nearly all methods that attempt to detect changes in temporal data, such as anomaly detection and change-point detection methods, need to establish an underlying trend of the measure over time. Substantial variation in the daily measures of screen time could complicate the detection of changes potentially due to STB episodes. With that, there were very few suicide attempts in this cohort during the monitoring period (data not reported in this paper), highlighting the clinical team’s effectiveness in ensuring participant safety. Finally, the crucial question of the longitudinal association between screen time and STB, as well as the identification of STB episodes, will be addressed in future work, with this paper laying the groundwork by introducing the necessary method.

Another finding was the significant impact of DST changes on screen use behavior. When DST started, a decrease in screen-on time probability was observed between 2:24 AM and 4:19 AM UTC, with average probabilities dropping from 0.26 to 0.17. Conversely, when DST ended, an increase in screen-on time probability occurred between 2:22 AM and 3:19 AM UTC, with average probabilities rising from 0.15 to 0.27. A further decrease was observed between 10:52 AM and 11:44 AM UTC, with average screen-on probabilities dropping from 0.15 to 0.06. These findings suggest that sleep adjustments related to DST shifts do influence screen use behaviors.

Our study highlights the advantages of phone log–derived measures of screen time compared to traditional self-report surveys. As noted by Harris et al [30-33], self-report scales often lack internal consistency and test-retest reliability, rely on the participant’s memory, and may result in both over- and underestimation of screen time. In contrast, phone logs provide an objectively measured representation of an individual’s smartphone screen time, making them particularly useful for identifying trends or patterns in screen time, especially in longitudinal studies and free-living settings.

Our framework derives 4 daily measures of screen time: total screen-on time, average screen-on bout duration, average screen-off bout duration, and screen-on bout count to enable us to characterize a phone use behavior. The framework allows adaptation for different time resolutions, allowing researchers to explore subday-level (including minute-level) use patterns. For instance, they can investigate diurnal variations in screen time, identify atypical use periods, or study the impact of specific events on phone interactions. In future work, we propose to use sudden deviations from a participant’s routine of phone use (and screen time in particular) to identify potential STB episodes, with the goal of ultimately developing clinical interventions.

Finally, to the best of our knowledge, this is the first presentation of freely available code for preprocessing and analysis of raw phone state logs [24]. Using open-source software and sharing publicly the analysis code are essential for ensuring transparency

and replicability of the results, promoting accountability, and building trust in the research process.

Limitations

The primary limitation is the lack of a gold standard for identifying missing smartphone state logs. Though related mental health research often fails to report data quality, accounting for missingness is crucial for accurate screen time estimates and valid comparisons [16]. We proposed a heuristic for iOS to identify missingness at the minute level based on battery level changes and a method for both platforms to label data quality by day. Additionally, we evaluated the sensitivity of screen time measures to missing log imputation and, inherently related, the choice of maximum screen-on bout duration threshold (Figure S1 in Multimedia Appendix 1). Results showed substantial variability in participants' mean total screen-on time and screen-on bout duration based on the chosen screen-on duration cap. Given the lack of a principled approach for selecting an optimal cap, researchers should be mindful that this choice can substantially influence measure estimates. We do not have a definitive reason for choosing the 30-minute threshold instead of other options. However, we believe this choice strikes a good balance by indicating when a long screen time session has likely occurred while minimizing the impact of outliers on our daily estimates. Our code is openly available for transparency.

We acknowledge the potential for differences in daily measures resulting from difference in how we identified on-screen bouts for iOS versus Android devices. In additional analyses (Tables S2 and S3 in Multimedia Appendix 1), we quantified the differences in daily measures between Android versus iOS across age groups (adolescents and adults). However, the observed difference sizes were relatively small, and no statistical significance between iOSs was found. It is important to note that the absence of statistical significance does not necessarily imply the absence of a true effect. Nevertheless, given the substantial sample size, we deemed the results sufficient for our analysis.

In the sensitivity analysis for time-varying effect of a daylight time change on minute-level screen time (Figure S2 in Multimedia Appendix 1), for the ±7-day model, the 95% joint confidence bands do overlap with 0 value for the functional coefficient in the DST end model. We note that the ±7-day model's insignificant results might be due to either small sample size or using conservative 95% joint confidence bands to make inferences about the subset of the whole functional domain. Future studies could consider joint confidence bands for a prespecified restricted subset of the functional domain to increase power.

Conclusions

Passively collected smartphone logs allowed us to estimate daily measures of screen time characteristics in a large sample of adolescents and adults with suicidal thinking over a half year–long monitoring period. Our work demonstrates the feasibility of this approach, opening doors for further research on the associations between daily screen time, mental health, and other factors.

Acknowledgments

The authors would like to thank the participants and their parents and guardians who devoted their time to participate in the study. The authors thank Peter Lund Kristensen for providing us with further explanation and engaging in the discussion of the details of his paper, which we relied on in this work. The authors thank Eli Jones for his assistance in determining the characteristics of smartphone state logs across phone operating systems. This research was supported by the National Institute of Mental Health (grant U01MH116928) and by the Fuss Family Research Fund. RGF is supported by the National Institute of Mental Health (grant K23MH132766). KLZ is supported by the National Institute of Mental Health (K23MH120439). KHB is supported by the National Institute of Mental Health (K23MH120436).

Data Availability

A project repository with all data preprocessing and analysis code is publicly available on GitHub (onnela-lab/stb-beiwe-screen-time).

Authors' Contributions

MK contributed to the study concept and design, conducted data preprocessing and analysis, and wrote the first manuscript draft. DH contributed to the study concept and design and the data preprocessing and analysis concept. ZC managed operations related to data access and processing infrastructure. AJM was involved in the concept and design and provided a critical review of the manuscript. EMK, KHB, KLZ, RGF, DD, AH, AD, RJB, SAB, and JWS contributed to data collection and provided a critical review of the manuscript. MKN provided funding, contributed to the study concept and design, and provided a critical review of the manuscript. JPO contributed to the study concept and design and the data preprocessing and analysis concept and provided major edits to the manuscript draft. All authors reviewed the final manuscript.

Conflicts of Interest

MKN receives publication royalties from Macmillan, Pearson, and UpToDate. He has been a paid consultant in the past 3 years for Apple, Microsoft, and COMPASS Pathways and for legal cases regarding a death by suicide. He has stock options in Cerebral

Inc. He is an unpaid scientific advisor for Empatica, Koko, and TalkLife. JWS is a member of the Scientific Advisory Board of Sensorium Therapeutics (with equity) and has received grant support from Biogen, Inc. DD is the founder and CEO of Apoth. Other authors report no conflicts of interest.

Multimedia Appendix 1
Supplementary material containing further data on smartphone state log characteristics, estimated daily measures, screen-on time and bouts, the effect of daylight saving time, and additional details regarding our methodology.
[DOCX File, 626 KB - mhealth_v12i1e57439_app1.docx]

References

1. Onnela JP. Opportunities and challenges in the collection and analysis of digital phenotyping data. Neuropsychopharmacology 2021 Jan;46(1):45-54. [doi: 10.1038/s41386-020-0771-3] [Medline: 32679583]
2. Onnela JP, Rauch SL. Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. Neuropsychopharmacology 2016 Jun;41(7):1691-1696. [doi: 10.1038/npp.2016.7] [Medline: 26818126]
3. Nock MK, Borges G, Bromet EJ, Cha CB, Kessler RC, Lee S. Suicide and suicidal behavior. Epidemiol Rev 2008;30(1):133-154. [doi: 10.1093/epirev/mxn002] [Medline: 18653727]
4. About underlying cause of death, 2018-2022, single race. Centers for Disease Control and Prevention, National Center for Health Statistics. URL: http://wonder.cdc.gov/ucd-icd10-expanded.html [accessed 2024-02-06]
5. Kleiman EM, Nock MK. Real-time assessment of suicidal thoughts and behaviors. Curr Opin Psychol 2018 Aug;22:33-37. [doi: 10.1016/j.copsyc.2017.07.026] [Medline: 30122275]
6. Sedano-Capdevila A, Porras-Segovia A, Bello HJ, Baca-García E, Barrigon ML. Use of ecological momentary assessment to study suicidal thoughts and behavior: a systematic review. Curr Psychiatry Rep 2021 May 18;23(7):41. [doi: 10.1007/s11920-021-01255-7] [Medline: 34003405]
7. Kleiman EM, Turner BJ, Fedor S, Beale EE, Huffman JC, Nock MK. Examination of real-time fluctuations in suicidal ideation and its risk factors: results from two ecological momentary assessment studies. J Abnorm Psychol 2017 Aug;126(6):726-738. [doi: 10.1037/abn0000273] [Medline: 28481571]
8. Kleiman EM, Turner BJ, Fedor S, et al. Digital phenotyping of suicidal thoughts. Depress Anxiety 2018 Jul;35(7):601-608. [doi: 10.1002/da.22730] [Medline: 29637663]
9. Mou D, Kleiman EM, Fedor S, Beck S, Huffman JC, Nock MK. Negative affect is more strongly associated with suicidal thinking among suicidal patients with borderline personality disorder than those without. J Psychiatr Res 2018 Sep;104:198-201. [doi: 10.1016/j.jpsychires.2018.08.006] [Medline: 30103067]
10. Coppersmith DDL, Kleiman EM, Glenn CR, Millner AJ, Nock MK. The dynamics of social support among suicide attempters: a smartphone-based daily diary study. Behav Res Ther 2019 Sep;120:103348. [doi: 10.1016/j.brat.2018.11.016] [Medline: 30594300]
11. Oquendo MA, Galfalvy HC, Choo TH, et al. Highly variable suicidal ideation: a phenotypic marker for stress induced suicide risk. Mol Psychiatry 2021 Sep;26(9):5079-5086. [doi: 10.1038/s41380-020-0819-0] [Medline: 32576966]
12. Kleiman EM, Coppersmith DDL, Millner AJ, Franz PJ, Fox KR, Nock MK. Are suicidal thoughts reinforcing? A preliminary real-time monitoring study on the potential affect regulation function of suicidal thinking. J Affect Disord 2018 May;232:122-126. [doi: 10.1016/j.jad.2018.02.033] [Medline: 29481996]
13. Hallensleben N, Glaesmer H, Forkmann T, et al. Predicting suicidal ideation by interpersonal variables, hopelessness and depression in real-time. An ecological momentary assessment study in psychiatric inpatients with depression. Eur Psychiatry 2019 Feb;56(43–50):43-50. [doi: 10.1016/j.eurpsy.2018.11.003] [Medline: 30530103]
14. Nock MK, Kleiman EM, Abraham M, et al. Consensus statement on ethical & safety practices for conducting digital monitoring studies with people at risk of suicide and related behaviors. Psychiatr Res Clin Pract 2021;3(2):57-66. [doi: 10.1176/appi.prcp.20200029] [Medline: 34414359]
15. Porras-Segovia A, Molina-Madueño RM, Berrouiguet S, et al. Smartphone-based ecological momentary assessment (EMA) in psychiatric patients and student controls: a real-world feasibility study. J Affect Disord 2020 Sep 1;274:733-741. [doi: 10.1016/j.jad.2020.05.067] [Medline: 32664009]
16. Demirci K, Akgönül M, Akpinar A. Relationship of smartphone use severity with sleep quality, depression, and anxiety in university students. J Behav Addict 2015 Jun;4(2):85-92. [doi: 10.1556/2006.4.2015.010] [Medline: 26132913]
17. Hussain Z, Griffiths MD, Sheffield D. An investigation into problematic smartphone use: the role of narcissism, anxiety, and personality factors. J Behav Addict 2017 Sep 1;6(3):378-386. [doi: 10.1556/2006.6.2017.052] [Medline: 28849667]
18. Twenge JM, Hisler GC, Krizan Z. Associations between screen time and sleep duration are primarily driven by portable electronic devices: evidence from a population-based study of U.S. children ages 0-17. Sleep Med 2019 Apr;56:211-218. [doi: 10.1016/j.sleep.2018.11.009] [Medline: 30639033]
19. Rozgonjuk D, Levine JC, Hall BJ, Elhai JD. The association between problematic smartphone use, depression and anxiety symptom severity, and objectively measured smartphone use over one week. Comput Human Behav 2018 Oct;87:10-17. [doi: 10.1016/j.chb.2018.05.019]

20. Hicks JL, Althoff T, Sosic R, et al. Best practices for analyzing large-scale health data from wearables and smartphone apps. NPJ Digit Med 2019;2:45. [doi: 10.1038/s41746-019-0121-1] [Medline: 31304391]
21. Di J, Demanuele C, Kettermann A, et al. Considerations to address missing data when deriving clinical trial endpoints from digital health technologies. Contemp Clin Trials 2022 Feb;113:106661. [doi: 10.1016/j.cct.2021.106661] [Medline: 34954098]
22. Kiang MV, Chen JT, Krieger N, et al. Sociodemographic characteristics of missing data in digital phenotyping. Sci Rep 2021 Jul 29;11(1):15408. [doi: 10.1038/s41598-021-94516-7] [Medline: 34326370]
23. Onnela JP, Dixon C, Griffin K, et al. Beiwe: a data collection platform for high-throughput digital phenotyping. JOSS 2021;6(68):3417. [doi: 10.21105/joss.03417]
24. Onnela-lab/stb-beiwe-screen-time. GitHub Repository. URL: https://github.com/onnela-lab/stb-beiwe-screen-time [accessed 2024-02-06]
25. Kristensen PL, Olesen LG, Egebæk HK, Pedersen J, Rasmussen MG, Grøntved A. Criterion validity of a research-based application for tracking screen time on Android and iOS smartphones and tablets. Comput Hum Behav Rep 2022 Mar;5:100164. [doi: 10.1016/j.chbr.2021.100164]
26. Cui E, Leroux A, Smirnova E, Crainiceanu CM. Fast univariate inference for longitudinal functional models. J Comput Graph Stat 2022;31(1):219-230. [doi: 10.1080/10618600.2021.1950006] [Medline: 35712524]
27. Oudin A, Maatoug R, Bourla A, et al. Digital phenotyping: data-driven psychiatry to redefine mental health. J Med Internet Res 2023 Oct 4;25(1):e44502. [doi: 10.2196/44502] [Medline: 37792430]
28. Choi A, Ooi A, Lottridge D. Digital phenotyping for stress, anxiety, and mild depression: systematic literature review. JMIR Mhealth Uhealth 2024 May 23;12(1):e40689. [doi: 10.2196/40689] [Medline: 38780995]
29. Ettore E, Müller P, Hinze J, et al. Digital phenotyping for differential diagnosis of major depressive episode: narrative review. JMIR Ment Health 2023 Jan 23;10(1):e37225. [doi: 10.2196/37225] [Medline: 36689265]
30. Harris B, Regan T, Schueler J, Fields SA. Problematic mobile phone and smartphone use scales: a systematic review. Front Psychol 2020;11:672. [doi: 10.3389/fpsyg.2020.00672] [Medline: 32431636]
31. Vanden Abeele M, Beullens K, Roe K. Measuring mobile phone use: gender, age and real usage level in relation to the accuracy and validity of self-reported mobile phone use. Mob Med Comm 2013 May;1(2):213-236. [doi: 10.1177/2050157913477095]
32. Mireku MO, Mueller W, Fleming C, et al. Total recall in the SCAMP cohort: validation of self-reported mobile phone use in the smartphone era. Environ Res 2018 Feb;161:1-8. [doi: 10.1016/j.envres.2017.10.034] [Medline: 29096315]
33. Lee PH, Tse ACY, Wu CST, Mak YW, Lee U. Validation of self-reported smartphone usage against objectively-measured smartphone usage in Hong Kong Chinese adolescents and young adults. Psychiatr Investig 2021 Feb;18(2):95-100. [doi: 10.30773/pi.2020.0197] [Medline: 33517618]

Abbreviations

DST: daylight saving time
EDT: Eastern Daylight Time
EMA: ecological momentary assessment
FoS-GLMM: function-on-scalar generalized linear mixed-effects model
LMM: linear mixed-effects model
STB: suicidal thoughts and behaviors
UTC: Coordinated Universal Time

Edited by L Buis, M Sobolev; submitted 16.02.24; peer-reviewed by E Cui, X Lai; revised version received 18.08.24; accepted 21.08.24; published 11.10.24.

Please cite as:
Karas M, Huang D, Clement Z, Millner AJ, Kleiman EM, Bentley KH, Zuromski KL, Fortgang RG, DeMarco D, Haim A, Donovan A, Buonopane RJ, Bird SA, Smoller JW, Nock MK, Onnela JP
Smartphone Screen Time Characteristics in People With Suicidal Thoughts: Retrospective Observational Data Analysis Study
JMIR Mhealth Uhealth 2024;12:e57439
URL: https://mhealth.jmir.org/2024/1/e57439
doi:10.2196/57439

Mobile-Based Platform With a Low-Calorie Dietary Intervention Involving Prepackaged Food for Weight Loss for People With Overweight and Obesity in China: Half-Year Follow-Up Results of a Randomized Controlled Trial

Xi Wang[1*], MM; Suyuan Wang[2*], MMPH; Lingyu Zhong[3], MM; Chenghui Zhang[1], MD; Yanhong Guo[1], MM; Mingxia Li[1], MM; Li Zhao[4], DrPH; Shuming Ji[5], MMPH; Junjie Pan[6], MD; Yunhong Wu[1,2], MD

[1]Department of Endocrinology, Hospital of Chengdu Office of People's Government of Tibetan Autonomous Region, Chengdu, China

[2]Tibet Autonomous Region Clinical Research Center for High-Altitude Stress Endocrinology and Metabolism Disease, Hospital of Chengdu Office of People's Government of Tibetan Autonomous Region, Chengdu, China

[3]Department of Clinical Nutrition, Hospital of Chengdu Office of People's Government of Tibetan Autonomous Region, Chengdu, China

[4]Department of Health Policy and Management, West China School of Public Health and West China Fourth Hospital, Sichuan University, Chengdu, China

[5]Department of Clinical Research Management, West China Hospital of Sichuan University, Sichuan University, Chengdu, China

[6]Department of Cardivascular Medicine, Huashan Hospital Affilicated to Fudan University, Shanghai, China

[*]these authors contributed equally

Corresponding Author:
Yunhong Wu, MD
Department of Endocrinology, Hospital of Chengdu Office of People's Government of Tibetan Autonomous Region, , Chengdu, , China

Abstract

Background: Obesity is a rapidly increasing health problem in China, causing massive economic and health losses annually. Many techniques have emerged to help people with obesity better adhere to intervention programs and achieve their weight loss goals, including food replacement and internet-delivered weight loss consultations. Most studies on weight loss interventions mainly focused on the change in body weight or BMI; however, body fat, especially visceral fat mass, is considered the main pathogenic factor in obesity. In China, more reliable evidence is required on this topic. Moreover, it is unclear whether an integrated weight loss program combining food replacement products, mobile app-based platforms, and daily body composition monitoring using a wireless scale is useful and practical in China.

Objective: In this 2-arm, parallel-designed, randomized study, we explored the effectiveness and safety of the Metawell (Weijian Technologies Inc) weight loss program in China, which combines prepackaged biscuits, a wireless scale, and a mobile app.

Methods: Participants in the intervention group were guided to use food replacement products and a scale for weight loss and monitoring, whereas participants in the control group received printed material with a sample diet and face-to-face education on weight loss at enrollment. The intervention lasted for 3 months, and follow-up visits were conducted at months 3 and 6 after enrollment. Dual-energy x-ray absorptiometry and quantitative computed tomography were used to assess body fat. A multilevel model for repeated measurements was used to compare differences between the 2 groups.

Results: In total, 220 patients were randomly assigned to intervention (n=110) and control (n=110) groups. Participants in the intervention group had significantly greater decreases in BMI, total body fat, visceral adipose area, and subcutaneous adipose area (all P<.001) than those in the control group. However, the rate of change in lean mass was not significantly different between the 2 groups (P=.62). Further, 35 participants in the intervention group reported adverse events. Constipation was the most frequently reported adverse event (11/110), followed by dizziness (6/110), hypoglycemia (4/110), fatigue (3/110), and gastritis (3/35).

Conclusions: The Metawell program was effective for weight loss. After the intervention, participants in the intervention group lost more body weight and body fat while retaining muscle mass than those in the control group.

Trial Registration: Chinese Clinical Trial Registry ChiCTR1900021630; https://www.chictr.org.cn/showproj.html?proj=36183

International Registered Report Identifier (IRRID): RR2-10.1136/bmjopen-2020-048106

(JMIR Mhealth Uhealth 2024;12:e47104) doi:10.2196/47104

KEYWORDS

weight loss; obesity; body fat; food replacement product; meal replacement; weight; obese; RCT; randomized; mHealth; mobile health; mobile app; mobile application; mobile phone

Introduction

In China, the prevalence of obesity has increased rapidly from 3.1% in 2004 to 8.1% in 2018 [1]. Obesity is closely related to a series of diseases such as cardiovascular diseases, diabetes, and hypertension and leads to physical impairments and a decrease in the quality of life [2]. The economic losses caused by overweight and obesity in China are estimated to be US $12.97 billion [3]. Moreover, an elevated BMI was estimated to be the seventh highest risk factor for death in China, equating to nearly 0.7 million deaths in 2017 [4].

Nutrition therapy is considered to be the foundation of obesity management [5]. Meal replacement products provide an easy way for participants to follow a weight loss program, and rapid weight loss during the initial stage of the intervention can help to build the confidence of participants, thus leading to greater weight loss. The Doctor Referral of Overweight People to Low Energy Total Diet Replacement Treatment (DROPLET) study revealed that total meal replacement programs provided by general practitioners are effective for weight loss [6], and a meta-analysis showed that partial or total meal replacements were associated with greater weight loss [7,8].

The delivery of lifestyle intervention programs may also have an impact on their effectiveness in weight loss. Whether the use of the internet or a mobile app platform to deliver the intervention will lead to greater weight loss is controversial. A meta-analysis showed that the use of mobile apps can improve the effect of weight loss interventions [9], whereas other studies did not find a statistically significant effect for mobile apps [10].

Most studies of weight loss interventions focused on the change in body weight or BMI; however, body fat, especially visceral fat mass, is considered the main pathogenic factor in obesity [11]. We found 1 randomized study focused on the change in body composition during a weight loss intervention in China [12]. However, the sample size was limited, and the body composition results are not sufficiently accurate considering their measuring method; therefore, we still need more reliable evidence on this topic. Furthermore, it is unknown whether an integrated weight loss program combining food replacement products, a mobile app-based platform, and daily monitoring of body composition using a wireless scale would be useful and practical in China. Thus, we performed this study to evaluate the safety and efficacy of the Metawell program, a weight loss program using prepackaged biscuits, a Bluetooth body composition scale, and a mobile app for uploading body composition data and delivering meal guidance from health care professionals daily for weight loss and changes in body composition and body fat distribution.

Methods

Study Design

This was a single-center, open-label, 2-arm, parallel-designed, randomized controlled trial; all participants were recruited from Chengdu City via social media platforms and face-to-face visits, and assessments will take place at enrollment and at 3 and 6 months after enrollment. The trial started on April 11, 2019. Participants were randomly assigned to the intervention and control groups at a ratio of 1:1. This trial was registered in the Chinese Clinical Trial Registry (ChiCTR1900021630). After our trial started, several participants older than 65 years showed a strong willingness to participate; after discussion, we decided to raise the upper limit of the participant aged from 65 to 75 years.

Ethical Considerations

This trial was approved by the ethics committee of the Hospital of Chengdu Office of People's Government of the Tibetan Autonomous Region (2019-No.01). All participants have signed the informed consent; all data are anonymized. All participants will be informed that they will receive ¥200 (a currency exchange rate of ¥6.92=US $1 is applicable) compensation for each follow-up visit they attend.

Eligibility Criteria

The inclusion criteria were those aged between 18 and 75 years with a BMI ≥25 and <40 kg/m^2 and one or more of the following: a history of hypertension or either systolic blood pressure >120 mm Hg or diastolic blood pressure >80 mm Hg, abdominal circumference >96 cm (90 cm for women), fasting triglyceride level >1.69 mmol/L, a history of type 2 diabetes mellitus managed with lifestyle (not on insulin or oral medications) or fasting blood glucose level >5.6 mmol/L, and high-density lipoprotein cholesterol level <1.04 mmol/L (1.3 mmol/L for women).

Participants with one or more of the following conditions were excluded: a history of coronary artery disease; diabetes mellitus managed with insulin or any oral hypoglycemic pill for diabetes; glucose intolerance or fasting glucose level ≥8 mmol/L; congestive heart failure; familial hypercholesterolemia, including familial hypertriglyceridemia, fasting low-density lipoprotein cholesterol level >4.2 mmol/L, fasting triglyceride level >6.8 mmol/L, and current use of lipid-lowering agents; a past medical history of hypothyroidism, Cushing syndrome, eating disorders, gout in the past 6 months, confirmed episodes of hypoglycemia, pregnancy, advanced liver disease, renal insufficiency, or any other major chronic medical condition; and smokers who planned to quit smoking in the following 12 months. Participants with hypertension were included only if they were taking fewer than 3 antihypertensive medications, did not have changes in the dose of their blood pressure medications in the last month, and had systolic blood pressure

<160 mm Hg and diastolic blood pressure <100 mm Hg. Finally, people who could not use smartphones were excluded.

Randomization, Masking, and Blinding

The list of random numbers was generated using R with a certain random seed, the results of which were sequentially sealed in opaque envelopes and were unsealed while participants signed the consent and finished eligibility screening. This study was an open-label study; patients and researchers were not blinded.

Interventions

Details of the interventions had been described in our previously published protocol [13]. In general, participants in the intervention group would download a mobile app and be guided to use food replacement products and a scale for weight loss and monitoring. Their diet, body composition, and urine ketone will be monitored every day, and trained instructors will guide them remotely. During the weight loss stage (0 - 3 months), participants will be guided to take food replacement biscuits along with a selection of healthy recipes (such as seaweed soup, skimmed milk, spiced beef, grains, vegetables) and advised to take multivitamin tablets daily. Detailed nutritional information on diet replacement biscuits is listed in Table S1 in Multimedia Appendix 1, and the overall intake of energy for the intervention group was restricted to 800 - 1200 kcal/day; the instructor will guide participants to adjust their recipe based on their health condition and the speed of weight loss. During the weight maintenance stage (3 - 6 months), participants will stop using the biscuit, and they will be told to monitor their weight and upload data. When noticeable weight regain is detected, the practitioner will initiate a 2 - to 3-day weight loss intervention using the same protocol as that used in the weight loss stage to maintain the participant's body weight. Based on obesity management guidelines, a calorie intake of 800 - 1200 is generally safe. Furthermore, on analyzing the data of over 250,000 Chinese using this program for weight loss, we found no significant safety concerns.

Meanwhile, participants in the control group would receive printed material with a sample diet, as well as face-to-face education on weight loss at enrollment. The printed sample diet used for the control group included a daily energy intake of 1500 kcal for men and 1200 kcal for women. Doctors and the dietician in our research group (XW, L Zhong, and YW) were responsible for monitoring participants' safety.

Outcomes

The primary outcome was weight loss as a result of the intervention, defined as the reduction in BMI after 6 months. The following two secondary outcomes were also analyzed: (1) the proportion of participants achieving a body weight reduction of greater or equal to 10% of baseline body weight and (2) the change in body fat mass and lean mass, as well as changes in the distribution of adipose tissue.

Assessment and Follow-Up

Participants were followed up at 3 and 6 months after enrollment. During each follow-up assessment, body weight was measured using an ultrasonic height-weight scale (DHM-200, Dinghengkeji), and physical activity was assessed using the International Physical Activity Questionnaire long form. Dual-energy x-ray absorptiometry (DXA) was used to assess the body composition of the participants using a GE Lunar Prodigy scanner (GE Healthcare). A 16-slice computed tomography (CT) scanner (Aquilion 16; TOSHIBA) was used to perform abdominal scanning, and the distribution of adipose tissue was analyzed with QCTPRO software (Mindways Software) using a slice between L2 and L3. DXA and CT were performed only at baseline and at the 6-month follow-up.

Sample Size

According to the DROPLET study [6], we hypothesized that the average absolute difference in weight loss between the trial and control groups would be at least 4 kg in weight, with an SD of 9 kg. Therefore, 100 participants per group will provide more than 90% power to detect this difference, with an α error of .05. The sample size would be 110 participants per group considering an estimated 10% attrition rate.

Statistical Analysis

The intention-to-treat dataset was used to analyze the primary and secondary outcomes. The normality of data was tested using Kolmogorov-Smirnov test. All of our continuous data had nonnormality according to test results, so they are presented as median (IQR). Categorical data are presented as count (%). Baseline information was compared between the 2 groups, Mann-Whitney *U* test and chi-square test were used for continuous and categorical variables, respectively. A multilevel model for repeated measurements was used to test the interaction of the intervention and time against BMI as the primary outcome and fat mass, lean mass, visceral adipose area, and subcutaneous adipose area as the secondary outcomes. Variants of different individuals were treated as the first level and different follow-up times were treated as the second level. After reviewing previous studies such as he DROPLET study and incorporating with clinical experience of clinicians in our group, baseline body weight, physical activity, sex, and age were adjusted. The rate of missing data was relatively low in this study; therefore, missing data were not considered in the analysis.

We performed a post hoc subgroup analysis to determine whether there were any differences among participants of different age and sex groups. Baseline characteristics were compared using SPSS (version 24.0; IBM Corp), and multilevel model analysis was performed using MLwiN (version 2.30; Centre for Multilevel Modelling, University of Bristol). *P* values <.05 were considered statistically significant.

Results

Overview

Participants were recruited between April 2019 and January 2020. A total of 329 participants were registered for screening and 220 were randomly assigned to the intervention and control groups (n=110 each; Figure 1). The median age of the participants was 33 (IQR 28-41) years and 57.8% (n=125) were women. The median BMI was 27.9 (IQR 4) kg/m^2. There were no differences in any of the baseline characteristics between the 2 groups (Table 1). At the 6-month follow-up, we followed 103

participants in the intervention group and 94 in the control group (Figure 1).

Figure 1. Flowchart of this study.

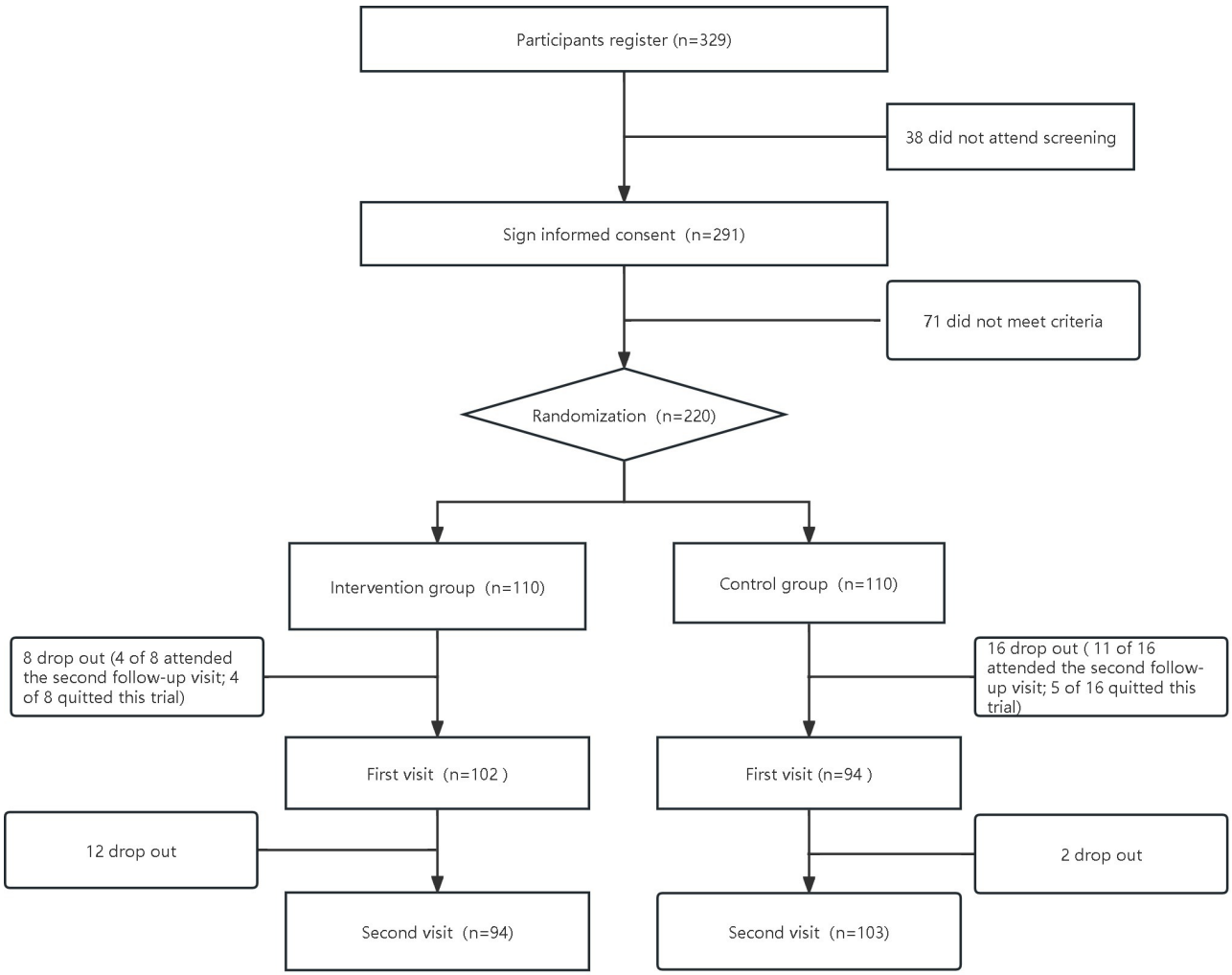


Table . Baseline information of the participants.

	Overall (n=220)	Intervention group (n=110)	Control group (n=110)	*P* value
Sex, n (%)				.63
Male	95 (42.22)	48 (44.86)	47 (41.59)	
Female	125 (57.78)	59 (55.14)	66 (58.41)	
Age (years), median (IQR)	33 (28-41)	32 (28- 39)	33 (28-41)	.48
Height (cm), median (IQR)	164.5 (158.0-171.0)	165.0 (159.0-170.5)	164 (158.0-171.0)	.85
Weight (kg), median (IQR)	76.4 (68.9-86.1)	77.4 (70.1-85.3)	75.2 (67.8-86.2)	.49
BMI (kg/m^2), median (IQR)	27.9 (26.2-29.9)	28.2 (26.3-30.9)	27.7 (26.0-29.7)	.39
Waist circumference (cm), median (IQR)	97.4 (92.4-102.8)	97.7 (93.0-102.5)	97.1 (92.1-102.9)	.57
Hip circumference (cm), median (IQR)	105.4 (101.3-110.0)	105.5 (100.2-109.8)	106.4 (102.1-110.0)	.18
Subcutaneous adipose (cm^2), median (IQR)	201.5(158.7-271.1)	208.8 (167.5-275.7)	196.9 (155.1-267.8)	.33
Visceral adipose (cm^2), median (IQR)	167.9 (123.2-230.7)	171.6 (127.4-215.4)	163.3 (120.5-237.1)	.54
Total fat mass (kg), median (IQR)	29.4 (25.1-34.9)	29.8 (25.6-34.7)	29.35 (24.6-35.6)	.96
Total lean mass (kg), median (IQR)	46.1 (39.1-55.0)	47.5 (39.1-54.9)	43.83 (39.1-55.1)	.67
Physical activity (METs[a]/week), median (IQR)	1062.5 (258.0-2715.0)	1284 (346.5-2806.0)	928.5 (102.5-2331.0)	.33

[a]MET: metabolic equivalent of task.

Primary Outcome

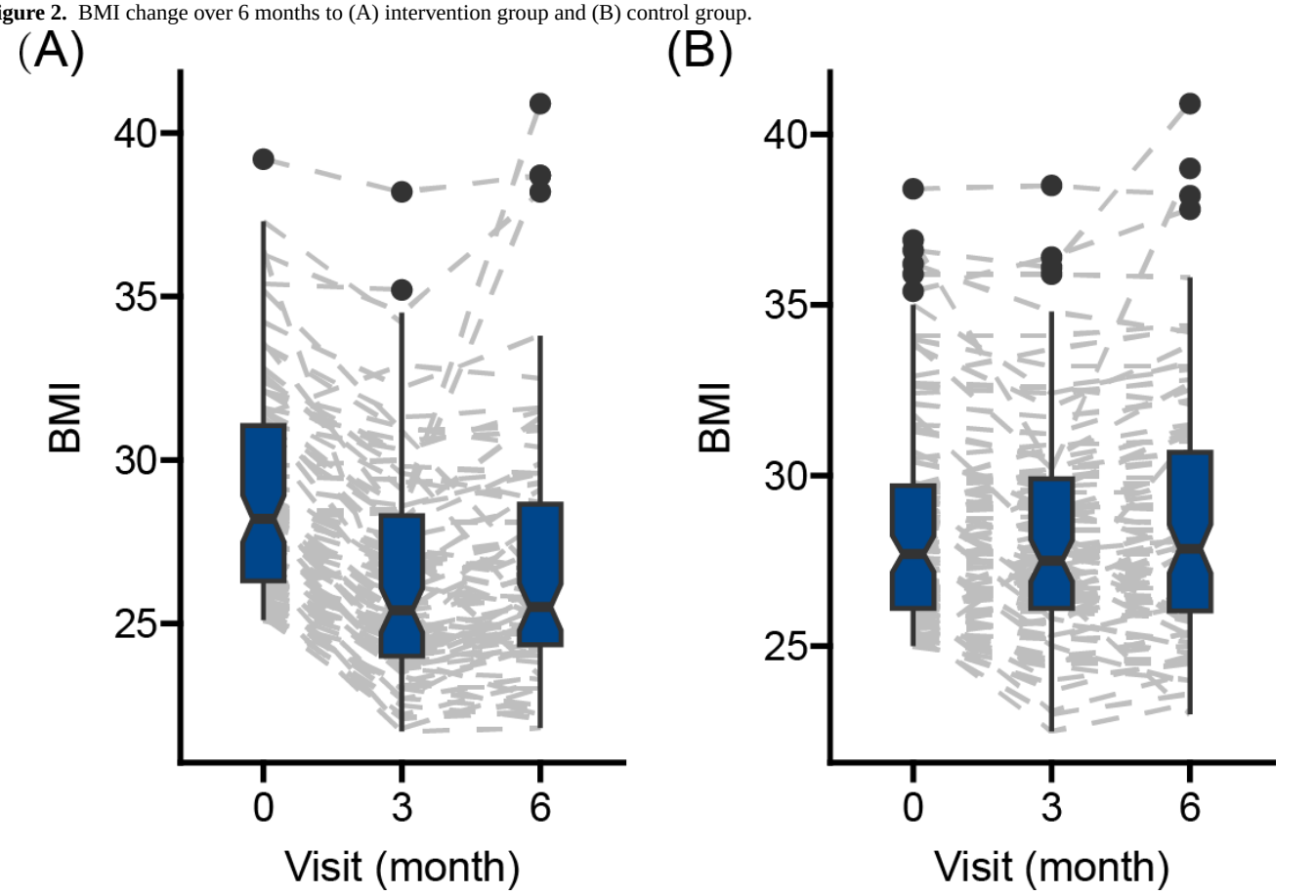
At 6 months, 31 (28%) and 3 (3%) participants in the intervention and control groups, respectively, achieved a weight loss of at least 10% compared with their baseline weight. The median change in BMI was –1.90 (IQR –0.8 to –3.2) kg/m^2 for the intervention group and –0.03 (IQR –0.9 to –0.4) kg/m^2 for the control group (Table 2).

Table . Primary and secondary outcomes by group.

Outcomes	Intervention group	Control group
3 months		
BMI change (kg/m^2), median (IQR)	–2.3 (–3.2 to –1.3)	–0.2 (–0.9 to 0.4)
At least 10% weight lost, n (%)	37 (33.63)	5 (4.54)
6 months		
BMI change (kg/m^2), median (IQR)	–1.9 (–0.8 to –3.2)	–0.03 (–0.9 to 0.4)
At least 10% weight lost, n (%)	31 (28.18)	3 (2.72)
Body fat mass change (kg), median (IQR)	–3.9 (–7.3 to –1.3)	0.1 (–1.9 to 2.6)
Body lean mass change (kg), median (IQR)	–0.8 (–2.1 to 0.2)	–0.4 (–1.7 to 0.6)
Subcutaneous adipose change (cm^2), median (IQR)	–37.6 (–71.4 to –10.4)	–11.2 (–34.9 to 18.8)
Visceral adipose change (cm^2), median (IQR)	–15.1 (-44.3 to 2.4)	7.8 (–12.6 to 37.9)

The change in BMI was different between the 2 groups before ($\beta_{group\times time}$=–1.214, *P*<.001) and after ($\beta_{group\times time}$=–1.238, *P*<.001) adjustment for sex, age, baseline body weight, and baseline physical activity level (Figure 2; Table S2 in Multimedia Appendix 2).

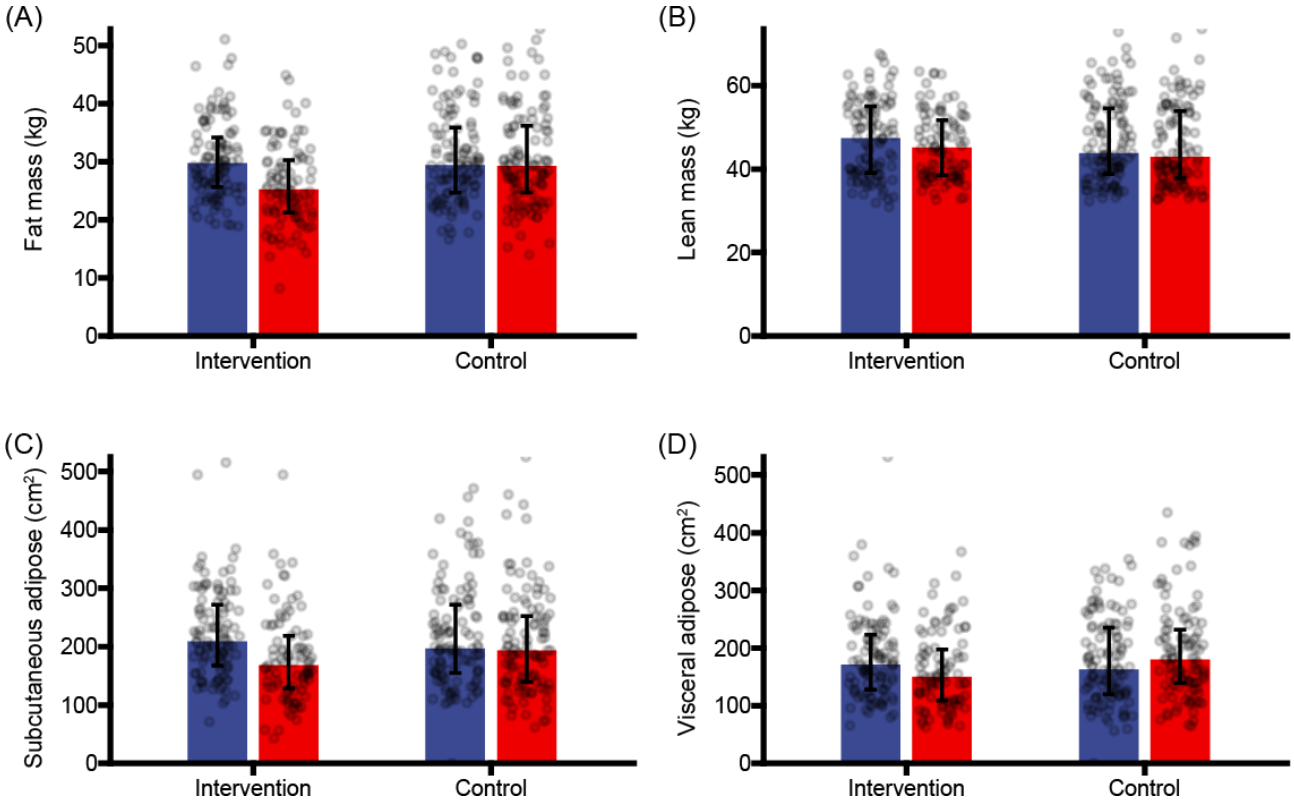
Figure 2. BMI change over 6 months to (A) intervention group and (B) control group.



Secondary Outcomes

After the intervention, the total body fat mass decreased by 5.07 kg and the total body lean mass decreased by 3.13 kg in the intervention group. In the control group, body fat mass increased by 0.06 kg, and body lean mass decreased by 0.27 kg (Figure 3). The change rate of body fat mass between the 2 groups was statistically significant ($\beta_{group\times time}$=–4.127, P<.001), whereas that of total muscle mass was not significant ($\beta_{group\times time}$=–0.177, P=.62; Table S2 in the Multimedia Appendix 2).

Figure 3. Change in body composition before (blue) and after (red) the intervention in different groups: (A) total body fat mass, (B) total lean mass, (C) subcutaneous adipose area, and (D) visceral adipose area.



Subcutaneous and visceral adipose areas decreased by 53 and 31 cm^2, respectively, in the intervention group (Figure 3). In the control group, subcutaneous adipose area decreased by 1 cm^2 and visceral adipose area increased by 19 cm^2. The rates of change in subcutaneous adipose area ($\beta_{group\times time}$=–40.898, P<.001) and visceral adipose area ($\beta_{group\times time}$=–40.409, P<.001) were different between the 2 groups (Table S3 in Multimedia Appendix 3).

Subgroup Analysis

Subgroup analyses were performed to determine whether the intervention was effective in the different sex and age groups. The decrease rate in BMI was greater in the intervention group in both male and female participants ($\beta_{group\times time}$=–1.781 for male sex and $\beta_{group\times time}$=–0.833 for female sex; Table S4 in Multimedia Appendix 4). Furthermore, for participants in all age groups, the rate of decrease in BMI was greater in the intervention group ($\beta_{group\times time}$=–0.754 for age ≤30 years, $\beta_{group\times time}$=–0.833 for age 31 - 40 years, and $\beta_{group\times time}$=–2.055 for age ≥41 years; Table S4 in Multimedia Appendix 4).

Adverse Events

In the control group, only 2 participants reported 4 adverse events (AEs) during the follow-up visits, whereas 35 participants reported AEs in the intervention group (Table 3). The most common AE in the intervention group was constipation (11/110); 10% of the participants in the intervention group reported varying degrees of constipation during the follow-up visits. Other common AEs were dizziness (6/35), hypoglycemia (4/110), fatigue (3/110), and gastritis (3/110). All AEs were classified as mild to moderate, none of these participants were hospitalized because of AEs.

Table . Numbers of participants reporting adverse events.

	Intervention group, n	Control group, n	Total, n
Constipation	11	1	12
Hypoglycemia	4	0	4
Diarrhea	1	1	2
Fatigue	3	1	4
Dizziness	6	1	7
Sour regurgitation	1	0	1
Hair loss	1	0	1
Flustered	1	0	1
Rhinitis	2	0	2
Gastritis	3	0	3
Menstrual disorder	1	0	1
Elevated urinary ketones	1	0	1

Discussion

Principal Results

This investigation demonstrated that the implementation of a weight loss program combining food replacement products, remote guidance by practitioners, and daily monitoring of body composition is effective for weight loss and reducing fat mass, especially visceral fat, without significant muscle loss in people with overweight and obesity in China.

Comparison With Prior Work

According to previous investigations, more than half of patients with obesity were not motivated to adhere to weight loss interventions; the most frequently reported barrier was a lack of accessibility to weight loss diets and a lack of support and external control [14]. Furthermore, the need for intensive follow-up with motivational interviews was highlighted. In our program, a series of methods were used to address these barriers, including using food replacement products, intensive guidance delivered via a mobile app, and a wireless scale provided to participants for their daily monitoring of body composition. Food replacement products can be a useful tool in weight loss programs, providing an easy way for participants to adhere to the diet intervention [14]. Mobile apps enable frequent interactions between health care professionals and patients and allow intensive motivational interviews. Additionally, more frequent body composition monitoring is closely associated with better adherence. Several studies have revealed that the use of food replacement products, mobile apps, and wireless scales can lead to greater weight loss [7,15-17]. Our study confirms that the combination of these techniques can lead to greater weight loss than traditional educational materials.

Although BMI is a good marker for obesity screening, the threshold varies among different ethnic and age groups. Moreover, considering that most comorbidities of obesity are mediated by excess fat, endocrine and immune responses of adipose tissue are distributed throughout different organs [18,19]. Hence, BMI is not suitable to reflect the complexity of obesity. In recent years, some studies on weight loss interventions have investigated the change in body composition and adipose tissue distribution; however, most used bioelectrical impedance analysis for the measurements, which is less accurate than DXA or CT [20]. In our study, we used DXA and quantitative CT for a more accurate assessment of body composition changes, and we confirmed that the intervention program could reduce the amount of total fat, visceral fat, and subcutaneous fat compared with the control program. Therefore, we can be more confident that patients with obesity can achieve health benefits through the loss of body fat by adhering to this intervention program.

We also found that the change rate in lean mass was not statistically significant (P=.62) between the 2 groups. In contrast, previous studies suggested that severe calorie-restricting interventions would lead to a greater loss of lean mass [21]. The use of a food replacement product might be a reason why the participants in the intervention group did not experience a significant loss of lean mass, despite having a restricted intake of calories. A systematic review by Coleman et al [22] indicated that food replacement products can minimize the loss of lean mass by reducing the variety of foods. Furthermore, our intervention provided a relatively high protein intake level while restricting calorie intake. Increasing the daily intake of protein can prevent muscle loss while decreasing the amount of body fat [23].

The safety of food replacement products is a concern for health care professionals, and their use is not recommended by most guidelines because of the lack of data on safety and effectiveness. Our study demonstrated that, although the risk was slightly higher in the intervention group, the side effects of this product were mild to moderate. The most common AE was constipation, which is consistent with a previous study on total food replacement products and low-carbohydrate diets [6]. To assess the safety of our intervention program, we also analyzed the change in spinal bone mineral density (BMD), we found the BMD decreased by about 2% (from mean 1.11, SD 0.10 g/cm^2 to mean 1.09, SD 0.09 g/cm^2, P<.01) in the

intervention group (data not shown). However, considering that the least significant change of our DXA measurement was 5.3%, which is much higher than the decreasing rate we found in our study, we believe that this change was not clinically significant. Therefore, food replacement programs are relatively safe for weight loss in healthy individuals.

The follow-up of some participants was delayed owing to the COVID-19 pandemic and the lockdown in Chengdu City; however, considering that our participants did not have access to our weight loss intervention during the lockdown and according to relevant studies [24,25], we expected their body weight to increase, resulting in the underestimation of the effect size; therefore, we did not treat this situation.

Limitations and Strengths

To the best of our knowledge, this is the first study using DXA and quantitative CT to accurately and systematically evaluate change in body fat, visceral adipose tissue, and subcutaneous adipose tissue during weight loss intervention, which is more relative to participants' health than BMI [26,27]. This is the main highlight of our study.

The main limitation of our study is that the follow-up period was relatively short to observe the long-term effects of this intervention. Meanwhile, we found a statistically significant (although without clinical significance) decrease in BMD in the intervention group, it remains unclear whether there would be any lasting impact on BMD after the intervention ended. Therefore, an extended follow-up will be performed to investigate whether participants will regain their body weight and whether their BMD will continue to decrease. Furthermore, we only explored the effectiveness of this weight loss program without performing a cost-effectiveness analysis. We will collect and analyze related data in the following study. Moreover, this study's results should be treated as exploratory outcomes, and further studies are required to verify them.

Conclusions

The Metawell program was effective for weight loss. After the intervention, participants in the intervention group lost more body weight and body fat while retaining muscle mass than those in the control group.

Acknowledgments

We would like to present our special thanks to Dr Francisco Lopez-Jimenez, Dr Lilach O Lerman, and Dr Amir Lerman for their help in the designation of our study, and Miss Li Qiuyue for helping us in recruiting participants. This study was funded by Weijian Technologies Inc and the National Scientific Foundation of China (82273748).

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

Conceptualization, resources, supervision, methodology, project administration, writing of the original draft, and review and editing were done by XW. Methodology, writing of the original draft, investigation, data curation, and visualization was handled by SW. Methodology, supervision, review and editing, and investigation was by L Zhong. Methodology, supervision, review and editing, and investigation was performed by CZ. Supervision, review and editing, and investigation was conducted by YG. Supervision, review and editing, and project administration was done by ML. Formal analysis was done by L Zhao and SJ. Review and editing was handled by JP. Conceptualization, resources, supervision, methodology, project administration, review and editing, and resources were performed by YW.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Nutritional information of Yufit biscuit.
[DOCX File, 19 KB - mhealth_v12i1e47104_app1.docx]

Multimedia Appendix 2
The effect of the intervention on the change in BMI.
[DOCX File, 17 KB - mhealth_v12i1e47104_app2.docx]

Multimedia Appendix 3
The effect of the intervention on body composition changes.
[DOCX File, 17 KB - mhealth_v12i1e47104_app3.docx]

Multimedia Appendix 4

Results of the subgroup analysis.
[DOCX File, 18 KB - mhealth_v12i1e47104_app4.docx]

Checklist 1
CONSORT (Consolidated Standards for Reporting Trials) checklist.
[XLSX File, 13 KB - mhealth_v12i1e47104_app5.xlsx]

References

1. Wang L, Zhou B, Zhao Z, et al. Body-mass index and obesity in urban and rural China: findings from consecutive nationally representative surveys during 2004-18. Lancet 2021 Jul 3;398(10294):53-63. [doi: 10.1016/S0140-6736(21)00798-4] [Medline: 34217401]
2. Guh DP, Zhang W, Bansback N, Amarsi Z, Birmingham CL, Anis AH. The incidence of co-morbidities related to obesity and overweight: a systematic review and meta-analysis. BMC Public Health 2009 Mar 25;9:88. [doi: 10.1186/1471-2458-9-88] [Medline: 19320986]
3. Zhang J, Shi XM, Liang XF. Economic costs of both overweight and obesity among Chinese urban and rural residents, in 2010. Zhonghua Liu Xing Bing Xue Za Zhi 2013 Jun;34(6):598-600. [Medline: 24125612]
4. Zhou M, Wang H, Zeng X, et al. Mortality, morbidity, and risk factors in China and its provinces, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2019 Sep 28;394(10204):1145-1158. [doi: 10.1016/S0140-6736(19)30427-1] [Medline: 31248666]
5. Wharton S, Lau DCW, Vallis M, et al. Obesity in adults: a clinical practice guideline. CMAJ 2020 Aug 4;192(31):E875-E891. [doi: 10.1503/cmaj.191707] [Medline: 32753461]
6. Astbury NM, Aveyard P, Nickless A, et al. Doctor referral of overweight people to low energy total diet replacement treatment (DROPLET): pragmatic randomised controlled trial. BMJ 2018 Sep 26;362:k3760. [doi: 10.1136/bmj.k3760] [Medline: 30257983]
7. Astbury NM, Piernas C, Hartmann-Boyce J, Lapworth S, Aveyard P, Jebb SA. A systematic review and meta-analysis of the effectiveness of meal replacements for weight loss. Obes Rev 2019 Apr;20(4):569-587. [doi: 10.1111/obr.12816] [Medline: 30675990]
8. Hartmann-Boyce J, Ordóñez-Mena JM, Theodoulou A, et al. Impact of program characteristics on weight loss in adult behavioral weight management interventions: systematic review and component network meta-analysis. Obesity (Silver Spring) 2022 Sep;30(9):1778-1786. [doi: 10.1002/oby.23505] [Medline: 35918886]
9. Dao KP, De Cocker K, Tong HL, Kocaballi AB, Chow C, Laranjo L. Smartphone-delivered ecological momentary interventions based on ecological momentary assessments to promote health behaviors: systematic review and adapted checklist for reporting ecological momentary assessment and intervention studies. JMIR mHealth uHealth 2021 Nov 19;9(11):e22890. [doi: 10.2196/22890] [Medline: 34806995]
10. Brindal E, Hendrie GA, Freyne J, Noakes M. Incorporating a static versus supportive mobile phone app into a partial meal replacement program with face-to-face support: randomized controlled trial. JMIR mHealth uHealth 2018 Apr 18;6(4):e41. [doi: 10.2196/mhealth.7796] [Medline: 29669704]
11. Yuan X, Wang J, Yang S, et al. Effect of intermittent fasting diet on glucose and lipid metabolism and insulin resistance in patients with impaired glucose and lipid metabolism: a systematic review and meta-analysis. Int J Endocrinol 2022 Mar 24;2022:6999907. [doi: 10.1155/2022/6999907] [Medline: 35371260]
12. Cai J, Shao L, Zhao S, Liu W, Liu P. The effects of three weight management methods on body composition and serum lipids of overweight and obese people. Front Nutr 2022 Dec 8;9:1073576. [doi: 10.3389/fnut.2022.1073576] [Medline: 36570146]
13. Wang X, Wang S, Zhang C, et al. Internet-based platform for a low-calorie dietary intervention involving prepackaged food for weight loss in overweight and obese individuals in China: protocol for a randomised controlled trial. BMJ Open 2022 Jan 21;12(1):e048106. [doi: 10.1136/bmjopen-2020-048106] [Medline: 35063952]
14. Trujillo-Garrido N, Santi-Cano MJ. Motivation and limiting factors for adherence to weight loss interventions among patients with obesity in primary care. Nutrients 2022 Jul 17;14(14):2928. [doi: 10.3390/nu14142928] [Medline: 35889885]
15. Min J, Kim SY, Shin IS, Park YB, Lim YW. The effect of meal replacement on weight loss according to calorie-restriction type and proportion of energy intake: a systematic review and meta-analysis of randomized controlled trials. J Acad Nutr Diet 2021 Aug;121(8):1551-1564. [doi: 10.1016/j.jand.2021.05.001] [Medline: 34144920]
16. Zhou M, Zhang N, Zhang Y, et al. Effect of mobile-based lifestyle intervention on weight loss among the overweight and obese elderly population in China: a randomized controlled trial. Int J Environ Res Public Health 2021 Aug 21;18(16):8825. [doi: 10.3390/ijerph18168825] [Medline: 34444573]
17. Chew HSJ, Koh WL, Ng JSHY, Tan KK. Sustainability of weight loss through smartphone apps: systematic review and meta-analysis on anthropometric, metabolic, and dietary outcomes. J Med Internet Res 2022 Sep 21;24(9):e40141. [doi: 10.2196/40141] [Medline: 36129739]

18. Frühbeck G, Busetto L, Dicker D, et al. The ABCD of obesity: an EASO position statement on a diagnostic term with clinical and scientific implications. Obes Facts 2019;12(2):131-136. [doi: 10.1159/000497124] [Medline: 30844811]
19. Mechanick JI, Hurley DL, Garvey WT. Adiposity-based chronic disease as a new diagnostic term: the American Association of Clinical Endocrinologists and American College of Endocrinology position statement. Endocr Pract 2017 Mar;23(3):372-378. [doi: 10.4158/EP161688.PS] [Medline: 27967229]
20. Faria SL, Faria OP, Cardeal MDA, Ito MK. Validation study of multi-frequency bioelectrical impedance with dual-energy X-ray absorptiometry among obese patients. Obes Surg 2014 Sep;24(9):1476-1480. [doi: 10.1007/s11695-014-1190-5] [Medline: 24464546]
21. Seimon RV, Wild-Taylor AL, Keating SE, et al. Effect of weight loss via severe vs moderate energy restriction on lean mass and body composition among postmenopausal women with obesity: the TEMPO diet randomized clinical trial. JAMA Netw Open 2019 Oct 2;2(10):e1913733. [doi: 10.1001/jamanetworkopen.2019.13733] [Medline: 31664441]
22. Coleman CD, Kiel JR, Mitola AH, Langford JS, Davis KN, Arterburn LM. Effectiveness of a Medifast meal replacement program on weight, body composition and cardiometabolic risk factors in overweight and obese adults: a multicenter systematic retrospective chart review study. Nutr J 2015 Aug 6;14:77. [doi: 10.1186/s12937-015-0062-8] [Medline: 26245279]
23. Guo X, Xu Y, He H, et al. Effects of a meal replacement on body composition and metabolic parameters among subjects with overweight or obesity. J Obes 2018 Dec 26;2018:2837367. [doi: 10.1155/2018/2837367] [Medline: 30687550]
24. Heinberg LJ, Steffen K. Social isolation and loneliness during the COVID-19 pandemic: impact on weight. Curr Obes Rep 2021 Sep;10(3):365-370. [doi: 10.1007/s13679-021-00447-9] [Medline: 34297344]
25. Tan ST, Tan CX, Tan SS. Changes in dietary intake patterns and weight status during the COVID-19 lockdown: a cross-sectional study focusing on young adults in Malaysia. Nutrients 2022 Jan 10;14(2):280. [doi: 10.3390/nu14020280] [Medline: 35057460]
26. Nuttall FQ. Body mass index: obesity, BMI, and health: a critical review. Nutr Today 2015 May;50(3):117-128. [doi: 10.1097/NT.0000000000000092] [Medline: 27340299]
27. Piché ME, Poirier P, Lemieux I, Després JP. Overview of epidemiology and contribution of obesity and body fat distribution to cardiovascular disease: an update. Prog Cardiovasc Dis 2018;61(2):103-113. [doi: 10.1016/j.pcad.2018.06.004] [Medline: 29964067]

Abbreviations

AE: adverse event
BMD: bone mineral density
CT: computed tomography
DROPLET: Doctor Referral of Overweight People to Low Energy Total Diet Replacement Treatment
DXA: dual-energy x-ray absorptiometry

Edited by L Buis; submitted 16.03.23; peer-reviewed by A Button; revised version received 12.05.24; accepted 03.06.24; published 28.10.24.

Please cite as:
Wang X, Wang S, Zhong L, Zhang C, Guo Y, Li M, Zhao L, Ji S, Pan J, Wu Y
Mobile-Based Platform With a Low-Calorie Dietary Intervention Involving Prepackaged Food for Weight Loss for People With Overweight and Obesity in China: Half-Year Follow-Up Results of a Randomized Controlled Trial
JMIR Mhealth Uhealth 2024;12:e47104
URL: https://mhealth.jmir.org/2024/1/e47104
doi:10.2196/47104

An Evaluation of the Effect of App-Based Exercise Prescription Using Reinforcement Learning on Satisfaction and Exercise Intensity: Randomized Crossover Trial

Cailbhe Doherty[1,2], BSc, PhD; Rory Lambe[1,2], BSc; Ben O'Grady[1,2], BSc; Diarmuid O'Reilly-Morgan[2], BSc; Barry Smyth[2], BSc, PhD; Aonghus Lawlor[2], BSc, PhD; Neil Hurley[2], BSc, PhD; Elias Tragos[2], BSc, PhD

[1]School of Public Health, Physiotherapy and Sports Science, University College Dublin, Dublin, Ireland

[2]Insight SFI Research Centre for Data Analytics, O'Brien Centre for Science, University College Dublin, Dublin, Ireland

Corresponding Author:
Cailbhe Doherty, BSc, PhD
School of Public Health, Physiotherapy and Sports Science, University College Dublin, , Dublin, , Ireland

Abstract

Background: The increasing prevalence of sedentary lifestyles has prompted the development of innovative public health interventions, such as smartphone apps that deliver personalized exercise programs. The widespread availability of mobile technologies (eg, smartphone apps and wearable activity trackers) provides a cost-effective, scalable way to remotely deliver personalized exercise programs to users. Using machine learning (ML), specifically reinforcement learning (RL), may enhance user engagement and effectiveness of these programs by tailoring them to individual preferences and needs.

Objective: The primary aim was to investigate the impact of the Samsung-developed i80 BPM app, implementing ML for exercise prescription, on user satisfaction and exercise intensity among the general population. The secondary objective was to assess the effectiveness of ML-generated exercise programs for remote prescription of exercise to members of the public.

Methods: Participants were randomized to complete 3 exercise sessions per week for 12 weeks using the i80 BPM mobile app, crossing over weekly between intervention and control conditions. The intervention condition involved individualizing exercise sessions using RL, based on user preferences such as exercise difficulty, selection, and intensity, whereas under the control condition, exercise sessions were not individualized. Exercise intensity (measured by the 10-item Borg scale) and user satisfaction (measured by the 8-item version of the Physical Activity Enjoyment Scale) were recorded after the session.

Results: In total, 62 participants (27 male and 42 female participants; mean age 43, SD 13 years) completed 559 exercise sessions over 12 weeks (9 sessions per participant). Generalized estimating equations showed that participants were more likely to exercise at a higher intensity (intervention: mean intensity 5.82, 95% CI 5.59 - 6.05 and control: mean intensity 5.19, 95% CI 4.97 - 5.41) and report higher satisfaction (RL: mean satisfaction 4, 95% CI 3.9-4.1 and baseline: mean satisfaction 3.73, 95% CI 3.6-3.8) in the RL model condition.

Conclusions: The findings suggest that RL can effectively increase both the intensity with which people exercise and their enjoyment of the sessions, highlighting the potential of ML to enhance remote exercise interventions. This study underscores the benefits of personalized exercise prescriptions in increasing adherence and satisfaction, which are crucial for the long-term effectiveness of fitness programs. Further research is warranted to explore the long-term impacts and potential scalability of RL-enhanced exercise apps in diverse populations. This study contributes to the understanding of digital health interventions in exercise science, suggesting that personalized, app-based exercise prescriptions may be more effective than traditional, nonpersonalized methods. The integration of RL into exercise apps could significantly impact public health, particularly in enhancing engagement and reducing the global burden of physical inactivity.

Trial Registration: ClinicalTrials.gov NCT06653049; https://clinicaltrials.gov/study/NCT06653049

(JMIR Mhealth Uhealth 2024;12:e49443) doi:10.2196/49443

KEYWORDS

reinforcement learning; exercise therapy; personal satisfaction; satisfaction; physiotherapy; physical therapy; exercise intensity; mobile apps; randomized controlled trial; crossover trial; apps; exercise; physical activity; mobile phone

Introduction

Every year, hundreds of randomized controlled trials evaluating the effects of exercise in diverse population groups are published. This research has demonstrated the medicinal capacities of exercise for numerous chronic diseases [1]. Consequently, exercise is considered an essential component in the management plans of many health care interventions [1-3]. Regular exercise has also been shown to have wide-ranging health benefits for nonpathological groups; for instance, there is evidence to suggest that a sedentary lifestyle may be an even stronger predictor of mortality than smoking, hypertension, and diabetes [4]. National and international campaigns, advertisements, and public health guidelines have been developed to increase public awareness of the benefits of exercise [5,6]. Despite this, inactivity (and the increased risk of morbidity and mortality associated with it) remains a significant public health concern [7,8]. This raises the question: if people know that regular physical activity and exercise are good for them, why are so many inactive?

Even in randomized controlled trials, low levels of adherence to exercise sometimes belie the treatment effect to such an extent that results in comparison to a nonexercising control group are statistically insignificant [9,10]. Adherence to exercise, which is defined as the degree to which the target intensity and volume are achieved [11,12], is likely to be worse among the general population, who are not enthusiastic volunteers in a research study and who are being closely supervised by a research team [13].

Many theories and models have been proposed to explain why adherence to exercise is suboptimal [14,15]. A recent umbrella review identified a number of key factors for improving adherence to exercise [16]. Among the 14 factors that were identified, individualizing the exercise program, making sure that it integrates easily into participants' daily living schedules, continually monitoring and providing feedback on progress (and adapting the exercise program accordingly), ensuring that users have an active role in goal setting, and using technology to deliver the exercise intervention were deemed to be important in improving adherence [16]. Patient education was also considered crucial to increase self-efficacy, enhancing the knowledge about what they can do and what they can change to improve their overall health [17-19].

Mobile technologies (eg, smartphone apps and wearable activity trackers) are a cost-effective, scalable way of delivering exercise programs to users that incorporate these factors. Over 60% of adults worldwide own a smartphone, with worldwide penetration rates highest in the United States (where >80% of the population uses a smartphone) [20]. In addition to being able to deliver interventions through wireless internet and messaging connectivity, smartphones have in-built tools like global positioning systems, inertial measurement units, and cameras that can objectively measure several exercise parameters [21-23]. Smartphones also have powerful computation and communication capabilities that enable the use of machine learning (ML) and artificial intelligence to individualize each user's exercise program.

Reinforcement learning (RL) represents a compelling method within ML for tailoring and adapting exercise programs to individual users. In RL, a decision-making agent performs actions that lead to preferred states within its environment. Each action transitions the environment to a new state, following which the agent receives feedback—either positive or negative reinforcement. This feedback helps to refine the agent's "policy," which is essentially a strategy that maps states to actions aimed at maximizing cumulative rewards over time.

For instance, consider the use of RL within a smartphone app designed for exercise prescription. Here, the agent could consider variables such as user satisfaction or perceived exertion (ie, the session's intensity) as states. The agent's actions could involve adjusting various exercise parameters, such as the number of sets, repetitions, or rest periods in a training session, or altering the types of recommended exercises. The reward function would assess the degree of user satisfaction or exertion, aiming to minimize discrepancies and optimize user experience.

RL is particularly well-suited to the sequential decision-making required in exercise prescription. It operates in a cycle where the agent proposes an exercise session, the user completes it and provides feedback, and the agent uses this information to tailor future sessions. Although several frameworks exist for automating exercise prescription based on user demographics, fitness levels, engagement behaviors, and preferences [24-27], comprehensive studies evaluating the effectiveness of fully computerized, app-based exercise prescription remain limited [28-31]. This gap underscores the need for further research to validate and optimize RL apps in this field.

Therefore, the aim of this study was to test a smartphone app that generates adaptive exercise regimes, incorporating RL to personalize the composition of exercises within sessions based on user satisfaction. We compared user satisfaction between sessions generated by the RL system to a control condition that administered a generic exercise to the user, irrespective of their preferences. The primary outcome for this aim was users' satisfaction with exercise, which was defined by an abbreviated 8-item version of the Physical Activity Enjoyment Scale (PACES-8) [32,33].

Our primary hypotheses were that users would report higher satisfaction and demonstrate greater adherence to exercise programs generated using RL compared with sessions that are generated randomly or using predefined (ie, nonpersonalized) templates. Our secondary hypothesis was that users would demonstrate higher levels of exertion, measured using Borg's rate of perceived exertion (RPE) scale, during sessions generated with an RL approach.

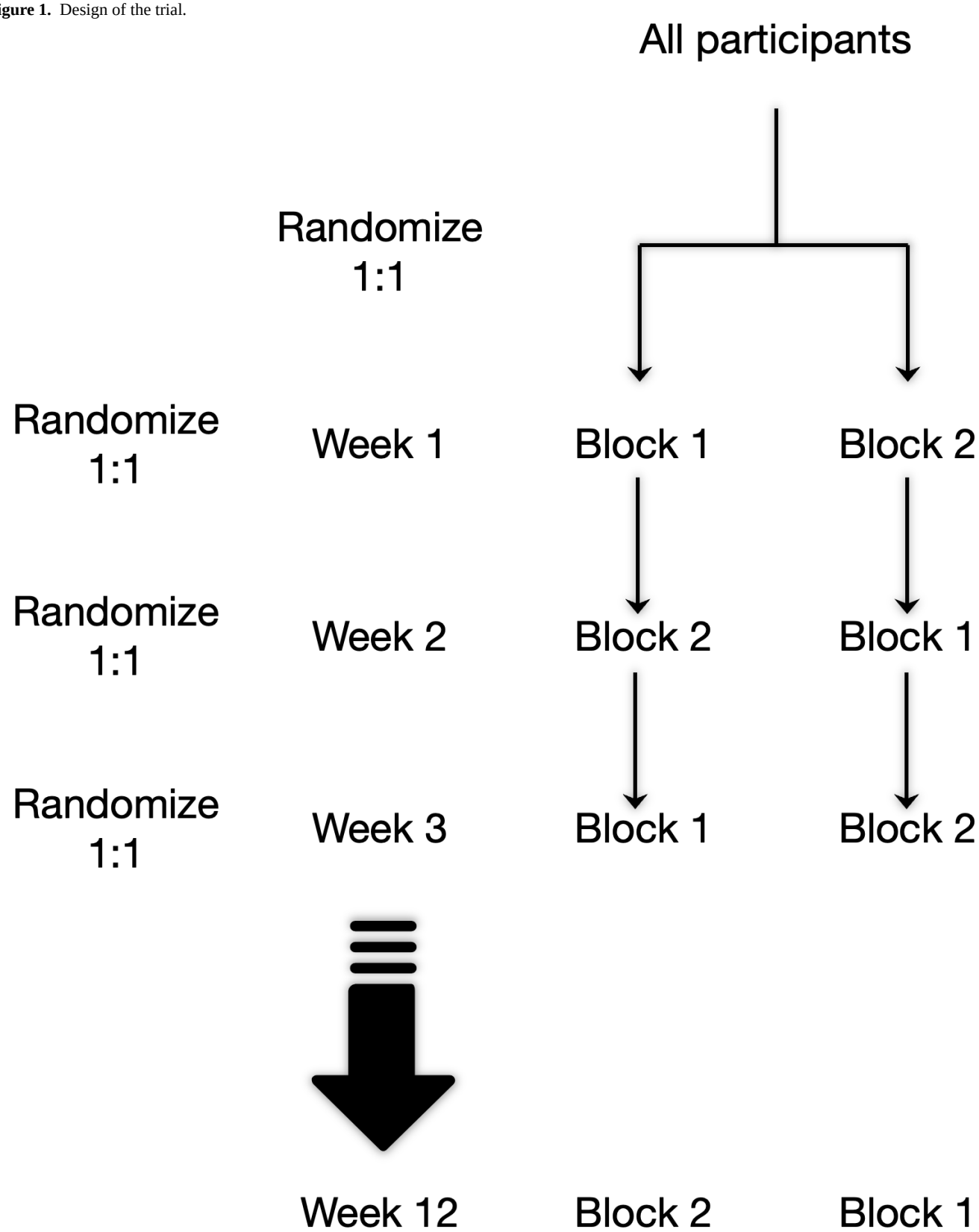
Methods

Study Design

The protocol for this study was developed using the SPIRIT (Standard Protocol Items: Recommendations for Interventional Trials) checklist [34] The design was a 12-week, assessor-blinded, randomized crossover trial, with the primary end point being user satisfaction after each exercise session. However, unlike a typical crossover trial, participants

transitioned “back and forth” between experimental conditions. Specifically, each participant alternated between the intervention and control conditions at the end of each 1-week cycle. The order in which participants completed each condition (RL condition and control condition) was altered on a weekly basis; each participant was administered condition-generated exercise sessions for 1 week. Each of these 1-week cycles was comprised of 3 workout sessions of approximately 20 minutes in duration, containing >30 exercises (each exercise lasted <30 seconds). The only difference between the workout sessions within each condition was the specific exercises that were recommended (and the order in which they were completed); all other parameters of exercise were held consistent. After a 1-week cycle had elapsed, each cluster of participants completing the protocol under each of the 2 conditions crossed over to the opposing condition, with each crossover marking the start of a new 1-week cycle, regardless of whether participants actually completed the sessions within that cluster. This design is illustrated in Figure 1.

Figure 1. Design of the trial.



Ethical Considerations

The University College Dublin Human Research Ethics Committee approved this study (LS-21 - 34-Tragos-Lawlor). Written informed consent was obtained, and health screening was conducted for all participants before they enrolled in the trial. The trial was not prospectively registered. Recruitment was conducted between July 6, 2022, and August 29, 2022, and the trial was completed on November 16, 2022. During data collection, each participant was given a participant ID number when they registered with the app. The data were stored using these ID numbers. The key to the code that matched participants' full name to their ID number was saved in an electronic password-protected file, which was stored on an encrypted drive.

The authorship team had sole access rights to these key codes and ID numbers during the trial. Data acquired via the smartphone app were stored on a password-protected University College Dublin server. After data collection, data were anonymized by permanently deleting the file with the participants' identification key. Participants were not compensated for their participation.

Population

Participants were recruited from Dublin, Ireland, and its environs via word of mouth and social media. Male and female participants aged 18 to 65 years were recruited. To be eligible for inclusion, participants were required to be healthy, recreationally active adults. This was defined as having engaged in aerobic activity for a total of 80 minutes at moderate intensity, less than or equal to twice per week. Prior to randomization, participants were screened for eligibility by the "exercise preparticipation health screening questionnaire for exercise professionals." Exclusion criteria included an inability to exercise due to physical disability or motor impairment, having a severe cognitive impairment, or an inability to read and write in English.

Blinding

Participants were blinded to their allocation throughout the course of the trial. The researcher conducting data analysis was also blinded to knowledge of the intervention.

Procedure

Upon expressing interest, potential participants received detailed information about the trial along with a health screening questionnaire. After completing the health screening, participants attended an information session via Zoom (Zoom Video Communications), where they were shown how to use the app and its various functions and also instructed on downloading it to their mobile devices. At this session, informed consent was obtained. Each participant was issued a Samsung Galaxy Fit 2 smartwatch (Samsung Electronics Co, Ltd) to use during the exercise sessions for recording heart rate data. Several studies have demonstrated the accuracy of the Samsung Gear Fit for heart rate measurement during a variety of activity types, making it a suitable choice for this study [35-37]. Participants were allowed to keep the fitness tracker after the trial concluded. Immediately following randomization, baseline data collection was carried out through the app.

Interventions

Participants were asked to complete 3 exercise sessions weekly over 12 weeks using the i80 BPM smartphone app developed by Samsung Electronics Co, Ltd. This app offers video-guided exercise programs and includes a library of 161 exercises. Using RL, i80 BPM customizes and adapts exercise sessions to suit the user's preferences and abilities. The duration of these sessions, ranging from 10 to 30 minutes, was selected by the user. Users could create personalized session plans varying from 1 to 12 weeks and choose between aerobic and muscle-strengthening exercises, targeting specific muscle groups. The RL model selected exercises for each session based on these preferences, ensuring a diverse range of possible sessions.

Aside from the RL model's app, all other parameters remained constant across both the intervention and control conditions. These included the exercises available, the app's interface, and options for session duration and focus (aerobic vs muscle strengthening). The key difference was the individualization of exercise sessions using the RL model in the intervention condition, whereas the control condition used generic, nonpersonalized sessions. Further details on the RL model are discussed elsewhere [38], but in summary, the model used within the i80 BPM app aimed to optimize user satisfaction by personalizing the exercise sequence based on user feedback and evolving preferences. The app's RL framework used a decision-making agent within an environment shaped by user interactions and fitness profiles. The model was trained to maximize expected rewards, which incorporated metrics of performance and user feedback. In this framework, "actions" referred to the selection of exercises from the app's database. The "state" included information such as the user's fitness level and exercise preferences along with session specifics like previously recommended exercises and user feedback. The "reward function" balanced various elements including exercise diversity, suitability to the user's fitness level, and feedback after each session to ensure that the recommendations met the user's goals and responses to past workout sessions. The RL model used a neural network to map states to action probabilities, optimizing the sequence of recommended exercises to maximize cumulative rewards—reflecting both user satisfaction and session effectiveness [38].

RL Intervention

The RL intervention used the app's full functionality, meaning that exercise sessions were tailored to the user by the RL model, as outlined earlier.

Baseline Control Condition

The control condition used the same i80 BPM app as described earlier; however, the RL model was not applied. Instead, generic exercise sessions were provided for the user, irrespective of their preferences.

Crossover Design

As crossover design was used in this trial, participants transitioned between the intervention and control conditions. At the end of each 1-week cycle, participants transitioned to the opposing condition regardless of how many sessions they had completed. Therefore, each participant could complete exercise sessions that were adapted by the RL model (intervention) as well as sessions that were not adapted (control). This design was chosen to maximize the amount of data captured for each condition in anticipation of a high number of dropouts or waning adherence as the trial progressed.

Outcome Measures

Our primary outcome, user satisfaction, was determined after each session and was collected via the smartphone app. Satisfaction was measured using an abbreviated PACES-8 [32,33].

Our secondary outcome was perceived exertion, measured using the Borg scale, which was administered via the smartphone app

at the end of each session; participants were cued to "rate your perceived exertion on a scale from 2 to 10, where 2 means 'really easy' and 10 means 'maximal exertion.'" Additionally, heart rate data were collected using the Galaxy Fit 2 smartwatch provided to each participant [37]. This device uses photoplethysmography to measure heart rate continuously during the exercise sessions. The heart rate data captured by the smartwatch were automatically relayed to the i80 BPM app, which was used to record and analyze these measurements as part of the session data. See Table 1 for all outcomes and questionnaires.

Table . Outcome measures.

		Baseline	Weekly	During each session	After each session
Satisfaction					
	Physical Activity Enjoyment Scale-8				✓
	Rating of perceived exertion				✓
	Heart rate			✓	
	Exercise skipped			✓	
Behavioral engagement					
	Exercise altered			✓	
	Total time spent on the app		✓		
	Number of sessions completed		✓		
User profiling					
	Demographics (eg, age and sex)	✓			
	International Physical Activity Questionnaire	✓			

Public Involvement

During an initial feasibility trial, members of the public worked with us to evaluate the sessions generated by the RL system and the mobile app user interface and were asked to assess the burden and time commitment of the study as part of a user-centered design approach [19]. During this feasibility study, participants also completed a web-based survey to establish their current activity levels, their experience with different health and fitness smartphone apps, how they currently exercise, and their self-efficacy. Participants were then asked to use the app over an 8-week period (between June and October 2021) and were asked to provide feedback on their experience. Analysis of participants' feedback and the app use logs was undertaken by the project team, and the app was further iterated based on this feedback. The version of the app that was used in this study (which took place between June and October 2022) integrated all participants' feedback from this initial feasibility trial.

Power Calculations and Sample Size

A sample size of minimum 40 participants was estimated based on the following factors: the sample size recruited as part of the initial feasibility study (we previously recruited 36 participants over an 8-week trial period), an estimated mean difference of 8 on the PACES-8 scale based on previous work for 80% power, and the rule of thumb of at least 10 events for variable (or measures of user satisfaction and behavioral and physiological engagement) [39]. Controlling for 15% dropout, we aimed to recruit a total of 42 participants.

Data Management and Statistical Analysis

Overview

All collected data were anonymized before analysis, with each participant assigned a unique identification code that was used in place of personal identifiers. Data were stored on secure, encrypted servers accessible only to authorized personnel involved in the study. All electronic communications and data transmissions involving participant information were encrypted. Participant consent forms and other sensitive documents were stored separately from the study data, in locked cabinets within secure facilities.

Baseline Characteristics

Descriptive statistics were used to summarize user demographics and scores on the questionnaires.

Primary and Secondary Hypotheses

The primary outcome measure, user satisfaction, was assessed using the PACES-8 questionnaire, while the secondary outcome, perceived exertion, was measured using the Borg scale. Both outcomes were evaluated at the session level, meaning each exercise session contributed to the analysis rather than summarizing data at the participant level.

To account for the repeated measures design and the correlation of observations within participants, generalized estimating equations (GEEs) were used to assess differences between the RL and control conditions for both primary and secondary outcomes. The dependent variables in the models were PACES satisfaction scores and Borg perceived exertion scores. The independent variables included condition (RL vs control) and trial week (treated as a continuous variable to account for time effects). Covariates adjusted for in the models included age, gender, and baseline physical activity level (categorized as low, moderate, or high). User ID was included as a subject effect to account for within-subject correlation. An independent working correlation matrix was specified for the GEE models, with an identity link function used for both models. Normal distribution was assumed for the outcomes. Model fit was assessed using the quasi-likelihood under independence model criterion and the corrected quasi-likelihood under independence model criterion, with lower values indicating better model fit. The main effects of condition and trial week were estimated, and their significance was assessed using Wald chi-square tests. All statistical analyses were performed using SPSS Statistics software (version 29; IBM Corp).

Results

In total, 69 participants were initially recruited for the study (27 male and 42 female participants); however, only 62 (24 male and 38 female participants) completed at least 1 exercise session. The mean age of participants was 42.8 (SD 13.3) years, with an average height of 1.7 (SD 2.4) m, body mass of 75.7 (SD 21.1) kg, and BMI of 25.8 (SD 7.8) kg/m^2. These 62 participants completed 559 exercise sessions between them (9 sessions per participant). A total of 16% (n=11) adhered to at least 50% (n=18) of the sessions they were assigned to complete (ie, 18 of 36), while 34% (n=23) adhered to at least 25% (n=9) of sessions. In total, 268 sessions were completed that were comprised of exercises generated by the RL model, while 291 sessions were completed that were comprised of exercises generated by the baseline model.

The PACES-8 GEE estimated a main effect for condition. An analysis of the parameter estimates revealed that participants were more satisfied with exercise sessions generated by the RL model (RL condition: mean satisfaction 4, 95% CI 3.9-4.1 and baseline condition: mean satisfaction 3.73, 95% CI 3.6-3.8). The main effect for condition was significant (P=.02), with a parameter estimate of B=−0.261 (SE 0.109), indicating a modest but meaningful difference in satisfaction scores. The 95% CI for this estimate ranged from −0.475 to −0.047.

The intensity GEE estimated significant main effects for condition (P<.01) and trial week (P=.048). The intensity was significantly higher in sessions generated by the RL model compared to the baseline model (RL condition: mean intensity 5.82, 95% CI 5.59 - 6.05 and baseline condition: mean intensity 5.19, 95% CI 4.97 - 5.41). Analysis of the parameter estimates indicated that participants exercised at higher intensities during RL sessions (B=−0.633, SE 0.179; P<.01).

We conducted an exploratory analysis of the separate questions from the PACES-8 questionnaire. Specifically, separate GEEs assessed differences in responses to individual items on the PACES-8 questionnaire. Significant main effects for condition were found for items related to pleasure (item #1), invigoration (item #4), gratification (item #5), exhilaration (item #6), stimulation (item #7), and how refreshing each session was (item #8; P=.02 for each). Summary statistics for the primary outcome measures of satisfaction, difficulty, heart rate, the number of exercises completed, and the duration for exercise sessions generated by the baseline and RL models are presented in Table 2.

Table . Summary statistics for overall satisfaction and heart rate for the baseline and reinforcement learning (RL) models and the average scores for the individual PACES-8[a] items.

Outcome	Baseline, mean (95% CI)	RL model, mean (95% CI)	*P* value
PACES-8	3.73 (3.6-3.8)	4 (3.9-4.1)	.02
Heart rate	106.81 (105-107)	107.24 (106-108)	—[b]
Intensity	5.19 (4.97-5.41)	5.82 (5.59-6.05)	<.01
Duration (minutes)	17 (17-18)	20 (20-21)	—
PACES-8_pleasure[c]	3.66 (3.6-3.7)	3.88 (3.8-4.0)	.047
PACES-8_fun[d]	3.53 (3.4-3.6)	3.72 (3.6-3.8)	.08
PACES-8_pleasant[e]	3.53 (3.4-3.6)	3.71 (3.6-3.8)	.05
PACES-8_invigorating[f]	3.75 (3.7-3.8)	4.15 (4.1-4.2)	<.01
PACES-8_gratifying[g]	3.84 (3.7-3.9)	4.17 (4.1-4.3)	<.01
PACES-8_exhilarating[h]	3.74 (3.6-3.8)	4.04 (3.9-4.1)	.02
PACES-8_stimulating[i]	3.85 (3.7-3.9)	4.13 (4.0-4.2)	.02
PACES-8_refreshing[j]	3.84 (3.7-3.9)	4.06 (4.0-4.1)	.02

[a]PACES-8: 8-item version of the Physical Activity Enjoyment Scale.

[b]Not available.

[c]PACES-8_pleasure: item related to the pleasure derived from the exercise.

[d]PACES-8_fun: item related to the fun experienced during the exercise session.

[e]PACES-8_pleasant: item related to how pleasant the exercise session felt.

[f]PACES-8_invigorating: item related to the invigorating nature of the exercise session.

[g]PACES-8_gratifying: item related to the gratification from the exercise session.

[h]PACES-8_exhilarating: item related to the exhilaration felt during the exercise session.

[i]PACES-8_stimulating: item related to how stimulating the exercise session was.

[j]PACES-8_refreshing: item related to how refreshing the exercise session felt.

Discussion

Principal Findings

The aim of this randomized crossover trial was to investigate whether exercise sessions generated using RL were associated with better user satisfaction compared with a control condition. In this trial, exercise recommendations were delivered via the same smartphone app over the course of a 12-week “back and forth” crossover period, in which participants were subjected to each experimental condition in a 1-week cycle, each comprising 3 exercise sessions of 20 minutes in duration. Whether participants actually completed the sessions within a cycle did not affect their weekly crossover to an alternative condition. The primary outcome of user satisfaction was determined using the PACES-8 questionnaire, which has been previously used in diverse population groups [32,33,40,41] and as part of smartphone app testing protocols [42]. The secondary outcome, perceived exertion, was determined after each exercise session using the Borg scale.

During the trial, 62 (62% female) participants completed 559 exercise sessions (approximately 9 sessions per participant). Our results showed that participants were more satisfied with exercise sessions generated by the RL model (RL condition: mean satisfaction 4, 95% CI 3.9-4.1 and baseline condition: mean satisfaction 3.73, 95% CI 3.6-3.8). By incorporating trial week as a covariate in our analysis, we were able to show that participants were generally more satisfied with exercise sessions generated by the RL model compared to the control condition. Additionally, participants reported exercising at a higher intensity during sessions generated using the RL model (RL condition: mean intensity 5.82, 95% CI 5.59 - 6.05 and baseline condition: mean intensity 5.19, 95% CI 4.97 - 5.41).

Comparison to Prior Work

Our findings have import practical implications. Despite the well-known benefits of regular physical activity, many individuals struggle to stick to an exercise routine over the long term [43,44]. This is often due to a lack of motivation or a mismatch between the exercise program and the individual’s personal preferences and abilities [45-47]. RL-generated sessions, however, make it possible to personalize exercise programs, tailoring them to an individual’s unique preferences, abilities, and goals. This could potentially improve motivation and adherence to the exercise routine, ultimately leading to better health outcomes.

In particular, the ubiquity of smartphones means that app-based exercise prescription, enabled by RL algorithms, has the potential to greatly improve public health by increasing motivation and adherence to exercise routines. RL is a type of

artificial intelligence that involves training a model to make decisions that maximize a reward. This is typically done through a process of trial and error, where the model receives a reward for taking certain actions and a penalty for others. Over time, the model learns to make the decisions that are most likely to result in the maximum reward. The process of RL is similar to how animals and humans learn to perform tasks, and it has been used to solve a wide range of problems in fields such as robotics, gaming, and finance [48,49]. The key advantage of RL is its ability to learn from experience and adapt to new situations without needing to be explicitly programmed. This makes it an effective approach for solving complex, dynamic problems such as how to devise an exercise program.

The use of ML methods, including RL, to generate is a burgeoning research area. Previous research has demonstrated the effectiveness of using ML for tailoring interventions in these apps. For instance, Aguilera et al [50] observed a significant improvement in step count among individuals with diabetes and depression over the short term when ML was used to tailor SMS text messages within a self-management app. Additionally, RL has been successfully applied to enhance adherence to exercise in patients with diabetes through the distribution of individualized service messages that encourage physical activity [51]. The application of ML has even shown promise in the general population, with personalized daily step goals leading to increased physical activity compared to static step goals [52].

Our trial is the first known study to use an RL model specifically for personalizing exercise sessions and assessing its impact on participant satisfaction. While prior studies have used RL to promote exercise adherence through personalized messaging, the outcomes were mostly centered around physical activity measures such as step count [52,53]. Although our study and these previous studies differ in design and primary outcomes, a common thread can be observed regarding the potential of RL to encourage physical activity and exercise, despite the challenges inherent in smartphone app delivery. Furthermore, there is a relative scarcity of evidence concerning the use of RL specifically for exercise, as opposed to physical activity. RL has found utility in wider health care contexts, such as individualizing the difficulty level in a virtual reality rehabilitation game [54], aiding in weight loss maintenance [55], and even selecting and delivering drugs [56-58].

These studies, alongside the findings from our trial, reinforce the emerging potential of RL in health-related behaviors and outcomes. The results of this crossover trial contribute to our understanding of the potential for adaptive exercise interventions to improve adherence to an exercise intervention. Our hypothesis that users would be more satisfied with sessions generated by the RL system was confirmed. The effect size, calculated as the mean difference normalized by the SD (Cohen *d*), was approximately 0.24. This represents a small to moderate effect size, suggesting that while the RL model led to higher satisfaction, the magnitude of this difference, though statistically significant, might not be substantial in practical terms. These findings underscore the importance of considering both statistical and practical significance in evaluating the efficacy of health interventions. The practical implications suggest that while the RL model enhances satisfaction, efforts to refine the model to maximize its impact on user engagement and satisfaction should continue. Future research should focus on identifying and integrating additional predictors of satisfaction into the RL model, potentially increasing the effect size and thereby the practical impact of the technology. This research could also inform the development of RL systems for exercise prescription in patient populations (such as those with cardiovascular or metabolic diseases). This could lead to better health outcomes and reduced health care costs, as regular physical activity and exercise are known to prevent a wide range of chronic diseases [59]. Additionally, app-based exercise prescription has the potential to be more accessible and convenient than traditional forms of exercise prescription, making it easier for individuals to incorporate physical activity into their daily lives. On this basis, RL in smartphone app–based exercise prescription has the potential to greatly benefit public health; this study is the first step in setting the empirical foundation for follow-on research to investigate and establish the potential of RL-based exercise prescription in larger and more diverse populations.

Typically, in exercise science research investigating the use of mobile technologies to deliver the exercise intervention, participants are randomized to either an intervention group or a control group [60,61]. These groups are subsequently monitored for a predefined period as they use the app and are measured at the end of the intervention period to determine the health benefits of the exercise intervention, for example, based on a decrease in body weight or BMI, an increase in fat-free mass, or alterations in certain blood biomarkers [62,63]. However, the health benefits of exercise in diverse population groups are now well established. This is one of the primary reasons why we adopted a session-focused, “back and forth” crossover design (which was more robust to reducing levels of participation over the course of the trial), with satisfaction as our primary outcome. By cycling between each intervention condition on a weekly basis, participants acted as their own controls and ensured that we have a representative dataset for both the RL and control conditions. Rather than user-associated outcomes such as body weight at the start and conclusion of the intervention period being the primary experimental end point then, we focused on session-level outcomes like satisfaction and RPE. Specifically, each exercise session was characterized by the aggregated statistics of satisfaction and RPE for our entire cohort. This allowed us to build a profile for each session and evaluate how profiles are altered for each experimental condition under which the sessions are completed (and also to provide continuous input for the RL system).

However, despite its novelty and the strength of its methodological design, this study is not without limitations. Ours was a relatively small sample of homogenous participants; therefore, our findings may not be generalizable to other populations. It is also important to note the potential for selection bias in our recruitment strategy. Participants were recruited through word of mouth and social media, which might not provide a fully representative sample of the target population. This approach could attract individuals who are more technologically savvy, potentially more health-conscious, or with specific demographic characteristics that are

overrepresented in social media networks. Consequently, our findings might not be generalizable to all segments of the population. While the crossover study design was statistically efficient, it meant that we were unable to compare 2 separate groups in the long term for our primary and secondary outcomes. It was also not possible to follow a double-blind design, as researchers needed to be aware of participants' group allocation throughout the intervention period to manage follow-up and ensure that participants did not have any issues using the app; the researchers and technical personnel had to monitor our server for troubleshooting throughout the trial. Finally, we observed that participants completed an average of 9 sessions over the 12-week trial period, equating to less than 1 session per week. This level of engagement is notably lower than the intended 3 sessions per week per participant outlined in the study protocol. Several factors could contribute to this lower-than-expected adherence rate, including potential barriers faced by participants such as time constraints, lack of motivation, or the possibility that the app did not fully engage users as anticipated. Understanding and addressing these factors are crucial for the future development of app-based exercise prescriptions, as enhanced user engagement is key to achieving the health benefits of regular physical activity. The insights gained could also inform the broader field of digital health interventions aimed at improving lifestyle behaviors.

Conclusions

This randomized crossover trial evaluated the utility of a type of artificial intelligence called RL for devising an app-based system for exercise prescription. Our findings demonstrated that exercise sessions generated using RL were associated with higher satisfaction. Additionally, our participants completed sessions in the RL condition at a higher intensity than in the control condition. Taken together, our results suggest that there is significant scope for using RL algorithms to train a model to recommend exercises that are tailored to an individual's unique goals, preferences, and abilities. Further research should investigate the potential for exercise prescription using RL to empower the general population to be healthier by providing personalized and engaging exercise programs that are tailored to an individual's unique goals, preferences, and abilities. Such an approach may help individuals take control of their own health and fitness and provide them with the tools and support they need to maintain a regular exercise routine. By making exercise more accessible, engaging, and personalized, RL has the potential to improve public health by encouraging more individuals to incorporate physical activity into their daily lives.

Acknowledgments

This publication has emanated from research conducted with the financial support of the Science Foundation Ireland (grant 12/RC/2289_P2) and by Samsung Electronics c. Ltd., Seoul, Republic of Korea. For the purpose of open access, the author has applied a Creative Commons public copyright license to any author-accepted manuscript version arising from this submission.

Conflicts of Interest

None declared.

Editorial Notice

This randomized study was only retrospectively registered. The editor granted an exception from ICMJE rules mandating prospective registration of randomized trials, because the risk of bias appears low and the study was considered formative, guiding the development of the application. However, readers are advised to carefully assess the validity of any potential explicit or implicit claims related to primary outcomes or effectiveness, as retrospective registration does not prevent authors from changing their outcome measures retrospectively.

Checklist 1
CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) checklist (version 1.6.1).
[PDF File, 324 KB - mhealth_v12i1e49443_app1.pdf]

References

1. Pedersen BK, Saltin B. Exercise as medicine—evidence for prescribing exercise as therapy in 26 different chronic diseases. Scand Med Sci Sports 2015 Dec;25(S3):1-72. [doi: 10.1111/sms.12581]
2. Lavie CJ, O'Keefe JH, Sallis RE. Exercise and the heart—the harm of too little and too much. Curr Sports Med Rep 2015;14(2):104-109. [doi: 10.1249/JSR.0000000000000134]
3. Lavie CJ, Ozemek C, Carbone S, Katzmarzyk PT, Blair SN. Sedentary behavior, exercise, and cardiovascular health. Circ Res 2019 Mar;124(5):799-815. [doi: 10.1161/CIRCRESAHA.118.312669]
4. Ross R, Blair SN, Arena R, et al. Importance of assessing cardiorespiratory fitness in clinical practice: a case for fitness as a clinical vital sign: a scientific statement from the American Heart Association. Circulation 2016 Dec 13;134(24):27881567. [doi: 10.1161/CIR.0000000000000461]
5. Blair SN, Sallis RE, Hutber A, Archer E. Exercise therapy—the public health message. Scand Med Sci Sports 2012 Aug;22(4):22429265. [doi: 10.1111/j.1600-0838.2012.01462.x]

6. Thomas MM, Phongsavan P, McGill B, O'Hara BJ, Bauman AE. A review of the impact of physical activity mass media campaigns on low compared to high socioeconomic groups. Health Educ Res 2018 Oct 1;33(5):429-446. [doi: 10.1093/her/cyy032]
7. Booth FW, Roberts CK, Laye MJ. Lack of exercise is a major cause of chronic diseases. Compr Physiol 2012 Apr;2(2):1143-1211. [doi: 10.1002/cphy.c110025] [Medline: 23798298]
8. Arocha Rodulfo JI. Sedentarismo, la enfermedad del siglo xxi. Clín Invest Arterioscl 2019 Sep;31(5):233-240. [doi: 10.1016/j.arteri.2019.04.004]
9. Johnston CA, Moreno JP, Hernandez DC, et al. Levels of adherence needed to achieve significant weight loss. Int J Obes 2019 Jan;43(1):125-131. [doi: 10.1038/s41366-018-0226-7]
10. Beaudry R, Kruger C, Liang Y, Parliament M, Haykowsky M, McNeely M. Effect of supervised exercise on aerobic capacity in cancer survivors: adherence and workload predict variance in effect. WJMA 2015;3(1):43. [doi: 10.13105/wjma.v3.i1.43]
11. Hawley-Hague H, Horne M, Skelton DA, Todd C. Review of how we should define (and measure) adherence in studies examining older adults' participation in exercise classes. BMJ Open 2016 Jun 23;6(6):e011560. [doi: 10.1136/bmjopen-2016-011560] [Medline: 27338884]
12. Slade SC, Dionne CE, Underwood M, Buchbinder R. Consensus on Exercise Reporting Template (CERT): explanation and elaboration statement. Br J Sports Med 2016 Dec;50(23):1428-1437. [doi: 10.1136/bjsports-2016-096651]
13. Lemstra M, Bird Y, Nwankwo C, Rogers M, Moraros J. Weight loss intervention adherence and factors promoting adherence: a meta-analysis. Patient Prefer Adherence 2016;10(1547-59):1547-1559. [doi: 10.2147/PPA.S103649] [Medline: 27574404]
14. Sirur R, Richardson J, Wishart L, Hanna S. The role of theory in increasing adherence to prescribed practice. Physiother Can 2009;61(2):68-77. [doi: 10.3138/physio.61.2.68] [Medline: 20190989]
15. Lai B, Young HJ, Bickel CS, Motl RW, Rimmer JH. Current trends in exercise intervention research, technology, and behavioral change strategies for people with disabilities. Am J Phys Med Rehabil 2017;96(10):748-761. [doi: 10.1097/PHM.0000000000000743]
16. Collado-Mateo D, Lavín-Pérez AM, Peñacoba C, et al. Key factors associated with adherence to physical exercise in patients with chronic diseases and older adults: an umbrella review. Int J Environ Res Public Health 2021 Feb 19;18(4):2023. [doi: 10.3390/ijerph18042023] [Medline: 33669679]
17. Bhat AA, DeWalt DA, Zimmer CR, Fried BJ, Callahan LF. The role of helplessness, outcome expectation for exercise and literacy in predicting disability and symptoms in older adults with arthritis. Patient Educ Couns 2010 Oct;81(1):73-78. [doi: 10.1016/j.pec.2009.12.004]
18. Dewalt DA, Berkman ND, Sheridan S, Lohr KN, Pignone MP. Literacy and health outcomes: a systematic review of the literature. J Gen Intern Med 2004 Dec;19(12):1228-1239. [doi: 10.1111/j.1525-1497.2004.40153.x] [Medline: 15610334]
19. Keogh A, Argent R, Doherty C, et al. Breaking down the digital fortress: the unseen challenges in healthcare technology—lessons learned from 10 years of research. Sensors (Basel) 2024 Jun 11;24(12):3780. [doi: 10.3390/s24123780] [Medline: 38931564]
20. Smartphone-penetration-worldwide-by-country. Statista. 2022. URL: https://www.statista.com/statistics/539395/smartphone-penetration-worldwide-by-country [accessed 2024-10-30]
21. Althoff T, Sosič R, Hicks JL, King AC, Delp SL, Leskovec J. Large-scale physical activity data reveal worldwide activity inequality. Nature New Biol 2017 Jul 20;547(7663):336-339. [doi: 10.1038/nature23018] [Medline: 28693034]
22. Case MA, Burwick HA, Volpp KG, Patel MS. Accuracy of smartphone applications and wearable devices for tracking physical activity data. JAMA 2015 Feb 10;313(6):625. [doi: 10.1001/jama.2014.17841]
23. Hekler EB, Buman MP, Grieco L, et al. Validation of physical activity tracking via android smartphones compared to ActiGraph accelerometer: laboratory-based and free-living validation studies. JMIR Mhealth Uhealth 2015 Apr 15;3(2):e36. [doi: 10.2196/mhealth.3505] [Medline: 25881662]
24. Kostopoulos K, Chouvarda I, Koutkias V, Kokonozi A, van Gils M, Maglaveras N. An ontology-based framework aiming to support personalized exercise prescription: application in cardiac rehabilitation. Presented at: 2011 33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society; Aug 30 to Sep 3, 2011; Boston, MA, United States. [doi: 10.1109/IEMBS.2011.6090456]
25. Lofaro D, Groccia MC, Guido R, Conforti D, Caroleo S, Fragomeni G. Machine learning approaches for supporting patient:specific cardiac rehabilitation programs. Presented at: Proceedings of the 2016 Computing in Cardiology Conference (CinC); Sep 14, 2016; Vancouver, BC, Canada. [doi: 10.22489/CinC.2016.047-209]
26. Lo CL, Tseng HT. Predicting rehabilitation treatment helpfulness to stroke patients: a supervised learning approach. Artif Intell Res 2017;6(2):1. [doi: 10.5430/air.v6n2p1]
27. Hansen D, Dendale P, Coninx K, et al. The European Association of Preventive Cardiology Exercise Prescription in Everyday Practice and Rehabilitative Training (EXPERT) tool: a digital training and decision support system for optimized exercise prescription in cardiovascular disease. Concept, definitions and construction methodology. Eur J Prev Cardiol 2017 Jul;24(10):1017-1031. [doi: 10.1177/2047487317702042]
28. Azar KMJ, Lesser LI, Laing BY, et al. Mobile applications for weight management: theory-based content analysis. Am J Prev Med 2013 Nov;45(5):583-589. [doi: 10.1016/j.amepre.2013.07.005] [Medline: 24139771]

29. Bardus M, van Beurden SB, Smith JR, Abraham C. A review and content analysis of engagement, functionality, aesthetics, information quality, and change techniques in the most popular commercial apps for weight management. Int J Behav Nutr Phys Act 2016 Mar 10;13(35):35. [doi: 10.1186/s12966-016-0359-9] [Medline: 26964880]
30. Mercer K, Li M, Giangregorio L, Burns C, Grindrod K. Behavior change techniques present in wearable activity trackers: a critical analysis. JMIR Mhealth Uhealth 2016 Apr 27;4(2):e40. [doi: 10.2196/mhealth.4461] [Medline: 27122452]
31. Breslin J, Vickey T, Williams A. Fitness—there's an app for that: review of mobile fitness apps. Int J Sport Soc 2013;3(4):109-127. [doi: 10.18848/2152-7857/CGP/v03i04/53952]
32. McArthur LH, Raedeke TD. Race and sex differences in college student physical activity correlates. Am J Health Behav 2009;33(1):80-90. [doi: 10.5993/AJHB.33.1.8]
33. Raedeke TD. The relationship between enjoyment and affective responses to exercise. J Appl Sport Psychol 2007 Feb 14;19(1):105-115. [doi: 10.1080/10413200601113638]
34. Chan AW, Tetzlaff JM, Altman DG, et al. SPIRIT 2013 statement: defining standard protocol items for clinical trials. Ann Intern Med 2013 Feb 5;158(3):200-207. [doi: 10.7326/0003-4819-158-3-201302050-00583] [Medline: 23295957]
35. Kim C, Kim SH, Suh MR. Accuracy and validity of commercial smart bands for heart rate measurements during cardiopulmonary exercise test. Ann Rehabil Med 2022 Aug;46(4):209-218. [doi: 10.5535/arm.22050] [Medline: 36071003]
36. Spinsante S, Porfiri S, Scalise L. Accuracy of heart rate measurements by a smartwatch in low intensity activities. Presented at: 2019 IEEE International Symposium on Medical Measurements and Applications (MeMeA); Jun 26-28, 2019; Istanbul, Turkey. [doi: 10.1109/MeMeA.2019.8802216]
37. Xie J, Wen D, Liang L, Jia Y, Gao L, Lei J. Evaluating the validity of current mainstream wearable devices in fitness tracking under various physical activities: comparative study. JMIR Mhealth Uhealth 2018 Apr 12;6(4):e94. [doi: 10.2196/mhealth.9754] [Medline: 29650506]
38. Tragos E, O'Reilly-Morgan D, Geraci J, et al. Keeping people active and healthy at home using a reinforcement learning-based fitness recommendation framework. Presented at: Thirty-Second International Joint Conference on Artificial Intelligence {IJCAI-23}; Aug 19, 2023; Macau, SAR China URL: https://www.ijcai.org/proceedings/2023 [accessed 2024-11-20] [doi: 10.24963/ijcai.2023/692]
39. Terwee CB, Bot SDM, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007 Jan;60(1):34-42. [doi: 10.1016/j.jclinepi.2006.03.012]
40. Bird ML, Clark B, Millar J, Whetton S, Smith S. Exposure to "exergames" increases older adults' perception of the usefulness of technology for improving health and physical activity: a pilot study. JMIR Serious Games 2015 Nov 27;3(2):e8. [doi: 10.2196/games.4275] [Medline: 26614263]
41. Forsyth L, Bonacci J, Childs C. A pilot randomised control trial of the efficacy of stability-based training with visualisation for people with chronic ankle instability. Med Biol Eng Comput 2022 Apr;60(4):1199-1209. [doi: 10.1007/s11517-022-02533-z] [Medline: 35247167]
42. Taylor ME, Close JCT, Lord SR, et al. Pilot feasibility study of a home-based fall prevention exercise program (StandingTall) delivered through a tablet computer (iPad) in older people with dementia. Australas J Ageing 2020 Sep;39(3):e278-e287. [doi: 10.1111/ajag.12717] [Medline: 31538401]
43. Li CL, Lai YC, Tseng CH, Lin JD, Chang HY. A population study on the association between leisure time physical activity and self-rated health among diabetics in Taiwan. BMC Public Health 2010 May 26;10(1):277. [doi: 10.1186/1471-2458-10-277] [Medline: 20500890]
44. Artistico D, Pinto AM, Douek J, Black J, Pezzuti L. The value of removing daily obstacles via everyday problem-solving theory: developing an applied novel procedure to increase self-efficacy for exercise. Front Psychol 2013;4:20. [doi: 10.3389/fpsyg.2013.00020] [Medline: 23372560]
45. Chard S. Qualitative perspectives on aquatic exercise initiation and satisfaction among persons with multiple sclerosis. Disabil Rehabil 2017 Jun 19;39(13):1307-1312. [doi: 10.1080/09638288.2016.1194897]
46. Sun T, Xu Y, Xie H, Ma Z, Wang Y. Intelligent personalized exercise prescription based on an eHealth promotion system to improve health outcomes of middle-aged and older adult community dwellers: pretest-posttest study. J Med Internet Res 2021 May 24;23(5):e28221. [doi: 10.2196/28221] [Medline: 34028359]
47. Sutin AR, Stephan Y, Luchetti M, Artese A, Oshio A, Terracciano A. The five-factor model of personality and physical inactivity: a meta-analysis of 16 samples. J Res Pers 2016 Aug;63:22-28. [doi: 10.1016/j.jrp.2016.05.001] [Medline: 29056783]
48. Abdulhameed SA, Lupenko S. Potentials of reinforcement learning in contemporary scenarios. Sci J TNTU 2022;2(106):92-100. [doi: 10.33108/visnyk_tntu2022.02.092]
49. Charpentier A, Élie R, Remlinger C. Reinforcement learning in economics and finance. Comput Econ 2021;62(1):425-462. [doi: 10.1007/s10614-021-10119-4]
50. Aguilera A, Figueroa CA, Hernandez-Ramos R, et al. mHealth app using machine learning to increase physical activity in diabetes and depression: clinical trial protocol for the DIAMANTE Study. BMJ Open 2020 Aug 20;10(8):e034723. [doi: 10.1136/bmjopen-2019-034723] [Medline: 32819981]

51. Zhou M, Fukuoka Y, Mintz Y, et al. Evaluating machine learning-based automated personalized daily step goals delivered through a mobile phone app: randomized controlled trial. JMIR Mhealth Uhealth 2018 Jan 25;6(1):e28. [doi: 10.2196/mhealth.9117] [Medline: 29371177]
52. Yom-Tov E, Feraru G, Kozdoba M, Mannor S, Tennenholtz M, Hochberg I. Encouraging physical activity in patients with diabetes: intervention using a reinforcement learning system. J Med Internet Res 2017 Oct 10;19(10):e338. [doi: 10.2196/jmir.7994] [Medline: 29017988]
53. Zhou M, Mintz Y, Fukuoka Y, et al. Personalizing mobile fitness apps using reinforcement learning. CEUR Workshop Proc 2018 Mar 7;2068. [Medline: 32405286]
54. Ávila-Sansores S, Orihuela-Espina F, Enrique-Sucar L. Patient tailored virtual rehabilitation. In: Converging Clinical and Engineering Research on Neurorehabilitation: Springer Berlin Heidelberg; 2013:879-883. [doi: 10.1007/978-3-642-34546-3_143]
55. Forman EM, Kerrigan SG, Butryn ML, et al. Can the artificial intelligence technique of reinforcement learning use continuously-monitored digital data to optimize treatment for weight loss? J Behav Med 2019 Apr;42(2):276-290. [doi: 10.1007/s10865-018-9964-1] [Medline: 30145623]
56. Parbhoo S. A Reinforcement Learning Design for HIV Clinical Trials [PhD thesis]. : Faculty of Science, School of Computer Science, University of the Witwatersrand; 2014 URL: https://wiredspace.wits.ac.za/server/api/core/bitstreams/c45179b7-e1e5-47cd-876a-665ed1f3019f/content [accessed 2024-11-05]
57. Daskalaki E, Diem P, Mougiakakou SG. Personalized tuning of a reinforcement learning control algorithm for glucose regulation. Annu Int Conf IEEE Eng Med Biol Soc 2013;2013:3487-3490. [doi: 10.1109/EMBC.2013.6610293] [Medline: 24110480]
58. Escandell-Montero P, Chermisi M, Martínez-Martínez JM, et al. Optimization of anemia treatment in hemodialysis patients via reinforcement learning. Artif Intell Med 2014 Sep;62(1):47-60. [doi: 10.1016/j.artmed.2014.07.004]
59. Physical activity—key facts. World Health Organization. 2023. URL: https://www.who.int/news-room/fact-sheets/detail/physical-activity#:~:text=Regular%20physical%20activity%20is%20proven,of%20life%20and%20well%2Dbeing [accessed 2024-10-30]
60. Hurling R, Catt M, Boni MD, et al. Using internet and mobile phone technology to deliver an automated physical activity program: randomized controlled trial. J Med Internet Res 2007 Apr 27;9(2):e7. [doi: 10.2196/jmir.9.2.e7] [Medline: 17478409]
61. Allen JK, Stephens J, Dennison Himmelfarb CR, Stewart KJ, Hauck S. Randomized controlled pilot study testing use of smartphone technology for obesity treatment. J Obes 2013;2013(1):151597. [doi: 10.1155/2013/151597] [Medline: 24392223]
62. Spring B, Pellegrini CA, Pfammatter A, et al. Effects of an abbreviated obesity intervention supported by mobile technology: the ENGAGED randomized clinical trial. Obesity (Silver Spring) 2017 Jul;25(7):1191-1198. [doi: 10.1002/oby.21842] [Medline: 28494136]
63. Lahtio H, Rintala A, Immonen J, Sjögren T. The effectiveness of physical activity-promoting web- and mobile-based distance weight loss interventions on body composition in rehabilitation settings: systematic review, meta-analysis, and meta-regression analysis. J Med Internet Res 2022 Mar 24;24(3):e25906. [doi: 10.2196/25906] [Medline: 35323126]

Abbreviations

GEE: generalized estimating equation
ML: machine learning
PACES-8: 8-item version of the Physical Activity Enjoyment Scale
RL: reinforcement learning
RPE: rate of perceived exertion
SPIRIT: Standard Protocol Items: Recommendations for Interventional Trials

Edited by L Buis; submitted 29.05.23; peer-reviewed by C Jacob, H Takeuchi; revised version received 30.07.24; accepted 26.08.24; published 26.11.24.

Please cite as:
Doherty C, Lambe R, O'Grady B, O'Reilly-Morgan D, Smyth B, Lawlor A, Hurley N, Tragos E
An Evaluation of the Effect of App-Based Exercise Prescription Using Reinforcement Learning on Satisfaction and Exercise Intensity: Randomized Crossover Trial
JMIR Mhealth Uhealth 2024;12:e49443
URL: https://mhealth.jmir.org/2024/1/e49443
doi:10.2196/49443

The Real-World Impact of App-Based Mindfulness on Headspace Members With Moderate and Severe Perceived Stress: Observational Study

Christine Callahan, PhD; Justin Kimber, PhD; Emily Hu, PhD; Leah Tanner, MSPH; Sarah Kunkle, MPH

Headspace, Santa Monica, CA, United States

Corresponding Author:
Christine Callahan, PhD
Headspace, , Santa Monica, CA, , United States

Abstract

Background: Perceived stress in the United States has drastically increased since the COVID-19 pandemic and is associated with negative mental health outcomes such as depression and anxiety. Digital mental health (DMH) interventions are efficacious tools to address negative mental health outcomes and have helped reduce the severity of psychological symptoms, such as anxiety, depression, and perceived stress, compared to waitlist controls. Although DMH tools have been studied in controlled settings, less is known about the real-world evidence of such interventions.

Objective: This study aimed to (1) characterize patterns in baseline perceived stress and changes in perceived stress among Headspace members with moderate and severe baseline perceived stress and (2) examine associations between engagement with Headspace content and changes in perceived stress (ie, evaluate whether there is a dose-response relationship).

Methods: We evaluated real-world perceived stress and engagement data at 2 time points among Headspace app members with baseline moderate and severe perceived stress. Perceived stress was measured using the Perceived Stress Scale (PSS-10) and engagement using active days and active minutes engaged with Headspace as well as the number of user sessions. Descriptive statistics were computed for all variables. Correlations between baseline and follow-up scores, percent change in PSS-10 scores, days between PSS-10 use, active days, active days per week, active minutes, active minutes per day, sessions, and sessions per week were evaluated. We used *t* tests to investigate differences in the abovementioned parameters between (1) participants who did and those who did not see improvements in PSS-10 scores (yes vs no improvement) and (2) participants who saw ≥30% improvement versus those who saw a <30% improvement in PSS-10 scores.

Results: Overall, 21,088 Headspace members were included in these analyses. On average, members saw a 23.52% decrease in PSS-10 scores from baseline to follow-up. On average, members had 2.42 (SD 1.76) active days per week and 25.89 (SD 33.40) active minutes per day, and completed 7.11 (SD 8.34) sessions per week. *t* tests suggest that members who saw improvements in PSS-10 scores from baseline to follow-up had significantly higher baseline PSS-10 scores (Cohen *d*=0.56), more active days per week (Cohen *d*=0.33), and more sessions per week (Cohen *d*=0.27) than those who did not see improvements in PSS-10 scores (all *P*<.001). Additional *t* tests suggest that members with ≥30% improvement in PSS-10 scores had significantly higher baseline PSS-10 scores (Cohen *d*=0.35), more active days per week (Cohen *d*=0.36), and more sessions per week (Cohen *d*=0.31) than those with a >30% improvement (all *P*<.001).

Conclusions: Real-world use of Headspace is associated with decreased perceived stress. Furthermore, data suggest that more engagement, specifically weekly active days and sessions, is associated with a greater likelihood of stress reduction.

(JMIR Mhealth Uhealth 2024;12:e52968) doi:10.2196/52968

KEYWORDS

digital mental health; meditation; real word evidence; mental health; app; application; stress; observational study; depression; anxiety; tool; tools; psychological; psychological symptom; engagement; stress reduction; PSS-10

Introduction

Stress in the United States has increased, with significant impacts from the COVID-19 pandemic. While the pandemic itself caused a drastic increase in stress from 2019 to 2020 [1], stress levels still remain high and continue to impact the majority of Americans [2]. Higher perceived stress is associated with poorer mental health outcomes such as depression and anxiety [3]. Further meta-analytic research indicates small to medium effects of the relationship between perceived stress and mental health outcomes, suggesting that peoples' appraisal of stressful situations in their lives might be a predictor of mental health outcomes [4,5]. Higher perceived stress also has economic costs, and studies have estimated that perceived stress, primarily

work-related perceived stress, accounts for US $221 million to US $187 billion in both direct (eg, medical) and indirect costs (eg, absenteeism, burnout, and decreased productivity) [6]. This is important in the larger context of mental health costs and outcomes and how people manage stressors daily in their lives. Digital mental health (DMH) interventions may provide an accessible, scalable way to mitigate perceived stress.

DMH interventions are scalable and accessible; incorporate evidence-based practices (ie, mindfulness meditation and cognitive behavioral therapy techniques); are efficacious for a range of mental health concerns including anxiety, depression, and posttraumatic stress disorder [3]; and have shown to be effective in a wide range of populations with all levels of mental health concerns (ie, mild, moderate, and severe), including college students, employees, graduate trainees, and rural communities [7-9]. Studies suggest that digital mindfulness-based interventions significantly reduce perceived stress [10-12]. Specifically, a randomized controlled trial evaluated an app-based mindfulness tool among individuals with moderate and severe baseline perceived stress and suggested a 30.12% decrease in perceived stress from baseline to 8 weeks (intervention completion), with these reductions retained at 12-week follow-up (a 31.24% decrease) [12]. In addition to randomized controlled trials, current meta-analytic and systematic review data suggest medium effect sizes for DMH interventions for perceived stress from baseline to post intervention [4,5]. Beyond clinical outcomes, DMH interventions improve access to mental health care and provide individuals with effective, cost-effective care readily available via a mobile app or website.

While DMH interventions in clinical trials have been shown to be effective, less is known about their use in real-world settings. Real-world evidence builds upon clinical trials to improve our understanding of an intervention's efficacy in a person's daily functioning, providing data on the effectiveness and accessibility of DMH products as well as the external validity of interventions. As DMH interventions are delivered via a mobile device or computer to use within one's own environments, real-world evidence highlights intervention effectiveness outside of controlled settings. As such, there is an increasing need to understand the real-world effectiveness of widely used commercial apps and DMH interventions that help improve mental health outcomes for the overall population.

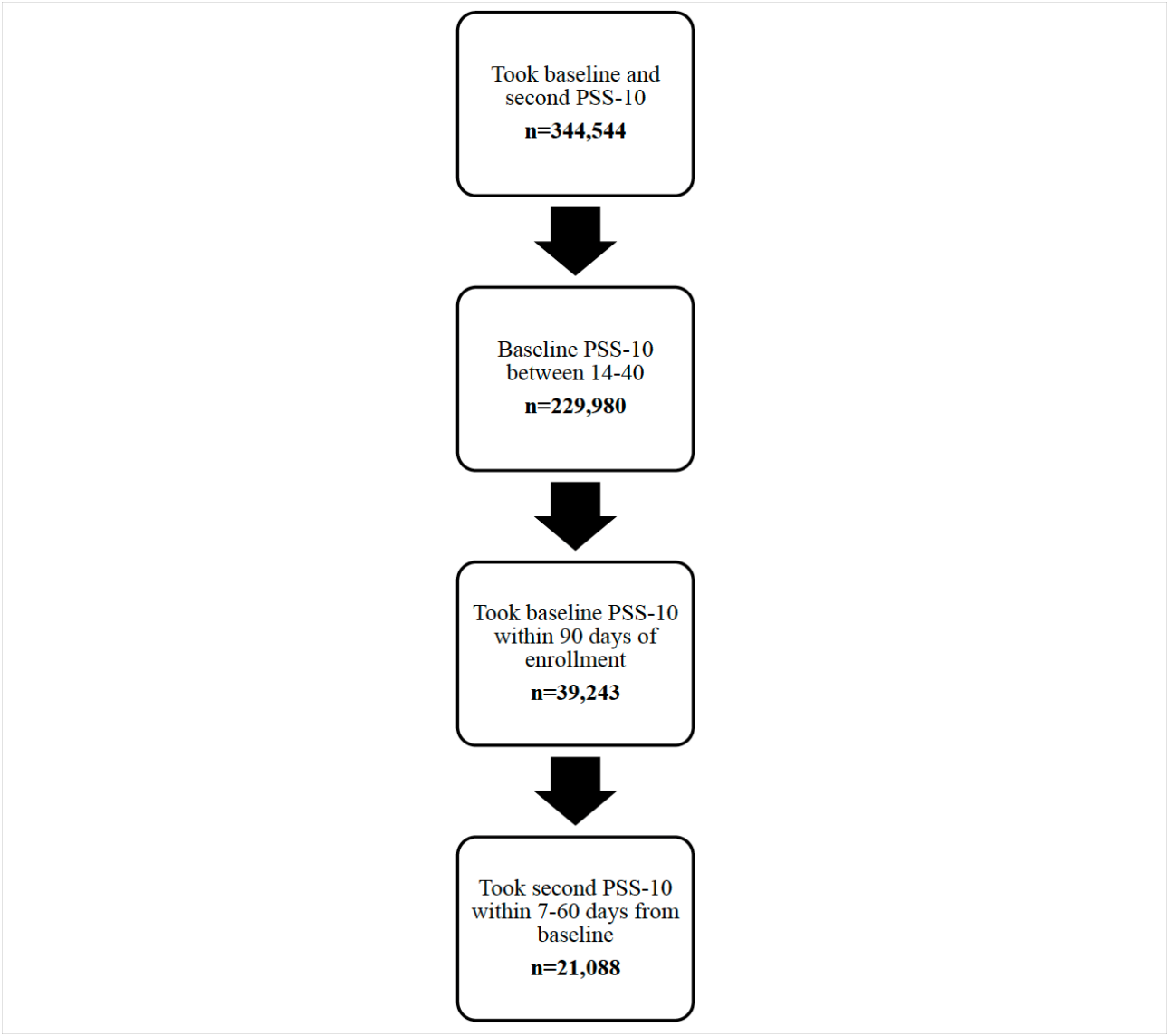
Headspace is a popular and evidence-based DMH platform with over 100 million downloads and supported by >50 published peer-reviewed studies. The Headspace app offers a range of services, most notably to this study, mindfulness and meditation-based content that teaches coping strategies to manage daily stressors. Previous clinical trials on the efficacy of the Headspace app show evidence of improved mindfulness, focus, stress, sleep quality, burnout, resilience, anxiety, depression, and quality of life. Given the reach and scale of Headspace's membership, there is an opportunity to better understand real-world outcomes beyond clinical trials. As such, this study aims to use real-world data to evaluate perceived stress among Headspace members. To accomplish this goal, this study's aims are to (1) characterize patterns in baseline perceived stress and changes in perceived stress among Headspace members with moderate and severe baseline perceived stress and (2) examine associations between engagement with Headspace content and changes in perceived stress (ie, to evaluate if there is a dose-response relationship).

Methods

Study Design and Participants

This real-world observational study examined perceived stress among Headspace members. This study followed the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) framework for reporting observational studies [13]. Individuals were included in this study if they enrolled on Headspace between March 2020 and January 2023, completed a baseline and follow-up perceived stress questionnaire (the Perceived Stress Scale [PSS-10]), completed the baseline PSS-10 within 90 days of enrollment, completed the follow-up PSS-10 at >7 and <60 days from the baseline, and reported moderate or severe perceived stress levels at baseline (Figure 1).

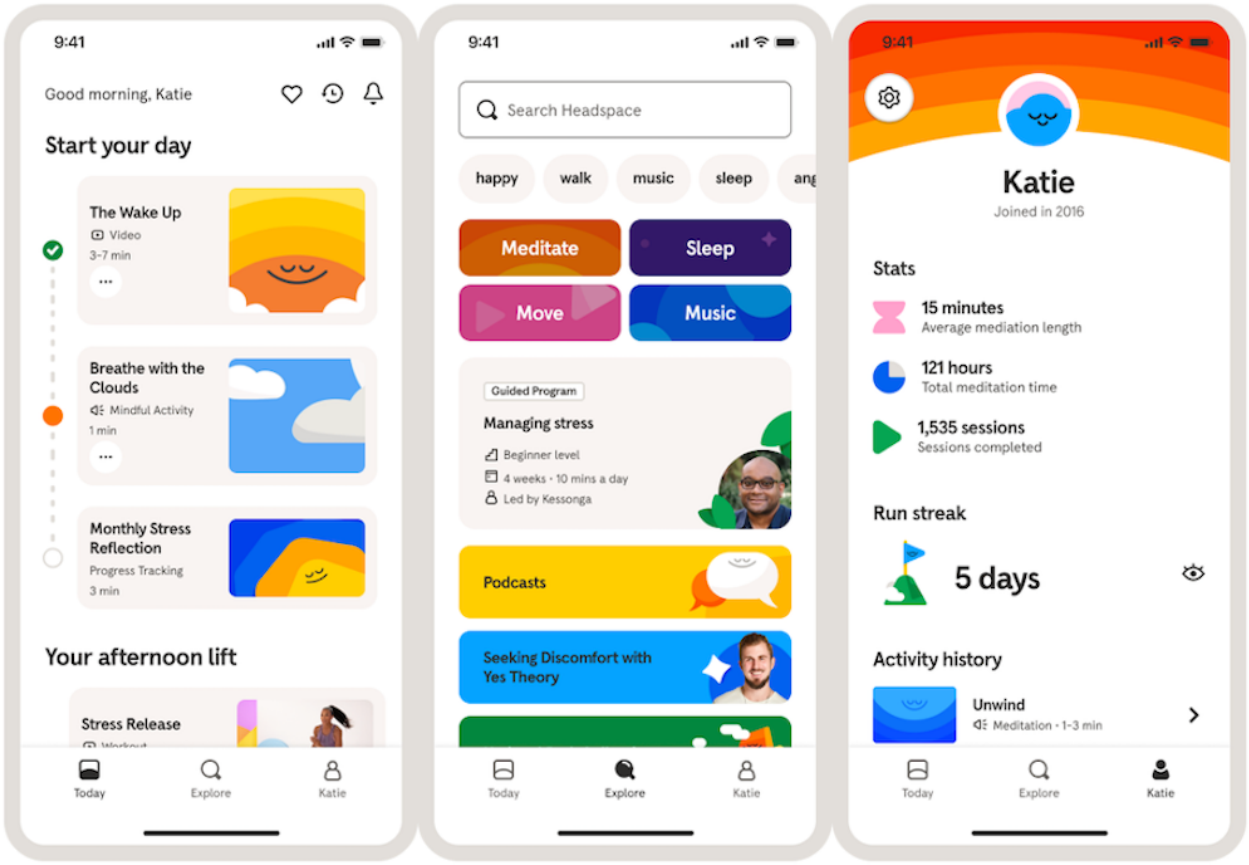
Figure 1. Sample size filter for the current study. PSS-10: Perceived Stress Scale.



The Headspace App

Study participants had access to all Headspace app offerings and used the app at their own discretion. The Headspace app includes evidence-based, expert-led content including guided mindfulness, meditation with animated guided videos, progressive muscle relaxation, psychoeducation, and gratitude exercises. Mindfulness content ranges from content for general self-care to that specific to mental health disorders such as anxiety, stress, and sleep disorders. When participants open the Headspace app, they are first directed to the "Today" tab, which contains personalized, daily content recommendations to encourage health habit formation throughout the day. Content on the "Today" tab includes a breathing exercise, educational video, and 3 meditations (1 each for the morning, afternoon, and evening). All other content can be found in the "Explore" tab, which includes a search bar; content organized into large categories of meditate, sleep, move, and music at the top; and as they scroll further, more specific categories and courses such as Beginning Meditation, Mindfulness at Work, Mindful Eating, etc. Finally, all participants have a personalized "Profile" tab, which tracks activity history and overall app statistics (minutes meditated, sessions completed, and days in a row of content engagement). Study participants had access to all Headspace app offerings and navigated the app to choose their own content (Figure 2).

Figure 2. The "Today," "Explore," and "Profile" tabs in the Headspace app shown from left to right.



Ethical Considerations

This study represents a secondary analysis of preexisting deidentified data [14]. The study team does not have access to the participants' identifying information and will not contact the participants. Therefore, in accordance with the US Department of Health and Human Services' regulations (45 CRF 46.104), this study is deemed exempt from institutional review board oversight and informed consent. Participants are notified through the Headspace terms of service that their deidentified data may be used for research.

Measures

Perceived stress was measured using the PSS-10 [15]. Participants were prompted to complete the PSS-10 through the Headspace app and answered 10 questions about their perceived feelings and thoughts regarding stress with a 5-point Likert scale (0=never to 4=very often). Scores range from 0 - 40, with higher scores indicating higher perceived stress. The scale has high test-retest reliability (Cronbach α=.85). PSS-10 scores were categorized to indicate low (<14), moderate (14-26), and high (27-40) levels of perceived stress [15]. Those with low levels of perceived stress at baseline were excluded from this study. Study participants completed a baseline and follow-up PSS-10. Total scores for each time point were reported, along with the raw change in PSS-10 score and percent change in PSS-10 score. Participants completed the PSS-10 within the Headspace app, and final scores were extracted directly from the app. Scores were calculated and extracted only for those who completed all questions on the PSS-10 (ie, incomplete data were excluded from this study).

Engagement was assessed using active days and minutes engaged with Headspace as well as the number of sessions started. Additionally, ratios for active days per week, active minutes per day, and sessions per week were calculated to further define engagement. Engagement data are tracked for all members within the Headspace app and were extracted directly from the app.

Statistical Analysis

Means, SDs, medians, and 95% CIs were computed for baseline and follow-up PSS-10 scores, raw change in PSS-10 score, percent change in PSS-10 score, days between PSS-10 assessments, active days, active days per week, active minutes, active minutes per day, sessions, and sessions per week. Frequencies and percentages were computed for baseline and follow-up PSS-10 scores (moderate vs high).

Mean percent change in PSS-10 scores was segmented out by active days per week, active minutes per day, and sessions per week to delineate the relationship between app engagement and percent change in PSS-10 score. Correlations between baseline and follow-up PSS-10 scores, percent change in PSS-10 score, days between PSS-10 assessments, active days, active days per week, active minutes, active minutes per day, sessions, and sessions per week were evaluated.

We used *t* tests to evaluate the differences in baseline PSS-10 scores, follow-up PSS-10 scores, percent change in PSS-10

score, active days, active days per week, active minutes, active minutes per day, sessions, and sessions per week between (1) those who saw and those who did not see improvements in PSS-10 scores (yes vs no improvement) and (2) those who saw a ≥30% improvement versus those who saw a <30% improvement. A 30% improvement threshold was chosen to coincide with a clinically meaningful change in perceived stress [16]. *P* values of <.05 were considered significant, and effect sizes were reported using the Cohen *d* to determine the magnitude of significance.

Results

Sample and Perceived Stress Descriptives

Overall, 344,544 Headspace members completed 2 PSS-10s between March 2020 and January 2023. Of them, 21,088 met this study's inclusion criteria and were included as study participants (Figure 1).

At baseline, 15,127 (71.73%) participants reported moderate levels of perceived stress and 5961 (28.27%) reported severe levels of perceived stress with a mean baseline PSS-10 score of 23.14 (SD 5.69). On average, members completed their follow-up PSS-10 33.25 (SD 11.17) days after their first PSS-10 with a mean follow-up PSS-10 score of 20.41 (SD 6.47). At follow-up, 2878 (13.56%) participants reported mild to moderate levels of perceived stress, 14,376 (68.17%) reported moderate levels, and 3834 (18.18%) reported high levels. On average, members saw a 23.52% decrease in PSS-10 scores from baseline to follow-up, and 13,692 (64.93%) participants saw a decrease in their PSS-10 score (Table 1).

Table . Descriptive statistics for engagement and primary outcomes.

Parameters		Mean (SD)	Median	95% CI
PSS-10[a]				
	Baseline	23.14 (5.69)	23	23.06 to 23.21
	Follow-up	20.41 (6.47)	20	20.33 to 20.50
	Raw change in the PSS-10 score	–2.72 (6.06)	–3	–2.80 to –2.64
	Percent change in the PSS-10 score	–23.52 (62.60)	–13.04	–24.37 to –22.68
	Days between PSS-10 assessments	33.25 (11.17)	31	33.10 to 33.40
Engagement metrics				
	Active days	18.04 (11.42)	17	17.89 to 18.20
	Active days per week	2.42 (1.76)	2.01	2.39 to 2.44
	Active minutes	547.4 (1004.81)	281.95	533.83 to 560.96
	Active minutes per day	25.89 (33.40)	16.68	25.44 to 26.34
	Sessions	49.73 (48.14)	35	49.08 to 50.38
	Sessions per week	7.11 (8.34)	4.33	7.00 to 7.23

[a]PSS-10: Perceived Stress Scale.

Engagement

On average, members engaged with Headspace content for 18.08 (SD 11.42) days and 547.40 (SD 1004.81) minutes and started 49.73 (SD 48.14) sessions. On average, members had 2.42 (SD 1.76) active days per week, 25.89 (SD 33.40) active minutes per day, and completed 7.11 (SD 8.34) sessions per week (Table 1). The percent change in PSS-10 score by active days per week, active minutes per day, and sessions per week are presented in Table 2, with data suggesting peak changes in PSS-10 scores at 7 active days per week, 11 - 15 active minutes per day, and 19 - 20 sessions per week.

Table . Mean percent change in PSS-10[a] scores by active days per week, active minutes per day, and sessions per week.

		Mean percent change in PSS-10 score
Active days per week		
	1	–21.11
	2	–21.55
	3	–23.48
	4	–28.94
	5	–23.69
	6	–32.61
	7	–43.24
Active minutes per day		
	0-5	–19.86
	6-10	–21.33
	11-15	–24.41
	16-20	–24.26
	21-25	–23.84
	26-30	–25.65
	>30	–21.33
Sessions per week		
	1-2	–17.18
	2-4	–21.75
	5-6	–25.88
	7-8	–23.25
	9-10	–27.56
	11-12	–27.89
	13-14	–29.64
	15-16	–32.21
	17-18	–35.94
	19-20	–41.71
	>20	–25.39

[a]PSS-10: Perceived Stress Scale.

Engagement and Perceived Stress

Correlations between engagement and perceived stress are reported in Table 3. The results of *t* tests investigating the association between engagement and perceived stress are reported in Table 4. Participants who demonstrated an improvement in perceived stress had significantly higher baseline PSS-10 scores ($t_{16,575}$=–40.08; *P*<.001; Cohen *d*=0.56) and significantly more active days ($t_{14,774}$=–9.00; *P*<.001; Cohen *d*=0.13), active days per week ($t_{17,152}$=–24.18; *P*<.001; Cohen *d*=0.33), sessions ($t_{16,515}$=–11.18; *P*<.001; Cohen *d*=0.16), and sessions per week ($t_{19,025}$=–20.53; *P*<.001; Cohen *d*=0.27) than those who did not demonstrate an improvement.

Participants with a ≥30% improvement in perceived stress had significantly higher baseline PSS-10 scores ($t_{12,794}$=–23.52 *P*<.001; Cohen *d*=0.35), more active days per week ($t_{11,401}$=–23.10; *P*<.001; Cohen *d*=0.36), and more sessions per week ($t_{10,150}$=–18.87; *P*<.001; Cohen *d*=0.31) than those with a <30% improvement. Although effect sizes suggest smaller relationships, *t* tests also suggest that participants with a ≥30% improvement in PSS-10 scores had significantly more active days ($t_{12,984}$=–12.02; *P*<.001; Cohen *d*=0.17), more active minutes ($t_{13,569}$=–2.61; *P*<.001; Cohen *d*=0.04), and more sessions ($t_{11,539}$=–12.68; *P*<.001; Cohen *d*=0.19) than those with a <30% improvement in PSS-10 score.

Table . Pearson correlations between PSS-10[a] scores and engagement outcomes.

	Baseline PSS-10 score	Follow-up PSS-10 score	Percent change in PSS-10 score	Days between PSS-10 assessments	Active days	Active days per week	Active minutes	Active minutes per day	Sessions	Sessions per week
Baseline PSS-10 score										
r	N/A[b]	0.51	–0.14	–0.03	–0.12	0.05	0.01	0.04	–0.03	0.08
P value	N/A	*<.001*[c]	*<.001*	*<.001*	*<.001*	*<.001*	.66	*<.001*	*<.001*	*<.001*
Follow-up PSS-10 score										
r	0.51	N/A	0.51	0.03	–0.17	–0.13	–0.01	0.03	–0.11	–0.09
P value	*<.001*	N/A	*<.001*	*<.001*	*<.001*	*<.001*	.05	*<.001*	*<.001*	*<.001*
Percent change in PSS-10 score										
r	–0.14	0.51	N/A	0.03	–0.05	–0.12	–0.01	–0.01	–0.07	–0.10
P value	*<.001*	*<.001*	N/A	*<.001*	*<.001*	*<.001*	.18	.97	*<.001*	*<.001*
Days between PSS-10 assessments										
r	–0.03	0.03	0.03	N/A	0.33	–0.11	0.12	–0.01	0.16	–0.13
P value	*<.001*	*<.001*	*<.001*	N/A	*<.001*	*<.001*	*<.001*	.12	*<.001*	*<.001*
Active days										
r	–0.12	–0.17	–0.05	0.33	N/A	0.63	0.46	0.21	0.73	0.43
P value	*<.001*	*<.001*	*<.001*	*<.001*	N/A	*<.001*	*<.001*	*<.001*	*<.001*	*<.001*
Active days per week										
r	0.05	–0.13	–0.12	–0.11	0.63	N/A	0.31	0.19	0.58	0.78
P value	*<.001*	*<.001*	*<.001*	*<.001*	*<.001*	N/A	*<.001*	*<.001*	*<.001*	*<.001*
Active minutes										
r	0.01	−0.01	−0.01	0.12	0.46	0.31	N/A	0.87	0.52	0.36
P value	.66	.05	.18	*<.001*	*<.001*	*<.001*	N/A	*<.001*	*<.001*	*<.001*
Active minutes per day										
r	0.04	0.03	−0.01	−0.01	0.21	0.19	0.87	N/A	0.35	0.30
P value	*<.001*	*<.001*	.97	.12	*<.001*	*<.001*	*<.001*	N/A	*<.001*	*<.001*
Sessions										
r	–0.03	–0.11	–0.07	0.16	0.73	0.58	0.52	0.35	N/A	0.78
P value	*<.001*	*<.001*	*<.001*	*<.001*	*<.001*	*<.001*	*<.001*	*<.001*	N/A	*<.001*
Sessions per week										
r	0.08	–0.09	–0.10	–0.13	0.43	0.78	0.36	0.30	0.78	N/A
P value	*<.001*	*<.001*	*<.001*	*<.001*	*<.001*	*<.001*	*<.001*	*<.001*	*<.001*	N/A

[a]PSS-10: Perceived Stress Scale.

[b]N/A: not applicable.

[c]Italicized values are significant at $P<.05$.

Table . Differences in PSS-10[a] scores and engagement metrics between participants who saw an improvement in their PSS-10 score and those who did not and between participants who saw a ≥30% improvement in the PSS-10 score and those who saw a <30% improvement in their PSS-10 score.

	PSS-10 score improvement, mean (SD)	No PSS-10 score improvement, mean (SD)	95% CI	*t* test (*df*)	*P* value	Cohen *d*	≥30% improvement in PSS-10 score, mean (SD)	<30% improvement in PSS-10 score, mean (SD)	95% CI	*t* test (*df*)	*P* value	Cohen *d*
Baseline PSS-10 score	24.22 (5.69)	21.14 (5.12)	–3.23 to –2.93	–40.08 (16,576)	*<.001* [b]	0.56	24.49 (5.59)	22.52 (5.63)	–2.12 to –1.80	–23.52 (14,364)	*<.001*	0.35
Follow-up PSS-10 score	18.16 (5.73)	24.60 (5.62)	6.28 to 6.60	78.87 (15,418)	*<.001*	1.13	15.13 (4.52)	22.81 (4.52)	7.54 to 7.83	104.68 (14,980)	*<.001*	1.42
Percent change in PSS-10 score	–43.42 (69.50)	13.32 (11.84)	55.55 to 57.94	93.07 (15,121)	*<.001*	1.01	–74.19 (90.34)	–0.53 (17.34)	71.45 to 75.85	65.60 (6682)	*<.001*	1.40
Total active days	18.57 (11.28)	17.07 (11.62)	–1.82 to –1.17	–9.00 (14,774)	*<.001*	0.13	19.43 (11.21)	17.41 (11.46)	–2.35 to –1.69	–12.02 (14,236)	*<.001*	0.17
Active days per week	2.62 (1.82)	2.04 (1.57)	–0.63 to –0.53	–24.18 (17,152)	*<.001*	0.33	2.85 (1.88)	2.22 (1.66)	–0.68 to –0.57	–23.10 (13,437)	*<.001*	0.36
Total active minutes	557.59 (969.71)	528.52 (1066.57)	–58.31 to 0.17	–1.95 (13,969)	.05	0.03	573.54 (956.53)	535.53 (1025.86)	–66.52 to −9.51	–2.61 (14,700)	*.01*	0.04
Active minutes per day	26.10 (32.33)	25.50 (35.29)	–1.56 to 0.38	–1.20 (14,062)	.23	0.02	26.25 (31.60)	25.73 (34.18)	–1.46 to 0.43	–1.07 (14,540)	.28	0.01
Total sessions	52.36 (49.64)	44.85 (44.85)	–8.83 to –6.20	–11.18 (16,515)	*<.001*	0.16	56.21 (51.62)	46.79 (46.18)	–10.87 to –7.96	–12.68 (13,549)	*<.001*	0.19
Sessions per week	7.90 (9.00)	5.65 (6.72)	–2.47 to –2.04	–20.53 (19,025)	*<.001*	0.27	8.88 (9.82)	6.31 (7.44)	–2.83 to –2.30	–18.87 (12,439)	*<.001*	0.31

[a]PSS-10: Perceived Stress Scale.

[b]Italicized values are significant at $P<.05$.

Discussion

This study focused on using real-world data from Headspace members to evaluate changes in perceived stress and its association with app engagement. Our findings suggest that participants experienced a significant reduction in perceived stress scores and those who used Headspace more frequently experienced greater reductions in perceived stress. This study builds on prior clinical trials and provides real-world evidence supporting the use of Headspace to improve perceived stress.

On average, participants experienced a 23.52% reduction in perceived stress scores from baseline to the follow-up assessment (ie, approximately a 30-day period). Furthermore, 64.93% of participants saw a reduction in their PSS-10 scores from baseline to follow-up, suggesting that Headspace improved perceived stress for a majority of individuals. Shifts in perceived stress levels also provide evidence supporting the use of Headspace, as 10% of members who reported high levels of perceived stress at baseline no longer met those criteria at follow-up (ie, reported moderate or mild levels of perceived stress at follow-up). These shifts in individual perceived stress are important for self-management behaviors (eg, mindfulness or self-regulation) that are associated with improved patient well-being and mental health outcomes [17]. Thus, it is likely that shifts in perceived stress might be an important predictor of patients' mental health outcomes. These findings are consistent with previous clinical research examining the impact of DMH interventions to improve perceived stress [10,18,19] and provide additional real-world evidence to support those claims.

We also identified relationships between app engagement and changes in perceived stress. Specifically, correlations suggest that active days per week and active sessions per week were associated with a higher percent change in PSS-10 scores. App-based mindfulness interventions of longer duration, such as those from 4 to 12 weeks, have shown improvements in not only mindfulness and perceived stress but also depression, anxiety, and overall well-being [11,20,21]. Few app-based mindfulness interventions are less than 4 weeks in duration, supporting the use of longer interventions. While this study revealed that a higher number of active days (total and per week) is associated with larger improvements in perceived stress, our

data did not reveal a relationship between active minutes (total and per week) and percent change in perceived stress. These data support the findings of a previous study comparing 10 and 30 minutes of daily mindfulness, which reported no difference in mindfulness or psychological distress between the dosage groups [22]. This study's findings, in addition to those of previous studies investigating mindfulness dosage, may suggest that active days are more important than total time spent practicing mindfulness or engaging with the app. Future research is necessary to further understand these relationships to more accurately suggest mindfulness dosage for clinical practice and research.

Identifying changes in perceived stress by active days per week, active minutes per day, and sessions per week provide more nuanced data informing the duration and dosage of mindfulness interventions. The largest percent change in perceived stress occurred with 7 active days per week; however, data show a sizable increase at 4 active days per week with a leveling off in those active for more days. We also note peak percent change in PSS-10 score at 11 - 15 active minutes per day and 19 - 20 sessions per week. These data indicate that consistent Headspace use results in more prominent improvements in perceived stress. Previous research has reported that the psychological status of approximately 25% of patients improved after 1 psychotherapy session, with steadier improvements occurring over 8 weeks [23]. This study's findings indicate a similar pattern that emphasizes a shift in perceived stress from baseline to the second administration of the PSS-10 during the first 3 weeks of engaging with the Headspace platform.

Higher engagement, specifically more active days and sessions, was significantly associated with a higher likelihood of improving stress (both overall and ≥30%), suggesting that frequent engagement with the Headspace app might be related to improved outcomes. This is an important finding given the breadth of research on increased patient engagement with psychotherapy and improved mental health outcomes [24]. In particular, the effect sizes for active days per week and sessions per week of –0.33 and 0.27 in the group that saw improvements in perceived stress and the group that did not see an improvement and 0.36 and 0.31 in the group that saw a ≥30% improvement in perceived stress and the group that saw a <30% improvement are notable. These results build on previously reported correlations, suggesting that higher engagement days and a higher number of sessions are associated with a higher percent change in PSS-10 scores and provide further support for the intervention dosage, suggesting multiple bouts of mindfulness each day [22]. Previous research highlights positive mental health outcomes for people who establish consistent health behaviors. In light of higher costs and access to mental health care, these preliminary findings suggest that consistently engaging with Headspace may decrease perceived stress, thus supporting the use of DMH in real-world settings to accessibly improve outcomes.

Strengths, Limitations, and Future Research

This study has several limitations and strengths. As this study was the first real-world study investigating the Headspace app, our data are largely focused on descriptive and group mean differences. While this information is important to establish overall benchmarks for this study, we are limited in the types of questions (eg, prospective) we can ask with this study's format and recognize that the current statistical approach does not allow for causal inferences. Additionally, these data did not include the demographic information of Headspace members; therefore, we were unable to investigate the impact of demographics on the study outcomes. Future research should examine longitudinal data and changes in perceived stress, while accounting for the demographic characteristics of Headspace members and engagement factors. As a real-world evidence study, we were unable to identify specific programming used by members; we focused only on overall engagement outcomes. Future studies should include more in-depth engagement outcomes to better understand how certain programming is used and impacts perceived stress. Finally, data for this study were collected during the COVID-19 pandemic (March 2020 and January 2023). As noted, stress greatly increased among individuals in the United States during this time, which may have impacted PSS-10 scores (ie, increased scores). However, members were only included if they completed their 2 PSS-10s between 7 and 60 days, and the average number of days between PSS-10 assessments was approximately 1 month, suggesting that if members completed their baseline PSS-10 at the beginning of the pandemic, when stress levels were higher nationwide, they would have completed their follow-up at a similar time period within the pandemic.

A primary strength of the study is the methodology incorporated to understand changes in perceived stress and engagement on the platform. The use of a real-world evidence methodology often helps researchers and clinicians to observe the feasibility and generalizability of interventions in daily functioning. Given the current sample size, which is over 20,000 participants, it is evident that the Headspace platform provides benefits to people who frequently engage with the platform. Furthermore, this study's large sample size also offers strong evidence for the generalizability of our findings in the real world. The large sample size coupled with the current methodology also allowed us to establish reliable findings for our current platform; therefore, establishing a foundational understanding of how much change is possible on the platform while understanding overall engagement trends.

Conclusions

Our findings suggest that members using Headspace experienced significant reductions in perceived stress in a real-world setting. Furthermore, data suggest that members who engaged with the platform more regularly were more likely to experience improvements in perceived stress. This study is the first to provide real-world evidence of the DMH Headspace platform aimed to reduce participants' perceived stress. Our results have implications for clinical practice, which include incorporating mediation and mental health psychoeducation as an adjunct to psychotherapy or as a preventive intervention to reduce stress.

Authors' Contributions

CC contributed to the study design, data analysis, manuscript writing, and manuscript review. JK contributed to the study design, manuscript writing, and manuscript review. EH contributed to the study design, data analysis, manuscript writing, and manuscript review. LT contributed to the data analysis and manuscript review. SK contributed to the study design, data analysis, manuscript writing, and manuscript review.

Conflicts of Interest

CC, LT, and SK are currently employed by Headspace. JK and EH were employed by Headspace at the time of data collection.

References

1. Stress in America 2020: a national mental health crisis. American Psychological Association. 2020. URL: https://www.apa.org/news/press/releases/stress/2020/report-october [accessed 2023-12-15]
2. Bethune S. Stress in America 2023: a nation recovering from collective trauma. American Psychological Association. 2023. URL: https://www.apa.org/news/press/releases/stress/2023/collective-trauma-recovery [accessed 2023-12-15]
3. Fleming T, Bavin L, Lucassen M, Stasiak K, Hopkins S, Merry S. Beyond the trial: systematic review of real-world uptake and engagement with digital self-help interventions for depression, low mood, or anxiety. J Med Internet Res 2018 Jun 6;20(6):e199. [doi: 10.2196/jmir.9275] [Medline: 29875089]
4. Gál É, tefan S, Cristea IA. The efficacy of mindfulness meditation apps in enhancing users' well-being and mental health related outcomes: a meta-analysis of randomized controlled trials. J Affect Disord 2021 Jan 15;279:131-142. [doi: 10.1016/j.jad.2020.09.134] [Medline: 33049431]
5. Jayewardene WP, Lohrmann DK, Erbe RG, Torabi MR. Effects of preventive online mindfulness interventions on stress and mindfulness: a meta-analysis of randomized controlled trials. Prev Med Rep 2017 Mar;5:150-159. [doi: 10.1016/j.pmedr.2016.11.013] [Medline: 28050336]
6. Hassard J, Teoh KRH, Visockaite G, Dewe P, Cox T. The cost of work-related stress to society: a systematic review. J Occup Health Psychol 2018 Jan;23(1):1-17. [doi: 10.1037/ocp0000069] [Medline: 28358567]
7. Flett JAM, Hayne H, Riordan BC, Thompson LM, Conner TS. Mobile mindfulness meditation: a randomised controlled trial of the effect of two popular apps on mental health. Mindfulness 2019 May;10(5):863-876. [doi: 10.1007/s12671-018-1050-9]
8. Schulte-Frankenfeld PM, Trautwein FM. App-based mindfulness meditation reduces perceived stress and improves self-regulation in working university students: a randomised controlled trial. Appl Psychol Health Well Being 2022 Nov;14(4):1151-1171. [doi: 10.1111/aphw.12328] [Medline: 34962055]
9. Wolever RQ, Bobinet KJ, McCabe K, et al. Effective and viable mind-body stress reduction in the workplace: a randomized controlled trial. J Occup Health Psychol 2012 Apr;17(2):246-258. [doi: 10.1037/a0027278] [Medline: 22352291]
10. Morledge TJ, Allexandre D, Fox E, et al. Feasibility of an online mindfulness program for stress management--a randomized, controlled trial. Ann Behav Med 2013 Oct;46(2):137-148. [doi: 10.1007/s12160-013-9490-x] [Medline: 23632913]
11. Krusche A, Cyhlarova E, King S, Williams JMG. Mindfulness online: a preliminary evaluation of the feasibility of a web-based mindfulness course and the impact on stress. BMJ Open 2012;2(3):e000803. [doi: 10.1136/bmjopen-2011-000803] [Medline: 22614170]
12. Huberty J, Green J, Glissmann C, Larkey L, Puzia M, Lee C. Efficacy of the mindfulness meditation mobile app "Calm" to reduce stress among college students: randomized controlled trial. JMIR Mhealth Uhealth 2019 Jun 25;7(6):e14273. [doi: 10.2196/14273] [Medline: 31237569]
13. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Ann Intern Med 2007 Oct 16;147(8):573-577. [doi: 10.7326/0003-4819-147-8-200710160-00010] [Medline: 17938396]
14. Exemptions (2018 requirements). US Department of Health and Human Services. 2021. URL: https://www.hhs.gov/ohrp/regulations-and-policy/regulations/45-cfr-46/common-rule-subpart-a-46104/index.html [accessed 2023-12-15]
15. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983 Dec;24(4):385-396. [Medline: 6668417]
16. Eskildsen A, Dalgaard VL, Nielsen KJ, et al. Cross-cultural adaptation and validation of the Danish consensus version of the 10-item Perceived Stress Scale. Scand J Work Environ Health 2015 Sep 1;41(5):486-490. [doi: 10.5271/sjweh.3510] [Medline: 26111225]
17. McAndrew LM, Crede M, Maestro K, Slotkin S, Kimber J, Phillips LA. Using the common-sense model to understand health outcomes for medically unexplained symptoms: a meta-analysis. Health Psychol Rev 2019 Dec;13(4):427-446. [doi: 10.1080/17437199.2018.1521730] [Medline: 30196755]
18. Leung C, Pei J, Hudec K, Shams F, Munthali R, Vigo D. The effects of nonclinician guidance on effectiveness and process outcomes in digital mental health interventions: systematic review and meta-analysis. J Med Internet Res 2022 Jun 15;24(6):e36004. [doi: 10.2196/36004] [Medline: 35511463]

19. Heber E, Lehr D, Ebert DD, Berking M, Riper H. Web-based and mobile stress management intervention for employees: a randomized controlled trial. J Med Internet Res 2016 Jan 27;18(1):e21. [doi: 10.2196/jmir.5112] [Medline: 26818683]
20. Truhlar LM, Durand C, Cooper MR, Goldsmith CA. Exploring the effects of a smartphone-based meditation app on stress, mindfulness, well-being, and resilience in pharmacy students. Am J Health Syst Pharm 2022 Nov 22;79(23):2159-2165. [doi: 10.1093/ajhp/zxac240] [Medline: 36070430]
21. Komariah M, Ibrahim K, Pahria T, Rahayuwati L, Somantri I. Effect of mindfulness breathing meditation on depression, anxiety, and stress: a randomized controlled trial among university students. Healthcare (Basel) 2022 Dec 22;11(1):26. [doi: 10.3390/healthcare11010026] [Medline: 36611488]
22. Fincham GW, Mavor K, Dritschel B. Effects of mindfulness meditation duration and type on well-being: an online dose-ranging randomized controlled trial. Mindfulness (N Y) 2023;14(5):1171-1182. [doi: 10.1007/s12671-023-02119-2] [Medline: 37304656]
23. Kadera SW, Lambert MJ, Andrews AA. How much therapy is really enough? A session-by-session analysis of the psychotherapy dose-effect relationship. J Psychother Pract Res 1996;5(2):132-151. [Medline: 22700273]
24. Gold C, Solli HP, Krüger V, Lie SA. Dose-response relationship in music therapy for people with serious mental disorders: systematic review and meta-analysis. Clin Psychol Rev 2009 Apr;29(3):193-207. [doi: 10.1016/j.cpr.2009.01.001] [Medline: 19269725]

Abbreviations

DMH: digital mental health
PSS-10: Perceived Stress Scale
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

Edited by L Buis; submitted 20.09.23; peer-reviewed by C Leung, T Pace; revised version received 18.12.23; accepted 22.12.23; published 04.03.24.

Please cite as:
Callahan C, Kimber J, Hu E, Tanner L, Kunkle S
The Real-World Impact of App-Based Mindfulness on Headspace Members With Moderate and Severe Perceived Stress: Observational Study
JMIR Mhealth Uhealth 2024;12:e52968
URL: https://mhealth.jmir.org/2024/1/e52968
doi:10.2196/52968

Identifying Preferred Appearance and Functional Requirements of Aged Care Robots Among Older Chinese Immigrants: Cross-Sectional Study

Ching-Ju Chiu[1], PhD; Yi-Hsuan Lo[2], BSc; Jed Montayre[3,4], PhD; Hammoda Abu-Odah[3,4], PhD; Mei-Lan Chen[5], PhD; Ivy Yan Zhao[3,4], PhD

[1]Institute of Gerontology, College of Medicine, National Cheng Kung University, Tainan, Taiwan

[2]Department of Statistics, College of Management, National Cheng Kung University, Tainan, Taiwan

[3]World Health Organization Collaborating Centre for Community Health Services, School of Nursing, The Hong Kong Polytechnic University, Hong Kong, China

[4]School of Nursing, The Hong Kong Polytechnic University, Hong Kong, China

[5]School of Nursing, Byrdine F Lewis College of Nursing and Health Professions, Georgia State University, Atlanta, GA, United States

Corresponding Author:
Ivy Yan Zhao, PhD
World Health Organization Collaborating Centre for Community Health Services, School of Nursing, The Hong Kong Polytechnic University, , Hong Kong, , China

Abstract

Background: Older Chinese immigrants constitute the largest older Asian ethnic population in New Zealand. Aging in a foreign land can be complex, presenting increasing challenges for gerontology scholars, practitioners, and policy makers. Older Chinese immigrants are more susceptible to experiencing loneliness and social isolation compared to native older people, primarily due to language, transportation, and cultural barriers. These factors subsequently impact their physical and mental health. With advancements in robotic technology, aged care robots are being applied to support older people with their daily living needs. However, studies on using robots with older immigrants living in the community are sparse. Their preferences for the appearance and function of aged care robots are unclear, which impacts the acceptance and usability of robots, highlighting the need for a user-centered design approach.

Objective: This study aims to explore older Chinese immigrants' needs and preferences toward the appearance and function of aged care robots and to examine their relationships with the demographic characteristics of participants.

Methods: A cross-sectional design was used in this study, which was undertaken between March and May 2020. A total of 103 participants completed a web-based survey.

Results: The average age of participants was 68.7 (SD 5.5) years. The results suggest that 41.7% (n=43) of the 103 participants preferred a humanlike adult appearance, while 32% (n=33) suggested an animallike appearance. These participants reported higher scores in both rigorousness and friendliness compared to others who preferred different robot appearances. Participants expressed a greater preference for the functions of housework assistance (n=86, 83.5%), language translation (n=79, 76.7%), health monitoring (n=78, 75.7%), facial expressions (n=77, 74.8%), news reading (n=66, 64.1%), and security monitoring (n=65, 63.1%). These preferences were found to be significantly associated with marital status, financial status, and duration of immigration.

Conclusions: To support immigrant populations to age well in a foreign country and address the growing shortage of health and social professionals, it is important to develop reliable robotic technology services that are tailored based on the needs and preferences of individuals. We collected and compared the perspectives of immigrant and nonimmigrant participants on using robots to support aging in place. The results on users' needs and preferences inform robotic technology services, indicating a need to prioritize older Chinese immigrants' preference toward aged care robots that perform housework assistance, language translation, and health and safety monitoring, and robots with humanlike features.

(JMIR Aging 2023;6:e48646) doi:10.2196/48646

KEYWORDS

robotic technology services; appearance; function; aged care; immigrant; Chinese; robot; robots; robotic; robotics; older adults; elderly; preference; cross sectional; cross-sectional; survey; healthy aging; aging in place; social; isolation; companion; companionship; Asian; Asian population; population; population studies; aging

Introduction

Older Chinese immigrants are more prone to experiencing loneliness and social isolation than native older people due to language, transportation, and cultural barriers, which subsequently impact their physical and mental health [1]. In 2017, 11 million people emigrated from China to their destination countries [2]. New Zealand is one of the most popular host countries for Chinese immigrants, and as a result, it faces the growing needs of an increasingly diverse aging population. According to the New Zealand Census results in 2018, there were 247,770 Chinese immigrants 45 years or older, with 23,625 (9.5%) of them 65 years or older [3]. Compared to the statistics from the 2013 census, there was an increase of 76,359 (44.5%) Chinese immigrants [4]. Older Chinese immigrants are the largest older Asian ethnic population in New Zealand [3].

Aging in a foreign land can be complex, posing increasing challenges for gerontology scholars, practitioners, and policy makers. Self-supported aging in place has been reported as a benefit for enhancing older people's health and quality of life, as it supports the continuity of the environment and promotes independent living within the community [5]. However, a large and rapidly growing social and health workforce shortage in New Zealand has been unable to meet the increasing needs of older Chinese immigrants to access social and care services in the community [6]. With innovations and advancements in computer systems, robotic technology and information and communication technologies have been applied to support older people with their daily living needs.

In a New Zealand study, a daily care robot was used at home to assist older community-dwelling adults who had different aging-related health needs. The robot's purpose was to remind them of daily activities, and it showed promising potential in old age care, especially in providing reminders for taking medication [5]. Most existing studies on using robots to support older people have focused on dementia care and cognitive training, and have been undertaken in dementia care units. For example, a recent study in Italy reported that a humanoid robot called NAO effectively supported memory training among 24 patients with mild cognitive impairment, enhancing their therapeutic compliance and reducing symptoms of depression [7]. The study reported significant changes in prose memory and verbal fluency measures [7]. Additionally, the social robot Paro (an animal seal robot) was tested in Taiwan with 20 older adults in a long-term care facility for 8 weeks using a single-group pre-post quasi-experimental design and showed a statistical decrease in depression and loneliness and an increase in quality of life among the participants [8]. Our pilot study in Hong Kong has reported good feasibility and acceptance of using a humanoid social robot called KaKa among older Chinese adults and their family caregivers in their homes [9]. However, studies of using robots among older immigrants living in the community are sparse. Their preferences for the appearance and function of aged care robots are unclear, which impact the acceptance and usability of robots, and therefore, a user-centered design is required. It is imperative to understand users' needs and preferences before designing and developing a robot to meet their specific care needs [10].

Aged care robots, including health care assistive robots and socially assistive robots, should be tailored for older people to be easy to use, flexible, and able to support natural older people–robot interaction [10]. In particular, the design should consider those people with less experience in using technological devices [10]. Gaseiger and colleagues [5] reported that older people living alone at home accepted a robot as a companion, and the functions of an aged care robot should be more personalized to meet older people's health and social needs. A cross-sectional survey among middle-aged and older Taiwanese living in the community revealed that female participants preferred an animallike robot, while male participants favored a humanlike robot [11]. The most popular functions of a robot recommended by those participants included dancing, singing, storytelling, and news reading [11]. Moreover, the New Zealand study indicated that older Chinese immigrants were more likely to accept the companionship of robots when they were feeling lonely, yet more evidence is needed regarding their preferences for robot features [12]. Therefore, this study aimed to explore older Chinese immigrants' needs and preferences toward the appearance and function of aged care robots. Additionally, it sought to examine the relationships between these needs and preferences and the demographic characteristics of participants.

Methods

Participants

Adults 60 years or older, as defined by the World Health Organization [13], who self-identified their ethnicity as Chinese; held a permanent resident visa in New Zealand or were New Zealand citizens; were able to read, write, and understand traditional or simplified Chinese; were able to access the internet; and had completed the web-based survey were eligible to participate. There were no specific exclusion criteria. We screened each participant's eligibility through their individual demographic information.

Instrument

We used a web-based survey design tool, SurveyCake, to create a structured anonymous survey written in the traditional Chinese language. A simplified Chinese version was also made available as an alternative option. At the beginning of the survey, there was a 5-minute video using images from various online sources that was made for research purposes only. The video introduced different types of aged care robots, including health care assistive robots and socially assistive robots, each with a variety of appearances and features. The content of the video was presented in Mandarin with traditional Chinese subtitles, providing participants with a general idea about the types of robots and their capabilities. Following the video, participants were asked to complete the survey, which included four sections: (1) demographic information, (2) five personality traits, (3) eHealth literacy, and (4) preference for robot appearance and functions. Screenshots of the video are shown in 1-3.

Figure 1. Screenshot of a video depicting an older adult chatting with a robot.

Figure 2. Screenshot of a video depicting a robot's companionship role for an older lonely adult.

Figure 3. Screenshot of a video depicting a humanlike infant robot.

Participants were asked 18 questions regarding demographic information, including age, gender, education level, etc. The 15-item International Personality Item Pool, Five Personality Scale (extroversion, friendliness, rigorousness, emotional stability, and intelligence/imagination) was used [14]. The eHealth literacy of participants (ie, the internet use and search skills, ability to find reliable web-based content, and confidence in their abilities to search the internet) was assessed by the eHealth Literacy Scale [15]. Participants' preferred or favorite appearance (ie, animallike, humanlike infant, humanlike adult, or another form) and functions of the robots (eg, assisting with housework, health monitoring, and instant language translation) were collected. The survey was developed based on our previous study in Taiwan on middle-aged and older Chinese adults [11].

Ethical Considerations

The study protocol was approved by the institutional review board of the affiliated university (No. A-ER-105-509). Before participating in the web-based survey, each participant had to provide informed consent. They were introduced to the aims and content of the study, potential risks and benefits, and the right to withdraw from the study at any time. Each participant was asked to click a box to confirm their willingness to proceed. All respondents in this study have ticked this box, and their responses were anonymous.

Procedure

The study was undertaken between March and May 2020 using a cross-sectional design. The recruitment was supported by a local social service organization for older people in Auckland and several Chinese community groups from different regions of Auckland. Following the ethics approval, an electronic version of the flyer (in both traditional and simplified Chinese), advertising the study and linking to the web-based survey, was circulated by social workers and community group leaders via word of mouth, WeChat groups, and WhatsApp groups. No incentives were offered for participation in this web-based survey study. The completeness check was done by two team members after the questionnaire had been submitted.

Data Analysis

Descriptive analyses and inferential statistics were performed on R (version 4.1.1; R Foundation for Statistical Computing). Descriptive statistics were used for analyzing the demographic information, five personalities, eHealth literacy, and preference for robot appearance and functions. For each item, we calculated descriptive statistics as appropriate (eg, mean and SD or frequency and percentage). ANOVA and χ^2 test were adopted to analyze the correlations between demographic factors and preferences for robot functions.

Results

Demographic Information

A total of 103 older Chinese immigrants completed the survey, resulting in a response rate of 89.6% (103/115). Among the participants, the minimum age was 60 years, and the maximum age was 87 years, with an average age of 68.7 (SD 5.5) years. Of the participants, 74 were female and 29 were male. Most of the 103 participants attained a bachelor's degree or above (n=72, 70%), and had a good self-reported financial status (mean 22.7, SD 4.7). Most of the participants were married or had a partner (n=88, 85.4%) and lived with family or others (n=90, 87.4%). Most of them immigrated to New Zealand to reunite with family (n=83, 80.6%), and about half of them had lived in this country for more than 10 years. Participants reported higher scores in

rigorousness and friendliness than other personalities. Details on the data distribution are shown in Table 1.

Table . Demographic information of participants (N=103).

Demographic variables		Values
Age (years), mean (SD)		68.7 (5.5)
Gender, n (%)		
	Female	74 (71.8)
	Male	29 (28.2)
Level of education, n (%)		
	Under bachelor's degree	31 (30.1)
	Bachelor's degree or above	72 (69.9)
Marital status, n (%)		
	Married or have a partner	88 (85.4)
	Unmarried, widowed, or no partner	15 (14.6)
Live alone, n (%)		
	Yes	13 (12.6)
	No	90 (87.4)
Type of occupation, n (%)		
	Nontechnical	15 (14.6)
	Semitechnical/technical	29 (28.2)
	Professional/management	52 (50.5)
	Others	7 (6.8)
Whether or not retired, n (%)		
	Retired	95 (92.2)
	Employed	8 (7.8)
Reasons for immigration[a], n (%)		
	Job opportunity	7 (6.8)
	Family reunion	83 (80.6)
	Retirement	18 (17.5)
Duration of immigration (years), n (%)		
	<10	51 (49.5)
	≥10	52 (50.5)
Original living place, n (%)		
	Mainland of China	74 (71.8)
	Taiwan	16 (15.5)
	Hong Kong and Macau	9 (8.7)
	Other Asian countries or regions	4 (3.9)
Self-rated financial status (range 1-5), mean (SD)		3.6 (0.9)
Five personality scale (range 0-15), mean (SD)		
	Extroversion	10.3 (2.1)
	Friendliness	11 (2.1)
	Rigorousness	11.1 (2.1)
	Emotional stability	10.5 (2.4)
	Intelligence/imagination	9.7 (1.8)
eHealth literacy (range 8-40), mean (SD)		22.7 (4.7)

[a]Participants could provide multiple answers to this question.

Participants' Preferences for Appearance of Aged Care Robots

Table 2 shows that most of the 103 participants, both female and male, preferred a humanlike adult appearance (n=43, 41.7%), and their second preference was an animallike appearance (n=33, 32%). The remaining participants reported their preferences for a humanlike infant appearance (n=20, 19.4%) and other forms (n=6, 5.8%). Participants who preferred humanlike adult or infant and animallike appearances reported higher scores in rigorousness and friendliness than other personalities. Participants who desired other appearances, rather than animal and humanlike appearances, were more likely to report high scores in extroversion, friendliness, rigorousness, and emotional stability as well as in eHealth literacy. Participants who were married or had a partner were more likely to choose humanlike adult and animallike appearances.

Table . Participants' preference for the appearance of a robot (N=103).

		Preference for robot's appearance[a]			
		Animallike	Humanlike infant	Humanlike adult	Other forms
Participants, n (%)		33 (32.0)	20 (19.4)	43 (41.7)	6 (5.8)
Gender, n (%)					
	Female (n=74)	25 (33.8)	17 (23.0)	28 (37.8)	4 (5.4)
	Male (n=29)	8 (27.6)	3 (10.3)	15 (51.7)	2 (6.9)
Personality (range 0-15), mean (SD)					
	Extroversion	10.5 (2.2)	10.8 (2.0)	9.9 (2.2)	12.3 (2.0)
	Friendliness	11.2 (2.4)	11.3 (1.7)	11.0 (2.0)	12.3 (2.0)
	Rigorousness	11.4 (2.1)	11.1 (1.9)	11.2 (2.1)	12.5 (1.9)
	Emotional stability	10.7 (2.6)	10.9 (2.5)	10.4 (2.2)	12.3 (2.4)
	Intelligence/imagination	10.1 (2.0)	9.6 (1.5)	9.9 (1.7)	9.8 (1.5)
eHealth literacy (range 8-40), mean (SD)		23.6 (3.1)	21.7 (4.1)	23.6 (4.3)	25.3 (3.3)
Marital status, n (%)					
	Married or have a partner (n=88)	32 (36.4)	17 (19.3)	40 (45.5)	6 (6.8)
	Unmarried, widowed, or no partner (n=15)	1 (6.7)	3 (20.0)	3 (20.0)	0 (0.0)

[a]Participants could provide multiple answers to this question.

Participants' Most Preferred Functions of Aged Care Robots

Participants' most preferred functions of aged care robots were housework assistance, language translation, health monitoring, facial expressions, news reading, and security monitoring. By analyzing the correlations between the demographic factors and the preferred functions of a robot, participants who were married or had a partner were more likely to choose functions of facial expressions (71/88, 81%; *P*=.002), news reading (61/88, 69%; *P*=.02), and security monitoring (60/88, 68%; *P*=.02) than their counterparts. High financial status showed a significant correlation with the function of facial expressions (mean 3.7, SD 0.7; *P*=.001). Participants who had immigrated for fewer than 10 years were more likely to use the function of facial expressions (43/51, 84%; *P*=.05) than people who had immigrated for 10 years or more. Detailed information is shown in Table 3.

Table . Participants' most preferred six functions of robots (N=103).[a]

		Preference for robot functions[b]					
		Housework assistance	Language translation	Health monitor	Facial expressions	News reading	Security monitor
Participants, n (%)		86 (83.5)	79 (76.7)	78 (75.7)	77 (74.8)	66 (64.1)	65 (63.1)
Marital status, n (%)							
	Married or have partner (n=88)	72 (81.8)	70 (79.5)	69 (78.4)	71 (80.7)	61 (69.3)	60 (68.2)
	Unmarried, widowed, or no partner (n=15)	14 (93.3)	9 (60.0)	9 (60.0)	6 (40.0)	5 (33.3)	5 (33.3)
	P value	.46	.19	.23	*.002*	*.02*	*.02*
Financial status							
	Mean (SD)	3.6 (0.9)	3.6 (0.8)	3.6 (0.8)	3.7 (0.7)	3.6 (0.8)	3.6 (0.8)
	P value	.86	.49	.41	*.001*	.50	.42
Duration of immigration (years), n (%)							
	<10 (n=51)	41 (80.4)	41 (80.4)	41 (80.4)	43 (84.3)	34 (66.7)	32 (62.7)
	≥10 (n=52)	45 (86.5)	38 (73.1)	37 (71.2)	34 (65.4)	32 (61.5)	33 (63.5)
	P value	.57	.52	.39	.05	.74	>.99

[a]Statistically significant results ($P<.05$) are in italics.

[b]Participants could provide multiple answers to this question.

Discussion

Principal Findings

With the advancement of robotic technology, care robots are being used to assist older adults with their daily living needs. However, there is a lack of research on the use of robots with older immigrants residing in the community and their preferences regarding the appearance and functionality of aged care robots. Therefore, this study aimed to investigate the needs and preferences of older Chinese immigrants toward the appearance and function of aged care robots, and examine the relationship between these preferences and the demographic characteristics of the participants. The findings revealed that older Chinese immigrants favored a humanlike adult appearance for the robots. Additionally, participants with different marital status, financial status, and duration of immigration had varying needs for the robot's functionalities.

Robot Appearance

In this study, the most popular appearance of a robot rated by older Chinese immigrants were those with a humanlike adult appearance, where no difference was identified between female and male participants. Our findings are consistent with Chiu et al's [11] study that the favorite appearance of robots among middle-aged and older Taiwanese people was a humanlike adult appearance, but the correlation to marital status was less significant. Prakash and Rogers [16] reported distinctive differences in preferences for a robot appearance between young and older adults, whereas older adults had a higher preference for humanlike robots. The authors suggested that the differences might be attributed to their experiences with robots [16], and older adults felt comfortable talking with robots with human traits such as eyes and a mouth [17]. Moreover, a rural study in China reported that older adults perceived small-sized robots as more friendly, and steel machinelike robots were less preferred [18].

According to the research on human-robot interaction, the appearance and morphology of a robot are known to be important in increasing the acceptance and use of and interaction with a robot among older adults [19]. However, there is less evidence showing that robots were developed based on the characteristics of older adults. This study uniquely found that older Chinese immigrants who preferred humanlike adult or infant appearances reported higher scores in rigorousness and friendliness. This knowledge may help inform future robot designs for older adults from a morphological perspective [19].

User Needs and Preferences

Results of the study showed that participants with different marital status, financial status, and duration of immigration had different needs for robot functions. Designing robot services to support older people must be based on individually collected information [20]. In this study, older Chinese immigrants mostly desired family service functions (ie, housework assistance), language translation, health monitoring, facial expressions, news reading, and security monitoring. The findings were different from the reported results among middle-aged and older Taiwanese, where the most preferred functions were the skill and recreation functions, followed by family services (ie, housework) and then health status monitoring [11]. The difference might be explained by the immigrant context of individual circumstances and the purpose of immigration. In

this study, over 80% of the participants immigrated to reunite with their adult children, which is aligned with Zhao et al's [21] study that found that most Chinese late-life immigrants relocated to New Zealand to share house chores with their adult children or look after their grandchildren. The burden of housework adversely impacted their health and became a risk factor for their experiences of loneliness and social isolation [1,12,21]. The same issue was also observed among other Asian immigrant groups due to the value of filial piety [22]. Assistive functions of robots were required by participants to relieve their workload and address the language barrier in a host country.

The evidence of this study found that the safety- and health-monitoring functions of the robot were regarded as essential for participants to meet their health and well-being needs, and our data supports Chiu et al's [11] findings that different ages were significantly related to the preference for the safety-monitoring function of the robot. Most of our participants lived independently and expected to maintain their independence, which is consistent with Park et al's [20] study that early detection of emergencies by using robot technology to assist with community-dwelling older adults' daily living is necessary. Moreover, living with others was significantly correlated to participants' preference for the health-monitoring robot function, which might be justified by several studies in New Zealand on Asian immigrant groups that have barriers to access health services due to their language, culture, and transportation barriers, and they intended to stay healthy and avoid becoming a burden to their family [1,21,23].

Moreover, participants who were married or had a partner, had good financial status, and had immigrated fewer than 10 years ago were more likely to choose the function of facial expressions. The finding is consistent with the previous study that older adults with lower technology acceptance preferred friendly and familiar robot designs with humanlike facial features [19]. However, the needs of those who were single with lower education levels might be underreported in this study, as the majority of our participants had higher education levels and were married or had a partner.

Limitations and Future Work

This study recognizes several limitations. First, due to the web-based survey, we might have excluded potential participants without access to the internet, computer, cellphone, etc, or who were not able to answer the survey on the web. Second, it is possible that the responses of participants were biased due to the self-reported data. Third, the data collection was undertaken in Auckland. The generalizability of the study results for the whole of New Zealand and other destination countries is limited. In the future, larger representative samples are needed to further investigate the needs and preferences of using robots in the later phase of life and to generalize the relationships between demographic factors, characteristics of older adults, and preferences for robots among immigrant populations. Mixed methods and co-design research methods are recommended to gain in-depth insights into end users' needs and preferences for robots to support their functions and independence in old age.

Conclusion

To support immigrant populations to age well in a foreign country and to fill the gaps of increasing shortages in the health and social workforce, it is important to develop reliable robotic technology services that are tailored based on the needs and preferences of individuals. We collected and compared the opinions on using robots to support aging in place among immigrant and nonimmigrant groups. The results of users' needs and preferences would inform robotic technology services to prioritize older Chinese immigrants' preference toward housework assistance, language translation, health and safety monitoring, and robots with humanlike features.

Data Availability

The data sets generated or analyzed during this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

None declared.

References

1. Zhao IY, Holroyd E, Garrett N, Wright-St Clair VA, Neville S. Chinese late-life immigrants' loneliness and social isolation in host countries: an integrative review. J Clin Nurs 2023 May;32(9-10):1615-1624. [doi: 10.1111/jocn.16134] [Medline: 37083151]
2. Department of Economic and Social Affairs. Migration and population change - drivers and impacts. United Nations. 2017 Dec. URL: www.un.org/en/development/desa/population/migration/publications/populationfacts/docs/MigrationPopFacts20178.pdf [accessed 2023-10-30]
3. 2018 census totals by topic – national highlights (updated). Stats NZ. 2020 Apr 30. URL: www.stats.govt.nz/information-releases/2018-census-totals-by-topic-national-highlights [accessed 2023-10-30]
4. 2013 census ethnic group profiles. Stats NZ. 2014. URL: www.stats.govt.nz/tools/2013-census-ethnic-group-profiles/ [accessed 2021-08-13]
5. Gasteiger N, Ahn HS, Fok C, Lim J, Lee C, MacDonald BA, et al. Older adults' experiences and perceptions of living with Bomy, an assistive dailycare robot: a qualitative study. Assist Technol 2022 Jul 4;34(4):487-497. [doi: 10.1080/10400435.2021.1877210] [Medline: 33544067]

6. Goodyear-Smith F, Ashton T. New Zealand health system: universalism struggles with persisting inequities. Lancet 2019 Aug 3;394(10196):432-442. [doi: 10.1016/S0140-6736(19)31238-3] [Medline: 31379334]
7. Pino O, Palestra G, Trevino R, De Carolis B. The humanoid robot NAO as trainer in a memory program for elderly people with mild cognitive impairment. Int J Soc Robotics 2020 Jan;12(1):21-33. [doi: 10.1007/s12369-019-00533-y]
8. Chen SC, Moyle W, Jones C, Petsky H. A social robot intervention on depression, loneliness, and quality of life for Taiwanese older adults in long-term care. Int Psychogeriatr 2020 Aug;32(8):981-991. [doi: 10.1017/S1041610220000459] [Medline: 32284080]
9. Leung AYM, Zhao IY, Lin S, Lau TK. Exploring the presence of humanoid social robots at home and capturing human-robot interactions with older adults: experiences from four case studies. Healthcare (Basel) 2022 Dec 22;11(1):39. [doi: 10.3390/healthcare11010039] [Medline: 36611499]
10. Di Nuovo A, Broz F, Wang N, Belpaeme T, Cangelosi A, Jones R, et al. The multi-modal interface of robot-era multi-robot services tailored for the elderly. Intel Serv Robotics 2018 Jan;11(1):109-126. [doi: 10.1007/s11370-017-0237-6]
11. Chiu CJ, Hsieh S, Li CW. Needs and preferences of middle-aged and older adults in Taiwan for companion robots and pets: survey study. J Med Internet Res 2021 Jun 11;23(6):e23471. [doi: 10.2196/23471] [Medline: 34347621]
12. Chiu CJ, Lo YH, Ho MH, Montayre J, Zhao IY. Association between loneliness and acceptance of using robots and pets as companions among older Chinese immigrants during the COVID-19 pandemic. Australas J Ageing 2022 Sep;41(3):414-423. [doi: 10.1111/ajag.13075] [Medline: 35438833]
13. World report on ageing and health. World Health Organization. 2015. URL: apps.who.int/iris/bitstream/handle/10665/186463/9789240694811_eng.pdf [accessed 2023-10-30]
14. Li R, Chen Y. The development of a shortened version of IPIP Big Five Personality Scale and the testing of its measurement invariance between middle-aged and older people. J Educ Res Dev 2016;12(4):87-119.
15. Chung SY, Nahm ES. Testing reliability and validity of the eHealth Literacy Scale (eHEALS) for older adults recruited online. Comput Inform Nurs 2015 Apr;33(4):150-156. [doi: 10.1097/CIN.0000000000000146] [Medline: 25783223]
16. Prakash A, Rogers WA. Why some humanoid faces are perceived more positively than others: effects of human-likeness and task. Int J Soc Robot 2015 Apr 1;7(2):309-331. [doi: 10.1007/s12369-014-0269-4] [Medline: 26294936]
17. Lazar A, Thompson HJ, Piper AM, Demiris G. Rethinking the design of robotic pets for older adults. Presented at: DIS '16: Proceedings of the 2016 ACM Conference on Designing Interactive Systems; Jun 4-8, 2016; Brisbane, QLD, Australia p. 1034-1046. [doi: 10.1145/2901790.2901811]
18. Liu SX, Shen Q, Hancock J. Can a social robot be too warm or too competent? Older Chinese adults' perceptions of social robots and vulnerabilities. Comput Hum Behav 2021 Dec;125(106942):106942. [doi: 10.1016/j.chb.2021.106942]
19. Oh S, Oh YH, Ju DY. Understanding the preference of the elderly for companion robot design. In: Chen J, editor. Advances in Human Factors in Robots and Unmanned Systems: Proceedings of the AHFE 2019 International Conference on Human Factors in Robots and Unmanned Systems, July 24-28, 2019, Washington DC, USA: Springer; 2020:92-103. [doi: 10.1007/978-3-030-20467-9_9]
20. Park YH, Chang HK, Lee MH, Lee SH. Community-dwelling older adults' needs and acceptance regarding the use of robot technology to assist with daily living performance. BMC Geriatr 2019 Aug 5;19(1):208. [doi: 10.1186/s12877-019-1227-7] [Medline: 31382887]
21. Zhao IY, Holroyd E, Wright-St Clair VA, Wang SS, Garrett N, Neville S. Feeling a deep sense of loneliness: Chinese late-life immigrants in New Zealand. Australas J Ageing 2022 Sep;41(3):448-456. [doi: 10.1111/ajag.13108] [Medline: 35791051]
22. Montayre J, Saravanakumar P, Zhao I, Holroyd E, Adams J, Neville S. Holding on and letting go: views about filial piety among adult children living in New Zealand. J Clin Nurs 2022 Oct;31(19-20):2797-2804. [doi: 10.1111/jocn.16098] [Medline: 34687112]
23. Wright-St Clair VA, Nayar S. Older Asian immigrants' participation as cultural enfranchisement. J Occup Sci 2016 Aug 15;24(1):64-75. [doi: 10.1080/14427591.2016.1214168]

Edited by H LaMonica; submitted 02.05.23; peer-reviewed by A Poulsen, ES Nahm; revised version received 10.10.23; accepted 10.10.23; published 08.11.23.

Please cite as:
Chiu CJ, Lo YH, Montayre J, Abu-Odah H, Chen ML, Zhao IY
Identifying Preferred Appearance and Functional Requirements of Aged Care Robots Among Older Chinese Immigrants: Cross-Sectional Study
JMIR Aging 2023;6:e48646
URL: https://aging.jmir.org/2023/1/e48646
doi:10.2196/48646

Original Paper

Use and Engagement With Low-Intensity Cognitive Behavioral Therapy Techniques Used Within an App to Support Worry Management: Content Analysis of Log Data

Paul Farrand[1,2*], PhD; Patrick J Raue[3*], PhD; Earlise Ward[4*], PhD; Dean Repper[5*], MSc; Patricia Areán[3*], PhD

[1]Clinical Education, Development and Research, Faculty of Health and Life Sciences, University of Exeter, Exeter, United Kingdom

[2]Department of Psychology, Faculty of Health and Life Sciences, University of Exeter, Exeter, United Kingdom

[3]AIMS CENTER, Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA, United States

[4]School of Medicine and Public Health, Carbone Comprehensive Cancer Center, University of Wisconsin-Madison, Madison, WI, United States

[5]Trent PTS, Improving Access to Psychological Therapies, Derby, United Kingdom

[*]all authors contributed equally

Corresponding Author:
Paul Farrand, PhD
Clinical Education, Development and Research
Faculty of Health and Life Sciences
University of Exeter
Sir Henry Wellcome Building for Mood Disorders Reserach
Perry Road
Exeter, EX4 4QG
United Kingdom
Phone: 44 01392725793
Email: p.a.farrand@exeter.ac.uk

Abstract

Background: Low-intensity cognitive behavioral therapy (LICBT) has been implemented by the Improving Access to Psychological Therapies services across England to manage excessive worry associated with generalized anxiety disorder and support emotional well-being. However, barriers to access limit scalability. A solution has been to incorporate LICBT techniques derived from an evidence-based protocol within the Iona Mind Well-being app for Worry management (IMWW) with support provided through an algorithmically driven conversational agent.

Objective: This study aims to examine engagement with a mobile phone app to support worry management with specific attention directed toward interaction with specific LICBT techniques and examine the potential to reduce symptoms of anxiety.

Methods: Log data were examined with respect to a sample of "engaged" users who had completed at least 1 lesson related to the Worry Time and Problem Solving in-app modules that represented the "minimum dose." Paired sample 2-tailed *t* tests were undertaken to examine the potential for IMWW to reduce worry and anxiety, with multivariate linear regressions examining the extent to which completion of each of the techniques led to reductions in worry and anxiety.

Results: There was good engagement with the range of specific LICBT techniques included within IMWW. The vast majority of engaged users were able to interact with the cognitive behavioral therapy model and successfully record types of worry. When working through Problem Solving, the conversational agent was successfully used to support the user with lower levels of engagement. Several users engaged with Worry Time outside of the app. Forgetting to use the app was the most common reason for lack of engagement, with features of the app such as completion of routine outcome measures and weekly reflections having lower levels of engagement. Despite difficulties in the collection of end point data, there was a significant reduction in severity for both anxiety (t_{53}=5.5; *P*<.001; 95% CI 2.4-5.2) and low mood (t_{53}=2.3; *P*=.03; 95% CI 0.2-3.3). A statistically significant linear model was also fitted to the Generalized Anxiety Disorder–7 ($F_{2,51}$=6.73; *P*<.001), while the model predicting changes in the Patient Health Questionnaire–8 did not reach significance ($F_{2,51}$=2.33; *P*=.11). This indicates that the reduction in these measures was affected by in-app engagement with Worry Time and Problem Solving.

Conclusions: Engaged users were able to successfully interact with the LICBT-specific techniques informed by an evidence-based protocol although there were lower completion rates of routine outcome measures and weekly reflections. Successful interaction

with the specific techniques potentially contributes to promising data, indicating that IMWW may be effective in the management of excessive worry. A relationship between dose and improvement justifies the use of log data to inform future developments. However, attention needs to be directed toward enhancing interaction with wider features of the app given that larger improvements were associated with greater engagement.

(JMIR Mhealth Uhealth 2024;12:e47321) doi:10.2196/47321

KEYWORDS

cognitive behavioral therapy; low-intensity; mCBT; app; log data; worry management; CBT; management; application; therapy; implementation; treatment; symptoms; anxiety; worry; engagement

Introduction

Excessive worry represents a core characteristic associated with generalized anxiety disorder (GAD) [1], characterized as 6 or more months of chronic worry about several different events and little belief worry can be controlled, and is associated with general somatic symptoms of anxiety [2]. It is highly pervasive in high-income countries, with a lifetime prevalence rate of 7.8% in the United States [3] and a median age of onset between 24 and 50 years, and is more common in women and people who are unemployed [4]. Excessive worry is deleterious to the individual, and if it manifests into GAD, it places a significant burden on society and employers with an average of 6.3 days per month of work absenteeism [5]. Furthermore, treatment is associated with increased service use [6], thereby placing a significant burden on primary care settings within both the United Kingdom [7] and United States [8].

Despite excessive worry impacting on the individual quality of life with progression to GAD representing a major public health problem [9], treatment availability remains limited. The treatment gap for GAD has been reported to be as high as 67% in the United Kingdom and 70% in the United States [10]. In an attempt to address the treatment gap [11], innovations in the delivery of evidence-based psychological therapy have been witnessed, for example, by broadening the workforce delivering cognitive behavioral therapy (CBT) for late-life GAD with no differences in effectiveness demonstrated when delivered by bachelor-level lay providers compared with PhD-level experienced therapists [12]. Further attempts to address the treatment gap have been addressed within the Improving Access to Psychological Therapies (IAPT) program, implementing low-intensity cognitive behavioral therapy (LICBT) self-help for the management of mild to moderate common mental health problems alongside therapist-delivered high-intensity CBT for moderate to severe presentations [13].

While improved access has been facilitated through the IAPT program, predictions indicate that access will only be increased to 25% of the community prevalence of depression and anxiety disorders by 2023-2024 [14]. Furthermore, between 2021 and 2022, only 37% of patients completed a course of therapy [15]. Difficulties in reducing the treatment gap are further dependent on a number of individual-level barriers such as stigma, desire to handle the problem independently, and limited willingness to disclose problems [11].

Greater implementation of digital health technologies such as smartphone apps [16] has potential to address barriers at the level of the individual. Furthermore, with high use of smartphones, for example, 81% of adults within the United States [17], apps offer the promise of delivering mental well-being interventions at scale and derive cost savings with respect to delivery and increased productivity within the workplace [18]. The expansion of apps to manage challenges with well-being has predominantly been based on CBT [19]. CBT is particularly well suited to inform mental well-being apps with emphasis placed on active engagement with specific techniques rather than exclusive reliance on a supportive relationship with a behavioral health coach or therapist [13]. This is especially salient with regard to an approach based on "collaborative empiricism," whereby after engagement with specific CBT techniques, people are encouraged to explore outcomes for themselves [20]. In doing so, a better understanding of the way their mental health difficulty is affecting them can be derived through an appreciation of the cognitive behavioral model.

Despite CBT providing a compatible evidence-based approach for mental well-being apps with the potential to deliver at scale to close the treatment gap, implementation and uptake remain limited [21]. This is partly dependent on poor engagement with apps relying on factors such as poor usability, techniques inconsistent with user expectations, and poor health literacy [22]. Exploring ways to enhance engagement is of significance given that greater engagement has been reported to predict improvements in mental well-being [23]. In particular, focusing on obstacles and difficulties encountered in completing LICBT techniques is of significance given their effectiveness [24], while not dismissing common factors to establish a good "therapeutic relationship" generating a sense of genuineness, warmth, and collaborative working [25]. Focusing on both common and specific LICBT techniques used within the Iona Mind Well-being app for Worry management (IMWW) is therefore of importance given that the combination of both is crucial for bringing about therapeutic change [26].

This paper reports log data [27] to examine engagement with IMWW based on LICBT to help in the management of excessive worry. It has been proposed that rather than looking at overall engagement regarding areas such as number of sessions engaged with or session duration, it is better to focus attention on specific user interactions [28]. The focus of this paper is therefore directed toward appreciating engagement and interaction with specific LICBT techniques and wider features of IMWW to examine potential enhancements alongside wider usability. Furthermore, the relationship between engagement and outcomes

will be explored to inform ongoing development to maximize effectiveness.

Methods

Design

Areas representing the focus of log data analysis have been informed by recommendations identified as useful when seeking to address the effectiveness of eHealth technology alongside behavioral and theoretical models [27]. Analysis was only undertaken on data collected regarding "engaged users" and their log data associated with engagement with the LICBT techniques. To be considered an engaged user, the user was required to have completed at least 1 lesson related to the Worry Time and Problem Solving in-app modules within any time period. These criteria represented the "minimum dose" [29] necessary for the user to be considered to have engaged enough to be able to understand the in-app CBT techniques and apply them outside of the app to manage excessive worry. This approach to represent "engaged users" has previously been adopted. For example, with respect to a feasibility trial examining internet-administered LICBT for parents of children treated for cancer [30].

A χ^2 test of homogeneity was undertaken to compare demographic data provided by the engaged users and those who failed to engage with specific LICBT techniques to receive a minimum dose.

Sample

Over 6 months (July 19, 2022, to February 19, 2023), 956 adults 18 years and older with a GAD-7 score of 6 and above downloaded and opened IMWW and completed the sign-up process. Of these, 803 (84%) adults did not engage sufficiently with the app to be considered an "engaged user," resulting in 153 (16%) users engaging with the app sufficiently to be considered to have received a minimum dose (Table 1). Approximate data regarding the continent the user was accessing the app from were automatically collected by the app from the time zone set on the user's phone and therefore collected on all 956 adults.

χ^2 tests of homogeneity indicate that differences between engaged users and those who had downloaded the app but failed to receive a minimum dose were not significant at the 5% level across any of the demographic variables. In addition to the collection of demographic data, 41 of 153 (27%) engaged users responded to a question regarding receipt of other therapy, of whom 34 (83%) users indicated that they were not.

Table 1. Demographic questionnaire responses completed (N=956).

Variable	Engaged users (n=153)	Not received minimum dose (n=803)
Gender, n (%)[a]		
Women	43 (72)	50 (85)
Men	14 (23)	9 (15)
Other	3 (5)	0 (0)
Age range (years; n=53), n (%)		
18-24	18 (34)	16 (30)
25-34	20 (38)	19 (36)
35-44	9 (17)	12 (23)
45-54	2 (4)	6 (11)
55-64	1 (2)	0 (0)
≥65	3 (6)	0 (0)
Continent, n (%)		
Americas	78 (51)	369 (46)
Europe	33 (22)	196 (24)
Asia	28 (18)	161 (20)
Africa	5 (3)	22 (3)
Australasia	4 (3)	32 (4)
Unknown	5 (3)	23 (3)

[a]A total of 60 engaged users and 59 users who did not receive the minimum dose responded.

Iona Mind Well-Being App for Worry Management

The IMWW is, in part, informed by the techniques described in the LICBT written self-help intervention for managing excessive worry [31], based on the CBT protocol for the management of GAD [32]. LICBT is recommended for the management of GAD [33] and is one of the most commonly adopted written self-interventions used within the IAPT program [34]. The focus of IMWW is explained during onboarding where the user is required to explicitly acknowledge its purpose as a

well-being tool. Users wishing to continue engagement acknowledge that they understand conditions related to use and consent to have their data processed.

Collection of Demographic Data

Demographic data were not used to inform the delivery or functionality of the app. Therefore, a screen requesting demographic data, or a question regarding receipt of other therapy, was only presented once the user had engaged with IMWW on 2 occasions at least 6 hours apart following enablement, and no other higher priority messages were pending. If higher priority messages were pending, the request to provide demographic data was repeatedly postponed to the following day until the user had supplied all data, completed specific questions, or declined the request to open the screen (Table 1). Due to the optionality and logic surrounding whether demographic data were requested from users who downloaded the app, such data were therefore not requested from all users and were not prioritized over other more useful app functionality. Collecting demographic data was not prioritized given that answering such questions on an app can increase the risk of disengagement [35].

Supporting Interaction

Interaction is supported by an algorithmically driven chatbot simulating a "conversation" between the conversational agent (CA) and the user. Users interact by entering raw text or selecting a predetermined response. The type of response depends on the type of content being delivered and varies between selecting a button from a list of options or entering free text when a personalized response is requested. User feedback is collected at the level of a user message supplied within the "conversation" through selecting a "thumbs up" or "thumbs down" icon.

Key principles associated with user-centered systems design were adopted to inform the development of IMWW [36]. The app was developed to manage excessive worry and support emotional wellness with an "SOS" button prominently displayed for users finding themselves in significant emotional distress. If selected, signposting information to a comprehensive list of local and international crisis helplines is presented alongside mindfulness practices to assist with mood stabilization. Before engaging, the user is further reminded that IMWW is not intended to deliver treatment but rather is a tool to support well-being and is not designed for anyone who has been diagnosed with a psychiatric disorder. Consequently, it is stressed that the app is not to be used outside of the context of a well-being self-help aid.

Progress Through IMWW

Engagement begins with users landing on the Today home screen and progressing through 6 educational modules through which they learn about and interact with the LICBT techniques (Multimedia Appendix 1).

Educational modules are chronologically ordered and unlocked as engagement is initiated. The order in which they are unlocked is dependent on user choice, reflecting whether they wish to initially address practical or hypothetical worries. However, the user is able to move forward and backward between these specific factors to address the different types of worries where preferred. After onboarding, the user is given the opportunity to complete the GAD-7 and Patient Health Questionnaire–8 (PHQ-8) weekly during their weekly review, with scores presented on a progress screen. However, following the completion of these measures during onboarding, subsequent completion is voluntary.

Home Screen

Informed by the CBT protocol for the management of GAD, which can also be used in the context of improving emotional well-being by supporting the management of worries [37], the Today (Home) screen supports the user to record their worries. This screen also presents a timer that counts down to the user's scheduled worry time and offers tips to complete Worry Time (Multimedia Appendix 1). Should the user not have completed the lesson, a placeholder is displayed inviting them to learn more about Worry Time. The screen has been designed to make it as easy as possible for users to record worries, plan Worry Time, and access the CBT content. A navigation bar at the bottom of the screen links to the screens related to the LICBT techniques through which users can engage depending on preference.

CBT Techniques

To promote engagement, the IMWW is informed by collaborative empiricism where the user actively engages with techniques associated with CBT [38]. Collaborative empiricism has been identified as core to the therapeutic relationship [39], supporting "learning by doing" fundamental to CBT [40]. Accordingly, LICBT techniques are presented as skills to be mastered through regular practice. The user is initially introduced to the CBT model followed by supporting them to record and categorize their worries. The user is then provided with the choice of Problem Solving or Worry Time to address practical or hypothetical worries, respectively.

CBT Model

An interactive CBT Five Areas model (introduced in October 2022) is presented [41], and the user is encouraged to interact and identify a current situation in the "here and now." In response to this current situation, the user is also encouraged to interact with boxes reflecting "Thoughts" that go through their head, "Behaviors" engaged in, "Physical Feelings," and to recognize "Emotions" (Multimedia Appendix 1). Additionally, the model helps them appreciate ways in which the specific LICBT techniques presented to address practical and hypothetical worries may be helpful. The model serves as psychoeducation, enabling the user to recognize the interaction between each of these areas, understand the nature of their worry, and appreciate factors that maintain their worry behavior and the impact of physical symptoms associated with anxiety.

Worry Diary

The primary function of the Worry Diary is to enable the user to actively add new worries as they arise throughout the day and as a record of worries for subsequent review (Multimedia Appendix 1). Prompts and predefined categories are used to

enable the user to differentiate between practical worries that have a solution and hypothetical worries that do not.

Problem Solving

Where practical worries are logged, the user is guided to list potential solutions, consider strengths and weaknesses for each solution, and select the most appropriate one. A time to try the solution out is then optionally scheduled by the user with a reminder given to complete it at the chosen time. After the chosen time has passed or 30 hours elapsed, on opening the app, the CA will ask the user to review how their solution went. Prompts ask the user if the problem was resolved and if not request further information regarding the challenges encountered. Advice is offered where problems have been encountered, putting the solution into action (eg, to break the problem down or work through and apply another solution).

Worry Time

Unless explicitly overridden by the user, all worries identified as hypothetical are displayed only during Worry Time at a time determined by the user and are blurred out at all other times. Users are reminded that their worry time is starting with a push notification to their phone. If the user opens the app during Worry Time, they are prompted to work through the time they have set aside to worry with the CA. If they choose to do so, the CA will list out all user worries and request that these be worried about for the specified period of time. Subsequently, the CA will review each worry with the user, asking them if it still remains an issue or if it now better represents a practical worry. Worry Time represents a form of cognitive exposure with users exposed to hypothetical worries written down during the day. This is proposed to overcome avoidance behavior and reduce intolerance of uncertainty when it is recognized that there are no solutions to the worry [37].

Maintaining Engagement

A chatbot informed by theoretically driven techniques is adopted to help establish a “therapeutic approach” to maintain and promote user engagement embedded within IMWW [38]. Such techniques help to establish an approach based on collaborative empiricism [39], whereby the user is encouraged to explore outcomes arising from engagement for themselves.

Conversational Agent

Support is omnipresent throughout the engagement and comes in the form of an algorithmically driven chatbot stimulating engagement between IMWW, the CA, and users. This helps them overcome difficulties encountered with the specific LICBT techniques and uses common factors to maintain engagement. Upon recognition that difficulties are experienced with any of the specific techniques, the CA is deployed to enable users to work through the specific techniques. If the user reports difficulties in trying out a solution they have planned, the CA will ask questions to determine the nature of the difficulty encountered and direct them to the appropriate parts of the app. For example, the user would be directed back to Problem Solving should they need to break the problem down, or Worry Diary if the worry appears to be hypothetical rather than practical (Table 1).

On other occasions, the CA provides the user with helpful tips and advice or the opportunity to ask FAQs to navigate difficulties experienced. For example, 2 days after learning about Worry Time, the CA will check back in with the user and ask how the exercise has been going. Depending on user response, advice will be given. For example, if the user forgets to engage with Worry Time, they are reminded to turn on their notifications and set an alarm on their phone to serve as a prompt. Consistent with the delivery of CBT, during engagement with the LICBT techniques, the CA brings the user back to the CBT model to reinforce their understanding of the intervention and maintain motivation for continued engagement.

Common Factors

The CA uses nontherapeutic common factor skills in the form of “therapeutic empathy” to instill a sense of hopefulness and encouragement to maximize engagement with the specific factors linked to symptom reduction [42]. Statements include those demonstrating an empathic stance highlighting a desire to help alongside empathic attunement where statements demonstrate an appreciation of the user’s emotional experience [42]. When recognizing that the user is experiencing difficulties in engaging or is not improving, the CA uses empathy to maintain engagement.

Behavior Change Techniques

Within the module on recording worries, Behavioral Contracting [43] encourages the user to sign an agreement to consistently engage with IMWW throughout the 6-week program with a separate Goal Setting lesson guiding the user to set approach, rather than avoidance, goals [44]. Behavior “Push” notifications serve to prompt or maintain behavior change while engaging with the app. Furthermore, constructs derived from self-determination theory [45] promote autonomy and intrinsic motivation that serve to facilitate collaborative empiricism [37].

Monitoring Progress

Throughout engagement, a progress screen presents a summary of the user’s app use and engagement with in-app lessons, previously entered goals, and scores regarding symptom severity associated with anxiety (GAD-7) [46] and low mood (PHQ-8) [47]; it also presents links to the settings page, which houses operational features such as typing speed (Multimedia Appendix 1). Given a potential association between providing feedback and improved outcomes, all data collected are repeatedly presented to the user throughout engagement [48].

Weekly Reflection

Consistent with face-to-face CBT [39], on a weekly basis, the CA prompts the user to reflect on their engagement with IMWW and the features found most helpful (Table 1). Using reflective learning within the app facilitates learning, with the CA encouraging engagement to promote self-discovery [49]. Furthermore, during the weekly reflection, the CA requests information on the LICBT techniques engaged with and highlights those found most helpful. In the event an identified technique was not engaged with, the CA also requested information as to the main reasons from a range of options provided.

Data Collection and Analysis

Log Data

Consistent with the aims of the study, analysis was undertaken on log data collected from engaged users to reflect their engagement with the specific LICBT techniques. Log data were collected by IMWW automatically logging the actions of each engaged user and requests to complete surveys throughout the use of the app. From these data, summary statistics for use in this paper were extracted. Progress of engaged users through IMWW was monitored and informed by data regarding the number of sessions completed, session duration, weekly reflections, and completion of LICBT techniques alongside summary statistics recorded. Engagement with IMWW was explored with respect to the number of users who reached the CBT model and interacted with it, text entered into each area, completion of the LICBT techniques, and general input and behavior during the engagement. Specific worry management techniques were examined with respect to the number of worries entered and the proportion classified as practical or hypothetical problems. With respect to Problem Solving, data analysis included the number of users who completed the lesson, the number of times the in-app tool was used to solve a practical problem, the number of practical problems entered, the number of users prompted to follow up on their problem-solving with the CA, and the number who engaged with it. Furthermore, analysis was undertaken on the number of engaged users who sought to manage hypothetical worries by learning about Worry Time, set a time for Worry Time, and started an in-app session alongside the number of hypothetical worries entered being recorded.

User responses from the Weekly Reflection conversation within IMWW were also collected and analyzed to gauge general engagement with the specific techniques. As a proxy for behavior change approaches adopted to maintain engagement within the app, the number of users who were delivered at least 1 push notification and the number of those who interacted were also examined. For engaged users completing more than 1 GAD-7 or PHQ-8 at assessment, the log of assessments and the number of times IMWW was used for more than 10 seconds, which is defined as a "session," were analyzed.

Potential Effectiveness

To examine the potential effectiveness of IMWW for engaged users, separate paired samples 2-tailed *t* tests were undertaken to examine the difference between outcome data collected regarding the severity of anxiety (GAD-7) and low mood (PHQ-8). This analysis was only undertaken for the 54 of 153 (35%) engaged users who completed the outcome measures during onboarding and at the end of the engagement.

Impact of Engagement on Potential Effectiveness

A multivariate linear regression model was used to investigate the impact that engagement with IMWW had on improvement in anxiety and low mood. In particular, the extent to which specific features were used to complete therapeutic exercises impacted on scores over time. Engagement with, and completion of, Worry Time and Problem Solving was expected to lead to improvements in the symptoms of anxiety, and hence a model to analyze this was specified. Because there are multiple discrete interventions being applied within IMWW and the dependent variable is not univariate, the multiple regression $y=\beta_0+\beta_1x_1+\beta_2x_2+$ was adopted. Within this model, y is the change in GAD-7 or PHQ-8 from the initial score at onboarding to the final input during progress review, x_1 is the binary variable indicating whether the user completed Problem Solving and resolved their problem, x_2 is the binary variable indicating whether the user completed at least 1 instance of Worry Time in-app, and is the stochastic error term. Additional controls were added to the model to examine the extent to which the number of in-app sessions completed, and the number of worries, problems, and solutions recorded predicted improvement in GAD-7 and PHQ-8. All models met OLS model assumptions associated with multicollinearity, heteroskedasticity, and normality of residuals.

Ethical Considerations

Users were only able to download IMWW after agreeing to Iona Mind's Terms of Service and Privacy policy, which required them to acknowledge that they understand conditions related to use and consent to have their anonymized data processed. Being based on anonymous, routinely collected log data from a nonclinical population, research ethics was not required for this study.

Results

General Engagement

Analysis of log data collected from the 153 engaged users indicated engagement with 1108 sessions (mean 7.2, SD 7.7) with an average session length of 6.2 (SD 6.2) minutes. The number of sessions and session length varied significantly across users with a median session length of 4.5 minutes and 6 being the median number of sessions (Multimedia Appendix 2).

Engagement With LICBT Techniques

Since inclusion (October 2022), 36 users started filling out the CBT model to reflect their current difficulties with anxiety, and of these, 31 (86%) users completed all areas in an average of 2.4 minutes (SD 1.7; median 1.7 minutes). The lesson on the Worry Time technique was successfully completed by almost all users (147/153, 96%). However, of these users, only 50 (33%) were observed to have performed Worry Time at their chosen time using the in-app tools.

Problem Solving was engaged with by 114 of 153 (74.5%) users; however, only 89 of 153 (58.2%) users actually completed the lesson. This indicates that 25 of 114 (21.9%) users engaged with the in-app tools to problem-solve one of their practical worries without completing the lesson. This behavior is permitted within the IMWW user experience because the user is able to choose the specific features of the app they wish to engage with. The majority of app features start in an unlocked state to encourage exploration and self-discovery. Only 42 of 153 (27%) users completed a follow-up conversation to review their solutions and progress using the Problem Solving protocol on their worries. During the interaction, engaged users recorded a total of 720 worries (mean 4.7, SD 6.1), and a median of 3

worries were recorded for each user. Of the worries recorded, 399 (55%) were categorized by the user as practical, 306 (43%) as hypothetical, and only 15 (2%) worries were not categorized. With respect to practical worries, 244 (61%) worries were problem-solved using the in-app tools with at least 1 possible solution added.

Weekly Reflection

In response to the CA asking the user to reflect on their experience of engaging with IMWW, in-app Weekly Reflections were completed by 58 of 153 (38%) users who recorded 206 responses (mean 3.6) identifying LICBT techniques engaged with, alongside 48 responses identifying the technique found most helpful (Table 2).

During the Weekly Reflection, 27 of 58 (47%) individual users reflected on engaging with Worry Time, of whom 19 (70%) were observed to have used the in-app tooling to complete it at their chosen time. A total of 8 of 58 (14%) users therefore engaged with Worry Time without using the in-app tools. In addition to asking which features of IMWW the user had engaged with, the CA also asked which feature they found most helpful. The users were asked this question during the weekly review, and for each weekly review, they could give at most 1 response.

Table 2. Weekly reflection techniques engaged with and found most helpful.

Technique	Engaged with[a] (n=206), n (%)	Most helpful (n=48), n (%)
Journaling worries	50 (24)	16 (33)
Worry time	46 (22)	11 (23)
Problem-solving	35 (17)	10 (21)
Avoiding worry behaviors	28 (14)	4 (8)
Watching out for different worry types (Worry categorization)	17 (8)	4 (8)
CBT[b] model[c]	8 (4)	1 (2)

[a]Users can respond multiple times.

[b]CBT: cognitive behavioral therapy.

[c]Introduced in October 2022.

Maintaining Engagement

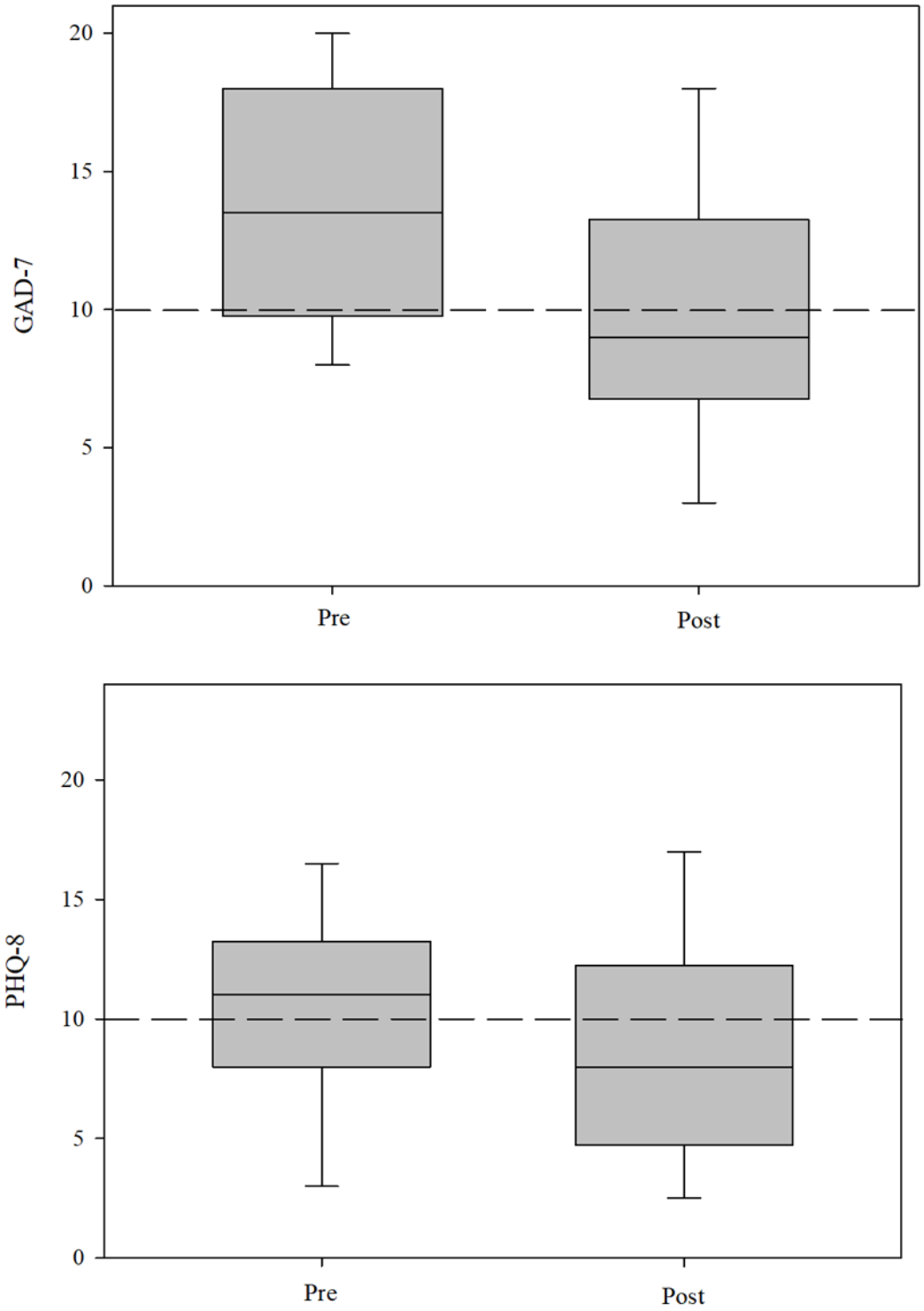
To maintain engagement with IMWW, 142 of 153 (93%) users were sent at least 1 push notification with 113 of 153 (74%) users responding. An average of 84 (SD 75) push notifications were sent to each engaged user throughout their engagement, although the quantity of push notifications per user varied substantially with use pattern and duration.

Potential Effectiveness

Separate paired sample 2-tailed *t* tests were conducted to examine the difference between the GAD-7 and PHQ-8 scores for 54 of 153 (35%) engaged users who completed the measures during onboarding and the final score provided. There was a significant reduction in both anxiety (t_{53}=5.5; *P*<.001; 95% CI 2.4-5.2) and low mood (t_{53}=2.3; *P*=.03; 95% CI 0.2-3.3), with severity dropping from moderate to mild in both instances (Figure 1).

Examination of individual-level data indicates that the vast majority of users (43/53, 81%) experienced a reduction in anxiety between baseline and final observation with the score of 2 (4%) users remaining unchanged. The majority of users (35/53, 66%) also saw a reduction in PHQ-8 with no difference arising for 4 (8%) users. Deterioration in GAD-7 was experienced by 9 (17%) users and rose to 15 (28%) users for low mood.

Figure 1. Pre-post mean differences (95% CI) for GAD-7 and PHQ-8. GAD-7: Generalized Anxiety Disorder–7; PHQ-8: Patient Health Questionnaire–8.



Impact of Engagement on Potential Effectiveness

The multivariate linear regression predicting changes in GAD-7 based on engagement reached statistical significance ($F_{2,51}$=6.73; P=.03), but the model predicting changes in the PHQ-8 did not ($F_{2,51}$=2.33; P=.11). Two-sided 2-tailed t tests were performed on the slope estimates in the models. The model predicting changes in GAD-7 estimates that the marginal effect of a user completing in-app Worry Time (β_2) is a –3.3 change in GAD-7 and is significant (P=.02). The constant β_0 and the marginal effect of the user successfully completing Problem Solving β_1 had respective values of –1.6 and –3.0. However, they failed to reach significance at α=.05 (P_0=.07, P_1=.08). The model had an R^2 of 0.21. Furthermore, an improvement in the GAD-7 and PHQ-8 was not predicted by the number of in-app sessions completed (P=.09) or the number of worries (P=.36), problems (P=.27), and solutions (P=.16) recorded.

Discussion

Principal Findings

While engaged users represented a minority of those who downloaded IMWW, a large number of these interacted with the LICBT techniques associated with the CBT protocol to manage excessive worry and support emotional well-being [31]. The vast majority of those who engaged completed all areas presented with the CBT model and were able to successfully record worries and categorize them as practical or hypothetical. The CA was commonly used to help engaged users overcome difficulties when engaging with practical worries. Worry Time was engaged with to a much lesser extent within the app; however, several users reported engaging with it outside of the app. Forgetting to engage with the LICBT techniques was identified as the most common reason for lack of engagement, while experiencing them as too difficult to comprehend was only reported by a small minority of engaged users. The change in the user GAD-7 score was predominantly explained by engagement with the LICBT techniques as opposed to the number of times they used IMWW.

Poor engagement with an app following download is not uncommon, with only 14% of people often using it the following day [50] and even lower rates typically associated with mental health apps [22]. Despite using common factors and behavior change techniques, however, only a minority of users who engaged with IMWW had enough engagement with the LICBT techniques to be considered engaged users. This is of some concern given that users failing to engage to a point where they have received an adequate dose to bring about recovery may serve as a barrier to seeking further support.

While engagement following download was poor, log data identified that engaged users had moderate to good levels of interaction and fidelity [51], with the CBT model alongside recording and categorizing worries. Fidelity and interaction with Problem Solving were also good, potentially arising from support provided through the CA. When engaging with Problem Solving, the CA was commonly used to support users to overcome difficulties in engaging with the LICBT techniques and to encourage continued engagement. There was less within-app engagement with Worry Time; however, some users reported engaging with it outside of the app. Engagement with the LICBT techniques included within IMWW may therefore have been greater than log data alone suggest. This supports the additional benefits of exploring out-of-app engagement with specific techniques to get a full appreciation of interaction [28]. Exploring ways to promote out-of-app engagement is of benefit given that engagement with techniques in face-to-face CBT between support sessions as “homework” is identified as important to improve clinical outcomes related to anxiety [52].

Although there were moderate levels of interaction with LICBT techniques used within IMWW, exploring additional ways to enhance interaction across all techniques and promote prolonged engagement would be highly beneficial. Enhancing engagement through approaches such as involving personalized support, guidance, and feedback regarding engagement has also been associated with improved effectiveness for mental well-being digital tools [48]. Furthermore, recommendations to enhance out-of-app homework compliance to deliver better outcomes have also been proposed [53]. These include ensuring that app content is congruent to the therapeutic approach adopted, learning is consolidated through engagement, and emphasis is placed on completion. Additionally, recommendations include ensuring that the app is tailored to specific populations and building connections with others has been identified as supporting engagement with homework [53]. Within IMWW the CA was used to enhance engagement through the use of common factor skills to encourage and motivate the user. However, greater focus needs to be directed toward maximizing the ability of the CA to enhance engagement within and outside of the app.

Maximizing engagement may be achieved by implementing mental well-being apps for use adjuvant to health professional support and integrated into clinical settings [54]. Benefits associated with providing support are recognized by the National Institute of Health and Care Excellence recommendations for supported LICBT for anxiety and depression [14]. This has resulted in Psychological Practitioner support for LICBT adopted by the IAPT program implemented across England [14]. Support enables the patient to engage with the interventions by using personalized common factor skills, monitor progress, and provide encouragement during weekly support sessions. However, it does not include a therapeutic role in the delivery of LICBT techniques within the clinical sessions [13].

However, nonprofessional forms of support have also been demonstrated to enhance engagement and improve outcomes with LICBT. For example, group support within community settings is provided by trained volunteers with varying backgrounds [55]. Furthermore, forms of support through technology such as web-based communities providing constructive peer support [55] and discussion forums [56] have been identified to enhance engagement with digital tools [54]. Potentially, therefore, using IMWW adjuvant to some form of minimal-contact support provided by a practitioner, volunteers within community organizations, or mediated through technology offers promise to result in enhanced effectiveness at reduced delivery costs.

With respect to outcomes, the average level of anxiety and low mood improved among users who engaged with IMWW to a degree they would be considered to have received a minimum dose of the LICBT techniques [29]. That anxiety and low mood are identified to share mechanisms has led to recommendations to combine techniques within a single app to reduce the commitment needed by users to maximize engagement [19]. However, when exploring recovery at the level of the individual user, the low mood of several more users deteriorated compared with anxiety. However, it would remain possible to develop a single app that included protocol-informed LICBT techniques to target low mood or anxiety once the main emotional difficulty being experienced was determined.

Strengths and Limitations

Providing a clear description of the LICBT techniques contained within IMWW informed by a theoretical basis represents a real strength of the paper. This has enabled the analysis of log data

to be interpreted with respect to interactions with the techniques. Clearer conclusions regarding the relationship between engagement and outcomes regarding the management of symptoms associated with anxiety were therefore able to be reached. This facilitates specific targeting of future development work on IMWW to ensure greater levels of engagement to derive improved outcomes.

There was a large difference between the number of people who downloaded IMWW and those who interacted with at least 1 lesson related to Worry Time and Problem Solving for them to be considered engaged users. While it is known that the background demographics of these 2 groups did not significantly differ, it is unclear as to why a large number of those who downloaded IMWW never went on to engage with one of these specific LICBT techniques. Unfortunately, reasons behind failing to engage with IMWW were not requested, and therefore the extent to which poor usability may have been a relevant factor is unknown. As the use of digital health technologies continues to increase [16], understanding the usability of apps is of increasing interest [57]. Future research exploring log data could therefore consider using a measure of usability, such as the mHealth Usability Questionnaire [58], alongside the collection of log data to gain a better understanding of the way in which an app is used alongside potential barriers to usability.

Finally, while data regarding outcomes can be considered promising with respect to IMWW as a tool to support worry management, this study does not enable definitive conclusions regarding effectiveness to be reached. As a consequence of the lack of clear end points when using log data, users can stop using the app at any time without completing outcome measures. This makes it difficult to reach conclusions regarding effectiveness. The use of multivariate regression with terms to represent proxy use of techniques was adopted to compensate for this. However, this cannot be considered to represent a substitute for the collection of clear and reliably collected end point data within a trial design comparing IMWW with an appropriate control [59]. Furthermore, reaching conclusions regarding effectiveness is further confounded given that only a minority of engaged users responded to a question regarding the current receipt of treatment.

Conclusions

While a large number of people downloaded IMWW, only a minority engaged with the app to be considered engaged users. Of these users, however, analysis of log data identified good interaction with the LICBT techniques associated with an evidence-based protocol to support worry management [31]. Although there were good levels of interaction, exploring additional ways to promote interaction with the LICBT techniques and other features of the app to result in prolonged engagement remains beneficial. This could involve adopting a “user-centric” design process whereby potential users are directly involved in ongoing development [22]. Considering log data as part of a user-centric design process may enhance engagement to a point where more users receive an appropriate “dose” to bring about improvement [29]. Log data can therefore be used to inform ongoing development to maximize engagement and protocol fidelity [51]. This is significant given the relationship between engagement and effectiveness. While effectiveness data associated with IMWW can only be seen as promising, capturing log data will serve to enhance ongoing intervention development. A high-quality randomized controlled trial would then enable definitive conclusions regarding effectiveness to be reached [54]. This would help address concerns that the current level of evidence derived from poor-quality trials does not enable recommendations regarding apps to enhance mental well-being to be reached [60].

Authors' Contributions

PF, PA, and PJR conceptualized and designed the project with EW and DR providing theoretical input informing the background of the paper. PF wrote the initial draft of the paper with all authors contributing to the development of the paper, interpretation of the analysis, and editing of the final manuscript. PF was not involved in any part of data analysis.

Conflicts of Interest

PF is currently on a paid sabbatical with Iona Mind from the University of Exeter. All other authors have confirmed they have no conflicts of interest to declare.

Multimedia Appendix 1
Progress through the Iona Mind Well-being app for Worry management.
[DOCX File , 951 KB - mhealth_v12i1e47321_app1.docx]

Multimedia Appendix 2
Number of sessions of varying length.
[PNG File , 60 KB - mhealth_v12i1e47321_app2.png]

References

1. Diagnostic and Statistical Manual of Mental Disorders 5th edition. Washington, DC: American Psychiatric Association; 2022.
2. Wittchen HU. Generalized anxiety disorder: prevalence, burden, and cost to society. Depress Anxiety 2002;16(4):162-171 [FREE Full text] [doi: 10.1002/da.10065] [Medline: 12497648]

3. Ruscio AM, Hallion LS, Lim CCW, Aguilar-Gaxiola S, Al-Hamzawi A, Alonso J, et al. Cross-sectional comparison of the epidemiology of DSM-5 generalized anxiety disorder across the globe. JAMA Psychiatry 2017;74(5):465-475 [FREE Full text] [doi: 10.1001/jamapsychiatry.2017.0056] [Medline: 28297020]
4. Bandelow B, Michaelis S. Epidemiology of anxiety disorders in the 21st century. Dialogues Clin Neurosci 2015;17(3):327-335 [FREE Full text] [doi: 10.31887/DCNS.2015.17.3/bbandelow] [Medline: 26487813]
5. Hoge EA, Ivkovic A, Fricchione GL. Generalized anxiety disorder: diagnosis and treatment. BMJ 2012;345:e7500. [doi: 10.1136/bmj.e7500] [Medline: 23187094]
6. Calleo J, Stanley MA, Greisinger A, Wehmanen O, Johnson M, Novy D, et al. Generalized anxiety disorder in older medical patients: diagnostic recognition, mental health management and service utilization. J Clin Psychol Med Settings 2009;16(2):178-185 [FREE Full text] [doi: 10.1007/s10880-008-9144-5] [Medline: 19152056]
7. Kendall T, Cape J, Chan M, Taylor C, Guideline Development Group. Management of generalised anxiety disorder in adults: summary of NICE guidance. BMJ 2011;342:c7460 [FREE Full text] [doi: 10.1136/bmj.c7460] [Medline: 21270081]
8. Huffman JC, Niazi SK, Rundell JR, Sharpe M, Katon WJ. Essential articles on collaborative care models for the treatment of psychiatric disorders in medical settings: a publication by the academy of psychosomatic medicine research and evidence-based practice committee. Psychosomatics 2014;55(2):109-122 [FREE Full text] [doi: 10.1016/j.psym.2013.09.002] [Medline: 24370112]
9. Kessler RC. The epidemiology of pure and comorbid generalized anxiety disorder: a review and evaluation of recent research. Acta Psychiatr Scand Suppl 2000;102(s406):7-13. [Medline: 11131470]
10. Kohn R, Saxena S, Levav I, Saraceno B. The treatment gap in mental health care. Bull World Health Organ 2004;82(11):858-866 [FREE Full text] [Medline: 15640922]
11. Collins KA, Westra HA, Dozois DJA, Burns DD. Gaps in accessing treatment for anxiety and depression: challenges for the delivery of care. Clin Psychol Rev 2004;24(5):583-616 [FREE Full text] [doi: 10.1016/j.cpr.2004.06.001] [Medline: 15325746]
12. Stanley MA, Wilson NL, Amspoker AB, Kraus-Schuman C, Wagener PD, Calleo JS, et al. Lay providers can deliver effective cognitive behavior therapy for older adults with generalized anxiety disorder: a randomized trial. Depress Anxiety 2014;31(5):391-401 [FREE Full text] [doi: 10.1002/da.22239] [Medline: 24577847]
13. Farrand P. Low intensity cognitive behavioral therapy: revolution not evolution. In: Farrand P, editor. Low-intensity CBT Skills and Interventions: A Practitioner's Manual. London: SAGE Publications; 2020:5-20.
14. The improving access to psychological therapies manual. National Collaborating Centre for Mental Health. 2022. URL: https://www.england.nhs.uk/wp-content/uploads/2018/06/the-nhs-talking-therapies-manual-v6.pdf [accessed 2023-12-01]
15. Improving Access to Psychological Therapies (IAPT) programme. Nuffield Trust. 2022. URL: https://www.nuffieldtrust.org.uk/public/resource/improving- [accessed 2023-12-01]
16. Marzano L, Hollis C, Cipriani A, Malhi GS. Digital technology: coming of age? Evid Based Ment Health 2017 Nov;20(4):97 [FREE Full text] [doi: 10.1136/eb-2017-102821] [Medline: 29056607]
17. Smartphone ownership is growing rapidly around the world, but not always equally. Pew Research Center. 2019. URL: http://tinyurl.com/3erjjbjt [accessed 2023-12-01]
18. Darden M, Espie CA, Carl JR, Henry AL, Kanady JC, Krystal AD, et al. Cost-effectiveness of digital cognitive behavioral therapy (Sleepio) for insomnia: a Markov simulation model in the United States. Sleep 2021;44(4):zsaa223 [FREE Full text] [doi: 10.1093/sleep/zsaa223] [Medline: 33151330]
19. Bakker D, Kazantzis N, Rickwood D, Rickard N. Mental health smartphone apps: review and evidence-based recommendations for future developments. JMIR Ment Health 2016;3(1):e7 [FREE Full text] [doi: 10.2196/mental.4984] [Medline: 26932350]
20. Dattilio FM, Hanna MA. Collaboration in cognitive-behavioral therapy. J Clin Psychol 2012;68(2):146-158 [FREE Full text] [doi: 10.1002/jclp.21831] [Medline: 23616295]
21. Torous J, Firth J. Bridging the dichotomy of actual versus aspirational digital health. World Psychiatry 2018;17(1):108-109 [FREE Full text] [doi: 10.1002/wps.20464] [Medline: 29352531]
22. Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. Evid Based Ment Health 2018;21(3):116-119 [FREE Full text] [doi: 10.1136/eb-2018-102891] [Medline: 29871870]
23. Bakker D, Rickard N. Engagement with a cognitive behavioural therapy mobile phone app predicts changes in mental health and wellbeing: MoodMission. Aust Psychol 2020;54(4):245-260 [FREE Full text] [doi: 10.1111/ap.12383]
24. Mulder R, Murray G, Rucklidge J. Common versus specific factors in psychotherapy: opening the black box. Lancet Psychiatry 2017;4(12):953-962 [FREE Full text] [doi: 10.1016/S2215-0366(17)30100-1] [Medline: 28689019]
25. Symons Z. Specific factors: the importance of what you do and how you do it. In: Farrand P, editor. Low-intensity Cognitive Behaviour Therapy Skills and Interventions: A Practitioner's Manual. London: SAGE Publications; 2020:79-92.
26. Cuijpers P, Reijnders M, Huibers MJH. The role of common factors in psychotherapy outcomes. Annu Rev Clin Psychol 2019;15:207-231 [FREE Full text] [doi: 10.1146/annurev-clinpsy-050718-095424] [Medline: 30550721]

27. Sieverink F, Kelders S, Poel M, van Gemert-Pijnen L. Opening the black box of electronic health: collecting, analyzing, and interpreting log data. JMIR Res Protoc 2017;6(8):e6452 [FREE Full text] [doi: 10.2196/resprot.6452] [Medline: 28784592]
28. O'Brien HL, Morton E, Kampen A, Barnes SJ, Michalak EE. Beyond clicks and downloads: a call for a more comprehensive approach to measuring mobile-health app engagement. BJPsych Open 2020;6(5):e86 [FREE Full text] [doi: 10.1192/bjo.2020.72] [Medline: 32778200]
29. Lu SC, Xu M, Wang M, Hardi A, Cheng AL, Chang SH, et al. Effectiveness and minimum effective dose of app-based mobile health interventions for anxiety and depression symptom reduction: systematic review and meta-analysis. JMIR Ment Health 2022;9(9):e39454 [FREE Full text] [doi: 10.2196/39454] [Medline: 36069841]
30. Thiblin E, Woodford J, Reuther C, Lundgren J, Lutvica N, von Essen L. Internet-administered, low-intensity cognitive behavioral therapy for parents of children treated for cancer: a feasibility trial (ENGAGE). Cancer Med 2023;12(5):6225-6243 [FREE Full text] [doi: 10.1002/cam4.5377] [Medline: 36404407]
31. Farrand P, Woodford J, Small F. Managing your worries. University of Exeter. 2019. URL: http://tinyurl.com/bdd2z522 [accessed 2023-12-01]
32. Dugas MJ, Gagnon F, Ladouceur R, Freeston MH. Generalized anxiety disorder: a preliminary test of a conceptual model. Behav Res Ther 1998;36(2):215-226 [FREE Full text] [doi: 10.1016/s0005-7967(97)00070-3] [Medline: 9613027]
33. Generalised Anxiety Disorder and Panic Disorder in Adults (CG113). London: National Institute for Health and Care Excellence; 2020.
34. Farrand P, Dawes A, Doughty M, Phull S, Saines S, Winter S, et al. Development and application of criteria to evaluate written CBT self-help interventions adopted by improving access to psychological therapies services. Cogn Behav Ther 2022;15:E28 [FREE Full text] [doi: 10.1017/S1754470X22000241]
35. Mavletova A, Couper MP. A meta-analysis of breakoff rates in mobile web surveys. In: Toninelli D, Pinter R, Pedraza P, editors. Mobile Research Methods: Opportunities and Challenges of Mobile Research Methodologies. London: Ubiquity Press; 2015:81-98.
36. Gulliksen J, Göransson B, Boivie I, Blomkvist S, Persson J, Cajander Å. Key principles for user-centred systems design. Beh Inf Technol 2003;22(6):397-409. [doi: 10.1080/01449290310001624329]
37. Dugas MJ, Ladouceur R. Treatment of GAD. targeting intolerance of uncertainty in two types of worry. Behav Modif 2000;24(5):635-657. [doi: 10.1177/0145445500245002] [Medline: 11036732]
38. Tee J, Kazantzis N. Collaborative empiricism in cognitive therapy: a definition and theory for the relationship construct. Clin Psychol Sci Pract 2011;18(1):47-61. [doi: 10.1111/j.1468-2850.2010.01234.x]
39. Kazantzis N, Beck JS, Dattilio FM, Dobson KS, Rapee RM. Collaborative empiricism as the central therapeutic relationship element in CBT: an expert panel discussion at the 7th international congress of cognitive psychotherapy. Int J Cogn Ther 2013;6(4):386-400. [doi: 10.1521/ijct.2013.6.4.386]
40. Beck JS. Cognitive Behavior Therapy: Basics and Beyond. New York: Guilford Publications; 2020.
41. Williams C, Garland A. A cognitive–behavioural therapy assessment model for use in everyday clinical practice. Adv Psychiatr Treat 2018;8(3):172-179 [FREE Full text] [doi: 10.1192/apt.8.3.172]
42. Thwaites R, Bennett-Levy J. Conceptualizing empathy in cognitive behaviour therapy: making the implicit explicit. Behav Cogn Psychother 2007;35(05):591-612. [doi: 10.1017/s1352465807003785]
43. Neale AV. Behavioural contracting as a tool to help patients achieve better health. Fam Pract 1991;8(4):336-342 [FREE Full text] [doi: 10.1093/fampra/8.4.336] [Medline: 1800196]
44. Bailey RR. Goal setting and action planning for health behavior change. Am J Lifestyle Med 2019;13(6):615-618 [FREE Full text] [doi: 10.1177/1559827617729634] [Medline: 31662729]
45. Deci EL, Ryan RM. Intrinsic Motivation and Self-Determination in Human Behavior. New York: Plenum; 1985.
46. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006;166(10):1092-1097 [FREE Full text] [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
47. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
48. Krägeloh CU, Czuba KJ, Billington DR, Kersten P, Siegert RJ. Using feedback from patient-reported outcome measures in mental health services: a scoping study and typology. Psychiatr Serv 2015;66(3):224-241 [FREE Full text] [doi: 10.1176/appi.ps.201400141] [Medline: 25727110]
49. Brockbank A, McGill I. Facilitating Reflective Learning in Higher Education. New York: Open University Press; 2007.
50. Owen JE, Jaworski BK, Kuhn E, Makin-Byrd KN, Ramsey KM, Hoffman JE. mHealth in the wild: using novel data to examine the reach, use, and impact of PTSD coach. JMIR Ment Health 2015;2(1):e7 [FREE Full text] [doi: 10.2196/mental.3935] [Medline: 26543913]
51. Lambert JD, Elliott LR, Taylor AH, Farrand P, Haase AM, Greaves CJ. A novel method for assessing design fidelity in web-based behavioral interventions. Health Psychol 2021;40(3):217-225. [doi: 10.1037/hea0001046] [Medline: 33630643]
52. Lebeau RT, Davies CD, Culver NC, Craske MG. Homework compliance counts in cognitive-behavioral therapy. Cogn Behav Ther 2013;42(3):171-179 [FREE Full text] [doi: 10.1080/16506073.2013.763286] [Medline: 23419077]

53. Tang W, Kreindler D. Supporting homework compliance in cognitive behavioural therapy: essential features of mobile apps. JMIR Ment Health 2017;4(2):e20 [FREE Full text] [doi: 10.2196/mental.5283] [Medline: 28596145]
54. Weisel KK, Fuhrmann LM, Berking M, Baumeister H, Cuijpers P, Ebert DD. Standalone smartphone apps for mental health-a systematic review and meta-analysis. NPJ Digit Med 2019;2:118 [FREE Full text] [doi: 10.1038/s41746-019-0188-8] [Medline: 31815193]
55. Williams C, McClay CA, Matthews L, McConnachie A, Haig C, Walker A, et al. Community-based group guided self-help intervention for low mood and stress: randomised controlled trial. Br J Psychiatry 2018;212(2):88-95 [FREE Full text] [doi: 10.1192/bjp.2017.18] [Medline: 29436324]
56. Parikh SV, Huniewicz P. E-health: an overview of the uses of the internet, social media, apps, and websites for mood disorders. Curr Opin Psychiatry 2015;28(1):13-17. [doi: 10.1097/YCO.0000000000000123] [Medline: 25420193]
57. Chan AHY, Honey MLL. User perceptions of mobile digital apps for mental health: acceptability and usability—an integrative review. J Psychiatr Ment Health Nurs 2022;29(1):147-168 [FREE Full text] [doi: 10.1111/jpm.12744] [Medline: 33604946]
58. Zhou L, Bao J, Setiawan IMA, Saptono A, Parmanto B. The mHealth App Usability Questionnaire (MAUQ): development and validation study. JMIR Mhealth Uhealth 2019;7(4):e11500 [FREE Full text] [doi: 10.2196/11500] [Medline: 30973342]
59. Lutz J, Offidani E, Taraboanta L, Lakhan SE, Campellone TR. Appropriate controls for digital therapeutic clinical trials: a narrative review of control conditions in clinical trials of digital therapeutics (DTx) deploying psychosocial, cognitive, or behavioral content. Front Digit Health 2022;4:823977 [FREE Full text] [doi: 10.3389/fdgth.2022.823977] [Medline: 36060538]
60. Bleijenberg N, de Man-van Ginkel JM, Trappenburg JCA, Ettema RGA, Sino CG, Heim N, et al. Increasing value and reducing waste by optimizing the development of complex interventions: enriching the development phase of the Medical Research Council (MRC) framework. Int J Nurs Stud 2018;79:86-93 [FREE Full text] [doi: 10.1016/j.ijnurstu.2017.12.001] [Medline: 29220738]

Abbreviations

CA: conversational agent
CBT: cognitive behavioral therapy
GAD-7: Generalized Anxiety Disorder–7
IAPT: improving access to psychological therapies
IMWW: Iona Mind Well-being app for Worry management
LICBT: low-intensity cognitive behavioral therapy
PHQ-8: Patient Health Questionnaire–8

Edited by L Buis; submitted 21.03.23; peer-reviewed by C Rummel-Kluge; comments to author 26.05.23; revised version received 19.06.23; accepted 28.11.23; published 10.01.24.

Please cite as:
Farrand P, Raue PJ, Ward E, Repper D, Areán P
Use and Engagement With Low-Intensity Cognitive Behavioral Therapy Techniques Used Within an App to Support Worry Management: Content Analysis of Log Data
JMIR Mhealth Uhealth 2024;12:e47321
URL: https://mhealth.jmir.org/2024/1/e47321
doi:10.2196/47321
PMID:38029300

Original Paper

A Behaviorally Informed Mobile App to Improve the Nutritional Quality of Grocery Shopping (SwapSHOP): Feasibility Randomized Controlled Trial

Carmen Piernas[1,2], MSc, PhD; Charlotte Lee[1], MSc, DPhil; Alice Hobson[1], MSc; Georgina Harmer[1], MSc; Sarah Payne Riches[1], MPH, DPhil; Michaela Noreik[1,3], MSc, PhD; Susan A Jebb[1], PhD

[1]Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, United Kingdom

[2]Department of Biochemistry and Molecular Biology II, Center for Biomedical Research (CIBM), Institute of Nutrition and Food Technology (INYTA), University of Granada, Granada, Spain

[3]Faculty of Food and Nutrition Sciences, Hochschule Niederrhein, Mönchengladbach, Germany

Corresponding Author:
Carmen Piernas, MSc, PhD
Nuffield Department of Primary Care Health Sciences
University of Oxford
200 Woodstock Rd
Oxford, OX26GG
United Kingdom
Phone: 44 07752599266
Email: carmen.piernas-sanchez@phc.ox.ac.uk

Abstract

Background: Interventions targeting the nutritional quality of grocery shopping have the potential to help improve diet and health outcomes.

Objective: This study aims to assess the feasibility and acceptability of receiving advice on healthier food purchases through SwapSHOP, a behaviorally informed smartphone app that allows users to scan barcodes of grocery products from the United Kingdom, providing nutritional information and personalized swap suggestions to encourage healthier purchases.

Methods: We randomized adult volunteers in a 6-arm parallel-group controlled feasibility trial. Participants used the SwapSHOP app to record their grocery shopping during a 2-week run-in period and were individually randomized in a 3:1 ratio to either intervention or control arms within 3 strata related to a nutrient of concern of their choice: saturated fat (SFA), sugar, or salt. Participants randomized to the intervention received the SwapSHOP app with a healthier swap function, goal setting, and personalized feedback. Participants in the control group were instructed to use a simpler version of the app to log all their food purchases without receiving any guidance or advice. The primary outcome was the feasibility of progression to a full trial, including app use and follow-up rates at 6 weeks. The secondary outcomes included other feasibility outcomes, process and qualitative measures, and exploratory effectiveness outcomes to assess changes in the nutrient content of the purchased foods.

Results: A total of 112 participants were randomized into 3 groups: SFA (n=38 intervention and n=13 control), sugar (n=40 intervention and n=15 control), and salt (n=5 intervention and n=1 control, not analyzed). The 2 progression criteria were met for SFA and sugar: 81% (30/37) and 87% (34/39) of intervention participants in the SFA and sugar groups, respectively, used the app to obtain healthier swaps, and 89% (68/76) of intervention participants and 96% (23/24) of control participants completed follow-up by scanning all purchases over the follow-up period. The process and qualitative outcomes suggested that the intervention was acceptable and has the potential to influence shopping behaviors. There were reductions of −0.56 g per 100 g (95% CI −1.02 to −0.19) in SFA and −1 g per 100 g (95% CI −1.97 to −0.03) in total sugars across all food purchases in the intervention groups.

Conclusions: People were willing to use the SwapSHOP app to help reduce sugar and SFA (but not salt) in their grocery shopping. Adherence and follow-up rates suggest that a full trial is feasible. Given the suggestive evidence indicating that the intervention resulted in reductions in sugars and SFA, a definitive trial is necessary to target improvements in health outcomes.

Trial Registration: International Standard Randomised Controlled Trial Number ISRCTN13022312; https://doi.org/10.1186/ISRCTN13022312

(JMIR Mhealth Uhealth 2024;12:e45854) doi:10.2196/45854

KEYWORDS

swaps; mobile app; supermarket; food purchases; diet; randomized controlled trial; mobile phone

Introduction

Background

Consumption of saturated fat (SFA), sugars, and salt in the United Kingdom and most high-income countries currently exceeds dietary recommendations for good health [1]. These nutrients of concern contribute to the burden of diabetes and cardiovascular disease, either directly or through their effects on blood cholesterol, blood pressure, insulin sensitivity, and body weight [2-8]. Despite decades of nutrition education and efforts to promote healthier behaviors, significant and sustained dietary changes are yet to be observed. Evidence suggests that people need more than general knowledge of dietary recommendations to change behavior [9-11]; however, there is limited evidence for individual-level interventions that are feasible and effective in supporting dietary change at the population level.

Food purchasing is a key antecedent of food consumption, and improving the nutritional quality of food purchases presents a clear opportunity to intervene. Grocery stores account for 71% of the weekly expenditure on food and drinks, including a substantial proportion of foods high in SFA, sugar, and salt [12]. For many foods, there are usually alternatives with less SFA, sugar, or salt, which are functionally similar. This variation can be attributed partly to the natural differences in recipes, such as those found in ready meals or variances in ingredient costs. In addition, it can also be linked to product reformulation driven by consumer demand, such as the introduction of low-fat yogurts, or policy-driven actions, such as changes in soft drink formulations [13].

Systematic reviews have identified some intervention components to support individual dietary change, including providing information, tailored dietary advice, self-monitoring, and personalized feedback [9,14,15]. Specifically, recommending healthier swaps at the point of purchase or as part of individually tailored regular feedback has shown potential to help improve the nutritional quality of grocery shopping [16-20]. Technological advances such as smartphone apps can facilitate this by providing scalable, lower-cost support to help people make healthier choices while shopping.

A systematic review of smartphone apps identified only 2 studies that tested apps that can provide healthier alternatives at the point of choice [21]. Although there is evidence of similar apps in the scientific literature, these previous apps mostly provide healthier swaps to consumers without the option of self-monitoring the quality of their grocery shopping, and many apps have limitations regarding how information is presented to consumers or the limited comprehensiveness of the databases covering the UK food market [22,23]. Our recent proof-of-concept study tested the functionality of the SaltSwap app to help find lower-salt foods when grocery shopping, which included behavioral components such as nutritional information, prompts to lower-salt options, goal setting on swaps, feedback on swaps and salt reduction, and history of swaps [20].

Objectives

For this study, we have further developed the SaltSwap app into SwapSHOP, a new app that also provides swaps for SFA, sugars, and salt. This study aimed to assess the feasibility and acceptability of receiving dietary advice through the SwapSHOP app. It is a stand-alone intervention that allowed users to scan barcodes of grocery products from major UK stores to obtain nutritional information and suggestions for personalized swaps with lower SFA, sugar, or salt and also enabled users to set goals and monitor their shopping behavior.

Methods

Study Design

This was a prospectively registered (International Standard Randomised Controlled Trial Number: ISRCTN13022312) randomized 6-arm parallel-group controlled feasibility trial conducted in the United Kingdom. After giving written informed consent to participate in the study and completing screening and baseline assessment, participants entered a 2-week run-in period where they used a basic version of the SwapSHOP app (no swap or behavioral functionality shown) to record their grocery shopping. They were individually randomized to 1 of the intervention arms or control following a 3:1 ratio (intervention: control) within 3 strata related to a nutrient of concern of their choice (SFA, sugar, or salt). Individuals participated in the study for 6 weeks from screening to the final follow-up.

Ethical Considerations

This study was reviewed and approved by the University of Oxford Medical Sciences Interdivisional Research Ethics Committee (R67216/RE001). Informed consent was obtained from all the participants.

Participants and Procedures

Participants were recruited between April and October 2021 through the community (eg, word of mouth) and through social networks and media, online newsletters and newspapers, and electronic mailing lists. Participants were eligible if they were aged >18 years, were willing and able to give informed consent, were English speaking and able to understand the demands of the study, were looking for support to improve their diet, owned a smartphone with access to the internet and an email account, were willing to download and use a smartphone app to scan and track their grocery shopping for the 6 weeks during their participation in the study, were responsible for at least some of their household grocery shopping, and shopped regularly (eg, at least once a week in a physical store or online at any of the 6 major food retailers in the United Kingdom: Tesco, Sainsbury's, Waitrose, Asda, Morrisons, or Iceland). People were not eligible if they were already on a clinician-supervised diet or a restricted diet, were currently using apps to monitor the quality of their food shopping (excluding apps to track and monitor food intake), were currently participating in another study aimed at dietary change or asking them to change the way

they shop for food, or were planning on going away from home (holiday or other) for >2 consecutive weeks following enrollment in the study.

Eligible participants were invited to complete a baseline web-based questionnaire that collected demographic information, relevant self-reported medical history, and details about their shopping behaviors. Participants also reported the nutrient (SFA, sugar, or salt) they were most concerned about in their diet, which was used as a stratification factor during the randomization process. Participants then entered a 2-week run-in period during which they were asked to record all their food purchases, either by scanning the barcode of purchased grocery products or manually inputting what was purchased using a simpler version of the SwapSHOP app. Only participants who completed this task with good engagement with the app were randomized. Good engagement was defined as scanning products in at least 2 shops, each with products from at least 3 different predefined food categories (eg, fresh meat, chilled ready meals, and soft drinks), and scanning products from ≥5 of the food categories across the 2 weeks.

Randomization and Blinding

After the 2-week run-in period, participants were individually randomized to either an intervention arm or control following a 3:1 ratio (intervention:control) within each stratum depending on their chosen nutrient of concern (sugar, salt, or SFA). Participants who did not complete the baseline data collection were not randomized. Randomization was performed with a computer using a random allocation sequence (concealed to the investigators) and was stratified by the nutrient of concern using block randomization with block sizes of 4 and 8.

It was not possible to blind the participants to the intervention because of the nature of the intervention; however, randomization was performed remotely via computer-generated randomization, and the researchers were not aware of the treatment group until after randomization was complete.

Intervention

Participants randomized to the intervention arms were able to access an enhanced version of the SwapSHOP app, which included a healthier swap function, goal setting, and personalized feedback. Participants scanned the barcode of grocery products to receive nutritional information about the product using the UK traffic light system and were presented with healthier alternatives or swaps that were lower in SFA, sugar, or salt (depending on group allocation). The swaps that appeared after scanning the original product were chosen from a list of products within the same food category, which were ranked from a larger to a smaller reduction in the specific nutrient that was initially chosen. Participants could also set goals for the number of swaps they wanted to make in each shopping trip and record the products they swapped in the app. The app had specific functions to provide feedback on the achievement of their goals and feedback on the overall nutrient reduction achieved by making swaps. The app also had a section to display the ranking of all purchased foods contributing the most to each nutrient of concern (Figure S1 in Multimedia Appendix 1).

The SwapSHOP app included a comprehensive database of >70,000 grocery products available in January 2021 within major UK grocery stores (including Morrisons, Sainsbury's, Tesco, Waitrose, Iceland, and Asda) [24] and a bespoke system for categorizing products and selecting suitable alternative swap suggestions. SwapSHOP was based on a previous version developed exclusively for salt, and its theoretical basis is grounded in the Behavior Change Wheel [20,25], a framework that integrates the elements of capability, opportunity, and motivation, which are key for behavior change. A range of behavior change techniques from the taxonomy groups *goals and planning* and *feedback and monitoring* that were incorporated into this intervention have been associated with successful dietary change [9,26].

To enable assessment of changes in the nutritional composition of the shopping baskets, participants received weekly reminders to continue scanning and recording all their purchased products for the remaining 4 weeks of follow-up (with a minimum of 2 full weeks of grocery shopping data during the follow-up period).

Comparator

Participants randomized to the control arm were asked to continue using the simpler version of the app with no product nutrition information, swaps, or behavioral components, solely to record all their food purchases as part of the outcome assessment at the end of the 6-week follow-up.

Outcome Measures

Primary Outcomes

The primary objective of this study was to determine the feasibility of a larger study to evaluate the effectiveness of an app that offers healthier swaps while grocery shopping to improve the nutritional quality of food purchasing with respect to sugar, salt, or SFA. The primary outcomes were prespecified progression criteria as follows:

1. App use: at least 70% of participants in the active intervention group use the app to obtain swaps on at least 1 occasion by the end of the second week after randomization.
2. Follow-up rate: at least 60% of participants in total (intervention and control) complete follow-up by scanning all their purchases for a minimum of 2 weeks over the entire follow-up period (4 weeks).

Secondary Outcomes

Feasibility outcomes included (1) recruitment rates: total recruited (including number signed up, eligible, consented, and randomized), (2) time needed for recruitment of the final sample, (3) outcome reporting: number of participants who failed to scan their purchases for a minimum of 2 weeks during follow-up, and number of participants who failed to complete the end-of-study questionnaires.

Process Evaluation and Qualitative Outcome Measures

A summary of app-related use (within-app automatic recording) and acceptability measures was collected through the end-of-study questionnaires at follow-up: (1) average number

of shopping trips where the app was used to scan products to obtain a swap per week; (2) number of occasions the app was used to scan products for a swap per trip; (3) number of swaps made overall per week and per shopping trip; (4) nutrient (SFA, sugar, and salt) reduction per swap made; (5) use of specific functionality, for example, goal setting and feedback; (6) end-of-study questionnaires with free text to understand app use and acceptability of the swaps, app functionality (eg, if app scans most products), and if this prompted other behaviors such as reading nutrition labels.

Exploratory Effectiveness Outcomes

Measures included changes in the nutrient content (SFA and sugar in grams per 100 g) of household food purchases recorded in the app between baseline and follow-up in the intervention group compared with the control group.

Statistical Analysis and Sample Size

This study was planned as a feasibility study. A sample size of 120 (n=approximately 30 participants per intervention group and n=approximately 10 participants per control group) was prespecified as sufficient to enable testing of the trial methodology and outcome measures using parametric statistical models. The protocol was prospectively registered (International Standard Randomised Controlled Trial Number: ISRCTN13022312).

Baseline characteristics overall and by trial arm were summarized as means and SDs for continuous variables and counts and percentages for categorical variables. Baseline characteristics were coded as age (years), gender (man, woman, other, or not specified), ethnic group (Asian, Black, White, mixed, or other), education (no formal qualifications, secondary education, or higher education), income (>£15,000, £15,000-£24,999, £25,000-£39,999, £40,000-£75,000, and >£75,000; GBP £1=US $1.38), household size (1, 2-4, and ≥5), grocery shopping habits (eg, spending >£25) on groceries more than once a week, once a week, once a fortnight, once a month, and less than once a month), and relevant self-reported health history (eg, concerns related to weight, high blood pressure, diabetes, and heart disease).

Descriptive statistics were used to present the primary and secondary outcomes using all available data, regardless of whether the participants completed the trial or withdrew.

For the exploratory effectiveness measures, we used data from products purchased over the 2 weekly periods recorded at the beginning and end of the trial, with available information on weight or volume as well as the nutrient content (grams per 100 g) to estimate changes in SFA and sugar for food purchases recorded in the app. The excluded items in this analysis included those categorized as fresh fruits, vegetables, or with no or minimal nutrient content (eg, sugar-free gum) as well as products that were manually entered in the app or those missing nutrient or volume information. We used linear regression models to investigate changes in the nutrient content of food purchases between baseline and follow-up in the intervention and control groups. Tests for linear regression assumptions were conducted and met. The main models included adjustment for baseline values of the dependent variable. Potential confounding because of imbalance in baseline characteristics was investigated, and the following variables were adjusted in sensitivity analyses: age (years), gender (man, woman, other, or not specified), ethnic group (Asian, Black, White, mixed, and other), and income (>£15,000, £15,000-£24,999, £25,000-£39,999, £40,000-£75,000, and >£75,000). StataSE (version 16; StataCorp) was used for all the analyses. A *P* value of <.05 was set to denote statistical significance.

A descriptive analysis of the qualitative outcome measures was conducted using the method by Braun and Clarke [27] for thematic analysis in NVivo 1 software (Lumivero). Each response was line-by-line coded, and codes were inductively constructed based on the aim of the study. We then organized codes into subthemes using the One Sheet of Paper technique and produced top-level themes [28].

Patient and Public Involvement

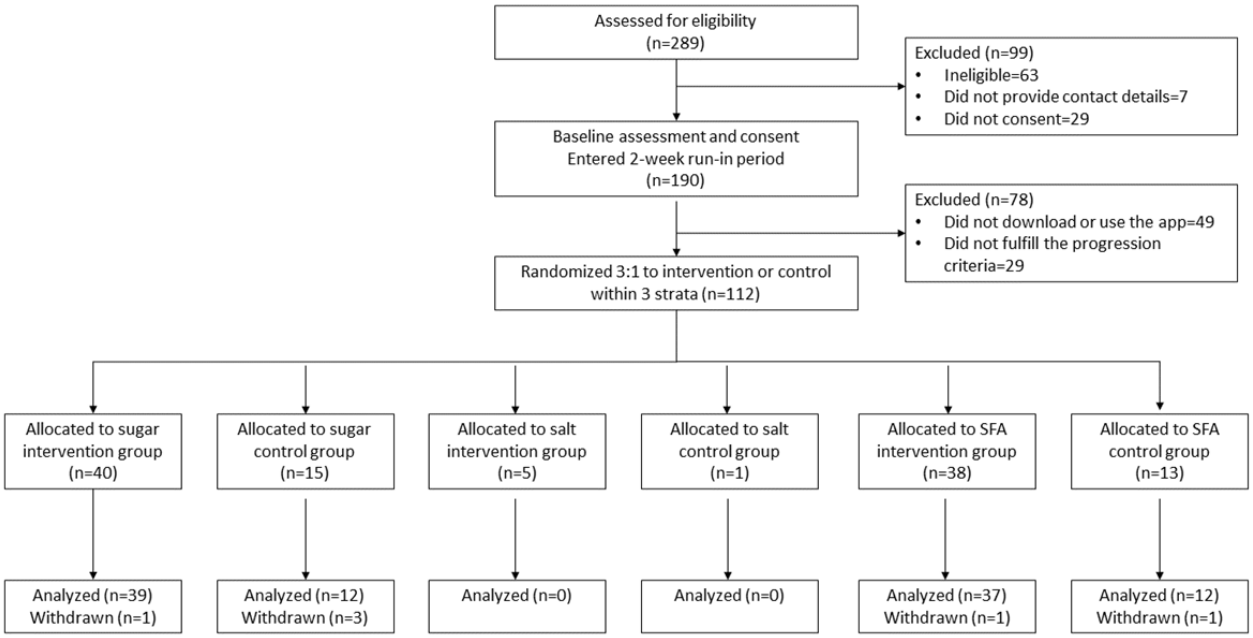
The SwapSHOP app was based on a previous version developed exclusively for salt reduction [20], which was conceived and tested using input from a patient and public involvement panel. People told us that they would also value an app to help change other aspects of their diet. The current app was also beta tested by members of the public recruited from the community.

Results

Participant Characteristics

Study participants were recruited between April and October 2021. Of the 289 interested participants, 190 (65.7%) were eligible, provided consent, and entered the 2-week run-in period. Of these 190 participants, 78 (41%) were excluded because they did not download or use the app (n=49, 63%) or did not fulfill the progression criteria based on their engagement with the app (n=29, 37%). The final sample of 38.8% (112/289) of participants successfully completed the task and were randomized (Figure 1). A total of 51 participants were randomized to the SFA group (n=38 to intervention and n=13 to control), 55 participants were randomized to the sugar group (n=40 to intervention and n=15 to control), and 6 participants were randomized to the salt group (n=5 to intervention and n=1 to control). Of the randomized participants, 100 (89%) completed the study and were analyzed as follows: 49 (96%) in the SFA group and 51 (93%) in the sugar group. A total of 2 participants in the SFA group and 4 participants in the sugar group were withdrawn from the study and excluded from the analysis because their data indicated fraudulent activity (eg, fake phone numbers and implausible shopping patterns). Data from 6 participants who were randomized to the salt group were not analyzed, as this group did not reach the target sample by the end of the recruitment period.

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) diagram. SFA: saturated fat.



Participants in the analytic sample were, on average, aged 35 (SD 12) years, mostly women (80/100, 80%) of White ethnicity (79/100, 79%), and from higher education backgrounds (73/100, 73%; Table 1). Participants mostly lived in households with 2 to 4 other members, shopped once (61/100, 61%) or more than once a week (26/100, 26%), shopped mostly in larger grocery stores (89/100, 98%), or shopped on the internet (39/100, 39%). Most participants looked at SFA or sugar information on nutrition labels always, often, or sometimes and approximately half of the sample (54/100, 54%) reported concerns related to weight among the relevant medical conditions.

Table 1. Baseline characteristics of participants in the analytic sample[a].

	Total population (n=100)	Sugar group		Saturated fat group	
		Control (n=12)	Intervention (n=39)	Control (n=12)	Intervention (n=37)
Age (y), mean (SD)	35 (12)	36 (8)	35 (10)	32 (10)	37 (16)
Gender (woman), n (%)	80 (80)	9 (75)	33 (85)	9 (75)	29 (78)
Ethnic group, n (%)					
Black or Asian	9 (9)	1 (8)	6 (15)	0 (0)	2 (5)
White	79 (79)	10 (83)	27 (69)	12 (100)	30 (81)
Mixed, other, or not specified	12 (12)	1 (8)	6 (15)	0 (0)	5 (14)
Education, n (%)					
No qualifications or not specified	2 (2)	0 (0)	0 (0)	0 (0)	2 (5)
Secondary education	25 (25)	0 (0)	12 (31)	3 (25)	10 (27)
Higher education	73 (73)	12 (100)	27 (69)	9 (75)	25 (68)
Individual income[b], n (%)					
<£15,000	28 (28)	3 (25)	9 (23)	2 (17)	14 (38)
£15,000-£24,999	24 (24)	4 (33)	10 (26)	2 (17)	8 (22)
£25,000-£39,999	31 (31)	4 (33)	10 (26)	8 (67)	9 (24)
£40,000-£75,000	11 (11)	1 (8)	6 (15)	0 (0)	4 (11)
>£75,000	1 (1)	0 (0)	1 (3)	0 (0)	0 (0)
Household size, n (%)					
1	29 (29)	2 (17)	11 (28)	3 (25)	13 (35)
2-4	64 (64)	8 (67)	24 (62)	9 (75)	23 (62)
≥5	7 (7)	2 (17)	4 (10)	0 (0)	1 (3)
Frequency of grocery shopping (≥ £25 per trip), n (%)					
More than once a week	26 (26)	7 (58)	8 (21)	5 (42)	6 (16)
Once a week	61 (61)	4 (33)	26 (67)	7 (58)	24 (65)
Once a fortnight	11 (11)	1 (8)	3 (8)	0 (0)	7 (19)
Once a month	1 (1)	0 (0)	1 (3)	0 (0)	0 (0)
Less than once a month	1 (1)	0 (0)	1 (3)	0 (0)	0 (0)
Type of grocery shop usually visited, n (%)					
Supermarkets	98 (98)	12 (100)	38 (97)	11 (92)	37 (100)
Online supermarkets	30 (39)	5 (42)	11 (28)	4 (33)	10 (27)
Corner shop or convenience store	32 (32)	1 (8)	14 (36)	3 (25)	14 (38)
Greengrocers or fruit and vegetable shop	12 (12)	3 (25)	5 (13)	0 (0)	4 (11)
Butchers or meat market	11 (11)	0 (0)	6 (15)	2 (17)	3 (8)
Other fresh food markets	17 (17)	4 (33)	7 (18)	0 (0)	6 (16)
Looking at salt in nutrition labels, n (%)					
Always or often	14 (14)	2 (17)	3 (8)	5 (42)	4 (11)
Sometimes	24 (24)	3 (25)	12 (31)	1 (8)	8 (22)
Rarely or never	62 (62)	7 (58)	24 (62)	6 (50)	25 (68)
Looking at sugar in nutrition labels, n (%)					
Always or often	42 (42)	2 (17)	19 (49)	7 (58)	14 (38)

	Total population (n=100)	Sugar group		Saturated fat group	
		Control (n=12)	Intervention (n=39)	Control (n=12)	Intervention (n=37)
Sometimes	30 (30)	6 (50)	10 (26)	2 (17)	12 (32)
Rarely or never	28 (28)	4 (33)	10 (26)	3 (25)	11 (30)
Looking at fat in nutrition labels, n (%)					
Always or often	29 (29)	1 (8)	6 (15)	8 (67)	14 (38)
Sometimes	44 (44)	7 (58)	21 (54)	1 (8)	15 (41)
Rarely or never	27 (27)	4 (33)	12 (31)	3 (25)	8 (22)
Relevant health conditions, n (%)					
Concerns related to weight	54 (54)	7 (58)	18 (46)	5 (42)	24 (65)
High blood pressure	5 (5)	0 (0)	2 (5)	0 (0)	3 (8)
Diabetes	11 (11)	2 (17)	4 (10)	1 (8)	4 (11)
Heart disease	4 (4)	0 (0)	1 (3)	1 (8)	2 (5)
Relevant current medications, n (%)					
High blood pressure	4 (4)	0 (0)	1 (3)	0 (0)	3 (8)
Diabetes	6 (6)	0 (0)	2 (5)	1 (8)	3 (8)
Heart disease	1 (1)	0 (0)	0 (0)	1 (8)	0 (0)

[a]The number of participants recruited for the salt group did not reach the target sample size. Data from the participants in this group were not analyzed further.

[b]GBP £1=US $1.38.

Primary Outcomes

The number of participants recruited for the salt group did not reach the target sample size despite additional recruitment attempts, suggesting that a full-scale trial for salt reduction would not be feasible. Data from the participants in this group were not analyzed further.

For participants randomized to the SFA and sugar groups, the 2 progression criteria were met above the set thresholds (Table 2). Overall, most intervention participants (30/37, 81% in the SFA group and 34/39, 87% in the sugar group) used the app to obtain swaps on at least 1 occasion by the end of the second week after randomization. In addition, 89% (68/76) of the intervention participants and 96% (23/24) of the control participants completed follow-up by scanning all purchases for a minimum of 2 weeks over the entire follow-up period.

Table 2. Primary and secondary outcomes—progression criteria and feasibility outcomes[a,b].

	Total population, n (%)			Sugar group, n (%)		Saturated fat group, n (%)	
	Total (N=100)	Control (n=24)	Intervention (n=76)	Control (n=12)	Intervention (n=39)	Control (n=12)	Intervention (n=37)
Primary outcomes—progression criteria							
Participants used the app to obtain healthier swaps on at least 1 occasion by the end of the second week after randomization	N/A[c]	N/A	64 (84)	N/A	34 (87)	N/A	30 (81)
Participants completed follow-up by scanning all purchases for a minimum of 2 weeks over the entire follow-up period	91 (91)	23 (96)	68 (89)	11 (92)	36 (92)	12 (100)	32 (86)
Secondary outcomes—feasibility outcomes							
Signed-up participants	289 (100)	N/A	N/A	N/A	N/A	N/A	N/A
Eligible participants	226 (78)	N/A	N/A	N/A	N/A	N/A	N/A
Consented participants	197 (68)	N/A	N/A	N/A	N/A	N/A	N/A
Completed baseline assessments	141 (49)	N/A	N/A	N/A	N/A	N/A	N/A
Randomized participants	106 (100)	28 (100)	78 (100)	15 (100)	40 (100)	13 (100)	38 (100)
Failed to complete follow-up or end-of-study questionnaire	16 (15)	0 (0)	16 (21)	0 (0)	9 (23)	0 (0)	7 (18)

[a]Percentage of eligible, consented, and completed participants was calculated based on the number of signed-up participants originally; the number of randomized participants (total population) excluded those randomized to the salt group, as these were not analyzed.

[b]The number of participants recruited for the salt group did not reach the target sample size. Data from the participants in this group were not analyzed further.

[c]N/A: not applicable.

Secondary Outcomes

The secondary outcomes provided evidence that a larger trial aiming at reducing sugars or SFA would recruit enough participants that adhere to the trial procedures (ie, completing baseline assessments) within the set time frames and that 84% (16/100 failed to complete follow-up) of those randomized would complete follow-up assessments.

Process Evaluation and Qualitative Outcomes

Participants randomized to the SwapSHOP intervention used the app regularly (Table 3). The average number of shopping trips where the app was used to obtain a swap was 5 in the SFA group (92% of shopping trips) and 5.4 in the sugar group (83% of shopping trips). The average number of occasions when the app was used to scan products for a swap in each shopping trip was 2.5 times in the SFA group and 3.3 in the sugar group. Overall, participants set goals averaging approximately 2 swaps per shopping trip. The results showed that individual product swaps were associated with an average reduction in total sugars (−12.5 g of sugar per 100 g, SD 6.97) or in SFA (−4 g of SFA per 100 g, SD 2.16) in the sugar and SFA groups, respectively.

Table 3. Secondary outcomes—process evaluation measures among participants using the swap function[a,b].

	Sugar group (n=34), mean (SD)	Saturated fat group (n=30), mean (SD)
Average number of shopping trips where the app was used to obtain a swap	5.41 (5.14)	5.07 (3.39)
Percentage of total shopping trips where the app was used to obtain a swap	83.15 (24.74)	91.62 (18.83)
Occasions the app was used to scan products for a swap per shopping trip	3.31 (2.30)	2.49 (1.94)
Average swap goals set in the app per shopping trip	2.14 (1.15)	2.01 (1.54)
Sugar reduction per swap (grams per 100 g)	−12.47 (6.97)	−0.48 (6.70)
Saturated fat reduction per swap (grams per 100 g)	−1.62 (2.57)	−4.02 (2.16)

[a]Sugar and SFA reduction per swap per 100 g were calculated as the average change in nutrient per 100 g of product across all swaps made during the intervention period.

[b]The number of participants recruited for the salt group did not reach the target sample size. Data from the participants in this group were not analyzed further.

The results of the qualitative research relating to the acceptability of the app, the feedback and swaps provided through the app, and the usefulness and comprehensiveness of the app were summarized into 4 themes (Table S1 in Multimedia Appendix 1). Overall, most participants shared positive experiences of the intervention, noting that the app was helpful and the swap suggestions were acceptable. They valued the novelty of the app and the traffic light food labeling system, which encouraged them to read food labels. The key barriers to usability were that some grocery stores were not supported by the app and that there was poor product coverage within the app database. Although participants could manually enter their food purchases to find alternatives, they reported that this was time consuming and a barrier to engagement. Some participants also found it difficult to locate the suggested swaps in the store or were frustrated that the suggested swaps were not available in the store or that the product database in the app did not include own-label products in some stores. One significant obstacle to accepting the intervention was the specificity of certain swap suggestions. These suggestions were not always direct replacements or tailored to personal or household dietary preferences, which made it challenging for shoppers to act upon the prompts. In other instances, some participants found it challenging to reduce one nutritional component if it involved an increase in another, and they expressed a desire for a swap suggestion that recognized the overall healthiness of the product. Participants suggested that more information on the fiber composition of swaps, recommended portion sizes, and price comparison would inform their purchasing behavior. Most participants noted that other visual self-monitoring techniques would improve engagement with the intervention.

Exploratory Effectiveness Outcomes

Baseline and follow-up data on food purchases with available information on volume and nutrient content were available for 86% (44/51) of the participants in the sugar group and 82% (40/49) of the participants in the SFA group. This analysis included all food purchases recorded in the app but excluded purchases that were entered manually (22%) or those with no, minimal, or missing nutritional information; 16%). The degree of missingness in the food purchasing data was comparable between the groups (manually entered products: 23% intervention vs 19% control group and missing nutrient information: 15% intervention vs 16% control group).

There was evidence of changes in the intended direction in both the intervention groups (Table 4). The sugar group reduced total sugars in their grocery purchases by −1 (95% CI −1.97 to −0.03) g/100 g, whereas the control group reduced total sugars in their grocery purchases by 0.32 (−1.47 to 2.11) g/100 g, though the differences between groups were not significant (−0.68, 95% CI −1.94 to 0.58 g/100 g; *P*=.28 adjusted for baseline values). The SFA group reduced total SFA in food purchases by −0.56 (95% CI −1.02 to −0.19) g/100 g, and the control group increased total SFA in food purchases by 0.52 (95% CI −0.19 to 1.22) g/100 g, with a significant between-group difference of −1.05 (95% CI −1.83 to −0.27) g/100 g and *P*=.009 adjusted for baseline values. These results were robust in sensitivity analyses adjusted for age, gender, ethnicity, and income (Table S2 in Multimedia Appendix 1).

Table 4. Secondary outcomes—exploratory effectiveness measures[a,b].

	Baseline, mean (SD)	Follow-up, mean (SD)	Change, mean (95% CI)	Change adjusted[b], mean (95% CI)	Between-group difference[b], intervention vs control	
					Mean (95% CI)	*P* value
Purchased sugar (g/100 g)						
Sugar group (n=34)	5.13 (2.63)	4.13 (2.34)	−1.00 (−1.97 to −0.03)	−0.86 (−1.67 to −0.05)	−0.68 (−1.94 to 0.58)	.28
Control (n=10)	4.26 (1.49)	4.58 (1.86)	0.32 (−1.47 to 2.11)	−0.18 (−1.14 to 0.79)	N/A[c]	N/A
Purchased SFA[d] (g/100 g)						
SFA group (n=28)	2.13 (1.18)	1.58 (0.98)	−0.56 (−1.02 to −0.10)	−0.55 (−0.89 to −0.22)	−1.05 (−1.83 to −0.27)	.009
Control (n=12)	2.10 (0.84)	2.61 (1.22)	0.52 (−0.19 to 1.22)	0.50 (−0.20 to 1.20)	N/A	N/A

[a]Linear regression adjusted for baseline values.

[b]The number of participants recruited for the salt group did not reach the target sample size. Data from the participants in this group were not analyzed further.

[c]Not applicable.

[d]SFA: saturated fat.

Discussion

Principal Findings

Participants in this study were willing and able to use the SwapSHOP app as intended, and the research methods ran as planned, with high levels of adherence and follow-up recorded. This provided preliminary evidence of effectiveness to support dietary change to lower the intake of sugars and SFAs. There was little evidence that this general population sample was motivated to reduce salt intake using the SwapSHOP app. Further improvements to the app, especially enhancing the coverage of product data and the specificity of the swap algorithm, are needed to provide a higher quality user experience.

Comparison With Prior Work

Previous studies have shown the potential of instore swaps to support healthier choices, but there is very little evidence from interventions involving habitual shopping in physical stores. A quasi-experimental study evaluated the impact of the Change4Life Smart Swaps campaign to promote changes to lower fat or sugar foods when grocery shopping, showing that a higher percentage of participants in the intervention group reported choosing healthier options at the end of the study [29]. A total of 3 randomized controlled trials used a smartphone app to help reduce salt intake by promoting lower-salt swaps at the point of choice, showing changes in purchased salt in the intended direction but without evidence of changes in salt intakes [20,23,30]. Our previous study conducted in primary care settings provided individually tailored regular feedback on food shopping and offered lower-SFA swaps to patients with raised low-density lipoprotein cholesterol. The study showed modest but nonsignificant reductions in SFA consumption and SFA in purchased foods [18]. Other randomized controlled trials conducted in online retail environments, both real and experimental platforms, have shown that offering healthier swaps at the point of choice helps improve the nutritional quality of food shopping [16,17,19].

This SwapShop intervention showed promising signs of early effectiveness, with observed reductions of −1 g/100 g (95% CI −1.97 to −0.03) in total sugars and −0.56 (95% CI −1.02 to −0.19) in SFA in all food purchases in their respective intervention group, though this feasibility study was not powered to detect an effect in purchased nutrients. Data also show that smartphone app use is increasing across a wide range of demographic and age groups [31], and consumers are showing an increasing interest in healthier options [32]. However, although the app may be a useful tool for promoting healthier food choices through personalized advice and support, it is unlikely to be sufficient to achieve dietary recommendations. It is plausible that the impact could be enhanced by interventions to encourage product reformulation to offer greater availability of healthier alternatives. Moreover, price, promotions, positioning, and availability strategies within supermarkets have all been found to be major determinants of food choices [33-36], and these structural interventions are likely to be complementary.

Strengths and Limitations

This study's strength lies in its randomized design and the process evaluation used to investigate the intervention's pathway to impact. The qualitative components helped to provide context-specific information about the usability of the app and the acceptability of the swap suggestions. The SwapSHOP app had undergone extensive testing before this trial, given that it is an enhanced version of a previously tested app (SaltSwap) that was specifically designed to help people with hypertension choose lower-salt foods when grocery shopping in combination with face-to-face advice from a health care professional [20,37]. Furthermore, beta testing helped refine the app in line with the intended user feedback. SwapSHOP incorporated several behavioral elements with proven evidence to support dietary change [9,14,15], allowing people to set goals to swap to foods,

lower the nutrients of concern, identify major sources of SFA and sugar in their shopping, and provide feedback on achievement of their goal as well as on the overall nutrient reduction achieved through making swaps. A systematic review of interventions using apps to support dietary behavioral improvements suggested greater benefits of multicomponent interventions compared with single-component interventions [38]. Unlike SwapSHOP, the 2 other existing smartphone apps provided information on healthier swaps but did not offer the option of self-monitoring the nutritional quality of food shopping and had limitations regarding how information is presented to consumers or the limited comprehensiveness of the databases covering the UK food market [21,22].

Although the SwapSHOP app included the major UK grocery retailers, participants particularly reported issues related to the product database, the limited range of UK grocery stores, and the improvements needed in the swapping function to create a list of healthier options that are similar to the original product and to accommodate dietary preferences. These aspects must be addressed in future versions of the app to maximize usability and acceptability. Purchased products that were recorded manually by the participants or had missing nutrient information were excluded from the analysis of changes in total purchases, limiting the robustness of the exploratory effectiveness outcome results, although this should not differ by trial arm.

Another limitation is that the study recruited a small, self-selected sample of people who were motivated to take action to improve their diet quality. A large proportion of participants reported receiving higher education and living in less deprived geographical areas than the national average [39]. A lower socioeconomic status is related to poorer dietary quality and health outcomes; hence, the observed results may not represent a wider population with lower adherence to dietary recommendations [40]. Ethnicity is also related to food choice [41]. Although this study mostly included people of White backgrounds, the app database included a wide range of products available in the UK market, thus covering different dietary preferences that apply to other ethnic groups.

Conclusions

SwapSHOP is a behaviorally informed smartphone app that allows users to scan barcodes of grocery products from major UK supermarkets, providing tailored nutritional information and suggesting personalized swaps to support dietary change. This study provided evidence of feasibility as a stand-alone intervention to support motivated individuals wanting to reduce their SFA or sugar intake as well as preliminary evidence of effectiveness to support healthier food purchases. Given the low cost and scalability of this intervention, after further refinement of the app technology and expanded market coverage, a definitive trial is warranted to assess the potential of this tool to improve health outcomes.

Acknowledgments

This study was funded by the National Institute for Health Research (NIHR) Applied Research Collaborations Oxford. SAJ, MN, and AH were funded by the NIHR Oxford Biomedical Research Centre (BRC) Obesity, Diet and Lifestyle Theme. CP, GH, and SAJ were funded by the NIHR Applied Research Collaborations Oxford. CL was funded by the Engineering and Physical Sciences Research Council and the NIHR Oxford BRC. SPR was supported by a British Heart Foundation Clinical Research Training Fellowship (FS/16/34/32211). CP is currently funded by grant RYC2020-028818-I, MCIN/AEI/10.13039/501100011033 and “ESF Investing in your future” (Ministry of Science and Innovation, Spain). The funders had no role in designing the study, data collection, analysis, interpretation of data, writing the report, or the decision to submit the manuscript for publication. The authors would like to thank Peter Scarborough and Richard Harrington for their help and support in using the FoodDB database of food products included in the SwapSHOP app. The University of Oxford is the owner of the SwapSHOP app.

Data Availability

The data sets generated during and/or analyzed during this study are available on reasonable request. Access to the study data and the statistical code will be reviewed and granted upon request by the Nuffield Department of Primary Care PRimDISC committee (primdisc@phc.ox.ac.uk)

Authors' Contributions

All authors contributed to the concept and design of the study. CP, AH, CL, MN, GH, and SPR developed the study protocols and materials, performed the experiments, and collected the study data. CP and CL analyzed and interpreted the data. CP wrote the manuscript and generated the tables and figures. All authors discussed the results and implications and commented on the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary tables and figures.
[DOCX File , 692 KB - mhealth_v12i1e45854_app1.docx]

Multimedia Appendix 2
CONSORT eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 11373 KB - mhealth_v12i1e45854_app2.pdf]

References

1. NDNS: results from years 7 and 8 (combined). United Kingdom Government. 2018 Mar 16. URL: https://www.gov.uk/government/statistics/ndns-results-from-years-7-and-8-combined [accessed 2023-12-03]
2. Te Morenga LA, Howatson AJ, Jones RM, Mann J. Dietary sugars and cardiometabolic risk: systematic review and meta-analyses of randomized controlled trials of the effects on blood pressure and lipids. Am J Clin Nutr 2014 Jul;100(1):65-79 [FREE Full text] [doi: 10.3945/ajcn.113.081521] [Medline: 24808490]
3. Mozaffarian D, Micha R, Wallace S. Effects on coronary heart disease of increasing polyunsaturated fat in place of saturated fat: a systematic review and meta-analysis of randomized controlled trials. PLoS Med 2010 Mar 23;7(3):e1000252 [FREE Full text] [doi: 10.1371/journal.pmed.1000252] [Medline: 20351774]
4. He FJ, Pombo-Rodrigues S, Macgregor GA. Salt reduction in England from 2003 to 2011: its relationship to blood pressure, stroke and ischaemic heart disease mortality. BMJ Open 2014 Apr 14;4(4):e004549 [FREE Full text] [doi: 10.1136/bmjopen-2013-004549] [Medline: 24732242]
5. Forouhi NG, Krauss RM, Taubes G, Willett W. Dietary fat and cardiometabolic health: evidence, controversies, and consensus for guidance. BMJ 2018 Jun 13;361:k2139 [FREE Full text] [doi: 10.1136/bmj.k2139] [Medline: 29898882]
6. Hooper L, Martin N, Abdelhamid A, Davey Smith G. Reduction in saturated fat intake for cardiovascular disease. Cochrane Database Syst Rev 2020 May 19;5:CD011737. [doi: 10.1002/14651858.CD011737] [Medline: 26068959]
7. Imamura F, O'Connor L, Ye Z, Mursu J, Hayashino Y, Bhupathiraju SN, et al. Consumption of sugar sweetened beverages, artificially sweetened beverages, and fruit juice and incidence of type 2 diabetes: systematic review, meta-analysis, and estimation of population attributable fraction. BMJ 2015 Jul 21;351:h3576 [FREE Full text] [doi: 10.1136/bmj.h3576] [Medline: 26199070]
8. Siervo M, Lara J, Chowdhury S, Oggioni C, Ashor DA, Mathers JC. Effects of dietary approaches to stop hypertension (DASH) diet on cardiovascular risk factors: a systematic review and meta-analysis. Proc Nutr Soc 2015 Apr 15;74(OCE1):E138. [doi: 10.1017/S0029665115001536]
9. Michie S, Abraham C, Whittington C, McAteer J, Gupta S. Effective techniques in healthy eating and physical activity interventions: a meta-regression. Health Psychol 2009 Nov;28(6):690-701. [doi: 10.1037/a0016136] [Medline: 19916637]
10. Teasdale N, Elhussein A, Butcher F, Piernas C, Cowburn G, Hartmann-Boyce J, et al. Systematic review and meta-analysis of remotely delivered interventions using self-monitoring or tailored feedback to change dietary behavior. Am J Clin Nutr 2018 Feb 01;107(2):247-256 [FREE Full text] [doi: 10.1093/ajcn/nqx048] [Medline: 29529158]
11. Celis-Morales C, Livingstone KM, Petermann-Rocha F, Navas-Carretero S, San-Cristobal R, O'Donovan CB, et al. Frequent nutritional feedback, personalized advice, and behavioral changes: findings from the European Food4Me internet-based RCT. Am J Prev Med 2019 Aug;57(2):209-219. [doi: 10.1016/j.amepre.2019.03.024] [Medline: 31248745]
12. Neal B, Sacks G, Swinburn B, Vandevijvere S, Dunford E, Snowdon W, et al. Monitoring the levels of important nutrients in the food supply. Obes Rev 2013 Oct;14 Suppl 1:49-58. [doi: 10.1111/obr.12075] [Medline: 24074210]
13. Pell D, Mytton O, Penney TL, Briggs A, Cummins S, Penn-Jones C, et al. Changes in soft drinks purchased by British households associated with the UK soft drinks industry levy: controlled interrupted time series analysis. BMJ 2021 Mar 10;372:n254 [FREE Full text] [doi: 10.1136/bmj.n254] [Medline: 33692200]
14. Atkins L, Michie S. Designing interventions to change eating behaviours. Proc Nutr Soc 2015 May;74(2):164-170. [doi: 10.1017/S0029665115000075] [Medline: 25998679]
15. Broekhuizen K, Kroeze W, van Poppel MN, Oenema A, Brug J. A systematic review of randomized controlled trials on the effectiveness of computer-tailored physical activity and dietary behavior promotion programs: an update. Ann Behav Med 2012 Oct;44(2):259-286. [doi: 10.1007/s12160-012-9384-3] [Medline: 22767052]
16. Huang A, Barzi F, Huxley R, Denyer G, Rohrlach B, Jayne K, et al. The effects on saturated fat purchases of providing internet shoppers with purchase- specific dietary advice: a randomised trial. PLoS Clin Trials 2006 Sep 22;1(5):e22 [FREE Full text] [doi: 10.1371/journal.pctr.0010022] [Medline: 17013429]
17. Koutoukidis DA, Jebb SA, Ordóñez-Mena JM, Noreik M, Tsiountsioura M, Kennedy S, et al. Prominent positioning and food swaps are effective interventions to reduce the saturated fat content of the shopping basket in an experimental online supermarket: a randomized controlled trial. Int J Behav Nutr Phys Act 2019 Jun 07;16(1):50 [FREE Full text] [doi: 10.1186/s12966-019-0810-9] [Medline: 31174547]
18. Piernas C, Aveyard P, Lee C, Tsiountsioura M, Noreik M, Astbury NM, et al. Evaluation of an intervention to provide brief support and personalized feedback on food shopping to reduce saturated fat intake (PC-SHOP): a randomized controlled trial. PLoS Med 2020 Nov 5;17(11):e1003385 [FREE Full text] [doi: 10.1371/journal.pmed.1003385] [Medline: 33151934]
19. Payne Riches S, Aveyard P, Piernas C, Rayner M, Jebb SA. Optimising swaps to reduce the salt content of food purchases in a virtual online supermarket: a randomised controlled trial. Appetite 2019 Feb 01;133:378-386 [FREE Full text] [doi: 10.1016/j.appet.2018.11.028] [Medline: 30502442]

20. Payne Riches S, Piernas C, Aveyard P, Sheppard JP, Rayner M, Albury C, et al. A mobile health salt reduction intervention for people with hypertension: results of a feasibility randomized controlled trial. JMIR Mhealth Uhealth 2021 Oct 21;9(10):e26233 [FREE Full text] [doi: 10.2196/26233] [Medline: 34673535]
21. Mauch CE, Wycherley TP, Laws RA, Johnson BJ, Bell LK, Golley RK. Mobile apps to support healthy family food provision: systematic assessment of popular, commercially available apps. JMIR Mhealth Uhealth 2018 Dec 21;6(12):e11867 [FREE Full text] [doi: 10.2196/11867] [Medline: 30578213]
22. Dunford E, Trevena H, Goodsell C, Ng KH, Webster J, Millis A, et al. FoodSwitch: a mobile phone app to enable consumers to make healthier food choices and crowdsourcing of national food composition data. JMIR Mhealth Uhealth 2014 Aug 21;2(3):e37 [FREE Full text] [doi: 10.2196/mhealth.3230] [Medline: 25147135]
23. Eyles H, McLean R, Neal B, Jiang Y, Doughty RN, McLean R, et al. A salt-reduction smartphone app supports lower-salt food purchases for people with cardiovascular disease: findings from the SaltSwitch randomised controlled trial. Eur J Prev Cardiol 2017 Sep;24(13):1435-1444 [FREE Full text] [doi: 10.1177/2047487317715713] [Medline: 28631933]
24. Harrington RA, Adhikari V, Rayner M, Scarborough P. Nutrient composition databases in the age of big data: foodDB, a comprehensive, real-time database infrastructure. BMJ Open 2019 Jun 27;9(6):e026652 [FREE Full text] [doi: 10.1136/bmjopen-2018-026652] [Medline: 31253615]
25. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011 Apr 23;6:42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
26. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95 [FREE Full text] [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
27. Braun V, Clarke V. Reflecting on reflexive thematic analysis. Qual Res Sport Exerc Health 2019 Jun;11(4):589-597 [FREE Full text] [doi: 10.1080/2159676x.2019.1628806]
28. Ziebland S, McPherson A. Making sense of qualitative data analysis: an introduction with illustrations from DIPEx (personal experiences of health and illness). Med Educ 2006 May;40(5):405-414. [doi: 10.1111/j.1365-2929.2006.02467.x] [Medline: 16635119]
29. Wrieden WL, Levy LB. 'Change4Life Smart Swaps': quasi-experimental evaluation of a natural experiment. Public Health Nutr 2016 Sep;19(13):2388-2392 [FREE Full text] [doi: 10.1017/S1368980016000513] [Medline: 27002189]
30. Dorsch MP, Cornellier ML, Poggi AD, Bilgen F, Chen P, Wu C, et al. Effects of a novel contextual just-in-time mobile app intervention (LowSalt4Life) on sodium intake in adults with hypertension: pilot randomized controlled trial. JMIR Mhealth Uhealth 2020 Aug 10;8(8):e16696 [FREE Full text] [doi: 10.2196/16696] [Medline: 32663139]
31. Adults' media use and attitudes report 2022. Ofcom. 2022 Mar 30. URL: https://www.ofcom.org.uk/__data/assets/pdf_file/0020/234362/adults-media-use-and-attitudes-report-2022.pdf [accessed 2022-09-24]
32. Ogundijo DA, Tas AA, Onarinde BA. Factors influencing the perception and decision-making process of consumers on the choice of healthier foods in the United Kingdom: a systematic review using narrative synthesis. Int J Food Sci Technol 2021 Nov 25;57(2):881-897. [doi: 10.1111/ijfs.15478]
33. Hartmann-Boyce J, Bianchi F, Piernas C, Payne Riches S, Frie K, Nourse R, et al. Grocery store interventions to change food purchasing behaviors: a systematic review of randomized controlled trials. Am J Clin Nutr 2018 Jun 01;107(6):1004-1016 [FREE Full text] [doi: 10.1093/ajcn/nqy045] [Medline: 29868912]
34. Cameron AJ, Charlton E, Ngan WW, Sacks G. A systematic review of the effectiveness of supermarket-based interventions involving product, promotion, or place on the healthiness of consumer purchases. Curr Nutr Rep 2016 Jul 18;5:129-138. [doi: 10.1007/s13668-016-0172-8]
35. Shaw S, Ntani G, Baird J, Vogel CA. A systematic review of the influences of food store product placement on dietary-related outcomes. Nutr Rev 2020 Dec 01;78(12):1030-1045 [FREE Full text] [doi: 10.1093/nutrit/nuaa024] [Medline: 32483615]
36. Karpyn A, McCallops K, Wolgast H, Glanz K. Improving consumption and purchases of healthier foods in retail environments: a systematic review. Int J Environ Res Public Health 2020 Oct 16;17(20):7524 [FREE Full text] [doi: 10.3390/ijerph17207524] [Medline: 33081130]
37. Payne Riches S, Piernas C, Aveyard P, Sheppard JP, Rayner M, Jebb SA. The salt swap intervention to reduce salt intake in people with high blood pressure: protocol for a feasibility randomised controlled trial. Trials 2019 Oct 11;20(1):584 [FREE Full text] [doi: 10.1186/s13063-019-3691-y] [Medline: 31604477]
38. Schoeppe S, Alley S, Van Lippevelde W, Bray NA, Williams SL, Duncan MJ, et al. Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: a systematic review. Int J Behav Nutr Phys Act 2016 Dec 07;13(1):127 [FREE Full text] [doi: 10.1186/s12966-016-0454-y] [Medline: 27927218]
39. Index of multiple deprivation 2019. JSNA Bitesize. 2019 Oct. URL: https://insight.oxfordshire.gov.uk/cms/system/files/documents/20191004_bitesize_IMD2019overview.pdf [accessed 2022-07-01]
40. Yau A, Adams J, Monsivais P. Time trends in adherence to UK dietary recommendations and associated sociodemographic inequalities, 1986-2012: a repeated cross-sectional analysis. Eur J Clin Nutr 2019 Jul;73(7):997-1005 [FREE Full text] [doi: 10.1038/s41430-018-0347-z] [Medline: 30446763]

41. Leung G, Stanner S. Diets of minority ethnic groups in the UK: influence on chronic disease risk and implications for prevention. Nutr Bull 2011 May;365(2):161-198. [doi: 10.1111/j.1467-3010.2011.01889.x]

Abbreviations

SFA: saturated fat

Edited by L Buis; submitted 19.01.23; peer-reviewed by N Mungoli, E Dunford; comments to author 16.03.23; revised version received 29.04.23; accepted 27.11.23; published 11.01.24.

Please cite as:
Piernas C, Lee C, Hobson A, Harmer G, Payne Riches S, Noreik M, Jebb SA
A Behaviorally Informed Mobile App to Improve the Nutritional Quality of Grocery Shopping (SwapSHOP): Feasibility Randomized Controlled Trial
JMIR Mhealth Uhealth 2024;12:e45854
URL: https://mhealth.jmir.org/2024/1/e45854
doi:10.2196/45854
PMID:38206671

Original Paper

Effectiveness of Metaverse Space–Based Exercise Video Distribution in Young Adults: Randomized Controlled Trial

Rami Mizuta[1], MSc, PT; Noriaki Maeda[1], PT, PhD; Tsubasa Tashiro[1*], PT, PhD; Yuta Suzuki[2], PT, PhD; Sayo Kuroda[1], BA, PT; Ayano Ishida[1], BA, PT; Sakura Oda[1], BA, PT; Tomoya Watanabe[1], BA, PT; Yuki Tamura[1], BA, PT; Makoto Komiya[1], PT, PhD; Yukio Urabe[1*], PT, PhD

[1]Department of Sport Rehabilitation, Graduate School of Biomedical and Health Sciences, Hiroshima University, Hiroshima, Japan

[2]Department of Physical Therapy, Faculty of Rehabilitation, Kyushu Nutrition Welfare University, Fukuoka, Japan

[*]these authors contributed equally

Corresponding Author:
Yukio Urabe, PT, PhD
Department of Sport Rehabilitation
Graduate School of Biomedical and Health Sciences
Hiroshima University
1-2-3 Kasumi, Minami-ku
Hiroshima, 734-8553
Japan
Phone: 81 082 257 5405
Email: yurabe@hiroshima-u.ac.jp

Abstract

Background: In response to the serious lack of physical activity among young adults, recent attempts have been made to encourage young people to exercise through exercise video distribution. However, merely distributing videos does not lead to improved physical activity levels. Metaverse space, which enables web-based interaction through avatars, allows users to watch exercise videos in the same space as other avatars.

Objective: This study explored whether exercise video distribution using metaverse space is effective in improving physical activity levels, along with mental health and locomotive function, among young people.

Methods: In this parallel-group randomized controlled trial participants were recruited using printed poster displays. A total of 48 young adults aged between 18 and 30 years were assigned to 3 groups of 16 each: the metaverse, YouTube, and control group. To encourage exercise, the metaverse group was given an exercise video each week with a load of around 4-8 metabolic equivalents of tasks (METs) for 8 videos delivered in the metaverse space. The YouTube group was sent a URL on YouTube every week to view exercise videos with the same content as the metaverse group. The control group was given no special instructions. The intervention period was 8 weeks. Pre- and postintervention physical activity, well-being, locomotive syndrome risk tests, and social capital were measured. Although this study was not blinded to the participants, the measurers did not know to which group the participants belonged. Mixed model repeated-measures analyses and a post hoc Wilcoxon signed rank sum test were performed to detect the effects of the intervention in all groups.

Results: The results of the mixed model repeated-measures analyses showed a significant interaction between groups and before and after the intervention for total physical activity (metaverse group: pre 737.1, SD 609.5 METs/week, post 1575.4, SD 1071.8 METs/week; YouTube group: pre 661.7, SD 710.7 METs/week, post 911.9, SD 1103.3 METs/week; and control group: pre 930.6, SD 665.1 METs/week, post 844.7, SD 701.8 METs/week; P=.04) but none for the indicators of well-being (P=.40), locomotive function scale (P=.17), and social capital (P=.23). A post hoc test showed a significant increase in physical activity in the metaverse group before and after the intervention (P=.006).

Conclusions: This study is the first to show that delivering exercise videos through metaverse space can increase physical activity in young adults by providing a gathering space for individuals similarly motivated for exercise practice. This way, the sense of isolation during exercise is reduced compared with merely distributing videos on YouTube. The use of metaverse space in health promotion is likely to spread, and this study provides a useful reference for its exploration.

Trial Registration: ClinicalTrials.gov NCT06019156; https://ichgcp.net/clinical-trials-registry/NCT06019156

(JMIR Mhealth Uhealth 2024;12:e46397) doi:10.2196/46397

KEYWORDS

exercise video distribution; exercise; metaverse; physical activity; web-based intervention

Introduction

Physical inactivity among young people is a serious problem, with 40.3% of men and 66.1% of women in their 20s reporting not exercising even once a week, according to the National Health and Nutrition Survey, 2019 [1]. Surprisingly, a study of 100 college students aged between 18 and 23 years found that 65% of them were in the high-risk group for locomotive syndrome, a condition that reduces physical function and mobility such as standing and walking. A study of young Japanese adults aged between 18 and 20 years revealed that 45.9% of them had a prevalence of frailty-prefrailty [2]. As Japan's aging society progresses, the number of people requiring nursing care also increases, and the reasons for this need are often locomotive disorders such as falls and joint diseases [3,4]. Therefore, establishing exercise habits at a young age and maintaining and improving locomotive function are important.

The spread of COVID-19 has led to a lack of exercise and a change to a more sedentary lifestyle, which has focused attention on exercise interventions over home-based web-based video delivery methods. This promotion of physical activity through exercise videos was intended to help address people's lack of motivation to engage in physical activity when gymnasiums and recreation centers had to close and physical distance was required [5]. An exercise intervention study using YouTube reported that when a new exercise video of approximately 5 minutes was delivered once a week for 12 weeks, physical activity and frequency of muscle training increased in the exercise intervention group [6]. Hence, the distribution of exercise videos through YouTube, the most accessible tool for young people, has shown some effectiveness and has become commonplace even as of 2023, when COVID-19–related restrictions have eased.

However, this conventional promotion of physical activity through the distribution of exercise videos has elements that fail to encourage behavioral changes. Self-determination theory, as it relates to exercise implementation, begins with the fulfillment of 3 basic human psychological needs: competency (the experience of achievement and efficacy), autonomy (the experience of motivation), and relatedness (the experience of connection with others) [7]. However, traditional exercise video distribution using YouTube and other media has focused on individual video-watching, which fails to provide opportunities for connecting with other people. In fact, another study of an exercise intervention using YouTube reported improvements in mental health with no increase in physical activities [8]. Studies showed connection with others as an important factor in exercise implementation and that being in an exercise group was more effective in increasing physical activity than exercising alone [9,10]. Further, social encouragement by those around an individual and exercising together had a positive impact on the acquisition of exercise habits [11,12]. In addition, people who shared their step count information using a special app tended to exercise the next day when other people "liked" or commented on their step counts [13]. Moreover, the group that was provided with peer interaction and social support showed positive effects on locomotive functions, such as improved results in stand-up tests, compared with the group that lacked these supportive activities [14]. Thus, adding an element of social community when promoting exercise is more effective.

The metaverse is an internet-based 3D virtual world. For the generation of digital natives born in the 1990s and 2000s, it is expected that the metaverse will increasingly become a space where they spend part of their daily lives [15]. In recent years, the metaverse space has gained attention in the health care field [16-18], and a bibliometric analysis of virtual and augmented reality showed that the metaverse has been adapted for diagnosis, surgical treatment, and rehabilitation of pain, stroke, anxiety, depression, fear, cancer, and neurodegenerative diseases, with satisfactory results [16]. In a metaverse space, people can interact with and encourage others on the web through their avatars, which are their own alter egos, and create a community of people who gather in the same space. Social distance is known to interfere with mental health [19], and the metaverse's ability to recreate a space for social interaction is expected to help improve the treatment of mental health symptoms such as anxiety, stress, and eating disorders [20-23]. Therefore, the use of the metaverse has the potential to improve mental health. Moreover, encouraging physical activity within the metaverse space has the potential to promote behavioral responses from users, such as motivating people to adopt a healthier lifestyle [24]. Therefore, exercise training using the metaverse is expected to develop as a method to promote physical activity among young people [16]; however, there is insufficient knowledge on whether physical activity can be promoted by distributing exercise videos in the metaverse space. This study verified whether exercise video distribution using an app that provides metaverse space is effective in improving physical activity levels, mental health, and locomotive function among young people. We hypothesized that the distribution of exercise videos in the metaverse space would increase physical activity and have positive effects related to mental health and locomotive function.

Methods

Study Design and Participants

This study was a parallel-group, randomized controlled trial. The participants were recruited among Hiroshima University students between August 10 and September 9, 2022. For recruitment, we displayed a poster explaining the study and indicated a contact person available for questions for those considering participation. Several people expressed their willingness to participate. Participants were offered an incentive of US $6.93. Written consent was obtained after we explained the study details face-to-face. All data were collected in person.

This study's specific participation criteria were as follows: individuals (1) aged 18 and 30 years at the time of obtaining consent, (2) who owned a smartphone or computer and had internet access, and (3) who did not belong to an exercise community such as a sports club. The exclusion criteria were as follows: individuals (1) who had a history of a disease that prohibited exercise, (2) whose physical activity was >3000 metabolic equivalents of tasks (METs)/week because a person with >3000 METs/week is defined as active [25-27], and (3) who were (or possibly) pregnant.

The sample size was calculated using G*power 3.1.9.2 (version 3.1.9.7; Heinrich-Heine-University Düsseldorf). When the effect size was calculated using raw data from a previous study with a similar design, which used a web-based exercise intervention with a physical activity measure as the main outcome [28], a large effect size of 0.34 was obtained with a partial η^2 of 0.108. When setting α error probability to .05, power (1–β error probability) to .8, number of groups to 3, and number of measurements to 2, the required sample size was 27 participants, with 9 participants per group. A total of 48 participants were required when the dropout rate was approximately 40% [8].

After completion of the baseline prequestionnaire and measurements, 3 participants who met the exclusion criteria and were physically active for >3000 METs/week were excluded. The participants were divided into 3 groups: "metaverse space-based exercise video distribution group (metaverse group)," "YouTube-based exercise video distribution group (YouTube group)," and "no video distribution group (control group)" in a 1:1:1 ratio. The allocation was performed by segregating blocks of size 3 to generate the sequence [29], and the allocation order was hidden until after each group was allocated. A total of 3 separate staff members performed each of the 3 tasks: generating the random allocation sequence, enrolling participants, and assigning participants to the interventions. We explained to all participants the nature of the 3 groups while obtaining their consent. After allocation, the metaverse and YouTube groups received an 8-week intervention in which the exercise videos were distributed, and a control group was established. The intervention period was from October 3 to November 27, 2022. Thereafter, a postquestionnaire survey and measurements were conducted for all 3 groups.

Intervention

Procedure

The recruitment ran for approximately a month, from August 10 to September 9, 2022, and the premeasurements were taken during a 2-week period from September 12 to 25, 2022. The 8-week intervention was conducted from October 3 to November 27, 2022. The postmeasurements were conducted over a period of 1 week (from December 5 to 11, 2022).

Metaverse Group

The participants in the metaverse group had access to Metaverse space (Spatial Systems Inc), where they could watch exercise videos whenever they wanted (Figure 1). The researcher created a metaverse space for watching exercise videos and set up new videos each week. The URL to access them was shared only with participants in the metaverse group. The unique feature of the metaverse was that users could enter a room to watch exercise videos with their avatars, which were their own alter egos and other avatars. The participants watched exercise videos, which were created by a physical therapist to increase their physical activity, for approximately 5 minutes [6,8]. Then, we instructed them to do the exercise while watching videos. The load setting was approximately 4-8 METs to train the trunk and the upper and lower extremities. A new video was released once a week [6], and the participants could watch 8 videos for 8 weeks. They were informed in advance that new videos would be updated every Monday and, as a reminder, were sent a SMS text message when a new video was available. A web-based survey regarding exercise video viewing frequency was administered, and adherence was tracked weekly.

Figure 1. The metaverse space in this study.

YouTube Group

The YouTube group was sent a message with the URL to access the exercise video every Monday. The YouTube group could watch videos at their preferred times by clicking on a link. The content and time of the video were identical to those of the metaverse group. As in the metaverse group, adherence to exercise implementation was confirmed through a web-based questionnaire, and video update reminders were sent every week.

Control Group

Participants were informed by email that they were assigned to the control group simultaneously with the start of the intervention in the Metaverse and YouTube groups. We asked the participants in the control group to spend 8 weeks as usual, without any special instructions or delivery of videos. This group was established as an indicator of the amount of physical activity that may be significantly affected by the infection status and government directives during the COVID-19 pandemic.

Outcome Measurements

Basic Information

The participants were asked and tested for age, height, weight, BMI, lifestyle, sex, living status, and frailty to characterize the basic information. The weight and frailty tests were measured by a physical therapist, and the other parameters were measured using a questionnaire.

Lifestyle

Breslow's 7 health practice scores were used to assess lifestyle [30,31]. The 7 parameters were: smoking (presence or absence), alcohol intake (high or low: consumption of ≥5 or <5 alcoholic beverages), sleep (≥7 hours or <7 hours), breakfast (eating or not eating), eating between meals (eating or not eating), exercise (≥2 times, 30 minutes/week or <2 times, 30 minutes/week), and proper weight (BMI <25 kg/m^2 or ≥25 kg/m^2).

Living Status

Regarding their living status, the participants were asked whether they lived alone or with others.

Frailty Test

The Japanese Cardiovascular Health Study criteria questionnaire was designed to indicate potential frailty, and its contents are as follows [32,33]: (1) have you lost 2 kg or more in the past 6 months? (2) Do you engage in moderate levels of physical exercise or sports aimed at improving health? (3) In the past 2 weeks, have you felt tired without reason? (4) Hand grip strength of <26 kg in men or <18 kg in women, and (5) gait speed of <1 m/second. Grip strength was measured using a hand dynamometer (TKK 5401; Takei). Each participant performed 2 trials, and the best value from both trials was used for the analysis. Walking speed was calculated by performing a 10-meter walk test and dividing the measured value by 10 to obtain the results per meter. All questions were rated 1 or 0, and higher scores indicated a higher possibility of frailty. Frailty, prefrailty, and robustness were defined by the number of applicable items: 3-5, 1-2, and 0.

Primary and Secondary Outcomes

The study assessed physical activity, well-being, locomotive syndrome risk tests, and social capital before and after the intervention for 8 weeks.

Primary Outcome: Physical Activity

The short form of the International Physical Activity Questionnaire was used to measure the effect of an intervention using exercise videos with different distribution methods (Metaverse and YouTube) on physical activity [34]. Total physical activity was measured as the average amount of vigorous physical activity, the amount of moderate activity, and the amount of walking (METs minutes/week). A MET is defined as the amount of energy required for a person to sit quietly. In the premeasurement, the participants answered for the most recent week, and in the postmeasurement, they responded for the week after the intervention was completed.

Secondary Outcomes: Well-being, Locomotive Syndrome Risk Tests, and Social Capital

Well-Being

To assess well-being and mental health, we used the Japanese version of the World Health Organization-Five Well-being Index [35]. Responses to each item were rated on a 6-point scale from 0 to 5 and consisted of 5 questions. The highest score is 25, with higher scores indicating better well-being.

Locomotive Syndrome Risk Test

We used the 25-question geriatric locomotive function scale, proposed by the Japanese Orthopedic Association [33,36], to measure levels of locomotive function based on normal daily activities and pain. All the items were answered on a scale of 0-4. The higher the score, the more impaired the motor function, which indicates a decline in mobility and interference with social life.

Social Capital

The social capital questions consisted of 11 items related to civic participation, social cohesion, and reciprocity, with a perfect score of 11 and higher scores indicating stronger social connections [37]. We included this question item because of the possibility that the sense of social participation could change in the metaverse group.

Supplemental Outcomes: Impression of Each Intervention (Metaverse and YouTube Groups)

We interviewed each of the metaverse and YouTube groups regarding their impressions of the intervention using exercise videos after the 8-week intervention. User experience is the experience of a user with the use of a product, system, or service. From the review of user experience, we selected the following list of questions: novelty, relatedness (connection with others), motivation, excitement, satisfaction, delight, comfort, attractiveness, expectation, and fulfillment [38]. The participants answered each question with a "good," "rather good," "rather poor," or "poor" response.

Statistical Analysis

Primary and secondary outcomes were analyzed by performing a mixed model repeated-measures ANOVA. This analysis was used to detect the effect of interventions between the metaverse, YouTube, and control groups. We assessed the main effect as well as group and time interactions on the outcome measure. A mixed model repeated-measures analysis is an intention-to-treat analysis with unbiased estimates that considers all available data from participants enrolled in the study [39]. The effect size *r* for the interaction effect of a mixed model repeated-measures analysis was calculated using F statistics. As a posttest, the Wilcoxon signed rank sum test was performed after confirming nonnormality to compare the pre- and postintervention outcomes of each group. The z-statistic was used to calculate the effect size *r*. Supplemental outcomes were analyzed by performing chi-square tests to compare the intervention impressions between the metaverse and YouTube groups. Note that to use the chi-square test, “rather good” was included in “good” and “rather poor” was included in “poor,” thus reorganizing responses into 2 groups. The effect size was calculated using Cramer *V*.

Statistical analyses were performed using SPSS (version 28.0; IBM Corp). The significance level was set at 5%.

Ethical Considerations

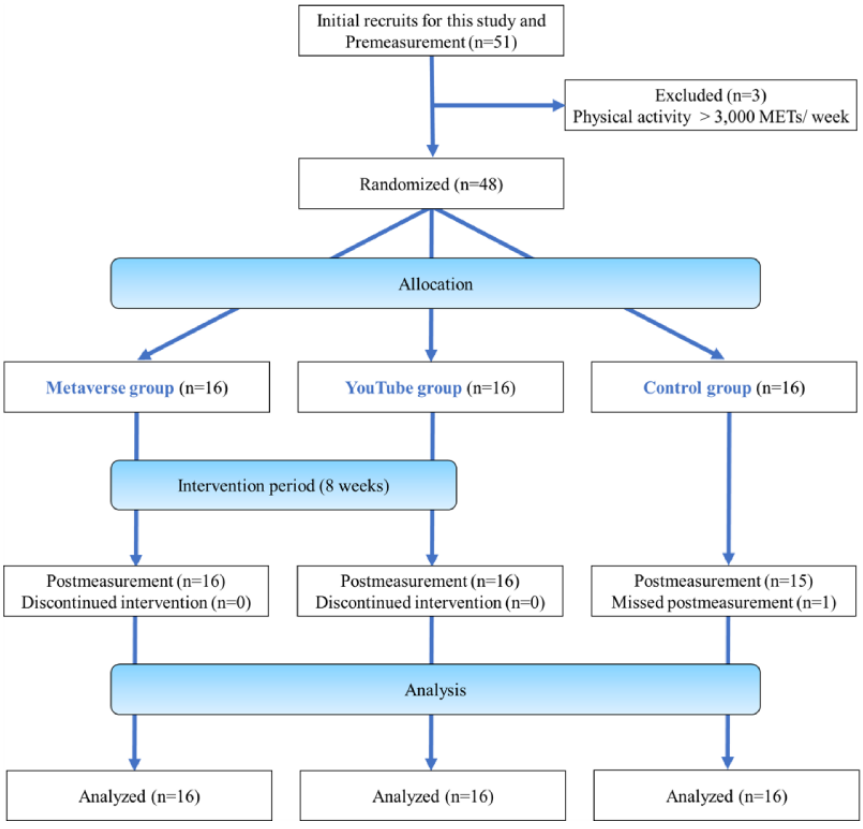
This study was approved by the Ethical Committee for Clinical Research at Hiroshima University (C2022-0004) and registered with UMIN (UMIN000048046). We followed the guidelines of the Consolidated Standards for Reporting Trials [40]. There were no significant changes in the method used after the study’s initiation.

Results

Participants

Figure 2 shows the participant flowchart. A total of 51 individuals expressed interest in the study; however, 3 were excluded because their physical activity was >3000 METs/week. All the participants in the metaverse and YouTube groups completed the intervention. Participants were also interviewed weekly during the 8-week intervention period to monitor their exercise implementation. The average frequency of exercise while watching the exercise video per week was 4.1 (SD 3.9) and 2.6 (SD 1.6) for the metaverse and YouTube groups, respectively (response rates: metaverse group 100% and YouTube group 99%). For the postmeasurement, a participant in the control group could not participate owing to illness.

Figure 2. Study flowchart. MET: metabolic equivalents of task.



Basic information, Lifestyle, Living Status, and Frailty Test

Table 1 presents the participants’ demographic characteristics (mean 22.4, SD 2.4 years; 32/48, 67% women). Prefrailty accounted for 41/48 (85%) participants because we recruited those who were not members of a sports club and whose physical activity was <3000 METs/week.

Table 1. Participant's demographics at baseline.

Variables	Total	Metaverse group (n=16)	YouTube group (n=16)	Control group (n=16)
Age (years), mean (SD)	22.4 (2.4)	22.9 (1.4)	22.5 (2.0)	22.0 (3.4)
Height (cm), mean (SD)	164.0 (7.9)	162.4 (6.6)	163.9 (6.2)	165.7 (10.3)
Weight (kg), mean (SD)	57.7 (11.6)	56.7 (9.8)	57.3 (10.2)	59.0 (14.8)
BMI (kg/m²), mean (SD)	21.3 (2.9)	21.4 (2.5)	21.2 (2.5)	21.3 (3.6)
Lifestyle, mean (SD)	4.6 (0.9)	4.6 (0.9)	4.4 (0.7)	4.7 (1.1)
Sex, n (%)				
Male	16 (33)	6 (38)	6 (38)	5 (31)
Female	32 (67)	10 (63)	10 (63)	11 (69)
Living status, n (%)				
Alone	31 (65)	11 (69)	10 (63)	10 (63)
With others	17 (35)	5 (31)	6 (38)	6 (38)
Frailty test, n (%)				
Frailty	2 (4)	1 (6)	1 (6)	0 (0)
Prefrailty	41 (85)	13 (81)	14 (88)	14 (88)
Robust	5 (10)	2 (13)	1 (6)	2 (13)

Physical Activity, Well-Being, Locomotive Syndrome Risk Test, and Social Capital

Table 2 shows the estimation results of the fixed effects from the mixed model repeated-measure analysis. Only total physical activity showed a significant interaction ($F_{2, 45.042}$=3.338, P=.04; effect size=0.263). The main effects for the time showed a significant difference in vigorous activity ($F_{1, 45.083}$=6.921, P=.01), locomotive function scale ($F_{1, 44.895}$=9.557, P=.003), and social capital ($F_{1, 44.650}$=5.085, P=.03).

Table 3 shows the results of the post hoc test comparison of the difference in total physical activity before and after the intervention in each group. In the metaverse group, a significant difference was detected between pre- and postintervention periods (P=.006; effect size=0.682).

Table 2. Estimate of the fixed effects from the mixed model repeated-measures analysis, (N=16).

Variables	Metaverse group	YouTube group	Control group	Main effect				Interaction effect		
				Time		Group		Time*Group		
				F test (*df*)	*P* value	*F* test (*df*)	*P* value	*F* test (*df*)	*P* value	Effect size
Physical activity: total (METs[a] minutes/week), mean (SD)				5.153 (1, 45.051)	.03	1.383 (2, 45.297)	.26	3.338 (2, 45.042)	.04	0.263
Pre	737.1 (609.5)	661.7 (710.7)	930.6 (665.1)							
Post	1575.4 (1071.8)	911.9 (1103.3)	844.7 (701.8)							
Physical activity: vigorous (METs minutes/week), mean (SD)				6.921 (1, 45.083)	.01	3.140 (2, 45.257)	.053	1.917 (2, 45.075)	.16	0.202
Pre	237.5 (423.4)	120 (231.9)	137.5 (253.8)							
Post	745 (677.4)	345 (613.90)	162.7 (493.3)							
Physical activity: moderate (METs minutes/week)				2.004 (1, 44.498)	.16	0.994 (2, 44.623)	.38	0.748 (2, 44.490)	.48	0.129
Pre	211 (270.3)	162.5 (319.2)	311.3 (458.2)							
Post	425 (347.1)	235 (411.5)	317.6 (351.9)							
Physical activity: walking (METs minutes/week)				0.013 (1, 45.119)	.91	0.192 (2, 45.253)	.83	1.488 (2, 45.110)	.24	0.179
Pre	288.6 (255.8)	379.2 (441.9)	481.8 (552.1)							
Post	636.5 (686.8)	331.9 (368.1)	364.4 (284.0)							
Well-being				2.449 (1, 44.726)	.12	1.189 (2, 45.186)	.31	0.934 (2, 44.718)	.40	0.143
Pre	15.4 (3.9)	16.4 (4.1)	18.4 (4.1)							
Post	17.6 (5.5)	17.3 (3.9)	18.5 (4.8)							
Locomotive function scale				9.557 (1, 44.895)	.003	0.429 (2, 45.308)	.65	1.821 (2, 44.887)	.17	0.197
Pre	3.9 (3.4)	3.8 (3.3)	3.5 (3.7)							
Post	1.1 (1.9)	2.6 (2.9)	2.9 (3.3)							
Social capital				5.085 (1, 44.650)	.03	2.348 (2, 45.142)	.11	1.519 (2, 44.642)	.23	0.181
Pre	5.9 (1.9)	4.9 (1.8)	6.2 (1.3)							
Post	7.1 (2.4)	5.7 (2.0)	6.2 (1.5)							

[a]MET: metabolic equivalents of task.

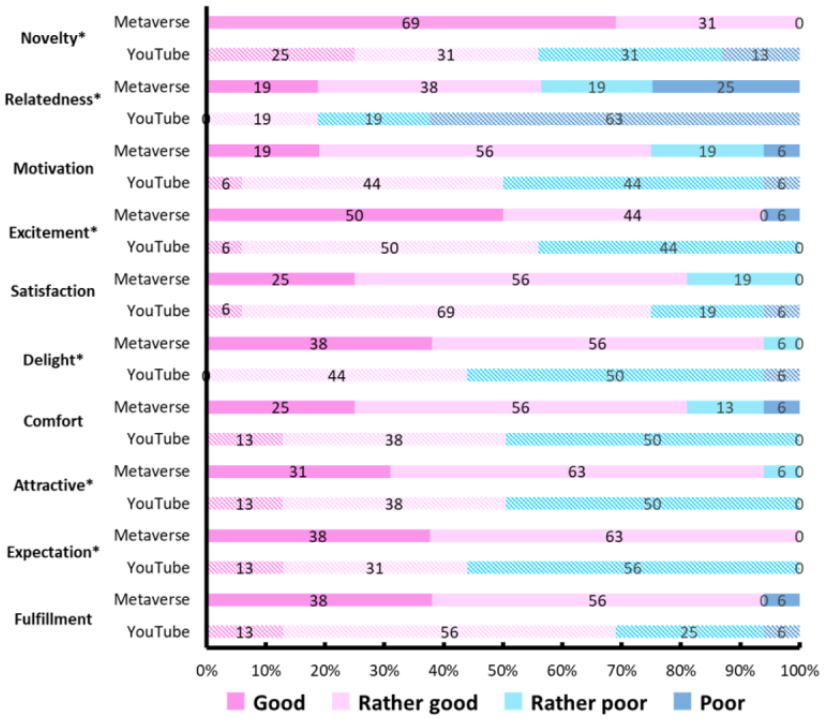
Table 3. Differences between pre- and postintervention outcome measures (N=48).

Variables	Physical activity: total (metabolic equivalents of tasks minutes/week)
Metaverse group	
Pre, mean (SD)	737.1 (609.5)
Post, mean (SD)	1575.4 (1071.8)
P value	.006
Effect size	0.682
YouTube group	
Pre, mean (SD)	661.7 (710.7)
Post, mean (SD)	911.9 (1103.3)
P value	.23
Effect size	0.298
Control group	
Pre, mean (SD)	930.6 (665.1)
Post, mean (SD)	844.7 (701.8)
P value	.57
Effect size	–0.142

Comparison of Impressions of the Intervention Between the Metaverse and YouTube Groups

Multimedia Appendix 1 and Figure 3 show the results of a chi-square test on impressions of the intervention between the metaverse and YouTube groups. Significant differences were observed between the 2 groups in novelty (*P*=.003; effect size=0.529), relatedness (*P*=.03; effect size=0.387), excitement (*P*=.01; effect size=0.433), delight (*P*=.002; effect size=0.539), attractiveness (*P*=.006; effect size=0.487), and expectation (*P*<.001; effect size=0.626). The results did not reveal significant differences in motivation, satisfaction, comfort, or fulfillment.

Figure 3. Impressions of intervention in the Metaverse and YouTube groups. *P<.05.



Discussion

This study examined whether an exercise intervention using exercise videos in the metaverse space has a positive impact on physical activity. The main results show a significant interaction between groups and time (pre- and postintervention) in total physical activity, with post hoc analysis showing a significant increase in total physical activity in the metaverse group after compared with before the intervention and no significant change in the YouTube and control groups. This study is the first to show that using the metaverse space to deliver exercise videos can promote increased physical activity.

The current results support the hypothesis that the use of the metaverse increases physical activity. The exercise videos used in the metaverse, and the YouTube groups are identical, and the metaverse group differs from the YouTube group in that the exercise videos can be viewed in a shared space with other avatars in the metaverse space. Studies have noted that interacting with others is more effective than exercising alone at increasing physical activity and sustaining exercise [9-13]. Metaverse characteristics include the ability to form communities within virtual reality and the persistence of virtual reality regardless of individual user access [15]. In this study, no change in social capital was detected before and after the intervention; however, the percentage of those who had an impression of relatedness (connection with others) was also significantly higher in the metaverse group (9/16, 56%) than in the YouTube group (3/16, 19%)—indicating that more people felt connected to others when using the metaverse than when using YouTube. Additionally, strategies to increase physical activity include improving the physical environment, such as its construction, and the social environment, including social support [41]. The use of the metaverse resulted in the persistence of a purposeful social space for exercise that could be accessed at any time. Therefore, the existence of the metaverse space is an aspect of environmental improvement for exercise for participants who previously had no exercise habits, leading to increased exercise implementation and physical activity.

The reason indicators other than physical activity did not improve in the metaverse group is also an important area to consider. This study asked about well-being as an item that could be related to mental health. Exercise is known to promote the secretion of estradiol and serotonin and regulate cortisol levels, an internal change that reduces depression and other depressive symptoms [42-44]. Additionally, reducing loneliness and encouraging social interaction decreases the risk of depression and mental health problems [45]. Therefore, we expected to observe mental health improvements in the metaverse group; however, there were no significant differences among the 3 groups. The standard for suspicion of mental problems is a cutoff value of ≤11 points for well-being [35,46]. However, the mean values for participant well-being did not correspond to the cutoff values. Thus, mental health changes were unlikely to be observed even if the amount of physical activity increased because the group did not initially have low well-being.

The locomotive function scale, a measure of body pain and locomotor function, showed an improving trend in the metaverse group; however, the difference was not significant. A previous study found pain relief after an 8-week exercise program, including strength training of the trunk, in participants with low back pain that appeared to be caused by inactivity [47]. Several studies have reported muscle hypertrophy occurring within 6-8 weeks [48]. These facts indicate that if training is managed meticulously in the right way, the body can be changed in a period of 8 weeks. However, it is possible that the unmonitored, 8-week intervention period did not result in sufficient changes in motor function, even though the amount of physical activity increased. To improve motor function, training to improve muscle strength and endurance should be continued at least 3 times per week [49].

As for the differences in impressions of the intervention between the metaverse and YouTube groups, many factors such as novelty, relatedness, excitement, delight, attractiveness, and expectation were more positive in the metaverse group than in the YouTube group. Regarding novelty, a previous study stated that sufficient novelty for the target audience leads to successful health support using the app [50]. Even for young adults who were relatively familiar with and used apps and social networking services on a regular basis, the metaverse seemed to be a novelty, which may have supported the implementation of the exercise in this study. Delight is an important factor in many areas of life [51]. Studies have reported that delight is the most important predictor of commitment and participation in general youth and elite sports [52,53]. This amplification of delight using a metaverse space to distribute videos is meaningful in promoting exercise. Interestingly, the metaverse group found the intervention more attractive. A previous study that used apps to promote healthy lifestyles, including regular exercise, found that a combination of different services, such as health-related quizzes and the ability to record health status, led to successful interventions [54]. Exercise video-delivered studies have also shown that simply delivering videos does not increase physical activity [8]. The essence of using the metaverse may have increased the attractiveness of exercise video distribution in this study.

This study examined the effectiveness of exercise video distribution using metaverse space by comparing the previously used YouTube and a control population. This study's strength lies in making the first attempt to distribute exercise videos using metaverse space. Furthermore, it shows the potential of using metaverse space to improve the physical activity level of young adults.

However, this study has some limitations. First, selection bias may have occurred. We recruited participants who were not members of an athletic club and who did not exercise heavily. However, it is possible that the target population was composed of students who were interested in exercising, as suggested by their interest in participating in this study. Additionally, we recruited participants from a single university. The risk of selection bias may be reduced by conducting a study on a larger scale and involving more institutions. However, the 3 groups were randomly assigned in this study, and the required, precalculated sample size was secured. Second, it was difficult

to accurately determine the time and frequency of interactions between participants in the metaverse space. If a system can be implemented to monitor the time of entry into the metaverse space, the frequency of interactions, and so on, the mechanism for promoting physical activity through the use of the metaverse space could be made more visible. Third, participants could receive the intervention at the time of their choosing; however, the start time of the intervention was not monitored. The effectiveness of the intervention could be moderated or altered by the time of the intervention, and future researchers should accurately monitor the start time of the intervention. We note that the main outcome, the calculation of physical activity, did not depend on the starting time of the intervention. Final, there was room for improvement in the comfort of the interventions in the metaverse group. This study used metaverse space in the manner in which the videos were arranged. Metaverse space (Spatial, Spatial Systems, Inc) has the advantage of being a service that anyone can use free of charge; however, if industry and academia collaborate to devise a system more suited to the distribution of exercise videos and the promotion of exercise, it could have a more positive effect on physical activity and other aspects.

In the near future, services allowing access to the metaverse will likely continue to expand. Therefore, the number of opportunities for the general public to experience the metaverse will further increase. The fact that 100% of the metaverse group in this study had high expectations is a hopeful part of the development of using the metaverse to improve physical activity. This research provides a basis for establishing further methods of using the metaverse to promote exercise.

Acknowledgments

We thank all the participants. We also thank Editage for the English language editing. This work was supported by the Japan Science and Technology Agency through the establishment of university fellowships toward the creation of science and technology innovation (JPMJFS2129).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Comparison of impressions on intervention between the metaverse and YouTube groups.
[PDF File (Adobe PDF File), 160 KB - mhealth_v12i1e46397_app1.pdf]

Multimedia Appendix 2
CONSORT-eHealth (V 1.6.1).
[PDF File (Adobe PDF File), 1312 KB - mhealth_v12i1e46397_app2.pdf]

References

1. 2019 National Health and Nutrition Survey Report. Ministry of Health, Labour and Welfare. 2019. URL: https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/kenkou_iryou/kenkou/eiyou/r1-houkoku_00002.html [accessed 2022-12-10]
2. Uesugi Y, Kanaya S, Nakanishi H, Naito Y. The relationship between locomotive syndrome risk, gait pattern, and standing posture in young Japanese women: a cross-sectional study. Healthcare (Basel) 2020;8(4):565 [FREE Full text] [doi: 10.3390/healthcare8040565] [Medline: 33333912]
3. Director-General for Statistics and Information Policy, Ministry of Health, Labour and Welfare. From comprehensive survey of living conditions, 2019. Graphical Review of Japanese Household. 2021. URL: https://www.mhlw.go.jp/toukei/list/dl/20-21-h29.pdf [accessed 2022-12-10]
4. Iida H, Seki T, Takegami Y, Osawa Y, Kato D, Takemoto G, et al. Association between locomotive syndrome and fall risk in the elderly individuals in Japan: the yakumo study. J Orthop Sci 2022 [FREE Full text] [doi: 10.1016/j.jos.2022.11.023] [Medline: 36522244]
5. Lippi G, Henry BM, Sanchis-Gomar F. Physical inactivity and cardiovascular disease at the time of coronavirus disease 2019 (COVID-19). Eur J Prev Cardiol 2020;27(9):906-908 [FREE Full text] [doi: 10.1177/2047487320916823] [Medline: 32270698]
6. McDonough DJ, Helgeson MA, Liu W, Gao Z. Effects of a remote, YouTube-delivered exercise intervention on young adults' physical activity, sedentary behavior, and sleep during the COVID-19 pandemic: randomized controlled trial. J Sport Health Sci 2022;11(2):145-156 [FREE Full text] [doi: 10.1016/j.jshs.2021.07.009] [Medline: 34314877]
7. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000;55(1):68-78. [doi: 10.1037//0003-066x.55.1.68] [Medline: 11392867]
8. Fukui K, Suzuki Y, Kaneda K, Kuroda S, Komiya M, Maeda N, et al. Do "stay-at-home exercise" videos induce behavioral changes in college students? A randomized controlled trial. Sustainability 2021;13(21):11600 [FREE Full text] [doi: 10.3390/su132111600]

9. Hillsdon M, Foster C, Thorogood M. Interventions for promoting physical activity. Cochrane Database Syst Rev 2005(1):CD003180 [FREE Full text] [doi: 10.1002/14651858.CD003180.pub2] [Medline: 15674903]
10. Kassavou A, Turner A, French DP. Do interventions to promote walking in groups increase physical activity? A meta-analysis. Int J Behav Nutr Phys Act 2013;10:18 [FREE Full text] [doi: 10.1186/1479-5868-10-18] [Medline: 23388115]
11. Darlow SD, Xu X. The influence of close others' exercise habits and perceived social support on exercise. Psychol Sport Exerc 2011;12(5):575-578. [doi: 10.1016/j.psychsport.2011.04.004]
12. Gellert P, Ziegelmann JP, Warner LM, Schwarzer R. Physical activity intervention in older adults: does a participating partner make a difference? Eur J Ageing 2011;8(3):211 [FREE Full text] [doi: 10.1007/s10433-011-0193-5] [Medline: 28798651]
13. Wang L, Guo X, Wu T, Lv L, Zhang Z. Short-term effects of social encouragement on exercise behavior: insights from China's wanbu network. Public Health 2017;148:25-29 [FREE Full text] [doi: 10.1016/j.puhe.2017.03.004] [Medline: 28404530]
14. Thomas GN, Macfarlane DJ, Guo B, Cheung BMY, McGhee SM, Chou KL, et al. Health promotion in older Chinese: a 12-month cluster randomized controlled trial of pedometry and "peer support". Med Sci Sports Exerc 2012;44(6):1157-1166 [FREE Full text] [doi: 10.1249/MSS.0b013e318244314a] [Medline: 22143109]
15. Kye B, Han N, Kim E, Park Y, Jo S. Educational applications of metaverse: possibilities and limitations. J Educ Eval Health Prof 2021;18:32 [FREE Full text] [doi: 10.3352/jeehp.2021.18.32] [Medline: 34897242]
16. Petrigna L, Musumeci G. The metaverse: a new challenge for the healthcare system: a scoping review. J Funct Morphol Kinesiol 2022;7(3):63 [FREE Full text] [doi: 10.3390/jfmk7030063] [Medline: 36135421]
17. Chengoden R, Victor N, Huynh-The T, Yenduri G, Jhaveri RH, Alazab M, et al. Metaverse for healthcare: a survey on potential applications, challenges and future directions. IEEE Access 2023;11:12765-12795 [FREE Full text] [doi: 10.1109/access.2023.3241628]
18. Plechatá A, Makransky G, Böhm R. Can extended reality in the metaverse revolutionise health communication? NPJ Digit Med 2022;5(1):132 [FREE Full text] [doi: 10.1038/s41746-022-00682-x] [Medline: 36056245]
19. Marroquín B, Vine V, Morgan R. Mental health during the COVID-19 pandemic: effects of stay-at-home policies, social distancing behavior, and social resources. Psychiatry Res 2020;293:113419 [FREE Full text] [doi: 10.1016/j.psychres.2020.113419] [Medline: 32861098]
20. Usmani SS, Sharath M, Mehendale M. Future of mental health in the metaverse. Gen Psychiatr 2022;35(4):e100825 [FREE Full text] [doi: 10.1136/gpsych-2022-100825] [Medline: 36189180]
21. Goharinejad S, Goharinejad S, Hajesmaeel-Gohari S, Bahaadinbeigy K. The usefulness of virtual, augmented, and mixed reality technologies in the diagnosis and treatment of attention deficit hyperactivity disorder in children: an overview of relevant studies. BMC Psychiatry 2022;22(1):4 [FREE Full text] [doi: 10.1186/s12888-021-03632-1] [Medline: 34983446]
22. Freeman D, Reeve S, Robinson A, Ehlers A, Clark D, Spanlang B, et al. Virtual reality in the assessment, understanding, and treatment of mental health disorders. Psychol Med 2017;47(14):2393-2400 [FREE Full text] [doi: 10.1017/S003329171700040X] [Medline: 28325167]
23. de Carvalho MR, de Santana Dias TR, Duchesne M, Nardi AE, Appolinario JC. Virtual reality as a promising strategy in the assessment and treatment of bulimia nervosa and binge eating disorder: a systematic review. Behav Sci (Basel) 2017;7(3):43 [FREE Full text] [doi: 10.3390/bs7030043] [Medline: 28698483]
24. Clark DB, Tanner-Smith EE, Killingsworth SS. Digital games, design, and learning: a systematic review and meta-analysis. Rev Educ Res 2016;86(1):79-122 [FREE Full text] [doi: 10.3102/0034654315582065] [Medline: 26937054]
25. Maugeri G, Castrogiovanni P, Battaglia G, Pippi R, D'Agata V, Palma A, et al. The impact of physical activity on psychological health during Covid-19 pandemic in Italy. Heliyon 2020;6(6):e04315 [FREE Full text] [doi: 10.1016/j.heliyon.2020.e04315] [Medline: 32613133]
26. Guidelines for the Data Processing and Analysis of the International Physical Activity Questionnaire. URL: https://www.tmu-ph.ac/news/data/180327_1.pdf [accessed 2022-09-01]
27. Cheng HL. A simple, easy-to-use spreadsheet for automatic scoring of the International Physical Activity Questionnaire (IPAQ) short form. ResearchGate 2016 [FREE Full text] [doi: 10.13140/RG.2.2.21067.80165]
28. Mandic D, Bjegovic-Mikanovic V, Vukovic D, Djikanovic B, Stamenkovic Z, Lalic NM. Successful promotion of physical activity among students of medicine through motivational interview and Web-based intervention. Peer J 2020;8:e9495 [FREE Full text] [doi: 10.7717/peerj.9495] [Medline: 32714663]
29. Urbaniak GC, Plous S. Research randomizer. URL: https://www.randomizer.org/ [accessed 2022-09-27]
30. Belloc NB, Breslow L. Relationship of physical health status and health practices. Prev Med 1972;1(3):409-421. [doi: 10.1016/0091-7435(72)90014-x] [Medline: 5085007]
31. Tenjin K, Sekine M, Yamada M, Tatsuse T. Relationship between parental lifestyle and dietary habits of children: a cross-sectional study. J Epidemiol 2020;30(6):253-259. [doi: 10.2188/jea.JE20190015] [Medline: 31155541]
32. Satake S, Shimada H, Yamada M, Kim H, Yoshida H, Gondo Y, et al. Prevalence of frailty among community-dwellers and outpatients in Japan as defined by the Japanese version of the cardiovascular health study criteria. Geriatr Gerontol Int 2017;17(12):2629-2634. [doi: 10.1111/ggi.13129] [Medline: 29265757]

33. Yasuda T. Identifying preventative measures against frailty, locomotive syndrome, and sarcopenia in young adults: a pilot study. J Phys Ther Sci 2021;33(11):823-827 [FREE Full text] [doi: 10.1589/jpts.33.823] [Medline: 34776616]
34. Lee PH, Macfarlane DJ, Lam TH, Stewart SM. Validity of the International Physical Activity Questionnaire Short Form (IPAQ-SF): a systematic review. Int J Behav Nutr Phys Act 2011;8:115 [FREE Full text] [doi: 10.1186/1479-5868-8-115] [Medline: 22018588]
35. Inagaki H, Ito K, Sakuma N, Sugiyama M, Okamura T, Awata S. Reliability and validity of the simplified Japanese version of the WHO-Five well-being index (S-WHO-5-J). Nihon Koshu Eisei Zasshi 2013;60(5):294-301 [FREE Full text] [Medline: 23942026]
36. Yamada K, Ito YM, Akagi M, Chosa E, Fuji T, Hirano K, et al. Reference values for the locomotive syndrome risk test quantifying mobility of 8681 adults aged 20-89 years: a cross-sectional nationwide study in Japan. J Orthop Sci 2020;25(6):1084-1092 [FREE Full text] [doi: 10.1016/j.jos.2020.01.011] [Medline: 32173180]
37. Saito M, Kondo N, Aida J, Kawachi I, Koyama S, Ojima T, et al. Development of an instrument for community-level health related social capital among Japanese older people: the JAGES project. J Epidemiol 2017;27(5):221-227 [FREE Full text] [doi: 10.1016/j.je.2016.06.005] [Medline: 28169107]
38. Law ELC, van Schaik P, Roto V. Attitudes towards User Experience (UX) measurement. Int J Hum Comput Stud 2014;72(6):526-541 [FREE Full text] [doi: 10.1016/j.ijhcs.2013.09.006]
39. Molenberghs G, Verbeke G. A review on linear mixed models for longitudinal data, possibly subject to dropout. Statistical Modelling 2016 Aug 18;1(4):235-269. [doi: 10.1177/1471082X0100100402]
40. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMJ 2010;340:c332 [FREE Full text] [doi: 10.1136/bmj.c332] [Medline: 20332509]
41. Kahn EB, Ramsey LT, Brownson RC, Heath GW, Howze EH, Powell KE, et al. The effectiveness of interventions to increase physical activity. A systematic review. Am J Prev Med 2002;22(4 Suppl):73-107. [doi: 10.1016/s0749-3797(02)00434-8] [Medline: 11985936]
42. Chakravorty SG, Halbreich U. The influence of estrogen on monoamine oxidase activity. Psychopharmacol Bull 1997;33(2):229-233. [Medline: 9230635]
43. Rocha BA, Fleischer R, Schaeffer JM, Rohrer SP, Hickey GJ. 17 Beta-estradiol-induced antidepressant-like effect in the forced swim test is absent in estrogen receptor-beta knockout (BERKO) mice. Psychopharmacology (Berl) 2005;179(3):637-643. [doi: 10.1007/s00213-004-2078-1] [Medline: 15645223]
44. Yang HJ, Koh E, Kang Y. Susceptibility of women to cardiovascular disease and the prevention potential of mind-body intervention by changes in neural circuits and cardiovascular physiology. Biomolecules 2021;11(5):708 [FREE Full text] [doi: 10.3390/biom11050708] [Medline: 34068722]
45. Matthews T, Danese A, Wertz J, Odgers CL, Ambler A, Moffitt TE, et al. Social isolation, loneliness and depression in young adulthood: a behavioural genetic analysis. Soc Psychiatry Psychiatr Epidemiol 2016;51(3):339-348 [FREE Full text] [doi: 10.1007/s00127-016-1178-7] [Medline: 26843197]
46. Christensen KS, Haugen W, Sirpal MK, Haavet OR. Diagnosis of depressed young people--criterion validity of WHO-5 and HSCL-6 in Denmark and Norway. Fam Pract 2015;32(3):359-363 [FREE Full text] [doi: 10.1093/fampra/cmv011] [Medline: 25800246]
47. Kalichman L, Carmeli E, Been E. The association between imaging parameters of the paraspinal muscles, spinal degeneration, and low back pain. Biomed Res Int 2017;2017:2562957 [FREE Full text] [doi: 10.1155/2017/2562957] [Medline: 28409152]
48. Krzysztofik M, Wilk M, Wojdała G, Gołaś A. Maximizing muscle hypertrophy: a systematic review of advanced resistance training techniques and methods. Int J Environ Res Public Health 2019;16(24):4897 [FREE Full text] [doi: 10.3390/ijerph16244897] [Medline: 31817252]
49. Landi F, Marzetti E, Martone AM, Bernabei R, Onder G. Exercise as a remedy for sarcopenia. Curr Opin Clin Nutr Metab Care 2014;17(1):25-31. [doi: 10.1097/MCO.0000000000000018] [Medline: 24310054]
50. Simons LPA, Hampe JF, Guldemond NA. Designing healthy living support: mobile applications added to hybrid (e)Coach solution. Health Technol 2013;3(1):85-95. [doi: 10.1007/s12553-013-0052-9]
51. Kawabata M, Mallett CJ. Progressing the construct of enjoyment: conceptualizing enjoyment as a proactive process. Discov Psychol 2022;2(1):2 [FREE Full text] [doi: 10.1007/s44202-021-00015-1]
52. Jones MV. Emotion regulation and performance. In: Murphy SM, editor. The Oxford Handbook of Sport and Performance Psychology. Oxford, UK: Oxford University Press; 2012:154-172.
53. Biddle SJH, Mutrie N, Gorely T. Psychology of Physical Activity: Determinants, Well-Being and Interventions. 3rd Edition. New York: Routledge; 2015.
54. Simons LPA, Foerster F, Bruck PA, Motiwalla L, Jonker CM. Microlearning mApp raises health competence: hybrid service design. Health Technol (Berl) 2015;5(1):35-43 [FREE Full text] [doi: 10.1007/s12553-015-0095-1] [Medline: 26097799]

Abbreviations

MET: metabolic equivalents of task

Edited by L Buis; submitted 10.02.23; peer-reviewed by L Wang, R Martínez Romero; comments to author 03.08.23; revised version received 04.09.23; accepted 28.11.23; published 16.01.24.

Please cite as:
Mizuta R, Maeda N, Tashiro T, Suzuki Y, Kuroda S, Ishida A, Oda S, Watanabe T, Tamura Y, Komiya M, Urabe Y
Effectiveness of Metaverse Space–Based Exercise Video Distribution in Young Adults: Randomized Controlled Trial
JMIR Mhealth Uhealth 2024;12:e46397
URL: https://mhealth.jmir.org/2024/1/e46397
doi:10.2196/46397
PMID:38227355

Original Paper

Mobile App Intervention of a Randomized Controlled Trial for Patients With Obesity and Those Who Are Overweight in General Practice: User Engagement Analysis Quantitative Study

Vera Helen Buss[1*], PhD; Margo Barr[1*], PhD; Sharon M Parker[1*], MPH; Alamgir Kabir[1*], PhD; Annie Y S Lau[2*], PhD; Siaw-Teng Liaw[3*], PhD; Nigel Stocks[4*], MD; Mark F Harris[1*], MD

[1]Centre for Primary Health Care and Equity, University of New South Wales, Sydney, Australia

[2]Australian Institute of Health Innovation, Macquarie University, Sydney, Australia

[3]School of Population Health, University of New South Wales, Sydney, Australia

[4]Adelaide Medical School, University of Adelaide, Adelaide, Australia

[*]all authors contributed equally

Corresponding Author:
Margo Barr, PhD
Centre for Primary Health Care and Equity
University of New South Wales
AGSM Building
High Street, Kensington Campus
Sydney, 2052
Australia
Phone: 61 290656041
Email: Margo.barr@unsw.edu.au

Related Article:

This is a corrected version. See correction statement: https://mhealth.jmir.org/2024/1/e58507

Abstract

Background: The Health eLiteracy for Prevention in General Practice trial is a primary health care–based behavior change intervention for weight loss in Australians who are overweight and those with obesity from lower socioeconomic areas. Individuals from these areas are known to have low levels of health literacy and are particularly at risk for chronic conditions, including diabetes and cardiovascular disease. The intervention comprised health check visits with a practice nurse, a purpose-built patient-facing mobile app (mysnapp), and a referral to telephone coaching.

Objective: This study aimed to assess *mysnapp* app use, its user profiles, the duration and frequency of use within the Health eLiteracy for Prevention in General Practice trial, its association with other intervention components, and its association with study outcomes (health literacy and diet) to determine whether they have significantly improved at 6 months.

Methods: In 2018, a total of 22 general practices from 2 Australian states were recruited and randomized by cluster to the intervention or usual care. Patients who met the main eligibility criteria (ie, BMI>28 in the previous 12 months and aged 40-74 years) were identified through the clinical software. The practice staff then provided the patients with details about this study. The intervention consisted of a health check with a practice nurse and a lifestyle app, a telephone coaching program, or both depending on the participants' choice. Data were collected directly through the app and combined with data from the 6-week health check with the practice nurses, the telephone coaching, and the participants' questionnaires at baseline and 6-month follow-up. The analyses comprised descriptive and inferential statistics.

Results: Of the 120 participants who received the intervention, 62 (52%) chose to use the app. The app and nonapp user groups did not differ significantly in demographics or prior recent hospital admissions. The median time between first and last app use was 52 (IQR 4-95) days, with a median of 5 (IQR 2-10) active days. App users were significantly more likely to attend the 6-week health check (2-sided Fisher exact test; $P<.001$) and participate in the telephone coaching (2-sided Fisher exact test; $P=.007$)

than nonapp users. There was no association between app use and study outcomes shown to have significantly improved (health literacy and diet) at 6 months.

Conclusions: Recruitment and engagement were difficult for this study in disadvantaged populations with low health literacy. However, app users were more likely to attend the 6-week health check and participate in telephone coaching, suggesting that participants who opted for several intervention components felt more committed to this study.

Trial Registration: Australian New Zealand Clinical Trials Registry ACTRN12617001508369; https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=373505

International Registered Report Identifier (IRRID): RR2-10.1136/bmjopen-2018-023239

(JMIR Mhealth Uhealth 2024;12:e45942) doi:10.2196/45942

KEYWORDS

health literacy; primary health care; mobile application; overweight; vulnerable populations; health behavior; mHealth; obesity; weight loss; mysnapp app; mobile phone

Introduction

Problem Statement

Obesity is a major contributor to disease burden, increasing the risk of chronic conditions, including ischemic heart disease, stroke, diabetes mellitus, chronic kidney disease, and hypertensive heart disease [1]. According to the Global Burden of Disease Study 2017 [2], high BMI was the cause of 4.72 million deaths and 148 million disability-adjusted life-years worldwide, making it the fourth leading risk for mortality in 2017. In 2017 to 2018, an estimated 36% of the Australian adult population were overweight (ie, BMI 25.0-29.9) and 31% of them had obesity (ie, BMI≥30.0) [3]. The proportion of people who are overweight or hose with obesity is higher in populations from lower socioeconomic backgrounds [3]. In 2017-2018, 72% of Australian adults residing in the lowest socioeconomic areas were overweight or had obesity compared to 62% from the highest, after adjusting for age [3]. People from the lowest socioeconomic areas were 1.9 times more likely to have diabetes in 2020 and 1.6 times more likely to have self-reported coronary heart disease in 2017-2018 than those from the highest socioeconomic areas [4].

Rationale for the Study

Other research has shown that mobile app-based interventions can facilitate weight loss in individuals who are overweight and those with obesity, but it requires regular app use. For example, Patel et al [5] reported that consistent weight self-monitoring via a mobile app could lead to clinically meaningful weight loss. However, the study classified only a quarter of participants as consistent trackers, which they defined as self-monitoring weight and diet on at least 6 days per week for at least 75% of the study weeks [5]. Their study highlighted that consistent tracking was crucial, but only a minority of participants did so. Similarly, Laing et al [6] found that providing access to a weight loss app to primary care patients who are overweight and those with obesity did not lead to significant weight loss compared to usual care. Only one-third of them logged into the app in the sixth month of the intervention, in which the median number of logins was 0 (IQR 0-2). The authors concluded that prescribing self-monitoring apps for caloric counting may be successful in primary care patients who are particularly motivated to lose weight [6]. Chin et al [7] analyzed user data from a popular commercial weight loss app and found that in a multivariate logistic regression model, the frequency of entering body weight and consumption of dinner particularly was associated with successful weight loss in app users. Considering other studies, the focus of this study was understanding how participants used a mobile app within the Health eLiteracy for Prevention in General Practice (HeLP-GP) trial and if its use led to improvements in health literacy and diet.

Description of the Intervention

The HeLP-GP trial was a behavior change intervention developed for implementation in Australian general practices aimed at Australians who are overweight and those with obesity from lower socioeconomic areas to help them reduce their weight. The intervention was based on the 5As framework (assess, advise, agree, assist, and arrange) [8]. It included health check visits with a practice nurse based on the 5As framework, the use of a purpose-built patient-facing mobile app called *mysnapp*, and referral to health coaching via the “Get Healthy” information and coaching service [9]. The *mysnapp* is based on a web-based platform developed by Lau et al [10].

The trial was a pragmatic, 2-arm, unblinded cluster randomized controlled trial, which continued for 12 months. Primary outcomes included changes in weight, blood pressure, health literacy, and eHealth literacy [11,12]. Secondary outcomes included lipids, diet (fruit and vegetable intake), level of physical activity, quality of life, advice received, and referral for diet, physical activity, and weight loss [12]. Participants who received the intervention could choose to use the mobile app and access the telephone coaching program. The HeLP-GP trial assessed the intervention’s effectiveness [12]. The intervention led to significant improvements at 6 months compared to the controls for health literacy (mean DiD 0.22, 95% CI 0.01-0.44) and diet (mean DiD 0.98, 95% CI 0.50-1.47). There were no associations with any of the other outcomes [12].

Objectives

The overall aim of this study, within the HeLP-GP trial, was to assess *mysnapp* app use, engagement, its association with other intervention components, and its association with study outcomes shown to have significantly improved (health literacy and diet) at 6 months.

Our objectives were to (1) explore differences in demographics and hospital admissions between participants who used the app and those who did not, (2) examine the duration and frequency of app use (app engagement) by participants overall and by module, (3) assess the association among app use, app engagement, and participation in other intervention components, and (4) examine the association between app use and app engagement on study outcomes that were shown to be significantly improved at 6 months (ie, health literacy and diet).

Methods

Ethical Considerations

The University of New South Wales Human Research Ethics Committee (HC17474) approved the trial. The University of Adelaide Human Research Ethics Committee ratified this approval. All participants provided consent to take part in this study.

Intervention

The methodology of the randomized controlled trial, of which this study is a subanalysis, was published previously [13] and prospectively registered with the Australian New Zealand Clinical Trials Registry (ACTRN12617001508369). In 2018, a total of 22 general practices were recruited from 2 Australian states, New South Wales (South West and Central Sydney) and South Australia (Adelaide), and randomized by cluster to the HeLP-GP intervention (11 practices) or usual care (11 practices). General practices were recruited through the local Primary Health Networks. Practices were located in local government areas with Socio-Economic Indexes for Areas scores [14] equal to or below the eighth decile. The Australian Bureau of Statistics reported that these are usually associated with lower health literacy levels in the population [15], with health literacy being defined by the Australian Institute of Health and Welfare in their latest Health Literacy Report as "how people access, understand and use health information in ways that benefit their health" [16]. In total, 4 strata based on the practice size (<5 general practice [GPs] and ≥5 GPs) and the state were created and then we randomly allocated practices to each stratum's intervention or usual care group. The intervention comprised a practice nurse–led health check; additionally, participants could choose whether to take up a lifestyle app, a telephone coaching program, or both. Potential participants were identified using the GPs' software. The general practitioners of the intervention sites also assessed their patients for eligibility. The eligible patients were provided with trial information and consent forms by the reception staff. Recruitment occurred between October 2018 and September 2019.

At the baseline health check, the practice nurses helped participants with the *mysnapp* setup and access the coaching program. They entered the participant's height, weight, waist circumference, and blood pressure into the app and set the health goals with the participant. For 6 weeks, the participants received a nutrition-related and a physical activity–related text message weekly. These were preprepared to be sent automatically each week and provided direct advice and a web link for further information. In addition, the telephone coaching program provided free, confidential health support to participants to reach personalized lifestyle goals concerning diet, physical activity, alcohol, and body weight [17]. The coaching was available in multiple languages through an interpreter service. The practice nurses conducted a 6-week health check in which they reviewed and revised the participants' health goals. Additionally, general practitioners conducted a 12-week health review. Text messages reminded participants to attend these follow-up visits.

Participants

Individuals were eligible for this study if they were aged 40-74 years, had a BMI of ≥28 and blood pressure levels recorded in the clinical software within the last 12 months, spoke English or Arabic, and had access to a smartphone or tablet. Potential participants were ineligible if they fulfilled any of the following exclusion criteria: recent weight loss (ie, >5% in the past 3 months), taking weight loss drugs (ie, orlistat or phentermine), diagnosed with insulin-dependent diabetes or cardiovascular disease (ie, angina, myocardial infarction, heart failure, heart valve disease, or stroke), cognitive impairment, or physical impairment disallowing them to perform moderate physical activity.

mysnapp Design

The *mysnapp* content was based on a web-based platform designed to help individuals control and maintain their health data and information to manage their health [10]. Research by Webb et al [18] and DiFilippo et al [19] into behavior change through mobile and electronic platforms informed the app design, including goal setting and self-monitoring, and additional methods to interact with individuals, mainly text messaging. The *mysnapp* app consisted of 4 core modules that allowed users to (1) set physical activity– and diet-based goals, (2) monitor their progress over the past 6 weeks, (3) take notes in a diary, and (4) learn about healthy eating and physical activity. Users could choose from the following goal options: set daily servings of fruits or vegetables or physical activity minutes; aim to drink fewer soft drinks, eat smaller portions, or eat fewer snacks or takeaway foods. In the self-monitoring module, they entered how many days of the week they achieved their goals. The educational material consisted of short text summaries and fact sheets about healthy foods, portion sizes, discretionary beverage consumption, physical activity benefits in English or Arabic, and links to simple exercise videos on YouTube.

Study Measures

App Use Measure

Data were collected on app use, specifically, when the study participants in the intervention group had an app account set up.

App Engagement Measures

Data were collected on the participants' app use directly through *mysnapp*. Each month, a cumulative data report was created about app logins and interactions with the different app modules from each participant for 12 months. App engagement included active days, duration of app use, and frequency of accessing app modules.

Other Intervention Component Measures

The data from the 6-week health check with the practice nurses (ie, occurrence) and the telephone coaching (ie, occurrence and completion status) were the other intervention component measures.

Outcome Measures

The participants' questionnaires at baseline and 6-month follow-up (ie, self-reported fruit and vegetable intake, and health literacy) were used.

Specifically, the diet questions were as follows: (1) How many servings of fruit do you usually eat each day? A serving is 1 medium-sized fruit such as an apple or 2 small-sized fruits or 1 cup of fruit pieces. (2) How many servings of vegetables do you usually eat each day? One serving is half a cup of cooked vegetables or 1 cup of salad vegetables. With the diet score being the portions of fruit intake (between 0 and a maximum of 2 per day) plus portions of vegetable intake (between 0 and a maximum of 5 per day) with a range of 0 to 7 based on the sum of fruit and vegetable scores. This diet measure has been validated against food frequency questionnaires [20].

Specifically, the Health Literacy Questionnaire domain 8 questions were used [11]: (1) find information about health problems; (2) find health information from several …. ; (3) get information about health so you…; (4) get health information in words you…; and (5) get health information by yourself. There is a 5-point response option scale for each question (cannot do or always difficult, usually difficult, sometimes difficult, usually easy or always easy). The scores are reported as averages for the domain (with a range between 1 and 5) with high scores representing higher health literacy.

Table 1 contains definitions for study measures. Duration of app use, active days, and consistent use had preset maximum values (365 days or 52 weeks); the values were capped when they exceeded the maximum.

Table 1. Measures and their definitions.

Measure	Type of variable	Explanation or definition
App use measures		
App user	Binary and input variable	Study participants in the intervention group who had an app account set up
App engagement measures		
Duration of app use	Continuous and input variable; maximum value: 365 days	Number of days between the first and last time a participant accessed the app
Active days	Continuous and input variable; maximum value: 365 days	Number of days a participant accessed the app
Consistent app use	Continuous and input variable; maximum value: 52 weeks	Number of consecutive weeks a participant accessed ≥1 time the app starting from the first app use
App module use	Binary and input variable	Participant accessed ≥1 the corresponding app module (goal setting, progress tracking, diary, or education)
Frequency of accessing app modules	Continuous and input variable	Number of times a participant accessed the corresponding app module (goal setting, progress tracking, diary, or education)
Other intervention component measures		
Practice nurse-led health check	Categorical and input variable	Attended and not attended
Telephone coaching	Categorical and input variable	Completed, not completed, and not participated
Outcome measures		
Health literacy	Continuous and output variable	Health literacy, specifically the self-reported ability to find good quality health information, according to domain 8 of the Health Literacy Questionnaire [11], at baseline and 6-month follow-up. The scores were reported as averages for the domain (with a range between 1 and 5) with high scores representing higher health literacy.
Diet score	Continuous and output variable	Self-reported daily fruit and vegetable intake at baseline and 6-month follow-up. Diet score was the portions of fruit intake (between 0 and a maximum of 2 per day) plus portions of vegetable intake (between 0 and a maximum of 5 per day) with a range between 0 and 7 based on the sum of fruit and vegetable scores.

Data Analysis

Descriptive and inferential analyses in RStudio (with the programming language *R*; *R* Foundation for Statistical Computing) using a significance level of .05 for all statistical tests were conducted. Normally distributed continuous variables were summarized using the mean and SD, and nonnormally distributed continuous variables with median and IQR. Box

plots compared continuous variables across the categories of nonnumerical variables [21]. Normality was tested using the Shapiro-Wilk normality test [22-24]. The 2-sided Welch *t* test was performed to compare the means of continuous variables between 2 subgroups (eg, participants using *mysnapp* versus those not using it) for normally distributed continuous variables [25]. Alternatively, the Wilcoxon signed rank test with continuity correction comparing the medians of nonnormally distributed continuous variables between 2 subgroups was used [26,27]. The Kruskal-Wallis rank-sum test was performed for more than 2 subgroups and nonnormally distributed continuous variables [28]. Pearson chi-square test with Yates continuity correction was used to test for associations between 2 categorical variables and the 2-sided Fisher exact test was used when there were less than 5 participants in any cell of the contingency table of expected frequencies [29-31].

For objective 4, we used 1-sided tests to assess whether app use versus nonapp use, or app engagement was associated with health literacy or diet between baseline and 6-month follow-up. Correlations between the app engagement and health literacy or diet score were measured with the Kendall rank correlation test (if variables did not follow a normal distribution) or Pearson product-moment correlation test (if they followed a normal distribution) [32,33].

Results

App Users

In total, 120 participants received the intervention, of which 62 (52%) people chose to use *mysnapp*. Among the 62 app users, 38 (61%) also opted for telephone coaching. Table 2 shows the results for the first objective, comparing the demographic characteristics of the participants who chose not to use *mysnapp* to those who decided to use it. There were no significant differences between app users and nonapp users.

Table 2. Demographic characteristics of participants in the intervention group (N=120).

Variables	Nonapp users (n=58)	App users (n=62)	Test statistics for differences between groups	*P* value
Age (years), mean (SD)	58 (8)	61 (9)	t_{115}=–1.56	.12
Women, n (%)	28 (48)	32 (52)	χ^2_1<0.1	.86
Born in Australia, n (%)	27 (47)	39 (63)	χ^2_1=2.6	.11
Preferred language is English, n (%)	54 (93)	58 (94)	OR[a] 0.93, 95% CI 0.16-5.26	>.99
Hospital admission in past 12 months, n (%)	15 (26)	12 (19)	χ^2_1=0.4	.53
Location New South Wales, n (%)	50 (86)	49 (79)	χ^2_1=0.6	.43

[a]OR: odds ratio.

App Engagement

The median duration of app use was 52 (IQR 4-95) days. Further, 2 participants used *mysnapp* weekly throughout the 12 months (Table 3). Active days ranged from 1 to 117 days, with a median of 5 (IQR 2-10) days. The median number of weeks participants consistently used *mysnapp* from baseline was 1 (IQR 1-2). Of the 62 app users, 60 (97%) opened the goal setting module, 55 (89%) the education module, 39 (63%) the progress tracking module, and 25 (39%) the diary. Table 3 shows the consistency of app use and how many modules the app users accessed over the entire period of the intervention. Of the 19 app users who had opened 3 of the 4 modules, 17 (89%) had accessed the goal setting, progress tracking, and education modules. Among the 16 who had opened 2 modules, 14 (88%) had accessed the goal setting and education modules.

Table 3. Consistency of app use and frequency of accessing app modules (n=62).

Variables and values	Participants, n (%)
Consistent app use (weeks)	
1	45 (73)
2-4	10 (16)
5-19	5 (8)
20-52	2 (3)
Number of modules accessed	
0	1 (2)
1	5 (8)
2	16 (26)
3	19 (31)
4	21 (34)
Frequency of accessing the goal setting module	
0-3	54 (87)
4-7	6 (10)
8-15	2 (3)
>15	0 (0)
Frequency of accessing the progress tracking module	
0-3	43 (69)
4-7	11 (18)
8-15	6 (10)
>15	2 (3)
Frequency of accessing the diary module	
0-3	48 (77)
4-7	5 (8)
8-15	4 (6)
>15	5 (8)
Frequency of accessing the education module	
0-3	41 (66)
4-7	12 (19)
8-15	6 (10)
>15	3 (5)

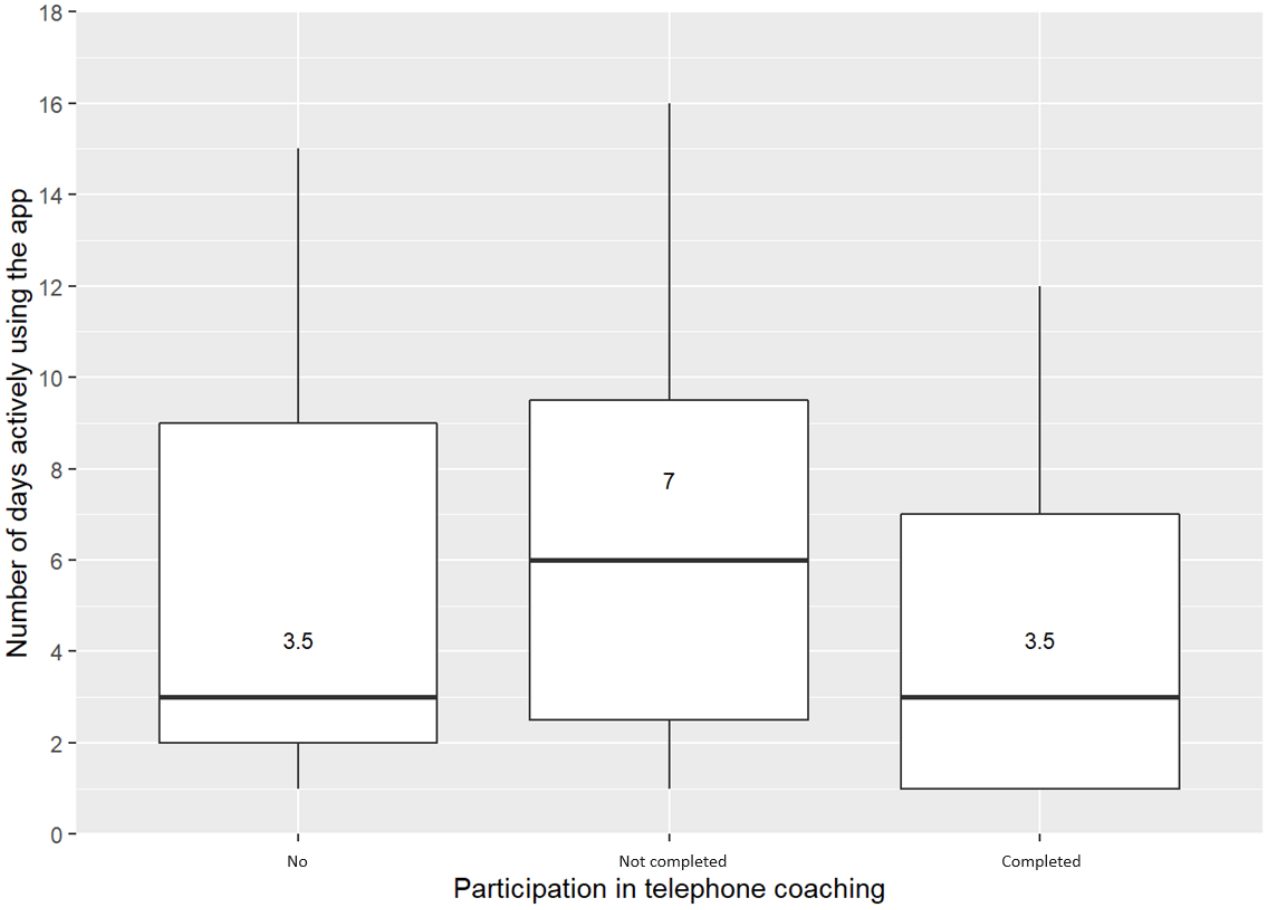
Association With Other Intervention Components

The difference in telephone coaching uptake between the app and nonapp users was statistically significant (Freeman-Halton extension of 2-sided Fisher exact test $P<.001$, Table 4). The median number of days using *mysnapp* for the app users who completed the telephone coaching was 3.5 (IQR 1-7) days, for the app users who did not complete the telephone coaching it was 7 (IQR 2.5-9.5) days, and for the app users who did not undertake the telephone coaching it was 3.5 (IQR 2-9) days (Figure 1). The difference in median active days by telephone coaching completion status was not statistically significant ($\chi^2_{19}=13.2$, $P=.83$).

Table 4. Association of app use with other intervention components (N=120).

Other intervention components and status	Nonapp users (n =58), n (%)	App users (n=62), n (%)	Test for differences between groups
Telephone coaching program			Freeman-Halton extension of Fisher exact test (2-tailed) *P*<.001
Not participated	47 (81)	24 (39)	
Not completed	8 (14)	16 (26)	
Completed	3 (5)	22 (35)	
A 6-week health check			Fisher exact test (2-tailed) *P*=.007
Not attended	54 (93)	46 (74)	
Attended	4 (7)	16 (26)	

Figure 1. Box plots of the number of days actively using mysnapp depending on the participation in telephone coaching; outliers excluded (two for no telephone coaching: 30 and 105 days, one for not completed: 117 days, and one for completed: 105 days).



The difference in the attendance rate of the 6-week health check between app users and nonusers was significant (2-sided Fisher exact test *P*=.007, Table 4). Those app users who attended the 6-week health check with the practice nurse did not have significantly more active days using *mysnapp* (median active days for 6-week health check attendees: 6, IQR 2-10 days, and for nonattendees: 4, IQR 2-10 days; *W*=374, *P*=.46).

Impact of App Use and App Engagement on Behavioral and Biomedical Outcome Measures

Differences in outcome measures between app users and nonusers, and app engagement were not significant (Tables 5 and 6) for study outcomes which were shown to be significantly improved at 6 months (ie, health literacy and diet).

Table 5. Health literacy and diet score at 2 time points for app and nonapp users, test for significant changes, and sensitivity analysis.

Outcome variable and measure	App users (n=62)		Nonapp users (n=58)		Test statistic[a]
	Baseline	6 months	Baseline	6 months	
HLQ[b] domain 8					
Data available, n (%)	52 (84)	44 (76)	50 (86)	20 (34)	N/A[c]
Median (IQR)	4 (4-5)	4 (4-5)	4 (4-4)	4 (4-5)	*W*=230.5, *P*=.10
Diet score					
Data available, n (%)	57 (92)	54 (93)	46 (74)	20 (34)	N/A
Mean (SD)	3 (2-5)	4 (4-5)	3 (2-4)	4 (3-5)	t_{36}=0.32, *P*=.37

[a]Test for greater change in app users versus nonapp users from baseline to 6 months.

[b]HLQ: Health Literacy Questionnaire.

[c]N/A: not applicable.

Table 6. Correlation between app engagement and change in health literacy or diet score.

Outcome variable and measure for app use	Test statistics for differences
HLQ[a] domain 8	
Active days	*z*=–0.24, *P*=.81, τ=–0.03
Consistent app use	*z*=0.43, *P*=.67, τ=0.06
Diet score	
Active days	*z*=0.55, *P*=.58, τ=0.07
Consistent app use	*z*=0.43, *P*=.67, τ=0.06

[a]HLQ: Health Literacy Questionnaire.

Discussion

Principal Results

The overall aim of this study was to assess *mysnapp* app use within the HeLP-GP trial and its association with study outcomes shown to have significantly improved (health literacy and diet) at 6 months. With regard to the specific objectives, (1) there were no significant differences in demographics between participants who used *mysnapp* and those who did not; (2) among app users, the median duration of app use was 52 days, with a median of 5 active days; (3) more participants who chose to use *mysnapp* also attended the 6-week health check with the practice nurse and opted for telephone coaching; and (4) there was no association between app use and study outcomes shown to have significantly improved (health literacy and diet) at 6 months.

Length and Frequency of App Use and Module Access

Turner-McGrievy et al [34] aimed to identify the best criteria for defining adherence to dietary self-monitoring with mobile devices when predicting weight loss. They found that adherence, defined as the number of days participants tracked at least 2 meal times, explained the most variance in weight loss at 6 months [34]. We were not able to measure this because the diary, available for recording meals, could also be used for other reasons such as activities, appointments, plans for the future, and thoughts about progress. In the study by Jacobs et al [35], they analyzed data from 7680 users of a commercial weight loss app; high adherence to self-monitoring (ie, logging at least 1 food event within a reasonable time after a meal) was associated with increased weight loss. However, they also found that app users with higher adherence rates had significantly lower body weight at baseline than those with lower adherence rates [35]. The analysis only comprised people who entered data in the app at least once a week for 12 weeks. In our study, 4.9% (n=3) of the app users were still entering data at week 12. Analyzing data from the same commercial app, Carey et al [36] found significant differences in 7 different engagement measures (ie, number of articles read, meals logged, steps recorded, messages to coach, exercise logged, weigh-ins, and days with 1 meal logged per week) between app users with moderate or high weight loss (ie, 5%-10% or >10% body weight loss, respectively) and individuals with no change in body weight (ie, ±1% body weight). Their analysis indicated that people with moderate to high weight loss engaged with all app sections [36]. In our study, only 34% (n=21) of the app users had accessed all of the modules.

Impact of App Use and App Engagement on Behavioral and Biomedical Outcome Measures

Other studies showed promising results for weight loss apps, for example, Carter et al [37], Patel et al [5], and Antoun et al [38]. Specifically, Carter et al [37] conducted a pilot study of 128 volunteers who are overweight comparing a smartphone app (My Meal Mate) with a website and paper diary. They found the mean weight loss over 6 months for the app was higher (4.6 kg, 95% CI 3.0-6.2) than for the diary group (2.9 kg, 95% CI

1.1-4.7) or the website group (1.3 kg, 95% CI 0.1-2.7). Antoun et al [38] in their review of 34 studies that evaluated the use of smartphones for weight loss found an overall mean loss of 2.8 kg (95% CI 2.6-3.0) at 6 months. Patel et al [5] found that consistent tracking was associated with greater weight loss than inconsistent tracking at 6 months (2.1 kg, 95% CI 0.3-4.0). A difference between these studies and ours was that they did not specifically target disadvantaged populations with low health literacy. Therefore, their apps were more complex than ours. In contrast, Lanpher et al [39] developed a weight loss intervention suitable for individuals with low health literacy. A computer algorithm automatically allocated the self-monitoring goals (eg, no sugary drinks, no snacking after dinner, eating 5 fruits and vegetables a week). Participants reported whether they achieved the goals via interactive voice response calls [39]. The algorithm decided which goals to assign next based on previous adherence to goals so that individuals would rather receive goals to which they were receptive [40]. They also received tailored skills training through verbal calls and materials, one-on-one counseling calls, and a membership at the gym [39]. The results showed that the intervention group maintained or lost weight over 12 months, independent of their level of health literacy [38].

Bennett et al [40] extended the intervention to comprise a mobile app. They evaluated its effectiveness in a randomized controlled trial including socioeconomically disadvantaged patients with increased cardiovascular risk by comparing the intervention to usual care [40]. The app used interactive voice responses or text messaging to simplify self-monitoring, like in the previous study. Additionally, participants received in-person coaching and personalized feedback messages immediately after entering data [40]. The intervention group achieved meaningful weight loss, with more than 40% of participants reducing their body weight by at least 5% compared to 17% of participants in the usual care group [39]. Comparing this intervention to ours raises the question of whether the way people had to select and track their goals in our app contributed to the low engagement and the nonsignificant findings. Locke and Latham [41] explained that goal commitment, goal importance, self-efficacy, feedback, and task complexity act as moderators between goals and performance. Potentially, the app did not sufficiently address all 5 moderators.

Association With Other Intervention Components

This study showed that *mysnapp* users were more likely to attend the 6-week face-to-face health check with the practice nurse and to participate in the telephone coaching program than nonusers. Potentially, these individuals were more motivated to lose weight and, therefore, more willing to engage in the other intervention components. Another explanation could be that participants who opted for several intervention components felt more committed to study participation and, therefore, made more use of the individual intervention components. Griauzde et al [42] proposed a similar hypothesis in their mobile health–based prediabetes intervention study; they assumed that participants who received a more robust intervention were more committed to the study and subsequently more likely to complete the 12-week survey. Hutchesson et al [43] concluded that adding nondigital components, such as face-to-face visits and telephone coaching, to mobile health interventions can improve participants' accountability even though these additional features may not be necessary for the intervention's effectiveness.

Limitations

The plan for the randomized controlled trial was to recruit 800 study participants; however, only 215 individuals were able to be recruited (120 in the intervention and 95 in the control group) [13]. Further, despite targeting low socioeconomic areas, this study failed to recruit many participants with low health literacy. One needs to be cautious when interpreting the results of this study due to the small sample size and the high dropout. Despite considerable efforts and additional time to recruit participating practices and patients, the anticipated sample size was not achieved. Research by Perkins et al [44] has shown an ongoing issue with recruitment through Australian general practices. Another problem with the study was that the uptake of intervention components was determined by the clinician and patient. Thus, some chose to just have the app and others to just have the phone coaching. Additionally, the study may not be generalizable to other settings. Since recruitment was from 2 Australian urban areas, results could differ in rural areas or other urban areas. Diet score and health literacy level were self-reported, posing a risk of bias. Further, caution is required when interpreting the results in the context of low health literacy because the baseline health literacy levels were higher than anticipated [12]. According to data from the National Health Survey 2018, the health literacy level in this study's sample was comparable to that of Australians who are overweight or those with obesity in the general population [45]. A potential explanation is that this study's requirements (randomization, completing the questionnaire, and undertaking the health check) stopped people with low health literacy from participating. This rationale is in line with results from Kripalani et al [46], who found that people with low health literacy or numeracy were significantly less interested in participating in research.

Conclusions

There was no association between app use and study outcomes shown to have significantly improved (health literacy and diet) at 6 months. Recruitment and engagement were difficult for this study in disadvantaged populations with low health literacy. A potential explanation could be related to the self-selection of the goals and the weekly submission of the goal achievements. The practice nurses assisted participants at the beginning with the selection of goals. However, these may not have been relevant to participants, and nurses did not receive specific training in selecting meaningful goals for individuals.

However, app users were more likely to attend the 6-week health check and participate in telephone coaching, suggesting that participants who opted for several intervention components felt more committed to this study.

Acknowledgments

The authors would like to acknowledge the contribution to this research by An Tran, Carmel McNamara, Elizabeth Denney-Wilson, Katrina Paine, and Shoko Saito. We also acknowledge Louise Thomas for contributing to the trial protocol and early development. We are grateful for the partnership and support of the South Western Sydney, Adelaide and Nepean and Blue Mountains Primary Health Networks and the Australian Institute of Health Innovation. We would like to acknowledge the general practices and their staff and patients for participating in the research and consumers affiliated with Adelaide PHN for piloting *mysnapp*. This work was supported by the National Health and Medical Research Council of Australia (grant APP1125681, 2017).

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1102 KB - mhealth_v12i1e45942_app1.pdf]

References

1. Lin X, Xu Y, Xu J, Pan X, Song X, Shan L, et al. Global burden of noncommunicable disease attributable to high body mass index in 195 countries and territories, 1990-2017. Endocrine 2020;69(2):310-320 [FREE Full text] [doi: 10.1007/s12020-020-02352-y] [Medline: 32488838]
2. GBD 2017 Risk Factor Collaborators. Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018;392(10159):1923-1994 [FREE Full text] [doi: 10.1016/S0140-6736(18)32225-6] [Medline: 30496105]
3. Overweight and obesity. Australian Institute of Health Welfare. 2020. URL: http://tinyurl.com/24tcj7f6 [accessed 2023-12-28]
4. Health across socioeconomic groups. Australian Institute of Health Welfare. 2020. URL: http://tinyurl.com/5xp2a37d [accessed 2023-12-28]
5. Patel ML, Brooks TL, Bennett GG. Consistent self-monitoring in a commercial app-based intervention for weight loss: results from a randomized trial. J Behav Med 2020;43(3):391-401 [FREE Full text] [doi: 10.1007/s10865-019-00091-8] [Medline: 31396820]
6. Laing BY, Mangione CM, Tseng CH, Leng M, Vaisberg E, Mahida M, et al. Effectiveness of a smartphone application for weight loss compared with usual care in overweight primary care patients: a randomized, controlled trial. Ann Intern Med 2014;161(10 Suppl):S5-S12 [FREE Full text] [doi: 10.7326/M13-3005] [Medline: 25402403]
7. Chin SO, Keum C, Woo J, Park J, Choi HJ, Woo JT, et al. Successful weight reduction and maintenance by using a smartphone application in those with overweight and obesity. Sci Rep 2016;6(1):34563 [FREE Full text] [doi: 10.1038/srep34563] [Medline: 27819345]
8. Schlair S, Moore S, McMacken M, Jay M. How to deliver high-quality obesity counseling in primary care using the 5As framework. J Clin Outcomes Manag 2012;19(5):221-229 [FREE Full text]
9. Get Healthy Service. NSW Government. URL: https://www.gethealthynsw.com.au/ [accessed 2023-12-28]
10. Lau AY, Dunn AG, Mortimer N, Gallagher A, Proudfoot J, Andrews A, et al. Social and self-reflective use of a web-based personally controlled health management system. J Med Internet Res 2013;15(9):e211 [FREE Full text] [doi: 10.2196/jmir.2682] [Medline: 24064188]
11. Osborne RH, Batterham RW, Elsworth GR, Hawkins M, Buchbinder R. The grounded psychometric development and initial validation of the Health Literacy Questionnaire (HLQ). BMC Public Health 2013;13(1):658 [FREE Full text] [doi: 10.1186/1471-2458-13-658] [Medline: 23855504]
12. Parker SM, Barr M, Stocks N, Denney-Wilson E, Zwar N, Karnon J, et al. Preventing chronic disease in overweight and obese patients with low health literacy using eHealth and teamwork in primary healthcare (HeLP-GP): a cluster randomised controlled trial. BMJ Open 2022;12(11):e060393 [FREE Full text] [doi: 10.1136/bmjopen-2021-060393] [Medline: 36450426]
13. Parker SM, Stocks N, Nutbeam D, Thomas L, Denney-Wilson E, Zwar N, et al. Preventing chronic disease in patients with low health literacy using eHealth and teamwork in primary healthcare: protocol for a cluster randomised controlled trial. BMJ Open 2018;8(6):e023239 [FREE Full text] [doi: 10.1136/bmjopen-2018-023239] [Medline: 29866737]
14. Socio-Economic Indexes for Areas (SEIFA), Australia. Australian Bureau of Statistics. 2023. URL: http://tinyurl.com/5xp2a37d [accessed 2023-12-28]
15. Health literacy, Australia. Australian Bureau of Statistics. 2006. URL: http://tinyurl.com/zdpzcnb7 [accessed 2023-12-28]
16. Health literacy. Australian Institute of Health and Welfare. 2022. URL: https://www.aihw.gov.au/reports/australias-health/health-literacy [accessed 2023-08-20]

17. Free telephone-based health coaching. Get Healthy Information and Coaching Service. 2022. URL: https://www.gethealthynsw.com.au/ [accessed 2022-05-05]
18. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
19. DiFilippo KN, Huang WH, Andrade JE, Chapman-Novakofski KM. The use of mobile apps to improve nutrition outcomes: a systematic literature review. J Telemed Telecare 2015;21(5):243-253 [FREE Full text] [doi: 10.1177/1357633X15572203] [Medline: 25680388]
20. Cook A, Roberts K, O'Leary F, Allman-Farinelli MA. Comparison of single questions and brief questionnaire with longer validated food frequency questionnaire to assess adequate fruit and vegetable intake. Nutrition 2015;31(7-8):941-947 [FREE Full text] [doi: 10.1016/j.nut.2015.01.006] [Medline: 26003391]
21. McGill R, Tukey JW, Larsen WA. Variations of box plots. Am Stat 1978;32(1):12-16 [FREE Full text] [doi: 10.2307/2683468]
22. Royston JP. An extension of Shapiro and Wilk's W test for normality to large samples. J R Stat Soc Ser C Appl Stat 1982;31(2):115-124 [FREE Full text] [doi: 10.2307/2347973]
23. Royston JP. Algorithm AS 181: the W test for normality. Appl Stat 1982;31(2):176-180 [FREE Full text] [doi: 10.2307/2347986]
24. Royston P. Remark AS R94: a remark on algorithm AS 181: the w-test for normality. J R Stat Soc Ser C Appl Stat 1995;44(4):547-551 [FREE Full text] [doi: 10.2307/2986146]
25. Student. The probable error of a mean. Biometrika 1908;6(1):1-25 [FREE Full text] [doi: 10.2307/2331554]
26. Bauer DF. Constructing confidence sets using rank statistics. J Am Stat Assoc 1972;67(339):687-690 [FREE Full text] [doi: 10.1080/01621459.1972.10481279]
27. Hollander M, Wolfe DA. A distribution-free rank sum test (Wilcoxon). In: Nonparametric Statistical Methods. New York, United States of America: John Wiley & Sons; 1973:68-75.
28. Hollander M, Wolfe DA. A distribution-free test (Kruskal-Wallis). In: Nonparametric Statistical Methods. New York, United States of America: John Wiley & Sons; 1973:115-120.
29. Hope ACA. A simplified monte carlo significance test procedure. J R Stat Soc Series B Stat Methodol 1968;30(3):582-598 [FREE Full text] [doi: 10.1111/j.2517-6161.1968.tb00759.x]
30. Patefield WM. Algorithm AS 159: an efficient method of generating random R × C tables with given row and column totals. J R Stat Soc Series C Appl Stat 1981;30(1):91-97 [FREE Full text] [doi: 10.2307/2346669]
31. Agresti A. An Introduction to Categorical Data Analysis. 2 Edition. New York, United States of America: John Wiley & Sons; 2007:38.
32. Best DJ, Roberts DE. Algorithm AS 89: the upper tail probabilities of Spearman's rho. J R Stat Soc Series C Appl Stat 1975;24(3):377-379 [FREE Full text] [doi: 10.2307/2347111]
33. Hollander M, Wolfe DA. A distribution-free test for independence (Kendall). In: Nonparametric Statistical Methods. New York, United States of America: John Wiley and Sons; 1973:185-194.
34. Turner-McGrievy GM, Dunn CG, Wilcox S, Boutté AK, Hutto B, Hoover A, et al. Defining adherence to mobile dietary self-monitoring and assessing tracking over time: tracking at least two eating occasions per day is best marker of adherence within two different mobile health randomized weight loss interventions. J Acad Nutr Diet 2019;119(9):1516-1524 [FREE Full text] [doi: 10.1016/j.jand.2019.03.012] [Medline: 31155473]
35. Jacobs S, Radnitz C, Hildebrandt T. Adherence as a predictor of weight loss in a commonly used smartphone application. Obes Res Clin Pract 2017;11(2):206-214 [FREE Full text] [doi: 10.1016/j.orcp.2016.05.001] [Medline: 27292942]
36. Carey A, Yang Q, DeLuca L, Toro-Ramos T, Kim Y, Michaelides A. The relationship between weight loss outcomes and engagement in a mobile behavioral change intervention: retrospective analysis. JMIR mHealth uHealth 2021;9(11):e30622 [FREE Full text] [doi: 10.2196/30622] [Medline: 34747706]
37. Carter MC, Burley VJ, Nykjaer C, Cade JE. Adherence to a smartphone application for weight loss compared to website and paper diary: pilot randomized controlled trial. J Med Internet Res 2013;15(4):e32 [FREE Full text] [doi: 10.2196/jmir.2283] [Medline: 23587561]
38. Antoun J, Itani H, Alarab N, Elsehmawy A. The effectiveness of combining nonmobile interventions with the use of smartphone apps with various features for weight loss: systematic review and meta-analysis. JMIR mHealth uHealth 2022;10(4):e35479 [FREE Full text] [doi: 10.2196/35479] [Medline: 35394443]
39. Lanpher MG, Askew S, Bennett GG. Health literacy and weight change in a digital health intervention for women: a randomized controlled trial in primary care practice. J Health Commun 2016;21(Suppl 1):34-42 [FREE Full text] [doi: 10.1080/10810730.2015.1131773] [Medline: 27043756]
40. Bennett GG, Steinberg D, Askew S, Levine E, Foley P, Batch BC, et al. Effectiveness of an app and provider counseling for obesity treatment in primary care. Am J Prev Med 2018;55(6):777-786 [FREE Full text] [doi: 10.1016/j.amepre.2018.07.005] [Medline: 30361140]
41. Locke EA, Latham GP. Building a practically useful theory of goal setting and task motivation. A 35-year odyssey. Am Psychol 2002;57(9):705-717. [doi: 10.1037//0003-066x.57.9.705] [Medline: 12237980]

42. Griauzde D, Kullgren JT, Liestenfeltz B, Ansari T, Johnson EH, Fedewa A, et al. A mobile phone-based program to promote healthy behaviors among adults with prediabetes who declined participation in free diabetes prevention programs: mixed-methods pilot randomized controlled trial. JMIR mHealth uHealth 2019;7(1):e11267 [FREE Full text] [doi: 10.2196/11267] [Medline: 30626566]
43. Hutchesson MJ, Rollo ME, Krukowski R, Ells L, Harvey J, Morgan PJ, et al. eHealth interventions for the prevention and treatment of overweight and obesity in adults: a systematic review with meta-analysis. Obes Rev 2015;16(5):376-392 [FREE Full text] [doi: 10.1111/obr.12268] [Medline: 25753009]
44. Perkins D, Harris MF, Tan J, Christl B, Taggart J, Fanaian M. Engaging participants in a complex intervention trial in Australian general practice. BMC Med Res Methodol 2008;8(1):55 [FREE Full text] [doi: 10.1186/1471-2288-8-55] [Medline: 18700984]
45. National health survey: health literacy. Australian Bureau of Statistics. 2019. URL: http://tinyurl.com/9229hha9 [accessed 2022-07-19]
46. Kripalani S, Heerman WJ, Patel NJ, Jackson N, Goggins K, Rothman RL, et al. Association of health literacy and numeracy with interest in research participation. J Gen Intern Med 2019;34(4):544-551 [FREE Full text] [doi: 10.1007/s11606-018-4766-2] [Medline: 30684202]

Abbreviations

GP: general practice
HeLP-GP: Health eLiteracy for Prevention in General Practice

Edited by L Buis; submitted 22.01.23; peer-reviewed by L Bell, J Job; comments to author 20.04.23; revised version received 21.08.23; accepted 19.12.23; published 09.02.24.

Please cite as:
Buss VH, Barr M, Parker SM, Kabir A, Lau AYS, Liaw ST, Stocks N, Harris MF
Mobile App Intervention of a Randomized Controlled Trial for Patients With Obesity and Those Who Are Overweight in General Practice: User Engagement Analysis Quantitative Study
JMIR Mhealth Uhealth 2024;12:e45942
URL: https://mhealth.jmir.org/2024/1/e45942
doi:10.2196/45942
PMID:38335014

Original Paper

The Association of Macronutrient Consumption and BMI to Exhaled Carbon Dioxide in Lumen Users: Retrospective Real-World Study

Shlomo Yeshurun[1], PhD; Tomer Cramer[1], PhD; Daniel Souroujon[1,2], MD, MPH; Merav Mor[1], PhD

[1]Metaflow Ltd, Tel-Aviv, Israel

[2]School of Public Health, Tel Aviv University, Tel-Aviv, Israel

Corresponding Author:
Shlomo Yeshurun, PhD
Metaflow Ltd
HaArba'a St 30
Tel-Aviv, 6473926
Israel
Phone: 972 37684062
Email: shlomoyesh@gmail.com

Abstract

Background: Metabolic flexibility is the ability of the body to rapidly switch between fuel sources based on their accessibility and metabolic requirements. High metabolic flexibility is associated with improved health outcomes and a reduced risk of several metabolic disorders. Metabolic flexibility can be improved through lifestyle changes, such as increasing physical activity and eating a balanced macronutrient diet. Lumen is a small handheld device that measures metabolic fuel usage through exhaled carbon dioxide (CO_2), which allows individuals to monitor their metabolic flexibility and make lifestyle changes to enhance it.

Objective: This retrospective study aims to examine the postprandial CO_2 response to meals logged by Lumen users and its relationship with macronutrient intake and BMI.

Methods: We analyzed deidentified data from 2607 Lumen users who logged their meals and measured their exhaled CO_2 before and after those meals between May 1, 2023, and October 18, 2023. A linear mixed model was fitted to test the association between macronutrient consumption, BMI, age, and gender to the postprandial CO_2 response, followed by a 2-way ANOVA.

Results: The model demonstrated significant associations (P<.001) between CO_2 response after meals and both BMI and carbohydrate intake (BMI: β=–0.112, 95% CI –0.156 to –0.069; carbohydrates: β=0.046, 95% CI 0.034-0.058). In addition, a 2-way ANOVA revealed that higher carbohydrate intake resulted in a higher CO_2 response compared to low carbohydrate intake ($F_{2,2569}$=24.23; P<.001), and users with high BMI showed modest responses to meals compared with low BMI ($F_{2,2569}$=5.88; P=.003).

Conclusions: In this study, we show that Lumen's CO_2 response is influenced both by macronutrient consumption and BMI. The results of this study highlight a distinct pattern of reduced metabolic flexibility in users with obesity, indicating the value of Lumen for assessing postprandial metabolic flexibility.

(JMIR Mhealth Uhealth 2024;12:e56083) doi:10.2196/56083

KEYWORDS

app; applications; association; BMI; body mass index; carbohydrate; carbon dioxide; consumption; correlate; correlation; diet; dietary; exhalation; exhale; food; Lumen; macronutrient; meal; metabolic flexibility; metabolic; metabolism; mHealth; mobile health; nutrient; nutrition; nutritional; obese; obesity; postprandial; prandial; retrospective; weight

Introduction

Background

The presence of obesity is the leading risk factor for metabolic disorders such as type 2 diabetes (T2D) and cardiovascular diseases (CVDs), and it has been linked to a reduced life expectancy [1,2]. Metabolic syndrome represents a collection of metabolic abnormalities that includes obesity as well as insulin resistance, hypertension, and dyslipidemia [3]. Changing the macronutrient distribution, such as low-fat or low-carbohydrate diets, has been proposed for the treatment and management of metabolic syndrome [4], and they have been shown to improve several clinical features of metabolic syndrome, including weight loss and cardiovascular risk [5,6].

Metabolic flexibility is the ability to switch between fuel sources (such as carbohydrates and fats) in response to their availability and metabolic demands, which is crucial for maintaining overall health and well-being [7,8]. Lifestyle modifications, such as exercise and a balanced diet, can therefore enhance metabolic flexibility, and high metabolic flexibility is associated with improved health outcomes, including a reduced risk of metabolic syndrome [7]. In addition, it was found to be impaired in individuals with obesity and T2D compared to lean and healthy individuals [9,10]. Assessment of metabolic flexibility can be conducted through postprandial changes (following a meal or any other insulin stimulation) in the respiratory exchange ratio (RER) from the metabolic cart [11,12], which estimates the body's preference for macronutrient oxidation. Accordingly, several studies showed different RER responses between participants with high metabolic flexibility compared with participants with low flexibility for fasting, a high-carbohydrate diet, and a high-fat diet [13]. A recent study has shown that RER is greater after high-carbohydrate overfeeding than high-fat overfeeding, and impaired metabolic flexibility, as measured by ΔRER, is associated with greater weight gain over the following 6- and 12-month periods [14]. Furthermore, ΔRER is significantly elevated after high-carbohydrate intake compared with high-fat intake, indicating that metabolic flexibility is influenced by macronutrient composition [15]. These studies show that metabolic flexibility can be assessed based on RER from the metabolic cart.

However, metabolic carts are expensive, time-consuming, and only available in health care laboratories [16]. Thus, a small handheld device was developed to measure metabolic fuel usage by analyzing exhaled carbon dioxide (CO_2). The Lumen device is a portable breath-analyzer that measures metabolic fuel use through exhaled CO_2 and was found to be in agreement with the RER measured by the metabolic cart [17]. Furthermore, Lumen was found to be able to detect different metabolic responses to low- or high-carb lifestyles through CO_2 changes [18].

A variety of mobile health apps, including Lumen, which are designed for mobility and ubiquity, have demonstrated promising results in enhancing metabolism and facilitating weight management [19]. Using real-time data tracking, these tools empower individuals to make more informed health decisions [20]. As such, mobile apps such as Lumen have the potential to make a positive difference in people's health and well-being [21].

Objective

This study examined data collected from Lumen users' exhaled CO_2 measurements taken before and after logging a meal. The objective of this analysis was to investigate the postprandial response of Lumen's CO_2 measurements to meals logged by Lumen users and to understand the association of this response with macronutrient consumption as well as users' age, gender, and BMI.

Methods

Ethical Considerations

This study was determined to be exempt from institutional board review (IRB) under category 2, as detailed in 45 CFR 46.104(d) and the standard operating procedure of the Biomedical Research Alliance of New York (BRANY), by the BRANY Social, Behavioral, and Educational Research (SBER) IRB on May 9, 2023 (BRANY IRB File 23-119-1476).

Study Design

This is a retrospective observational study based on deidentified data collected from users of the Lumen device and app.

Participants

Participants' data were collected retrospectively between May 1, 2023, and October 18, 2023. All users in the analysis were aged 18 years or older. Since only 12 users who were underweight (BMI≤18.5) were found in the database, they were removed from the analysis.

Data Sources

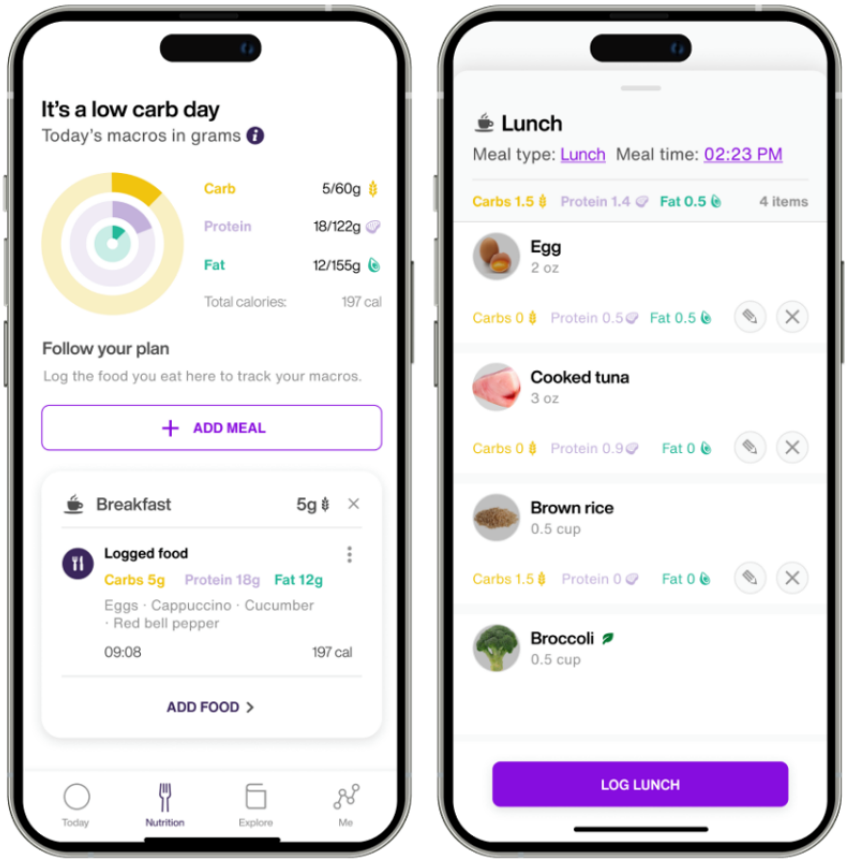
As part of the onboarding process, users were required to specify their gender, date of birth, height, and weight. The use of the Lumen app includes an optional morning fasted measurement with the Lumen device, as well as recommended pre- and postmeal measurements throughout the day and a bedtime measurement before sleeping. The app provides nutritional recommendations based on the user's personal preferences as well as their morning CO_2 measurement regarding the amount of macronutrients one should consume during the day. The app also allows users to log their meals—whether they are breakfast, lunch, dinner, or snacks. The macronutrient composition of these meals is determined by the nutritional database of Nutritionix [22], which includes all nutritional data available from the United States Department of Agriculture's (USDA) Food Composition Database, restaurant chain data, and foods added by Nutritionix's dietitians. An example of the food log feature in the app is shown in Figure 1.

Exhaled CO_2 measurements were obtained using the Lumen device (Metaflow Ltd). The Lumen mobile app guides the user through each phase of the Lumen maneuver (inhale, breath hold, and exhale), as previously described [17].

In this study, premeal and fasting measurements were defined as "premeal %CO_2," while postmeal and bedtime measurements

were considered to be the "postmeal $\%CO_2$" measure, with a maximum of 210 minutes between pre- and postmeal measurements selected for the analysis. The relative change in $\%CO_2$ from premeal to postmeal was calculated for the final analysis. In addition, meals that were tagged as either breakfast, lunch, or dinner were selected for the analysis, while snacks were removed. Further data were excluded from the analysis using the Tukey method for outlier removal with an IQR of 1.5 for all macronutrients.

Figure 1. Lumen mobile app screenshots detailing the logged meals and their macronutrient distribution.



Statistical Analysis

Data were analyzed using Python JupyterLab (version 3.6.3; Project Jupyter), and all statistical analyses were conducted with the Python programming language, using custom scripts and the *statsmodels* package [23]. Figures were made with GraphPad Prism (version 10.1.0 for Windows, GraphPad Software) [24].

A linear mixed model (LMM) was fitted in order to test the relationship between the quantity of macronutrients (carbohydrates, fats, and proteins) and personal information (gender, age, and BMI) to the outcome variable of percentage of change in $\%CO_2$ from premeal to postmeal, where users' unique ID was adjusted for as a random effect [25]. A 2-way ANOVA was used to test the differences between different groups of statistically significant variables in the LMM. Among all the analyses performed, a 2-sided $P \leq .05$ was considered statistically significant.

Results

Participants

Overall, a total of 2607 users who completed 6671 coupled pre- and postmeal sessions from 6207 logged meals were used in the final analysis, with most users contributing only 1 session into the analysis (median 1, IQR 1-2 sessions per user).

Descriptive

A total of 81.97% (2137/2607) of the participants were women, and men and women did not differ in their ages or BMIs. A summary of the characteristics of the participants in the study is presented in Table 1.

Additionally, the macronutrient composition of the meals consumed by the participants was primarily carbohydrates, followed by proteins and fats, as detailed in Table 2.

Table 1. Sample characteristics of users (N=2607).

Characteristics	Values
Gender (female), n (%)	2137 (81.97)
Age (years), mean (SD)	47.6 (9.4)
BMI (kg/m^2), mean (SD)	28.8 (6.0)

Table 2. Macronutrient composition of meals (N=6207).

Parameters	Values, mean (SD)
Carbohydrates (grams)	32.5 (20.5)
Proteins (grams)	29.7 (14.7)
Fats (grams)	22.5 (13.7)

Evaluation Outcomes

The overall CO_2 response after meal consumption at various time points is shown in Figure 2. As expected, %CO_2 measured by Lumen 30 to 210 minutes following meal intake increased significantly compared to premeal measurement.

The relationship between postprandial CO_2 response and macronutrient intake, as recorded in the Lumen app and users' personal information, was examined using Linear Mixed Modeling (LMM). The users' unique ID was incorporated as a random effect in the analysis. The model revealed a statistically significant relationship between the CO_2 response and the BMI of the users (P<.001), as well as their carbohydrate intake (P<.001). Lower BMI predicted a more significant increase in postprandial CO_2 response, and increased carbohydrate intake predicted a greater postprandial CO_2 response (Table 3). In contrast, their postprandial CO_2 response was not significantly influenced by their age, gender, fat intake, or protein intake (all P>.05; Table 3).

As the LMM analysis demonstrated significant associations between BMI, carbohydrate intake, and postprandial CO_2 response, these parameters were then divided into 3 categories each. BMI was classified as healthy (18.5-25 kg/m^2), overweight (25-30 kg/m^2), and obese (≥30 kg/m^2), while carbohydrate intake was classified as low (0-30 grams), medium (30-60 grams), and high (≥60 grams).

To examine how the CO_2 response differed between each BMI category for each carbohydrate intake class, we conducted a 2-way ANOVA. In this analysis, only the last session of each user was used to eliminate the effect within the same user's sessions. In agreement with the results from the LMM, this analysis revealed a significant effect of users' BMI category ($F_{2,2569}$=5.88; P=.003) and level of carbohydrate intake ($F_{2,2569}$=24.23; P<.001), where users with obesity tended to have a lower postprandial CO_2 response than those with healthy BMI, and high consumption of carbohydrates resulted in a higher CO_2 response (Figure 3). Nevertheless, users with high BMI had similar elevated CO_2 response as overweight and healthy users when consuming meals high in carbohydrates.

Figure 2. Exhaled carbon dioxide (CO_2) response to meal at different time intervals (N=6671). Results represent mean (95% CI).

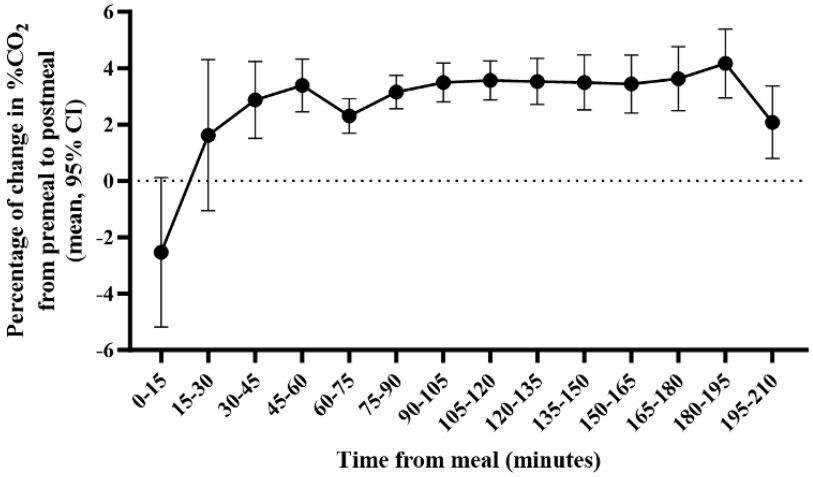
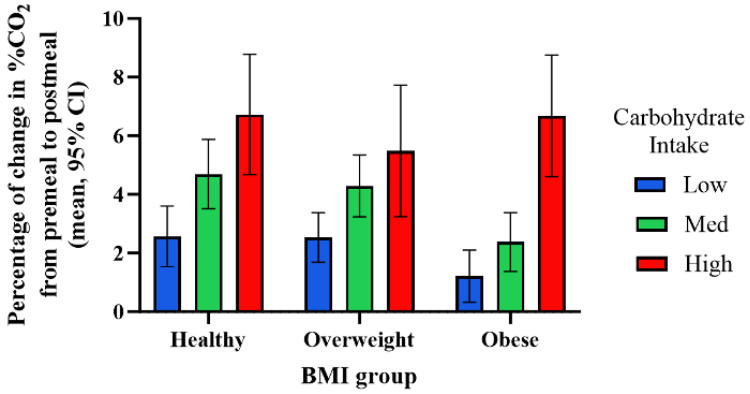


Table 3. Determinants of CO_2 response to food intake (N=6671).

Variables	β (95% CI)	*z* statistics	*P* value
Intercept	5.221 (3.242 to 7.199)	5.172	<.001
BMI	–0.112 (–0.156 to –0.069)	–5.026	<.001
Age	–0.021 (–0.049 to 0.007)	–1.483	.14
Gender	–0.612 (–1.324 to 0.100)	–1.683	.09
Carbohydrates	0.046 (0.034 to 0.058)	7.496	<.001
Fat	0.018 (–0.001 to 0.038)	1.816	.07
Protein	0.016 (–0.002 to 0.035)	1.755	.08

Figure 3. Carbon dioxide (CO_2) response to different carbohydrates intake in different BMI categories (N=2607). Results represent mean (95% CI).



Discussion

Principal Findings

This study showed how the Lumen device can detect the $\%CO_2$ response to meals with a different macronutrient composition (mixed meals). Notably, this response is influenced by the carbohydrate consumption in each meal as well as the BMI of the user. Lumen's measured exhaled $\%CO_2$ was higher when carbohydrate intake was higher, in accordance with previous studies describing the association between RER and carbohydrates [11,14,15,18]. In addition, it reveals metabolic inflexibility in individuals with obesity as they show a reduced $\%CO_2$ response compared with healthy individuals, which again is in accordance with the literature on metabolic flexibility [8,9,26].

Comparison With Previous Work

A recent prospective study has shown that a high-carbohydrate diet results in an elevation of Lumen's $\%CO_2$ compared with a low-carbohydrate diet [18]. In this retrospective analysis of Lumen users, we show similar results, as $\%CO_2$ is elevated when carbohydrate consumption is increased, which highlights Lumen's ability to detect changes in carbohydrate intake. Although previous studies that included high-fat and high-protein consumption resulted in elevated RER [14,27,28], we did not observe this in this study. Further study with a larger sample size is warranted to check for the effect of dietary fats and proteins on exhaled $\%CO_2$.

The duration of $\%CO_2$ elevation has shown a consistent plateau, spanning from 45 minutes postmeal to the last checkup at 210 minutes, where a subsequent decline was observed. Lumen users often analyze their postmeal breath between 60 and 120 minutes, so understanding how $\%CO_2$ responds over longer periods is challenging. Woerle et al [29] showed carbohydrate oxidation up to 360 minutes after meal consumption, long after glucose and insulin levels in the plasma peaked. Future prospective studies should address when $\%CO_2$ decreases after a mixed macronutrient meal and how it changes compared to other metabolites.

The impact of postprandial measurement of glucose levels on metabolic disease prevention has been discussed in previous studies [30]. Novel technological advancements have made it available to consumers, mostly with recent developments in continuous glucose monitoring (CGM) [31]. Other metabolic indicators, such as insulin, lipids, metabolomics, and indirect calorimetry, have also been used to assess metabolic flexibility and were found to be predictive of body weight, metabolic syndrome, T2D, and CVDs [11,25,32,33]. As of today, CGM devices need replacing every 2 weeks and are somewhat invasive, whereas most of the other aforementioned measurements can only be performed in a laboratory setting. In contrast, the Lumen device can be used at home and measures an individual's metabolic state with a simple breath maneuver, which was found to be in agreement with the RER measured in the metabolic cart [17]. Moreover, compared to the metabolic cart, the portability of the Lumen device enables the collection of CO_2 measurements from a large number of participants after

consuming mixed macronutrient meals, which provides a higher level of external validity.

In addition to the effect of carbohydrate consumption, BMI was also found to be a key parameter in the postprandial $\%CO_2$ response, as a higher BMI resulted in a reduced response. These results are in accordance with the literature from the metabolic cart, where participants with obesity and T2D showed low ΔRER after insulin stimulation [9,10].

In this analysis, the majority of participants were women (2137/2607, 81.97%), in line with similar nutrition and weight management studies finding that women participate 3 times more often than men [34,35]. Age and gender were both statistically insignificant in their effect on the postprandial $\%CO_2$ response in the LMM, while some studies mention that metabolic flexibility is reduced with age [36]. Although progesterone is directly associated with CO_2 [37,38], our analysis did not reveal any influence of gender on metabolic flexibility, possibly since women may have been in different phases of their menstrual cycle or at menopause. Future studies should investigate the effects of gender and the menstrual cycle on Lumen's $\%CO_2$.

This study shows the potential of the Lumen device and mobile app to improve outcomes for people with metabolic disorders. There is increasing evidence showing that mobile health technologies can improve metabolic outcomes, in particular for lifestyle modifications in T2D and weight loss [39-41]. While most of these technologies incorporate a mobile app only, many studies have shown the benefits of an app accompanied by a device capable of providing real-time feedback on how those lifestyle modifications affect their measurements, which can improve engagement with them [42-45]. In a recent pilot clinical study, Lumen device and app usage improved several metabolic parameters in prediabetic patients after 3 months [46]. In light of these findings, the potential public health implications of mobile and ubiquitous health tools such as Lumen are noteworthy, showing great promise for addressing the epidemic of metabolic dysfunction across the globe in the near future.

Limitations

Despite the large data set of real-world evidence that was used in this analysis, several limitations need to be mentioned. First, due to its retrospective and observational nature, we cannot identify causal relationships between any of the variables in this study. Second, since this analysis is based on real-world data, the macronutrient consumption of each meal may be miscalculated or incomplete (recall bias), and a misuse of the device might occur as well. Moreover, the heterogeneous composition of the meals makes it difficult to determine if other factors in the meal might influence the $\%CO_2$ as well, in particular the type of carbohydrate and the fiber composition. In addition, Lumen users' macronutrient intake might not be representative of most typical diets, as the Lumen app guides them toward a specific diet, primarily low in carbohydrate, and with most of them on a weight loss journey. As this study has specific user characteristics, caution should be exercised in interpreting these conclusions for a broader audience. Furthermore, this analysis did not consider the fasted $\%CO_2$ levels, which, together with the postprandial levels, could also be useful in the metabolic flexibility assessment [47]. Nonetheless, we believe that due to the nature of the outcome variable, which is the percentage of change from premeal to postmeal, it might have limited to no effect on the current results.

Lastly, postprandial $\%CO_2$ might be affected by other confounding variables, including preexisting conditions; medication use; dietary restrictions; the menstrual cycle; and lifestyle factors such as stress, sleep, and physical activity, which were not controlled for in this study. Future studies, both prospective and retrospective, should aim to address these limitations, particularly through controlled prospective investigations involving diverse populations and retrospective analyses that comprehensively consider potential covariates.

Conclusions

In conclusion, Lumen's ability to evaluate metabolic flexibility following mixed meals was demonstrated through this retrospective analysis of postprandial exhaled $\%CO_2$, in which increased $\%CO_2$ was specific to carbohydrate consumption but not to fat consumption. Furthermore, a moderate $\%CO_2$ response was also observed among users with a high BMI, which suggests metabolic inflexibility among this group. As such, we propose Lumen as an accessible tool allowing individuals to make informed dietary choices conducive to metabolic health.

Acknowledgments

This study was funded by Metaflow Ltd. The authors would like to express their thanks to all app users who contributed their data to this study and to the Lumen team for their support. We would also like to thank all the reviewers for their timely and constructive inputs to this manuscript.

Authors' Contributions

SY, TC, and MM led the design and conceptualization of the study. MM was the principal supervisor of the study. SY led the curation of the data and conducted data interpretation and formal analysis. SY wrote the first draft of the manuscript. TC and DS contributed to the data analysis. All authors critically revised the manuscript and gave their final approval.

Conflicts of Interest

MM is the chief scientific officer and the cofounder of MetaFlow Ltd. DS is a consultant at MetaFlow. SY and TC are paid employees at MetaFlow.

References

1. Powell-Wiley TM, Poirier P, Burke LE, Després JP, Gordon-Larsen P, Lavie CJ, et al. Obesity and cardiovascular disease: a scientific statement from the American Heart Association. Circulation 2021;143(21):e984-e1010 [FREE Full text] [doi: 10.1161/CIR.0000000000000973] [Medline: 33882682]
2. Cercato C, Fonseca FA. Cardiovascular risk and obesity. Diabetol Metab Syndr 2019;11:74 [FREE Full text] [doi: 10.1186/s13098-019-0468-0] [Medline: 31467596]
3. Alberti KGMM, Eckel RH, Grundy SM, Zimmet PZ, Cleeman JI, Donato KA, et al. Harmonizing the metabolic syndrome: a joint interim statement of the International Diabetes Federation Task Force on Epidemiology and Prevention; National Heart, Lung, and Blood Institute; American Heart Association; World Heart Federation; International Atherosclerosis Society; and International Association for the Study of Obesity. Circulation 2009;120(16):1640-1645 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.109.192644] [Medline: 19805654]
4. Castro-Barquero S, Ruiz-León AM, Sierra-Pérez M, Estruch R, Casas R. Dietary strategies for metabolic syndrome: a comprehensive review. Nutrients 2020;12(10):2983 [FREE Full text] [doi: 10.3390/nu12102983] [Medline: 33003472]
5. Ge L, Sadeghirad B, Ball GDC, da Costa BR, Hitchcock CL, Svendrovski A, et al. Comparison of dietary macronutrient patterns of 14 popular named dietary programmes for weight and cardiovascular risk factor reduction in adults: systematic review and network meta-analysis of randomised trials. BMJ 2020;369:m696 [FREE Full text] [doi: 10.1136/bmj.m696] [Medline: 32238384]
6. Danesi F, Mengucci C, Vita S, Bub A, Seifert S, Malpuech-Brugère C, et al. Unveiling the correlation between inadequate energy/macronutrient intake and clinical alterations in volunteers at risk of metabolic syndrome by a predictive model. Nutrients 2021;13(4):1377 [FREE Full text] [doi: 10.3390/nu13041377] [Medline: 33923923]
7. Smith RL, Soeters MR, Wüst RCI, Houtkooper RH. Metabolic flexibility as an adaptation to energy resources and requirements in health and disease. Endocr Rev 2018;39(4):489-517 [FREE Full text] [doi: 10.1210/er.2017-00211] [Medline: 29697773]
8. Goodpaster BH, Sparks LM. Metabolic flexibility in health and disease. Cell Metab 2017;25(5):1027-1036 [FREE Full text] [doi: 10.1016/j.cmet.2017.04.015] [Medline: 28467922]
9. Kelley DE, Goodpaster B, Wing RR, Simoneau JA. Skeletal muscle fatty acid metabolism in association with insulin resistance, obesity, and weight loss. Am J Physiol 1999;277(6):E1130-E1141 [FREE Full text] [doi: 10.1152/ajpendo.1999.277.6.E1130] [Medline: 10600804]
10. Kelley DE, Mandarino LJ. Fuel selection in human skeletal muscle in insulin resistance: a reexamination. Diabetes 2000;49(5):677-683 [FREE Full text] [doi: 10.2337/diabetes.49.5.677] [Medline: 10905472]
11. Yu EA, Le NA, Stein AD. Measuring postprandial metabolic flexibility to assess metabolic health and disease. J Nutr 2021;151(11):3284-3291 [FREE Full text] [doi: 10.1093/jn/nxab263] [Medline: 34293154]
12. Glaves A, Díaz-Castro F, Farías J, Ramírez-Romero R, Galgani JE, Fernández-Verdejo R. Association between adipose tissue characteristics and metabolic flexibility in humans: a systematic review. Front Nutr 2021;8:744187 [FREE Full text] [doi: 10.3389/fnut.2021.744187] [Medline: 34926544]
13. Galgani JE, Fernández-Verdejo R. Pathophysiological role of metabolic flexibility on metabolic health. Obes Rev 2021;22(2):e13131. [doi: 10.1111/obr.13131] [Medline: 32815226]
14. Begaye B, Vinales KL, Hollstein T, Ando T, Walter M, Bogardus C, et al. Impaired metabolic flexibility to high-fat overfeeding predicts future weight gain in healthy adults. Diabetes 2020;69(2):181-192 [FREE Full text] [doi: 10.2337/db19-0719] [Medline: 31712321]
15. Gribok A, Leger JL, Stevens M, Hoyt R, Buller M, Rumpler W. Measuring the short-term substrate utilization response to high-carbohydrate and high-fat meals in the whole-body indirect calorimeter. Physiol Rep 2016;4(12):e12835 [FREE Full text] [doi: 10.14814/phy2.12835] [Medline: 27354539]
16. Delsoglio M, Achamrah N, Berger MM, Pichard C. Indirect calorimetry in clinical practice. J Clin Med 2019;8(9):1387 [FREE Full text] [doi: 10.3390/jcm8091387] [Medline: 31491883]
17. Lorenz KA, Yeshurun S, Aziz R, Ortiz-Delatorre J, Bagley JR, Mor M, et al. A handheld metabolic device (Lumen) to measure fuel utilization in healthy young adults: device validation study. Interact J Med Res 2021;10(2):e25371 [FREE Full text] [doi: 10.2196/25371] [Medline: 33870899]
18. Roberts J, Dugdale-Duwell D, Lillis J, Pinto JM, Willmott A, Yeshurun S, et al. The efficacy of a home-use metabolic device (Lumen) in response to a short-term low and high carbohydrate diet in healthy volunteers. J Int Soc Sports Nutr 2023;20(1):2185537 [FREE Full text] [doi: 10.1080/15502783.2023.2185537] [Medline: 36862060]
19. Ghelani DP, Moran LJ, Johnson C, Mousa A, Naderpoor N. Mobile apps for weight management: a review of the latest evidence to inform practice. Front Endocrinol (Lausanne) 2020;11:412 [FREE Full text] [doi: 10.3389/fendo.2020.00412] [Medline: 32670197]

20. Kang HS, Exworthy M. Wearing the future-wearables to empower users to take greater responsibility for their health and care: scoping review. JMIR Mhealth Uhealth 2022;10(7):e35684 [FREE Full text] [doi: 10.2196/35684] [Medline: 35830222]
21. Han M, Lee E. Effectiveness of mobile health application use to improve health behavior changes: a systematic review of randomized controlled trials. Healthc Inform Res 2018;24(3):207-226 [FREE Full text] [doi: 10.4258/hir.2018.24.3.207] [Medline: 30109154]
22. Nutrition API. Nutritionix. URL: https://www.nutritionix.com/business/api [accessed 2024-02-08]
23. Seabold S, Perktold J. Statsmodels: econometric and statistical modeling with Python. 2010 Presented at: Proceedings of the 9th Python in Science Conference; 2010; Austin p. 92-96. [doi: 10.25080/majora-92bf1922-011]
24. PRISM: analyze, graph and present your scientific work. GraphPad. URL: https://www.graphpad.com/ [accessed 2024-03-12]
25. Lépine G, Tremblay-Franco M, Bouder S, Dimina L, Fouillet H, Mariotti F, et al. Investigating the postprandial metabolome after challenge tests to assess metabolic flexibility and dysregulations associated with cardiometabolic diseases. Nutrients 2022;14(3):472 [FREE Full text] [doi: 10.3390/nu14030472] [Medline: 35276829]
26. Galgani JE, Moro C, Ravussin E. Metabolic flexibility and insulin resistance. Am J Physiol Endocrinol Metab 2008;295(5):E1009-e1017 [FREE Full text] [doi: 10.1152/ajpendo.90558.2008] [Medline: 18765680]
27. Hollstein T, Basolo A, Ando T, Krakoff J, Piaggi P. Reduced adaptive thermogenesis during acute protein-imbalanced overfeeding is a metabolic hallmark of the human thrifty phenotype. Am J Clin Nutr 2021;114(4):1396-1407 [FREE Full text] [doi: 10.1093/ajcn/nqab209] [Medline: 34225360]
28. Flint A, Helt B, Raben A, Toubro S, Astrup A. Effects of different dietary fat types on postprandial appetite and energy expenditure. Obes Res 2003;11(12):1449-1455 [FREE Full text] [doi: 10.1038/oby.2003.194] [Medline: 14694208]
29. Woerle HJ, Meyer C, Dostou JM, Gosmanov NR, Islam N, Popa E, et al. Pathways for glucose disposal after meal ingestion in humans. Am J Physiol Endocrinol Metab 2003;284(4):E716-e725 [FREE Full text] [doi: 10.1152/ajpendo.00365.2002] [Medline: 12475753]
30. Blaak EE, Antoine JM, Benton D, Björck I, Bozzetto L, Brouns F, et al. Impact of postprandial glycaemia on health and prevention of disease. Obes Rev 2012;13(10):923-984 [FREE Full text] [doi: 10.1111/j.1467-789X.2012.01011.x] [Medline: 22780564]
31. Cappon G, Vettoretti M, Sparacino G, Facchinetti A. Continuous glucose monitoring sensors for diabetes management: a review of technologies and applications. Diabetes Metab J 2019;43(4):383-397 [FREE Full text] [doi: 10.4093/dmj.2019.0121] [Medline: 31441246]
32. Bermingham KM, Mazidi M, Franks PW, Maher T, Valdes AM, Linenberg I, et al. Characterisation of fasting and postprandial NMR metabolites: insights from the ZOE PREDICT 1 study. Nutrients 2023;15(11):2638 [FREE Full text] [doi: 10.3390/nu15112638] [Medline: 37299601]
33. Newman JW, Krishnan S, Borkowski K, Adams SH, Stephensen CB, Keim NL. Assessing insulin sensitivity and postprandial triglyceridemic response phenotypes with a mixed macronutrient tolerance test. Front Nutr 2022;9:877696 [FREE Full text] [doi: 10.3389/fnut.2022.877696] [Medline: 35634390]
34. Crane MM, Jeffery RW, Sherwood NE. Exploring gender differences in a randomized trial of weight loss maintenance. Am J Mens Health 2017;11(2):369-375 [FREE Full text] [doi: 10.1177/1557988316681221] [Medline: 27923968]
35. Robertson C, Avenell A, Boachie C, Stewart F, Archibald D, Douglas F, et al. Should weight loss and maintenance programmes be designed differently for men? a systematic review of long-term randomised controlled trials presenting data for men and women: the ROMEO project. Obes Res Clin Pract 2016;10(1):70-84 [FREE Full text] [doi: 10.1016/j.orcp.2015.04.005] [Medline: 25937165]
36. Houtkooper RH, Williams RW, Auwerx J. Metabolic networks of longevity. Cell 2010;142(1):9-14 [FREE Full text] [doi: 10.1016/j.cell.2010.06.029] [Medline: 20603007]
37. Dutton K, Blanksby BA, Morton AR. CO2 sensitivity changes during the menstrual cycle. J Appl Physiol (1985) 1989;67(2):517-522. [doi: 10.1152/jappl.1989.67.2.517] [Medline: 2507498]
38. Hadziomerović D, Moeller KT, Licht P, Hein A, Veitenhansel S, Kusmitsch M, et al. The biphasic pattern of end-expiratory carbon dioxide pressure: a method for identification of the fertile phase of the menstrual cycle. Fertil Steril 2008;90(3):731-736 [FREE Full text] [doi: 10.1016/j.fertnstert.2007.06.078] [Medline: 17936278]
39. Wu X, Guo X, Zhang Z. The efficacy of mobile phone apps for lifestyle modification in diabetes: systematic review and meta-analysis. JMIR Mhealth Uhealth 2019;7(1):e12297 [FREE Full text] [doi: 10.2196/12297] [Medline: 30664494]
40. Islam MM, Poly TN, Walther BA, Li YCJ. Use of mobile phone app interventions to promote weight loss: meta-analysis. JMIR Mhealth Uhealth 2020;8(7):e17039 [FREE Full text] [doi: 10.2196/17039] [Medline: 32706724]
41. Alharbi NS, Alsubki N, Jones S, Khunti K, Munro N, de Lusignan S. Impact of information technology-based interventions for type 2 diabetes mellitus on glycemic control: a systematic review and meta-analysis. J Med Internet Res 2016;18(11):e310 [FREE Full text] [doi: 10.2196/jmir.5778] [Medline: 27888169]
42. Kumbara AB, Iyer AK, Green CR, Jepson LH, Leone K, Layne JE, et al. Impact of a combined continuous glucose monitoring-digital health solution on glucose metrics and self-management behavior for adults with type 2 diabetes: real-world, observational study. JMIR Diabetes 2023;8:e47638 [FREE Full text] [doi: 10.2196/47638] [Medline: 37590491]

43. Yost O, DeJonckheere M, Stonebraker S, Ling G, Buis L, Pop-Busui R, et al. Continuous glucose monitoring with low-carbohydrate diet coaching in adults with prediabetes: mixed methods pilot study. JMIR Diabetes 2020;5(4):e21551 [FREE Full text] [doi: 10.2196/21551] [Medline: 33325831]
44. Fundoiano-Hershcovitz Y, Bacher D, Ritholz MD, Horwitz DL, Manejwala O, Goldstein P. Blood pressure monitoring as a digital health tool for improving diabetes clinical outcomes: retrospective real-world study. J Med Internet Res 2022;24(2):e32923 [FREE Full text] [doi: 10.2196/32923] [Medline: 35133284]
45. Bian RR, Piatt GA, Sen A, Plegue MA, De Michele ML, Hafez D, et al. The effect of technology-mediated diabetes prevention interventions on weight: a meta-analysis. J Med Internet Res 2017;19(3):e76 [FREE Full text] [doi: 10.2196/jmir.4709] [Medline: 28347972]
46. Buch A, Yeshurun S, Cramer T, Baumann A, Sencelsky Y, Sagi SZ, et al. The effects of metabolism tracker device (Lumen) usage on metabolic control in adults with prediabetes: pilot clinical trial. Obes Facts 2023;16(1):53-61 [FREE Full text] [doi: 10.1159/000527227] [Medline: 36195053]
47. Li-Gao R, de Mutsert R, Rensen PCN, van Klinken JB, Prehn C, Adamski J, et al. Postprandial metabolite profiles associated with type 2 diabetes clearly stratify individuals with impaired fasting glucose. Metabolomics 2018;14(1):13 [FREE Full text] [doi: 10.1007/s11306-017-1307-7] [Medline: 29249917]

Abbreviations

BRANY: Biomedical Research Alliance of New York
CGM: continuous glucose monitoring
CO_2: carbon dioxide
CVD: cardiovascular disease
IRB: institutional board review
LMM: linear mixed model
RER: respiratory exchange ratio
SBER: Social, Behavioral, and Educational Research
T2D: type 2 diabetes
USDA: United States Department of Agriculture

Edited by L Buis; submitted 09.01.24; peer-reviewed by A Scarry, D Kirk, P Gnaganrella; comments to author 31.01.24; revised version received 21.02.24; accepted 29.02.24; published 01.04.24.

Please cite as:
Yeshurun S, Cramer T, Souroujon D, Mor M
The Association of Macronutrient Consumption and BMI to Exhaled Carbon Dioxide in Lumen Users: Retrospective Real-World Study
JMIR Mhealth Uhealth 2024;12:e56083
URL: https://mhealth.jmir.org/2024/1/e56083
doi:10.2196/56083
PMID:38439744

Original Paper

An mHealth Intervention Promoting Physical Activity and Healthy Eating in a Family Setting (SMARTFAMILY): Randomized Controlled Trial

Kathrin Wunsch[1*], PhD; Janis Fiedler[1*], PhD; Sebastian Hubenschmid[2], MSc; Harald Reiterer[2], PhD; Britta Renner[3], PhD; Alexander Woll[1], PhD

[1]Institute of Sports and Sports Science, Karlsruhe Institute of Technology, Karlsruhe, Germany

[2]Department of Computer and Information Science, University of Konstanz, Konstanz, Germany

[3]Department of Psychology, University of Konstanz, Konstanz, Germany

[*]these authors contributed equally

Corresponding Author:
Kathrin Wunsch, PhD
Institute of Sports and Sports Science
Karlsruhe Institute of Technology
Engler-Bunte-Ring 15
Karlsruhe, 76131
Germany
Phone: 49 721608 ext 45431
Email: kathrin.wunsch@kit.edu

Abstract

Background: Numerous smartphone apps are targeting physical activity (PA) and healthy eating (HE), but empirical evidence on their effectiveness for the initialization and maintenance of behavior change, especially in children and adolescents, is still limited. Social settings influence individual behavior; therefore, core settings such as the family need to be considered when designing mobile health (mHealth) apps.

Objective: The purpose of this study was to evaluate the effectiveness of a theory- and evidence-based mHealth intervention (called SMARTFAMILY [SF]) targeting PA and HE in a collective family–based setting.

Methods: A smartphone app based on behavior change theories and techniques was developed, implemented, and evaluated with a cluster randomized controlled trial in a collective family setting. Baseline (t_0) and postintervention (t_1) measurements included PA (self-reported and accelerometry) and HE measurements (self-reported fruit and vegetable intake) as primary outcomes. Secondary outcomes (self-reported) were intrinsic motivation, behavior-specific self-efficacy, and the family health climate. Between t_0 and t_1, families of the intervention group (IG) used the SF app individually and collaboratively for 3 consecutive weeks, whereas families in the control group (CG) received no treatment. Four weeks following t_1, a follow-up assessment (t_2) was completed by participants, consisting of all questionnaire items to assess the stability of the intervention effects. Multilevel analyses were implemented in R (R Foundation for Statistical Computing) to acknowledge the hierarchical structure of persons (level 1) clustered in families (level 2).

Results: Overall, 48 families (CG: n=22, 46%, with 68 participants and IG: n=26, 54%, with 88 participants) were recruited for the study. Two families (CG: n=1, 2%, with 4 participants and IG: n=1, 2%, with 4 participants) chose to drop out of the study owing to personal reasons before t_0. Overall, no evidence for meaningful and statistically significant increases in PA and HE levels of the intervention were observed in our physically active study participants (all P>.30).

Conclusions: Despite incorporating behavior change techniques rooted in family life and psychological theories, the SF intervention did not yield significant increases in PA and HE levels among the participants. The results of the study were mainly limited by the physically active participants and the large age range of children and adolescents. Enhancing intervention effectiveness may involve incorporating health literacy, just-in-time adaptive interventions, and more advanced features in future app development. Further research is needed to better understand intervention engagement and tailor mHealth interventions to individuals for enhanced effectiveness in primary prevention efforts.

Trial Registration: German Clinical Trials Register DRKS00010415; https://drks.de/search/en/trial/DRKS00010415

International Registered Report Identifier (IRRID): RR2-10.2196/20534

(JMIR Mhealth Uhealth 2024;12:e51201) doi:10.2196/51201

KEYWORDS

mobile app; telemedicine; behavior change; health behavior; family; primary prevention; exercise; diet; food and nutrition; randomized controlled trial; accelerometer; wearable electronic devices; social-cognitive determinants; just-in-time adaptive intervention; digital intervention; mobile phone

Introduction

Background

Physical activity (PA) and healthy eating (HE) are protective factors for general health and can also enhance health [1]. In contrast, a lack of PA, too much sedentary behavior (eg, excessive screen time and nonactive media use), and an unhealthy diet are serious concerns and increase the risk of health conditions across all ages [2-5]. Although preschool children seem to show high adherence to PA guidelines [6], they have low adherence to screen time guidelines [7]. Research revealed that children, as they become older, and adolescents do not engage sufficiently in PA [8] and frequently make unhealthy food choices [9,10]. Worldwide, 81% of children and adolescents (and 23% of adults) do not meet recommendations on PA levels and HE, that is, fruit and vegetable intake (FVI) [11]. In this regard, a dose-response relationship (ie, the relationship between the dose of PA and the effect observed) was detected, with even slight increases in PA leading to physiological and psychological health benefits in children and adolescents [12,13] and in adults. Hence, health behavior interventions should aim at sustaining PA and HE levels for those adhering to guidelines or at increasing these behaviors in those who do not adhere to guidelines [7,11]. As longitudinal studies showed that behavioral patterns in childhood and adolescence have a low-to-moderate influence on PA patterns in adulthood [14-19], there is a need for interventions already targeting children and adolescents to promote a sustainable and healthier lifestyle.

To achieve this goal, it is important to recognize that health-related behaviors, such as PA and HE, are influenced by social contexts, such as the family environment, and are shaped by social relationships and connections [20,21]. Leisure time PA of children is, for example, directly linked to parental PA levels [22,23], and the eating behavior of children is dependent on their parents' food choices [24-26]. Therefore, addressing behavioral changes embedded in social contexts might be a promising approach for facilitating an individual's behavior change. When focusing on children and adolescents, the most imprinting social context is daily family life. Family meals, for example, are often an important part of everyday life in families, and there is accumulating evidence that this collective behavior is associated with a better overall diet quality and BMI [27-29]. In a similar vein, there is some evidence that family-based PA is positively associated with individual PA levels [30]. It has been shown that supportive interactions within a family and shared values about health behavior affect PA engagement [31] and the eating behavior of children [32]. In addition, results of intervention studies indicate that social support is significantly associated with the continuation of exercise programs [33-37] as well as participation in weight loss interventions [38-40].

Families are difficult to reach with typical in-person interventions, as the daily lives of family members are highly different regarding their content and time schedule. Hence, mobile health (mHealth) technologies might be the means of choice, as they are increasingly used as a delivery mode for health behavior change interventions across different age groups. Specifically, smartphone-based apps offer great promise for enhancing PA and HE as well as making health care more accessible and scalable, more cost-effective, and more equitable [41,42]. Reviews and meta-analyses support the view that app‐based mobile interventions are effective and highly promising for changing PA [36,43] and nutrition behaviors [44], especially when implemented in a family setting, including parental involvement [45]. Moreover, a systematic review of economic evaluations of mHealth solutions found a consistent overall reporting of positive economic outcomes (eg, increase in life-years gained, cost savings, and cost-effectiveness) [46].

Regarding functional principles of such mHealth apps, reviews indicate that the strategies or central *building blocks* of app-based interventions mainly encompassed 4 behavior change technique (BCT) clusters [47], including goal setting, feedback and self-monitoring, information, and social support provision, which coincides with successful conventional individual and group-based interventions [44,48,49]. On the basis of causal analyses, an umbrella review found that the effectiveness of interventions was increased by engaging social support, targeting both PA and HE, and using well-defined or established BCTs. However, because mobile interventions distinguish themselves by being interactive, adaptive, time sensitive, and intraindividually dynamic, more dynamic concepts, including the timing of feedback and tailoring tasks and goals to individual progress and capacities as specified in persuasive technology and gamification approaches, might be essential ingredients of effective and focused mobile interventions [44,50,51].

Moreover, mobile interventions are capable of fulfilling the abovementioned demands to be embedded into a social system so that all members of this system, that is, a family, can simultaneously and collectively take part in an intervention and share their goals and progress. However, currently available apps for health promotion are almost exclusively tailored to an individual level [52]. Motivation for behavior performance is higher when the individual is embedded in a social system of mutual appreciation and importance, according to self-determination theory (SDT) [53,54], which was successfully used in PA interventions by enhancing autonomous motivation and fulfilling the 3 basic psychological needs: "autonomy," "competence," and "relatedness" [53]. As healthy or unhealthy

behavioral patterns are developed and maintained in social contexts, embedding a mHealth intervention in a family setting and targeting all family members might be promising and corresponds to assumptions of family-as-systems approaches [55]. An umbrella review about digital interventions for health behavior change in PA and HE found a lack of studies focusing on the social contexts, while BCTs and a theoretical foundation were associated with higher intervention effectiveness [56]. Another review of the effect of family-based mHealth interventions found heterogeneous results in a limited research body and called for more experimental studies in this area [57]. Overall, there is a lack of randomized controlled mHealth studies that use important key facets of effective interventions such as BCTs and a theoretical foundation while considering the social system of participants.

Objectives

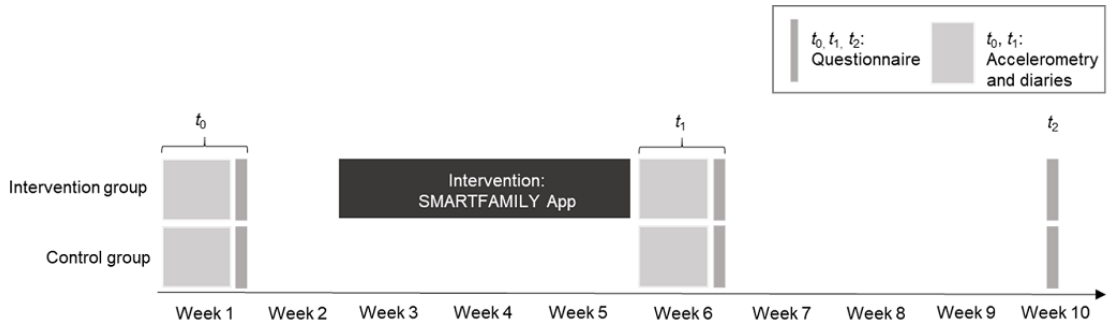
This study aimed to evaluate the effectiveness of a mobile stand-alone smartphone intervention app that incorporates the family as a social system of high relevance for its single members. We designed and evaluated a multicomponent mHealth intervention that aimed to improve PA and HE in families. The development of the app was based on behavior change theories, including SDT, and the use of BCTs. The behavior of children and parents was targeted to induce family- and individual-based behavior changes. Specifically, family members were using the SMARTFAMILY (SF) app individually and cooperatively. The complete study protocol and more details are available in the study by Wunsch et al [58]. It was hypothesized that the mHealth intervention would positively influence PA variables (steps and moderate to vigorous PA [MVPA]) and the HE variable (FVI) in the whole family.

Methods

Study Design

The study was conducted and described according to the corresponding study protocol and the CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) checklist that provides guidance for standardized reporting of eHealth and mHealth interventions (Multimedia Appendix 1) [36,41-44,46,53,54,59-66]. This is a cluster randomized controlled trial with two groups: (1) an intervention group (IG) receiving the app intervention and (2) a nonintervention control group (CG) that neither received material nor was contacted during the intervention period. The outline of the SF trial is presented in Figure 1. Assessment of outcomes was completed at baseline (t_0), after the 3-week intervention (or no-intervention) period (postintervention; t_1), and 4 weeks after the measurement (follow-up; t_2). As the study protocol is freely accessible [58], the study design and measurements will only be described very briefly. For a complete overview, please refer to the study protocol.

Figure 1. Detailed study design.



All eligible members of each family participated in an initial visit to the research facility, where they received instructions on how to use various tools that would be used throughout the study. These included wearable devices called accelerometers, which recorded PA levels, as well as paper diaries for tracking their behavior over time. At the end of this initial phase, participants also answered questions related to their habits and behaviors in the past week, which served as a baseline. This information was shared with those participating in the intervention so they could set family goals for themselves based on this starting point.

Participants assigned to the IG were given specially designed smartphones equipped with our SF app. The app was created in an iterative process, incorporating feedback from both the target audience and experts. The development process also drew upon insights from previous research conducted as part of the SMARTACT project, as well as behavioral theories. The programing of the apps was carried out by the Human-Computer Interaction Workgroup at the University of Konstanz, as a component of the SMARTACT project. Study staff ensured that all participants had access to similar technology by providing them with these phones. Providing additional study smartphones was also recommended by the ethics committee for data security reasons. The study staff also carefully explained the app's functionality to families, handed out a study manual, and were available to assist with any issues that arose. In addition, the accelerometers worn by participants were synced wirelessly with the smartphones through Bluetooth low-energy connections.

When the 3-week intervention began, families in the IG were instructed to establish collective weekly objectives for both PA and HE. The 3-week period was chosen due to practical concerns, as we examined families in their natural setting. In Germany, a continual school period lasts a maximum of 6 to 8

weeks, followed by a vacation period. To conduct the core assessments, including pre- and posttesting accelerometry during 1 continual school period, an intervention period >3 weeks was not feasible for the study design. Longer intervention periods would inevitably mean that there is a cofounding between the assessment period (school time vs vacation). Specifically, they aimed to accumulate a certain number of steps, engage in appropriate amounts of moderate or vigorous PA, plan enjoyable family activities together, and consume adequate portions of fruits and vegetables as well as joint family meals. Families were instructed to formulate the goals together based on their previous performance from the baseline assessments. The goals were set on 1 smartphone for the whole family, and the app notified the family every Sunday to set new goals. Participants were encouraged in the initial explanation phase by the study staff to strive for progressively more ambitious but realistic goals based on their previous behavior. As part of the program, participants revisited and adjusted their goals regularly to keep challenges manageable while promoting continuous improvement. All smartphones were retrieved after the intervention period ended. Therefore, the IG had neither access to the app during t_1 nor afterward.

The app included 10 BCTs [47] for the IG and no BCT for the CG. Intervention BCTs were behavioral goal setting, prompt review of behavioral goals, prompt self-monitoring of behavior, providing feedback on performance, planning social support or social change, prompt identification as a role model or position advocate, setting graded tasks, shaping, prompt rewards contingent on effort or progress toward behavior, and providing rewards contingent on successful behavior.

Once the intervention or control period ended, participants underwent 2 more testing sessions (t_1 and t_2). Our investigation used a single-blind design, and the survey instruments were evaluated for user-friendliness and technical reliability beforehand.

Eligibility Criteria

Households comprising at least 1 adult caregiver and 1 child aged >10 years residing together could participate in the research. If applicable, other siblings were welcome to join the project as well. Everyone taking part had to possess basic proficiency in handling mobile devices, be physically able to participate in PA independently, and communicate effectively in German. However, exceptions might have been made for younger siblings who met this requirement.

Ethical Considerations

Before commencement, participants (including minors), parents, or legal guardians signed informed consent forms. Full ethics approval was granted by the University of Konstanz and the Karlsruhe Institute of Technology. This research adhered to the guidelines outlined in the Helsinki Declaration. Furthermore, during the testing sessions, ethical standards were upheld, and personal data were protected.

Families who completed the study were eligible to participate in a prize draw at the end of the study, having the chance to win one of three family-tickets for a big theme park in Germany.

Randomization and Blinding

This study used a cluster randomized controlled design comparing 2 distinct groups: an IG that received the intervention and a CG that did not receive any treatment. Consenting families were randomly assigned to either arm through a straightforward allocation scheme suitable for cluster trials [60]. While members of the IG were aware of the mHealth aspects of the study, participants in the CG were merely informed about contributing to an epidemiological examination of PA and overall health. Accordingly, all participants wore accelerometers for 1 week twice across a 10-week duration to ensure accurate and dependable measurements. In addition, participants completed various questionnaires during this period.

Participants

Participants were recruited in schools, school holiday programs, music schools, and sports clubs via personal communication, newspapers, social media, and email distribution lists of the Karlsruhe Institute of Technology.

Overall, 48 families (CG: n=22, 46%, with 68 participants and IG: n=26, 54%, with 88 participants) were recruited for the study. Two families (CG: n=1, 2%, with 4 participants and IG: n=1, 2%, with 4 participants) chose to drop out of the study owing to personal reasons before t_0.

Measurements

PA Measures

Device-Measured PA (Accelerometry)

Hip-worn (right side) 3-axial accelerometers (Move 3 or Move 4, Movisens GmbH) were used to continuously record PA levels. These accelerometers are scientific research instruments with a measurement range of +16 g to −16 g, an output rate of 64 Hz, physical dimensions of 62.3×38.6×11.5 mm, a weight of 25 g, and custom epoch lengths (ie, 10 s). Data were recorded in a raw format (64 Hz) and were summarized afterward in the epoch lengths of choice. Epoch length was chosen as 10 seconds as shorter epoch lengths are believed to be more appropriate to estimate vigorous PA and assess PA in children owing to intermittent movement behavior [67,68]. Validity has been evaluated for a previous version of the accelerometer (Move 2), which uses comparable digital signal processing as the Move 3 or Move 4 has been considered accurate for assessing steps [69] and energy estimation [70,71] in adults. Handling the accelerometer was explained and demonstrated by a study instructor. Participants were instructed to wear the accelerometer during wake time and remove it only during a shower, swimming, or certain sports involving bodily contact to minimize the probability of injuries. The outcomes for the accelerometer that were used for this study were MVPA (>3.0 metabolic equivalents [MET]) and steps for all participants. MET values were calculated based on activity class (based on acceleration and barometric signals), which determines the estimation model. Afterward, movement acceleration, altitude change, and demographics were combined in the model for the MET estimation [71]. Accelerometer data were included if a minimum wear time of at least 8 hours a day for at least 4 of the 7 days during the measured week was obtained. Nonwear

time was calculated in 30-second intervals. The nonwear time detection was based on an algorithm that used accelerometry and temperature signals over a 10-minute window to distinguish between wear time, nonwear time, and sleep, as described elsewhere [72]. For valid measurements, the average of MVPA and steps per valid day was multiplied by 7 to estimate the total minutes per week.

Self-Reported PA

At the end of each measurement week, adults were asked to fill short version of the International Physical Activity Questionnaire in German [73], asking retrospectively for activities during the previous week. The results of the question relating to minutes spent in moderate (comprising moderate activity and walking) and vigorous PA were calculated for this study by multiplying the reported number of days with the reported duration of the indicated activity per day and then adding the values of moderate and vigorous PA to MVPA per week. Children completed the 60-Minute Screening Measure [74] for MVPA, which yields the number of days with at least 60 minutes of MVPA according to the (now outdated) World Health Organization guidelines [75]. In addition, a PA diary was used for all participants at t_0 and t_1, which is not included in the current examination owing to noncomparable results [68]. Parents and children were instructed to fill out their self-reported questionnaires and diaries independently.

FVI Measures

FVI was assessed using a single item asking for the total amount of fruits and vegetables consumed within a typical week [76] in the questionnaire as well as using a description of detailed food consumption during t_0 and t_1 by indicating the time of the meal, ingredients, portions of FVI, and whether the meal was consumed within the family or alone in a diary.

Secondary Outcomes

Demographics

In the t_0 questionnaire, demographic information of the participants was collected, including sex, age, height, and weight.

Health Status

Perceived general health was assessed using a single item [76].

Intrinsic Motivation Toward PA

According to the SDT, activity-related self-determination was assessed using the Behavioral Regulation in Exercise Questionnaire (BREQ-2) [77]. BREQ-2 assesses the manifestation of the 5 different regulation modes by the SDT, reflected by the subscales of amotivation (4 items), external (4 items), introjected (3 items), identified (3 items), and intrinsic (4 items) regulations. Responses were made on a 4-point Likert scale, ranging from 0=*not true*, 1=*rather not true*, 2=*rather true* to 3=*true*. For the analysis, a relative autonomy index was formed. It is derived from the subscales and gives an index of the degree to which respondents feel self-determined. The index is obtained by applying a weighting to each subscale and then summing these weighted scores. In other words, each subscale score is multiplied by its weighting, and then these weighted scores are summed. Higher, positive scores indicate greater relative autonomy; lower, negative scores indicate more controlled regulation.

Intrinsic Motivation Toward HE

For assessing dietary-related intrinsic motivation, the Regulation of Eating Behavior Scale [78] was used. The dimension "integrated regulation" was omitted, resulting in a total of 5 subscales, coded from 0 to 3. A sum score was built analogous to the BREQ-2.

Self-Efficacy for PA and HE

Activity-related self-efficacy and dietary-related self-efficacy were assessed using the health-specific self-efficacy scales, comprising 5 items for each behavior-related dimension [79]. Participants were asked how certain they were to handle different health-specific barriers. Responses were given on a 4-point Likert scale, ranging from 1=very uncertain, 2=uncertain, 3=certain to 4=very certain. A sum score was built for both scales.

Family Health Climate

Shared perceptions and cognitions regarding health behaviors were assessed by use of the Family Health Climate (FHC) Scale [80], comprising the 2 separate scales for PA (FHC-PA) and nutrition (FHC Nutrition [FHC-NU]). FHC-PA contains 14 items, which are assigned to the 3 subscales of value (5 items), cohesion (5 items), and information (4 items). FHC-NU includes 17 items, comprising the 4 subscales of value (4 items), cohesion (5 items), consensus (3 items), and communication (5 items). Responses for each dimension were scored on a 4-point Likert scale ranging from 0=*not true*, 1=*rather not true*, 2=*rather true* to 3=*true*. Sum scores were built for both scales.

Joint PA and Meals Within the Family

Joint PA and nutrition were assessed using a single item that referred to the number of activities and meals in which the whole family was involved during the past week. The mean value per family was used for the analysis.

Statistical Analysis

The analyses were run with different packages of R (R Foundation for Statistical Computing) [81] and RStudio (Posit) [82]. The package *ggplot2* was used for visualizations [83] following the instructions by Allan et al [84]. Mixed models were calculated using the package *lmerTest* [85] with participants (level 1) nested in families (level 2) to acknowledge the hierarchical structure of the data. The result tables of the regression analyses were generated using the package *sjPlot* [86]. In total, 6 final models were calculated, 1 with each measurement method and outcome parameter (1 steps, 3 MVPA, and 2 FVI intake per wk) as dependent variables. Assumptions were checked using the visualization of the *performance* package [87]. A hierarchical approach was used for the inclusion of the control variables, and the model fit was assessed with the Akaike information criterion for sensitivity analysis. In addition, 2 models were calculated for joint PA and nutrition based on the family mean values without the random factor family because it was defined as "all family members were present."

The predictor group (ie, control=0 and intervention=1)×time (dummy coded with t_0 as reference for t_1 and t_2) was included in the models to evaluate the interaction effect (main effect) of the intervention on the 8 outcome parameters. To assess sensitivity regarding the additional variables, the secondary outcome parameters self-efficacy, intrinsic motivation, and the FHC were added either referring to PA or FVI depending on the outcome, and the control variables health status, population (adult=0 and children=1), sex (0=male and 1=female), and nonwear time per week—only for the device-measured PA models—were tested for the inclusion in the random effect models. In addition, the inclusion of random slopes and random intercepts were evaluated based on the model fit. The level of statistical significance was set a priori to $\alpha<.05$ with no correction for multiple comparisons.

Results

Data Availability and Participant Characteristics

In total, 46 families with 148 participants (74/148, 50% adults: 45/74, 61% female and 29/74, 59% male; and 74/148, 50% children: 38/74, 51% female and 36/74, 49% male) participated in the study. The mean ages of adults and children were 47.8 (SD 5.0) and 13.3 (SD 2.7) years, while the average BMI values were 24.8 (SD 4.1) and 19.0 (SD 3.3), respectively. Testing took place between December 1, 2017, and January 31, 2020. Technical issues with the app during the intervention, insufficient wear time of the accelerometer (ie, <4 days with >8 hours wear time), and missing data for the self-reported items led to the inclusion of a differing number of participants for each calculated model (depending on the outcome variables; Multimedia Appendices 2 and 3). Participant characteristics of the 46 families separated by group (CG vs IG), population (children vs adults), and sex (male vs female) are displayed in Table 1. Descriptive results for all included outcomes and predictors can be found in Table 2.

Table 1. Participant characteristics of the 46 families included in the SMARTFAMILY trial.

Characteristics	Control group		Intervention group	
	Children (n=32)	Adults (n=32)	Children (n=42)	Adults (n=42)
Sex, n (%)				
Male	15 (47)	11 (34)	21 (50)	18 (43)
Female	17 (53)	21 (66)	21 (50)	24 (57)
Age (y), mean (SD)				
Male	13.1 (3.23)	49.1 (5.02)	13.4 (2.48)	50.2 (5.58)
Female	13.4 (2.61)	47.0 (3.90)	13.5 (2.58)	46.1 (4.70)
BMI (kg/m^2), mean (SD)				
Male	18.7 (2.59)	25.7 (3.30)	19.0 (4.06)	26.6 (5.35)
Female	19.2 (3.14)	24.7 (3.63)	18.9 (3.00)	23.3 (3.01)

Table 2. Descriptive results of the 46 families included in the SMARTFAMILY trial. Data were assessed for the three measurement weeks: baseline (t_0), after intervention (t_1), and 4 weeks after intervention at follow-up (t_2).

Measurements	Control group, mean (SD)						Intervention group, mean (SD)					
	t_0		t_1		t_2		t_0		t_1		t_2	
	Child	Adult	Child	Adult	Child	Adult	Child	Adult	Child	Adult	Child	Adult
Steps (counts/wk)	62,800 (21,800)	60,000 (24,000)	58,900 (23,600)	59,500 (27,700)	—[a]	—	66,200 (22,400)	66,700 (24,100)	64,000 (24,600)	66,500 (32,800)	—	—
MVPA[b] (min/wk)	721 (303)	674 (282)	652 (303)	675 (338)	—	—	711 (307)	745 (237)	609 (294)	716 (316)	—	—
Nonwear time (min/wk)	4660 (789)	4010 (389)	4910 (677)	4140 (537)	—	—	4720 (538)	4120 (618)	4540 (596)	4190 (731)	—	—
IPAQ[c] MVPA (min/wk)	X[d]	975 (916)	X	826 (1010)	X	772 (737)	X	1060 (1030)	X	1080 (839)	X	999 (678)
KIKA[e] MVPA (active d/wk)	4.46 (1.67)	X	3.92 (1.52)	X	4.25 (1.68)	X	4.19 (1.86)	X	4.00 (1.76)	X	3.57 (1.87)	X
FVI[f] quest (portions/wk)	10.2 (8.94)	12.3 (6.54)	9.77 (8.01)	12.5 (8.13)	8.84 (6.82)	13.5 (7.88)	14.8 (11.0)	16.4 (9.77)	14.4 (9.29)	18.2 (11.3)	23.7 (55.9)	17.3 (10.9)
FVI diary (portions/wk)	12.2 (5.73)	14.3 (8.14)	11.3 (8.00)	15.0 (9.77)	—	—	17.1 (12.0)	16.2 (8.80)	17.3 (11.2)	20.3 (11.6)	—	—
Health	4.09 (0.86)	3.74 (0.73)	4.07 (0.77)	3.61 (0.74)	3.96 (0.82)	3.74 (0.76)	4.21 (0.58)	3.95 (0.70)	4.14 (0.65)	3.94 (0.64)	4.25 (0.57)	3.85 (0.62)
M intrinsic NU[g] (RAI)[h]	24.9 (16.1)	23.6 (17.7)	21.5 (14.3)	25.9 (14.9)	21.1 (14.9)	24.0 (18.2)	22.8 (16.4)	29.8 (12.6)	27.0 (17.6)	32.2 (14.3)	25.9 (16.3)	31.7 (13.8)
M intrinsic PA[i] (RAI)	38.1 (17.7)	36.4 (16.9)	35.8 (17.2)	33.6 (15.6)	36.1 (14.4)	32.5 (16.7)	39.2 (13.7)	39.0 (15.5)	40.0 (14.5)	39.1 (12.4)	38.8 (15.2)	39.3 (13.7)
FHC[j] NU[k]	30.3 (6.85)	32.1 (8.83)	27.9 (7.82)	29.1 (7.48)	28.4 (7.77)	30.0 (8.82)	30.9 (8.64)	34.9 (6.81)	31.6 (8.38)	35.1 (7.72)	31.7 (8.52)	36.0 (6.87)
FHC PA[l]	20.9 (7.15)	22.2 (7.18)	20.2 (5.83)	19.5 (5.77)	21.6 (6.88)	19.5 (7.32)	22.8 (7.63)	25.5 (7.84)	22.3 (7.82)	23.4 (8.31)	23.2 (6.55)	24.2 (7.85)
Self-efficacy NU (RAI)	49.0 (10.1)	42.9 (11.7)	48.1 (7.86)	49.1 (9.60)	49.2 (10.3)	46.2 (10.8)	47.7 (12.5)	49.8 (11.4)	47.2 (14.8)	49.9 (10.2)	50.2 (11.5)	50.1 (10.3)
Self-efficacy PA (RAI)	50.9 (9.32)	44.8 (8.39)	50.6 (9.03)	43.7 (9.51)	51.5 (10.6)	45.4 (9.15)	49.4 (9.38)	47.8 (11.9)	49.7 (9.08)	49.6 (12.1)	49.9 (10.7)	51.3 (10.9)
Joint PA	0.586 (0.907)	0.429 (0.634)	0.556 (0.641)	0.815 (1.14)	0.690 (0.930)	0.520 (0.586)	0.946 (1.31)	0.917 (1.42)	0.972 (0.910)	0.719 (0.772)	1.00 (1.00)	0.935 (0.929)
Joint NU	8.97 (5.28)	7.55 (4.20)	7.67 (4.10)	7.39 (3.57)	7.14 (4.03)	8.39 (4.46)	7.97 (4.30)	8.45 (4.08)	6.95 (3.69)	7.21 (3.86)	7.94 (4.56)	8.03 (4.76)

[a]Not measured at t_3.

[b]MVPA: moderate to vigorous physical activity.

[c]IPAQ: International Physical Activity questionnaire.

[d]The International Physical Activity questionnaire was used for adults, and the 60-Minute Screening Measure was used for children.

[e]KIKA: 60-Minute Screening Measure.

[f]FVI: fruit and vegetable intake.

[g]M intrinsic NU: intrinsic motivation toward nutrition.

[h]RAI: relative autonomy index.

[i]M intrinsic PA: intrinsic motivation toward physical activity.

[j]FHC: Family Health Climate.

[k]NU: nutrition.

[l]PA: physical activity.

All control variables (except sex for the PA questionnaire of children) improved the model fit based on the Akaike information criterion and were therefore included in the final sensitivity models. Random slopes were not supported by the data, but random intercepts were used for all models. Sensitivity analysis showed no difference in patterns for the effectiveness of the intervention (Tables S1-S6 in Multimedia Appendix 3). Therefore, only the main models are reported. Data and code are available at the Open Science Framework [88].

Effect of the Intervention on PA

Results of the multilevel models indicate no significant main effect for the interaction of the group with time in the PA outcomes (Tables S1-S4 in Multimedia Appendix 2). The only significant main effect was found between t_0 and t_1 for the self-reported 60-Minute Screening Measure in children (P=.03; β=−.35). Figures 2 and 3 display the descriptive results of the main effects for device-measured PA outcomes, MVPA, and step count. As displayed by the gray dotted lines in Figure 2, both mean and median values are clearly above the recommendation for MVPA [89] for both children (ie, 60 min/d on average, 420 min/wk) and adults (ie, 150 min/wk) in both groups and at both measurement periods. For steps, mean and medium values are shortly below the commonly used 10,000 steps per day goal [90,91] for all participants (Figure 3).

Figure 2. Interaction effect of group×time for device-measured physical activity (PA) for the parameter minutes of moderate to vigorous PA (MVPA) per week. Displayed are the mean MVPA (y-axis) of 109 participants during 1 week of baseline measurement (t_0) and 1 week of postintervention measurement after a 3-week intervention and waiting period (t_1) for the control group (green) and the intervention group (red), stratified by children and adults. The gray dashed lines represent the PA recommendations for children (420 min/wk) and adults (150 min/wk).

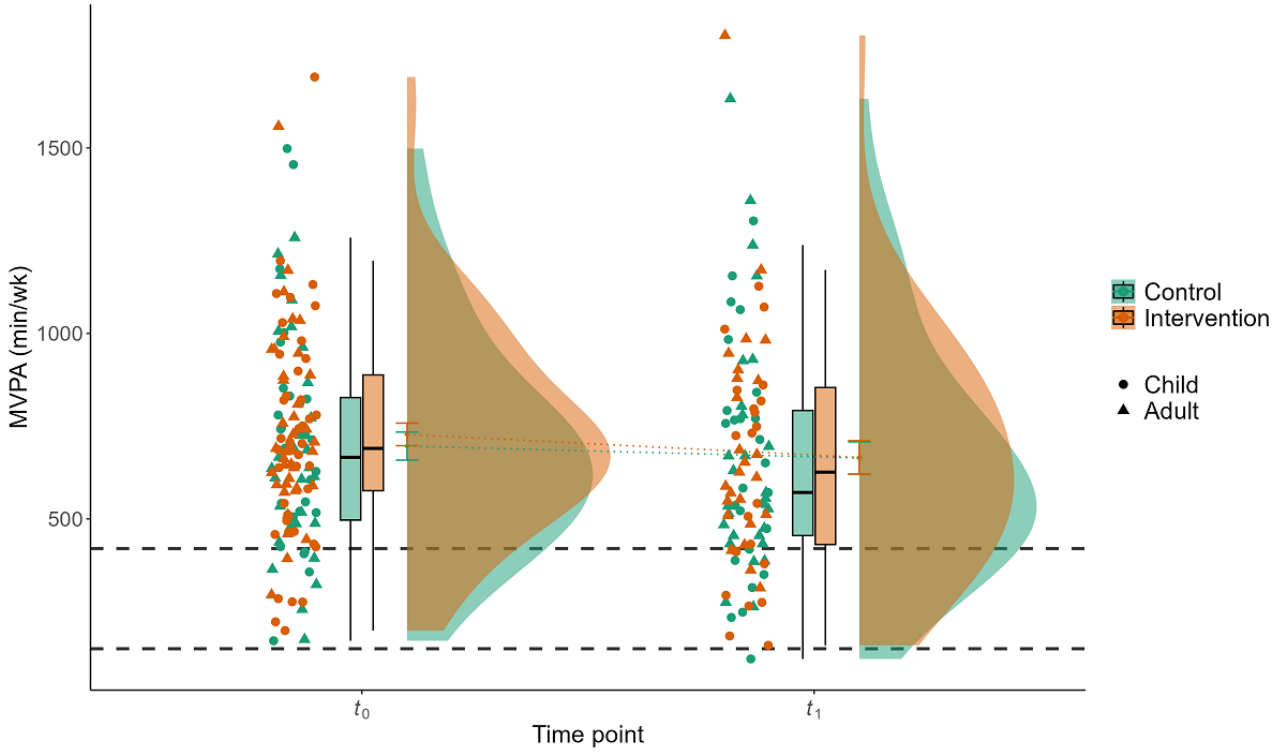
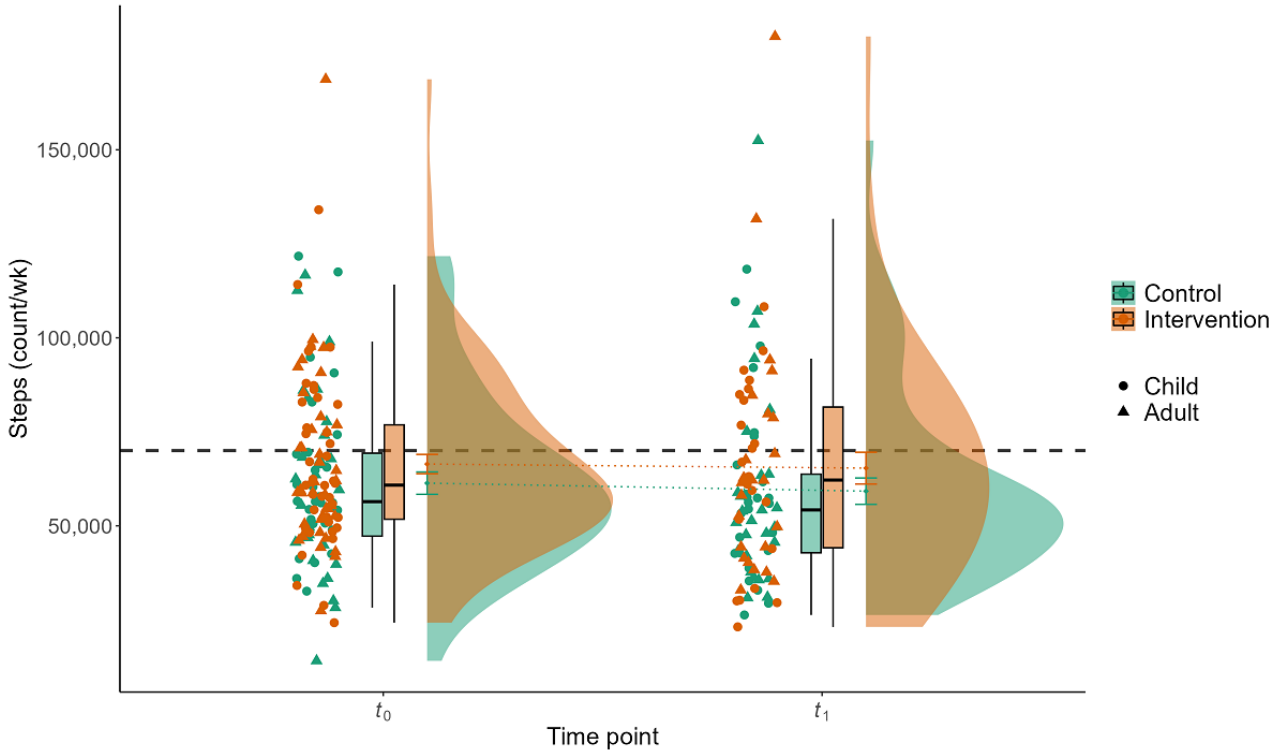


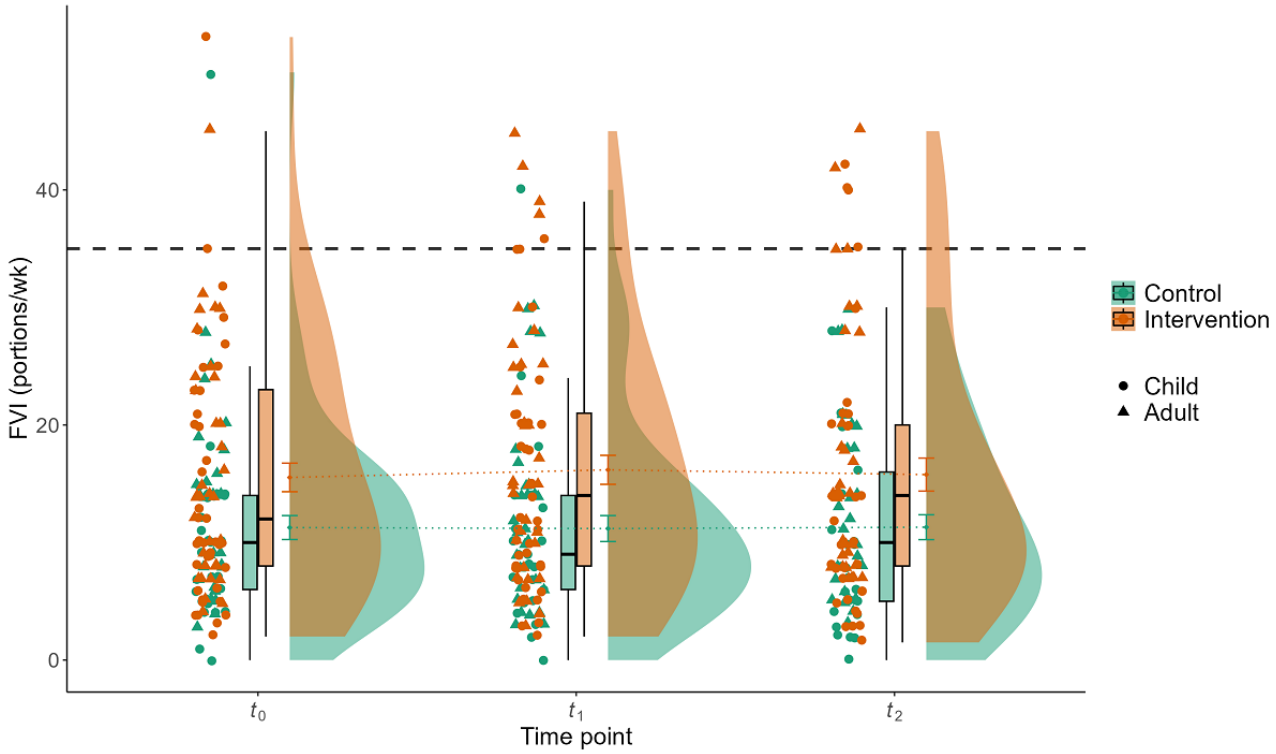
Figure 3. Interaction effect of group×time for device-measured physical activity for the parameter steps per week (steps). Displayed are the mean step count (y-axis) of 109 participants during 1 week of baseline measurement (t_0) and 1 week of postintervention measurement after a 3-week intervention and waiting period (t_1) for the control group (green) and the intervention group (red), stratified by children and adults. The gray dashed line represents the commonly used step recommendation of 10,000 steps per day (70,000 steps/wk).



Effect of the Intervention on FVI

Results of the multilevel models indicate no significant interaction of group×time concerning FVI (Tables S5 and S6 in Multimedia Appendix 2). The only observed significant main effect was a group effect for FVI from the questionnaire (P=.049; β=.23). Figure 4 displays the descriptive results for self-reported FVI per week assessed by the questionnaire. Here, both mean and median values are clearly below the recommended FVI of 35 portions per week [92].

Figure 4. Interaction effect of group×time for the parameter fruit and vegetable intake (FVI) per week assessed by questionnaire. Displayed are the mean FVI (y-axis) of 118 participants related to the week of baseline measurement (t_0), the week of postintervention measurement after a 3-week intervention and waiting period (t_1), and the week of follow-up measurement (t_2) for the control group (green) and the intervention group (red), stratified by children and adults. The gray dashed line represents the recommendation for daily FVI of 5 portions (35 portions/wk).



Effect of the Intervention on Joint PA and Meals

Results of the linear model indicate no significant effect for group×time but a significant main effect for group (P<.001; β=.25) in joint PA. For joint meals, a significant effect for group× time (t_2) was observed (P=.04; β=.16). In addition, a significant reduction of joint meals between t_0 and t_1 was observed (P=.001; β=−.31). All results are displayed in Tables S7 and S8 in Multimedia Appendix 2.

Discussion

Principal Findings

The SF trial evaluated the effectiveness of a mHealth intervention to increase PA and HE in families. Extending previous research, the behavior of children and parents was targeted to induce individual behavior changes that are anchored in daily family life. Moreover, besides a theoretical foundation, several BCTs were included, which contribute to the fulfillment of basic psychological needs according to the SDT [53,54]. Overall, no evidence for meaningful and statistically significant increases in PA and HE levels of the intervention was observed in our physically active sample. Guidelines for weekly MVPA were highly over exceeded in this study sample. However, PA levels were sustained and did not decrease, independent of group affiliation. Intervention effectiveness will be re-evaluated in the SMART*FAMILY2.0* trial, which includes gamification features, health literacy, a just-in-time adaptive intervention, and more sophisticated app features.

PA and HE

Neither hypotheses, that PA or HE would increase as a result of the intervention, were supported by any self-reported or device-based measure. This is in line with previous digital health studies that also found heterogeneous results, with a majority of studies, however, revealing at least some significant benefit of interventions [44,56]. However, it needs to be noted that the current sample was exceedingly high in their amount of PA, which might have impacted the current null findings. Regarding PA, participants exceeded the recommended amount of 420 minutes of MVPA per week for children and adolescents by 230 minutes on average but fell short of the recommended amount of 35 portions of fruits and vegetables by 21 portions. Although the current sample can be classified as being highly active in a national and international comparison [93], they cannot be classified as equally healthy regarding their eating behavior. In a German-wide representative sample, 17% of 6- to 11-year-olds and 20% of 12- to 17-year-olds reached the recommendation on fruit consumption, 7% and 23%, respectively, for vegetable consumption [94]. Comparable studies are rare. A systematic review and meta-analysis of the efficacy of mHealth interventions to improve PA only focused on inactive participants, as the largest effects are to be expected in this group [95]. In the review [95], only 2 studies were found that implemented a stand-alone mHealth intervention. However, the results were in favor of the interventions for inactive participants. Another review found only 50% of mHealth interventions to be effective [96].

Another impacting factor in our investigation might have been that the intervention duration was rather short; however, it was

determined by the short intervals between school holidays, which were omitted within the measurement periods. Hence, the intervention was only carried out within school periods to depict "normal" days. The structured-day hypothesis postulates that health behaviors differ between structured days and school vacation times or weekends. A meta-analysis has shown that the largest effects can be expected for interventions that last over several weeks [97], which might have had a positive influence on intervention effectiveness. However, mHealth intervention studies even revealed significant behavior change effects with intervention durations of only 1 [98], 2 [99], and 3 weeks [100]. These results collectively suggest that both health behaviors should be considered separately but can be addressed together in interventions to harness their synergistic impact on health, a notion supported by other studies as well [101,102]. Although not everyone might need an intervention in both factors, few fulfill both PA and FVI guidelines.

Family-Based Interventions

Studies focusing on mHealth family-based interventions are rare, especially those including randomized controlled trials. Results for parent- and child-based digital interventions are heterogeneous regarding their reported effectivity, often combined with nondigital interventions, and focused on the behavior of the children instead of the behavior of the whole family [103,104]. A recent review pointed to the effectiveness of digital interventions for obesity prevention, including the enhancement of HE and PA in children, but only 2 of the included studies focused on mHealth interventions [105]. The first study that focused on children with obesity used an app for self-monitoring of weight and goal setting, which led to a greater reduction of weight than the standard care group after 6 months [106]. The other study aimed to improve fundamental movement skills in 3- to 6-year-old children and found an improvement in these skills after a 2-month intervention period compared to a CG [107]. Of the 7 studies included in another review [57], 2 pilot studies reported significant improvements in PA; 3 studies found evidence for some improvements in PA measures, for example, collaborative PA of children and parents; and the remaining 2 studies found no evidence for an effect. Interestingly, 1 study pointed out that the adolescent dropout rate was 12.2 times higher if their parents stopped using the app.

In this regard, analyzing family behavior and dyadic relationships will be a promising approach for future investigations. Studies suggest that family meal practices and values can support HE [108] and that the frequency of shared family meals is significantly related to nutritional health in children and adolescents [27-29]. The results of this study do not support the assumption that joint PA or joint meals (ie, some type of "quality time" within the family) impacts PA or HE behavior. However, it needs to be acknowledged that joint PA was rather low in our sample (only about 1 joint activity per wk).

In summary, the results of this study point to an urgent need to better understand intervention engagement, both quantitatively and qualitatively, to design effective interventions [109,110]. If researchers have a better idea about what works when and for whom, then mHealth interventions can be tailored more specifically and facilitate the benefit of individualization, which was found to be related to enhanced effectiveness [51]. This knowledge can be used for just-in-time adaptive interventions, which aim to support in the exact moment when support is needed and behavior change is realistic [50,111]. Overall, research on mHealth interventions for families remains limited, particularly in the realm of primary prevention. However, theoretical frameworks emphasize the significance and potential of such interventions for promoting PA and HE [56,105]. An important consideration for designing future evaluations of interventions is to additionally account for methods that go beyond pre-post follow-up designs to account for the timelines and complexity of mHealth interventions on a minute-to-minute basis. This would allow important insights for the analysis of continuous longitudinal data in ambulatory assessment studies [112]. In addition, future studies would benefit from recruiting participants with more diversity in their activity and BMI levels and from either restricting the age of children and parents more clearly or tailoring the mHealth intervention specifically to the age of participants [51] to factor the different needs, for example, while growing up [113]. It needs to be noted that already active participants are still a valid target for interventions to maintain a sufficient PA level throughout life. Future primary prevention–focused interventions should, however, clearly define a PA level–specific aim. An example would be to aim for PA maintenance in sufficiently active participants and PA increase in insufficiently active ones and to frequently evaluate PA levels to allow for early interventions. However, both strategies need different accompanying BCTs (adjusted for maintenance purposes), which should also be accounted for.

Strengths and Limitations

The main strengths of the SF intervention are as follows: it collaboratively targets the family, it is designed as a cluster randomized controlled trial, it is theory based, and it incorporates 10 different BCTs. Furthermore, the goal-setting feature in the app is ad libitum selected by the family to fit their schedule and preferences with guidance from the results of the first measurement week. This enables the families to set self-selected goals, which have been found to increase motivation and adherence [114,115]. Another strength regarding the evaluation is that multiple outcome measures of self-reported and device-measured PA were considered. This is important as these measures are known to yield different results, and including multiple measurement tools improves the plausibility of the results [68]. Finally, using advanced statistical methods to consider the nested structure of the data by applying multilevel analyses enhanced the accuracy of the results, as it considered the variance based on the clustering of participants in families.

However, some limitations must be acknowledged. The actual sample size differed from the planned size by 6 families and 8 participants, falling short of the intended 52 families and 156 participants as determined by the power analysis in the study protocol [58]. This varied from analysis to analysis owing to missing values in distinct variables. In addition, the total population of the IG and CG were not perfectly balanced, and the level of significance was not adjusted for multiple primary outcomes. However, the observed *P* values for the interactions

in this paper are not close to .05, except for the group×time point interaction for joint meals per week and joint physical activities. Therefore, the abovementioned issues have likely no large impact on the study results. Regarding our sample, family sizes and ages within families have been very diverse. However, there is a lack of knowledge about how these composite family structures influence results regarding behavior change or the accomplishment of healthy lifestyles. For example, it might be assumed that older parents might be more aware of healthy food choices, as they consider healthy nutrition as being more important for themselves than their younger counterparts [116]. Advanced paternal age is associated with an increased risk of eating disorders in children, whereas younger paternal age is associated with a decreased risk of eating disorders in children [117]. A further restriction might be the age range, especially for children and adolescents. As this study includes the whole family, children of different ages and with different needs and perceptions were addressed similarly by the app, which might have affected intervention effects. Future interventions should aim to address the individual needs of the family members and tailor the intervention more specifically to the participants. Regarding the assessment of theoretical assumptions, it would have been interesting if the included influenced the effectiveness of our intervention. This cannot be answered by this study, as we only assessed the influence of covariables on PA and not on the effectiveness. Hence, the key constructs of the SDT [53,54] might separately be assessed in future studies. Another important factor concerning the sample of this study, which could explain the lack of significant effects for PA, was probably the highly active sample with approximately 650 minutes of MVPA per week per person, which already fulfilled the guidelines of the World Health Organization for PA [118]. This is in contrast to recent research about PA guideline fulfillment, which reliably reveals only small proportions of participants fulfilling the PA recommendations, with lower amounts corresponding to increasing ages [8,119]. Hence, future interventions should either aim at recruiting inactive participants to be able to detect changes in PA or should adjust statistical analyses and intervention design for monitoring PA maintenance over time. Our sample also deviated from the general German population regarding body composition and HE behavior [94]. With BMIs between 18 and 19 in children, and between 23 and 26 in adults, our study participants were of normal weight. Studies have shown that intervention effects can be expected to be higher on participants with overweight or obesity, both for HE and PA interventions [120]. This might be a probable explanation for the absence of intervention effects (ie, there was no or only little room for improvement in our sample). An additional aspect was that the participants had to use the provided smartphones instead of their own for equality and data security reasons, which can be burdensome. However, previous research showed no differences in engagement between participants with their own smartphones versus additional smartphones [121]. If a program aims for long-term sustainability beyond a scientific scope, the use of an additional phone must be considered very carefully. Another potentially limiting factor is the comparably short duration of the intervention. On the basis of literature regarding behavior change theories (ie, the transtheoretical model) [122,123], an intervention duration of 3 weeks might not have been sufficient [44], as the largest effects are expected for interventions that last over several weeks [97] in behavioral studies. However, mHealth intervention studies even revealed significant behavior change effects with intervention durations of only 1 [98], 2 [99], and 3 weeks [100]. In a similar vein, a recent meta-analysis of mobile apps for diet showed that interventions with longer duration were not generally more effective [44].

Conclusions

Taken together, the evaluation of the SF trial expands the existing body of evidence as it investigated the influence of a theory-based mHealth intervention targeting PA and HE in a collective family–based setting. However, no significant evidence for the effectiveness of the trial has been found. This finding, however, is not unique to mHealth interventions [56] and might be attributable to an initially active sample with a normal BMI. Therefore, future evaluations of interventions should also (1) consider methods that go beyond pre-post follow-up designs to account for the timeliness and complexity of mHealth interventions, (2) consider recruiting participants of all activity and weight levels, and (3) control for or restrict ages of children and parents.

Acknowledgments

The authors thank Dr Jens-Peter Gnam, Laura Berkhan, Jelena Bezold, Dr Christina Niermann, Dr Tobias Eckert, Nicolas Mahler, Fabian Fernandez-Pachecho, and Adam Kaczmarcyk for their support during the trial as well as all partners involved in the research association SMARTACT. This research was supported by the Federal Ministry of Education and Research (*Bundesministerium für Bildung und Forschung*; BMBF) within the project SMARTACT (BMBF grant: FKZ 01EL1820C). The authors also acknowledge support by the KIT-Publication Fund of the Karlsruhe Institute of Technology.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) checklist.
[DOCX File , 190 KB - mhealth_v12i1e51201_app1.docx]

Multimedia Appendix 2
Detailed results of the main analyses.
[DOCX File , 34 KB - mhealth_v12i1e51201_app2.docx]

Multimedia Appendix 3
Detailed results of the sensitivity analyses.
[DOCX File , 36 KB - mhealth_v12i1e51201_app3.docx]

References

1. Biddle SJ, Gorely T, Stensel DJ. Health-enhancing physical activity and sedentary behaviour in children and adolescents. J Sports Sci 2004 Aug;22(8):679-701. [doi: 10.1080/02640410410001712412] [Medline: 15370482]
2. Owen N, Healy GN, Matthews CE, Dunstan DW. Too much sitting: the population health science of sedentary behavior. Exerc Sport Sci Rev 2010 Jul;38(3):105-113 [FREE Full text] [doi: 10.1097/JES.0b013e3181e373a2] [Medline: 20577058]
3. Owen N, Sparling PB, Healy GN, Dunstan DW, Matthews CE. Sedentary behavior: emerging evidence for a new health risk. Mayo Clin Proc 2010 Dec;85(12):1138-1141 [FREE Full text] [doi: 10.4065/mcp.2010.0444] [Medline: 21123641]
4. Tremblay MS, LeBlanc AG, Kho ME, Saunders TJ, Larouche R, Colley RC, et al. Systematic review of sedentary behaviour and health indicators in school-aged children and youth. Int J Behav Nutr Phys Act 2011 Sep 21;8:98 [FREE Full text] [doi: 10.1186/1479-5868-8-98] [Medline: 21936895]
5. de Rezende LF, Rodrigues Lopes M, Rey-López JP, Matsudo VK, Luiz C. Sedentary behavior and health outcomes: an overview of systematic reviews. PLoS One 2014;9(8):e105620 [FREE Full text] [doi: 10.1371/journal.pone.0105620] [Medline: 25144686]
6. Bourke M, Haddara A, Loh A, Carson V, Breau B, Tucker P. Adherence to the World Health Organization's physical activity recommendation in preschool-aged children: a systematic review and meta-analysis of accelerometer studies. Int J Behav Nutr Phys Act 2023 Apr 26;20(1):52 [FREE Full text] [doi: 10.1186/s12966-023-01450-0] [Medline: 37101226]
7. McArthur BA, Volkova V, Tomopoulos S, Madigan S. Global prevalence of meeting screen time guidelines among children 5 years and younger: a systematic review and meta-analysis. JAMA Pediatr 2022 Apr 01;176(4):373-383 [FREE Full text] [doi: 10.1001/jamapediatrics.2021.6386] [Medline: 35157028]
8. Janssen I. Physical activity epidemiology. In: The Oxford Handbook of Exercise Psychology. Oxford, UK: Oxford University Press; 2012.
9. Nielsen SJ, Rossen LM, Harris DM, Odgen CL. Fruit and vegetable consumption of U.S. Youth, 2009-2010. NCHS Data Brief 2014 Jul(156):1-8 [FREE Full text] [Medline: 25027507]
10. Walther J, Aldrian U, Stüger HP, Kiefer I, Ekmekcioglu C. Nutrition, lifestyle factors, and mental health in adolescents and young adults living in Austria. Int J Adolesc Med Health 2014;26(3):377-386. [doi: 10.1515/ijamh-2013-0310] [Medline: 24803606]
11. Waxman A, World Health Assembly. WHO global strategy on diet, physical activity and health. Food Nutr Bull 2004 Sep;25(3):292-302. [doi: 10.1177/156482650402500310] [Medline: 15460274]
12. Granger E, Di Nardo F, Harrison A, Patterson L, Holmes R, Verma A. A systematic review of the relationship of physical activity and health status in adolescents. Eur J Public Health 2017 May 01;27(suppl_2):100-106. [doi: 10.1093/eurpub/ckw187] [Medline: 28340201]
13. Janssen I, Leblanc AG. Systematic review of the health benefits of physical activity and fitness in school-aged children and youth. Int J Behav Nutr Phys Act 2010 May 11;7:40 [FREE Full text] [doi: 10.1186/1479-5868-7-40] [Medline: 20459784]
14. Malina RM. Tracking of physical activity and physical fitness across the lifespan. Res Q Exerc Sport 1996 Sep;67(3 Suppl):S48-S57. [doi: 10.1080/02701367.1996.10608853] [Medline: 8902908]
15. Andersen LB, Haraldsdóttir J. Tracking of cardiovascular disease risk factors including maximal oxygen uptake and physical activity from late teenage to adulthood. An 8-year follow-up study. J Intern Med 1993 Sep 05;234(3):309-315. [doi: 10.1111/j.1365-2796.1993.tb00748.x] [Medline: 8354982]
16. Raitakari OT, Porkka KV, Taimela S, Telama R, Räsänen L, Viikari JS. Effects of persistent physical activity and inactivity on coronary risk factors in children and young adults. The Cardiovascular Risk in Young Finns Study. Am J Epidemiol 1994 Aug 01;140(3):195-205. [doi: 10.1093/oxfordjournals.aje.a117239] [Medline: 8030623]
17. Twisk JW, Van Mechelen W, Kemper HC, Post GB. The relation between “long-term exposure” to lifestyle during youth and young adulthood and risk factors for cardiovascular disease at adult age. J Adolesc Health 1997 Apr;20(4):309-319. [doi: 10.1016/s1054-139x(96)00183-8]
18. Fogelholm M. How physical activity can work? Int J Pediatr Obes 2008;3 Suppl 1:10-14. [doi: 10.1080/17477160801896481] [Medline: 18278627]
19. Telama R, Yang X, Viikari J, Välimäki I, Wanne O, Raitakari O. Physical activity from childhood to adulthood: a 21-year tracking study. Am J Prev Med 2005 Apr;28(3):267-273. [doi: 10.1016/j.amepre.2004.12.003] [Medline: 15766614]

20. Umberson D, Crosnoe R, Reczek C. Social relationships and health behavior across life course. Annu Rev Sociol 2010 Aug 01;36:139-157 [FREE Full text] [doi: 10.1146/annurev-soc-070308-120011] [Medline: 21921974]
21. Lawler M, Heary C, Nixon E. Peer support and role modelling predict physical activity change among adolescents over twelve months. J Youth Adolesc 2020 Jul;49(7):1503-1516. [doi: 10.1007/s10964-019-01187-9] [Medline: 31898771]
22. Yang X, Kukko T, Hirvensalo M, Biddle SJ, Rovio SP, Pahkala K, et al. Longitudinal associations between parental and offspring's leisure-time physical activity: the Young Finns Study. Scand J Med Sci Sports 2022 Jan;32(1):223-232. [doi: 10.1111/sms.14066] [Medline: 34617340]
23. Kaseva K, Hintsa T, Lipsanen J, Pulkki-Råback L, Hintsanen M, Yang X, et al. Parental physical activity associates with offspring's physical activity until middle age: a 30-year study. J Phys Act Health 2017 Jul;14(7):520-531. [doi: 10.1123/jpah.2016-0466] [Medline: 28290745]
24. Hursti UK. Factors influencing children's food choice. Ann Med 1999;31(sup1):26-32. [doi: 10.1080/07853890.1999.11904396] [Medline: 28850280]
25. Borah-Giddens J, Falciglia GA. A meta-analysis of the relationship in food preferences between parents and children. J Nutr Educ 1993 May;25(3):102-107. [doi: 10.1016/s0022-3182(12)80565-6]
26. Thorsteinsdottir S, Olsen A, Olafsdottir AS. Fussy eating among children and their parents: associations in parent-child dyads, in a sample of children with and without neurodevelopmental disorders. Nutrients 2021 Jun 25;13(7):2196 [FREE Full text] [doi: 10.3390/nu13072196] [Medline: 34202394]
27. Berge JM. A review of familial correlates of child and adolescent obesity: what has the 21st century taught us so far? Int J Adolesc Med Health 2009;21(4):457-483 [FREE Full text] [doi: 10.1515/ijamh.2009.21.4.457] [Medline: 20306760]
28. Hammons AJ, Fiese BH. Is frequency of shared family meals related to the nutritional health of children and adolescents? Pediatrics 2011 Jun;127(6):e1565-e1574 [FREE Full text] [doi: 10.1542/peds.2010-1440] [Medline: 21536618]
29. Dallacker M, Hertwig R, Mata J. The frequency of family meals and nutritional health in children: a meta-analysis. Obes Rev 2018 May;19(5):638-653. [doi: 10.1111/obr.12659] [Medline: 29334693]
30. Cleland V, Timperio A, Salmon J, Hume C, Telford A, Crawford D. A longitudinal study of the family physical activity environment and physical activity among youth. Am J Health Promot 2011;25(3):159-167. [doi: 10.4278/ajhp.090303-QUAN-93] [Medline: 21192744]
31. Viner RM, Ozer EM, Denny S, Marmot M, Resnick M, Fatusi A, et al. Adolescence and the social determinants of health. The Lancet 2012 Apr;379(9826):1641-1652. [doi: 10.1016/s0140-6736(12)60149-4]
32. Neumark-Sztainer D, MacLehose R, Loth K, Fulkerson JA, Eisenberg ME, Berge J. What's for dinner? Types of food served at family dinner differ across parent and family characteristics. Public Health Nutr 2014 Jan;17(1):145-155 [FREE Full text] [doi: 10.1017/S1368980012004594] [Medline: 23083836]
33. Dishman RK, Sallis JF, Orenstein DR. The determinants of physical activity and exercise. Public Health Rep 1985;100(2):158-171 [FREE Full text] [Medline: 3920714]
34. Martin JE, Dubbert PM. Exercise applications and promotion in behavioral medicine: current status and future directions. J Consult Clin Psychol 1982;50(6):1004-1017. [doi: 10.1037/0022-006x.50.6.1004]
35. van Sluijs EM, Kriemler S, McMinn AM. The effect of community and family interventions on young people's physical activity levels: a review of reviews and updated systematic review. Br J Sports Med 2011 Sep;45(11):914-922 [FREE Full text] [doi: 10.1136/bjsports-2011-090187] [Medline: 21836175]
36. Coughlin SS, Whitehead M, Sheats JQ, Mastromonico J, Smith S. A review of smartphone applications for promoting physical activity. Jacobs J Community Med 2016;2(1):021 [FREE Full text] [Medline: 27034992]
37. Lindsay Smith G, Banting L, Eime R, O'Sullivan G, van Uffelen JG. The association between social support and physical activity in older adults: a systematic review. Int J Behav Nutr Phys Act 2017 Apr 27;14(1):56 [FREE Full text] [doi: 10.1186/s12966-017-0509-8] [Medline: 28449673]
38. Murphy JK, Williamson DA, Buxton AE, Moody SC, Absher N, Warner M. The long-term effects of spouse involvement upon weight loss and maintenance. Behav Ther 1982 Nov;13(5):681-693. [doi: 10.1016/s0005-7894(82)80024-5]
39. Dubbert PM, Wilson G. Goal-setting and spouse involvement in the treatment of obesity. Behav Res Ther 1984;22(3):227-242. [doi: 10.1016/0005-7967(84)90003-2] [Medline: 6466273]
40. Brownell KD, Heckerman CL, Westlake RJ, Hayes SC, Monti PM. The effect of couples training and partner co-operativeness in the behavioral treatment of obesity. Behav Res Ther 1978;16(5):323-333. [doi: 10.1016/0005-7967(78)90002-5] [Medline: 743074]
41. Whittaker R. Key issues in mobile health and implications for New Zealand. Health Care Inform Rev Online 2012;16(2):2-7.
42. König LM, Sproesser G, Schupp HT, Renner B. Describing the process of adopting nutrition and fitness apps: behavior stage model approach. JMIR Mhealth Uhealth 2018 Mar 13;6(3):e55 [FREE Full text] [doi: 10.2196/mhealth.8261] [Medline: 29535078]
43. Conroy DE, Yang CH, Maher JP. Behavior change techniques in top-ranked mobile apps for physical activity. Am J Prev Med 2014 Jun;46(6):649-652. [doi: 10.1016/j.amepre.2014.01.010] [Medline: 24842742]
44. Villinger K, Wahl DR, Boeing H, Schupp HT, Renner B. The effectiveness of app-based mobile interventions on nutrition behaviours and nutrition-related health outcomes: a systematic review and meta-analysis. Obes Rev 2019 Oct;20(10):1465-1484 [FREE Full text] [doi: 10.1111/obr.12903] [Medline: 31353783]

45. Rose T, Barker M, Maria Jacob C, Morrison L, Lawrence W, Strömmer S, et al. A systematic review of digital interventions for improving the diet and physical activity behaviors of adolescents. J Adolesc Health 2017 Dec;61(6):669-677 [FREE Full text] [doi: 10.1016/j.jadohealth.2017.05.024] [Medline: 28822682]
46. Iribarren SJ, Cato K, Falzon L, Stone PW. What is the economic evidence for mHealth? A systematic review of economic evaluations of mHealth solutions. PLoS One 2017;12(2):e0170581 [FREE Full text] [doi: 10.1371/journal.pone.0170581] [Medline: 28152012]
47. Michie S, Ashford S, Sniehotta FF, Dombrowski SU, Bishop A, French DP. A refined taxonomy of behaviour change techniques to help people change their physical activity and healthy eating behaviours: the CALO-RE taxonomy. Psychol Health 2011 Nov;26(11):1479-1498. [doi: 10.1080/08870446.2010.540664] [Medline: 21678185]
48. Michie S, Abraham C, Whittington C, McAteer J, Gupta S. Effective techniques in healthy eating and physical activity interventions: a meta-regression. Health Psychol 2009 Nov;28(6):690-701. [doi: 10.1037/a0016136] [Medline: 19916637]
49. Middelweerd A, Mollee JS, van der Wal CN, Brug J, Te Velde SJ. Apps to promote physical activity among adults: a review and content analysis. Int J Behav Nutr Phys Act 2014 Jul 25;11:97 [FREE Full text] [doi: 10.1186/s12966-014-0097-9] [Medline: 25059981]
50. Wunsch K, Eckert T, Fiedler J, Woll A. Just-in-time adaptive interventions in mobile physical activity interventions- A synthesis of frameworks and future directions. Eur Health Psychol 2022;22(4):834-842.
51. Baumann H, Fiedler J, Wunsch K, Woll A, Wollesen B. mHealth interventions to reduce physical inactivity and sedentary behavior in children and adolescents: systematic review and meta-analysis of randomized controlled trials. JMIR Mhealth Uhealth 2022 May 11;10(5):e35920 [FREE Full text] [doi: 10.2196/35920] [Medline: 35544294]
52. Pina LR, Sien SW, Ward T, Yip JC, Munson SA, Fogarty J, et al. From personal informatics to family informatics: understanding family practices around health monitoring. In: Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing. 2017 Presented at: CSCW '17; February 25-March 1, 2017; Portland, OR. [doi: 10.1145/2998181.2998362]
53. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000;55(1):68-78. [doi: 10.1037/0003-066X.55.1.68]
54. Deci EL, Ryan RM. Self-determination theory. In: Handbook of Theories of Social Psychology: Volume 1. Thousand Oaks, CA: SAGE Publications Ltd; 2012.
55. Cox MJ, Paley B. Understanding families as systems. Curr Dir Psychol Sci 2016 Jun 24;12(5):193-196. [doi: 10.1111/1467-8721.01259]
56. Fiedler J, Eckert T, Wunsch K, Woll A. Key facets to build up eHealth and mHealth interventions to enhance physical activity, sedentary behavior and nutrition in healthy subjects - an umbrella review. BMC Public Health 2020 Oct 23;20(1):1605 [FREE Full text] [doi: 10.1186/s12889-020-09700-7] [Medline: 33097013]
57. Rodríguez-González P, Hassan MA, Gao Z. Effects of family-based interventions using mobile apps on youth's physical activity: a systematic review. J Clin Med 2022 Aug 17;11(16):4798 [FREE Full text] [doi: 10.3390/jcm11164798] [Medline: 36013037]
58. Wunsch K, Eckert T, Fiedler J, Cleven L, Niermann C, Reiterer H, et al. Effects of a collective family-based mobile health intervention called "SMARTFAMILY" on promoting physical activity and healthy eating: protocol for a randomized controlled trial. JMIR Res Protoc 2020 Nov 11;9(11):e20534 [FREE Full text] [doi: 10.2196/20534] [Medline: 33174849]
59. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
60. Esserman D, Allore HG, Travison TG. The method of randomization for cluster-randomized trials: challenges of including patients with multiple chronic conditions. Int J Stat Med Res 2016 Jan 08;5(1):2-7 [FREE Full text] [doi: 10.6000/1929-6029.2016.05.01.1] [Medline: 27478520]
61. Butscher S, Wang Y, Müller J, Ziesemer K, Villinger K, Wahl D, et al. Lightweight visual data analysis on mobile devices - providing self-monitoring feedback. In: Proceedings of the 1st International Workshop on Valuable Visualization of Healthcare Information. 2016 Presented at: VVH 2016; June 7, 2016; Bari, Italy. [doi: 10.1145/2909132.2927474]
62. Wahl D, Villinger K, Sproesser G, Schupp HT, Renner B. The behavioural signature of snacking – a visual analysis. In: Proceedings of the 31st conference of the EHPS: Innovative Ideas in Health Psychology. 2017 Presented at: EHPS 2017; August 29-September 2, 2017; Padua, Italy.
63. Wahl DR, Villinger K, Blumenschein M, König LM, Ziesemer K, Sproesser G, et al. Why we eat what we eat: assessing dispositional and in-the-moment eating motives by using ecological momentary assessment. JMIR Mhealth Uhealth 2020 Jan 07;8(1):e13191 [FREE Full text] [doi: 10.2196/13191] [Medline: 31909719]
64. Wunsch K, Gnam JP, Renner B, Woll A, Cleven L. A family-based m-health intervention to promote physical activity and healthy eating. In: Proceedings of the 32nd Annual Conference of the European Health Psychology Society. 2018 Presented at: EHPS 2018; August 21-25, 2018; Galway, Ireland.
65. Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007 May;39(2):175-191. [doi: 10.3758/bf03193146] [Medline: 17695343]

66. Schoeppe S, Alley S, Van Lippevelde W, Bray NA, Williams SL, Duncan MJ, et al. Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: a systematic review. Int J Behav Nutr Phys Act 2016 Dec 07;13(1):127 [FREE Full text] [doi: 10.1186/s12966-016-0454-y] [Medline: 27927218]
67. Burchartz A, Anedda B, Auerswald T, Giurgiu M, Hill H, Ketelhut S, et al. Assessing physical behavior through accelerometry – State of the science, best practices and future directions. Psychol Sport Exerc 2020 Jul;49:101703. [doi: 10.1016/j.psychsport.2020.101703]
68. Fiedler J, Eckert T, Burchartz A, Woll A, Wunsch K. Comparison of self-reported and device-based measured physical activity using measures of stability, reliability, and validity in adults and children. Sensors (Basel) 2021 Apr 10;21(8):2672 [FREE Full text] [doi: 10.3390/s21082672] [Medline: 33920145]
69. Anastasopoulou P, Härtel S, Hey S. A comparison of two commercial activity monitors for measuring step counts during different everyday life walking activities. Int J Sports Sci Eng 2013;7(1):31-35.
70. Anastasopoulou P, Tubic M, Schmidt S, Neumann R, Woll A, Härtel S. Validation and comparison of two methods to assess human energy expenditure during free-living activities. PLoS One 2014;9(2):e90606 [FREE Full text] [doi: 10.1371/journal.pone.0090606] [Medline: 24587401]
71. Härtel S, Gnam JP, Löffler S, Bös K. Estimation of energy expenditure using accelerometers and activity-based energy models—validation of a new device. Eur Rev Aging Phys Act 2010 Dec 02;8(2):109-114. [doi: 10.1007/s11556-010-0074-5]
72. Barouni A, Ottenbacher J, Schneider J, Feige B, Riemann D, Herlan A, et al. Ambulatory sleep scoring using accelerometers-distinguishing between nonwear and sleep/wake states. PeerJ 2020;8:e8284 [FREE Full text] [doi: 10.7717/peerj.8284] [Medline: 31915581]
73. Mäder U, Martin BW, Schutz Y, Marti B. Validity of four short physical activity questionnaires in middle-aged persons. Med Sci Sports Exerc 2006 Jul;38(7):1255-1266. [doi: 10.1249/01.mss.0000227310.18902.28] [Medline: 16826022]
74. Prochaska JJ, Sallis JF, Long B. A physical activity screening measure for use with adolescents in primary care. Arch Pediatr Adolesc Med 2001 May;155(5):554-559. [doi: 10.1001/archpedi.155.5.554] [Medline: 11343497]
75. Global recommendations on physical activity for health. World Health Organization. 2010 Jan 1. URL: https://www.who.int/publications/i/item/9789241599979 [accessed 2024-04-16]
76. DeSalvo KB, Fisher WP, Tran K, Bloser N, Merrill W, Peabody J. Assessing measurement properties of two single-item general health measures. Qual Life Res 2006 Mar;15(2):191-201. [doi: 10.1007/s11136-005-0887-2] [Medline: 16468076]
77. Markland D, Tobin V. Modification to the behavioural regulation in exercise questionnaire to include an assesment of amotivation. J Sport Exerc Psychol 2004 Jun;26(2):191-196. [doi: 10.1123/jsep.26.2.191]
78. Pelletier LG, Dion SC, Slovinec-D'Angelo M, Reid R. Why do you regulate what you eat? Relationships between forms of regulation, eating behaviors, sustained dietary behavior change, and psychological adjustment. Motiv Emot 2004 Sep;28(3):245-277 [FREE Full text] [doi: 10.1023/B:MOEM.0000040154.40922.14]
79. Hahn A, Renner B, Schwarzer R. Risiko und Gesundheitsverhalten: Dokumentation der Meßinstrumente des Forschungsprojekts "Berlin risk appraisal and health motivation study" (BRAHMS). Berlin, Germany: Freie Universität Berlin; 1996.
80. Niermann C, Krapf F, Renner B, Reiner M, Woll A. Family health climate scale (FHC-scale): development and validation. Int J Behav Nutr Phys Act 2014 Mar 05;11(1):30 [FREE Full text] [doi: 10.1186/1479-5868-11-30] [Medline: 24593840]
81. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing. 2021. URL: https://www.R-project.org/ [accessed 2024-04-16]
82. Allaire JJ. RStudio: integrated development environment for R. The Comprehensive R Archive Network. URL: https://www.r-project.org/conferences/useR-2011/abstracts/180111-allairejj.pdf [accessed 2024-04-16]
83. Wickham H. ggplot2: Elegant Graphics for Data Analysis. New York, NY: Springer; 2009.
84. Allen M, Poggiali D, Whitaker K, Marshall TR, Kievit RA. Raincloud plots: a multi-platform tool for robust data visualization. Wellcome Open Res 2019 Apr 1;4:63. [doi: 10.12688/wellcomeopenres.15191.1]
85. Kuznetsova A, Brockhoff PB, Christensen RH. lmerTest package: tests in linear mixed effects models. J Stat Softw 2017;82(13):1-26. [doi: 10.18637/jss.v082.i13]
86. Lüdecke D. sjPlot - Data visualization for statistics in social science. The Comprehensive R Archive Network. 2023. URL: https://CRAN.R-project.org/package=sjPlot [accessed 2024-04-16]
87. Lüdecke D, Ben-Shachar MS, Patil I, Waggoner P, Makowski D. performance: an R package for assessment, comparison and testing of statistical models. J Open Source Softw 2021 Apr;6(60):3139. [doi: 10.21105/joss.03139]
88. SMARTFAMILY 1.0 main results. OSF Home. 2023. URL: https://osf.io/rxk6m/ [accessed 2024-04-19]
89. WHO guidelines on physical activity and sedentary behaviour. World Health Organization. 2020 Nov 25. URL: https://www.who.int/publications/i/item/9789240015128 [accessed 2024-04-16]
90. Tudor-Locke C, Craig CL, Brown WJ, Clemes SA, De Cocker K, Giles-Corti B, et al. How many steps/day are enough? For adults. Int J Behav Nutr Phys Act 2011 Jul 28;8:79 [FREE Full text] [doi: 10.1186/1479-5868-8-79] [Medline: 21798015]
91. Paluch AE, Bajpai S, Bassett DR, Carnethon MR, Ekelund U, Evenson KR, Steps for Health Collaborative. Daily steps and all-cause mortality: a meta-analysis of 15 international cohorts. Lancet Public Health 2022 Mar;7(3):e219-e228 [FREE Full text] [doi: 10.1016/S2468-2667(21)00302-9] [Medline: 35247352]

92. 10 guidelines of the German Nutrition Society (DGE) for a wholesome diet. Deutsche Gesellschaft für Ernährung. URL: https://www.dge.de/fileadmin/dok/english/10-guidelines/10-guidelines-wholesome-diet-dge.pdf [accessed 2024-02-16]
93. Bann D, Scholes S, Fluharty M, Shure N. Adolescents' physical activity: cross-national comparisons of levels, distributions and disparities across 52 countries. Int J Behav Nutr Phys Act 2019 Dec 30;16(1):141 [FREE Full text] [doi: 10.1186/s12966-019-0897-z] [Medline: 31888652]
94. Rabenberg M, Mensink GB. Fruit and vegetable consumption today. GBE kompakt 2011;2(6):1-8 [FREE Full text]
95. Zhang M, Wang W, Li M, Sheng H, Zhai Y. Efficacy of mobile health applications to improve physical activity and sedentary behavior: a systematic review and meta-analysis for physically inactive individuals. Int J Environ Res Public Health 2022 Apr 18;19(8):4905 [FREE Full text] [doi: 10.3390/ijerph19084905] [Medline: 35457775]
96. Emberson MA, Lalande A, Wang D, McDonough DJ, Liu W, Gao Z. Effectiveness of smartphone-based physical activity interventions on individuals' health outcomes: a systematic review. Biomed Res Int 2021;2021:6296896 [FREE Full text] [doi: 10.1155/2021/6296896] [Medline: 34409104]
97. de Greeff JW, Bosker RJ, Oosterlaan J, Visscher C, Hartman E. Effects of physical activity on executive functions, attention and academic performance in preadolescent children: a meta-analysis. J Sci Med Sport 2018 May;21(5):501-507. [doi: 10.1016/j.jsams.2017.09.595] [Medline: 29054748]
98. Garde A, Umedaly A, Abulnaga SM, Robertson L, Junker A, Chanoine JP, et al. Assessment of a mobile game ("MobileKids Monster Manor") to promote physical activity among children. Games Health J 2015 Apr;4(2):149-158. [doi: 10.1089/g4h.2014.0095] [Medline: 26181809]
99. Sirriyeh R, Lawton R, Ward J. Physical activity and adolescents: an exploratory randomized controlled trial investigating the influence of affective and instrumental text messages. Br J Health Psychol 2010 Nov;15(Pt 4):825-840. [doi: 10.1348/135910710X486889] [Medline: 20156396]
100. Fukuoka Y, Vittinghoff E, Jong SS, Haskell W. Innovation to motivation--pilot study of a mobile phone intervention to increase physical activity among sedentary women. Prev Med 2010;51(3-4):287-289 [FREE Full text] [doi: 10.1016/j.ypmed.2010.06.006] [Medline: 20600263]
101. Vale S, Silva P, Santos R, Soares-Miranda L, Mota J. Compliance with physical activity guidelines in preschool children. J Sports Sci 2010 Apr;28(6):603-608. [doi: 10.1080/02640411003702694] [Medline: 20397094]
102. Jago R, Ness AR, Emmett P, Mattocks C, Jones L, Riddoch CJ. Obesogenic diet and physical activity: independent or associated behaviours in adolescents? Public Health Nutr 2010 May;13(5):673-681 [FREE Full text] [doi: 10.1017/S1368980009992151] [Medline: 19954571]
103. Hammersley ML, Jones RA, Okely AD. Parent-focused childhood and adolescent overweight and obesity eHealth interventions: a systematic review and meta-analysis. J Med Internet Res 2016 Jul 21;18(7):e203 [FREE Full text] [doi: 10.2196/jmir.5893] [Medline: 27443862]
104. Zarnowiecki D, Mauch CE, Middleton G, Matwiejczyk L, Watson WL, Dibbs J, et al. A systematic evaluation of digital nutrition promotion websites and apps for supporting parents to influence children's nutrition. Int J Behav Nutr Phys Act 2020 Feb 10;17(1):17 [FREE Full text] [doi: 10.1186/s12966-020-0915-1] [Medline: 32041640]
105. Chai LK, Farletti R, Fathi L, Littlewood R. A rapid review of the impact of family-based digital interventions for obesity prevention and treatment on obesity-related outcomes in primary school-aged children. Nutrients 2022 Nov 15;14(22):4837 [FREE Full text] [doi: 10.3390/nu14224837] [Medline: 36432522]
106. Johansson L, Hagman E, Danielsson P. A novel interactive mobile health support system for pediatric obesity treatment: a randomized controlled feasibility trial. BMC Pediatr 2020 Sep 23;20(1):447 [FREE Full text] [doi: 10.1186/s12887-020-02338-9] [Medline: 32967638]
107. Trost SG, Brookes DS. Effectiveness of a novel digital application to promote fundamental movement skills in 3- to 6-year-old children: a randomized controlled trial. J Sports Sci 2021 Feb;39(4):453-459. [doi: 10.1080/02640414.2020.1826657] [Medline: 32985373]
108. de Wit JB, Stok FM, Smolenski DJ, de Ridder DD, de Vet E, Gaspar T, et al. Food culture in the home environment: family meal practices and values can support healthy eating and self-regulation in young people in four European countries. Appl Psychol Health Well Being 2015 Mar;7(1):22-40. [doi: 10.1111/aphw.12034] [Medline: 25346476]
109. Kelders SM, van Zyl LE, Ludden GD. The concept and components of engagement in different domains applied to eHealth: a systematic scoping review. Front Psychol 2020;11:926 [FREE Full text] [doi: 10.3389/fpsyg.2020.00926] [Medline: 32536888]
110. Mclaughlin M, Delaney T, Hall A, Byaruhanga J, Mackie P, Grady A, et al. Associations between digital health intervention engagement, physical activity, and sedentary behavior: systematic review and meta-analysis. J Med Internet Res 2021 Feb 19;23(2):e23180 [FREE Full text] [doi: 10.2196/23180] [Medline: 33605897]
111. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-Time Adaptive Interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. Ann Behav Med 2018 May 18;52(6):446-462 [FREE Full text] [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
112. Reichert M, Giurgiu M, Koch E, Wieland LM, Lautenbach S, Neubauer AB, et al. Ambulatory assessment for physical activity research: state of the science, best practices and future directions. Psychol Sport Exerc 2020 Sep;50:101742 [FREE Full text] [doi: 10.1016/j.psychsport.2020.101742] [Medline: 32831643]

113. Salmon J, Booth ML, Phongsavan P, Murphy N, Timperio A. Promoting physical activity participation among children and adolescents. Epidemiol Rev 2007 May 02;29(1):144-159. [doi: 10.1093/epirev/mxm010] [Medline: 17556765]
114. Locke E, Latham G. The application of goal setting to sports. J Sport Psychol 1985;7(3):205-222. [doi: 10.1123/jsp.7.3.205]
115. Latham GP, Locke EA. New developments in and directions for goal-setting research. Eur Psychol 2007 Jan;12(4):290-300. [doi: 10.1027/1016-9040.12.4.290]
116. Hjartåker A, Lund E. Relationship between dietary habits, age, lifestyle, and socio-economic status among adult Norwegian women. The Norwegian Women and Cancer Study. Eur J Clin Nutr 1998 Aug;52(8):565-572. [doi: 10.1038/sj.ejcn.1600608] [Medline: 9725656]
117. Javaras KN, Rickert ME, Thornton LM, Peat CM, Baker JH, Birgegård A, et al. Paternal age at childbirth and eating disorders in offspring. Psychol Med 2017 Feb;47(3):576-584 [FREE Full text] [doi: 10.1017/S0033291716002610] [Medline: 27808013]
118. Bull FC, Al-Ansari SS, Biddle S, Borodulin K, Buman MP, Cardon G, et al. World Health Organization 2020 guidelines on physical activity and sedentary behaviour. Br J Sports Med 2020 Dec;54(24):1451-1462 [FREE Full text] [doi: 10.1136/bjsports-2020-102955] [Medline: 33239350]
119. Friel CP, Duran AT, Shechter A, Diaz KM. U.S. children meeting physical activity, screen time, and sleep guidelines. Am J Prev Med 2020 Oct;59(4):513-521 [FREE Full text] [doi: 10.1016/j.amepre.2020.05.007] [Medline: 32863080]
120. Pocock SJ, Assmann SE, Enos LE, Kasten LE. Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. Stat Med 2002 Oct 15;21(19):2917-2930. [doi: 10.1002/sim.1296] [Medline: 12325108]
121. Ziesemer K, König LM, Boushey CJ, Villinger K, Wahl DR, Butscher S, et al. Occurrence of and reasons for "missing events" in mobile dietary assessments: results from three event-based ecological momentary assessment studies. JMIR Mhealth Uhealth 2020 Oct 14;8(10):e15430 [FREE Full text] [doi: 10.2196/15430] [Medline: 33052123]
122. Prochaska JO, Velicer WF. The transtheoretical model of health behavior change. Am J Health Promot 1997;12(1):38-48. [doi: 10.4278/0890-1171-12.1.38] [Medline: 10170434]
123. Wunsch K. Transtheoretical model. In: Hackfort D, Schinke RJ, editors. Dictionary of Sport Psychology: Sport, Exercise, and Performing Arts. Cambridge, MA: Academic Press; 2019.

Abbreviations

BCT: behavior change technique
BREQ-2: Behavioral Regulation in Exercise Questionnaire
CG: control group
CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
FHC: Family Health Climate
FHC-NU: Family Health Climate–Nutrition
FHC-PA: Family Health Climate–Physical Activity
FVI: fruit and vegetable intake
HE: healthy eating
IG: intervention group
MET: metabolic equivalents
mHealth: mobile Health
MVPA: moderate to vigorous physical activity
PA: physical activity
SDT: self-determination theory
SF: SMARTFAMILY

Edited by L Buis; submitted 24.07.23; peer-reviewed by S Yang, MA Rusandi, R Newton; comments to author 29.09.23; revised version received 27.10.23; accepted 27.02.24; published 26.04.24.

Please cite as:
Wunsch K, Fiedler J, Hubenschmid S, Reiterer H, Renner B, Woll A
An mHealth Intervention Promoting Physical Activity and Healthy Eating in a Family Setting (SMARTFAMILY): Randomized Controlled Trial
JMIR Mhealth Uhealth 2024;12:e51201
URL: https://mhealth.jmir.org/2024/1/e51201
doi:10.2196/51201
PMID:38669071

Original Paper

Efficacy of the Flo App in Improving Health Literacy, Menstrual and General Health, and Well-Being in Women: Pilot Randomized Controlled Trial

Adam C Cunningham[1], PhD; Carley Prentice[1], MPH; Kimberly Peven[1,2], PhD; Aidan Wickham[1], PhD; Ryan Bamford[1], MS; Tara Radovic[1,3], MSc; Anna Klepchukova[1], MD; Maria Fomina[4], MSc; Katja Cunningham[5], PhD; Sarah Hill[5], PhD; Liisa Hantsoo[6], PhD; Jennifer Payne[7], MD; Liudmila Zhaunova[1*], PhD; Sonia Ponzo[1,8*], PhD

[1]Flo Health UK Limited, London, United Kingdom

[2]Maternal, Adolescent, Reproductive & Child Health Centre, London School of Hygiene & Tropical Medicine, London, United Kingdom

[3]Department of Psychology and Ergonomics, Technische Universitaet Berlin, Berlin, Germany

[4]Department of Behavioural Science and Health, University College London, London, United Kingdom

[5]Department of Psychology, Texas Christian University, Fort Worth, TX, United States

[6]Department of Psychiatry & Behavioral Sciences, Johns Hopkins University, Baltimore, MD, United States

[7]Department of Psychiatry and Neurobehavioral Sciences, University of Virginia, Charlottesville, VA, United States

[8]Institute of Health Informatics, University College London, London, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Liudmila Zhaunova, PhD
Flo Health UK Limited
27 Old Gloucester Street
London, WC1N 3AX
United Kingdom
Phone: 44 7770032146
Email: l_zhaunova@flo.health

Abstract

Background: Reproductive health literacy and menstrual health awareness play a crucial role in ensuring the health and well-being of women and people who menstruate. Further, awareness of one's own menstrual cycle patterns and associated symptoms can help individuals identify and manage conditions of the menstrual cycle such as premenstrual syndrome (PMS) and premenstrual dysphoric disorder (PMDD). Digital health products, and specifically menstrual health apps, have the potential to effect positive change due to their scalability and ease of access.

Objective: The primary aim of this study was to measure the efficacy of a menstrual and reproductive health app, Flo, in improving health literacy and health and well-being outcomes in menstruating individuals with and without PMS and PMDD. Further, we explored the possibility that the use of the Flo app could positively influence feelings around reproductive health management and communication about health, menstrual cycle stigma, unplanned pregnancies, quality of life, work productivity, absenteeism, and body image.

Methods: We conducted 2 pilot, 3-month, unblinded, 2-armed, remote randomized controlled trials on the effects of using the Flo app in a sample of US-based (1) individuals who track their cycles (n=321) or (2) individuals who track their cycles and are affected by PMS or PMDD (n=117).

Results: The findings revealed significant improvements at the end of the study period compared to baseline for our primary outcomes of health literacy (cycle tracking: D^{-}=1.11; t_{311}=5.73, P<.001; PMS or PMDD: D^{-}=1.20; t_{115}=3.76, P<.001) and menstrual health awareness (D^{-}=3.97; t_{311}=7.71, P<.001), health and well-being (D^{-}=3.44; t_{311}=5.94, P<.001), and PMS or PMDD symptoms burden (D^{-}=–7.08; t_{115}=–5.44, P<.001). Improvements were also observed for our secondary outcomes of feelings of control and management over health (D^{-}=1.01; t_{311}=5.08, P<.001), communication about health (D^{-}=0.93; t_{311}=2.41, P=.002), menstrual cycle stigma (D^{-}=–0.61; t_{311}=–2.73, P=.007), and fear of unplanned pregnancies (D^{-}=–0.22; t_{311}=–2.11, P=.04) for

those who track their cycles, as well as absenteeism from work and education due to PMS or PMDD (D^{-}=–1.67; t_{144}=–2.49, P=.01).

Conclusions: These pilot randomized controlled trials demonstrate that the use of the Flo app improves menstrual health literacy and awareness, general health and well-being, and PMS or PMDD symptom burden. Considering the widespread use and affordability of the Flo app, these findings show promise for filling important gaps in current health care provisioning such as improving menstrual knowledge and health.

Trial Registration: OSF Registries osf.io/pcgw7; https://osf.io/pcgw7 ; OSF Registries osf.io/ry8vq; https://osf.io/ry8vq

(JMIR Mhealth Uhealth 2024;12:e54124) doi:10.2196/54124

KEYWORDS

digital health; health literacy; menstrual cycle; period tracking app; women's health; PMS; PMDD; tracking; app; application; tracking app; tracking application; menstrual; women; efficacy; general health; wellbeing; randomized controlled trial; awareness; symptoms; manage; management; premenstrual; premenstrual syndrome; premenstrual dysphoric disorder; reproductive; reproductive health; health management; communication; pregnancy; quality of life; productivity; education; functionality

Introduction

Many women of childbearing age lack foundational knowledge about their menstrual cycle despite it being a vital indicator of women's health [1]. Nearly half of all women are unaware of the average length of a regular menstrual cycle and around 40% are unfamiliar with the ovulatory cycle [1,2]. Poor knowledge about menstrual health can hinder women's ability to make informed decisions about their health [3], negatively impact their daily sociocultural activities [4-6], and contribute to misconceptions and taboos that compromise their overall physical and mental well-being [7,8]. For example, a recent systematic review found that poor knowledge regarding menstrual health may lead to negative experiences such as shame and the lack of confidence to engage in social activities, ultimately increasing mental burden [9]. Further, low menstrual literacy has been associated with inadequate self-care behaviors [10] and reduced satisfaction with medical visits, attributed to ineffective communication between patients and health care providers [11]. This is further evidenced by the results of the UK Women's Health Strategy Survey, where as many as 1 in 4 women reported not feeling comfortable talking to health care professionals about their menstrual cycle [12]. Interventions aimed at improving menstrual cycle knowledge are therefore likely to be associated with improvements in a wide range of domains relevant to women's physical and mental health.

A good example of where improved menstrual and reproductive health literacy may have beneficial outcomes is in the case of premenstrual syndrome (PMS) and premenstrual dysphoric disorder (PMDD). PMS and PMDD, a more severe form of PMS, are common, with a worldwide prevalence of 48% for PMS [13,14] and 3% to 8% for PMDD [14]. Affected individuals report a broad range of psychological and physical symptoms occurring in the luteal phase of the menstrual cycle, which can cause social or economic dysfunction and affective and functional impairment [15,16]. Severe PMS symptoms are associated with poorer ratings of quality of life and higher levels of stress [14,17,18]. Similarly, the health burden of PMDD has been deemed greater than type 2 diabetes and hypertension in terms of reported pain [14,18]. Both PMS and PMDD have a proven negative impact on workplace productivity and in some cases even lead to absenteeism [19,20]. Despite this heavy burden on individuals and society, these conditions are often poorly treated and overlooked in research. Evidence suggests that the symptoms of PMS and PMDD may be dismissed by health care professionals, and even by the affected women themselves, who might consider them as a normal part of the menstrual cycle [21,22], highlighting how poor health literacy can result in delays in seeking care for reproductive health conditions [1,23,24].

While there is a notable lack of research into topics related to women's health [25], there is an acute shortage of research that validates the effectiveness of educational interventions for menstrual health literacy. Most studies have relatively small sample sizes or focus on limited populations or age ranges [26,27]. Digital, internet, and mobile-based health technologies can provide novel ways to deliver interventions for menstrual health and have gained popularity due to their scalability, availability, and anonymity [28,29], with the latter being crucial for menstrual and mental health [30-35]. Health apps focusing on menstrual and reproductive health can help individuals feel more prepared for their menstrual cycle symptoms, be more aware of their body signals throughout the cycle, and be better able to manage associated symptoms [36]. Additionally, evidence suggests that self-guided interventions using mobile phones can successfully equip individuals experiencing PMS and PMDD with strategies to promote and maintain healthy behavior, improve PMS symptoms, and reduce their interference with everyday life activities [37,38]. Although preliminary studies show promising results, research on the efficacy of such health solutions remains limited.

The Flo smartphone app [39] (Flo Health) is a period tracking and reproductive health app. It provides personalized menstruation and ovulation predictions, symptom forecasts based on user input, and expert-reviewed content accessible via the library or the health assistant (chatbot). The users of the Flo app also interact anonymously on a community platform called "Secret Chats," which facilitates conversations on stigmatized topics such as sexual and menstrual health [29,30,33].

The primary objective of these 2 pilot randomized controlled trials (RCTs) was to measure the efficacy of the Flo app in improving menstrual-related health literacy and awareness and well-being outcomes in (1) individuals who reported an interest

in using the app to track their cycles and symptoms (trial 1) or (2) individuals who reported an interest in using the app to track their cycles and symptoms and reported being affected by symptoms of conditions of the menstrual cycle such as PMS and PMDD (trial 2). The secondary objective was to investigate the effect of app use in these populations on overall feelings of health management, communication about health (eg, with health care providers or with their partner), menstrual cycle stigma, fear of unplanned pregnancies, quality of life, productivity at work, absenteeism, and body image. For trial 1, we hypothesized that the use of the Flo app would significantly improve menstrual-related health literacy and awareness and general health and well-being outcomes in the interventional group compared to the waitlist control group; for trial 2, we hypothesized that PMS- or PMDD-related health literacy and PMS or PMDD symptom burden would significantly decrease in the interventional group compared to the waitlist control group. We also hypothesized that all secondary outcomes would either significantly improve (feelings of health management, communication about health with health care providers or with a partner, quality of life, productivity at work, and body image) or decrease (menstrual cycle stigma, fear of unplanned pregnancies, and absenteeism from work or school due to PMS or PMDD symptoms) compared to the waitlist control group.

Methods

Study Design

We intended to conduct 2 fully powered, 3-month, unblinded, 2-armed RCTs, with a between-subjects component (group: intervention or control) and a within-subjects component (time point; see Multimedia Appendix 1 for the CONSORT [Consolidated Standards of Reporting Trials] checklist).

Trial 1 aimed to investigate the efficacy of the Flo app in improving health literacy and health and well-being outcomes in participants who report their current reproductive health aim as (1) cycle tracking, (2) trying to conceive (TTC), or (3) pregnancy tracking. Trial 2 aimed to investigate the efficacy of the Flo app in improving health literacy and PMS and PMDD symptom burden outcomes in participants who report moderate-to-severe PMS or PMDD symptoms. Trials 1 and 2 have been registered on the Open Science Framework [40,41].

Due to recruitment issues in trial 1, very low numbers of eligible participants were recruited in the pregnancy tracking group, and an extremely low number of participants completed the study in the TTC group. We therefore deviated from the preregistered study design by excluding these groups from the further analyses and focused only on the cycle tracking group. Additionally, as both trials had relatively low numbers of study participants who followed the study protocol, we converted our trials into feasibility pilots to inform further full RCTs with improved recruitment strategies. All questionnaire measures were self-reported through a web-based questionnaire platform (Survey Monkey).

Ethical Considerations

Trials 1 and 2 were approved by an independent ethical review board (WCG IRB, ethics applications 20222535 and 20222597, respectively), using protocols specific to each population. All participants provided informed, electronic consent via a checkbox prior to their enrollment in the study, and data were deidentified. Participants who fully completed all data collection activities received a free yearly subscription to Flo Premium (US $39.99) and were entered into a lottery to win up to US $600 in Amazon Vouchers. This study was funded by Flo Health UK LTD. The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008.

Recruitment

For both trials, computer and internet-literate participants were recruited using participant pools (SurveySwap, PureSpectrum, and Prolific) as well as via social media and mailing lists. Recruitment took place between July and December 2022, and potential participants were screened for eligibility via screening SurveyMonkey surveys (screening stage). It may have been possible for the same participant to respond under different email addresses, but these addresses would also need to be registered to the Flo app.

Inclusion criteria for trial-1 participants included being aged 18-40 years, being a US resident, being fluent in English, not being a current or former user of Flo app, having a cycle tracking health goal, scoring less than 80% on the cycle tracking health literacy quiz (see the *Measures* section), and scoring less than 80% on the general health and well-being survey (see the *Measures* section). Participants were excluded from trial 1 if they were currently using any form of hormonal contraception (eg, pill, injection, implant, and hormonal coil) as these would influence or stop their natural menstrual cycle. Participants in trial 1 were additionally excluded if they became pregnant during the study period.

Participants were included in trial 2 if they were older than 18 years, scored at least moderate on the Premenstrual Screening Tool (PSST) [42] (see the *Measures* section), scored less than 80% on the PMS or PMDD health literacy quiz (see the *Measures* section), and had never used Flo before. Further, participants were excluded from trial 2 if they were currently taking antidepressants or were currently or had previously received psychiatric treatment or had been diagnosed with a psychiatric disorder other than PMDD at the time of recruitment.

Participants were excluded from any trial if they had their last period more than 90 days prior or were pregnant at the time of recruitment.

Randomization and Blinding

If participants met inclusion criteria at the screening stage, they were then randomized by minimization using a Python script based on the following factors: cycle tracking or PMS or PMDD health literacy quiz scores, age, yearly household income in US $, highest level of education completed, race or ethnicity, and reproductive health disorders (Multimedia Appendix 2). Trial-1 participants were additionally minimized on general health and well-being scores, along with the number of children and number of pregnancies that did not result in a live birth, to control for differences in health literacy and awareness that

might result from previous experience of conception and pregnancy. Trial-2 participants were additionally minimized on PSST score and current day of cycle, to attempt to control for differences in PMS or PMDD symptom burden due to the menstrual cycle phase.

Participants were either allocated to the trial-specific intervention condition and instructed to use the Flo Premium app for 12 weeks (trial 1) and 3 months (trial 2) or to the waitlist control condition, which received the intervention at the end of the study. An example of the Flo app user interface is shown in Figure 1.

Figure 1. Example of the Flo user interface including the main screen, tiles linking to educational content, cycle day information, and the symptom logging interface.

Intervention Groups

Both trial participants were provided with detailed tutorials via email on how to use the Flo app and instructed to interact freely with the app for the whole duration of the study. All participants were provided with the premium version of the app for 1 year. In comparison to the free offering, premium app users have access to the full range of in-app educational content, all symptom- and cycle-related chatbots, as well as all community features.

Waitlist Control Groups

Participants who were randomized into the control group were not provided with access to the Flo app during the study period and were instructed not to use the Flo app in any form for the duration of the study. At the end of the study, control participants were provided with the premium version of the app for 1 year.

Measures

Demographics

Demographic information for both trials was collected using a demographic questionnaire, which asked about previous Flo use, US residency, age, race or ethnicity, household income and composition, education, gender identity, reproductive health, contraception use, and timing of last period (Multimedia Appendix 3). Trial-1 participants were additionally asked about English proficiency, parity, prior pregnancy outcomes, and reproductive health goals (cycle tracking, TTC, or pregnancy). Trial-2 participants were additionally asked about employment, present pregnancy status, use of antidepressants, diagnosis of a psychiatric disorder, past or present receival of any treatment for a psychiatric disorder, and perimenopause or menopause symptoms.

Primary Outcomes

Menstrual and PMS- or PMDD-Related Health Literacy

Menstrual-related health literacy was examined in trial-1 participants using a set of 12 questions regarding general knowledge about the menstrual cycle, such as typical bleeding length, how the length of the menstrual cycle is calculated, and how menstrual symptoms such as pain can be relieved (Multimedia Appendix 4). Knowledge about PMS or PMDD in trial-2 participants was measured through a set of questions investigating PMS and PMDD incidence, symptoms, and methods that can be used to help manage symptoms (Multimedia Appendix 5). The questions used were developed internally by Flo, using input from Flo Health's medical advisors.

Self-Reported Menstrual Health Awareness

Menstrual health awareness was only examined in the trial-1 participants using 7 questions selected from the 33-item Health Literacy Instrument for Adults [43] and modified to be specifically relevant to cycle tracking (Multimedia Appendix 6).

Self-Reported General Health and Well-Being

The trial-1 participants answered a set of 14 questions, selected both from domain 2 of the World Health Organization Quality of Life Brief Survey [44] and designed internally by Flo, using input from Flo Health's medical advisors. These questions assessed self-reported physical, mood, and menstrual symptoms experienced over the previous 3 months, as well as satisfaction with overall health, sex life, personal relationships, and support from friends (Multimedia Appendix 7).

PMS and PMDD Symptom Burden

PMS and PMDD symptom burden and the extent to which PMS or PMDD symptoms interfere with daily life were only assessed in trial-2 participants using the PSST [42] (Multimedia Appendix 8).

Secondary Outcomes

Overview

Feelings of control and management over health, communication and emotional well-being, menstrual cycle stigma (Menstrual Attitude Questionnaire [45]), fear of unplanned pregnancy, and body image and appreciation (Body Appreciation Scale-2) [46] were assessed (Multimedia Appendix 9) in trial-1 participants. Absenteeism due to PMS or PMDD symptoms, work productivity using the Stanford Presenteeism Scale-6 [47], and quality of life and satisfaction using the Quality of Life Enjoyment Satisfaction Questionnaire [48] were additionally assessed in trial-2 participants (Multimedia Appendix 10).

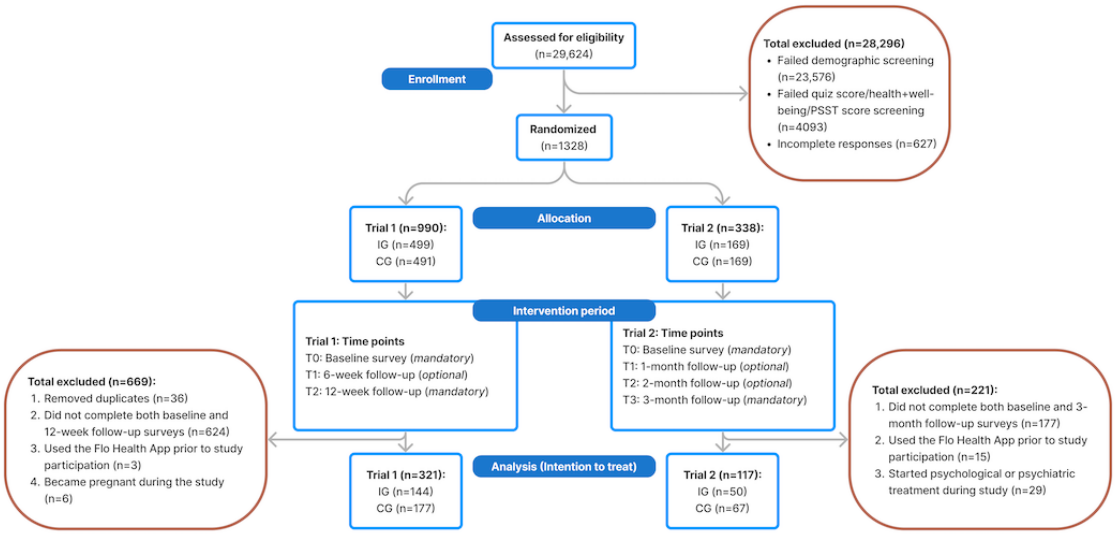
App Engagement

Data on app use by the intervention groups were also calculated, including the mean total number of sessions (log-ins), the mean length of sessions in seconds, and the mean number of days of app use (active days).

Data Collection Time Points

Trial-1 (cycle tracking) participants responded to questionnaires at 3 time points: screening or baseline, 6 weeks after baseline, and 12 weeks after baseline (Figure 2). The screening survey included demographic information, menstrual health literacy, and general health and well-being questions. The baseline and 6-week surveys included menstrual health awareness and all cycle tracking secondary outcome measures. The 12-week survey included the menstrual health literacy, general health and well-being questions, and all cycle tracking secondary outcome measures.

Figure 2. Participant flow diagram through trial 1 (cycle tracking) and trial 2 (premenstrual syndrome or premenstrual dysphoric disorder) intention-to-treat analysis. IG: intervention group; CG: control group; PSST: Premenstrual Screening Tool.



Trial-2 (PMS or PMDD) participants responded to survey questions at screening or baseline, 1 month after baseline, 2 months after baseline, and 3 months after baseline (Figure 2). The screening questionnaire included demographic information and primary outcomes. The baseline survey included all secondary outcome measures, while the month-1 and month-2 questionnaires included PMS or PMDD symptom burden, as well as all secondary outcome measures. The month-3

questionnaire included all primary and secondary outcome measures.

The difference in data collection time points between trials 1 and 2 was due to the need to measure PMS or PMDD symptoms monthly in trial 2.

Statistical Analyses

Data Exclusion

Any participant who did not meet the inclusion criteria based on responses from the screening survey was excluded from the study. To reduce the effect of bias due to noncompliance, we aimed to conduct an intention-to-treat (ITT) analysis and a per-protocol (PP) analysis. In the ITT analysis, participants were analyzed according to the group (intervention or control) they were randomized to, as long as the inclusion criteria were met and the screening, baseline, and final surveys were completed. In the PP analysis, all participants who completed all data collection waves and used the app at least once a month throughout the study duration were included.

Data Analysis

Main analysis of all outcomes was conducted using a Mixed Models for Repeated Measures (MMRM) approach. Models were constructed using an unstructured correlation where the outcome of interest was predicted by time point and condition status, including the interaction between time point and condition. Age, highest education level completed, and household income were included as fixed effects. Individual ID and time point were included as random effects. Preplanned pairwise contrasts between baseline outcome scores and final time point scores were used to assess change in outcome scores.

All data analyses were conducted using statistical and data science packages in R (version 4.3.0l R Foundation for Statistical Computing) [49] running on Mac OS X 13.3.1. Packages used included *arsenal* version 3.6.3 [50], *emmeans* version 1.8.6 [51], *ggpubr* version 0.6.0 [52], *gtools* version 3.9.4 [53], *gtsummary* version 1.7.1 [54], *longpower* version 1.0.24 [55-60], *mmrm* version 0.2.2 [61], *patchwork* version 1.1.2 [62], and *tidyverse* version 2.0.0 [63].

Power

Using MMRM analysis with 4 data collection waves as is the case in the PMS or PMDD group, an attrition rate of 10% per wave, and a SD of the outcome of interest of 2.5 points, the minimum sample size required to detect a change in score of 1 point with 80% power would be 126 participants in each group. Using 3 data collection waves as in the cycle tracking group, 115 participants assigned to each condition would be required. Due to difficulties in recruiting samples to meet these target group sizes, we converted both trials from full RCTs to feasibility pilot trials.

Results

Baseline Sample Characteristics

A total of 438 participants (194 intervention and 244 control) were included in the ITT sample, of which 321 were included in trial 1 (cycle tracking; 144 intervention and 177 control) and 117 were included in trial 2 (PMS or PMDD; 50 intervention and 67 control; Figure 2). The PP sample was comprised of 233 participants (43 intervention and 190 control), of which 174 were included from trial 1 (17 intervention and 157 control) and 59 were included from trial 2 (26 intervention and 33 control; Multimedia Appendix 11). Due to a low number of participants completing all protocol steps, and therefore being eligible for the PP analysis, we present the results and interpretation of the ITT analysis in the main text and avoid interpretation of our underpowered PP results. However, the results of the PP analyses can be found in Multimedia Appendices 8-17.

In the ITT trial-1 and trial-2 samples, we found no differences in demographic characteristics (including age, income, education, or race or ethnicity). At baseline, PMS- or PMDD-related health literacy scores were higher in the intervention group in trial 2. No other primary outcome measures differed between groups in either trial at baseline (Table 1). Demographic and baseline measures for the PP analyses are presented in Multimedia Appendix 12.

Participants in the ITT trial-1 intervention group had a mean session length of 234.45 (SD 181.98) seconds, a mean total number of sessions of 29.94 (SD 20.94), and a mean number of active days of 15.44 (SD 10.95). Participants in the ITT trial-2 intervention group had a mean session length of 146.98 (SD 122.85) seconds, a mean total number of sessions of 70.35 (SD 55.63), and a mean number of active days of 36.08 (SD 24.50). App engagement statistics for the PP sample are shown in Multimedia Appendix 13.

The mean time between completing the screening and baseline surveys was 8.56 (SD 9.86) days in the ITT trial 1 analysis and 9.45 (SD 9.18) days in the ITT trial 2 analysis.

Table 1. Intention-to-treat analysis demographic and summary statistics for primary outcome measures at baseline.

Characteristics	Trial 1 (cycle tracking)				Trial 2 (PMS[a] or PMDD[b])			
	Control (n=177)	Intervention (n=144)	Total (N=321)	*P* value	Control (n=67)	Intervention (n=50)	Total (N=117)	*P* value
Age group (years), n (%)				.82[c]				.26[c]
18-24	18 (10.2)	14 (9.7)	32 (10)		9 (13.4)	7 (14)	16 (13.7)	
25-34	81 (45.8)	71 (49.3)	152 (47.4)		16 (23.9)	20 (40)	36 (30.8)	
35-44	78 (44.1)	59 (41)	137 (42.7)		33 (49.3)	17 (34)	50 (42.7)	
45-54	0 (0)	0 (0)	0 (0)		9 (13.4)	6 (12)	15 (12.8)	
Race or ethnicity, n (%)				.43[d]				.86[d]
American Indian or Alaskan Native	2 (1.1)	5 (3.5)	7 (2.2)		3 (4.5)	0 (0)	3 (2.6)	
Asian or Asian American	8 (4.5)	9 (6.2)	17 (5.3)		5 (7.5)	2 (4)	7 (6)	
Biracial or multiracial	10 (5.6)	4 (2.8)	14 (4.4)		3 (4.5)	3 (6)	6 (5.1)	
Black or African American	30 (16.9)	20 (13.9)	50 (15.6)		6 (9)	7 (14)	13 (11.1)	
Hispanic, Latino, or Spanish origin	14 (7.9)	12 (8.3)	26 (8.1)		9 (13.4)	6 (12)	15 (12.8)	
Native Hawaiian or Other Pacific Islander	1 (0.6)	0 (0)	1 (0.3)		0 (0)	0 (0)	0 (0)	
White, European American, or Caucasian	110 (62.1)	94 (65.3)	204 (63.6)		41 (61.2)	32 (64.0)	73 (62.4)	
Other	2 (1.1)	0 (0)	2 (0.6)		0 (0)	0 (0)	0 (0)	
Household income (US $), n (%)				.36[c]				.20[c]
Missing data	1 (0.6)	5 (3.5)	6 (1.9)		0 (0)	0 (0)	0 (0)	
Under $15,000	30 (17)	20 (14.4)	50 (15.9)		4 (6.0)	6 (12)	10 (8.5)	
Between $15,000 and $29,999	33 (18.8)	22 (15.8)	55 (17.5)		9 (13.4)	8 (16)	17 (14.5)	
Between $30,000 and $49,999	40 (22.7)	46 (33.1)	86 (27.3)		14 (20.9)	10 (20)	24 (20.5)	
Between $50,000 and $74,999	37 (21)	16 (11.5)	53 (16.8)		18 (26.9)	14 (28)	32 (27.4)	
Between $75,000 and $99,999	23 (13.1)	13 (9.4)	36 (11.4)		10 (14.9)	6 (12)	16 (13.7)	
Between $100,000 and $150,000	7 (4)	17 (12.2)	24 (7.6)		9 (13.4)	4 (8)	13 (11.1)	
Over $150,000	6 (3.4)	5 (3.6)	11 (3.5)		3 (4.5)	2 (4)	5 (4.3)	
Highest education level, n (%)				.30[c]				.60[d]
Missing data	2 (1.1)	0 (0)	2 (0.6)		0 (0)	0 (0)	0 (0)	
Incomplete or complete secondary education	90 (51.4)	62 (43.1)	152 (47.6)		27 (40.3)	17 (34)	44 (37.6)	
Some postsecondary education, certificate, or associate degree	29 (16.6)	33 (22.9)	62 (19.4)		10 (14.9)	15 (30)	25 (21.4)	
Bachelor's degree or further education	56 (32)	49 (34)	105 (32.9)		30 (44.8)	18 (36)	48 (41)	
Health literacy				.90[e]				.03[e]

Characteristics	Trial 1 (cycle tracking)				Trial 2 (PMS[a] or PMDD[b])			
	Control (n=177)	Intervention (n=144)	Total (N=321)	*P* value	Control (n=67)	Intervention (n=50)	Total (N=117)	*P* value
Mean (SD)	9.045 (2.205)	9.014 (2.109)	9.031 (2.159)		8.866 (2.014)	9.680 (1.789)	9.214 (1.956)	
Range	3-14	2-13	2-14		4-12	6-12	4-12	
Menstrual awareness				.91[e]				N/A[f]
Mean (SD)	34.712 (6.178)	34.632 (5.926)	34.676 (6.057)		N/A	N/A	N/A	
Range	13-49	16-49	13-49		N/A	N/A	N/A	
Health and well-being				.25[c]				N/A
Mean (SD)	44.689 (10.398)	45.972 (9.206)	45.265 (9.887)		N/A	N/A	N/A	
Range	16-67	19-64	16-67		N/A	N/A	N/A	
PSST[g] score				N/A				.01[e]
Mean (SD)	N/A	N/A	N/A		33.119 (9.338)	37.280 (7.874)	34.897 (8.948)	
Range	N/A	N/A	N/A		9-54	15-54	9-54	

[a]PMS: premenstrual syndrome.

[b]PMDD: premenstrual dysphoric disorder.

[c]Trend test for ordinal variables.

[d]Pearson chi-square test.

[e]Linear model ANOVA.

[f]N/A: not applicable.

[g]PSST: Premenstrual Screening Tool.

Primary Outcomes

Menstrual Health Literacy (Trial 1)

Our MMRM analysis in the ITT trial-1 sample revealed significant main effects of time point (β=.523; *P*=.003) and highest education level (β=.488; *P*=.001) on menstrual health literacy, such that menstrual health literacy scores were higher postintervention and in individuals with higher levels of education (Table S1 in Multimedia Appendix 14). We found a significant group by time point interaction (β=.585; *P*=.02), such that greater improvement in menstrual health literacy was observed in the intervention group compared to the control group, with an estimated mean improvement between prescreen and the end of the intervention of 1.1 (SE 0.193) points in the intervention group and to 0.52 (SE 0.173) in the control group (Figure 3A; Table 2). There were no significant associations between menstrual health literacy scores and age group or household income. Similar improvements in estimated mean health literacy scores were seen in the PP analysis (Multimedia Appendix 15).

Figure 3. Estimated mean health literacy scores and 95% CIs from Mixed Models for Repeated Measures analysis at screening and end of intervention for the (A) intention-to-treat trial 1 (cycle tracking) and (B) intention-to-treat trial 2 (PMS or PMDD). PMDD: premenstrual dysphoric disorder; PMS: premenstrual syndrome.

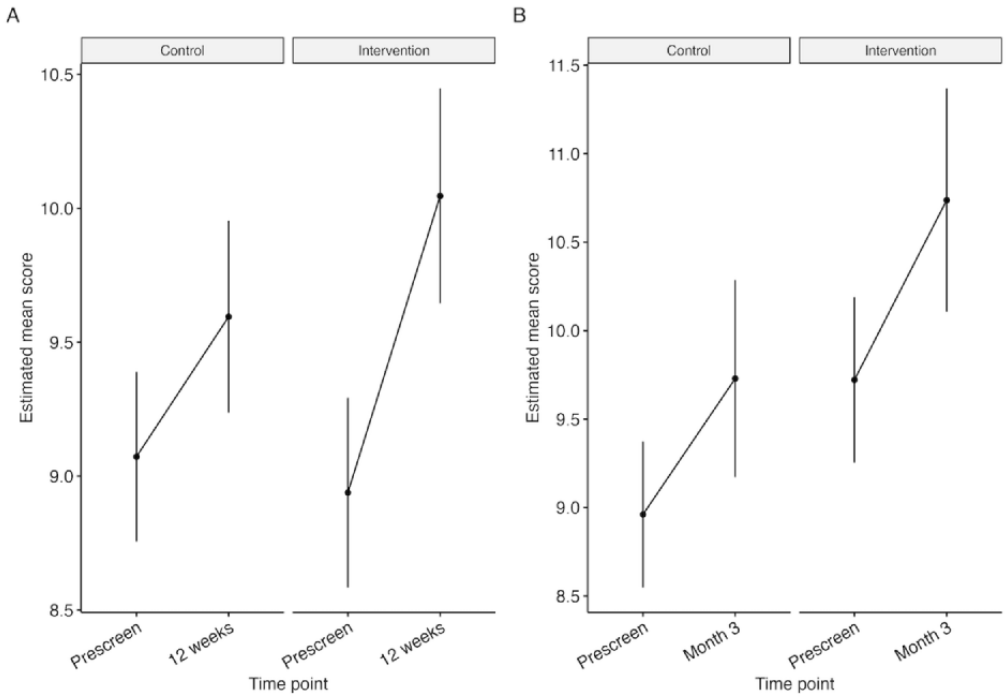


Table 2. Estimated mean differences in Menstrual Health literacy in the ITT[a] trial-1 analysis (cycle tracking) and PMS[b] or PMDD[c] health literacy scores in the ITT trial-2 analysis (PMS or PMDD) between baseline and the end of the intervention.

Trial	Control or intervention	Estimated mean difference (SE)	*T* ratio (*df*)	*P* value
Trial 1 (cycle tracking)	Control	0.523 (0.173)	3.026 (311)	.003
Trial 1 (cycle tracking)	Intervention	1.108 (0.193)	5.729 (311)	<.001
Trial 2 (PMS or PMDD)	Control	0.642 (0.276)	2.327 (115)	.02
Trial 2 (PMS or PMDD)	Intervention	1.200 (0.319)	3.758 (115)	<.001

[a]ITT: intention-to-treat.

[b]PMS: premenstrual syndrome.

[c]PMDD: premenstrual dysphoric disorder.

PMS- or PMDD-Related Health Literacy (Trial 2)

Estimated mean PMS or PMDD health literacy score also improved in both the ITT trial-2 intervention and control groups between baseline and end of intervention (Figure 3B; Table 2). We observed significant main effects of the highest education level (β=.430; *P*=.04), condition (β=.819; *P*=.03), and time point (β=.642; *P*=.02; Table S2 in Multimedia Appendix 14). Individuals with higher education had higher average health literacy scores, participants in the intervention group had higher average health literacy scores than controls at both time points, and health literacy scores increased after the intervention in both groups. No association between household income or age group and PMS or PMDD health literacy score was found. Estimated mean health literacy scores increased by 1.2 (SE 0.319) points in the intervention group compared to 0.64 (SE 0.276) in the control group (Table 2). Similar improvements were found in the PP analysis (Multimedia Appendix 15).

Self-Reported Menstrual Health Awareness (Trial 1)

In the ITT trial-1 analysis, we observed significant associations between menstrual health awareness and age group (β=1.053; *P*=.03) and time point (β=1.845; *P*<.001; Table S3 in Multimedia Appendix 14), where individuals in the older age groups had higher menstrual health awareness scores, and menstrual health awareness scores were higher after the intervention. We also found a significant interaction between treatment group and time point (β=2.126; *P*=.002), indicating that individuals in the intervention group had a greater improvement in menstrual health awareness scores (Figure 4). No association was found between highest education level, household income or treatment group, and menstrual health awareness scores. Estimated mean scores increased by around 4 (SE 0.515) points in the intervention group, as compared to 1.8 (SE 0.461) points in the control group (Table 3). Similar improvements in menstrual health awareness were observed in the PP analysis (Multimedia Appendix 15).

Figure 4. Estimated mean self-reported menstrual health awareness scores and 95% CIs from ITT MMRM analysis at baseline and 12 weeks after baseline for trial 1 (cycle tracking).

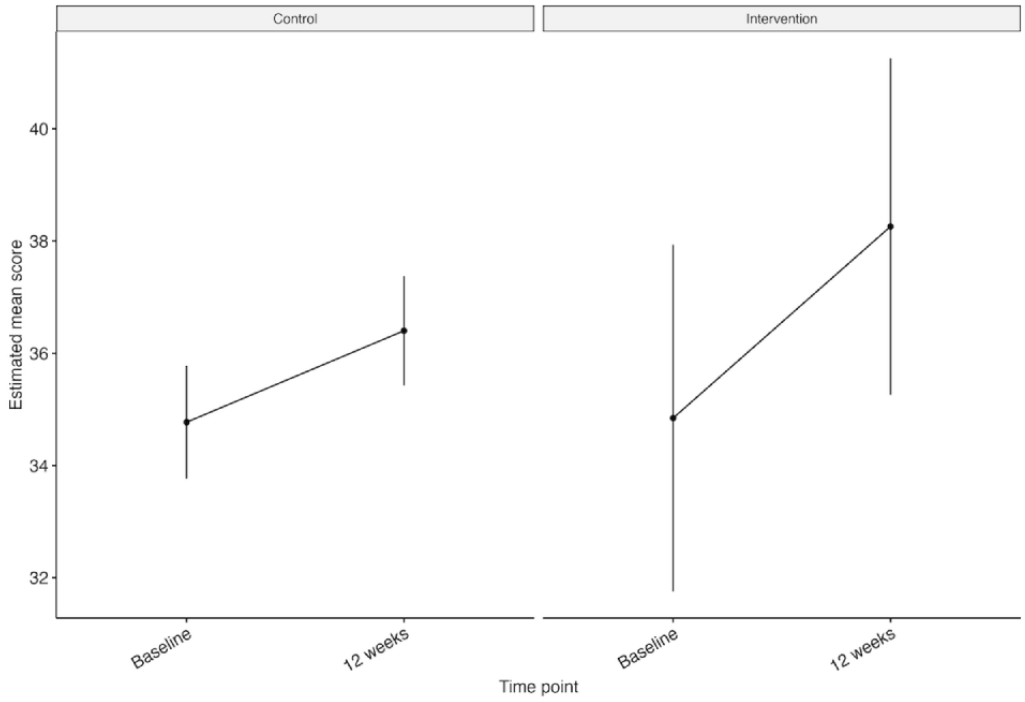


Table 3. Estimated mean differences in menstrual health awareness from trial 1 (cycle tracking).

Control or intervention	Estimated mean difference (SE)	*T* ratio (*df*=311)	*P* value
Control	1.845 (0.461)	4.005	<.001
Intervention	3.971 (0.515)	7.706	<.001

Self-Reported General Health and Well-Being (Trial 1)

Our ITT MMRM analysis revealed significant associations between household income and general health and well-being scores (β=1.444; *P*<.001; Table S4 in Multimedia Appendix 14), with individuals with higher income having better health and well-being scores. We found significant interactions between the treatment group and the 6-week time point (β=1.618; *P*=.03) and treatment group and 12-week time point (β=2.674; *P*=.001), indicating at both time points that individuals in the intervention group had greater improvement in health and well-being scores than in the control group. No association was found between age group, treatment group, or 6- and 12-week time points, and health and well-being scores. Estimated mean health and well-being scores improved by 3.44 (SE 0.579) points in the intervention group as compared to 0.76 (SE 0.517) points in the control group, with a significant improvement in estimated mean health and well-being scores between screening and the end of intervention observed in the intervention group alone (Figure 5; Table 4). Improvements in health and well-being scores were found in the intervention condition only in the PP analysis (Multimedia Appendix 15).

Figure 5. Estimated mean general health and well-being scores and 95% CIs from intention-to-treat Mixed Models for Repeated Measures approach analysis at screening, 6 weeks, and 12 weeks after baseline for trial 1 (cycle tracking).

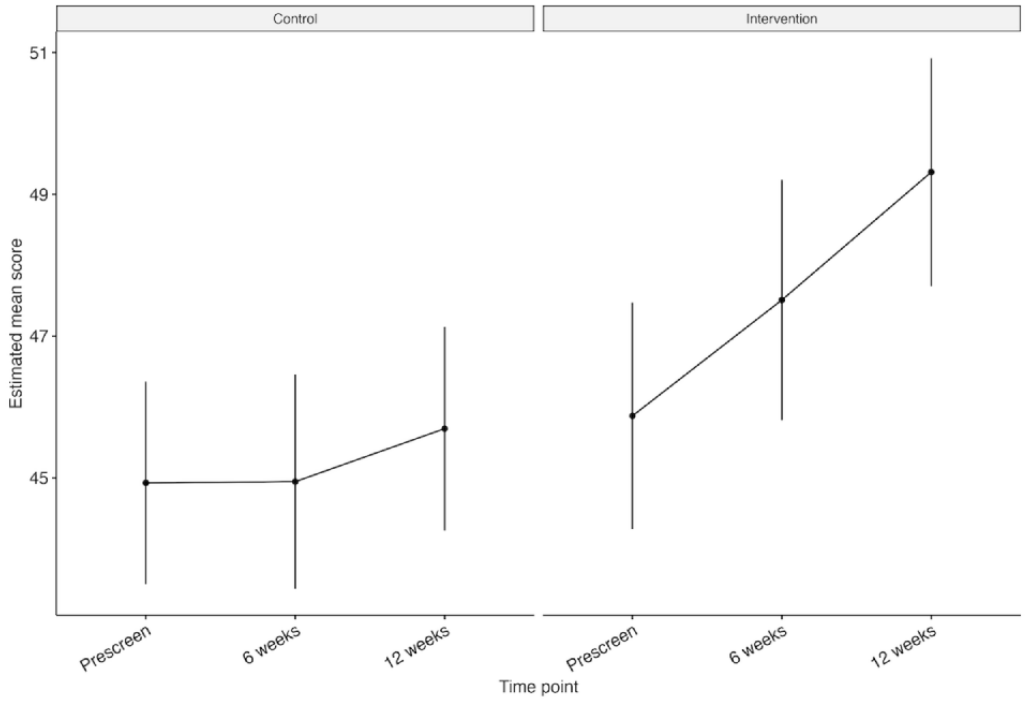


Table 4. Estimated mean differences in general health and well-being scores between screening and 12 weeks after baseline for the ITT[a] trial-1 (cycle tracking) analysis.

Control or intervention	Estimated mean difference (SE)	*T* ratio (*df*=311)	*P* value
Control	0.764 (0.517)	1.478	.14
Intervention	3.439 (0.579)	5.944	<.001

[a]ITT: intention-to-treat.

PMS or PMDD Symptom Burden (Trial 2)

We found significant main effects of group (β=4.094; *P*=.014), month-1 time point (β=–4.246; *P*<.001), month-2 time point (β=–5.133; *P*<.001), and month-3 time point (β=–5.239; *P*<.001; Table S5 in Multimedia Appendix 14). There were no significant interactions, and no associations were found between age group, highest education, household income, or treatment group with PMS or PMDD symptom burden. Estimated mean PSST scores improved by 7.1 (SE 1.300) points in the intervention group as compared to 5.2 (SE 1.123) points in the control group (Figure 6; Table 5). PMS or PMDD symptom burden also improved in both groups in the PP analysis (Multimedia Appendix 15).

Figure 6. Intention-to-treat Mixed Models for Repeated Measures estimated mean Premenstrual Screening Tool scores at each timepoint for intention-to-treat trial 2 (premenstrual syndrome or premenstrual syndrome).

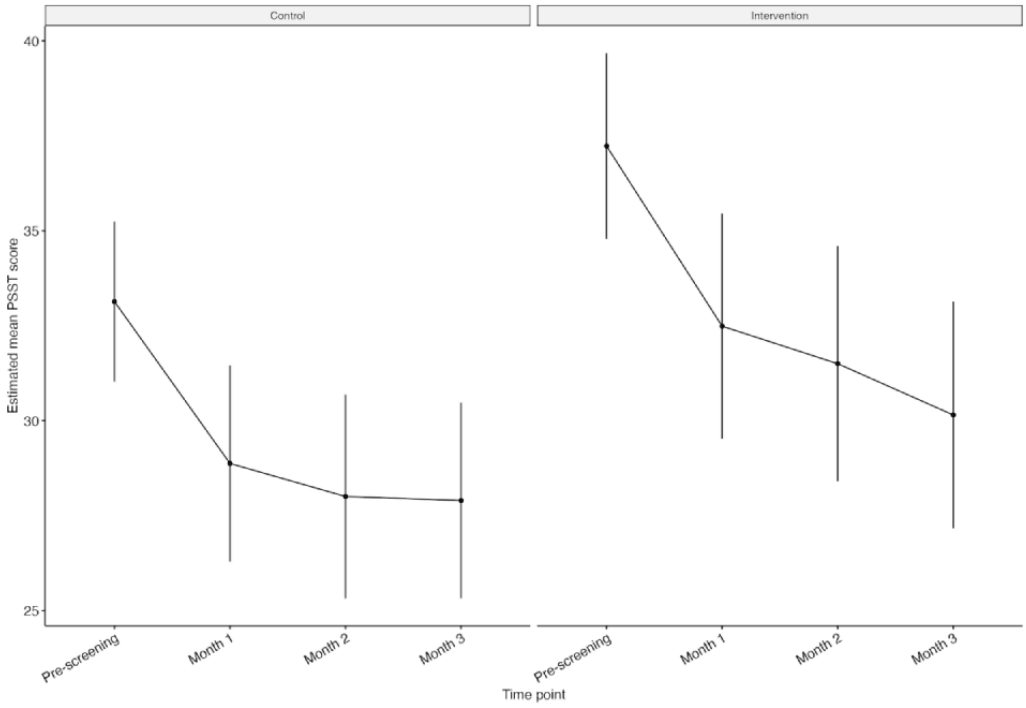


Table 5. Estimated mean differences in PSST[a] score between screening and month 3 for the ITT[b] trial-2 (PMS or PMDD) analysis.

Condition	Estimated mean difference (SE)	*T* ratio (*df*=115)	*P* value
Control	–5.239 (1.123)	–4.663	<.001
Intervention	–7.080 (1.300)	–5.444	<.001

[a]PSST: Premenstrual Screening Tool.

[b]ITT: intention-to-treat.

Secondary Outcomes

In trial 1, we observed improvements in many of our secondary outcomes after 12 weeks of using Flo in our ITT analyses. Communication and emotional score improved only in the intervention group by 0.93 points (*P*=.02). Similarly, the ratings of menstrual cycle stigma and fear of unplanned pregnancy improved by 0.61 points (*P*=.01) and 0.22 points (*P*=.04), respectively, in the intervention group only. Finally, the score of feeling of control and management of health improved in both the intervention and control groups after using Flo for 12 weeks by 1.00 (*P*<.001) and 0.45 points (*P*=.01), respectively. However, we saw no improvement in body image scores in either the intervention or control groups.

In the trial-2 ITT analysis, we observed improvements in absenteeism due to PMS or PMDD symptoms, with participants in the intervention group taking fewer absent days from work or education (estimated mean difference=–1.67, SE 0.319; *P*=.04), while no improvement was seen in the control group. We also saw no improvement in quality of life or work productivity scores in either of the trial-2 intervention or control groups.

Full ITT analysis results for trial-1 and trial-2 secondary outcomes are presented in Multimedia Appendix 16, while full PP results are presented in Multimedia Appendix 17.

Discussion

Principal Findings

In this study, we present the findings of 2 pilot, 2-armed RCTs on the effect of using the Flo app in (1) individuals who track their cycles and (2) individuals who track their cycles and are affected by PMS or PMDD. Our ITT analysis demonstrated improvements postintervention in all primary outcomes, namely health literacy and awareness, health and well-being, and PMS or PMDD symptom burden. As for secondary outcomes, positive effects were observed for control and management over health, communication about health, menstrual cycle stigma, and fear of unplanned pregnancies for those who track their cycles as well as absenteeism for PMS or PMDD. Quality of life, body image, and workplace productivity did not show improvement.

Comparison With Prior Work

The primary hypothesis of this study was that using the Flo app would increase health literacy and awareness around menstrual and reproductive health. We found that menstrual (trial 1) and PMS or PMDD (trial 2) health literacy scores as well as

menstrual health awareness (trial 1) significantly improved in both intervention and control groups, albeit with larger improvements in the intervention groups.

These results are consistent with preliminary observational data, which find that the Flo app can help women and people who menstruate feel more prepared throughout their menstrual cycles, helping them understand their bodies and symptoms [36,64]. These results are also consistent with previous research finding that psychoeducational and information-based interventions can improve health literacy around PMS and PMDD [37,38] and symptom scores in PMS [65]. Low health literacy has a negative impact on both individuals and health care systems alike. Poor health literacy has been associated with poorer well-being [66] and self-care [67] while contributing to higher health care costs [68], resulting from underuse of preventive services [69]. Further, the lack of knowledge and low symptom awareness often lead to delays in seeking care [70], leading to worse health outcomes. Thus, easily accessible tools addressing reproductive and menstrual health literacy have the potential to fill in the gaps in current provisioning.

Health and well-being scores (measured via a shorter version of the World Health Organization Quality of Life Brief Survey) improved in the trial-1 cycle tracking intervention group but not in the control group. PMS or PMDD symptom severity ratings (indicated by a drop in PSST scores at the final time point compared to baseline) improved in both the control and intervention groups, with a larger improvement in the intervention group. These findings are in line with previous research demonstrating that educational programs can help reduce the negative impact of menstrual and PMS or PMDD symptoms on women's lives [71,72]. Tracking symptoms over time and identifying patterns may help individuals feel better prepared and equipped to deal with their symptoms and manage their conditions. It is also possible that the observed improvements in health and well-being may have been driven by interaction with community-based tools such as Flo's Secret Chats, which could help individuals normalize their menstrual cycle experiences and decrease feelings of loneliness or social exclusion [29,30,33,73,74]. Improvements in PMS or PMDD symptom burden observed in the control condition could be due to participants using the app without our knowledge, motivation through exposure to study materials to think about how they may better manage any symptoms they experience, or use of other out-of-app menstrual health and PMS or PMDD resources to improve symptom management.

While the lack of existing literature on the efficacy of digital health interventions for menstrual health makes comparative analysis difficult, the magnitude of changes observed in our primary outcomes is comparable both to a previous RCT for a PMS educational app [65] and to app-based interventions for other health conditions such as sleep difficulties [75] or depression and anxiety [76,77].

In terms of secondary outcomes, questions addressing feelings of control over one's own reproductive health showed improvements in both cycle tracking intervention and control groups, with a stronger effect observed in the intervention groups. Again, improvements in the control group are likely due to unidentified or anonymous use of the app (using the anonymous mode) by control group participants or motivation to seek out external resources and think about health management. Positive effects were observed in the cycle tracking intervention group only for outcomes including communication about health, stigma around menstrual health, and fears of unplanned pregnancies. Being able to effectively communicate with health care providers and share feelings and concerns about one's own health are particularly important for menstrual and reproductive health. The findings from a large governmental investigation (UK Women's Health Strategy Survey) highlighted how women and people who menstruate do not feel heard when talking to their health care provider, with 1 in 4 declaring not feeling comfortable sharing information about their menstrual cycle [12]. This could contribute to delays in diagnosis as women may not be bringing up concerning symptoms in a timely manner. Further, the finding that feelings of stigma were reduced highlights the potential of solutions like Flo app to help normalize discussions around symptoms and overall reproductive and menstrual health. Finally, fears of unplanned pregnancies may have been reduced due to the exposure to content aimed at educating users about effective contraception methods. Absenteeism was measured in the PMS or PMDD group and showed a reduced number of days taken in the intervention group only. This finding is in line with previous research highlighting the impact of conditions of the menstrual cycle on workplace productivity [36,78-81] and further suggests that psychoeducational tools may help individuals better manage their symptoms [64].

We did not observe improvements in other secondary outcomes. This could be due to a number of factors. First, while we hypothesized that better knowledge and improved management of the menstrual cycle would translate into improvements in quality of life, workplace productivity, and body image, these are complex and multifaceted outcomes. Each of these outcomes is heavily influenced by health or social and economic factors beyond those relevant to menstrual health that Flo app is not targeted to address fully. Second, while 3-month interventions tend to be the norm in the digital health field [82], quality of life, workplace productivity, and body appreciation may require longer interventions for smaller effects to be detectable and sustained over time. Finally, it has been shown that studies targeting multifactorial health behaviors tend to be less efficacious [29]. Due to this, further research, with a better-powered sample and longer timeframes, is needed to explore whether solutions like Flo app can help tackle these issues.

Limitations and Future Research

This study does have limitations. First, as stated above, we cannot be sure that participants in the control group did not use Flo app. While our ITT analyses excluded individuals in the control condition who we identified as using the app before or during the study period, this identification was carried out using email addresses provided by study participants. If control participants signed up for the app using a different email address than the one provided to study staff, or used the Anonymous Mode feature, we would not be able to identify these individuals and exclude them on the basis of app use. Hence, any

improvements in the control condition of both trials could be due to the fact that some control participants did use the Flo app, despite instructions not to. Further, participants in the control group may also have been motivated to look for other sources of knowledge on reproductive health by virtue of taking part in this study, hence improving their final scores.

Our primary findings are based on an ITT analysis, which should be considered conservative, as we included all intervention participants with some exposure to the intervention. This means that any null findings could be due to noncompliance, or low usage of the app in intervention participants. Encouragingly, however, our PP analyses do broadly show similar improvements in outcomes as in the ITT analyses.

Additionally, the intervention period was only 3 months. While this was appropriate for this kind of pilot RCT, a longer interventional period may have allowed for the detection of additional effects, particularly in the secondary and exploratory outcomes.

Furthermore, due to participant dropout, we did not achieve an adequately powered sample for the PMS or PMDD ITT analysis and all PP analyses, and therefore these results should be considered as needing further validation. Follow-up studies should investigate methods of reducing participant dropout and improving rates of completion of measures. As a remote study, some attrition is expected, but closer monitoring of participants, additional reminders or prompts to complete data collection, along with personalized and improved incentives could allow for a greater proportion of participants to fully complete the study. In-person recruitment and data collection should also be considered to aid data quality and completeness and prevent app use by participants in the control group.

Finally, the study population was limited to individuals from the United States who were English speakers and may not be fully representative of other populations worldwide. Future research should prioritize inclusion of a diverse study population by including participants from different cultural, linguistic and socioeconomic backgrounds in order to allow for robust and generalizable results.

Conclusions

Low menstrual and reproductive health literacy and awareness can have a significant negative impact on the overall health and well-being of individuals. Further, health literacy and awareness of one's own symptoms can greatly influence the course of conditions of the menstrual cycle and help improve knowledge of how individuals can manage them. Digital tools such as the Flo app are widely accessible and provide users with a wealth of content and tools that can be successfully turned into actionable insights. The findings from this study show how increased health literacy, symptom tracking, and exposure to a community of like-minded individuals can improve menstrual and reproductive health knowledge, symptom management, and overall health outcomes.

Acknowledgments

The authors thank the participants who took part in the study.

Conflicts of Interest

SP, ACC, AW, KP, CP, TR, RB, LZ, and AK are employees of Flo Health. LH and JP are consultants for Flo Health. SP, ACC, AW, KP, AK, and LZ hold equity interests in Flo Health.

Multimedia Appendix 1
CONSORT checklist.
[PDF File (Adobe PDF File), 1075 KB - mhealth_v12i1e54124_app1.pdf]

Multimedia Appendix 2
Trial-1 and trial-2 minimization and randomization categories.
[DOCX File , 8 KB - mhealth_v12i1e54124_app2.docx]

Multimedia Appendix 3
Trial-1 and trial-2 demographic questions.
[DOCX File , 9 KB - mhealth_v12i1e54124_app3.docx]

Multimedia Appendix 4
Trial 1 specific health literacy quiz questions.
[DOCX File , 9 KB - mhealth_v12i1e54124_app4.docx]

Multimedia Appendix 5
Trial 2 specific health literacy quiz questions.
[DOCX File , 8 KB - mhealth_v12i1e54124_app5.docx]

Multimedia Appendix 6
Trial-1 menstrual health awareness questions.
[DOCX File , 7 KB - mhealth_v12i1e54124_app6.docx]

Multimedia Appendix 7
Trial-1 general health and well-being questions.
[DOCX File , 7 KB - mhealth_v12i1e54124_app7.docx]

Multimedia Appendix 8
Trial-2 PMS or PMDD symptom burden (PSST score). PMDD: premenstrual dysphoric disorder; PMS: premenstrual syndrome; PSST: Premenstrual Screening Tool.
[DOCX File , 7 KB - mhealth_v12i1e54124_app8.docx]

Multimedia Appendix 9
Trial-1 secondary outcomes.
[DOCX File , 8 KB - mhealth_v12i1e54124_app9.docx]

Multimedia Appendix 10
Trial-2 secondary outcomes.
[DOCX File , 8 KB - mhealth_v12i1e54124_app10.docx]

Multimedia Appendix 11
Flow chart of participants in the per-protocol analysis.
[DOCX File , 215 KB - mhealth_v12i1e54124_app11.docx]

Multimedia Appendix 12
Per-protocol analysis demographics and baseline summary statistics for primary outcome measures.
[DOCX File , 13 KB - mhealth_v12i1e54124_app12.docx]

Multimedia Appendix 13
App engagement statistics for per-protocol analyses.
[DOCX File , 7 KB - mhealth_v12i1e54124_app13.docx]

Multimedia Appendix 14
Full MMRM model results for trial-1 and trial-2 intention-to-treat analyses. MMRM: Mixed Models for Repeated Measures approach.
[DOCX File , 23 KB - mhealth_v12i1e54124_app14.docx]

Multimedia Appendix 15
Per-protocol estimated mean differences in primary outcome measures for trials 1 and 2.
[DOCX File , 8 KB - mhealth_v12i1e54124_app15.docx]

Multimedia Appendix 16
Intention-to-treat estimated mean differences in secondary outcomes for trials 1 and 2.
[DOCX File , 9 KB - mhealth_v12i1e54124_app16.docx]

Multimedia Appendix 17
Per-protocol estimated mean differences in secondary outcomes for trials 1 and 2.
[DOCX File , 9 KB - mhealth_v12i1e54124_app17.docx]

References

1. Lundsberg LS, Pal L, Gariepy AM, Xu X, Chu MC, Illuzzi JL. Knowledge, attitudes, and practices regarding conception and fertility: a population-based survey among reproductive-age United States women. Fertil Steril 2014;101(3):767-774 [FREE Full text] [doi: 10.1016/j.fertnstert.2013.12.006] [Medline: 24484995]
2. Ayoola AB, Zandee GL, Adams YJ. Women's knowledge of ovulation, the menstrual cycle, and its associated reproductive changes. Birth 2016;43(3):255-262. [doi: 10.1111/birt.12237] [Medline: 27157718]

3. Hennegan J. Inserting informed choice into global menstrual product use and provision. Lancet Public Health 2019;4(8):e361-e362 [FREE Full text] [doi: 10.1016/S2468-2667(19)30126-4] [Medline: 31324420]
4. Belayneh Z, Mekuriaw B. Knowledge and menstrual hygiene practice among adolescent school girls in southern Ethiopia: a cross-sectional study. BMC Public Health 2019;19(1):1595 [FREE Full text] [doi: 10.1186/s12889-019-7973-9] [Medline: 31783741]
5. Mohammed S, Larsen-Reindorf RE. Menstrual knowledge, sociocultural restrictions, and barriers to menstrual hygiene management in Ghana: evidence from a multi-method survey among adolescent schoolgirls and schoolboys. PLoS One 2020;15(10):e0241106 [FREE Full text] [doi: 10.1371/journal.pone.0241106] [Medline: 33091080]
6. Tshomo T, Gurung MS, Shah S, Gil-Cuesta J, Maes P, Wangdi R, et al. Menstrual hygiene management-knowledge, attitudes, and practices among female college students in Bhutan. Front Reprod Health 2021;3:703978 [FREE Full text] [doi: 10.3389/frph.2021.703978] [Medline: 36303967]
7. Lahme AM, Stern R, Cooper D. Factors impacting on menstrual hygiene and their implications for health promotion. Glob Health Promot 2018;25(1):54-62. [doi: 10.1177/1757975916648301] [Medline: 27380769]
8. Wall LL, Teklay K, Desta A, Belay S. Tending the 'monthly flower:' a qualitative study of menstrual beliefs in Tigray, Ethiopia. BMC Womens Health 2018;18(1):183 [FREE Full text] [doi: 10.1186/s12905-018-0676-z] [Medline: 30424763]
9. Barrington DJ, Robinson HJ, Wilson E, Hennegan J. Experiences of menstruation in high income countries: a systematic review, qualitative evidence synthesis and comparison to low- and middle-income countries. PLoS One 2021;16(7):e0255001 [FREE Full text] [doi: 10.1371/journal.pone.0255001] [Medline: 34288971]
10. Armour M, Parry K, Al-Dabbas MA, Curry C, Holmes K, MacMillan F, et al. Self-care strategies and sources of knowledge on menstruation in 12,526 young women with dysmenorrhea: a systematic review and meta-analysis. PLoS One 2019;14(7):e0220103 [FREE Full text] [doi: 10.1371/journal.pone.0220103] [Medline: 31339951]
11. Rubinsky V, Gunning JN, Cooke-Jackson A. "I thought I was dying:" (un)supportive communication surrounding early menstruation experiences. Health Commun 2020;35(2):242-252. [doi: 10.1080/10410236.2018.1548337] [Medline: 30460861]
12. Results of the 'women's health—let's talk about it' survey. GOV.UK. 2022. URL: https://tinyurl.com/msdm9cat [accessed 2023-09-13]
13. Direkvand-Moghadam A, Sayehmiri K, Delpisheh A, Sattar K. Epidemiology of Premenstrual Syndrome (PMS)-a systematic review and meta-analysis study. J Clin Diagn Res 2014;8(2):106-109 [FREE Full text] [doi: 10.7860/JCDR/2014/8024.4021] [Medline: 24701496]
14. Halbreich U, Borenstein J, Pearlstein T, Kahn LS. The prevalence, impairment, impact, and burden of premenstrual dysphoric disorder (PMS/PMDD). Psychoneuroendocrinology 2003;28(Suppl 3):1-23. [doi: 10.1016/s0306-4530(03)00098-2] [Medline: 12892987]
15. Walsh S, Ismaili E, Naheed B, O'Brien S. Diagnosis, pathophysiology and management of premenstrual syndrome. Obstet Gynaecol 2015;17(2):99-104. [doi: 10.1111/tog.12180]
16. Lustyk MKB, Gerrish WG. Premenstrual syndrome and premenstrual dysphoric disorder: issues of quality of life, stress and exercise. In: Preedy VR, Watson RR, editors. Handbook of Disease Burdens and Quality of Life Measures. New York, NY: Springer; 2010:1951-1975.
17. Heinemann LAJ, Do Minh T, Filonenko A, Uhl-Hochgräber K. Explorative evaluation of the impact of premenstrual disorder on daily functioning and quality of life. Patient 2010;3(2):125-132. [doi: 10.2165/11533750-000000000-00000] [Medline: 22273361]
18. Yang M, Wallenstein G, Hagan M, Guo A, Chang J, Kornstein S. Burden of premenstrual dysphoric disorder on health-related quality of life. J Womens Health (Larchmt) 2008;17(1):113-121. [doi: 10.1089/jwh.2007.0417] [Medline: 18240988]
19. Dean BB, Borenstein JE. A prospective assessment investigating the relationship between work productivity and impairment with premenstrual syndrome. J Occup Environ Med 2004;46(7):649-656. [doi: 10.1097/01.jom.0000131796.62115.84] [Medline: 15247803]
20. Chawla A, Swindle R, Long S, Kennedy S, Sternfeld B. Premenstrual dysphoric disorder: is there an economic burden of illness? Med Care 2002;40(11):1101-1112. [doi: 10.1097/01.MLR.0000032191.26152.90] [Medline: 12409855]
21. Rapkin AJ, Winer SA. Premenstrual syndrome and premenstrual dysphoric disorder: quality of life and burden of illness. Expert Rev Pharmacoecon Outcomes Res 2009;9(2):157-170. [doi: 10.1586/erp.09.14] [Medline: 19402804]
22. Reid RL. Premenstrual dysphoric disorder (formerly premenstrual syndrome). In: Feingold KR, Anawalt B, Blackman MR, Boyce A, Chrousos G, Corpas E, et al, editors. Endotext. South Dartmouth, MA: MDText.com, Inc; 2000.
23. Ghant MS, Sengoba KS, Vogelzang R, Lawson AK, Marsh EE. An altered perception of normal: understanding causes for treatment delay in women with symptomatic uterine fibroids. J Womens Health (Larchmt) 2016;25(8):846-852 [FREE Full text] [doi: 10.1089/jwh.2015.5531] [Medline: 27195902]
24. Simpson CN, Lomiguen CM, Chin J. Combating diagnostic delay of endometriosis in adolescents via educational awareness: a systematic review. Cureus 2021;13(5):e15143 [FREE Full text] [doi: 10.7759/cureus.15143] [Medline: 34164243]
25. Mirin AA. Gender disparity in the funding of diseases by the U.S. National Institutes of Health. J Womens Health (Larchmt) 2021;30(7):956-963 [FREE Full text] [doi: 10.1089/jwh.2020.8682] [Medline: 33232627]

26. Evans RL, Harris B, Onuegbu C, Griffiths F. Systematic review of educational interventions to improve the menstrual health of young adolescent girls. BMJ Open 2022;12(6):e057204 [FREE Full text] [doi: 10.1136/bmjopen-2021-057204] [Medline: 35676001]
27. Hennegan J, Montgomery P. Do menstrual hygiene management interventions improve education and psychosocial outcomes for women and girls in low and middle income countries? a systematic review. PLoS One 2016;11(2):e0146985 [FREE Full text] [doi: 10.1371/journal.pone.0146985] [Medline: 26862750]
28. Carolan S, Harris PR, Cavanagh K. Improving employee well-being and effectiveness: systematic review and meta-analysis of web-based psychological interventions delivered in the workplace. J Med Internet Res 2017;19(7):e271 [FREE Full text] [doi: 10.2196/jmir.7583] [Medline: 28747293]
29. Howarth A, Quesada J, Silva J, Judycki S, Mills PR. The impact of digital health interventions on health-related outcomes in the workplace: a systematic review. Digit Health 2018;4:2055207618770861 [FREE Full text] [doi: 10.1177/2055207618770861] [Medline: 29942631]
30. Babbar K, Martin J, Ruiz J, Parray AA, Sommer M. Menstrual health is a public health and human rights issue. Lancet Public Health 2022;7(1):e10-e11 [FREE Full text] [doi: 10.1016/S2468-2667(21)00212-7] [Medline: 34717798]
31. Bradbury A. Mental health stigma: the impact of age and gender on attitudes. Community Ment Health J 2020;56(5):933-938. [doi: 10.1007/s10597-020-00559-x] [Medline: 32002711]
32. Horsfall J, Cleary M, Hunt GE. Stigma in mental health: clients and professionals. Issues Ment Health Nurs 2010;31(7):450-455. [doi: 10.3109/01612840903537167] [Medline: 20521914]
33. Johnston-Robledo I, Chrisler JC. The menstrual mark: menstruation as social stigma. Sex Roles 2011;68(1-2):9-18. [doi: 10.1007/s11199-011-0052-z]
34. Ponzo S, Morelli D, Kawadler JM, Hemmings NR, Bird G, Plans D. Efficacy of the digital therapeutic mobile app BioBase to reduce stress and improve mental well-being among university students: randomized controlled trial. JMIR Mhealth Uhealth 2020;8(4):e17767 [FREE Full text] [doi: 10.2196/17767] [Medline: 31926063]
35. Cook RJ, Dickens BM. Reducing stigma in reproductive health. Int J Gynaecol Obstet 2014;125(1):89-92 [FREE Full text] [doi: 10.1016/j.ijgo.2014.01.002] [Medline: 24513258]
36. Ponzo S, Wickham A, Bamford R, Radovic T, Zhaunova L, Peven K, et al. Menstrual cycle-associated symptoms and workplace productivity in US employees: a cross-sectional survey of users of the Flo mobile phone app. Digit Health 2022;8:20552076221145852 [FREE Full text] [doi: 10.1177/20552076221145852] [Medline: 36544535]
37. Weise C, Kaiser G, Janda C, Kues JN, Andersson G, Strahler J, et al. Internet-based cognitive-behavioural intervention for women with premenstrual dysphoric disorder: a randomized controlled trial. Psychother Psychosom 2019;88(1):16-29. [doi: 10.1159/000496237] [Medline: 30783069]
38. Borji-Navan S, Mohammad-Alizadeh-Charandabi S, Esmaeilpour K, Mirghafourvand M, Ahmadian-Khooinarood A. Internet-based cognitive-behavioral therapy for premenstrual syndrome: a randomized controlled trial. BMC Womens Health 2022;22(1):5 [FREE Full text] [doi: 10.1186/s12905-021-01589-7] [Medline: 34996424]
39. Flo. URL: https://flo.health/ [accessed 2024-04-08]
40. Efficacy of the Flo app in improving health literacy, mood, quality of life and productivity at work in women with moderate to severe premenstrual symptoms: a randomized controlled trial. OSF. 2022. URL: https://osf.io/pz9ks/?view_only=346986293449442582 1d345e70469788 [accessed 2023-10-16]
41. Flo reproductive health tracking application use and women's reproductive health, knowledge, and well-being. OSF Registries. URL: https://osf.io/pcgw7?view_only=f0de2708daa040ec94a813ba88975797 [accessed 2023-10-16]
42. Steiner M, Macdougall M, Brown E. The Premenstrual Symptoms Screening Tool (PSST) for clinicians. Arch Womens Ment Health 2003;6(3):203-209. [doi: 10.1007/s00737-003-0018-4] [Medline: 12920618]
43. Tavousi M, Haeri-Mehrizi A, Rakhshani F, Rafiefar S, Soleymanian A, Sarbandi F, et al. Development and validation of a short and easy-to-use instrument for measuring health literacy: the Health Literacy Instrument for Adults (HELIA). BMC Public Health 2020;20(1):656 [FREE Full text] [doi: 10.1186/s12889-020-08787-2] [Medline: 32397970]
44. WHOQOL-BREF. World Health Organization. URL: https://www.who.int/tools/whoqol/whoqol-bref [accessed 2023-09-12]
45. Brooks-Gunn J, Ruble DN. The menstrual attitude questionnaire. Psychosom Med 1980;42(5):503-512. [doi: 10.1097/00006842-198009000-00005] [Medline: 7465737]
46. Tylka TL, Wood-Barcalow NL. The body appreciation scale-2: item refinement and psychometric evaluation. Body Image 2015;12:53-67. [doi: 10.1016/j.bodyim.2014.09.006] [Medline: 25462882]
47. Koopman C, Pelletier KR, Murray JF, Sharda CE, Berger ML, Turpin RS, et al. Stanford presenteeism scale: health status and employee productivity. J Occup Environ Med 2002;44(1):14-20. [doi: 10.1097/00043764-200201000-00004] [Medline: 11802460]
48. Endicott J, Nee J, Harrison W, Blumenthal R. Quality of life enjoyment and satisfaction questionnaire: a new measure. Psychopharmacol Bull 1993;29(2):321-326. [Medline: 8290681]
49. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing URL: https://www.R-project.org [accessed 2024-03-22]
50. Heinzen E, Sinnwell J, Atkinson E, Gunderson T, Dougherty G. Arsenal: an Arsenal of `R' functions for large-scale statistical summaries. CRAN. 2021. URL: https://CRAN.R-project.org/package=arsenal [accessed 2024-03-22]

51. Lenth RV. emmeans: estimated marginal means, aka least-squares means. CRAN. 2023. URL: https://tinyurl.com/5n7n8c7n [accessed 2024-03-22]
52. Kassambara A. ggpubr: `ggplot2' based publication ready plots. CRAN. 2023. URL: https://tinyurl.com/yjkp7mh6 [accessed 2024-03-22]
53. Bolker B, Warnes GR, Lumley T. gtools: various R Programming tools. CRAN. 2022. URL: https://tinyurl.com/2rhxueds [accessed 2024-03-22]
54. Sjoberg DD, Whiting K, Curry M, Lavery JA, Larmarange J. Reproducible summary tables with the gtsummary package. R J 2021;13(1):570-580 [FREE Full text] [doi: 10.32614/RJ-2021-053]
55. Diggle PJ, Heagerty P, Liang KY, Zeger SL. Analysis of longitudinal data. Oxford Statistical Science Series, 2nd Edition. Oxford: Oxford University Press; 2002.
56. Donohue MC, Gamst AC, Edland SD. longpower: power and sample size calculations for linear mixed models. Version R Package Version 2023:23.
57. Hu N, Mackey H, Thomas R. Power and sample size for random coefficient regression models in randomized experiments with monotone missing data. Biom J 2021;63(4):806-824. [doi: 10.1002/bimj.202000184] [Medline: 33586212]
58. Iddi S, Donohue MC. Power and sample size for longitudinal models in R—the longpower package and Shiny app. R J 2022;14(1):264-282 [FREE Full text] [doi: 10.32614/rj-2022-022]
59. Liu G, Liang KY. Sample size calculations for studies with correlated observations. Biometrics 1997;53(3):937-947. [Medline: 9290224]
60. Lu K, Luo X, Chen PY. Sample size estimation for repeated measures analysis in randomized clinical trials with missing data. Int J Biostat 2008;4(1):9 [FREE Full text] [doi: 10.2202/1557-4679.1098] [Medline: 22462117]
61. Bove DS, Dedic J, Kelkhoff D, Kunzmann K, Lang BM, Li L, et al. mmrm: mixed models for repeated measures. CRAN. 2022. URL: https://CRAN.R-project.org/package=mmrm [accessed 2024-03-22]
62. Pedersen TL. patchwork: the composer of plots. CRAN. 2022. URL: https://CRAN.R-project.org/package=patchwork [accessed 2024-03-22]
63. Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, et al. Welcome to the Tidyverse. J Open Source Softw 2019;4(43):1686 [FREE Full text] [doi: 10.21105/joss.01686]
64. Zhaunova L, Bamford R, Radovic T, Wickham A, Peven K, Croft J, et al. Characterization of self-reported improvements in knowledge and health among users of Flo period tracking app: cross-sectional survey. JMIR Mhealth Uhealth 2023;11:e40427 [FREE Full text] [doi: 10.2196/40427] [Medline: 37099370]
65. Asadi DM, Tajrishi KZ, Gharaei B. Mindfulness training intervention with the Persian version of the mindfulness training mobile app for premenstrual syndrome: a randomized controlled trial. Front Psychiatry 2022;13:922360 [FREE Full text] [doi: 10.3389/fpsyt.2022.922360] [Medline: 35782413]
66. Ratnayake P, Hyde C. Mental health literacy, help-seeking behaviour and wellbeing in young people: implications for practice. Educ Dev Psychol 2020;36(1):16-21. [doi: 10.1017/edp.2019.1]
67. Schillinger D, Grumbach K, Piette J, Wang F, Osmond D, Daher C, et al. Association of health literacy with diabetes outcomes. JAMA 2002;288(4):475-482 [FREE Full text] [doi: 10.1001/jama.288.4.475] [Medline: 12132978]
68. Weiss BD, Palmer R. Relationship between health care costs and very low literacy skills in a medically needy and indigent Medicaid population. J Am Board Fam Pract 2004;17(1):44-47 [FREE Full text] [doi: 10.3122/jabfm.17.1.44] [Medline: 15014052]
69. Scott TL, Gazmararian JA, Williams MV, Baker DW. Health literacy and preventive health care use among Medicare enrollees in a managed care organization. Med Care 2002;40(5):395-404. [doi: 10.1097/00005650-200205000-00005] [Medline: 11961474]
70. Bennett CL, Ferreira MR, Davis TC, Kaplan J, Weinberger M, Kuzel T, et al. Relation between literacy, race, and stage of presentation among low-income patients with prostate cancer. J Clin Oncol 1998;16(9):3101-3104. [doi: 10.1200/JCO.1998.16.9.3101] [Medline: 9738581]
71. Chau JP, Chang AM. Effects of an educational programme on adolescents with premenstrual syndrome. Health Educ Res 1999;14(6):817-830 [FREE Full text] [doi: 10.1093/her/14.6.817] [Medline: 10585388]
72. Taghizadeh Z, Shirmohammadi M, Feizi A, Arbabi M. The effect of cognitive behavioural psycho-education on premenstrual syndrome and related symptoms. J Psychiatr Ment Health Nurs 2013;20(8):705-713. [doi: 10.1111/j.1365-2850.2012.01965.x] [Medline: 22957993]
73. Gaybor J. Everyday (online) body politics of menstruation. Fem Media Stud 2020;22(4):898-913 [FREE Full text] [doi: 10.1080/14680777.2020.1847157]
74. Solberg LB. The benefits of online health communities. Virtual Mentor 2014;16(4):270-274 [FREE Full text] [doi: 10.1001/virtualmentor.2014.16.04.stas1-1404] [Medline: 24735575]
75. Huberty JL, Green J, Puzia ME, Larkey L, Laird B, Vranceanu AM, et al. Testing a mindfulness meditation mobile app for the treatment of sleep-related symptoms in adults with sleep disturbance: a randomized controlled trial. PLoS One 2021;16(1):e0244717 [FREE Full text] [doi: 10.1371/journal.pone.0244717] [Medline: 33411779]

76. Huberty J, Puzia ME, Green J, Vlisides-Henry RD, Larkey L, Irwin MR, et al. A mindfulness meditation mobile app improves depression and anxiety in adults with sleep disturbance: analysis from a randomized controlled trial. Gen Hosp Psychiatry 2021;73:30-37 [FREE Full text] [doi: 10.1016/j.genhosppsych.2021.09.004] [Medline: 34537477]
77. Kosasih FR, Yee VTS, Toh SHY, Sündermann O. Efficacy of intellect's self-guided anxiety and worry mobile health programme: a randomized controlled trial with an active control and a 2-week follow-up. PLOS Digit Health 2023;2(5):e0000095 [FREE Full text] [doi: 10.1371/journal.pdig.0000095] [Medline: 37224139]
78. Armour M, Ferfolja T, Curry C, Hyman MS, Parry K, Chalmers KJ, et al. The prevalence and educational impact of pelvic and menstrual pain in Australia: a national online survey of 4202 young women aged 13-25 years. J Pediatr Adolesc Gynecol 2020;33(5):511-518. [doi: 10.1016/j.jpag.2020.06.007] [Medline: 32544516]
79. Schoep ME, Adang EMM, Maas JWM, De Bie B, Aarts JWM, Nieboer TE. Productivity loss due to menstruation-related symptoms: a nationwide cross-sectional survey among 32 748 women. BMJ Open 2019;9(6):e026186 [FREE Full text] [doi: 10.1136/bmjopen-2018-026186] [Medline: 31248919]
80. Sharma A, Taneja DK, Sharma P, Saha R. Problems related to menstruation and their effect on daily routine of students of a medical college in Delhi, India. Asia Pac J Public Health 2008;20(3):234-241. [doi: 10.1177/1010539508316939] [Medline: 19124317]
81. Titilayo A, Agunbiade OM, Banjo O, Lawani A. Menstrual discomfort and its influence on daily academic activities and psychosocial relationship among undergraduate female students in Nigeria. Tanzan J Health Res 2009;11(4):181-188 [FREE Full text] [doi: 10.4314/thrb.v11i4.50173] [Medline: 20734697]
82. Buhi ER, Trudnak TE, Martinasek MP, Oberne AB, Fuhrmann HJ, McDermott RJ. Mobile phone-based behavioural interventions for health: a systematic review. Health Educ J 2012;72(5):564-583. [doi: 10.1177/0017896912452071]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
ITT: intention-to-treat
MMRM: Mixed Models for Repeated Measures approach
PMDD: premenstrual dysphoric disorder
PMS: premenstrual syndrome
PP: per-protocol
PSST: Premenstrual Screening Tool
RCT: randomized controlled trial
TTC: trying to conceive

Edited by L Buis; submitted 30.10.23; peer-reviewed by S Buttigieg, J Sun; comments to author 30.01.24; revised version received 13.02.24; accepted 27.02.24; published 02.05.24.

Please cite as:
Cunningham AC, Prentice C, Peven K, Wickham A, Bamford R, Radovic T, Klepchukova A, Fomina M, Cunningham K, Hill S, Hantsoo L, Payne J, Zhaunova L, Ponzo S
Efficacy of the Flo App in Improving Health Literacy, Menstrual and General Health, and Well-Being in Women: Pilot Randomized Controlled Trial
JMIR Mhealth Uhealth 2024;12:e54124
URL: https://mhealth.jmir.org/2024/1/e54124
doi:10.2196/54124
PMID:38696773

Original Paper

Effectiveness of Mobile-Based Progressive and Fixed Physical Activity on Depression, Stress, Anxiety, and Quality of Life Outcomes Among Adults in South Korea: Randomized Controlled Trial

Ye Hoon Lee[1], PhD; Hyungsook Kim[2,3], PhD; Juhee Hwang[1], BS; Sihyeon Noh[1]

[1]Division of Global Sport Industry, Hankuk University of Foreign Studies, Gyeonggi-do, Republic of Korea

[2]Department of Data Science, Hanyang University, Seoul, Republic of Korea

[3]Hanyang Digital Healthcare Center, Hanyang University, Seoul, Republic of Korea

Corresponding Author:
Hyungsook Kim, PhD
Department of Data Science
Hanyang University
222, Wangsimni-ro
Seongdong-gu
Seoul, 04763
Republic of Korea
Phone: 82 222204751
Email: khsook12@hanyang.ac.kr

Abstract

Background: Depression acts as a significant obstacle to the overall well-being of individuals. Given the significant consequences, timely recognition and proactive steps to manage symptoms of depression become essential. Such actions not only reduce personal distress but also play a crucial role in reducing its far-reaching impact on society as a whole.

Objective: In response to this concern, the objective of this study was to explore the use of mobile-based interventions as a possible remedy. More specifically, this study aimed to investigate the effectiveness of 2 types of physical activity (PA), progressive and fixed, within a mobile-based app on depression, perceived stress, anxiety, physical health, and psychological health, aiming to contribute to the optimization of mental health benefits.

Methods: Participants (N=60; mean age 25.29, SD 6.10 years) were recruited using a combination of web-based and offline methods, and the study lasted for 8 weeks. The baseline and posttest questionnaires were administered to all participants. The participants were randomly assigned to 1 of the 3 groups: progressive group (n=20; performing mobile-based progressive PA), fixed group (n=20; performing mobile-based fixed intensity PA), and control group C (n=20). Data analysis involved comparing scores between the experimental and control groups using a one-way ANOVA, paired sample *t* tests (2-tailed), and repeated measures ANOVA with a 3 (group)×2 (time) design.

Results: The findings revealed significant improvements in mental health indicators among participants engaged in both fixed and progressive PA groups compared with the control group. However, the fixed PA group demonstrated more significant reductions in symptoms. Specifically, the progressive PA group showed significant reductions in depression ($F_{1,36}$=6.941; P=.01; η_p^2=0.16) and perceived stress ($F_{1,36}$=5.47; P=.03; η_p^2=0.13), while the fixed PA group exhibited significant reductions in depression ($F_{1,37}$=5.36; P=.03; η_p^2=0.12), perceived stress ($F_{1,37}$=7.81; P=.008; η_p^2=0.17), and general anxiety disorder ($F_{1,37}$=5.45; P=.03; η_p^2=0.13) compared with the control group.

Conclusions: This study underscores the potential of mobile-based PA in improving mental health outcomes. The findings offer significant insights for mental health professionals and researchers aiming to optimize mental well-being through innovative mobile therapies.

Trial Registration: Clinical Research Information Service KCT0009100; https://tinyurl.com/mr33fmur

(JMIR Mhealth Uhealth 2024;12:e55578) doi:10.2196/55578

KEYWORDS

depressive symptoms; mental health; mobile-based exercise; non–face-to-face physical activity; progressive exercise; mobile phone

Introduction

Background

Depressive disorder, commonly known as depression, is a widespread mental condition characterized by persistent feelings of sadness, loss of interest or pleasure in activities, and prolonged symptom duration [1]. Unlike normal mood fluctuations, depression significantly impairs an individual's ability to function effectively in daily life, affecting relationships, work, and engagement in once-enjoyed activities. It can also affect concentration, memory, decision-making, and motivation, making simple tasks challenging [2]. Notably, approximately 280 million people worldwide experience depression, with a prevalence of 3.8% in the population, including 5% of adults (4% in men and 6% in women) and 5.7% among adults aged ≥60 years [3,4]. Consequently, this creates an economic burden, resulting in annual losses of approximately US $36.6 billion [5]. South Korea, which has the greatest suicide rate among Organisation for Economic Co-operation and Development countries, is a nation that encounters unique challenges in this regard. Depression rates in Korea are increasing annually, and the annual social and economic costs associated with suicide, which include the potential reduction in income, amount to approximately ₩ 4.83 trillion or KRW 4.83 trillion (equivalent to US $3.67 billion) [6]. These figures emphasize the critical nature of directing national focus toward the assessment and treatment of depression. Despite the availability of effective treatments for mental disorders, inadequate investment in mental health care, limited availability of trained health care professionals, and social stigma surrounding mental disorders contribute to barriers that impede access to effective care [7,8].

Physical activity (PA) is widely recognized as a significant contributor to mental well-being and a protective factor against depressive symptoms [9-15]. Regular participation in PA not only decreases the frequency and intensity of health problems and chronic illnesses but also improves the general standard of living, playing a vital role in maintaining good health [16]. The underlying mechanisms of the beneficial impacts of PA on depression can be outlined as follows: first, PA stimulates the secretion of endorphins, which are natural substances that elevate mood, relieve pain and tension, and induce a state of relaxation and overall contentment, and it also plays a vital function in the regulation of neurotransmitters, including serotonin, dopamine, and norepinephrine. These neurotransmitters are essential for regulating mood and emotions, ultimately contributing to enhanced mental well-being. Second, PA also serves as a means of alleviating stress, leading to a decrease in overall stress levels and enhancing persons' ability to effectively manage daily obstacles. Third, regular PA not only improves physical fitness and body image but also boosts self-esteem and self-confidence, which in turn has the potential to decrease depressive symptoms. Fourth, PA has been linked to better sleep quality and can help regulate sleep patterns, addressing one of the common issues faced by those with depression [17-24]. Thus, regular PA can alleviate the symptoms of depression and improve overall quality of life by addressing various aspects of human functioning. Indeed, the body of scientific research supporting the benefits of PA continues to expand, highlighting its positive effects on mental health recovery and enhancement [9-16]. For instance, Marques et al [9] conducted a 4-year follow-up analysis on 32,392 European adults from 14 countries and revealed that participating in moderate or vigorous PA was linked to reduced depression scores in both genders. This association remains significant even when accounting for self-rated health, sociodemographic factors, and chronic diseases, indicating a negative correlation between PA and depression symptoms. Even a modest frequency of PA, such as twice a week, was found to lead to cumulative benefits [25].

Despite the recognized importance of PA in mental health recovery, significant limitations to participating in PA exist in traditional face-to-face PA. These limitations, which revolve around accessibility issues, create barriers that hinder people's engagement in PA [26-28]. One aspect of accessibility is psychological distance, which encompasses factors such as a lack of interest, lack of perseverance, and limited PA knowledge. These psychological barriers contribute to individuals feeling distant from engaging in PA, reducing their motivation to participate [28-30]. Aside from psychological barriers, geographical distance poses a hindrance to involvement. Proximity to PA facilities and the timing of PA programs can significantly impact convenience for participants, adding to the psychological distance they experience [27]. Face-to-face programs typically operate on fixed schedules, which may not align with everyone's availability. This lack of flexibility also poses challenges for individuals with busy lifestyles or conflicting commitments, making it difficult for them to commit to specific classes or training sessions. The financial aspects further emphasize the constraints of face-to-face PA programs. Face-to-face PA programs typically involve additional expenses, such as fees for gym memberships, classes, or personal training. These financial responsibilities can serve as obstacles, especially for individuals with constrained budgets or restricted financial means [31]. Collectively, these factors serve as obstacles to face-to-face PA engagement, hindering persons from experiencing the advantages of PA for mental health improvement and enhancement. Considering these limitations, it is crucial to identify alternative choices that provide enhanced accessibility, flexibility, and affordability to ensure that individuals participate in PA and benefit from its favorable effects on their mental well-being.

Mobile-based PA programs are structured interventions that use mobile technology, such as smartphones, tablets, wearable activity trackers, and PDAs. These programs deliver personalized plans for PA, including aerobic exercises, strength training, and flexibility routines. They also incorporate real-time monitoring and tracking features to assess participants' progress

[32]. Compared with traditional face-to-face PA approaches, mobile-based PA programs offer several advantages. First, it tackles the accessibility concerns by enabling individuals to participate in PA at any time and in any location, thus eliminating the limitations of time and place that are associated with face-to-face PA. Studies indicate that mobile-based programs can successfully encourage participation in PA without the requirement of physical gym facilities [33]. In addition, the use of mobile technology allows mobile-based PA programs to provide systematic supervision through customized workout plans and tailored PA activities that cater to the specific demands and fitness levels of individuals [34]. This process offers them prompt feedback and encouragement to adhere to their fitness objectives. Finally, the use of mobile PA apps provides a cost-efficient method for reducing expenses connected to facility use and health concerns [35] while it significantly improves physical health variables, such as managing weight; self-confidence; and mental health aspects, such as decreasing depression, stress, and increasing enjoyment of life [36-41]. For example, Murray et al [39] found that college students who engaged in approximately 150 minutes of moderate to vigorous PA per week over an 8-week period showed significant decreases in depression and anxiety. In general, the mobile-based PA program offers the ability to overcome the restrictions of accessibility, cost, and interest that are commonly associated with in-person PA.

Recognizing the beneficial effects of mobile-based PA programs on mental well-being, it is imperative to further investigate and analyze the impact of different types of PA on mental health. One of the fundamental principles of fitness training is progressive exercise, which involves a gradual increase in the intensity, duration, or complexity of the exercise program over time. The objective of progressive exercises is to continually challenge the body and induce adaptation, leading to improvements in strength, endurance, flexibility, and other desired fitness outcomes [16]. While this type of PA has proven to effectively enhance physical health among individuals, research has also demonstrated the effectiveness of progressive PA in affecting various indicators of mental health [42,43]. For example, Singh et al [42] conducted a 10-week randomized controlled trial involving older adults aged ≥60 years and found that progressive resistance training significantly reduced depression measures. In addition to its impact on the symptoms of depression, previous literature has revealed that individuals engaging in regular PA, including progressive PA, experienced improved mood, well-being, and quality of life compared with those with sedentary lifestyles [43]. These studies offer compelling evidence supporting the positive effects of progressive PA on mental well-being, encompassing reductions in depressive and anxiety symptoms, improvements in mood, and an overall enhancement of quality of life. While research has already demonstrated the benefits of progressive PA in traditional settings, there remains a need for further investigation to explore its effectiveness and implementation in the context of mobile-based PA. Remarkably, no studies have specifically examined the application of progressive exercise principles within a mobile-based PA framework. Understanding the potential impact of progressive PA within mobile-based interventions could have substantial implications for promoting mental health and well-being in an increasingly digitally connected world.

Objective

In summary, previous studies have successfully demonstrated that regular PA has a positive impact on alleviating mental issues such as depression, anxiety, and stress and improving the quality of life. However, it is important to note that most previous research has predominantly focused on face-to-face PA interventions or mobile-based PA characterized by a fixed intensity. Unfortunately, limited studies have ventured into exploring the potential benefits of mobile-based PA through apps or the implementations of progressive PA approaches. To address this research gap, this study aimed to investigate the effects of both progressive and fixed PA interventions, using a mobile app, on depression, stress, anxiety, and quality of life of adults in South Korea. We hypothesized that both progressive and fixed PA interventions could lead to improvements in mental health indicators such as depression, perceived stress, and anxiety, as well as enhance quality of life indicators related to physical and psychological health. Through this exploration of the benefits of mobile-based PA programs and their potential to positively influence mental health outcomes, we aimed to have a deeper understanding of how technology-enabled interventions can play a significant role in promoting well-being in modern society.

Methods

Participants

The recruitment of participants was meticulously designed to ensure a representative and diverse sample. In order to do this, a blend of web-based and offline recruitment techniques was used, with each method playing a role in the random sampling procedure. Web-based recruitment efforts targeted a diverse pool of potential participants by leveraging various digital platforms such as social media groups, forums, and university websites. Our objective was to attract individuals with diverse backgrounds who were interested in contributing to mental health research. Offline recruitment strategies complemented the web-based efforts by reaching individuals who may not have regular access to internet resources. Promotional materials were disseminated in public venues such as university counseling centers, local community centers, and coffee shops. These venues were selected to ensure visibility and accessibility to individuals lacking access to internet platforms. By using a combination of web-based and offline recruitment techniques, we aimed to create a sample that reflects the diversity of the target population. The recruitment period lasted from the first to the fourth week of March 2023, and both web-based and offline advertisements were used. The participants completed the baseline and posttest questionnaires in the first weeks of April and June 2023, respectively.

Eligibility criteria were predetermined; participants should be (1) healthy adults, aged between 18 and 65 years to ensure a sample representative of the adult population; (2) fluent in reading, writing, and conversing in Korean to ensure participants could fully understand and engage with study materials and instructions; (3) having no contraindications for PA as assessed

by the Physical Activity Readiness Questionnaire (PAR-Q) [44] to ensure participant safety during the intervention period; (4) willing to be randomized into an intervention that may require up to 2 hours per week of their time to ensure participant commitment and adherence to the study protocol; (5) familiar with using a smart mobile phone to ensure participants could adequately perceive and respond to visual and auditory stimuli presented during the study; (6) having normal (or corrected-to-normal) vision and hearing; and (7) having no history of psychosis to minimize potential risks associated with mental health conditions during the intervention period.

Sample Size Calculation

In this study, the appropriate sample size was determined using the G* power calculator (version 3.1.9.4; Heinrich-Heine-Universität) [45]. The aim was to achieve a significance level of .05 and a statistical power of 95%, with an effect size of 0.36, based on the data from a prior study that investigated the impact of mobile-based PA interventions on depression [41]. The calculations indicated that 30 participants were required to reach the desired power of 0.95. Accounting for a 57% dropout rate observed in computer-based psychological treatment [46], the total sample size needed was calculated to be 47 participants to maintain a power of 0.95. Therefore, including 60 participants was anticipated to provide sufficient statistical power for the study.

Procedure

Initially, a total of 60 participants were recruited. Following the fulfillment of inclusion criteria, the process of participant allocation used a simple block randomization methodology. This approach involved using the resources available on the open-source website to assign a total of 60 participants into 3 distinct groups: the progressive group, the fixed group, and the control group. Each group comprised 20 participants. The procedure involved the creation of 10 blocks, each containing 6 participants, as the designated units for randomization. Participants were allocated to the 3 groups within each block using randomization, guaranteeing an equal distribution of participants among the groups throughout each block. At first, the groups had the following gender distribution: the progressive group consisted of 10 men and 10 women, the fixed group consisted of 14 men and 6 women, and the control group consisted of 12 men and 8 women. After the random selection, 2 women from the fixed group expressed a preference for the control group due to their disinclination for PA, while 2 men from control group wished to switch to the fixed group in order to participate in PA. Consequently, the research protocol allowed these participants to transfer between groups, leading to the following revised composition: the progressive group consisted of 10 men and 10 women, the fixed group consisted of 16 men and 4 women, and the control group comprised 10 men and 10 women.

After the randomization, the researchers contacted the participants via the Webex platform (Cisco) for both baseline and posttest evaluations at baseline and 8 weeks. Initially, the participants were instructed to fill out the web-based questionnaire, which encompassed an explanation of the study's objectives, privacy safeguards, and procedures to ensure confidentiality during the baseline evaluation, informed consent, and Physical Activity Readiness Questionnaire (PAR-Q) [44]. During the intervention phase, the researchers consistently spoke with the participants on a weekly basis to ensure that they followed the instructions and to track their advancement. They supervised the completion of the assigned PA and provided assistance for any personal queries. To validate the participants' activities, the researchers received verification photos via messenger, showing the number, duration, and type of PA sessions for each week. To uphold consistency and anonymity, all surveys were collected without personal identifiers, and each participant used a unique ID for completing the surveys during both the baseline and postevaluation phases.

Ethical Considerations

This study received ethical approval from the institutional review board of Hankuk University of Foreign Studies (approval number HIRB-202306-HR-001). Before commencing the study, all necessary approvals were obtained to ensure compliance with ethical standards and the protection of participants' rights. The objectives and methods of the study were clearly explained to potential participants, and only those who provided written informed consent and willingly chose to participate were included.

Instrument

Depression

Depression was measured using the Korean version of the Patient Health Questionnaire-9 (PHQ-9). The PHQ-9 has been reported to be a reliable and valid tool for screening and assessing depressive symptoms [47]. The inventory has 9 items pertaining to symptoms of depression, including diminished interest or pleasure, mood fluctuations, disruptions in sleep patterns, exhaustion, and alterations in food behavior. The scoring system ranges from 0 (not at all) to 3 (nearly every day) for each item. A greater cumulative score on the PHQ-9 signifies a higher level of depression symptoms, whereas a lower score signifies a lower level of symptoms. The Cronbach α coefficients were 0.86 for the study sample and 0.81 in the initial validation [47].

Perceived Stress

The assessment of perceived stress was conducted by using the Brief Encounter Psychosocial Instrument-Korean version (BEPSI-K). BEPSI-K is a tool that has been created using the dynamic interaction model of stress and its detrimental effects on health [48]. This is a Korean version of the original BEPSI that has been tested and confirmed to be reliable and legitimate [49]. The scoring system for each item is based on a scale that spans from 1 (not at all) to 5 (always). Consequently, elevated scores on the BEPSI-K signify an increased perception of stress, whereas lower scores indicate a decreased experience of stress. In our investigation, the Cronbach α coefficient obtained from our sample was 0.81, which is somewhat higher than the value of 0.80 reported in the original validation [48].

Anxiety

We used the Generalized Anxiety Disorder-7 (GAD-7) to evaluate the symptoms associated with generalized anxiety

disorder. The GAD-7 is beneficial because of its simplicity and limited number of components, which allows for easy access [50]. The scale has 7 items that evaluate anxiety symptoms. Each item is rated on a scale ranging from 0 (not at all) to 3 (nearly every day). The GAD-7 scores are directly proportional to the frequency and severity of symptoms associated with generalized anxiety disorder. Higher scores indicate a greater number and intensity of symptoms, whereas lower scores indicate a lesser number of symptoms. The Cronbach α coefficient in our study sample was 0.90, although the original validation study reported a value of 0.92 [50].

Quality of Life

The study used the Korean version of the World Health Organization Quality of Life-Brief assessment questionnaire to measure the quality of life. The World Health Organization Quality of Life-Brief is a reliable and valid tool developed by the WHO in 1996 as a shortened version of the WHOQOL-100, which scientifically measures and evaluates the quality of life [51]. It comprises 5 domains: physical health, psychological health, social relationships, environmental factors, and overall quality of life. This study used 13 items, including 7 items related to physical health and 6 items related to psychological health. These items assess various aspects, such as physical satisfaction, need for treatment, sleep quality, and negative emotions. Scores range from 1 (not at all) to 5 (very much)**.** The Cronbach α values were 0.79 (physical health) and 0.81 (psychological health) for the study sample, and 0.71 (physical health) and 0.83 (psychological health) in the original validation [51]. In our study sample, the Cronbach α values were 0.79 for physical health and 0.81 for psychological health, while the original validation reported values of 0.71 for physical health and 0.83 for psychological health.

Mobile-Based PA Program

The program used in this study is a personalized health care app that provides artificial intelligence–based wellness services in the areas of PA, diet, nutrient, and mental health. The app offers various PA videos tailored to different goals including 32 stretching movements, 44 full-body aerobic PA, 20 upper-body strengthening PA, and 18 lower-body strengthening PA, categorized according to difficulty (easy, moderate, and difficult). Thus, participants could select desired movements and adjust the number of sets to create personalized PA lists. All participants, irrespective of their group assignment, were granted equal access to the complete app. They were provided the freedom to choose and combine PA according to their individual preferences and needs and had the flexibility to adjust the number of sets as desired.

In the context of this app-driven intervention, participants in both the progressive and fixed groups were instructed to engage in PA for 8 weeks, with a minimum of 3 sessions per week and a minimum PA duration of 15 minutes per day. For the progressive group, participants were guided to participate in progressive PA, which involved a daily and weekly gradual increase in difficulty level, duration, and intensity. For example, the daily program of the progressive group mostly involved performing 2 to 3 movements at each difficulty level, with the number of repetitions increasing as the difficulty level increased. That is, the routine started with easier movements and became more difficult over time, with more repeats and higher levels of difficulty during each session. Furthermore, the progressive group performed 12 sets in weeks 1 to 2, a total of 18 sets in weeks 3 to 4, a total of 27 sets in weeks 5 to 6, and 36 sets in weeks 7 to 8. Finally, the progressive group was encouraged to gradually increase the duration of PA, starting with a minimum of 15 minutes in the first week and progressing up to approximately 30 minutes by the eighth week. This methodology facilitated a seamless transition toward progressive PA engagement.

The fixed group in our study followed a different approach compared with the progressive group. Instead of experiencing a gradual increase in intensity, frequency, duration, and difficulty levels, participants in the fixed group engaged in a consistent and predetermined intensity level and set of PA throughout the 8-week period. As for the intensity levels, they were guided to choose and proceed with a moderate intensity PA in the app. The participants freely chose and performed 6 types of PAs that they desired at the specified intensity level. They also performed the PA with a fixed number of sets and repetitions for each session, maintaining a PA frequency of approximately 16 to 17 minutes per session, 3 times a week, from the first to the eighth week. This steady and consistent PA routine was maintained without the progressive changes in difficulty levels and repetition experienced by the progressive group. In contrast, the control group did not participate in any structured PA during the study period.

Covariate

In this study, gender was incorporated as a covariate variable in the statistical analysis to consider its potential impact on mental health outcomes, such as depression, anxiety, and perceived stress. Research has consistently demonstrated gender differences in the occurrence, presentation, and management of mental health disorders [52,53]. By controlling for gender, we aimed to mitigate the potential confounding effects of gender-related factors on the study outcomes, thereby enhancing the internal validity of the expected findings. Furthermore, by implementing gender control, we aimed to distinguish the specific impacts of the independent variable, specifically, the PA intervention type, on the dependent variables, facilitating more accurate conclusions from the findings.

Data Analysis

Initially, we conducted descriptive analysis to obtain frequencies and percentages for categorical variables and means and SDs for continuous variables. Subsequently, we assessed differences between groups at baseline and posttest evaluations by using a one-way ANOVA for continuous variables and a chi-square test for categorical variables. When we found statistically significant differences between the groups through the initial one-way ANOVA, we proceeded with post hoc tests for further analysis. We used the appropriate Tukey honestly significant difference test. By performing post hoc tests, we aimed to explore and interpret the significant differences observed during the initial analysis, contributing to a more nuanced interpretation of the study results. To evaluate the effects of the mobile-based PA, we compared and analyzed self-report questionnaire scores

between the experimental and control groups using SPSS Statistics for Windows (version 22.0; IBM Corp). Initially, we conducted paired sample *t* tests within each group to compare the baseline and postevaluation measurements separately for both groups. Subsequently, we performed a 3 (group)×2 (time) repeated measures ANOVA (RMANOVA) using baseline and postevaluation measurements, with the group as the independent variable. To control for their influence on the study outcomes [52], we included gender as a covariate in our analysis. Upon finding significant differences in the 3×2 RMANOVA, we further examined the individual group comparisons using a 2 (group)×2 (time) RMANOVA. Specifically, we compared the progressive PA group with the control group, as well as the fixed PA group with the control group, to identify which specific group differences contributed to the observed effects.

Results

Demographic Characteristics at Baseline

Figure 1 illustrates the CONSORT (Consolidated Standards of Reporting Trials) diagram, presenting the participant flow throughout the study. The initial 60 participants were divided into 3 groups: progressive group, fixed group, and control group. The progressive group followed a mobile-based progressive PA program, with a daily and weekly gradual increase in the intensity and difficulty levels. More specifically, 20 (33%) participants were assigned to each of the experimental groups A and B, while 20 (33%) participants were allocated to the control group C. Notably, 98% (39/40) of the participants in progressive and fixed groups and 100% (20/20) of the participants in the control group completed the posttest assessment.

In this study, a comprehensive statistical analysis was performed to investigate potential demographic differences between the study groups. The results indicated no statistically significant variations in gender (P=.10), age (P=.78), height (P=.06), weight (P=.15), BMI (P=.31), and internet self-efficacy (P=.58) among the groups. In addition, the level of internet self-efficacy remained consistent throughout the various study groups. Table 1 displays the demographic attributes of the participants.

Figure 1. Study flow diagram. PA: physical activity.

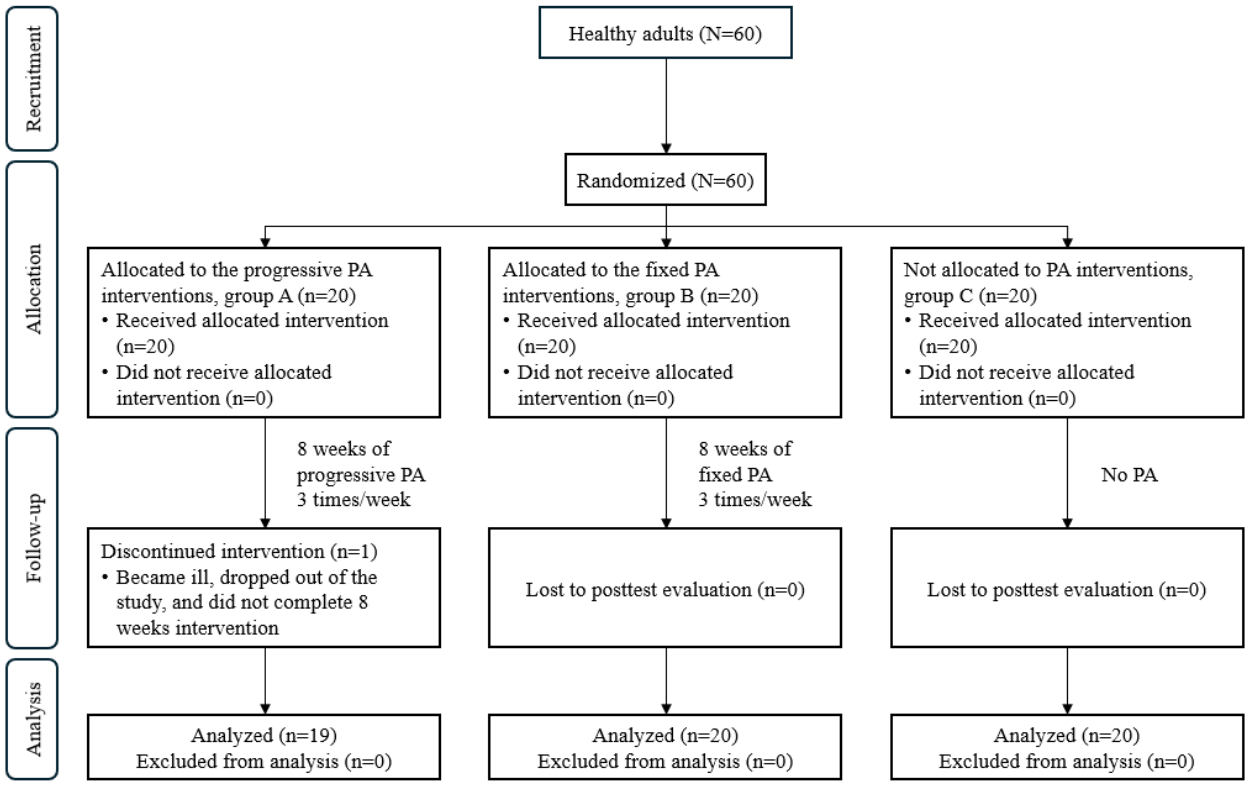


Table 1. Demographic characteristics.

Characteristics	Progressive group (n=19)	Fixed group (n=20)	Control group (n=20)	*P* value
Categorical variables, n (%)				
Gender				.10
Man	10 (53)	16 (80)	10 (50)	
Woman	9 (47)	4 (20)	10 (50)	
Education				.51
High-school graduation	1 (5)	1 (5)	4 (20)	
College students	12 (63)	13 (65)	11 (55)	
Bachelor's degree	4 (21)	5 (25)	5 (25)	
Graduate degree	2 (11)	1 (5)	0 (0)	
Experience in mobile-based physical activity app				.10
Yes	10 (53)	4 (20)	7 (35)	
No	9 (47)	16 (80)	13 (65)	
Continuous variables, mean (SD)				
Age (y)	25.11 (4.62)	26.05 (6.28)	24.7 (7.28)	.78
Height (cm)	170.61 (7.82)	172.52 (6.01)	167.00 (7.28)	.06
Weight (kg)	70.88 (13.35)	70.14 (15.85)	63.47 (14.42)	.15
BMI (kg/m^2)	24.50 (3.80)	22.27 (6.58)	22.44 (3.90)	.31
Internet self-efficacy	4.00 (1.20)	3.65 (1.39)	4.05 (1.36)	.58

Differences Observed in the 3 Groups

This study used a one-way ANOVA to examine the variables of depression, stress, anxiety, and physical and psychological health among 3 unique groups at both the baseline and postevaluation stages. The Levene test was conducted to assess whether the variances were equal across the 3 groups (PHQ-9, *P*=.92; BEPSI-K, *P*=.57; GAD-7, *P*=.70; physical health, *P*=.97; and psychological health, *P*=.51). Therefore, there was no significant evidence of a difference in variances among the groups at the baseline evaluation.

The characteristics of the participants in the progressive group (n=19), fixed group (n=20), and control group C (n=20) at baseline are summarized in Table 2. The results revealed that no statistically significant differences existed between the 3 groups for any of the variables at baseline.

Table 3 provides a summary of the intergroup comparisons following the intervention. The analysis revealed statistically significant differences between groups as determined by one-way ANOVA for depression ($F_{2,56}$=5.06; *P*=.01; η_p^2=0.15) and GAD-7 ($F_{2,56}$=4.13; *P*=.02; η_p^2=0.12). A Tukey post hoc test revealed significant differences between the progressive group and control group for depression (mean difference −3.27; *P*=.01) and for GAD-7 (mean difference −0.33; *P*=.03). There was no statistically significant difference between the groups for perceived stress ($F_{2,56}$=3.12; *P*=.05), physical health ($F_{2,56}$=3.08; *P*=.05), and psychological health ($F_{2,56}$=1.04; *P*=.36).

Table 2. Baseline evaluation.

	Progressive group (n=19), mean (SD)	Fixed group (n=20), mean (SD)	Control group (n=20), mean (SD)	*F* test (*df*)	*P* value	Cohen *d*
PHQ-9[a]	4.15 (4.21)	4.75 (4.27)	4.85 (4.62)	0.14 (57)	.87	0.004
BEPSI-K[b]	1.65 (0.66)	1.83 (0.57)	1.72 (0.53)	0.44 (57)	.64	0.015
GAD-7[c]	0.33 (0.45)	0.45 (0.52)	0.44 (0.60)	0.29 (57)	.75	0.010
WHOQOL[d]—physical health	3.11 (0.46)	2.80 (0.48)	2.77 (0.50)	2.83 (57)	.07	0.091
WHOQOL—psychological health	3.37 (0.49)	3.11 (0.62)	3.15 (0.59)	1.18 (57)	.31	0.040

[a]PHQ-9: Patient Health Questionnaire-9.

[b]BEPSI-K: Brief Encounter Psychosocial Instrument-Korean version.

[c]GAD-7: Generalized Anxiety Disorder-7.

[d]WHOQOL: World Health Organization Quality of Life.

Table 3. Posttest evaluation.

	Progressive group (n=19), mean (SE)	Fixed group (n=20), mean (SE)	Control group (n=20), mean (SE)	Progressive vs control group, GMD[a] (95% CI)	Fixed vs control group, GMD (95% CI)	Progressive vs fixed group, GMD (95% CI)
PHQ-9[b]	1.52 (1.61)	2.25 (1.97)	4.80 (5.24)	−3.27 (−5.88 to −0.65)[c]	−2.55 (−5.13 to 0.03)	−0.72 (−3.33 to 1.89)
BEPSI-K[d]	1.41 (0.42)	1.39 (0.39)	1.68 (0.86)	−0.39 (−0.84 to 0.06)	−0.41 (−0.85 to 0.03)	0.02 (−0.42 to 0.47)
GAD-7[e]	0.16 (0.17)	0.19 (0.25)	0.50 (0.63)	−0.33 (−0.64 to −0.02)[c]	−0.31 (−0.61 to −0.02)	−0.03 (−0.34 to 0.28)
WHOQOL[f]—physical health	3.12 (0.53)	2.72 (0.48)	2.95 (0.54)	0.39 (−.01 to 0.71)	0.30 (−0.09 to 0.70)	0.09 (−0.31 to 0.50)
WHOQOL—psychological health	3.40 (0.52)	3.28 (0.63)	3.11 (0.69)	0.29 (−0.36 to 0.60)	0.17 (−0.30 to 0.64)	0.09 (−0.32 to 0.50)

[a]GMD: group mean difference.

[b]PHQ-9: Patient Health Questionnaire-9.

[c]*P*=.009 for Patient Health Questionnaire-9 and *P*=.02 for Generalized Anxiety Disorder-7.

[d]BEPSI-K: Brief Encounter Psychosocial Instrument-Korean version.

[e]GAD-7: Generalized Anxiety Disorder-7.

[f]WHOQOL: World Health Organization Quality of Life.

Effectiveness of Mobile-Based PA on the Outcomes

Table 4 and Figure 2 illustrate the changes observed in the experiment and control groups from baseline to program completion.

There was a significant positive interaction found between group and time in relation to depression ($F_{2,55}$=3.76; *P*=.03; η_p^2=0.12) and perceived stress ($F_{2,55}$=4.77; *P*=.01; η_p^2=0.15). These results suggest that depression and perceived stress levels changed over time, and the impact of the intervention was dependent on the group assignment. However, no significant interaction was observed for GAD-7 ($F_{2,55}$=2.67; *P*=.08; η_p^2=0.09), physical health ($F_{2,55}$=1.72; *P*=.19; η_p^2=0.06), and psychological health ($F_{2,55}$=1.48; *P*=.24; η_p^2=0.05) between the groups. Specifically, the post hoc analysis showed a significant difference in the change of depression ($F_{1,36}$=6.94; *P*=.01; η_p^2=0.16) and perceived stress ($F_{1,36}$=5.47; *P*=.02; η_p^2=0.13) between progressive group and control group over time. Further, a significant difference in the change of depression ($F_{1,37}$=5.36; *P*=.02; η_p^2=0.12), perceived stress ($F_{1,37}$=7.81; *P*=.01; η_p^2=0.17), and GAD-7 ($F_{1,37}$=5.45; *P*=.02; η_p^2=0.13) between the fixed group and control group was also observed across time.

Table 4. Differences in the outcome variables across time (n=59).

Outcome	Evaluation					
	Baseline, mean (SD)	Posttest, mean (SD)	Difference			
			Values, mean (95% CI)	*t* test (*df*)	*P* value	Cohen *d*
PHQ-9[a]						
Progressive	4.15 (4.21)	1.52 (1.61)	2.63 (0.52 to 4.73)	2.54 (56)	.02	0.82
Fixed	4.75 (4.27)	2.25 (1.97)	2.50 (0.36 to 4.63)	2.37 (56)	.02	0.75
Control	4.85 (4.62)	4.80 (5.24)	0.05 (−3.11 to 3.21)	0.03 (56)	.98	0.01
BEPSI-K[b]						
Progressive	1.65 (0.66)	1.41 (0.42)	0.24 (−0.12 to 0.60)	1.34 (56)	.18	0.43
Fixed	1.83 (0.57)	1.39 (0.39)	0.44 (0.12 to 0.76)	2.82 (56)	.008	0.90
Control	1.82 (0.91)	1.74 (0.78)	−0.08 (−0.52 to 0.36)	−0.36 (56)	.71	0.09
GAD-7[c]						
Progressive	0.33 (0.45)	0.16 (0.17)	0.17 (−0.05 to 0.39)	1.55 (56)	.13	0.50
Fixed	0.45 (0.51)	0.19 (0.25)	0.26 (0.00 to 0.52)	2.05 (56)	.04[b]	0.64
Control	0.44 (0.60)	0.50 (0.63)	−0.05 (−0.45 to 0.33)	−0.29 (56)	.77	0.09
WHOQOL[d]**—physical health**						
Progressive	3.11 (0.46)	3.12 (0.53)	−0.01 (−0.33 to 0.32)	.96 (56)	.96	0.02
Fixed	2.80 (0.48)	3.02 (0.56)	−0.22 (−0.56 to 0.10)	−1.37 (56)	.17	0.42
Control	2.77 (0.50)	2.72 (0.48)	0.05 (−0.26 to 0.37)	0.36 (56)	.72	0.10
WHOQOL—psychological health						
Progressive	3.37 (0.49)	3.40 (0.52)	−0.02 (−0.35 to 0.30)	.87 (56)	.87	0.06
Fixed	3.11 (0.62)	3.28 (0.63)	−0.16 (−0.56 to 0.23)	−0.83 (56)	.40	0.27
Control	3.15 (0.59)	3.11 (0.69)	0.03 (−0.38 to 0.44)	0.16 (56)	.87	0.06

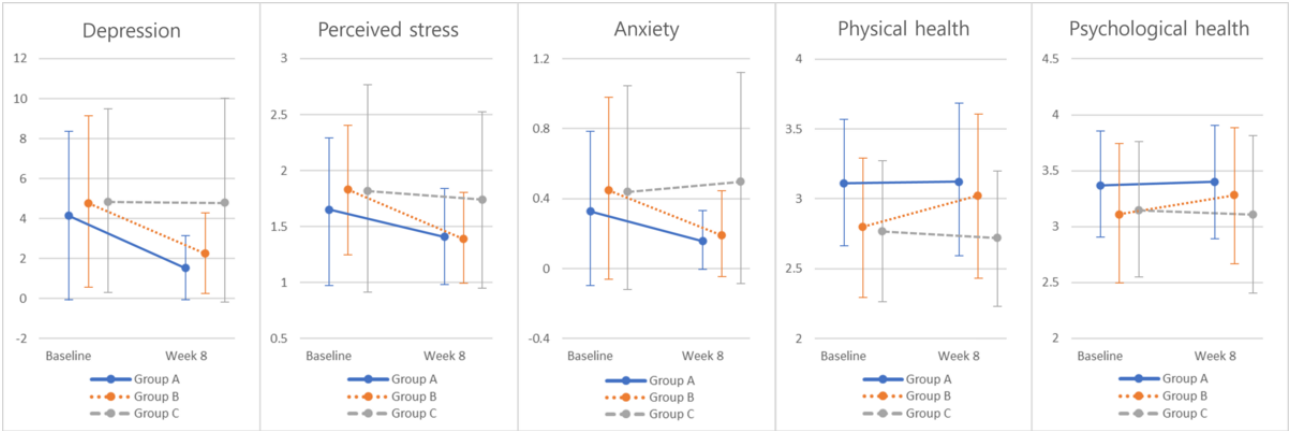
[a]PHQ-9: Patient Health Questionnaire-9.

[b]BEPSI-K: Brief Encounter Psychosocial Instrument-Korean version.

[c]GAD-7: Generalized Anxiety Disorder-7.

[d]WHOQOL: World Health Organization Quality of Life.

Figure 2. The results of the interactional analysis.



Discussion

Principal Findings

This study investigated the impact of progressive and fixed PA programs delivered through a mobile app on mental health indicators (depression, perceived stress, and anxiety) and quality of life (physical and psychological health) among South Korean adults. Following the intervention, the findings of this study indicated that a significant positive interaction effect existed between group and time for depression and perceived stress, implying that the effectiveness of the interventions in reducing the outcomes may have varied depending on the specific group in which they participated. Thus, we conducted additional statistical analyses to examine the impact of each PA type on the outcome variables in the experimental groups and compare the variables with the control group, leading to the following discussion.

First, we observed that participants in both the progressive and fixed PA groups showed a greater reduction in depression and perceived stress compared with those in the control group. Engaging in PA, including progressive PA, has been shown to stimulate the release of endorphins, which are neurotransmitters associated with feelings of well-being and happiness [17,54]. Furthermore, regular PA promotes neuronal growth and connectivity, particularly in brain regions that are involved in mood regulation, such as the hippocampus and prefrontal cortex [20,21,55]. These alterations in neurochemical and neurobiological processes may help reduce depression symptoms [54,55]. In addition, engaging in PA via a mobile app might function as a diversion from detrimental ideas and excessive contemplation, which are frequently linked to depression [56]. By prioritizing PA at home and attaining specific objectives and significant achievements, individuals can shift their focus away from depressed symptoms and cultivate a more positive mindset. As a result, those who participate in both the progressive and fixed PA groups may observe a more significant decrease in depression and perceived stress compared with those who do not engage in PA. Previous literature has found the significant impact of mobile-based PA programs on mental health [36-41], and this aligns with the findings of this study. Theoretical implications of these findings in academia endorse the use of both regular fixed PA and progressive PA through mobile apps as potential therapies for mitigating depression and stress. These findings contribute to a growing body of research on the effectiveness of mobile-based therapies in enhancing mental well-being.

Along with the finding of significant impact of both progressive and fixed PA on mental health indicators, a critical finding of this study is the superior improvement observed in the outcome variables within the fixed PA group compared with the progressive PA group. This finding is somewhat different from those of the previous literature exploring the impact of progressive and fixed PA on physical health outcomes [57-59]. However, this study found that fixed PA showed better results in terms of mental health indicators. One possible explanation could be related to motivation and adherence to the prescribed PA program. In this study, the fixed group encompassed a relatively higher proportion of men (16/20, 80%) compared with that in the other groups (control and progressive groups: 10/20, 50%), while 2 men from the control group switched to the fixed group, suggesting their high motivation for fixed PA engagement. Previous literature has consistently reported a positive association between motivation and mental well-being [60,61]. For example, a meta-analysis on student motivation [60] identified that amotivation (ie, lack of motivation) was positively associated with depression and anxiety, whereas regulation (ie, self-directed value-driven motivation) and intrinsic motivation were negatively associated with only anxiety. This aligns with the result of this study. Furthermore, Lauderdale et al [62] noted that male college students had more intrinsic motivations than women college students. Thus, it is possible that the fixed group includes participants with higher motivation than the progressive group, which in turn demonstrates better improvement in mental health outcomes, specifically with regard to anxiety.

Another possible explanation may be related to the psychological and physical efforts incurred during the intervention period. The progressive PA group may have faced challenges related to the increasing complexity and intensity of their PA program over time. Some participants might have been reluctant to increase the duration and intensity but felt obligations to adhere to the prescribed protocol due to their commitments to the research protocol. Furthermore, engaging in approximately 30 minutes of PA at weeks 7 and 8 could have induced more physical exertion and stress in the progressive group compared with the fixed group who engaged in approximately 17 minutes of PA. Indeed, grounded upon the reversal theory [63], Kerr and Kuk [64] found that compared with running at a low intensity, running at a high intensity resulted in significantly higher scores for physical stress and exertion among participants who ran 1.7 km. Thus, the gradual progression and high-intensity PA could have led to psychological and physical strain, potentially affecting their mental conditions and generating conflicts with the mental health enhancement. In contrast, participants in the fixed PA group may have experienced less or no psychological strain due to the consistent and predictable nature of their PA protocols. Knowing exactly what to expect in terms of PA duration, intensity, and routine may have provided a sense of structure and control, which may lead to psychological comfort during the intervention phase. Thus, in this context, although progressive PA still had significant impacts on mental health outcomes, the effect may not have been as pronounced as the effect in the fixed group, which did not incur psychological strains and emotional challenges related to engaging in PA at different and higher intensities and durations. It is important to note that this interpretation is only speculative, and further investigation into the psychological mechanisms underlying both progressive and fixed PA protocols on anxiety is warranted. In addition, although the progressive group did not demonstrate statistically significant reductions in anxiety during the 8-week intervention period, it is worth noting that the paired sample *t* test revealed a decrease in the anxiety level from 0.33 to 0.16 (P=.13). This suggests a potential trend toward improvement in anxiety within the progressive group, albeit not reaching statistical significance.

Although this study found that the fixed PA group showed better results in mental health outcomes than the progressive PA group, the main advantage of this study is the demonstration of the efficacy of a mobile-based progressive PA program in diminishing different mental health markers, such as depressive symptoms and perceived stress. Notably, this study is the first to apply the principle of progressive exercise in the mobile PA app context. Prior studies have primarily explored the impact of progressive PA on mental health [42,43] and overall health indicators [57-59] within traditional face-to-face PA contexts. In this study, the researchers expanded the research paradigm by introducing the principles of progressive exercise into the mobile context and specifically investigated the effects of progressive exercise principles in comparison with the control group. Thus, this study contributes significantly to our understanding of the effectiveness of different types of mobile-based PA programs in enhancing mental health. The incorporation of progressive exercise principles into mobile PA programs offers the potential and additional avenues for individuals to experience mental health benefits.

Finally, this study found that neither progressive nor fixed mobile phone–based PA programs had a significant impact on the participants' quality of life (P=.19 for physical health and P=.24 for psychological health), as measured by physical and psychological health. While there was a slight increase in the average scores for both physical and psychological health in both progressive and fixed groups, these increases were not statistically significant. There are several possible explanations for the lack of significant results. First, the duration of the intervention might not have been sufficiently long to produce significant changes in the participants' quality of life. Quality of life is a multifaceted construct that may require a longer period of sustained PA to observe meaningful improvements compared with the negative mental health indicators of depression, stress, and anxiety [65,66]. A meta-analysis of 16 studies examined the impact of PA on the quality of life among patients with cancer [65], and the average duration of each study was 17.18 minutes, ranging from 3 weeks to 16 months. Second, the level of adherence to the prescribed PA programs varied among the participants. Variations in adherence may have influenced the outcomes, as participants who did not consistently engage in PA may not have experienced the full potential benefits.

Limitations and Future Research Directions

Although this study has made valuable findings, it is important to acknowledge its limitations. First, deviation from the initial randomization strategy creates intricacies that may impact the internal validity of the investigation. Although participant choice and ethical responsiveness are prioritized, it is important to acknowledge that these transfers could introduce confounding variables that may affect the results of the study. For example, a higher proportion of men in the fixed group than that in the progressive group, along with the transfer of men to the fixed group, might lead to differences in motivation levels between the groups. This could possibly result in bias in favor of the fixed group over the progressive group. Future research may explore the differential responses of men and women to PA interventions and explore the underlying motivational factors influencing engagement, thus enabling the development of tailored interventions that optimize effectiveness for each gender. Notwithstanding these deviations, measures were implemented to guarantee methodological rigor, such as a meticulous random sampling procedure during the recruitment phase, transparent documentation of the transfers, supplementary analysis to assess their potential impacts, and inclusion of gender as a controlled variable in the statistical analysis. Despite significant efforts to reduce the negative effects of this deviation, it is advisable for readers to consider these complexities while interpreting the study's results.

In a similar vein, if individual differences such as participants' motivational level have the potential to influence the results of PA interventions, it is important to consider individual and contextual differences as moderators in the relationship between PA protocols and mental health outcomes. These individual and contextual differences may include personality traits [67], self-esteem [68], resilience [69], and group dynamics [70], all of which have been shown to significantly impact mental health indicators. Future research may explore these individual differences as moderators or examine the effects of PA interventions on mental health across various individual differences.

Furthermore, potential biases may arise throughout the recruitment process. For example, selection bias may occur if certain groups of individuals are more likely to engage in a study than others [71]. Individuals with a specific interest in mental health or PA may be more inclined to volunteer, leading to a sample that is not representative. In addition, recruitment bias can occur when some recruitment strategies attract specific demographic groups more than others [72]. For example, web-based advertisements might predominantly reach individuals who frequently visit university-related websites, while distributing flyers at community centers may be more accessible to participants who visit such places more often.

In addition, the inability to monitor participant adherence to the assigned PA protocols represents a potential limitation. Although participants were required to capture and submit their PA records generated by the app, the lack of real-time tracking made it challenging to assess the extent to which participants complied with the prescribed PA protocols. Thus, future studies should implement robust adherence tracking mechanisms to provide a more comprehensive understanding of participant engagement and its relationship to intervention effectiveness.

Finally, this study primarily focused on assessing the PA intervention outcomes in a healthy population. Consequently, the protocols and dosages used in this study may not be directly generalizable to populations with preexisting mental health concerns or diagnoses. Individuals with mental health conditions may have unique needs and considerations when engaging in PA interventions, including tailored exercise prescriptions, supervision, and support mechanisms. Therefore, caution should be exercised when extrapolating the findings of this study to clinical populations, and future research should aim to investigate the efficacy of PA interventions specifically tailored to individuals with mental health concerns.

Practical Implications

This study underscores the significance of mobile-based PA in generating positive mental health outcomes in adults. Consequently, it is crucial to actively promote and facilitate the adoption of such programs. This can be achieved through diverse approaches, including awareness campaigns, educational materials, and partnerships with fitness professionals or organizations. By emphasizing the convenience and accessibility of mobile-based programs, individuals can be motivated to incorporate PA routines into their daily lives using smartphones or tablets at their preferred times and locations [73].

In order to encourage mobile-based progressive PA programs, governments must engage in efforts such as providing financing or subsidies for the creation and promotion of mobile PA apps. Engaging in collaboration with app developers, fitness experts, and mental health physicians is vital to guarantee the caliber and efficacy of these programs. Furthermore, governments have the capability to incorporate mobile PA programs into public health campaigns, emphasizing their positive impact on mental health and overall well-being.

Finally, fitness facilities might assume a crucial function by collaborating with mobile app developers or developing their own fitness applications. These centers can grant members access to these apps as a component of their membership packages or provide specific programs that integrate in-person instruction with mobile-based PA. Conducting workshops or events to enlighten persons about the advantages of mobile-based progressive PA and enhance their efficacy can also be advantageous.

Conclusions

This study aimed to fill the existing gap in knowledge regarding the efficacy and application of progressive and fixed PA programs in mobile-based environments, with the ultimate purpose of maximizing mental health advantages. This study produced noteworthy results that provided insights into the influence of this intervention on mental health outcomes. The progressive and fixed PA groups showed a remarkable reduction in depression and perceived stress levels compared with the control group. In addition, only the fixed group showed significant decreases in GAD-7 scores.

These findings offer valuable insights into the potential advantages of engaging in mobile-based PA in enhancing the mental health outcomes of adults. In addition, this study highlights the effectiveness of progressive PA in reducing depressive symptoms and perceived stress, even in mobile-based settings, indicating a specific type of program that yields positive outcomes. Consequently, further development and implementation of innovative strategies that optimize the convenience and effectiveness of mobile-based interventions are warranted to support individuals in achieving better mental health.

Acknowledgments

This research was supported by the Bio & Medical Technology Development Program of the National Research Foundation (NRF) and funded by the Korean government (Ministry of Science and ICT; NRF-2021M3A9E4080780) and Hankuk University of Foreign Studies (2023).

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1319 KB - mhealth_v12i1e55578_app1.pdf]

References

1. American Psychiatric Association. The Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition. Washington, DC: American Psychiatric Association; 2013.
2. Depression and other common mental disorders: global health estimates. World Health Organization. 2024 Jan 3. URL: https://www.who.int/publications/i/item/depression-global-health-estimates [accessed 2024-01-03]
3. GBD results. Institute of Health Metrics and Evaluation. URL: https://vizhub.healthdata.org/gbd-results/ [accessed 2024-05-30]
4. Woody CA, Ferrari AJ, Siskind DJ, Whiteford HA, Harris MG. A systematic review and meta-regression of the prevalence and incidence of perinatal depression. J Affect Disord 2017 Sep;219:86-92. [doi: 10.1016/j.jad.2017.05.003] [Medline: 28531848]
5. Kessler RC, Petukhova M, Sampson NA, Zaslavsky AM, Wittchen HU. Twelve-month and lifetime prevalence and lifetime morbid risk of anxiety and mood disorders in the United States. Int J Methods Psychiatr Res 2012 Sep;21(3):169-184 [FREE Full text] [doi: 10.1002/mpr.1359] [Medline: 22865617]
6. Lee JG, Seonwoo SM, Choi MJ, Kim DH, Park GM, Go JS, et al. Calculation of socioeconomic costs of depression in Korea in 2019. J Korean Soc Biol Ther Psychiatry 2021;27(3):237-244. [doi: 10.22802/jksbtp.2021.27.3.237]

7. Evans-Lacko S, Aguilar-Gaxiola S, Al-Hamzawi A, Alonso J, Benjet C, Bruffaerts R, et al. Socio-economic variations in the mental health treatment gap for people with anxiety, mood, and substance use disorders: results from the WHO World Mental Health (WMH) surveys. Psychol Med 2017 Nov 27;48(9):1560-1571. [doi: 10.1017/s0033291717003336]
8. Depressive disorder (depression). World Health Organization. 2023 Mar 31. URL: https://www.who.int/news-room/fact-sheets/detail/depression [accessed 2023-03-31]
9. Marques A, Bordado J, Peralta M, Gouveia ER, Tesler R, Demetriou Y, et al. Cross-sectional and prospective relationship between physical activity and depression symptoms. Sci Rep 2020 Sep 30;10(1):16114 [FREE Full text] [doi: 10.1038/s41598-020-72987-4] [Medline: 32999306]
10. Schuch FB, Vasconcelos-Moreno MP, Borowsky C, Zimmermann AB, Rocha NS, Fleck MP. Exercise and severe major depression: effect on symptom severity and quality of life at discharge in an inpatient cohort. J Psychiatr Res 2015 Feb;61:25-32. [doi: 10.1016/j.jpsychires.2014.11.005] [Medline: 25439084]
11. Bridle C, Spanjers K, Patel S, Atherton NM, Lamb SE. Effect of exercise on depression severity in older people: systematic review and meta-analysis of randomised controlled trials. Br J Psychiatry 2012 Sep 02;201(3):180-185. [doi: 10.1192/bjp.bp.111.095174] [Medline: 22945926]
12. Kvam S, Kleppe CL, Nordhus IH, Hovland A. Exercise as a treatment for depression: a meta-analysis. J Affect Disord 2016 Sep 15;202:67-86. [doi: 10.1016/j.jad.2016.03.063] [Medline: 27253219]
13. Matias TS, Lopes MV, da Costa BG, Silva KS, Schuch FB. Relationship between types of physical activity and depression among 88,522 adults. J Affect Disord 2022 Jan 15;297:415-420. [doi: 10.1016/j.jad.2021.10.051] [Medline: 34715174]
14. Pearce M, Garcia L, Abbas A, Strain T, Schuch FB, Golubic R, et al. Association between physical activity and risk of depression: a systematic review and meta-analysis. JAMA Psychiatry 2022 Jun 01;79(6):550-559 [FREE Full text] [doi: 10.1001/jamapsychiatry.2022.0609] [Medline: 35416941]
15. Sabe M, Chen C, Sentissi O, Deenik J, Vancampfort D, Firth J, et al. Thirty years of research on physical activity, mental health, and wellbeing: a scientometric analysis of hotspots and trends. Front Public Health 2022 Aug 9;10:943435 [FREE Full text] [doi: 10.3389/fpubh.2022.943435] [Medline: 36016904]
16. Liguori G, American College of Sports Medicine (ACSM). ACSM's Guidelines for Exercise Testing and Prescription, 11th Edition. Alphen aan den Rijn, Netherlands: Wolters Kluwer; 2021.
17. Fuss J, Steinle J, Bindila L, Auer MK, Kirchherr H, Lutz B, et al. A runner's high depends on cannabinoid receptors in mice. Proc Natl Acad Sci U S A 2015 Oct 20;112(42):13105-13108 [FREE Full text] [doi: 10.1073/pnas.1514996112] [Medline: 26438875]
18. Rhodes RE, Janssen I, Bredin SS, Warburton DE, Bauman A. Physical activity: health impact, prevalence, correlates and interventions. Psychol Health 2017 Aug 30;32(8):942-975. [doi: 10.1080/08870446.2017.1325486] [Medline: 28554222]
19. Stults-Kolehmainen MA, Sinha R. The effects of stress on physical activity and exercise. Sports Med 2014 Jan;44(1):81-121 [FREE Full text] [doi: 10.1007/s40279-013-0090-5] [Medline: 24030837]
20. Krogh J, Rostrup E, Thomsen C, Elfving B, Videbech P, Nordentoft M. The effect of exercise on hippocampal volume and neurotrophines in patients with major depression--a randomized clinical trial. J Affect Disord 2014 Aug;165:24-30. [doi: 10.1016/j.jad.2014.04.041] [Medline: 24882173]
21. Firth J, Stubbs B, Vancampfort D, Schuch F, Lagopoulos J, Rosenbaum S, et al. Effect of aerobic exercise on hippocampal volume in humans: a systematic review and meta-analysis. Neuroimage 2018 Feb 01;166:230-238 [FREE Full text] [doi: 10.1016/j.neuroimage.2017.11.007] [Medline: 29113943]
22. Marosi K, Bori Z, Hart N, Sárga L, Koltai E, Radák Z, et al. Long-term exercise treatment reduces oxidative stress in the hippocampus of aging rats. Neuroscience 2012 Dec 13;226:21-28. [doi: 10.1016/j.neuroscience.2012.09.001] [Medline: 22982624]
23. Zheng H, Liu Y, Li W, Yang B, Chen D, Wang X, et al. Beneficial effects of exercise and its molecular mechanisms on depression in rats. Behav Brain Res 2006 Mar 15;168(1):47-55 [FREE Full text] [doi: 10.1016/j.bbr.2005.10.007] [Medline: 16290283]
24. Heijnen S, Hommel B, Kibele A, Colzato LS. Neuromodulation of aerobic exercise-a review. Front Psychol 2015;6:1890 [FREE Full text] [doi: 10.3389/fpsyg.2015.01890] [Medline: 26779053]
25. Garber CE, Blissmer B, Deschenes MR, Franklin BA, Lamonte MJ, Lee IM, et al. American College of Sports Medicine position stand. Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: guidance for prescribing exercise. Med Sci Sports Exerc 2011 Jul;43(7):1334-1359. [doi: 10.1249/MSS.0b013e318213fefb] [Medline: 21694556]
26. Committee on Environmental Health, Tester JM. The built environment: designing communities to promote physical activity in children. Pediatrics 2009 Jun;123(6):1591-1598. [doi: 10.1542/peds.2009-0750] [Medline: 19482771]
27. Office of the Surgeon General, U. S. Department Human Services, Office Surgeon. The Surgeon General's Vision for a Healthy and Fit Nation. Scotts Valley, CA: CreateSpace Independent Publishing Platform; 2012.
28. Raza A, Pulakka A, Hanson LL, Westerlund H, Halonen JI. Distance to sports facilities and low frequency of exercise and obesity: a cross-sectional study. BMC Public Health 2022 Nov 07;22(1):2036 [FREE Full text] [doi: 10.1186/s12889-022-14444-7] [Medline: 36344975]

29. Ani R, Zheng J. Proximity to an exercise facility and physical activity in China. Southeast Asian J Trop Med Public Health 2014 Nov;45(6):1483-1491. [Medline: 26466435]
30. Panter JR, Jones AP. Associations between physical activity, perceptions of the neighbourhood environment and access to facilities in an English city. Soc Sci Med 2008 Dec;67(11):1917-1923. [doi: 10.1016/j.socscimed.2008.09.001] [Medline: 18835074]
31. Safe, affordable, and accessible physical activity. Harvard T.H. Chan. URL: https://tinyurl.com/4j3a4va2 [accessed 2024-05-30]
32. AlSwayied G, Guo H, Rookes T, Frost R, Hamilton FL. Assessing the acceptability and effectiveness of mobile-based physical activity interventions for midlife women during menopause: systematic review of the literature. JMIR Mhealth Uhealth 2022 Dec 09;10(12):e40271 [FREE Full text] [doi: 10.2196/40271] [Medline: 36485026]
33. Buckingham SA, Williams AJ, Morrissey K, Price L, Harrison J. Mobile health interventions to promote physical activity and reduce sedentary behaviour in the workplace: a systematic review. Digit Health 2019;5:2055207619839883 [FREE Full text] [doi: 10.1177/2055207619839883] [Medline: 30944728]
34. Firth J, Torous J, Nicholas J, Carney R, Rosenbaum S, Sarris J. Can smartphone mental health interventions reduce symptoms of anxiety? A meta-analysis of randomized controlled trials. J Affect Disord 2017 Aug 15;218:15-22 [FREE Full text] [doi: 10.1016/j.jad.2017.04.046] [Medline: 28456072]
35. Hutchesson MJ, Gough C, Müller AM, Short CE, Whatnall MC, Ahmed M, et al. eHealth interventions targeting nutrition, physical activity, sedentary behavior, or obesity in adults: a scoping review of systematic reviews. Obes Rev 2021 Oct 23;22(10):e13295. [doi: 10.1111/obr.13295] [Medline: 34159684]
36. Haller N, Lorenz S, Pfirrmann D, Koch C, Lieb K, Dettweiler U, et al. Individualized web-based exercise for the treatment of depression: randomized controlled trial. JMIR Ment Health 2018 Oct 12;5(4):e10698 [FREE Full text] [doi: 10.2196/10698] [Medline: 30314962]
37. D'Amore C, Reid JC, Chan M, Fan S, Huang A, Louie J, et al. Interventions including smart technology compared with face-to-face physical activity interventions in older adults: systematic review and meta-analysis. J Med Internet Res 2022 Oct 31;24(10):e36134 [FREE Full text] [doi: 10.2196/36134] [Medline: 36315229]
38. Mizuta R, Maeda N, Tashiro T, Suzuki Y, Kuroda S, Ishida A, et al. Effectiveness of metaverse space-based exercise video distribution in young adults: randomized controlled trial. JMIR Mhealth Uhealth 2024 Jan 16;12:e46397 [FREE Full text] [doi: 10.2196/46397] [Medline: 38227355]
39. Murray A, Marenus M, Cahuas A, Friedman K, Ottensoser H, Kumaravel V, et al. The impact of web-based physical activity interventions on depression and anxiety among college students: randomized experimental trial. JMIR Form Res 2022 Apr 01;6(4):e31839 [FREE Full text] [doi: 10.2196/31839] [Medline: 35363151]
40. Lambert JD, Greaves CJ, Farrand P, Price L, Haase AM, Taylor AH. Web-based intervention using behavioral activation and physical activity for adults with depression (The eMotion Study): pilot randomized controlled trial. J Med Internet Res 2018 Jul 16;20(7):e10112 [FREE Full text] [doi: 10.2196/10112] [Medline: 30012547]
41. Kim H, Lee K, Lee YH, Park Y, Park Y, Yu Y, et al. The effectiveness of a mobile phone-based physical activity program for treating depression, stress, psychological well-being, and quality of life among adults: quantitative study. JMIR Mhealth Uhealth 2023 Jun 26;11:e46286 [FREE Full text] [doi: 10.2196/46286] [Medline: 37358905]
42. Singh NA, Clements KM, Fiatarone MA. A randomized controlled trial of progressive resistance training in depressed elders. J Gerontol A Biol Sci Med Sci 1997 Jan 01;52(1):M27-M35. [doi: 10.1093/gerona/52a.1.m27] [Medline: 9008666]
43. Singh NA, Stavrinos TM, Scarbek Y, Galambos G, Liber C, Fiatarone Singh MA. A randomized controlled trial of high versus low intensity weight training versus general practitioner care for clinical depression in older adults. J Gerontol A Biol Sci Med Sci 2005 Jun;60(6):768-776. [doi: 10.1093/gerona/60.6.768] [Medline: 15983181]
44. Thomas S, Reading J, Shephard RJ. Revision of the Physical Activity Readiness Questionnaire (PAR-Q). Can J Sport Sci 1992 Dec;17(4):338-345. [Medline: 1330274]
45. Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Method 2007 May;39(2):175-191. [doi: 10.3758/BF03193146]
46. Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. Clin Psychol Rev 2012 Jun;32(4):329-342. [doi: 10.1016/j.cpr.2012.02.004] [Medline: 22466510]
47. Park SJ, Choi HR, Choi JH, Kim KW, Hong JP. Reliability and validity of the Korean version of the Patient Health Questionnaire-9 (PHQ-9). Anxiety Mood 2010;6(2):119-124.
48. Frank SH, Zyzanski SJ. Stress in the clinical setting: the Brief Encounter Psychosocial Instrument. J Fam Pract 1988 May;26(5):533-539. [Medline: 3367118]
49. Kim JW, Kim JS, Lim JJ, Kim YS, Lee O, Choi BY, et al. Relationship between physical activity and depression according to the presence of disease in Koreans over the age of 40. Korean J Sport Sci 2020 Mar;31(1):24-34. [doi: 10.24985/kjss.2020.31.1.24]
50. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
51. Min SK, Kim KI, Lee CI, Jung YC, Suh SY, Kim DK. Development of the Korean versions of WHO quality of life scale and WHOQOL-BREF. Qual Life Res 2002 Sep;11(6):593-600. [doi: 10.1023/a:1016351406336] [Medline: 12206580]

52. Addis ME, Mahalik JR. Men, masculinity, and the contexts of help seeking. Am Psychol 2003 Jan;58(1):5-14. [doi: 10.1037/0003-066x.58.1.5] [Medline: 12674814]
53. Kuehner C. Why is depression more common among women than among men? Lancet Psychiatry 2017 Feb;4(2):146-158 [FREE Full text] [doi: 10.1016/S2215-0366(16)30263-2] [Medline: 27856392]
54. Kandola A, Ashdown-Franks G, Hendrikse J, Sabiston CM, Stubbs B. Physical activity and depression: towards understanding the antidepressant mechanisms of physical activity. Neurosci Biobehav Rev 2019 Dec;107:525-539. [doi: 10.1016/j.neubiorev.2019.09.040] [Medline: 31586447]
55. He SB, Tang WG, Tang WJ, Kao XL, Zhang CG, Wong XT. Exercise intervention may prevent depression. Int J Sports Med 2012 Jul;33(7):525-530. [doi: 10.1055/s-0032-1306325] [Medline: 22504906]
56. Lazarus RS, Folkman S. Stress, Appraisal, and Coping. Cham, Switzerland: Springer; 1984.
57. Taylor NF, Dodd KJ, Damiano DL. Progressive resistance exercise in physical therapy: a summary of systematic reviews. Phys Ther 2005 Nov 01;85(11):1208-1223. [doi: 10.1093/ptj/85.11.1208]
58. O'Bryan SJ, Giuliano C, Woessner MN, Vogrin S, Smith C, Duque G, et al. Progressive resistance training for concomitant increases in muscle strength and bone mineral density in older adults: a systematic review and meta-analysis. Sports Med 2022 Aug;52(8):1939-1960 [FREE Full text] [doi: 10.1007/s40279-022-01675-2] [Medline: 35608815]
59. Aube D, Wadhi T, Rauch J, Anand A, Barakat C, Pearson J, et al. Progressive resistance training volume: effects on muscle thickness, mass, and strength adaptations in resistance-trained individuals. J Strength Cond Res 2022 Mar 01;36(3):600-607. [doi: 10.1519/JSC.0000000000003524] [Medline: 32058362]
60. Howard JL, Bureau J, Guay F, Chong JX, Ryan RM. Student motivation and associated outcomes: a meta-analysis from self-determination theory. Perspect Psychol Sci 2021 Nov;16(6):1300-1323. [doi: 10.1177/1745691620966789] [Medline: 33593153]
61. Sheehan RB, Herring MP, Campbell MJ. Associations between motivation and mental health in sport: a test of the hierarchical model of intrinsic and extrinsic motivation. Front Psychol 2018 May 8;9:707 [FREE Full text] [doi: 10.3389/fpsyg.2018.00707] [Medline: 29867672]
62. Lauderdale ME, Yli-Piipari S, Irwin CC, Layne TE. Gender differences regarding motivation for physical activity among college students: a self-determination approach. Phys Educ 2015 Oct 06;72(5):153-172. [doi: 10.18666/TPE-2015-V72-I5-4682]
63. Apter MJ. Reversal theory as a theory of the emotions. In: Apter MJ, Kerr JH, Cowles MP, editors. Advances in Psychology. Amsterdam, The Netherlands: Elsevier; 1988.
64. Kerr JH, Kuk G. The effects of low and high intensity exercise on emotions, stress and effort. Psychol Sport Exerc 2001 Jul;2(3):173-186. [doi: 10.1016/S1469-0292(00)00021-2]
65. O'Brien KK, Tynan AM, Nixon SA, Glazier RH. Effectiveness of Progressive Resistive Exercise (PRE) in the context of HIV: systematic review and meta-analysis using the Cochrane Collaboration protocol. BMC Infect Dis 2017 Apr 12;17(1):268 [FREE Full text] [doi: 10.1186/s12879-017-2342-8] [Medline: 28403830]
66. Gerritsen JK, Vincent AJ. Exercise improves quality of life in patients with cancer: a systematic review and meta-analysis of randomised controlled trials. Br J Sports Med 2016 Jul;50(13):796-803. [doi: 10.1136/bjsports-2015-094787] [Medline: 26719503]
67. Bucher MA, Suzuki T, Samuel DB. A meta-analytic review of personality traits and their associations with mental health treatment outcomes. Clin Psychol Rev 2019 Jun;70:51-63. [doi: 10.1016/j.cpr.2019.04.002] [Medline: 30981042]
68. Gao S, Zhang X, Xu X. A meta-analysis of the relationship between self-esteem and mental health: the sample of chinese college students. Adv Psychol Sci 2015;23(9):1499-1507. [doi: 10.3724/sp.j.1042.2015.01499]
69. Hu T, Zhang D, Wang J. A meta-analysis of the trait resilience and mental health. Pers Individ Dif 2015 Apr;76:18-27. [doi: 10.1016/j.paid.2014.11.039]
70. Beauchamp MR, Hulteen RM, Ruissen GR, Liu Y, Rhodes RE, Wierts CM, et al. Online-delivered group and personal exercise programs to support low active older adults' mental health during the COVID-19 pandemic: randomized controlled trial. J Med Internet Res 2021 Jul 30;23(7):e30709 [FREE Full text] [doi: 10.2196/30709] [Medline: 34328433]
71. Winship C, Mare RD. Models for sample selection bias. Annu Rev Sociol 1992 Aug;18:327-350. [doi: 10.1146/annurev.so.18.080192.001551]
72. Brierley G, Brabyn S, Torgerson D, Watson J. Bias in recruitment to cluster randomized trials: a review of recent publications. J Eval Clin Pract 2012 Aug;18(4):878-886. [doi: 10.1111/j.1365-2753.2011.01700.x] [Medline: 21689213]
73. Ludwig K, Arthur R, Sculthorpe N, Fountain H, Buchan DS. Text messaging interventions for improvement in physical activity and sedentary behavior in youth: systematic review. JMIR Mhealth Uhealth 2018 Sep 17;6(9):e10799 [FREE Full text] [doi: 10.2196/10799] [Medline: 30224335]

Abbreviations

BEPSI-K: Brief Encounter Psychosocial Instrument-Korean version
CONSORT: Consolidated Standards of Reporting Trials
GAD-7: Generalized Anxiety Disorder-7

PA: physical activity
PHQ-9: Patient Health Questionnaire-9
RMANOVA: repeated measures ANOVA

Edited by L Buis; submitted 17.12.23; peer-reviewed by C Chen, W Chiu, M Taheri; comments to author 22.02.24; revised version received 19.03.24; accepted 22.04.24; published 12.06.24.

Please cite as:
Lee YH, Kim H, Hwang J, Noh S
Effectiveness of Mobile-Based Progressive and Fixed Physical Activity on Depression, Stress, Anxiety, and Quality of Life Outcomes Among Adults in South Korea: Randomized Controlled Trial
JMIR Mhealth Uhealth 2024;12:e55578
URL: https://mhealth.jmir.org/2024/1/e55578
doi:10.2196/55578
PMID:38865705

Original Paper

Improvement and Maintenance of Clinical Outcomes in a Digital Mental Health Platform: Findings From a Longitudinal Observational Real-World Study

Lydia G Roos[1,2,3], PhD; Sara J Sagui-Henson[4], PhD; Cynthia Castro Sweet[4], PhD; Camille E Welcome Chamberlain[4], MSc; Brooke J Smith[4], MSc

[1]Department of Psychiatry and Biobehavioral Sciences, University of California, Los Angeles, Los Angeles, CA, United States

[2]School of Medicine, Stanford University, Stanford, CA, United States

[3]EvolveWell Research Partners, Cincinnati, OH, United States

[4]Modern Health, San Francisco, CA, United States

Corresponding Author:
Lydia G Roos, PhD
Department of Psychiatry and Biobehavioral Sciences
University of California, Los Angeles
760 Westwood Plaza
Los Angeles, CA, 90024
United States
Phone: 1 8435446005
Email: lroos@mednet.ucla.edu

Abstract

Background: Digital mental health services are increasingly being provided by employers as health benefit programs that can improve access to and remove barriers to mental health care. Stratified care models, in particular, offer personalized care recommendations that can offer clinically effective interventions while conserving resources. Nonetheless, clinical evaluation is needed to understand their benefits for mental health and their use in a real-world setting.

Objective: This study aimed to examine the changes in clinical outcomes (ie, depressive and anxiety symptoms and well-being) and to evaluate the use of stratified blended care among members of an employer-sponsored digital mental health benefit.

Methods: In a large prospective observational study, we examined the changes in depressive symptoms (9-item Patient Health Questionnaire), anxiety symptoms (7-item Generalized Anxiety Disorder scale), and well-being (5-item World Health Organization Well-Being Index) for 3 months in 509 participants (mean age 33.9, SD 8.7 years; women: n=312, 61.3%; men: n=175, 34.4%; nonbinary: n=22, 4.3%) who were newly enrolled and engaged in care with an employer-sponsored digital mental health platform (Modern Health Inc). We also investigated the extent to which participants followed the recommendations provided to them through a stratified blended care model.

Results: Participants with elevated baseline symptoms of depression and anxiety exhibited significant symptom improvements, with a 37% score improvement in depression and a 29% score improvement in anxiety (*P* values <.001). Participants with baseline scores indicative of poorer well-being also improved over the study period (90% score improvement; *P*=.002). Furthermore, over half exhibited clinical improvement or recovery for depressive symptoms (n=122, 65.2%), anxiety symptoms (n=127, 59.1%), and low well-being (n=82, 64.6%). Among participants with mild or no baseline symptoms, we found high rates of maintenance for low depressive (n=297, 92.2%) and anxiety (n=255, 86.7%) symptoms and high well-being (n=344, 90.1%). In total, two-thirds of the participants (n=343, 67.4%) used their recommended care, 16.9% (n=86) intensified their care beyond their initial recommendation, and 15.7% (n=80) of participants underused care by not engaging with the highest level of care recommended to them.

Conclusions: Participants with elevated baseline depressive or anxiety symptoms improved their mental health significantly from baseline to follow-up, and most participants without symptoms or with mild symptoms at baseline maintained their mental health over time. In addition, engagement patterns indicate that the stratified blended care model was efficient in matching individuals with the most effective and least costly care while also allowing them to self-determine their care and use combinations of services that best fit their needs. Overall, the results of this study support the clinical effectiveness of the platform for improving

and preserving mental health and support the utility and effectiveness of stratified blended care models to improve access to and use of digitally delivered mental health services.

(JMIR Mhealth Uhealth 2024;12:e48298) doi:10.2196/48298

KEYWORDS

digital mental health; employee health; depression; anxiety; well-being; mobile phone

Introduction

Background

The prevalence of mental health needs in the United States has been on an upward trend in recent years, with 21% of adults meeting criteria for a mental, behavioral, or emotional disorder in 2020, up from 18% in 2010 [1]. National data indicate that <50% of people with mental health concerns are able to access mental health services [1]. Traditional models of mental health care are inadequate, as they rely heavily on high-cost providers delivering scheduled, time-limited encounters, and training programs are decades away from adequately closing the provider shortage gaps [2,3]. In addition to a pervasive shortage of mental health professionals to provide needed care [4], issues related to cost, accessibility, and stigma also prevent individuals from accessing evidence-based care to address mental health concerns [1]. Thus, innovative models for mental health care that are scalable, resource sensitive, and acceptable to individuals are needed to sufficiently improve the provision of robust mental health care in the United States.

Innovative and flexible models of mental health care leverage technology and telecommunications to provide more rapid and scalable access to a myriad of mental health services, from self-guided "self-help" techniques to access to providers who deliver secure, remote care [5]. An advantageous feature of digital mental health platforms is their flexibility in offering a variety of care modalities, enabling users to exercise their preferences in accessing care in a way that best fits their needs and comfort level. Stepped care delivery models further accelerate improvements in mental health care access and affordability. There are currently 2 models: progressive and stratified. A progressive model recommends the lowest-intensity intervention first to all individuals and intensifies care if or when symptoms do not improve [6]. This is the prevailing system used by the United Kingdom National Health Service's Talking Therapies program (formerly called Improving Access to Psychological Therapies) [7]. The evidence supporting this model suggests that patients' baseline symptomatology does not impact the efficacy of low-intensity or high-intensity treatments [8,9].

However, recent research suggests the Improving Access to Psychological Therapies program may not adequately support or match the complexity of patients' presenting mental health issues [10]. There are several criticisms [11] of the progressive approach, including (1) low-intensity interventions may not be suitable or acceptable for everyone; (2) patients who do not respond to low-intensity treatment may develop negative attitudes toward treatment or be deterred from undergoing further treatment; (3) engaging in high-intensity treatment after a minimal intervention may be unnecessarily burdensome; and (4) those with greater clinical needs may have to wait longer to receive a more effective level of care and, in the interim, experience an exacerbation of symptoms and additional impairment.

As an alternative, a *stratified* model considers patient characteristics, preferences, and baseline mental health symptoms to identify and deliver the most clinically effective yet least burdensome and least costly initial intervention from a range of care modalities of different intensities [12]. Stratifying care with personalized recommendations is thought to be more patient-centered and is responsive to key drawbacks of the progressive stepped care approach [11]. In some stratified systems, more specifically blended care models, patients can access multiple modalities simultaneously; that is, they can use their recommended treatment modality as well as additional modalities of lower intensity than their recommended starting point (eg, digital tools plus provider sessions, as opposed to digital-only or provider-only session). While advantageous from a delivery perspective and found to be effective [13,14], stratified, blended models can be more difficult to evaluate because of the complexity and variety of care pathways offered to patients and the variability of "blends" that patients may use at different points in time.

Research Questions

Prior findings regarding the clinical superiority of progressive models over stratified models are mixed [15,16]. These mixed results and the criticisms of progressive stepped care suggest that by incorporating patient-level factors to match individuals with the most effective yet least costly mental health services, stratified, blended models can offer more personalized care and increase access while optimizing resources. In this study, we examined the effectiveness of and engagement in a digital mental health platform that uses a stratified blended care approach to deliver therapy, coaching, and self-guided digital services. Specifically, we tested two research questions: (1) Was this approach clinically effective, that is, did participants with elevated baseline mental health symptoms significantly improve their mental health as defined by depression, anxiety, and well-being scores, and did participants without elevated baseline symptoms maintain good mental health from baseline to 3-month follow-up? (2) Was this approach effective in stratifying resources, that is, did people follow, underuse, or overuse mental health services at the levels they were recommended?

Methods

Design and Participants

This investigation was conducted as part of a larger prospective, observational study of individuals who received services through

an employer-sponsored digital mental health benefits platform (Modern Health Inc). The study time frame was September 20, 2021, through May 31, 2022. Participants were eligible if they were aged ≥18 years; were based in the United States; were onboarded with the employer-sponsored mental health benefit; had access to a smartphone, a tablet, or a computer; and had engaged with at least 1 piece of digital content or matched with a coach or therapist (see the Intervention section for more detailed descriptions of the services).

Ethical Considerations

This study protocol was reviewed and approved by the Western Clinical Group Institutional Review Board (protocol no 1316167). Participants provided informed consent to participate in this investigation. The Western Clinical Group Institutional Review Board authorized a waiver of documentation of consent for the team to collect consent through secure electronic methods.

All data were deidentified for the purpose of analyses. Participants were compensated with a US $25 digital gift card upon completion of each of the 3 surveys in this investigation.

Procedures

Participants registered for an account through a mobile app or a website and completed onboarding assessments, including questions designed to assess participants' areas of focus and care modality preferences, as well as validated measures to assess depressive and anxiety symptoms and well-being (described in the Study Measures section). A proprietary algorithm factored in a combination of the participant's clinical acuity, their modality preference, and their topic of focus to recommend an initial care pathway (eg, digital programs, coaching, and therapy). Participants were not required to follow the recommendation; instead, it was offered as an appropriate suggested starting point. Participants who were recommended therapy also had access to coaching and digital content, and those who were recommended coaching also had access to digital content as part of their recommendations (Figure 1). Participants could also self-refer or be referred by a provider to a different combination of care from their recommended combinations.

Figure 1. Stratified blended care model incorporating care preferences and clinical assessments into personalized care recommendations. GAD-7: 7-item Generalized Anxiety Disorder scale; PHQ-9: 9-item Patient Health Questionnaire; WHO-5: 5-item World Health Organization Well-Being Index.

Eligible members were invited to complete a screener for the study via email, which collected their demographic information (age, gender identity, race, and ethnicity). All screeners were sent within 2 weeks of onboarding, with most members receiving the screener approximately 1 week after onboarding. During this time, the members were able to engage with the digital mental health services outlined in the Intervention section. A total of 2 factors determined the length of time it took to send the screener: research staff availability and demographic balancing. Limits were set such that enrolled study participants reflected the current distribution of age, gender, ethnic/racial identity, and mental health symptom acuity observed in the platform's commercial population. Out of the 8786 individuals who were eligible and invited to participate, 950 (10.81%) enrolled, provided informed consent, and completed the baseline survey, hosted by Qualtrics (**Qualtrics** International Inc). They were then emailed a link to complete a follow-up survey 12 weeks after the baseline survey. The baseline and follow-up surveys each took 30 to 45 minutes to complete.

Intervention

Digital Health Services

Participants could engage in all the following digital mental health services. All services were paid for by the participant's employer, at no cost to the individual.

Telecoaching and Teletherapy

Coaches certified by the International Coaching Federation accredited program provided telecoaching services, and therapists who were licensed and had an advanced degree in clinical psychology or a related field (eg, PhD, PsyD, licensed clinical social worker, licensed marriage and family therapist, or licensed professional counselor) provided teletherapy services to participants. All visits were conducted via a secured videoconferencing platform. Participants could also communicate with their therapist or coach through in-app

messaging. All coaches had at least 150 hours of coaching experience, were vetted by a provider management team to ensure their work aligned with evidence-based practices, and completed an additional 6 hours of training from the company clinical strategy team in evidence-based techniques (eg, cognitive behavioral approaches) and culturally centered care. Coaches were also vetted and trained on how to assess for high-risk situations that may require a participant's referral to a therapist or crisis resource.

Therapists were selected for their use of evidence-based practices, such as cognitive behavioral therapy and acceptance and commitment therapy. All coaches and therapists were trained on the company's proprietary model of care. The number of therapy and coaching sessions attended by participants depended on the number of sessions covered by their employer, as well as on personal preferences and their level of need.

Self-Guided Digital Content

All participants had unlimited access to a digital library of mental health programs and resources that they could access at any time. These included short (2 minutes each) daily exercises; interactive programs and podcasts (2 to 15 minutes each); mindfulness exercises such as meditations and breathing exercises (2 to 15 minutes each); and self-paced structured educational lessons (several chapters of content, akin to self-help workbooks, that are paced to be completed over several weeks). Digital programs were developed and designed by an in-house team of clinical psychologists and covered topics such as emotions, relationships, professional life, healthy lifestyles, and finances. Engagement across all digital resources was combined in analyses to represent total digital program engagement.

Study Measures

Demographic Information

Participants self-reported demographic characteristics such as age; gender identity (to select all that applied from a list: agender, genderqueer or genderfluid, Māhū [third gender], man, muxe, nonbinary, questioning or unsure, 2-spirit, woman, prefer to self-describe, and prefer not to say); and race and ethnicity (to select all that applied from a list: American Indian or Alaska Native; Asian; Black or African American; Hawaiian or Pacific Islander; Hispanic, Latinx, or Spanish; White [not Hispanic or Latinx]; multiracial) during the screener. On the basis of participants' demographics, 3 categories were used in analyses: men, women, and nonbinary (all other categories except "prefer to self-describe" or "prefer not to say" collapsed).

Depressive Symptoms

The 9-item Patient Health Questionnaire (PHQ-9) [17] was used to assess the presence of depression symptoms over the past 2 weeks at baseline and follow-up. Participants responded on a 4-point scale (0="not at all" to 3="nearly every day"). Possible ranges for scores include 0-27, with higher scores indicating a higher severity of depression symptomatology. The clinically validated cutoff for probable depression ("high risk") is ≥10 [17], and clinical improvement was indicated when participants' scores decreased by >6 points [18].

Anxiety Symptoms

The 7-item Generalized Anxiety Disorder Questionnaire (GAD-7) [19] was used to assess the presence of anxiety symptoms over the past 2 weeks at baseline and follow-up. Participants responded on a 4-point scale (0="not at all" to 3="nearly every day"). The possible range was 0-21, with higher scores indicating a higher severity of anxiety symptomatology. The clinical cutoff score for probable anxiety disorder ("high risk") is ≥8 [20], and improvement was indicated when participants' scores decreased by >4 points [21].

Well-Being

The 5-item World Health Organization Well-Being Index (WHO-5) [22] was used to assess well-being over the past 2 weeks at baseline and follow-up. Participants responded on a 6-point scale (0="at no time" to 5="all of the time"). Scores are summed and multiplied by 4, giving a total range of 0-100, with higher scores indicating greater subjective well-being. The clinical cutoff indicating low well-being ("high risk") is ≤28; recovery was indicated when the baseline score was <28 and the 3-month score was >28. Prior research has defined clinical improvement in well-being as an increase of at least 10 points [22,23].

Platform Engagement

To operationalize care engagement, we categorized participants based on whether they (1) engaged in care at the level recommended, (2) underused care, or (3) overused care. We defined engaging or following the care recommendation as a participant using their level of recommended care (with the ability to use anything below that level of care). Overusing care occurred when participants used a higher level of care than what they were initially recommended (intensified their care above what was originally recommended, regardless of whether that use step-up was self-referred or referred by a provider). Because participants who were originally recommended therapy as their care modality could not step up their care any higher, only participants who were recommended coaching or digital content or coaching could overuse care. Finally, we defined the underuse of care as participants using a lower level of care than their recommendation and not using any higher care modality. Because participants who were recommended digital content could not use a lower level of care, only participants who were recommended therapy or coaching could underuse it.

Statistical Analysis

Analysis of participant demographics and preliminary analyses were conducted using descriptive statistics and frequencies. We used McNemar χ^2 tests and paired sample 2-tailed *t* tests to examine the clinical effectiveness of the platform, that is, mental health improvement, recovery, or maintenance. Specifically, we used McNemar χ^2 tests to assess whether the percentage of participants categorized as high risk in each mental health outcome significantly decreased from baseline to follow-up. We used paired sample *t* tests to assess whether changes in each outcome (measured continuously) were significantly improved from baseline to follow-up.

We also reported the percentage of participants who reliably improved, recovered, and maintained their mental health. For participants who met the clinical cutoff for outcomes at baseline ("high risk"), we examined improvement or recovery in symptoms from baseline to the 3-month follow-up. Improvement in each measure was indicated when participants' scores changed by established clinical thresholds (see the Study Measures section), and recovery in each measure was indicated when participants met or exceeded the clinical cutoff at baseline (ie, were categorized as "high risk") but did not meet the clinical cutoff at follow-up (ie, were categorized as "low risk"). Maintenance was indicated when participants remained below the clinical cutoff from baseline ("low risk") to follow-up. Continuous variables (ie, depressive and anxiety symptoms and well-being at baseline and follow-up) were examined for kurtosis and skewness; all values were between −1 and 1. Thus, original values were used in analyses.

We assessed our research question regarding use-care recommendations using frequencies and descriptive statistics. We report the percentage of participants who engaged in their recommended services, the percentage who overused services, and the percentage who underused services.

Results

Study Participants and Preliminary Analyses

A total of 950 members completed the baseline survey, of which 696 (73.2%) completed the follow-up survey. Of the 696 with full data, 528 (75.9%) members engaged with the platform (eg, used digital content, had teletherapy, or had a telecoaching visit) at least once between baseline and follow-up. There were 10 (n=528, 1.9%) participants who were not included in data analyses because they were recommended a care modality for which we did not have engagement data (ie, group psychoeducation sessions), and 9 (n=528, 1.7%) participants were not provided with a recommended care plan for reasons unknown. There were no significant baseline clinical differences between people who did not engage with anything on the Modern Health app after baseline (n=168) and those who did engage. As engagement with the app at least once during the analytic time frame was necessary for inclusion in the study, the remaining analyses included the final 509 (53.6% of enrolled) participants for every outcome except for anxiety, for which there were missing data, that yielded a total of 506 (53.2% of enrolled) for anxiety analyses.

The *t* test and χ^2 analyses comparing baseline data from individuals who met final eligibility criteria (509/950, 53.6%) versus those who did not meet eligibility criteria (441/950, 46.4%) revealed no significant differences in age, gender identity, or race and ethnicity at baseline, and the groups were not significantly different on depression, anxiety, or well-being scores when assessed continuously (*P* values >.10). The participants who met the final eligibility criteria were significantly less likely to meet the clinical cutoffs for depressive and anxiety symptoms and low well-being (*P* values <.001).

The descriptive statistics of the sample are provided in Table 1.

Of the 509 participants, 342 (67.2%) engaged with digital content on the app at least once. A total of 159 (31.2%) participants attended at least 1 therapy visit, and 296 (58.2%) participants attended at least 1 coaching visit. The participants that attended therapy or coaching visits typically saw 1 provider over the course of the study (149/159, 93.7% and 283/296, 95.6%, respectively). The maximum number of therapists and coaches seen by any 1 participant were 3 and 2, respectively.

Table 1. Descriptive statistics of sample[a] (n=509).

	Values	Participants at baseline, n (%)	Participants at 3-month follow-up, n (%)
Age (years), mean (SD)	33.9 (8.7)	—[b]	—
Race/ethnicity, n (%)			
Asian	91 (17.9)	—	—
American Indian or Alaska Native	1 (0.2)	—	—
Black	33 (6.5)	—	—
Hispanic, Latinx, or Spanish origin	44 (8.6)	—	—
White (non-Hispanic or Latinx)	305 (59.9)	—	—
Multiracial	33 (6.5)	—	—
Gender identity, n (%)			
Women	312 (61.3)	—	—
Men	175 (34.4)	—	—
Nonbinary	22 (4.3)	—	—
Scores[c]			
Depressive symptoms above clinical cutoff	—	187 (36.7)	106 (20.8)
Anxiety symptoms above clinical cutoff	—	215 (42.2)	156 (30.8)
Well-being below clinical cutoff	—	127 (25)	94 (18.5)

[a]n=509 for depressive symptoms and well-being and n=506 for anxiety symptoms.

[b]—: not available.

[c]Possible ranges for scores include 0 to 27 for depressive symptoms, 0-21 for anxiety symptoms, and 0-100 for well-being. Clinical cutoffs at baseline were ≥10 for depressive symptoms, ≥8 for anxiety symptoms, and ≤28 for well-being.

Improvement and Recovery in Mental Health Symptoms Among Participants at Higher Risk at Baseline

All improvement, recovery, and change in mental health results among participants who met the clinical cutoff on each measure at baseline are presented in Tables 2 and 3.

Participants at a higher risk for depressive symptoms at baseline reported a statistically significant 37% improvement in PHQ-9 scores at follow-up, on average, with 65.2% (122/187) experiencing clinically meaningful improvement or recovery (P<.001). Participants at a higher risk for anxiety symptoms at baseline reported a statistically significant 29% improvement in GAD-7 scores at follow-up, on average, with 59.1% (127/215) experiencing clinically meaningful improvement or recovery (P<.001). Participants at a higher risk for lower well-being at baseline reported a statistically significant 90% improvement in WHO-5 scores at follow-up, on average, with 65.6% (82/127) experiencing clinically meaningful improvement or recovery (P=.002). Post hoc sensitivity analyses revealed that our models among higher-risk participants were sensitive to detect small effect sizes (Cohen d_z=0.20 for depressive symptoms, 0.19 for anxiety symptoms, and 0.25 for well-being), with 80% power and α=.05.

Table 2. Clinical improvement, recovery, and change in mental health from baseline to follow-up among participants at higher risk at baseline (n=509).

Baseline symptoms[a]	Improvement, n (%)	Recovery, n (%)	Improvement and recovery, n (%)	Improvement or recovery, n (%)
Depressive symptoms (n=187)	83 (44.4)	106 (56.7)	67 (35.8)	122 (65.2)
Anxiety symptoms (n=215)	107 (49.8)	98 (45.6)	78 (36.3)	127 (59.1)
Well-being (n=127)	76 (59.8)	71 (55.9)	65 (51.2)	82 (64.6)

[a]Possible ranges for scores include 0-27 for depressive symptoms, 0-21 for anxiety symptoms, and 0-100 for well-being.

Table 3. Change in mental health from baseline to follow-up among participants at higher risk at baseline (n=509).

Baseline symptoms[a]	Baseline, mean (SD)[b]	Follow-up, mean (SD)[c]	Cohen *d*	*t* test (*df*)	*P* value
Depressive symptoms (n=187)	14.13 (3.43)	8.91 (4.25)	1.16	15.81 (186)	<.001
Anxiety symptoms (n=215)	12.68 (3.69)	8.99 (4.76)	0.75	11.02 (214)	<.001
Well-being (n=127)	20.44 (6.80)	38.80 (19.75)	−0.93	−10.50 (126)	<.001

[a]Possible ranges for scores include 0-27 for depressive symptoms, 0-21 for anxiety symptoms, and 0-100 for well-being.

[b]Improvement was indicated when depressive and anxiety symptom scores decreased by ≥6 points, and ≥4 points, respectively, and when well-being scores increased by ≥10 points.

[c]Recovery was indicated when participants met or exceeded the clinical cutoff at baseline, but not at follow-up.

Maintenance of Mental Health Symptoms Among Participants at Lower Risk at Baseline

All maintenance and change in mental health results among participants who had mild or no symptoms on each measure at baseline are presented in Table 4.

Participants at a lower risk for depressive symptoms at baseline reported a small, significant improvement in PHQ-9 scores at follow-up, on average, with 92.2% (297/322) maintaining their low symptom status. Participants at lower risk for anxiety symptoms at baseline reported no significant change (and no escalation) in the GAD-7 scores at follow-up, on average, with 86.7% (255/291) maintaining their low symptom status. Participants at a lower risk for poorer well-being at baseline reported a small, significant improvement in the WHO-5 scores at follow-up, on average, with 90.1% (344/382) maintaining their low symptom status. Post hoc sensitivity analyses revealed that our models among lower-risk participants were sensitive to detect very small effect sizes (Cohen d_z=0.16 for depressive symptoms, 0.16 for anxiety symptoms, and 0.14 for well-being), with 80% power and α=.05.

Table 4. Maintenance and change in mental health from baseline to follow-up among participants at lower risk at baseline (n=509).

Baseline symptoms[a]	Maintenance[b], n (%)	Baseline, mean (SD)	Follow-up, mean (SD)	Cohen *d*	*t* test (*df*)	*P* value
Depressive symptoms (n=322)	297 (92.2)	4.82 (2.88)	4.39 (3.42)	0.13	2.26 (321)	.02
Anxiety symptoms (n=291)	255 (86.7)	3.79 (2.24)	4.00 (3.57)	−0.06	−1.01 (290)	.31
Well-being (n=382)	344 (90.1)	51.84 (14.07)	58.55 (18.83)	−0.40	−7.90 (381)	<.001

[a]Possible ranges for scores include 0-27 for depressive symptoms, 0-21 for anxiety symptoms, and 0-100 for well-being.

[b]Maintenance was indicated when participants did not meet or exceed the clinical cutoff at baseline or follow-up.

Engagement in Recommended Care Plan

Of the 509 participants, 99 (19.4%) were recommended therapy and all lower-level services, 362 (71.1%) were recommended coaching and lower-level services, and 48 (9.4%) were recommended digital content only. Most participants (343/509, 67.4%) engaged with the level of care recommended to them; that is, they engaged at least once with their recommended care modality and did not step up above their recommended care. Specifically, of the 99 participants who were recommended therapy, 84 (85%) met with a therapist at least once; of the 362 participants who were recommended coaching, 224 (61.9%) met at least once with a coach; and of the 48 participants who were recommended digital content, 35 (73%) engaged with at least 1 piece of digital content.

A total of 16.9% (86/509) of the participants overused care beyond their original recommendation; that is, they used a care modality of higher intensity than they were recommended. Of the 362 participants who were recommended coaching, 73 (20.2%) intensified their care to meet with a therapist. Of the 48 participants who were recommended digital content, 13 (27%) intensified their care to access coaching and 2 (4%) participants intensified their care to access a therapist.

In total, 15.7% (80/509) of the participants underused care; that is, they did not engage with the care they were recommended or with a higher-intensity care. Specifically, 15 (15%) of the 99 participants who were recommended therapy declined the invitation to connect with a therapist, and 65 (10.8%) of the 362 participants who were recommended coaching declined the invitation to connect with a coach. A post hoc chi-square analysis comparing the likelihood of clinical improvement or recovery among groups of underusers, overusers, and those who engaged with their recommended level of care found no significant differences in outcomes between groups.

Discussion

Principal Findings

We examined the clinical effectiveness of and engagement in a digital mental health platform that uses a stratified blended care model to deliver mental health services. We found significant improvements in depressive, anxiety, and well-being symptoms among participants with elevated baseline symptoms and high rates of maintaining low symptoms and well-being among participants with lower clinical risk at baseline. Between 60% and 66% of the participants experienced clinically

meaningful improvement or recovery in depressive, anxiety, or well-being symptoms over 3 months. We observed the greatest improvements in well-being (90% score increase), followed by depressive (37% score reduction) and anxiety symptoms (29% score reduction). These results are similar to the published rates of recovery in stepped care systems, which range from 40% to 60% [24].

Among participants with a lower baseline risk (defined as having scores that did not meet the clinical cutoff for that measure), mental health symptoms did not escalate above the clinical threshold or significantly worsen overall; anxiety symptoms remained stable, and depressive symptoms and well-being slightly improved on average. Between 87% and 92% of the participants experienced maintenance of good mental health at follow-up. Given that prevention and well-being promotion are cost-effective for mental health care and provide a positive return on investment for payers [25], our results lend further support to digital mental health services as being helpful for maintaining good mental health. Examining both symptom improvement and prevention of escalating symptoms is crucial to evaluating whether a model of mental health care is improving population health.

Most participants engaged with their personalized care recommendation, with only around one-third overusing or underusing services. Two-thirds (343/509, 67.4%) of the participants used the recommended level of services; that is, they engaged with their recommended care services but did not step up to use higher-intensity care. Rates of meeting the care recommendation were the highest for therapy, with 85% (84/99) of people who were recommended therapy having at least 1 visit with a therapist.

We also found that less than one-fifth (86/509, 16.9%) of the participants overused care; that is, they engaged with a care modality of a higher intensity than the one they were initially recommended. Although we did not have data available on referral pathways, participants could use a higher intensity of care through 2 channels: self-referral by contacting member services through the platform with their request and provider referral where a coach could refer a participant to a therapist if they had a demonstrated clinical need. Interestingly, patterns of overuse matched the intensity hierarchy of services: 20% (73/362) of the people who were recommended coaching moved up 1 level of intensity to therapy, 27% (13/48) of the people who were recommended digital content moved up 1 level to coaching, and only 4.2% (2/48) of the people who were recommended digital content moved up 2 levels to therapy. This further supports the accuracy of the initial care recommendations of this model. In addition, the fairly low rates of overuse are encouraging, considering long-held beliefs about psychotherapy as the gold standard for mental health treatment at all levels of care [26].

Finally, less than one-fifth (80/509, 15.7%) of the participants underused care; that is, they only engaged with lower-intensity care modalities than their recommendation. The rate was slightly higher for those who were recommended therapy (of the 99 participants, 15 (15%) did not have a visit with a therapist), while among those who were recommended coaching, 10.8% (65/362) did not have a visit with a coach or therapist. These rates of treatment nonadherence are consistent with those observed in psychotherapy research [27] and are better than those observed in previous research on digital interventions [28]. A post hoc analysis did not find statistically significant differences in improvement or recovery between people who followed their recommended care and those who overused or underused care. However, the lack of significance could be an artifact of low statistical power, given the skewed proportion of participants in the underuse and overuse categories. Future research that examines reasons for underusing or overusing care qualitatively would provide further insight into why this occurs for some people and may allow programs to stratify individuals into levels of care more appropriately.

It is difficult to compare the effectiveness of stratified blended care approaches in real-world contexts with that of other studies in the literature because most research has been conducted within highly controlled clinical trials, which lack ecological validity (eg, see Andrews et al [13] and Ho et al [29]). However, in one trial comparing stratified and progressive stepped care models, 76% of eligible screened patients engaged in treatment (high intensity or low intensity depending on personalized treatment recommendations) [16]. Thus, the engagement rate observed here is similar to that in research settings with high internal validity, suggesting that this platform is relatively accurate in recommending effective care. Our observational results of stratified blended care engagement patterns indicate that this model was efficient in matching individuals with the most effective and least costly care while also allowing them to self-determine their care and use combinations of services that best fit their needs.

Limitations

Although this investigation demonstrated improvements in depression, anxiety, and well-being over time in a sample of digital mental health platform users, the observational nature of this study presents a limitation. The inclusion of a comparison or control group in an experimental design would confirm the causality of the observed changes. The 3-month time frame of this investigation also limits our conclusions to short-term gains; a longer-term follow-up period is needed to determine the persistence of improvements.

Only 9.25% (950/10,270) of the individuals who were eligible and invited to participate enrolled in this study. The reasons for this are not known but may be due to a combination of factors, such as the time commitment needed, as each survey took between 30 and 45 minutes to complete. In addition, although study materials highlighted the separation between the study and their employer, most invitations to participate were distributed to workplace email addresses, which may have given the impression to participants that their employers would be involved or aware of their participation in this investigation. Although the retention rate of this investigation was 73.26% (696/950) overall, the sample was limited to those who used at least 1 piece of content or sought a provider. More individuals were registered for the platform but did not use any care. This is common in real-world settings, as individuals may register for services without the intention of using them immediately.

There were no significant differences on key demographic characteristics and baseline symptoms between the sample who met eligibility criteria and those who did not, which helps to bolster the generalizability of our findings.

In addition, we were unable to discern between self-referrals and provider referrals for the 16.9% (86/509) of the members who sought more intensive services beyond their initial care recommendation. It is possible that the 17% include some individuals who were appropriately identified for intensification of services (ie, a coach detected additional symptoms, or a member disclosed additional pertinent information that was not detected in the assessment algorithm that determines care recommendations).

Finally, most of the sample (312/509, 61.3%) identified as women, and only 34.4% (175/509) identified as men. Furthermore, the percentage of nonbinary people in this sample was higher than the overall US population [30] at 4.3%. There are several possibilities for the disproportionate numbers of women and nonbinary people enrolling in this study. First, the disproportionate number of nonbinary people in the sample may reflect a younger population, who are more likely to identify as transgender or nonbinary than older populations (5% vs 1.6% overall) [30]. In addition, we provided multiple nonbinary options for participants as opposed to a single all-encompassing option (eg, “nonbinary”), which may have encouraged identification. Finally, women and people who are lesbian, gay, bisexual, transgender, or questioning, including people who identify as nonbinary, are more likely to experience mental illness [31,32]. Women are more likely to seek help than men (due in part to societal expectations of stoicism and self-reliance for men, as well as mental health stigma), which might have affected their likelihood of signing up for the Modern Health app and participating in research [33]. Regardless of the reason for the lack of men in this sample, it is possible that the results here may be different among other populations, and we encourage future research to examine these potential differences.

Conclusions

Overall, the results of this study lend support to the utility and effectiveness of the stratified blended care model used in this study to improve access to and use of mental health services. In a resource-constrained ecosystem, using a stratified blended model, such as the one evaluated, can make efficient use of limited and costly services while centering the individual’s needs, preferences, and receptivity to mental health care. The deliberate allocation of resources preserved the more intensive and costly resources for those who are most likely to benefit from them while providing beneficial care at all levels. As our results indicate, meaningful clinical improvements can be gained through stratified blended care while honoring the individual’s personal preference for how they want to engage in care.

Research continues throughout the field of mental health to determine the therapeutic approaches, techniques, and tools that can be adapted and disseminated for digital delivery while preserving safety, quality, validity, and efficacy [34]. As progress continues, the constraints of the traditional tertiary care model of mental health will eventually give way to a more comprehensive approach that can serve the full spectrum of mental health from primary prevention to treatment.

Acknowledgments

The authors would like to thank Dr Myra Altman, the Modern Health Clinical Strategy and Research team, the larger Modern Health organization, as well as the members and providers for their contributions to this work. The authors declare that no funds, grants, or other support was received to support the preparation of this manuscript.

Data Availability

The data sets generated during and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

This study was conceived and designed by LGR, SJS-H, CCS, and CEWC. Data were acquired by SJS-H, CEWC, and BJS. LGR contributed to data analysis, while LGR, SJS-H, CCS, and BJS interpreted the data. The manuscript was drafted by LGR, SJS-H, CCS, and CEWC and critically revised by LGR, SJS-H, CCS, CEWC, and BJS. All authors provided critical feedback and edits to each version of the manuscript. All authors read and approved the final manuscript and agreed to be accountable for all aspects of the work.

Conflicts of Interest

SJS-H, CCS, CEWC, and BJS are employed by Modern Health and receive salary and stock options. LGR declares no conflict of interest.

References

1. Key substance use and mental health indicators in the United States: results from the 2020 national survey on drug use and health. Substance Abuse and Mental Health Services Administration. 2021. URL: https://tinyurl.com/2vsjeaxc [accessed 2024-05-01]
2. Satiani A, Niedermier J, Satiani B, Svendsen DP. Projected workforce of psychiatrists in the United States: a population analysis. Psychiatr Serv 2018 Jun 01;69(6):710-713. [doi: 10.1176/appi.ps.201700344] [Medline: 29540118]

3. APA fact sheet series on psychologist supply and demand projections 2015-2030: geographic patterns in supply and demand. American Psychological Association. 2018. URL: https://www.apa.org/workforce/publications/supply-demand/geographic-patterns.pdf [accessed 2024-05-01]
4. State-level projections of supply and demand for behavioral health occupations: 2016-2030. U.S. Department of Health and Human Services. 2018 Sep. URL: https://bhw.hrsa.gov/sites/default/files/bureau-health-workforce/data-research/state-level-estimates-report-2018.pdf [accessed 2024-05-01]
5. Torous J, Jän Myrick K, Rauseo-Ricupero N, Firth J. Digital mental health and COVID-19: using technology today to accelerate the curve on access and quality tomorrow. JMIR Ment Health 2020 Mar 26;7(3):e18848 [FREE Full text] [doi: 10.2196/18848] [Medline: 32213476]
6. Bower P, Gilbody S. Stepped care in psychological therapies: access, effectiveness and efficiency. Narrative literature review. Br J Psychiatry 2005 Jan;186:11-17. [doi: 10.1192/bjp.186.1.11] [Medline: 15630118]
7. Clark DM. Realizing the mass public benefit of evidence-based psychological therapies: the IAPT program. Annu Rev Clin Psychol 2018 May 07;14(1):159-183 [FREE Full text] [doi: 10.1146/annurev-clinpsy-050817-084833] [Medline: 29350997]
8. Bower P, Kontopantelis E, Sutton A, Kendrick T, Richards DA, Gilbody S, et al. Influence of initial severity of depression on effectiveness of low intensity interventions: meta-analysis of individual patient data. BMJ 2013 Feb 26;346:f540 [FREE Full text] [doi: 10.1136/bmj.f540] [Medline: 23444423]
9. Chan SW, Adams M. Service use, drop-out rate and clinical outcomes: a comparison between high and low intensity treatments in an IAPT service. Behav Cogn Psychother 2014 Jan 02;42(6):747-759. [doi: 10.1017/s1352465813000544]
10. Martin C, Iqbal Z, Airey ND, Marks L. Improving Access to Psychological Therapies (IAPT) has potential but is not sufficient: how can it better meet the range of primary care mental health needs? Br J Clin Psychol 2022 Mar;61(1):157-174. [doi: 10.1111/bjc.12314] [Medline: 34124792]
11. Richards DA, Hill JJ, Gilbody S, Coventry PA, Gask L. Chapter 28 - Collaborative care and stepped care. In: Gask L, Kendrick T, Peveler R, Chew-Graham CA, editors. Primary Care Mental Health. Cambridge, MA: Cambridge University Press; 2018.
12. Haaga DA. Introduction to the special section on stepped care models in psychotherapy. J Consult Clin Psychol 2000 Aug;68(4):547-548. [Medline: 10965628]
13. Andrews B, Klein B, Van Nguyen H, Corboy D, McLaren S, Watson S. Efficacy of a digital mental health biopsychosocial transdiagnostic intervention with or without therapist assistance for adults with anxiety and depression: adaptive randomized controlled trial. J Med Internet Res 2023 Jun 12;25:e45135 [FREE Full text] [doi: 10.2196/45135] [Medline: 37307046]
14. Serio C, Gabarda A, Uyar-Morency F, Silfee V, Ludwig J, Szigethy E, et al. Strengthening the impact of digital cognitive behavioral interventions through a dual intervention: proficient motivational interviewing-based health coaching plus in-application techniques. JMIR Form Res 2022 May 11;6(5):e34552 [FREE Full text] [doi: 10.2196/34552] [Medline: 35544323]
15. Boyd L, Baker E, Reilly J. Impact of a progressive stepped care approach in an improving access to psychological therapies service: an observational study. PLoS One 2019;14(4):e0214715 [FREE Full text] [doi: 10.1371/journal.pone.0214715] [Medline: 30964883]
16. Delgadillo J, Ali S, Fleck K, Agnew C, Southgate A, Parkhouse L, et al. Stratified care vs stepped care for depression: a cluster randomized clinical trial. JAMA Psychiatry 2022 Feb 01;79(2):101-108 [FREE Full text] [doi: 10.1001/jamapsychiatry.2021.3539] [Medline: 34878526]
17. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
18. Liness S, Beale S, Lea S, Byrne S, Hirsch CR, Clark DM. Multi-professional IAPT CBT training: clinical competence and patient outcomes. Behav Cogn Psychother 2019 Mar 28;47(6):672-685. [doi: 10.1017/s1352465819000201]
19. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
20. Plummer F, Manea L, Trepel D, McMillan D. Screening for anxiety disorders with the GAD-7 and GAD-2: a systematic review and diagnostic metaanalysis. Gen Hosp Psychiatry 2016 Mar;39:24-31. [doi: 10.1016/j.genhosppsych.2015.11.005] [Medline: 26719105]
21. Toussaint A, Hüsing P, Gumz A, Wingenfeld K, Härter M, Schramm E, et al. Sensitivity to change and minimal clinically important difference of the 7-item Generalized Anxiety Disorder Questionnaire (GAD-7). J Affect Disord 2020 Mar 15;265:395-401. [doi: 10.1016/j.jad.2020.01.032] [Medline: 32090765]
22. Topp CW, Østergaard SD, Søndergaard S, Bech P. The WHO-5 Well-Being Index: a systematic review of the literature. Psychother Psychosom 2015;84(3):167-176 [FREE Full text] [doi: 10.1159/000376585] [Medline: 25831962]
23. Bech P, Lunde M, Bech-Andersen G, Lindberg L, Martiny K. Psychiatric outcome studies (POS): does treatment help the patients? A Popperian approach to research in clinical psychiatry. Nord J Psychiatry 2007 Jul 12;61 Suppl 46(sup46):4-34. [doi: 10.1080/08039480601151238] [Medline: 17365777]

24. Firth N, Barkham M, Kellett S. The clinical effectiveness of stepped care systems for depression in working age adults: a systematic review. J Affect Disord 2015 Jan 01;170:119-130 [FREE Full text] [doi: 10.1016/j.jad.2014.08.030] [Medline: 25240141]
25. Le LK, Esturas AC, Mihalopoulos C, Chiotelis O, Bucholc J, Chatterton ML, et al. Cost-effectiveness evidence of mental health prevention and promotion interventions: a systematic review of economic evaluations. PLoS Med 2021 May 11;18(5):e1003606 [FREE Full text] [doi: 10.1371/journal.pmed.1003606] [Medline: 33974641]
26. Schleider JL, Dobias ML, Mullarkey MC, Ollendick T. Retiring, rethinking, and reconstructing the norm of once-weekly psychotherapy. Adm Policy Ment Health 2021 Jan 28;48(1):4-8 [FREE Full text] [doi: 10.1007/s10488-020-01090-7] [Medline: 32989621]
27. Leeuwerik T, Cavanagh K, Strauss C. Patient adherence to cognitive behavioural therapy for obsessive-compulsive disorder: a systematic review and meta-analysis. J Anxiety Disord 2019 Dec;68:102135. [doi: 10.1016/j.janxdis.2019.102135] [Medline: 31704633]
28. Fleming T, Bavin L, Lucassen M, Stasiak K, Hopkins S, Merry S. Beyond the trial: systematic review of real-world uptake and engagement with digital self-help interventions for depression, low mood, or anxiety. J Med Internet Res 2018 Jun 06;20(6):e199 [FREE Full text] [doi: 10.2196/jmir.9275] [Medline: 29875089]
29. Ho FY, Yeung WF, Ng TH, Chan CS. The efficacy and cost-effectiveness of stepped care prevention and treatment for depressive and/or anxiety disorders: a systematic review and meta-analysis. Sci Rep 2016 Jul 05;6:29281 [FREE Full text] [doi: 10.1038/srep29281] [Medline: 27377429]
30. Brown A. About 5% of young adults in the U.S. say their gender is different from their sex assigned at birth. Pew Research Center. 2022 Jun 7. URL: https://www.pewresearch.org/short-reads/2022/06/07/about-5-of-young-adults-in-the-u-s-say-their-gender-is-different-from-their-sex-assigned-at-birth/ [accessed 2023-09-19]
31. Rodgers J, Cuevas AG, Williams DR, Kawachi I, Subramanian SV. The relative contributions of behavioral, biological, and psychological risk factors in the association between psychosocial stress and all-cause mortality among middle- and older-aged adults in the USA. Geroscience 2021 Apr;43(2):655-672 [FREE Full text] [doi: 10.1007/s11357-020-00319-5] [Medline: 33511488]
32. Newcomb ME, Hill R, Buehler K, Ryan DT, Whitton SW, Mustanski B. High burden of mental health problems, substance use, violence, and related psychosocial factors in transgender, non-binary, and gender diverse youth and young adults. Arch Sex Behav 2020 Feb;49(2):645-659 [FREE Full text] [doi: 10.1007/s10508-019-01533-9] [Medline: 31485801]
33. Mental health treatment among adults: United States, 2020. Centers for Disease Control and Prevention. 2021. URL: https://tinyurl.com/4zpjc3z3 [accessed 2024-05-02]
34. Schueller SM, Torous J. Scaling evidence-based treatments through digital mental health. Am Psychol 2020 Nov;75(8):1093-1104 [FREE Full text] [doi: 10.1037/amp0000654] [Medline: 33252947]

Abbreviations

GAD-7: 7-item Generalized Anxiety Disorder scale
PHQ-9: 9-item Patient Health Questionnaire
WHO-5: 5-item World Health Organization Well-Being Index

Edited by L Buis; submitted 18.04.23; peer-reviewed by L Balcombe, S Rennick-Egglestone, A Bucher; comments to author 14.06.23; revised version received 20.09.23; accepted 26.03.24; published 24.06.24.

Please cite as:
Roos LG, Sagui-Henson SJ, Castro Sweet C, Welcome Chamberlain CE, Smith BJ
Improvement and Maintenance of Clinical Outcomes in a Digital Mental Health Platform: Findings From a Longitudinal Observational Real-World Study
JMIR Mhealth Uhealth 2024;12:e48298
URL: https://mhealth.jmir.org/2024/1/e48298
doi:10.2196/48298
PMID:38913405

Original Paper

Long-Term Efficacy of a Mobile Mental Wellness Program: Prospective Single-Arm Study

Meaghan McCallum[1], PhD; Matthew Baldwin[1], PhD; Paige Thompson[1], MPH; Kelly Blessing[1], BS; Maria Frisch[1], PhD; Annabell Ho[1], PhD; Matthew Cole Ainsworth[1], PhD; Ellen Siobhan Mitchell[1], PhD; Andreas Michaelides[1], PhD; Christine N May[1], PhD

Academic Research, Noom, Inc, New York City, NY, United States

Corresponding Author:
Meaghan McCallum, PhD
Academic Research
Noom, Inc
450 West 33rd Street
New York City, NY, 10001
United States
Phone: 1 8882665071
Email: meaghanm@noom.com

Abstract

Background: Rising rates of psychological distress (symptoms of depression, anxiety, and stress) among adults in the United States necessitate effective mental wellness interventions. Despite the prevalence of smartphone app–based programs, research on their efficacy is limited, with only 14% showing clinically validated evidence. Our study evaluates Noom Mood, a commercially available smartphone-based app that uses cognitive behavioral therapy and mindfulness-based programming. In this study, we address gaps in the existing literature by examining postintervention outcomes and the broader impact on mental wellness.

Objective: Noom Mood is a smartphone-based mental wellness program designed to be used by the general population. This prospective study evaluates the efficacy and postintervention outcomes of Noom Mood. We aim to address the rising psychological distress among adults in the United States.

Methods: A 1-arm study design was used, with participants having access to the Noom Mood program for 16 weeks (N=273). Surveys were conducted at baseline, week 4, week 8, week 12, week 16, and week 32 (16 weeks' postprogram follow-up). This study assessed a range of mental health outcomes, including anxiety symptoms, depressive symptoms, perceived stress, well-being, quality of life, coping, emotion regulation, sleep, and workplace productivity (absenteeism or presenteeism).

Results: The mean age of participants was 40.5 (SD 11.7) years. Statistically significant improvements in anxiety symptoms, depressive symptoms, and perceived stress were observed by week 4 and maintained through the 16-week intervention and the 32-week follow-up. The largest changes were observed in the first 4 weeks (29% lower, 25% lower, and 15% lower for anxiety symptoms, depressive symptoms, and perceived stress, respectively), and only small improvements were observed afterward. Reductions in clinically relevant anxiety (7-item generalized anxiety disorder scale) and depression (8-item Patient Health Questionnaire depression scale) criteria were also maintained from program initiation through the 16-week intervention and the 32-week follow-up. Work productivity also showed statistically significant results, with participants gaining 2.57 productive work days from baseline at 16 weeks, and remaining relatively stable (2.23 productive work days gained) at follow-up (32 weeks). Additionally, effects across all coping, sleep disturbance (23% lower at 32 weeks), and emotion dysregulation variables exhibited positive and significant trends at all time points (15% higher, 23% lower, and 25% higher respectively at 32 weeks).

Conclusions: This study contributes insights into the promising positive impact of Noom Mood on mental health and well-being outcomes, extending beyond the intervention phase. Though more rigorous studies are necessary to understand the mechanism of action at play, this exploratory study addresses critical gaps in the literature, highlighting the potential of smartphone-based mental wellness programs to lessen barriers to mental health support and improve diverse dimensions of well-being. Future research should explore the scalability, feasibility, and long-term adherence of such interventions across diverse populations.

(JMIR Mhealth Uhealth 2024;12:e54634) doi:10.2196/54634

KEYWORDS

mHealth; psychological distress; Noom Mood; digital mental wellness programs; mobile phone

Introduction

Psychological distress, encompassing nonspecific symptoms of stress, anxiety, and depression, has been on the rise in the United States since 1999. More specifically, a 40% increase in prevalence was found from 2000 to 2018 [1-4] and was exacerbated by the COVID-19 pandemic [5-9]. One 2020 meta-analysis found the prevalence of stress in the general population to be between 29.6% and 33.7% [5]. This distress correlates with personal and occupational impairment, chronic medical disorders, adverse health behaviors, and increased mortality risk from diseases such as cardiovascular disease, cancer, and liver disease [10-12]. Depression and anxiety symptoms also predict absenteeism (missed work days) and presenteeism (low productivity days), costing an estimated 3% to 4% of the gross national product [13-16].

Despite the high prevalence, a large portion of affected adults do not receive necessary mental health support. Common barriers to mental health support in the United States include a desire for autonomy in the process, fear of social stigma, and lack of access (financially and structurally) [17-19]. Smartphone app–based mental wellness programs offer a promising solution, and the benefits of mobile mental wellness programs include increased convenience, adherence, personalization, social support, and cost-effectiveness [20-22]. However, research on their effectiveness is lacking [23], with only 14% showing clinically validated evidence of effectiveness [18], particularly among commercially available programs.

In response to the scarcity of clinical validation among these programs, we previously evaluated the feasibility and short-term efficacy of Noom Mood [24], a commercially available, smartphone-based mental wellness program that provides educational content, skills-based activities, and nonclinical coaching support to help users manage stress and symptoms of anxiety. In this study, we extend our initial findings (at 4 weeks) to the full length of the intervention (at 16 weeks) and postintervention (at 32 weeks) in a larger sample. We also aim to contribute to important gaps within the literature by increasing our understanding of the efficacy and outcome maintenance associated with mobile mental wellness programs.

Research on the maintenance of outcomes after completing mobile mental wellness programs is currently scarce and is essential to inform postintervention support and establish long-term efficacy. Economides and colleagues [25] provide some evidence that a reduction in anxiety and depressive symptoms can be maintained at 6 months post intervention. However, their analysis was retrospective, and follow-up assessments were limited. High-quality, long-term postintervention data are needed to better understand the full clinical utility of mobile mental wellness programming for the general population [26]. In addition, there are gaps in the existing literature looking at overall mental wellness, beyond mental health-related symptoms (eg, emotion regulation, coping, or quality of life). Despite theoretical and empirical assertions that smartphone-based programs can indirectly help improve mental wellness, such outcomes have not been widely studied [27,28].

This 1-arm prospective study examines the efficacy and postintervention outcomes of a commercially available mental wellness program on a broad range of mental health outcomes: symptoms of anxiety, stress, and depression, as well as well-being, quality of life, coping skills, emotion regulation, sleep, and workplace presenteeism and absenteeism. In doing so, this study contributes to the empirical foundation and clinical utility of such programming for the general population.

Methods

Participants

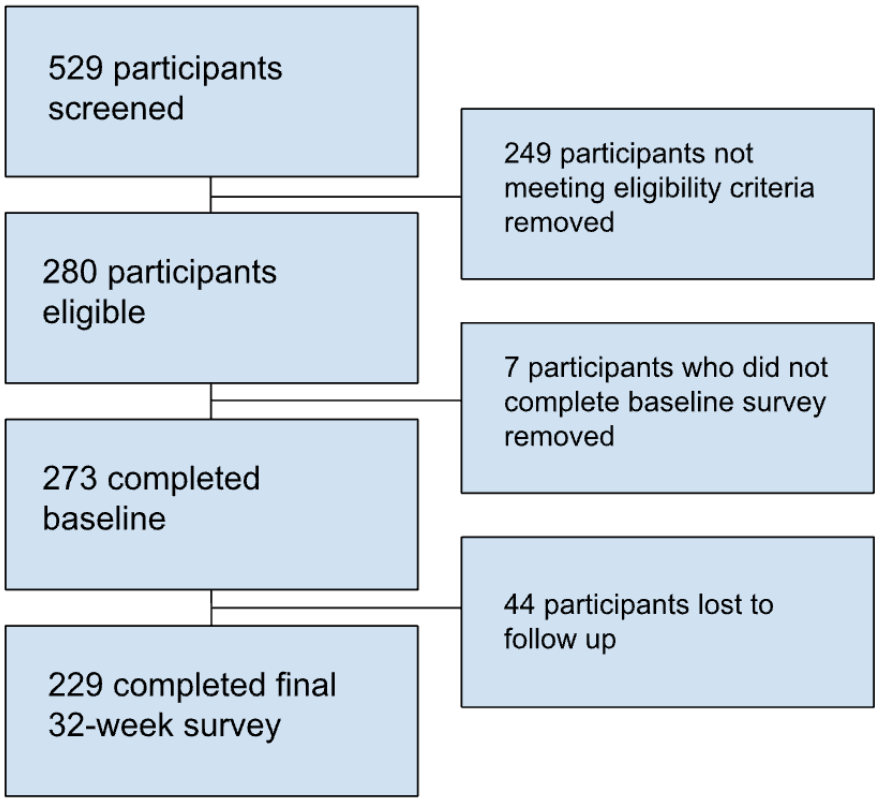
A total of 273 adults participated in this study. All participants were recruited from social media (ie, Facebook and Instagram) between May and September 2022, with advertisements targeting a 50% (50/100) female audience. Before enrollment, an online screening questionnaire confirmed eligibility. A total of 529 participants were recruited and screened for this study.

Participants were eligible for participation in this study if they were located in the United States, English-speaking, aged older than 18 years, used an Apple operating system mobile device, and agreed to not alter current mental health treatment, if any. Given the online nature of this study, they were required to verify their identity by uploading a photo of their state-issued identification card. They were also required to complete a decisional balance exercise to ensure a thorough understanding of and willingness to participate in this study. This exercise, based on the work of Goldberg and Kiernan [29], aimed to increase study retention rates by discussing the nature, design, and importance of this study and required potential participants to consider all the pros and cons of participation before deciding whether or not to participate.

Participants were excluded if they had a current or past diagnosis of psychotic disorder or bipolar disorder; if they had started, stopped, or changed doses of psychotropic medication in the past 2 months; or had started or stopped psychotherapy within the past 2 months. They were excluded if they had used the Noom Mood program previously, or were currently using the Noom Weight program.

In total, 351 participants were eligible for participation. Eligible participants were provided informed consent forms and were required to download the app within 14 days to be considered fully enrolled. The final sample included 273 participants who downloaded the program and provided consent (see Figure 1).

Figure 1. Study flow CONSORT diagram. CONSORT: Consolidated Standards of Reporting Trials.



Ethical Considerations

This study was approved by the Advarra Institutional Review Board (Pro00062644).

Power Analysis and Sample Size Calculation

The sample size was determined based on an estimated medium effect on postintervention scores on our primary outcome, the 7-item generalized anxiety disorder scale (GAD-7), from baseline to end of the program, 95% power, and α set at .05. The effect size is based on an ITT analysis of the GAD-7 (d=0.70) from a previous Noom Mood feasibility study [24]. The sample size calculation was conducted using G*Power [30] matched pairs t tests (2-tailed). Our recruitment target also accounted for 48% attrition, based on a recent meta-analysis of dropout rates in clinical trials with smartphone apps [31]. Therefore, this study aimed to recruit at least 247 participants in total.

Recruitment and Eligibility

The eligibility criteria of this study included the ability to understand and provide informed consent, willingness to not alter current mental health treatment (ie, psychotropic medications or psychotherapy), aged at least 18 years, English speaking, located in the United States, iOS user (because certain program features were only available on iPhones at the time of study launch), and successful identity verification. The ineligibility criteria included current or past diagnosis of psychotic disorder or bipolar disorder, past use of the Noom Mood program, starting, changing dose, or stopping psychotropic medication within the past 2 months, starting or stopping psychotherapy within the last 2 months, or current use of the Noom Weight program.

Procedure

Participants were invited to complete the baseline questionnaire after consenting to participate in this study. As part of this study, participants were required to redeem a unique code and download the Noom Mood app within 14 days of consent, however, there were no further stipulations on participants' level of engagement with the app. As part of this study, all participants had access to the full Noom Mood program, which consists of 16 weeks of content, at no cost. Participants were surveyed 4, 8, 12, 16, and 32 weeks after baseline. All surveys were administered online via OpenClinica, a HIPAA (Health Insurance Portability and Accountability Act)-compliant electronic data capture platform [32].

Program

Noom Mood is a commercially available, mobile app–based, mobile mental wellness program that focuses on stress and anxiety management as well as improving overall mental wellness. The program is based on cognitive behavioral therapy, dialectical behavior therapy, acceptance and commitment therapy, and mindfulness-based stress reduction with an emphasis on building a toolkit of coping skills that can be applied to daily life. Participants have access to several program features, including a daily curriculum consisting of psychoeducational articles for users to read, personalized coaching offered through in-app messaging, weekly skills-based activities, and a mood-logging feature. In a previous study of feasibility, acceptability, and preliminary outcomes, significant improvements in anxiety symptoms, perceived stress, depressive

feelings, emotion regulation, and optimism were found from baseline to 4-week follow-up. Additionally, learning coping skills (eg, breathing and cognitive reframing techniques), interacting with the program features, and gaining awareness of personal emotions and thought patterns were the features of the program that participants reported benefiting most from [24].

Measures

Overview

The primary outcomes of this study were self-reported measures of symptoms of anxiety, stress, depression, perceived health and quality of life, and mental wellness, detailed below.

Anxiety Symptoms (GAD 7)

The GAD-7 [33] assesses the extent to which individuals experience symptoms of anxiety (eg, "feeling nervous, anxious, or on edge") on a scale of 0 ("not at all") to 3 ("nearly every day"). Clinical severity was determined using a normed cutoff of 10 for this measure.

Depressive Symptoms (8-Item Patient Health Questionnaire Depression Scale)

The 8-item Patient Health Questionnaire depression scale [34] assesses the extent to which participants experience feelings of depression (eg, "feeling down, depressed, or hopeless" or "little interest or pleasure in doing things") on a scale of 0 ("not at all") to 3 ("nearly every day"). Clinical severity was determined using a normed cutoff of 10 for this measure.

Perceived Stress (10-Item Perceived Stress Scale)

The 10-item Perceived Stress Scale [35] assesses the frequency with which individuals experience various symptoms of stress (eg, "how often have you felt that you were unable to control the important things in your life?") on a scale of 0 ("never") to 4 ("very often").

Mental Wellness (14-Item Warwick Edinburgh Mental Well-Being Scale)

The 14-Item Warwick Edinburgh Mental Well-Being scale [36] measures mental wellness. The questions (eg, I have been feeling "optimistic, useful, relaxed") are answered using a scale of 0 ("none of the time") to 5 ("all of the time").

Absenteeism, Presenteeism, and Productivity Loss (Sheehan Disability Scale, Modified)

The Sheehan Disability Scale [37] is a scale that measures impairment in functioning ("how many days in the last week did your symptoms cause you to miss school or work or leave you unable to carry out normal daily responsibilities?"). Further, 2 items in particular were used for this study. Absenteeism was measured with the days lost item ("on how many days in the last month did your symptoms cause you to miss school or work or leave you unable to carry out your normal daily responsibilities?"); participants provided a number from 0 to 30. Presenteeism was measured with the days unproductive item ("on how many days in the last month did you feel so impaired by your symptoms, that even though you went to school or work, your productivity was reduced?") Participants provided a number from 0 to 30. Productivity loss was calculated as the sum of absenteeism and presenteeism, similar to previous work [38,39].

Quality of Life (10-Item Patient-Reported Outcomes Measurement Information System Global)

The 10-item Patient-Reported Outcomes Measurement Information System Global scale [40] measures general health care–related quality of life (eg, "how would you rate your physical health?") on a scale of 1 ("poor") to 5 ("excellent").

Sleep (Pittsburgh Sleep Quality Index)

The Pittsburgh Sleep Quality Index [41] is a 9-item scale used to measure sleep quality and patterns in adults. The first 4 questions (for instance, "when have you usually gone to bed?") are free-response questions. The following 4 questions (eg, "during the past month, how often have you taken medicine [prescribed or 'over the counter'] to help you sleep?") use a scale of 0 ("not during the past month") to 3 ("three or more times per week"). The final question ("during the past month, how would you rate your sleep quality overall?") uses a scale of 0 "very good" to 3 "very bad."

Emotion Regulation (Difficulties in Emotion Regulation Scale)

The Difficulties in Emotion Regulation Scale [42] is an 18-item scale that measures emotional dysregulation (eg, "I pay attention to how I feel") and uses a scale of 1 "almost never" to 5 "almost always."

Statistical Analysis

All analyses were linear mixed-effects models. Dummy codes reflecting comparisons between baseline and all other time points were included as predictors of each primary outcome. Random intercepts were specified for all models. For count data (eg, absenteeism or presenteeism), a Poisson mixed-effects model was used. For categorical outcomes (eg, clinical cutoffs), logistic mixed-effects models were used. Maximum-likelihood estimation was used to account for missing data, which was relatively low (see below).

Results

Demographics and Participation

Participant demographic characteristics are presented in Table 1. Of the 273 participants, the mean age was 40.5 (SD 11.7) years. Nearly half (124/273, 45.4%) identified as female, 53.8% (147/273) as male, and 0.7% (2/273) as other, within a predominantly White, non-Hispanic sample (85%). A substantial portion of the sample possessed a college degree (103/273, 37.7%) or graduate degree (81/273, 29.7%). Income distribution spanned a wide range, with the majority reporting annual incomes of US $25,000-$50,000 (44/273, 16.1%), US $50,000-$100,000 (98/273, 35.9%), or US $100,000-$200,000 (70/273, 25.6%). Attrition was minimal and unlikely to introduce bias, with a maximum of 7% missing data reported at all time points.

Table 1. Participant demographics (N=273).

Demographics	Value
Age (years), mean (SD)	40.5 (11.8)
Gender, n (%)	
Male	147 (53.8)
Female	124 (45.4)
Other	2 (0.7)
Ethnicity, n (%)	
Hispanic	23 (8.4)
Not Hispanic	245 (90.1)
Prefer not to say or N/A[a]	5 (1.8)
Race, n (%)	
Asian or Pacific Islander	38 (13.9)
Black or African American	26 (13.1)
White	215 (85.3)
Other	8 (2.9)
Prefer not to say or N/A	4 (1.5)
Income, n (%)	
Less than US $25,000	23 (8.4)
US $25,000-$50,000	44 (16.2)
US $50,000-$100,000	98 (36)
US $100,000-$200,000	70 (25.7)
More than US $200,000	22 (8.1)
Prefer not to say or N/A	16 (5.5)
Education, n (%)	
High school, GED[b], or less education	15 (5.5)
Some college or Associate degree	73 (26.7)
College graduate	103 (37.7)
Graduate degree	81 (29.7)
Prefer not to say	1 (0.4)

[a]N/A: not applicable.

[b]GED: General Educational Development.

Retention

Of the 273 participants who on a baseline survey with retention rates at the 4-week, 8-week, 12-week, 16-week, and 32-week surveys, respectively, were 266/273 (97.4% retention rate from baseline), 255/273 (93.4% retention rate from baseline), 245/273 (89.7% retention rate from baseline), 250/273 (91.6% retention rate from baseline), and 229/273 (83.9% retention rate from baseline).

Psychological Distress

Statistically significant changes in anxiety symptoms, depression symptoms, perceived stress, well-being, and quality of life were observed by week 4 and maintained through program end (16 weeks) and follow-up (32 weeks, $P<.001$). The largest changes occurred between baseline and 4 weeks, with only small incremental improvements afterward, as shown in Figure 2. Mixed-effects regression models, using random intercepts, were used to assess the impact of time on the aforementioned variables, as depicted in Figure 2 and Tables 2 and 3. All comparisons between program end (16-weeks) and follow-up (32-weeks) were not statistically significant ($.29\leq P$ values $\leq.99$).

Figure 2. The effects of time on anxiety symptoms, depression symptoms, perceived stress, well-being, and quality of life. Points are jittered to avoid overplotting. bl: baseline; GAD-7: 7-item Generalized Anxiety Disorder Scale; PHQ-8: 8-item Patient Health Questionnaire depression scale; PROMIS: Patient-Reported Outcomes Measurement Information System; PSS: Perceived Stress Scale; WEMWBS: Warwick Edinburgh Mental Well-Being scale.

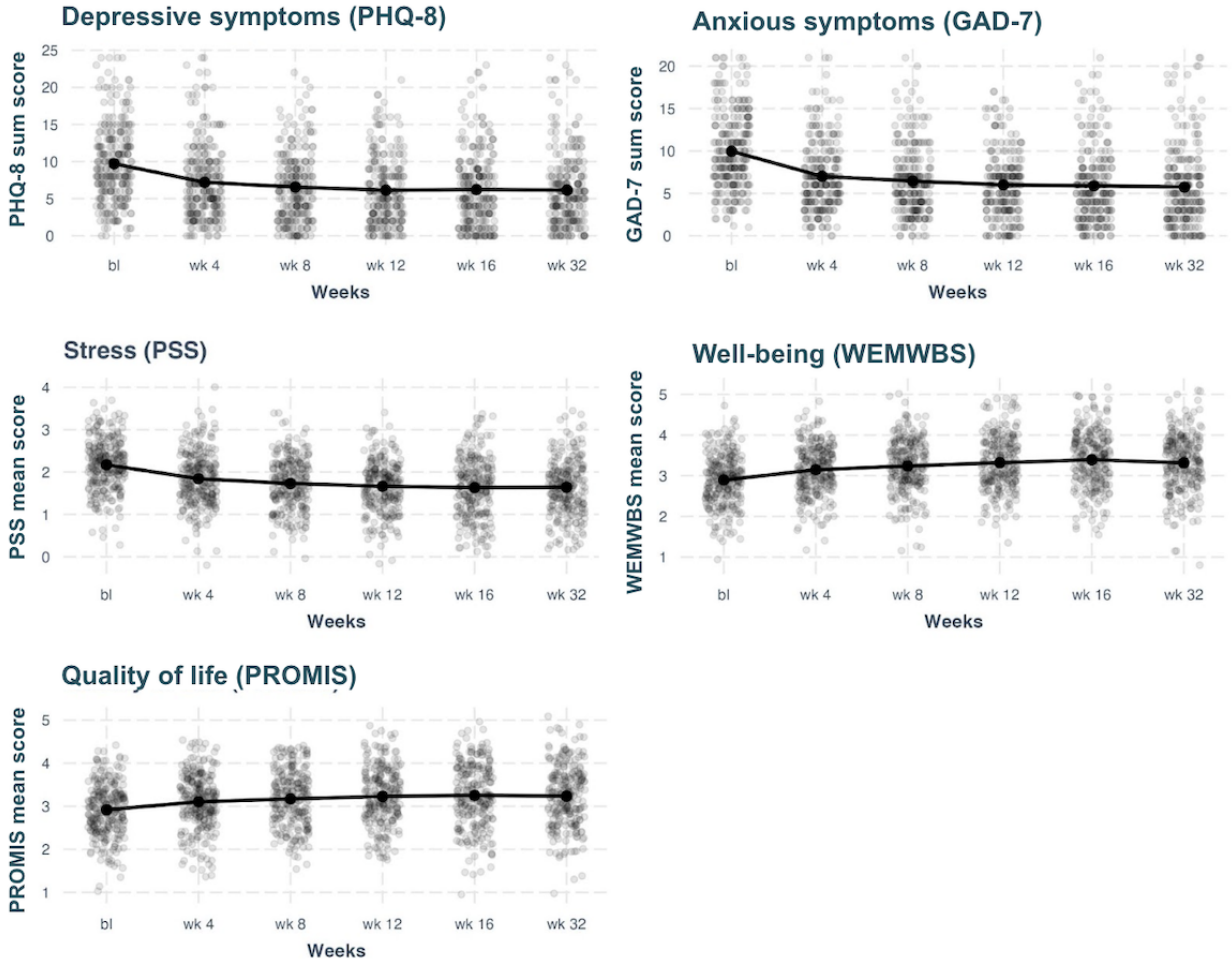


Table 2. Changes in anxiety symptoms, depression symptoms, perceived stress, well-being, and quality of life at follow-up time points versus baseline (predictors).

	Depressive symptoms			Anxious symptoms			Stress			Well-being			Quality of life		
	Estimates	Standard β	*P* value	Estimates	Standard β	*P* value	Estimates	Standard β	*P* value	Estimates	Standard β	*P* value	Estimates	Standard β	*P* value
Predictors															
Intercept	9.72	.55	<.001	9.96	.70	<.001	2.17	.60	<.001	2.89	–.49	<.001	2.92	–.35	<.001
BL[a] vs wk 4	–2.51	–.50	<.001	–2.95	–.66	<.001	–0.33	–.51	<.001	0.25	.38	<.001	0.19	.28	<.001
BL vs wk 8	–3.16	–.63	<.001	–3.52	–.78	<.001	–0.44	–.68	<.001	0.35	.52	<.001	0.25	.38	<.001
BL vs wk 12	–3.58	–.71	<.001	–3.96	–.88	<.001	–0.5	–.78	<.001	0.43	.64	<.001	0.31	.47	<.001
BL vs wk 16	–3.50	–.70	<.001	–4.07	–.91	<.001	–0.54	–.84	<.001	0.50	.75	<.001	0.34	.51	<.001
BL vs wk 32	–3.56	–.71	<.001	–4.22	–.94	<.001	–0.53	–.82	<.001	0.42	.64	<.001	0.32	.48	<.001

[a]BL: baseline.

Table 3. Changes in anxiety symptoms, depression symptoms, perceived stress, well-being, and quality of life at follow-up time points versus baseline (random effects).

	Depressive symptoms	Anxious symptoms	Stress	Well-being	Quality of life
Random effects					
σ^2	8.60	8.53	0.14	0.17	0.11
Γ_{00}	$15.56_{accesscode}$	$9.55_{accesscode}$	$0.24_{accesscode}$	$0.25_{accesscode}$	$0.32_{accesscode}$
ICC[a]	0.64	0.53	0.64	0.60	0.74
N	$273_{accesscode}$	$273_{accesscode}$	$272_{accesscode}$	$273_{accesscode}$	$273_{accesscode}$
Observations	1552	1555	1545	1548	1545
Marginal R^2/conditional R^2	0.065/0.667	0.109/0.579	0.088/0.669	0.062/0.629	0.032/0.743

[a]ICC: intraclass correlation coefficient.

Clinical Severity

At baseline, nearly half of all participants met clinically relevant criteria for anxiety (130/273, 47.6%) and depression (130/273, 47.6%) symptoms. A mixed-effects binomial logistic regression with a random intercept was conducted to analyze changes in clinical severity. Reductions in the probability of meeting clinically relevant criteria for anxiety or depression were significant across all time points (P<.001), as shown in Tables 4 and 5. By week 4, the model estimates that only 10% and 12% met the criteria for anxiety or depression, respectively. By week 8, these estimated percentages further decreased to 8% and 9%. The substantial reduction in participants meeting clinical criteria for anxiety or depression was maintained on average through program end (16 weeks) and follow-up (32 weeks), as illustrated in Figure 3.

Table 4. Changes in anxiety symptom severity and depression symptom severity at follow-up time points versus baseline (predictors).

	Depressive symptom severity			Anxious symptom severity		
	Odds ratios	Standard β	*P* value	Odds ratios	Standard β	*P* value
Predictors						
Intercept	0.78	.78	.31	0.83	.83	.37
BL[a] vs wk 4	0.18	.18	<.001	0.14	.14	<.001
BL vs wk 8	0.13	.13	<.001	0.11	.11	<.001
BL vs wk 12	0.09	.09	<.001	0.08	.08	<.001
BL vs wk 16	0.10	.10	<.001	0.08	.08	<.001
BL vs wk 32	0.08	.08	<.001	0.08	.08	<.001

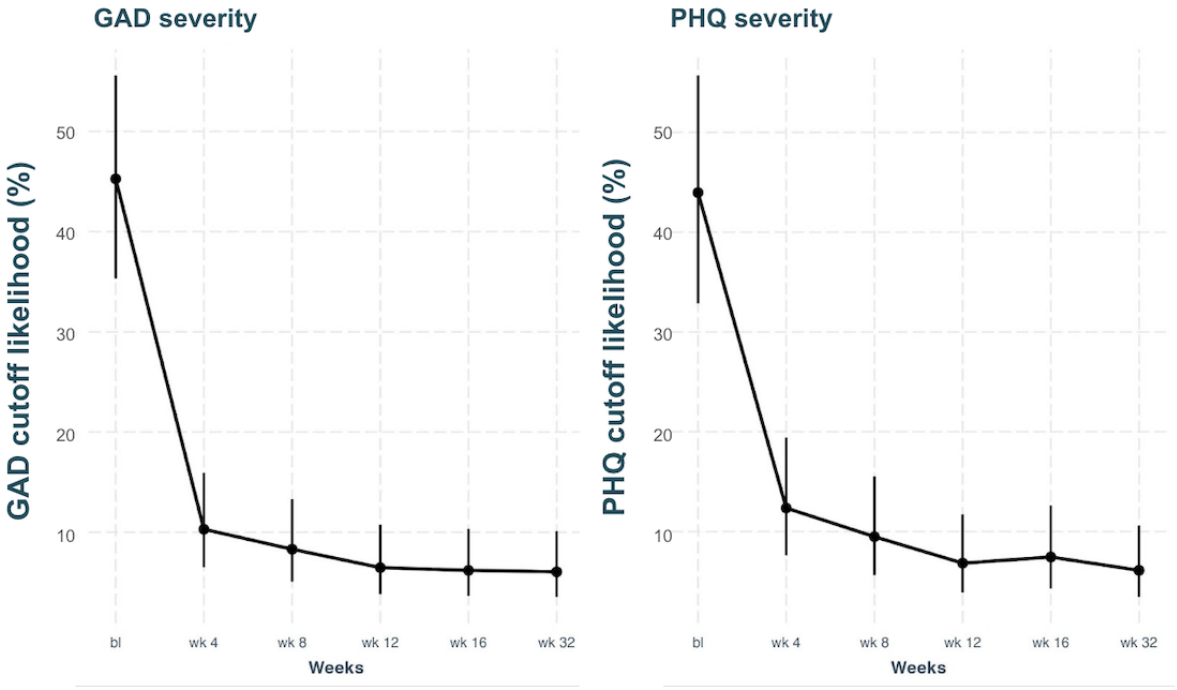
[a]BL: baseline.

Table 5. Changes in anxiety symptom severity and depression symptom severity at follow-up time points versus baseline (random effects).

	Depressive symptom severity	Anxious symptom severity
Random effects		
σ^2	3.29	3.29
Γ_{00}	$6.35_{accesscode}$	$4.45_{accesscode}$
ICC[a]	0.66	0.58
N	$273_{accesscode}$	$273_{accesscode}$
Observations	1552	1555
Marginal R^2/conditional R^2	0.072/0.683	0.098/0.617

[a]ICC: intraclass correlation coefficient.

Figure 3. Reductions in the probability of meeting a clinically relevant cutoff score of 10 for GAD and PHQ. Error bars represent SEs. bl: baseline; GAD: Generalized Anxiety Disorder Scale; PHQ: Patient Health Questionnaire depression scale.



Overall Well-Being

Secondary effects of the intervention on coping, emotion regulation, sleep, and workplace absenteeism or presenteeism were examined via separate models for each variable. Absenteeism and presenteeism days were aggregated to form a comprehensive productivity loss outcome. A 0-inflated Poisson mixed-effects model with a random intercept was used for absenteeism and presenteeism. By program end (16 weeks), participants gained 2.57 productive work days from baseline, with this number remaining relatively stable (2.23 days) at follow-up (32 weeks). Effects across all well-being variables exhibited positive and significant trends at all time points (P<.001), as summarized in Tables 6 and 7 and displayed in Figure 4.

Table 6. Changes to well-being at follow-up time points versus baseline (predictors).

	Coping			Emotion dysregulation			Sleep disturbance			Absenteeism			Presenteeism			Productivity loss		
	Estimates	Standard β	*P* value	Estimates	Standard β	*P* value	Estimates	Standard β	*P* value	Incidence rate ratios	Standard β	*P* value	Incidence rate ratios	Standard β	*P* value	Incidence rate ratios	Standard β	*P* value
Predictors																		
Intercept	2.71	–.42	<.001	2.49	.38	<.001	931	.37	<.001	0.60	.60	.001	3.92	3.92	<.001	5.04	5.04	<.001
BL[a] vs wk 16										0.66	.66	<.001	0.66	.66	<.001	0.67	.67	<.001
BL vs wk 4	0.34	.51	<.001	–0.30	–.44	<.001	–1.83	–.44	<.001	0.59	.59	<.001	0.63	.63	<.001	0.63	.63	<.001
BL vs wk 32										0.58	.58	<.001	0.44	.44	<.001	0.48	.48	<.001
BL vs wk 8	0.43	.63	<.001	–0.40	–.58	<.001	–2.20	–.52	<.001	0.51	.51	<.001	0.45	.45	<.001	0.46	.46	<.001
BL vs wk 12	0.41	.61	<.001	–0.37	–.54	<.001	–2.23	–.53	<.001	0.40	.40	<.001	0.46	.46	<.001	0.46	.46	<.001

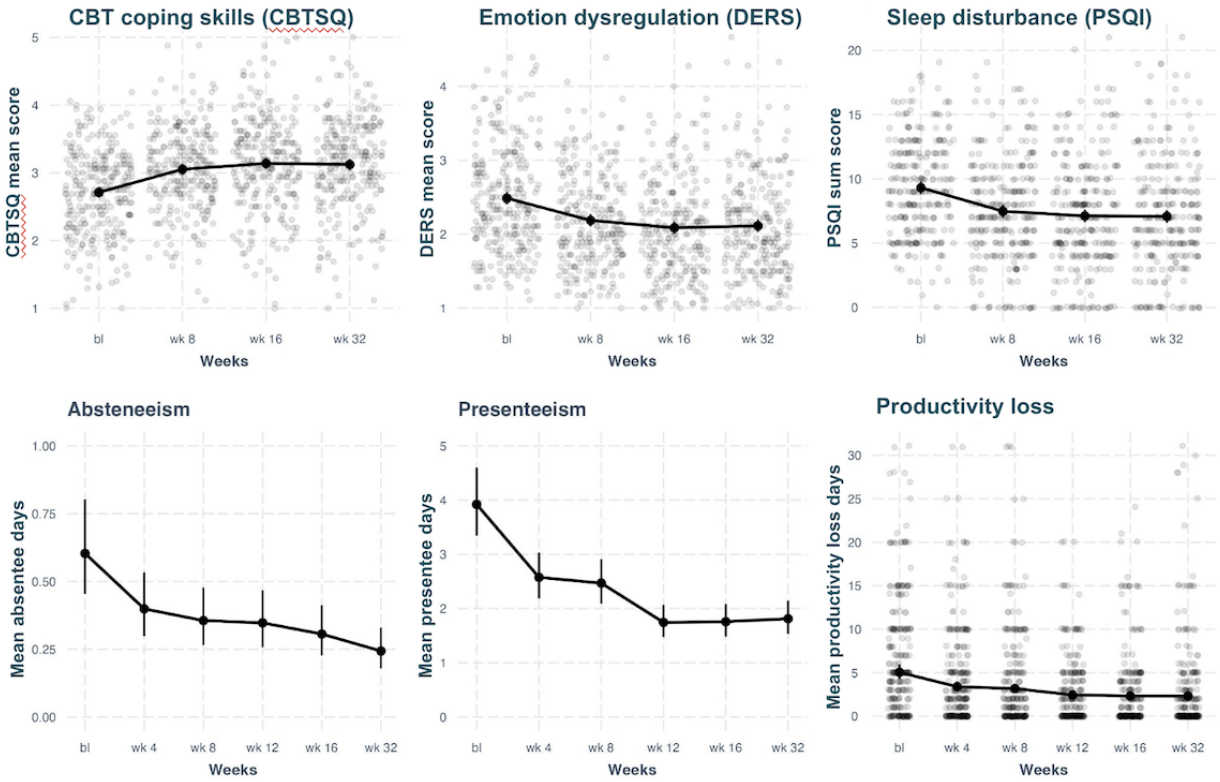
[a]BL: baseline.

Table 7. Changes to well-being at follow-up time points versus baseline (random effects).

	Coping	Emotion dysregulation	Sleep disturbance	Absenteeism	Presenteeism	Productivity loss
Random effects						
σ^2	0.19	0.17	7.15	1.28	0.35	0.27
Γ_{00}	$0.24_{accesscode}$	$0.28_{accesscode}$	$9.62_{accesscode}$	$3.64_{accesscode}$	$1.53_{accesscode}$	$1.65_{accesscode}$
ICC[a]	0.56	0.63	0.57	0.74	0.82	0.86
N	$273_{accesscode}$	$273_{accesscode}$	$273_{accesscode}$	$273_{accesscode}$	$273_{accesscode}$	$273_{accesscode}$
Observations	1030	1030	1100	1478	1479	1476
Marginal R^2/conditional R^2	0.068/0.593	0.055/0.649	0.048/0.594	0.016/0.743	0.044/0.824	0.040/0.863

[a]ICC: intraclass correlation coefficient.

Figure 4. Plots of mixed-effects models for changes to well-being over time. bl: baseline; CBT: cognitive behavioral therapy; CBTSQ: Cognitive Behavioral Therapy Skills Questionnaire; DERS: Difficulties in Emotion Regulation Scale; PSQI: Pittsburgh Sleep Quality Index.



Discussion

Principal Findings

This study contributes to research on mobile mental wellness interventions by examining the efficacy and postintervention outcomes of the Noom Mood program for the general population. The investigation addresses gaps in the existing literature by examining the impact of this type of intervention on mental health and well-being outcomes both postintervention and after a 16-week maintenance period. Using a prospective 1-arm design, our study evaluated the effects of Noom Mood on an array of mental health indicators and well-being measures, spanning from baseline assessment to a follow-up assessment 32 weeks later (16 weeks after the conclusion of the intervention phase). Promising results were seen in terms of improvements in mental health, workplace productivity, and well-being outcomes.

Our study benefited from several notable strengths. First, a major strength of this study was the achievement of retention rates ranging from 85% to 98%. Low retention rates are a widely acknowledged barrier to understanding and applying the results of digital health research. Other digital studies have often found the most successful retention rates (48%-56%) to still remain suboptimal [31,43]. The higher retention rates observed within this study are likely related to the screening practices used, such as the decisional balance exercise, which are designed to both select for and enhance participant motivation and commitment at this study's start [29]. An additional strength of this study is the generalizability of the results. Given that we did not select individuals with a certain score on mental health measures, the results support the intervention's ability to reduce distress and enhance well-being for this quite large group of individuals. Furthermore, we did not use a certain engagement threshold for this study, meaning that participants were able to use the app as much or as little as they liked, and the results remained promising. Finally, this type of mobile app–based intervention has a large potential for scalability.

Mental Health

Mobile mental wellness apps have shown promising potential in providing support and resources for individuals experiencing stress, symptoms of depression, and symptoms of anxiety [21,44,45]. Further, a study by Economides et al [25] reported substantial and sustained improvements in depression and anxiety symptoms 6 months after an 8-week therapist-led smartphone-based intervention for elevated symptoms of anxiety and depression. Likewise, a recent systematic review and meta-analysis of mobile mental wellness interventions suggested the high potential of these interventions to improve (or significantly support the improvement of) mental health challenges such as depression and anxiety [28]. These findings underscore the potential for a prolonged positive impact of mental wellness interventions on major mental health dimensions.

Similarly, a recent systematic review and meta-analysis of 9 randomized controlled trials (RCTs) including 1837 participants also found small to medium effect sizes of mobile mental wellness apps on anxious symptoms, though when compared

with an active control only small effect sizes were found [21]. Similarly, a meta-analysis of 18 RCTs including 3414 participants found that depressive symptoms were reduced significantly following a digital mental health intervention when compared with a control condition; effect sizes were small, however, when compared to active control conditions [45].

Our findings showed statistically significant reductions within 4 weeks of program initiation for both anxiety and depression symptoms, assessed using the GAD-7 and 8-item Patient Health Questionnaire depression scales respectively. These initial improvements are maintained throughout the program, from the intervention's conclusion at 16 weeks to the 32-week follow-up assessment. Outside of symptom reduction, we observed a significant reduction in the proportion of participants meeting clinically relevant anxiety and depression symptom criteria at all assessment points. Our findings align with other published results, however, the outcomes of the meta-analyses highlight the need for a study comparing Noom Mood with an active control condition.

Workplace Productivity and Accessibility

Previous studies have shown a clear connection between mental health and workplace productivity [13,46-49]. Further, a cross-sectional study found the mean days of absenteeism and presenteeism were significantly higher among participants with moderate or high psychological distress compared to low distress [47], and a recent multi-arm, pilot RCT examining the outcome of digital interventions among working adults showed promising preliminary findings indicating improved depressive symptoms, well-being, and functioning following a digital intervention [48].

In parallel, this study showed the potential of a mental wellness intervention to reduce both symptoms of depression and decreased workplace absences along with improved workplace focus. The personalized nature and convenience of mobile interventions allow them to be positioned as promising tools for reaching individuals with limited access to conventional mental health services.

Furthermore, to rationalize employer investment in mobile mental wellness programs, it is necessary to demonstrate their effectiveness in not only improving employee health and well-being but also in saving money. Previous studies have indicated that digital mental wellness interventions can reduce weekly costs by US $155.82 over 8 weeks [49]. Given the anticipated costs to employers associated with anxiety, stress, and depressive symptoms, there is a clear indication that scalable digital interventions such as Noom Mood have the potential to demonstrate significant cost savings within the workplace.

Well-Being

Several studies have cited the scarcity of research looking at various dimensions of well-being within the context of self-guided mobile mental wellness interventions [27,28]. The literature that does exist is somewhat ambivalent, with some studies finding improvements in well-being as a result of a digital mental wellness intervention and some not. For instance, a recent systematic review and meta-analysis found that digital mental wellness apps, when compared to controls, showed small effects for reducing mental health symptoms and improving well-being, but a medium effect for emotion regulation [28]. Another recent RCT found that the group using a mindfulness app for 4 weeks saw small reductions in stress and depression, but students who practiced more did not experience additional improvement in well-being [50].

Our study reveals enhancements in well-being dimensions including coping skills, emotion regulation, and sleep quality, and suggests that mobile mental wellness programs are not only beneficial in terms of symptom reduction but also support a variety of well-being dimensions. However, given the existing literature, and the limitations of this study's design, more rigorous testing is needed to better understand this association.

Limitations

Limitations of this study include the absence of a control group, lack of sample diversity, reliance on self-report measures, potential selection bias from recruitment via social media ads, and the potential impact of high retention rates on outcomes' representativeness. Importantly, results should be interpreted with caution as this study's design did not use a control group, meaning results could in part be due to the passage of time alone. This study advances understanding of the effects of a commercially available mental wellness program on various mental health outcomes and contributes potential insights into the program's utility for the general population; however, robust research is needed to affirm the program's efficacy, particularly across diverse populations.

Conclusions

In conclusion, this single-arm prospective study examined 16-week and 32-week outcomes of the Noom Mood self-guided smartphone-based mental wellness program, including a 16-week maintenance period. Our outcomes showed positive impacts on mental health and well-being dimensions, both in terms of symptom reduction and broader psychosocial aspects, though outcomes are limited in scope due to the lack of a control arm in this study's design. The results highlight significant improvements across all mental health and well-being outcomes at program end (16 weeks), which extend to postprogram follow-up (32 weeks). Significant improvements emerge as early as the first follow-up assessment, conducted 4 weeks after program initiation. These findings contribute to the evidence base of mental wellness programs and address gaps in the literature concerning maintenance and the comprehensive effects of a digital self-guided mental wellness intervention.

Future studies must include an RCT study design and a study population spanning diverse groups, accounting for race and ethnicity, education level, and income level. Additionally, future research should aim to understand the specific mechanisms underlying the observed improvements as well as explore the program's scalability and feasibility within real-world settings. Further research investigating the program's potential for long-term adherence and sustained outcomes may additionally provide valuable insights for both practitioners and researchers. Lastly, a direct comparison of the effectiveness of the Noom Mood program with similar interventions may shed light on the

relative advantages and disadvantages of different approaches in the field of mobile mental wellness.

Acknowledgments

MM and ESM conceived this study. MM, PT, KB, and MCA carried out study management. MB analyzed data. PT, MF, and AH wrote this paper. AM and CNM provided supervision. All authors were involved in revising this paper and had final approval of the submitted version.

Conflicts of Interest

Authors MM, MB, PT, KB, MCA, and AM are employees at Noom Inc in the academic research department and have received salary and stock options for their employment. Authors MF, AH, ESM, and CNM were employees at Noom at the time of this study. This study was funded by Noom, which did not (other than the specific authors listed above) play a role in this study's design, execution, analyses, interpretation of the data, or the decision to publish the results.

References

1. Daly M. Prevalence of psychological distress among working-age adults in the United States, 1999-2018. Am J Public Health 2022 Jul;112(7):1045-1049. [doi: 10.2105/AJPH.2022.306828] [Medline: 35617656]
2. Viertiö S, Kiviruusu O, Piirtola M, Kaprio J, Korhonen T, Marttunen M, et al. Factors contributing to psychological distress in the working population, with a special reference to gender difference. BMC Public Health 2021;21(1):611 [FREE Full text] [doi: 10.1186/s12889-021-10560-y] [Medline: 33781240]
3. Javaid SF, Hashim IJ, Hashim MJ, Stip E, Samad MA, Ahbabi AA. Epidemiology of anxiety disorders: global burden and sociodemographic associations. Middle East Curr Psychiatry 2023;30(1):44. [doi: 10.1186/s43045-023-00315-3]
4. Bandelow B, Michaelis S. Epidemiology of anxiety disorders in the 21st century. Dialogues Clin Neurosci 2015;17(3):327-335 [FREE Full text] [doi: 10.31887/DCNS.2015.17.3/bbandelow] [Medline: 26487813]
5. Salari N, Hosseinian-Far A, Jalali R, Vaisi-Raygani A, Rasoulpoor S, Mohammadi M, et al. Prevalence of stress, anxiety, depression among the general population during the COVID-19 pandemic: a systematic review and meta-analysis. Global Health 2020;16(1):57 [FREE Full text] [doi: 10.1186/s12992-020-00589-w] [Medline: 32631403]
6. McGinty EE, Presskreischer R, Anderson KE, Han H, Barry CL. Psychological distress and COVID-19-related stressors reported in a longitudinal cohort of US adults in April and July 2020. JAMA 2020;324(24):2555-2557 [FREE Full text] [doi: 10.1001/jama.2020.21231] [Medline: 33226420]
7. Ettman C, Abdalla S, Cohen G, Sampson L, Vivier PM, Galea S. Prevalence of depression symptoms in US adults before and during the COVID-19 pandemic. JAMA Netw Open 2020;3(9):e2019686 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.19686] [Medline: 32876685]
8. Wang Y, Kala M, Jafar T. Factors associated with psychological distress during the coronavirus disease 2019 (COVID-19) pandemic on the predominantly general population: a systematic review and meta-analysis. PLoS One 2020;15(12):e0244630 [FREE Full text] [doi: 10.1371/journal.pone.0244630] [Medline: 33370404]
9. Lahav Y. Psychological distress related to COVID-19—the contribution of continuous traumatic stress. J Affect Disord 2020;277:129-137 [FREE Full text] [doi: 10.1016/j.jad.2020.07.141] [Medline: 32818776]
10. Bobo WV, Grossardt BR, Virani S, St Sauver JL, Boyd CM, Rocca WA. Association of depression and anxiety with the accumulation of chronic conditions. JAMA Netw Open 2022;5(5):e229817 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.9817] [Medline: 35499825]
11. Fiksdal A, Hanlin L, Kuras Y, Gianferante D, Chen X, Thoma MV, et al. Associations between symptoms of depression and anxiety and cortisol responses to and recovery from acute stress. Psychoneuroendocrinology 2019;102:44-52 [FREE Full text] [doi: 10.1016/j.psyneuen.2018.11.035] [Medline: 30513499]
12. Hockey M, Rocks T, Ruusunen A, Jacka FN, Huang W, Liao B, et al. Psychological distress as a risk factor for all-cause, chronic disease- and suicide-specific mortality: a prospective analysis using data from the National Health Interview Survey. Soc Psychiatry Psychiatr Epidemiol 2022;57(3):541-552 [FREE Full text] [doi: 10.1007/s00127-021-02116-7] [Medline: 34363488]
13. Johnston D, Harvey S, Glozier N, Calvo R, Christensen H, Deady M. The relationship between depression symptoms, absenteeism and presenteeism. J Affect Disord 2019;256:536-540. [doi: 10.1016/j.jad.2019.06.041] [Medline: 31280078]
14. Hendriks SM, Spijker J, Licht CM, Hardeveld F, de Graaf R, Batelaan NM, et al. Long-term work disability and absenteeism in anxiety and depressive disorders. J Affect Disord 2015;178:121-130. [doi: 10.1016/j.jad.2015.03.004] [Medline: 25805404]
15. Bailey SK, Haggarty J, Kelly S. Global absenteeism and presenteeism in mental health patients referred through primary care. Work 2015;53(2):399-408. [doi: 10.3233/WOR-152172] [Medline: 26519010]
16. Leka S, Jain A. Health impact of psychosocial hazards at work: an overview. World Health Organization. 2010. URL: https://iris.who.int/bitstream/handle/10665/44428/?sequence=1 [accessed 2024-06-05]

17. Mental illness. National Institute of Mental Health. 2023. URL: https://www.nimh.nih.gov/health/statistics/mental-illness [accessed 2024-06-04]
18. Wang K, Varma DS, Prosperi M. A systematic review of the effectiveness of mobile apps for monitoring and management of mental health symptoms or disorders. J Psychiatr Res 2018;107:73-78. [doi: 10.1016/j.jpsychires.2018.10.006] [Medline: 30347316]
19. Sareen J, Jagdeo A, Cox BJ, Clara I, ten Have M, Belik S, et al. Perceived barriers to mental health service utilization in the United States, Ontario, and the Netherlands. Psychiatr Serv 2007;58(3):357-364. [doi: 10.1176/ps.2007.58.3.357] [Medline: 17325109]
20. Walker ER, Cummings JR, Hockenberry JM, Druss BG. Insurance status, use of mental health services, and unmet need for mental health care in the United States. Psychiatr Serv 2015;66(6):578-584 [FREE Full text] [doi: 10.1176/appi.ps.201400248] [Medline: 25726980]
21. Firth J, Torous J, Nicholas J, Carney R, Rosenbaum S, Sarris J. Can smartphone mental health interventions reduce symptoms of anxiety? A meta-analysis of randomized controlled trials. J Affect Disord 2017;218:15-22 [FREE Full text] [doi: 10.1016/j.jad.2017.04.046] [Medline: 28456072]
22. Müller AM, Maher CA, Vandelanotte C, Hingle M, Middelweerd A, Lopez ML, et al. Physical activity, sedentary behavior, and diet-related ehealth and mhealth research: bibliometric analysis. J Med Internet Res 2018;20(4):e122 [FREE Full text] [doi: 10.2196/jmir.8954] [Medline: 29669703]
23. Hutchesson MJ, Gough C, Müller AM, Short CE, Whatnall MC, Ahmed M, et al. eHealth interventions targeting nutrition, physical activity, sedentary behavior, or obesity in adults: A scoping review of systematic reviews. Obes Rev 2021;22(10):e13295. [doi: 10.1111/obr.13295] [Medline: 34159684]
24. McCallum M, Ho AS, Mitchell ES, May CN, Behr H, Ritschel L, et al. Feasibility, acceptability, and preliminary outcomes of a cognitive behavioral therapy-based mobile mental well-being program (Noom Mood): single-arm prospective cohort study. JMIR Form Res 2022;6(4):e36794 [FREE Full text] [doi: 10.2196/36794] [Medline: 35436218]
25. Economides M, Ranta K, Nazander A, Hilgert O, Goldin PR, Raevuori A, et al. Long-term outcomes of a therapist-supported, smartphone-based intervention for elevated symptoms of depression and anxiety: quasiexperimental, pre-postintervention study. JMIR mHealth uHealth 2019;7(8):e14284 [FREE Full text] [doi: 10.2196/14284] [Medline: 31452521]
26. Linardon J, Cuijpers P, Carlbring P, Messer M, Fuller-Tyszkiewicz M. The efficacy of app-supported smartphone interventions for mental health problems: a meta-analysis of randomized controlled trials. World Psychiatry 2019;18(3):325-336 [FREE Full text] [doi: 10.1002/wps.20673] [Medline: 31496095]
27. Bakker D, Kazantzis N, Rickwood D, Rickard N. Mental health smartphone apps: review and evidence-based recommendations for future developments. JMIR Ment Health 2016;3(1):e7 [FREE Full text] [doi: 10.2196/mental.4984] [Medline: 26932350]
28. Eisenstadt M, Liverpool S, Infanti E, Ciuvat RM, Carlsson C. Mobile apps that promote emotion regulation, positive mental health, and well-being in the general population: systematic review and meta-analysis. JMIR Ment Health 2021;8(11):e31170 [FREE Full text] [doi: 10.2196/31170] [Medline: 34747713]
29. Goldberg J, Kiernan M. Innovative techniques to address retention in a behavioral weight-loss trial. Health Educ Res 2005;20(4):439-447. [doi: 10.1093/her/cyg139] [Medline: 15598664]
30. Faul F, Erdfelder E, Lang A, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007;39(2):175-191. [doi: 10.3758/bf03193146] [Medline: 17695343]
31. Torous J, Lipschitz J, Ng M, Firth J. Dropout rates in clinical trials of smartphone apps for depressive symptoms: a systematic review and meta-analysis. J Affect Disord 2020;263:413-419. [doi: 10.1016/j.jad.2019.11.167] [Medline: 31969272]
32. Alavi CB, Massman JD. Selecting an electronic data capture system. Urol Pract 2016;3(3):236-241. [doi: 10.1016/j.urpr.2015.06.007] [Medline: 37592499]
33. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
34. Kroenke K, Strine TW, Spitzer RL, Williams JB, Berry JT, Mokdad AH. The PHQ-8 as a measure of current depression in the general population. J Affect Disord 2009;114(1-3):163-173. [doi: 10.1016/j.jad.2008.06.026] [Medline: 18752852]
35. Lee E. Review of the psychometric evidence of the Perceived Stress Scale. Asian Nurs Res (Korean Soc Nurs Sci) 2012;6(4):121-127 [FREE Full text] [doi: 10.1016/j.anr.2012.08.004] [Medline: 25031113]
36. Stewart-Brown S, Platt S, Tennant A, Maheswaran H, Parkinson J, Weich S, et al. The Warwick-Edinburgh Mental Well-Being Scale (WEMWBS): a valid and reliable tool for measuring mental well-being in diverse populations and projects. J Epidemiol Community Health 2011;65(Suppl 2):A38-A39. [doi: 10.1136/jech.2011.143586.86]
37. Leon AC, Olfson M, Portera L, Farber L, Sheehan DV. Assessing psychiatric impairment in primary care with the Sheehan Disability Scale. Int J Psychiatry Med 1997;27(2):93-105. [doi: 10.2190/T8EM-C8YH-373N-1UWD] [Medline: 9565717]
38. Brouwer W, Verbooy K, Hoefman R, van Exel J. Production losses due to absenteeism and presenteeism: the influence of compensation mechanisms and multiplier effects. Pharmacoeconomics 2023;41(9):1103-1115 [FREE Full text] [doi: 10.1007/s40273-023-01253-y] [Medline: 36856941]

39. Brunner B, Igic I, Keller AC, Wieser S. Who gains the most from improving working conditions? health-related absenteeism and presenteeism due to stress at work. Eur J Health Econ 2019;20(8):1165-1180 [FREE Full text] [doi: 10.1007/s10198-019-01084-9] [Medline: 31309366]
40. Elsman EBM, Roorda LD, Crins MHP, Boers M, Terwee CB. Dutch reference values for the patient-reported outcomes measurement information system scale v1.2 - Global Health (PROMIS-GH). J Patient Rep Outcomes 2021;5(1):38 [FREE Full text] [doi: 10.1186/s41687-021-00314-0] [Medline: 33978855]
41. Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res 1989;28(2):193-213. [doi: 10.1016/0165-1781(89)90047-4] [Medline: 2748771]
42. Victor SE, Klonsky ED. Validation of a brief version of the Difficulties in Emotion Regulation Scale (DERS-18) in five samples. J Psychopathol Behav Assess 2016;38(4):582-589. [doi: 10.1007/s10862-016-9547-9]
43. Fillipo IRG, Pullmann MD, Hull TD, Zech J, Wu J, Litvin B, et al. Participant retention in a fully remote trial of digital psychotherapy: comparison of incentive types. Front Digit Health 2022;4:963741 [FREE Full text] [doi: 10.3389/fdgth.2022.963741] [Medline: 36148211]
44. Lecomte T, Potvin S, Corbière M, Guay S, Samson C, Cloutier B, et al. Mobile apps for mental health issues: meta-review of meta-analyses. JMIR mHealth uHealth 2020;8(5):e17458 [FREE Full text] [doi: 10.2196/17458] [Medline: 32348289]
45. Firth J, Torous J, Nicholas J, Carney R, Pratap A, Rosenbaum S, et al. The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. World Psychiatry 2017;16(3):287-298 [FREE Full text] [doi: 10.1002/wps.20472] [Medline: 28941113]
46. Stratton E, Jones N, Peters SE, Torous J, Glozier N. Digital mhealth interventions for employees: systematic review and meta-analysis of their effects on workplace outcomes. J Occup Environ Med 2021;63(8):e512-e525. [doi: 10.1097/JOM.0000000000002267] [Medline: 34050095]
47. Elotla SF, Gaafar SEM, Ameen AE, AM F. Association between mental health distress and work productivity: a cross-sectional study. Egypt J Occup Med 2021;45(2):51-68. [doi: 10.21608/EJOM.2021.170562]
48. Taylor RW, Male R, Economides M, Bolton H, Cavanagh K. Feasibility and preliminary efficacy of digital interventions for depressive symptoms in working adults: multiarm randomized controlled trial. JMIR Form Res 2023;7:e41590 [FREE Full text] [doi: 10.2196/41590] [Medline: 37327027]
49. Huberty JL, Espel-Huynh HM, Neher TL, Puzia ME. Testing the pragmatic effectiveness of a consumer-based mindfulness mobile app in the workplace: randomized controlled trial. JMIR mHealth uHealth 2022;10(9):e38903 [FREE Full text] [doi: 10.2196/38903] [Medline: 36169991]
50. Lahtinen O, Aaltonen J, Kaakinen J, Franklin L, Hyönä J. The effects of app-based mindfulness practice on the well-being of university students and staff. Curr Psychol 2023;42(6):4412-4421 [FREE Full text] [doi: 10.1007/s12144-021-01762-z] [Medline: 33967565]

Abbreviations

GAD-7: 7-item Generalized Anxiety Disorder Scale
HIPAA: Health Insurance Portability and Accountability Act
RCT: randomized controlled trial

Edited by L Buis; submitted 17.11.23; peer-reviewed by A Serlachius, G Jonathan; comments to author 19.01.24; revised version received 21.02.24; accepted 22.05.24; published 27.06.24.

Please cite as:
McCallum M, Baldwin M, Thompson P, Blessing K, Frisch M, Ho A, Ainsworth MC, Mitchell ES, Michaelides A, May CN
Long-Term Efficacy of a Mobile Mental Wellness Program: Prospective Single-Arm Study
JMIR Mhealth Uhealth 2024;12:e54634
URL: https://mhealth.jmir.org/2024/1/e54634
doi:10.2196/54634
PMID:

Original Paper

mHealth Apps for Dementia, Alzheimer Disease, and Other Neurocognitive Disorders: Systematic Search and Environmental Scan

Suad Ali[1,2,3], MSc; Hira Alizai[1], BHS; Delal Jemal Hagos[1], BSc; Sindy Ramos Rubio[1], BHS; Dale Calabia[1], BSc, BHS; Penelope Serrano Jimenez[1], BHS; Vinuu Aarif Senthil[1], BHS; Lora Appel[1,2,4], PhD

[1]Faculty of Health, York University, Toronto, ON, Canada

[2]Knowledge, Innovation, Talent, Everywhere, OpenLab, University Health Network, Toronto, ON, Canada

[3]Women's Brain Health Initiative, Toronto, ON, Canada

[4]Michael Garron Hospital, Toronto, ON, Canada

Corresponding Author:
Lora Appel, PhD
Faculty of Health
York University
4700 Keele Street
Toronto, ON, M3J 1P3
Canada
Phone: 1 416 736 2100
Email: lora.appel@yorku.ca

Abstract

Background: Lifestyle behaviors including exercise, sleep, diet, stress, mental stimulation, and social interaction significantly impact the likelihood of developing dementia. Mobile health (mHealth) apps have been valuable tools in addressing these lifestyle behaviors for general health and well-being, and there is growing recognition of their potential use for brain health and dementia prevention. Effective apps must be evidence-based and safeguard user data, addressing gaps in the current state of dementia-related mHealth apps.

Objective: This study aims to describe the scope of available apps for dementia prevention and risk factors, highlighting gaps and suggesting a path forward for future development.

Methods: A systematic search of mobile app stores, peer-reviewed literature, dementia and Alzheimer association websites, and browser searches was conducted from October 19, 2022, to November 2, 2022. A total of 1044 mHealth apps were retrieved. After screening, 152 apps met the inclusion criteria and were coded by paired, independent reviewers using an extraction framework. The framework was adapted from the Silberg scale, other scoping reviews of mHealth apps for similar populations, and background research on modifiable dementia risk factors. Coded elements included evidence-based and expert credibility, app features, lifestyle elements of focus, and privacy and security.

Results: Of the 152 apps that met the final selection criteria, 88 (57.9%) addressed modifiable lifestyle behaviors associated with reducing dementia risk. However, many of these apps (59/152, 38.8%) only addressed one lifestyle behavior, with mental stimulation being the most frequently addressed. More than half (84/152, 55.2%) scored 2 points out of 9 on the Silberg scale, with a mean score of 2.4 (SD 1.0) points. Most of the 152 apps did not disclose essential information: 120 (78.9%) did not disclose expert consultation, 125 (82.2%) did not disclose evidence-based information, 146 (96.1%) did not disclose author credentials, and 134 (88.2%) did not disclose their information sources. In addition, 105 (69.2%) apps did not disclose adherence to data privacy and security practices.

Conclusions: There is an opportunity for mHealth apps to support individuals in engaging in behaviors linked to reducing dementia risk. While there is a market for these products, there is a lack of dementia-related apps focused on multiple lifestyle behaviors. Gaps in the rigor of app development regarding evidence base, credibility, and adherence to data privacy and security standards must be addressed. Following established and validated guidelines will be necessary for dementia-related apps to be effective and advance successfully.

(JMIR Mhealth Uhealth 2024;12:e50186) doi:10.2196/50186

KEYWORDS

dementia; Alzheimer disease; mHealth; mobile health; apps; lifestyle behaviors; mobile phone

Introduction

Background

Dementia, a condition with global impact and no current cure, has led to a rise in mobile health (mHealth) apps catering to individuals and caregivers [1-3]. These apps offer features such as location tracking [4], medication reminders [4], education [3], support for caregivers [3], and planning and information sharing across one's circle of care including with health care providers [5]. Some also focus on screening and diagnostic capabilities, with the ability to analyze large amounts of personal data (eg, changes in voice) that would not have been possible to collect previously [6].

At the same time, research suggests that lifestyle behaviors such as physical activity, sleep, stress, mental stimulation, diet, and social interaction play a crucial role in dementia prevention [7]. Research shows that engaging in these behaviors can significantly decrease the risk of dementia. A recent meta-analysis showed that adults in their 40s who participated in physical activities had a considerably decreased risk of dementia years later [8]. One study found that increased levels of stress and exposure to >2 stressful life events significantly increased the risk of all-cause dementia, while levels of neuroticism caused a higher risk of dementia and Alzheimer disease [9]. Another study found that adults participating in mentally stimulating leisure activities had a significantly decreased risk of developing dementia and other cognitive impairments later in life [10]. According to a random-effect model, individuals with sleep disorders had a higher likelihood of having dementia than those with regular sleep participants [11]. Diet has also been found to be a lifestyle behavior of significance, with one study finding that higher adherence to a Mediterranean diet is adversely related to cognitive decline, dementia, or Alzheimer disease [12]. Similarly, a meta-analysis evaluating social relationships and their impact on the risk of dementia found that social engagement helps prevent dementia. These relationships act as a stress buffer and source of information that help develop positive health behaviors and optimal use of health services [13].

In addition, there is a mounting body of research associating sensory loss and dementia, recognizing a connection between an individual's vision, hearing, dual sensory (vision and hearing) impairment, smell, and touch with their cognitive health [14,15]. Fischer et al [16] have observed an association between the impairment of the auditory, visual, and olfactory faculties and cognitive decline in older adults. This study suggests that hearing loss and cognitive decline may be magnified by several probable mechanisms including social isolation, a lifestyle behavior associated with dementia risk [16,17]. mHealth apps show potential in supporting the maintenance of users' sensory health, as observed in a successful pilot that used a mobile app to support electronic devices in the implementation of a home-based cognitive-multisensory-physical exercise program for participants with dementia [18]. These findings demonstrate the potential of exploring if and how dementia-related apps address sensory decline to maintain users' brain health and prevent cognitive decline.

Currently, there are many popular apps that focus on these lifestyle behaviors and have gained tremendous popularity including MapMyRun that focuses on tracking physical activity and MyFitnessPal, another fitness app focused on tracking diet and calorie intake [19,20]. Although these apps address a number of the lifestyle behaviors (mentioned in the previous paragraphs) that have been linked to the likelihood of developing dementia, few apps have been designed to target these functions specifically for persons with or at risk of developing dementia. Given the movement toward designing mHealth apps to promote healthy lifestyle behaviors (eg, improved diet, exercise habits, and memory-training games), these apps could be leveraged to promote such behaviors earlier in life, with the hopes that this can prevent or delay the onset of dementia or manage symptoms once they present [21,22]. Understanding the breadth of what is available in terms of which lifestyle behavior or combination of behaviors that currently available mHealth apps leverage presents an opportunity to create mHealth apps to address existing gaps and needs for targeted populations.

However, as commercially available smartphone apps can be developed and published by anyone, it is essential to understand which of these are evidence-based so that the public may safely use these apps as a tool for dementia education, prevention, or overall well-being promotion [23]. Since these products may also be eligible for medical device accreditation, or at the very least are considered mHealth or wellness products, we need to understand how they are being marketed, what security and privacy measures are in place to protect their users, and whether the information or tools provided are evidence-based [23]. In particular, these privacy and transparency factors must be strongly considered when leveraging existing or new health apps for persons with dementia or when marketing a certain app as a tool for the prevention of dementia or promotion of well-being. Older adults were found to express concern for their personal information getting hacked while using technology, leading to a hesitancy toward its use [24]. In the case of people with dementia, their personal health information gathered by mHealth apps is at greater risk of being breached [25]. Specifically, many app users with dementia live with impaired cognitive capacity that may interfere with their ability to understand the details of an app's privacy policies, which are usually presented in confusing or complex language [25].

Objective

Our objective with this systematic search and environmental scan was to map out the landscape of smartphone apps intended for use by the general public relating to dementia prevention and its risk factors. We have reported on the breadth and depth of available dementia-related mHealth apps and importantly identified gaps that could be addressed by future apps.

Methods

Systematic Search and Environmental Scan

This study was conducted according to PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [26]. Using selected search terms, available mHealth apps directed at dementia and Alzheimer disease and other neurocognitive symptoms were searched for through a systematic search of mobile app stores and supplemented by an environmental scan of peer-reviewed and gray literature. The environmental scan allowed for the identification of a larger subset of relevant apps from external resources to examine the current state of existing mHealth apps directed at dementia and Alzheimer disease and other neurocognitive symptoms [27]. Apps were then screened for eligibility using inclusion and exclusion criteria. Those apps that met the inclusion criteria were then systematically evaluated using an extraction framework developed for this study, available in Multimedia Appendix 1.

Search Strategies

A systematic search of mHealth apps was conducted by the research team from October 19, 2022, to November 2, 2022. The search terms *Dementia*, *Alzheimer's*, and *Mild Cognitive Impairment* OR *Cognitive Decline* were entered into the 4 most popular mobile app stores in North America: App Store (Apple), Google Play store, Samsung store, and Microsoft store. These terms were used to identify apps available in these stores that were directed at dementia and Alzheimer disease and other neurocognitive symptoms, with the research team recording up to the first 100 app results as they were generated.

In addition to app stores, the following 3 data sources were searched:

1. The websites of recognized dementia and Alzheimer associations and advocacy groups across North America that listed any recommended or advertised mobile apps: Alzheimer's Association, Alzheimer's Family Association, Alzheimer's Disease International, Alzheimer Society of Canada, Alzheimer Society of Simcoe County, Dementia Society of Canada, National Aphasia Association, and The Ontario Caregiver Association
2. The top 100 article results and listed relevant mobile apps for the search query "Top <dementia/ Alzheimers/ MCI/MCD> apps" on the Google search engine
3. Peer-reviewed meta-analyses, systematic reviews, and scoping reviews on the topic of "Dementia OR Alzheimers OR MCI/MCD and mobile app"

Once duplicates were removed, HA, DJH, SRR, DC, PSJ, and VAS independently screened the apps in pairs based on our inclusion and exclusion criteria, and conflicts were resolved by a third reviewer among them.

Eligibility Criteria

Apps that indicated any relation to dementia and Alzheimer disease or other neurocognitive symptoms in either title or description were included (eg, dementia-related news, tips for caregivers, and reminder systems). Furthermore, apps available on either the App Store, Google Play, Samsung Galaxy, or Microsoft app stores or a combination of any of the above were included. Textbox 1 defines the inclusion and exclusion criteria for this study in detail.

Textbox 1. App eligibility criteria.

Inclusion criteria

- Available in English
- Title or description of app indicated the purpose or audience related to dementia and Alzheimer disease or other neurocognitive symptoms, using information available on app stores, developer websites, or other web sources

Exclusion criteria

- Not available in English
- Title, app store description, or other information available on other app stores, developer website, or other web sources did not indicate that app purpose or audience is related to dementia and Alzheimer disease or other neurocognitive symptoms
- Unavailable on an app store during screening phase

Data Extraction and Charting Framework

A data extraction framework was developed, partially drawing from the study by Giunti et al [28], where the authors conducted a systematic search of app stores for mHealth apps focused on multiple sclerosis, another neurodegenerative condition. Several elements were modified to fit the context of the nature of dementia risk and lived experience with the condition, which both cope with declining cognitive function. This included the addition of lifestyle behaviors associated with dementia risk as a major coding theme, following background research into the area. The data extraction framework was also designed to set a focus on the credibility of reviewed apps, incorporating coding themes to assess sources consulted, currency, and transparency.

Included apps were independently coded by SA, HA, DJH, SRR, DC, PSJ, and VAS in pairs using the extraction framework described in Multimedia Appendix 1, which included elements such as evidence-based and expert credibility, app features, purpose, lifestyle elements of focus, and privacy and security. In addition, the Silberg scale was incorporated into the data extraction framework, serving as an established evaluation scale designed to assess the accountability of web-based health information, which in this study applies to mHealth apps [29,30]. The Silberg scale was also chosen as it allowed for screening

mHealth apps without downloading them and instead used information available through app store descriptions, which included intended use and audiences, available features, and privacy practices, along with information from developer websites when available. While all 9 of the Silberg scale's evaluation points were adapted from the framework by Jeon et al [29], they were separated from their originally assigned categories and reassigned to an appropriate theme under the new framework.

Data Analysis

The coding process was performed independently by SA, HA, DJH, SRR, DC, PSJ, and VAS in pairs over multiple iterations, with the researchers meeting to discuss connections between the codes and agree on emerging themes at each stage. The final data set was analyzed by examining the responses of the paired coders or, in some instances, the response of a third coder when present to resolve any conflicts. The consolidated extracted data are available in Multimedia Appendix 2. Descriptive statistics were calculated for numeric responses, which included number of downloads and app store rating. Qualitative data were analyzed using thematic analysis.

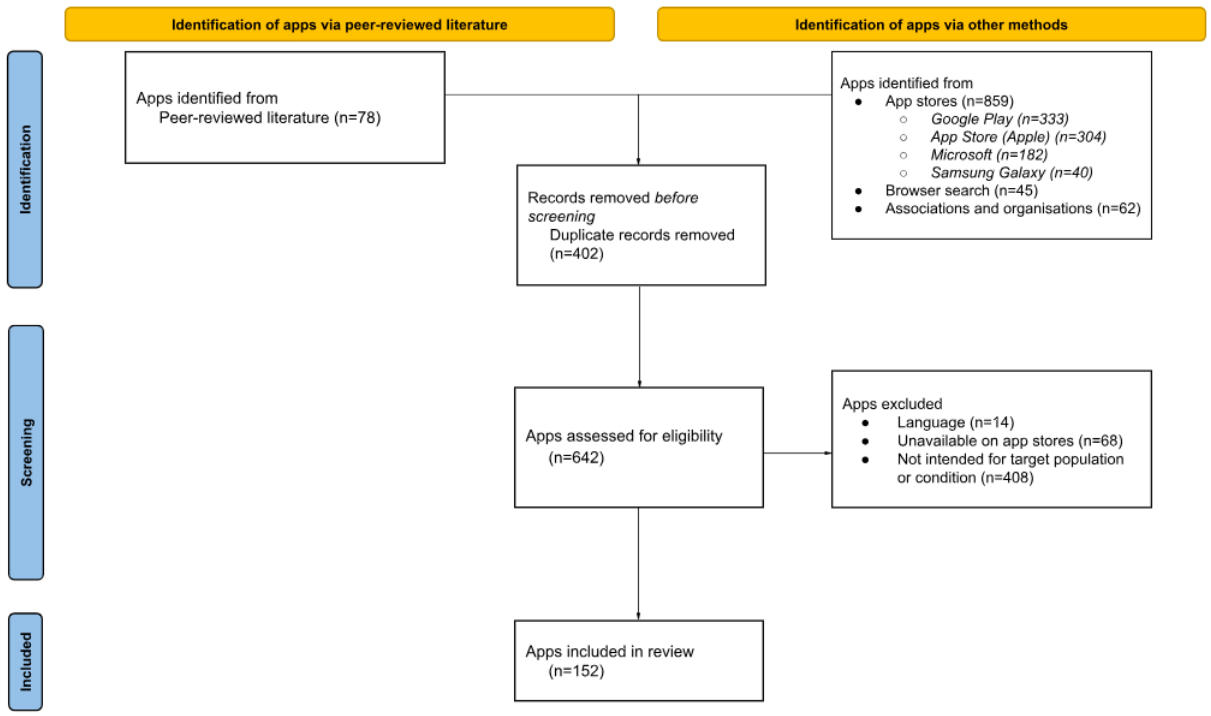
Results

Overview

A total of 1044 apps were found after the initial search for mHealth apps in 4 popular mobile app stores (n=859, 82.27%) and in peer-reviewed literature (n=78, 7.47%). Apps were also found from nonscholarly sources (dementia and Alzheimer association and advocacy websites: n=62, 5.9%; websites accessed through a browser search: n=45, 4.3%). Following this, 402 (38.5%) duplicates were removed, and 642 (61.5%) apps were screened by HA, DJH, SRR, DC, PSJ, and VAS based on our inclusion and exclusion criteria.

Of these 642 screened apps, 14 (2.2%) were not offered in English, 408 (63.6%) of the apps were determined to not be intended for the target populations or condition, and 68 (10.6%) were unavailable on any of the 4 app stores. With the application of these exclusion criteria, 152 (23.7%) apps remained for full data extraction and charting. The PRISMA flow diagram represented in Figure 1 shows this process at each stage.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram of study process.



Themes

The following results are presented thematically using the extraction framework details provided in Multimedia Appendix 1, which includes a breakdown of each theme and its objective, along with the operational definition of each variable used within each theme.

Evidence-Based and Expert Credibility

At minimum, 120 (78.9%) of the 152 apps assessed did not disclose any expert credibility or use of evidence, such as providing references, citing information sources used, or disclosing that in the process of the app's development, there was consultation with at least 1 expert, which includes people with lived experience and caregivers. Table 1 illustrates the number of apps that have disclosed these elements.

Table 1. Number of apps that satisfy app elements (n=152).

App element	Apps, n (%)
Variables related to evidence-based and expert credibility disclosed	
Consulted with at least one expert	32 (21.1)
Evidence-based information	27 (17.8)
Information sources	18 (11.8)
Funding from a medical or health organization	16 (10.5)
References	8 (5.3)
Author credentials	6 (3.9)
Common mHealth app purposes supported	
Education	53 (34.9)
Reminder features	28 (18.4)
Diagnostic tools	25 (16.4)
Habit tracking	13 (8.6)
Progress tracking	7 (4.6)
Goal setting	6 (3.9)
Professional supervision	4 (2.6)
Lifestyle elements associated with reducing dementia risk	
Mental stimulation	59 (38.8)
Social stimulation	26 (17.1)
Exercise	10 (6.6)
Stress management	10 (6.6)
Eating and nutrition	7 (4.6)
Sleep	3 (2.0)
Various elements of transparency disclosed	
Application ownership	143 (94.1)
Author affiliation	20 (13.2)
Authors credited	11 (7.2)
Sponsorship	8 (5.3)

Out of the 152 apps, most apps (n=114, 75%) were created by miscellaneous app developers, in other words, developers who were not stated to be associated with any organization or institution. Researchers and health care organizations made up the next 2 largest categories of app creators, observed in 12 (7.9%) and 11 (7.2%) apps, respectively. Members of the public created 8 (5.3%) apps, while medical app developers created 5 (3.3%) apps. Nonprofit non-health care organizations created 2 (1.3%) apps.

Purpose

Of the 152 apps, approximately one-third (n=53, 34.9%) were categorized under “Health & Fitness” by app stores. Following this, “Medical” and “Educational” comprised the next 2 largest categories at 40 (26.3%) and 18 (11.8%) apps, respectively. Multimedia Appendix 3 presents the wide range of categories assigned by 4 app stores and the number of apps tagged under each category.

Table 2 illustrates the wide range of intended users of the assessed apps. Note that some apps were observed to have targeted >1 group of users. Of the 152 apps, 105 (69%) apps were intended for people with dementia, making them the most frequently targeted audience. Informal caregivers at 79 (52%) apps and the general public (ie, people without neurocognitive symptoms) at 65 (42.8%) apps comprised the next 2 largest groups.

Table 2. Intended app users, up to 3 groups per app (n=152).

Intended user group	Apps, n (%)
People with dementia	105 (69.1)
Informal caregivers	79 (52)
General public	65 (42.8)
Friends or family (noncaregivers)	38 (25)
Clinicians or other providers	27 (17.8)
Dementia Advocates	3 (1.9)
Not stated	2 (1.3)

Of the common mHealth app purposes, the most common was providing education on the health topic (ie, dementia), applicable among 53 (34.9%) out of the total 152 apps assessed. As illustrated in Table 1, the least common was providing professional supervision to users, applicable among 4 (2.6%) apps.

Lifestyle Elements of Focus

Of the lifestyle elements important for the prevention of dementia, the most common category was mental stimulation, addressed by 59 (38.8%) out of the total 152 apps assessed. The least common element was sleep, addressed by 3 (2%) apps. Table 1 illustrates these findings. Meanwhile, 65 (42.8%) apps did not focus on any of the lifestyle elements. In addition, only 6 (3.9%) apps addressed sensory health. Table 3 also presents the number of modifiable lifestyle behaviors addressed by the included mHealth apps.

Table 3. Number of lifestyle behaviors of focus per app (n=152).

How many lifestyle behaviors does an app focus on?	Apps, n (%)
0	65 (42.8)
1	67 (44)
2	12 (7.9)
3	4 (2.6)
4	2 (1.3)
5	1 (0.6)
6	1 (0.6)
7	0 (0)

App Features

Observed in 57 (37.5%) out of the 152 assessed apps, providing communication avenues with various parties was the most frequently observed app feature, while 45 (29.6%) apps incorporated game-based elements. The least embraced feature, found in only 2 (1.3%) apps, was the ability to link to social media.

In addition, there was a lack of mention of features designed for people with dementia such as visuals supplementing texts, voice recognition, and customized accessibility features in the app descriptions.

Currency of Apps (as of April 2023)

Of the 152 apps, 136 (89.5%) apps specified their creation date or year and the date of their last update. At the time of analysis in April 2023, only 14 (9.2%) were modified within the previous month.

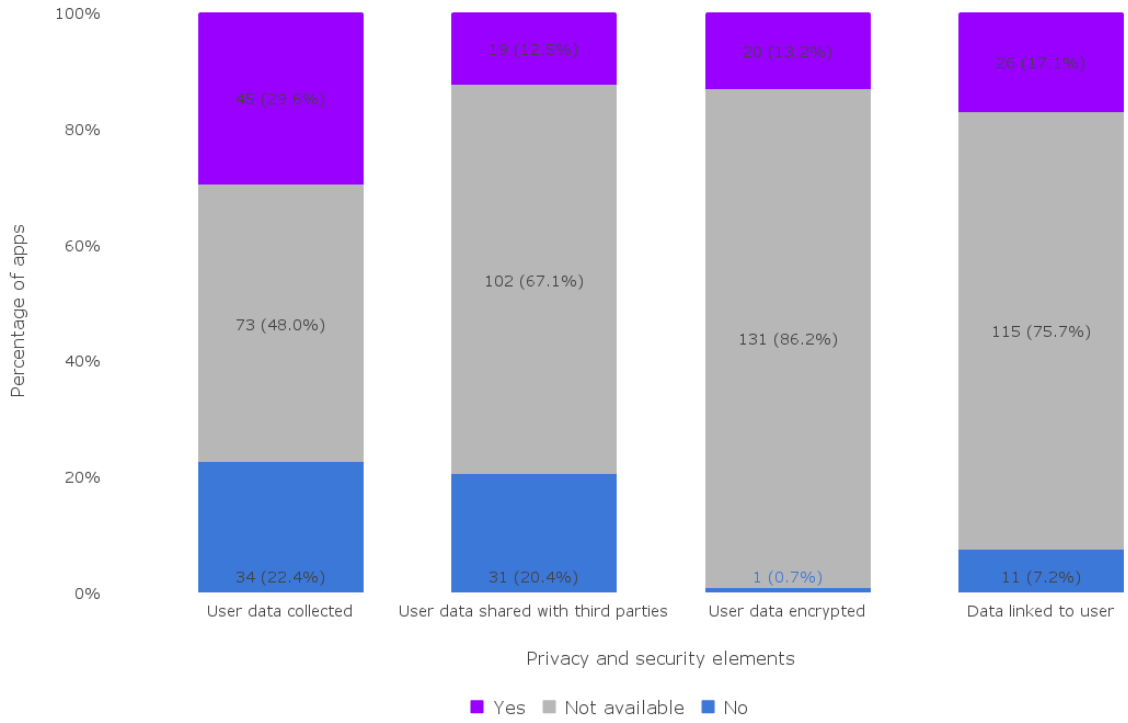
Transparency

Most apps (143/152, 94.1%) had disclosed app ownership, as seen in Table 1, and most apps (132/152, 86.8%) did not disclose any author affiliations or sponsorships or credited any authors or lack thereof.

Privacy and Security

The most disclosed facet of user privacy and security was whether user data were being collected. This was disclosed by 79 (52%) out of 152 apps, of which 45 (57%) apps collected user data. The vast majority of apps (131/152, 86.2%), however, did not disclose if any user data were encrypted at all, with an average median of 105 (69.2%) apps that did not disclose the privacy and security of their apps. Figure 2 visualizes these observations.

Figure 2. Privacy and security of the assessed apps (n=152).

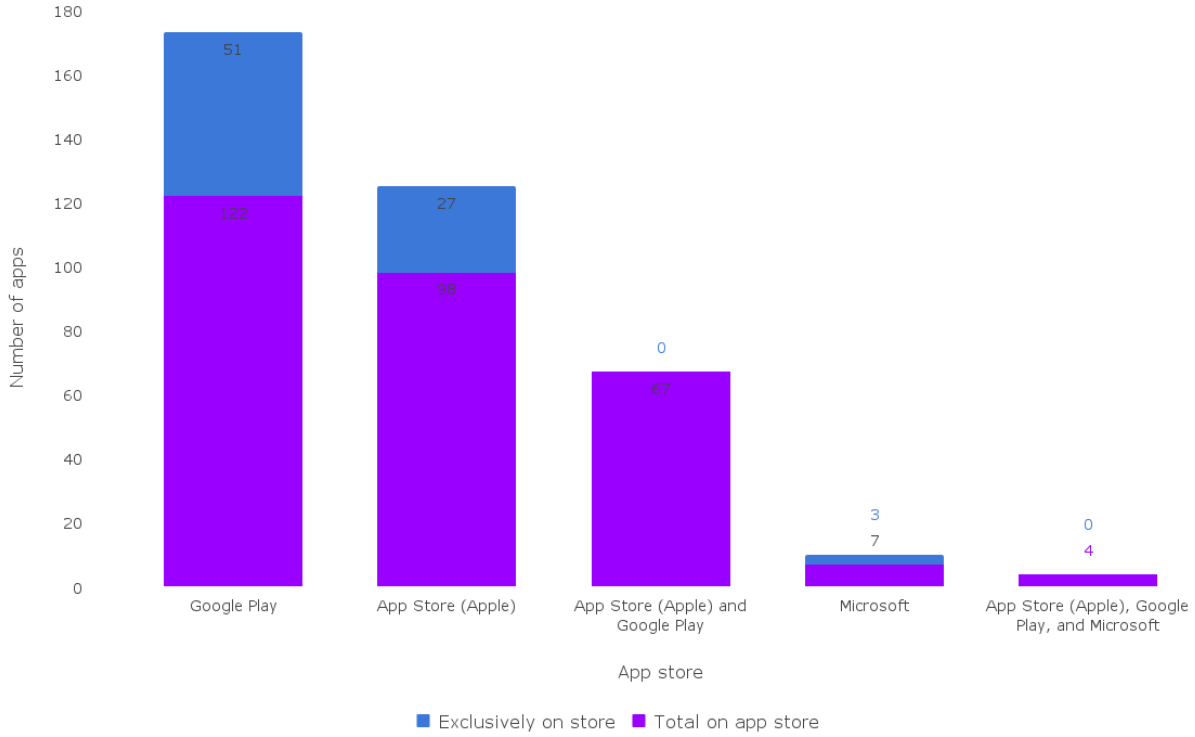


App Availability

Of the 152 apps, most apps (n=111, 73%) were free. Free with in-app purchases was the second largest category at 25 (16.4%) apps, followed by 13 (8.6%) premium or paid apps. The smallest category stood with 3 (2%) “freemium” apps, referring to those with basic or limited features available for free download but offering a paid subscription for premium features once installed.

Many apps were available on >1 app store. Figure 3 presents the number of apps available by store, which include the Google Play store, App Store, Microsoft app store, Samsung Galaxy store, and combinations of these.

Figure 3. App availability by app store (n=152).



User Reception

Due to the structure of some app stores (ie, certain elements of app information are or are not made available and provided by app stores), 120 (78.9%) out of the 152 assessed apps had no ratings available. The next highest group of app ratings was 4 to 4.9 stars, given to 14 (9.2%) apps. Only 7 (4.6%) assessed apps received a rating of 5 stars.

A total of 77 (50.7%) apps had no download information available due to the structure of some app stores (eg, the App Store did not present this element of app information at the time of charting), while 20 (13.2%) had between 100 and 999 downloads. The next highest number of downloads were those apps with <100 downloads as observed in 19 (12.5%) apps, then between 1000 and 4999 downloads observed in 18 (11.8%) apps. Only 4 (2.6%) apps had between 5000 and 9999 downloads, while 10 (6.6%) apps had between 10,000 and 49,999 downloads. Another 4 (2.6%) apps had between 50,000 and 999,999 downloads.

Silberg Scale

As the Silberg scale aims to assess the accountability of web-based health information, with each point representing 1 of 9 elements of accountability disclosed [29,30], a low score suggests that these apps possess gaps in accountability to users. More than half of the apps (84/152, 55.3%) evaluated scored 2 out of 9 Silberg scale points (mean score 2.4, SD 1.0 points), as visualized in Table 4.

Table 4. Silberg scale points accrued by apps (n=152).

Silberg scale points	Apps, n (%)
0	4 (2.6)
1	12 (7.9)
2	84 (55.2)
3	33 (21.7)
4	13 (8.6)
5	4 (2.6)
6	1 (0.7)
7	1 (0.7)
8	0 (0)
9	0 (0)

Of the 152 apps, 4 (2.6%) apps did not score any Silberg scale points (ie, received total scores of 0), and none of the included apps achieved a perfect score. The Care4Dementia app totaled to 7 points, forgoing 2 elements of app accountability: crediting the app's authors and indicating whether an update had been made within the last month (as of the screening and review stages). Meanwhile, the Memory Lane Games app received a score of 6, forgoing the same elements as the Care4Dementia app along with failing to disclose author credentials.

Discussion

Principal Findings

This study aimed to assess the scope of mHealth apps intended for use by the general public relating to dementia prevention and its risk factors and to determine any gaps pertaining to factors including accountability of information, addressing modifiable lifestyle behaviors, and privacy and security. The findings of this systematic search and environmental scan of mHealth apps have found that most lack sufficient accountability of information, particularly for products that target health and can store sensitive user information. Most apps assessed did not satisfactorily disclose their evidence base, lacked comprehensive focus on >1 modifiable lifestyle behavior related to reducing dementia risk, and did not disclose adherence to privacy and security measures or established privacy guidelines. However, apps did show promising diversity in its target audiences, as a number of apps were designed for caregivers of people with dementia (and more broadly, people with neurocognitive symptoms); friends and family; clinicians; or the general public (ie, people without neurocognitive symptoms). These gaps highlight opportunities for researchers and clinicians to establish participatory design frameworks for developing rigorous mHealth apps, as the interest in dementia prevention grows and is projected to continue to grow.

In this study, our goal was to describe the breadth and depth of dementia-related mHealth apps to understand whether these apps demonstrate factors such as credibility, privacy, and transparency in order to be leveraged as a tool for the prevention of dementia. In addition, this study sought to explore whether dementia-related apps specifically addressed any of the above-discussed lifestyle behaviors attributed to maintaining brain health. In the following sections, we will report on the state of apps intended for the public relating to the prevention of dementia by the themes identified above.

Themes

Evidence-Based and Expert Credibility

Most apps did not report author credentials, give information on whether peer-reviewed papers or other sources were used, or provide references. When creating apps that address people with dementia, there should be clear guidelines for developers to follow to ensure that a minimum standard is met in terms of how credible the information in the app is, such as including the source of evidence-based information presented [31]. The absence of these factors could impact the overall effectiveness of the app and its uptake by patients or could mislead and misinform individuals looking to use apps to improve their brain health.

Similarly, most assessed apps (114/152, 75%) did not specify the app developer association, meaning their affiliations were not clearly defined. Meanwhile, only a small subset of assessed apps (23/152, 15.1%) had been created by developers categorized as researchers or health care organizations. This reinforces the need for more credibility and reporting of evidence-based information that could benefit users. Furthermore, only a few indicated having consultations with an expert, which included people with dementia or caretakers of people with dementia, whose lived experiences could be strongly leveraged to take input from, as apps are developed to influence behaviors and habits [32]. The lack of consultation with experts [33] or people with lived experience could impact the quality of the app and the information delivered, making these apps less relevant to those who could best benefit from it. These gaps may mislead consumers in the target populations or provide them with misinformation that could further affect their condition onset or prognosis.

Purpose

Of the 152 apps assessed, 96 (62.5%) were categorized under "Health & Fitness," "Medical," and "Educational," with most apps intended for persons with neurocognitive symptoms, informal caregivers, and the general public (ie, people without neurocognitive symptoms). Meanwhile, only a small subset of these apps actually provides information pertaining to dementia; offers diagnostics tools; and provides reminder, goal-setting, progress-tracking, or habit-tracking features. The lack of diagnostic tools and ability for professionals to remotely monitor users align with the minimal evidence-based information or clinical credibility available in these apps. There appears to be a gap in what these apps that claim to target persons with dementia (and more broadly, people with neurocognitive symptoms) realistically provide to benefit this population. Education, reminder features, goal setting, and tracking would be beneficial tools to provide through apps not only for the target population but also for informal and formal caregivers as well (ie, apps could help set reminders for medications while providing education on certain habits that individuals can commit to improving their condition) [34].

Lifestyle Elements of Focus

Apps were screened for whether they focused on one of the modifiable lifestyle behaviors associated with reducing the risk of developing dementia, which include exercise, sleep, stress management, mental stimulation, eating and nutrition, and social stimulation. More than 2 in 5 apps did not address any lifestyle behavior, while 57.2% (87/152) addressed at least 1. When apps did address modifiable lifestyle behaviors, it was most common to focus on 1 behavior at a time while few addressed a combination of 2 behaviors per app, as demonstrated in Table 1. The lifestyle behavior of focus was mental stimulation; a greater focus on mental stimulation in these apps is in line with the number of apps that had contained game-based elements. This complements the observations of many studies that dementia prevention mHealth apps focused on mental stimulation in maintaining cognitive vitality in healthy

individuals and those with mildly impaired cognition, including improved memory and enhanced quality of life [35].

Research suggests that activities addressing these lifestyle behaviors could in fact improve brain health and lower the chances of developing dementia. In a study exploring physical exercise, diet, smoking, alcohol consumption, cognitive stimulation, and social stimulation, Jia et al [36] observed that positively maintaining any one of these lifestyle behaviors contributes to slowing memory decline. Mamalaki et al [37] also observed the individual impacts of diet, physical activity, sleep, and engagement in activities of daily living (including social stimulation) in maintaining cognitive vitality and reducing the risk of developing dementia. Both studies highlighted the combined effects of addressing multiple lifestyle behaviors to further slow cognitive decline, with the upkeep of more behaviors resulting in stronger prevention [36,37]. It is noteworthy to mention that both studies found diet as the modifiable lifestyle behavior with the largest associated effect on maintaining cognitive vitality and reducing the risk of decline. Thus, if apps are to target patients and their caregivers, it would be most useful to use this evidence-based information to allow for habit development and tracking in activities related to exercise, reminders to maintain sensory well-being, or prompts that could structure stress management. In particular, delivering interventions related to the prevention of dementia using the mHealth app format can expand the availability of these products to hard-to-reach populations [38].

App Features

Apps were also categorized based on whether they offered game-based elements, communication avenues, and links to social media. These features are attractive and can address feelings of loneliness for patients. While game-based apps can be a therapeutic experience, this population may have difficulty in carrying out the tasks or activities required by some of these game apps. Apps may also be a promising way to address social isolation and loneliness [39], providing a channel for in-app users to connect among themselves or presenting features to link with family, friends [40], or the general public for connection or support.

One subset of the intended users of these apps, that is, persons with dementia, requires specific design features for these apps to be beneficial and accessible. Such features include prompts (eg, redirection prompts while completing an activity), visuals supplementing the text, reminder systems, and voice recognition, along with the option to customize accessibility features because dementia affects each of those with the condition differently [41,42]. If these game-based apps are targeted at patients, they may be inaccessible for them due to impaired cognitive abilities where they cannot complete multistep cognitive tasks, remember information, or answer questions in a short time span as they may be cognitively demanding. These features are clearly lacking, making them less accessible to the target users. Thus, the gap remains that consulting the target populations during the development of these apps to identify useful elements would likely serve as the best use of time and resources to yield a positive impact on patients.

Currency of Apps (as of April 2023)

Information on currency, year of creation, and last update date were available for most of the apps. This practice of transparency around app currency is important as new research is released regularly and thus would be important for the consumer to understand when the app itself was created along with the date of the information that the app is based on.

Transparency

A minority of apps credited authors to acknowledge their work, reported author affiliation to demonstrate connection to an organization or academic center, or disclosed sponsorship on whether a specific organization provided funding for the development of the app. This results in poor transparency, as it does not provide consumers with the necessary information to identify the basis of the app or disclose whether there is a possible conflict of interest because of a sponsorship, author, or organization affiliations. Moreover, it is crucial to disclose all these pieces of information, as simply reporting sources or authors while failing to report sponsoring organizations obscures acknowledgment of possible bias embedded in the information presented and the intention behind mHealth apps [43,44]. Providing this information to users could add credibility to the app, which could help an individual decide whether or not such an app will serve their needs and whether it is trustworthy and safe to use [23,45]. Furthermore, the availability of this information is crucial to ensure patient safety and suitability on part of professionals who may recommend such mHealth apps for clinical use [23,45]. By contrast, users are less aware of and less preoccupied with considering an app's evidence base and instead emphasize the quality of its consumer-oriented, practical use such as data use, battery drainage, loading speed, and presence of advertisements when deciding whether to continue its use [43].

Privacy and Security

In regard to privacy and security of these apps, most apps did not provide information regarding the collection of user data, data sharing with third parties, user data encryption, or data linking to users. This demonstrates a gap in the security of app-user data, which may be at risk of being compromised due to weak data governance guidelines that are not disclosed clearly or at all. Close attention should be paid to ensuring the privacy and security of mHealth apps, especially when dealing with susceptible target groups such as people living with mild cognitive impairment [46]. One relevant guideline that can be followed is the *mHealth Data Security, Privacy, and Confidentiality: Guidelines for Program Implementers and Policy Makers*, which details best practices and considerations for mHealth app development. These best practices cover the data life cycle stages of data capture and storage, access, transfer, and disposal, as well as address risks associated with operating systems, mobile devices, networks, and mHealth data storage [47]. The *Privacy and Code of Conduct on Mobile Health Apps* developed by the European Commission [48] lays out principles for mHealth app development and function, including considerations for user consent, app security measures, and use of user data for secondary purposes such as research and third-party sharing among others. While these guidelines

are not official law or policy, they provide detailed considerations that will assist mHealth app developers in creating and sustaining responsible data governance.

App Availability

For apps to be useful, they should be accessible to the targeted population. Commercialization of the apps significantly impacts this, as almost three-quarters of assessed apps were offered for free, while about one-tenth involved a purchase for the basic or the full version of the app. Furthermore, 44.1% (67/152) of the apps were offered both on the App Store and Google Play app store, whereas an exclusive 33.5% (51/152) were offered only on Google Play as opposed to 17.8% (27/152) of apps that were exclusively available on the App Store. Interestingly, only a small number of the apps (7/152, 4.6%) were available on the Microsoft app store, and none were offered on the Samsung Galaxy app store. A needs-based analysis to identify which app features are most useful for patients and caregivers would be appropriately supplemented by identifying the type of smartphones that are accessible to this population.

User Reception

Measured by the number of downloads, user reception is important when it comes to measuring meaningful use of apps. While not all apps had information pertaining to their exact number of downloads, 39 (25.7%) out of the 152 assessed apps had <999 downloads, of which 19 (49%) had <100 downloads per app. This illustrates a very low use or download rate of apps that are categorized and are targeted for persons with neurocognitive symptoms.

Silberg Scale

Apps were also scored using the Silberg scale, which is an established validated measure to capture the extent of accountability of health information presented on the web, including through mHealth apps [29,30]. An app can score up to a total of 9 Silberg scale points, wherein the higher the total Silberg score, the more the accountability of the information presented by or used to create the app is communicated [49,50]. Meanwhile, a low total Silberg score implies the opposite, suggesting that the app in question does not transparently disclose the credibility of the app's information, which may therefore lack accountability toward users and the public. Within this study, the mean Silberg scale score of 2.4 (SD 1.0) across all reviewed apps points toward a trend skewed toward low accountability of information used and presented by the apps (ie, in communicating the authorship, attribution, disclosure, and currency of their information). The observed skewed trend of apps with low Silberg scale scores raises concern for a lack of accountability these apps take in disclosing the credibility of their foundations, their reliability, usefulness to targeted users, and the efficacy and effectiveness of their results, if any.

While the apps assessed in this study produced low Silberg scale scores, this does not appear to be unique in this regard, as other nondementia mHealth apps have observed similar findings. Several studies assessing mHealth apps using the Silberg scale found that most apps scored <5 points [29,51-53]. In the case of a study assessing the (then) current state of mHealth apps that addressed postnatal depression, the authors observed a mean Silberg score of 3.0 points and suggested the need for the involvement of health care professionals in developing mHealth apps [51]. In another study, an analysis of obesity management mHealth apps observed a mean score of 4.55 points along with a low percentage of apps disclosing information sources and references [29]. The assessed obesity management apps were found to present poor accountability of the information, highlighting the need to improve the current state of mHealth app information quality. The systematic review of mHealth apps by Feldman et al [53] that had addressed peripartum mood disorders reached a similar conclusion, finding <30% of the assessed apps indicated the use of research for app development. Overall, the authors found that most apps reviewed in their study were of poor quality, as indicated by low Silberg scores, and suggested their poor accountability of information as a hindrance to users in making an informed decision about their quality [53]. Overall, there is a general trend of low Silberg scale scores observed across mHealth apps, which is a gap that requires addressing by developers.

The observations noted in this section point toward the need for app developers and app stores to fill in the gaps when presenting descriptions of the products they are offering. The most significant gaps observed in targeted apps included (1) a lack of prevention-oriented apps addressing >1 modifiable lifestyle behavior (diet, sleep, exercise, stress management, mental stimulation, social stimulation, and sensory health) associated with dementia risk; (2) an absence of apps created using evidence-based information and credible authors and an absence of transparent disclosure of these sources; (3) inadequate rigor of general mHealth data privacy and security procedures and a complete lack of disclosed specific guidelines created for people with cognitive impairment; and (4) a lack of evidence-based, usable app features designed for the needs of people with declining or limited cognitive abilities.

Limitations

While efforts were made to make this study systematic and comprehensive, several limitations should be acknowledged. This study used PRISMA with the goal of adhering to rigorous standards and to thoroughly structure its data search, screening, and extraction. However, PRISMA guidelines are intended for systematic reviews and meta-analyses evaluating peer-reviewed literature, contrasting with this study's environmental scan component along with its evaluation of mobile apps. In addition, a limited selection of app stores were searched from the available pool of North American app stores, and reviewed apps had to be available in English, and thus, this study may not have captured the entirety of dementia-related apps available on the global market. While peer-reviewed journal articles, browser searches, and dementia and Alzheimer association and organization websites were initial sources for identifying available apps, the search using these sources was not exhaustive. There are also additional search terms that could have been used to widen the search results such as "memory," "brain training," and "cognit*." As a result, there is a possibility that some potential apps for identification may have been missed, and we recommend taking this into account for future research.

Due to the dynamic nature of app stores, some apps included in this study's final review have been pulled out of app stores since the time of analysis. In addition, the structural differences among the 4 app stores resulted in some details being available on 1 or some app stores (ie, number of app downloads, user data collection, sharing of user data with third parties, user data encryption, and linking data to user) and not on other app stores. Because of this, some data points were extracted as "N/A" and do not present a full picture of the variable.

Apps were not downloaded, meaning that features and functionality could not be assessed to the fullest extent. Thus, relying on app store descriptions as the primary source for information, supplemented with other sources available on the web, we could not extract the data regarding the full extent of app features, purposes, audiences, and other elements in the screening and data extraction phases. Furthermore, some descriptions encountered provided only scarce information, impacting the scope of data extraction for individual app screeners at times. However, second- and third-screener corroboration was conducted in response to this and ensured complete data extraction.

Finally, in the process of screening for mHealth apps, we encountered a challenge in accurately distinguishing between those targeted for users with mild cognitive impairment and dementia due to the inherent complexities and overlapping characteristics of these conditions. Recognizing the importance of this distinction in clinical contexts, it is crucial to acknowledge that our screening criteria erred on the side of inclusivity, resulting in a broader scope of apps included in our review. While this may initially seem to widen the spectrum of included apps, it inherently contributes to a more comprehensive and nuanced report of the current landscape of available apps. By encompassing a broader range of users experiencing neurocognitive symptoms, our study provides a holistic view of the mHealth app landscape, offering valuable insights into potential digital interventions for individuals at varying stages of cognitive decline. This approach enhances the practical relevance of our findings and acknowledges the diverse needs of users in the evolving field of digital therapeutics.

Conclusions

As the prevalence of dementia continues to grow, so does the burden of disease on patients, caregivers, the health system, and the greater community. This could be combated by leveraging mHealth apps to address and mediate the symptoms and effects of these diseases. Specifically, the literature suggests addressing a number of modifiable lifestyle behaviors as an effective way to improve brain health. In order to address these behaviors through mHealth apps, the apps should not only be accessible in terms of availability but also offer features suited for individuals living with declining or impaired cognitive vitality. Furthermore, these apps should be transparent in the foundation of information, its evidence base, and the involvement of professionals or people with lived experience in its development. Thus, providing app designers and developers with a guideline for what is required and necessary to fulfill the needs of patients and caregivers, along with minimal standards for what an app must have to be appropriate for this susceptible population, would be ideal. Such a guideline would be instrumental in leveraging mHealth apps to maintain cognitive vitality, socially connect, and provide therapeutic avenues for persons with dementia (and more broadly, those with neurocognitive symptoms) or their caregivers.

Acknowledgments

This study was supported by the funding (2122-HQ-000257) from the Public Health Agency of Canada. The authors would like to thank Thuraya Marjan for her contributions to the data analysis and peer-review stages. The authors also thank the Women's Brain Health Initiative for their input and support throughout this study. The views expressed in this study are those of the authors and do not necessarily reflect the official views of the Public Health Agency of Canada or the Government of Canada.

Authors' Contributions

LA conceived the study and designed the protocol. DC, DJH, HA, PSJ, SRR, and VAS screened the apps. DC, DJH, HA, PSJ, SA, SRR, and VAS conducted data extraction as first and second coders and resolved conflicts emerging in coded data as third reviewers. SA, HA, DJH, and SRR wrote the first draft of the manuscript. LA, SA, and HA edited and revised the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Extraction framework, organized by themes.
[DOCX File , 17 KB - mhealth_v12i1e50186_app1.docx]

Multimedia Appendix 2
Full set of extracted and coded data from included mobile health apps (n=152).
[XLSX File (Microsoft Excel File), 62 KB - mhealth_v12i1e50186_app2.xlsx]

Multimedia Appendix 3

Apps by app store categories (n=152).
[DOCX File , 9 KB - mhealth_v12i1e50186_app3.docx]

References

1. Dementia. World Health Organization. 2023 Mar 15. URL: https://www.who.int/news-room/fact-sheets/detail/dementia [accessed 2023-06-13]
2. Poon AN, Xiang Y, Zavalishina Y, Ayanian S, Aitken CF, Procter AC, et al. Systematic review estimating the burden of dementia in the WHO Southeast Asia Region using Bayesian and frequentist approaches. J Glob Health 2020 Dec;10(2):020701 [FREE Full text] [doi: 10.7189/jogh.10.020701] [Medline: 33282225]
3. Rathnayake S, Moyle W, Jones C, Calleja P. mHealth applications as an educational and supportive resource for family carers of people with dementia: an integrative review. Dementia (London) 2019;18(7-8):3091-3112. [doi: 10.1177/1471301218768903] [Medline: 29631492]
4. Sriram V, Jenkinson C, Peters M. Informal carers' experience of assistive technology use in dementia care at home: a systematic review. BMC Geriatr 2019 Jun 14;19(1):160 [FREE Full text] [doi: 10.1186/s12877-019-1169-0] [Medline: 31196003]
5. Matthew-Maich N, Harris L, Ploeg J, Markle-Reid M, Valaitis R, Ibrahim S, et al. Designing, implementing, and evaluating mobile health technologies for managing chronic conditions in older adults: a scoping review. JMIR Mhealth Uhealth 2016 Jun 09;4(2):e29 [FREE Full text] [doi: 10.2196/mhealth.5127] [Medline: 27282195]
6. Thomas C, Keselj V, Cercone N, Rockwood K, Asp E. Automatic detection and rating of dementia of Alzheimer type through lexical analysis of spontaneous speech. In: Proceedings of the IEEE International Conference Mechatronics and Automation. 2005 Presented at: ICMA 2005; July 29-August 01, 2005; Niagara Falls, ON URL: http://vlado.cs.dal.ca/~vlado/papers/icma05.pdf
7. Jaqua E, Biddy E, Moore C, Browne G. The impact of the six pillars of lifestyle medicine on brain health. Cureus 2023 Feb;15(2):e34605 [FREE Full text] [doi: 10.7759/cureus.34605] [Medline: 36883088]
8. Ahlskog JE, Geda YE, Graff-Radford NR, Petersen RC. Physical exercise as a preventive or disease-modifying treatment of dementia and brain aging. Mayo Clin Proc 2011 Sep;86(9):876-884 [FREE Full text] [doi: 10.4065/mcp.2011.0252] [Medline: 21878600]
9. Franks KH, Bransby L, Saling MM, Pase MP. Association of stress with risk of dementia and mild cognitive impairment: a systematic review and meta-analysis. J Alzheimers Dis 2021;82(4):1573-1590. [doi: 10.3233/JAD-210094] [Medline: 34366334]
10. Yates LA, Ziser S, Spector A, Orrell M. Cognitive leisure activities and future risk of cognitive impairment and dementia: systematic review and meta-analysis. Int Psychogeriatr 2016 Nov;28(11):1791-1806. [doi: 10.1017/S1041610216001137] [Medline: 27502691]
11. Shi L, Chen SJ, Ma MY, Bao YP, Han Y, Wang YM, et al. Sleep disturbances increase the risk of dementia: a systematic review and meta-analysis. Sleep Med Rev 2018 Aug;40:4-16. [doi: 10.1016/j.smrv.2017.06.010] [Medline: 28890168]
12. van de Rest O, Berendsen AA, Haveman-Nies A, de Groot LC. Dietary patterns, cognitive decline, and dementia: a systematic review. Adv Nutr 2015 Mar;6(2):154-168 [FREE Full text] [doi: 10.3945/an.114.007617] [Medline: 25770254]
13. Kuiper JS, Zuidersma M, Oude Voshaar RC, Zuidema SU, van den Heuvel ER, Stolk RP, et al. Social relationships and risk of dementia: a systematic review and meta-analysis of longitudinal cohort studies. Ageing Res Rev 2015 Jul;22:39-57. [doi: 10.1016/j.arr.2015.04.006] [Medline: 25956016]
14. Brenowitz WD, Kaup AR, Yaffe K. Incident dementia and faster rates of cognitive decline are associated with worse multisensory function summary scores. Alzheimers Dement 2020 Oct;16(10):1384-1392 [FREE Full text] [doi: 10.1002/alz.12134] [Medline: 32657033]
15. Hwang PH, Longstreth WTJ, Brenowitz WD, Thielke SM, Lopez OL, Francis CE, et al. Dual sensory impairment in older adults and risk of dementia from the GEM Study. Alzheimers Dement (Amst) 2020;12(1):e12054 [FREE Full text] [doi: 10.1002/dad2.12054] [Medline: 32671180]
16. Fischer ME, Cruickshanks KJ, Schubert CR, Pinto AA, Carlsson CM, Klein BE, et al. Age-related sensory impairments and risk of cognitive impairment. J Am Geriatr Soc 2016 Oct;64(10):1981-1987 [FREE Full text] [doi: 10.1111/jgs.14308] [Medline: 27611845]
17. Schubert CR, Cruickshanks KJ, Fischer ME, Chen Y, Klein BE, Klein R, et al. Sensory impairments and cognitive function in middle-aged adults. J Gerontol A Biol Sci Med Sci 2017 Aug 01;72(8):1087-1090 [FREE Full text] [doi: 10.1093/gerona/glx067] [Medline: 28535277]
18. Villamil-Cabello E, Meneses-Domínguez M, Fernández-Rodríguez Á, Ontoria-Álvarez P, Jiménez-Gutiérrez A, Fernández-Del-Olmo M. A pilot study of the effects of individualized home dual task training by mobile health technology in people with dementia. Int J Environ Res Public Health 2023 Apr 11;20(8):5464 [FREE Full text] [doi: 10.3390/ijerph20085464] [Medline: 37107746]
19. Fletcher O. 'Friendly' and 'noisy surveillance' through MapMyRun during the COVID-19 pandemic. Geoforum 2022 Jul;133:11-19 [FREE Full text] [doi: 10.1016/j.geoforum.2022.05.004] [Medline: 35601807]

20. Laing BY, Mangione CM, Tseng CH, Leng M, Vaisberg E, Mahida M, et al. Effectiveness of a smartphone application for weight loss compared with usual care in overweight primary care patients: a randomized, controlled trial. Ann Intern Med 2014 Nov 18;161(10 Suppl):S5-12 [FREE Full text] [doi: 10.7326/M13-3005] [Medline: 25402403]
21. Kao C, Liebovitz DM. Consumer mobile health apps: current state, barriers, and future directions. PM R 2017 May;9(5S):S106-S115. [doi: 10.1016/j.pmrj.2017.02.018] [Medline: 28527495]
22. Rosa I, Lages M, Grilo C, Barros R, Guarino MP. mHealth applications to monitor lifestyle behaviors and circadian rhythm in clinical settings: current perspective and future directions. Front Public Health 2022;10:862065 [FREE Full text] [doi: 10.3389/fpubh.2022.862065] [Medline: 35923965]
23. About the Digital Health Assessment Framework (DHAF). American College of Physician. URL: https://www.acponline.org/practice-resources/business-resources/telehealth-guidance-and-resources/pilot-test-orcha-acp-library-of-weight-management-apps/about-the-digital-health-assessment-framework-dhaf [accessed 2024-01-22]
24. Ye B, How TV, Chu CH, Mihailidis A. Dementia care apps for people with dementia and informal caregivers: a systematic review protocol. Gerontology 2021;67(5):633-638 [FREE Full text] [doi: 10.1159/000514838] [Medline: 33774646]
25. Rosenfeld L, Torous J, Vahia IV. Data security and privacy in apps for dementia: an analysis of existing privacy policies. Am J Geriatr Psychiatry 2017 Aug;25(8):873-877. [doi: 10.1016/j.jagp.2017.04.009] [Medline: 28645535]
26. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
27. Charlton P, Kean T, Liu RH, Nagel DA, Azar R, Doucet S, et al. Use of environmental scans in health services delivery research: a scoping review. BMJ Open 2021 Nov 10;11(11):e050284 [FREE Full text] [doi: 10.1136/bmjopen-2021-050284] [Medline: 34758992]
28. Giunti G, Guisado Fernández E, Dorronzoro Zubiete E, Rivera Romero O. Supply and demand in mHealth apps for persons with multiple sclerosis: systematic search in app stores and scoping literature review. JMIR Mhealth Uhealth 2018 May 23;6(5):e10512 [FREE Full text] [doi: 10.2196/10512] [Medline: 29792295]
29. Jeon E, Park HA, Min YH, Kim HY. Analysis of the information quality of korean obesity-management smartphone applications. Healthc Inform Res 2014 Jan;20(1):23-29 [FREE Full text] [doi: 10.4258/hir.2014.20.1.23] [Medline: 24627815]
30. Silberg WM. Assessing, controlling, and assuring the quality of medical information on the internet. JAMA 1997 Apr 16;277(15):1244-1245. [doi: 10.1001/jama.1997.03540390074039]
31. Choi SK, Yelton B, Ezeanya VK, Kannaley K, Friedman DB. Review of the content and quality of mobile applications about Alzheimer's disease and related dementias. J Appl Gerontol 2020 Jun;39(6):601-608 [FREE Full text] [doi: 10.1177/0733464818790187] [Medline: 30049239]
32. Kuo CY, Stachiv I, Nikolai T. Association of late life depression, (non-) modifiable risk and protective factors with dementia and Alzheimer's disease: literature review on current evidences, preventive interventions and possible future trends in prevention and treatment of dementia. Int J Environ Res Public Health 2020 Oct 14;17(20):7475 [FREE Full text] [doi: 10.3390/ijerph17207475] [Medline: 33066592]
33. Nouri R, R Niakan Kalhori S, Ghazisaeedi M, Marchand G, Yasini M. Criteria for assessing the quality of mHealth apps: a systematic review. J Am Med Inform Assoc 2018 Aug 01;25(8):1089-1098 [FREE Full text] [doi: 10.1093/jamia/ocy050] [Medline: 29788283]
34. Pérez-Jover V, Sala-González M, Guilabert M, Mira JJ. Mobile apps for increasing treatment adherence: systematic review. J Med Internet Res 2019 Jun 18;21(6):e12505 [FREE Full text] [doi: 10.2196/12505] [Medline: 31215517]
35. Lee J, Lim JM. Factors associated with the experience of cognitive training apps for the prevention of dementia: cross-sectional study using an extended health belief model. J Med Internet Res 2022 Jan 14;24(1):e31664 [FREE Full text] [doi: 10.2196/31664] [Medline: 35029540]
36. Jia J, Zhao T, Liu Z, Liang Y, Li F, Li Y, et al. Association between healthy lifestyle and memory decline in older adults: 10 year, population based, prospective cohort study. BMJ 2023 Jan 25;380:e072691 [FREE Full text] [doi: 10.1136/bmj-2022-072691] [Medline: 36696990]
37. Mamalaki E, Charisis S, Anastasiou CA, Ntanasi E, Georgiadi K, Balomenos V, et al. The longitudinal association of lifestyle with cognitive health and dementia risk: findings from the HELIAD study. Nutrients 2022 Jul 08;14(14):2818 [FREE Full text] [doi: 10.3390/nu14142818] [Medline: 35889774]
38. Eggink E, Hafdi M, Hoevenaar-Blom MP, Song M, Andrieu S, Barnes LE, et al. Prevention of dementia using mobile phone applications (PRODEMOS): protocol for an international randomised controlled trial. BMJ Open 2021 Jun 09;11(6):e049762 [FREE Full text] [doi: 10.1136/bmjopen-2021-049762] [Medline: 34108173]
39. Sharma M, Batra K, Flatt J. Testing the multi-theory model (MTM) to predict the use of new technology for social connectedness in the COVID-19 pandemic. Healthcare (Basel) 2021 Jul 01;9(7):838 [FREE Full text] [doi: 10.3390/healthcare9070838] [Medline: 34356216]
40. Brown A, O'Connor S. Mobile health applications for people with dementia: a systematic review and synthesis of qualitative studies. Inform Health Soc Care 2020 Oct 01;45(4):343-359. [doi: 10.1080/17538157.2020.1728536] [Medline: 32237937]

41. Joddrell P, Astell AJ. Implementing accessibility settings in touchscreen apps for people living with dementia. Gerontology 2019;65(5):560-570. [doi: 10.1159/000498885] [Medline: 31018201]
42. Øksnebjerg L, Woods B, Waldemar G. Designing the ReACT App to support self-management of people with dementia: an iterative user-involving process. Gerontology 2019;65(6):673-685 [FREE Full text] [doi: 10.1159/000500445] [Medline: 31266031]
43. Grundy Q. A review of the quality and impact of mobile health apps. Annu Rev Public Health 2022 Apr 05;43:117-134 [FREE Full text] [doi: 10.1146/annurev-publhealth-052020-103738] [Medline: 34910582]
44. Wykes T, Schueller S. Why reviewing apps is not enough: transparency for trust (T4T) principles of responsible health app marketplaces. J Med Internet Res 2019 May 02;21(5):e12390 [FREE Full text] [doi: 10.2196/12390] [Medline: 31045497]
45. Albrecht UV. Transparency of health-apps for trust and decision making. J Med Internet Res 2013 Dec 30;15(12):e277 [FREE Full text] [doi: 10.2196/jmir.2981] [Medline: 24449711]
46. Muchagata J, Teles S, Vieira-Marques P, Abrantes D, Ferreira A. Dementia and mHealth: on the way to GDPR compliance. In: 12th International Conference on Biomedical Engineering Systems and Technologies. 2019 Presented at: BIOSTEC 2019; February 22-24, 2019; Prague, Czech Republic. [doi: 10.1007/978-3-030-46970-2_19]
47. Spigel L, Wambugu S, Villella C. mHealth data security, privacy, and confidentiality: guidelines for program implementers and policymakers. MEASURE Evaluation. 2018. URL: https://www.measureevaluation.org/resources/publications/ms-17-125a.html [accessed 2023-07-13]
48. Privacy code of conduct on mobile health apps. European Commission. URL: https://digital-strategy.ec.europa.eu/en/policies/privacy-mobile-health-apps [accessed 2023-07-13]
49. Thomas C, Simmons E, Musbahi A, Small P, Courtney M. A contemporary review of smart phone applications in bariatric and metabolic surgery: an underdeveloped support service. Obes Surg 2023 Jun;33(6):1866-1875. [doi: 10.1007/s11695-023-06566-7] [Medline: 37067684]
50. Zhang H, Wang T, Zhang Z, Lin B, Mei Y, Zhang Y, et al. The current status of stroke-related smartphone applications available to adopt in China: a systematic review study. Medicine (Baltimore) 2020 Jul 02;99(27):e20656 [FREE Full text] [doi: 10.1097/MD.0000000000020656] [Medline: 32629637]
51. Zhang MW, Ho RC, Loh A, Wing T, Wynne O, Chan SW, et al. Current status of postnatal depression smartphone applications available on application stores: an information quality analysis. BMJ Open 2017 Nov 14;7(11):e015655 [FREE Full text] [doi: 10.1136/bmjopen-2016-015655] [Medline: 29138195]
52. Xu ZF, Luo X, Shi J, Lai Y. Quality analysis of smart phone sleep apps in China: can apps be used to conveniently screen for obstructive sleep apnea at home? BMC Med Inform Decis Mak 2019 Nov 15;19(1):224 [FREE Full text] [doi: 10.1186/s12911-019-0916-7] [Medline: 31730487]
53. Feldman N, Back D, Boland R, Torous J. A systematic review of mHealth application interventions for peripartum mood disorders: trends and evidence in academia and industry. Arch Womens Ment Health 2021 Dec;24(6):881-892 [FREE Full text] [doi: 10.1007/s00737-021-01138-z] [Medline: 33929636]

Abbreviations

mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by L Buis; submitted 25.08.23; peer-reviewed by S Amadoru, N Martin, S Hahn-Goldberg; comments to author 11.12.23; revised version received 30.01.24; accepted 23.04.24; published 03.07.24.

Please cite as:
Ali S, Alizai H, Hagos DJ, Rubio SR, Calabia D, Serrano Jimenez P, Senthil VA, Appel L
mHealth Apps for Dementia, Alzheimer Disease, and Other Neurocognitive Disorders: Systematic Search and Environmental Scan
JMIR Mhealth Uhealth 2024;12:e50186
URL: https://mhealth.jmir.org/2024/1/e50186
doi:10.2196/50186
PMID:

Original Paper

Effect of a Smartphone App (S-Check) on Actual and Intended Help-Seeking and Motivation to Change Methamphetamine Use Among Adult Consumers of Methamphetamine in Australia: Randomized Waitlist-Controlled Trial

Krista J Siefried[1,2,3], RN, BSc, PhD; Florence Bascombe[1,4,5,6], BSc(Hons), RN, MSc; Brendan Clifford[1,2,3,7], BSc, BN, MPH, PhD; Zhixin Liu[8], PhD; Peter Middleton[2,7], MA; Frances Kay-Lambkin[9], BSc(Hons), PhD; Jack Freestone[10,11], MPH; Daniel Herman[12], BPsychSci(Hons), MPsych; Michael Millard[13], MBBS; Maureen Steele[2], BA; Liam Acheson[1,2,3], BMus(Hons), PhD; Carl Moller[1], BCom, BA, BSc(Hons), MPH, MPsych (Clin Neuro), PhD; Nicky Bath[14], MSocSci; Nadine Ezard[1,2,3,7], BA, MPH, MBBS, PhD

[1]The National Centre for Clinical Research on Emerging Drugs, University of New South Wales, Randwick, Australia
[2]St Vincent's Hospital Alcohol and Drug Service, Sydney, Australia
[3]The National Drug and Alcohol Research Centre, The University of New South Wales, Sydney, Australia
[4]Institute for Global Health, University College London, London, United Kingdom
[5]Central and North-West London NHS Foundation Trust, London, United Kingdom
[6]University College London Hospitals NHS Foundation Trust, London, United Kingdom
[7]New South Wales Drug and Alcohol Clinical Research and Improvement Network, New South Wales Ministry of Health, Sydney, Australia
[8]Healthdirect Australia, Sydney, Australia
[9]School of Medicine and Public Health, College of Health, Medicine and Wellbeing, University of Newcastle, Newcastle, Australia
[10]ACON, Sydney, Australia
[11]The Kirby Institute, University of New South Wales, Sydney, Australia
[12]The Practice Healthcare, Sydney, Australia
[13]Clinical Research Unit for Anxiety and Depression, St Vincent's Hospital, Sydney, Australia
[14]LGBTIQ+ Health Australia, Sydney, Australia

Corresponding Author:
Krista J Siefried, RN, BSc, PhD
The National Centre for Clinical Research on Emerging Drugs
University of New South Wales
UNSW Randwick Campus
22/32 King Street
Randwick, 2031
Australia
Phone: 61 2 9065 7808
Email: krista.siefried@svha.org.au

Abstract

Background: Interventions are required that address delays in treatment-seeking and low treatment coverage among people consuming methamphetamine.

Objective: We aim to determine whether a self-administered smartphone-based intervention, the "S-Check app" can increase help-seeking and motivation to change methamphetamine use, and determine factors associated with app engagement.

Methods: This study is a randomized, 28-day waitlist-controlled trial. Consenting adults residing in Australia who reported using methamphetamine at least once in the last month were eligible to download the app for free from Android or iOS app stores. Those randomized to the intervention group had immediate access to the S-Check app, the control group was wait-listed for 28 days before gaining access, and then all had access until day 56. Actual help-seeking and intention to seek help were assessed by the modified Actual Help Seeking Questionnaire (mAHSQ), modified General Help Seeking Questionnaire, and motivation to

change methamphetamine use by the modified readiness ruler. χ^2 comparisons of the proportion of positive responses to the mAHSQ, modified General Help Seeking Questionnaire, and modified readiness ruler were conducted between the 2 groups. Logistic regression models compared the odds of actual help-seeking, intention to seek help, and motivation to change at day 28 between the 2 groups. Secondary outcomes were the most commonly accessed features of the app, methamphetamine use, feasibility and acceptability of the app, and associations between S-Check app engagement and participant demographic and methamphetamine use characteristics.

Results: In total, 560 participants downloaded the app; 259 (46.3%) completed eConsent and baseline; and 84 (32.4%) provided data on day 28. Participants in the immediate access group were more likely to seek professional help (mAHSQ) at day 28 than those in the control group (n=15, 45.5% vs n=12, 23.5%; χ^2_1=4.42, *P*=.04). There was no significant difference in the odds of actual help-seeking, intention to seek help, or motivation to change methamphetamine use between the 2 groups on the primary logistic regression analyses, while in the ancillary analyses, the imputed data set showed a significant difference in the odds of seeking professional help between participants in the immediate access group compared to the waitlist control group (adjusted odds ratio 2.64, 95% CI 1.19-5.83, *P*=.02). For participants not seeking help at baseline, each minute in the app increased the likelihood of seeking professional help by day 28 by 8% (ratio 1.08, 95% CI 1.02-1.22, *P*=.04). Among the intervention group, a 10-minute increase in app engagement time was associated with a decrease in days of methamphetamine use by 0.4 days (regression coefficient [β] –0.04, *P*=.02).

Conclusions: The S-Check app is a feasible low-resource self-administered intervention for adults in Australia who consume methamphetamine. Study attrition was high and, while common in mobile health interventions, warrants larger studies of the S-Check app.

Trial Registration: Australian New Zealand Clinical Trials Registry ACTRN12619000534189; https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=377288&isReview=true

(JMIR Mhealth Uhealth 2024;12:e55663) doi:10.2196/55663

KEYWORDS

methamphetamine; smartphone app; behavior change; help-seeking; motivation to change; mHealth; mobile health; app; apps; application; applications; smartphone; smartphones; motivation; motivational; RCT; randomized; controlled trial; controlled trials; drug; drugs; substance use; engagement; substance abuse; mobile phone

Introduction

Frequent use of methamphetamines is associated with adverse physical and mental health consequences [1,2], and Australia reports one of the highest rates of methamphetamine use and dependence in the world [2]. Since 2010, associated harms have been steadily increasing [3]. Psychosocial therapies are the primary mode of treatment for methamphetamine use disorder [4], especially as no pharmacotherapy has yet been shown to be effective [5]. Interventions such as cognitive behavioral therapy show modest effects [6], while contingency management has been demonstrated to be effective for outcomes associated both with methamphetamine use and with harm reduction for the duration of the intervention, but lacks widespread uptake [7]. Access to treatment is associated with more positive outcomes [8], but despite this, a gap of up to 10 years has been shown between the onset of problems related to methamphetamine use and presentation to health services for assistance [9].

Delays in access to health services can be due to a perception of "nonproblematic" methamphetamine use among people with methamphetamine use disorder [10,11]. Mental, physical, and social consequences can result from untreated persistent methamphetamine use disorder [9]. The development of a low-threshold intervention to encourage awareness of problems related to methamphetamine use could potentially change behavior and reduce the gaps between use, problems related to use, and treatment access. Health interventions supported by mobile devices (mHealth) are a promising mode of delivering early interventions to people who use drugs. As demonstrated in other mental health and substance use disorders, mHealth offers advantages in privacy and access over traditional face-to-face services and may facilitate help-seeking [12-16]. There are no mHealth interventions such as smartphone apps that have been shown to be effective in encouraging behavioral changes in people who use methamphetamine [17].

The S-Check intervention is targeted toward people who use methamphetamine, who do not necessarily identify problems related to its use [18], and are curious about the relationship between their use and other aspects of their health and well-being. Developed in 2011, the S-Check intervention consists of 4 sessions: a psychosocial assessment with a counsellor (based on Smout et al [19]); a physical health assessment with a medical officer; physical health feedback from a medical officer; and a biopsychosocial feedback, goal setting, and care planning session with a counsellor. The S-Check intervention aims to reduce harms associated with methamphetamine use, support behavior change, and encourage treatment seeking where indicated. When evaluated in 2018 [11], a majority of participants reported that the S-Check intervention was their first contact with general health services for issues related to methamphetamine use. Participants identified harm reduction, including education and advice about safer consumption practices, and the accessible and approachable environment to track their methamphetamine use patterns, as benefits of the model [11].

The S-Check intervention was translated into a smartphone app ("the S-Check app") that uses the eHealth Behavior Management Model [20] to encourage help-seeking for methamphetamine use. The eHealth Behavior Management Model incorporates the transtheoretical model [20] and the behavioral intent aspect of the theory of planned behavior [21] to create systems that tailor health communication to meet the needs of individual app users through a set of predeveloped messages based on readiness to change. The S-Check app allows app users to assess the risks and harms associated with their methamphetamine use in 5 domains (sexual, lifestyle, physical, psychological, and cognitive health), track their use of methamphetamine and other health measures with an Ecological Momentary Intervention [22], and receive tailored feedback based on persuasive communication principles [20,23,24]. The S-Check app incorporates principles of efficacious behavioral change models [25,26] to reinforce engagement and positive change by encouraging S-Check app users to attend to messages; requesting further inputs from app users; and motivating app users toward change. The S-Check app was piloted with 10 consumers and found to be acceptable and usable by participants [27].

Following the development of the S-Check intervention and the translation to a usable smartphone app, we aimed to assess the effectiveness of the S-Check app to motivate help-seeking among people who use methamphetamine and its effect on motivation to change methamphetamine use. To do this, we undertook a randomized waitlist-controlled trial of the S-Check app.

Methods

Trial Design

This study was a digital randomized 28-day waitlist-controlled trial. Participants were allocated 1:1 to one of two study arms: immediate access to the S-Check app (intervention group) or access after a 28-day waitlist period (waitlist control group). The primary end point was day 28, and the primary outcome was between-group differences in help-seeking. Both groups had the option to use the S-Check app through to day 56. Optional qualitative interviews were conducted with consenting participants and will be reported elsewhere. This study is reported in line with the CONSORT (Consolidated Standards of Reporting Trials) guidelines [28], a checklist is available in Table S1 in Multimedia Appendix 1.

Ethical Considerations

This study was approved by the Human Research Ethics Committee of St Vincent's Hospital, Sydney, Australia (HREC/18/SVH/196).

Participants

Inclusion criteria were adults (aged ≥18 years) living in Australia, who had self-reported methamphetamine use at least once in the month before study enrollment and private access to a smartphone (Android or Apple) to day 56 of follow-up. There were no exclusion criteria. Participants were recruited anonymously. This study was promoted through posters and leaflets at community (eg, people who use drugs) and partnering (eg, needle and syringe programs [NSP], medically supervised injecting center, sexual health clinics) organizations, alcohol and other drug service websites, and social media (including location-based dating apps).

Intervention

The S-Check app was available to download free of charge from Android and iOS app stores. Eligibility screening and provision of the participant information sheet and consent form were completed electronically within the app as an anonymous automated consent procedure [29].

Once accessed, the S-Check app home screen presented a dashboard (Figure 1) to access the S-Check app resources. These included 6 structured self-assessment tools (methamphetamine use, sexual health, social health, psychological well-being, physical health, and cognitive health; Table 1), provision of automated feedback (Figure 2), generation of a referral letter when appropriate to services or further support, relevant up-to-date information and resources about methamphetamine use, and a tool to track methamphetamine use and health impacts of methamphetamine over time (Figure 3).

Data were stored on a secure, encrypted password-protected Customer Relationships Management database integrated into the S-Check app's unique Application Programming Interface.

Figure 1. S-Check app dashboard.

Table 1. S-Check app self-assessment layers and questionnaires.

Scale	S-Check app layer	Outcome measured	Questions, n
Self-assessment module: methamphetamine use			
Based on ASSIST[a]-Lite List [30]	My use	Frequency of substance use (methamphetamine)	8
SDS[b] [31]	How I feel about my use	Methamphetamine dependence	5
Self-assessment module: psychological well-being			
Derived from The Mental State Examination–Mood [32]	My mood rating	Predominant emotion over the past 2 weeks	1
World Health Organization 5 Well-Being Index Scale [33]	My general well-being	Mental well-being	5
Kessler Psychological Distress Scale (K10) [34]	Am I feeling ok?	Mental distress	10
Patient Health Questionnaire [35]	My mental health	Depression	9
Generalized Anxiety Disorder 7 [36]	Stress and anxiety	Screen for or severity of anxiety	7
Self-assessment module: social health and lifestyle			
Derived from the World Health Organization Quality of Life Scale [37]	Overall satisfaction	Quality of life	3
Derived from S-Check Biopsychosocial assessment [18]	Day-to-day activities	Impact of use on social factors	10
Self-assessment module: physical health			
Derived from S-Check biopsychosocial and physical health assessment [18]	How you take your [methamphetamine], your medical history, your family's health, and your body	Effect of drug use on physical health	17
Self-assessment module: sexual health			
Derived from S-Check biopsychosocial assessment [18]	ChemSex, sexual behavior, and sex and drugs	Effect of drug use on sexual health	6
Self-assessment module: cognitive health			
Quality of life in neurological disorders—cognitive function questionnaire [38]	Cognitive health	Cognitive function	8
A 16-item version of the Prodromal Questionnaire (psychosis screener) [39]	Reality check	Risk of psychosis	16

[a]ASSIST: Alcohol, Smoking and Substance Involvement Screening Test.

[b]SDS: Severity of Dependence Scale.

Figure 2. S-Check app generated tailored feedback.

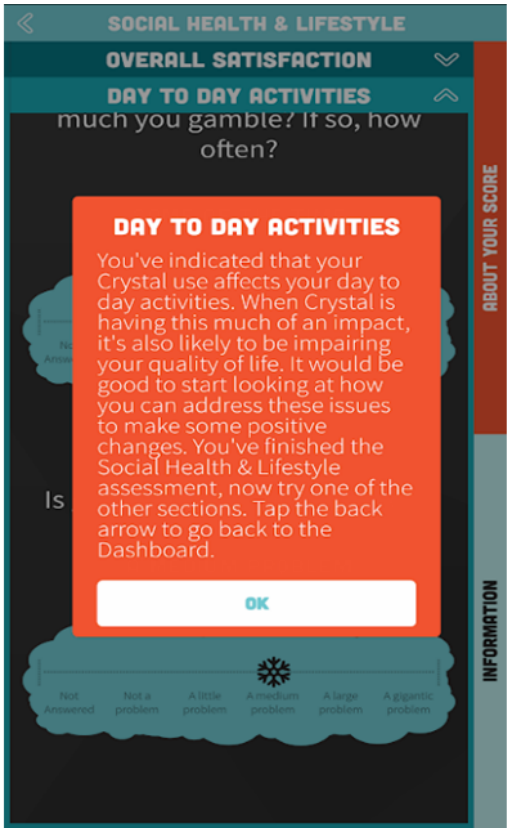
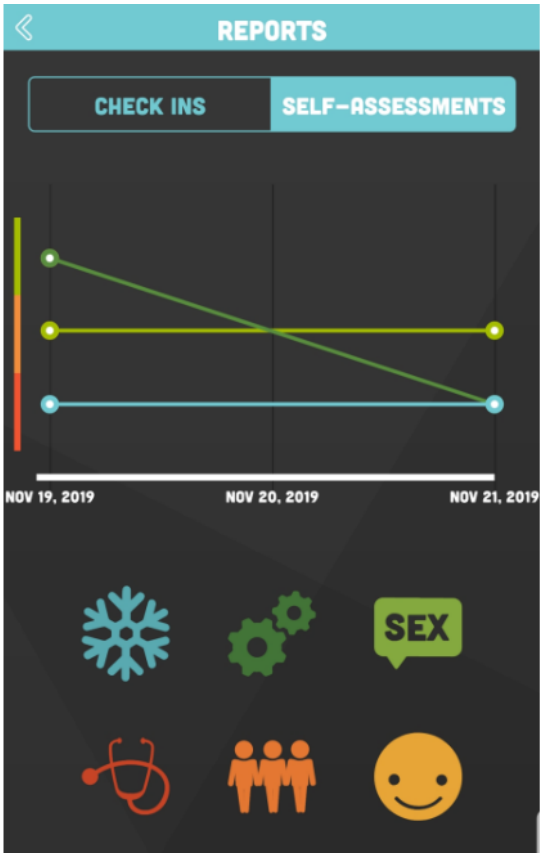


Figure 3. Tracking health over time.



Outcomes and Measures

Help-seeking was measured by (1) the modified Actual Help Seeking Questionnaire (mAHSQ; recent help-seeking) [40], (2) the modified General Help Seeking Questionnaire (mGHSQ; intention to seek help) [40], and (3) the modified readiness ruler (mRR; motivation to change behavior). AHSQ and GHSQ are validated to measure intention for and actual help-seeking in mHealth interventions addressing substance use [41] and mental health [42-45]. Both problem-type and the help-source for AHSQ and GHSQ were modified for methamphetamine use [40] over 4 weeks, and for ease of use in the app, response options were collapsed into "professional" and "nonprofessional" help sources. The mAHSQ has 2 binary measures (yes or no) for help-seeking for a problem related to the use of methamphetamine in the past 4 weeks from (1) a professional source (eg, doctor, counsellor, psychologist, specialist, telephone or web-based counseling, etc) and (2) a nonprofessional source (eg, partner, friends, parents, other family members, colleagues, internet) [41,46]. The mGHSQ scale captures intent to seek professional and nonprofessional help in the next 4 weeks, rating each source of help on a 7-point Likert-type scale ranging from 1 (extremely unlikely) to 7 (extremely likely). The mRR is a measure that has demonstrated higher predictive validity than longer measures for readiness to change among people with substance use disorder [47-49]. Participants were asked: "how ready are you to change your use of [nickname of methamphetamine]?" with responses rated and scored on a 10-point Likert-type scale ranging from 1 (not ready) to 10 (trying hard to change). Primary outcomes were measured at baseline (day 0), day 28 (primary end point), and repeated at day 56 for both intervention and control groups.

Secondary outcomes were the most commonly accessed features of the S-Check app (number of times participants accessed each section); number of days of methamphetamine use in the past 28 days (collected at baseline, day 28, and day 56); feasibility and acceptability assessed by the number of participants retained to days 28 and 56 and a qualitative substudy (to be reported elsewhere). Associations were examined between S-Check app engagement over this study's period (total minutes in-app use) and participant characteristics collected within the S-Check app: age, gender, sexual orientation, education, age of first use of methamphetamine, injection of methamphetamine in the last 4 weeks, treatment history, days of methamphetamine use in the past 28 days, and recent treatment.

For completion of the day 28 and day 56 surveys, participants in both the waitlist and control group were reimbursed with an AUS $25 (approximately US $19 at the time of enrollment) digital grocery voucher, redeemable through the S-Check app.

Sample Size

We determined the sample size required to evaluate the outcome of actual help-seeking at a power of 0.8 to detect a small to medium effect size (d=0.25), and α set at .05. A minimum of 510 participants was required [50] as we anticipated 50% attrition, and 255 participants to remain enrolled to the primary end point (day 28).

Randomization and Blinding

Randomization was automated within the S-Check app using a JavaScript (Refsnes Data) function whereby a random-access token was generated and participants were allocated to either immediate access to the S-Check app or a waitlist control group. Since participants were automatically enrolled and consented within the app, there was no risk of unblinding to the investigators.

Statistical Methods

We compared baseline characteristics between the 2 groups using χ^2 for categorical variables, and the t test (2-tailed) or Mann-Whitney U test for continuous variables.

χ^2 comparisons for the proportion of positive responses to the mAHSQ, mGHSQ (dichotomized to <5 being no intention to seek help and ≥5 being intention to seek help), and mRR (dichotomized to >8 indicating motivation to change) were conducted between the 2 groups at baseline and day 28.

For the mAHSQ, a logistic regression model compared the odds of recent help-seeking at day 28 between the intervention and the control groups, adjusted with the recent help-seeking status at baseline included as a covariate in the model. For the mGHSQ a logistic regression compared the intention of help-seeking score (derived from the mGHSQ dichotomized to <5 being no intention to seek help and ≥5 being intention to seek help) at day 28 between the intervention and control groups, adjusted with the intention to seek help status at baseline included as a covariate in the model. For the mRR, a logistic regression compared the motivation to change behavior (derived from the mRR with a score >8 indicating motivation to change) at day 28 between the 2 groups, adjusted for the corresponding baseline score, by including it as a covariate in the model.

Descriptive statistics were used to report participant access to layers within the app, methamphetamine use, and study retention at days 28 and 56. To analyze the association between demographics and methamphetamine use characteristics and S-Check app use, total engagement time for the intervention group was aggregated from the range of S-Check app use activities (from time of download to day 28). Multiple linear regression was applied to examine which baseline characteristics were predictive of more time engaged with the S-Check app, where engagement time as a dependent variable was log-transformed. The effect of time engaged with the S-Check app on help-seeking (mAHSQ and mGHSQ) and motivation to change (mRR) at day 28 were also examined. Dashboard activity time greater than 60 minutes was treated as outliers, as they were unlikely to accurately reflect true app engagement duration and be imputed with the sample median.

To handle missing data at follow-up, the analysis was conducted initially and primarily based on the retained cohort of observed data, following the recommendation of Jakobson et al [51]. Ancillary analysis based on the full intention-to-treat (ITT) cohort using multiple imputations by chained equations was performed under the missing at-random assumption. Variables in the multiple imputation model included the response to the primary outcome measure at baseline, as well as demographic

and methamphetamine use variables in the collected data. The statistical significance level was set at 2-sided .05. Analyses were performed using SAS (SAS Institute) and R (R Foundation for Statistical Computing).

Results

Participants

In total, 560 participants downloaded the app, of which 259 (46.3%) completed the eConsent process and baseline survey. Of these, 84 (32.4%) provided data on day 28. Further, 175 did not provide data at day 28, resulting in a 68% attrition from baseline. A participant flow diagram is presented in Figure 4.

Baseline participant features, summarized by intervention group or control group, are presented in Table 2. There was no significant difference observed between the 2 groups. The majority of the sample (n=159, 61.6%) had not previously accessed treatment services for methamphetamine use, while 55.8% (n=144) had injected methamphetamine in the prior 4 weeks.

Figure 4. Participant flow diagram.

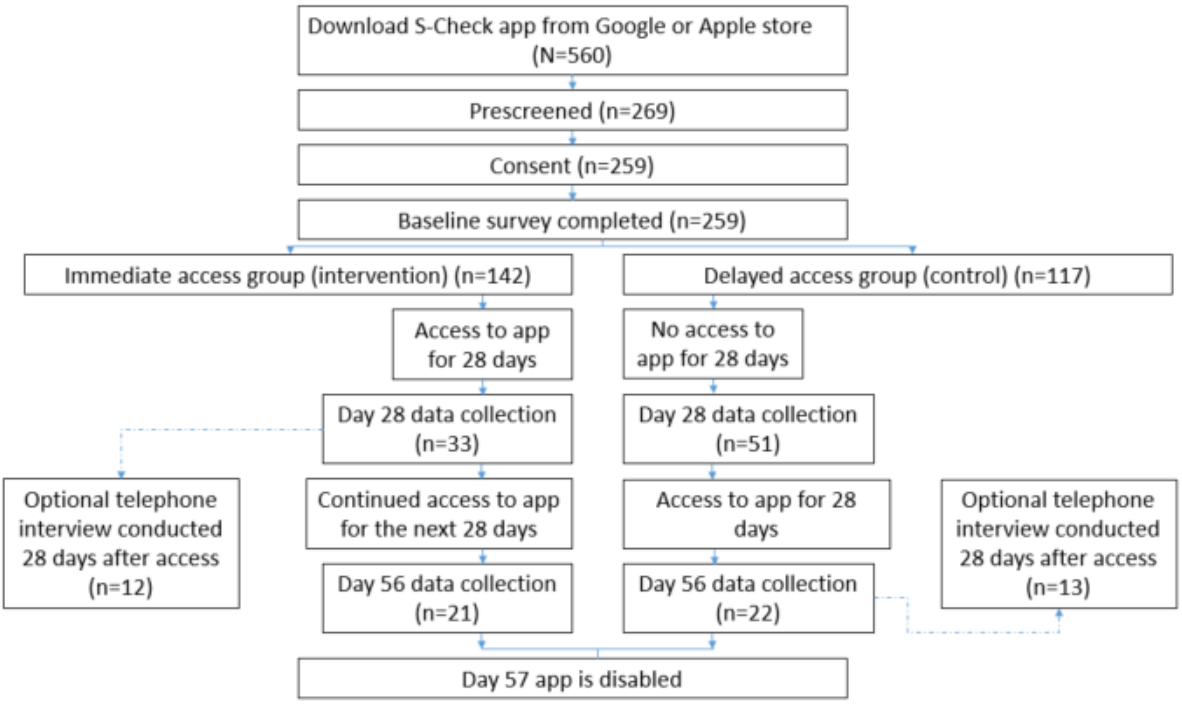


Table 2. Baseline characteristics of the sample.

Variable	Total eligible participants (n=259)			Participants retained to day 28 (n=84)		
	Control (n=117)	Intervention (n=142)	Total (n=258)[a]	Control (n=50)	Intervention (n=33)	Total, n
Age (years), mean (SD)	38.2 (10.2)	38.9 (11.1)	38.7 (10.7)	39.3 (10.9)	40.7 (10.2)	83
Sexuality and gender identity, n (%)						
Cis gay man	58 (50)	73 (51.4)	131 (50.8)	27 (54)	19 (57.6)	46
Cis heterosexual woman	19 (16.4)	20 (14.1)	39 (15.1)	5 (10)	7 (21.2)	12
Cis heterosexual man	13 (11.2)	23 (16.2)	36 (13.9)	4 (8)	3 (9.1)	7
Nonbinary queer or different	9 (7.8)	10 (7)	19 (7.4)	4 (8)	1 (3)	5
Cis bisexual woman	8 (6.9)	7 (4.9)	15 (5.8)	5 (10)	1 (3)	6
Cis bisexual man	6 (5.2)	8 (5.6)	14 (5.4)	3 (6)	2 (6.1)	5
Cis lesbian	2 (1.7)	1 (0.7)	3 (1.2)	2 (4)	0 (0)	2
Transgender heterosexual woman	1 (0.9)	0 (0)	1 (0.4)	0 (0)	0 (0)	0
Sex assigned at birth, n (%)						
Male	83 (71.6)	110 (77.5)	193 (74.8)	37 (74)	24 (72.7)	61
Female	33 (28.5)	32 (22.5)	65 (25.2)	13 (26)	9 (27.3)	22
Intersex, n (%)						
No	114 (98.3)	134 (94.4)	248 (96.1)	49 (98)	32 (97)	81
Prefer not to answer	2 (1.7)	3 (2.1)	5 (1.9)	1 (2)	0 (0)	1
Yes	0 (0)	5 (3.5)	5 (1.9)	0 (0)	1 (3)	1
Education, n (%)						
Did not complete high school	10 (8.6)	16 (11.3)	26 (10.1)	4 (8)	2 (6.1)	6
High school	26 (22.4)	41 (28.9)	67 (26)	13 (26)	9 (27.3)	22
Trade certificate or diploma	35 (30.2)	37 (26.1)	72 (27.9)	13 (26)	12 (36.4)	25
Undergraduate degree	27 (23.3)	25 (17.6)	52 (20.2)	11 (22)	7 (21.2)	18
Postgraduate degree	18 (15.5)	23 (16.2)	41 (15.9)	9 (18)	3 (9.1)	12
MA[b] use, mean (SD)						
MA use in the past 28 days (days)	15.7 (8.8)	14.3 (8.9)	14.9 (8.9)	15.8 (8.9)	15.4 (8.2)	—[c]
Age of first MA use (years)	26.4 (9.3)	26.5 (10.6)	26.4 (10)	28.9 (9)	27.5 (10.9)	—
History of treatment for problems related to MA use, n (%)						
No	74 (63.3)	85 (59.9)	159 (61.6)	35 (68.6)	24 (72.7)	59
Yes	43 (36.8)	57 (40.1)	100 (38.8)	16 (31.4)	9 (27.3)	25
History of injecting MA in past 4 weeks, n (%)						
No	50 (42.7)	65 (45.8)	115 (44.6)	21 (41.2)	9 (27.3)	30
Yes	67 (57.3)	77 (54.2)	144 (55.8)	30 (58.8)	24 (72.7)	54

[a]Baseline and day 28 demographic data on age, gender identity, sexuality, sex assigned at birth, intersex status, education, and days of methamphetamine use in the previous 28 days is missing for 1 participant.

[b]MA: methamphetamine.

[c]Not applicable.

Primary Outcomes

The primary outcomes for help-seeking were assessed using the 84 participants retained to day 28 including 33 (39.2%) participants in the intervention group and 51 (60.7%) participants in the control group.

Table 3 compares actual help-seeking (mAHSQ), intention to seek help (mGHSQ), and motivation to change (mRR) between the 2 groups at baseline and day 28. A greater proportion of participants in the intervention group recently sought

professional help (mAHSQ) at day 28 than in the control group (n=15, 45.5% vs n=12, 23.5%; χ^2_1=4.42; *P*=.04).

Regression analyses for mAHSQ, mGHSQ, and mRR are presented in Table 4. In the primary regression analyses, there was no statistically significant difference in the odds of actual help-seeking, intention to seek help, or motivation to change between the immediate access group and the waitlist control group on day 28.

Table 3. Help-seeking, intention to seek help, and motivation to change.

	Actual help-seeking (mAHSQ[a])				Intention to seek help (mGHSQ[b])[c]				Motivation to change (mRR[d])[e]	
	Professional	*P* value	Nonprofes-sional	*P* value	Professional	*P* value	Nonprofes-sional	*P* value	Motivation to change	*P* value
Baseline, n (%)		.23		.22		.19		.43		≥.99
Control	11 (21.6)		15 (29.4)		30 (58.8)		23 (45.1)		17 (33.3)	
Intervention	11 (33.3)		14 (42.4)		24 (72.7)		23 (36.4)		11 (33.3)	
Day 28, n (%)		.04		.31		.45		.90		.09
Control	12 (23.5)		22 (43.1)		32 (62.8)		27 (52.9)		14 (27.4)	
Intervention	15 (45.5)		18 (54.6)		28 (54.6)		27 (51.5)		15 (45.4)	

[a]mAHSQ: modified Actual Help Seeking Questionnaire.

[b]mGHSQ: modified General Help Seeking Questionnaire.

[c]Modified General Help Seeking Questionnaire categorized as 1-4, no intention to seek help; 5-7, intention to seek help.

[d]mRR: modified readiness ruler.

[e]Modified readiness ruler categorized as <8, not motivated to change; ≥8, motivated to change.

Table 4. Logistic regression for mAHSQ[a], mGHSQ[b], and mRR[c].

Outcome measure and outcome response set	Adjusted odds ratio[d] (95% CI)	*P* value
mAHSQ		
Professional help-seeking–actual	2.72 (0.86-8.64)	.09
Nonprofessional help-seeking–actual	1.21 (0.44-3.67)	.71
mGHSQ[e]		
Professional help-seeking–intention	0.49 (0.17-1.32)	.17
Nonprofessional help-seeking–intention	1.04 (0.42-2.64)	.93
mRR[f]		
Motivation to change	2.4 (0.89-6.47)	.08

[a]mAHSQ: modified Actual Help Seeking Questionnaire.

[b]mGHSQ: modified General Help Seeking Questionnaire.

[c]mRR: modified readiness ruler.

[d]Adjusted for baseline response.

[e]Score on a 7-point Likert scale categorized as 1-4, no intention to seek help; 5-7, intending to seek help.

[f]mRR scored as <8, not motivated to change; ≥8, motivated to change.

Ancillary Analysis

Based on the multiple imputed data set, consisting of 10 baseline characteristics, an ancillary analysis was conducted. While overall the results were consistent with our observed case analysis, there was a significant difference in the odds of recent professional help-seeking between the immediate access group compared to the waitlist control group on this analysis (adjusted odds ratio [aOR] 2.64, 95% CI 1.19-5.83, *P*=.02; Table S2 in Multimedia Appendix 1).

Secondary Outcomes

The most commonly accessed features of the app over this study's period are presented in Table S3 in Multimedia Appendix 1. The most accessed features were the dashboard feature, information, and methamphetamine tracking resources.

There were no between-group changes in days of methamphetamine use at day 28, as presented in Table S4 in Multimedia Appendix 1.

In a multivariable regression analysis with app engagement time as the response variable, participants who identified as gay

males had 68% more app engagement time than all other participants (ratio 1.68, 95% CI 1.04-2.98, *P*=.03). It was not possible to analyze S-Check app engagement by other sexualities and gender identities due to small numbers. Greater days of methamphetamine use at baseline was also found to be a predictor of S-Check app engagement, with 3% more app engagement time for each additional day of methamphetamine use (ratio 1.03, 95% CI 1.00-1.06, *P*=.03). Participants who intended to seek professional help (mGHSQ) at baseline had 17% more S-Check app engagement than those who did not (ratio 1.17, 95% CI 1.03-1.32, *P*=.01). After adjusting for baseline days of methamphetamine use, a 10-minute increase in app engagement time over the 28-day intervention was associated with a decrease in days of methamphetamine use by 0.4 days (regression coefficient [β] –0.04, *P*=.02; intervention group: n=33). In a subgroup analysis of participants who were not seeking help at baseline, for every additional minute of S-Check app use, there was an 8% increase in the likelihood of seeking professional help at day 28 (ratio 1.08, 95% CI 1.02-1.22, *P*=.04).

Discussion

Principal Findings

This study demonstrated that compared to waitlist controls, almost twice the proportion of participants with immediate access to the S-Check app sought professional help by day 28. Importantly, among participants not seeking help at baseline, each minute in the app was associated with seeking professional help by day 28, and there was an association between increased S-Check app use and decreased methamphetamine use over the 28 days among those in the immediate access group.

This study has limitations. We aimed to recruit 510 participants but we were only able to achieve a sample of 259. Retention to the primary end point was 32%, less than the expected 50%, which was likely unrealistic, given emerging evidence for this type of study suggests that when participants are enrolled in mHealth studies after downloading an app, retention is generally <30% [52]. Future research may include methods whereby participants are recruited before downloading the app; however, this design would affect the generalizability of the results to the population recruited. We learned important lessons about implementing an app-based trial. We underestimated the resources required to support a study of this nature. Keeping this study open longer may have increased our sample size. Budgeting for longer periods of recruitment and advertising may support higher enrollment in future trials. Apps to address substance use disorder should be co-designed with the community anticipated to benefit from their use. To this end, the stepped process of consultation with affected populations in transitioning from a clinical intervention to a smartphone app is a strength of the S-Check app. Further refinement of this resource or subsequent apps to address methamphetamine use should include active engagement with people who will use the app, to ensure alignment of what will benefit this population and what research and developers are delivering.

There is a possibility that this sample was already motivated toward change. Given the design of the app encourages people to understand more about their methamphetamine use and its impact, those who downloaded it likely had some interest in how their methamphetamine use was affecting them. In selecting a primary outcome, future studies of the S-Check app may therefore seek to examine how people use the app to better understand their methamphetamine use, identify which aspects of their use may be problematic for them, and how the app provides earlier insights into the relationships between methamphetamine use and aspects of their life. In the present study, assessing help-seeking and motivation to change is difficult given fixed time points and in binary measures as is demanded of a randomized controlled trial, while change is a fluid process. Engaging with the app may in itself have addressed concerns for participants, reducing the need or motivation to seek external help. Further exploration of the place of a self-administered information app such as the S-Check app within the broader treatment landscape in support of people who consume methamphetamine is warranted. A nuanced understanding of progress along the change continuum, an aim of the app, is required in future studies.

Sample characteristics limit the generalizability of the results to the broader Australian population of people who use methamphetamine. The majority of participants recruited to our study identified as male (69.9%) while 64% of past-month methamphetamine use occurs among men in Australia [53]. Our sample had a large proportion of participants who had undertaken higher level education (35.9% had a University degree), as compared to underlying population data (24%) [53]. In our study, neither ethnicity nor employment data were collected, therefore we cannot comment on the generalizability of these findings to a range of communities affected by methamphetamine use. This study recruited a large proportion of participants identifying as gay men (n=131, 50.8% of the sample), reflecting the recruitment strategy through targeted social media channels. The underrepresentation of women, transgender people, or other populations may be attributed to our recruitment strategy, but we do not have any way of knowing whether this is due to the design or content of the app. It is also possible that people did not want to use an app that would be associated with their drug use, particularly given the legal implications of drug use. Nevertheless, the anonymous nature of the S-Check app may reach participants who are reluctant to seek traditional or face-to-face treatment services due to barriers such as high cost, treatment access delays (eg, long waitlists), travel time to services, or fear of stigma.

The S-Check app recruited participants who had predominantly never sought treatment for methamphetamine use previously, and who reported injecting methamphetamine in the 4 weeks to baseline, a potential surrogate for severity of dependence. Furthermore, a higher proportion of those who were retained in the intervention group were people who injected methamphetamine. Previous studies have demonstrated greater methamphetamine-related service use among people who inject methamphetamine [54,55]. This may be due to the lack of treatment options for methamphetamine use disorders, whereas harm reduction services are more accessible, such as NSPs to reduce injection-related harms [56]. Our recruitment strategy included advertising through NSPs, and this may reflect that.

However, our results suggest that the S-Check app can act as a feasible low-resource, early intervention for participants who may be reluctant to seek traditional or face-to-face services, or are not linked to services such as those by referral points that may be made available at NSPs.

While it was not designed to be a motivator for reducing methamphetamine use, participants who spent more time engaged with the S-Check app demonstrated a reduction in methamphetamine use from baseline to day 28. This finding is consistent with other mHealth apps that included self-monitoring for alcohol and other substance use [57], reinforcing the theory that supportive self-monitoring influences behavior change [58,59]. Importantly, participants who had more frequent methamphetamine use engaged for longer with the S-Check app than those with less frequent use; and those participants who reported injecting methamphetamine were more likely to be retained to day 28 if they were randomized to the intervention group than those who did not report injecting. These findings highlight a potential treatment gap that could be filled by an mHealth intervention such as the S-Check app, or at the very least the potential of the S-Check app to be an adjunct to other treatment and harm reduction services and an important and accessible resource for when those are not available. The S-Check app was intended as an early intervention but may be an attractive resource for those people who are more frequent or injecting consumers of methamphetamine as well.

Previous research has highlighted a need for targeted interventions to improve pathways to professional support among people who use methamphetamine [55]. While we collected baseline data on 259 participants, 560 people downloaded the app. Of those who completed screening, only 10 were ineligible. This suggests that the S-Check app was highly accessible, while perhaps recruitment into the research study was not favorable; for example, it is unknown if potential participants did not proceed due to the chance of being randomized to the waitlist control. While not statistically significant, a higher proportion of those randomized to the waitlist went on to complete the day 28 assessment. It cannot be inferred from our data whether this is related to satisfaction of those randomized to immediately access the app, or perhaps reflecting the context of the waitlist group wanting to access the app.

Future directions include exploring the incorporation of the S-Check app as part of a stepped-care approach to clinical treatment, within a range of clinical settings, or as a motivational and self-awareness tool to be provided to people waiting for care. Primary care and sexual health service use is high among people who use methamphetamine, but opportunities to intervene are often missed. A health economic analysis was not part of the scope of this project and as such more work is required to test and analyze the costing and implementation of the app. Further development of the S-Check app and dissemination more broadly as a freely accessible resource to support people in understanding their methamphetamine use is planned.

Conclusions

This study demonstrates the effectiveness of the S-Check app in motivating help-seeking, particularly among treatment-naïve populations. The S-Check app was shown to be a feasible, low-resource, self-administered intervention for adults in Australia who use methamphetamine. Using supportive self-monitoring, the S-Check app assists people who use methamphetamine to identify problem use, shifting people along the motivational pathway, and promoting treatment seeking. As an early intervention, the S-Check app has the potential to facilitate the first step in a stepped-care approach to treatment, reducing treatment delays and avoiding more serious adverse health outcomes. Further work is required to demonstrate the generalizability of the S-Check app among people who use methamphetamine in Australia and to increase access to other internet-enabled devices.

Acknowledgments

The authors wish to thank all participants who took part in this study. In addition, we wish to acknowledge contributions from the investigators and research team (Ms Jessica Wilson, Mr Duncan Graham, Dr Emily Li, Ms Victoria Malone, Mr Quoc Nguyen, Ms Lauren Rogers, Dr Samineh Sanatkar) and the S-Check study advisory group (Mr Thomas Capell-Hattam, Dr Hila Haskelberg, Ms Ruth Hennessy, Dr Suzie Hudson, Ms Debbie Kaplan, Ms Chris Keyes, Associate Professor Anna McNulty). This study was funded by a New South Wales Health Early Intervention and Innovation Fund Grant (number A280560), the National Centre for Clinical Research on Emerging Drugs (which is funded by the Australian Department of Health and Aged Care), and the St Vincent's Hospital Sydney Inclusive Health Program.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT (Consolidated Standards of Reporting Trials) checklist and tables.
[DOCX File , 24 KB - mhealth_v12i1e55663_app1.docx]

Multimedia Appendix 2
CONSORT eHEALTH checklist V 1.6.1.

[PDF File (Adobe PDF File), 1146 KB - mhealth_v12i1e55663_app2.pdf]

References

1. Hillhouse MP, Marinelli-Casey P, Gonzales R, Ang A, Rawson RA, Methamphetamine Treatment Project Corporate Authors. Predicting in-treatment performance and post-treatment outcomes in methamphetamine users. Addiction 2007;102 Suppl 1:84-95. [doi: 10.1111/j.1360-0443.2007.01768.x] [Medline: 17493057]
2. Farrell M, Martin NK, Stockings E, Bórquez A, Cepeda JA, Degenhardt L, et al. Responding to global stimulant use: challenges and opportunities. Lancet 2019;394(10209):1652-1667 [FREE Full text] [doi: 10.1016/S0140-6736(19)32230-5] [Medline: 31668409]
3. Man N, Sisson SA, McKetin R, Chrzanowska A, Bruno R, Dietze PM, et al. Trends in methamphetamine use, markets and harms in Australia, 2003-2019. Drug Alcohol Rev 2022;41(5):1041-1052. [doi: 10.1111/dar.13468] [Medline: 35604870]
4. American Psychiatric Association. Diagnostic and statistical manual of mental disorders. Arlington, Virginia: American Psychiatric Association; 2013.
5. Siefried KJ, Acheson LS, Lintzeris N, Ezard N. Pharmacological treatment of methamphetamine/amphetamine dependence: a systematic review. CNS Drugs 2020;34(4):337-365 [FREE Full text] [doi: 10.1007/s40263-020-00711-x] [Medline: 32185696]
6. Stuart AM, Baker AL, Denham AM, Lee NK, Hall A, Oldmeadow C, et al. Psychological treatment for methamphetamine use and associated psychiatric symptom outcomes: a systematic review. J Subst Abuse Treat 2020;109:61-79 [FREE Full text] [doi: 10.1016/j.jsat.2019.09.005] [Medline: 31856953]
7. Brown HD, DeFulio A. Contingency management for the treatment of methamphetamine use disorder: a systematic review. Drug Alcohol Depend 2020;216:108307. [doi: 10.1016/j.drugalcdep.2020.108307] [Medline: 33007699]
8. Brensilver M, Heinzerling KG, Swanson AN, Shoptaw SJ. Placebo-group responders in methamphetamine pharmacotherapy trials: the role of immediate establishment of abstinence. Exp Clin Psychopharmacol 2012;20(5):430-435 [FREE Full text] [doi: 10.1037/a0029210] [Medline: 22867036]
9. Lee NK, Harney AM, Pennay AE. Examining the temporal relationship between methamphetamine use and mental health comorbidity. Adv Dual Diagn 2022;5(1):23-31 [FREE Full text] [doi: 10.1108/17570971211225145]
10. Quinn B, Stoové M, Papanastasiou C, Dietze P. An exploration of self-perceived non-problematic use as a barrier to professional support for methamphetamine users. Int J Drug Policy 2013;24(6):619-623. [doi: 10.1016/j.drugpo.2013.05.015] [Medline: 23867050]
11. Brener L, Lea T, Rance J, Wilson H, Bryant J, Ezard N. Providing a model of health care service to stimulant users in Sydney. Drugs: Educ, Prev Policy 2016;25(2):130-137. [doi: 10.1080/09687637.2016.1221061]
12. Kazemi DM, Borsari B, Levine MJ, Li S, Lamberson KA, Matta LA. A systematic review of the mHealth interventions to prevent alcohol and substance abuse. J Health Commun 2017;22(5):413-432 [FREE Full text] [doi: 10.1080/10810730.2017.1303556] [Medline: 28394729]
13. Dick S, Whelan E, Davoren MP, Dockray S, Heavin C, Linehan C, et al. A systematic review of the effectiveness of digital interventions for illicit substance misuse harm reduction in third-level students. BMC Public Health 2019;19(1):1244 [FREE Full text] [doi: 10.1186/s12889-019-7583-6] [Medline: 31500618]
14. Luxton DD, McCann RA, Bush NE, Mishkind MC, Reger GM. mHealth for mental health: integrating smartphone technology in behavioral healthcare. Prof Psychol: Res Pract 2011;42(6):505-512. [doi: 10.1037/a0024485]
15. Giroux I, Goulet A, Mercier J, Jacques C, Bouchard S. Online and mobile interventions for problem gambling, alcohol, and drugs: a systematic review. Front Psychol 2017;8:954 [FREE Full text] [doi: 10.3389/fpsyg.2017.00954] [Medline: 28649211]
16. Poushter J. Smartphone ownership and internet usage continues to climb in emerging economies. Pew Research Center 2016;22(1):1-44. [doi: 10.4135/9781412963947.n376]
17. Chapman C, Champion KE, Birrell L, Deen H, Brierley M, Stapinski LA, et al. Smartphone apps about crystal methamphetamine ("Ice"): systematic search in app stores and assessment of composition and quality. JMIR mHealth uHealth 2018;6(11):e10442 [FREE Full text] [doi: 10.2196/10442] [Medline: 30463834]
18. Sullivan F, Bascombe F, Ezard N. St vincent's s-check clinic model of care. Sydney: The University of New South Wales and St Vincent's Health Australia. Technical Report Number 2019/02. 2019. URL: https://nccred.org.au/wp-content/uploads/2019/12/2019-02_NCCRED_S-checkModelofCare.pdf [accessed 2024-06-06]
19. Smout MF, Longo M, Harrison S, Minniti R, Cahill S, Wickes W, et al. The psychostimulant check-up: a pilot study of a brief intervention to reduce illicit stimulant use. Drug Alcohol Rev 2010;29(2):169-176. [doi: 10.1111/j.1465-3362.2009.00133.x] [Medline: 20447225]
20. Bensley RJ, Mercer N, Brusk JJ, Underhile R, Rivas J, Anderson J, et al. The eHealth behavior management model: a stage-based approach to behavior change and management. Prev Chronic Dis 2004;1(4):A14 [FREE Full text] [Medline: 15670446]
21. Ajzen I. The theory of planned behavior. Organ Behav Hum Decis Processes 1991;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-t]

22. Heron KE, Smyth JM. Ecological momentary interventions: incorporating mobile technology into psychosocial and health behaviour treatments. Br J Health Psychol 2010;15(Pt 1):1-39 [FREE Full text] [doi: 10.1348/135910709X466063] [Medline: 19646331]
23. Trull TJ, Ebner-Priemer UW. Using experience sampling methods/ecological momentary assessment (ESM/EMA) in clinical assessment and clinical research: introduction to the special section. Psychol Assess 2009;21(4):457-462 [FREE Full text] [doi: 10.1037/a0017653] [Medline: 19947780]
24. Salazar A, de Sola H, Failde I, Moral-Munoz JA. Measuring the quality of mobile apps for the management of pain: systematic search and evaluation using the mobile app rating scale. JMIR mHealth uHealth 2018;6(10):e10718 [FREE Full text] [doi: 10.2196/10718] [Medline: 30361196]
25. Tomczyk S, Schomerus G, Stolzenburg S, Muehlan H, Schmidt S. Ready, willing and able? An investigation of the theory of planned behaviour in help-seeking for a community sample with current untreated depressive symptoms. Prev Sci 2020;21(6):749-760 [FREE Full text] [doi: 10.1007/s11121-020-01099-2] [Medline: 32140825]
26. Gajecki M, Berman AH, Sinadinovic K, Rosendahl I, Andersson C. Mobile phone brief intervention applications for risky alcohol use among university students: a randomized controlled study. Addict Sci Clin Pract 2014;9(1):11 [FREE Full text] [doi: 10.1186/1940-0640-9-11] [Medline: 24985342]
27. Herman D, Ferrari M, Moran-Pryor A, Malone V, Ezard N. Usability and acceptability of S-Check app: a harm reduction and early intervention smartphone application to engage people who use methamphetamine and are not-in-treatment. 2018 Presented at: NADA Conference Exploring Therapeutic Interventions; June 7-8, 2018; Sydney, NSW.
28. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. Ann Intern Med 2010;152(11):726-732 [FREE Full text] [doi: 10.7326/0003-4819-152-11-201006010-00232] [Medline: 20335313]
29. Nguyen Q, Middleton P, Herman D, Li K, Grundy E, Li E. What's in an app? Incorporating an automated consent procedure to recruit those who use methamphetamine to a harm reduction and early intervention smartphone-based application clinical trial'. Drug Alcohol Rev 2019;38(S1):S77-S77 [FREE Full text]
30. Ali R, Meena S, Eastwood B, Richards I, Marsden J. Ultra-rapid screening for substance-use disorders: the Alcohol, Smoking and Substance Involvement Screening Test (ASSIST-Lite). Drug Alcohol Depend 2013;132(1-2):352-361. [doi: 10.1016/j.drugalcdep.2013.03.001] [Medline: 23561823]
31. Gossop M, Darke S, Griffiths P, Hando J, Powis B, Hall W, et al. The Severity of Dependence Scale (SDS): psychometric properties of the SDS in English and Australian samples of heroin, cocaine and amphetamine users. Addiction 1995;90(5):607-614. [doi: 10.1046/j.1360-0443.1995.9056072.x] [Medline: 7795497]
32. NA. Mental State Examination. In: Clinical Practice Guidelines. Melbourne: The Royal Children's Hospital Melbourne; 2018.
33. Wellbeing measures in primary health care/the DEPCARE project: report on a WHO meeting. World Health Organization. 1998. URL: https://iris.who.int/bitstream/handle/10665/349766/WHO-EURO-1998-4234-43993-62027-eng.pdf [accessed 2024-06-06]
34. Kessler RC, Barker PR, Colpe LJ, Epstein JF, Gfroerer JC, Hiripi E, et al. Screening for serious mental illness in the general population. Arch Gen Psychiatry 2003;60(2):184-189. [doi: 10.1001/archpsyc.60.2.184] [Medline: 12578436]
35. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
36. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
37. The Whoqol Group. The world health organization quality of life assessment (WHOQOL): development and general psychometric properties. Soc Sci Med 1998;46(12):1569-1585. [doi: 10.1016/s0277-9536(98)00009-4] [Medline: 9672396]
38. Cella D, Lai J, Nowinski CJ, Victorson D, Peterman A, Miller D, et al. Neuro-QOL: brief measures of health-related quality of life for clinical research in neurology. Neurology 2012;78(23):1860-1867 [FREE Full text] [doi: 10.1212/WNL.0b013e318258f744] [Medline: 22573626]
39. Ising HK, Veling W, Loewy RL, Rietveld MW, Rietdijk J, Dragt S, et al. The validity of the 16-item version of the prodromal questionnaire (PQ-16) to screen for ultra high risk of developing psychosis in the general help-seeking population. Schizophr Bull 2012;38(6):1288-1296 [FREE Full text] [doi: 10.1093/schbul/sbs068] [Medline: 22516147]
40. Wilson C, Deane F, Ciarrochi J, Rickwood D. Measuring help-seeking intentions: properties of the general help-seeking questionnaire. Canadian J Counselling 2005;39(1):15-28. [doi: 10.1037/t42876-000]
41. Tait RJ, McKetin R, Kay-Lambkin F, Bennett K, Tam A, Bennett A, et al. Breakingtheice: a protocol for a randomised controlled trial of an internet-based intervention addressing amphetamine-type stimulant use. BMC Psychiatry 2012;12:67 [FREE Full text] [doi: 10.1186/1471-244X-12-67] [Medline: 22731926]
42. Beiwinkel T, Hey S, Bock O, Rössler W. Supportive mental health self-monitoring among smartphone users with psychological distress: protocol for a fully mobile randomized controlled trial. Front Public Health 2017;5:249 [FREE Full text] [doi: 10.3389/fpubh.2017.00249] [Medline: 28983477]

43. Cheng VWS, Davenport T, Johnson D, Vella K, Mitchell J, Hickie IB. Naturalistic evaluation of a sport-themed mental health and wellbeing app aimed at men (MindMax), that incorporates applied video games and gamification. Internet Interv 2020;20:100306 [FREE Full text] [doi: 10.1016/j.invent.2020.100306] [Medline: 32181146]
44. Gulliver A, Griffiths KM, Christensen H, Mackinnon A, Calear AL, Parsons A, et al. Internet-based interventions to promote mental health help-seeking in elite athletes: an exploratory randomized controlled trial. J Med Internet Res 2012;14(3):e69 [FREE Full text] [doi: 10.2196/jmir.1864] [Medline: 22743352]
45. Rickwood D, Thomas K, Bradford S. Review of help-seeking measures in mental health: an evidence check rapid review brokered by the Sax Institute for Beyondblue. Sax Institute for Beyondblue 2012:1-35.
46. Rickwood DJ, Braithwaite VA. Social-psychological factors affecting help-seeking for emotional problems. Soc Sci Med 1994;39(4):563-572. [doi: 10.1016/0277-9536(94)90099-x] [Medline: 7973856]
47. Bowling A. Just one question: if one question works, why ask several? J Epidemiol Community Health 2005;59(5):342-345 [FREE Full text] [doi: 10.1136/jech.2004.021204] [Medline: 15831678]
48. LaBrie JW, Quinlan T, Schiffman JE, Earleywine ME. Performance of alcohol and safer sex change rulers compared with readiness to change questionnaires. Psychol Addict Behav 2005;19(1):112-115. [doi: 10.1037/0893-164X.19.1.112] [Medline: 15783287]
49. Hesse M. The readiness ruler as a measure of readiness to change poly-drug use in drug abusers. Harm Reduct J 2006;3:3 [FREE Full text] [doi: 10.1186/1477-7517-3-3] [Medline: 16436208]
50. Cohen J. A power primer. Psychol Bull 1992;112(1):155-159. [doi: 10.1037//0033-2909.112.1.155] [Medline: 19565683]
51. Jakobsen JC, Gluud C, Wetterslev J, Winkel P. When and how should multiple imputation be used for handling missing data in randomised clinical trials - a practical guide with flowcharts. BMC Med Res Methodol 2017;17(1):162 [FREE Full text] [doi: 10.1186/s12874-017-0442-1] [Medline: 29207961]
52. Staiger PK, O'Donnell R, Liknaitzky P, Bush R, Milward J. Mobile apps to reduce tobacco, alcohol, and illicit drug use: systematic review of the first decade. J Med Internet Res 2020;22(11):e17156 [FREE Full text] [doi: 10.2196/17156] [Medline: 33231555]
53. National Drug Strategy Household Survey 2019. Canberra: Australian Institute of Health and Welfare; 2019.
54. McKetin R, Kelly E. Socio-demographic factors associated with methamphetamine treatment contact among dependent methamphetamine users in Sydney, Australia. Drug Alcohol Rev 2007;26(2):161-168. [doi: 10.1080/09595230601146652] [Medline: 17364851]
55. Quinn B, Stoové M, Dietze P. Factors associated with professional support access among a prospective cohort of methamphetamine users. J Subst Abuse Treat 2013;45(2):235-341. [doi: 10.1016/j.jsat.2013.02.003] [Medline: 23522848]
56. Nielsen S, Bruno R, Lintzeris N, Fischer J, Carruthers S, Stoové M. Pharmaceutical opioid analgesic and heroin dependence: how do treatment-seeking clients differ in Australia? Drug Alcohol Rev 2011;30(3):291-299. [doi: 10.1111/j.1465-3362.2011.00302.x] [Medline: 21545560]
57. Aharonovich E, Stohl M, Cannizzaro D, Hasin D. HealthCall delivered via smartphone to reduce co-occurring drug and alcohol use in HIV-infected adults: a randomized pilot trial. J Subst Abuse Treat 2017;83:15-26 [FREE Full text] [doi: 10.1016/j.jsat.2017.09.013] [Medline: 29129192]
58. Michie S, Whittington C, Hamoudi Z, Zarnani F, Tober G, West R. Identification of behaviour change techniques to reduce excessive alcohol consumption. Addiction 2012;107(8):1431-1440. [doi: 10.1111/j.1360-0443.2012.03845.x] [Medline: 22340523]
59. Michie S, Abraham C, Whittington C, McAteer J, Gupta S. Effective techniques in healthy eating and physical activity interventions: a meta-regression. Health Psychol 2009;28(6):690-701. [doi: 10.1037/a0016136] [Medline: 19916637]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
ITT: intention-to-treat
mAHSQ: modified Actual Help Seeking Questionnaire
mGHSQ: modified General Help Seeking Questionnaire
mRR: modified readiness ruler
NSP: needle and syringe program

Edited by L Buis; submitted 19.12.23; peer-reviewed by S Kershaw, AJ Saxon; comments to author 18.01.24; revised version received 23.01.24; accepted 10.05.24; published 03.07.24.

Please cite as:
Siefried KJ, Bascombe F, Clifford B, Liu Z, Middleton P, Kay-Lambkin F, Freestone J, Herman D, Millard M, Steele M, Acheson L, Moller C, Bath N, Ezard N
Effect of a Smartphone App (S-Check) on Actual and Intended Help-Seeking and Motivation to Change Methamphetamine Use Among Adult Consumers of Methamphetamine in Australia: Randomized Waitlist-Controlled Trial
JMIR Mhealth Uhealth 2024;12:e55663
URL: https://mhealth.jmir.org/2024/1/e55663
doi:10.2196/55663
PMID:

Original Paper

Deconstructing Fitbit to Specify the Effective Features in Promoting Physical Activity Among Inactive Adults: Pilot Randomized Controlled Trial

Keisuke Takano[1*], PhD; Takeyuki Oba[1*], PhD; Kentaro Katahira[1], PhD; Kenta Kimura[1], PhD

Human Informatics and Interaction Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Ibaraki, Japan

[*]these authors contributed equally

Corresponding Author:
Keisuke Takano, PhD
Human Informatics and Interaction Research Institute
National Institute of Advanced Industrial Science and Technology (AIST)
1-1-1 Higashi, Tsukuba
Ibaraki, 305-8566
Japan
Phone: 81 298491456
Email: keisuke.takano@aist.go.jp

Abstract

Background: Wearable activity trackers have become key players in mobile health practice as they offer various behavior change techniques (BCTs) to help improve physical activity (PA). Typically, multiple BCTs are implemented simultaneously in a device, making it difficult to identify which BCTs specifically improve PA.

Objective: We investigated the effects of BCTs implemented on a smartwatch, the Fitbit, to determine how each technique promoted PA.

Methods: This study was a single-blind, pilot randomized controlled trial, in which 70 adults (n=44, 63% women; mean age 40.5, SD 12.56 years; closed user group) were allocated to 1 of 3 BCT conditions: self-monitoring (feedback on participants' own steps), goal setting (providing daily step goals), and social comparison (displaying daily steps achieved by peers). Each intervention lasted for 4 weeks (fully automated), during which participants wore a Fitbit and responded to day-to-day questionnaires regarding motivation. At pre- and postintervention time points (in-person sessions), levels and readiness for PA as well as different aspects of motivation were assessed.

Results: Participants showed excellent adherence (mean valid-wear time of Fitbit=26.43/28 days, 94%), and no dropout was recorded. No significant changes were found in self-reported total PA (dz<0.28, P=.40 for the self-monitoring group, P=.58 for the goal setting group, and P=.19 for the social comparison group). Fitbit-assessed step count during the intervention period was slightly higher in the goal setting and social comparison groups than in the self-monitoring group, although the effects did not reach statistical significance (P=.052 and P=.06). However, more than half (27/46, 59%) of the participants in the precontemplation stage reported progress to a higher stage across the 3 conditions. Additionally, significant increases were detected for several aspects of motivation (ie, integrated and external regulation), and significant group differences were identified for the day-to-day changes in external regulation; that is, the self-monitoring group showed a significantly larger increase in the sense of pressure and tension (as part of external regulation) than the goal setting group (P=.04).

Conclusions: Fitbit-implemented BCTs promote readiness and motivation for PA, although their effects on PA levels are marginal. The BCT-specific effects were unclear, but preliminary evidence showed that self-monitoring alone may be perceived demanding. Combining self-monitoring with another BCT (or goal setting, at least) may be important for enhancing continuous engagement in PA.

Trial Registration: Open Science Framework; https://osf.io/87qnb/?view_only=f7b72d48bb5044eca4b8ce729f6b403b

(JMIR Mhealth Uhealth 2024;12:e51216) doi:10.2196/51216

KEYWORDS

wearable activity tracker; mHealth; mobile health; motivation; physical activity; lifestyle; smartwatch; wearables; Fitbit; exercise; fitness; BCT; behavior change technique; behavior change; motivation; adherence; engagement

Introduction

Background

Promoting physical activity (PA) is an urgent mission in modern society, as physical inactivity is a known risk factor for a wide range of illnesses such as diabetes, cardiovascular diseases, and depression [1,2]. Behavior change techniques (BCTs) and lifestyle interventions have been studied extensively, and systematic reviews and meta-analyses have found that, overall, behavioral interventions are effective in promoting PA [3-5]. Notably, many behavioral interventions have been delivered digitally through web applications or smartphone apps (or augmented by online support tools) for recording, giving feedback on, or coaching to promote PA. Thus, eHealth and mobile health (mHealth) interventions are expected to play an important role in promoting PA, especially during and after the COVID-19 pandemic [6,7].

Key players in digital health care research and business are wearable activity trackers, which are (consumer) devices that monitor PA and provide feedback to the wearer. These devices include pedometers, fitness trackers, smartwatches, and smart rings, which are often linked to web applications or smartphone apps that offer individualized dashboards, health messages, social support, and tips to continue exercising regularly (eg, [8,9]). Meta-analyses [10-12] have suggested that activity trackers significantly improve PA with standardized mean difference (SMD)=0.3-0.6, equating to approximately 1800-1850 extra steps per day.

Wearable activity trackers have been found to be well accepted by various populations, including adolescents [13], older adults with chronic diseases [14,15], and cancer survivors [16]. Users typically report positive experiences with activity trackers, which are perceived to be easy to use and useful to track daily (physical) activity. Thematic analyses highlighted enhanced self-awareness among users, which increases motivation to improve or maintain PA [14,15]. Indeed, a pilot trial showed that activity trackers help improve cardiometabolic risk variables in patients with diabetes and could potentially be a motivation tool to increase PA [17,18], particularly with appropriate features and implementations [19].

Fitbit devices are some of the most widely used wearable activity trackers in health care research. A recent narrative review [20] identified 75 published trials using Fitbit, in which a variety of behavior change strategies (eg, competition or challenges, self-monitoring, social support, and goal setting) were implemented. Through these trials, considerable heterogeneity was found in the use of Fitbit. Typically, these trials were designed to investigate the overall effect of a Fitbit-based intervention versus an intervention without Fitbit (or treatment as usual). If not limited to Fitbit, it is often the case with behavioral interventions in which multiple BCTs (ie, 5-10 different techniques [9]) are implemented simultaneously, which makes it difficult to identify which BCTs are active components to improve PA and motivation for PA. Therefore, we aimed to investigate the effects of different BCTs implemented using the Fitbit Sense smartwatch to determine how each technique would improve PA and motivation for PA through a mobile intervention.

The following three BCTs (indexed by the BCT taxonomy v1 [5]) were targeted in this study: (1) self-monitoring, which encouraged participants to monitor and record their step counts each day (*2. Feedback and monitoring* and *2.3 Self-monitoring of behavior*); (2) goal setting, which provided participants with an explicit daily step count goal and feedback on the goal progress (*1.1 Goal setting* and *2.7 Feedback on outcomes of behavior*); and (3) social comparison, which drew the participants' attention to their peers' step counts to allow comparison with their own performance (*6.2 Social comparison*). Note that the latter 2 conditions included self-monitoring by nature; without self-monitoring, feedback on goal progress or social comparisons could not be implemented.

We regarded self-monitoring as the control condition (or treatment as usual) because it is the most prevalent BCT implemented in mHealth interventions, and there is robust evidence that self-monitoring increases PA and improves dietary behavior [21,22]. A systematic review on digital implementations of self-monitoring [23] distinguished active self-monitoring, which requires users to reflect on the current state and physical functioning, from passive self-monitoring, which relies on sensors automatically recording (locomotor) acceleration, heart rate, and other physical or environmental information. In this study, we used active self-monitoring (ie, participants manually entered their step count of the day on an online questionnaire platform). This decision was made because (1) we wanted to be sure that each participant consciously checked their steps of the day and (2) the systematic review [23] identified in the literature no intervention study that included wearables to combine active self-monitoring with passive self-monitoring of additional physical information.

Furthermore, a meta-analysis [24] suggested that interventions combining self-monitoring with at least 1 other self-regulatory technique (eg, goal setting or feedback) are more effective than other interventions. A goal setting BCT is implemented in most commercial PA apps—95% set specific and measurable goals, which allow users to evaluate whether and to what extent they achieved the goal [25]. We followed this basic, core implementation, although other components of goal settings (eg, tailoring goal difficulty, scheduling, and action planning) were not considered in this study.

A systematic review [26] found that most of the social features implemented in mHealth interventions involve the delivery of social support or social comparison, which are often combined with the self-monitoring of behavior. The researchers also meta-analyzed the efficiency of mHealth interventions with social features, showing a moderate-to-large size effect, albeit

with large heterogeneity (SMD=0.96; I^2=99.6%). A recent analysis of the use patterns of commercial PA apps in a community sample suggested that people typically use only 2 functions within an app, often a combination of self-monitoring and one other function [27]. Given that users highly appreciate the simplicity of an app [28], it is of theoretical and practical importance to clarify how large the additive values of goal setting and social comparison are compared with self-monitoring alone.

Objectives

In summary, we aimed to test the effects of the 3 BCTs implemented using the Fitbit smartwatch to resolve an answered question in the literature, that is, how each BCT would contribute to improvement in PA levels and motivation in the form of mHealth intervention. A 2 × 3 factorial design was used, with assessment occasions (pre- vs postintervention time points) as the within-person factor and groups (self-monitoring vs goal setting vs social comparison) as the between-person factor. During the 4-week intervention, participants wore a Fitbit Sense to receive 1 of the 3 interventions (ie, self-monitoring, goal setting, or social comparison) and assess their levels of PA (in the form of step count) for 24 hours. They also responded to daily online questionnaires asking how motivated they were regarding PA. The primary outcome was self-reported PA levels assessed before and after the intervention. The secondary outcomes included (1) the stage of change for PA [29], (2) intrinsic and extrinsic motivations for PA, (3) Fitbit-assessed step count, and (4) day-to-day motivations for PA [30]. We expected that the goal setting and social comparison groups would show larger increases in the levels and motivations for PA compared with the self-monitoring group. We did not have a clear hypothesis for a direct comparison between goal setting and social comparison, which was tested in an exploratory manner.

Methods

Participants

Invitations were spread to local inhabitants in and around Ibaraki Prefecture (a middle-north region in Japan; February 2023) who had been registered in a database of potential participants. The study was advertised (online) as one that assessed psychological and behavioral data in daily life. The advertisement explicitly stated that participants would be asked to wear the Fitbit for 1 month, but they were blinded to the intervention conditions. In total, 140 individuals were assessed for eligibility, and 70 (50%) met the following inclusion criteria: (1) being aged 18-64 years; (2) expecting no major life change (eg, in a living place or job) before or during the study; (3) being not pregnant or not lactating; (4) being in good health (ie, having no medical, neurological, psychiatric, or other physical illnesses); (5) having not been diagnosed with dysautonomia, depression, mania, bipolar disorder, or facial paralysis in the past; (6) not taking tranquilizers or antihypertensive drugs regularly; (7) not taking hormone replacement therapy or not taking oral contraceptives; and (8) not exercising regularly. Regular PA levels (criterion 8) were assessed using the Stage of Change (SoC) questionnaire for PA (see the *Measures* section), and those who identified themselves as being at the precontemplation, contemplation, or preparation stages were eligible for participation. We assessed SoC instead of PA levels because we targeted motivation and readiness for PA, not only the actual engagement in PA behavior. Among the 70 excluded participants, 60 (86%) were physically more active than criterion 8 (belonging to the action or higher stages of change), and 10 (14%) could not find time for the briefing session of the study. All participants were assumed to have good internet literacy (necessary support was provided at the briefing session by an experimenter).

The sample size was determined after the recommendations of a pilot randomized controlled trial (eg, [31]). We did not perform a rigorous power analysis because we regarded this study as a pilot trial, and there was no good way to determine an expected effect size that could be used for a priori power analysis, particularly for the comparison of goal setting and social comparison. However, the current sample size (n=23 or 24 per group) allowed for the detection of an effect of Cohen d=0.87 with α=.05, and power=0.80. A meta-analysis [26] reported SMD=0.96 as the effect of social features implemented in mobile interventions on PA levels albeit high heterogeneity being documented.

Interventions

Participants were randomly allocated to 1 of 3 groups and received a smartphone and a Fitbit Sense at the beginning of the study (briefing session). The Fitbit app was installed on the smartphones with varying settings for each intervention condition.

In the self-monitoring group, participants received a smartphone with the Fitbit app, with all notification functions turned off, except for low battery warnings. Participants were allowed to see the home display, showing the step count of the day and other physiological information, including sleep, heart rate, and skin temperature. However, the app indicated no step goals to achieve, social functions were disabled, and no push notifications were sent to the devices during the intervention period. Participants received a chat message on their smartphone each evening (at 9 PM) asking them to read the step count of the day displayed on the Fitbit and manually record it on an online questionnaire platform.

Participants in the goal setting group were explicitly informed of the goal to be achieved each day (ie, 9000 and 8500 steps for men and women, respectively) by an experimenter during the briefing session. These goals were set according to the Japanese national health recommendations [32]. During the briefing session, the experimenter instructed the participants to enter the step goal into their Fitbit app to monitor their progress at any time, and the Fitbit Sense notified each participant of their daily goal achievement using vibration.

The participants in the social comparison group were explicitly instructed to read and record the step counts of their peers on the Fitbit app each evening. On the Fitbit app, 5 peers (called “friends”) had been registered before distribution. Participants were not explicitly told who these friends were, but each friend was displayed on the app with a unique participant ID (eg, AIST_01 and AIST_02). These 5 friends were the researchers

and research assistants of this project, who wore the Fitbit Sense throughout the period of data collection. The average step count for the 5 friends was 6405-11,449 steps per day. Each participant was paired with the same 5 friends so that there was no variability in the friends' activity levels between participants. We asked participants to check their friends' steps each day but did not explicitly ask participants to outperform their friends.

Measures

We administered the following three types of assessments: (1) pre- and postintervention assessments in the lab, (2) daily in-intervention assessments at home using self-report scales for motivation, and (3) accelerometry (implemented in Fitbit) for step counts (Table 1).

Table 1. Overview of the measures and response items.

Type and timing of assessment and measure	Variable and description
Self-report (in the lab, at pre- and postintervention time points)	
Stage of Change questionnaire (at the eligibility screening)	Stage of change, readiness for PA[a]
International Physical Activity Questionnaire—Short Form	Levels of total PA
Self-determined Motivation Scale for Exercise	Different motivations for regular PA (intrinsic motivation, external regulation, introjected regulation, identified regulation, integrated regulation, and amotivation)
Self-report (at home, each evening)	
Modified Situational Motivational Scale	Identified regulation, intrinsic motivation, and external regulation
Step count of the day (all 3 groups)	Read and record the step count of the day, displayed on Fitbit
Goal achievement of the day (goal setting group only)	Record whether the goal (9000 or 8500 steps) was achieved
Step count of a peer (social comparison group only)	Record the step count of the Fitbit friend whose count was the highest among the friend group
Fitbit (at home, at any time)	
Step count	Accelerometer based

[a]PA: physical activity.

SoC Questionnaire

Readiness for PA was assessed using the Japanese version [33,34] of the SoC questionnaire [29,35]. Participants were classified into 1 of 5 stages (ie, precontemplation, contemplation, preparation, action, and maintenance), selecting the most applicable statement among the following: "I currently do not exercise and do not intend to start exercising in the future" (pre-contemplation); "I currently do not exercise but I am thinking about starting to exercise in the next six months" (contemplation); "I currently exercise some, but not regularly" (preparation); "I currently exercise regularly, but have only begun doing so within the last six months" (action); and "I currently exercise regularly and have done so for longer than six months" (maintenance). We instructed the participants to consider regular exercise as exercising twice or more per week for 20 minutes or longer.

International Physical Activity Questionnaire—Short Form

Total PA was assessed using the International Physical Activity Questionnaire—Short Form (IPAQ-SF) [36,37]. The IPAQ-SF includes three dimensions: (1) walking, (2) moderate-intensity activity, and (3) vigorous-intensity activity. Sedentary time was not included in the analyses. Participants reported the number of days and duration spent on each dimension of PA over an average week. The reported weekly minutes of PA (duration×days for each PA dimension) were transformed into hourly metabolic equivalents per week (METs-h/w). This scaling allowed us to explore how many participants adhered to the Japanese public health guidelines for PA, which is 23 METs-h/w for adults [32].

Self-determined Motivation Scale for Exercise

The baseline and postintervention levels of motivation for PA were assessed using the revised version of the Self-determined Motivation Scale for Exercise [38], which was developed based on self-determination theory [39]. This 22-item self-report questionnaire covers the following six types of motivations or regulatory styles concerning PA: (1) intrinsic motivation (eg, "Exercising itself is fun"), (2) external regulation (eg, "I exercise because other people will be pleased with me"), (3) introjected regulation (eg, "I feel guilty if I do not exercise"), (4) identified regulation (eg, "I think it is a good way to improve myself"), (5) integrated regulation (eg, "It is essential to my identity and sense of self"), and (6) amotivation (eg, "I do not know why I exercise"). Each construct is assessed using 3- or 4-item subscales, and each item is rated on a 5-point scale (1=not at all true and 5=very true). Cronbach α was >.73 for each subscale.

Daily Measures: Motivation

A chat message was sent to the participants' smartphones in the evening (at 9 PM). The message included a link to an online survey platform (Exkuma [40]) where participants rated their motivation for PA. The link was to expire at 2 AM, but because of a technical problem, responses were erroneously accepted even after this due time. All such delayed responses were

excluded from analyses. The evening questionnaire included the following 3 items: identified regulation ("I'm feeling that exercising is important for me"), intrinsic motivation ("I'm feeling that exercising is fun"), and external motivation ("I'm feeling that I have to exercise"). Each item is rated on a 7-point scale (1=not at all and 7=very much). Motivation items were adapted from the Situational Motivational Scale [41] for physical exercise [42]. We adjusted the wording for daily or momentary assessments (ie, asking how one is feeling right now) and selected the items with the highest factor loadings in each dimension to minimize the burden on the participants [30]. In addition, we did not include items on amotivation, as our focus was on a pro-PA psychological state, not on the barriers to PA.

Procedure

Interested participants completed an online screening survey for eligibility. Those identified as eligible were then invited to an online system to schedule an in-person briefing session, where groups of 1-4 participants were formed arbitrarily. These groups were randomly allocated to 1 of the 3 intervention conditions (1:1:1) following a random sequence generated by a trial analyst using R (version 4; R Foundation for Statistical Computing). Experimenters (who conducted the recruitment and assessments) were not involved in this process. Block randomization (batch per group size) was used to prevent imbalances in group allocation. Single blinding was used; participants were not aware of whether they were in an active or control condition or what interventions other participants received. During the briefing session, the participants received instructions about the study, provided informed consent, and completed baseline questionnaires. At the end of this session, the participants received a smartphone (Android OS) paired with a Fitbit Sense. As the data collected via these devices were tagged with participant IDs (but not with their real names), there was no security concerns about personal information.

During the 4-week intervention period, participants were asked to wear a Fitbit Sense on their nondominant wrist for 24 hours. Participants were instructed not to change the settings of the Fitbit app. Each evening, all participants recorded their daily step count on an online questionnaire platform. Those in the social comparison group were also asked to record the highest step count of the day among their friends on the Fitbit app. There was no in-person component in the interventions although technical support (eg, for app login and online questionnaires) was provided by an experimenter at any point of the interventions.

Immediately after the 4-week interventions, the participants were invited to the lab again, returned their smartphones and Fitbit, and completed the postintervention questionnaires. They also took part in a brief in-person interview asking whether participants experienced any adverse events during the intervention period and how they liked Fitbit and each intervention component (ie, self-monitoring, goal setting, or social comparison). At the end of the debriefing session, the participants received 49,000 JPY (US $350) as compensation. Postintervention checks were conducted on the devices right after the interventions were completed. These checks reassured us that (1) the step goals were not changed on the system (8500 or 9000 steps in the goal setting group and no goals in the other 2 groups), and similarly, (2) the friend list was not changed (5 peers in the social comparison group, and no peers in the other 2 groups). The data collection was ended in May 2023.

Ethical Considerations

The study protocol was approved by the Ethics Committee of the National Institute of Advanced Industrial Science and Technology (approval ID: 2022-1240). The protocol was preregistered with the Open Science Framework before data collection.

Statistical Analyses

The primary outcome (ie, self-reported PA level) was analyzed using a 2-level multilevel model with the groups as between-person predictors, time (pre- vs postintervention time points) as the within-person predictor, and cross-level interactions. Groups were dummy coded, with the self-monitoring group as the reference level. Random intercepts and slopes were assumed, which allowed the baseline level of PA and rate of change in PA to vary across individuals. No correlation was assumed between the intercept and the slope. Each model was estimated using the restricted maximum likelihood estimation method. The same or similar models were used for the secondary outcomes. An exception is that the analyses of day-level variables (ie, daily step count and motivations reported each evening) had a continuous time predictor (ie, days since the start of the intervention) instead of pre- and postintervention time points. All randomized participants were included in the analyses on an intention-to-treat basis. The effect sizes (Cohen *d* for between-group differences and *dz* for within-group changes) were calculated [43]. All analyses were performed using R (version 4; R Foundation for Statistical Computing) with the specific packages *lme4* [44], *lmerTest* [45], and *fitbitr* [46].

Sensitivity analyses were planned to examine the influences of missingness on (1) the Fitbit-assessed step count and (2) evening levels of motivation for PA. First, for the main analysis on step count, we excluded days with steps <2000 as outliers. This cutoff was determined on the distribution observed in the current data (5.6% [110/1960] of records met <2000 steps). Also, 2000 steps can be reliably regarded as too few given that a daily step count of 5000 is operationalized as a sedentary lifestyle [47], and thus, measurement issues were suspected (eg, Fitbit was not worn appropriately). Second, as a sensitivity analysis, we reran the analysis while filtering out the days with missing heart rate (ie, days with <50% of valid heart rate records) instead of days with <2000 steps. Note that heart rate was not a target of our analyses but was measured by Fitbit; also, missingness in heart rate does not always indicate invalid wearing as temporary dysfunctions in the device or sensor may result in missingness. Third, we checked compliance in responding to daily questionnaires to see whether any participants should be excluded because of too few valid responses. However, we found excellent compliance in the current data (24.96/28 days, 89%, 1747/1960; see the *Results* section for details), and thus, data from all participants were used for statistical analyses.

Results

Demographics and Descriptive Statistics at the Baseline

The demographic characteristics of the participants are shown in Table 2. At the baseline assessment, the 70 participants were physically inactive: 46 (66%) in the precontemplation stage, 15 (21%) in the contemplation stage, and 9 (13%) in the preparation stage. In the IPAQ-SF, 46 (66%) participants did not meet the national health guidelines for PA (23 METs-h/w). Among the remaining participants (n=24), 10 (42%) were classified under category 3 (health-enhancing PA). As these highly active individuals were not the target of this study (and responded inconsistently to the SoC questionnaire and IPAQ-SF), we performed sensitivity analyses after excluding them at baseline (see Tables S4 and S5 in Multimedia Appendix 1 for details; overall, we identified no significant changes in the results that influenced the conclusion). No significant group differences were found in age, gender, education level, or other socioeconomic factors.

Table 2. Demographics and descriptives at baseline.

Variable	Self-monitoring (n=23)	Goal setting (n=23)	Social comparison (n=24)	Test on group differences, *F* test (*df*) or chi-square (*df*)	*P* value
Age (years), mean (SD)	40.1 (11.4)	39.2 (12.8)	42.1 (13.7)	0.31 (2, 67)	.73
Women, n (%)	15 (65.2)	17 (73.9)	12 (50.0)	2.96 (2)	.23
BMI, mean (SD)	25.0 (4.5)	23.5 (5.5)	25.3 (4.1)	0.97 (2, 67)	.39
Married, n (%)	13 (56.5)	11 (47.8)	17 (70.8)	2.62 (2)	.27
Having a child, n (%)	13 (56.5)	13 (56.5)	14 (58.3)	0.02 (2)	.99
Education level, n (%)				0.91 (4)	.92
Middle or high school	10 (43.5)	11 (47.8)	13 (54.2)		
College or vocational school	6 (26.1)	5 (21.7)	6 (25)		
University and above	7 (30.4)	7 (30.4)	5 (20.8)		
Job, n (%)	17 (73.9)	19 (82.6)	20 (83.3)	0.78 (2)	.67
Household income, n (%)				9.11 (8)	.33
<3 million JPY (US $19,000)	7 (30.4)	3 (13.0)	6 (25)		
3-5 million JPY (US $19,000-31,600)	8 (34.8)	13 (56.5)	7 (29.2)		
5-7 million JPY (US $31,600-44,300)	5 (21.7)	4 (17.4)	7 (29.2)		
7-10 million JPY (US $44,300-63,300)	2 (8.7)	0 (0)	3 (12.5)		
>10 million JPY (US $63,300)	1 (4.4)	3 (13)	1 (4.2)		
Total PA[a], mean (SD)	25.6 (33.7)	30.4 (46.9)	55.4 (128.1)	0.90 (2, 67)	.41
Stage of change, n (%)				0.01 (4)	1.00
Precontemplation	15 (65.2)	15 (65.2)	16 (66.7)		
Contemplation	5 (21.7)	5 (21.7)	5 (20.8)		
Preparation	3 (13)	3 (13)	3 (12.5)		
SMSE-2[b]					
Intrinsic motivation	11.9 (3.2)	11.4 (3.8)	11.1 (3.7)	0.26 (2, 67)	.77
Integrated regulation	9.4 (3.6)	7.9 (2.7)	8.1 (3.6)	1.27 (2, 67)	.29
Identified regulation	13.7 (3.9)	13.8 (3.7)	13.0 (4.4)	0.27 (2, 67)	.76
Introjected regulation	10.4 (3.9)	8.5 (3.8)	8.0 (2.8)	2.97 (2, 67)	.06
External regulation	6.6 (2.2)	5.2 (2.3)	5.5 (3.0)	1.95 (2, 67)	.15
Amotivation	6.9 (2.7)	6.5 (3.4)	6.7 (2.7)	0.10 (2, 67)	.90

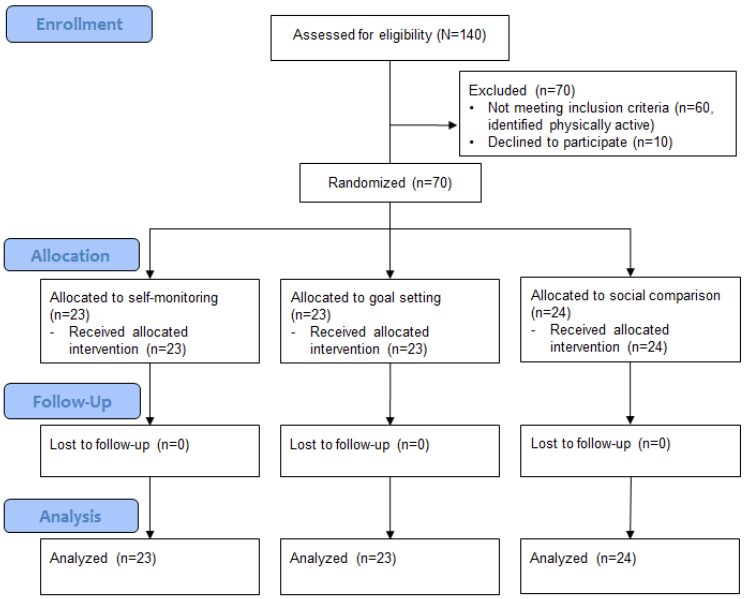
[a]Total PA: total physical activity in metabolic equivalents for hours per week, as assessed using the International Physical Activity Questionnaire—Short Form.

[b]SMSE-2: Self-determined Motivation Scale for Exercise 2.

Compliance and Adherence

All participants completed the pre- and postintervention assessments and received the intended interventions (see Figure 1 for the Consolidated Standards of Reporting Trials flow diagram). The compliance rate was 89% (mean 24.96, SD 5.75 out of 28 days) for the daily online questionnaires. The mean valid-wear time of Fitbit was 26.43 days (94%) out of the entire assessment period (28 days) after the invalid days (when <2000 steps were recorded) were left out. As another measure of adherence to wearing Fitbit, we checked the presence of heart rate signals in the data. This check identified 30 days to be excluded across 70 participants, and the remaining mean 27.57 (SD 1.04; 99%) days were submitted to the sensitivity analysis. No adverse event was recorded.

Figure 1. Consolidated Standards of Reporting Trials flow diagram.



We examined the correlations between compliance and baseline variables (Table S1 in Multimedia Appendix 1). The older participants showed higher compliance (day-to-day assessments; r=0.30) and adherence (Fitbit wear time; r=0.39). Those with children or jobs showed higher adherence (r=0.35 and 0.31, respectively). Individuals with higher baseline levels of integrated regulations had a lower compliance rate (r=–0.35).

Changes From the Pre- to Postintervention Assessments

The results of a 2 × 3 multilevel model (pre- vs postintervention time point; self-monitoring vs goal setting vs social comparison) of self-reported total PA revealed neither a significant main effect of time nor a significant interaction effect between time and group (Table 3). The same multilevel model analyses were repeated for SoC and motivation variables as outcomes, which showed no significant time-group interaction effects. However, significant effects of time were observed for the following secondary outcomes (Table S2 in Multimedia Appendix 1): SoC (dz=0.50, 95% CI 0.25-0.75, P=.006), integrated regulation (dz=0.35, 95% CI 0.11-0.59, P=.03), and external regulation (dz=0.42, 95% CI 0.17-0.66, P=.003). These time effects suggest that participants reported progress to higher stages at the end of the interventions; typically, people who had been in the precontemplation stage progressed toward the contemplation stage (see Table S3 in Multimedia Appendix 1). Similarly, participants experienced increases in integrated regulation (ie, they recognized more personal values and sense of self in PA) and external regulation (ie, they gained more motivation that was controlled by external sources and actions such as pleasing their family), but these effects did not differ between the BCT conditions.

Table 3. Changes in the levels of physical activity and other outcomes from the pre- to postintervention assessment.

Outcome and group[a]	Value, mean (SD)	Standardized within-group change (*dz*)	Standardized mean group difference (vs self-monitoring)	
			Cohen *d*	*P* value
Total PA[b] (METs-h/w)[c]				
Self-monitoring	–4.97 (27.82)	–0.18	__[d]	—
Goal	5.48 (46.49)	0.12	0.28	.33
Social	22.12 (80.80)	0.27	0.45	.15
All groups	7.75 (56.91)	0.14	—	—
SoC[e]				
Self-monitoring	0.48 (0.73)	0.65	—	—
Goal	0.52 (0.90)	0.58	0.05	.86
Social	0.21 (0.78)	0.27	–0.37	.25
All groups	0.40 (0.81)	0.50	—	—
Intrinsic motivation				
Self-monitoring	0.83 (3.26)	0.25	—	—
Goal	0.96 (2.42)	0.40	0.05	.88
Social	–0.58 (3.35)	0.17	–0.44	.12
All groups	0.39 (3.08)	0.13	—	—
Integrated regulation				
Self-monitoring	1.17 (2.85)	0.41	—	—
Goal	1.09 (2.29)	0.47	–0.03	.91
Social	0.38 (2.41)	0.16	–0.31	.28
All groups	0.87 (2.52)	0.35	—	—
Identified regulation				
Self-monitoring	1.39 (3.35)	0.41	—	—
Goal	0.96 (3.02)	0.32	–0.14	.68
Social	–0.13 (4.26)	0.03	–0.40	.15
All groups	0.73 (3.60)	0.20	—	—
Introjected regulation				
Self-monitoring	0.57 (3.87)	0.15	—	—
Goal	0.65 (3.75)	0.17	0.02	.93
Social	1.83 (3.06)	0.60	0.37	.23
All groups	1.03 (3.57)	0.29	—	—
External regulation				
Self-monitoring	1.65 (2.62)	0.63	—	—
Goal	1.48 (1.81)	0.82	–0.08	.82
Social	0.21 (3.15)	0.07	–0.51	.06
All groups	1.10 (2.64)	0.42	—	—
Amotivation				
Self-monitoring	–0.43 (2.59)	0.17	—	—
Goal	–0.57 (2.64)	0.21	–0.05	.87
Social	0.63 (2.60)	0.24	0.42	.17
All groups	–0.11 (2.63)	0.04	—	—

[a]n=23, 23, and 24 for the self-monitoring, goal setting (goal), and social comparison (social) conditions, respectively. Stage of change coded as follows: precontemplation=1, contemplation=2, preparation=3. *P* values for group differences were calculated using multilevel modeling (ie, group-time interaction effects; see Table S2 in Multimedia Appendix 1 for details).

[b]PA: physical activity.

[c]METs-h/w: metabolic equivalents for hours per week

[d]Not available.

[e]SoC: Stage of Change.

Day-Level Outcomes

We performed multilevel modeling with days, groups, and their interactions as predictors. The results showed no significant day-group interactions in predicting Fitbit-assessed step count, identified regulation, or intrinsic motivation (Figure 2; Table 4). However, external motivation was significantly predicted by the interaction between the day and group dummy of self-monitoring versus goal setting. This interaction effect implies that the self-monitoring group experienced a significant increase in external motivation (*b*=0.019, SE=0.008; $t_{68.24}$=2.436, *P*=.02), whereas the goal setting group showed no change over time (*b*=–0.003, SE=0.011; $t_{66.51}$=–0.516, *P*=.61). Over the intervention period, the Fitbit-assessed step count remained slightly higher in the goal setting and social comparison conditions than in the self-monitoring condition (Figure 2D), although these main effects of the condition did not reach statistical significance (*P*=.051 and *P*=.06; Table 4). The results were unchanged when heart rate was used to define invalid wearing time (the main effects were not significant; *P*=.053 for goal setting and *P*=.17 for social comparison).

Figure 2. Levels of motivations and Fitbit-assessed step count as a function of day.

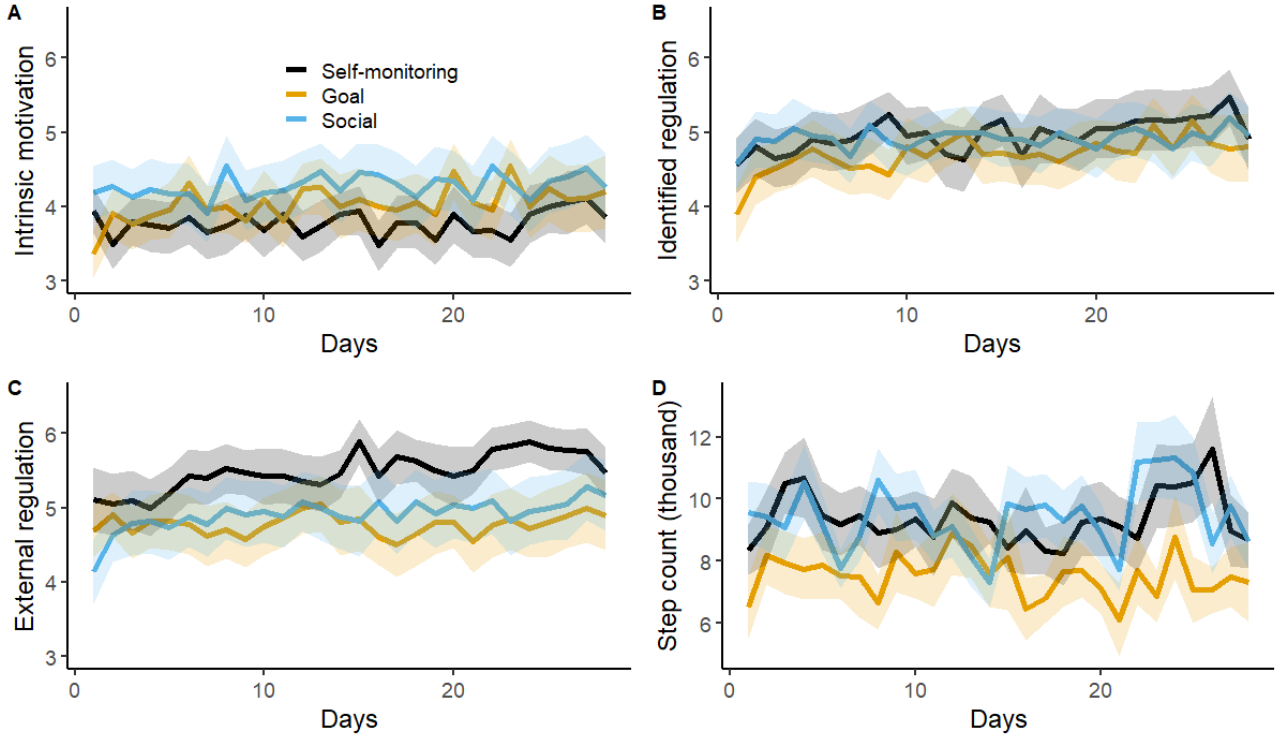


Table 4. Estimates of multilevel models predicting the day-level outcomes[a].

IV[b]	Estimate	SE	2-tailed *t* test (*df*)	*P* value	95% CI
DV[c]: identified regulation					
Time	0.011	0.008	1.307 (67.01)	.20	–0.005 to 0.027
Goal	–0.287	0.499	–0.575 (66.19)	.57	–1.257 to 0.684
Social	0.087	0.493	0.176 (66.13)	.86	–0.873 to 1.047
Time×Goal	0.001	0.012	0.047 (66.43)	.96	–0.022 to 0.023
Time×Social	–0.006	0.011	–0.500 (65.67)	.62	–0.028 to 0.017
DV: intrinsic motivation					
Time	0.011	0.008	1.403 (68.00)	.17	–0.004 to 0.027
Goal	0.175	0.503	0.348 (65.93)	.73	–0.803 to 1.153
Social	0.426	0.498	0.856 (65.87)	.40	–0.542 to 1.394
Time×Goal	–0.003	0.011	–0.236 (67.27)	.81	–0.025 to 0.019
Time×Social	–0.006	0.011	–0.516 (66.51)	.61	–0.027 to 0.016
DV: external regulation					
Time	0.019	0.008	2.436 (68.24)	.02	0.004 to 0.033
Goal	–0.385	0.519	–0.741 (66.25)	.46	–1.394 to 0.625
Social	–0.454	0.514	–0.883 (66.19)	.38	–1.453 to 0.545
Time×Goal	–0.022	0.011	–2.105 (67.47)	.04	–0.043 to –0.002
Time×Social	–0.002	0.010	–0.169 (66.70)	.87	–0.022 to 0.018
DV: Fitbit-assessed step count					
Time	–0.020	0.022	–0.923 (64.93)	.36	–0.063 to 0.022
Goal	0.238	0.120	1.990 (66.59)	.051	0.006 to 0.471
Social	0.221	0.114	1.933 (66.96)	.06	–0.002 to 0.444
Time×Goal	0.026	0.029	0.905 (63.13)	.37	–0.030 to 0.082
Time×Social	0.038	0.028	1.377 (64.72)	.17	–0.016 to 0.093

[a]The step count was log transformed. Goal, Social=dummies representing the goal setting or social comparison conditions vs self-monitoring. A total of 1934 and 1960 observations were used for the motivation outcomes and step count, respectively.

[b]IV: independent variable.

[c]DV: dependent variable.

Discussion

Principal Findings

This study examined the effects of different BCTs implemented in Fitbit on subjective and objective levels of PA as well as motivations for PA. Three BCTs (ie, self-monitoring, goal setting, and social comparison) were offered via the smartwatch for 4 weeks, during which participants lived in their daily lives. Overall, participants showed excellent adherence to the interventions (older participants and those with a child or job showed even better adherence), and there were no dropouts at the end of the interventions. Although we found no significant increase in total PA (as assessed by the IPAQ-SF) or Fitbit-assessed step count, participants reported that the stage progressed typically from precontemplation to contemplation across the 3 BCT conditions. In addition, significant increases were detected in several aspects of motivation (ie, integrated and external regulation), and the day-to-day level of external regulation showed a larger increase in self-monitoring than in goal-setting conditions.

Overall, these results suggest that the effects of Fitbit-delivered BCTs on PA are marginal. There are several possible reasons for the detection of smaller-than-expected effects. First, we used an active control condition (ie, self-monitoring offered by a wearable device), unlike previous trials that compared the presence and absence of an activity tracker [10], which may have mitigated the detectable between-condition effects. Second, changes from the pre- to the postintervention assessment showed extremely large heterogeneity. This heterogeneity can be partly attributed to large individual differences (or SD) in baseline PA levels. Notably, the social comparison group showed an average increase of 22.1 (SD 80.9) METs-h/w, which is regarded as a sufficient increase in PA according to the national health recommendation (23 METs-h/w) [32]. However, a large SD prevented the effect from reaching statistical significance, which is in line with a recent meta-analysis of the heterogeneous effects

of social factors in behavior change interventions [26]. This may indicate that personalization (delivering social comparisons to people who appreciate social factors) is important for maximizing the intervention effect [48]. Third, we did not have a baseline period for the Fitbit assessment; on receipt of the Fitbit, the participants started each intervention immediately. Interestingly, the step count was slightly higher in the goal setting and social comparison groups than in the self-monitoring group over the course of the intervention (P=.051 and P=.06). These effects suggest that participants had already adapted their behavior on the first day of the interventions, which resulted in an overall elevation in step count, albeit with nonsignificant time-group interactions. These minor effects should be interpreted carefully, but it is possible that self-regulatory and social factors contribute a meaning addition to self-monitoring in PA promotion [24].

Our aim was to experimentally investigate the effects of mobile implementations of BCTs, and therefore, we decided to keep each BCT manipulation as simple as possible. However, our implementations might feel too simple compared with the features and functions equipped in commercial PA apps. For example, we set rigid goals for each participant (ie, 8000 or 9500 steps each day), but goal setting theory suggests that goals should be adequate and tailored to the individual skills (ie, doable but challenging) [25,49]. Also, other components could be considered as effective additions in real-life implementation, such as action planning, goal evaluation, and goal reevaluation. Similarly, social comparison could be provided in another form—for example, some participants commented in the postintervention interview that they would be more motivated if the leaderboard displayed the steps of close others (eg, real friends and family members). Several trials have tested the efficacy of collaboration and competition within a gamification intervention for PA, which gives a meaningful context of social comparison and may help enhance motivation to engage in the intervention [50,51].

We detected significant increases in the levels of integrated and external regulations in the postintervention assessment. These increases were not moderated by the BCT conditions, indicating that each group experienced almost equal increases in the 2 types of regulation. Similarly, day-level external regulation increased over the 4-week interventions, which was qualified by the significant time-group interaction. This interaction suggests that participants in the self-monitoring group experienced a larger increase in external regulation than those in the goal setting group. A cross-sectional survey of the Mexican general population [52] found that external regulation was predominant in the precontemplation stage, whereas intrinsic, integrated, and identified regulations increased as the stage progressed (the action and maintenance stage). The detected increase in integrated regulation may be a good sign, but the increase in external regulation may have both positive (eg, people may have found external rewards to start PA and to progress from the precontemplation stage) and negative aspects (eg, this may not directly lead to habit formation). Furthermore, the day-to-day increase in external regulation may suggest that participants in the self-monitoring group felt some pressure and tension regarding PA, as the questionnaire item stated that they felt that they have to exercise [30]. Thus, the significant interaction effect can be interpreted as the goal setting technique preventing participants from feeling obliged to engage in PA.

Limitations

These findings should be interpreted with consideration of several important limitations. First, we based the eligibility assessment on a 1-item SoC questionnaire, which is not an optimal screening tool. Approximately one-third of eligible participants were sufficiently active to exceed the national health recommendation (23 METs-h/w) at baseline. In the postintervention interview, some participants reported that they had not been aware of their activity level at eligibility screening but realized how active they were when responding to each item of the IPAQ-SF. Future research should consider a sufficiently detailed eligibility assessment for PA and preferably use an objective measure, such as a pedometer or accelerometer. Second, it was not possible to strictly control how the participants used each Fitbit feature during the intervention period. For example, we did not prevent participants from setting their own step goals, even in the self-monitoring or social comparison conditions. Although we did not detect any significant violations in the log of Fitbit usage, it might be appropriate to add a system to lock or unlock a particular function in the app and device to reduce potential contamination. Third, we did not conduct a follow-up assessment to test whether the intervention effects were maintained even after the end of the intervention. This decision was made because we could not give away the Fitbit, but it would be valid to leave the Fitbit with participants even after the end of the intervention if we wanted to observe a long-lasting follow-up effect. Fourth, during the study period, motivation was rated or reported each night together with the step count of the day. This assessment timing may have influenced how participants responded to the questionnaires—it is possible that people are more motivated for PA during the daytime. However, we decided not to administer multiple questionnaires (eg, in the morning; at random moments) within a day to reduce the burden on participants and to achieve high compliance.

Conclusions

Nevertheless, our results suggest that activity trackers are useful for increasing readiness and motivation for PA, although significant heterogeneity was identified. The addition of goal setting and social comparison potentially enhances step count, and goal setting may prevent feelings of pressure or tension from engaging in PA. However, to fully detect BCT-specific effects, a proper trial (eg, with a larger sample size, longer intervention and assessment periods, and different populations) is warranted. It is difficult to evaluate the effects of each BCT and their complex combinations (eg, [53]), but future research should expand to other BCTs often used in mHealth practices, such as messaging and rewards.

Acknowledgments

We thank Shingo Terada for his extensive support with data collection. This study was supported by JSPS KAKENHI (23H01051).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary tables.
[DOCX File , 31 KB - mhealth_v12i1e51216_app1.docx]

Multimedia Appendix 2
CONSORT checklist.
[PDF File (Adobe PDF File), 1199 KB - mhealth_v12i1e51216_app2.pdf]

References

1. Hallal PC, Andersen LB, Bull FC, Guthold R, Haskell W, Ekelund U, Lancet Physical Activity Series Working Group. Global physical activity levels: surveillance progress, pitfalls, and prospects. Lancet 2012;380(9838):247-257. [doi: 10.1016/S0140-6736(12)60646-1] [Medline: 22818937]
2. Rhodes RE, Janssen I, Bredin SSD, Warburton DER, Bauman A. Physical activity: health impact, prevalence, correlates and interventions. Psychol Health 2017;32(8):942-975. [doi: 10.1080/08870446.2017.1325486] [Medline: 28554222]
3. Bradley T, Campbell E, Dray J, Bartlem K, Wye P, Hanly G, et al. Systematic review of lifestyle interventions to improve weight, physical activity and diet among people with a mental health condition. Syst Rev 2022;11(1):198 [FREE Full text] [doi: 10.1186/s13643-022-02067-3] [Medline: 36085250]
4. Howlett N, Trivedi D, Troop N, Chater A. Are physical activity interventions for healthy inactive adults effective in promoting behavior change and maintenance, and which behavior change techniques are effective? a systematic review and meta-analysis. Transl Behav Med 2019;9(1):147-157 [FREE Full text] [doi: 10.1093/tbm/iby010] [Medline: 29506209]
5. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The Behavior Change Technique Taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013;46(1):81-95 [FREE Full text] [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
6. Dor-Haim H, Katzburg S, Revach P, Levine H, Barak S. The impact of COVID-19 lockdown on physical activity and weight gain among active adult population in Israel: a cross-sectional study. BMC Public Health 2021;21(1):1521 [FREE Full text] [doi: 10.1186/s12889-021-11523-z] [Medline: 34362319]
7. Giuntella O, Hyde K, Saccardo S, Sadoff S. Lifestyle and mental health disruptions during COVID-19. Proc Natl Acad Sci U S A 2021;118(9):e2016632118 [FREE Full text] [doi: 10.1073/pnas.2016632118] [Medline: 33571107]
8. Chaddha A, Jackson EA, Richardson CR, Franklin BA. Technology to help promote physical activity. Am J Cardiol 2017;119(1):149-152. [doi: 10.1016/j.amjcard.2016.09.025] [Medline: 27889045]
9. Lyons EJ, Lewis ZH, Mayrsohn BG, Rowland JL. Behavior change techniques implemented in electronic lifestyle activity monitors: a systematic content analysis. J Med Internet Res 2014;16(8):e192 [FREE Full text] [doi: 10.2196/jmir.3469] [Medline: 25131661]
10. Ferguson T, Olds T, Curtis R, Blake H, Crozier AJ, Dankiw K, et al. Effectiveness of wearable activity trackers to increase physical activity and improve health: a systematic review of systematic reviews and meta-analyses. Lancet Digit Health 2022;4(8):e615-e626 [FREE Full text] [doi: 10.1016/S2589-7500(22)00111-X] [Medline: 35868813]
11. Laranjo L, Ding D, Heleno B, Kocaballi B, Quiroz JC, Tong HL, et al. Do smartphone applications and activity trackers increase physical activity in adults? systematic review, meta-analysis and metaregression. Br J Sports Med 2021;55(8):422-432. [doi: 10.1136/bjsports-2020-102892] [Medline: 33355160]
12. Larsen RT, Wagner V, Korfitsen CB, Keller C, Juhl CB, Langberg H, et al. Effectiveness of physical activity monitors in adults: systematic review and meta-analysis. BMJ 2022;376:e068047 [FREE Full text] [doi: 10.1136/bmj-2021-068047] [Medline: 35082116]
13. Ridgers ND, Timperio A, Brown H, Ball K, Macfarlane S, Lai SK, et al. Wearable activity tracker use among Australian adolescents: usability and acceptability study. JMIR Mhealth Uhealth 2018;6(4):e86 [FREE Full text] [doi: 10.2196/mhealth.9199] [Medline: 29643054]
14. Brickwood KJ, Williams AD, Watson G, O'Brien J. Older adults' experiences of using a wearable activity tracker with health professional feedback over a 12-month randomised controlled trial. Digit Health 2020;6:2055207620921678 [FREE Full text] [doi: 10.1177/2055207620921678] [Medline: 32426152]
15. Mercer K, Giangregorio L, Schneider E, Chilana P, Li M, Grindrod K. Acceptance of commercially available wearable activity trackers among adults aged over 50 and with chronic illness: a mixed-methods evaluation. JMIR Mhealth Uhealth 2016;4(1):e7 [FREE Full text] [doi: 10.2196/mhealth.4225] [Medline: 26818775]

16. Rossi A, Frechette L, Miller D, Miller E, Friel C, van Arsdale A, et al. Acceptability and feasibility of a fitbit physical activity monitor for endometrial cancer survivors. Gynecol Oncol 2018;149(3):470-475. [doi: 10.1016/j.ygyno.2018.04.560] [Medline: 29692337]
17. Lindgren T, Hooper J, Fukuoka Y. Perceptions and experiences of women participating in a digital technology-based physical activity intervention (the mPED trial): qualitative study. JMIR Public Health Surveill 2019;5(4):e13570 [FREE Full text] [doi: 10.2196/13570] [Medline: 31859677]
18. Pelletier C, Gagnon MP, Alméras N, Després JP, Poirier P, Tremblay A, et al. Using an activity tracker to increase motivation for physical activity in patients with type 2 diabetes in primary care: a randomized pilot trial. Mhealth 2021;7:59 [FREE Full text] [doi: 10.21037/mhealth-20-154] [Medline: 34805390]
19. Peterson NE, Bate DA, Macintosh JL, Trujillo Tanner C. Wearable activity trackers that motivate women to increase physical activity: mixed methods study. JMIR Form Res 2023;7:e48704 [FREE Full text] [doi: 10.2196/48704] [Medline: 38096000]
20. St Fleur RG, St George SM, Leite R, Kobayashi M, Agosto Y, Jake-Schoffman DE. Use of fitbit devices in physical activity intervention studies across the life course: narrative review. JMIR Mhealth Uhealth 2021;9(5):e23411 [FREE Full text] [doi: 10.2196/23411] [Medline: 34047705]
21. Burke LE, Wang J, Sevick MA. Self-monitoring in weight loss: a systematic review of the literature. J Am Diet Assoc 2011;111(1):92-102 [FREE Full text] [doi: 10.1016/j.jada.2010.10.008] [Medline: 21185970]
22. Dombrowski SU, Sniehotta FF, Avenell A, Johnston M, MacLennan G, Araújo-Soares V. Identifying active ingredients in complex behavioural interventions for obese adults with obesity-related co-morbidities or additional risk factors for co-morbidities: a systematic review. Health Psychology Review 2012;6(1):7-32. [doi: 10.1080/17437199.2010.513298]
23. Bartels SL, van Knippenberg RJM, Dassen FCM, Asaba E, Patomella AH, Malinowsky C, et al. A narrative synthesis systematic review of digital self-monitoring interventions for middle-aged and older adults. Internet Interv 2019;18:100283 [FREE Full text] [doi: 10.1016/j.invent.2019.100283] [Medline: 31890630]
24. Michie S, Abraham C, Whittington C, McAteer J, Gupta S. Effective techniques in healthy eating and physical activity interventions: a meta-regression. Health Psychol 2009;28(6):690-701. [doi: 10.1037/a0016136] [Medline: 19916637]
25. Baretta D, Bondaronek P, Direito A, Steca P. Implementation of the goal-setting components in popular physical activity apps: review and content analysis. Digit Health 2019;5:2055207619862706 [FREE Full text] [doi: 10.1177/2055207619862706] [Medline: 31360535]
26. Tong HL, Laranjo L. The use of social features in mobile health interventions to promote physical activity: a systematic review. NPJ Digit Med 2018;1:43 [FREE Full text] [doi: 10.1038/s41746-018-0051-3] [Medline: 31304323]
27. Oba T, Takano K, Katahira K, Kimura K. Use patterns of smartphone apps and wearable devices supporting physical activity and exercise: large-scale cross-sectional survey. JMIR Mhealth Uhealth 2023;11:e49148 [FREE Full text] [doi: 10.2196/49148] [Medline: 37997790]
28. Mendiola MF, Kalnicki M, Lindenauer S. Valuable features in mobile health apps for patients and consumers: content analysis of apps and user ratings. JMIR Mhealth Uhealth 2015;3(2):e40 [FREE Full text] [doi: 10.2196/mhealth.4283] [Medline: 25972309]
29. Marcus BH, Simkin LR. The stages of exercise behavior. J Sports Med Phys Fitness 1993;33(1):83-88. [Medline: 8350613]
30. Østerlie O, Løhre A, Haugan G. The Situational Motivational Scale (SIMS) in physical education: a validation study among Norwegian adolescents. Cogent Educ 2019;6(1):A.
31. Whitehead AL, Julious SA, Cooper CL, Campbell MJ. Estimating the sample size for a pilot randomised trial to minimise the overall trial sample size for the external pilot and main trial for a continuous outcome variable. Stat Methods Med Res 2016;25(3):1057-1073 [FREE Full text] [doi: 10.1177/0962280215588241] [Medline: 26092476]
32. Physical activity standards for health promotion. Ministry of Health, Labor and Welfare. 2013. URL: https://www.e-healthnet.mhlw.go.jp/information/policy/guidelines_2013.html [accessed 2023-01-23]
33. Oka K. Stages of change for exercise behavior and self-efficacy for exercise among middle-aged adults. Nihon Koshu Eisei Zasshi 2003;50(3):208-215. [Medline: 12704833]
34. Oka K, Hirai K, Tsutsumi T. Psychological factors associated with physical inactivity among middle-aged adults: decisional balance for exercise. Jpn J Behav Med 2003;9(1):23-30. [doi: 10.11331/jjbm.9.23]
35. Reed GR, Velicer WF, Prochaska JO, Rossi JS, Marcus BH. What makes a good staging algorithm: examples from regular exercise. Am J Health Promot 1997;12(1):57-66. [doi: 10.4278/0890-1171-12.1.57] [Medline: 10170436]
36. Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Ainsworth BE, et al. International physical activity questionnaire: 12-country reliability and validity. Med Sci Sports Exerc 2003;35(8):1381-1395. [doi: 10.1249/01.MSS.0000078924.61453.FB] [Medline: 12900694]
37. Murase N, Katsumura T, Ueda C, Inoue S, Shimomitsu T. Validity and reliability of Japanese version of International Physical Activity Questionnaire. J Health Welf Stat 2002;49(11):1-9.
38. Matsumoto H, Taniguchi A, Nishida J. A revised self-determined motivation scale for exercise with integrated regulation inclusion. J Health Psychol Res 2021;34(1):13-22. [doi: 10.11560/jhpr.200608129]
39. Ryan RM, Deci EL. Intrinsic and extrinsic motivations: classic definitions and new directions. Contemp Educ Psychol 2000;25(1):54-67. [doi: 10.1006/ceps.1999.1020] [Medline: 10620381]

40. Exkuma. URL: https://exkuma.com/ [accessed 2023-07-25]
41. Lonsdale C, Sabiston CM, Taylor IM, Ntoumanis N. Measuring student motivation for physical education: examining the psychometric properties of the Perceived Locus of Causality Questionnaire and the Situational Motivation Scale. Psychology of Sport and Exercise 2011;12(3):284-292. [doi: 10.1016/j.psychsport.2010.11.003]
42. Rodrigues F, Cid L, Faustino T, Monteiro D. The situational motivation scale in the exercise context: construct validity, factor structure, and correlational analysis. Curr Psychol 2021;42:4811-4820.
43. Lakens D. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. Front Psychol 2013;4:863 [FREE Full text] [doi: 10.3389/fpsyg.2013.00863] [Medline: 24324449]
44. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using. J. Stat. Soft 2015;67(1):48. [doi: 10.18637/jss.v067.i01]
45. Kuznetsova A, Brockhoff PB, Christensen RHB. Package: tests in linear mixed effects models. J. Stat. Soft 2017;82(13):1-27. [doi: 10.18637/jss.v082.i13]
46. Kaye M. Fitbitr: Interface with the 'Fitbit' API. 2023. URL: https://cran.r-project.org/web/packages/fitbitr/index.html [accessed 2023-06-07]
47. Bassett DR, Toth LP, LaMunion SR, Crouter SE. Step counting: a review of measurement considerations and health-related applications. Sports Med 2017;47(7):1303-1315 [FREE Full text] [doi: 10.1007/s40279-016-0663-1] [Medline: 28005190]
48. Mauch CE, Edney SM, Viana JNM, Gondalia S, Sellak H, Boud SJ, et al. Precision health in behaviour change interventions: a scoping review. Prev Med 2022;163:107192 [FREE Full text] [doi: 10.1016/j.ypmed.2022.107192] [Medline: 35963310]
49. Locke EA, Latham GP. New directions in goal-setting theory. Curr Dir Psychol Sci 2016;15(5):265-268. [doi: 10.1111/j.1467-8721.2006.00449.x]
50. Lewey J, Murphy S, Zhang D, Putt ME, Elovitz MA, Riis V, et al. Effectiveness of a text-based gamification intervention to improve physical activity among postpartum individuals with hypertensive disorders of pregnancy: a randomized clinical trial. JAMA Cardiol 2022;7(6):591-599 [FREE Full text] [doi: 10.1001/jamacardio.2022.0553] [Medline: 35442393]
51. Patel MS, Small DS, Harrison JD, Fortunato MP, Oon AL, Rareshide CAL, et al. Effectiveness of behaviorally designed gamification interventions with social incentives for increasing physical activity among overweight and obese adults across the united states: the STEP UP randomized clinical trial. JAMA Intern Med 2019;179(12):1624-1632 [FREE Full text] [doi: 10.1001/jamainternmed.2019.3505] [Medline: 31498375]
52. Zamarripa J, Castillo I, Baños R, Delgado M, Álvarez O. Motivational regulations across the stages of change for exercise in the general population of monterrey (Mexico). Front Psychol 2018;9:2368 [FREE Full text] [doi: 10.3389/fpsyg.2018.02368] [Medline: 30559693]
53. Schroé H, van Dyck D, de Paepe A, Poppe L, Loh WW, Verloigne M, et al. Which behaviour change techniques are effective to promote physical activity and reduce sedentary behaviour in adults: a factorial randomized trial of an e- and m-health intervention. Int J Behav Nutr Phys Act 2020;17(1):127 [FREE Full text] [doi: 10.1186/s12966-020-01001-x] [Medline: 33028335]

Abbreviations

BCT: behavior change technique
IPAQ-SF: International Physical Activity Questionnaire—Short Form
METs-h/w: hourly metabolic equivalents per week
PA: physical activity
SMD: standardized mean difference
SMSE: Self-determined Motivation Scale for Exercise
SoC: Stage of Change

Edited by L Buis; submitted 26.07.23; peer-reviewed by S Phillips, KL Alfrey; comments to author 08.01.24; revised version received 28.02.24; accepted 22.03.24; published 12.07.24.

Please cite as:
Takano K, Oba T, Katahira K, Kimura K
Deconstructing Fitbit to Specify the Effective Features in Promoting Physical Activity Among Inactive Adults: Pilot Randomized Controlled Trial
JMIR Mhealth Uhealth 2024;12:e51216
URL: https://mhealth.jmir.org/2024/1/e51216
doi:10.2196/51216
PMID:

Original Paper

Mobile App Intervention to Reduce Substance Use, Gambling, and Digital Media Use in Vocational School Students: Exploratory Analysis of the Intervention Arm of a Randomized Controlled Trial

Kristin Grahlher[1], MSc; Matthis Morgenstern[2], Prof Dr, PhD; Benjamin Pietsch[2], PhD; Elena Gomes de Matos[3,4], PhD; Monika Rossa[3], BSc; Kirsten Lochbühler[3], PhD; Anne Daubmann[5], Dipl-Stat; Rainer Thomasius[1], MD, Prof Dr Med; Nicolas Arnaud[1], PhD

[1]German Centre for Addiction Research in Childhood and Adolescence, University Medical Centre Hamburg-Eppendorf, Hamburg, Germany

[2]Institute for Therapy and Health Research, IFT-Nord, Kiel, Germany

[3]IFT Institut für Therapieforschung, Munich, Germany

[4]Institute for Clinical Psychology and Psychotherapy, Charlotte Fresenius University, Munich, Germany

[5]Institute for Medical Biometry and Epidemiology, University Medical Centre Hamburg-Eppendorf, Hamburg, Germany

Corresponding Author:
Nicolas Arnaud, PhD
German Centre for Addiction Research in Childhood and Adolescence
University Medical Centre Hamburg-Eppendorf
Martinistr 52
Hamburg, 20246
Germany
Phone: 49 40741059307
Fax: 49 40741056571
Email: n.arnaud@uke.de

Abstract

Background: During adolescence, substance use and digital media exposure usually peak and can become major health risks. Prevention activities are mainly implemented in the regular school setting, and youth outside this system are not reached. A mobile app (“Meine Zeit ohne”) has been developed specifically for vocational students and encourages participants to voluntarily reduce or abstain from a self-chosen addictive behavior including the use of a substance, gambling, or a media-related habit such as gaming or social media use for 2 weeks. Results from a randomized study indicate a significant impact on health-promoting behavior change after using the app. This exploratory study focuses on the intervention arm of this study, focusing on acceptance and differential effectiveness.

Objective: The aims of this study were (1) to examine the characteristics of participants who used the app, (2) to explore the effectiveness of the mobile intervention depending on how the app was used and depending on participants’ characteristics, and (3) to study how variations in app use were related to participants’ baseline characteristics.

Methods: Log data from study participants in the intervention group were analyzed including the frequency of app use (in days), selection of a specific challenge, and personal relevance (ie, the user was above a predefined risk score for a certain addictive behavior) of challenge selection (“congruent use”: eg, a smoker selected a challenge related to reducing or quitting smoking). Dichotomous outcomes (change vs no change) referred to past-month substance use, gambling, and media-related behaviors. The relationship between these variables was analyzed using binary, multilevel, mixed-effects logistic regression models.

Results: The intervention group consisted of 2367 vocational students, and 1458 (61.6%; mean age 19.0, SD 3.5 years; 830/1458, 56.9% male) of them provided full data. Of these 1458 students, 894 (61.3%) started a challenge and could be included in the analysis (mean 18.7, SD 3.5 years; 363/894, 40.6% female). Of these 894 students, 466 (52.1%) were considered frequent app users with more than 4 days of active use over the 2-week period. The challenge area most often chosen in the analyzed sample was related to social media use (332/894, 37.1%). A total of 407 (45.5%) of the 894 students selected a challenge in a behavioral domain of personal relevance. The effects of app use on outcomes were higher when the area of individual challenge choice was equal to the area of behavior change, challenge choice was related to a behavior of personal relevance, and the individual risk of engaging in different addictive behaviors was high.

Conclusions: The domain-specific effectiveness of the program was confirmed with no spillover between behavioral domains. Effectiveness appeared to be dependent on app use and users' characteristics.

Trial Registration: German Clinical Trials Register DRKS00023788; https://tinyurl.com/4pzpjkmj

International Registered Report Identifier (IRRID): RR2-10.1186/s13063-022-06231-x

(JMIR Mhealth Uhealth 2024;12:e51307) doi:10.2196/51307

KEYWORDS

prevention; vocational students; adolescents; mobile intervention; voluntary commitment; substance use; internet-related problems; mobile phone; adolescent; youths; student; students; use

Introduction

Prevalence and Prevention of Substance Use in Vocational Schools

Substance use and other addictive behaviors such as gambling, screen-related gaming, and excessive use of digital media are widespread among adolescents and young adults across different cultures [1]. In Germany, the prevalence of substance-related and web-related addictive disorders peaks during adolescence [2-5]. Research suggests that vocational students are particularly vulnerable to substance use and addictive behaviors compared with age-matched peers in the general population [6].

Schools are an important setting for facilitating prevention programs because a large number of young people can be reached simultaneously [7]. However, vocational schools have been largely neglected concerning empirically evaluated prevention activities in Germany [8]. Approximately 50% of all graduates from conventional schools in Germany enter vocational training [9] and are thus not reached by available programs during the critical time of adolescence. This seems to be a serious limitation for national public health goals. Established prevention programs cannot be implemented in the vocational school setting without adaptation, as this setting differs from the normal school setting due to the combination of practical training in a company and vocational education at school. Obstacles to implementation might be limited class cohesion, time, and teacher involvement (see reference [9] for a detailed description of the vocational school setting in Germany).

Mobile Interventions

Digital interventions and mobile technologies have the potential to overcome these barriers, especially with a vast majority of young people owning smartphones [10]. Mobile prevention programs offer a location- and time-independent as well as cost-effective way to deliver health-promoting content efficiently and in doses that can be easily integrated into daily life [11-13]. However, the participation rates (ie, the reach) of such mobile prevention programs vary [14-17]. Furthermore, in a recent review, Diestelkamp et al [18] report that despite growing evidence for the effectiveness of mobile substance use interventions, it is largely unclear which specific elements can be effectively implemented using mobile technology and how the extent to which individuals use this content may be associated with different outcomes.

Studies across different outcomes and delivery modes imply that multiple factors can influence the use and outcomes of digital interventions and that usage metrics vary across different types of interventions [19]. Previous research has shown that intensity of use (eg, number of log-ins) and active user engagement (eg, module completion) can influence intervention outcomes [19-21]. For example, a recent feasibility study investigating user engagement with a mobile smoke-free app showed a positive association between increased app use and cigarette abstinence [22]. Furthermore, a study focusing on a mobile intervention for heavy drinking and smoking among college students found that receiving more modules of the intervention was significantly associated with a lower likelihood of any drinking during the 14-day assessment period and a significant reduction in smoking at 1-month follow-up [23]. Thus, on a general level, usage patterns of digital interventions appear to play an important role in their effectiveness. However, it is still very unclear which specific usage metrics have an impact and how large the potential impact is, depending on the design of the intervention, its objectives, and its target population. It is also not clear how users' background characteristics, such as sociodemographic factors, psychological characteristics, previous experiences with technologies, and the level of health symptoms, may influence intervention use and positive health outcomes.

Meine Zeit ohne Intervention

In a recent randomized study [24,25], we transferred the school-based prevention approach of a voluntary commitment into an app-based program for vocational students ("Meine Zeit ohne" [MZo]; IFT-Nord Institut fuer Therapie- und Gesundheitsforschung GmbH). The app (see Multimedia Appendix 1 for app screenshots) was developed for implementation in vocational schools. In the trial, students in the intervention group (IG) received access to the MZo app, which was available for download at the Google Play Store and the Apple App Store (described in detail by Pietsch et al [25]). It includes an introduction to the overall theme of habits and on potential risky health behaviors concerning substance and media use as well as gambling. The app features a short explanation video demonstrating the use and goals of the app. Participants were invited to select 1 behavioral goal for the next 2 weeks: complete abstinence or substantial reduction regarding their use of cannabis, smoking (cigarettes or e-products), alcohol, digital social media use, gaming, gambling, or "another habit." For the options of cannabis, alcohol, and gambling, only an abstinence goal could be selected. After choosing their behavioral goal and starting the challenge, participants received

daily push notifications at fixed times (in the morning and at noon) to remind them to rate their subjective confidence in reaching their selected behavioral goal for the next 24 hours and to enter whether they succeeded in achieving their goal on the previous day. Following the last rating after 2 weeks, students could download and share a certificate regardless of their challenge results. The results suggest that this app-based intervention is feasible and effective in a vocational school setting [25]. Vocational students in the intervention arm classes were significantly (odds ratio [OR] 1.24, 95% CI 1.05-1.46; *P*=.01) more likely to report health-promoting changes such as reductions in substance use behavior and gambling as well as daily screen time for gaming or social media 30 days after the end of the app challenge.

As the effects on health-promoting change for a specific area were larger when students had chosen that particular challenge area, it seems that variations in *how* the MZo app was used (eg, the frequency of use and choice of challenge) and personal characteristics of users may have an impact on the intervention effects. In addition, as suggested by 2 recent studies, it may be interesting to explore whether the effects of a specific behavior reduction or abstinence are limited to that specific behavior or have positive ("spillover") effects on other behaviors [26,27].

This paper presents an exploratory analysis, which focuses only on the IG and investigates who could be reached by the app and how the mobile intervention was used. The aims of this study were (1) to examine the characteristics of participants who used the app, (2) to explore the effectiveness of the mobile intervention depending on how the app was used and depending on individual participants' characteristics, and (3) to study how variations in app use were related to participants' baseline characteristics.

Methods

Ethical Considerations

Approval for the study was obtained from the ethics committee of the Center for Psychosocial Medicine at the University Medical Center Hamburg Eppendorf (approval number: LPEK-0121) and the responsible school authorities at each study site (Center for Education Monitoring and Quality Development at schools in Hamburg, Institut für Bildungsmonitoring und Qualitätsentwicklung; the Center for Prevention at the Institute for Quality Development at Schools in Schleswig-Holstein, Institut für Qualitätsentwicklung an Schulen Schleswig-Holstein; and the Bavarian State Ministry for Education and Cultural Affairs) prior to data collection. The study was conducted in accordance with the CONSORT (Consolidated Standards of Reporting Trials) guidelines and complied with the principles laid down in the Declaration of Helsinki [28]. It was registered in the German Clinical Trials Register public database (DRKS00023788), and a detailed study protocol was published [24]. Written informed consent was obtained from all participants prior to study enrollment. Participants did not receive any financial compensation for participating in the study. All substantial protocol deviations or modifications were communicated to the ethics committee and German Clinical Trials Register.

Study Design and Participants

The MZo evaluation study was a 2-arm multicenter, cluster-randomized, waitlist-controlled trial (randomized controlled trial [RCT]) to test the effectiveness of a low-threshold, mobile app–based intervention in a sample of vocational students in Germany. The MZo app consisted of a 2-week intervention period followed by a 30-day follow-up period. Data were collected through web-based questionnaires and app use logs, starting in March 2021 and ending in April 2022. The RCT was conducted in 3 study centers in Germany—Munich, Kiel, and Hamburg.

At these sites, schools were consecutively recruited with the support of local school authorities or directly using digital and printed information materials, school conferences, etc. After initial agreement from school principalities, research staff, social school workers, or principals contacted the teachers at the participating school and informed them about the study's aims and procedures. In the participating vocational schools, teachers and school principals selected classes to participate in the study. Participation in the challenge and the survey was voluntary. Since the study was conducted during the COVID-19 pandemic, enrollment, data assessment, and the introduction of the app could not take place in the classroom as initially planned but were done on the web if necessary. Randomization was done at the class level and performed before the baseline assessment. To achieve this aim, 2 classes were each paired into similar dyads based on the three class characteristics: (1) frequency of in-school education, (2) educational area (eg, technical or IT, services, and trade), and (3) year of training (eg, first, second, and third) [25]. Paired classes were then randomized into the IG or control group. Researchers had no access to the app data during the intervention.

Measures

Overview

This analysis used baseline measurements, follow-up measures (all self-reported), and mobile app use log data from study participants in the IG. Study variables are described in detail below. Sociodemographic data were collected at baseline via web-based questionnaires and included sex or gender, age, migration status, income, and highest level of education.

App usage metrics were collected through usage log files recorded by the mobile phone app on an external server. Usage log files included, for each participating user, the registration code to match the app data with the data collected in the questionnaires, information about the challenge (eg, challenge choice, decision to choose a reduction or abstinence goal, and reduction goal in units or time), confidence, as well as success ratings in meeting their challenge goal for each day. The log files were retrieved from the server at the end of the intervention period. Usage metrics could be divided into 3 categories, which are summarized in Textbox 1. The definitions of the variables frequent use and congruent use are described in detail below.

Textbox 1. Description of usage metrics.

- Selected challenge: Users selecting a specific challenge (cannabis, smoking or vaping, alcohol, gaming, social media, gambling, and "other").
- Frequent use: Users using the app frequently.
- Congruent use: Users selecting a challenge congruent to his or her personal relevance.

Frequent Use

The number of days of use was determined by the total number of days on which participants made a confidence rating entry in the app. The confidence rating variable was chosen because it had to be rated on the respective day. Thus, this variable determined whether or not participants were actively using the app on a certain day during the challenge. As the number of days of active use showed a left-skewed distribution, further analysis was done using the binary variable frequent use based on a median split (eg, above or below 4 days of active use). Thus, nonfrequent use was defined when users actively used the app for up to 4 days, and frequent use was defined when users actively used the app for more than 4 days.

Congruent Use

The congruent use variable identified users who selected a challenge that met their personal relevance. Personal relevance was defined for each behavioral area (tobacco, e-cigarettes, alcohol, cannabis, gambling, gaming, and digital media) when a predefined threshold for the frequency of that behavior was exceeded at baseline assessment. For cannabis use, personal relevance was defined as a positive 30-day prevalence. For alcohol, cigarettes, and e-products, personal relevance was defined as consumption at least once a week (ie, on more than 4 days) in the past 30 days. For social media, gaming, and gambling, personal relevance was defined using the median split. Thus, for social media, personal relevance was determined if social media use exceeded 4 hours per day, and for gaming, if gaming days per month exceeded 3 days. For gambling, personal relevance was defined as at least 1 gambling day per month. Congruent use was determined when personal relevance for a particular substance-related behavior or behavioral addictive–related behavior was met and a participant actually chose a challenge for the same behavior. For students without a personally relevant behavior (ie, no behavior exceeding these thresholds), congruent use was set when the challenge "other" was selected. For example, congruent use was set when someone who smoked more than 4 days a month selected the smoking challenge.

Personal Relevance Score

To further explore how the level of personal risk regarding substance-related and addictive behavior influenced health-promoting change, a continuous variable named personal relevance score was created based on the personal relevance variables for tobacco, e-cigarettes, alcohol, cannabis, gambling, and gaming as well as digital media (for definitions of thresholds, see the *Congruent Use* section). The intention was to quantify the number of personal relevant habits. The personal relevance score was the sum of personal relevant behaviors for each participant; for example, when a participant smoked cigarettes at least once a week and drank alcohol at least once a week, his or her personal relevance score was set to 2. The values of the personal relevance score ranged between 0 (no area of personal relevance) and 7 (7 areas of personal relevance).

Outcome Variables: Health-Promoting Change Variables

For this analysis, the dichotomous health-promoting change variables and a variable called general adverse health behavior improvement (GAHBI) were created according to Pietsch et al [25]. For alcohol, gambling, and cannabis use, a health-promoting change was defined as a positive 30-day prevalence at baseline compared with a zero 30-day prevalence at follow-up. For cigarettes and e-products, change was defined as a reduction of at least 50% in the monthly number of cigarettes or e-product units from baseline to follow-up. For social media and gaming, change was defined as a reduction of at least 20 minutes in daily screen time. The GAHBI was coded as health-promoting change if there was at least 1 change in the abovementioned areas.

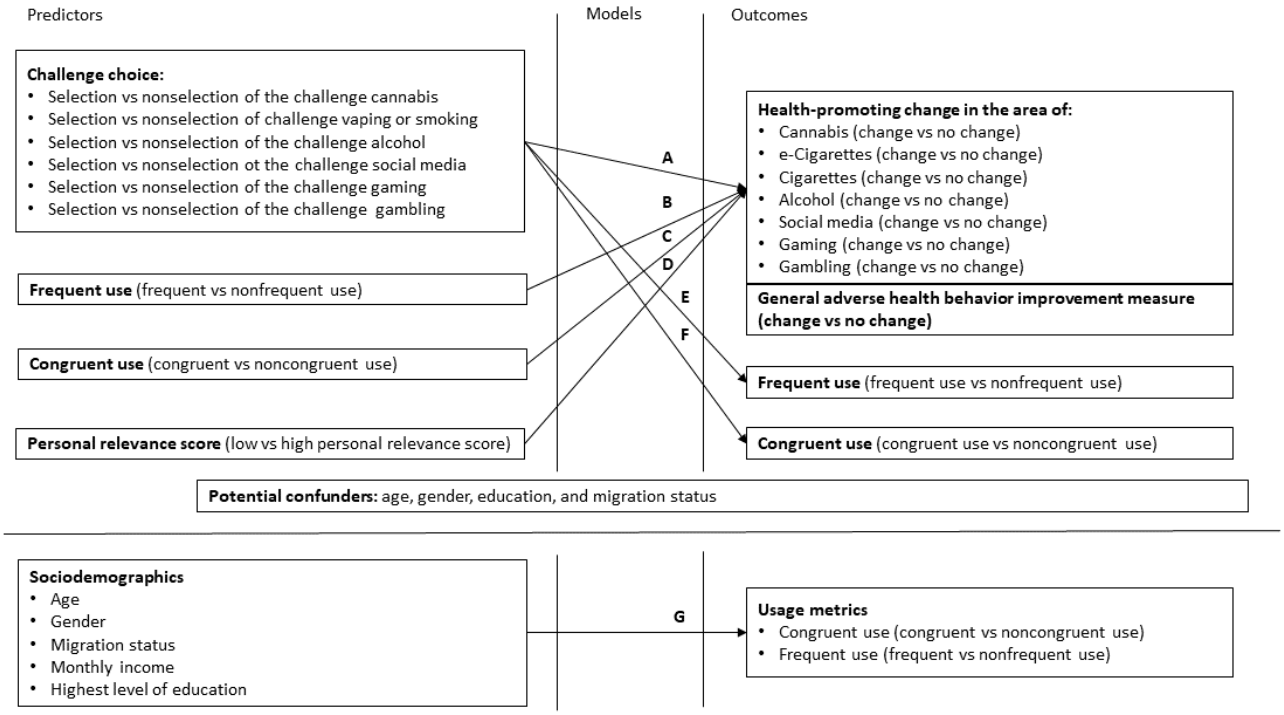
Statistical Analysis

Statistical analysis was performed using SAS for Windows (version 9.4 TS Level 1M7; SAS Institute Inc). Statistical significance was set at $P \leq .05$ (but exact P values are reported). Descriptive statistics were calculated using SPSS Statistics Release (version 26; IBM Corp). Differences in baseline characteristics between app users and nonapp users were assessed using Bonferroni-corrected 2-tailed t tests and chi-square tests. The relationship between the predictors (challenge choice, frequent use, congruent use, and personal relevance score) and the outcome variables (change in all dichotomous health-promoting change variables as well as the GAHBI [change vs no change]) was analyzed using binary multilevel mixed-effects logistic regression models (SAS Glimmix), accounting for the clustered structure of the data with clustering occurring within federal states, schools, and classes (see models A-D in Figure 1). For these results, ORs with corresponding 95% CIs are reported. Age, sex or gender, education, and migration status as potential confounders were included as fixed effects to the models. For the models for alcohol, cannabis, and gambling, the number of included confounders had to be reduced in order to achieve convergence. For the models for alcohol and cannabis, age and migration status were added, and for the model for gambling, migration status only.

The relationship between the usage metric of congruent use (yes vs no) and the selected challenge as well as frequent use was analyzed using binary multilevel mixed-effects logistic regression models (see models E and F in Figure 1). This was also done to analyze the relationship between the usage metric of frequent use and the selected challenge as well as congruent use. ORs, 95% CI, and P values were calculated and reported for all outcomes. As the usage metrics were important variables

in this study, they were further tested for associations with the sociodemographic baseline variables age, gender, migration status, monthly income, and highest level of education using multilevel mixed-effects logistic regression models (see model G in Figure 1).

Figure 1. Main analyses.

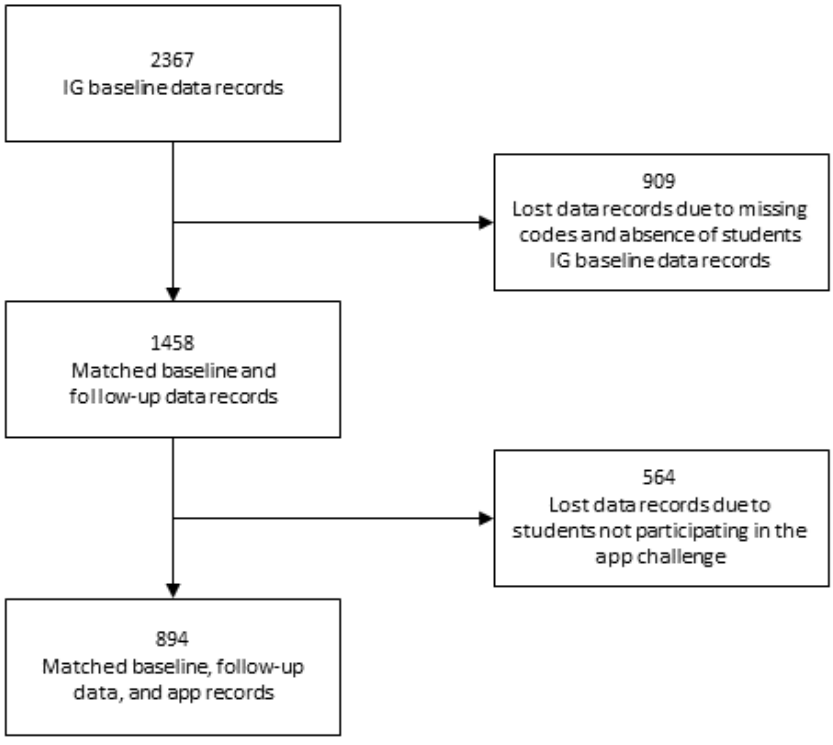


Results

Overview

Altogether, 2367 (51.6%) of 4591 students were randomized to the IG. The baseline and follow-up data of 1458 (61.6%) of the 2367 students in the IG could be matched by the corresponding codes of participants in the baseline and follow-up assessments. A total of 909 students were lost to follow-up due to absence or missing survey codes. Of these 1458 students, 894 (61.3%) started a challenge. Thus, 894 data records with baseline, follow-up, and app data could be included in the analysis (Figure 2).

Figure 2. Participant flowchart. IG: intervention group.



Sample Characteristics

The included sample (n=2367) consisted of 1308 (55.3%) male participants, 1040 (43.9%) female participants, and 19 (0.8%) participants who did not identify as binary. Their mean age at baseline was 19.2 (SD 4.2) years, and 786 (33.2%) reported a migrant background. Reported 30-day prevalence rates were 817 (34.5%) for smoking, 1536 (64.9%) for alcohol use, 294 (12.4%) for gambling, and 379 (16%) for cannabis use. In addition, IG students used social media for an average of 222.2 (SD 185.5) minutes per day and reported playing video games for an average of 99.6 (SD 140.1) minutes per day. The analysis revealed several differences between app users (students who used the app at least 1 time) and nonapp users (students who did not use the app at all) at the sociodemographic level. Students who used the app on average had a younger age and were more likely to have no migrant background, to have a medium monthly income, to have a middle school diploma, and to have vocational training in commerce, industry, and technology (Table 1).

Table 1. Description of the overall IG[a] sample, the IG sample with matched baseline and follow-up data records, and differences between app users and nonapp users in the matched IG sample assessed with Bonferroni-corrected 2-tailed t tests and chi-square tests.

Sociodemographics	Overall IG (n=2367)	Matched IG (n=1458)	App users (n=894)	Nonapp users (n=564)	*P* value
Age (years), mean (SD)	19.2 (4.2)	19.0 (3.5)	18.7 (3.5)	19.4 (3.6)	<.001
Sex, n (%)					
Male	1308 (55.3)	830 (56.9)	525 (58.7)	305 (54.9)	—[b]
Female	1040 (43.9)	618 (42.4)	363 (40.6)	255 (45.2)	.22
Other	19 (0.8)	10 (0.7)	6 (0.7)	4 (0.7)	—
Migration background, n (%)					
No	1581 (66.8)	1053 (72.2)	673 (75.3)	380 (67.4)	—
Yes	786 (33.2)	405 (27.8)	221 (24.7)	184 (32.6)	.001
Monthly income (€)[c], n (%)					
>1000	345 (14.6)	189 (13)	120 (13.4)	69 (12.2)	—
600-999	1184 (50)	790 (54.2)	506 (56.6)	284 (50.4)	.01
<600	838 (35.4)	479 (32.9)	268 (30)	211 (37.4)	—
Highest education, n (%)					
High school diploma	473 (20)	292 (20)	154 (17.2)	138 (24.5)	—
Middle school diploma	1247 (52.7)	837 (57.4)	578 (64.7)	259 (45.9)	<.001
Below middle school diploma	647 (27.3)	329 (22.6)	162 (18.1)	167 (29.6)	—
Educational area, n (%)					
Vocational preparation	120 (5.1)	46 (3.2)	25 (2.8)	21 (3.7)	—
Commerce, industry, and technology	1307 (55.2)	886 (60.8)	608 (68)	278 (49.3)	—
Economics and management	554 (23.4)	290 (19.9)	127 (14.2)	163 (28.9)	<.001
General school-based education	386 (16.3)	236 (16.2)	134 (15)	102 (18.1)	—
Number of personal relevant behaviors, mean (SD)	2.11 (1.4)	2.07 (1.3)	2.08 (1.3)	2.06 (1.3)	.74

[a]IG: intervention group.

[b]Not applicable.

[c]As of June 12, 2024, a currency exchange rate of €1=US $1.075 is applicable.

Usage Metrics and Personal Relevance Score

A total of 894 (61.3%) of the 1458 students in the IG sample with matched data started the challenge (Figure 2). The mean number of days of active use was 5.62 (SD 4.76). With up to 4 days of active use, 417 (46.6%) of the 894 students using the app were nonfrequent users, 466 (52.1%) were frequent users with more than 4 days of active use, and 11 (1.2%) selected a challenge but had no days of active use. A total of 85 (9.5%) actively used the app throughout the entire intervention period of 14 days.

The challenge area most often chosen in the matched sample (n=894) was social media (n=332, 37.1%), followed by “other challenge” (n=159, 17.8%), alcohol (n=128, 14.3%), smoking or vaping (n=127, 14.2%), gaming (n=79, 8.8%), and cannabis (n=38, 4.3%). The least chosen challenge was gambling with 31 (4.2%) students. A total of 407 (45.5%) students chose a challenge congruent with their personal relevance. Of these 407

students, 126 (31%) selected a challenge in the category social media, 84 (20.6%) in the category smoking or vaping, 79 (19.4%) in the category alcohol, 61 (15%) in the category gaming, 23 (5.7%) in the category cannabis, 19 (4.7%) in the category "other challenge," and 15 (3.7%) in the category gambling.

The mean number of personal relevant behaviors was 1.88 (SD 1.32). Of all 894 students, 296 (33.1%) had a personal relevance score of 1, followed by 245 (27.4%) with a personal relevance score of 2, 245 (27.4%) with a personal relevance score above 2, and 108 (12.1%) with a personal relevance score of 0.

Association Between Challenge Choice and Health-Promoting Change or Congruent Use

ORs for health-promoting change in some areas were associated with the choice of challenge, particularly when the area of the challenge and health-promoting change were the same (eg, smoking or vaping challenge and health-promoting change in smoking; see model A in Figure 1). The results are displayed in Table 2. The ORs for health-promoting change increased for cigarettes (OR 3.51, 95% CI 2.14-5.76; *P*<.001) and e-cigarettes (OR 2.87, 95% CI 1.69-4.85; *P*<.001) for users who chose the smoking or vaping challenge. For users choosing the social media challenge, the ORs for a health-promoting change for social media increased (OR 1.66, 95% CI 1.25-2.20; *P*=.001) but decreased for a health-promoting change for smoking (cigarettes: OR 0.55, 95% CI 0.34-0.89; *P*=.001 and e-cigarettes: OR 0.56, 95% CI 0.33-0.92; *P*=.02). ORs for health-promoting change also increased for gaming (OR 1.77, 95% CI 1.09-2.89; *P*=.02) for users who chose the respective challenge. For users who chose the challenge "other challenge," ORs for health-promoting change increased for alcohol (OR 2.15, 95% CI 1.13-4.09; *P*=.02) but decreased for GAHBI (OR 0.63, 95% CI 0.42-0.96; *P*=.03) and social media (OR 0.66, 95% CI 0.45-0.97; *P*=.03). ORs were not reported when the number of participants with a health-promoting change for a particular model was less than 5.

Furthermore, the likelihood for the choice of a congruent challenge (see model F in Figure 1) was higher for users who chose the challenges smoking or vaping (OR 2.74, 95% CI 1.80-4.16; *P*<.001), alcohol (OR 2.17, 95% CI 1.45-3.26; *P*<.001), and gaming (OR 4.57, 95% CI 2.58-8.10; *P*<.001) and lower for users who chose the challenges social media (OR 0.57, 95% CI 0.34-0.77; *P*=.001) and "other challenge" (OR 0.10, 95% CI 0.06-0.17; *P*<.001).

Table 2. ORs[a] with a 95% CI for health-promoting change for students who chose a challenge corresponding to the health-promoting change area (see model A in Figure 1).

Health-promoting change	Participants, n	Events, n	OR (95% CI)	*P* value
Cigarettes	894	32	3.51 (2.14-5.76)	<.001
e-Cigarettes	894	26	2.87 (1.69-4.85)	<.001
Social media	894	185	1.66 (1.25-2.20)	<.001
Gaming	894	34	1.77 (1.09-2.89)	.02
Alcohol	894	9	1.08 (0.53-2.21)	.84
Cannabis	—[b]	—	—	—
Gambling	—	—	—	—

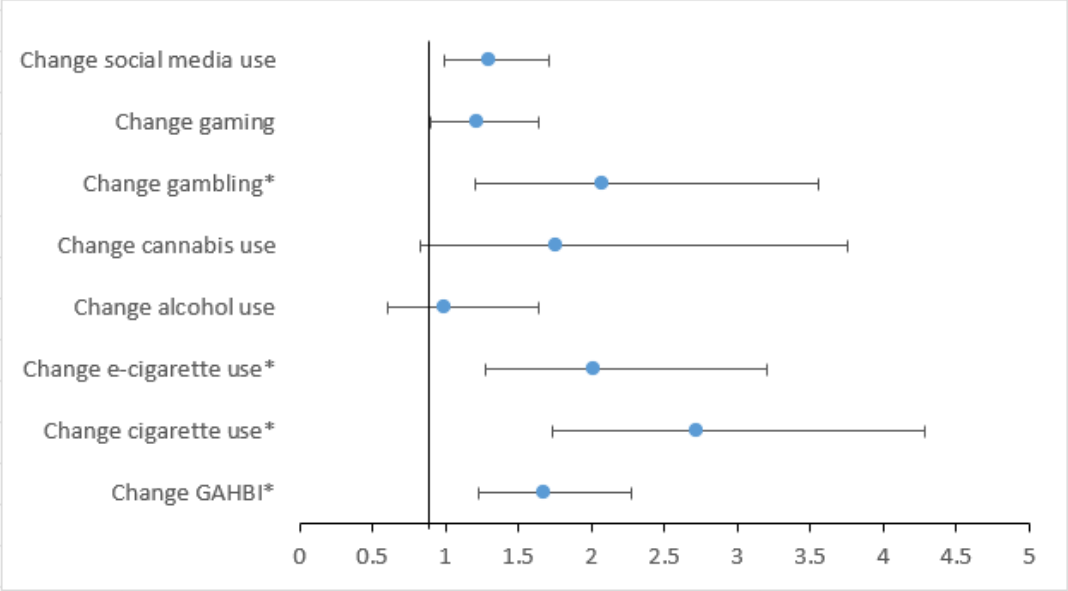
[a]OR: odds ratio.

[b]Not available.

Association Between Congruent Use or Frequent Use and Health-Promoting Change Variables

Congruent use was associated with increased odds of health-promoting change (see model C in Figure 1) for cigarettes (OR 2.72, 95% CI 1.73-4.29; *P*<.001), e-cigarettes (OR 2.02, 95% CI 1.28-3.20; *P*=.003), gambling (OR 2.07, 95% CI 1.21-3.55; *P*=.008), and GAHBI (OR 1.67, 95% CI 1.23-2.27; *P*=.001). Thus, students with a personal relevance score for cigarettes who chose cigarettes as their challenge (congruent use) were 2.72 times more likely to report a health-promoting change in cigarettes than students with a personal relevance score for cigarettes who did not choose smoking but another challenge (incongruent use). The results are displayed in Figure 3. Frequent use showed lower odds for health-promoting change (see model B in Figure 1) for gambling (OR 0.47, 95% CI 0.26-0.86; *P*=.009; Multimedia Appendix 2).

Figure 3. Health-promoting change for students choosing a congruent challenge compared to students choosing an incongruent challenge (adjusted odds ratios with 95% CI). GAHBI: general adverse health behavior improvement. *P<.05.



Association Between the Personal Relevance Score and Health-Promoting Change

The odds for health-promoting change increased with each additional unit of the personal relevance score (see model D in Figure 1) except for alcohol (Table 3). Thus, there was an increase in the odds for health-promoting change between personal relevance scores of 0 and 1 and a further increase in odds between personal relevance scores of 1 and 2, meaning that those with a personal relevance score of 7 (personal relevance score for alcohol, cigarettes, e-cigarettes, cannabis, social media, gaming, and gambling) were most likely to show a health-promoting change. To sum it up, the more problematic consumption habits a student had, the more likely he or she was to have a health-promoting change, except for a health-promoting change for alcohol.

Table 3. ORs[a] with a 95% CI for health-promoting change increasing with each additional unit of the personal relevance score, except for alcohol.

Health-promoting change	Participants, n	Events, n	OR (95% CI)	*P* value
Cigarettes	894	99	1.63 (1.40-1.92)	<.001
e-Cigarettes	894	90	1.73 (1.47-2.04)	<.001
Social media	894	429	1.27 (1.14-1.41)	<.001
Gaming	894	248	1.25 (1.12-1.41)	<.001
Alcohol	894	71	0.97 (0.80-1.17)	.72
Cannabis	894	30	1.93 (1.49-2.50)	<.001
Gambling	894	63	2.43 (1.99-2.98)	<.001
GAHBI[b]	894	649	1.69 (1.46-1.95)	<.001

[a]OR: odds ratio.

[b]GAHBI: general adverse health behavior improvement.

Associations Between Congruent Use and Sociodemographic Variables

Congruent use was found to be a relevant predictor of health-promoting changes and was therefore analyzed for associations with participants' sociodemographic characteristics such as age, gender, income, migration, and highest level of education. With the exception of migrant background, none of the included socioeconomic variables were found to be associated with congruent use. The odds for congruent use were lower for students with a migrant background compared to students with no migrant background (OR 0.67, 95% CI 0.49-0.94; *P*=.02).

Discussion

Principal Findings

The aim of this study was to examine the reach of a low-threshold mobile prevention program and its effectiveness depending on participant characteristics and actual app use. To achieve this aim, metrics describing the use of the mobile intervention and a relevance score related to substance- and

addiction-related personal risk and their associations with health-promoting change were analyzed.

The exploratory study revealed three main findings: (1) approximately 6 out of 10 students (894/1458, 61.3%) used the app and were thus reached by the prevention intervention; (2) almost half (407/894, 45.5%) of the app users used the app in congruence with personal relevance, suggesting a relatively high awareness of their personal habitual behaviors; and (3) as health-promoting changes were influenced by congruent use, choice of challenge, and personal relevance score, it can be argued that these associations provide empirical support for differential effectiveness of the app.

The proactive invitation to participate in the program and in the study in vocational school classes, combined with the offer of a low-threshold mobile intervention, reached almost 2 out of 3 students (894/1458, 61.3%) in the matched IG sample who downloaded the MZo app and started a challenge. This participation rate appears to be substantial, given the fact that the MZo program was not integrated into regular school lessons, but students were encouraged to download the app to their personal smartphones. Moreover, the introduction of the program was rather brief, and about one-quarter of the students were introduced to the program on the web. Nevertheless, the participation rate matches other mobile phone intervention programs with similar target groups in Switzerland and Germany, focusing on reducing problematic drinking or supporting smoking cessation, which attracted between 50% and 70% of the students [14-17,29]. However, a recent review suggests that local and personal recruitment of users leads to higher adherence or program participation [30]. Due to limited access to schools during the COVID-19 pandemic, 24.1% (1106/4591) of the students were recruited on the web. A higher rate of students might have been reached if recruitment had entirely been done during face-to-face school visits. In terms of who used the app, the results showed that app users were on average younger and more likely to have no migrant background; to have a medium monthly income; and to have a vocational training in commerce, industry, and technology. Furthermore, there was an association between congruent use of the app and no migrant background.

Results indicate that frequent use had no influence on health-promoting change. It could be argued that for this low-threshold mobile app, active use was not a necessary requirement for health-promoting change, as users received reminders to continue the challenge twice a day, and no further content about the challenge or the goal achievement was provided by the app once the challenge was started. This distinguishes the app from other mobile prevention apps [22,31].

Results also indicate that the odds for health-promoting behavior change were higher when the area of individual challenge choice was equal to the area of behavior change, challenge choice was congruent with a behavior of personal relevance (eg, a regular smoker chose the smoking challenge), and the individual risk of engaging in addictive behaviors was high. Congruent use almost tripled the odds for a health-promoting change for cigarettes and doubled the odds for a health-promoting change for e-cigarettes and gambling. Moreover, congruent use was the only variable in the analysis that provided an increase in the odds for general adverse health behavior improvement. This effect was based on the health-promoting changes in cigarettes and e-cigarettes.

As students' challenge choices were self-selected, these results indicate that a certain awareness of problematic habitual behaviors in combination with a motivation to change these behaviors may lead to health-promoting changes. Recent studies confirm that awareness and intrinsic motivation are important factors for behavioral change [32,33]. However, in contrast to Brailovskaia et al [26,27], the effects of a specific behavior reduction or abstinence were limited to the corresponding health-promoting change area. Thus, a "spillover effect" from a challenge in one behavioral domain did not generalize to other outcomes in this study. The results also suggest that it is more difficult to achieve the goal of a health-promoting change when it is defined as complete abstinence from a substance (eg, alcohol and cannabis). One reason for this may be that the app was designed as a low-threshold intervention targeting habitual behaviors rather than an intervention providing therapeutic support to change behavior. In this study, 1 in 3 adolescents reported to drink alcohol regularly [4]. Therefore, total alcohol abstinence might be an unrealistic goal for this group. The same may be true for cannabis, especially for the 8.6% of the 18- to 25-year-old adults who reported to consume it regularly [4].

Furthermore, health-promoting change for social media and gaming, although defined as reduction goals, seems to be more difficult to achieve than the reduction of cigarettes or e-cigarettes. One reason for this may be that the data collection for this study took place during the COVID-19 pandemic. Limited real-life contact with peers due to COVID-19 restrictions led to an increased frequency and duration of digital media use [34]. Thus, it could be argued that reducing gaming and social media time may have been more challenging during the COVID-19 pandemic because there were fewer social contact alternatives.

The results also show that a challenge choice increased the odds for the respective health-promoting change areas for cigarettes, e-cigarettes, social media, and gaming, except for alcohol. This can be seen as an indicator that the app effectively supported users who were willing to change a particular behavior. These results also confirm that reducing consumption is easier to achieve than total abstinence (eg, alcohol). The effectiveness of the app is further demonstrated by the fact that the personal relevance score was associated with increased odds for health-promoting change for cigarettes, e-cigarettes, cannabis, gambling, gaming, and social media. It is likely that students with a higher degree of problematic health behavior as indicated by a high relevance score were more aware of their behavior and thus more motivated to change. On the other hand, it could also be expected that these users would be less likely to achieve a health-promoting change with such a minimalistic support program, as these users are likely to have greater problems with controlling addictive behaviors. It would be informative to know more about the motivational background of the users in more detail. However, this requires qualitative data, for example, in the context of mixed method studies, which appears useful when testing new complex interventions.

Limitations

This study has a number of limitations: (1) the attrition rate was high and biased (1473/2367, 62.2%), partly due to the anonymization procedure, which allowed linking of baseline, follow-up, and app data without collecting any personal information. (2) As all data were based on self-reported data, social desirability and a recall bias may have influenced the results of the study. (3) Data collection for the study took place during the COVID-19 pandemic in 2021-2022, with varying restrictions in schools concerning data collection and social contact. These restrictions also affected the lives, substance use, and other behaviors of the students and thus the results of the study and their generalizability. (4) The number of health-promoting change events for alcohol, cannabis, and gambling was low, so these results should be interpreted with caution. (5) The results concerning the personal relevance score should also be interpreted with caution. As the formal risk screeners for addictive behaviors did not show relevant results, the personal relevance score was pragmatically created with the aim of showing habitual behavior. (6) As a convenience sample was recruited from school classes and students participated voluntarily, the results cannot be generalized to vocational school students in Germany.

Comparisons With Prior Work

The main finding of this study, the influence of congruent use and challenge choice on the intervention outcomes, partially confirms previous findings of Brailovskaia et al [26,27]. On the one hand, health promotion can be achieved through behavioral abstinence or reduction, but on the other hand, in contrast to Brailovskaia et al [26,27], the results of this study indicate that the effects of a specific behavior reduction or abstinence are limited to that specific behavior. The odds of health-promoting change increased either when students chose a challenge corresponding to their personal relevance or when the challenge choice met the respective area of health-promoting change. However, it was not investigated whether there were other positive health effects, for example, on physical activity or on mental health. Future research should further deepen the knowledge of the effects of voluntary commitment on health-related outcomes. The finding that congruent use and individual choice can play an important role in the effectiveness of the MZo intervention supplements previous research discussing how the most appropriate metrics of use may differ between different types and aims of interventions and how the exploration of use can help to understand which metrics are most associated with effectiveness [19,22,31]. In addition, the results of the study point toward harm or risk reduction rather than abstinence-related prevention for adolescents. Furthermore, there was an influence of the personal relevance score on health-promoting change. Future research should therefore focus on an in-depth exploration of the most appropriate measures of use of different interventions, on how these mechanisms of use influence the effectiveness of different interventions, and on how these mechanisms are influenced by intervention design and personal characteristics. Looking at the role of habitual behavior in the effectiveness of the app could be another area for future research.

To improve future versions of the MZo program, the reach of the program could be extended by embedding it in regular school curricula. This may increase awareness of the MZo program but would also require higher teacher involvement. Given the current results, congruent use of the app should be more strongly encouraged to increase the effectiveness of the program. Moreover, individual suitability may be increased by providing additional individual options, such as reduction goals in the areas of gambling, alcohol, and cannabis.

Conclusions

The MZo program appears to be promising in supporting behavioral change across a spectrum of addictive behaviors. The prevention approach based on voluntary commitment can be introduced to vocational students with minimal effort and offers autonomous and mobile use with a low threshold for participating schools and students. It has the potential to reach and raise awareness among adolescents before habits become addictions.

Acknowledgments

The authors would like to thank Ludwig Kraus, Reiner Hanewinkel, Nele Fritsch, Simone Leuckfeld, Freia Biedenweg, Laura Borwieck, and Johanna Weymann for their contributions to the study. This work was supported by the Federal Ministry of Health Germany (ZMVI1-2519DSM216). Moreover, the authors thank the anonymous reviewer for their comments, which significantly improved the paper.

Authors' Contributions

KG and NA conceptualized the study’s aims and wrote the initial draft of the manuscript. KG, NA, and BP completed data preparation. KG completed the data analysis. AD supervised the analyses, advised on the choice of statistical methods, and supported the interpretation of the results. All authors contributed to interpretations of the results and to revisions and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Screenshots of the Meine Zeit ohne app.
[PNG File , 290 KB - mhealth_v12i1e51307_app1.png]

Multimedia Appendix 2
Odds ratio with 95% CI for health-promoting change for high-use intensity.
[DOCX File , 15 KB - mhealth_v12i1e51307_app2.docx]

References

1. ESPAD report 2019: results from the European school survey project on alcohol and other drugs. European Monitoring Centre for Drugs and Drug Addiction (EU Body or Agency). 2020. URL: https://www.emcdda.europa.eu/publications/joint-publications/espad-report-2019_en [accessed 2024-04-30]
2. Arnaud N, Wartberg L, Simon-Kutscher K, Thomasius R, IMAC-Mind Consortium. Prevalence of substance use disorders and associations with mindfulness, impulsive personality traits and psychopathological symptoms in a representative sample of adolescents in Germany. Eur Child Adolesc Psychiatry 2024 Feb;33(2):451-465 [FREE Full text] [doi: 10.1007/s00787-023-02173-0] [Medline: 36853515]
3. Atzendorf J, Rauschert C, Seitz NN, Lochbühler K, Kraus L. The use of alcohol, tobacco, illegal drugs and medicines: an estimate of consumption and substance-related disorders in Germany. Dtsch Arztebl Int 2019 Sep 02;116(35-36):577-584 [FREE Full text] [doi: 10.3238/arztebl.2019.0577] [Medline: 31587705]
4. Orth B, Merkel C. Der Substanzkonsum Jugendlicher und junger Erwachsener in Deutschland. Ergebnisse des Alkoholsurveys 2021 zu Alkohol, Rauchen, Cannabis und Trends [Substance use among adolescents and young adults in Germany. Results of the Alcohol Survey 2021 on alcohol consumption, smoking behaviour, cannabis consumption and trends]. BZgA-Forschungsbericht. 2022. URL: https://www.bzga.de/forschung/studien/abgeschlossene-studien/studien-ab-1997/suchtpraevention/der-substanzkonsum-jugendlicher-und-junger-erwachsener-in-deutschland/ [accessed 2024-04-30]
5. Wartberg L, Kriston L, Thomasius R. The prevalence and psychosocial correlates of internet gaming disorder. Dtsch Arztebl Int 2017 Jun 23;114(25):419-424 [FREE Full text] [doi: 10.3238/arztebl.2017.0419] [Medline: 28683852]
6. Montag J, Hanewinkel R, Morgenstern M. Verbreitung und Korrelate des Substanzkonsums unter 5,688 auszubildenden an beruflichen Schulen. Prevalence and correlates of substance use in 5,688 vocational school students [Article in German]. Gesundheitswesen 2015;77(6):411-417. [doi: 10.1055/s-0034-1382043] [Medline: 25244547]
7. Onrust SA, Otten R, Lammers J, Smit F. School-based programmes to reduce and prevent substance use in different age groups: what works for whom? systematic review and meta-regression analysis. Clin Psychol Rev 2016;44:45-59 [FREE Full text] [doi: 10.1016/j.cpr.2015.11.002] [Medline: 26722708]
8. Morgenstern M, Seidel AK, Hanewinkel R. Suchtprävention an beruflichen Schulen. Addiction prevention at vocational schools [Article in German]. Bildung und Beruf 2019;2:50-57 [FREE Full text]
9. Auswärtiges Amt. The German vocational training system: an overview. Federal Foreign Office. URL: https://www.germany.info/us-en/welcome/wirtschaft/03-Wirtschaft/-/1048296 [accessed 2023-03-27]
10. Feierabend S, Rathgeb T, Kheredmand H, Glöckler S. JIM-Studie 2022. MPFS. 2022. URL: https://www.mpfs.de/studien/jim-studie/2022/ [accessed 2024-04-30]
11. Iribarren SJ, Akande TO, Kamp KJ, Barry D, Kader YG, Suelzer E. Effectiveness of mobile apps to promote health and manage disease: systematic review and meta-analysis of randomized controlled trials. JMIR Mhealth Uhealth 2021;9(1):e21563 [FREE Full text] [doi: 10.2196/21563] [Medline: 33427672]
12. Lee M, Lee H, Kim Y, Kim J, Cho M, Jang J, et al. Mobile app-based health promotion programs: a systematic review of the literature. Int J Environ Res Public Health 2018;15(12):2838 [FREE Full text] [doi: 10.3390/ijerph15122838] [Medline: 30551555]
13. Zhao J, Freeman B, Li M. Can mobile phone apps influence people's health behavior change? an evidence review. J Med Internet Res 2016;18(11):e287 [FREE Full text] [doi: 10.2196/jmir.5692] [Medline: 27806926]
14. Haug S, Boumparis N, Wenger A, Schaub MP, Castro RP. Efficacy of a mobile app-based coaching program for addiction prevention among apprentices: a cluster-randomized controlled trial. Int J Environ Res Public Health 2022;19(23):15730 [FREE Full text] [doi: 10.3390/ijerph192315730] [Medline: 36497804]
15. Haug S, Castro RP, Filler A, Kowatsch T, Fleisch E, Schaub MP. Efficacy of an internet and SMS-based integrated smoking cessation and alcohol intervention for smoking cessation in young people: study protocol of a two-arm cluster randomised controlled trial. BMC Public Health 2014;14:1140 [FREE Full text] [doi: 10.1186/1471-2458-14-1140] [Medline: 25369857]
16. Haug S, Castro PR, Kowatsch T, Filler A, Schaub MP. Efficacy of a technology-based, integrated smoking cessation and alcohol intervention for smoking cessation in adolescents: results of a cluster-randomised controlled trial. J Subst Abuse Treat 2017;82:55-66 [FREE Full text] [doi: 10.1016/j.jsat.2017.09.008]
17. Haug S, Schaub MP, Venzin V, Meyer C, John U. Efficacy of a text message-based smoking cessation intervention for young people: a cluster randomized controlled trial. J Med Internet Res 2013;15(8):e171 [FREE Full text] [doi: 10.2196/jmir.2636] [Medline: 23956024]

18. Diestelkamp S, Schulz AL, Thomasius R. Technologiebasierte interventionen zur alkoholprävention bei kindern und jugendlichen. Technology-based interventions for alcohol prevention among children and adolescents [Article in German]. Bundesgesundheitsbl 2021;64(6):714-721 [FREE Full text] [doi: 10.1007/s00103-021-03338-5] [Medline: 33950365]
19. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011 Aug 05;13(3):e52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]
20. Donkin L, Hickie IB, Christensen H, Naismith SL, Neal B, Cockayne NL, et al. Rethinking the dose-response relationship between usage and outcome in an online intervention for depression: randomized controlled trial. J Med Internet Res 2013;15(10):e231 [FREE Full text] [doi: 10.2196/jmir.2771] [Medline: 24135213]
21. Castro RP, Haug S, Debelak R, Jakob R, Kowatsch T, Schaub MP. Engagement with a mobile phone-based life skills intervention for adolescents and its association with participant characteristics and outcomes: tree-based analysis. J Med Internet Res 2022;24(1):e28638 [FREE Full text] [doi: 10.2196/28638] [Medline: 35044309]
22. Schmidt CA, Romine JK, Bell ML, Armin J, Gordon JS. User participation and engagement with the See Me Smoke-Free mHealth app: prospective feasibility trial. JMIR Mhealth Uhealth 2017;5(10):e142 [FREE Full text] [doi: 10.2196/mhealth.7900] [Medline: 28993302]
23. Witkiewitz K, Desai SA, Bowen S, Leigh BC, Kirouac M, Larimer ME. Development and evaluation of a mobile intervention for heavy drinking and smoking among college students. Psychol Addict Behav 2014;28(3):639-650 [FREE Full text] [doi: 10.1037/a0034747] [Medline: 25000269]
24. Arnaud N, Weymann J, Lochbühler K, Pietsch B, Rossa M, Kraus L, et al. Effectiveness of an app-based intervention to reduce substance use, gambling, and digital media use in vocational school students: study protocol for a randomized controlled trial. Trials 2022;23(1):277 [FREE Full text] [doi: 10.1186/s13063-022-06231-x] [Medline: 35395932]
25. Pietsch B, Arnaud N, Lochbühler K, Rossa M, Kraus L, Gomes de Matos E, et al. Effects of an app-based intervention program to reduce substance use, gambling, and digital media use in adolescents and young adults: a multicenter, cluster-randomized controlled trial in vocational schools in Germany. Int J Environ Res Public Health 2023;20(3):1970 [FREE Full text] [doi: 10.3390/ijerph20031970] [Medline: 36767337]
26. Brailovskaia J, Ströse F, Schillack H, Margraf J. Less Facebook use—more well-being and a healthier lifestyle? an experimental intervention study. Comput Hum Behav 2020;108:106332. [doi: 10.1016/j.chb.2020.106332]
27. Brailovskaia J, Meier-Faust J, Schillack H, Margraf J. A two-week gaming abstinence reduces internet gaming disorder and improves mental health: an experimental longitudinal intervention study. Comput Hum Behav 2022;134:107334. [doi: 10.1016/j.chb.2022.107334]
28. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA 2013 Nov 27;310(20):2191-2194. [doi: 10.1001/jama.2013.281053] [Medline: 24141714]
29. Haug S, Schaub MP, Venzin V, Meyer C, John U, Gmel G. A pre-post study on the appropriateness and effectiveness of a web- and text messaging-based intervention to reduce problem drinking in emerging adults. J Med Internet Res 2013;15(9):e196 [FREE Full text] [doi: 10.2196/jmir.2755] [Medline: 23999406]
30. Jakob R, Harperink S, Rudolf AM, Fleisch E, Haug S, Mair JL, et al. Factors influencing adherence to mHealth apps for prevention or management of noncommunicable diseases: systematic review. J Med Internet Res 2022;24(5):e35371 [FREE Full text] [doi: 10.2196/35371] [Medline: 35612886]
31. Mattila E, Lappalainen R, Välkkynen P, Sairanen E, Lappalainen P, Karhunen L, et al. Usage and dose response of a mobile acceptance and commitment therapy app: secondary analysis of the intervention arm of a randomized controlled trial. JMIR Mhealth Uhealth 2016;4(3):e90 [FREE Full text] [doi: 10.2196/mhealth.5241] [Medline: 27468653]
32. Taylor VA, Moseley I, Sun S, Smith R, Roy A, Ludwig VU, et al. Awareness drives changes in reward value which predict eating behavior change: probing reinforcement learning using experience sampling from mobile mindfulness training for maladaptive eating. J Behav Addict 2021;10(3):482-497 [FREE Full text] [doi: 10.1556/2006.2021.00020] [Medline: 34264854]
33. Mata J, Silva MN, Vieira PN, Carraça EV, Andrade AM, Coutinho SR, et al. Motivational "spill-over" during weight control: increased self-determination and exercise intrinsic motivation predict eating self-regulation. Health Psychol 2009;28(6):709-716. [doi: 10.1037/a0016764] [Medline: 19916639]
34. Paschke K, Austermann MI, Simon-Kutscher K, Thomasius R. Adolescent gaming and social media usage before and during the COVID-19 pandemic: interim results of a longitudinal study. SUCHT 2021;67(1):13-22. [doi: 10.1024/0939-5911/a000694]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
GAHBI: general adverse health behavior improvement
IG: intervention group
MZo: Meine Zeit ohne
OR: odds ratio

RCT: randomized controlled trial

Edited by L Buis; submitted 27.07.23; peer-reviewed by P Caponnetto, O Siljeholm, Y Spinzy; comments to author 18.01.24; revised version received 11.03.24; accepted 22.03.24; published 23.07.24.

Please cite as:
Grahlher K, Morgenstern M, Pietsch B, Gomes de Matos E, Rossa M, Lochbühler K, Daubmann A, Thomasius R, Arnaud N
Mobile App Intervention to Reduce Substance Use, Gambling, and Digital Media Use in Vocational School Students: Exploratory Analysis of the Intervention Arm of a Randomized Controlled Trial
JMIR Mhealth Uhealth 2024;12:e51307
URL: https://mhealth.jmir.org/2024/1/e51307
doi:10.2196/51307
PMID:

Original Paper

Evaluating the Effects of a Self-Help Mobile Phone App on Worry and Rumination Experienced by Young Adults: Randomized Controlled Trial

Daniel Edge[1], BSc, PostDip, MRES, DClinPsy; Edward Watkins[1], BA, MSc, ClinPsy, Prof Dr; Alexandra Newbold[1], BSc, PhD; Thomas Ehring[2], Prof Dr; Mads Frost[3], PhD; Tabea Rosenkranz[2], PhD

[1]Mood Disorders Centre, School of Psychology, University of Exeter, Exeter, United Kingdom

[2]Department of Psychology, Ludwig-Maximilians-University of Munich, Munich, Germany

[3]Monsenso A/S, Copenhagen, Denmark

Corresponding Author:
Daniel Edge, BSc, PostDip, MRES, DClinPsy
Mood Disorders Centre
School of Psychology
University of Exeter
Washington Singer Building Perry Road
Washington Singer Laboratories
Exeter, EX4 4QG
United Kingdom
Phone: 44 01392724621
Email: de292@exeter.ac.uk

Abstract

Background: Delivery of preventative interventions via mobile phone apps offers an effective and accessible way to address the global priority of improving the mental health of adolescents and young adults. A proven risk factor for anxiety and depression is elevated worry and rumination, also known as repetitive negative thinking (RNT).

Objective: This was a prevention mechanism trial that aimed to investigate whether an RNT-targeting self-help mobile phone app (MyMoodCoach) reduces worry and rumination in young adults residing in the United Kingdom. A secondary objective was to test whether the app reduces symptoms of anxiety and depression and improves well-being.

Methods: A web-based, single-blind, 2-arm parallel-group randomized controlled trial was conducted with 236 people aged between 16 and 24 years, who self-reported high levels of worry or rumination. Eligible participants were randomized to an active intervention group (usual practice, plus up to 6 weeks of using the RNT-targeting mobile app, n=119) or a waitlist control group (usual practice with no access to the app until after 6 weeks, n=117). The primary outcome was changes in worry and rumination 6 weeks after randomization. Secondary outcomes included changes in well-being and symptoms of anxiety and depression after 6 weeks and changes in all measures after 12 weeks.

Results: Participants randomly allocated to use the RNT-targeting self-help app showed significantly lower levels of rumination (mean difference –2.92, 95% CI –5.57 to –0.28; P=.03; η_p^2=0.02) and worry (mean difference –3.97, 95% CI –6.21 to –1.73; P<.001; η_p^2=0.06) at 6-week follow-up, relative to the waitlist control. Similar differences were observed for well-being (P<.001), anxiety (P=.03), and depression (P=.04). The waitlist control group also showed improvement when given access to the app after 6 weeks. Improvements observed in the intervention group after 6 weeks of using the app were maintained at the 12-week follow-up point.

Conclusions: The MyMoodCoach app had a significant positive effect on worry and rumination, well-being, anxiety, and depression in young adults, relative to waitlist controls, providing proof-of-principle that an unguided self-help app can effectively reduce RNT. This app, therefore, has potential for the prevention of anxiety and depression although longer-term effects on incidence need to be directly evaluated.

Trial Registration: ClinicalTrials.gov NCT04950257; https://www.clinicaltrials.gov/ct2/show/NCT04950257

International Registered Report Identifier (IRRID): RR2-10.1186/s12888-021-03536-0

(JMIR Mhealth Uhealth 2024;12:e51932) doi:10.2196/51932

KEYWORDS

worry; rumination; repetitive negative thinking; prevention-mechanism; well-being; depression; anxiety; mobile-based interventions; mobile phone; mobile health application; app; application

Introduction

There is growing concern about the early onset of mental health disorders in adolescents and young adults [1]. Poor mental health during adolescence can severely affect a young person's future life chances and can have a significant negative impact on health, education, and employment in later life [2,3]. Depression and anxiety are 2 of the most common mental health problems in adolescents and young adults, which significantly contribute toward global disability and produce a high economic burden [4]. In 2020, the prevalence of depression and anxiety disorders in adolescents globally was estimated at between 25% and 31% [5].

Although effective treatments for depression and anxiety have been developed and delivered in many countries, the evidence suggests that the overall prevalence of both disorders remains largely unchanged [6]. Because the prevalence of these common mental health disorders is so high, it is not possible for traditional delivery models of psychological intervention, such as a course of face-to-face sessions, to fully address the global need for treatment as it is unlikely there will ever be a sufficient number of trained mental health professionals [7]. Further, even if acute treatments were made widely available, as a consequence of the high levels of recurrence and relapse for depression and anxiety, it has been estimated that only a partial reduction of the overall prevalence could be achieved [6,8]. Effective prevention approaches are, therefore, considered essential in lowering overall prevalence rates as they have the potential to reduce both initial and recurrent episodes, thereby decreasing the demand for acute treatment [9]. Given the substantial increase in the incidence of depression and anxiety during mid-adolescence peaking in young adulthood [2], adolescents and young adults are considered an important group for which prevention of poor mental health is urgently needed [10-12].

While prevention interventions for anxiety and depression already exist for adolescents and young people, systematic reviews suggest that effect sizes are relatively small [13-15]. Further, most prevention interventions are delivered in person and require considerable input from trained professionals, such as teachers and therapists, which can increase costs and limit their availability [16]. While there is some evidence that supported interventions may have better adherence and efficacy than unguided interventions, their coverage and volume of delivery are constrained because each individual supporting the intervention has a finite capacity, such that these interventions cannot be made available to all who may benefit. As such, to ensure the large-scale coverage needed for effective mental health prevention approaches, there is potential value in supplementing supported interventions with ones that can be delivered to multiple users simultaneously and do not require additional input from practitioners (ie, nonconsumable self-help interventions [17]). The use of such nonconsumable self-help interventions would substantively increase the scalability and availability of preventative interventions making them more suitable for use as a public health approach at a population level.

The use of the internet and mobile phones has been increasingly explored as an avenue that may be able to increase the accessibility of prevention approaches and reduce the cost of intervention [18]. While internet and mobile-based interventions have many similarities, their method of delivery can vary. Mobile-based interventions, for example, are typically delivered through dedicated mobile apps and accessed on mobile phones, whereas internet-based interventions use web-based (or compatible) platforms designed to be accessed via laptops or personal computers. Internet-based interventions tend to deliver therapeutic content made up of pages working through a specific topic and more closely resemble a structured therapeutic session, for example, taking a set amount of time to complete and working through set content. Mobile apps, however, are based on more flexible use of "bite-sized" information and exercises, with less information on the screen, and allowing the user to "swipe" through available material to access the content they desire. The use of mobile phones is near ubiquitous in young people in the United Kingdom with an estimated 99% of those aged between 16 and 24 years using a smartphone on a regular basis in 2023 [19]. Further, mobile apps can help integrate behavioral changes into daily life—the app is always available to the user, making it well-suited for changing unhelpful habits. Mobile apps, therefore, provide a unique avenue in which to engage young adults in health-related activities, as well as promote good mental health and prevention strategies [20,21]. Despite a huge increase in the number of mobile-based mental health apps over the last 10 years [22], only a small number have been developed with scientific rigor. Further, many do not use established treatment principles nor have they been rigorously tested in robust well-powered randomized controlled trials (RCTs) [23]. While emerging evidence suggests that mobile-based apps can deliver efficacious treatment interventions for anxiety and depression [24,25], there have not been many trials examining their use for well-being promotion and prevention of poor mental health in young people specifically [26,27].

A psychological process thought to be key when promoting good well-being and preventing mental health problems is repetitive negative thinking (RNT) [28]. RNT is defined as a pattern of thinking that is repetitive, difficult to manage, and focused on negative content, which can significantly impact an individual's well-being and emotional functioning [28-30]. RNT encompasses various thought patterns, but the exemplars of RNT are worry and rumination which have been identified as robust risk factors for several mental health disorders [31]. Worry is described as a relatively uncontrollable chain of negative thinking about the future in the form of "What if" type

questions. Such thoughts can focus on typical everyday activities (such as work and relationships), as well as more catastrophic concerns (such as worrying that you may get hit by a falling tree) [32]. Rumination is a form of dysfunctional, negative thinking, which focuses on analyzing the causes and consequences of negative events. This can involve dwelling on past events and continually going over and over why things went wrong [32]. The degree of rumination experienced by an individual has been shown to predict the onset and duration of major depressive episodes [33,34], as well as the severity of depressive symptoms [35,36]. Rumination has also been shown to mediate the effects of other identified risk factors, such as neuroticism and stressful life events, on the onset of depressive episodes [37]. Worry has been identified as having strong associations with symptoms of depression and anxiety [38]. There is also evidence that increased levels of worry are predictive of greater symptom severity for both anxiety and depression [39] and that daily worrying predicts subsequent increases in anxiety [40]. There is extensive evidence that worry and rumination share a common process including evidence that they are highly correlated with each other and that this common factor accounts for their relationship with anxiety and depression, which underpins the conceptualization of the wider process of RNT [28,29]. There is robust evidence for RNT as a transdiagnostic risk factor, which is common across many psychological mood disorders including depression and anxiety [32,41]. The induction of RNT has been found to exacerbate symptoms of anxiety and depression such as negative thinking, delayed decision-making speed, poor problem-solving, and negative affect [42-45]. There is extensive evidence that RNT predicts future levels of depressive and anxiety-related symptoms, as well as the onset of depression [28,46]. As rumination and worry have been shown to be proximal risk factors that affect onset, maintenance, and relapse, RNT is considered an underlying mechanism for both anxiety and depression and is identified as an important target for preventative interventions [41,47].

There are several interventions that have been designed specifically to reduce RNT and improve general mental well-being, including those which have been trialed as digital interventions delivered via the internet. An RCT of a preventative intervention designed to target excessive levels of RNT found that both a group-based and internet version of a rumination-focused Cognitive Behavioral Therapy program significantly reduced RNT (d=0.53-0.89) and symptom levels of anxiety and depression (d=0.36-0.72) compared to a waitlist group [48]. Another trial involving 235 high-risk participants found that an internet-based version of the program reduced the risk of depression by 34% relative to controls and that participants showed a significant improvement in RNT and depressive symptoms in the short to medium term [49]. While there is promising evidence, therefore, that preventative interventions focused on reducing RNT are effective when delivered via the internet, there is limited research exploring the efficacy of RNT-targeting interventions delivered via a mobile phone app.

The Assessing and Enhancing Emotional Competence for Well-being in the Young project (ECoWeB) aimed to address the gap in large-scale trial evidence for preventative apps by developing and evaluating a self-guided mobile phone app (MyMoodCoach). The app for this large-scale trial was designed to promote emotional well-being and prevent mental health problems in adolescents and young adults through engaging and personalized tools that train psychological skills [50]. Specifically, this app was personalized by targeting the 2 most problematic components of emotional competence skills (from a set of 4: emotional regulation, emotional appraisal achievement context, emotional appraisal social context, and emotional perception and knowledge) for the user, which were identified in their baseline assessments [50]. One of these components was addressed via a module focused on targeting RNT for individuals identified as having elevated levels of RNT at baseline (reflecting poor emotional regulation and emotional competence skills). The trial described in the current paper is an off-shoot study of this larger project, specifically focused on testing the value of an app that targets RNT. In this study, we tested a standalone variant of the app used in the large-scale trial that included only the module focused on targeting RNT, aiming to shift users toward more adaptive emotional regulation skills.

Developing trials to evaluate preventative interventions can take considerable time and resources. However, while outcomes for prevention interventions typically focus on reducing the incidence (ie, new cases) of a specific mental health problem, prevention mechanism trials concentrate on the underlying etiological processes. This allows for a more detailed examination of proximal mechanisms. Prevention mechanism trials, therefore, offer an efficient way of establishing whether interventions can reduce risk factors associated with pathology, thereby evaluating their potential for prevention [51]. Since worry and rumination are established risk factors for anxiety and depression, this trial was a prevention mechanism trial to test whether the use of an app targeting RNT can reduce worry and rumination in young people, thus evaluating its potential as a prevention intervention. Our primary hypothesis was that people allocated to the RNT-targeting mobile app would show significantly lower levels of rumination and worry relative to those allocated to the waitlist control arm. A second hypothesis was that those allocated to the mobile app would also report significantly lower symptoms of anxiety and depression and higher well-being relative to waitlist control.

Methods

Study Design

A superiority 2-arm parallel-group RCT was conducted comparing an active intervention arm (usual practice plus up to 6 weeks of using the RNT-targeting mobile app) with a waitlist control arm (usual practice with access to the app only after a 6-week wait). As it was not possible to blind participants to their allocated groups, this was a single-blind RCT with only the researcher blind to participant group allocation. Further details can be found in the trial protocol [52]. This study was conducted according to the CONSORT (Consolidated Standards of Reporting Trials) guidelines [53,54] and extensions for nonpharmacological treatment interventions, as well as

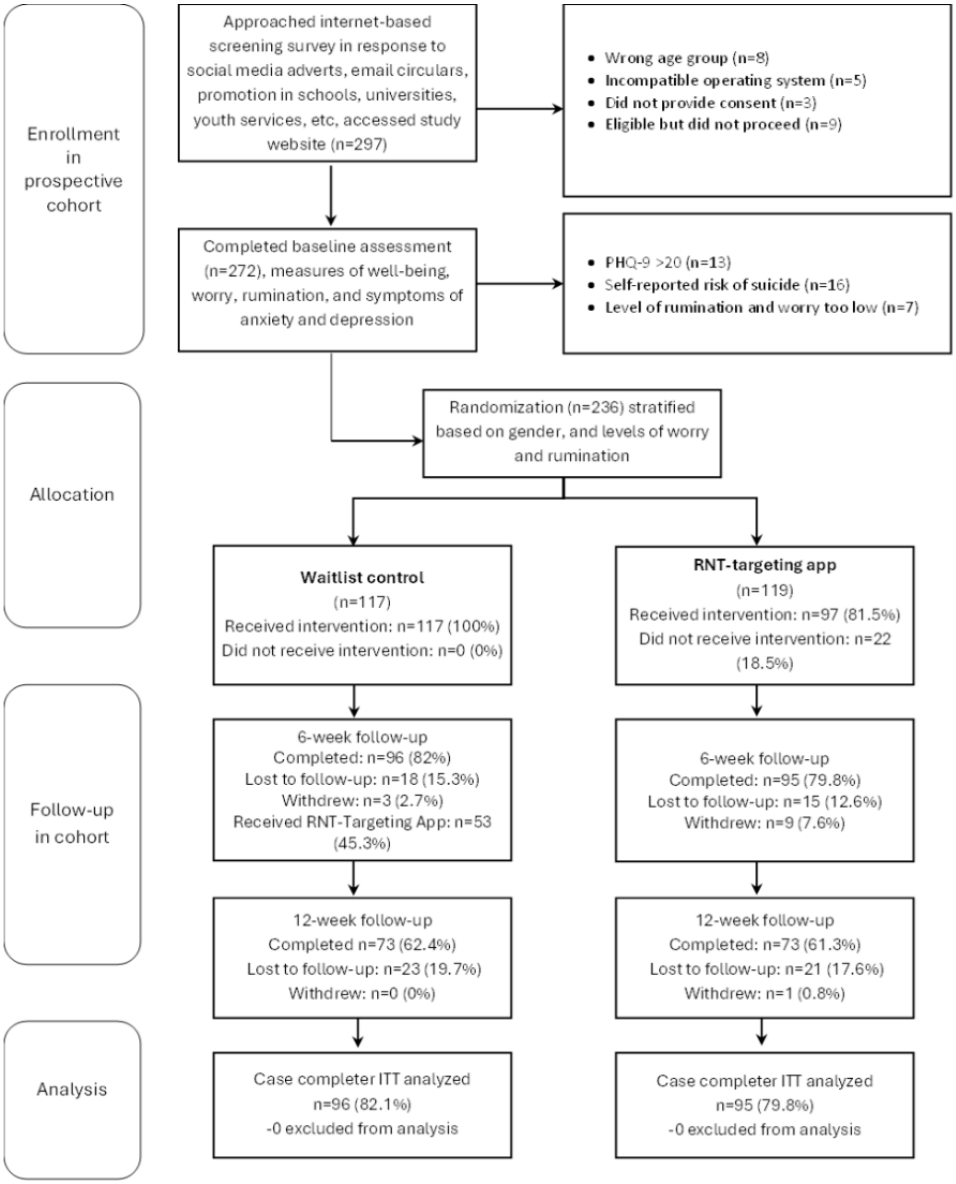
CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile HEalth Applications and on Line Tele Health) guidelines for improving and standardizing evaluation reports of web-based and mobile health interventions (Multimedia Appendix 1) [55].

Recruitment

We recruited young adults aged between 16 and 24 residing in the United Kingdom. The study was advertised on social media platforms, including Facebook (Meta Platforms) and Instagram (Meta Platforms), as well as an internal research recruitment system (Sona) used by the University of Exeter. Only participants who reported elevated levels of RNT at baseline were eligible for this trial. This was defined as scoring above the 50th percentile (ie, top half of scale) on either the Ruminative Response Scale (RRS; >34) or the Penn State Worry Questionnaire (PSWQ; >41). Participants also had to be aged between 16 and 24 years, currently residing in the United Kingdom, possess basic literacy in English and have access to a smartphone (either Android or iOS). Recruitment commenced on May 14, 2021, and ceased on October 11, 2021.

Participants were excluded at baseline if they reported elevated symptoms of depression indicating they required more specialist treatment, defined as having a score of 20 or higher on the Patient Health Questionnaire-9 (PHQ-9) [56] because this suggests a severe level of symptoms where a more intensive treatment from a mental health professional would be indicated. Participants who reported elevated symptoms of anxiety were included and automatically provided with additional information for accessing support if required. Other exclusion criteria included active suicidality or self-reported to be currently receiving treatment for a mental health problem (ie, psychological therapy, counseling, or psychiatric medication) at baseline. Those who self-reported having a current diagnosis of clinical depression, bipolar disorder, or psychosis were also excluded. A summary is given in the CONSORT flow diagram (Figure 1).

Figure 1. CONSORT flow diagram for the trial. CONSORT: Consolidated Standards of Reporting Trials; ITT: intention-to-treat; PHQ-9: Patient Health Questionnaire-9; RNT: repetitive negative thinking.



Screening and Consent Procedure

Potential participants were directed to a website providing further information about the study and were prompted to answer some prescreening questions, regarding their age and current experiences of mental health. Individuals who were not suitable at this stage (eg, outside of the specified age range) were automatically directed to a web page explaining why they were not suitable for the trial. Participants who passed this prescreening stage were then provided with an information sheet and asked to consent to provide contact details and complete a baseline assessment where suitability for the trial (ie, level of RNT and symptoms for anxiety and depression) was assessed more thoroughly. Following the baseline assessment, participants who met all eligibility criteria were then asked to consent to take part in the trial before being randomized into one of the 2 arms. Everyone accessing the trial website was provided with contact details for the research team.

Randomization

Randomization was conducted independently using a pregenerated computerized allocation algorithm. Randomization was in a 1:1 ratio and stratified across each arm according to the sex (male, female, and nonbinary) and level of RNT (50th-75th percentile vs 75th percentile or higher). Participants scoring in the highest quartile (at or above the 75th percentile) on either the RRS or PSWQ and in the top tercile on the other measure were identified as having high RNT, based on criteria used by Topper et al [48], which have previously been found to predict increased risk for subsequent depression [10].

Participant Flow

Overall, 297 people completed the prescreening process of whom 272 went on to complete baseline questionnaires and have their suitability for the trial assessed more thoroughly. In total, 236 participants met the eligibility criteria and were randomized to either receive access to the digital RNT-targeting self-help app straight away (n=119) or after a period of 6 weeks (n=117). The rate of follow-up attrition was 19.07% (n=45; intervention arm, n=24; and waitlist control, n=21) at 6 weeks and 38.14% (n=90; intervention arm, n=46; and waitlist control, n=44) at 12 weeks. These calculations included 11 (4.6%) participants who contacted the research team requesting to withdraw from the study, 1 of who did so after completing their 6-week follow-up. There were no missing data from any of the participants who completed their surveys. Further details are given in the CONSORT flow diagram (Figure 1).

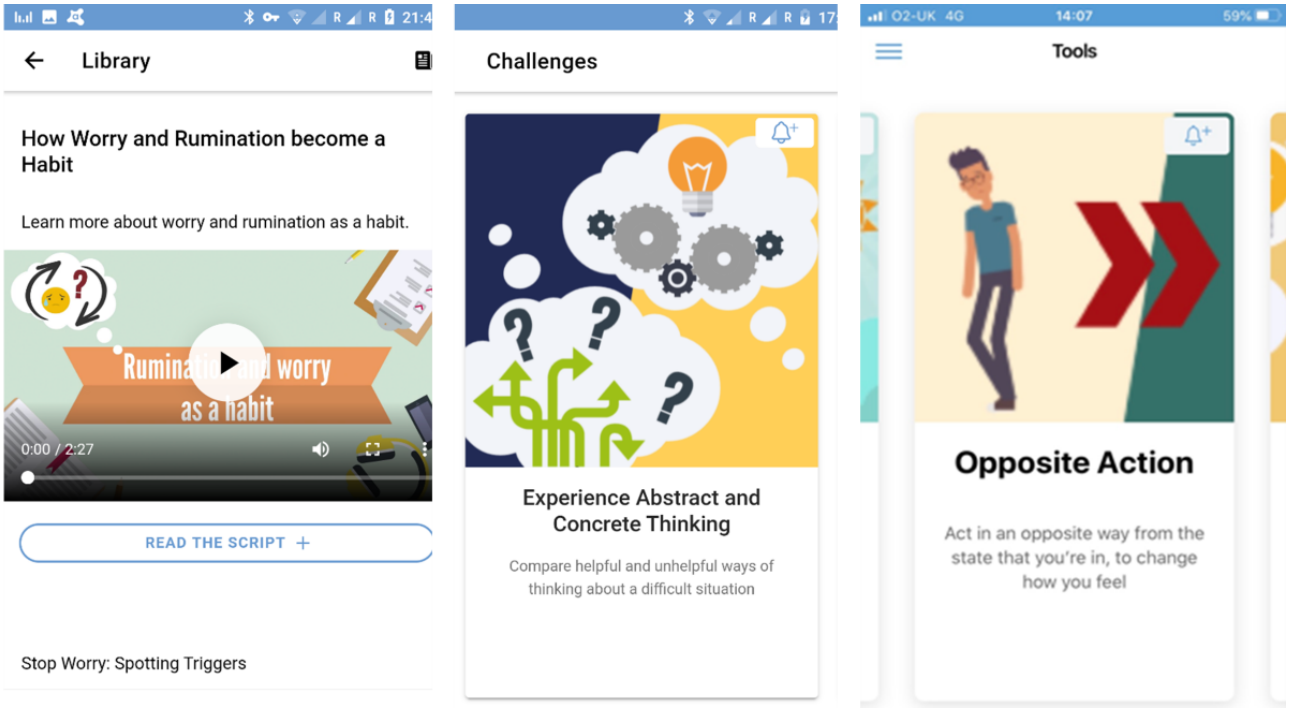
Interventions

Digital RNT-Targeting Self-Help App (Intervention Arm)

The self-help app used in this trial (called MyMoodCoach) was a focused version of the app evaluated in the main ECoWeB trial [50]. This standalone version was not personalized to each user and was specifically focused on reducing RNT to help improve emotion regulation, that is, it only included the emotional regulation rumination-focused module. The app included self-monitoring, psychoeducation, and active self-help exercises based on RNT-specific strategies from an evidence-based, rumination-focused cognitive behavioral therapy program intervention [48,49,57-59], adapting these evidence-based interventions from an internet-delivery context to an app format. Core elements of the intervention were designed to break the ruminative habit and enable users to shift toward a more helpful processing style. This involved coaching participants to spot warning signs for rumination and worry, and then plan alternative strategies. These included being more active, slowing things down, breaking tasks down, opposite action, relaxation, concrete thinking, becoming absorbed, self-compassion, and assertiveness. Participants were prompted to practice alternative strategies in response to their warning signs. The user interface includes text, pictures, audio recordings, animations, audio exercises, and questionnaires with tailored automated feedback. There was also a self-monitoring component which prompted users 5 times a day to rate their mood and level of rumination. The app featured a menu structure including a dashboard to monitor notifications and progress, a library function that had psychoeducation and explanatory animated videos, and an explore function to graph the self-monitoring responses made by the participant. The app also included challenges that provide learning exercises (eg, behavioral exercises) and tools that are brief strategies that young people can use at the moment when they need them (for example, compassion and relaxation exercises). The app was entirely automated (ie, self-guided) and designed for use on both iOS and Android phones. The app was accessed for free via each participant's smartphone app store.

Changes from the proven internet versions of the intervention [9,11] included information condensed so that "bite-sized" content reflecting 1 point of information was presented per swipe of the mobile phone screen and could be consumed in brief moments of time (rather than by scrolling down web pages each with multiple points of information); greater flexibility of use such that users could select any element of the app in any order from the menu (eg, use of tools or challenges) unlike the internet-delivered interventions, which tended to be more modular and arranged in a fixed structure and order, as a "lesson" or "session" that might take an hour to work through in its entirety. Some screenshots for the app are provided in Figure 2 and further details can be accessed via Multimedia Appendix 2.

Figure 2. Screenshots from the MyMoodCoach app from left to right: screen from the library of visual resources, screen within the challenges menu to enter the abstract and concrete thinking challenge, screen within the tools menu to enter the opposite action tool.



Waitlist Control Group

The waitlist control group received access to the RNT-targeting digital self-help app after a 6-week wait.

Baseline and Follow-Up Assessments

All outcome measures were completed by participants independently via a website at baseline, with follow-up assessments at 6-weeks and 12-weeks post randomization. Demographic information including, age, gender, employment status, and ethnicity, was collected only at baseline. Additional feedback about experiences of using the mobile app was only collected at the 12-week follow-up point. Automated emails were sent to participants at each follow-up point with a link to complete their survey. Further reminders were sent manually via email and text message if there was no response to the automated reminders.

As an incentive, all participants who completed their follow-up surveys were entered into a prize draw. Participants recruited via the University of Exeter Sona system could choose whether to be entered into the prize draw or receive course credits. Out of those eligible to participate in the trial, 82 (32.7%) chose to receive course credit. The researcher who was blind to treatment allocation was the only team member involved in sending out reminders to complete follow-up surveys. Other members of the team were available to follow up on issues of risk and answer technical queries from participants. Throughout the trial process, there were no deviations from the planned procedure with no incidents of unblinding being reported.

Outcomes

The primary outcome for this trial was changes in the levels of rumination and worry at the primary end point (6 weeks after randomization). Rumination was measured using the RRS [36]. The measure consists of 22 items that assess the tendency for an individual to respond to depressed moods with a focus on either the self, depressive symptoms, or negative consequences. Items are scored from 1 (almost never) to 4 (almost always) and total scores can range from 22 to 88 with higher scores indicating a higher level of rumination. The RRS has good internal consistency in an adolescent population (α=.88) [60] and test-retest reliability ranges from moderate to high (r=0.47 for over 1 year and r=0.80 for over 5 months) [33,61].

Worry was measured using the PSWQ [62], a 16-item self-assessment that assesses the intensity, tendency, and uncontrollability of worrying thoughts using a 5-point Likert scale for each item. Items are scored from 1 (not at all typical of me) to 5 (very typical of me) and total scores can range from 16 to 80 with higher scores indicating greater levels of worry. The measure has high internal consistency in both clinical and nonclinical samples (α=.88-.95), as well as good test-retest reliability (r=0.72-0.93 for 2 weeks and 1 month, respectively) [62].

Secondary outcomes for this study included changes in mental well-being and symptoms of anxiety and depression at the primary end point, as well as changes in all measures between 6-week and 12-week follow-up. Mental well-being was measured using the Warwick-Edinburgh Mental Well-being Scale (WEMWBS) [63]. The WEMWBS has demonstrated high internal consistency in both student and general populations (α=.89-.91), with good test-retest reliability (intraclass correlation coefficient=0.83; P<.01) [63]. The Generalized Anxiety Disorder questionnaire (GAD-7) [64] was used to assess symptoms of anxiety. Symptoms of depression were assessed using the PHQ-9 [65]. Both measures have demonstrated good internal consistency (α=.86-.92), as well as good test-retest reliability (0.83-0.84) [66].

To assess user's perspective on the app, participants were asked to rate the app on its ease of use, look and feel, features, and content on a 5-point Likert scale of 1 (terrible) to 5 (brilliant), as well as provide qualitative feedback about the positives and negatives of using the app. These questions were devised by the lead researcher and adapted from components of the technology acceptance model (TAM) and user experience (UX) design approaches [67,68].

Statistical Analysis

Power

An estimated minimum clinically important difference (MCID) for the primary outcome of the RRS was used to calculate the sample size for this study. One recommended approach for identifying the MCID is half of the SD for the respective index [69]. A conservative estimate of the normative SD for the RRS was identified based on previous research (mean 47.19, SD 7.58 to mean 48.89, SD 8.51) [48]. A reduction of 4 points on the RRS, therefore, was identified as an appropriate MCID. Using an α of .05 with 90% power, for an MCID of 4 required 85 participants per group (170 in total). Allowing for a 20% follow-up attrition rate, we aimed to recruit a minimum of 204 participants (102 per arm).

Analysis Plan

Data were analyzed using SPSS (version 28; IBM Corp) and the statistical analytical plan was finalized prior to data being unblinded. The primary analysis was based on intention-to-treat and all participants, regardless of their level of engagement, were included [70]. Missing data were handled via multiple imputations (MI) using a linear regression model with a monotone imputation method. As multiple imputations had equivalent results to case-completer analyses, only the case-completer intention-to-treat analysis results have been reported. Analysis of covariance (ANCOVA) at 6 weeks was used for both primary and secondary outcomes, with baseline scores as the covariate. ANCOVA was also used for both primary and secondary outcomes at 12 weeks with 6-week follow-up scores as the covariate.

Feedback about user experience was collated and analyzed using NVivo software (Lumivero) which identified common words and phrases occurring within the data to generate preliminary codes. The same software was then used to automatically group the identified codes into suggested themes. These preliminary themes were then checked and refined by the primary researcher before being finalized.

Ethical Considerations

Ethical approval for this study was obtained from the Ethics Committee of the School of Psychology, University of Exeter (eCLESPsy001977v5.1). All participants provided written informed consent. Data were collected in a pseudonymized manner and stored securely.

Results

Overview

Baseline demographics across both arms for the study are shown in Table 1. All participants were aged between 16 and 24 years and were predominantly White female students.

Table 1. Baseline demographics for each condition.

Baseline characteristics	RNT[a]-targeting app (n=119)	Waitlist control (n=117)
Age, mean (SD)	18.44 (2.01)	18.56 (2.4)
Sex, n (%)		
Male	15 (12.6)	16 (13.7)
Female	109 (84)	98 (83.8)
Non-binary or third gender	4 (3.4)	3 (2.6)
Ethnicity, n (%)		
White	95 (79.8)	96 (82.1)
Black, African, Caribbean, or Black British	2 (1.7)	1 (0.9)
Asian or Asian British	14 (11.8)	15 (12.8)
Mixed or other ethnic group	8 (6.7)	5 (4.3)
Employment, n (%)		
Employed (full-time)	5 (4.2)	9 (7.7)
Employed (part-time)	11 (9.2)	14 (12.0)
Unemployed	8 (6.7)	4 (3.4)
Student	95 (79.8)	90 (76.9)
History of mental health problems, n (%)		
Yes	18 (15.1)	19 (16.2)

[a]RNT: repetitive negative thinking.

Outcome at Primary End Point (6 Weeks)

Means and SDs for each measure and arm at each time point are shown in Table 2. One-way condition (intervention vs control) ANCOVAs found a significant main effect of intervention condition on rumination (Figures 3 and 4; $F_{1,188}$=4.72; P=.03; η_p^2=0.02), worry ($F_{1,188}$=12.24; P<.001; η_p^2=0.06), depression ($F_{1,188}$=4.14; P=.04; η_p^2=0.02), and anxiety ($F_{1,188}$=5.43; P=.02; η_p^2=0.03). These significant effects reflect that after adjusting for the covariate of baseline scores, participants randomized to use the RNT-targeting app in the intervention arm reported significantly lower rumination, worry, depression, and anxiety at 6-week follow-up than those randomized to the waitlist control (mean difference [rumination]: –2.92, 95% CI –5.57 to –0.28; mean difference [worry]: –3.97, 95% CI –6.21 to –1.73; mean difference [depression]: –1.34, 95% CI –2.63 to –0.04; mean difference [anxiety]: –1.46, 95% CI –2.7 to –0.23). Participants randomized to the intervention arm reported significantly higher well-being at 6 weeks than the waitlist control group ($F_{1,188}$=12.38; P<.001; η_p^2=0.06; adjusted mean difference 3.78, 95% CI 1.66-5.9).

Table 2. Means and SDs for all measures at each follow-up point.

	Baseline, mean (SD)		6-week follow-up, mean (SD)		12-week follow-up, mean (SD)	
	RNT[a]-targeting app (n=119)	Waitlist control (n=117)	RNT-targeting app (n=95)	Waitlist control (n=96)	RNT-targeting app (n=73)	Waitlist control (n=73)
RRS[b]	51.08 (11.62)	51.51 (10.76)	49.91 (11.52)	52.88 (11.68)	47.86 (10.78)	49.3 (11.1)
PSWQ[c]	59.87 (10.82)	60.24 (10.4)	56.81 (9.38)	60.65 (11.08)	55.18 (10.49)	58.04 (11.08)
WEMWBS[d]	43.56 (7.92)	44.09 (7.08)	46.24 (8.95)	42.96 (7.47)	47.62 (6.96)	45.86 (7.13)
GAD-7[e]	9.40 (4.67)	9.04 (4.79)	7.97 (4.62)	9.14 (5.1)	7.45 (3.78)	7.90 (4.77)
PHQ-9[f]	8.61 (4.25)	8.56 (4.48)	8.57 (5.16)	9.73 (5.19)	7.85 (4.33)	8.49 (4.3)

[a]RNT: repetitive negative thinking.

[b]RRS: Ruminative Response Scale.

[c]PSWQ: Penn State Worry Questionnaire.

[d]WEMWBS: Warwick-Edinburgh Mental Well-being Scale.

[e]GAD-7: Generalized Anxiety Disorder-7.

[f]PHQ-9: Patient Health Questionnaire-9.

Figure 3. Graph of mean RRS scores for RNT-targeting app and waitlist control participants with 95% CI error bars. RNT: repetitive negative thinking; RRS: Ruminative Response Scale.

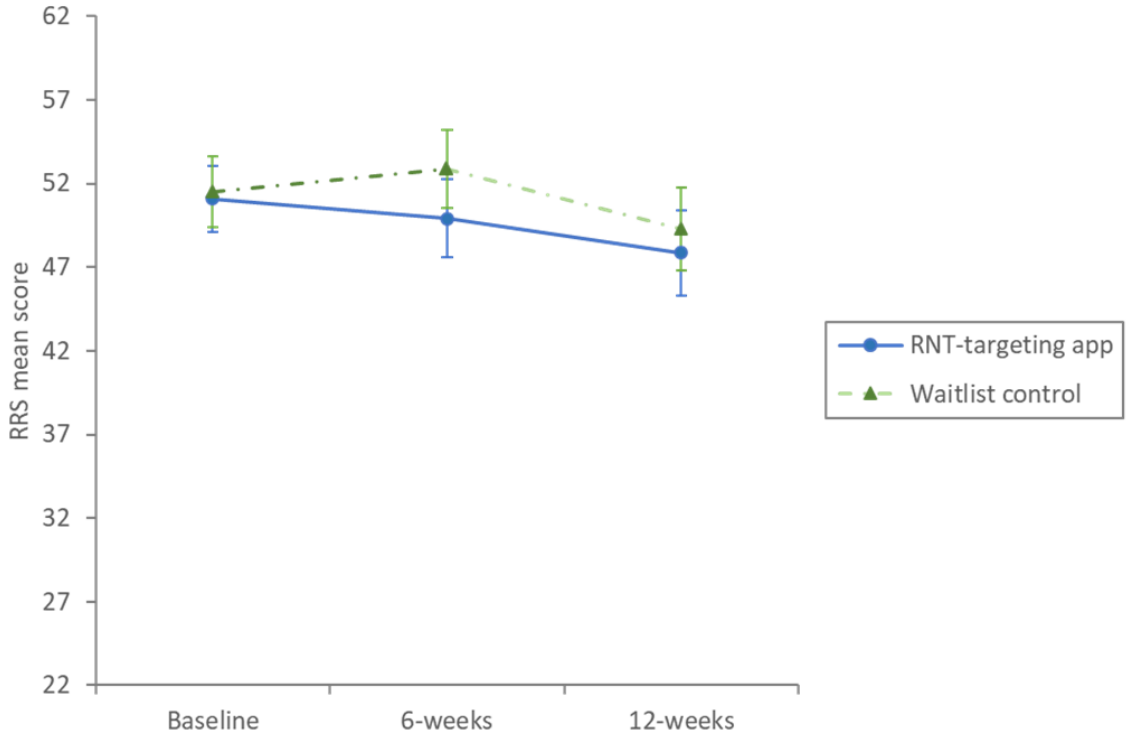
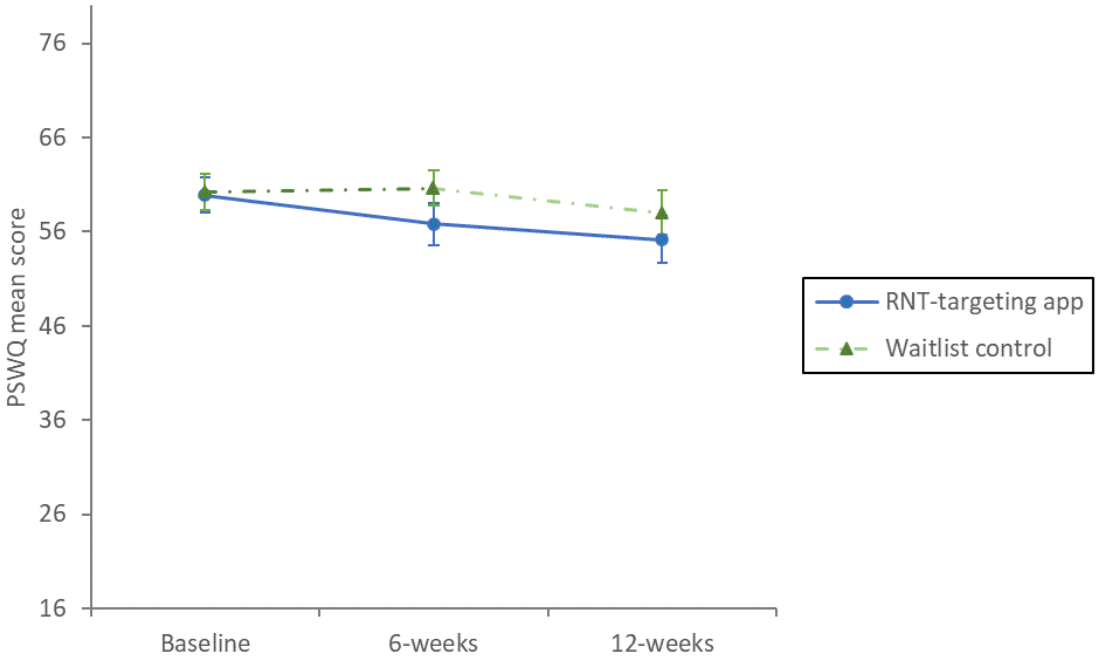


Figure 4. Graph of mean PSWQ scores for RNT-targeting app and waitlist control participants with 95% CI error bars. PSWQ: Penn State Worry Questionnaire; RNT: repetitive negative thinking.



Outcome at Follow-Up (12 Weeks)

At 12 weeks, participants who were initially allocated to use the RNT-targeting app maintained the changes that were made on all measures during the first 6 weeks, with no significant changes observed between the 2 follow-up points (see Table 3). After being introduced to the app at 6 weeks, participants in the waitlist control showed significant reductions in rumination, worry, anxiety, and depression and an increase in well-being (see Table 3). One-way ANCOVAs found no statistically significant differences between the 2 groups for any of the measures at the 12-week follow-up point ($F_{1, 143}$≤1.17; P≥.28).

Table 3. Within-participants contrast between the 6- and 12-week follow-up point.

	RNT[a]-targeting app			Waitlist control		
	F test (*df*)	*P* value	Effect size (η_p^2)	*F* test (*df*)	*P* value	Effect size (η_p^2)
RRS[b]	2.63 (1, 144)	.11	0.02	22.09 (1, 144)	<.001	0.13
PSWQ[c]	1.47 (1, 144)	.23	0.01	12.2 (1, 144)	<.001	0.09
WEMWBS[d]	0.93 (1, 144)	.34	0.01	20 (1, 144)	<.001	0.12
GAD-7[e]	1.12 (1, 144)	.30	0.01	9.8 (1, 144)	.002	0.07
PHQ-9[f]	2.82 (1, 144)	.10	0.02	9.8 (1, 144)	.002	0.07

[a]RNT: repetitive negative thinking.

[b]RRS: Ruminative Response Scale.

[c]PSWQ: Penn State Worry Questionnaire.

[d]WEMWBS: Warwick-Edinburgh Mental Well-being Scale.

[e]GAD-7: Generalized Anxiety Disorder-7.

[f]PHQ-9: Patient Health Questionnaire-9.

Adherence

Overall, 97 out of 119 (81.5%) participants in the intervention arm and 53 out of 96 (55.2%) in the waitlist control signed into the app. To estimate app usage, we calculated the mean number of days participants completed a daily mood rating (including those who never signed into the app). For those randomly allocated to use the RNT-targeting app, this was 8 (SD 10.92; range 0-40) days during the first 6-week period, and for the waitlist control, this was 3 (SD 6.31; range 0-31) days once given access to the app.

UX Findings

When identifying positive attributes about the app, participants consistently highlighted the mood tracker and tool features as being beneficial. Participants also identified that the app

provided them with useful skills to help improve their emotional management such as identifying unhelpful patterns, reflection, and analyzing their emotions more critically. Several participants noted that the app helped them feel relaxed and that it was most useful when they were experiencing heightened levels of distress or going through "a dark time." When identifying areas to improve, several features appeared to affect participants' motivation to regularly engage with it. A total of 9 comments highlighted that the app sent too many notifications which became very annoying; 11 comments alluded to the app being difficult to navigate and quite tiring to engage with on a regular basis; 8 comments noted experiencing frequent bugs and crashes; and 26 comments noted some design issues such as the range of emotions in the app being too limited, the questions used to evaluate daily events being repetitive, and the user interface being "too clinical" and not aesthetically pleasing. Overall, the average satisfaction rating from all participants who provided ratings regarding their experience (n=137) was 3.74 out of 5 (SD 0.73; ease of use: mean 3.99, SD 0.9; look and feel: mean 3.45, SD 1.08; features and functionality: mean 3.72, SD 0.87; content: mean 3.8, SD 0.94).

Discussion

Principal Findings

This was a prevention-mechanism trial that aimed to test whether an RNT-targeting self-help mobile phone app (MyMoodCoach) was effective at reducing worry and rumination in young adults. A further objective was to pilot the efficacy of the app in reducing symptoms of anxiety and depression and improving well-being.

As hypothesized, participants randomized to the MyMoodCoach app showed a significant decrease in both worry and rumination, relative to participants in the waitlist control condition. As hypothesized, participants randomized to the app also showed significant increases in well-being and reduction in anxiety and depression relative to the waitlist control group. All these changes were sustained at the 12-week follow-up point suggesting that any benefit of using the app is likely to persist over several months. Moreover, once they were given access to the app, participants in the waitlist control showed similar changes in RNT, anxiety, depression, and well-being, paralleling the benefits observed in the initial active intervention arm.

Patterns of change across the trial indicate that symptoms of depression during the first 6 weeks stayed constant in the intervention arm while worsening in the waitlist control group. This suggests that engaging in an intervention designed to target RNT may have a preventive effect on depressive symptoms consistent with the wider literature on rumination and worry as vulnerability factors for depression [39,71,72]. Further, this is consistent with previous RCTs that found that targeting RNT can have a medium-term preventive effect reducing the onset of depression [48,49]. The reduction of both anxiety and depression from a self-help app that explicitly targeted worry and rumination is also consistent with theoretical models proposing that RNT is a transdiagnostic risk factor that can affect symptoms associated with both disorders [41]. This research thus provides proof-of-principle of the value of targeting RNT as a mechanism to prevent anxiety and depression, consistent with the prevention mechanism trial approach. However, because we did not explicitly assess new onsets of major depression or anxiety disorders, and the follow-up was only short-term, longer-term follow-ups with an assessment of incidence are needed to rigorously test the potential of the self-help app targeting RNT as a prevention intervention. Despite app usage only being modest, it still had an effect. Further, the percentage of those not using the app (14%) was below the mean percentage of "non-users" reported in a recent meta-analytic review of adherence in mobile interventions (41.2%) [73].

In addition to its potential as a prevention intervention, the observed benefits of the app suggest there may be several clinical applications that could help reduce the burden on mental health services. As the intervention is focused on worry and rumination, the app could be offered to people who do not meet the severity threshold to access treatment for anxiety or depression, but still show high levels of RNT. It could also be offered as an interim approach for people on a waitlist for treatment as using the app may either reduce their associated symptoms or prevent their symptoms from getting worse. Moreover, as the app targets associated risk factors rather than a specific mental health problem, it may also help reduce the impact of stigma, which can prevent people from seeking support when distressed, and potentially increase engagement [74].

The study had several limitations. First, as all screening questionnaires were based on self-report, completed remotely, and no diagnostic interview was conducted, we cannot be sure that participants did not have a disorder at baseline. This also meant that we could not directly ascertain the effect of the app on the incidence of episodes of anxiety disorders and major depression. Second, participants were mainly White, female students, which may limit the generalizability of the findings, although we note a much greater proportion of females to males in the sample was expected given our inclusion criteria because elevated RNT tends to be much more common in females than males and has been found to partially account for the increased rates (nearly double) of depression and anxiety reported in females relative to males [75,76]. As such, this gender imbalance is not necessarily problematic for the purposes of developing an RNT-targeting preventative intervention for anxiety and depression but rather it accurately reflects the distribution of RNT in the target population. Future studies, however, will need to encourage more participation from other gender identities to confirm whether the app is effective for those who do not identify as female. Third, while the follow-up period in this study was appropriate for a prevention mechanism trial, a longer-term follow-up is needed to test the effect of the RNT-targeting app on preventing the incidence of episodes of major depression and generalized anxiety.

Conclusions

Despite some limitations, the MyMoodCoach app has the potential to offer a valuable contribution toward large-scale effective prevention for young people [77]. These results provide proof of principle that the intervention can effectively target

worry and rumination as possible prevention mechanisms for anxiety and depression in young people. As RNT is a well-established vulnerability factor, it is likely that this intervention will have a positive impact on the incidence of anxiety and depression in the medium-term.

Acknowledgments

Funding for this paper was provided by the Horizon 2020 Assessing and Enhancing Emotional Competence for Well-being in the Young (ECoWeB) project (grant 754657), which included this single-intervention study within the study protocol, which was peer-reviewed by the funding body. The study sponsor and funder have no influence on the conduct, delivery, or reporting of the research.

Authors' Contributions

EW and DE designed the trial. EW and AN coordinated the trial and randomization process. DE and EW prepared the statistical analysis plan and statistical analysis was conducted by DE. EW, TE, MF, and TR contributed toward the design and preparation of content for the repetitive negative thinking (RNT)–targeting app and helped coordinate the development of the app intervention. DE and EW contributed to the writing of the draft paper. AN, MF, TE, and TR reviewed and approved the final version.

Conflicts of Interest

MF is a cofounder and shareholder of Monsenso, whose solution delivered the app. EW is the developer of Rumination-focused cognitive-behavioural therapy, which is the intervention used in the app and receives payment for training workshops and royalties from the treatment manual published by Guilford Press. The other authors declare no conflicts of interests.

Multimedia Appendix 1
Completed CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile HEalth Applications and on Line Tele Health) form.
[PDF File (Adobe PDF File), 1154 KB - mhealth_v12i1e51932_app1.pdf]

Multimedia Appendix 2
Additional information about the RNT-targeting app.
[DOCX File , 9393 KB - mhealth_v12i1e51932_app2.docx]

References

1. Thapar A, Collishaw S, Pine DS, Thapar AK. Depression in adolescence. Lancet 2012;379(9820):1056-1067 [FREE Full text] [doi: 10.1016/S0140-6736(11)60871-4] [Medline: 22305766]
2. Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the national comorbidity survey replication. Arch Gen Psychiatry 2005;62(6):593-602. [doi: 10.1001/archpsyc.62.6.593] [Medline: 15939837]
3. Mehta N, Croudace T, Davies SC. Public mental health: evidenced-based priorities. Lancet 2015;385(9976):1472-1475. [doi: 10.1016/S0140-6736(14)61400-8] [Medline: 25217115]
4. Newby JM, McKinnon A, Kuyken W, Gilbody S, Dalgleish T. Systematic review and meta-analysis of transdiagnostic psychological treatments for anxiety and depressive disorders in adulthood. Clin Psychol Rev 2015;40:91-110 [FREE Full text] [doi: 10.1016/j.cpr.2015.06.002] [Medline: 26094079]
5. Silva SA, Silva SU, Ronca DB, Gonçalves VSS, Dutra ES, Carvalho KMB. Common mental disorders prevalence in adolescents: a systematic review and meta-analyses. PLoS One 2020;15(4):e0232007 [FREE Full text] [doi: 10.1371/journal.pone.0232007] [Medline: 32324835]
6. Jorm AF, Patten SB, Brugha TS, Mojtabai R. Has increased provision of treatment reduced the prevalence of common mental disorders? Review of the evidence from four countries. World Psychiatry 2017;16(1):90-99 [FREE Full text] [doi: 10.1002/wps.20388] [Medline: 28127925]
7. Kazdin AE, Blase SL. Rebooting psychotherapy research and practice to reduce the burden of mental illness. Perspect Psychol Sci 2011;6(1):21-37. [doi: 10.1177/1745691610393527] [Medline: 26162113]
8. Andrews G, Sanderson K, Corry J, Lapsley HM. Using epidemiological data to model efficiency in reducing the burden of depression*. J Ment Health Policy Econ 2000;3(4):175-186. [doi: 10.1002/mhp.96] [Medline: 11967454]
9. Ormel J, Cuijpers P, Jorm A, Schoevers RA. What is needed to eradicate the depression epidemic, and why. Ment Health Prev 2020;17:200177. [doi: 10.1016/j.mhp.2019.200177]
10. Patton GC, Sawyer SM, Ross DA, Viner RM, Santelli JS. From advocacy to action in global adolescent health. J Adolesc Health 2016;59(4):375-377 [FREE Full text] [doi: 10.1016/j.jadohealth.2016.08.002] [Medline: 27664465]

11. No health without public mental health. Royal College of Psychiatrists. 2010. URL: https://www.rcpsych.ac.uk/pdf/PS04_2010.pdf [accessed 2024-07-04]
12. Depression and other common mental disorders global health estimates. World Health Organization. 2017. URL: https://apps.who.int/iris/bitstream/handle/10665/254610/WHO-MSD-MER-2017.2-eng.pdf [accessed 2020-04-14]
13. Hetrick SE, Cox GR, Witt KG, Bir JJ, Merry SN. Cognitive behavioural therapy (CBT), third-wave CBT and interpersonal therapy (IPT) based interventions for preventing depression in children and adolescents. Cochrane Database Syst Rev 2016;2016(8):CD003380 [FREE Full text] [doi: 10.1002/14651858.CD003380.pub4] [Medline: 27501438]
14. Merry SN, Hetrick SE, Cox GR, Brudevold-Iversen T, Bir JJ, McDowell H. Psychological and educational interventions for preventing depression in children and adolescents. Cochrane Database Syst Rev 2011(12):CD003380. [doi: 10.1002/14651858.CD003380.pub3] [Medline: 22161377]
15. Stockings EA, Degenhardt L, Dobbins T, Lee YY, Erskine HE, Whiteford HA, et al. Preventing depression and anxiety in young people: a review of the joint efficacy of universal, selective and indicated prevention. Psychol Med 2016;46(1):11-26. [doi: 10.1017/S0033291715001725] [Medline: 26315536]
16. Kazdin AE. Technology-based interventions and reducing the burdens of mental illness: perspectives and comments on the special series. Cogn Behav Pract 2015;22(3):359-366. [doi: 10.1016/j.cbpra.2015.04.004]
17. Muñoz RF, Cooper LA. The COVID-19 pandemic and mental health-implementing evidence-based interventions to advance equity and reverse a worsening crisis. JAMA Heal Forum 2022;3(4):e221282 [FREE Full text] [doi: 10.1001/jamahealthforum.2022.1282] [Medline: 36218968]
18. Kazdin AE, Rabbitt SM. Novel models for delivering mental health services and reducing the burdens of mental illness. Clin Psychol Sci 2013;1(2):170-191. [doi: 10.1177/2167702612463566]
19. Share of smartphone users in the United Kingdom (UK) 2012-2023. Statista. 2024. URL: https://www.statista.com/statistics/300402/smartphone-usage-in-the-uk-by-age/ [accessed 2024-04-14]
20. Musiat P, Goldstone P, Tarrier N. Understanding the acceptability of e-mental health--attitudes and expectations towards computerised self-help treatments for mental health problems. BMC Psychiatry 2014;14:109 [FREE Full text] [doi: 10.1186/1471-244X-14-109] [Medline: 24725765]
21. Giota KG, Kleftaras G. Mental health apps: innovations, risks and ethical considerations. ETSN 2014;03(03):19-23. [doi: 10.4236/etsn.2014.33003]
22. Torous J, Roberts LW. Needed innovation in digital health and smartphone applications for mental health: transparency and trust. JAMA Psychiatry 2017;74(5):437-438. [doi: 10.1001/jamapsychiatry.2017.0262] [Medline: 28384700]
23. Bauer M, Glenn T, Geddes J, Gitlin M, Grof P, Kessing LV, et al. Smartphones in mental health: a critical review of background issues, current status and future concerns. Int J Bipolar Disord 2020;8(1):2 [FREE Full text] [doi: 10.1186/s40345-019-0164-x] [Medline: 31919635]
24. Firth J, Torous J, Carney R, Newby J, Cosco TD, Christensen H, et al. Digital technologies in the treatment of anxiety: recent innovations and future directions. Curr Psychiatry Rep 2018;20(6):44 [FREE Full text] [doi: 10.1007/s11920-018-0910-2] [Medline: 29779065]
25. Firth J, Torous J, Nicholas J, Carney R, Pratap A, Rosenbaum S, et al. The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. World Psychiatry 2017;16(3):287-298 [FREE Full text] [doi: 10.1002/wps.20472] [Medline: 28941113]
26. Deady M, Choi I, Calvo RA, Glozier N, Christensen H, Harvey SB. eHealth interventions for the prevention of depression and anxiety in the general population: a systematic review and meta-analysis. BMC Psychiatry 2017;17(1):310 [FREE Full text] [doi: 10.1186/s12888-017-1473-1] [Medline: 28851342]
27. Sander L, Rausch L, Baumeister H. Effectiveness of internet-based interventions for the prevention of mental disorders: a systematic review and meta-analysis. JMIR Ment Health 2016;3(3):e38 [FREE Full text] [doi: 10.2196/mental.6061] [Medline: 27535468]
28. Ehring T, Watkins ER. Repetitive negative thinking as a transdiagnostic process. Int J Cogn Ther 2008;1(3):192-205. [doi: 10.1521/ijct.2008.1.3.192]
29. Topper M, Molenaar D, Emmelkamp PM, Ehring T. Are rumination and worry two sides of the same coin? A structural equation modelling approach. J Exp Psychopathol 2014;5(3):363-381. [doi: 10.5127/jep.038813]
30. McEvoy PM, Salmon K, Hyett MP, Jose PE, Gutenbrunner C, Bryson K, et al. Repetitive negative thinking as a transdiagnostic predictor of depression and anxiety symptoms in adolescents. Assessment 2019;26(2):324-335. [doi: 10.1177/1073191117693923] [Medline: 29214855]
31. Taylor MM, Snyder HR. Repetitive negative thinking shared across rumination and worry predicts symptoms of depression and anxiety. J Psychopathol Behav Assess 2021;43(4):904-915. [doi: 10.1007/s10862-021-09898-9]
32. Harvey AG, Watkins ER, Mansell W, Shafran R. Cognitive Behavioural Processes Across Psychological Disorders: A Transdiagnostic Approach to Research and Treatment. UK: Oxford University Press; 2004.
33. Just N, Alloy LB. The response styles theory of depression: tests and an extension of the theory. J Abnorm Psychol 1997;106(2):221-229. [doi: 10.1037//0021-843x.106.2.221] [Medline: 9131842]
34. Robinson MS, Alloy LB. Negative cognitive styles and stress-reactive rumination interact to predict depression: a prospective study. Cognit Ther Res 2003;27(3):275-291. [doi: 10.1023/A:1023914416469]

35. Abela JRZ, Brozina K, Haigh EP. An examination of the response styles theory of depression in third- and seventh-grade children: a short-term longitudinal study. J Abnorm Child Psychol 2002;30(5):515-527. [doi: 10.1023/a:1019873015594] [Medline: 12403154]
36. Nolen-Hoeksema S, Morrow J. A prospective study of depression and posttraumatic stress symptoms after a natural disaster: the 1989 Loma Prieta earthquake. J Pers Soc Psychol 1991;61(1):115-121. [doi: 10.1037//0022-3514.61.1.115] [Medline: 1890582]
37. Spasojević J, Alloy LB. Rumination as a common mechanism relating depressive risk factors to depression. Emotion 2001;1(1):25-37. [doi: 10.1037/1528-3542.1.1.25] [Medline: 12894809]
38. Eisma MC, Boelen PA, Schut HAW, Stroebe MS. Does worry affect adjustment to bereavement? A longitudinal investigation. Anxiety Stress Coping 2017;30(3):243-252. [doi: 10.1080/10615806.2016.1229464] [Medline: 27575924]
39. Young CC, Dietrich MS. Stressful life events, worry, and rumination predict depressive and anxiety symptoms in young adolescents. J Child Adolesc Psychiatr Nurs 2015;28(1):35-42. [doi: 10.1111/jcap.12102] [Medline: 25808272]
40. Dickson KS, Ciesla JA, Reilly LC. Rumination, worry, cognitive avoidance, and behavioral avoidance: examination of temporal effects. Behav Ther 2012;43(3):629-640. [doi: 10.1016/j.beth.2011.11.002] [Medline: 22697450]
41. Kaplan DM, Palitsky R, Carey AL, Crane TE, Havens CM, Medrano MR, et al. Maladaptive repetitive thought as a transdiagnostic phenomenon and treatment target: an integrative review. J Clin Psychol 2018;74(7):1126-1136. [doi: 10.1002/jclp.22585] [Medline: 29342312]
42. Metzger RL, Miller ML, Cohen M, Sofka M, Borkovec TD. Worry changes decision making: the effect of negative thoughts on cognitive processing. J Clin Psychol 1990;46(1):78-88. [doi: 10.1002/1097-4679(199001)46:1<78::aid-jclp2270460113>3.0.co;2-r] [Medline: 2303569]
43. Lyubomirsky S, Nolen-Hoeksema S. Effects of self-focused rumination on negative thinking and interpersonal problem solving. J Pers Soc Psychol 1995;69(1):176-190. [doi: 10.1037//0022-3514.69.1.176] [Medline: 7643299]
44. McLaughlin KA, Borkovec TD, Sibrava NJ. The effects of worry and rumination on affect states and cognitive activity. Behav Ther 2007;38(1):23-38. [doi: 10.1016/j.beth.2006.03.003] [Medline: 17292692]
45. Hubbard NA, Faso DJ, Krawczyk DC, Rypma B. The dual roles of trait rumination in problem solving. Pers Individ Dif 2015;86:321-325. [doi: 10.1016/j.paid.2015.06.034]
46. Watkins ER. Constructive and unconstructive repetitive thought. Psychol Bull 2008;134(2):163-206 [FREE Full text] [doi: 10.1037/0033-2909.134.2.163] [Medline: 18298268]
47. Watkins ER, Roberts H. Reflecting on rumination: consequences, causes, mechanisms and treatment of rumination. Behav Res Ther 2020;127:103573. [doi: 10.1016/j.brat.2020.103573] [Medline: 32087393]
48. Topper M, Emmelkamp PM, Watkins E, Ehring T. Prevention of anxiety disorders and depression by targeting excessive worry and rumination in adolescents and young adults: a randomized controlled trial. Behav Res Ther 2017;90:123-136. [doi: 10.1016/j.brat.2016.12.015] [Medline: 28049069]
49. Cook L, Mostazir M, Watkins E. Reducing stress and preventing depression (RESPOND): randomized controlled trial of web-based rumination-focused cognitive behavioral therapy for high-ruminating university students. J Med Internet Res 2019;21(5):e11349 [FREE Full text] [doi: 10.2196/11349] [Medline: 31094331]
50. Newbold A, Warren FC, Taylor RS, Hulme C, Burnett S, Aas B, et al. Promotion of mental health in young adults via mobile phone app: study protocol of the ECoWeB (emotional competence for well-being in young adults) cohort multiple randomised trials. BMC Psychiatry 2020;20(1):458 [FREE Full text] [doi: 10.1186/s12888-020-02857-w] [Medline: 32962684]
51. Zalta AK, Shankman SA. Conducting psychopathology prevention research in the RDoC era. Clin Psychol (New York) 2016;23(1):94-104 [FREE Full text] [doi: 10.1111/cpsp.12144] [Medline: 27065571]
52. Edge D, Newbold A, Ehring T, Rosenkranz T, Frost M, Watkins ER. Reducing worry and rumination in young adults via a mobile phone app: study protocol of the ECoWeB (Emotional Competence for Well-Being in Young Adults) randomised controlled trial focused on repetitive negative thinking. BMC Psychiatry 2021;21(1):519 [FREE Full text] [doi: 10.1186/s12888-021-03536-0] [Medline: 34674669]
53. Grant S, Mayo-Wilson E, Montgomery P, Macdonald G, Michie S, Hopewell S, on behalf of the CONSORT-SPI Group. CONSORT-SPI 2018 explanation and elaboration: guidance for reporting social and psychological intervention trials. Trials 2018;19(1):406 [FREE Full text] [doi: 10.1186/s13063-018-2735-z] [Medline: 30060763]
54. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. Ann Intern Med 2010;152(11):726-732 [FREE Full text] [doi: 10.7326/0003-4819-152-11-201006010-00232] [Medline: 20335313]
55. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of web-based and mobile health interventions. J Med Internet Res 2011;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
56. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]

57. Watkins ER, Mullan E, Wingrove J, Rimes K, Steiner H, Bathurst N, et al. Rumination-focused cognitive-behavioural therapy for residual depression: phase II randomised controlled trial. Br J Psychiatry 2011;199(4):317-322. [doi: 10.1192/bjp.bp.110.090282] [Medline: 21778171]
58. Watkins ER, Taylor RS, Byng R, Baeyens C, Read R, Pearson K, et al. Guided self-help concreteness training as an intervention for major depression in primary care: a phase II randomized controlled trial. Psychol Med 2012;42(7):1359-1371 [FREE Full text] [doi: 10.1017/S0033291711002480] [Medline: 22085757]
59. Hvenegaard M, Moeller SB, Poulsen S, Gondan M, Grafton B, Austin SF, et al. Group rumination-focused cognitive-behavioural therapy (CBT) group CBT for depression: phase II trial. Psychol Med 2020;50(1):11-19 [FREE Full text] [doi: 10.1017/S0033291718003835] [Medline: 30630555]
60. Burwell RA, Shirk SR. Subtypes of rumination in adolescence: associations between brooding, reflection, depressive symptoms, and coping. J Clin Child Adolesc Psychol 2007;36(1):56-65. [doi: 10.1080/15374410709336568] [Medline: 17206881]
61. Nolen-Hoeksema S, Parker LE, Larson J. Ruminative coping with depressed mood following loss. J Pers Soc Psychol 1994;67(1):92-104. [doi: 10.1037//0022-3514.67.1.92] [Medline: 8046585]
62. Meyer TJ, Miller ML, Metzger RL, Borkovec TD. Development and validation of the Penn State Worry Questionnaire. Behav Res Ther 1990;28(6):487-495. [doi: 10.1016/0005-7967(90)90135-6] [Medline: 2076086]
63. Tennant R, Hiller L, Fishwick R, Platt S, Joseph S, Weich S, et al. The Warwick-Edinburgh Mental Well-Being Scale (WEMWBS): development and UK validation. Health Qual Life Outcomes 2007;5:63 [FREE Full text] [doi: 10.1186/1477-7525-5-63] [Medline: 18042300]
64. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
65. Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD. Prim Care Companion J Clin Psychiatry 1999;2(1):31. [doi: 10.1001/jama.282.18.1737]
66. Kroenke K, Spitzer RL, Williams JBW, Löwe B. The patient health questionnaire somatic, anxiety, and depressive symptom scales: a systematic review. Gen Hosp Psychiatry 2010;32(4):345-359. [doi: 10.1016/j.genhosppsych.2010.03.006] [Medline: 20633738]
67. Rahimi B, Nadri H, Lotfnezhad Afshar H, Timpka T. A systematic review of the technology acceptance model in health informatics. Appl Clin Inform 2018;9(3):604-634 [FREE Full text] [doi: 10.1055/s-0038-1668091] [Medline: 30112741]
68. Marques L, Matsubara PG, Nakamura WT, Ferreira BM, Wiese IS, Gadelha BF, et al. Understanding UX better: a new technique to go beyond emotion assessment. Sensors (Basel) 2021;21(21):7183 [FREE Full text] [doi: 10.3390/s21217183] [Medline: 34770490]
69. Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. Med Care 2003;41(5):582-592. [doi: 10.1097/01.MLR.0000062554.74615.4C] [Medline: 12719681]
70. White IR, Carpenter J, Horton NJ. Including all individuals is not enough: lessons for intention-to-treat analysis. Clin Trials 2012;9(4):396-407 [FREE Full text] [doi: 10.1177/1740774512450098] [Medline: 22752633]
71. Kertz SJ, Koran J, Stevens KT, Björgvinsson T. Repetitive negative thinking predicts depression and anxiety symptom improvement during brief cognitive behavioral therapy. Behav Res Ther 2015;68:54-63. [doi: 10.1016/j.brat.2015.03.006] [Medline: 25812825]
72. McLaughlin KA, Nolen-Hoeksema S. Rumination as a transdiagnostic factor in depression and anxiety. Behav Res Ther 2011;49(3):186-193 [FREE Full text] [doi: 10.1016/j.brat.2010.12.006] [Medline: 21238951]
73. Linardon J, Fuller-Tyszkiewicz M. Attrition and adherence in smartphone-delivered interventions for mental health problems: a systematic and meta-analytic review. J Consult Clin Psychol 2020;88(1):1-13. [doi: 10.1037/ccp0000459] [Medline: 31697093]
74. Clement S, Schauman O, Graham T, Maggioni F, Evans-Lacko S, Bezborodovs N, et al. What is the impact of mental health-related stigma on help-seeking? A systematic review of quantitative and qualitative studies. Psychol Med 2015;45(1):11-27. [doi: 10.1017/S0033291714000129] [Medline: 24569086]
75. Bekker MH, van Mens-Verhulst J. Anxiety disorders: sex differences in prevalence, degree, and background, but gender-neutral treatment. Gend Med 2007;4 Suppl B:S178-S193. [doi: 10.1016/s1550-8579(07)80057-x] [Medline: 18156102]
76. Robichaud M, Dugas MJ, Conway M. Gender differences in worry and associated cognitive-behavioral variables. J Anxiety Disord 2003;17(5):501-516. [doi: 10.1016/s0887-6185(02)00237-2] [Medline: 12941362]
77. Muñoz RF, Bunge EL, Chen K, Schueller SM, Bravin JI, Shaughnessy EA, et al. Massive open online interventions: a novel model for delivering behavioral- health services worldwide. Clin Psychol Sci 2015;4(2):194-205. [doi: 10.1177/2167702615583840]

Abbreviations

ANCOVA: analysis of covariance

CONSORT: Consolidated Standards of Reporting Trials
CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile HEalth Applications and on Line Tele Health
ECoWeB: Assessing and Enhancing Emotional Competence for Well-being in the Young project
GAD-7: Generalized Anxiety Disorder questionnaire
MCID: minimum clinically important difference
PHQ-9: Patient Health Questionnaire-9
PSWQ: Penn State Worry Questionnaire
RCT: randomized controlled trial
RNT: repetitive negative thinking
RRS: Ruminative Response Scale
TAM: technology acceptance model
UX: user experience
WEMWBS: Warwick-Edinburgh Mental Well-being Scale

Edited by L Buis; submitted 15.09.23; peer-reviewed by C Jacob, M Pritchard; comments to author 19.02.24; revised version received 25.04.24; accepted 06.05.24; published 13.08.24.

Please cite as:
Edge D, Watkins E, Newbold A, Ehring T, Frost M, Rosenkranz T
Evaluating the Effects of a Self-Help Mobile Phone App on Worry and Rumination Experienced by Young Adults: Randomized Controlled Trial
JMIR Mhealth Uhealth 2024;12:e51932
URL: https://mhealth.jmir.org/2024/1/e51932
doi:10.2196/51932
PMID:

Original Paper

The Impact of Air Pollution Information on Individuals' Exercise Behavior: Empirical Study Using Wearable and Mobile Devices Data

Yang Yang[1], PhD; Khim-Yong Goh[2], PhD; Hock Hai Teo[2], PhD; Sharon Swee-Lin Tan[2], PhD

[1]School of Business and Management, Royal Holloway, University of London, Egham, United Kingdom

[2]School of Computing, National University of Singapore, Singapore, Singapore

Corresponding Author:
Yang Yang, PhD
School of Business and Management
Royal Holloway, University of London
Egham Hill
Egham, TW20 0EX
United Kingdom
Phone: 44 1784 434 455
Email: y.yang@rhul.ac.uk

Abstract

Background: Physical exercise and exposure to air pollution have counteracting effects on individuals' health outcomes. Knowledge on individuals' real-time exercise behavior response to different pollution information sources remains inadequate.

Objective: This study aims to examine the extent to which individuals avoid polluted air during exercise activities in response to different air pollution information sources.

Methods: We used data on individuals' exercise behaviors captured by wearable and mobile devices in 83 Chinese cities over a 2-year time span. In our data set, 35.99% (5896/16,379) of individuals were female and 64% (10,483/16,379) were male, and their ages predominantly ranged from 18 to 50 years. We further augmented the exercise behavior data with air pollution information that included city-hourly level measures of the Air Quality Index and particulate matter 2.5 concentration (in μg/m^3), and weather data that include city-hourly level measures of air temperature (ºC), dew point (ºC), wind speed (m/s), and wind direction (degrees). We used a linear panel fixed effect model to estimate individuals' exercise-aversion behaviors (ie, running exercise distance at individual-hour, city-hour, or city-day levels) and conducted robustness checks using the endogenous treatment effect model and regression discontinuity method. We examined if alternative air pollution information sources could moderate (ie, substitute or complement) the role of mainstream air pollution indicators.

Results: Our results show that individuals exhibit a reduction of running exercise behaviors by about 0.50 km (or 7.5%; P<.001) during instances of moderate to severe air pollution, and there is no evidence of reduced distances in instances of light air pollution. Furthermore, individuals' exercise-aversion behaviors in response to mainstream air pollution information are heightened by different alternative information sources, such as social connections and social media user-generated content about air pollution.

Conclusions: Our results highlight the complementary role of different alternative information sources of air pollution in inducing individuals' aversion behaviors and the importance of using different information channels to increase public awareness beyond official air pollution alerts.

(JMIR Mhealth Uhealth 2024;12:e55207) doi:10.2196/55207

KEYWORDS

air pollution; information sources; exercise activity; wearable and mobile devices; econometric analysis

Introduction

Background

Air pollution remains an extremely serious environmental and health hazard [1]. Although exercising can improve one's health, doing so in polluted air can offset the health benefits and be detrimental to one's cardiovascular and respiratory health [2]. Providing air quality alert information has become an important initiative adopted by government authorities to protect the public from the risk of air pollution [3], with the anticipation that individuals will adjust their behaviors considering this information. Despite the availability of public health information, there is limited evidence to suggest whether individuals truly act upon this information and the extent to which they may adjust their outdoor exercise behaviors. Furthermore, with government agencies providing more detailed information, concerns have arisen over the danger of air pollution alerts losing their prominence, which can diminish the efficacy of information provision as a policy tool [4]. Thus, it is of immense interest to uncover individuals' actual reaction and to what extent individuals adjust their exercise behaviors in response to air pollution information.

Air quality alert information helps individuals to maximize the utility of their outdoor activities while accounting for the health risk from air pollution. Prior studies on the relationship between air pollution information and avoidance behavior primarily focus on the impact of mainstream or official air quality information sources [3,4]. Mainstream or official information on air pollution typically include particulate matter (PM) 2.5 concentration and Air Quality Index (AQI) which indicate the severity and potential health hazards of air pollution. In most countries, public agencies monitor and disseminate such information about a city's pollution parameters through AQI announcements on television, radio, newspaper, and website channels.

Beyond the mainstream sources, there are alternative information sources through which individuals can access air pollution information. Particularly, with the rise of social media, people can access alternative information on air pollution shared by the public on social media platforms and mobile chat apps [5]. Research has shown that social media messages are strongly correlated with AQI and are indicative of true particle pollution levels [6]. Despite the increasing impact of user-generated content on social media [7] and the availability of other information sources (eg, web–based social connections and mobile apps), there is a lack of knowledge in the literature on whether these alternative information sources may moderate (ie, substitute or complement) the effect of mainstream pollution information sources on individuals' exercise aversion behaviors.

Furthermore, current understanding of the impact of air pollution on individuals' exercise behavior is limited, and existing research evidence is mainly built on self-reported survey data [8]. These data can suffer from recall errors and social desirability bias [9], and is limited to provide detailed spatial and temporal attributes of exercise behaviors. However, air pollution levels can vary across different specific locations and fluctuate across time. Thus, it becomes critical to provide scientific evidence using real-time, disaggregate spatial and temporal data to understand the extent to which individuals avert exposure to polluted air in their physical activities in real time (ie, on an hourly basis), in response to different sources of air pollution information. According to recent reports on personal wellness and fitness, there is a rise in consumers' adoption of digital health services and technology-driven fitness programs. More consumers are also using wearable activity trackers, such as smart wristbands and fitness apps, which offer instant feedback on various exercise metrics [10]. Data from these wearable fitness devices and applications enable researchers to address the limitations previously mentioned.

Objective

We aimed to evaluate the extent to which individuals avoid polluted air during exercise activities, in response to different air pollution information sources. We also examined whether alternative air pollution information sources moderate the effect of mainstream information sources on individuals' exercise behaviors.

Methods

Data

In China, air pollution continues to persist at a concerning level and affects the economy and people's overall quality of life [11]. Research has shown that air pollution in northern regions of China led to a 5.5 year reduction in life expectancies of residents, compared with those in southern regions [12]. The alarming rate of air pollution makes China a valid and representative context to investigate the research questions.

We conducted empirical analyses using a novel data set of over 16,000 individuals' running exercise behaviors in 83 Chinese cities over a 2-year time span from January 2013 to November 2014, as captured by wearable devices and mobile fitness apps. This sample period is particularly appropriate because there was a confluence of multiple episodes of severe air pollution in China and the mass adoption of wearable and mobile devices from 2013 to 2014. The data on individuals' exercise activities were obtained from a China-based, web and mobile fitness platform that offers products (eg, smart wristband and mobile fitness apps) for tracking exercise activities which incorporate social network features. Users connect to the platform through a wearable device, web browser or mobile app, which enables them to track exercise activities, share their progress with peers, and leverage social networking features to foster support and encouragement among users. All data generated during physical exercise activities are synchronized to the platform's data servers and is viewable by users through the mobile app and website. This individual-level minute-by-minute exercise behavior data was further supplemented with hourly weather controls data and various air pollution information sources such as AQI and PM2.5 measures, social connections, Weibo tweets (a Chinese Twitter-like platform), and mobile apps.

Our exercise behavior data set includes information on individuals' user anonymous ID, sex, total cumulative distance (km) of exercise activity, reward points balance, exercise skill grade level, home and exercise city location, exercise distance

(km), exercise speed (km/h), exercise type (eg, running or walking), exercise tracking date and time (for both start and end of exercise session), tracking device type (eg, smart wristband or mobile fitness app), and calories expended. We removed, from our raw data set, exercise records or observations that had various issues such as duplicated records, missing exercise and home city locations, and outliers in terms of exercise distances (ie, top and bottom 0.5 percentiles, corresponding to likely errors in distances recorded that were >100 km or <0.01 km, for example). We focused only on running and excluded walking, as our data set cannot differentiate whether a physical activity was conducted indoor or outdoor, and walking can occur indoors from point to point. In contrast, running is a recreational physical activity that is mostly conducted outdoors in China [13], and thus is susceptible to the effects of air pollution. Our final data set consisted of 447,666 observations for 16,379 individuals with running exercise records across 83 cities in China over a roughly 2-year time span. In our final data set, 35.99% (5896/16,379) of users were female and 64% (10,483/16,379) were male. We do not have access to the users' age data; however, statistics from the focal fitness platform we analyzed suggest that their users' age ranged predominantly from 18 to 50 years [14]. Furthermore, in terms of the users' exercise behavior, the average exercise distance was around 6.61 km for the users in all the 83 cities, with an average exercise speed of around 8.74 km/h. At the weekly level, the average weekly exercise frequency was around 3.5 times, with an average distance of 7.13 km exercised each week.

We augmented the exercise behavior data with 2013 to 2014 air pollution information that includes city-hourly level measures of AQI and PM2.5 concentration (in $\mu g/m^3$) from both the Chinese Ministry of Environmental Protection (CN-MEP; provided data for 83 cities) and the US Department of State (US-DOS; it provided data for 5 cities with embassies located in Beijing, Chengdu, Guangzhou, Shanghai, and Shenyang). Finally, for control purposes in the empirical analysis, we acquired the 2013 to 2014 National Oceanic and Atmospheric Administration Integrated Surface Database's weather data for 83 cities that include city-hourly level measures of air temperature (ºC), dew point (ºC), wind speed (m/s), and wind direction (degrees). Table 1 shows the descriptive statistics for our data set at the individual user–level for both types of air quality indicator (ie, Air Quality Index-China [AQI-CN] and particulate matter 2.5-United States [PM2.5-US]) across the 5 cities of Beijing, Chengdu, Guangzhou, Shanghai and Shenyang, as well as across all 83 cities in our data set.

Table 1. Descriptive statistics (mean (SD)): individual-level data.

Variables	AQI-CN[a] (5 cities)	AQI-CN (83 cities)	PM2.5-US[b] (5 cities)
Exercise distance (km)	6.648 (4.423)	6.611 (4.413)	6.659 (4.438)
Air quality (PM2.5 or AQI)	89.983 (56.321)	88.339 (54.842)	63.515 (54.438)
50<air quality≤100	0.468 (0.499)	0.486 (0.500)	0.316 (0.465)
100<air quality≤150	0.177 (0.382)	0.182 (0.386)	0.105 (0.307)
150<air quality≤200	0.071 (0.257)	0.062 (0.242)	0.038 (0.190)
200<air quality≤300	0.046 (0.210)	0.037 (0.189)	0.021 (0.142)
300<air quality≤500	0.008 (0.089)	0.008 (0.092)	0.007 (0.081)
Air quality>500	—[c]	—	0.001 (0.023)
Exercise speed (km/h)	8.706 (5.714)	8.738 (12.398)	8.693 (4.067)
Temperature (ºC)	18.803 (9.181)	19.345 (9.013)	18.717 (9.226)
Dew point (ºC)	11.418 (11.470)	12.620 (10.678)	11.292 (11.485)
Wind speed (m/s)	27.417 (19.201)	26.815 (18.241)	27.530 (19.250)
City			
Beijing	0.350 (0.477)	0.137 (0.343)	0.357 (0.479)
Chengdu	0.081 (0.273)	0.032 (0.175)	0.078 (0.269)
Guangzhou	0.175 (0.380)	0.068 (0.252)	0.167 (0.373)
Shanghai	0.370 (0.483)	0.144 (0.351)	0.375 (0.484)
Shenyang	0.024 (0.153)	0.009 (0.096)	0.023 (0.149)
Observations	174,654	447,666	172,352
Number of individuals	7165	16,379	7146

[a]AQI-CN: Air Quality Index-China.

[b]PM2.5-US: particulate matter 2.5-United States.

[c]No recorded instance of air quality in this range.

Air Pollution Information Sources

In China, mainstream air quality and pollution information source is the official pollutant standards index such as the AQI (ie, CN-MEP) communicated by government agencies. Beyond the mainstream information source, there are multiple alternative sources or channels from which individuals can acquire air pollution information. First, social media Weibo tweets can serve as an alternative form of information on air pollution [6]. There is a strong correlation between the Air Discussion Index constructed based on Weibo tweets and the actual measured PM in the air [6,15]. Second, for the 5 cities of Beijing, Chengdu, Guangzhou, Shanghai, and Shenyang, AQI and PM2.5 information provided by US-DOS can serve as an alternative information source corresponding to the mainstream one provided by CN-MEP. Third, we consider the case of mobile apps as an alternative information channel for air quality information [16]. Specifically, we focus on the China AQI mobile app (renamed as Air Matters in 2016), which was the pioneer for this category of apps and released their version 4.0 with over 160 cities coverage on May 10, 2013 [17]. This version added monitoring stations data from >800 monitoring stations in over 160 cities, constituting a substantial 33% increase in city level coverage of air quality information for mobile phone users. Fourth, there can be anchoring effects where individuals possess alternative or self-referent local information such as the prevailing levels of air pollution in a city [18], which may not be known to new visitors exercising in a location that is not their home city. Finally, individuals can access air pollution information from their connected peers on the focal fitness platform. Prior studies suggest that there can be social contagion in exercise behaviors [19] and transmission of information through these social connections [7].

Statistical Analysis

Descriptive Analysis

Figures 1 and 2 show a spatial plot of the means and SDs of exercise distances by 5 different bins or ranges of AQI-CN (ie, AQI data from CN-MEP) measures in January 2013 across China's provincial-level administrative units. It is clear from Figure 1 that the mean running exercise distance is lower in the more heavily polluted provinces such as Hebei, Beijing, Tianjin, and Shandong (ie, the cluster of eastern provinces shaded in red), compared with the less polluted provinces along its southern coastline such as Guangdong, Fujian, Zhejiang, Shanghai, and Jiangsu. In terms of the SDs of exercise distance, Figure 2 shows that the variability of running exercise distance is higher in the less polluted provinces shaded in blue, light blue, and light green. Therefore, there is tentative descriptive evidence of individuals' exercise-aversion behaviors in response to air pollution in various Chinese provinces.

Figure 1. Mean of exercise distance. AQI: Air Quality Index.

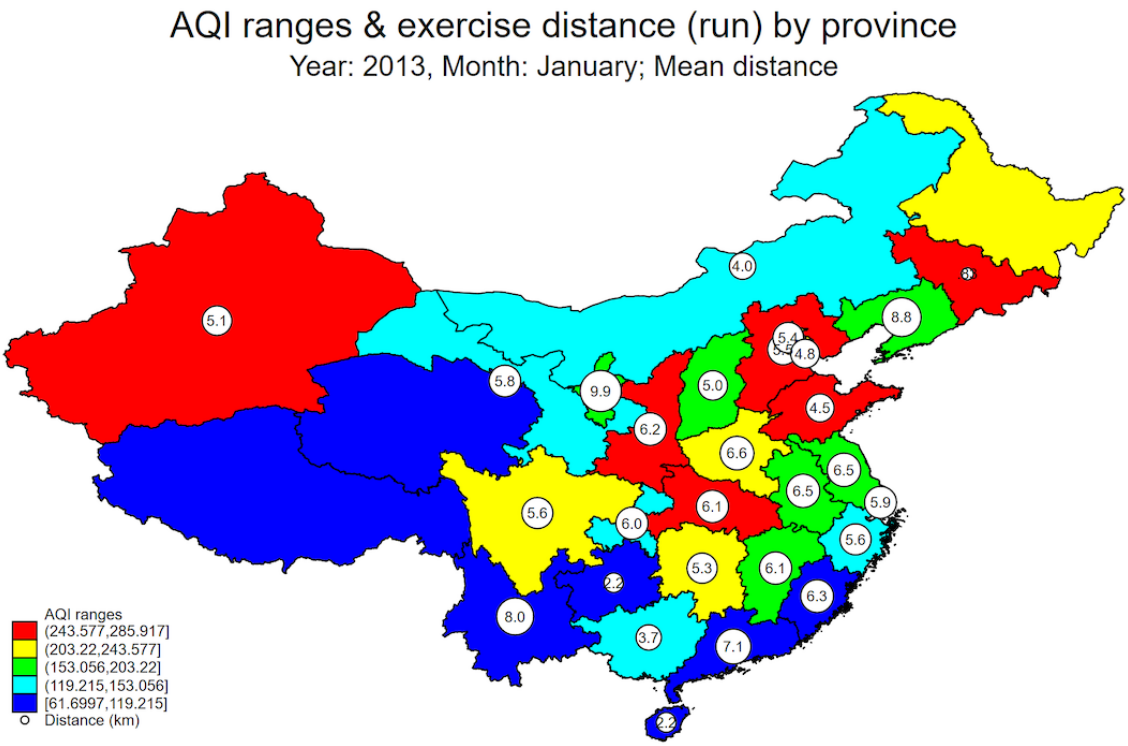
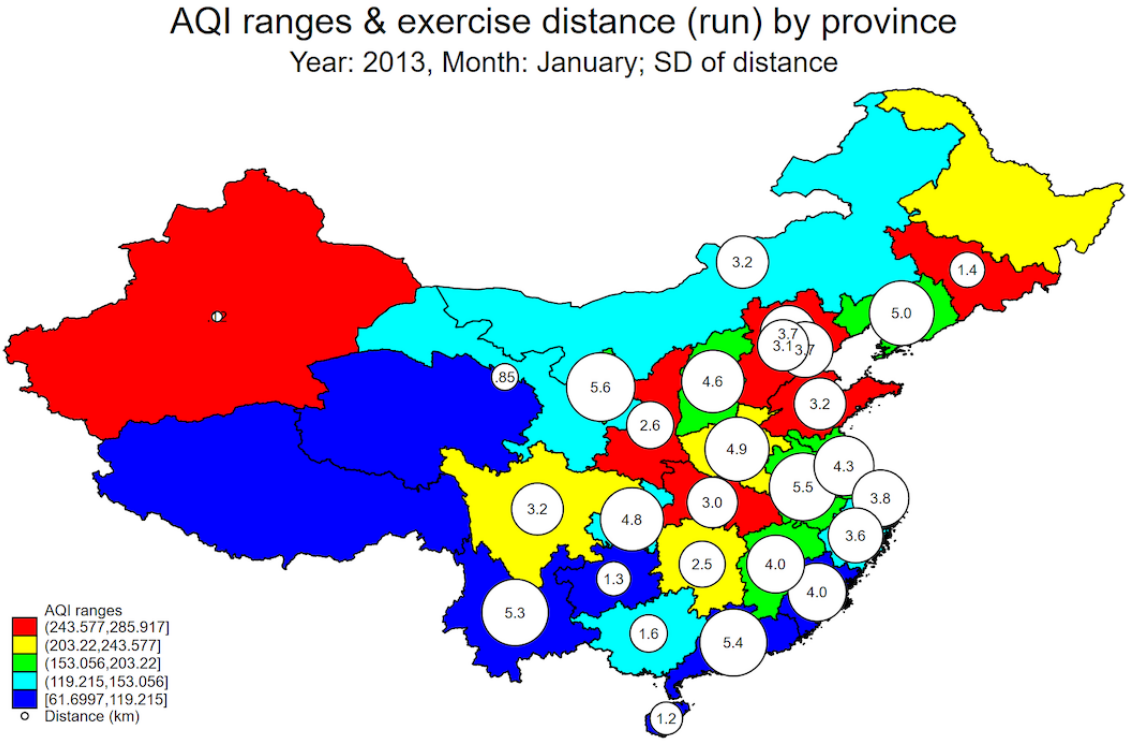


Figure 2. SD of exercise distance. AQI: Air Quality Index.



Linear Fixed Effect Model

We first specified a linear panel model specification for our focal variable of interest, that is, running exercise distance at different units of analysis (ie, individual-hour, city-hour, or city-day levels). Subscript *i* in equation (1) below refers to either an individual or a city (in a specific province shown in Figure 1), and *t* refers to the time unit of either hour or day.

$ExerDist_{it}$ captures the running exercise distance for an individual or a city at either the hourly or aggregated daily level. $AirQual_{it}$ is a vector of air quality indicators measuring the extent of air pollution at either the hourly or daily level, specific to the city where an individual conducts the running physical activities or exercises. We allowed for nonlinear effects of air pollution in our model specification by discretizing pollution levels through the use of dummy variables, which is a commonly used approach in the literature [20]. Specifically, we included dummy variables that correspond to the different pollution thresholds based on the CN-MEP standard of AQI (AQI 0-50: excellent, 51-100: good, 101-150: lightly polluted, 151-200: moderately polluted, 201-300: heavily polluted, and >300: severely polluted). In addition, in our model, we accounted for an individual's exercise speed ($ExerSpeed_{it}$), city-hour or day level temperature ($Temp_{it}$), dew point ($DewPt_{it}$), and wind speed ($WindSpeed_{it}$). Other controls include various dummy variables corresponding to the specific year, month, day of week, time of day, and city information during the incidence of an individual's running exercise activity. Furthermore, we performed various placebo tests to verify that it is indeed the air pollution levels in each specific city and specific date and hour that are driving the specific aversion responses of individuals' exercise behaviors (Table S1 in Multimedia Appendix 1).

Endogenous Treatment Effect Model

Daily and hourly variations (ie, the time unit of analysis for this study) of air pollutants can plausibly be considered as exogenous [21]. As air pollution is not randomly assigned to individuals, studies that attempt to compare aversion behaviors or health outcomes for populations exposed to varying pollution levels may not be adequately accounting for potential confounding factors. It is known that air quality is capitalized into real estate prices [22]. Thus, households with higher incomes or preferences for cleaner air are likely to sort into locations with better air quality. Failure to account for this sorting will lead to overestimates of the effects of pollution. Alternatively, pollution levels are typically higher in urban areas where there are often more educated individuals with better access to health care, leading to underestimates of the true effects of pollution.

To control for the potential endogeneity of mainstream air pollution indicators of PM2.5 and AQI measures, we used the endogenous treatment effect model [23] to model running exercise distance while accounting for the potential endogeneity (due to selection or sorting) of moderate to severe air pollution (ie, as indicated by the binary treatment dummy for AQI/PM2.5>150, ie, *AirPolluted*). The endogenous treatment effect model estimates an average treatment effect and the other parameters of a linear regression model for exercise distance, augmented with an endogenous binary-treatment variable for moderate to severe air pollution. The model specification is as follows:

where ε_{it} and u_{it} are bivariate normal with mean zero and covariance matrix Σ.

In our focal equation (2) of interest, $ExerDist_{it}$ captures the running exercise distance for an individual or a city at either the hourly or aggregated daily level. $AirPolluted_{it}$ is an individual or city-specific hourly or daily binary indicator for AQI-CN>150, that is, when air quality crosses into the moderate to severe air pollution range (ie, the same pollution threshold in literature [24], which have a substantial effect on individuals' decision to purchase health insurance). We also account for an individual's exercise speed ($ExerSpeed_{it}$) and weather conditions such as temperature ($Temp_{it}$), dew point ($DewPt_{it}$), and wind speed ($WindSpeed_{it}$). Appropriate control variables here include users' point balance and exercise skill grade on the focal fitness platform, and fixed effect dummies corresponding to the specific calendar year, month, day of week, and time of day information. In the treatment equation (3), the $AirPolluted_{it}$ indicator is a probit function of city-hour or day level wind speed ($WindSpeed_{it}$), wind direction ($WindDir_{it}$), air pressure ($AirPressure_{it}$), humidity ($Humidity_{it}$), and other controls such as calendar month and day of week fixed effects. Equations (2) and (3) above are estimated by maximum likelihood estimation.

Regression Discontinuity

To further bolster the confidence of our estimate for the causal effect of air pollution on exercise aversion behaviors, we used the regression discontinuity method [25,26] as a robustness check to estimate the effect of air pollution severity crossing specific AQI categorical thresholds on individuals' average running exercise distances at the city-hour level. We used the standard or continuity-based framework for regression discontinuity analysis, and specifically a sharp regression discontinuity design.

The regression discontinuity design can be used to isolate a treatment effect of interest from all other systematic differences between treated and control groups. Under appropriate assumptions [25,26], a comparison of individuals and cities for which the AQI indicators are barely below the moderate or severe pollution threshold and those for which the AQI indicators are barely above the same threshold will reveal the causal (local) effect of air pollution on exercise behaviors. If individuals and cities cannot systematically manipulate the air pollution indicators, observations just above and just below the cutoff will tend to be comparable in terms of all characteristics. Thus, right at the cutoff, the comparison is free of the complications introduced by systematic observed and unobserved differences between the treatment and control groups. Our regression discontinuity estimators are based on mean-squared error-optimal bandwidths. Further details of the regression discontinuity procedures and the related robustness checks are provided in Table S2 in Multimedia Appendix 2.

Moderation Effects Analysis: Alternative Information Sources

According to the Protection Motivation Theory, individuals' motivation to engage in protective behaviors is influenced by 2 cognitive appraisal processes, that is, threat appraisal and coping appraisal [27]. In the context of our study, the Protection Motivation Theory is helpful to understand how individuals assess the risk of air pollution to their health and their perceived ability to cope with these risks through personal actions, such as adjusting their outdoor exercise activities. The availability of alternative information sources or channels, such as social media platforms, play an important role in shaping these appraisals [28]. The alternative information sources can provide additional narratives or evidence about air pollution, which can affect the cognitive process of threat and coping appraisals. Specifically, the alternative information sources may amplify or attenuate individuals' perception of the threat (ie, the severity of the air pollution) and their belief in their capacity to take protective actions, such as modifying exercise routines. The interplay between alternative information sources and the cognitive appraisal processes emphasized by the Protection Motivation Theory underscores the necessity and significance of understanding the moderation effect of alternative information sources on the impact of mainstream air pollution information on individuals' exercise aversion behaviors. We further conduct empirical analyses using an individual-level linear fixed effect model in equation (1) to evaluate if alternative sources of air pollution information have interaction effects with mainstream air pollution information, such that they either mitigate or accentuate (ie, substitute or complement) the impact of mainstream air pollution indicators on individuals' exercise aversion behaviors.

First, we estimated if individuals' exercise averting response to the mainstream AQI-CN pollution measures will be accentuated with increasing volumes of social media Weibo tweets of Chinese language phrases associated with air pollution. Social media can serve as an alternative form of salient confirmatory or reinforcement information on air pollution. Social media Weibo tweets can amplify official messages and contribute to a sense of urgency or provide practical coping strategies for individuals [6], which may heighten perceived severity and vulnerability and likely lead to a stronger motivation to engage in protective behaviors.

Second, we evaluated whether individuals located in the 5 cities of Beijing, Chengdu, Guangzhou, Shanghai, and Shenyang with the alternative PM2.5-US information respond more dramatically in exercise aversion behaviors, compared with those in the other 78 cities with no such alternative pollution information. Information source and credibility significantly influence how individuals interpret and act on information about air pollution [29], with the US-DOS being regarded as a reliable source, potentially strengthening individuals' evaluation of the risk, and thus, promoting protective behaviors. We further estimated the effects using a neighboring matched-city approach (ie, matching Beijing to neighboring Tianjin, Chengdu to Chongqing, Guangzhou to Shenzhen, Shanghai to Hangzhou, and Shenyang to Dalian where there are no alternative PM2.5-US information).

Third, we evaluated if individuals' exercise aversion responses to the mainstream AQI-CN pollution indicator are affected by the increased 160 cities coverage of air quality information in the China AQI mobile app. The AQI mobile app can provide real-time data on pollution levels and health-related alerts about the specific risks of current air quality levels, which can heighten individuals' perception of the immediate risk and motivation to take protective action.

Fourth, we estimated whether individuals exercising in a location or city which is their home city alter their exercise behaviors to a larger extent during episodes of air pollution. Locals with anchoring knowledge for a specific home city have a coping advantage as they can refer to their experience to make informed decisions [30]. However, new visitors to a city may not possess local or home knowledge about the typical air pollution levels in a city and are likely to underestimate or overestimate the threat, affecting their coping appraisal and potential protective behaviors. This anchoring effect can therefore affect their perceived severity and coping appraisal.

Finally, we examined if individuals' exercise aversion behaviors are moderated by the numbers and types of social connections on the focal fitness platform. Information from connected peers often carries important weight because it comes from a trusted and relatable source [31], so the air pollution information shared by connected peers may increase individuals' awareness and perceived severity. We considered different types of social connections, that is, followers (incoming connections for the focal individual), followees (outgoing connections initiated by the focal individual), and mutual followers (individuals who are both the followers and followees for the focal individual).

Ethical Considerations

All the data in this study were collected in a manner that ensures the anonymity and privacy of individuals in our sample. Each user in our sample has a unique anonymous user ID, which is randomly generated by the focal fitness platform and has no connection to the participant's actual identity. This study was reviewed and approved by the Departmental Ethics Review Committee of School of Computing at the National University of Singapore. The DERC Case No. is SOC-22-03.

Results

Table 2 presents the model estimation results using the AQI-CN air quality indicators. Using a fixed effect model estimation approach at the individual-hour, city-hour, and city-day levels (ie, columns 1 to 3), we find significant evidence of individuals' pollution aversion behavior in terms of decreasing their running exercise distances when air pollution worsened. Such aversion behaviors in terms of decreasing their running exercise distances were more sensitive or in larger magnitudes during incidences of moderate to severe air pollution. At the individual and city-hourly level, when the air was moderately polluted (ie, 150<AQI-CN≤200), running exercise distances decreased by about 0.24 km (P<.001) to 0.32 km (P<.001) relative to when air quality was excellent (ie, 0<AQI-CN≤50). When the air was severely polluted (ie, 300<AQI-CN≤500), running distances dropped by about 0.72 km (P=.03) at the city-hour level and 0.73 km (P=.002) at the city-day level. The results from the endogenous treatment effect models show general consistency with those from the linear fixed effect models (Table 2). At the individual and city-hourly level, running distances decreased by 0.48 km (P<.001) to 0.87 km (P=.07) in episodes of moderate to severe air pollution. Furthermore, the results of regression discontinuity estimations confirm that there was a decrease in running exercise distance of 0.29 km (P=.07, without covariates) at the AQI-CN cutoff value of 150 for moderately polluted air conditions (Table S3 in Multimedia Appendix 2). Generally, these results on the extent of exercise aversion in terms of decreased running distances also corroborate those from the linear fixed effect and endogenous treatment effect models.

For the moderation effects of alternative information sources, first, we find that as the volumes of tweets (in millions) about the Chinese-language words for "haze" and "face mask" (which reflect health concerns and self-protection measures [12,32]) increase, there was a larger extent of decrease in running exercise distances during episodes of moderate to heavy air pollution. In particular, a 1 SD increase in the volume of tweets about "haze" and "face mask" (ie, 0.21 and 0.026 million tweets respectively) was associated with a drop of about 0.09 km (P=.02) to 0.12 km (P<.001) in running distances during heavily polluted episodes (ie, 200<AQI-CN≤300; Table 3).

Second, we find that relative to those in the other 78 cities, individuals exercising in the 5 cities of Beijing, Chengdu, Guangzhou, Shanghai, and Shenyang with alternative PM2.5-US information did tend to decrease running distances to a larger extent by about 0.25 km (P<.001) during light to moderate air pollution episodes (Table 3). Similarly, using a neighboring matched-city approach (ie, matching Beijing to neighboring Tianjin, Chengdu to Chongqing, Guangzhou to Shenzhen, Shanghai to Hangzhou, and Shenyang to Dalian where there are no alternative PM2.5-US information), the results further confirm that individuals in the 5 cities with alternative PM2.5-US information reduced running distances by 0.19 km (P=.002) to 0.22 km (P<.001) even in instances of low to light air pollution (Table 3).

Third, considering the case of mobile apps as an alternative channel for air quality information, we find that individuals exercising after the increased 160 cities coverage of air quality information in the China AQI mobile app, did significantly decrease their running exercise distances by about 0.30 km (P=.02) and 0.60 km (P=.004) during heavy and severe air pollution episodes respectively (Table 3).

Table 2. Fixed effect and endogenous treatment models for Air Quality Index-China ranges (Chinese Ministry of Environmental Protection) [model coefficient (*P* value) reported].

Variables[a]	(1) FE[b,c] (individual-hour level; 5 cities)	(2) FE[d] (city-hour level; 5 cities)	(3) FE[e] (city-day level; 5 cities)	(4) ET[c,f] (individual level; 5 cities)	(5) ET[d] (city-hour level; 5 cities)	(6) ET[e] (city-day level; 5 cities)
AQI-CN[g,h]>150	—[i]	—	—	−0.479 (<.001)	−0.872 (.07)	−0.469 (.14)
50<AQI-CN≤100	−0.082 (.002)	−0.047 (.12)	−0.201 (.07)	—	—	—
100<AQI-CN≤150	−0.163 (<.001)	−0.174 (.01)	−0.254 (.08)	—	—	—
150<AQI-CN≤200	−0.242 (<.001)	−0.319 (<.001)	−0.291 (<.001)	—	—	—
200<AQI-CN≤300	−0.130 (.02)	−0.353 (.08)	−0.595 (.03)	—	—	—
300<AQI-CN≤500	−0.205 (.15)	−0.715 (.03)	−0.731 (.002)	—	—	—
Exercise speed (km/h)	0.029 (.08)	0.116 (.002)	0.092 (.10)	0.054 (.10)	0.117 (.005)	0.144 (.04)
Temperature (ºC)	−0.002 (<.001)	−0.002 (.05)	0.001 (<.001)	−0.002 (.04)	−0.001 (.32)	0.001 (.32)
Dew point (ºC)	0.001 (.01)	−0.000 (<.001)	−0.003 (.003)	−0.000 (<.001)	0.001 (<.001)	−0.001 (.32)
Wind speed (m/s)	−0.001 (.31)	−0.002 (.05)	−0.004 (<.001)	−0.002 (.04)	0.001 (.32)	−0.003 (.32)
Constant	6.551 (<.001)	6.889 (<.001)	6.656 (<.001)	4.215 (<.001)	7.079 (<.001)	6.257 (<.001)
Observations	174,654	33,617	3269	162,858	30,923	2978
R^2 or log-likelihood	0.016	0.041	0.095	−519,371	−92,632	−6714
Number of individuals, n	7165	—[i]	—	7023	—	—

[a]Dependent variable: exercise distance (km).

[b]FE: fixed effect model.

[c]Controls for (1), (4): year, month, day of week, time of day, city dummies; individual's reward points, and exercise skill grade.

[d]Controls for (2), (5): year, month, day of week, time of day dummies.

[e]Controls for (3), (6): year, month, day of week dummies.

[f]ET: endogenous treatment model.

[g]AQI-CN: Air Quality Index-China.

[h]Covariates for (AQI-CN>150) treatment indicator: wind speed, wind direction, air pressure, humidity, month and day of week dummies.

[i]Not applicable.

Table 3. Moderation effects of alternative information sources (part 1) [model coefficient (*P* value) reported].

Variables[a]	(1) Moderation factor: Weibo (*haze*; 雾霾); 83 cities (all including BJ[b]-CD[c]-GZ[d]-SH[e]-SY[f])	(2) Moderation factor: Weibo *(face mask;* 口罩); 83 cities (all including BJ-CD-GZ-SH-SY)	(3) Moderation factor: particulate matter 2.5–United States; 83 cities (all including BJ-CD-GZ-SH-SY)	(4) Moderation factor: particulate matter 2.5–United States; 10 cities (BJ-CD-GZ-SH-SY, TJ[g]-CQ[h]-SZ[i]-HZ[j]-DL[k])	(5) Moderation factor: 160 cities coverage in mobile app; 83 cities (all including BJ-CD-GZ-SH-SY)
50<air quality≤100	0.021 (.29)	0.036 (.10)	0.107 (<.001)	0.141 (<.001)	0.056 (.41)
100<air quality≤150	−0.044 (.09)	−0.032 (.25)	0.073 (.02)	0.032 (.56)	0.167 (.03)
150<air quality≤200	−0.118 (<.001)	−0.018 (.66)	−0.002 (.96)	−0.107 (.21)	−0.151 (.11)
200<air quality≤300	−0.018 (.70)	0.040 (.43)	−0.044 (.41)	−0.199 (.06)	0.207 (.05)
300<air quality≤500	−0.077 (.44)	−0.218 (.04)	−0.157 (.10)	−0.733 (.02)	−0.047 (.75)
50<air quality≤100 * moderation factor[l]	−0.002 (.97)	−0.712 (.21)	−0.194 (<.001)	−0.222 (<.001)	−0.019 (.81)
100<air quality≤150 * moderation factor[l]	0.049 (.59)	−0.346 (.61)	−0.248 (<.001)	−0.194 (.002)	−0.157 (.08)
150<air quality≤200 * moderation factor[l]	0.005 (.97)	−3.370 (<.001)	−0.250 (<.001)	−0.141 (.13)	0.025 (.82)
200<air quality≤300 * moderation factor[l]	−0.408 (.02)	−4.572 (<.001)	−0.095 (.22)	0.072 (.54)	−0.297 (.02)
300<air quality≤500 * moderation factor[l]	−0.411 (.12)	0.758 (.74)	−0.004 (.98)	0.546 (.10)	−0.597 (.004)
Moderation factor	−0.000 (<.001)	0.000 (<.001)	0.203 (.004)	0.194 (.009)	0.016 (.81)
Constant	6.216 (<.001)	6.197 (<.001)	6.065 (<.001)	5.918 (<.001)	5.269 (<.001)
Controls included[m]	Yes	Yes	Yes	Yes	Yes
Observations, n	444,193	444,193	447,666	258,648	88,341
R^2	0.015	0.015	0.015	0.016	0.025
Number of individuals, n	16,336	16,336	16,379	10,114	6989
Roy-Zellner *F* test	1.745	4.871	7.566	6.005	2.980
Prob>*F*	0.121	0.000187	4.16e-07	1.49e-05	0.0109

[a]Dependent variable: exercise distance (km).

[b]BJ: Beijing.

[c]CD: Chengdu.

[d]GZ: Guangzhou.

[e]SH: Shanghai.

[f]SY: Shenyang.

[g]TJ: Tianjin.

[h]CQ: Chongqing.

[i]SZ: Shenzhen.

[j]HZ: Hangzhou.

[k]DL: Dalian.

[l]Moderation factor in the interaction term. The specific moderation factor for each column is indicated in the second row for moderation factor.

[m]Controls: temperature, dew point, wind speed; year, month, day of week, time of day, city dummies; individual's exercise speed, reward points, and exercise skill grade.

Fourth, in all 83 cities, individuals exercising in a city which is their home tend to be more responsive in decreasing the extent of running distances during polluted air. Such a moderation effect increased in magnitude across light to heavy and severe pollution episodes, with the drop in running distances ranging from 0.14 km (*P*=.003) to 0.40 km (*P*=.01) for individuals exercising in their home locations (Table 4). Similarly, individuals exercising in the 5 cities of Beijing, Chengdu, Guangzhou, Shanghai, and Shenyang exhibited this similar home-location moderation effect of 0.16 km (*P*=.01) to 0.60 km (*P*=.07) across light to heavy and severe pollution (Table 4).

Finally, we observe a negative moderation effect of social connections on the physical exercise aversion response to mainstream pollution information sources. Specifically, the numbers of followers, followees, and mutual connections for a focal individual accentuate the decrease in exercise behaviors during polluted air, even in good or lightly polluted conditions (Table 4). We note that the extent of this negative moderation effect of 1 additional social connection on individuals' exercise behaviors during lightly polluted conditions (ie, a roughly 0.15 km drop; *P*=.001) is almost double that during good air conditions (ie, 0.08 km decrease; *P*=.02).

Table 4. Moderation effects of alternative information sources (part 2) [model coefficient (*P* value) reported].

Variables[a]	(1) Moderation factor: same exercise and home location; 83 cities (all including BJ[b]-CD[c]-GZ[d]-SH[e]-SY[f])	(2) Moderation factor: same exercise and home location; 5 cities (BJ-CD-GZ-SH-SY)	(3) Moderation factor: social connections (followers); 5 cities (BJ-CD-GZ-SH-SY)	(4) Moderation factor: social connections (followees); 5 cities (BJ-CD-GZ-SH-SY)	(5) Moderation factor: social connections (mutual); 5 cities (BJ-CD-GZ-SH-SY)
50<air quality≤100	0.038 (.07)	−0.046 (.18)	−0.044 (.09)	−0.056 (.03)	−0.058 (.03)
100<air quality≤150	0.022 (.42)	−0.072 (.11)	−0.108 (.001)	−0.118 (<.001)	−0.120 (<.001)
150<air quality≤200	−0.085 (.02)	−0.282 (<.001)	−0.213 (<.001)	−0.225 (<.001)	−0.223 (<.001)
200<air quality≤300	0.001 (.98)	0.164 (.09)	−0.080 (.15)	−0.096 (.08)	−0.092 (.09)
300<air quality≤500	0.017 (.87)	0.266 (.37)	−0.142 (.31)	−0.136 (.33)	−0.151 (.28)
50<air quality≤100 * moderation factor[g]	−0.034 (.38)	−0.067 (.21)	−0.102 (.006)	−0.080 (.02)	−0.086 (.04)
100<air quality≤150 * moderation factor[g]	−0.142 (.003)	−0.161 (.01)	−0.154 (.002)	−0.149 (.001)	−0.160 (.004)
150<air quality≤200 * moderation factor[g]	−0.075 (.23)	0.042 (.63)	−0.077 (.11)	−0.052 (.24)	−0.068 (.17)
200<air quality≤300 * moderation factor[g]	−0.188 (.02)	−0.419 (<.001)	−0.136 (.06)	−0.107 (.09)	−0.138 (.08)
300<air quality≤500 * moderation factor[g]	−0.400 (.01)	−0.595 (.07)	−0.156 (.22)	−0.189 (.13)	−0.173 (.47)
Moderation factor	0.365 (.001)	0.592 (.049)	—[h]	—	—
Constant	5.874 (<.001)	6.284 (<.001)	6.547 (<.001)	6.551 (<.001)	6.550 (<.001)
Controls included[i]	Yes	Yes	Yes	Yes	Yes
Observations, n	447,666	174,654	174,654	174,654	174,654
R^2	0.015	0.016	0.016	0.016	0.016
Number of individuals, n	16,379	7165	7165	7165	7165
Roy-Zellner *F* test	2.271	4.723	2.162	2.693	2.013
Prob>*F*	0.0448	0.000261	0.0554	0.0195	0.0736

[a]Dependent variable: exercise distance (km).

[b]BJ: Beijing.

[c]CD: Chengdu.

[d]GZ: Guangzhou.

[e]SH: Shanghai.

[f]SY: Shenyang.

[g]Moderation factor in the interaction term. The specific moderation factor for each column is indicated in the second row for moderation factor.

[h]The main effect of social connections is included in the linear panel models, but its coefficient is dropped in the above fixed effect specification since the variable is nontime varying.

[i]Controls: temperature, dew point, wind speed; year, month, day of week, time of day, city dummies; individual's exercise speed, reward points, and exercise skill grade.

Discussion

Principal Findings

We examine the extent to which individuals avoid exposure to polluted air during exercise activities, in response to different sources of air pollution information, and whether alternative information sources moderate the effect of mainstream air pollution information sources on such exercise behaviors. Our findings show that (1) individuals exhibited a decrease in running exercise behaviors by about 0.50 km in instances of moderate to severe air pollution and (2) the alternative sources of air pollution information, such as social media user–generated content about air pollution and social connections, complement the role of mainstream air pollution information in inducing individuals' exercise-aversion behaviors.

Contrary to popular belief that people know how to cope with air pollution, such as reducing their outdoor physical exercising, our results show that individuals only reduce their running distances by about 0.50 km (or 7.5% on average across individuals in our sample) in instances of moderate to severe air pollution, and there is no causal evidence of reduced distances in instances of light air pollution. This averting response is less pronounced compared with that in previous studies based on different outcome metrics (eg, 8%-15% drop in outdoor facility attendances [21]; 14%-35% reduction in amount of cycling [33]). Individuals do seem to weigh the tradeoff of outdoor physical activities in the face of air pollution, but they did not exhibit a large extent of avoidance behaviors as expected. Drawing on the Protection Motivation Theory and related literature, there could be several potential explanations. First, individuals may not perceive air pollution as a severe health threat, or they have optimism bias about susceptibility to harm, which individuals consider themselves as less at risk than their peers regardless of their age and sex [34], and thus tend to view themselves as different from the susceptible group when assessing the risk of air pollution. If the perceived threat of air pollution is low, their motivation to change behavior such as reducing outdoor exercise will be low as well. Second, prior findings revealed that people who are healthy without asthma or cardiovascular disease are less likely to modify their behavior in the face of air pollution [35,36]. The participants in our sample, who are generally younger and in relatively good physical fitness [14,37], tend to underestimate the harmful impacts of air pollution during exercising, especially in the instances of light air pollution. Third, even if individuals recognize the risk of air pollution, they might feel incapable of changing their exercise routines due to personal constraints, such as lack of access to indoor exercise facilities or a strong preference for outdoor activities. Individuals might perform a cost-benefit analysis, either consciously or subconsciously, weighing the benefits of exercising outdoors against the potential health hazards of air pollution [38]. Individuals who perceive the benefits of outdoor exercise, such as enjoyment, social interactions, outweigh the potential negative impacts of exercising in polluted air, are less inclined to adjust their exercise behaviors.

Prior studies mainly focus on the air pollution information provided by official mainstream channels [3,4]. Our findings highlight the importance of alternative air pollution information sources beyond official channels in affecting individuals' perception of air pollution risk and helping them make informed decisions on reducing air pollution exposure during physical activities. Furthermore, our findings reveal the differential moderation impacts of alternative information sources on individuals' exercise behavioral responses during different air pollution conditions. In good to lightly polluted air conditions, the availability of alternative PM2.5-US information generally had the largest negative moderation impact. The PM2.5-US information offers an alternative perspective on air quality, potentially differing in methodology or perceived accuracy compared with local official sources based on AQI-CN. The existence of this alternative source allows individuals to cross-reference and assess the risks associated with air pollution more critically, which may amplify individuals' perceived severity of and vulnerability to air pollution episodes. The enhanced perception can potentially increase their motivation to engage in protective behaviors. In moderately to heavily polluted conditions, social media Weibo tweets of "face mask" and "haze" had the most significant and largest negative-moderation impact on the effect of mainstream pollution information on exercise aversion. Social media posts can amplify the perceived severity of and personal vulnerability to air pollution because information and personal experiences shared within one's social network can contribute to a heightened sense of risk. The visibility of collective attitudes and actions on social media strengthens the risk perception through social contagion [19], encouraging individuals to take protective behaviors. In the most severely polluted conditions, the availability of mobile app access to location-specific real-time air quality information had the most considerable negative moderation effect on exercise aversion behaviors. Given that many individuals use mobile phones to track their exercise activities, they are more likely to seek out air quality information using mobile apps before or during the exercises. Individuals use air quality mobile apps that provide real-time, localized air quality data, which can render the official air pollution information more personally relevant and actionable [39]. Such real-time and personalized data may lead to individuals to perceive the air pollution risk as more urgent and serious, prompting a stronger avoidance behavior in terms of reducing outdoor exercise.

Although prior literature has documented that exercise behaviors can be socially contagious [19], prior findings did not differentiate between the information transmission mechanism between different types of social connections. Our findings show that the efficacy of social connections as alternative sources of pollution information in affecting exercise aversion behaviors can rely on the type of social connection (ie, followers, followees, and mutual followers). Individuals tend to rely more on information about polluted air conditions from their followers and mutual followers and adjust the exercise aversion responses based on observational learning of peer behaviors [40]. Mutual followers, representing a stronger social bond, often share information perceived as more personalized or relevant. When individuals follow each other, the reciprocal

ties imply a higher level of engagement and trust, which makes information shared by mutual followers seem more relevant and credible [31]. Individuals tend to view their followers as peers who are similar to themselves or part of their in-group. This perception can increase the trust in the information shared by these followers, as it may be more personalized, relevant to the user's local context or specific interests, or validated by social proof from peers [41].

Limitations and Future Work

While our study represents an initial endeavor to examine the extent to which individuals adjust their exposure to polluted air during physical exercise, and respond in real time to diverse information sources, we contend that this study has several limitations. First, the data set was collected from individuals using wearable devices or mobile fitness apps linked to a web and mobile fitness platform. We acknowledge that these individuals in our research sample are likely to have come from a selected sample who may be more predisposed to physical exercise activities compared with the general population. In this sense, the degree of air pollution aversion identified in our study is likely to be a conservative estimate. This is because enthusiasts of sports or fitness might not alter their physical activity routines, even during periods of air pollution. To mitigate this issue, researchers in future work could aim for a more representative sample of the general population or leverage field experiments. Second, regarding the phenomenon of web and mobile fitness platform and their gamification features, we acknowledge that individuals' decisions in our data sample to engage in exercise activities could have been driven somewhat by the gamification elements on the platform. Nevertheless, in our empirical analysis at the individual user level, we control for these effects in our empirical models by including the user's reward points and exercise skill grade as certified by the platform. Investigating the role of gamification elements on the platform in driving and sustaining behaviors aimed at health and fitness enhancement presents an intriguing research opportunity [42]. Third, future research could beneficially identify the threshold where the health benefits of physical activity in polluted conditions become outweighed by its adverse effects, taking into account various pollution levels and population demographics (eg, healthy individuals vs those with specific susceptibilities).

Conclusions

In conclusion, this study is one of the firsts to uncover to what extent individuals weigh the exercise benefit-harm tradeoff to adapt their exercise activities in response to different air pollution information sources. The main strengths of this study include using comprehensive, disaggregate spatiotemporal data of both mainstream air pollution information sources and individuals' detailed exercise behavioral records from wearable and mobile devices, examining individuals' exercise-aversion behaviors in near real time on an hourly basis, and identifying and evaluating the impact of multiple alternative air pollution information sources. Furthermore, this study reveals the important role of different alternative sources of pollution information in inducing individuals' exercise aversion responses.

Our findings offer insights for public health management and personal health decision-making from several aspects. First, policy makers may need to reassess the messaging and strategies used to communicate the risks associated with outdoor activities during episodes of high air pollution. Health authorities can use more personalized communication strategies, such as targeted messages for at-risk populations such as children, older adults, or those with respiratory conditions. To increase public awareness, reliance on traditional media and official air pollution alerts are not sufficient, public health authorities should leverage multitiered communication strategies that account for various demographics, including varying access to technology, as well as diversified information channels to keep the public engaged and informed about the dangers of high air pollution levels, especially in relation to outdoor activities. For instance, public health authorities may establish partnerships with tech companies and app developers to ensure the dissemination of air quality data is prompt, precise, and user-friendly. This can help create more tailored and engaging content that effectively promotes protective behaviors. Furthermore, the public awareness campaigns can be designed to capitalize on the virality and user engagement aspects of social media and mobile fitness platforms, utilizing these platforms to generate interactive and shareable content that facilitates the message dissemination. The dynamic nature of social media and mobile apps allow for real-time updates, which can keep the public informed during rapidly changing air quality situations, enhancing the public's awareness and engagement in protective behaviors.

Second, our findings underscore the importance of disseminating location- and time-referent air quality information (eg, green space vs industrial area and 1-hour vs 24-hour AQI measurements) to better enable the public in evaluating the benefits and risks of health maintenance behaviors during air pollution episodes. Air quality can vary significantly within short distances and fluctuate throughout the day due to various factors such as traffic patterns, weather conditions, and industrial activities. By disseminating air quality information that is both location and time-referent, public health authorities can provide the public with a more comprehensive understanding of air pollution risks, which enables individuals to make better decisions about when and where to engage in outdoor activities. Specifically, by using mobile apps and related notifications, public health authorities can use location and time-specific air quality data and push real-time air quality updates to populations in high-risk areas. For instance, during a pollution spike in an industrial zone, authorities can advise residents on precautionary measures. Social media and mobile fitness platforms can serve as channels for sharing broader air quality trends and advisories. Such information can also be directly disseminated to individuals, citizens, and employees participating in increasingly popular wellness programs initiated and run by government agencies and commercial organizations (eg, insurance companies), and enabled by dedicated mobile apps or wearable devices for such programs.

Third, for the design of wearable devices and mobile apps, manufacturers of wearable devices and activity trackers can combine air pollution and health advisories with device features that monitor ambient air quality and users' bio-physical

indicators such as blood pressure that has been shown to be related to the extent of airborne pollutants. Mobile fitness app developers can also provide comprehensive real-time updates of mainstream and alternative information of air quality conditions through AQI measures, social connections, social media content, and geo-location advisories. For instance, with GPS information from mobile devices, app developers can issue time- and location-specific advisories based on real-time geo-locations of individuals and social connections engaged in physical activities or exercises, such as encouraging exercises in green spaces or parks, rather than alongside major roads with heavy vehicular traffic. With such real-time air quality data of different locations delivered through mobile apps or wearable devices, individuals can better assess the risks and make informed decisions about when and where to exercise outdoors. Future research should explore the long-term impact of repeated exposure to air pollution information and its impact on behavior over time, and explore whether individuals habituate to the alerts or if there are cumulative effects.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Robustness check using placebo tests.
[DOCX File , 32 KB - mhealth_v12i1e55207_app1.docx]

Multimedia Appendix 2
Robustness check using regression discontinuity method.
[DOCX File , 4294 KB - mhealth_v12i1e55207_app2.docx]

References

1. Cepeda M, Schoufour J, Freak-Poli R, Koolhaas CM, Dhana K, Bramer WM, et al. Levels of ambient air pollution according to mode of transport: a systematic review. Lancet Public Health 2017 Jan;2(1):e23-e34 [FREE Full text] [doi: 10.1016/S2468-2667(16)30021-4] [Medline: 29249477]
2. Giles LV, Koehle MS. The health effects of exercising in air pollution. Sports Med 2014 Feb 1;44(2):223-249. [doi: 10.1007/s40279-013-0108-z] [Medline: 24174304]
3. Neidell M. Air quality warnings and outdoor activities: evidence from Southern California using a regression discontinuity design. J Epidemiol Community Health 2010 Oct 12;64(10):921-926. [doi: 10.1136/jech.2008.081489] [Medline: 19822555]
4. Ward AL, Beatty TK. Who responds to air quality alerts? Environ Resource Econ 2015 Apr 29;65(2):487-511. [doi: 10.1007/S10640-015-9915-Z]
5. Huang G. PM_2 . opened a door to public participation addressing environmental challenges in China. Environ Pollut 2015 Feb;197:313-315. [doi: 10.1016/j.envpol.2014.12.001] [Medline: 25499795]
6. Jiang W, Wang Y, Tsou MH, Fu X. Using social media to detect outdoor air pollution and monitor air quality index (AQI): a geo-targeted spatiotemporal analysis framework with Sina Weibo (Chinese Twitter). PLoS One 2015 Oct 27;10(10):e0141185 [FREE Full text] [doi: 10.1371/journal.pone.0141185] [Medline: 26505756]
7. Goh KY, Heng CS, Lin Z. Social media brand community and consumer behavior: quantifying the relative impact of user- and marketer-generated content. Inf Syst Res 2013 Mar;24(1):88-107. [doi: 10.1287/isre.1120.0469]
8. An R, Zhang S, Ji M, Guan C. Impact of ambient air pollution on physical activity among adults: a systematic review and meta-analysis. Perspect Public Health 2018 Mar 22;138(2):111-121. [doi: 10.1177/1757913917726567] [Medline: 28829249]
9. Adams SA, Matthews CE, Ebbeling CB, Moore CG, Cunningham JE, Fulton J, et al. The effect of social desirability and social approval on self-reports of physical activity. Am J Epidemiol 2005 Feb 15;161(4):389-398 [FREE Full text] [doi: 10.1093/aje/kwi054] [Medline: 15692083]
10. Newsome AM, Reed R, Sansone J, Batrakoulis A, McAvoy C, W. Parrott M. 2024 ACSM worldwide fitness trends: future directions of the health and fitness industry. ACSMs Health Fit J 2024;28(1):14-26. [doi: 10.1249/fit.0000000000000933]
11. Air pollution in China. World Health Organization. URL: https://www.who.int/china/health-topics/air-pollution [accessed 2023-09-28]
12. Chen Y, Ebenstein A, Greenstone M, Li H. Evidence on the impact of sustained exposure to air pollution on life expectancy from China's Huai River policy. Proc Natl Acad Sci U S A 2013 Aug 06;110(32):12936-12941 [FREE Full text] [doi: 10.1073/pnas.1300018110] [Medline: 23836630]
13. Lü J, Liang L, Feng Y, Li R, Liu Y. Air pollution exposure and physical activity in China: current knowledge, public health implications, and future research needs. Int J Environ Res Public Health 2015 Nov 20;12(11):14887-14897 [FREE Full text] [doi: 10.3390/ijerph121114887] [Medline: 26610539]
14. Chen L, Zhang Z, Long Y. Association between leisure-time physical activity and the built environment in China: Empirical evidence from an accelerometer and GPS-based fitness app. PLoS One 2021 Dec 31;16(12):e0260570 [FREE Full text] [doi: 10.1371/journal.pone.0260570] [Medline: 34971553]

15. Tao Z, Kokas A, Zhang R, Cohan DS, Wallach D. Inferring atmospheric particulate matter concentrations from Chinese social media data. PLoS One 2016 Sep 20;11(9):e0161389 [FREE Full text] [doi: 10.1371/journal.pone.0161389] [Medline: 27649530]
16. Delmas MA, Kohli A. Can apps make air pollution visible? Learning about health impacts through engagement with air quality information. J Bus Ethics 2019 Jun 13;161(2):279-302. [doi: 10.1007/S10551-019-04215-7]
17. Meet Air Matters: the app alerting millions on air quality. All Tech Asia. 2017. URL: https://medium.com/act-news/meet-air-matters-the-app-alerting-millions-on-air-quality-17ae4f7a1f51 [accessed 2021-01-18]
18. Aprahamian F, Chanel O, Luchini S. Modeling starting point bias as unobserved heterogeneity in contingent valuation surveys: an application to air pollution. Am J Agric Econ 2007 May;89(2):533-547. [doi: 10.1111/J.1467-8276.2006.01020.X]
19. Aral S, Nicolaides C. Exercise contagion in a global social network. Nat Commun 2017 Apr 18;8:14753 [FREE Full text] [doi: 10.1038/ncomms14753] [Medline: 28418379]
20. Graff Zivin J, Neidell M. Environment, health, and human capital. J Econ Lit 2013 Sep 01;51(3):689-730. [doi: 10.1257/jel.51.3.689]
21. Zivin JG, Neidell M. Days of haze: environmental information disclosure and intertemporal avoidance behavior. J Environ Econ Manage 2009;58(2):119-128. [doi: 10.3386/W14271]
22. Zheng S, Cao J, Kahn ME, Sun C. Real estate valuation and cross-boundary air pollution externalities: evidence from Chinese cities. J Real Estate Finan Econ 2013 Apr 16;48(3):398-414. [doi: 10.1007/S11146-013-9405-4]
23. Wooldridge JM. Econometric Analysis of Cross Section and Panel Data, second edition. Cambridge, MA: MIT press; 2010.
24. Chang TY, Huang W, Wang Y. Something in the air: pollution and the demand for health insurance. Rev Econ Stud 2018;85(3):1609-1634. [doi: 10.1093/restud/rdy016]
25. Imbens GW, Lemieux T. Regression discontinuity designs: a guide to practice. J Econom 2008 Feb;142(2):615-635. [doi: 10.1016/j.jeconom.2007.05.001]
26. Lee DS, Lemieux T. Regression discontinuity designs in economics. J Econ Lit 2010 Jun 01;48(2):281-355. [doi: 10.1257/jel.48.2.281]
27. Rogers RW. A protection motivation theory of fear appeals and attitude change1. J Psychol 1975 Sep 02;91(1):93-114. [doi: 10.1080/00223980.1975.9915803] [Medline: 28136248]
28. Wang S, Paul MJ, Dredze M. Social media as a sensor of air quality and public response in China. J Med Internet Res 2015 Mar 26;17(3):e22 [FREE Full text] [doi: 10.2196/jmir.3875] [Medline: 25831020]
29. Tan T, Junghans C, Varaden D. Empowering community health professionals for effective air pollution information communication. BMC Public Health 2023 Dec 20;23(1):2547 [FREE Full text] [doi: 10.1186/s12889-023-17462-1] [Medline: 38124041]
30. Cheek NN, Coe-Odess S, Schwartz B. What have I just done? Anchoring, self-knowledge, and judgments of recent behavior. Judgm Decis Mak 2023 Jan 01;10(1):76-85. [doi: 10.1017/S1930297500003193]
31. Song T, Tang Q, Huang J. Triadic closure, homophily, and reciprocation: an empirical investigation of social ties between content providers. Inf Syst Res 2019 Sep;30(3):912-926. [doi: 10.1287/isre.2019.0838]
32. Zhang J, Mu Q. Air pollution and defensive expenditures: evidence from particulate-filtering facemasks. J Environ Econ Manage 2018 Nov;92:517-536. [doi: 10.1016/j.jeem.2017.07.006]
33. Saberian S, Heyes A, Rivers N. Alerts work! Air quality warnings and cycling. Resour Energy Econ 2017 Aug;49:165-185. [doi: 10.1016/j.reseneeco.2017.05.004]
34. Weinstein ND. Unrealistic optimism about susceptibility to health problems: conclusions from a community-wide sample. J Behav Med 1987 Oct;10(5):481-500. [doi: 10.1007/bf00846146]
35. Wen XJ, Balluz L, Mokdad A. Association between media alerts of air quality index and change of outdoor activity among adult asthma in six states, BRFSS, 2005. J Community Health 2009 Feb 27;34(1):40-46. [doi: 10.1007/s10900-008-9126-4] [Medline: 18821001]
36. Lissåker CT, Talbott EO, Kan H, Xu X. Status and determinants of individual actions to reduce health impacts of air pollution in US adults. Arch Environ Occup Health 2016 Dec 02;71(1):43-48. [doi: 10.1080/19338244.2014.988673] [Medline: 25454076]
37. Zhang WY. China: fitness app user age distribution 2021. Statista. URL: https://www.statista.com/statistics/1245480/china-fitness-app-user-age-distribution/ [accessed 2024-03-18]
38. Guo E, Zhong H, Li J, Gao Y, Li J, Wang Z. The influence of air pollution on residents' outdoor exercise participation behaviour: evidence from China Family Panel Studies. PLoS One 2022 Aug 30;17(8):e0270994 [FREE Full text] [doi: 10.1371/journal.pone.0270994] [Medline: 36040990]
39. Cline RJ, Haynes KM. Consumer health information seeking on the Internet: the state of the art. Health Educ Res 2001 Dec;16(6):671-692. [doi: 10.1093/her/16.6.671] [Medline: 11780707]
40. Zhang J. The sound of silence: observational learning in the U.S. kidney market. Mark Sci 2010 Mar;29(2):315-335. [doi: 10.1287/mksc.1090.0500]
41. Hofer M, Aubert V. Perceived bridging and bonding social capital on Twitter: differentiating between followers and followees. Comput Human Behav 2013 Nov;29(6):2134-2142. [doi: 10.1016/j.chb.2013.04.038]

42. Yang Y, Goh KY, Teo HH, Tan SS. Compete with me? The impact of online gamified competition on exercise behavior. J Assoc Inf Syst 2023;24(3):912-935. [doi: 10.17705/1jais.00806]

Abbreviations

AQI: Air Quality Index
AQI-CN: Air Quality Index-China
CN-MEP: Chinese Ministry of Environmental Protection
PM: particulate matter

Edited by L Buis; submitted 11.12.23; peer-reviewed by E Chen, A Batrakoulis; comments to author 29.02.24; revised version received 22.04.24; accepted 10.07.24; published 10.09.24.

Please cite as:
Yang Y, Goh KY, Teo HH, Tan SSL
The Impact of Air Pollution Information on Individuals' Exercise Behavior: Empirical Study Using Wearable and Mobile Devices Data
JMIR Mhealth Uhealth 2024;12:e55207
URL: https://mhealth.jmir.org/2024/1/e55207
doi:10.2196/55207
PMID:39255029

Original Paper

Real-World Outcomes of a Digital Behavioral Coaching Intervention to Improve Employee Health Status: Retrospective Observational Study

Amani Fadzlina Abdul Aziz[1], BHumSc; Tiffanie Ong[1], BSc, PhD

Naluri Hidup Sdn Bhd, Kuala Lumpur, Malaysia

Corresponding Author:
Tiffanie Ong, BSc, PhD
Naluri Hidup Sdn Bhd
22, Persiaran Damansara Endah
Bukit Damansara
Kuala Lumpur, 50490
Malaysia
Phone: 60 123072922
Email: tiffanie@naluri.life

Abstract

Background: Chronic noncommunicable diseases (NCDs) account for major disability and premature mortality worldwide, with low- and middle-income countries being disproportionately burdened. Given the negative impact of NCDs on employee performance and work productivity, there is a rising need for stakeholders to identify effective workplace solutions that can improve employee health outcomes. As the workplace becomes more dispersed post pandemic, digital behavioral coaching offers a scalable, personalized, and cost-effective method of managing chronic disease risk factors among employees.

Objective: This study aimed to retrospectively evaluate the impact of a digital behavioral coaching program on year-to-year changes in employee health status in a cohort of Indonesian employees.

Methods: This retrospective real-world exploratory analysis of secondary health data followed 774 employees of an Indonesian company who completed company-sponsored health screenings between 2021 and 2022 and were given access to Naluri (Naluri Hidup Sdn Bhd), a holistic digital therapeutics platform offering digital behavioral health coaching and self-help tools. Participants were retrospectively classified as those who received active coaching (n=177), passive coaching (n=108), and no coaching (n=489). Linear mixed-effects models were used to evaluate the year-to-year changes in health outcomes across the 3 employee groups, with post hoc analyses evaluating within-group differences between the 2 time points and between-group differences at follow-up.

Results: Significant time×group interaction effects were detected for body weight, BMI, hemoglobin A_{1c}, low-density lipoprotein, total cholesterol, and systolic and diastolic blood pressure. Post hoc pairwise comparisons revealed significant improvements in hemoglobin A_{1c} (mean difference [M_{diff}]=–0.14, P=.008), high-density lipoprotein (M_{diff}=+2.14, P<.001), and total cholesterol (M_{diff}=–11.45, P<.001) for employees in the Active Coaching group between 2021 and 2022, with the other 2 groups reporting deteriorations in multiple health outcomes throughout the 2 time points. At follow-up, those who received active coaching between 2021 and 2022 reported significantly lower body weight (P<.001), BMI (P=.001), low-density lipoprotein (P=.045), and total cholesterol (P<.001) than the No Coaching group.

Conclusions: This study demonstrates real-world outcomes and implications supporting the use of workplace digital behavioral coaching in improving employee health status. Given the rising burden of NCDs in the Southeast Asian region, our findings underscore the role that workplace digital health interventions can play in preventing and managing chronic disease risk factors.

(JMIR Mhealth Uhealth 2024;12:e50356) doi:10.2196/50356

KEYWORDS

digital behavioral coaching; chronic disease management; digital health; mHealth; workplace interventions; mobile phone

Introduction

Chronic noncommunicable diseases (NCDs) are responsible for 74% of deaths worldwide [1], and approximately 38% (15.2 million) of these deaths occur prematurely, affecting people aged 30 to 70 years [2]. The World Health Organization (WHO) estimates that 69% of all deaths in the Southeast Asian region are due to NCDs [3], which roughly translates to 9 million of the regional population dying each year from NCDs, with 52% of deaths occurring prematurely in those younger than 70 years of age [3]. In Indonesia, NCDs are estimated to account for 76% of all deaths in the country [1], with cardiovascular diseases and diabetes accounting for an estimated 35% and 6% of proportional mortality, respectively [4]. Indonesia's rapid urbanization over the past decade has not only led to higher individual life expectancy but also higher rates of modifiable NCD risk factors such as unhealthy diet choices and sedentary lifestyles, ultimately contributing to the rise of an aging population at risk of developing NCDs [5].

Largely due to stretched and overburdened health care systems, low- and middle-income countries are disproportionately affected by the burden of NCDs [6]. The COVID-19 pandemic has also disrupted essential health service provision for NCDs in the region, further burdening health care systems and creating long-term impacts on health and mortality in the region [7]. Similarly, although Indonesia has been shifting its health care disease burden from communicable diseases to NCDs [5], the COVID-19 pandemic has largely disrupted the nation's NCDs services in primary health care [8], highlighting the need for a multi-stakeholder, multi-component, and preventative approach to managing NCDs in the nation. As health care systems worldwide continue to tackle the downstream impact of the COVID-19 pandemic, a growing body of evidence sheds light on the potential of digital health as a resource-efficient solution for health care systems and health care payers [9,10], making their real-world uptake more relevant for NCD prevention as traditional health care systems in the region continue to be overburdened.

Beyond health care, workplace-level interventions have a high magnitude of impact in addressing and managing chronic disease risk factors [11]. Working adults spend more than a third of their lives in the workplace [12], which makes them increasingly vulnerable to work-related risk factors such as sedentary work, work stress, and easy access to unhealthy food [13]. For this reason, the WHO recommends workplace health promotion programs as an effective avenue to target physical inactivity and unhealthy dietary habits—two important modifiable risk factors for obesity, hypertension, diabetes, and hypercholesterolemia [14]. Although workplace health improvement initiatives have traditionally been physical and face-to-face in nature, the COVID-19 pandemic has highlighted the need for digital offerings that provide more accessible, personalized, and diverse offerings, that are becoming increasingly more relevant as offices become more dispersed in a post–COVID-19 world. Given that unmanaged NCDs adversely affect employers through increased health care costs, rising premiums, and increased work absenteeism and presenteeism [15,16], there is a rising need for employers to effectively use the workplace as an avenue for health promotion and NCD prevention.

Major modifiable risk factors for NCDs are often behavioral in nature, such as tobacco use, excessive alcohol use, unhealthy diets, and physical inactivity [17]. For this reason, interventions that aim to prevent or self-manage chronic disease risk factors often rely on behavioral change frameworks to create long-term and sustainable behavioral changes [18-20]. Digital behavioral health coaching is a scalable and accessible method of delivering patient-centered health coaching, which is defined by an individual's interpersonal relationship with a certified coach that allows individuals to develop intrinsic motivation, create sustainable change, and increase accountability through the use of behavioral change theory, motivational strategies, and communication techniques [21]. Concurrently, there is growing evidence to support the effectiveness of digital workplace behavioral coaching interventions in improving various chronic disease risk factors [22], such as weight and BMI [23-27], waistline circumference [26,28], blood pressure (BP) [23,26,28], cholesterol levels [26,28], and blood glucose levels [28].

A majority of studies evaluating digital workplace interventions are based in Western or high-income countries. Different health system capacities, income levels, health literacy levels, and health-seeking cultures necessitate more studies that are localized within the region to more accurately evaluate the impact of digital workplace interventions in the Southeast Asian population [29]. This study leverages a retrospective analysis of secondary employee health data to fill in the current gap in digital workplace health interventions in the region, by exploring year-to-year changes in employee physical health status associated with different levels of coaching delivered by Naluri, a Southeast Asian digital therapeutics platform, as part of a workplace health improvement program implemented in an Indonesian mining company. We focus on changes in BMI, waistline circumference, lipoprotein and total cholesterol levels, hemoglobin A_{1c} (HbA_{1c}), BP—all of which are primary indicators of obesity, hypercholesterolemia, diabetes, and hypertension, respectively.

Methods

Study Design

This was a retrospective, observational, and exploratory study using secondary, longitudinal data from a cohort of employees at a prominent mining company in Indonesia.

Settings and Participants

Indonesian labor regulations stipulate that employers conduct regular mandatory health examinations to assess the physical health and ability of employees [30]. In 2021 and 2022, employees of a mining company were invited to complete a company-sponsored health screening that measured current employee health profiles such as weight, BP, blood glucose, and cholesterol levels. Participants of this study were those who had completed the annual company-sponsored health screening in 2021 and 2022, with the inclusion criteria set to employees with valid and matching health data across the 2 health screenings. Exclusion criteria were set to (1) employees with

missing health data, who did not complete either one or both health screenings in 2021 and 2022, (2) employees reporting an active pregnancy status during either or both health screenings in 2021 and 2022, and (3) employees whose health data were considered as unreasonable or improbable due to extreme outliers beyond clinically realistic levels.

After completing their health screening in 2021, employees were given access to receive digital health services provided by Naluri as part of their employer-provided employee wellness benefits. Employees were given the autonomy and freedom to voluntarily receive Naluri's digital health services, with no external coercion or pressure to join from their employers or from Naluri personnel. Based on the employee health profile from the 2021 health screening, 300 employees with the highest risk factor indicators for obesity, hypercholesterolemia, diabetes, and hypertension were then identified and invited to join a 16-week structured, intensive digital chronic disease management program (CDMP) delivered on the Naluri mobile app (Naluri Hidup Sdn Bhd) at 2 points throughout the year, that is, in April 2022, and later in June 2022. Employees were invited to join Naluri's CDMP if their 2021 health screening outcomes met 2 or more of the following criteria: (1) HbA_{1c}≥5.7%; (2) total cholesterol level≥201.08 mg/dL; (3) BMI≥23 kg/m^2; or (4) systolic BP≥120 mm Hg and diastolic BP≥80 mm Hg. At the end of 2022, all employees, regardless of use or engagement with Naluri services, were again invited by their employers to participate in an annual company-sponsored health screening.

Employees who fulfilled this study's inclusion criteria were retrospectively categorized into three groups, which were (1) Active Coaching group, comprising employees who were invited and voluntarily enrolled into a 16-week, active, coach-led structured intensive digital CDMP during the interval between the 2021 and 2022 health screening; (2) Passive Coaching group, comprising of employees who voluntarily accessed self-led, unstructured, and passive digital behavioral coaching on the Naluri app through their employee assistance program (EAP) during the interval between the 2021 and 2022 health screening; and (3) No Coaching group, which are employees who reported no use of Naluri digital health services during the interval between the 2021 and 2022 health screening.

Intervention

Active Coaching: Naluri CDMP

The Naluri CDMP is a 16-week intensive disease prevention program that uses a multidisciplinary team–based approach to improve physical health outcomes and chronic disease risk factors, primarily targeting indicators for obesity, hypercholesterolemia, diabetes, and hypertension. Naluri adopts the transtheoretical model of behavioral change [31] to help individuals adopt and maintain lifestyle habits that can sustainably improve health outcomes. The 16-week CDMP comprises 5 stages. First, Getting Started, where individuals get onboarded onto the program, build rapport with their assigned coaches, and set baselines and expectations. Second, Preparation, where individuals discover their current needs and motivation, discuss habit formation and planning, and set behavioral goals. Third, Action, which involves skill building across 3 domains—behavior, cognitive, and emotion regulation—to help individuals build healthy habits, adjust mindsets, and manage emotions. Fourth, Resilience, which focuses on empowering individuals to overcome challenges, with emphasis on resilience boosting modules such as dealing with relapses, setbacks, and weight plateaus. Finally, Maintaining Success, which focuses on maintenance and inspiring accountability to prepare individuals for long-term, sustainable, and independent healthy living. Throughout each stage of the program, participants had access to a multidisciplinary team of health coaches, such as dietitian, fitness coach, medical advisors, and more, led by a mental health coach (master's-level health or clinical psychologist or a certified counsellor), who follow a comprehensive curriculum rooted in cognitive behavioral therapy and motivational interviewing to provide dedicated and continuous holistic support to help goal setting and habit formation.

Each week of the CDMP starts with a weekly check-in with a respective coach (eg, psychologist, dietitian, fitness coach, or medical advisor), where each coach engages in individual text-based coaching over the Naluri mobile app. Coaches were provided a coaching framework to follow throughout the program, with coaching objectives set for each week of the CDMP. For example, in week 2, which falls under the "Preparation" stage, mental health coaches were tasked to assess for participants' current mindsets, behaviors, barriers, and health goals, dietitians were tasked to assess for past dietary attempts for weight management and current eating patterns, while fitness coaches were tasked to assess for participants' current and past exercise history. In addition, coaches were also tasked to assign participants to complete educational modules to supplement the objectives and goals for each week of the program. Throughout the 16 weeks, participants were also assigned to complete and update their thought, food, weight, and exercise journals, at a minimum of once per week. Although the CDMP involved a structured curriculum and framework, the coaching received by the participants were also personalized to tailor to their current health needs and goals.

Passive Coaching: Naluri EAP

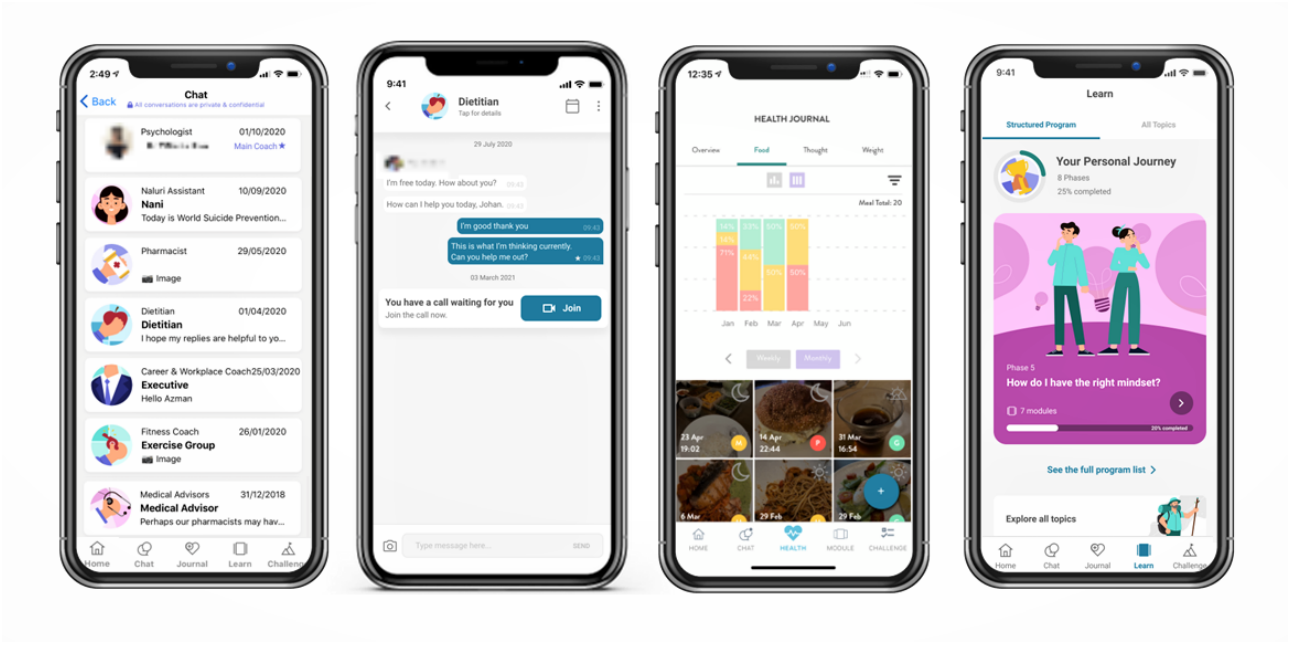
Naluri's EAP is a passive, self-led, and digitally-delivered service for employees seeking mental or physical health support. Although the Naluri EAP provides individuals with access to holistic digital health coaching from a multidisciplinary team of health care professionals, unlike the CDMP, the Naluri EAP is an unstructured and low-intensive service that allows individuals to personalize their own health improvement journey in a nonlinear fashion. The Active Coaching group received structured and proactive coaching by the CDMP whereas those in the Passive Coaching group were free to use any of Naluri's digital services as they saw fit, following no particular guidance or structure.

Naluri Mobile App

Both the Naluri CDMP and EAP are delivered digitally through the Naluri mobile app, a human-driven, artificial intelligence–augmented digital therapeutics platform that provides holistic digital care solutions aimed at predicting,

preventing, and managing cardiometabolic and mental health conditions (Figure 1). In addition to a team of on-demand multidisciplinary coaches to provide accessible and on-demand coaching, the Naluri app is also optimized with evidence-based digital tools that facilitate behavioral change through features that promote goal setting and self-monitoring. These include habit trackers and artificial intelligence–augmented health journals that allow users to log and monitor their food intake, thoughts and emotions, weight, and BP; educational modules on topics such as resilience, diet and nutrition, managing chronic care, and exercise libraries; and the ability to set personalized challenges, as well as join support groups.

Figure 1. Naluri mobile app platform. (A) Chat home page with a multidisciplinary team of coaches. (B) Example of chat screen for text-based coaching with dietitians. (C) Health journal showing food journal uploads and overview. (D) Self-paced lessons and educational modules.



Physical Outcomes

The physical health outcomes used in this study are sourced from secondary health data collected as part of an annual company-sponsored health screening for employees, which serves as the basis for this study's outcome measures. In addition to physical health outcomes, participants' self-reported ages and genders were also sourced from the data collected during the health screening.

We identify primary physical outcomes as changes in employee health profile, namely BMI, low-density lipoprotein (LDL) level, high-density lipoprotein (HDL) level, total cholesterol level, blood glucose level measured as HbA_{1c}, systolic and diastolic BPs, body weight, and waistline circumference from 2021 to 2022. All physical health outcomes were measured by health practitioners specially contracted to conduct employee physical health screenings, which were then shared with Naluri for further evaluation.

Statistical Analysis

All analyses were performed on RStudio version 2022.07.0+548, using R (version 4.2.1; R Core Team). Statistical tests were 2-sided and evaluated at a $P<.05$ significance threshold. We identify physical health outcomes at 2021 as "baseline" and physical health outcomes at 2022 as "follow-up." Descriptive statistics were calculated for sociodemographic variables, such as age and gender, and for all outcome variables, reporting means and SDs for continuous data, and frequencies and percentages for categorical data. We report standardized mean differences to evaluate for incomparability across the 3 groups on age, gender, and all outcome variables at baseline, where standardized mean differences >0.1 would indicate evidence of imbalance across the 3 groups [32].

Linear mixed-effects models (LMMs) were used to assess the year-to-year changes in individual health outcomes, with time, group, and time by group interaction as fixed effects, and participants as random effects. LMMs, which allow for comparisons between- and within-groups at multiple time points, are considered to have more statistical power than traditional methods of evaluating interventions such as mixed ANOVAs or ordinary least squares regressions [33]. As group randomization was not possible for this study, participants were modelled as a random intercept to account for individual differences across all group conditions [34]. In addition, all models included gender, age, and baseline outcomes as covariates. Baseline covariate adjustments are a reliable and statistically efficient method of accounting for selection bias in nonrandomized observational studies [35], more so since any baseline differences across the 3 groups in this study are likely an inherent property of group membership [36]. All models were implemented through the *lme4* package [37] using restricted maximum likelihood estimation, and the significance of fixed effects was evaluated through *P* values derived using Satterthwaite approximations from omnibus ANOVA tests on the *lmerTest* package [38], which has been shown to produce lower type I error rates [39]. Post hoc pairwise comparisons were conducted to explore the direction of significant interactions between group and time using the *emmeans* package

using the Tukey method to adjust for multiple comparisons [40]. Effect sizes were assessed using Cohen *d* statistics, which describes an effect size of 0.2 as small, 0.5 as medium, and 0.8 as large [41].

Ethical Considerations

This study was not preregistered. This study was approved by the Medical and Research Ethics Committee, Ministry of Health Malaysia (NMRR ID-23-01212-PQH). Implied consent was obtained from participants when they accepted the Naluri mobile app privacy policy during the creation of their user accounts, which includes a clause stating that anonymized user data may be used for research purposes. The data used in this study was transferred and stored in compliance with HIPAA (Health Insurance Portability and Accountability Act) requirements. All raw data were anonymized and deidentified by independent parties before extraction for this study. No compensation was offered to participants as part of this study.

Results

Participant Characteristics

A total of 774 employees fulfilled the inclusion criteria and were included as participants in this study. Of the 300 invited to receive active coaching, 177 employees voluntarily participated in the CDMP. Of the remaining employees, 108 reported the use of the Naluri mobile app and received passive coaching, whereas 489 employees reported no use of Naluri's digital services and thus received no coaching. The unequal sample sizes seen among the 3 groups reflect the real-world nature and voluntary enrollment of the employees into the offered digital health interventions.

The mean age of the overall sample was 38.51 (SD 8.32, range 21-81) years, and the majority of the participants identified as men (583/770, 75.71%). Table 1 presents the sample characteristics and baseline variables by group. Due to the nonrandomized nature of the study's design, significant imbalances were evident for all baseline demographic and outcome variables between the 3 participant groups.

Table 1. Participant characteristics and baseline variable (N=774).

Variable	Active coaching (n=177)	Passive coaching (n=108)	No coaching (n=489)	Active vs passive coaching, SMD[a]	Active vs no coaching, SMD	Passive vs no coaching, SMD
Age[b] (years), mean (SD, range)	41.14 (7.78, 25-58)	35.80 (8.01, 22-52)	38.16 (8.33, 21-68)	0.68	0.37	–0.29
Gender[b], n (%)				0.75	0.13	0.61
Women	29 (16.5)	53 (49.5)	105 (21.6)			
Men	147 (83.5)	54 (50.5)	382 (78.4)			
Body weight (kg), mean (SD)	77.58 (10.63)	69.37 (11.40)	67.41 (11.20)	0.75	0.93	0.17
BMI (kg/m^2), mean (SD)	28.87 (3.58)	26.84 (3.67)	25.20 (3.65)	0.56	1.02	0.45
HbA_{1c}[c] (%), mean (SD)	6.18 (1.56)	5.39 (0.61)	5.52 (0.88)	0.67	0.53	–0.17
LDL[d] (mg/dL), mean (SD)	148.34 (38.84)	117.39 (29.64)	125.59 (32.37)	0.90	0.64	–0.26
HDL[e] (mg/dL), mean (SD)	41.45 (8.74)	46.98 (12.17)	41.79 (9.90)	–0.52	–0.04	0.47
Total cholesterol (mg/dL), mean (SD)	230.05 (38.53)	190.69 (34.05)	199.66 (36.56)	1.09	0.81	–0.25
Systolic BP[f], mean (SD)	123.22 (12.67)	112.32 (13.72)	115.41 (11.49)	0.83	0.65	–0.25
Diastolic BP, mean (SD)	80.43 (6.60)	74.69 (6.87)	75.90 (6.93)	0.86	0.67	–0.18
Waist circumference (cm), mean (SD)	96.32 (8.91)	89.59 (11.06)	87.57 (10.13)	0.67	0.92	0.19

[a]SMD: standardized mean difference.

[b]For gender and age, total n=770 due to missing demographic values at baseline.

[c]HbA_{1c}: hemoglobin A_{1c} or glycated hemoglobin.

[d]LDL: low-density lipoprotein.

[e]HDL: high-density lipoprotein.

[f]BP: blood pressure.

Primary Health Outcomes

Table 2 presents descriptive statistics for all measured health outcomes at baseline and follow-up and omnibus results for fixed effects from the LMM analyses. As expected, the participants who were invited to receive active coaching reported poorer health outcomes in comparison to the other groups at

baseline. Participants who received active coaching within the year reported improvements in their health outcomes at follow-up, with the most noticeable improvements being total cholesterol levels (mean difference [M_{diff}]=–11.34 mg/dL). Those who engaged in passive coaching reported unchanging or slight deterioration in their health outcomes at follow-up, where slight improvements were only seen in HDL (M_{diff}=+1.89 mg/dL) and total cholesterol (M_{diff}=–0.78 mg/dL) levels. In comparison, participants who engaged in no coaching throughout the year recorded deteriorations in a majority of their health outcomes at follow-up, with the exception of improvements in HbA_{1c} (M_{diff}=–0.03%) and HDL levels (M_{diff}=+3.15 mg/dL). Results from LMMs showed a significant interaction effect of group×time for almost all health outcomes (P<.05) except for HDL (P=.09) and waist circumference (P=.06), indicating the presence of significant differences in the year-to-year change of health outcomes across the 3 groups.

Table 2. Means and SDs of health outcomes at baseline and follow-up by group with corresponding results of fixed effects from linear mixed models.

Variable and group	Baseline	Follow-up	Time		Group		Time×group	
	Mean (SD)	Mean (SD)	*F* test (*df*=1,1531)	*P* value	*F* test (*df*=2,1531)	*P* value	*F* test (*df*=2,1531)	*P* value
Body weight (kg)			3.83	*.05* [a]	9.02	*<.001*[a]	11.78	*<.001*[a]
Active Coaching	77.58 (10.63)	77.01 (11.21)						
Passive Coaching	69.37 (11.40)	70.09 (11.79)						
No Coaching	67.41 (11.20)	67.96 (11.48)						
BMI (kg/m^2)			0.51	.47	3.36	*.04*	4.86	*.008* [a]
Active Coaching	28.87 (3.58)	28.67 (3.74)						
Passive Coaching	26.84 (3.67)	27.00 (3.82)						
No Coaching	25.20 (3.65)	25.38 (4)						
HbA$_{1c}$[b] (%)			4.87	*.03* [a]	0.71	.49	3.68	*.03* [a]
Active Coaching	6.18 (1.56)	6.04 (1.30)						
Passive Coaching	5.39 (0.61)	5.39 (0.66)						
No Coaching	5.52 (0.88)	5.49 (0.80)						
LDL[c] (mg/dL)			36.51	*<.001*[a]	0.18	.84	6.86	*.001* [a]
Active Coaching	148.34 (38.84)	149.52 (39.38)						
Passive Coaching	117.39 (29.64)	126.61 (33.35)						
No Coaching	125.59 (32.37)	133.97 (32.23)						
HDL[d] (mg/dL)			64.07	*<.001*[a]	1.28	.28	2.32	*.01* [a]
Active Coaching	41.45 (8.74)	43.60 (8.06)						
Passive Coaching	46.98 (12.17)	48.87 (12.24)						
No Coaching	41.79 (9.90)	44.94 (10.46)						
Total cholesterol (mg/dL)			12.39	*<.001*[a]	0.84	.43	11.97	*<.001*[a]
Active Coaching	230.05 (38.53)	218.71 (41.94)						
Passive Coaching	190.69 (34.05)	189.81 (35.89)						
No Coaching	199.66 (36.56)	199.72 (35.48)						
Systolic BP[e]			13.28	*<.001*[a]	1.06	.35	5.57	*.004* [a]
Active Coaching	123.22 (12.67)	122.73 (10.79)						
Passive Coaching	112.32 (13.72)	116.00 (13.63)						
No Coaching	115.41 (11.49)	117.34 (11.93)						
Diastolic BP			0.13	.72	1.71	.18	5.18	*.006* [a]
Active Coaching	80.43 (6.60)	79.12 (5.16)						
Passive Coaching	74.69 (6.87)	75.81 (6.95)						
No Coaching	75.90 (6.93)	76.32 (6.85)						
Waist circumference (cm)			1.62	.20	0.14	.87	2.76	.06
Active Coaching	96.32 (8.91)	96.01 (8.52)						
Passive Coaching	89.59 (11.06)	89.92 (10.55)						
No Coaching	87.57 (10.13)	88.47 (10.49)						

[a]Indicate statistical significance.

[b]HbA$_{1c}$: hemoglobin A$_{1c}$ or glycated hemoglobin.

[c]LDL: low-density lipoprotein.

[d]HDL: high-density lipoprotein.

[e]BP: blood pressure.

Tables 3 and 4 present results from post hoc pairwise comparison of estimated marginal means from the LMMs, reporting contrasts for within-group differences in health outcomes over time, and contrasts for between-group differences at follow-up measurement. Those who received active coaching reported significant improvements in HbA_{1c}% (M_{diff}=–0.14, *P*=.008), HDL (M_{diff}=+2.14, *P*<.001), and total cholesterol levels (M_{diff}=–11.45, *P*<.001) between the years of 2021 and 2022. While the Active Coaching group did record improving trends in body weight, BMI, systolic and diastolic BPs, and waist circumference, these improvements were not substantial enough to be considered statistically significant. In comparison, participants who engaged in passive coaching only reported a significant improvement in HDL levels (M_{diff}=+1.95, *P*=.04) between the 2 time points, with the group recording deteriorations in a majority of their remaining outcomes, specifically in LDL levels (M_{diff}=+9.10, *P*=.001) and systolic BP (M_{diff}=+3.71, *P*=.005). Separately, employees who received no coaching at all between 2021 and 2022 also recorded larger deteriorations in year-to-year changes of health outcomes, with the No Coaching group recording significant deterioration in body weight (M_{diff}=+0.56), *P*<.001), LDL levels (M_{diff}=+8.47, *P*<.001), systolic BP (M_{diff}=+1.93, *P*=.001), and waist circumference (M_{diff}=+3.41, *P*=.009).

Pairwise contrasts of health outcomes at follow-up revealed significant between-group differences for body weight, BMI, LDL, and total cholesterol levels that are in favor of the Active Coaching group. Compared with the No Coaching group, those who received active coaching between 2021 and 2022 reported significantly lower body weight (*P*<.001), BMI (*P*=.001), LDL (*P*=.045), and total cholesterol (*P*<.001) at the follow-up screening. Similarly, the Active Coaching group also reported lower body weight than the Passive Coaching group (*P*<.001) at follow-up.

Table 3. Results of pairwise comparisons reporting estimated marginal means, mean differences, and SEs for within-group contrasts[a].

Variable and group	Baseline	Follow-up	Within-group contrasts[b]			
	EMM[c] (SE)	EMM (SE)	M_{diff}[d] (SE)	*t* test (*df*=767)	*P* value[e]	Cohen *d* (95% CI)
Body weight (kg)						
Active Coaching	70.05 (0.16)	69.47 (0.16)	0.58 (0.21)	2.70	.08	0.29 (0.08 to 0.50)
Passive Coaching	69.93 (0.20)	70.67 (0.19)	–0.74 (0.27)	–2.67	.08	–0.37 (–0.64 to –0.10)
No Coaching	70.04 (0.10)	70.60 (0.10)	–0.56 (0.13)	–4.38	*<.001*[f]	–0.28 (–0.41 to –0.16)
BMI (kg/m^2)						
Active Coaching	26.31 (0.08)	26.10 (0.08)	0.21 (0.11)	1.90	.40	0.21 (–0.01 to 0.41)
Passive Coaching	26.25 (0.10)	26.41 (0.10)	–0.16 (0.14)	–1.14	.87	–0.16 (–0.42 to 0.11)
No Coaching	26.29 (0.05)	26.47 (0.05)	–0.18 (0.07)	–2.80	.06	–0.18 (–0.31 to –0.05)
HbA_{1c}[g] (%)						
Active Coaching	5.72 (0.03)	5.57 (0.03)	0.14 (0.04)	3.45	*.008*[f]	0.37 (0.16 to 0.58)
Passive Coaching	5.61 (0.04)	5.62 (0.04)	–0.01 (0.05)	–0.12	>.99	–0.02 (–0.29 to 0.25)
No Coaching	5.63 (0.02)	5.61 (0.02)	0.02 (0.03)	0.89	.95	0.06 (–0.07 to 0.18)
LDL[h] (mg/dL)						
Active Coaching	131 (1.32)	132 (1.32)	–1.09 (1.78)	–0.61	.99	–0.07 (–0.27 to 0.14)
Passive Coaching	128 (1.63)	137 (1.63)	–9.10 (2.29)	–3.98	*.001*[f]	–0.54 (–0.81 to –0.28)
No Coaching	128 (0.83)	137 (0.83)	–8.47 (1.07)	–7.91	*<.001*[f]	–0.51 (–0.63 to –0.38)
HDL[i] (mg/dL)						
Active Coaching	42.80 (0.38)	44.90 (0.38)	–2.14 (0.52)	–4.12	*<.001*[f]	–0.44 (–0.65 to –0.23)
Passive Coaching	43.0 (0.48)	44.90 (0.48)	–1.95 (0.67)	–2.93	*.04*[f]	–0.40 (–0.67 to –0.13)
No Coaching	42.7 (0.24)	45.9 (0.24)	–3.17 (0.31)	–10.16	*<.001*[f]	–0.65 (–0.78 to –0.53)
Total cholesterol (mg/dL)						
Active Coaching	208 (1.53)	197 (1.53)	11.45 (2.04)	5.61	*<.001*[f]	0.60 (0.39 to 0.81)
Passive Coaching	203 (1.87)	202 (1.87)	1.10 (2.62)	0.42	>.99	0.06 (–0.21 to 0.33)
No Coaching	204 (0.95)	204 (0.95)	–0.09 (1.23)	–0.078	>.99	–0.005 (–0.13 to 0.12)
Systolic BP[j]						
Active Coaching	118 (0.60)	117 (0.60)	0.49 (0.81)	0.61	.99	0.06 (–0.14 to 0.27)
Passive Coaching	116 (0.75)	120 (0.75)	–3.71 (1.04)	–3.55	*.005*[f]	–0.49 (–0.75 to –0.22)
No Coaching	116 (0.38)	118 (0.38)	–1.93 (0.49)	–3.94	*.001*[f]	–0.25 (–0.38 to –0.13)
Diastolic BP						
Active Coaching	77.9 (0.37)	76.6 (0.37)	1.24 (0.51)	2.42	.15	0.26 (0.05 to 0.47)
Passive Coaching	76.2 (0.47)	77.3 (0.47)	–1.13 (0.66)	–1.72	.52	–0.24 (–0.50 to 0.03)
No Coaching	76.5 (0.22)	76.9 (0.22)	–0.42 (0.31)	–1.37	.74	–0.09 (–0.21 to 0.04)
Waist circumference (cm)						
Active Coaching	90.34 (0.33)	90.05 (0.33)	0.29 (0.45)	0.66	.99	0.07 (–0.14 to 0.28)
Passive Coaching	89.83 (0.41)	90.19 (0.41)	–0.36 (0.58)	–0.63	.99	–0.09 (–0.36 to 0.18)
No Coaching	89.61 (0.21)	90.53 (0.21)	–0.92 (0.27)	–3.41	*.009*[f]	–0.22 (–0.34 to –0.09)

[a]Results are averaged over the 2 levels of gender.

[b]Contrast: baseline – follow-up.

[c]EMM: estimated marginal means.

[d]M_{diff}: mean difference.

[e]*P* values are adjusted using the Tukey method for comparing a family of 6 estimates.

[f]Indicate statistical significance.

[g]HbA_{1c}: hemoglobin A_{1c} or glycated hemoglobin.

[h]LDL: low-density lipoprotein.

[i]HDL: high-density lipoprotein.

[j]BP: blood pressure.

Table 4. Results of pairwise comparisons reporting estimated marginal mean differences and SEs for between-group contrasts at follow-up.

Groups	Variables and between-group contrasts at follow-up[a]			
	M_{diff}[b] (SE)	*t* test (*df*)	*P* value[c]	Cohen *d* (95% CI)
Body weight (kg)				
No Coaching – Active Coaching	1.13 (0.18)	6.16 (1522)	*<.001* [d]	0.56 (0.39 to 0.74)
No Coaching – Passive Coaching	–0.07 (0.22)	–0.32 (1529)	>.99	–0.04 (–0.25 to 0.18)
Passive Coaching – Active Coaching	1.99 (0.25)	4.77 (1528)	*<.001* [d]	0.60 (0.35 to 0.84)
BMI (kg/m^2)				
No Coaching – Active Coaching	0.38 (0.09)	3.99 (1519)	*.001* [d]	0.37 (0.19 to 0.55)
No Coaching – Passive Coaching	0.07 (0.11)	0.62 (1529)	.99	0.07 (–0.15 to 0.28)
Passive Coaching – Active Coaching	0.31 (0.13)	2.39 (1527)	.16	0.30 (0.05 to 0.55)
HbA_{1c}[e] (%)				
No Coaching – Active Coaching	0.03 (0.04)	0.89 (1529)	.95	0.08 (–0.10 to 0.26)
No Coaching – Passive Coaching	–0.01 (0.04)	–0.35 (1530)	>.99	–0.04 (–0.25 to 0.17)
Passive Coaching – Active coaching	0.05 (0.05)	0.94 (1527)	.94	0.12 (–0.13 to 0.36)
LDL[f] (mg/dL)				
No Coaching – Active Coaching	4.34 (1.50)	2.89 (1528)	*.045* [d]	0.26 (0.08 to 0.44)
No Coaching – Passive Coaching	–0.09 (1.81)	–0.05 (1530)	>.99	–0.01 (–0.22 to 0.44)
Passive Coaching – Active Coaching	4.43 (2.11)	2.10 (1525)	.29	0.27 (0.02 to 0.51)
HDL[g] (mg/dL)				
No Coaching – Active Coaching	0.98 (0.43)	2.27 (1531)	.21	0.20 (0.03 to 0.37)
No Coaching – Passive Coaching	0.93 (0.53)	1.77 (1530)	.49	0.19 (–0.02 to 0.40)
Passive Coaching – Active Coaching	0.05 (0.61)	0.07 (1529)	>.99	0.01 (–0.26 to 0.25)
Total cholesterol (mg/dL)				
No Coaching – Active Coaching	7.24 (1.74)	4.17 (1525)	*<.001* [d]	0.38 (0.20 to 0.56)
No Coaching – Passive Coaching	1.73 (2.07)	0.84 (1530)	.96	0.09 (–0.12 to 0.30)
Passive Coaching – Active Coaching	5.51 (2.43)	2.26 (1523)	.21	0.29 (0.04 to 0.54)
Systolic BP[h]				
No Coaching – Active Coaching	0.67 (0.69)	0.98 (1528)	.93	0.09 (–0.09 to 0.26)
No Coaching – Passive Coaching	–1.57 (0.83)	–1.90 (1530)	.41	–0.21 (–0.42 to 0.007)
Passive Coaching – Active Coaching	2.24 (0.96)	2.33 (1527)	.18	0.29 (0.05 to 0.54)
Diastolic BP				
No Coaching – Active Coaching	0.26 (0.43)	0.60 (1529)	.99	0.05 (–0.12 to 0.23)
No Coaching – Passive Coaching	–0.43 (0.52)	–0.83 (1532)	.96	–0.09 (–0.30 to 0.12)
Passive Coaching – Active Coaching	0.68 (0.60)	1.14 (1529)	.87	0.14 (–0.10 to 0.39)
Waist circumference (cm)				
No Coaching – Active Coaching	0.49 (0.38)	1.27 (1522)	.80	0.12 (–0.06 to 0.29)
No Coaching – Passive Coaching	0.34 (0.46)	0.75 (1527)	.98	0.08 (–0.06 to 0.29)
Passive Coaching – Active Coaching	0.15 (0.53)	0.28 (1526)	>.99	0.03 (–0.21 to 0.28)

[a]Results are averaged over the 2 levels of gender.

[b]M_{diff}: mean difference.

[c]*P* values are adjusted using the Tukey method for comparing a family of 6 estimates.

[d]Indicate statistical significance.

[e]HbA_{1c}: hemoglobin A_{1c} or glycated hemoglobin.

[f]LDL: low-density lipoprotein.

[g]HDL: high-density lipoprotein.

[h]BP: blood pressure.

Discussion

Principal Findings

In this observational study analyzing retrospective data of 774 employees, receiving active coaching by the Naluri digital CDMP was associated with statistically significant improvement in year-to-year changes in health outcomes. We found significant group×time interaction effects for body weight, BMI, HbA_{1c}, LDL, total cholesterol, and systolic and diastolic BPs. Post hoc pairwise comparisons revealed that, of the 3 naturally occurring groups observed in this real-world study, those who engaged in passive coaching or no coaching at all recorded deteriorating trends in body weight, LDL, systolic BP, and waist circumference from 2021 to 202. Between-group contrasts also revealed that, at follow-up, employees who had received active coaching reported significant improvements in body weight, BMI, and total cholesterol compared with those who did not receive any form of coaching. Overall, our findings suggest that structured, intensive, and active digital health coaching was associated with an improvement in overall employee health status and minimizing chronic disease risk factors in an Indonesian workforce sample.

Our study adds to the growing evidence supporting the use of digital behavioral health coaching in improving health-related outcomes in the workplace. A systematic review of 22 randomized controlled trials found that digital health interventions in the workplace create a positive impact on health-related workplace outcomes such as diet, exercise, mental health, and job satisfaction [22]. Similarly, a systematic review of 22 studies found that digital health coaching was effective in managing and improving HbA_{1c}, weight loss, and BMI [42]. Although the effect sizes observed in this study were primarily small to medium in magnitude, our findings are similar to that of Nkhoma et al [43], which similarly found a small effect (Hedges g=0.29) in a meta-analysis of digital interventions on HbA_{1c}. Among studies looking into the Asian population, mobile-based digital workplace wellness interventions have been successful in reducing overall cardiovascular disease risk among hospital management workers in China [44], reducing the incidence of type 2 diabetes among male employees in India [45], and improving body weight, blood triglyceride, and systolic BP among university employees in Malaysia [46]. Although previous studies have demonstrated the effectiveness of workplace health promotion programs in improving employee metabolic health outcomes in Indonesia [47], to our knowledge, this study is the first to have looked into the real-world effects of workplace digital behavioral coaching among Indonesian employees.

Of the 3 groups of employees observed in our study, those who did not engage with or receive any form of digital coaching reported significantly poorer year-to-year health status compared with the other groups. Observational studies using propensity-matched controls have also demonstrated similar results, with Wilson et al [25] reporting year-to-year weight gain and an increase in fasting blood glucose in their propensity-matched control group. The trend in health outcomes among the participants in the No Coaching group offers a real-world glimpse of the year-to-year deterioration in health that can occur naturally and highlights the importance of preventative workplace health programs even for employees who are considered physically healthy. In a longitudinal study following a cohort of Singaporean employees, Sathish et al [48] observed a natural increase in physical inactivity, an increase in the prevalence of overweight or obesity, and significantly worsening dietary habits within 12 months. Similarly, large-scale studies have projected a 35.4% increase in diabetes prevalence by 2050 across the Southeast Asian region [49]. Given the intrinsic NCD health risks associated with traditional workplace environments [13], and the increasing stressors modern employees continue to face [50,51], our results reinforce the importance of workplace health interventions in reversing natural downward health trends and call attention to the health burdens that employers in the region will need to carry if no proactive action is taken.

Interestingly, despite having access to similar Naluri digital services as the Active Coaching group, we observed no change and even deterioration in a number of health outcomes among the group of employees who engaged in passive coaching. This may be attributed to the unstructured nature of the coaching received by this group of employees. Although limited evidence does suggest that unstructured and self-directed digital interventions can be effective in improving health outcomes [52,53], our findings are more aligned with the previous studies in this area that highlight the effectiveness of structured over unstructured digital interventions [54,55]. To this effect, our results highlight the strength of using Naluri's digital tools within the context of a structured and theory-driven intervention framework as compared with passive, unstructured, and self-driven use. Separately, participant-related characteristics such as motivation and self-determination have been shown to be important drivers of behavioral change in digital interventions [56]. The mere knowledge that employees in this group were not at high risk during the baseline screening may have precluded them from developing the motivation and self-determination needed to improve their health [57], similar to what is seen in individuals at the precontemplation stage [31].

Currently, there is limited evidence that supports the use of digital health programs for improving physical outcomes in a Southeast Asian population [58]. To our knowledge, our findings are the first in the region to use real-world data, thus building the body of evidence supporting the use of digital health for real-world NCD prevention. Despite the relative infancy of digital health in Southeast Asia, the region is also known as one

of the fastest adopters of mobile health technologies in the past decade [59], with Indonesians reporting one of the highest rates of smartphone health information-seeking behavior in the region [60]. To this effect, our findings highlight and reinforce the potential of digital health interventions as a cost-effective, resource-efficient, and scalable avenue for the prevention and management of NCDs in the region.

Limitations

The findings of this study should be interpreted within the context of several limitations. All employees included in the study were given access to their company's corporate-sponsored wellness initiatives within the period of 2021 to 2022, which were not limited to Naluri's digital health coaching. The potential use of additional health improvement initiatives outside of those described in this study may have inflated or concealed within- or between-group differences in health outcomes at follow-up. In addition, given that the sample groups were retrospectively determined based on health screening data, self-selection, and natural usage patterns, outcomes of the study are likely susceptible to selection bias, despite our best efforts to control for this statistically. Admittedly, the nonrandomized and noncontrolled design of this study limits proper inferences of causality as well as limits the representativeness of our findings.

Furthermore, the retrospective and observational nature of this study precludes us from controlling for measured or unmeasured confounding variables such as mental health status or the existence of comorbid physical health conditions [61]. The omission of these potential confounding variables, though unavoidable in retrospective, observational studies [61], can in fact lead to an overestimation or an underestimation of the observed associations we see between coaching status and employee health outcome at follow-up. However, the observational nature of our study does offer a reflection on the real-world outcomes and implications of offering digital coaching and health improvement initiatives in the workplace. Separately, we did not look into participant engagement and adherence to the interventions, and this hinders any inferences on the dose-response relationship between the intervention and our study outcomes [62]. Despite these limitations, the real-world setting of our study provides ecological validity to our findings, prompting the need for more real-world studies of digital health use in the region, more so given the low adoption of digital interventions outside of randomized controlled trials [63]. To further contextualize the effectiveness of digital health interventions in the workplace, there is a need for future studies to look into the long-term outcomes and real-world cost-effectiveness of these programs.

Conclusion

This study retrospectively investigated the year-to-year changes in employee health outcomes associated with a workplace digital behavioral coaching intervention. We found that an active coaching CDMP led to significant year-to-year improvements in multiple measures of employee health outcomes compared with those receiving passive coaching or no coaching at all. Despite methodological limitations, our study provides real-world evidence to support the use of digital workplace interventions in NCD prevention in the region.

Acknowledgments

The study authors wish to thank the cofounders of Naluri, Azran Osman-Rani, and Dr Jeremy Ting, for their support in driving this project. We further wish to thank the team at Naluri who were involved in the intervention delivery for this cohort of employees.

Data Availability

The data used and analyzed for this study are available from the corresponding author on reasonable request.

Authors' Contributions

AFAA and TO conceptualized the study. AFAA acquired and analyzed the data and wrote the original draft of the paper. TO reviewed and edited the manuscript. All authors approved the final version of the paper.

Conflicts of Interest

AFAA and TO are employed by Naluri Hidup Sdn Bhd, which created the mobile app and digital behavioral coaching intervention used in the study. The authors declare that this study received funding from Naluri Hidup Sdn Bhd.

References

1. World Health Organization. Noncommunicable Diseases Progress Monitor 2022. 2022. URL: https://www.who.int/publications-detail-redirect/9789240047761 [accessed 2023-03-23]
2. NCD Countdown 2030 collaborators. NCD countdown 2030: worldwide trends in non-communicable disease mortality and progress towards sustainable development goal target 3.4. Lancet 2018;392(10152):1072-1088 [FREE Full text] [doi: 10.1016/S0140-6736(18)31992-5] [Medline: 30264707]
3. World Health Organization. Noncommunicable Diseases in the South-East Asia. 2023. URL: https://www.who.int/southeastasia/health-topics/noncommunicable-diseases [accessed 2023-03-23]
4. World Health Organization. Noncommunicable Diseases Indonesia 2018 Country Profile. 2018. URL: https://cdn.who.int/media/docs/default-source/country-profiles/ncds/idn_en.pdf?sfvrsn=fbb0cdfb_33&download=true [accessed 2023-03-23]

5. Bloom DE, Chen S, McGovern M, Prettner K, Candeias V, Bernaert A, et al. The economics of non-communicable diseases in Indonesia. World Economic Forum. 2015. URL: https://www3.weforum.org/docs/WEF_The_Economics_of_non_Disease_Indonesia_2015.pdf [accessed 2015-04-20]
6. Allotey P, Davey T, Reidpath DD. NCDs in low and middle-income countries - assessing the capacity of health systems to respond to population needs. BMC Public Health 2014;14 Suppl 2(Suppl 2):S1 [FREE Full text] [doi: 10.1186/1471-2458-14-S2-S1] [Medline: 25082328]
7. Gadsden T, Downey LE, Vilas VDR, Peiris D, Jan S. The impact of COVID-19 on essential health service provision for noncommunicable diseases in the South-East Asia region: a systematic review. Lancet Reg Health Southeast Asia 2022;1:100010 [FREE Full text] [doi: 10.1016/j.lansea.2022.04.006] [Medline: 35769108]
8. Taher A, Gondhowiardjo SA. Comprehensive efforts to accelerate non-communicable disease services in the era of COVID-19 in Indonesia's suburban area. ASEAN Journal of Community Engagement 2022;6(1):152-168 [FREE Full text] [doi: 10.7454/ajce.v6i1.1167]
9. Gentili A, Failla G, Melnyk A, Puleo V, Tanna GLD, Ricciardi W, et al. The cost-effectiveness of digital health interventions: a systematic review of the literature. Front Public Health 2022;10:787135 [FREE Full text] [doi: 10.3389/fpubh.2022.787135] [Medline: 36033812]
10. Sweet CC, Jasik CB, Diebold A, DuPuis A, Jendretzke B. Cost savings and reduced health care utilization associated with participation in a digital diabetes prevention program in an adult workforce population. J Health Econ Outcomes Res 2020;7(2):139-147 [FREE Full text] [doi: 10.36469/jheor.2020.14529] [Medline: 32884964]
11. Tackling obesity in ASEAN: prevalence, impact, and guidance on interventions. The Economist Intelligence Unit. 2017. URL: https://www.evolveasia.org/wp-content/uploads/2019/11/EIU_Tackling-Obesity-in-ASEAN_Final-Report.pdf [accessed 2023-03-23]
12. Responding to the challenge of non-communicable diseases. International Labour Organization. 2021. URL: https://www.who.int/docs/default-source/ncds/uniatf-ilo-trifold-print-070920.pdf?sfvrsn=b4720cdb_2 [accessed 2023-03-23]
13. Pg Ismail PK, Koh D. Role of occupational health in managing non-communicable diseases in Brunei Darussalam. Glob Health Action 2014;7:25594 [FREE Full text] [doi: 10.3402/gha.v7.25594] [Medline: 25382684]
14. Preventing noncommunicable diseases in the workplace through diet and physical activity: WHO/World economic forum report of a joint event. World Health Organization, World Economic Forum. 2008. URL: https://www.who.int/publications/i/item/9789241596329 [accessed 2008-06-10]
15. Leonardi M, Scaratti C. Employment and people with non communicable chronic diseases: PATHWAYS recommendations and suggested actions for implementing an inclusive labour market for all and health in all sectors. Int J Environ Res Public Health 2018;15(8):1674 [FREE Full text] [doi: 10.3390/ijerph15081674] [Medline: 30087249]
16. World Health Organization. Noncommunicable Diseases: What Ministries of Labour and Employment Need to Know. 2016. URL: https://www.who.int/publications-detail-redirect/WHO-NMH-NMA-16.90 [accessed 2023-03-24]
17. Aveyard P, Hardeman W, Horne R. Promoting health behaviours at the individual level for NCD prevention and control. In: Banatvala N, Bovet P, editors. Noncommunicable Diseases. London: Routledge; 2023:355-361.
18. Dineen-Griffin S, Garcia-Cardenas V, Williams K, Benrimoj SI. Helping patients help themselves: a systematic review of self-management support strategies in primary health care practice. PLoS One 2019;14(8):e0220116 [FREE Full text] [doi: 10.1371/journal.pone.0220116] [Medline: 31369582]
19. Ehde DM, Dillworth TM, Turner JA. Cognitive-behavioral therapy for individuals with chronic pain: efficacy, innovations, and directions for research. Am Psychol 2014;69(2):153-166. [doi: 10.1037/a0035747] [Medline: 24547801]
20. White CA. Cognitive behavioral principles in managing chronic disease. West J Med 2001;175(5):338-342 [FREE Full text] [doi: 10.1136/ewjm.175.5.338] [Medline: 11694487]
21. Wolever RQ, Simmons LA, Sforzo GA, Dill D, Kaye M, Bechard EM, et al. A systematic review of the literature on health and wellness coaching: defining a key behavioral intervention in healthcare. Glob Adv Health Med 2013;2(4):38-57 [FREE Full text] [doi: 10.7453/gahmj.2013.042] [Medline: 24416684]
22. Howarth A, Quesada J, Silva J, Judycki S, Mills PR. The impact of digital health interventions on health-related outcomes in the workplace: a systematic review. Digit Health 2018;4:2055207618770861 [FREE Full text] [doi: 10.1177/2055207618770861] [Medline: 29942631]
23. Kempf K, Röhling M, Martin S, Schneider M. Telemedical coaching for weight loss in overweight employees: a three-armed randomised controlled trial. BMJ Open 2019;9(4):e022242 [FREE Full text] [doi: 10.1136/bmjopen-2018-022242] [Medline: 30975666]
24. Silberman JM, Kaur M, Sletteland J, Venkatesan A. Outcomes in a digital weight management intervention with one-on-one health coaching. PLoS One 2020;15(4):e0232221 [FREE Full text] [doi: 10.1371/journal.pone.0232221] [Medline: 32353035]
25. Wilson MG, Castro Sweet CM, Edge MD, Madero EN, McGuire M, Pilsmaker M, et al. Evaluation of a digital behavioral counseling program for reducing risk factors for chronic disease in a workforce. J Occup Environ Med 2017;59(8):e150-e155 [FREE Full text] [doi: 10.1097/JOM.0000000000001091] [Medline: 28650899]
26. Widmer RJ, Allison TG, Keane B, Dallas A, Lerman LO, Lerman A. Using an online, personalized program reduces cardiovascular risk factor profiles in a motivated, adherent population of participants. Am Heart J 2014;167(1):93-100. [doi: 10.1016/j.ahj.2013.09.019] [Medline: 24332147]

27. Lee Y, Lee NY, Lim HJ, Sung S. Weight reduction interventions using digital health for employees with obesity: a systematic review. Diabetes Metab Syndr Obes 2022;15:3121-3131 [FREE Full text] [doi: 10.2147/DMSO.S384450] [Medline: 36246516]
28. Widmer RJ, Allison TG, Keane B, Dallas A, Bailey KR, Lerman LO, et al. Workplace digital health is associated with improved cardiovascular risk factors in a frequency-dependent fashion: a large prospective observational cohort study. PLoS One 2016;11(4):e0152657 [FREE Full text] [doi: 10.1371/journal.pone.0152657] [Medline: 27092940]
29. Fritz M, Fromell H. How to dampen the surge of non-communicable diseases in Southeast Asia: insights from a systematic review and meta-analysis. Health Policy Plan 2022;37(1):152-167 [FREE Full text] [doi: 10.1093/heapol/czab138] [Medline: 34791261]
30. Adiratna Y, Astono S, Fertiaz M, Subhan, Sugistria CAO, Prayitno H, et al. The Indonesian national occupational safety and health profile 2022. Republic of Indonesia: Ministry of Manpower. 2022. URL: https://satudata.kemnaker.go.id/satudata-public/2022/10/files/publikasi/1675661836215_The%2520Indonesian%2520OSH%2520Profile%25202022_Compiled%252020230124_compressed.pdf [accessed 2024-04-08]
31. Prochaska JO, Velicer WF. The transtheoretical model of health behavior change. Am J Health Promot 1997;12(1):38-48. [doi: 10.4278/0890-1171-12.1.38] [Medline: 10170434]
32. Nguyen TL, Xie L. Incomparability of treatment groups is often blindly ignored in randomised controlled trials - a post hoc analysis of baseline characteristic tables. J Clin Epidemiol 2021;130:161-168. [doi: 10.1016/j.jclinepi.2020.10.012] [Medline: 33080343]
33. Hilbert S, Stadler M, Lindl A, Naumann F, Bühner M. Analyzing longitudinal intervention studies with linear mixed models. TPM-Testing, Psychometrics, Methodology in Applied Psychology 2019;26(1):101-119. [doi: 10.4473/TPM26.1.6]
34. Oberauer K. The importance of random slopes in mixed models for Bayesian hypothesis testing. Psychol Sci 2022;33(4):648-665 [FREE Full text] [doi: 10.1177/09567976211046884] [Medline: 35357978]
35. Elze MC, Gregson J, Baber U, Williamson E, Sartori S, Mehran R, et al. Comparison of propensity score methods and covariate adjustment: evaluation in 4 cardiovascular studies. J Am Coll Cardiol 2017;69(3):345-357 [FREE Full text] [doi: 10.1016/j.jacc.2016.10.060] [Medline: 28104076]
36. Barnicot K, Crawford M. Conclusions and questions from a non-randomised comparison of routine clinical services implementing different treatment models for borderline personality disorder. Psychol Med 2019;49(16):2812-2814. [doi: 10.1017/S0033291719002447] [Medline: 31551098]
37. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J. Stat. Softw 2015;67(1):1-48. [doi: 10.18637/jss.v067.i01]
38. Kuznetsova A, Brockhoff PB, Christensen RHB. lmerTest package: tests in linear mixed effects models. J. Stat. Soft 2017;82(13):1-26. [doi: 10.18637/jss.v082.i13]
39. Luke SG. Evaluating significance in linear mixed-effects models in R. Behav Res Methods 2017;49(4):1494-1502. [doi: 10.3758/s13428-016-0809-y] [Medline: 27620283]
40. Lenth RV, Bolker B, Buerkner P, Giné-Vázquez I, Herve M, Jung M, et al. emmeans: estimated marginal means, aka least-squares means. 2023. URL: https://cran.r-project.org/web/packages/emmeans/index.html [accessed 2023-06-27]
41. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Second Edition. New York: Routledge; 1988.
42. Gershkowitz BD, Hillert CJ, Crotty BH. Digital coaching strategies to facilitate behavioral change in type 2 diabetes: a systematic review. J Clin Endocrinol Metab 2021;106(4):e1513-e1520. [doi: 10.1210/clinem/dgaa850] [Medline: 33206975]
43. Nkhoma DE, Soko CJ, Bowrin P, Manga YB, Greenfield D, Househ M, et al. Digital interventions self-management education for type 1 and 2 diabetes: a systematic review and meta-analysis. Comput Methods Programs Biomed 2021;210:106370. [doi: 10.1016/j.cmpb.2021.106370] [Medline: 34492544]
44. Liu Z, Chen S, Zhang G, Lin A. Mobile phone-based lifestyle intervention for reducing overall cardiovascular disease risk in Guangzhou, China: a pilot study. Int J Environ Res Public Health 2015;12(12):15993-16004 [FREE Full text] [doi: 10.3390/ijerph121215037] [Medline: 26694436]
45. Ramachandran A, Snehalatha C, Ram J, Selvam S, Simon M, Nanditha A, et al. Effectiveness of mobile phone messaging in prevention of type 2 diabetes by lifestyle modification in men in India: a prospective, parallel-group, randomised controlled trial. Lancet Diabetes Endocrinol 2013;1(3):191-198. [doi: 10.1016/S2213-8587(13)70067-6] [Medline: 24622367]
46. Lee CL, Kaur H, Swee WCS. A structured lifestyle intervention combined with mobile health application to reduce chronic disease risk among employees at a university workplace. IeJSME 2022;16(1):3-15. [doi: 10.56026/imu.16.1.3]
47. Zahtamal Z, Rochmah W, Prabandari YS, Setyawati LK. Effects of multilevel intervention in workplace health promotion on workers' metabolic syndrome components. Kesmas: National Public Health Journal 2017;11(4):198-204. [doi: 10.21109/kesmas.v11i4.1279]
48. Sathish T, Dunleavy G, Soljak M, Visvalingam N, Nazeha N, Divakar U, et al. Risk factors for non-communicable diseases at baseline and their short-term changes in a workplace cohort in Singapore. Int J Environ Res Public Health 2019;16(22):4551 [FREE Full text] [doi: 10.3390/ijerph16224551] [Medline: 31752089]

49. GBD 2021 Diabetes Collaborators. Global, regional, and national burden of diabetes from 1990 to 2021, with projections of prevalence to 2050: a systematic analysis for the Global Burden of Disease Study 2021. Lancet 2023;402(10397):203-234 [FREE Full text] [doi: 10.1016/S0140-6736(23)01301-6] [Medline: 37356446]
50. Chodavadia P, Teo I, Poremski D, Fung DSS, Finkelstein EA. Prevalence and economic burden of depression and anxiety symptoms among Singaporean adults: results from a 2022 web panel. BMC Psychiatry 2023;23(1):104. [doi: 10.1186/s12888-023-04581-7] [Medline: 36782116]
51. Ratanasiripong P, Kaewboonchoo O, Bell E, Susilowati I, Isahak M, Harncharoen K, et al. Depression , anxiety and stress among small and medium enterprise workers in Indonesia , Malaysia , Thailand , and Vietnam. International Journal of Occupational Health and Public Health Nursing 2016;3(2):13-29 [FREE Full text]
52. Murray E, Sweeting M, Dack C, Pal K, Modrow K, Hudda M, et al. Web-based self-management support for people with type 2 diabetes (HeLP-Diabetes): randomised controlled trial in English primary care. BMJ Open 2017;7(9):e016009 [FREE Full text] [doi: 10.1136/bmjopen-2017-016009] [Medline: 28954789]
53. Tang JCH, Abraham C, Greaves CJ, Nikolaou V. Self-directed interventions to promote weight loss: a systematic review and meta-analysis. Health Psychol Rev 2016;10(3):358-372. [doi: 10.1080/17437199.2016.1172979] [Medline: 27091296]
54. Tate DF, Lutes LD, Bryant M, Truesdale KP, Hatley KE, Griffiths Z, et al. Efficacy of a commercial weight management program compared with a do-it-yourself approach: a randomized clinical trial. JAMA Netw Open 2022;5(8):e2226561 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.26561] [Medline: 35972742]
55. Heshka S, Anderson JW, Atkinson RL, Greenway FL, Hill JO, Phinney SD, et al. Weight loss with self-help compared with a structured commercial program: a randomized trial. JAMA 2003;289(14):1792-1798. [doi: 10.1001/jama.289.14.1792] [Medline: 12684357]
56. Chatterjee A, Prinz A, Gerdes M, Martinez S. Digital interventions on healthy lifestyle management: systematic review. J Med Internet Res 2021;23(11):e26931 [FREE Full text] [doi: 10.2196/26931] [Medline: 34787575]
57. O'Connor S, Hanlon P, O'Donnell CA, Garcia S, Glanville J, Mair FS. Understanding factors affecting patient and public engagement and recruitment to digital health interventions: a systematic review of qualitative studies. BMC Med Inform Decis Mak 2016;16(1):120 [FREE Full text] [doi: 10.1186/s12911-016-0359-3] [Medline: 27630020]
58. El Benny M, Kabakian-Khasholian T, El-Jardali F, Bardus M. Application of the eHealth literacy model in digital health interventions: scoping review. J Med Internet Res 2021;23(6):e23473 [FREE Full text] [doi: 10.2196/23473] [Medline: 34081023]
59. WHO Global Observatory for eHealth. mHealth: New Horizons for Health Through Mobile Technologies: Second Global Survey on eHealth. 2011. URL: https://apps.who.int/iris/handle/10665/44607 [accessed 2023-06-27]
60. Wang X, Shi J, Lee KM. The digital divide and seeking health information on smartphones in Asia: survey study of ten countries. J Med Internet Res 2022;24(1):e24086 [FREE Full text] [doi: 10.2196/24086] [Medline: 35023845]
61. Gardner AL, Charlesworth M. How to write a retrospective observational study. Anaesthesia 2023;78(4):521-525 [FREE Full text] [doi: 10.1111/anae.15831] [Medline: 35918805]
62. McVay MA, Bennett GG, Steinberg D, Voils CI. Dose-response research in digital health interventions: concepts, considerations, and challenges. Health Psychol 2019;38(12):1168-1174 [FREE Full text] [doi: 10.1037/hea0000805] [Medline: 31580127]
63. Fleming T, Bavin L, Lucassen M, Stasiak K, Hopkins S, Merry S. Beyond the trial: systematic review of real-world uptake and engagement with digital self-help interventions for depression, low mood, or anxiety. J Med Internet Res 2018;20(6):e199 [FREE Full text] [doi: 10.2196/jmir.9275] [Medline: 29875089]

Abbreviations

BP: blood pressure
CDMP: chronic disease management program
EAP: employee assistance program
HbA1c: hemoglobin A1c
HDL: high-density lipoprotein
HIPAA: Health Insurance Portability and Accountability Act
LDL: low-density lipoprotein
LMM: linear mixed-effects model
Mdiff: mean difference
NCD: noncommunicable disease
WHO: World Health Organization

Edited by L Buis; submitted 03.07.23; peer-reviewed by H Malini, C Curatoli, J Thielecke; comments to author 29.03.24; revised version received 09.04.24; accepted 07.06.24; published 10.09.24.

Please cite as:
Abdul Aziz AF, Ong T
Real-World Outcomes of a Digital Behavioral Coaching Intervention to Improve Employee Health Status: Retrospective Observational Study
JMIR Mhealth Uhealth 2024;12:e50356
URL: https://mhealth.jmir.org/2024/1/e50356
doi:10.2196/50356
PMID:39255013

Original Paper

Feasibility and Engagement of a Mobile App Preparation Program (Kwit) for Smoking Cessation in an Ecological Context: Quantitative Study

Luz Adriana Bustamante Perez[1], MSc; Lucia Romo[1,2], MSc, PhD; Oulmann Zerhouni[1,3], MSc, PhD

[1]Laboratoire EA 4430-Clinique Psychanalyse Developpement, Department of Psychology, University of Paris Nanterre, Nanterre, France

[2]Inserm-Le Centre de Recherche en Epidémiologie et Santé des Populations 1018 UPS, Paris, France

[3]Université de Rouen, Rouen, France

Corresponding Author:
Luz Adriana Bustamante Perez, MSc
Laboratoire EA 4430-Clinique Psychanalyse Developpement, Department of Psychology, University of Paris Nanterre
200 avenue de la République
Nanterre, 92001
France
Phone: 33 783192547
Email: adriana.bustamante93@gmail.com

Abstract

Background: Mobile health apps can facilitate access to effective treatment and therapeutic information services. However, the real-world effectiveness of mobile apps for smoking cessation and their potential impact in everyday settings remain unclear.

Objective: In an ecological context, this study aimed to estimate the engagement rate of a mobile app–based smoking cessation preparation program and its potential impact on users' willingness, ability, and readiness to quit smoking.

Methods: A total of 2331 "organic users" (ie, users who discover and install a mobile app on their own, without any prompts) chose 1 of 2 program versions of the mobile app (Kwit): the basic version or the premium version. Both versions were identical in design, with 4 more evidence-based content items and strategies in the premium version. Outcomes were analyzed based on automated data registered in the app (engagement rate, motivation to quit, motivation type, motivation levels, and satisfaction level). Mann-Whitney and χ^2 tests were used to compare the results of both groups.

Results: As expected, in the ecological context, a high dropout rate was observed at different moments. A significant difference was observed between the 2 versions (n=2331; χ^2_1=5.4; *P*=.02), with a proportionally higher engagement rate in the premium version (premium=4.7% vs basic=2%). Likewise, differences were also observed between the 2 groups in terms of reasons to quit (n=2331; χ^2_4=19; *P*≤.001; V=0.08), motivation type (n=2331; χ^2_7=14.7; *P*=.04), and motivation level. Users of the app's premium version more frequently reported "well-being" (23.3% vs 17.9%) and "planning a pregnancy" (7.4% vs 4.4%) as their primary reasons for quitting smoking compared to those with the basic version. Moreover, they reported being more likely to be driven in the smoking cessation process by intrinsic motivation (premium=28% vs basic=20.4%), as well as feeling significantly more willing (*z* score=156,055; *P*≤.001; Cohen *d*=0.15), able (*z* score=172,905; *P*=.04; Cohen *d*=0.09), and ready (*z* score=166,390; *P*=.005; Cohen *d*=0.12) to stop smoking than users who had the basic version before completion of the preparation program. Among participants who finished each version of the program (premium: 9/189, 4.8%; basic: 47/2142, 2.19%), significant improvements in motivation levels were observed in both groups, although in different areas for each group (willingness levels for the premium group and ability for the basic group).

Conclusions: These results suggest that even in ecological contexts where engagement rates are meager, the Kwit preparation program can address ambivalence by increasing willingness to change, self-confidence, and readiness to quit among its users, especially those who feel less able to do so. Further development and evaluations are needed to better understand determinants for regular mobile health apps.

(JMIR Mhealth Uhealth 2024;12:e51025) doi:10.2196/51025

KEYWORDS

smoking cessation; digital intervention; behavior change techniques; attrition rate; mobile app; preparation program; motivation; quit smoking; ecological settings; mobile phone

Introduction

Tobacco Consumption

The World Health Organization (WHO) considers smoking an epidemic that affects >1 billion people worldwide. It is a public health issue for many countries because tobacco use is one of the leading risk factors for a significant number of deaths and disabilities worldwide, and although preventable, the economic and social costs of disease burden remain very high [1]. On the basis of WHO recommendations, France's national tobacco reduction program successfully addressed tobacco consumption, as the quit rate has been decreasing significantly since implementing several tobacco control policies in 2014 (ie, the reimbursement of approximately EUR €150 [US $166] per year for nicotine replacement products, monitoring of health warnings, media restrictions, and taxation of cigarette packs) [2].

Although all these policies have had some success in increasing smoking cessation, the long-term abstinence rate is still extremely low despite 60% of worldwide smokers expressing their willingness to quit. In France, at least 29.9% of daily smokers tried to stop for at least a week the previous year without long-term success [2]. In addition, results from double-masked clinical trials show that the willingness to quit and actual reductions in smoking behavior are limited by the content of nicotine present in cigarettes, even in participants who were not initially interested in quitting [3,4]. Complex factors must be considered for long-term abstinence, such as the social influence of marketing misinformation about tobacco consumption, nicotine pharmacokinetics, everyday cue conditioning, withdrawal symptoms, and a comprehensive discussion of treatment options and goals [5].

Treatment for Smoking Cessation

Multiple treatment approaches have been explored because of the complexity of tobacco dependence. To date, 2 major treatment categories have been used to treat tobacco use disorder: pharmacological treatments (eg, nicotine replacement therapy, bupropion, and varenicline) and nonpharmacological approaches (eg, motivational interviewing, cognitive and behavioral therapies, and acceptance and commitment therapy) [6]. In addition, new treatment goals have been proposed as alternatives to total and abrupt abstinence. These include gradually reducing cigarette consumption or using methods like snus and vaping to lower the risks associated with smoking, particularly for individuals with high nicotine dependence [4,5].

Both WHO and the French Ministry of Health recognize the impact of mobile health (mHealth) by promoting communication on social networks and using new technologies (eg, helplines, government mobile apps, and specific websites) to facilitate access to information and, thus, health care for smoking cessation [7,8].

mHealth Apps for Smoking Cessation

Overview

mHealth refers to emerging technologies for accessing health and medical services through mobile devices [9]. This includes mobile eHealth apps (known as mHealth apps), which are software that provide health and wellness services, designed exclusively for mobile devices (smartphones and tablets) [10]. With nearly 70% of the world's population having a smartphone and access to the internet, mHealth apps can promote access to smoking cessation treatments [11]. mHealth apps can facilitate access to effective treatments and therapeutic information services by (1) adapting and translating "active principles" (the term used to refer to the various strategies and practices of evidence-based behavioral and cognitive therapies) into a digital format; (2) facilitating communication with health professionals, allowing a more flexible and personalized relationship (chat and social groups); and (3) encouraging a sense of responsibility and commitment to one's health through "nudges," such as positive reinforcement through messages or notifications, habit tracking, regular feedback, and audiovisual support [12].

Despite the many benefits of using mobile apps for smoking cessation, research is still in its infancy. It faces specific challenges: the quality of content, the range of potential uses, and engagement with mHealth apps for smoking cessation.

Content Quality of Mobile Apps for Smoking Cessation

According to a systematic review and meta-analysis, there is a relationship between the number of cognitive and behavioral techniques used in smoking cessation programs and short- and long-term effectiveness [9]. However, according to a recent review of the content analysis of mobile apps for smoking cessation in France, in 2020, (1) the average quality of mobile apps was 3.5 out of 5 (median 3.1; range 1-3) based on the Mobile App Rating Scale, (2) most of the apps made little use of the "active ingredients" of evidence-based therapies (between 4 and 38 out of 93 indexed cognitive behavioral therapy [CBT] techniques), and (3) there was a lack of information that delivers proper advice regarding the use of approved pharmacotherapy or the implementation of behavioral techniques specific to helping people prepare for smoking cessation[10]. In contrast, only some apps have been extensively studied and made available in France [11]; however, they still need to be presented using a proper taxonomy of behavior change techniques to facilitate further research on their mobile app content quality.

The Range of Potential Uses of an mHealth App

Current studies primarily address smoking cessation mobile apps as self-help material used as a stand-alone treatment by their users [12-14], not focusing on other treatment goals of smoking cessation interventions such as facilitating risk reduction and relapse prevention [5,15,16]. Moreover, mobile apps can also be used as a complementary tool that enables continual accompaniment of the patient throughout their treatment by health professionals. The transtheoretical change

model framework developed by framework of Prochaska and Prochaska [17] will be used to illustrate this purpose.

This model states that a long-term change (ie, smoking cessation) results from different stages. Each stage has specific challenges; once addressed successfully, the person can continue their change journey by following the next stage. The first stages of change (precontemplation, contemplation, and preparation) are about raising awareness of the target behavior and creating an action plan by identifying the benefits of quitting, decreasing the perceived difficulties of change, and acknowledging the pharmacological and psychological therapy. Hence, mobile apps can highlight therapeutic information on dependency and how pharmaceutical aids work. Mobile apps can also be used as an ecological momentary assessment tool through which it is possible to monitor cravings in real time to collect information that helps to formulate hypotheses on the precipitating and maintaining factors that lead to tobacco consumption [18]. On the basis of the previous assessment, it is possible to individualize therapeutic interventions by understanding the individual experience of craving through repeated and ecological measures [15].

In the action phase, the person can change the target behavior [17]. Therefore, to manage cravings differently instead of smoking, a person can experiment with new methods that can be transitional and complementary, such as the use of pharmacological (ie, nicotine replacement therapy or medications) and psychological substitute strategies (ie, avoiding a specific place or person). In this phase, mobile apps can be a tool that prompts experimentation of new strategies by giving different types of feedback on behavior (ie, monitoring nicotine consumption, meditation exercises, distraction activities, etc) as well as being a motivational levier that highlights the benefits of smoking cessation or its reduction in different aspects of the user's life (health benefits and savings).

In the maintenance phase, "relapse" is perceived as the norm and not the exception, so it can be a learning opportunity in which the goal is to anticipate and prevent high-risk situations as well as navigate through the moment of the onset of craving by creating a relapse prevention plan [17]. Mobile apps can be a tool that helps monitor emotional changes and track withdrawal symptoms, both very common to relapse. On the basis of previous use, the mobile app can also provide feedback indicating treatment effectiveness and the risk of relapses.

As a result of these insights, Kwit app—a French mobile app initially conceived for maintaining smoking cessation—released the Kwit's 9-step preparation program for smoking cessation (9s-Kwit's program) in July 2021. A new feature specifically designed for users self-identifying as being in the preparation phase of smoking cessation according to the transtheoretical model of change developed by Prochaska and Prochaska [17], this program is based on cognitive and behavioral therapy for smoking cessation as well as dialectical and behavioral therapy, motivational interviewing, and self-determination theory.

Engagement With Digital Interventions Such as mHealth Apps for Smoking Cessation in Real Settings

As in any treatment protocol, "adherence to therapies is an important determinant of treatment success" [19]. The smoking cessation intervention attrition rate could be as high as 77% [20]. Conversely, mobile apps for chronic diseases face similar challenges as other mHealth apps, with attrition rates ranging from 43% (95% CI 16-63) in randomized controlled trials to 49% (95% CI 27-70) in observational studies [21]. To date, the dropout rate of some smoking cessation apps tested in a controlled scenario could vary from 19.3% after 28 days of use [22] to 39.5% after 87.3 days [23]. However, it remains to be seen what the attrition rate of smoking cessation mobile apps in ecological contexts would be [24-27], particularly with organic users. In this study, "organic users" or "organic installs"—terms primarily used in mobile app marketing—refers to individuals who discover and install a mobile app on their own, without any influence from paid marketing. Crucially, for the purposes of our research, these organic users are characterized by their autonomy to download the app and pursue smoking cessation. This autonomy distinguishes them significantly from users who participate in studies for compensation or who are mandated to use the app by health care institutions. Therefore, it is hypothesized that organic users will exhibit different behaviors and levels of engagement in smoking cessation compared to their nonorganic counterparts.

In conclusion, even if an app includes all the therapeutic guidelines, it can only be effective if used regularly [10,24,27]. Understanding how users engage periodically with their health treatment is challenging for mHealth apps [24,27,28]. Therefore, it is essential to identify and assess the main factors that favor using mobile apps for smoking cessation. To this end, the various research methods should ensure ecological validity in a design of evaluation setup that matches the user's real work context [29] and ensures that the behavioral responses obtained represent the natural behaviors of people who want to quit smoking with the help of a mobile app [30].

The aims of this study were, therefore, to (1) explore, in an ecological context, the engagement rate of a 9s-Kwit's program on a mobile app and (2) estimate the impact of the program on the motivation level among users who finished the program.

We hypothesized that (1) given the ecological context of this study, we will observe a remarkably high dropout rate between 80% and 95% of the users participating in the study. In addition, we sought to examine some key factors' role in the Kwit program's engagement rate. Specifically, (2) we hypothesized that the attrition rate within the intervention could be explained by the version of the program (basic and premium), the lack of relevance and understandability of the program content, the primary motivation to quitting smoking (ie, "money" or "family"), the lower levels of internalization to stop smoking (external motivation), and the lower level of motivation to quit before starting the program. Finally, we hypothesized that (3) users' perception of their willingness, ability, and readiness to stop smoking will significantly improve after finishing the 9s-Kwit's program.

Methods

Overview

This observational study was conducted from July 4, 2021, to July 28, 2021, in an ecological context. It represents the natural behaviors of people who downloaded and used the app to prepare for smoking cessation. In detail, this ecological context includes the following (1) the participants were organic users (users that installed the app because of their search results and without having encountered paid advertisements before onboarding); (2) only information essential for the everyday use of the mobile app was collected, and only data required for the study were exported to be analyzed (ie, no demographic data were collected before onboarding); (3) no compensation was offered to the users for their participation in the study; and (4) they checked a nonpreemptive box to give their consent to the collection and analysis of data related to their use of the app for research and improvement purposes. If an organic user chooses not to allow Kwit app to use their data for research, this preference is recorded in the server. This choice does not affect the quality of the services offered by the Kwit app. All users, regardless of their data use choice, have the option to access the premium version of the app. This ensures that there is no ethical compromise in the provision of services.

Ethical Considerations

The local French Ethical Committee (South-East) validated the protocol study on March 25, 2021. The research identification number is 2020-A02733-36, and the committee's reference is CPP 20.10.02.44945 [31]. This study adheres to the ethical principles outlined in the Declaration of Helsinki. This feasibility study is part of a prospective study registered on April 8, 2022 (ClinicalTrials.gov NCT05318651). It aims to identify critical determinants of smoking cessation mobile app use among smokers seeking to quit.

Participants

Participants were aged at least 18 years and had a compatible smartphone (iOS 13.3 and above) with regular internet service. In total, 2 groups of participants were followed according to the type of program version they had access to, basic or premium.

Before participating in the program, participants were required to create an account that pseudonymized their personal information, including billing information where applicable. The Kwit app offered 2 versions: premium and basic. The choice between them was influenced by whether the participant opted for a 7-day free trial.

The premium version was available to users who opted for the free trial and provided full access to all features of the program. In contrast, the basic version limited the user's access to only 4 levels of the program (steps 0, 1, 2, and 8). It is important to note that both versions visually presented the same program dashboard screen to ensure a consistent user experience. The user journey leading to the 9s-Kwit's program dashboard is detailed in Multimedia Appendix 1.

In terms of pricing structure, the premium version of the Kwit app was offered through a 7-day free trial during the study period from July 4, 2021, to July 28, 2021. This trial period gave users the flexibility to discontinue the service if they wished. After the trial period, if the subscription was continued, it was considered an agreement to a monthly membership fee of EUR €9 (US $9.96). This pricing was strategically set to approximate the cost of a pack of cigarettes in France at the time, aligning the cost of the app with a tangible smoking expense and making the value proposition relatable and practical for users.

Intervention

The 9-step Kwit's program was developed after 2 years of research and development, as part of a PhD thesis, in collaboration with the University of Paris Nanterre. The content of this program was defined and coordinated by the cognitive behavioral psychologist and PhD student to be then executed by a team composed of a user experience designer, 3 software developers with at least 10 years of experience, a researcher in public health and clinical research, and 4 interns with a psychology master's degree. The 9s-Kwit's program has been developed for commercialization and designed for direct delivery as a mobile app intervention on Apple and Google Stores. This program is based primarily on cognitive and behavioral therapy for smoking cessation but is also inspired by dialectical and behavioral therapy, motivational interviewing, and the self-determination theory.

The 9s-Kwit's program comprises 9 steps from step 0 (s0) to step 8 (s8). Through different activities, each step aims to explore a specific aspect of tobacco use that has already been identified as essential to consider in treating smoking cessation [9,32]. A presentation of the 9s-Kwit's program's home screen, a summary of its contents, and an example of the activity screen per step are presented in Multimedia Appendix 2.

The presentation of the activity starts as follows: each activity starts with an informative screen that introduces the activity's main goals, then the activity itself, and ends with a feedback screen (Multimedia Appendix 3). When the user finishes an activity, the next one is unblocked. Different types of feedback coexist to inform the user that an activity is completed: (1) the color of the activity title changes (ie, in s0 from gray to blue), (2) information about the date and hour is presented instead of the average duration of the activity, (3) the icon next to the title activity changes from a lock to an arrow to show that the next activity is then accessible, and (4) the activities that are already done have a summary card if you click on it. Once all activities of the same step are completed, the next step is unlocked.

The initial step (s0) consists of becoming aware of one's motivation to engage in a change process, but to change the target behavior, motivation alone is not enough; it is crucial to get out of "self-pilot" mode and adopt an "observer attitude" to understand the association between the target behavior, its antecedents, and its consequences. Therefore, in the first step (s1), users are encouraged to monitor their behavior and create a baseline of the contexts and intensity of cravings and the inner strategies that have helped them cope. The behavioral baseline provides users with an objective measure that assists them in

recognizing patterns of cravings and helpful strategies. The second step (s2) focuses on the different components of dependency, and by using the Horn scale, it is possible to explain how opposing stimuli can induce people to smoke (ie, relaxation vs stimulation). On the basis of the results, users receive advice on new behavioral strategies. The third step (s3) proposes that users categorize cravings according to their short-term benefits (pleasure or relief) and then some mindfulness exercises. The fourth step (s4) defines a goal and a road map adapted to the user's resources. Steps 5, 6, and 7 teach new coping strategies for the 3 dependence types (s5: environmental, s7: psychical, and s7: psychological). The final step (s8) invites users to define a quit date and acknowledge the cognitive barriers to quitting. The program encourages being kind to oneself when slips and relapses occur and celebrating each small step toward the desired change. None of the program versions used in this study have the gamification layer where users earn points at the end of each activity, have their avatar, or any educational reading content or community layout that now exists in the platform. Table 1 details all activities of each of the 9s-Kwit's program according to the taxonomy on behavior change techniques developed by Michie et al [33].

Table 1. Overview of the behavioral and cognitive techniques (BCTs) and goals presented by step (s) and activity (a) of the 9s-Kwit's program based on the taxonomy developed by Michie et al [33].

ABS[a,b]	Type of activity	BCT number and label[a]
s0a1	Identify main reason to quit: health, well-being, economy, family, and planning a pregnancy	1.9 Commitment
s0a2	Identify the level of willingness, ability, and readiness to quit smoking using a visual analog scale	1.9 Commitment
s0a3	Define a landmark that recalls reason to quit when craving arises	1.9 Commitment
s0a4	Identify the degree of internalization of abstinence motivation using the French Smoking Cessation Motivation Scale	1.9 Commitment
s1a1	Introduction of the "plus bottom": craving arousal monitoring for 24 hours: context, intensity, and action (let it go or smoke)	2.1 Self-monitoring of the behavior
s1a2	Mindfulness exercises (3 minutes) focus on observing automatic responses when cravings arise, encouraging an observer's attitude	8.2 Behavior substitution
s2a1	Horn scale: physical, psychological, and behavioral dependency assessment	2.7 Feedback on outcome of behavior
s2a2	Identify the relationship with smoking through an open self-questionary about when the behavior installs and the reason it maintains in the time	4.2 Information about antecedents
s3a1	Identification of the most difficult craving to overcome in the journey and it impacts on the body and cognition	5.3 Monitoring of emotional consequences
s3a2	Classify cravings into anchored (routine) and reflex (contextual) types	4.1 Instruction on how to perform a behavior
s3a3	Mindfulness exercises (3 minutes)	8.2 Behavior substitution
s4a1	Assessment of current resources available to execute an action plan	1.4 Action planning
s4a2	Three types of experiences were proposed to be completed within 24 hours based on S4a1 for craving management: (1) "Act consciously" mindfulness exercises: for those not willing to stop smoking; (2) "Choosing your cravings": based on S3a1, users anticipated strategies to reduce "anchored" cravings; and (3) "Overcoming all cravings:" users anticipated craving management strategies for all cravings	1.1 Goal setting behavior and outcome
s4a2	Three types of experiences were proposed to be completed within 24 hours based on S4a1 for craving management: (1) "Act consciously" mindfulness exercises: for those not willing to stop smoking; (2) "Choosing your cravings": based on S3a1, users anticipated strategies to reduce "anchored" cravings; and (3) "Overcoming all cravings:" users anticipated craving management strategies for all cravings	11.1 Regulation by pharmacological support
s5a1	Assess nicotine dependency with the Fagerström Test for Nicotine and give proper feedback about dependency level and offer recommendations for pharmacological therapy options that could complement the app, if necessary	8.2 Behavioral substitution
s6a1	Introduction of new features of the "plus bottom": tracking nicotine substitutes or the vape consumption, mindfulness exercises, drinking water, and breathing exercises. Advise on how to avoid exposure to specific social and contextual cues	12.1 Restructuring the physical and social environment
s7a1	Identify barriers to quitting, such as learned helplessness, fear of withdrawal symptoms, and automatic behaviors related to craving management	13.3 Incompatible beliefs
s7a2	Recognize a behavior as it is and not as part of user's identity	13.4 Valued self-identity
s8a1	Identifying barriers to quit and propose some solutions to overcome it: the fear to stop smoking (eg, weight, stress, and concentration)	8. Behavior rehearsal
s8a2	Identify the level of willingness, ability, and readiness to quit smoking using a visual analog scale of s0a1 and give proper feedback	13.3 Incompatible beliefs

ABS[a,b]	Type of activity	BCT number and label[a]
s8a3	Recognize old behavior goals and consolidate experience through feedback. Define a future quit date	1.11 Review behavior goals

[a]This program content is created based primarily on cognitive and behavioral therapy for smoking cessation as well as dialectical and behavioral therapy, motivational interviewing, and self-determination theories.

[b]Activity by step=first activity of the step 0 (s0a1).

Outcome Variables and Measurements

In total, 4 different types of measurements were used: ecological momentary assessment to measure engagement rate and version of the program ("a method of data collection whereby a record is made each time a predefined event occurs" [18]); a visual analog scale (VAS) to measure user's perception of their willingness, ability, and readiness to quit smoking (the VAS slider ranged from 0 to 10); a Likert-type scale to measure the internalization degree to stop smoking, which can range from none (amotivation) to completely internalized (intrinsic motivation); and a multiple-choice question to measure user's relevance and understandability of the program content and the users' main reason to quit. Examples can be found in Multimedia Appendix 3.

Engagement Rate Toward the 9s-Kwit's program

The engagement rate is the ratio of users who completed the program from the first use (step 0) to the last activity proposed in 9s-Kwit's program (step 8).

Motives for Quitting Smoking

On the basis of previous studies on reasons for quitting smoking [34-36], users were asked to choose 1 of 5 reasons: health, wellness, money, family, and planning a pregnancy, using the following statement: "My main reason for quitting smoking is..." (Multimedia Appendix 3).

Motivation Level: The Willingness, Ability, and Readiness to Quit Scale

Using a VAS, users rated their perception of their willingness, ability, and readiness to quit smoking at the beginning (S0) and the end of the program (S0). For this purpose, each question was presented on a single screen and in the following order: (1) "To what extent this change is a priority for you right now?" (*willingness*), (2) "To what extent are you confident in your ability to change right now?" (*ability*), and (3) "To what extent do you feel ready to take action?" (*readiness*). This way, users could move a slider from 0 (lowest) to 10 (highest) on each screen (Multimedia Appendix 3). For statistical analysis purposes, scores <3 are defined as "low," scores between 4 and 7 are defined as "moderate," and scores >7 are defined as "high."

Nature of Motivation to Quit Smoking

The French Smoking Cessation Motivation Scale (F-SCMS) is a self-reported measure based on the self-determination theory, demonstrating good internal consistency (α=.86; ωh=0.7; ωt=0.89) and content validity (common-fit index=0.905, standardized root mean square residual=0.045, and root mean square error of approximation=0.087). The 18 items are divided into 6 subscales (each composed of 3 items) corresponding to the degree of internalization that participants have toward quitting smoking. The subscales are presented from no internalization to complete internalization of a behavioral change process: (1) amotivation, (2) external, (3) introjected, (4) mixed, (5) identified, (6) integrated, and (7) intrinsic motivation. We have previously validated the scale (F-SCMS) used in this study and explained the internalization process in a separate publication [37].

The scale was presented as follows within the app. Each screen was composed from top to bottom in the following order: (1) the statement "Right now," (2) the item appearing as a card, and (3) the list with 5 Likert-type response options ranging from 1 (does not match at all) to 5 (matches exactly). Once the user answered all the items, they could receive predefined feedback corresponding to their degree of internalization [37].

Perceived Content Relevance of s0

The last activity of s0 of the program was the evaluation by the users of the comprehensibility and relevance of the contents presented during this step. The question was presented as "You have completed Step X. How would you define it?" with the following 3 options: (a) *understandable and relevant*, (b) *understandable but irrelevant*, and (c) *not understandable and irrelevant.*

Statistical Analysis

We performed descriptive analyses to characterize the 2 groups' baseline samples and outcomes of interest. We conducted the Mann-Whitney test and a χ^2 test of association to compare differences between both groups at the beginning of the program in terms of the main reason to quit, the motivation level to quit smoking, the nature of the motivation, and the perceived content relevance at the beginning of the program. We conducted a Student *t* test (2-tailed) to estimate the program's impact on the motivation levels of the program completion for each group. For each test, a *P* value <.05 will indicate statistical significance. Analyses were conducted using Jamovi V2.3.8.

Results

Overview

A total of 2331 users started the preparation program. Overall, 91.89% (2142/2331) of the initial sample used the basic version of the program, and 8.1% (189/2331) used the premium version. As a reminder, the app's premium version allowed access to each program step's activities (from s0 to s8), whereas participants with the basic version had access only to s0, s1, s2, and s8 activities.

Table 2 presents the distribution of participants of both versions of the program in terms of (1) motives for quitting smoking, (2) motivation level to quit smoking, (3) motivation nature, and

(4) perceived content relevance of step 0. In general, participants' main reason for quitting smoking was health (1173/2331, 50.32%), followed by savings (450/2331, 19.31%), well-being (428/2331, 18.36%), family (172/2331, 7.38%), and planning a pregnancy (108/2331, 4.63%). At baseline, 80.65% (1880/2331) of the total sample was moderately motivated to quit smoking according to the Willingness, Ability, and Readiness to Quit (WAR) scale mean score, and 50.45% (1176/2331) fell into the 2 highest categories of internalization level for smoking cessation—integrated motivation (687/2331, 29.47%) and intrinsic motivation (489/2331, 20.98%)—according to the F-SCMS scale. Upon completion of step 0, the content of this step was considered by 71.55% (1668/2331) of the users as understandable and relevant (option A), by 10.98% (256/2331) as understandable but irrelevant (option B), by 0.82% (19/2331) as incomprehensible and irrelevant (option C), and 16.65% (388/2331) of the initial sample did not answer the question.

Table 2. The quantity and proportion of participants engaged in each activity of step 0[a] (motives for quitting smoking, level of motivation to quit smoking, nature of motivation, and perceived content relevance).

	Total (N=2331), n (%)	Basic step 0 (n=2142), n (%)	Premium step 0 (n=189), n (%)	Chi square (*df*)	*P* value
s0a1[b]: Motives for quitting smoking				19 (4)	.001
Health	1173 (50.3)	1073 (50)	100 (52.9)		
Well-being	428 (18.4)	384 (17.9)	41 (23.3)		
Savings	450 (19.3)	434 (20.3)	14 (8.5)		
Family	172 (7.4)	157 (7.3)	14 (7.9)		
Planning a pregnancy	108 (4.6)	94 (4.4)	14 (7.4)		
__[c]	0	0	0		
s0a2[d]: Level of motivation to quit smoking				6.48 (3)	.09
Low	144 (6.2)	140 (6.5)	4 (2.1)		
Moderate	1880 (80.7)	1724 (80.5)	156 (82.5)		
High	235 (10)	212 (9.9)	23 (12.2)		
—	72 (3.1)	66 (3.1)	6 (3.1)		
s0a4[e]: Nature of motivation to quit smoking				14.7 (7)	.04
Amotivation	81 (3.5)	78 (3.6)	3 (1.6)		
External	224 (9.6)	205 (9.6)	19 (10.1)		
Introjected	88 (3.8)	82 (3.8)	6 (3.2)		
Mixed	250 (10.7)	228 (10.6)	22 (11.6)		
Identified	155 (6.6)	138 (6.4)	17 (9)		
Integrated	687 (29.5)	635 (29.6)	52 (27.5)		
Intrinsic	489 (21)	436 (20.4)	53 (28)		
—	357 (15.3)	340 (15.9)	17 (9)		
s0: perceived content relevance				9.70 (3)	.02
Option A[f]	1668 (71.6)	1516 (70.8)	152 (80.4)		
Option B[g]	256 (11)	238 (11.1)	18 (9.5)		
Option C[h]	19 (0.8)	19 (0.9)	0		
—	388 (16.6)	369 (17.2)	19 (10)		

[a]s0: first step of 9s-Kwit's program.

[b]s0a1: first activity of step 0.

[c]Not applicable.

[d]s0a2: second activity of step 0.

[e]s0a4: fourth activity of step 0.

[f]Indicates understandable and relevant.

[g]Indicates understandable but irrelevant.

[h]Indicates not understandable and irrelevant.

Motives for Quitting Smoking

According to the χ^2 test of association and Cramér value (n=2331; χ^2_4=19; *P*≤.001; V=0.08), users of each version of the program have reported proportionally different reasons for quitting smoking. Those who choose to get the premium version of the program were less likely to choose savings as their main reason to quit smoking (premium=8.5% vs basic=20.3%) but more likely to choose the well-being (premium=23.3% vs basic=17.9%) and planning a pregnancy (premium=7.4% vs basic=4.4%) options.

Level of Motivation to Quit Smoking

Concerning the total score of the WAR scale, 3 levels of motivation were calculated based on the average score of 3 subscales: low (0-3), medium (4-7), and high (8-10). Most users (1880/2331, 80.7%) reported a moderate motivation to quit smoking at the beginning of the preparation program (s0; 1880/2331, 80.65%; mean 5.85 out of 10, SD 1.68). There was no significant difference between basic (n=2076) and premium (n=183) users when comparing these 3 categories (n=2331; χ^2_3=6.48; *P*=.09) at s0a2. Nevertheless, when the ordinal data were compared using the Mann-Whitney test, the 2 groups differed significantly but with a very slight effect within the total score and every subscale. Users who had the premium version of the app reported feeling significantly more willing (*z* score=56,055; *P*≤.001; Cohen *d*=0.15), capable (*z* score=172,905; *P*=.04; Cohen *d*=0.09), and ready (*z* score=166,390; *P*=.005; Cohen *d*=0.12) to stop smoking than users who had the basic version before completion of the preparation program.

Nature of Motivation to Engage With a Smoking Cessation Process

The internalization degree of motivation to quit smoking differs significantly between basic and premium users according to the χ^2 test of association (n=2331; χ^2_7=14.7; *P*=.04). Premium users are less likely to score in the “amotivated” profile (premium=1.6% vs basic=3.6%) and more likely to be driven by intrinsic motivation (premium=28% vs basic=20.4%) to engage in a smoking cessation process. In addition, the proportion of participants for whom no answer was received is lower in the premium group than in the basic group (premium=17/183, 9% vs basic=340/2076, 15.9%).

Perceived Content Relevance of s0

The content assessment of s0 differs between premium and basic users according to the χ^2 test of association (n=2331; χ^2_3=9.7; *P*≤.02), and by contrast, there are fewer premium users for whom a response is missing (–7.2%).

Engagement Rate Toward the 9s-Kwit’s program for Smoking Cessation in an Ecological Context

Overview

As shown in Figure 1, the engagement rate with the program drops significantly at 3 different moments. The biggest dropout concerns both samples (basic and premium) in between activities s0a4 and s1a1 (basic=–90.3% and premium=–70.1%). The second dropout was observed in the basic user sample between s2a2 and S8a1 (–4.2%), and the third dropout was within the premium sample from S4a2 to S5a1 (–12.5%). The proportion of participants who started each program activity is detailed in Table 3.

Figure 1. Engagement rate of the basic and the premium sample by activity for each step of the 9s-Kwit's program. a: activity; s: step.

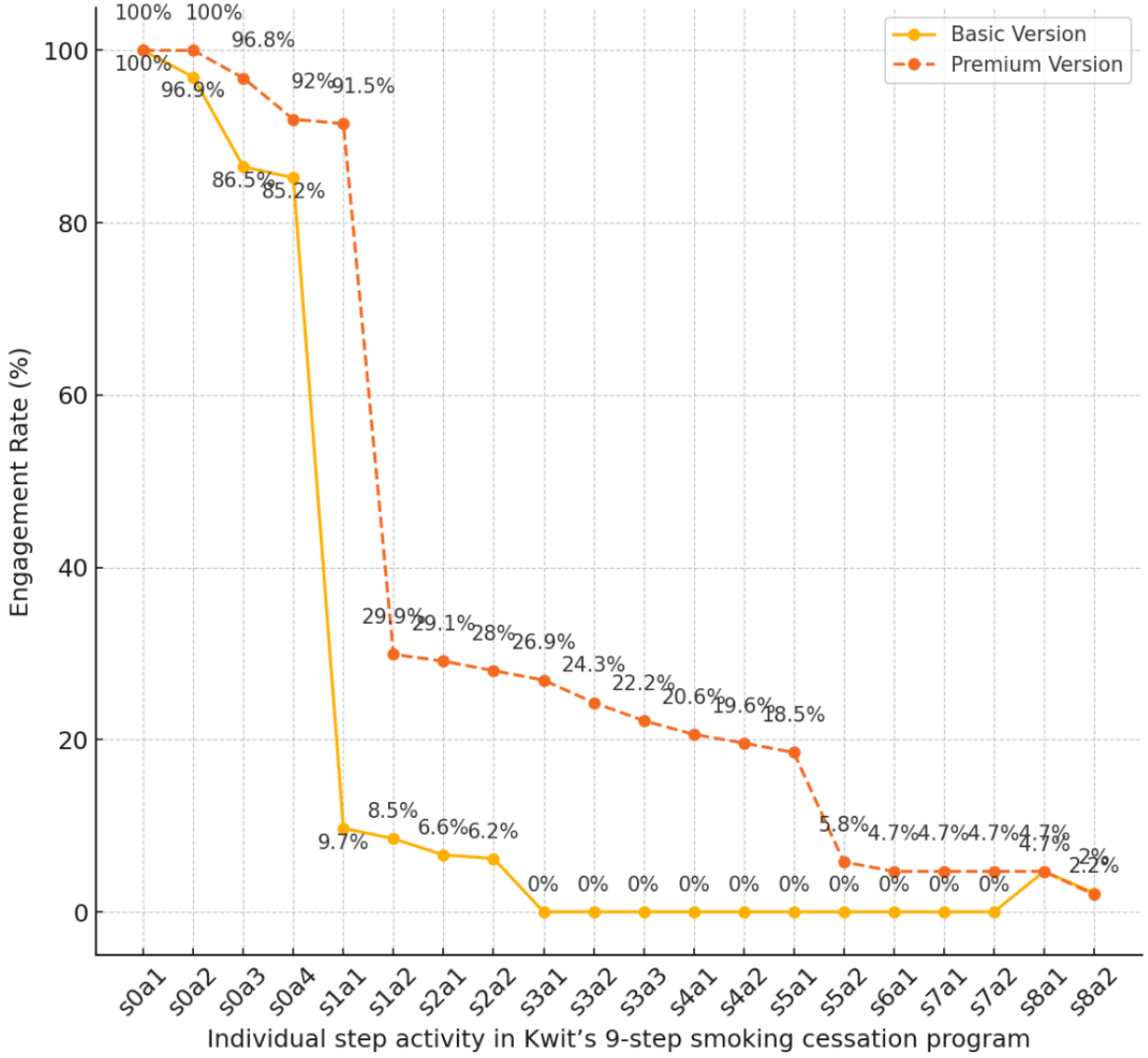


Table 3. Proportion of participants who completed each activity of the 9s-Kwit's program.

Activities by steps[a]	Total sample started, n (%)	Basic sample started, n (%)	Premium sample started, n (%)	Chi-square test of association and Cramér value (*df*)	*P* value
s0a1	2331 (100)	2142 (100)	189 (100)	—[b]	—
s0a2	2259 (96.9)	2076 (96.9)	183 (96.8)	—	—
s0a3	2238 (96)	1853 (86.5)	174 (92)	—	—
s0a4	1998 (85.7)	1825 (85.2)	173 (91.5)	7.8 (1)	.005
s1a1	265 (11.4)	209 (9.7)	56 (29.9)	68.1 (1)	≤.001
s1a2	236 (10.1)	181 (8.5)	55 (29.1)	—	—
s2a1	195 (8.4)	142 (6.6)	53 (28)	—	—
s2a2	184 (7.9)	133 (6.2)	51 (26.9)	—	—
s3a1	46 (1.9)	0	46 (24.3)	—	—
s3a2	42 (1.8)	0	42 (22.2)	—	—
s3a3	39 (1.6)	0	39 (20.6)	—	—
s4a1	37 (1.6)	0	37 (19.6)	—	—
s4a2	35 (1.5)	0	35 (18.5)	—	—
s5a1	11 (0.5)	0	11 (5.8)	—	—
s6a1	9 (0.4)	0	9 (4.7)	—	—
s7a1	9 (0.4)	0	9 (4.7)	—	—
s7a2	9 (0.4)	0	9 (4.7)	—	—
s8a1	59 (2.5)	50 (2.4)	9 (4.7)	—	—
s8a2	57 (2.4)	48 (2.2)	9 (4.7)	—	—
s8a3	54 (2.3)	44 (2)	9 (4.7)	5.43 (1)	.02

[a]The percentage in brackets is calculated in relation to the initial number of participants per group.

[b]Not applicable.

In addition, based on the χ^2 test of association and Cramér value, there is a significant difference between basic and premium users in terms of the engagement rate at the end of step 0 (s0a4; n=2331; χ^2_1=7.8; *P*=.005), the beginning of step 1 (s1a1; n=2331; χ^2_1=68.1; *P*≤.001), and the end of the preparation program (S8a3; n=2331; χ^2_1=5.43; *P*=.02). Premium users were more likely than basic users to finish the last activity of step 0 (s0a4; premium=91.5% vs basic=85.2%), start the new step (s1a1; premium=26.9% vs basic=9.7%), and complete all the program's activity until the definition of the quit date (S8a3; premium=4.7% vs basic=2%).

The Impact of the Program Completion on Users Who Finished the Program on Motivation Levels Toward Quitting

From an initial sample of 2331 ecological users, only 57 (2.4%) completed activity s8a2, which assesses the motivation levels regarding WAR scale. Table 4 presents the WAR scale and subscales scores from the beginning (s0a2) to the end of the program (s8a2) of the 2 users' groups who finished the program (57/2331, 2.4%) as well as the Mann-Whitney score.

Table 4. Description of the Total Willingness, Ability, and Readiness To Quit (WAR) scale scores and subscales at the start (s0a2) and end (s8a2) of the 9s-Kwit's program (n=57).

WAR scale	Total sample		Basic sample (n=48)		Premium sample (n=9)		Mann-Whitney test[a]		*P* value
	Values, mean (SD)	Values, median	Values, mean (SD)	Values, median	Values, mean (SD)	Values, median	U	Cohen *d*[b]	
Willingness									
s0a2	7.46 (1.89)	8.00	7.40 (1.87)	7.00	7.78 (2.11)	8.00	188	0.12	.54
s8a2	7.75 (1.91)	8.00	7.63 (1.94)	8.00	8.44 (1.67)	9.00	160	0.26	.21
Ability									
s0a2	5.33 (1.89)	5.00	5.02 (1.71)	5.00	7.00 (2.06)	7.00	94	0.56	.007
s8a2	6.25 (2.14)	6.00	5.98 (2.09)	6.00	7.67 (1.94)	7.00	118	0.45	.03
Readiness									
s0a2	6.30 (1.99)	6.00	6.23 (1.86)	6.00	6.67 (2.69)	6.00	201	0.06	.75
s8a2	6.86 (2.06)	7.00	6.77 (2.05)	6.50	7.33 (2.12)	7.00	184	0.15	.48
Total									
s0a2	6.36 (1.37)	6.33	6.22 (1.22)	6.00	7.15 (1.68)	6.33	190	0.12	.57
s8a2	6.95 (1.68)	6.67	6.79 (1.64)	6.67	7.81 (1.73)	7.60	144	0.32	.13

[a]*P* values for the Mann-Whitney test were statistically significant at $P<.05$.

[b]Cohen *d* effect size indicating the magnitude of the effect, where a small effect is approximately $d=0.2$, a medium effect is $d=0.5$, and a large effect is $d=0.8$.

To estimate the program's impact on the motivation levels of the program completion, Student *t* tests were conducted within each group.

At the beginning of the program, users who finished (57/2331, 2.4%) reported feeling, on average, highly willing to quit smoking (7.46/10), moderately confident in their abilities to do so (5.33/10), and moderately ready to start the quitting journey (6.30/10). A Mann-Whitney *U* test was performed to assess whether scores on the WAR scale and its subscales differed significantly between users of each group both at the beginning and at the end of the program. The results indicated that only subscale Ability differs significantly between both groups' samples (basic and premium), with a moderate effect size ($U=6.86$; *z* score=118; $P=.02$; $d=0.45$) before (s0a2) and after the program completion (s8a2). Users who had the app's premium version reported feeling significantly more capable of smoking cessation than users who had the basic version before (basic=5.02; premium=7) and after (basic=5.98; premium=7.33) the completion of the preparation program.

Premium users who finished 9s-Kwit's program reported, with a high effect size, significantly higher scores than at the beginning of the program on the Willingness subscale ($t_8=2.83$; $P=.02$; Cohen $d=0.9$) and the total score of the WAR scale ($t_8=3.16$; $P=.01$; Cohen $d=1.05$). Conversely, users who completed the basic version of the program (4 steps) reported, with moderate effect size, significantly higher end-of-program scores in the Ability subscale ($t_{47}=3.17$; $P=.003$; Cohen $d=0.46$) and a total score of the WAR scale ($t_{47}=2.92$; $P=.005$; Cohen $d=0.42$), and with low effect size, significantly higher end-of-program scores in the Readiness subscale ($t_{47}=1.857$; $P=.07$; Cohen $d=0.26$).

Discussion

Overview

This study aimed, in an ecological context, to explore the engagement rate of a 9s-Kwit's program on a mobile app (Kwit app); to examine some moderating agents that could contribute to the engagement rate we observed; and to estimate the impact of the program on the motivation level among users who finished the program. An ecological context refers to research conducted in natural, real-world settings as opposed to controlled laboratory environments. This approach is crucial for several reasons. First, it embeds our investigation in real-life scenarios where participants interact with the Kwit app as part of their daily routines. This allows us to gain authentic insights into user behaviors, app engagement levels, and the program's impact, mirroring genuine use patterns more accurately than laboratory conditions could. Such insights are invaluable for designing interventions that are not only effective in theory but also practical and beneficial for users in their real-world context.

Principal Findings

Overview

A total of 2331 users, all of whom were active smokers without a set quit date, started the 9s-Kwit's program. Most users (2142/2331, 91.89%) opted for the basic version of the program, whereas 8% (n=189) selected the premium one. Only 2.3% (54/189) of the initial sample reached the last step of the program (s8), which consisted of defining a quit date.

Notably, there was a marked difference in completion rates between the 2 versions: 4.7% (9/189) of premium users finished all program steps, which was higher than the 2.95% (44/2142)

completion rate observed in the basic version. This significant disparity in engagement rates (n=2331; χ^2_1=5.4; *P*=.02) highlights a greater commitment among premium users, despite them having more steps to complete.

To better understand the attrition phenomenon, we monitored user engagement at each program activity and observed 3 prominent dropouts. The first and most significant was observed between the last activity of step 0 (s0a4) and the beginning of step 1 (s1a1), with higher attrition rate among basic users compared to premium users (Table 2). The second dropout was observed in the basic user sample between s2a2 and s8a1 (–4.2%), and the third dropout was within the premium sample from activity s4a2 to s5a1 (–12.5%). These dropout patterns can be attributed to 3 main factors: (1) specific user characteristics (ie, areas of motivation and perceived content relevance), (2) the specific attributes of the app's user experience and the cognitive effort required to engage with different activities, and (3) typical attrition rates for smoking cessation interventions and the regular use of mHealth solutions.

Specific User Characteristics at the Beginning of the Program

In this study, data collection was confined to what was essential for the program's use, focusing on motivation areas and perceived content relevance, without collecting demographic information or insights into users' health and technology literacy. Following the rationale that user motivation to quit and user satisfaction with the app determines user engagement, higher engagement rates were anticipated. Only 26.9% (56/2142) of premium users and 9.7% (209/189) of basic users progressed beyond the initial stage of the program (from s0a4 to s1a2). Notably, while >70% (1668/2331) of both groups rated the content as "understandable and relevant," there was a significant perceived relevance gap, with premium users rating the app content more relevant by >10 points compared to basic users, as detailed in Table 2. This discrepancy suggests that content relevance may play a substantial role in influencing engagement levels, particularly among different user groups.

Significant differences were also observed in the initial sample (N=2331) between the 2 groups in all motivational measures: the reasons for quitting smoking, the WAR scale, and the internalization level to quit smoking. Premium users, who were more likely to choose wellness and family planning as their primary reason for quitting, also showed greater willingness, ability, and readiness to quit smoking at baseline compared to basic users, who were primarily motivated by financial savings. Interestingly, there is a difference between the willingness to quit and the user's self-perception of their ability and readiness to quit. On average, both groups reported high levels of willingness to quit smoking (premium=7.78; basic=7.40) but a moderate perception of their ability (premium=7; basic=5.02) and readiness (premium=6.67; basic=6.23) to do so. In addition, higher levels of internalization were observed in the premium sample with fewer incomplete responses (premium=15.9%; basic=9%) and lower amotivation profile (premium=1.6%; basic=3.6%). They were more represented in profiles of identified (premium=9%; basic=6.4%) and intrinsic motivation (premium=28%; basic=20.4%).

In summary, while both groups reported being highly willing to quit smoking, premium users reported feeling more confident and prepared to start their smoking cessation process, more driven by internalized motivation and reasons to quit (ie, well-being vs savings), and more committed from the early stages of the program.

The Specific Attributes of the Mobile App User Experience

Despite being user centered and clinically guided, the 9s-Kwit's program design might contribute to user disengagement at specific points: (1) for both versions in the transition between the last activity of step 0 (s0a4) and the beginning of step 1 (s1a1); (2) for the basic version, between step 2 (s2a2) and step 8 (s8a1); and (3) for the premium version, between step 4 and step 5. The AIM-ACT framework helps explain this attrition: emotional and motivational mismatches (AIM) and attentional or contextual misalignments (ACT) could hinder user engagement, especially during key step changes. This suggests the need for aligning app activities more closely with users' affective states, goals, and contextual realities to reduce dropout rates [27].

The first problem arises in the transition of the app between the dashboard and the activity screen (Multimedia Appendix 2). When users complete all activities in a step, they are redirected to the dashboard, where an animation (an opening of a small padlock) indicates the next available step. This transition can create friction rather than facilitate smooth progression, potentially discouraging motivated users from starting new activities. Alternative design elements, such as more prominent visual cues (eg, a red circle above the start step) or direct navigation to the next activity screens, could improve user engagement. Second, for basic version users, the dashboard could be improved by more clearly indicating accessible steps, perhaps by visually distinguishing available and unavailable steps, thus reducing confusion and friction in progression. For example, it would have been possible to slightly blur the steps to which they did not have full access to highlight the user's path and decrease friction between step 2 and step 8.

The third issue involves the cognitive load of certain activities, particularly those that require more active engagement, such as reading or logging cravings for a long period. This could be the case for the dropout in between step 4 (s4) and step 5 (s5); 3 types of exercises were proposed in s4 to be performed for 24 hours, setting goals for craving management. The design of these activities, including their duration and nature, might impact users' perception of the effort-to-reward ratio as insufficient to continue engaging with the app program.

In general, using any game element, such as achieving certain milestones, will encourage learning new skills and knowledge by making learning more enjoyable and engaging and by enhancing self-esteem and motivation, which are highly associated with long-term smoking cessation [25,38]. The version of the Kwit app used in this review represents the first stage of an evolving series of modifications. This baseline version was designed to establish a solid evidence-based core, filling a gap where many mobile smoking cessation apps have lacked a scientific foundation [10,39].

The Standardized Attrition Rate for Smoking Cessation and mHealth Solutions

The dropout rate observed in this study could be because of the nature of tobacco use disorders and the selection criteria. Tobacco use disorders are well studied, with various effective treatments available; however, they typically exhibit dropout rates ranging from 10.8% to 77% in participants who had already defined a quit date as part of the selection criteria [20]. According to model developed by Prochaska and Prochaska [17], our participants would be at a lower stage because they would be just in the preparation phase.

Moreover, attrition is a known challenge in mHealth, with an average dropout rate of 43%, with higher rates in research under ecological conditions (98%) and lower rates in randomized controlled trials (up to 30%) [21].

In the case of studies on smoking cessation apps, the reporting of quit rates has yet to be standardized and is not always reported. This threatens the validity of some studies by introducing selection bias and making it difficult to compare the effect of such apps [11]. For instance, dropout rates in such studies range from 19.3% at 28 days [22] to 39.5% at 87.3 days of use [23] in controlled settings with English-speaking participants, while studies in ecological settings with French-speaking participants report rates of 65% at 30 days and 80% at 90 days of use [39] with a 1-year data collection period (vs less than a month for our study).

The Impact of the Program Completion on Users' Motivation Levels Toward Quitting

To estimate the program's impact on the motivation level, the analysis was based on 57 users who completed the activity s8a2 of the program: 2.2% of the users from the basic version (n=48) and 4.7% from the premium version (n=9). At the beginning of the program, both groups reported high levels of willingness to quit smoking, with an average willingness score of 7.46 out of 10. However, they reported only moderate levels of ability and preparedness, with average scores of 5.33 out of 10 for their perceived ability to quit and 6.30 out of 10 for their readiness to begin the quitting process.

After the program completion, substantial improvements in the WAR total score were observed in both groups, with different areas of improvement for each group. The premium users improved their willingness levels, whereas users with the basic version benefited in the ability and readiness category. Given the difficulty of assessing the effect of each individual component, we can only hypothesize about the program content's effect on WAR scores. It is worth noting that throughout her work, Michie has linked several techniques based on behavioral theories to different behavior change as smoking cessation to the point that there is now an ontology and a whole system that would allow this question to be answered in future studies [40,41]. Therefore, the content of the 9s-Kwit's program was mainly described through the taxonomy developed by Michie et al [32,33].

Given the intricate challenge of isolating and evaluating the effect of individual components within the program, our analysis regarding its impact on WAR scale scores must be approached with caution. As such, we posit that the program's influence on users' sense of competence and readiness likely stems from the implementation of various CBT techniques previously recognized as efficacious for smoking cessation in personal intervention settings [42].

Extensive research into personal interventions across varied demographics and environments has pinpointed specific CBT techniques—and their synergies—to be instrumental in elevating smoking cessation success rates. The 9s-Kwit's program incorporates these proven techniques, including goal setting for behaviors and outcomes, using problem-solving strategies, revisiting and refining behavioral goals, fostering commitment, delivering feedback on behavioral outcomes, enabling behavior substitution, facilitating behavioral practice and rehearsal, and reconfiguring both physical and social contexts to bolster behavior change efforts [42]. These components are strategically combined to amplify the probability of cessation among users.

Moreover, the program integrates principles from acceptance and commitment therapy right from the outset, urging users to align with their core values, such as family, well-being, or personal health. Using metaphors, these values are presented as guiding beacons during moments of temptation (cravings) [43]. The "plus" button feature distinctively separates sensations from actions, offering various strategies such as the 4Ds (Delay, Deep breath, Drink water, Do something else) and meditation to manage cravings effectively [44]. In the second stage, the Horn scale demystifies the multifaceted reasons behind smoking habits, proposing tailored strategies for each [6]. Culminating in stage 8, the program prompts users to confront their quitting fears, providing customized solutions to navigate these apprehensions successfully and asking them to set a quit date.

These results suggest that Kwit's preparation program can address ambivalence by increasing willingness to change, self-confidence, and readiness to quit among its users, especially those who felt less able to do so (users with the basic version). Self-efficacy or confidence in one's abilities has been identified as a determinant in long-term abstinence [45].

Strengths and Limitations

Strengths

The strength of this type of study is simultaneously one of its major limitations; being observational in an ecological context, it offers a unique "snapshot" of real-life behavior of a group of users of a mobile smoking cessation app, providing insight into how its use may impact motivation outcomes. This approach not only sheds light on the "science of attrition" by recognizing when intervention dropouts occur but also aids in identifying real-life adoption challenges, thereby facilitating the generation of hypotheses to enhance health care treatment adherence, particularly with the evolution of commercial mHealth apps [21,29]. To the best of our knowledge, only 2 other studies have explored using mHealth tools [46,47] to treat smokers who are ambivalent about quitting, and these studies focused on French-speaking smokers. This is the third study to examine this topic within an ecological context (ie, popular mobile apps) and with a large sample size [39,48].

Limitations

The low generalizability of the findings was a limitation for several reasons, outlined in the following sections.

Lack of Individual-Level Measures

The absence of detailed individual-level data, particularly on smoking habits, dependency levels, and mental health issues such as anxiety and depression, is a significant limitation. These factors are known to be associated with lower adherence to treatment [48,49]. In additional, the study did not collect detailed demographic information, such as age and gender, before the use of the 0-step preparation program. Conducted within an ecological context and adhering to General Data Protection Regulation, the study's data collection was limited to what was essential for app functionality, excluding pre-use demographic details.

Sample Size and Demographics

The study's sample is limited to iOS users, excluding Android users, who may differ in terms of quit attempts, quit date settings [50], and sociodemographics [51]. Given the global market share and diverse user base of Android, this limitation means the findings may not fully represent the broader population of smartphone users. The decision to focus on iOS users was driven by specific constraints at the time of the research design, making the results more reflective of the iOS user group.

Short Recruitment and Follow-Up Duration

The short duration of the recruitment phase and follow-up period is another limitation. Smoking cessation is a complex process often involving cycles of relapse and recovery. A brief follow-up may only capture a snapshot of this process, potentially overlooking critical data on sustained abstinence or long-term relapse rates. It may also fail to adequately assess dropout rates and reasons for disengagement, which are crucial for understanding and improving app efficacy.

Small Sample Size of Premium Users

The explanatory power of the study is also limited by the small sample size of premium users who completed the 9s-Kwit program. Although similar attrition rates have been reported in other ecological studies [39,48], the small number of premium users reduces the generalizability of the findings.

Future Perspectives

To address all these limitations and investigate user engagement over extended periods, we have established a new protocol in collaboration with the University Paris Nanterre, registered in ClinicalTrials.gov (NCT05318651). This initiative aims to provide a more comprehensive understanding of the program's impact over time.

Despite the limitations outlined, studies such as ours are critical for shedding light on 1 stage of the mHealth app design process. This research provided the Kwit design team with critical early insights that allowed for rapid iterative adjustments through A/B testing before launching the program on the Android platform and localizing it for different language audiences. As for 2024, the newly developed program integrates game-like features (awarding users points for successfully completing activities, offering different levels, and allowing for avatar customization), educational reading materials related to the topic of each step of the program, and a community design that was not previously accessible. It also introduces educational reading materials and a community design, which were not previously available on the platform. It remains to be seen how this proactive and responsive approach to app design not only optimizes user engagement but also avoids significant costs. It underscores the critical role of early and ongoing user feedback in refining mHealth interventions to ensure that they are both effective and engaging.

Rather than focusing on increasing abstinence as a stand-alone intervention, new studies should aim to improve the use of these and other interventions (as a complementary treatment rather than a comparative one) [48] and consider dropout rates as a measure of intervention quality. A meager retention rate highlights the current inability to satisfy users' expectations and needs of those seeking to quit [21]. To increase the duration and intensity of app use, qualitative studies are needed to obtain a user-centered perspective. In addition, small trials with users can help guide decision-making on design tweaks or new features. In contrast, through a quasi-experimental observational study based on the technology acceptance model, it would be possible to identify the determinants of engagement with a smoking cessation app [31].

Conclusions

This study on the first version of the 9s-Kwit's program provides crucial insights into the challenges and potential of mHealth apps in facilitating smoking cessation efforts. Despite a modest completion rate—with only 2.3% (57/2331) of the participants reaching the program's final step—the data unveil a significant difference in engagement levels between users of the basic and premium versions.

At the program's onset, participants displayed a strong willingness to quit, yet their self-assessed ability and readiness were only moderate. Notably, premium users reported feeling more capable of quitting than basic users—a sentiment that not only persisted but also intensified by the program's end. Subsequent evaluations indicated significant enhancements in willingness, ability, and readiness to quit across both user groups, with premium users showing an increased willingness and basic users experiencing considerable gains in self-assessed ability and readiness to quit.

These results imply that, despite the study's limitations, the 9s-Kwit's program may address user ambivalence, thereby enhancing willingness, confidence, and readiness to quit, particularly among those who initially felt less capable. Such improvements in self-efficacy are vital, as they play a pivotal role in achieving long-term smoking cessation success [45].

Ultimately, this research adds to the broader discourse on digital health interventions, underscoring the critical importance of adaptive design and user-centric approaches in developing effective mHealth solutions [24,26,27]. The observed attrition patterns highlight the intricate interaction between user characteristics, the cognitive demands of the app, and the inherent challenges of smoking cessation. This emphasizes the

essential need for early, iterative design processes that are informed by user feedback.

However, the study's conclusions must be approached with caution because of limitations such as the small sample size, the exclusive focus on iOS users, and the lack of comprehensive demographic data. By acknowledging these limitations, we have initiated a new research protocol with the University of Paris Nanterre (NCT05318651) to extend our investigation. This study aims to assess the long-term impact of the 9s-Kwit's program on a more diverse and extensive participant base.

Acknowledgments

The authors want to thank Mathias Lebot, Laurent Cruchet, Lucas Gallego and Edwan Rebuffe, who actively participated in the first phase of this study during their internship in their master's program in cognitive and behavioral clinical psychology. Mickael Ehrminger participated in the initial data curation. Lara Perez, user interface researcher and product designer, brought the content to life. The funders had no role in the study's design; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Conflicts of Interest

LAB undertook consultancy and research for a mobile app for smoking cessation (Kwit SAS) under a PhD contract involving the National Association of Research and Technology and the University of Paris Nanterre. All other authors declare no other conflicts of interest.

Multimedia Appendix 1
User flow for accessing Kwit's smoking cessation preparation program.
[PNG File , 61 KB - mhealth_v12i1e51025_app1.png]

Multimedia Appendix 2
The 9-step preparation program dashboard screen.
[PNG File , 98 KB - mhealth_v12i1e51025_app2.png]

Multimedia Appendix 3
Presentation of the scales.
[PNG File , 84 KB - mhealth_v12i1e51025_app3.png]

References

1. WHO report on the global tobacco epidemic 2021: addressing new and emerging products. World Health Organization. 2021 Jul 27. URL: https://www.who.int/publications/i/item/9789240032095 [accessed 2021-08-28]
2. Cogordan C, Quatremère G, Andler R, Guignard R, Richard JB, Nguyen-Thanh V. [Dialogue between general practitioner and patient regarding tobacco and alcohol consumption, from the patient's standpoint]. Rev Epidemiol Sante Publique 2020 Nov 26;68(6):319-326 [FREE Full text] [doi: 10.1016/j.respe.2020.10.005] [Medline: 33162267]
3. Donny EC, Denlinger RL, Tidey JW, Koopmeiners JS, Benowitz NL, Vandrey RG, et al. Randomized trial of reduced-nicotine standards for cigarettes. N Engl J Med 2015 Oct;373(14):1340-1349. [doi: 10.1056/nejmsa1502403]
4. Hatsukami DK, Carroll DM. Tobacco harm reduction: past history, current controversies and a proposed approach for the future. Prev Med 2020 Nov;140:106099 [FREE Full text] [doi: 10.1016/j.ypmed.2020.106099] [Medline: 32335031]
5. Palmer AM, Toll BA, Carpenter MJ, Donny EC, Hatsukami DK, Rojewski AM, et al. Reappraising choice in addiction: novel conceptualizations and treatments for tobacco use disorder. Nicotine Tob Res 2022 Jan 01;24(1):3-9 [FREE Full text] [doi: 10.1093/ntr/ntab148] [Medline: 34270729]
6. Arrêt de la consommation de tabac: du dépistage individuel au maintien de l'abstinence en premier recours. Haute Autorité de Santé. 2014. URL: https://www.has-sante.fr/jcms/c_1718021/fr/arret-de-la-consommation-de-tabac-du-depistage-individuel-au-maintien-de-l-abstinence-en-premier-recours [accessed 2019-09-11]
7. WHO launches Quit Tobacco App. World Health Organization. 2022 Feb 15. URL: https://www.who.int/southeastasia/news/detail/15-02-2022-who-launches-quit-tobacco-app [accessed 2023-05-24]
8. Cambon L, Bergman P, Le Faou A, Vincent I, Le Maitre B, Pasquereau A, et al. Study protocol for a pragmatic randomised controlled trial evaluating efficacy of a smoking cessation e-'Tabac Info Service': ee-TIS trial. BMJ Open 2017 Feb 24;7(2):e013604 [FREE Full text] [doi: 10.1136/bmjopen-2016-013604] [Medline: 28237958]
9. Brown J, Michie S, Geraghty AW, Yardley L, Gardner B, Shahab L, et al. Internet-based intervention for smoking cessation (StopAdvisor) in people with low and high socioeconomic status: a randomised controlled trial. Lancet Respir Med 2014 Dec;2(12):997-1006. [doi: 10.1016/s2213-2600(14)70195-x]

10. Bustamante LA, Gill Ménard C, Julien S, Romo L. Behavior change techniques in popular mobile apps for smoking cessation in France: content analysis. JMIR Mhealth Uhealth 2021 May 13;9(5):e26082 [FREE Full text] [doi: 10.2196/26082] [Medline: 33983130]
11. Bahadoor R, Alexandre JM, Fournet L, Gellé T, Serre F, Auriacombe M. Inventory and analysis of controlled trials of mobile phone applications targeting substance use disorders: a systematic review. Front Psychiatry 2021;12:622394 [FREE Full text] [doi: 10.3389/fpsyt.2021.622394] [Medline: 33692708]
12. Vilardaga R, Casellas-Pujol E, McClernon JF, Garrison KA. Mobile applications for the treatment of tobacco use and dependence. Curr Addict Rep 2019 Jun 9;6(2):86-97 [FREE Full text] [doi: 10.1007/s40429-019-00248-0] [Medline: 32010548]
13. Regmi K, Kassim N, Ahmad N, Tuah NA. Effectiveness of mobile apps for smoking cessation: a review. Tob Prev Cessat 2017 Apr 12;3(April):12 [FREE Full text] [doi: 10.18332/tpc/70088] [Medline: 32432186]
14. Whittaker R, McRobbie H, Bullen C, Rodgers A, Gu Y, Dobson R. Mobile phone text messaging and app-based interventions for smoking cessation. Cochrane Database Syst Rev 2019 Oct 22;10(10):CD006611 [FREE Full text] [doi: 10.1002/14651858.CD006611.pub5] [Medline: 31638271]
15. Auriacombe M, Fatséas M, Daulouède J, Tignol J. Le craving et nouvelle clinique de l'addiction: une perspective simplifiée et opérationnelle. Annales Médico-Psychologiques, Revue Psychiatrique 2018 Oct;176(8):746-749. [doi: 10.1016/j.amp.2018.08.014]
16. Hatsukami DK, Luo X, Jensen JA, al'Absi M, Allen SS, Carmella SG, et al. Effect of immediate vs gradual reduction in nicotine content of cigarettes on biomarkers of smoke exposure: a randomized clinical trial. JAMA 2018 Sep 04;320(9):880-891 [FREE Full text] [doi: 10.1001/jama.2018.11473] [Medline: 30193275]
17. Prochaska JO, Prochaska JM. Changing to Thrive: Using the Stages of Change to Overcome the Top Threats to Your Health and Happiness. Center City, MN: Hazelden Publishing; 2016.
18. Serre F, Fatseas M, Swendsen J, Auriacombe M. Ecological momentary assessment in the investigation of craving and substance use in daily life: a systematic review. Drug Alcohol Depend 2015 Mar 01;148:1-20. [doi: 10.1016/j.drugalcdep.2014.12.024] [Medline: 25637078]
19. Adherence to long-term therapies: evidence for action. World Health Organization. 2003. URL: https://iris.who.int/handle/10665/42682 [accessed 2021-08-28]
20. Belita E, Sidani S. Attrition in smoking cessation intervention studies: a systematic review. Can J Nurs Res 2015 Dec;47(4):21-40. [doi: 10.1177/084456211504700402] [Medline: 29509476]
21. Meyerowitz-Katz G, Ravi S, Arnolda L, Feng X, Maberly G, Astell-Burt T. Rates of attrition and dropout in app-based interventions for chronic disease: systematic review and meta-analysis. J Med Internet Res 2020 Sep 29;22(9):e20283 [FREE Full text] [doi: 10.2196/20283] [Medline: 32990635]
22. Ubhi HK, Michie S, Kotz D, Wong WC, West R. A mobile app to aid smoking cessation: preliminary evaluation of SmokeFree28. J Med Internet Res 2015 Jan 16;17(1):e17 [FREE Full text] [doi: 10.2196/jmir.3479] [Medline: 25596170]
23. Marler JD, Fujii CA, Utley DS, Tesfamariam LJ, Galanko JA, Patrick H. Initial assessment of a comprehensive digital smoking cessation program that incorporates a mobile app, breath sensor, and coaching: cohort study. JMIR Mhealth Uhealth 2019 Feb 04;7(2):e12609. [doi: 10.2196/12609] [Medline: 30670372]
24. Sobolev M, Anand A, Dziak JJ, Potter LN, Lam CY, Wetter DW, et al. Time-varying model of engagement with digital self reporting: evidence from smoking cessation longitudinal studies. Front Digit Health 2023 Apr 13;5:1144081 [FREE Full text] [doi: 10.3389/fdgth.2023.1144081] [Medline: 37122813]
25. Rajani NB, Bustamante L, Weth D, Romo L, Mastellos N, Filippidis FT. Engagement with gamification elements in a smoking cessation app and short-term smoking abstinence: quantitative assessment. JMIR Serious Games 2023 Feb 01;11:e39975 [FREE Full text] [doi: 10.2196/39975] [Medline: 36724003]
26. Alkhaldi G, Hamilton FL, Lau R, Webster R, Michie S, Murray E. The effectiveness of prompts to promote engagement with digital interventions: a systematic review. J Med Internet Res 2016 Jan 08;18(1):e6 [FREE Full text] [doi: 10.2196/jmir.4790] [Medline: 26747176]
27. Nahum-Shani I, Shaw SD, Carpenter SM, Murphy SA, Yoon C. Engagement in digital interventions. Am Psychol 2022 Oct;77(7):836-852 [FREE Full text] [doi: 10.1037/amp0000983] [Medline: 35298199]
28. Bricker JB, Mull KE, Santiago-Torres M, Miao Z, Perski O, Di C. Smoking cessation smartphone app use over time: predicting 12-month cessation outcomes in a 2-arm randomized trial. J Med Internet Res 2022 Aug 18;24(8):e39208 [FREE Full text] [doi: 10.2196/39208] [Medline: 35831180]
29. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
30. Moore TM, Seavey A, Ritter K, McNulty JK, Gordon KC, Stuart GL. Ecological momentary assessment of the effects of craving and affect on risk for relapse during substance abuse treatment. Psychol Addict Behav 2014 Jun;28(2):619-624. [doi: 10.1037/a0034127] [Medline: 24128286]
31. Bustamante L, Romo L. Use determinants of smoking cessation app. National Institutes of Health National Library of Medicine. 2022. URL: https://clinicaltrials.gov/study/NCT05318651 [accessed 2022-08-15]

32. Michie S, Hyder N, Walia A, West R. Development of a taxonomy of behaviour change techniques used in individual behavioural support for smoking cessation. Addict Behav 2011 Apr;36(4):315-319. [doi: 10.1016/j.addbeh.2010.11.016] [Medline: 21215528]
33. Michie S, Wood CE, Johnston M, Abraham C, Francis JJ, Hardeman W. Behaviour change techniques: the development and evaluation of a taxonomic method for reporting and describing behaviour change interventions (a suite of five studies involving consensus methods, randomised controlled trials and analysis of qualitative data). Health Technol Assess 2015 Nov;19(99):1-188 [FREE Full text] [doi: 10.3310/hta19990] [Medline: 26616119]
34. Pisinger C, Aadahl M, Toft U, Jørgensen T. Motives to quit smoking and reasons to relapse differ by socioeconomic status. Prev Med 2011 Jan;52(1):48-52. [doi: 10.1016/j.ypmed.2010.10.007] [Medline: 21047525]
35. McCaul KD, Hockemeyer JR, Johnson RJ, Zetocha K, Quinlan K, Glasgow RE. Motivation to quit using cigarettes: a review. Addict Behav 2006 Jan;31(1):42-56. [doi: 10.1016/j.addbeh.2005.04.004] [Medline: 15916861]
36. Gentry S, Craig J, Holland R, Notley C. Smoking cessation for substance misusers: a systematic review of qualitative studies on participant and provider beliefs and perceptions. Drug Alcohol Depend 2017 Nov 01;180:178-192. [doi: 10.1016/j.drugalcdep.2017.07.043] [Medline: 28910690]
37. Bustamante LA, Romo L. Validation of the French smoking cessation motivation scale with French smokers using a mobile app for smoking cessation. Eur J Investig Health Psychol Educ 2022 Aug 19;12(8):1179-1190 [FREE Full text] [doi: 10.3390/ejihpe12080082] [Medline: 36005231]
38. Rajani NB, Mastellos N, Filippidis FT. Impact of gamification on the self-efficacy and motivation to quit of smokers: observational study of two gamified smoking cessation mobile apps. JMIR Serious Games 2021 Apr 27;9(2):e27290 [FREE Full text] [doi: 10.2196/27290] [Medline: 33904824]
39. Affret A, Luc A, Baumann C, Bergman P, Le Faou AL, Pasquereau A, et al. Effectiveness of the e-Tabac Info Service application for smoking cessation: a pragmatic randomised controlled trial. BMJ Open 2020 Oct 27;10(10):e039515 [FREE Full text] [doi: 10.1136/bmjopen-2020-039515] [Medline: 33109670]
40. Marques MM, Wright AJ, Corker E, Johnston M, West R, Hastings J, et al. The behaviour change technique ontology: transforming the behaviour change technique taxonomy v1. Wellcome Open Res 2023 Jul 17;8:308 [FREE Full text] [doi: 10.12688/wellcomeopenres.19363.1] [Medline: 37593567]
41. Michie S, West R, Finnerty AN, Norris E, Wright AJ, Marques MM, et al. Representation of behaviour change interventions and their evaluation: development of the upper level of the behaviour change intervention ontology. Wellcome Open Res 2020;5:123 [FREE Full text] [doi: 10.12688/wellcomeopenres.15902.2] [Medline: 33614976]
42. Black N, Johnston M, Michie S, Hartmann-Boyce J, West R, Viechtbauer W, et al. Behaviour change techniques associated with smoking cessation in intervention and comparator groups of randomized controlled trials: a systematic review and meta-regression. Addiction 2020 Nov 16;115(11):2008-2020. [doi: 10.1111/add.15056] [Medline: 32196796]
43. Bricker JB, Mull KE, Kientz JA, Vilardaga R, Mercer LD, Akioka KJ, et al. Randomized, controlled pilot trial of a smartphone app for smoking cessation using acceptance and commitment therapy. Drug Alcohol Depend 2014 Oct 01;143:87-94 [FREE Full text] [doi: 10.1016/j.drugalcdep.2014.07.006] [Medline: 25085225]
44. Maglione MA, Maher AR, Ewing B, Colaiaco B, Newberry S, Kandrack R, et al. Efficacy of mindfulness meditation for smoking cessation: a systematic review and meta-analysis. Addict Behav 2017 Jun;69:27-34. [doi: 10.1016/j.addbeh.2017.01.022] [Medline: 28126511]
45. Rajani NB, Mastellos N, Filippidis FT. Self-efficacy and motivation to quit of smokers seeking to quit: quantitative assessment of smoking cessation mobile apps. JMIR Mhealth Uhealth 2021 Apr 30;9(4):e25030 [FREE Full text] [doi: 10.2196/25030] [Medline: 33929336]
46. McClure JB, Heffner JL, Krakauer C, Mun S, Klasnja P, Catz SL. Feasibility, acceptability, and potential impact of a novel mHealth app for smokers ambivalent about quitting: randomized pilot study. JMIR Mhealth Uhealth 2023 Jun 28;11:e46155 [FREE Full text] [doi: 10.2196/46155] [Medline: 37379059]
47. Houston TK, Chen J, Amante DJ, Blok AC, Nagawa CS, Wijesundara JG, et al. Effect of technology-assisted brief abstinence game on long-term smoking cessation in individuals not yet ready to quit: a randomized clinical trial. JAMA Intern Med 2022 Mar 01;182(3):303-312 [FREE Full text] [doi: 10.1001/jamainternmed.2021.7866] [Medline: 35072714]
48. Etter JF, Khazaal Y. The stop-tabac smartphone application for smoking cessation: a randomized controlled trial. Addiction 2022 May 19;117(5):1406-1415 [FREE Full text] [doi: 10.1111/add.15738] [Medline: 34738687]
49. Hock ES, Franklin M, Baxter S, Clowes M, Chilcott J, Gillespie D. Covariates of success in quitting smoking: a systematic review of studies from 2008 to 2021 conducted to inform the statistical analyses of quitting outcomes of a hospital-based tobacco dependence treatment service in the United Kingdom. NIHR Open Res 2023 May 19;3:28. [doi: 10.3310/nihropenres.13427.1]
50. Ubhi HK, Kotz D, Michie S, van Schayck OC, West R. A comparison of the characteristics of iOS and Android users of a smoking cessation app. Transl Behav Med 2017 Jun;7(2):166-171 [FREE Full text] [doi: 10.1007/s13142-016-0455-z] [Medline: 28168609]
51. Smartphones: so many apps, so much time. The Nielsen Company. 2014 Jul. URL: https://www.nielsen.com/insights/2014/smartphones-so-many-apps--so-much-time/ [accessed 2020-10-02]

Abbreviations

9s-Kwit's program: Kwit's 9-step preparation program for smoking cessation
CBT: cognitive behavioral therapy
F-SCMS: French Smoking Cessation Motivation Scale
mHealth: mobile health
VAS: visual analog scale
WAR: Willingness, Ability, and Readiness to Quit
WHO: World Health Organization

Edited by L Buis; submitted 19.07.23; peer-reviewed by G Dolatkhah, B Lignier; comments to author 12.01.24; revised version received 19.02.24; accepted 11.04.24; published 02.10.24.

Please cite as:
Bustamante Perez LA, Romo L, Zerhouni O
Feasibility and Engagement of a Mobile App Preparation Program (Kwit) for Smoking Cessation in an Ecological Context: Quantitative Study
JMIR Mhealth Uhealth 2024;12:e51025
URL: https://mhealth.jmir.org/2024/1/e51025
doi:10.2196/51025
PMID:39357053

Original Paper

Influencing Factors and Implementation Pathways of Adherence Behavior in Intelligent Personalized Exercise Prescription: Qualitative Study

Xuejie Xu[1], BSN; Guoli Zhang[1], BSN; Yuxin Xia[1], BSN; Hui Xie[1], PhD; Zenghui Ding[2], PhD; Hongyu Wang[3], PhD; Zuchang Ma[2], PhD; Ting Sun[1,2], MPhil

[1]School of Nursing, Bengbu Medical University, Bengbu, China

[2]Institute of Intelligent Machines, Hefei Institutes of Physical Sciences, Chinese Academy of Sciences, Hefei, China

[3]Department of Physical Education and Arts, Bengbu Medical University, Bengbu, China

Corresponding Author:
Ting Sun, MPhil
School of Nursing
Bengbu Medical University
Number 2600, Donghai Avenue
Bengbu, 233030
China
Phone: 86 18005529759
Email: cangwuge21@163.com

Abstract

Background: Personalized intelligent exercise prescriptions have demonstrated significant benefits in increasing physical activity and improving individual health. However, the health benefits of these prescriptions depend on long-term adherence. Therefore, it is essential to analyze the factors influencing adherence to personalized intelligent exercise prescriptions and explore the intrinsic relationship between individual behavioral motivation and adherence. This understanding can help improve adherence and maximize the effectiveness of such prescriptions.

Objective: This study aims to identify the factors influencing adherence behavior among middle-aged and older community residents who have been prescribed personalized exercise regimens through an electronic health promotion system. It also explores how these factors affect the initiation and maintenance of adherence behavior.

Methods: We used purposive sampling to conduct individual, face-to-face semistructured interviews based on the Transtheoretical Model (TTM) with 12 middle-aged and older community residents who had been following personalized exercise regimens for 8 months. These residents had received detailed exercise health education and guidance from staff. The interviews were recorded, transcribed verbatim, and analyzed using NVivo software through grounded theory. We then applied the TTM and multibehavioral motivation theory to analyze the factors influencing adherence. Additionally, the relationship between behavioral motivations and adherence was explored.

Results: Using the behavior change stages of the TTM, open coding yielded 21 initial categories, which were then organized into 8 main categories through axial coding: intrinsic motivation, extrinsic motivation, benefit motivation, pleasure motivation, achievement motivation, perceived barriers, self-regulation, and optimization strategies. Selective coding further condensed these 8 main categories into 3 core categories: “multitheory motivation,” “obstacle factors,” and “solution strategies.” Using the coding results, a 3-level model of factors influencing adherence to intelligent personalized exercise prescriptions was developed. Based on this, an implementation path for promoting adherence to intelligent personalized exercise prescriptions was proposed by integrating the model with the TTM.

Conclusions: Adherence to personalized exercise prescriptions is influenced by both facilitating factors (eg, multibehavioral motivation, optimization strategies) and obstructive factors (eg, perceived barriers). Achieving and maintaining adherence is a gradual process, shaped by a range of motivations and factors. Personalized solutions, long-term support, feedback mechanisms, and social support networks are essential for promoting adherence. Future efforts should focus on enhancing adherence by strengthening multibehavioral motivation, optimizing solutions, and addressing barriers to improve overall adherence.

(JMIR Mhealth Uhealth 2024;12:e59610) doi:10.2196/59610

KEYWORDS

exercise prescription; adherence behavior; qualitative; influence factors; Transtheoretical Model; multiple motivations of behavior

Introduction

Mounting evidence suggests that regular moderate physical activity significantly improves health-related quality of life, promotes community engagement and healthy aging, and may enhance cognition and mental health [1,2]. It also appears to mitigate the risk of at least 35 chronic diseases and reduce mortality [1,3-6]. Physical activity is currently recognized as a sustainable approach to promoting individual and community health and well-being [7]. Exercise prescription [8], a structured program designed to guide exercise for health promotion and the prevention and management of chronic diseases, has been applied to diverse populations, including healthy individuals, those with sports injuries, perioperative patients, individuals with chronic conditions, and those with disabilities [9-11]. However, many physicians struggle to provide physical activity guidance and prescribe scientifically effective exercise regimens due to a lack of knowledge, training, awareness, and understanding of exercise rehabilitation [3]. These challenges, combined with the absence of standardized implementation guidelines [12] and other barriers, hinder the promotion and integration of exercise prescriptions.

Most exercise guidelines [13-16] strongly recommend personalized and well-defined exercise prescriptions that specify the mode, intensity, frequency, and duration of exercise. Traditional exercise prescriptions often struggle to achieve scientific personalization, prompting the development of intelligent, personalized alternatives. By incorporating advanced technologies such as artificial intelligence and big data analysis, these modern approaches provide individuals with more precise, scientifically grounded, and customized exercise plans. For example, Netz et al [17,18] developed an innovative tool for remotely assessing balance, strength, and flexibility in middle-aged and older individuals, providing personalized exercise plans via smartphones. Lin et al [19] and Sun et al [20] created and validated distinct systems: a force gauge system that integrates exercise games with the Internet of Things and a cloud-based intelligent personalized exercise prescription system. Both systems are designed to deliver tailored exercise prescriptions for middle-aged and older individuals. Additionally, a multimodal data-driven artificial intelligence system has been developed and validated to generate personalized exercise prescriptions for patients with mental disorders [21]. These intelligent tools and systems effectively address the limitations of conventional exercise prescription methods. Beyond significantly improving exercise outcomes, reducing exercise-related risks, and enhancing individual health and quality of life, they also alleviate the burden caused by the shortage of professional sports rehabilitation resources, thereby improving the quality and accessibility of sports rehabilitation services [19-22]. Most research on intelligent personalized exercise prescriptions has focused on system design and development [23] and efficacy validation [17]. However, studies examining the factors that influence individual adherence behaviors and their underlying pathways are currently lacking. Recent systematic reviews and meta-analyses [24-26] suggest that while smart health interventions hold promise for improving exercise adherence, the specific factors and pathways driving adherence have not been fully explored or systematically explained. This gap makes it challenging for individuals to maintain long-term adherence. Therefore, it is crucial to further investigate and refine the factors, pathways, and theoretical frameworks related to adherence to personalized smart exercise prescriptions.

Motivation is a critical determinant in shaping individuals' behavioral intentions and driving behavioral changes, with various motivational factors regulating and influencing individual behavior [27-29]. Although motivation significantly impacts adherence behavior, theoretical limitations, a lack of empirical research, and the complexity of individual motivations contribute to a theoretical "black box" regarding how these motivations influence adherence to exercise prescriptions. The relationship between behavioral motivation and individual adherence behavior remains underexplored [30-32]. In health promotion, several theoretical models [33,34] have been used to study individual behavior. Among these, the Transtheoretical Model (TTM) has been widely applied in research on digital health behavior and behavior change [35-37]. By dividing behavior change into 5 stages—precontemplation, contemplation, preparation, action, and maintenance—the TTM provides a detailed framework for understanding the mechanisms and pathways of individual behavior change [38]. The TTM has shown promising results in various social studies focused on health behavior change and health promotion [39-41].

Therefore, this study uses qualitative methods to explore the factors influencing the adherence behavior of community-dwelling middle-aged and older residents who have followed personalized exercise prescriptions issued through an electronic health promotion system for 8 months. By integrating the TTM, the study analyzes behavioral intentions and changes at different stages of behavior change, identifies key factors in the adherence behavior change process, and constructs an implementation path model for adherence to intelligent personalized exercise prescriptions, driven by behavioral motivations and other influencing factors. This study reveals how behavioral motivations drive and sustain adherence to exercise prescriptions, expands the application of the TTM in digital health services, and provides insights for future research on adherence to intelligent personalized exercise prescriptions. Additionally, it offers practical strategies for enhancing and maintaining long-term adherence to these prescriptions. To our knowledge, this is the first qualitative study on adherence behavior to intelligent personalized exercise prescriptions.

Methods

Design

This study is part of a longitudinal research project aimed at examining the health impacts of intelligent personalized exercise

prescriptions on community-dwelling middle-aged and older residents. It uses a descriptive exploratory qualitative design, based on face-to-face semistructured interviews. The study strictly followed the COREQ (Consolidated Criteria for Reporting Qualitative Research) guidelines [42].

Participants

The participants in this study are a subsample from a longitudinal research project conducted by the research team, which aims to investigate the long-term effects of intelligent personalized exercise prescriptions on the health of middle-aged and older individuals. The inclusion criteria for the longitudinal study were age 50 years or older; no severe physical diseases or related complications; no cognitive or mental disorders among community-dwelling middle-aged and older residents; and exclusion of individuals with severe cardiovascular, pulmonary, or renal diseases, severe diabetes or related complications, fasting blood glucose of 13.3 mmol/L or higher with positive urine ketones, postprandial blood glucose of 19.4 mmol/L or higher, resting blood pressure of 180/110 mmHg or higher, or severe cognitive or mental disorders. Participants were initially recruited from the community via telephone and verbal invitations.

To investigate adherence behavior and the factors influencing adherence to intelligent personalized exercise prescriptions, we used purposive sampling to select participants for face-to-face semistructured interviews. The inclusion criteria were (1) community-dwelling middle-aged and older residents who had undergone home-based health checkups between 2021 and 2022 and received personalized exercise prescriptions through an electronic health promotion system administered by community staff; (2) participants who were provided with printed exercise materials and received detailed explanations and guidance from community staff; and (3) individuals whose exercise prescriptions had been active for at least 8 months. To gain a comprehensive understanding of exercise prescription adherence, we specifically recruited participants with neutral or negative attitudes toward intelligent personalized exercise prescriptions, as well as those who self-reported low adherence. The sample size was determined based on theoretical saturation, defined as the point at which no new issues or insights emerge and all relevant conceptual categories have been identified and explored [43]. Throughout the research process, the team continuously analyzed the dialogues and assessed the saturation level of the interview data at each stage. After the 12th interview, no new information emerged, all major concepts and categories had been thoroughly identified and explored, and theoretical saturation was reached. As a result, 12 eligible community-dwelling middle-aged and older residents participated in the face-to-face semistructured interviews.

Ethics Considerations

This study received ethical approval from the Human Research Ethics Committee of Bengbu Medical University (approval number 2022-103).

Intelligent Personalized Exercise Prescription Program

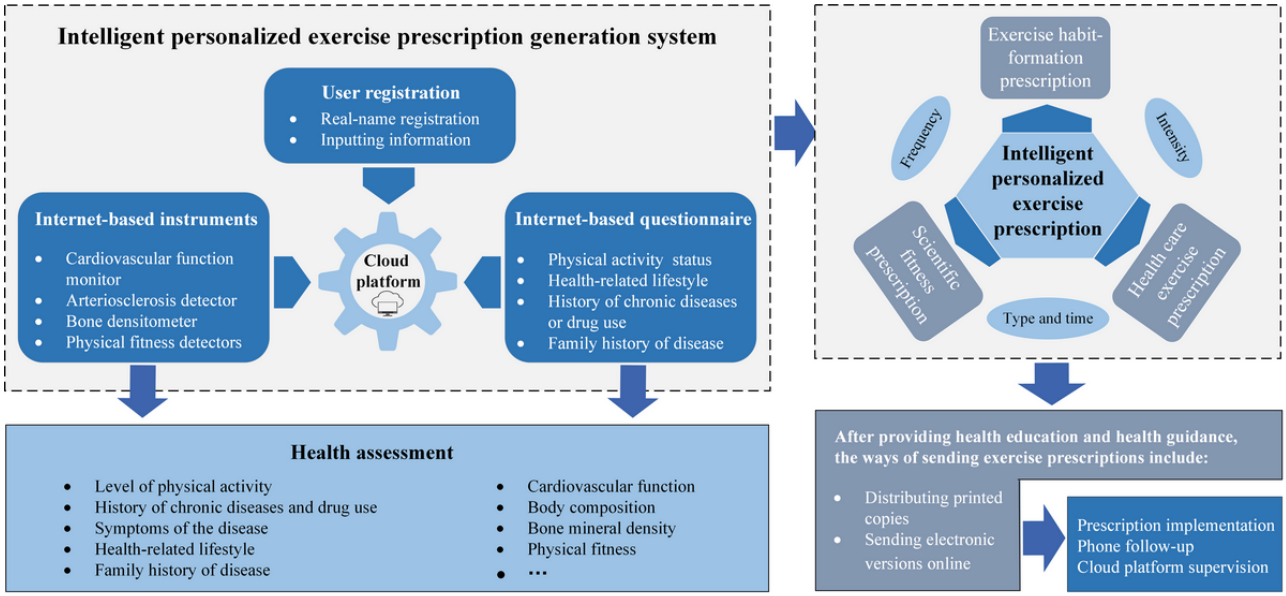
Details of the intelligent personalized exercise prescription system have been previously described [20]. The system generally consists of 4 components: user registration and information input, internet-based health monitoring devices, online questionnaire surveys, and a cloud platform.

Before generating a personalized exercise prescription, participants must register with their real names and provide personal information, including basic details (age, gender, height, and weight) and health history (past medical history and exercise habits). Next, participants' health data are collected using internet-based health monitoring devices. Additionally, participants must complete a series of online health questionnaires to provide the necessary data for generating personalized exercise prescriptions. Finally, all collected health data are uploaded to a cloud platform. The platform uses advanced data analysis techniques and artificial intelligence algorithms to evaluate participants' health data and generate personalized exercise prescriptions tailored to each individual.

Once the personalized exercise prescription is generated, community health service staff will provide face-to-face health education and guidance based on it. Depending on participants' needs, the prescription can be delivered in either electronic or paper form, promoting better adherence.

After the exercise prescription is implemented, community health service staff can access user information via the cloud platform and conduct regular follow-up calls. These calls aim to remind participants to adhere to the exercise prescription, answer their questions, and provide necessary support and guidance. Figure 1 illustrates the structure of the intelligent personalized exercise prescription system and the process of generating and applying the prescriptions.

Figure 1. Generation and application of intelligent personalized exercise prescription.



Procedures and Data Collection

This study strictly adheres to the COREQ [42]. To gather detailed data and gain a deeper understanding of participants' subjective experiences and true feelings, semistructured interviews were conducted. All interviews were carried out by 3 trained researchers, including the first author (XX), a graduate nursing student, and an undergraduate nursing student, between February and March 2023. Following purposive sampling to select interviewees, rapport was established between the interviewers and participants. Interviews were conducted in quiet, private spaces to minimize interruptions. Before the interviews, participants were briefed on the research objectives, methods, expected duration, and confidentiality principles. Informed consent was obtained, and participants signed consent forms. The interviews were audio-recorded by the interviewers, with additional on-site notes taken for further analysis. Additionally, interviewers observed and promptly documented participants' nonverbal behaviors, including tone, gestures, and facial expressions, during the interviews. Each interview lasted 20-30 minutes. Within 24 hours of completion, the recordings were transcribed into written transcripts and verified by the participants for content accuracy.

According to the main principles of the TTM, the occurrence and maintenance of individual behavior change occur across multiple stages. Therefore, before conducting the semistructured interviews, we determined the landmark events for each stage—"intention," "preparation," "action," and "maintenance"—based on the research objectives and the definitions and characteristics of these stages in the TTM. Relevant questions about stage transitions were formulated in the interview outline to ensure the interviews were systematic and scientifically rigorous. The interview outline included questions about the respondents' basic information; their perceptions and attitudes toward the intelligent personalized exercise prescription; factors that promote or hinder progression through specific stages of behavior change; and their activities, emotional states, needs, and suggestions within each stage. To ensure confidentiality, all participant information and data were anonymized and deidentified to prevent personal identification. Recordings and transcripts were stored in password-protected, secure files on the university server. The interviews concluded when no new information emerged.

Analysis

After the qualitative interviews were completed, the initial transcripts of the interview recordings—aligned with the interview questions corresponding to the 4 behavior change stages ("intention stage," "preparation stage," "action stage," and "maintenance stage") in the TTM—were imported into the qualitative analysis software NVivo 12 (Lumivero). The transcripts were then subjected to a 3-level coding process (Textbox 1) based on grounded theory [44,45] within the framework of the 4 behavior change stages.

Drawing on the TTM and the coding results, we analyzed the factors influencing the adherence behavior to intelligent personalized exercise prescriptions, proposed hypotheses for the pathways to achieving adherence, and developed an implementation path model for adherence to intelligent personalized exercise prescriptions. We validated the hypotheses through follow-up interviews with the 12 respondents and refined the theoretical model based on team discussions, ultimately finalizing it.

For clearer presentation and data referencing, the 12 participants were labeled P1 through P12. Data analysis and organization were primarily carried out by 5 researchers; 3 researchers (XX, GZ, and YX) sorted and classified the interview recordings for open, axial, and selective coding, while the remaining 2 (TS and HX) verified the phased collation and ensured accuracy and consistency in the 3-level coding process. The hypotheses and model framework for adherence behavior pathways to intelligent personalized exercise prescriptions were formulated by the research team, validated through follow-up interviews conducted by 2 researchers (XX and YX), and refined through group discussions.

Textbox 1. The 3-level coding process.

1. Open coding

This is the most basic level of abstraction in the coding process. Each interview transcript was read thoroughly, and the text was coded as it appeared, without unnecessary abstraction. By analyzing the text line by line, we identified and marked concepts and phenomena relevant to the research questions, ensuring that the coding remained original and authentic.

2. Axial coding

Building on open coding, we focused on the logical connections and relationships between the identified concepts. Through ongoing comparison and analysis, we refined these into higher-level main categories.

3. Selective coding

We further examined the relationships between the main categories, identifying a core category that served as the central focus of the coding system. This core category guided the analysis and discussion of the other categories.

Results

Overview

Effective interviews were conducted with 12 participants, and their general information is summarized in Table 1. The results of the 3-level coding process based on the 4 stages of behavior change in the TTM are as follows: (1) Open coding involves breaking down, reorganizing, and conceptualizing the transcribed interview materials. This process led to the identification of 21 categories (see Table 2), which highlight the direct factors influencing adherence to intelligent personalized exercise prescriptions. (2) In the axial coding stage, we examined the relationships and logical connections among the 21 categories identified during open coding. This led to the formation of 8 main categories (Table 3). (3) During selective coding, we integrated and refined the existing categories to extract a "core category." The categories of "hedonic motivation," "benefit motivation," "intrinsic motivation," "extrinsic motivation," and "achievement motivation" were consolidated into a single category, "multiple motivations of behavior." Similarly, "self-regulation" and "optimization solutions" were grouped under "problem-solving strategies," while "perceptual barriers" were categorized as "obstacle factors" (see Table 4). The open coding derived from the original interview data represents direct experiences and situational factors, forming the first level (bottom level). The 21 categories identified during open coding were organized into axial coding, which represents the preliminary classification of motivation and obstacle factors, making it the second level (middle level). Selective coding, which represents the core influencing factors derived from the systematic analysis of the main categories, forms the third level (top level). In summary, we constructed a 3-level model of factors influencing the adherence behavior to intelligent personalized exercise prescriptions, as illustrated in Figure 2.

Based on the TTM, Multiple Behavioral Motivation Theory, and the 3-level coding results from interviews on the 4 behavior change stages, we propose the following hypotheses for the path to adherence to intelligent personalized exercise prescriptions:

- Hypothesis 1: An individual's intrinsic motivations (such as disease susceptibility, perceived threat, and a sense of family responsibility) and extrinsic motivations (including family support, doctor's advice, recommendations from friends, and media publicity) contribute to the intention to adhere, leading to the "adherence intention stage."
- Hypothesis 2: Achievement motivation encourages the individual to reach the "adherence preparation stage," where they prepare for the actual adherence behavior to the exercise prescription.
- Hypothesis 3: Hedonic and benefit motivations drive the individual to enter the "adherence behavior stage," resulting in the initiation of actual adherence behaviors.
- Hypothesis 4: The presence of multiple motivations, coupled with problem-solving strategies, leads the individual to the "adherence maintenance stage," where they progress from partial to full adherence, achieving complete and sustained adherence to the exercise prescription.

The path hypotheses were verified and validated by 2 researchers (XX and YX) through follow-up interviews with the 12 respondents. Based on these findings, a model for the realization path of adherence behavior to intelligent personalized exercise prescriptions was further developed. This model consists of 4 stages of adherence: intention, preparation, action, and maintenance. In the intention stage, individuals are motivated by intrinsic and extrinsic factors, developing the intention to adhere, but without engaging in actual behavior. Driven by achievement motivation, they progress to the preparation stage, where they ready themselves for action and set goal expectations, though behavior has yet to occur. During the action stage, hedonic and benefit motivations play a key role in prompting individuals to take action and experience benefits, though full adherence may not yet be achieved. Finally, in the maintenance stage, supported by multiple motivations and problem-solving strategies, individuals transition from general adherents to complete adherents, fully adhering to the exercise prescription, as illustrated in Figure 3.

Table 1. General information on interviewees.

Number	Age (years)	Occupation	Gender	Educational level	Presence of chronic diseases	Type of exercise prescription	Previous exercise habits	Previous success or failure in health behavior change (eg, smoking cessation or weight loss)
P1	66	Shop assistant	Female	Junior high school	Yes	Scientific fitness mode	Yes	No
P2	71	Public official	Female	Junior high school	Yes	Scientific fitness mode	Yes	Weight loss success
P3	68	Teacher	Female	Vocational school	No	Health care mode	No	No
P4	72	Laborer	Male	Primary school	Yes	Scientific fitness mode	Yes	No
P5	63	Laborer	Female	Primary school	No	Health care mode	No	No
P6	68	Staff	Male	High school	No	Exercise habit–formation mode	No	Smoking and alcohol cessation success
P7	76	Staff	Male	Junior high school	No	Exercise habit–formation mode	No	No
P8	57	Pharmacy clerk	Female	High school	Yes	Scientific fitness mode	Yes	No
P9	60	Staff	Female	Junior high school	No	Exercise habit–formation mode	No	No
P10	69	Staff	Male	Junior high school	No	Health care mode	No	No
P11	72	Laborer	Female	High school	Yes	Scientific fitness mode	Yes	No
P12	68	Laborer	Male	High school	No	Health care mode	Yes	No

Table 2. Categories formed by open coding of interview content.

Behavior change stage, category, and initial concept	Representative original sentences
Intention stage	
Disease threat	
Encountering problems	*My cervical vertebra has some issues. Looking at my phone too much, due to my age, so I exercise to relieve the discomfort in my cervical vertebra.* [P5]
Worsening of symptoms	*My heart wasn't very good before. The doctor told me to try to increase the vitality of my heart. The lazier you are, the worse your heart condition gets.* [P7]
Susceptibility to disease	
Frailty due to aging	*As I've gotten older, my health has declined.* [P8]
Decline in function	*Now our physical functions are not as good as those of younger people. If we don't exercise more, it won't work.* [P8]
Hazards of prolonged sitting	*If you don't exercise and just lie down for a long time without moving, you'll definitely have health problems later on.* [P4]
Sense of family responsibility	
Reducing children's worries	*Although my children are grown up, I am still their support. If I am healthy, they will worry less.* [P9]
Promoting family health	*If I learn scientific exercise methods, I want to exercise with my family. Won't our family all be healthy then?* [P10]
Family support	
Support from children	*My children also strongly support me participating in this.* [P5]
Companionship of spouse	*Usually, my spouse and I play badminton together.* [P1]
Doctor's advice	
Doctor's notification	*The doctor told me to try to increase the vitality of my heart.* [P7]
Friend's recommendation	
Recommending participation	*My friend Yue ** joined first and spoke highly of your place, so she recommended me to join personalized exercises here.* [P6]
Media publicity	
Broadcasting publicity	*Initially, it was mainly through some radio broadcasts that I learned about the benefits of exercise.* [P5]
Preparation stage	
Technological awareness	
Scientifically expectations	*Your personalized exercise program must be well-organized and scientifically based.* [P1]
Organizing positive	*This exercise prescription must have benefits for the exercise of a specific body structure or physical fitness.* [P1]
Authoritative projects	*According to my daughter-in-law, she said, "Mom, there is a health program here in collaboration with a medical college and the University of Science and Technology of China. They provide free health checkups and tailor personalized exercise prescription for you. They also guide how to exercise. They are here to promote health. You should go and take a look. It's much better than those deceptive health products.*" [P2]
Exercise preparation	
Preparing sports equipment	*In order to adhere to the exercise prescription properly, I even prepared specific exercise clothing and shoes.* [P5]
Peer support	*At the beginning, I arranged to exercise together with a few people.* [P1]
Action stage	
Service hedonism	
Problem-solving	*It's great that you can explain any questions we don't understand so well.* [P3]
Thorough guidance	*You health providers are so thoughtful in explaining the examination reports and guiding us in exercise.* [P5]

Behavior change stage, category, and initial concept	Representative original sentences
Careful inspection	*Your staff are very meticulous in every examination, both before and after exercise.* [P4]
Professionalism of the service provider	*You have a lot of knowledge and explain things very well...* [P12]
Comprehensive reporting	*The reports you provide are very clear and comprehensive.* [P8]
Social hedonism	
Making friends	*During exercise, you can also make many friends (especially outdoors).* [P6]
Physiological benefits	
Symptom control	*Before, I had high blood sugar, but now my blood sugar is also very good.* [P6] *I used to have constipation, and it's still there, but the frequency has decreased...* [P5]
Physical improvement	*There has been an improvement in my body. I used to have high blood pressure and unstable blood pressure, but now it's very stable.* [P2]
Increasing physical activity	*Look, I can raise my arms very high now (demonstrating a posture).* [P2]
Psychological benefits	
Feeling relaxed	*After exercising, I feel my mood is more relaxed, my body feels better, and I have more energy than before.* [P3]
Energetic	*Now, both mentally and energetically, I am better than before.* [P2]
Experiencing well-being	*(After being guided on walking posture) When the heels touch the ground, I involuntarily straighten up, walking with chest out and stomach in, it feels particularly good, this kind of feeling...* [P6]
Personal benefits	
Acquiring knowledge	*Compared to before, I have learned how to exercise with proper posture, how to relax after exercising, and how to stretch...* [P2]
Free medical checkups	*They also conducted physical examinations for us, checking our bones and bone density, teaching us how to improve. They told me that my bone density had reached the bottom line and couldn't decline further, advising us to drink milk and eat eggs daily, as well as recommending specific vegetables. They explained it very thoroughly.* [P4]
Maintenance stage	
Self-management	
Perseverance	*I can persevere because that's my personality. Once I set my mind on something, I persist until I succeed, no matter what obstacles I face.* [P6]
Forming habits	*To develop a good exercise habit, I have to wake up at a certain time every day.* [P2]
Strict self-discipline	*I must be strict with myself and not make excuses like "it's cold today" or "it's windy" to forgive myself. I need to ensure that I accomplish what needs to be done.* [P2]
Flexibility adjustment	
Persisting to completion	*If I don't meet the planned exercise volume today, I'll make up for it tomorrow by sticking to the previous day's shortfall.* [P6]
Adapting flexibly	*I follow the exercise routine as advised by you. On rainy days, I use the treadmill at home, ensuring I meet the required exercise volume while exercising in a way I enjoy.* [P4]
Fulfillment of needs	
Adding videos	*It would be great if it could be turned into a video. The report is easy to misplace, and it's easier not to forget with a video, especially at our age...*[P1]
Enriching images	*I have a small suggestion. There are too few exercise images in our medical reports. It would be nice if they could be more varied. Otherwise, everything else is quite good.* [P10]
Positive feedback on results	
Good results upon retesting	*After exercising and reassessing my health, I found that I'm doing well. My grip strength is still good, and I'm almost reaching eighty kilograms. I'm still going strong. Hehehe (laughs happily).* [P4]
Ignition of exercise enthusiasm	
Eagerness to exercise	*After exercising for a while, I feel very enthusiastic about it, as if I'm eager to participate in sports from the bottom of my heart.* [P5]

Behavior change stage, category, and initial concept	Representative original sentences
Exercising voluntarily	*Having a workout plan makes me feel psychologically engaged. When it's time, I feel like I should exercise, and there's a sense of awareness, a kind of concern, that makes me want to work out.* [P6]
Pan-stage factors	
Personal limitations	
Struggling to persist	*Sometimes, doing exercise alone is difficult, so doing group activities helps.* [P7]
Lack of exercise interest	*However, later on, my legs felt better. Originally, I didn't like exercising and didn't do much exercise.* [P3]
Physical discomfort	*Because of my bad back, my legs hurt when I walk. So, I'm not good at walking and don't exercise much.* [P3]
Bodily fatigue	*Sometimes when I feel tired, I don't feel like exercising.* [P1]
Prone to forgetfulness	*I can't remember the exercise prescription because I'm old and forgetful. I used to stick to it at the beginning, but later, I did less, and I always forget the movements.* [P8]
External constraints	
Lack of resources	*There are no good places near my home. There is a small Friendship Square, but there is no track or facilities.* [P2]
Unable to free oneself	*The children need to come here for meals, buy groceries in the morning, and cook. It's usually fine, but sometimes it's overwhelming when I'm busy.* [P3]

Table 3. Main categories formed by interview content axial coding.

Behavior change stage, main category, and category	Definition
Intention stage	
Intrinsic motivation	
Disease threat	Perception of adverse health consequences resulting from one's current condition.
Susceptibility to disease	Adverse physical conditions, declining quality, and unhealthy lifestyle habits are more likely to trigger diseases.
Sense of family responsibility	Reduce the burden on the family and family members' worries and promote the health, harmony, and stable development of the family.
Extrinsic motivation	
Family support	Material or emotional support provided by family members.
Doctor's advice	Reasonable advice from doctors based on individual conditions to improve individual situations.
Friend's recommendation	Verbal introductions from friends that are hoped to be accepted and adopted.
Media publicity	Dissemination and transmission of information through information dissemination tools.
Preparation stage	
Achievement motivation	
Technological awareness	Understanding and familiarizing oneself with the intelligent personalized exercise prescription system.
Exercise preparation	Planning and preparing for upcoming exercise activities.
Action stage	
Hedonic motivation	
Service hedonism	Enjoy and experience high-quality, professional exercise testing and guidance services.
Social hedonism	Expand social connections, make friends, and enjoy the pleasure and benefits of social interaction.
Benefit motivation	
Physiological benefits	Control of adverse physiological symptoms of diseases and improvement in physical function or capability.
Psychological benefits	Improved mental state and psychological well-being, as well as increased energy.
Personal benefits	Personal access to free benefits, such as exercise and health knowledge, physical fitness, and health assessments.
Maintenance stage	
Self-regulation	
Self-management	Self-control and using inner strength for behavior change or habit formation.
Flexibility adjustment	Flexibly adjust according to individual circumstances within a certain range.
Optimization solutions	
Fulfilling needs	Fulfill the requirements arising from individual needs.
Positive feedback on results	Following the positive feedback from intelligent personalized exercise prescription to encourage better adherence.
Ignition of exercise enthusiasm	Stimulating or enhancing interest and enthusiasm for exercise through various means.
Pan-stage factors	
Perceptual barriers	
Personal limitation	Limitations imposed by individual conditions on accomplishing tasks.
External constraints	Limitations imposed by external factors beyond the individual on accomplishing tasks.

Table 4. Core categories formed by interview content selective coding.

Core category and main category	Definition
Multiple behavioral motivations	
Hedonic motivation	Behavioral motivation generated from the pursuit of the pleasant feelings brought by exercise.
Benefit motivation	Behavioral motivation arising from the actual benefits that can be obtained by following the exercise prescription.
Intrinsic motivation	Behavioral motivation originating from the individual's internal needs and perceptions.
Extrinsic motivation	Behavioral motivation resulting from the support and incentives of external factors.
Achievement motivation	Behavioral motivation arising from the expectation of achieving good performance or reaching specific goals by adhering to the exercise prescription.
Problem-solving strategies	
Self-regulation	Self-management and adjustment of an individual's behavior or thinking for better compliance with the exercise prescription.
Optimization solutions	Improvement measures and solutions for enhancing the compliance effect and adherence to the exercise prescription.
Obstacle factors	
Perceptual barriers	Difficulties and obstacles that an individual subjectively perceives as restricting their exercise in accordance with the exercise prescription.

Figure 2. The 3-level model of factors influencing the adherence behavior to intelligent personalized exercise prescription.

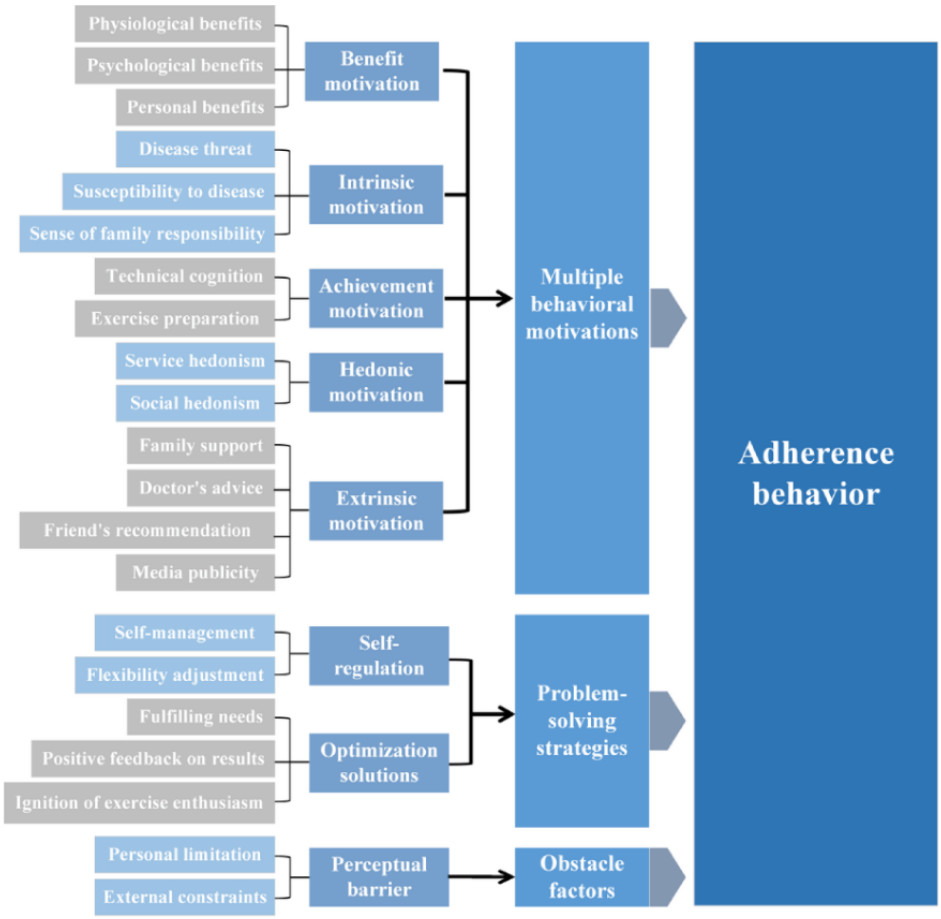
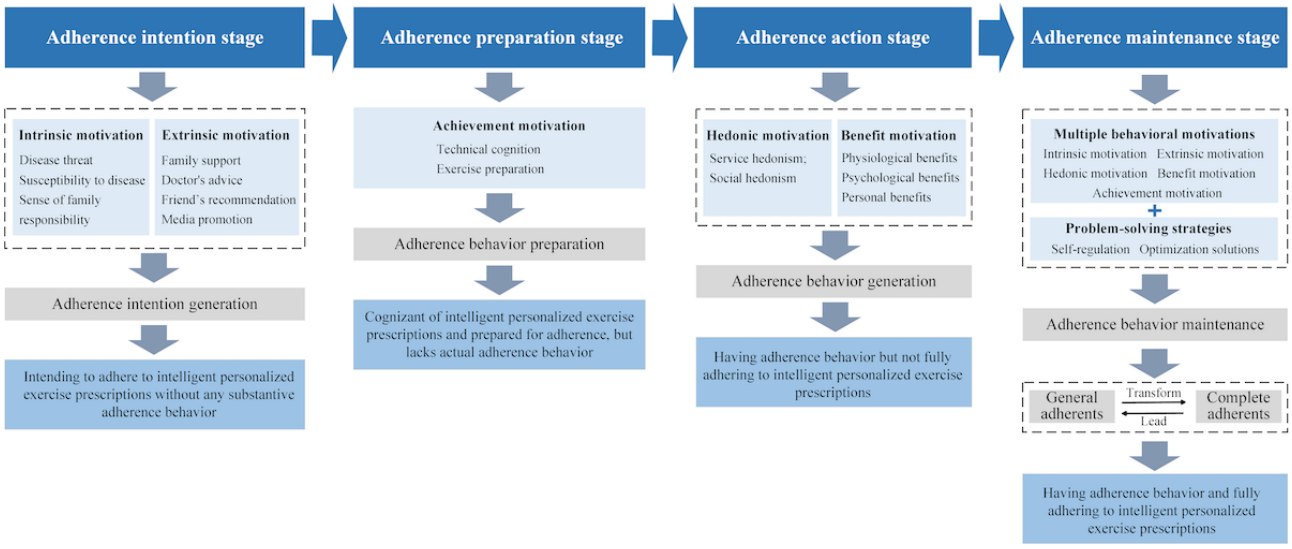


Figure 3. The implementation path model of adherence behavior to intelligent personalized exercise prescription based on the Transtheoretical Model and Multiple Behavioral Motivation Theory.



Factors Influencing Adherence Behavior With Intelligent Personalized Exercise Prescription and Pathways for Implementation

Multiple Behavioral Motivational Factors

Motivation is the central factor driving behavior change and serves as the intrinsic force behind individuals' actions [46,47]. This study explores how various motivations—hedonic, benefit, intrinsic, extrinsic, and achievement—affect participants' adherence to intelligent personalized exercise prescriptions. These motivations directly influence adherence behaviors [48]. (1) Intrinsic motivations (eg, perceived disease threat and susceptibility, sense of family responsibility) and extrinsic motivations (eg, family support, doctor recommendations, friend referrals, and media publicity) drive participants to develop an intention to adhere to the intelligent personalized exercise prescription. This phase, referred to as the "adherence intention stage," involves participants forming the intention to adhere, but without yet engaging in actual adherence behaviors. (2) After forming the intention to adhere, participants are driven by achievement motivation, aiming to reach self-set goals through adherence to the exercise prescription. At this stage, they begin preparing for adherence by seeking information about the exercise prescription and acquiring the necessary equipment. This phase, known as the "adherence preparation stage," builds upon the previous stage but does not yet involve actual adherence behavior. (3) When participants have a strong intention to adhere and are well-prepared, they enter the "adherence action stage" [49], where actual adherence behavior takes place. During this stage, high-quality health guidance, positive service experiences, and expanded social connections foster hedonic motivation. Additionally, the physiological and psychological benefits of adherence, increased health knowledge, and free health checkups provided by health care providers contribute to benefit motivation. Furthermore, hedonic and benefit motivations play a key role in sustaining and further developing adherence behavior during this phase. Therefore, the various motivations influencing participants in the context of intelligent personalized exercise prescriptions are crucial in the first 3 stages and serve as the primary drivers of adherence behavior.

Obstacles

"Perceptual barriers," or obstacles, are critical factors affecting adherence behavior. These barriers can undermine an individual's confidence and motivation, thereby hindering the sustainability of adherence. They negatively impact both the initiation and maintenance of adherence [50]. The findings by Kilgour et al [51] suggest that these obstacles stem from both individual limitations and external constraints. Individual limitations include disliking exercise, difficulty adhering due to personal traits, memory decline that causes the exercise prescription to be forgotten, and physical ailments or fatigue that prevent adherence. External constraints include insufficient exercise facilities, lack of necessary equipment, and a busy family life that interferes with following the exercise prescription. These factors impede the sustainability of adherence behavior and may lead participants to discontinue or abandon it. As obstacle factors do not occur exclusively in a specific stage of adherence behavior transformation but rather persist throughout each stage, they are classified as "pan-stage factors" during the coding process and the division of behavior change stages. These factors are independent of the 4 distinct stages of behavior change.

Solution Strategies

From the qualitative interview data, we extracted shared experiences, suggestions, and needs from participants, which were ultimately distilled into another key factor influencing adherence behavior: "problem-solving strategies." These strategies help promote the initiation and continuity of adherence behavior. "Problem-solving strategies" encompass participants' self-regulation and the optimization and enhancement of the exercise prescription intervention schemes. Participants can better promote adherence behavior through self-management, strict self-discipline, habit formation, resilience enhancement, and flexible adjustment of exercise time or location based on individual circumstances. Health service providers and intervention planners can address participants' needs by

incorporating relevant images and videos into the exercise prescriptions, thereby enhancing participants' motivation and enthusiasm for exercise. They can also regularly schedule or appropriately increase the frequency of health checkups, allowing participants to experience the positive health outcomes of adhering to the exercise prescription. This, in turn, generates positive feedback and promotes the sustained continuity of adherence behavior.

With diverse motivations and problem-solving strategies, adherence behavior progresses into the "adherence maintenance stage." In this phase, individuals who initially adhered gradually transition to full adherence. Fully adherent individuals then become role models, leading and motivating others to achieve complete adherence.

Discussion

Principal Findings

Overall, this study explores adherence behavior to intelligent personalized exercise prescriptions, identifies key influencing factors, and constructs a model that outlines the adherence pathway. The findings indicate that behavioral motivation is the primary driver of adherence, with motivations varying across stages. Intrinsic and extrinsic motivations predominantly influence the early stages, while benefit and hedonic motivations become more significant during the maintenance stage. This suggests that intervention strategies should be dynamically tailored to the individual's stage of behavior change. Additionally, self-regulation and optimization strategies play a crucial role in promoting adherence. Perceived barriers, such as individual limitations and external conditions, can hinder adherence. Moreover, diverse behavioral motivations drive the entire adherence process, facilitating the transition from intention to preparation, action, and ultimately to maintenance.

Enhancing Intrinsic and Extrinsic Motivations to Promote Adherence Intentions and Behavior Preparation

Consistent with the findings reported by Almagro et al [52], both intrinsic and extrinsic motivations contribute to the formation of adherence intentions, signaling the beginning of the intention stage. During this stage, individuals are motivated by intrinsic factors, such as perceived disease threat and susceptibility, as well as extrinsic factors, including support and encouragement from family and society, to adhere to intelligent personalized exercise prescriptions. However, individuals have not yet taken concrete action at this stage. To facilitate progression, it is essential to enhance health education and improve health literacy, thereby strengthening intrinsic motivation. Additionally, providing diverse forms of social support (eg, from family, friends, and doctors) can bolster extrinsic motivation [34,35]. Once individuals transition into the preparation stage, they develop a clearer understanding of their adherence goals and begin making preparations. At this stage, achievement motivation becomes the dominant driving force behind behavioral change.

Satisfying Individual Motivations and Aligning Feedback Mechanisms to Enhance the Maintenance of Adherence Behavior

When individuals begin to engage in adherence behavior, they enter the action stage. At this point, their hedonic and benefit motivations are partially fulfilled, which reinforces their adherence. However, full adherence may not yet be realized. Positive feedback from the outcomes of their behavior plays a critical role in sustaining adherence [53,54]. Therefore, it is essential to provide timely and encouraging feedback during this stage to support continued adherence. A positive feedback loop can enhance patients' exercise self-efficacy, thereby promoting sustained adherence. When individuals' diverse motivations are fulfilled and their adherence challenges are addressed through effective strategies, they transition fully from the action stage to the maintenance stage. In this stage, partial adherents gradually evolve into full adherents. These fully adherent individuals, in turn, can serve as role models, guiding and supporting those still in the action stage, helping them achieve full adherence.

Application of the TTM in Constructing Pathways for the Realization of Adherence Behavior

The precontemplation stage of the TTM refers to individuals who have no intention of changing their behavior within the next 6 months and are not considering any changes. This study, which uses a descriptive exploratory qualitative design within a longitudinal investigation of the effects of intelligent personalized exercise prescriptions on the health of middle-aged and older community residents, includes only participants who have already expressed an intention to change their behavior. Therefore, participants in the precontemplation stage are excluded from this study. As a result, the adherence pathway model developed for this research includes only the 4 stages of the TTM: intention, preparation, action, and maintenance. Additionally, integrating the Multiple Behavioral Motivation Theory with the TTM provides a comprehensive framework for understanding and enhancing adherence to exercise prescriptions. Future research should explore the interactions among various motivational factors and examine their specific effects on adherence behavior.

Enhancing Individual Self-Efficacy to Promote Effective Adherence Behavior

Through qualitative interviews and subsequent analysis, we found that participants with prior successful behavior change experiences (eg, quitting smoking, reducing alcohol consumption, or losing weight) exhibited a stronger intention to adhere to intelligent personalized exercise prescriptions. They also demonstrated better and more sustained adherence compared with others. This can be attributed to their previous successes, which enhanced their self-efficacy for behavior change [50,55,56]. Therefore, future interventions aimed at improving adherence should focus on boosting individual self-efficacy to facilitate effective and lasting adherence behaviors.

Strategies for Promoting, Implementing, and Optimizing Personalized Smart Exercise Prescription in the Community

To enhance the promotion, implementation, and optimization of personalized smart exercise prescriptions, we propose several strategies. First, targeted exercise health education and training activities should be organized within the community to improve exercise literacy and adherence intentions among middle-aged and older residents. Additionally, utilizing channels such as social media and community bulletin boards to disseminate information can help increase awareness and acceptance of personalized smart exercise prescriptions. Second, social cognitive methods have proven effective in promoting and maintaining health behavior changes [57]. Thus, using diverse social support strategies can enhance and sustain adherence to personalized smart exercise prescriptions. For example, creating a support network that includes family, friends, and health care providers can offer emotional support, encouragement, and supervision, helping individuals overcome challenges during exercise.

Additionally, establishing regular feedback mechanisms and offering appropriate incentives are crucial for enhancing adherence behavior [58]. Future efforts should consider utilizing smart health devices or apps to provide timely feedback on exercise performance and health improvements for individuals following personalized smart exercise prescriptions [59,60]. Based on this feedback, the exercise prescription can be continuously optimized and adjusted. For example, individuals without access to exercise spaces could be provided with home-based exercise plans. Additionally, adjusting the intensity, frequency, and type of exercises within the prescription can help overcome obstacles, enhance exercise achievements, and improve adherence to the personalized smart exercise prescription. Regarding incentives, implementing reward systems such as point-based rewards and fitness competitions [61] could stimulate both intrinsic and extrinsic motivations, thus promoting and maintaining adherence behavior. Finally, establishing a long-term tracking mechanism to regularly assess adherence to the personalized smart exercise prescription and overall health status is essential. Providing ongoing support and guidance is crucial for helping individuals maintain adherence behavior over the long term.

Study Limitations

First, this exploratory study has a limited sample size. Although small, the sample has reached theoretical saturation, meaning that adding more participants is unlikely to yield new insights. Second, despite using purposive sampling, selection bias may still be present, as participants with positive attitudes or better adherence to the personalized smart exercise prescription may have been more inclined to participate. To mitigate this, we specifically recruited participants with moderate to low attitudes toward personalized smart exercise prescriptions or those who self-reported low adherence. Additionally, the study focused exclusively on middle-aged and older community residents, excluding other potentially relevant groups, such as working adults or adolescents. This limitation affects the generalizability of the findings. Therefore, future research should include a broader range of populations for comparative analysis to provide a more comprehensive understanding of adherence behavior and its influencing factors across different groups.

Furthermore, qualitative research methods are inherently subjective. Although our study used rigorous qualitative analysis to minimize bias, incorporating quantitative methods could provide more objective data, thereby enhancing the credibility and scientific rigor of the results. Future research should adopt a mixed methods approach, combining both qualitative and quantitative data to gain a more comprehensive understanding. This study, as a qualitative component of a longitudinal investigation on the effects of intelligent personalized exercise prescriptions for middle-aged and older community residents, focuses on a specific cloud-based system. As a result, the generalizability of the findings may be limited. Future research should explore different intelligent personalized exercise prescription systems to validate the broader applicability of the findings. Additionally, the relatively low educational level of the participants may affect the representativeness of the results. Future studies should include participants with diverse educational backgrounds to ensure a more varied sample, gather comprehensive data, and enhance the representativeness of the findings.

Finally, the short duration of this study limited our ability to track long-term adherence behavior. Future research should incorporate long-term follow-ups to assess sustained changes in adherence and provide more comprehensive insights and practical guidance.

Conclusions

This study adopts a descriptive, exploratory, qualitative design nested within a longitudinal study examining the impact of intelligent personalized exercise prescriptions on the health of middle-aged and older community residents. Through individual, face-to-face, semistructured interviews with 12 participants who had adhered to the intelligent personalized exercise prescription for at least 8 months, received exercise health education materials, and received detailed explanations and guidance from staff, we identified 8 main categories and 3 core categories, integrating the TTM and various behavioral motivations. Furthermore, we developed a model illustrating the factors influencing adherence to intelligent personalized exercise prescriptions and the pathway of adherence behavior, based on the TTM and various behavioral motivations. This comprehensive approach deepens the understanding of both the positive and negative factors influencing adherence, as well as their implementation pathways. It also facilitates the development of targeted intervention strategies to enhance adherence, ultimately maximizing the health benefits of exercise.

Acknowledgments

This work was supported by the Major Project of the Anhui Provincial Department of Education (grant 2024AH040340), the Anhui Provincial Department of Education (grants SK2021A0432, 2022lhpysfjd063, and gxyq2022040), and the Research Innovation Project for Graduate Students of Bengbu Medical University 2023 (grant Byycx23020). The funding sources played no role in the study design, data collection, management, analysis, interpretation, report writing, or the decision to submit the manuscript for publication. Additionally, we express our gratitude to graduate student Mengyuan Li and undergraduate nursing student Jing Zhang for their contributions to collecting interview data.

Conflicts of Interest

None declared.

References

1. Colcombe S, Kramer AF. Fitness effects on the cognitive function of older adults: a meta-analytic study. Psychol Sci 2003 Mar;14(2):125-130. [doi: 10.1111/1467-9280.t01-1-01430] [Medline: 12661673]
2. Netz Y, Wu M, Becker BJ, Tenenbaum G. Physical activity and psychological well-being in advanced age: a meta-analysis of intervention studies. Psychol Aging 2005 Jun;20(2):272-284. [doi: 10.1037/0882-7974.20.2.272] [Medline: 16029091]
3. Barton C, King M, Dascombe B, Taylor N, de Oliveira Silva D, Holden S, et al. Many physiotherapists lack preparedness to prescribe physical activity and exercise to people with musculoskeletal pain: A multi-national survey. Phys Ther Sport 2021 May;49:98-105. [doi: 10.1016/j.ptsp.2021.02.002] [Medline: 33662892]
4. Piercy KL, Troiano RP. Physical activity guidelines for Americans from the US Department of Health and Human Services. Circ Cardiovasc Qual Outcomes 2018 Nov;11(11):e005263 [FREE Full text] [doi: 10.1161/CIRCOUTCOMES.118.005263] [Medline: 30571339]
5. Paffenbarger RS, Hyde RT, Wing AL, Hsieh CC. Physical activity, all-cause mortality, and longevity of college alumni. N Engl J Med 1986 Mar 06;314(10):605-613. [doi: 10.1056/NEJM198603063141003] [Medline: 3945246]
6. Badenhop DT, Cleary PA, Schaal SF, Fox EL, Bartels RL. Physiological adjustments to higher- or lower-intensity exercise in elders. Med Sci Sports Exerc 1983;15(6):496-502. [Medline: 6656559]
7. Lucini D, Malacarne M, Gatzemeier W, Pagani M. A simple home-based lifestyle intervention program to improve cardiac autonomic regulation in patients with increased cardiometabolic risk. Sustainability 2020 Sep 17;12(18):7671. [doi: 10.3390/su12187671]
8. Tipton CM. Historical perspective: the antiquity of exercise, exercise physiology and the exercise prescription for health. World Rev Nutr Diet 2008;98:198-245. [doi: 10.1159/000152989] [Medline: 18806470]
9. Zhu L, Wang Z, Zhu W. Construction of exercise prescription repository in the context of Healthy China Initiative. Journal of Sports Science 2020;40(1):4-15. [doi: 10.16469/j.css.202001001]
10. Zhu W. The past, present and future of exercise prescription. Sport Sci Res 2020;41(1):1-18. [doi: 10.12064/ssr.20200101]
11. Wang Z. Research and application progress of exercise prescriptions. Res Sports Sci 2021;35(3):40-49. [doi: 10.15877/j.cnki.nsic.20210601.001]
12. Curry J, Lind M, Short CE, Vandelanotte C, Evans HEL, Pearson M, et al. Evaluating a web-based computer-tailored physical activity intervention for those living with and beyond lung cancer (ExerciseGuide UK): protocol for a single group feasibility and acceptability study. Pilot Feasibility Stud 2022 Aug 13;8(1):182 [FREE Full text] [doi: 10.1186/s40814-022-01129-6] [Medline: 35964141]
13. Pelliccia A, Sharma S, Gati S, Bäck M, Börjesson M, Caselli S, et al. 2020 ESC guidelines on sports cardiology and exercise in patients with cardiovascular disease. Rev Esp Cardiol (Engl Ed) 2021 Jun;74(6):545. [doi: 10.1016/j.rec.2021.05.003] [Medline: 34020769]
14. Piercy KL, Troiano RP, Ballard RM, Carlson SA, Fulton JE, Galuska DA, et al. The physical activity guidelines for Americans. JAMA 2018 Nov 20;320(19):2020-2028 [FREE Full text] [doi: 10.1001/jama.2018.14854] [Medline: 30418471]
15. Zubin Maslov P, Schulman A, Lavie C, Narula J. Personalized exercise dose prescription. Eur Heart J 2018 Jul 01;39(25):2346-2355. [doi: 10.1093/eurheartj/ehx686] [Medline: 29301027]
16. Izquierdo M, Merchant RA, Morley JE, Anker SD, Aprahamian I, Arai H, et al. International exercise recommendations in older adults (ICFSR): expert consensus guidelines. J Nutr Health Aging 2021;25(7):824-853 [FREE Full text] [doi: 10.1007/s12603-021-1665-8] [Medline: 34409961]
17. Netz Y, Yekutieli Z, Arnon M, Argov E, Tchelet K, Benmoha E, et al. Personalized exercise programs based upon remote assessment of motor fitness: a pilot study among healthy people aged 65 years and older. Gerontology 2022;68(4):465-479. [doi: 10.1159/000517918] [Medline: 34515118]
18. Netz Y, Argov E, Yekutieli Z, Ayalon M, Tchelet K, Ben-Sira D, et al. Personalized multicomponent exercise programs using smartphone technology among older people: protocol for a randomized controlled trial. BMC Geriatr 2021 Oct 26;21(1):605 [FREE Full text] [doi: 10.1186/s12877-021-02559-2] [Medline: 34702168]

19. Lin C, Lin Y, Yeh C, Huang C, Kuo L, Su F. An exergame-integrated IoT-based ergometer system delivers personalized training programs for older adults and enhances physical fitness: a pilot randomized controlled trial. Gerontology 2023;69(6):768-782. [doi: 10.1159/000526951] [Medline: 36617413]
20. Sun T, Xu Y, Xie H, Ma Z, Wang Y. Intelligent personalized exercise prescription based on an eHealth promotion system to improve health outcomes of middle-aged and older adult community dwellers: pretest-posttest study. J Med Internet Res 2021 May 24;23(5):e28221 [FREE Full text] [doi: 10.2196/28221] [Medline: 34028359]
21. Tan M, Xiao Y, Jing F, Xie Y, Lu S, Xiang M, et al. Evaluating machine learning-enabled and multimodal data-driven exercise prescriptions for mental health: a randomized controlled trial protocol. Front Psychiatry 2024;15:1352420 [FREE Full text] [doi: 10.3389/fpsyt.2024.1352420] [Medline: 38287940]
22. Luo X, Chen X, Zhu X, Shu Y, He Y, Zhao C, et al. Application of sports APP in cardiac rehabilitation of patients with chronic heart failure. J Med Med Health 2023;39(17):2925-2928. [doi: 10.3969/j.issn.1009-5519.2023.17.010]
23. Ureña R, Chiclana F, Gonzalez-Alvarez A, Herrera-Viedma E, Moral-Munoz J. m-SFT: a novel mobile health system to assess the elderly physical condition. Sensors (Basel) 2020 Mar 06;20(5):1462 [FREE Full text] [doi: 10.3390/s20051462] [Medline: 32155931]
24. Jakob R, Harperink S, Rudolf AM, Fleisch E, Haug S, Mair JL, et al. Factors influencing adherence to mHealth apps for prevention or management of noncommunicable diseases: systematic review. J Med Internet Res 2022 May 25;24(5):e35371 [FREE Full text] [doi: 10.2196/35371] [Medline: 35612886]
25. Sevic A, Hashemi NS, Thørrisen MM, Strømstad K, Skarpaas LS, Storm M, et al. Effectiveness of eHealth interventions targeting employee health behaviors: systematic review. J Med Internet Res 2023 Apr 20;25:e38307 [FREE Full text] [doi: 10.2196/38307] [Medline: 37079369]
26. Bergevi J, Andermo S, Woldamanuel Y, Johansson U, Hagströmer M, Rossen J. User perceptions of eHealth and mHealth services promoting physical activity and healthy diets: systematic review. JMIR Hum Factors 2022 Jun 28;9(2):e34278 [FREE Full text] [doi: 10.2196/34278] [Medline: 35763339]
27. Rhodes RE, Janssen I, Bredin SSD, Warburton DER, Bauman A. Physical activity: health impact, prevalence, correlates and interventions. Psychol Health 2017 Aug;32(8):942-975. [doi: 10.1080/08870446.2017.1325486] [Medline: 28554222]
28. de Ridder D, Kroese F, Evers C, Adriaanse M, Gillebaart M. Healthy diet: health impact, prevalence, correlates, and interventions. Psychol Health 2017 Aug;32(8):907-941. [doi: 10.1080/08870446.2017.1316849] [Medline: 28447854]
29. Coumans JMJ, Lechner L, Bolman CAW, Oenema A. Clustering of motivational constructs based on self-determination theory for diet and physical activity and their associations with behaviour: a cross-sectional study. J Nutr Sci 2022;11:e22 [FREE Full text] [doi: 10.1017/jns.2022.12] [Medline: 35399551]
30. Bischof G, Bischof A, Rumpf H. Motivational interviewing: an evidence-based approach for use in medical practice. Dtsch Arztebl Int 2021 Feb 19;118(7):109-115 [FREE Full text] [doi: 10.3238/arztebl.m2021.0014] [Medline: 33835006]
31. Pynnönen K, Hassandra M, Tolvanen A, Siltanen S, Portegijs E, Rantanen T. Do the integrated theories of self-determination and planned behavior explain the change in active life engagement following a motivational counseling intervention among older people? Soc Sci Med 2023 Dec;339:116409 [FREE Full text] [doi: 10.1016/j.socscimed.2023.116409] [Medline: 37984181]
32. Delestre F, Debailly R, Galiby L, Hansel B. Determinants of motivation and adherence to a web application for health behaviour change: a qualitative study with cancer survivors. J Hum Nutr Diet 2023 Oct;36(5):1701-1712. [doi: 10.1111/jhn.13185] [Medline: 37272366]
33. Mohammadkhah F, Ramezankhani A, Atashpoosh A, Ahmady Moghadam F, Bakhtiar M, Khani Jeihooni A. Effect of training based on health belief model and behavioral intention on improving dental and oral self-care behavior in 9-12-year-old Iranian female students. BMC Oral Health 2022 Nov 19;22(1):515 [FREE Full text] [doi: 10.1186/s12903-022-02552-0] [Medline: 36402972]
34. Hagger MS, Smith SR, Keech JJ, Moyers SA, Hamilton K. Predicting social distancing intention and behavior during the COVID-19 pandemic: an integrated social cognition model. Ann Behav Med 2020 Oct 01;54(10):713-727 [FREE Full text] [doi: 10.1093/abm/kaaa073] [Medline: 32914831]
35. Hsu C, Peng C. What drives older adults' use of mobile registration apps in Taiwan? An investigation using the extended UTAUT model. Inform Health Soc Care 2022 Jul 03;47(3):258-273. [doi: 10.1080/17538157.2021.1990299] [Medline: 34672853]
36. Cao J, Kurata K, Lim Y, Sengoku S, Kodama K. Social acceptance of mobile health among young adults in Japan: an extension of the UTAUT model. Int J Environ Res Public Health 2022 Nov 17;19(22):15156 [FREE Full text] [doi: 10.3390/ijerph192215156] [Medline: 36429875]
37. Chen J, Wang T, Fang Z, Wang H. Research on elderly users' intentions to accept wearable devices based on the improved UTAUT model. Front Public Health 2022;10:1035398 [FREE Full text] [doi: 10.3389/fpubh.2022.1035398] [Medline: 36699866]
38. Prochaska JO, Velicer WF, Rossi JS, Goldstein MG, Marcus BH, Rakowski W, et al. Stages of change and decisional balance for 12 problem behaviors. Health Psychol 1994 Jan;13(1):39-46. [doi: 10.1037//0278-6133.13.1.39] [Medline: 8168470]

39. Xie K, Zhu Y, Ma Y, Chen Y, Chen S, Chen Z. Willingness of tea farmers to adopt ecological agriculture techniques based on the UTAUT extended model. Int J Environ Res Public Health 2022 Nov 21;19(22):15351 [FREE Full text] [doi: 10.3390/ijerph192215351] [Medline: 36430070]
40. Wu B, An X, Wang C, Shin HY. Extending UTAUT with national identity and fairness to understand user adoption of DCEP in China. Sci Rep 2022 Apr 27;12(1):6856 [FREE Full text] [doi: 10.1038/s41598-022-10927-0] [Medline: 35477990]
41. Schröder J, Bäuerle A, Jahre LM, Skoda E, Stettner M, Kleinschnitz C, et al. Acceptance, drivers, and barriers to use eHealth interventions in patients with post-COVID-19 syndrome for management of post-COVID-19 symptoms: a cross-sectional study. Ther Adv Neurol Disord 2023;16:17562864231175730 [FREE Full text] [doi: 10.1177/17562864231175730] [Medline: 37255668]
42. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
43. Saunders B, Sim J, Kingstone T, Baker S, Waterfield J, Bartlam B, et al. Saturation in qualitative research: exploring its conceptualization and operationalization. Qual Quant 2018;52(4):1893-1907 [FREE Full text] [doi: 10.1007/s11135-017-0574-8] [Medline: 29937585]
44. Xu Y, Yin Z, Liu H, Sun M, Guo Z, Liu B. Model construction of Chinese preservice physical education teachers' perception of social media: a grounded theory approach. Int J Environ Res Public Health 2023 Feb 22;20(5):1893 [FREE Full text] [doi: 10.3390/ijerph20053886] [Medline: 36900895]
45. Liu S. Analysis of the factors influencing the effectiveness of local government's purchase of older adults care services - a grounded theory study based on typical cases. Front Public Health 2023;11:1202472 [FREE Full text] [doi: 10.3389/fpubh.2023.1202472] [Medline: 37637803]
46. Flannery M. Self-determination theory: intrinsic motivation and behavioral change. Oncol Nurs Forum 2017 Mar 01;44(2):155-156. [doi: 10.1188/17.ONF.155-156] [Medline: 28222078]
47. Turner AR, Reed SM. Intrinsic motivation in exercise: a concept analysis. Nurs Forum 2022 Jan;57(1):136-143. [doi: 10.1111/nuf.12658] [Medline: 34558057]
48. Teixeira PJ, Carraça EV, Markland D, Silva MN, Ryan RM. Exercise, physical activity, and self-determination theory: a systematic review. Int J Behav Nutr Phys Act 2012 Jun 22;9:78 [FREE Full text] [doi: 10.1186/1479-5868-9-78] [Medline: 22726453]
49. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Quarterly 2003;27(3):425. [doi: 10.2307/30036540]
50. Bandura A, Freeman W, Lightsey R. Self-efficacy: the exercise of control. J Cogn Psychother 1999 Jan 01;13(2):158-166. [doi: 10.1891/0889-8391.13.2.158]
51. Kilgour AHM, Rutherford M, Higson J, Meredith SJ, McNiff J, Mitchell S, et al. Barriers and motivators to undertaking physical activity in adults over 70-a systematic review of the quantitative literature. Age Ageing 2024 Apr 01;53(4):afae080 [FREE Full text] [doi: 10.1093/ageing/afae080] [Medline: 38651329]
52. Almagro BJ, Sáenz-López P, Fierro-Suero S, Conde C. Perceived performance, intrinsic motivation and adherence in athletes. Int J Environ Res Public Health 2020 Dec 16;17(24):9441 [FREE Full text] [doi: 10.3390/ijerph17249441] [Medline: 33339278]
53. Klompstra L, Liljeroos M, Jaarsma T, Strömberg A. Experience of physical activity described by patients with heart failure who have received individualized exercise advice: a qualitative study. J Rehabil Med 2021 Jan 05;53(1):jrm00139 [FREE Full text] [doi: 10.2340/16501977-2771] [Medline: 33225374]
54. Warehime S, Dinkel D, Alonso W, Pozehl B. Long-term exercise adherence in patients with heart failure: a qualitative study. Heart Lung 2020;49(6):696-701 [FREE Full text] [doi: 10.1016/j.hrtlng.2020.08.016] [Medline: 32861888]
55. St Quinton T. Promoting physical activity through priming the content of motivation. Front Psychol 2017;8:1509 [FREE Full text] [doi: 10.3389/fpsyg.2017.01509] [Medline: 28928696]
56. Strecher VJ, DeVellis BM, Becker MH, Rosenstock IM. The role of self-efficacy in achieving health behavior change. Health Educ Q 1986;13(1):73-92. [doi: 10.1177/109019818601300108] [Medline: 3957687]
57. Bandura A. Health promotion by social cognitive means. Health Educ Behav 2004 Apr;31(2):143-164. [doi: 10.1177/1090198104263660] [Medline: 15090118]
58. de Buisonjé DR, Reijnders T, Cohen Rodrigues TR, Prabhakaran S, Kowatsch T, Lipman SA, et al. Investigating rewards and deposit contract financial incentives for physical activity behavior change using a smartphone app: randomized controlled trial. J Med Internet Res 2022 Oct 06;24(10):e38339 [FREE Full text] [doi: 10.2196/38339] [Medline: 36201384]
59. Sporrel K, De Boer RDD, Wang S, Nibbeling N, Simons M, Deutekom M, et al. The design and development of a personalized leisure time physical activity application based on behavior change theories, end-user perceptions, and principles from empirical data mining. Front Public Health 2020;8:528472 [FREE Full text] [doi: 10.3389/fpubh.2020.528472] [Medline: 33604321]
60. Brons A, Braam K, Broekema A, Timmerman A, Millenaar K, Engelbert R, et al. Translating promoting factors and behavior change principles into a blended and technology-supported intervention to stimulate physical activity in children with

asthma (Foxfit): design study. JMIR Form Res 2022 Jul 25;6(7):e34121 [FREE Full text] [doi: 10.2196/34121] [Medline: 35877162]
61. Jessen S, Mirkovic J, Ruland CM. Creating gameful design in mhealth: a participatory co-design approach. JMIR Mhealth Uhealth 2018 Dec 14;6(12):e11579 [FREE Full text] [doi: 10.2196/11579] [Medline: 30552080]

Abbreviations

COREQ: Consolidated Criteria for Reporting Qualitative Research
TTM: Transtheoretical Model

Edited by L Buis; submitted 17.04.24; peer-reviewed by Z Xie, R Cui, Q Lv, Y Zhou; comments to author 29.05.24; revised version received 23.07.24; accepted 11.11.24; published 05.12.24.

Please cite as:
Xu X, Zhang G, Xia Y, Xie H, Ding Z, Wang H, Ma Z, Sun T
Influencing Factors and Implementation Pathways of Adherence Behavior in Intelligent Personalized Exercise Prescription: Qualitative Study
JMIR Mhealth Uhealth 2024;12:e59610
URL: https://mhealth.jmir.org/2024/1/e59610
doi:10.2196/59610
PMID:

Original Paper

Feasibility and Efficacy of a Novel Mindfulness App Used With Matcha Green Tea in Generally Healthy Adults: Randomized Controlled Trial

Ryohei Tanaka-Kanegae[1], MS; Koji Yamada[2], MS, MBA; Chad M Cook[3], PhD; Traci M Blonquist[3], MS; Kristen D Taggart[3], MPH; Koichiro Hamada[1,2], DVM, PhD

[1]Saga Nutraceuticals Research Institute, Otsuka Pharmaceutical Co Ltd, Saga, Japan

[2]Otsuka Holdings Co Ltd, Tokyo, Japan

[3]Biofortis, Inc, Addison, IL, United States

Corresponding Author:
Ryohei Tanaka-Kanegae, MS
Saga Nutraceuticals Research Institute
Otsuka Pharmaceutical Co Ltd
5006-5 Aza Higashiyama, Omagari
Yoshinogari-cho, Kanzaki-gun
Saga, 842-0195
Japan
Phone: 81 952521522
Fax: 81 952532727
Email: tanaka.ryohei@otsuka.jp

Abstract

Background: Mindfulness practices, such as breathing meditation (BM), reduce stress and enhance mood. One such practice is mindful eating, where a practitioner focuses on the five senses while eating or drinking. A novel set of prototypes has been developed, incorporating principles of mindful eating. These prototypes include matcha green tea and a mobile app that provides audio guidance for meditation during the preparation and consumption of the beverage (hereafter referred to as guided tea meditation [GTM]).

Objective: This study assessed the feasibility and efficacy of GTM, evaluating meditation time, frequency, and prototype acceptability over 8 weeks, alongside changes in stress and mood. Additionally, other benefits of GTM were explored.

Methods: A comparator group was established in which participants performed traditional BM without an app or audio guide (active control). This unblinded randomized controlled trial involved 100 healthy American volunteers (n=49 GTM, n=51 BM). During the 8-week study period, participants were encouraged to perform either GTM or BM for 10 minutes daily. The meditation activity was self-reported the following day. Only the GTM group assessed the prototype acceptability. The Perceived Stress Scale-10 was used to measure stress levels, while the Two-Dimensional Mood Scale was used to evaluate mood changes. Other meditation benefits were explored using a questionnaire. All questionnaires were presented and completed via an app. An intention-to-treat analysis was performed.

Results: No significant between-group differences were found in total meditation time (P=.15) or frequency (P=.36). However, the weekly time and frequency of the GTM group remained above 50 minutes per week and 4 days per week, respectively. Over half of the GTM participants (≥28/49, ≥57%) accepted the prototype. The GTM group exhibited significant stress reductions at weeks 4 and 8 (both P<.001), similar to the BM group. Improvements in mood metrics were observed after a single GTM session on days 1 and 56, similar to the BM group. Moreover, increases in premeditation scores for relaxed and calm from day 1 to day 56 were significantly higher for the GTM group (P=.04 and .048, respectively). The majority of participants (≥25/49, ≥51%) assigned to GTM experienced positive changes in happiness, time management, quality of life, relationships, sleep, and work performance as they continued meditating. However, no significant between-group differences were found in these exploratory outcomes (P>.08).

Conclusions: We believe that GTM exhibits good feasibility. Meanwhile, GTM reduced stress, improved mood, and let the practitioners feel other benefits, similar to BM. Long-term practitioners of GTM may even feel more relaxed and calmer in the state of premeditation than those who practice BM.

Trial Registration: ClinicalTrials.gov NCT05832645; https://clinicaltrials.gov/study/NCT05832645

(JMIR Mhealth Uhealth 2024;12:e63078) doi:10.2196/63078

KEYWORDS

mindfulness; guided tea meditation; meditation; matcha; green tea; mobile app; smartphone; stress; mood; mHealth; mobile health; well-being; wellness

Introduction

Background

Stress management or reduction is an important skill for many individuals because of the increasing occurrence of stress, anxiety, and depression. Indeed, approximately one-third of the global population experiences stress [1]. The United States is one of the most stressed nations in the world, with more than three-fourths of American adults reporting symptoms of stress, including headache, tiredness, and sleep problems [2]. This problem was exacerbated by the COVID-19 pandemic and people still struggle to adapt to the rapid changes caused by the pandemic [3-5]. Stress not only affects mood and mental health but also physical health, either directly by disturbing the autonomic nervous and neuroendocrine systems or indirectly by causing changes in health behaviors [6]. Thus, it is important to assess and encourage stress-lowering techniques to achieve well-being and tranquility in life.

Mindfulness-based approaches have been increasingly used to help individuals respond more effectively to stress and other negative internal experiences. Participation in an 8-week mindfulness-based stress reduction (MBSR) program lowered stress and anxiety symptoms [7,8]. Breathing meditation (BM; referred to as sitting meditation with awareness of breathing in the MBSR program) is the main technique that program participants are encouraged to practice daily [9]. During the meditation, a practitioner pays attention to bodily sensations that accompany breathing, and whenever they notice their mind wandering, they note what diverts their attention and subsequently redirect it to breathing [10]. Studies have demonstrated the positive effects of a 10-minute BM on emotion regulation and attention control [11,12].

MBSR programs provide a variety of mindfulness techniques other than BM and participants are encouraged to seek as many opportunities as possible to practice these techniques in their daily lives [10]. One such technique is mindful eating, in which a practitioner pays attention to the sight, smell, taste, texture, and temperature of food, as well as their thoughts and feelings while eating [13]. Several studies have shown a correlation between the degree of beneficial traits or effects that practitioners can obtain from meditative training and the total time or frequency of meditation [12,14,15]. Therefore, it seems reasonable to use time spent eating and drinking, which is an essential part of living, as an opportunity to practice mindfulness.

Japan has a long-standing tradition of tea ceremony where matcha green tea is served to guests. The ceremony has been shown to have a calming effect on participants [16,17]. Regarding taste, matcha green tea tastes sweet and umami (savory), as well as bitter [18]. We believe this differentiates matcha green tea from other beverages such as regular green tea and coffee, enabling practitioners to cultivate their awareness while focusing on the different tastes. Hence, we developed a dedicated matcha green tea for mindful eating as a prototype. Moreover, a prototype mobile app that provides audio guidance to perform meditation while making and drinking the beverage was developed. Hereafter, the whole program for stress management provided by the novel set of prototypes is referred to as guided tea meditation (GTM).

There are several commercial mindfulness apps that guide mindful eating. However, to the best of our knowledge, they are mainly intended for weight control [19]. Their efficacy on mental health, such as stress levels and mood, is limited [20], and no commercial products exist that incorporate an app and food for cultivating mindfulness.

Objective

This study aimed to investigate the feasibility and efficacy of GTM. To assess feasibility, the time and frequency of meditation using the prototypes were evaluated over 8 weeks. Additionally, the acceptability of the prototypes was evaluated. To assess efficacy, changes in stress levels over 8 weeks and changes in mood before and after single and multiple meditation sessions were investigated. Moreover, other benefits of GTM were explored using a questionnaire. The findings of this study may facilitate the collaboration between the two different fields, mobile health and food science.

Methods

Ethical Considerations

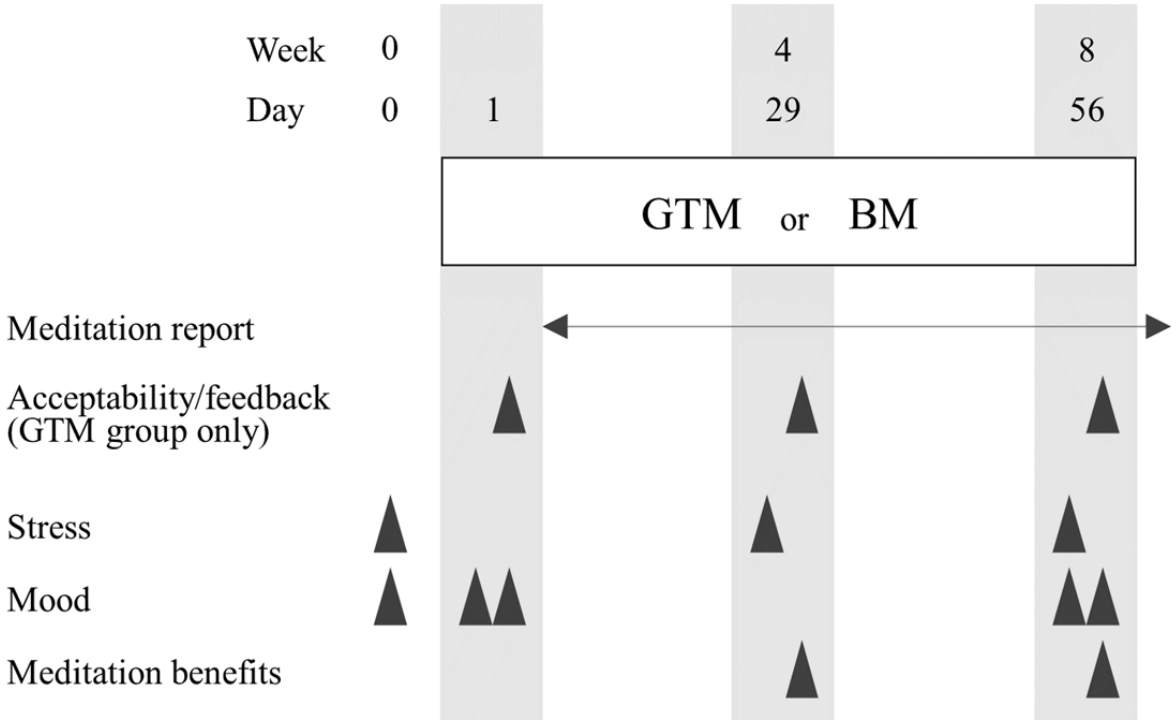
This study was approved by the Sterling IRB (Atlanta, Georgia; approval: 10829), registered at ClinicalTrials.gov (registration: NCT05832645), and conducted by Biofortis Clinical Research in accordance with the Declaration of Helsinki. All participants provided electronic consent before participating in the study. Data were anonymized and then analyzed. Email and Apple IDs were collected from the participants assigned to the GTM group for user authentication of the prototype app and stored with limited access. Other personally identifiable information was kept in strict confidence. Participants received a US $250 Visa gift card for completing all the surveys given below and were informed that the number and duration of the meditation sessions would not affect the stipend.

Trial Design

A comparator group was established in which participants performed traditional BM without an app or audio guide (active control), and this study used a randomized, unblinded, parallel-group design. Web-based screening was conducted to

assess eligibility and perceived stress levels. Given the results, eligible participants were asked to download the Castor Connect mobile app (Castor Connect) and allocated 1:1 to the GTM or BM group. Participants were encouraged to perform their assigned study intervention daily during the 56-day (8-week) intervention period. The duration was set according to the MBSR program. The meditation session performed on the previous day was self-reported by the participants via the mobile app. The acceptability of the prototypes was assessed and participant feedback was obtained only in the GTM group on days 1 (after the first use), 29, and 56 (after the last use). Perceived stress levels were assessed at weeks 0 (screening), 4 (day 29), and 8 (day 56). The mood was assessed before and after the first (day 1) and last (day 56) meditation sessions. A survey to explore the benefits of meditation was conducted at weeks 4 (day 29) and 8 (day 56). All questionnaires were presented and completed via the app. The study scheme and evaluation points are shown in Figure 1.

Figure 1. Study scheme. BM: breathing meditation; GTM: guided tea meditation.



Participants

The participants were recruited in May 2023 via email, websites, flyers, and Facebook. The main inclusion criteria were (1) aged 20-49 years; (2) born in the United States; (3) self-reported history of meditation, but no meditation within 30 days (one month) of screening; (4) access to an Apple ID, an email address linked to the Apple ID, and a dedicated smartphone (iPhone 8 or newer models with iOS 16 or newer versions) capable of downloading and running the app; (5) the ability to download, install, and update apps using a smartphone; (6) access to hot water to prepare the matcha green tea beverage; and (7) a willingness to maintain habitual diet (including supplements), lifestyle, and physical activity during the study. The third inclusion criterion was based on the product marketing strategy. The main exclusion criteria were as follows: (1) visual or hearing impairments that could not be corrected with glasses, contact lenses, or hearing aids; (2) self-reported taste or smell impairments within 30 days of screening; (3) a condition that prevented caffeine consumption (eg, caffeine sensitivity and underlying heart condition); and (4) contraindication, allergy, or sensitivity to any components in the study product or allergens present in the facility used to manufacture or pack the study product. A full list of the inclusion and exclusion criteria is available [21]. Based on their responses to the prescreening questionnaires, qualified participants proceeded to provide electronic informed consent for study participation. Eligible participants had a video call with the clinic staff to review the study instructions and confirm when to start the intervention.

Interventions for the GTM Group

Participants were instructed to perform GTM using the prototype, ekkomi center me (Otsuka Holdings Co Ltd). ekkomi center me consists of matcha green tea and a mobile app that delivers audio guidance to perform meditation while making and drinking the beverage. Audio guidance consists of three parts: (1) a guide for a relaxing introductory exercise, such as progressive muscle relaxation and guided imagery [22,23] (five variations); (2) a guide for meditation with matcha green tea (GTM); and (3) a guide for refocusing. Users could choose a voice actor out of three options and a background sound out of four options every time they used it. The total duration of all audio clips was designed to be approximately 10 minutes, considering time constraints. The contents of the app and audio clips were developed by a psychiatrist and mindfulness experts, as mentioned in the *Acknowledgments* section. Screenshots of the prototype app and an example of the audio script for GTM are provided in Multimedia Appendices 1 and 2, respectively. The flavor of matcha green tea was adjusted for a balanced taste profile with bitterness, sweetness, and umami, and the tea was

attached to a stick so that a user could easily make the tea beverage and pay attention to its complex but pleasant taste. The appearance of the prototype matcha stick is shown in Multimedia Appendix 3. The sticks were individually packed so that the participants could carry them.

Participants assigned to GTM received a packet that included the matcha sticks for 8 weeks and instructions to download the prototype app through TestFlight (attached to Multimedia Appendix 4). The participants were encouraged to perform GTM once a day for 8 weeks and were asked to report the time and frequency of meditation. Participants could choose when and where to meditate freely, and they were not prompted to meditate during the intervention period.

Interventions for the BM Group (as the Active Control)

Participants assigned to BM received a packet that included instructions to perform BM (Multimedia Appendix 5). They were encouraged to perform BM for at least 10 minutes per day according to the instructions. In traditional mindfulness-based programs such as MBSR, practitioners are encouraged to meditate for more than 45 minutes daily [9]. Therefore, the meditation time was not limited. Participants could choose when and where to meditate freely, and they were not prompted to meditate during the intervention period.

Outcomes

The total meditation time and frequency during the study period were calculated as the primary outcome measures based on the daily meditation reports of the participants. Frequency was defined as the number of days in which the participants completed at least one meditation session. Weekly meditation time and frequency were calculated and compared between the groups as secondary measures.

Changes in stress levels were assessed as secondary outcomes using the Perceived Stress Scale-10 (PSS-10) [24]. The total score ranges from 0 to 40, with higher scores indicating higher levels of stress. Mood metrics were also assessed as secondary outcomes using the Two-Dimensional Mood Scale [25]. In addition to the original 8 components (ie, calm, energetic, irritated, lethargic, listless, lively, nervous, and relaxed), measures of focus and anxiety were assessed using a 6-point Likert scale ranging from 0=Not at all to 5=Extremely. Changes from pre- to postmeditation scores and changes in premeditation scores over the study period were compared between groups. Other benefits of meditation were explored at weeks 4 and 8 by asking participants whether they noticed any positive changes in feelings of happiness, time management, physical discomfort, quality of life, relationships, sleep, and work performance. The answers were collected on a 5-point scale ranging from 1=Strongly Agree to 5=Strongly Disagree.

Sample Size

Because this was the first study to assess the feasibility and efficacy of GTM, the anticipated effect size was not available. We referred to two previous studies that evaluated mindfulness apps and recruited approximately 100 participants for two arms [20,26]. Under continuous outcome conditions with 100 participants (50 per arm), there was 80% power to detect an effect size of 0.57 using a 2-sided *t* test at a significance level of .05.

Randomization and Blinding

Participants were randomly assigned to one of the two intervention groups using a variable block randomization model with block sizes of 4 and 6 following an algorithm that automated randomization assignment in the Castor Platform. Randomization was 1:1 stratified by sex (male or female) and the screening score on the PSS-10 (0-17 [below-average stress] and 18-40 [above-average stress]). The average stress levels were based on a mean of approximately 17 in a sample of 2370 American men and women aged ≤45 years [27]. Hence, 100 eligible participants were enrolled in the GTM (n=49) or BM (n=51) group. The study staff and participants were unblinded to the group allocation.

Statistical Methods

An intention-to-treat (ITT) analysis was performed for all outcomes. The number of minutes of meditation over 56 days was compared between groups using a 2-tailed *t* test. Additionally, the response profile for all continuous outcomes measured over time (eg, stress levels and the weekly number of minutes meditated) was evaluated using a repeated-measures model. The outcome measured at each time point was included in the response vector, including the baseline. The covariance structure for the repeated measures was selected based on the minimization of the corrected Akaike information criteria. The model contained terms for time point, group, and time point by group interaction. Contrast or estimate statements were used to calculate the within-group change from baseline to each specified time point, as well as the between-group difference in the change from baseline. The frequency of meditation was evaluated using a generalized linear model following negative binomial regression and log link. The weekly frequency was explored using a generalized linear mixed model, where the model fit was similar to the repeated-measures model for continuous outcomes measured over time. To analyze within- and between-group changes in mood metrics, the Wilcoxon signed rank test and Wilcoxon rank sum test were used, respectively. Responses to the meditation benefit survey were recategorized (ie, strongly agree or agree, undecided, and disagree or strongly disagree) and compared between the groups using Fisher exact test.

In cases where meditation activity was not reported, the meditation time and frequency for the day were treated as zero. Missing data for stress and mood scores were not imputed, and only observed data were included in the analysis.

All tests were 2-sided and performed at a .05 significance level. Statistical analyses were performed using Statistical Analysis Systems Software (version 9.4; SAS Institute) and R (version 4.2.2; R Foundation for Statistical Computing).

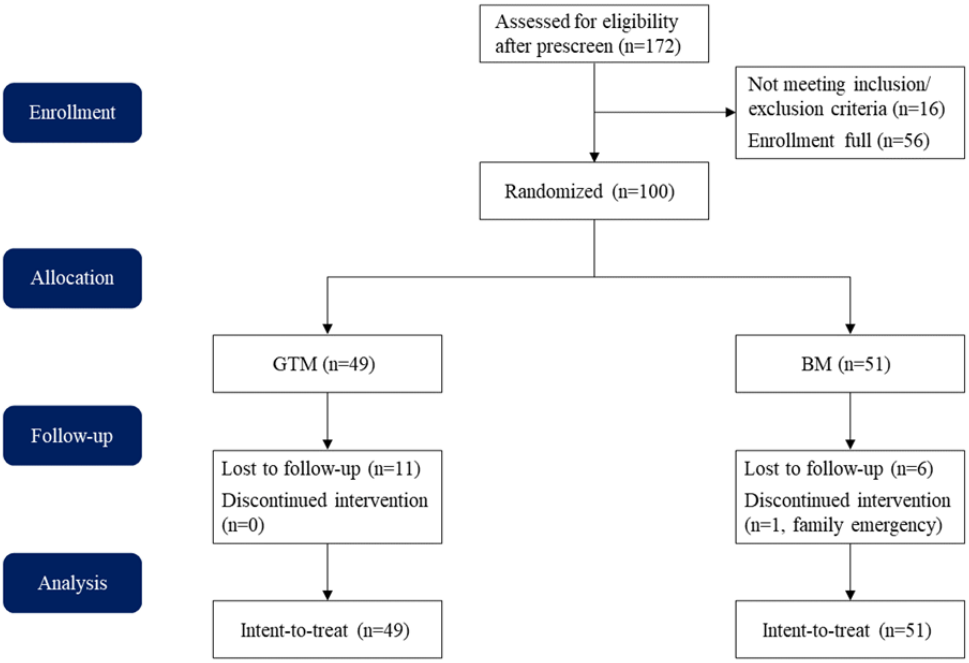
Results

Participants

The CONSORT (Consolidated Standards of Reporting Trials) participant flow diagram is shown in Figure 2, and the

CONSORT checklist is provided in Multimedia Appendix 6. The recruitment of test participants started in May 2023 and the follow-up ended in July 2023. Regarding retention, 38 (78%) out of 49 participants in the GTM group responded to follow-up surveys at 8 weeks, while 44 (86%) out of 51 participants responded in the BM group.

Figure 2. CONSORT (Consolidated Standards of Reporting Trials) participant flow diagram. BM: breathing meditation; GTM: guided tea meditation.



The participant demographics are presented in Table 1. Of the 100 participants, 81 (81%) participants were female and the average stress level was slightly higher than previously reported values for the US population (~17) [27]. Throughout the study, no adverse events were reported.

Table 1. Participant demographics.

Variable	GTM[a] (n=49)	BM[b] (n=51)	Overall (n=100)
Age (years), mean (SD)	34.5 (7.7)	33.3 (7.9)	33.9 (7.8)
Sex, n (%)			
Female	40 (82)	41 (80)	81 (81)
Male	9 (18)	10 (20)	19 (19)
PSS-10[c] group, n (%)			
Below-average stress	15 (31)	16 (31)	31 (31)
Above-average stress	34 (69)	35 (69)	69 (69)
PSS-10 screening score, mean (SD)	19.3 (4.9)	19.9 (5.5)	19.6 (5.2)

[a]GTM: guided tea meditation.

[b]BM: breathing meditation.

[c]PSS-10: Perceived Stress Scale-10.

Feasibility Assessment

Time and Frequency of Meditation (Prototype Use)

The total meditation time and frequency during the study period are presented in Table 2. Neither the time (mean 454.6, SD 219.4 minutes for GTM vs mean 521.0, SD 238.6 minutes for BM; *P*=.15) nor frequency (mean 37.4, SD 17.8 days for GTM vs mean 42.1, SD 16.1 days for BM; *P*=.36) significantly differed between groups.

Table 2. Total meditation time and frequency during the 8-week intervention period.

Variable	GTM[a] (n=49), mean (SD)	BM[b] (n=51), mean (SD)	Difference, mean (95% CI)[c]	*P* value
Time, total (minutes)	454.6 (219.4)	521.0 (238.6)	–66.5 (–157.5 to 24.6)	.15
Frequency, total (days)	37.4 (17.8)	42.1 (16.1)	0.9 (0.7 to 1.1)	.36

[a]GTM: guided tea meditation.

[b]BM: breathing meditation.

[c]The difference between groups is presented as the model-derived mean and 95% CI.

Weekly meditation time and frequency are listed in Multimedia Appendix 7. Regarding weekly meditation time, the interaction between week and group factors was significant (*P*=.02). There was a tendency that the weekly meditation minutes in the GTM group decreased in the latter half of the 8 weeks, whereas those of the BM group remained constant, and a significant difference was observed between the groups at week 5 (*P*=.04) and 8 (*P*=.045). Regarding frequency, no significant interaction between week and group was detected (*P*=.81), whereas a marginally significant difference was observed between groups at week 5 (*P*=.049).

Acceptability of and Feedback on the Prototypes

For the question "Was the app easy to understand and use?" 90% (n=44), 84% (n=41), and 71% (n=35) of the 49 participants answered "strongly agree or agree" on days 1 (after the first use of the prototype), 29, and 56 (after the last use), respectively.

For the question "Do you like the format and content of the audio guidance?" 82% (n=40), 69% (n=34), and 57% (n=28) of the participants answered "strongly agree or agree" on days 1, 29, and 56, respectively. Additionally, 45% (n=22) of the participants provided feedback that they would like more options in the audio clips.

For the question "Did you find the matcha green tea beverage tasty?" 80% (n=39) of the participants answered "strongly agree or agree" on day 1, with 59% (n=29) of them answering the same on days 29 and 56. Of the participants, 37% (n=18) gave feedback that they would like a sweeter taste, and 18% (n=9) commented that they would like more flavor options.

Efficacy Assessment

Perceived Stress Levels

Total PSS-10 scores are presented in Table 3. There was no significant interaction between week and group, indicating no significant difference in the response profiles between the groups (*P*=.29). Within the GTM group, a significant decrease in the PSS-10 score was detected between weeks 0 and 4 (estimate –4.8, 95% CI –6.5 to –3.1; *P*<.001) and between weeks 0 and 8 (–5.5, 95% CI –7.4 to –3.5; *P*<.001). Similarly, within the BM group, a significant difference was detected between weeks 0 and 4 (–4.5, 95% CI –6.1 to –2.9; *P*<.001) and between weeks 0 and 8 (–6.5, 95% CI –8.4 to –4.7; *P*<.001). No significant differences were detected between the groups in the changes from baseline to week 4 (*P*=.79) or week 8 (*P*=.44).

Table 3. Changes in perceived stress levels[a].

Week	GTM[b] (n=49), mean (SEM)	BM[c] (n=51), mean (SEM)	Between-group difference, mean (95% CI)[d]	*P* value
0	19.3 (0.7)	19.9 (0.8)	N/A[e]	N/A
4	14.3 (0.9)[f]	15.3 (0.9)[f]	–0.3 (–2.6 to 2.0)	.79
8	13.8 (0.9)[f]	13.5 (1.0)[f]	1.1 (–1.6 to 3.8)	.44

[a]Perceived stress levels were assessed using the Perceived Stress Scale-10.

[b]GTM: guided tea meditation.

[c]BM: breathing meditation.

[d]The difference between groups in the change from week 0 is presented as the model-derived mean and 95% CI.

[e]Not applicable.

[f]*P*<.001 versus week 0.

Mood

The pre- and postmeditation mood scores and changes are presented in Table 4 (data at the first meditation session on day 1) and Multimedia Appendix 8 (data at the last meditation session on day 56). During the first meditation session, the within-group change from the pre to postmeditation time point was significant for all items in both groups, except for energetic (GTM, *P*=.34; BM, *P*=.71) and lively (GTM, *P*=.12; BM, *P*=.12). Similarly, during the last meditation session, the within-group changes in both groups were significant, except for energetic (GTM, *P*=.42; BM, *P*=.12), lethargic (GTM, *P*=.95), and lively (GTM, *P*=.98). No significant between-group differences were detected in premeditation or postmeditation scores or changes at the first or last meditation session.

Changes in premeditation mood scores from baseline (day 1) to day 56 are presented in Table 5. Significant between-group differences were observed in the relaxed and calm levels (*P*=.04 and .048, respectively; Figure 3).

Table 4. Changes in mood before and after the first meditation session on day 1[a].

Variable	GTM[b]				BM[c]				Between-group *P* value for change
	Pre (n=42), mean (SEM)	Post (n=47), mean (SEM)	*P* value (pre vs post)	Change (post – pre), mean (SEM)	Pre (n=44), mean (SEM)	Post (n=50), mean (SEM)	*P* value (pre vs post)	Change (post – pre), mean (SEM)	
Anxious	1.9 (0.2)	0.9 (0.1)	<.001[d]	–1.0 (0.2)	1.7 (0.2)	0.7 (0.1)	<.001[d]	–1.0 (0.2)	.70
Calm	2.0 (0.2)	3.3 (0.2)	<.001[d]	1.3 (0.2)	2.2 (0.2)	3.1 (0.1)	<.001[d]	0.9 (0.2)	.15
Energetic	1.7 (0.2)	2.0 (0.2)	.34	0.2 (0.2)	1.8 (0.2)	1.8 (0.2)	.71	0.1 (0.2)	.66
Focused	1.8 (0.2)	2.8 (0.2)	<.001[d]	1.0 (0.2)	1.6 (0.2)	2.5 (0.1)	<.001[d]	0.9 (0.2)	.78
Irritated	1.4 (0.2)	0.4 (0.1)	<.001[d]	–1.0 (0.2)	1.6 (0.2)	0.5 (0.1)	<.001[d]	–1.0 (0.2)	.63
Lethargic	1.4 (0.2)	0.6 (0.1)	<.001[d]	–0.8 (0.1)	1.5 (0.2)	0.9 (0.2)	<.001[d]	–0.5 (0.1)	.17
Listless	0.9 (0.2)	0.6 (0.1)	.004[e]	–0.4 (0.1)	1.3 (0.2)	0.7 (0.1)	<.001[d]	–0.6 (0.2)	.25
Lively	1.5 (0.2)	1.8 (0.2)	.12	0.3 (0.2)	1.3 (0.1)	1.7 (0.2)	.12	0.3 (0.2)	.87
Nervous	1.4 (0.2)	0.6 (0.1)	.008[e]	–0.8 (0.2)	1.4 (0.2)	0.6 (0.1)	<.001[d]	–0.8 (0.1)	.38
Relaxed	1.5 (0.2)	3.3 (0.2)	<.001[d]	1.9 (0.2)	1.6 (0.2)	3.1 (0.2)	<.001[d]	1.6 (0.2)	.28

[a]Mood was assessed using the Two-Dimensional Mood Scale with minor modifications. Changes from pre- to postmeditation were calculated.

[b]GTM: guided tea meditation.

[c]BM: breathing meditation.

[d]*P*<.001.

[e]*P*<.01.

Table 5. Changes in premeditation mood scores during the 8-week intervention period[a].

Variable	GTM[b] (n=34), mean (SEM)	BM[c] (n=38), mean (SEM)	*P* value
Anxious	–0.9 (0.2)	–0.2 (0.3)	.10
Calm	1.2 (0.2)	0.6 (0.2)	.048[d]
Energetic	0.5 (0.3)	0.5 (0.2)	.72
Focused	0.7 (0.3)	0.6 (0.2)	.59
Irritated	–0.7 (0.2)	–0.6 (0.2)	.96
Lethargic	–0.3 (0.2)	–0.2 (0.1)	.49
Listless	–0.3 (0.2)	–0.4 (0.2)	.73
Lively	0.6 (0.3)	0.8 (0.2)	≥.99
Nervous	–0.7 (0.2)	–0.3 (0.2)	.10
Relaxed	1.4 (0.3)	0.6 (0.2)	.04[d]

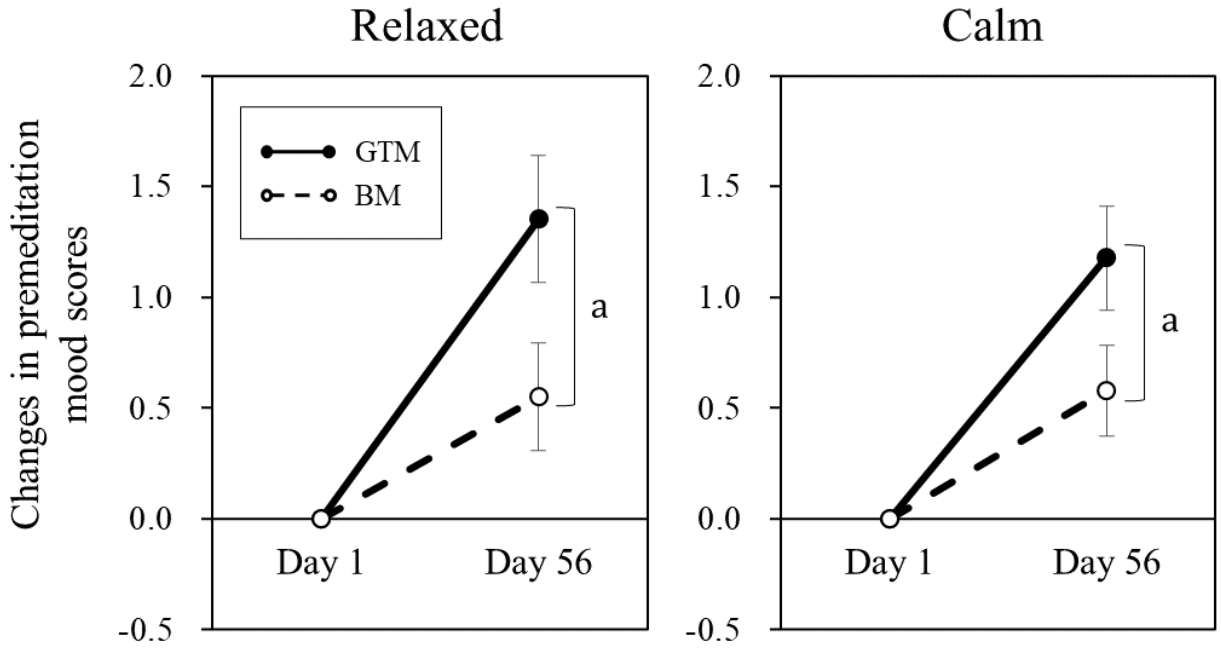
[a]Mood was assessed using the Two-Dimensional Mood Scale with minor modifications. Changes in premeditation scores from baseline (day 1) to day 56 were calculated.

[b]GTM: guided tea meditation.

[c]BM: breathing meditation.

[a]*P*<.05.

Figure 3. Changes in premeditation relaxed and calm scores from baseline (day 1) to day 56 were significantly greater in the GTM group. Mood was assessed using the Two-Dimensional Mood Scale. Data are presented as mean (SEM). a: *P*<.05. BM: breathing meditation; GTM: guided tea meditation.



Other Meditation Benefits

The categorized meditation benefit responses are shown in Multimedia Appendix 9. More than half of the participants in each group experienced positive changes in feelings of happiness (GTM: 32/49, 65%; BM: 34/51, 67%), time management (GTM: 25/49, 51%; BM: 37/51, 73%), quality of life (GTM: 37/49, 76%; BM: 39/51, 76%), relationships (GTM: 28/49, 57%; BM: 34/51, 67%), sleep (GTM: 26/49, 53%; BM: 33/51, 65%), and work performance (GTM: 30/49, 61%; BM: 31/51, 61%) at week 4. Similar results were observed at week 8. However, the groups displayed no significant differences in the distribution of responses at both time points (all *P*>.05).

Discussion

Principal Findings and Comparison With Prior Work

Although the performance of GTM was restricted to once (approximately 10 minutes long) a day, whereas the duration of BM was unrestricted, we expected that the meditation time and frequency in the GTM group would be equal to or higher than those in the BM group. As a result, the total meditation time and frequency did not significantly differ between the groups (Table 2), and weekly meditation time and frequency were even lower in the GTM group at weeks 5 and 8 (Table 3). However, the weekly time and frequency of the GTM group remained above 50 minutes per week and 4 days per week, respectively, throughout the study. Notably, these values, especially the frequency, are not lower than those reported for popular mindfulness apps on the market, although direct comparisons are not feasible due to differing test conditions between the studies [26,28-31]. On the other hand, the weekly meditation time in the BM group remained constant at approximately 65 minutes throughout the 8 weeks, although existing research consistently reports a decrease in engagement with mindfulness over time [32-35]. Moreover, the frequency in the BM group ranged from 5 to 6 times a week, indicating that the participants succeeded in performing BM almost every day as instructed. In general, meditating on a daily basis is considered difficult for most people due to several barriers [10,36]. However, it seemed that the population of this study, who had experience with meditation and discontinued it before study participation, did not perceive the daily implementation of BM as a burden. Although the BM group received only brief instructions on how to perform BM (Multimedia Appendix 5), none of the participants required additional instruction or assistance.

It should also be noted that 22% (11/49) and 14% (7/51) of the participants in the GTM and BM groups, respectively, were lost to follow-up or discontinued the intervention during the study. Although these attrition rates were not significantly higher than those of previous studies involving digital mental health interventions [20,36-38], attrition did affect time and frequency outcomes, as we adopted an ITT analysis. Considering participant feedback and decreases in weekly meditation time in the GTM group during the latter half of the 8 weeks, some participants may have become bored with the limited variety in the prototype offerings (5 types of audio scripts and 1 matcha flavor) and stopped using them. Downloading the prototype app through TestFlight (the procedures are shown in Multimedia Appendix 4) might be a hurdle for some participants. In addition, preparing for the environment to listen to audio, the matcha stick, and hot water might have been constraints for the GTM group. However, no obvious feedback was provided regarding this point, and more than half of the users showed acceptance of the prototype app, audio, and matcha throughout the 8 weeks. In conjunction with the meditation engagement data, we believe

that GTM exhibits good feasibility. However, further research is needed to expand the variety of offerings.

The effects of GTM on perceived stress levels and mood were investigated and compared with those of BM. The BM intervention that did not involve an app or audio guidance significantly reduced stress levels after 4 weeks. Similarly, GTM showed significant reductions in approximately 5 points of the PSS-10, which are similar to, or greater than, values reported by previous mobile health studies [26,28,39-41]. Direct comparisons with other mindfulness apps with audio guidance are desirable for future studies. A single session of BM significantly and consistently improved mood metrics after the first (day 1) and last (day 56) meditation sessions. Similar effects of GTM were observed and were sustained on day 56. Notably, the changes from day 1 to day 56 in the premeditation relaxed and calm scores were significantly greater in the GTM group. The greater changes in relaxed and calm levels observed in the GTM group may be attributed to the L-theanine contained in the matcha green tea. The relaxing effect of L-theanine has been demonstrated [42,43]. In the GTM group, the average premeditation score of relaxed and calm changed over 56 days, from 1.5 (SEM 0.2) to 2.9 (SEM 0.2) and from 2.0 (SEM 0.2) to 3.2 (SEM 0.2), respectively. In other words, the average participant in the prototype group experienced improvements in their relaxed and calm levels by more than one stage (eg, “somewhat” relaxed to “moderately” relaxed) in the state of premeditation. Elevated relaxed and calm levels may generate well-being and equanimity among practitioners, which needs to be investigated in future studies.

The majority of participants in both the BM and GTM groups experienced positive changes in happiness, time management, quality of life, relationships, sleep, and work performance as they continued the assigned meditation. However, these results should be interpreted with caution because the questionnaire used to assess these exploratory outcomes has not been validated. Future studies should conduct specific, validated surveys to assess each outcome.

Strengths and Limitations

A strength of this study is that a comparator group was set as the active control (traditional BM), whereas several relevant previous studies set a waitlist group as the control [26,38,44]. Another strength is that ITT analysis was applied to all outcomes, showing the stress-reducing and mood-improving effects of GTM. To the best of our knowledge, this is the first study demonstrating such effects of mindful eating–based practices by ITT analysis. Moreover, this study showed improvements in relaxed and calm levels following multiple GTM sessions were greater than those following multiple BM sessions.

This study does have some limitations. First, all data collected relied on a self-report method. Therefore, the reported meditation activity may not fully reflect actual activity, and data on the efficacy of meditation may be affected by expectation or placebo effects. After randomization, participants received information regarding whether their intervention was the intervention of interest (GTM) or the comparator. Second, correction for multiple comparisons was not conducted for mood metrics considering the exploratory nature of the study. To generalize the findings of this study, additional research needs to be conducted with a limited number of outcomes and a larger sample size. Third, matcha green tea consumption was limited to once a day, as this was the first long-term study to allow participants to consume the prototype matcha. Confirming the safety of the 8-week consumption of the matcha, we do not plan to limit the number of matcha consumption or GTM in commercial settings. Future studies should reflect real-world settings and be conducted with a longer evaluation period. Fourth, demographics such as the experience of meditation, the use of mindfulness apps, and socioeconomic status, which may have acted as confounders, were not investigated. Fifth, we did not assess trait mindfulness as our focus was on emotional state. Future studies should use validated questionnaires to assess mindfulness, such as the Five Facet Mindfulness Questionnaire and the Mindfulness Attention and Awareness Scale. Last, although the set of prototypes showed good feasibility and efficacy, we could not determine which parts of the offerings, to what degree, contributed to program adherence and efficacy.

Conclusions

Given the time and frequency of GTM and the acceptability of prototypes, we believe that GTM exhibits good feasibility. Meanwhile, GTM reduced stress, improved mood, and let the practitioners feel other benefits, similar to BM. Long-term practitioners of GTM may even feel more relaxed and calmer in the state of premeditation than those who practice BM. To further improve the feasibility and consolidate the efficacy of GTM, the offerings should be improved based on user feedback and compared with other mindfulness apps in future studies with a longer evaluation period.

Acknowledgments

The authors appreciate Dr Taisyu/Hironori Kawano for his insightful comments regarding the project and for creating the original version of the audio scripts. The authors thank Mindful Mamas for creating the audio clips for guided tea meditation and btrax, Inc for creating the initial prototype of the app. The authors also appreciate Adam Johnson, Todd Thompson, Renu Joshi, Sadia Zaidi, and other colleagues at Otsuka Precision Health, Inc for further development of the app, their cooperation in the study, and editing the manuscript. The authors thank Mr Taisei Kihara for creating the matcha stick. The authors are also grateful to Dr Dawn B Beckman and the clinical staff at Biofortis Clinical Research. This study was conducted using resources provided by Otsuka Holdings Co Ltd. The sponsor provided support in the form of salaries to KY and KH but did not have any additional role in the study design, data collection and analysis, or preparation of the manuscript.

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

RT-K worked on conceptualization, methodology, investigation, and writing of the original draft. KY worked on conceptualization, creation and preparation of the matcha stick, and project administration. CMC worked on investigation, data interpretation, and writing, reviewing, and editing of the manuscript. TMB performed data curation and formal analysis. KDT administered this study. KH worked on conceptualization, methodology, project management, and writing, reviewing, and editing of the manuscript. All authors reviewed the final manuscript.

Conflicts of Interest

KY and KH belong to Otsuka Holdings Co Ltd, the owner of ekkomi center me. RT-K belongs to Otsuka Pharmaceutical Co Ltd, a subsidiary of Otsuka Holdings Co Ltd. CMC, TMB, and KDT were employees of Biofortis Clinical Research, a Contract Research Organization that received funding from Otsuka Holdings Co Ltd to conduct the study.

Multimedia Appendix 1
Screen shots of the prototype app.
[PNG File , 1015 KB - mhealth_v12i1e63078_app1.png]

Multimedia Appendix 2
An example of the audio script for guided tea meditation.
[PNG File , 190 KB - mhealth_v12i1e63078_app2.png]

Multimedia Appendix 3
Appearance of the prototype matcha stick.
[PNG File , 141 KB - mhealth_v12i1e63078_app3.png]

Multimedia Appendix 4
Download instructions for the prototype app.
[PPTX File , 3781 KB - mhealth_v12i1e63078_app4.pptx]

Multimedia Appendix 5
Instructions on how to perform breathing meditation.
[DOCX File , 15 KB - mhealth_v12i1e63078_app5.docx]

Multimedia Appendix 6
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1115 KB - mhealth_v12i1e63078_app6.pdf]

Multimedia Appendix 7
Weekly meditation time and frequency.
[DOCX File , 17 KB - mhealth_v12i1e63078_app7.docx]

Multimedia Appendix 8
Changes in mood before and after the last meditation session on day 56.
[DOCX File , 18 KB - mhealth_v12i1e63078_app8.docx]

Multimedia Appendix 9
Categorized responses to the meditation benefit survey.
[DOCX File , 22 KB - mhealth_v12i1e63078_app9.docx]

References

1. Gallup 2019 global emotions report. Gallup. URL: https://www.gallup.com/analytics/248906/gallup-global-emotions-report-2019.aspx [accessed 2024-03-28]

2. Stress In America™ 2020: a national mental health crisis. American Psychological Association. 2020. URL: https://www.apa.org/news/press/releases/stress/2020/sia-mental-health-crisis.pdf [accessed 2024-03-28]
3. Weirich B, Ramirez C. Nature: a post-pandemic prescription. Nurse Lead 2022;20(5):473-479 [FREE Full text] [doi: 10.1016/j.mnl.2022.06.003] [Medline: 35935671]
4. Peters SE, Dennerlein JT, Wagner GR, Sorensen G. Work and worker health in the post-pandemic world: a public health perspective. Lancet Public Health 2022;7(2):e188-e194. [doi: 10.1016/s2468-2667(21)00259-0]
5. Liu JYW, Man DWK, Lai FHY, Cheung TC, Cheung AK, Cheung DS, et al. A health app for post-pandemic years (HAPPY) for people with physiological and psychosocial distress during the post-pandemic era: protocol for a randomized controlled trial. Digital Health 2023;9:20552076231210725 [FREE Full text] [doi: 10.1177/20552076231210725] [Medline: 37928335]
6. O'Connor DB, Thayer JF, Vedhara K. Stress and health: a review of psychobiological processes. Annu Rev Psychol 2021;72(1):663-688. [doi: 10.1146/annurev-psych-062520-122331] [Medline: 32886587]
7. Eberth J, Sedlmeier P. The effects of mindfulness meditation: a meta-analysis. Mindfulness 2012;3(3):174-189. [doi: 10.1007/s12671-012-0101-x]
8. Khoury B, Sharma M, Rush SE, Fournier C. Mindfulness-based stress reduction for healthy individuals: a meta-analysis. J Psychosom Res 2015;78(6):519-528. [doi: 10.1016/j.jpsychores.2015.03.009] [Medline: 25818837]
9. Santorelli S. Mindfulness-Based Stress Reduction (MBSR): Standards of Practice. Worcester, MA: Mindfulness in Medicine, Health Care & Society, University of Massachusetts Medical Schoo; 2014.
10. Kabat-Zinn J. Full Catastrophe Living: Using the Wisdom of Your Body and Mind to Face Stress, Pain, and Illness. New York, NY: Bantam Dell Publishing Group; 2013.
11. Keng SL, Tan HH. Effects of brief mindfulness and loving-kindness meditation inductions on emotional and behavioral responses to social rejection among individuals with high borderline personality traits. Behav Res Ther 2018;100:44-53. [doi: 10.1016/j.brat.2017.11.005] [Medline: 29179024]
12. Moore A, Gruber T, Derose J, Malinowski P. Regular, brief mindfulness meditation practice improves electrophysiological markers of attentional control. Front Hum Neurosci 2012;6:18 [FREE Full text] [doi: 10.3389/fnhum.2012.00018] [Medline: 22363278]
13. Tapper K. Mindful eating: what we know so far. Nutr Bull 2022;47(2):168-185. [doi: 10.1111/nbu.12559] [Medline: 36045097]
14. Verma G, Araya R. The effect of meditation on psychological distress among Buddhist monks and nuns. Int J Psychiatry Med 2011;40(4):461-468. [doi: 10.2190/pm.40.4.h]
15. Fish J, Brimson J, Lynch S. Mindfulness interventions delivered by technology without facilitator involvement: what research exists and what are the clinical outcomes? Mindfulness (N Y) 2016;7(5):1011-1023 [FREE Full text] [doi: 10.1007/s12671-016-0548-2] [Medline: 27642370]
16. Keenan BJ. The Japanese tea ceremony and stress management. Holist Nurs Pract 1996;10(2):30-37. [doi: 10.1097/00004650-199601000-00005] [Medline: 8550688]
17. Kochman J, Jakubczyk K, Antoniewicz J, Mruk H, Janda K. Health benefits and chemical composition of matcha green tea: a review. Molecules 2020;26(1):85 [FREE Full text] [doi: 10.3390/molecules26010085] [Medline: 33375458]
18. Kaneko S, Kumazawa K, Masuda H, Henze A, Hofmann T. Molecular and sensory studies on the umami taste of Japanese green tea. J Agric Food Chem 2006;54(7):2688-2694. [doi: 10.1021/jf0525232] [Medline: 16569062]
19. Lyzwinski LN, Edirippulige S, Caffery L, Bambling M. Mindful eating mobile health apps: review and appraisal. JMIR Mental Health 2019;6(8):e12820 [FREE Full text] [doi: 10.2196/12820] [Medline: 31441431]
20. Lyzwinski LN, Caffery L, Bambling M, Edirippulige S. The mindfulness app trial for weight, weight-related behaviors, and stress in university students: randomized controlled trial. JMIR Mhealth Uhealth 2019;7(4):e12210. [doi: 10.2196/12210]
21. A study investigating the effects of a novel mindfulness approach on frequency and duration of meditation and self-reported stress and mood in generally healthy adults. ClinicalTrials.gov. URL: https://classic.clinicaltrials.gov/ct2/show/NCT05832645 [accessed 2024-02-28]
22. Pelekasis P, Matsouka I, Koumarianou A. Progressive muscle relaxation as a supportive intervention for cancer patients undergoing chemotherapy: a systematic review. Palliative Supp Care 2016;15(4):465-473. [doi: 10.1017/s1478951516000870]
23. Roffe L, Schmidt K, Ernst E. A systematic review of guided imagery as an adjuvant cancer therapy. Psychooncology 2005;14(8):607-617 [FREE Full text] [doi: 10.1002/pon.889] [Medline: 15651053]
24. Cohen S. Perceived stress in a probability sample of the United States. In: Spacapan S, Oskamp S, editors. The Social Psychology of Health. Thousand Oaks, CA: SAGE Publications; 1988:31-67.
25. Sakairi Y, Nakatsuka K, Shimizu T. Development of the Two Dimensional Mood Scale for self‐monitoring and self‐regulation of momentary mood states. Jpn Psychol Res 2013;55(4):338-349. [doi: 10.1111/jpr.12021]
26. Huberty J, Green J, Glissmann C, Larkey L, Puzia M, Lee C. Efficacy of the mindfulness meditation mobile app "Calm" to reduce stress among college students: randomized controlled trial. JMIR Mhealth Uhealth 2019;7(6):e14273 [FREE Full text] [doi: 10.2196/14273] [Medline: 31237569]
27. Cohen S, Janicki‐Deverts D. Who's stressed? Distributions of psychological stress in the United States in probability samples from 1983, 2006, and 2009. J App Soc Pyschol 2012 Apr 16;42(6):1320-1334. [doi: 10.1111/j.1559-1816.2012.00900.x]

28. Callahan C, Kimber J, Hu E, Tanner L, Kunkle S. The real-world impact of app-based mindfulness on headspace members with moderate and severe perceived stress: observational study. JMIR Mhealth Uhealth 2024;12:e52968-e52968 [FREE Full text] [doi: 10.2196/52968] [Medline: 38488513]
29. Taylor H, Cavanagh K, Field AP, Strauss C. Health care workers' need for headspace: findings from a multisite definitive randomized controlled trial of an unguided digital mindfulness-based self-help app to reduce healthcare worker stress. JMIR Mhealth Uhealth 2022;10(8):e31744. [doi: 10.2196/31744] [Medline: 36006668]
30. Huberty JL, Espel-Huynh HM, Neher TL, Puzia ME. Testing the pragmatic effectiveness of a consumer-based mindfulness mobile app in the workplace: randomized controlled trial. JMIR Mhealth Uhealth 2022;10(9):e38903 [FREE Full text] [doi: 10.2196/38903] [Medline: 36169991]
31. Yang E, Schamber E, Meyer RML, Gold JI. Happier healers: randomized controlled trial of mobile mindfulness for stress management. J Altern Complementary Med 2018;24(5):505-513. [doi: 10.1089/acm.2015.0301] [Medline: 29420050]
32. Militello L, Sobolev M, Okeke F, Adler DA, Nahum-Shani I. Digital prompts to increase engagement with the headspace app and for stress regulation among parents: feasibility study. JMIR Form Res 2022;6(3):e30606 [FREE Full text] [doi: 10.2196/30606] [Medline: 35311675]
33. Lu Y, Remond J, Bunting M, Ilies R, Tripathi N, Narayanan J. An app-based workplace mindfulness intervention, and its effects over time. Front Psychol 2021;12:615137 [FREE Full text] [doi: 10.3389/fpsyg.2021.615137] [Medline: 33967887]
34. Parsons CE, Madsen MA, Jensen KL, Kæseler S, Fjorback LO, Piet J, et al. Smartphone monitoring of participants' engagement with home practice during mindfulness-based stress reduction: observational study. JMIR Mental Health 2020;7(1):e14467 [FREE Full text] [doi: 10.2196/14467] [Medline: 31934873]
35. Stecher C, Sullivan M, Huberty J. Using personalized anchors to establish routine meditation practice with a mobile app: randomized controlled trial. JMIR Mhealth Uhealth 2021;9(12):e32794 [FREE Full text] [doi: 10.2196/32794] [Medline: 34941558]
36. Rung AL, Oral E, Berghammer L, Peters ES. Feasibility and acceptability of a mobile mindfulness meditation intervention among women: intervention study. JMIR Mhealth Uhealth 2020;8(6):e15943 [FREE Full text] [doi: 10.2196/15943] [Medline: 32442147]
37. Gavrilova L, Zawadzki MJ. Examining how headspace impacts mindfulness mechanisms over an 8-week app-based mindfulness intervention. Mindfulness 2023;14(9):2236-2249. [doi: 10.1007/s12671-023-02214-4]
38. van Emmerik AAP, Berings F, Lancee J. Efficacy of a mindfulness-based mobile application: a randomized waiting-list controlled trial. Mindfulness (N Y) 2018;9(1):187-198 [FREE Full text] [doi: 10.1007/s12671-017-0761-7] [Medline: 29387266]
39. Coelhoso CC, Tobo PR, Lacerda SS, Lima AH, Barrichello CRC, Amaro E, et al. A new mental health mobile app for well-being and stress reduction in working women: randomized controlled trial. J Med Internet Res 2019;21(11):e14269. [doi: 10.2196/14269] [Medline: 31697244]
40. McCallum M, Baldwin M, Thompson P, Blessing K, Frisch M, Ho A, et al. Long-term efficacy of a mobile mental wellness program: prospective single-arm study. JMIR Mhealth Uhealth 2024;12:e54634 [FREE Full text] [doi: 10.2196/54634] [Medline: 38935946]
41. Kirk U, Staiano W, Hu E, Ngnoumen C, Kunkle S, Shih E, et al. App-based mindfulness for attenuation of subjective and physiological stress reactivity in a population with elevated stress: randomized controlled trial. JMIR Mhealth Uhealth 2023;11:e47371 [FREE Full text] [doi: 10.2196/47371] [Medline: 37831493]
42. Türközü D, Şanlier N. L-theanine, unique amino acid of tea, and its metabolism, health effects, and safety. Crit Rev Food Sci Nutr 2017;57(8):1681-1687. [doi: 10.1080/10408398.2015.1016141] [Medline: 26192072]
43. Wang L, Brennan M, Li S, Zhao H, Lange KW, Brennan C. How does the tea L-theanine buffer stress and anxiety. Food Sci Hum Wellness 2022;11(3):467-475. [doi: 10.1016/j.fshw.2021.12.004]
44. Godara M, Hecht M, Singer T. Training-related improvements in mental well-being through reduction in negative interpretation bias: a randomized trial of online socio-emotional dyadic and mindfulness interventions. J Affect Disord 2024;354:662-672 [FREE Full text] [doi: 10.1016/j.jad.2024.03.037] [Medline: 38484880]

Abbreviations

BM: breathing meditation
CONSORT: Consolidated Standards of Reporting Trials
GTM: guided tea meditation
ITT: intention-to-treat
MBSR: mindfulness-based stress reduction
PSS-10: Perceived Stress Scale-10

Edited by L Buis; submitted 24.06.24; peer-reviewed by PF Lee, A Eisingerich; comments to author 05.09.24; revised version received 25.10.24; accepted 05.11.24; published 10.12.24.

Please cite as:
Tanaka-Kanegae R, Yamada K, Cook CM, Blonquist TM, Taggart KD, Hamada K
Feasibility and Efficacy of a Novel Mindfulness App Used With Matcha Green Tea in Generally Healthy Adults: Randomized Controlled Trial
JMIR Mhealth Uhealth 2024;12:e63078
URL: https://mhealth.jmir.org/2024/1/e63078
doi:10.2196/63078
PMID:

Original Paper

Functions of Smartphone Apps and Wearable Devices Promoting Physical Activity: Six-Month Longitudinal Study on Japanese-Speaking Adults

Naoki Konishi[1], PhD; Takeyuki Oba[1], PhD; Keisuke Takano[1], PhD; Kentaro Katahira[1], PhD; Kenta Kimura[1], PhD

Human Informatics and Interaction Research Institute, The National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Ibaraki, Japan

Corresponding Author:
Keisuke Takano, PhD
Human Informatics and Interaction Research Institute
The National Institute of Advanced Industrial Science and Technology (AIST)
1-1-1 Higashi
Tsukuba, Ibaraki, 305-8566
Japan
Phone: 81 50 3522 4500
Email: keisuke.takano@aist.go.jp

Abstract

Background: Smartphone apps and wearable activity trackers are increasingly recognized for their potential to promote physical activity (PA). While studies suggest that the use of commercial mobile health tools is associated with higher PA levels, most existing evidence is cross-sectional, leaving a gap in longitudinal data.

Objective: This study aims to identify app-use patterns that are prospectively associated with increases in and maintenance of PA. The primary objective was to test whether continued app use is linked to adherence to the recommended PA levels (ie, 23 metabolic equivalent task [MET] hours per week for adults or 10 MET hours/week for individuals aged >65 years) during a follow-up assessment. The secondary objective was to explore which functions and features of PA apps predict changes in PA levels.

Methods: A 2-wave longitudinal survey was conducted, with baseline and follow-up assessments separated by 6 months. A total of 20,573 Japanese-speaking online respondents participated in the baseline survey, and 16,286 (8289 women; mean age 54.7 years, SD 16.8 years) completed the follow-up. At both time points, participants reported their current PA levels and whether they were using any PA apps or wearables. Each participant was classified into 1 of the following 4 categories: continued users (those using apps at both the baseline and follow-up; n=2150, 13.20%), new users (those who started using apps before the follow-up; n=1462, 8.98%), discontinued users (those who had used apps at baseline but not at follow-up; n=1899, 11.66%), and continued nonusers (those who had never used apps; n=10,775, 66.16%).

Results: The majority of continued users (1538/2150, 71.53%) either improved or maintained their PA at the recommended levels over 6 months. By contrast, discontinued users experienced the largest reduction in PA (−7.95 MET hours/week on average), with more than half failing to meet the recommended levels at the follow-up (n=968, 50.97%). Analyses of individual app functions revealed that both energy analysis (eg, app calculation of daily energy expenditure) and journaling (eg, users manually entering notes and maintaining an exercise diary) were significantly associated with increases in PA. Specifically, energy analysis was associated with an odds ratio (OR) of 1.67 (95% CI 1.05-2.64, P=.03), and journaling had an OR of 1.76 (95% CI 1.12-2.76, P=.01). By contrast, individuals who maintained the recommended PA levels at the follow-up were more likely to use the goal setting (OR 1.73, 95% CI 1.21-2.48, P=.003), sleep information (OR 1.66, 95% CI 1.03-2.68, P=.04), and blood pressure recording (OR 2.05, 95% CI 1.10-3.83, P=.02) functions.

Conclusions: The results highlight the importance of continued app use in both increasing and maintaining PA levels. Different app functions may contribute to these outcomes, with features such as goal setting and journaling playing a key role in increasing PA, while functions related to overall health, such as sleep tracking and blood pressure monitoring, are more associated with maintaining high PA levels.

(JMIR Mhealth Uhealth 2024;12:e59708) doi:10.2196/59708

KEYWORDS

mHealth; mobile health; smartphone app; physical activity; wearable activity tracker; longitudinal design; wearable; Japan; health promotion

Introduction

Background

Physical inactivity is highly prevalent in modern society, with an average global prevalence of 31% [1]. It is associated with noncommunicable diseases such as stroke, hypertension, and diabetes [2] and is now the fourth leading risk factor for mortality [3]. To promote physical activity (PA), behavior change techniques (BCTs) have been developed [4], which can be delivered either in-person (eg, by health care professionals) or digitally (eg, through smartphones and wearable activity trackers). These digital mobile health (mHealth) interventions are expected to play a pivotal role in promoting PA, particularly during and after the COVID-19 pandemic, when in-person contact is highly restricted. These digital PA tools typically offer measurement and monitoring (eg, activity logs), information and analysis (eg, progress and individual exercise data), and support and feedback (eg, advice on PA and goal setting) [5]. The most frequently implemented BCTs include self-monitoring, providing feedback on performance, and goal-setting [6]. Observational studies have found that fitness app users are more physically active than nonusers [7], with app users having approximately twice the odds of meeting aerobic PA guidelines compared with nonusers, even during the COVID-19 pandemic [8]. One important limitation in the literature is that most studies on PA apps (and health care apps in general) have used a cross-sectional design, meaning longitudinal evidence is still lacking. We aimed to fill this gap and investigate whether, and how, daily use of PA apps and wearables (not necessarily as part of a clinical intervention or treatment) is prospectively associated with increased levels of PA.

How Effective Is an mHealth Intervention? Evidence From Clinical Trials

A large number of randomized and nonrandomized trials have been published on this topic, not limited to daily use of commercial PA apps and wearables. To our knowledge, 3 umbrella reviews have been conducted, focusing on digital interventions for improving PA. An early umbrella review [9] synthesized 11 systematic reviews and meta-analyses on eHealth or mHealth interventions targeting PA, sedentary behavior, and healthy eating for healthy individuals. The authors concluded that the majority of eHealth/mHealth interventions were reported as effective, although high heterogeneity was observed across multiple studies. Another umbrella review [10] focused on interventions using wearable activity trackers to improve PA. A synthesis of 39 systematic reviews and meta-analyses indicated a moderate effect size (standardized mean difference 0.3-0.6). A more recent umbrella review [11] identified 17 systematic reviews and meta-analyses on digital interventions specifically targeting PA and sedentary behavior to prevent or manage noncommunicable diseases. The results suggest that digital interventions have a small to moderate effect on increasing PA, although heterogeneity is documented across multiple reviews. For example, 3 systematic reviews concluded that mHealth interventions are effective, particularly those involving gamification [12], personalization [13], or delivery in workplace settings [14]. By contrast, a meta-analysis [15] found no significant effects of mobile interventions on total PA, moderate to vigorous PA, or walking. Similarly, a review of mHealth interventions equipped with social features found a nonsignificant effect on PA outcomes [16].

Researchers have also explored specific components or features of apps and wearables that are key to improving PA. Apps and digital interventions offering richer content and a larger number of BCTs are found to be more favored by users [17] and are associated with better health outcomes [18,19], although users also appreciate simplicity (eg, ease of use) [20]. Researchers have also found that certain app features and characteristics are more favored than others, such as data export, usability, and cost [20]; tracking (eg, steps, heart rate, and ovulation) [21,22]; and health information and medical reminders [23]. However, the umbrella reviews [9,11] concluded that the evidence for the effectiveness of specific BCTs or combinations of BCTs in digital PA interventions is largely mixed. In the search for successful digital implementations of BCTs, meta-regressions and systematic reviews highlighted the importance of behavioral goals and self-monitoring [24]; SMS text messaging, personalization, goal setting and planning, and graded tasks [25]; goal setting, prompts/cues, feedback on behavior, and action planning [26]; and personalized goal setting with motivational feedback [27]. By contrast, several meta-analyses reported no significant associations between intervention efficacy and the number or types of BCTs implemented [28,29].

Note that these analyses typically targeted clinical trials focused on specific populations, such as patients, older adults, and individuals with low socioeconomic status. Only a few studies have investigated how the spontaneous use of commercial apps and wearables can help improve PA in a community sample. Investigating app use in uncontrolled settings is particularly important to assess the potential efficacy of PA apps and wearables on the market. Studies have highlighted substantial differences in user behavior between controlled clinical contexts and real-world settings. For example, the average retention rate of mHealth interventions in published randomized controlled trials is about 91% [25], which is surprisingly high compared with the user engagement observed with commercial health care apps (eg, 4%, the median percentage of daily active users of mental health apps) [30]. An exceptional longitudinal study [31] investigated user engagement with a commercial app that rewards users with digital incentives for walking. The results showed that 60% of participants engaged with the app for at least 6 months. Interestingly, users who actively engaged with the app experienced larger increases in daily step count compared with less frequent users. These findings highlight the importance of continuous, long-term use of PA apps or wearables for users to fully benefit from these digital tools.

Study Overview

The general purpose of this study was to identify app-use patterns prospectively associated with increased levels of PA. To achieve this, we conducted a 2-wave longitudinal survey with a 6-month interval. At each survey point, participants reported their current levels of PA and whether they were using any PA apps or wearables. Each participant was classified into 1 of 4 categories: continued users (those who used apps or wearables at both baseline and follow-up), new users (those who started using apps before the follow-up), discontinued users (those who used apps at baseline but not at follow-up), and continued nonusers (those who had never used apps). We were particularly interested in (1) whether continued and new users would increase their PA levels or maintain high-level PA at the 6-month follow-up, and (2) whether any features of the PA apps would predict increases in PA.

Our recent cross-sectional study found that users typically engaged with a limited number of functions within an app (median 2 functions, IQR 1-4 functions). Physically active users tended to use functions such as sensor information (eg, step count and heart rate), goal setting (eg, setting a daily step goal), energy analysis (eg, estimating calories burned), journaling (eg, manually recording daily exercise), and global positioning system (GPS)/maps. We did not have a specific hypothesis regarding the prospective effects of individual app functions, so the overall analyses were conducted in an exploratory manner. However, we expected that sensor information (closely related to self-monitoring/tracking in the BCT taxonomy) and other functions implementing regulatory techniques [32] would be associated with increases in PA. Additionally, we expected that the number of functions might be linked to PA increases, as some studies have suggested that the amount of app content or the number of implemented BCTs is associated with the efficacy of mobile interventions [18,19].

Objectives

In this study, we had 2 objectives: the primary objective was to investigate whether new and continued users would increase PA or maintain high levels of PA over 6 months. The secondary objective was to identify the functions and features of PA apps that would predict increases in and maintenance of PA.

Methods

Participants and Procedure

Participants were recruited from the respondents who completed the first (baseline) survey, the results of which have been published elsewhere [33,34]. The baseline survey was conducted in 2023. Invitations to participate were sent to potential participants (residents of Japan who were registered in a database for online surveys) [33,34]. At baseline, 20,573 online respondents completed questionnaires regarding general health and health-related behaviors, including PA levels and use of mHealth apps. The only eligibility criterion was age (>18 years), with no additional criteria. Proficiency in Japanese was assumed, as the survey was written in Japanese. All participants were invited to the second (follow-up) survey, which took place approximately 6 months after the baseline survey. The follow-up survey, which included the same questions on PA and mHealth use, was completed by 16,286 of the 20,573 participants. Data from those who completed the follow-up were used for statistical analyses. Because of the online nature of the surveys, the reasons for dropout could not be tracked. Specifically, we had no means of reminding participants about the follow-up survey other than via email. Tracking dropouts was technically impossible, as they no longer responded to our emails. Participants received a small compensation for each survey (an online shopping voucher valued at approximately US $0.31).

Ethics Approval

The study protocol was approved by the Ethics Committee of the National Institute of Advanced Industrial Science and Technology (approval ID 2022-1279).

Measures

International Physical Activity Questionnaire-Short Form

At each survey, participants reported how many days and minutes per day they engaged in (1) walking, (2) moderate-intensity activity, and (3) vigorous-intensity activity over an average week [35,36]. The reported duration and frequency of each activity were then multiplied and converted into metabolic equivalent task (MET) hours per week. To determine whether each participant had a sufficient level of PA, the MET score was dichotomized to represent adherence to the national PA criteria recommended by the Ministry of Health, Labour and Welfare of Japan: 23 MET hours/week for adults aged <65 years and 10 MET hours/week for older adults [37]. Although there has been debate about dichotomization, we chose this approach because using an established cutoff helps clarify whether each participant achieved a clinically meaningful level of PA. Another advantage is that it explicitly distinguished those who maintained low or high levels of PA over time, which could not be separated using a simple numeric change score.

Stages of Change Questionnaire

Participants completed the Japanese version [38] of the Stages of Change questionnaire for PA, based on the transtheoretical model [39,40]. The questionnaire asked participants to select the most applicable statement from the following 5 statements: *I currently do not exercise and do not intend to start exercising in the future* (Precontemplation); *I currently do not exercise but I am thinking about starting to exercise in the next 6 months* (Contemplation); *I currently exercise some, but not regularly* (Preparation); *I currently exercise regularly, but have only begun doing so within the last 6 months* (Action); and *I currently exercise regularly and have done so for longer than 6 months* (Maintenance). Regular exercise was explicitly defined in the questionnaire instructions as engaging in PA for at least 20 minutes, twice or more per week.

Use of Apps and Wearables

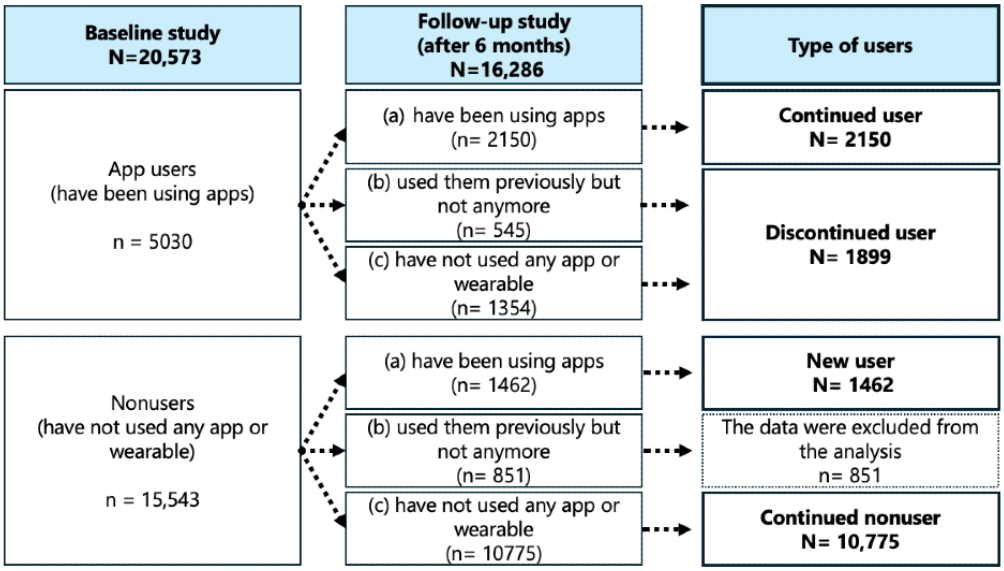
At baseline, participants provided a binary response indicating whether they used any apps or wearables to support their PA and exercise. Those who answered affirmatively were then asked for details on how they used the apps and wearables. The questions included the duration of app use, with the following

response options: <1 week, <1 month, <3 months, <6 months, <1 year, and ≥1 year. Participants were also asked about the functions and features of the apps they were using. They were presented with a list of 41 app functions (eg, sensor information, goal setting, and energy analysis) [33] and indicated any that applied to their usage [20,29]. However, as most of the listed functions were rarely used [33], we focused exclusively on the most frequently used functions for the current analysis: sensor information (eg, step count and heart rate), goal setting and progress tracking (eg, steps achieved), energy analysis (eg, estimated daily energy expenditure), weight recording, journaling (eg, manually entered diary or notes), GPS/maps, sleep information, reward points, and blood pressure recording.

At the 6-month follow-up, participants completed a similar questionnaire asking whether they were using apps and wearables. Unlike at baseline, participants were given 3 response options: (1) have been using apps and wearables for the past 6 months, (2) used them previously but no longer, and (3) have never used any app or wearable. Participants were also asked about the duration and frequency of app use (ie, how long and how often they had used/been using the app). However, questions regarding individual app functions and features were omitted due to limited space in the follow-up survey.

Responses from baseline and follow-up were interpreted as a 2×2 factorial matrix (user vs nonuser; baseline vs follow-up), classifying each participant into 4 categories (Figure 1): new users (those who began using apps before the follow-up), continued users (those who used apps or wearables at both baseline and follow-up), discontinued users (those who used apps at baseline but not at follow-up), and continued nonusers (those who never used apps). Participants who indicated at follow-up that they had used apps or wearables but were no longer using them (ie, those who selected option 2) were classified as discontinued users if they were identified as app users at baseline. If they were identified as nonusers at baseline, they were excluded from the statistical analyses for ease of interpretation. These participants were considered temporary app users—they may have used apps for a short period between baseline and follow-up but did not show significant changes in PA levels (P=.29; see Figure S1 in Multimedia Appendix 1 for details). We did not consider their usage comparable to that of the other participant groups.

Figure 1. Overview of user type classification.



Statistical Analyses

First, we explored demographic and descriptive differences between the 4 types of users based on gender, age, education level, household income, PA level, and readiness. Second, logistic regression analyses were conducted to examine how continued and discontinued app usage are associated with changes in PA levels. Two binary dependent variables were used to represent the following contrasts: (1) individuals who maintained underrecommended levels of PA (<23 or 10 METs) at the 6-month follow-up versus those who increased their PA to recommended or higher levels over time, and (2) individuals who maintained the recommended PA levels versus those who showed decreases and no longer met the recommended levels at follow-up. We also calculated simple change scores for PA (follow-up minus baseline) to clarify the magnitude of change experienced by each type of user over time. Finally, 2 independent logistic regression analyses were conducted to determine which app features were associated with changes in PA levels. The logistic regression models predicted the 2 binary dependent variables, specifically changes in (and maintenance of) adherence to the recommended PA levels, based on the individual 10 app features and functions used at baseline. All analyses were conducted using R (version 4.2.2; R Foundation) with the following specific packages: chisq.posthoc.test [41], finalfit [42], ggpubr [43], and tidyverse [44].

Results

Demographics

Table 1 shows the demographic characteristics for each type of app users (N=16,286). We identified 1462 (8.98%) new users, 2150 (13.20%) continued users, 1899 (11.66%) discontinued users, and 10,775 (66.16%) continued nonusers in the data set. A 1-way analysis of variance indicated significant age

differences between the user types (partial η^2=.009, *P*<.001), implying that new users were older than continued users and continued nonusers, and discontinued users were the youngest among the 4 types of users (*P*<.001; adjusted by the Tukey method). The descriptives per age group showed that, among older participants (eg, ≥60 years) who were identified as app users at baseline, 899 continued app use to follow-up, whereas 675 discontinued their use. Younger app users at baseline (eg, <30 years) were, however, more likely to discontinue than continue app use (n=250 vs 203). Chi-square tests revealed overall significant gender differences (*P*<.001). Residual analyses detected significant gender differences (*P*<.001) within continued users, men (1295/2150, 60.23%) versus women (855/2150, 39.77%). However, within continued nonusers, women (5770/10,775, 53.55%) were more dominant than men (5005/10,775, 46.45%; *P*<.001). Furthermore, continued users were the most prevalent among individuals with the highest household income (≥10 million JPY; 1 JPY=US $0.0065) and education level (university or above). At baseline, most continued users (1609/2150, 74.84%) reported that they had been using an app for longer than 6 months (1128/1899, 59.39%, for discontinued users). Similarly, most continued users (1654/2150, 76.93%) reported using a PA app once or more each day (1153/1899, 60.71%, for discontinued users).

Table 1. Demographic statistics of app users and nonusers.

Variable	Current app user		Current nonuser		Total (N=16,286)	*F* test or chi-square (*df*), *P* value
	Continued user (user at baseline; n=2150)	New user (nonuser at baseline; n=1462)	Discontinued user (user at baseline; n=1899)	Continued nonuser (nonuser at baseline; n=10,775)		
Age (years), mean (SD)	53.8 (16.3)	57.3 (16.8)	50.9 (17.3)	55.2 (16.7)	54.7 (16.8)	50.16[a] (3, 16,282), <.001
Age (years; categorical), n (%)						161.86 (9), <.001
<30	203 (9.44)	126 (8.62)	250 (13.16)	965 (8.96)	1544 (9.48)	
30-44	452 (21.02)	226 (15.46)	502 (26.43)	2093 (19.42)	3273 (20.10)	
45-59	596 (27.72)	333 (22.78)	472 (24.86)	2859 (26.53)	4260 (26.16)	
≥60	899 (41.81)	777 (53.15)	675 (35.55)	4858 (45.09)	7209 (44.27)	
Women, n (%)	855 (39.77)	749 (51.23)	915 (48.18)	5770 (53.55)	8289 (50.90)	142.56 (3), <.001
BMI, mean (SD)	22.4 (3.5)	22.3 (3.9)	22.3 (3.8)	22.0 (3.7)	22.1 (3.7)	7.27[a] (3, 16,282), <.001
Married, n (%)	1497 (69.63)	964 (65.94)	1213 (63.88)	6846 (63.54)	10,520 (64.60)	30.68 (3), <.001
Child/children[b], n (%)	1412 (65.67)	984 (67.31)	1211 (63.77)	6782 (62.94)	10,389 (63.79)	14.48 (3), .002
Education level, n (%)						231.79 (12), <.001
Middle school	27 (1.26)	36 (2.46)	38 (2.00)	290 (2.69)	391 (2.40)	
High school	517 (24.05)	453 (30.98)	531 (27.96)	3546 (32.91)	5047 (30.99)	
College or vocational school	382 (17.77)	287 (19.63)	441 (23.22)	2572 (23.87)	3682 (22.61)	
University or above	1210 (56.28)	673 (46.03)	864 (45.50)	4293 (39.84)	7040 (43.23)	
Other	14 (0.65)	13 (0.89)	25 (1.32)	74 (0.69)	126 (0.77)	
Job, n (%)	27 (1.26)	36 (2.46)	38 (2.00)	290 (2.69)	391 (2.40)	146.85 (3), (<.001)
Household income, n (%)						335.43 (15), (<.001)
<3 million JPY[c]	322 (14.98)	320 (21.89)	416 (21.91)	2487 (23.08)	3545 (21.77)	
3-5 million JPY	496 (23.07)	350 (23.94)	455 (23.96)	2656 (24.65)	3957 (24.30)	
5-7 million JPY	365 (16.98)	228 (15.60)	295 (15.53)	1550 (14.39)	2438 (14.97)	
7-10 million JPY	360 (16.74)	179 (12.24)	236 (12.43)	1135 (10.53)	1910 (11.73)	
≥10 million JPY	314 (14.60)	115 (7.87)	183 (9.64)	701 (6.51)	1313 (8.06)	
No response	293 (13.63)	270 (18.47)	314 (16.54)	2246 (20.84)	3123 (19.18)	
Physical activity (baseline, MET hours/week), median (IQR)	34.6 (16.5-67.1)	23.3 (9.9-49.6)	23.3 (7.7-53.5)	11.6 (0.6-33.0)	16.5 (3.3-41.7)	1239 (3), <.001
Physical activity (follow-up, MET hours/week[d]), median (IQR)	34.0 (16.2-65.2)	24.2 (11.6-51.7)	16.5 (2.5-43.1)	9.9 (0.0-29.7)	14.8 (1.1-38.1)	1419.6 (3), <.001

[a]*F* test.

[b]Data are from a follow-up survey.

[c]1 JPY=US $0.0065.

[d]Metabolic equivalent task hours per week.

Relationship Between App Use and Adherence to the Recommended PA Level

Table 2 illustrates changes in PA levels for each user type over 6 months. Continued nonusers generally maintained underrecommended levels of PA over time (5259/10,775, 48.81%). Among the 4 types of users, new users had the largest proportion of individuals who increased their PA to recommended levels (178/1462, 12.18%), although achieving these levels was uncommon in the current sample. Continued

users typically maintained the recommended PA levels over time (1327/2150, 61.72%), while discontinued users had the largest proportion of individuals who failed to adhere to the recommended levels at follow-up (330/1899, 17.38%).

Table 2. The number of participants who increased, decreased, or maintained physical activity over 6 months as a function of the app user types.a

PA change category (baseline → follow-up)	Current user		Current nonuser	
	Continued user (n=2150), n (%)	New user (n=1462), n (%)	Discontinued user (n=1899), n (%)	Continued nonuser (n=10,775), n (%)
Maintained underrecommended level (not adhered → not adhered)	390 (18.14)	370 (25.31)	638 (33.60)	5259 (48.81)
Increased (not adhered → adhered)	211 (9.81)	178 (12.18)	159 (8.37)	931 (8.64)
Decreased (adhered → not adhered)	222 (10.33)	144 (9.85)	330 (17.38)	1269 (11.78)
Maintained recommended level (adhered → adhered)	1327 (61.72)	770 (52.67)	772 (40.65)	3316 (30.77)

[a]Adherence to the recommended physical activity level is equal to or larger than 23 metabolic equivalent task hours per week for adults or 10 metabolic equivalent task hours per week for older adults aged ≥65 years.

Logistic regression analyses were performed to examine how the 4 types of users were associated with adherence to the recommended PA levels (23 or 10 MET hours/week) over 6 months (primary objective). Discontinued users were used as the reference group, as they showed the largest proportion of individuals who decreased their PA to underrecommended levels at follow-up. Results (Table 3) showed that, compared with discontinued users, continued users (odds ratio [OR] 2.171, 95% CI 1.71-2.76, *P*<.001) and new users (OR 1.93, 95% CI 1.50-2.48, *P*<.001) were more likely to increase PA to the recommended levels at follow-up. Continued nonusers, compared with discontinued users, were more likely to maintain underrecommended PA levels at follow-up (OR 0.71, 95% CI 0.59-0.86, *P*<.001). Another logistic regression analysis showed that continued users (OR 2.56, 95% CI 2.11-3.10, *P*<.001) and new users (OR 2.29, 95% CI 1.84-2.85, *P*<.001) were more likely to maintain the recommended PA levels compared with discontinued users. Continued nonusers did not significantly differ from discontinued users (OR 1.12, 95% CI 0.97-1.29, *P*=.11).

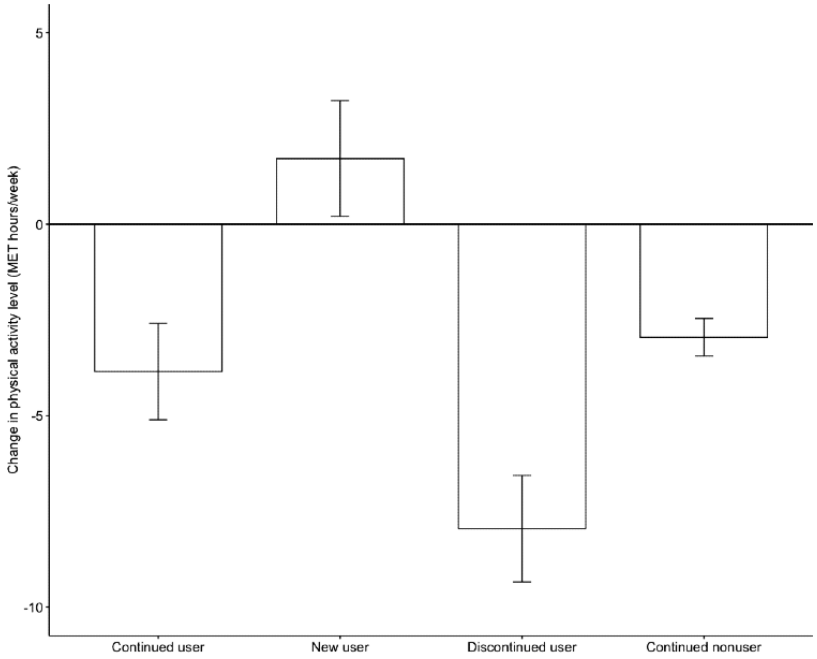
Table 3. Logistic regressions predicting physical activity changes based on 23 (or 10) metabolic equivalent task hours per week.

Outcome and predictor	Odds ratio (95% CI)	*P* value
Increased to versus maintained at below the recommended level (n=8136)		
Intercept (reference: discontinued user)	0.249 (0.209-0.297)	<.001
Continued user	2.171 (1.705-2.763)	<.001
New user	1.930 (1.504-2.477)	<.001
Continued nonuser	0.710 (0.589-0.857)	<.001
Maintained the recommended level versus decreased (n=8150)		
Intercept (reference: discontinued user)	2.339 (2.056-2.661)	<.001
Continued user	2.555 (2.109-3.096)	<.001
New user	2.286 (1.835-2.847)	<.001
Continued nonuser	1.117 (0.967-1.290)	.133

Changes in PA Level at the 6-Month Follow-Up

We then calculated the simple change scores for PA levels (ie, follow-up minus baseline in MET hours/week) to estimate the changes experienced by each user type over 6 months (Figure 2). New users were the only group to show increases in PA levels (mean 1.71, SD 57.76), which was significantly larger than the changes (decreases) observed in continued nonusers (mean –2.95, SD 50.76; $t_{1,780.5}$= 2.94, *P*=.003). Continued users showed a decrease in PA on average (mean –3.85, SD 58.53), while discontinued users exhibited even larger decreases (mean –7.95, SD 60.52; $t_{3,949.1}$=2.19, *P*=.03).

Figure 2. Change in physical activity level (in METs hours/week) at the 6-month follow-up among the 4 types of users and nonusers. The error bar indicates the SE. MET: metabolic equivalent of tasks.



Function-Wise Analyses Predicting Changes in PA

At baseline, app users most frequently reported using sensor information, followed by goal setting, goal progress, energy analysis, and weight recording (see Table S1 in Multimedia Appendix 1 for details). To explore which app functions are associated with increases in or maintenance of PA levels (secondary objective), we estimated 2 logistic regression models. These analyses targeted continued users exclusively (ie, individuals who reported using apps at both baseline and follow-up), with nonusers excluded. The results (Table 4) showed that energy analysis and journaling were significantly associated with increases in PA to the recommended levels (OR 1.67, 95% CI 1.05-2.64, *P*=.03 and OR 1.76, 95% CI 1.12-2.76, *P*=.01, respectively). Maintenance of the recommended levels (vs a decrease to underrecommended levels) was predicted by goal setting (OR 1.73, 95% CI 1.21-2.48, *P*=.003), sleep information (OR 1.66, 95% CI 1.03-2.68, *P*=.04), and blood pressure recording (OR 2.05, 95% CI 1.10-3.83, *P*=.02). We also tested the association between the number of app functions in use (at baseline) and changes in PA level (simple change score, ie, follow-up minus baseline) among continued users, which did not reach statistical significance (*r*=–0.02, *P*=.48).

Table 4. Logistic regression predicting physical activity change based on 23 metabolic equivalent task hours per week.

Outcome and predictor	OR (95% CI)	*P* value
Increased to versus maintained at below the recommended level (n=601)		
Show sensor info	1.008 (0.694-1.464)	.97
Goal setting	1.005 (0.649-1.557)	.98
Show goal progress	1.243 (0.789-1.957)	.35
Energy analysis	1.665 (1.051-2.637)	.03
Weight recording	0.785 (0.474-1.299)	.34
Journaling	1.755 (1.116-2.760)	.01
Global positioning system/maps	0.922 (0.567-1.500)	.74
Show sleep info	0.991 (0.624-1.573)	.97
Reward points	1.350 (0.790-2.308)	.27
Blood pressure recording	1.574 (0.795-3.117)	.19
Maintained the recommended level versus decreased (n=1549)		
Show sensor info	1.200 (0.878-1.640)	.25
Goal setting	1.729 (1.206-2.480)	.003
Show goal progress	1.307 (0.900-1.897)	.16
Energy analysis	0.975 (0.672-1.416)	.90
Weight recording	1.164 (0.766-1.770)	.48
Journaling	1.086 (0.748-1.575)	.66
Global positioning system/maps	1.484 (0.963-2.287)	.07
Show sleep info	1.658 (1.025-2.684)	.04
Reward points	1.127 (0.701-1.812)	.62
Blood pressure recording	2.046 (1.095-3.825)	.02

Discussion

Principal Findings

This study investigated whether continued use of commercial PA apps and wearables over 6 months contributes to increasing or maintaining PA levels. Overall, the results emphasize the importance of continued app use in maintaining PA levels. Most continued users (1538/2150, 71.53%) either maintained or improved their PA to the recommended levels over 6 months, whereas 51% (968/1899) of discontinued users failed to meet the recommended levels at follow-up (compared with 797/1899, 41.96%, who were nonadherent at baseline). New users were found to experience the largest increase in PA levels, while discontinued users showed the largest reduction among the 4 types of users. These results suggest that individuals who recently started using apps and wearables saw the greatest improvement in their PA levels. By contrast, continued app use helped maintain PA levels (despite slight decreases), whereas discontinuation led to a substantial reduction in PA levels, equivalent to a decrease of more than 1 hour of vigorous PA per week [45].

Characteristics of Continued Versus Discontinued Users

More than half (2150/4049, 53.09%) of the app users identified at baseline reported continuing to use the app at the 6-month follow-up. While this retention rate may seem high compared with the reported daily engagement rates for health care apps [30], it is comparable to the 60% active user engagement rate observed in a longitudinal study of a commercial PA app over 6 months [31]. Our analysis of the demographic characteristics of continued (vs discontinued) users indicated that continued users were older, more likely to be men, and had higher education levels and incomes. Previous cross-sectional studies have found that mHealth/eHealth users tend to be younger, more educated, and have higher (digital) health literacy than nonusers [46-48]. Educational attainment is thought to reflect both literacy and skills (including confidence with digital and smart devices) as well as social norms related to the perceived value of health [47]. In general, women are the dominant users of health care apps (for diet, nutrition, and self-care), while fitness apps tend to be more popular among men [46]. Older adults typically avoid new technologies and mHealth services [49]. However, our results showed that continued users were more prevalent than discontinued users among older participants, suggesting that older users were more likely to continue using apps. We do not have data to readily explain this unexpected finding.

However, given that most of the continued users in our data had already been using PA apps for a long time (>6 months) at baseline, even older users may have developed high self-efficacy and perceived ease of use, which could reduce technology anxiety. Studies have identified various facilitators and barriers to technology adoption among older adults (eg, personal experiences and subjective norms) [50], which could serve as a basis for future research to explore how older users successfully adapt and integrate mHealth tools into their daily routines.

App Functions Predictive of an Increase in PA

We found that energy analysis and journaling were predictive of increases in PA to the recommended levels over 6 months. Additionally, goal setting, sleep information, and blood pressure recording were commonly used by individuals who maintained the recommended PA levels at follow-up. In a previous cross-sectional analysis, we reported associations between PA levels and individual app functions [33], which indicated that individuals with health-enhancing PA levels typically used functions such as sensor information (eg, step count and heart rate), goal setting, goal progress, energy analysis, journaling, and GPS/maps. The current prospective analyses emphasize the particular importance of energy analysis and journaling in improving PA over time. While sensor information was commonly used by PA app users (Table S1 in Multimedia Appendix 1), the findings suggest that, in addition to the automatically recorded PA data (eg, step count), users may benefit from additional analyses of physiological data (eg, energy expenditure calculations) and more deliberate engagement with the app, such as journaling and manually logging daily exercise and PA.

Interestingly, maintenance of the recommended PA level was associated with the use of sleep information, blood pressure recording, and goal setting. This suggests that users who are already sufficiently active may value functions that support general health care, rather than those focused solely on fitness and exercise. It is also possible that some users were prompted to use PA apps due to specific health concerns. Published meta-analytic studies have identified key app components that enhance PA, including self-monitoring, goal setting and planning, prompts/cues, feedback on behavior, and action planning [24-27], most of which are specifically designed to support PA. It may be important for future research to broadly explore app functions and features (not limited to PA-related functions) to identify more effective combinations, especially when the focus is on maintaining rather than increasing PA levels. This could reinforce the usefulness of the stages of change model [39,51] (eg, to guide the best interventions for those in the action or maintenance stages) and highlight the importance of tailoring digital behavior interventions.

Another interesting finding from this analysis is that the number of app functions reported as being in use was not significantly associated with increases in PA over 6 months. While several studies have suggested that the amount of app content or the number of implemented BCTs is linked to the efficacy of mHealth interventions [18,19], this association has not always been replicated [29]. As our analyses utilized self-report data rather than actual logs of user behavior, we cannot exclude the possibility that participants may not have accurately or exhaustively reported all the functions they used. However, our findings suggest that (1) users may not be fully aware of every function available in an app (or, at least, they do not consciously use them all), and (2) they may not necessarily benefit from multifunctionality. Instead, a limited number of functions (eg, goal setting and journaling) may be more effective in improving PA. Indeed, it is known that users appreciate the simplicity of an app [20], and as Michie et al [32] found, interventions that combine self-monitoring with at least one regulatory technique (eg, goal setting) can form the most cost-effective, minimal set of interventions.

Limitations

The results reported here should be interpreted with caution due to several important limitations. First, we targeted Japanese-speaking adults exclusively, which may limit the generalizability of the findings. The apps and products available on the Japanese eHealth/mHealth market may differ from those in other regions and countries. While there are similarities in user behavior between Japan and Western countries, exploring country- or culture-specific aspects would be an interesting direction for future research. Second, we cannot rule out the possibility of sampling bias. As reported elsewhere [33], the current sample exhibited higher PA levels than the general population in Japan, likely because the study was advertised as a survey on PA and health. Additionally, attrition could introduce bias, as some participants (4287/20,573, 20.83%) dropped out by the follow-up. Third, the follow-up survey did not include questions regarding how participants used specific app functions and features (assessed only at baseline). As this study was part of a larger project, there was a limit on the length of each survey. Future research should explore how changes in PA influence the use patterns of individual app functions, which could provide insights into how tailoring and personalization can be incorporated during app use adaptation. Fourth, we relied exclusively on self-reported PA, which may not always align with objective measures, such as accelerometers, due to self-reporting bias and other assessment artifacts. Similarly, user behavior and individual app function usage can be monitored automatically or made publicly available (eg, [52]). However, a downside of such an approach is that the analysis would be limited to a specific app or platform, potentially sacrificing the generalizability of the results. Finally, we cannot rule out the possibility of selection bias. As the surveys were administered online, participants were likely familiar with the internet and possibly mobile technology as well. This could explain the unexpected finding that older individuals were more likely to continue using the app.

Conclusions

This study demonstrated that continued use of apps and wearables contributes to both increasing and maintaining PA levels over 6 months. The results also revealed that app features associated with increases in PA differ from those linked to the maintenance of PA, highlighting the importance of tailoring apps to users' PA levels and readiness. We believe these findings make a meaningful contribution to the literature by highlighting

the continued use of apps and wearables as key factors in enhancing and maintaining high levels of PA. Additionally, it is noteworthy that more than half of the users continued using the apps through the 6-month follow-up, despite the poor retention rates and barriers commonly reported in the literature (eg, [53,54]). This may suggest that attitudes toward and the acceptability of mHealth apps are changing, with the COVID-19 pandemic potentially serving as an opportunity. Health care practitioners could increasingly rely on app-based approaches in their intervention repertoires, although integrating mHealth into routine practice remains a challenge [55]. Unfortunately, we did not assess the barriers preventing users from continuous engagement (eg, [56,57]), and these should be explored in future research to identify effective strategies for maintaining active user engagement.

Acknowledgments

This study was supported by JSPS KAKENHI (grant 23H01051).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Changes in physical activity level among temporary users and use of individual app functions.
[DOCX File , 83 KB - mhealth_v12i1e59708_app1.docx]

References

1. Hallal PC, Andersen LB, Bull FC, Guthold R, Haskell W, Ekelund U, Lancet Physical Activity Series Working Group. Global physical activity levels: surveillance progress, pitfalls, and prospects. Lancet 2012 Jul 21;380(9838):247-257. [doi: 10.1016/S0140-6736(12)60646-1] [Medline: 22818937]
2. Rhodes RE, Janssen I, Bredin SSD, Warburton DER, Bauman A. Physical activity: health impact, prevalence, correlates and interventions. Psychol Health 2017 Aug;32(8):942-975. [doi: 10.1080/08870446.2017.1325486] [Medline: 28554222]
3. World Health Organization. Global recommendations on physical activity for health. WHO. Geneva, Switzerland: WHO; 2010 Jan 1. URL: https://www.who.int/publications/i/item/9789241599979 [accessed 2024-12-02]
4. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95 [FREE Full text] [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
5. Mollee JS, Middelweerd A, Kurvers RL, Klein MCA. What technological features are used in smartphone apps that promote physical activity? A review and content analysis. Pers Ubiquit Comput 2017 Jul 5;21(4):633-643. [doi: 10.1007/s00779-017-1023-3]
6. Middelweerd A, Mollee JS, van der Wal CN, Brug J, Te Velde SJ. Apps to promote physical activity among adults: a review and content analysis. Int J Behav Nutr Phys Act 2014 Jul 25;11:97 [FREE Full text] [doi: 10.1186/s12966-014-0097-9] [Medline: 25059981]
7. Barkley JE, Lepp A, Santo A, Glickman E, Dowdell B. The relationship between fitness app use and physical activity behavior is mediated by exercise identity. Computers in Human Behavior 2020 Jul;108:106313. [doi: 10.1016/j.chb.2020.106313]
8. Tong HL, Maher C, Parker K, Pham TD, Neves AL, Riordan B, et al. The use of mobile apps and fitness trackers to promote healthy behaviors during COVID-19: a cross-sectional survey. PLOS Digit Health 2022 Aug;1(8):e0000087 [FREE Full text] [doi: 10.1371/journal.pdig.0000087] [Medline: 36812578]
9. Fiedler J, Eckert T, Wunsch K, Woll A. Key facets to build up eHealth and mHealth interventions to enhance physical activity, sedentary behavior and nutrition in healthy subjects - an umbrella review. BMC Public Health 2020 Oct 23;20(1):1605 [FREE Full text] [doi: 10.1186/s12889-020-09700-7] [Medline: 33097013]
10. Ferguson T, Olds T, Curtis R, Blake H, Crozier AJ, Dankiw K, et al. Effectiveness of wearable activity trackers to increase physical activity and improve health: a systematic review of systematic reviews and meta-analyses. Lancet Digit Health 2022 Aug;4(8):e615-e626 [FREE Full text] [doi: 10.1016/S2589-7500(22)00111-X] [Medline: 35868813]
11. Mair JL, Salamanca-Sanabria A, Augsburger M, Frese BF, Abend S, Jakob R, et al. Effective behavior change techniques in digital health interventions for the prevention or management of noncommunicable diseases: an umbrella review. Ann Behav Med 2023 Sep 13;57(10):817-835 [FREE Full text] [doi: 10.1093/abm/kaad041] [Medline: 37625030]
12. Xu L, Shi H, Shen M, Ni Y, Zhang X, Pang Y, et al. The effects of mHealth-based gamification interventions on participation in physical activity: systematic review. JMIR Mhealth Uhealth 2022 Feb 03;10(2):e27794 [FREE Full text] [doi: 10.2196/27794] [Medline: 35113034]

13. Davis A, Sweigart R, Ellis R. A systematic review of tailored mHealth interventions for physical activity promotion among adults. Transl Behav Med 2020 Oct 12;10(5):1221-1232 [FREE Full text] [doi: 10.1093/tbm/ibz190] [Medline: 33044542]
14. Buckingham SA, Williams AJ, Morrissey K, Price L, Harrison J. Mobile health interventions to promote physical activity and reduce sedentary behaviour in the workplace: a systematic review. Digit Health 2019;5:2055207619839883 [FREE Full text] [doi: 10.1177/2055207619839883] [Medline: 30944728]
15. Direito A, Carraça E, Rawstorn J, Whittaker R, Maddison R. mHealth technologies to influence physical activity and sedentary behaviors: behavior change techniques, systematic review and meta-analysis of randomized controlled trials. Ann Behav Med 2017 Apr;51(2):226-239. [doi: 10.1007/s12160-016-9846-0] [Medline: 27757789]
16. Tong HL, Laranjo L. The use of social features in mobile health interventions to promote physical activity: a systematic review. NPJ Digit Med 2018;1:43 [FREE Full text] [doi: 10.1038/s41746-018-0051-3] [Medline: 31304323]
17. Paganini S, Terhorst Y, Sander LB, Catic S, Balci S, Küchler AM, et al. Quality of physical activity apps: systematic search in app stores and content analysis. JMIR Mhealth Uhealth 2021 Jun 09;9(6):e22587 [FREE Full text] [doi: 10.2196/22587] [Medline: 34106073]
18. Van Rhoon L, Byrne M, Morrissey E, Murphy J, McSharry J. A systematic review of the behaviour change techniques and digital features in technology-driven type 2 diabetes prevention interventions. Digit Health 2020;6:2055207620914427 [FREE Full text] [doi: 10.1177/2055207620914427] [Medline: 32269830]
19. de Leeuwerk ME, Bor P, van der Ploeg HP, de Groot V, van der Schaaf M, van der Leeden M, OPRAH consortium. The effectiveness of physical activity interventions using activity trackers during or after inpatient care: a systematic review and meta-analysis of randomized controlled trials. Int J Behav Nutr Phys Act 2022 May 23;19(1):59 [FREE Full text] [doi: 10.1186/s12966-022-01261-9] [Medline: 35606852]
20. Mendiola MF, Kalnicki M, Lindenauer S. Valuable features in mobile health apps for patients and consumers: content analysis of apps and user ratings. JMIR Mhealth Uhealth 2015 May 13;3(2):e40 [FREE Full text] [doi: 10.2196/mhealth.4283] [Medline: 25972309]
21. Peng W, Kanthawala S, Yuan S, Hussain SA. A qualitative study of user perceptions of mobile health apps. BMC Public Health 2016 Nov 14;16(1):1158 [FREE Full text] [doi: 10.1186/s12889-016-3808-0] [Medline: 27842533]
22. Amer SA, Bahumayim A, Shah J, Aleisa N, Hani BM, Omar DI. Prevalence and determinants of mobile health applications usage: a national descriptive study. Front Public Health 2022;10:838509 [FREE Full text] [doi: 10.3389/fpubh.2022.838509] [Medline: 35570952]
23. Xie Z, Nacioglu A, Or C. Prevalence, demographic correlates, and perceived impacts of mobile health app use amongst Chinese adults: cross-sectional survey study. JMIR Mhealth Uhealth 2018 Apr 26;6(4):e103 [FREE Full text] [doi: 10.2196/mhealth.9002] [Medline: 29699971]
24. Eckerstorfer LV, Tanzer NK, Vogrincic-Haselbacher C, Kedia G, Brohmer H, Dinslaken I, et al. Key elements of mHealth interventions to successfully increase physical activity: meta-regression. JMIR Mhealth Uhealth 2018 Nov 12;6(11):e10076 [FREE Full text] [doi: 10.2196/10076] [Medline: 30425028]
25. Laranjo L, Ding D, Heleno B, Kocaballi B, Quiroz JC, Tong HL, et al. Do smartphone applications and activity trackers increase physical activity in adults? Systematic review, meta-analysis and metaregression. Br J Sports Med 2021 Apr;55(8):422-432. [doi: 10.1136/bjsports-2020-102892] [Medline: 33355160]
26. Hardeman W, Houghton J, Lane K, Jones A, Naughton F. A systematic review of just-in-time adaptive interventions (JITAIs) to promote physical activity. Int J Behav Nutr Phys Act 2019 Apr 03;16(1):31 [FREE Full text] [doi: 10.1186/s12966-019-0792-7] [Medline: 30943983]
27. Triantafyllidis A, Filos D, Claes J, Buys R, Cornelissen V, Kouidi E, et al. Computerised decision support in physical activity interventions: a systematic literature review. Int J Med Inform 2018 Mar;111:7-16. [doi: 10.1016/j.ijmedinf.2017.12.012] [Medline: 29425636]
28. Western MJ, Armstrong MEG, Islam I, Morgan K, Jones UF, Kelson MJ. The effectiveness of digital interventions for increasing physical activity in individuals of low socioeconomic status: a systematic review and meta-analysis. Int J Behav Nutr Phys Act 2021 Nov 09;18(1):148 [FREE Full text] [doi: 10.1186/s12966-021-01218-4] [Medline: 34753490]
29. Iribarren SJ, Akande TO, Kamp KJ, Barry D, Kader YG, Suelzer E. Effectiveness of mobile apps to promote health and manage disease: systematic review and meta-analysis of randomized controlled trials. JMIR Mhealth Uhealth 2021 Jan 11;9(1):e21563 [FREE Full text] [doi: 10.2196/21563] [Medline: 33427672]
30. Baumel A, Muench F, Edan S, Kane JM. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. J Med Internet Res 2019 Sep 25;21(9):e14567 [FREE Full text] [doi: 10.2196/14567] [Medline: 31573916]
31. Mitchell M, Lau E, White L, Faulkner G. Commercial app use linked with sustained physical activity in two Canadian provinces: a 12-month quasi-experimental study. Int J Behav Nutr Phys Act 2020 Feb 25;17(1):24 [FREE Full text] [doi: 10.1186/s12966-020-00926-7] [Medline: 32098625]
32. Michie S, Abraham C, Whittington C, McAteer J, Gupta S. Effective techniques in healthy eating and physical activity interventions: a meta-regression. Health Psychol 2009 Nov;28(6):690-701. [doi: 10.1037/a0016136] [Medline: 19916637]

33. Oba T, Takano K, Katahira K, Kimura K. Use patterns of smartphone apps and wearable devices supporting physical activity and exercise: large-scale cross-sectional survey. JMIR Mhealth Uhealth 2023 Nov 22;11:e49148-e49148 [FREE Full text] [doi: 10.2196/49148] [Medline: 37997790]
34. Oba T, Takano K, Katahira K, Kimura K. Revisiting the transtheoretical model for physical activity: a large-scale cross-sectional study on Japanese-speaking adults. Ann Behav Med 2024 Feb 10;58(3):167-178 [FREE Full text] [doi: 10.1093/abm/kaad069] [Medline: 38166169]
35. Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Ainsworth BE, et al. International physical activity questionnaire: 12-country reliability and validity. Med Sci Sports Exerc 2003 Aug;35(8):1381-1395. [doi: 10.1249/01.MSS.0000078924.61453.FB] [Medline: 12900694]
36. Murase N, Katsumura T, Ueda C, Inoue S, Shimomitsu T. Validity and reliability of Japanese version of International Physical Activity Questionnaire. Journal of Health and Welfare Statistics 2002 Oct 1;49(11):1-9 [FREE Full text]
37. Ministry of Health, Labour and Welfare. Physical Activity Standards for Health Promotion. Ministry of Health, Labour and Welfare. 2013. URL: https://www.e-healthnet.mhlw.go.jp/information/policy/guidelines_2013.html [accessed 2023-01-23]
38. Oka K. [Stages of change for exercise behavior and self-efficacy for exercise among middle-aged adults]. Nihon Koshu Eisei Zasshi 2003 Mar;50(3):208-215. [Medline: 12704833]
39. Marcus BH, Simkin LR. The stages of exercise behavior. J Sports Med Phys Fitness 1993 Mar;33(1):83-88. [Medline: 8350613]
40. Reed GR, Velicer WF, Prochaska JO, Rossi JS, Marcus BH. What makes a good staging algorithm: examples from regular exercise. Am J Health Promot 1997;12(1):57-66. [doi: 10.4278/0890-1171-12.1.57] [Medline: 10170436]
41. Ebbert D. A post hoc analysis for Pearson's chi-squared test for count data. Comprehensive R Archive Network. 2019 Oct 21. URL: https://cran.r-project.org/web/packages/chisq.posthoc.test/vignettes/chisq_posthoc_test.html [accessed 2024-12-02]
42. Harrison E, Drake T, Pius R. finalfit: quickly create elegant regression results tables and plots when modelling. GitHub. 2024 Jul 24. URL: https://github.com/ewenharrison/finalfit [accessed 2024-12-02]
43. Kassambara A. ggpubr: 'ggplot2' based publication ready plots. The Comprehensive R Archive Network. 2023 Feb 10. URL: https://cran.r-project.org/web/packages/ggpubr/index.html [accessed 2024-12-02]
44. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, et al. Welcome to the Tidyverse. JOSS 2019 Nov;4(43):1686. [doi: 10.21105/joss.01686]
45. Haskell WL, Lee IM, Pate RR, Powell KE, Blair SN, Franklin BA, American College of Sports Medicine, American Heart Association. Physical activity and public health: updated recommendation for adults from the American College of Sports Medicine and the American Heart Association. Circulation 2007 Aug 28;116(9):1081-1093. [doi: 10.1161/CIRCULATIONAHA.107.185649] [Medline: 17671237]
46. Bol N, Helberger N, Weert JCM. Differences in mobile health app use: a source of new digital inequalities? The Information Society 2018 Apr 26;34(3):183-193. [doi: 10.1080/01972243.2018.1438550]
47. Carroll JK, Moorhead A, Bond R, LeBlanc WG, Petrella RJ, Fiscella K. Who uses mobile phone health apps and does use matter? A secondary data analytics approach. J Med Internet Res 2017 Apr 19;19(4):e125 [FREE Full text] [doi: 10.2196/jmir.5604] [Medline: 28428170]
48. Kontos E, Blake KD, Chou WYS, Prestin A. Predictors of eHealth usage: insights on the digital divide from the Health Information National Trends Survey 2012. J Med Internet Res 2014 Jul 16;16(7):e172 [FREE Full text] [doi: 10.2196/jmir.3117] [Medline: 25048379]
49. Cao L, Chongsuvivatwong V, McNeil EB. The sociodemographic digital divide in mobile health app use among clients at outpatient departments in Inner Mongolia, China: cross-sectional survey study. JMIR Hum Factors 2022 May 19;9(2):e36962 [FREE Full text] [doi: 10.2196/36962] [Medline: 35587367]
50. Zhang M. Older people's attitudes towards emerging technologies: a systematic literature review. Public Underst Sci 2023 Nov;32(8):948-968 [FREE Full text] [doi: 10.1177/09636625231171677] [Medline: 37204075]
51. Prochaska JO, Velicer WF. The transtheoretical model of health behavior change. Am J Health Promot 1997;12(1):38-48. [doi: 10.4278/0890-1171-12.1.38] [Medline: 10170434]
52. Su L, Anderson PL. User behavior of a publicly available, free-to-use, self-guided mHealth app for depression: observational study in a global sample. JMIR Form Res 2022 Oct 25;6(10):e35538 [FREE Full text] [doi: 10.2196/35538] [Medline: 36282559]
53. Alam MZ, Hu W, Kaium MA, Hoque MR, Alam MMD. Understanding the determinants of mHealth apps adoption in Bangladesh: a SEM-Neural network approach. Technology in Society 2020 May;61:101255. [doi: 10.1016/j.techsoc.2020.101255]
54. Fox G, Connolly R. Mobile health technology adoption across generations: narrowing the digital divide. Information Systems Journal 2018 Jan 29;28(6):995-1019. [doi: 10.1111/isj.12179]
55. Alsahli S, Hor SY, Lam M. Factors influencing the acceptance and adoption of mobile health apps by physicians during the COVID-19 pandemic: systematic review. JMIR Mhealth Uhealth 2023 Nov 08;11:e50419 [FREE Full text] [doi: 10.2196/50419] [Medline: 37938873]
56. Van Dijk JAGM. Digital divide: impact of access. In: Rössler P, Hoffner CA, Zoonen L, editors. The International Encyclopedia of Media Effects (1st Edition). West Sussex, UK: Wiley Blackwell; May 08, 2017:1-11.

57. Carroll LN, Smith SA, Thomson NR. Parents as teachers health literacy demonstration project: integrating an empowerment model of health literacy promotion into home-based parent education. Health Promot Pract 2015 Mar;16(2):282-290. [doi: 10.1177/1524839914538968] [Medline: 24957219]

Abbreviations

BCT: behavior change technique
GPS: global positioning system
MET: metabolic equivalent task
mHealth: mobile health
OR: odds ratio
PA: physical activity

Edited by G Badicu; submitted 19.04.24; peer-reviewed by A Jain, L Malisoux, O Taiwo; comments to author 13.07.24; revised version received 02.09.24; accepted 19.10.24; published 10.12.24.

Please cite as:
Konishi N, Oba T, Takano K, Katahira K, Kimura K
Functions of Smartphone Apps and Wearable Devices Promoting Physical Activity: Six-Month Longitudinal Study on Japanese-Speaking Adults
JMIR Mhealth Uhealth 2024;12:e59708
URL: https://mhealth.jmir.org/2024/1/e59708
doi:10.2196/59708
PMID:

Original Paper

Detection of Mild Cognitive Impairment Through Hand Motor Function Under Digital Cognitive Test: Mixed Methods Study

Aoyu Li[1], PhD; Jingwen Li[2], BD; Jiali Chai[3], MA; Wei Wu[4], MD; Suamn Chaudhary[3], MA; Juanjuan Zhao[3], PhD; Yan Qiang[3,5], PhD

[1]School of Software, Taiyuan University of Technology, Jinzhong, China

[2]School of Computer Science, Xijing University, Xian, China

[3]College of Computer Science and Technology (College of Data Science), Taiyuan University of Technology, Jinzhong, China

[4]Shanxi Provincial People's Hospital, Taiyuan, China

[5]School of Software, North University of China, Taiyuan, China

Corresponding Author:
Yan Qiang, PhD
College of Computer Science and Technology (College of Data Science)
Taiyuan University of Technology
No. 209, University Street, Yuji District
Shanxi Province
Jinzhong, 030024
China
Phone: 86 18635168680
Email: qiangyan@tyut.edu.cn

Abstract

Background: Early detection of cognitive impairment or dementia is essential to reduce the incidence of severe neurodegenerative diseases. However, currently available diagnostic tools for detecting mild cognitive impairment (MCI) or dementia are time-consuming, expensive, or not widely accessible. Hence, exploring more effective methods to assist clinicians in detecting MCI is necessary.

Objective: In this study, we aimed to explore the feasibility and efficiency of assessing MCI through movement kinetics under tablet-based "drawing and dragging" tasks.

Methods: We iteratively designed "drawing and dragging" tasks by conducting symposiums, programming, and interviews with stakeholders (neurologists, nurses, engineers, patients with MCI, healthy older adults, and caregivers). Subsequently, stroke patterns and movement kinetics were evaluated in healthy control and MCI groups by comparing 5 categories of features related to hand motor function (ie, time, stroke, frequency, score, and sequence). Finally, user experience with the overall cognitive screening system was investigated using structured questionnaires and unstructured interviews, and their suggestions were recorded.

Results: The "drawing and dragging" tasks can detect MCI effectively, with an average accuracy of 85% (SD 2%). Using statistical comparison of movement kinetics, we discovered that the time- and score-based features are the most effective among all the features. Specifically, compared with the healthy control group, the MCI group showed a significant increase in the time they took for the hand to switch from one stroke to the next, with longer drawing times, slow dragging, and lower scores. In addition, patients with MCI had poorer decision-making strategies and visual perception of drawing sequence features, as evidenced by adding auxiliary information and losing more local details in the drawing. Feedback from user experience indicates that our system is user-friendly and facilitates screening for deficits in self-perception.

Conclusions: The tablet-based MCI detection system quantitatively assesses hand motor function in older adults and further elucidates the cognitive and behavioral decline phenomenon in patients with MCI. This innovative approach serves to identify and measure digital biomarkers associated with MCI or Alzheimer dementia, enabling the monitoring of changes in patients' executive function and visual perceptual abilities as the disease advances.

(JMIR Mhealth Uhealth 2024;12:e48777) doi:10.2196/48777

KEYWORDS

mild cognitive impairment; movement kinetics; digital cognitive test; dual task; mobile phone

Introduction

Background

Mild cognitive impairment (MCI) is an intermediate stage between age-related cognitive decline and dementia, which affects the individual's cognitive, social, and mental aspects and can lead to emotional problems that affect daily living [1]. A recent study has revealed that the incidence of MCI is 6.7% for individuals aged 60 to 64 years, 8.4% for those aged 65 to 69 years, 10.1% for those aged 70 to 74 years, 14.8% for those aged 75 to 79 years, and 25.2% for those aged 80 to 84 years [2]. Previous evidence has shown that patients with MCI develop dementia or Alzheimer dementia (AD) at a rate of approximately 10% to 15% per year [3,4], significantly higher than the incidence of dementia in the general population, which is 1% to 2% per year. The increasing prevalence of dementia, coupled with a high conversion rate, presents challenges not only for those directly impacted by the condition, including individuals, caregivers, and families, but also for society.

Despite the high risk of dementia for people with MCI and the need for early intervention [5], there are currently no drugs or other treatments to modify the clinical course or delay the onset of dementia [6]. In clinical settings, the diagnostic process for MCI could be costly, often involving expensive and sometimes invasive or time-consuming examinations. In addition, the early symptoms of MCI are subtle, making it easy for patients to attribute them to the normal aging process rather than to cognitive impairment and leading to an early stage where patients may not undergo clinical examination. Therefore, achieving an effective MCI diagnosis remains one of the most challenging tasks in geriatric psychiatry [7].

Digital Drawing Tasks in MCI Detection

As a comprehensive activity, drawing requires various cognitive skills including orientation, selective and sustained attention, visual memory and reconstruction, visuospatial organization, and motor performance [8]. These requirements suggest that older adults may be susceptible to cognitive dysfunction, and an accurate assessment of their involvement in relevant activities might aid in identifying such disabilities [9]. With the development of portable devices and mobile computing, some researchers have attempted to improve drawing tests using digital methods to capture more information and influencing factors related to screening for MCI, such as working memory, attention allocation, cognitive flexibility, as well as visual and spatial processing. These factors play an essential role in the screening and assessment processes for MCI and help provide more comprehensive information about cognitive functioning, leading to a better understanding of cognitive health and possible problems in participants.

Recently, Müller et al [10] used a Windows Surface Pro 4 digitizer and a handheld stylus pen to evaluate the digital Clock Drawing Test (dCDT) and found a significant difference in time in air between the healthy older adults and the patient group. Afterward, Dion et al [11] discovered a difference in the "thinking time" percentage between the groups. The MCI group took 10 seconds longer to draw their clocks from memory, suggesting that the difference between those with and without MCI may lie in grouping and coordinating the necessary cognitive resources rather than severe drawing errors. Another study demonstrated differences in the time to completion, total pen stroke count, and higher-order decision-making latencies of dCDT between age groups, involving participants recruited as free of dementia and stroke [12]. As they transition from one part of the drawing to the next (eg, postclock delay), participants may use more diverse neurocognitive resources than simply processing speed. In addition to assessing the Clock Drawing Test (CDT), other drawing tasks were also studied. Kim et al [13] created a simplified Rey-Osterrieth Complex Figure Test (sRCFT) to assess digital pen strokes, spatial arrangement, and similarity of drawings in the drawing process, finding that these measures could serve as valuable digital biomarkers for studying visual structural dysfunction in AD. In addition, Müller et al [14] investigated the movement kinematics (ie, time in air, time on surface, and total time) of older adults with MCI when copying a 3D house to reflect their manual dexterity, visual space construction, and other cognitive abilities. Finally, in the digital tree drawing test, patients with MCI and early-stage AD used fewer colors and line widths, and their images displayed reduced contrast and heterogeneity [15].

Digital drawing tasks provide standardized test administration, more detailed feedback metrics, and automated scoring than traditional paper-based tests. However, these assessment tasks typically focus on the patient's overall performance, neglecting to capture detailed features (ie, clock face, numbers, and pointers in dCDT). In modern digital devices, tracking subtle cognitive changes in patients' selection of drawing areas, organization of shapes, and determining image sequences may offer a more practical approach to MCI detection.

Dual-Task Paradigm in MCI Detection

The dual-task paradigm is an experimental design approach for exploring the effects of multitasking on cognitive performance and brain function. In the dual-task paradigm, participants must perform ≥2 tasks simultaneously, involving different cognitive processes, such as attention, stimulus encoding, decision-making, working memory, response selection, and execution. This paradigm could detect changes in dual-task performance early in the progression of the most common neurodegenerative diseases [16]. Compared to other cognitive assessment tools, the dual-task paradigm is less influenced by the education level and is more applicable to participants with different educational backgrounds. In addition, the dual-task paradigm is quick, practical, and easy to apply in clinical practice, making it a promising method for cognitive assessment [17].

In studies related to the diagnosis of MCI, the analysis of dual-task gait tests has proven to be a valid method. The dual-task gait test is a commonly used clinical assessment that

requires participants to perform a cognitive task while simultaneously walking. In other words, participants must complete additional cognitive tasks such as counting, recalling words, or performing attention shifts while walking. For example, Montero-Odasso et al [18] conducted a dual-task gait test on 112 older adults with MCI and followed them for 6 years. Interestingly, high dual-task gait costs were associated with a 3.8-fold and a 2.4-fold increased risk of progression to dementia when counting backward and naming animals, respectively. Whitson et al [16] investigated gait-cognitive dual-task performance in 29 older adults and found that APOE 4 carriers exhibited more pronounced dual-task interference than low-risk participants. Compared to electronic sidewalks, Aoki et al [19] used a Kinect sensor to capture the whole-body movements of participants and extracted more substantial gait feature information by improving the motion performance capture method. Their findings revealed that a classifier based on dual-task gait features could detect older adults who scored low on the Mini-Mental State Examination (MMSE) test [20]. In addition, Ali et al [21] used the Vicon Nexus 2.8 motion capture system to obtain kinematic gait parameters such as velocity, peak knee extension angle, and dual-task cost for participants. Their study showed that the kinematic gait parameters of the dual-task peak knee extension angle for story recall gait could effectively differentiate the MCI group from the healthy control (HC) group.

Currently, methods for detecting MCI using the dual-task paradigm focus on gait-motor assessment or developing new screening tools. Commonly used screening methods for cognitive impairment in older adults (ie, CDT and Rey-Osterrieth Complex Figure Test [RCFT]) by neurologists, psychiatrists, and general practitioners still have good clinical outcomes due to their rapid administration, patient acceptability, and simple scoring rules [22,23]. Thus, integrating a dual-task paradigm into clinical screening tools may provide more alternatives to MCI detection.

Objective

We designed tablet-based “drawing and dragging” tasks, prototyped with clinically validated drawing tests, and explored the feasibility and efficiency of assessing MCI through movement kinetics under these tasks. In this study, the drawing task improves the traditional test clinicians use to fulfill the requirements for detecting richer cognitive domains. However, the dragging task is designed to meet the cognitive screening of older adults alone at home by providing a broader and more convenient means of self-examination and early warning. Finally, we hypothesized that movement kinetics features extracted from participants during drawing or dragging could effectively distinguish patients with MCI from healthy older adults.

Methods

Ethical Considerations

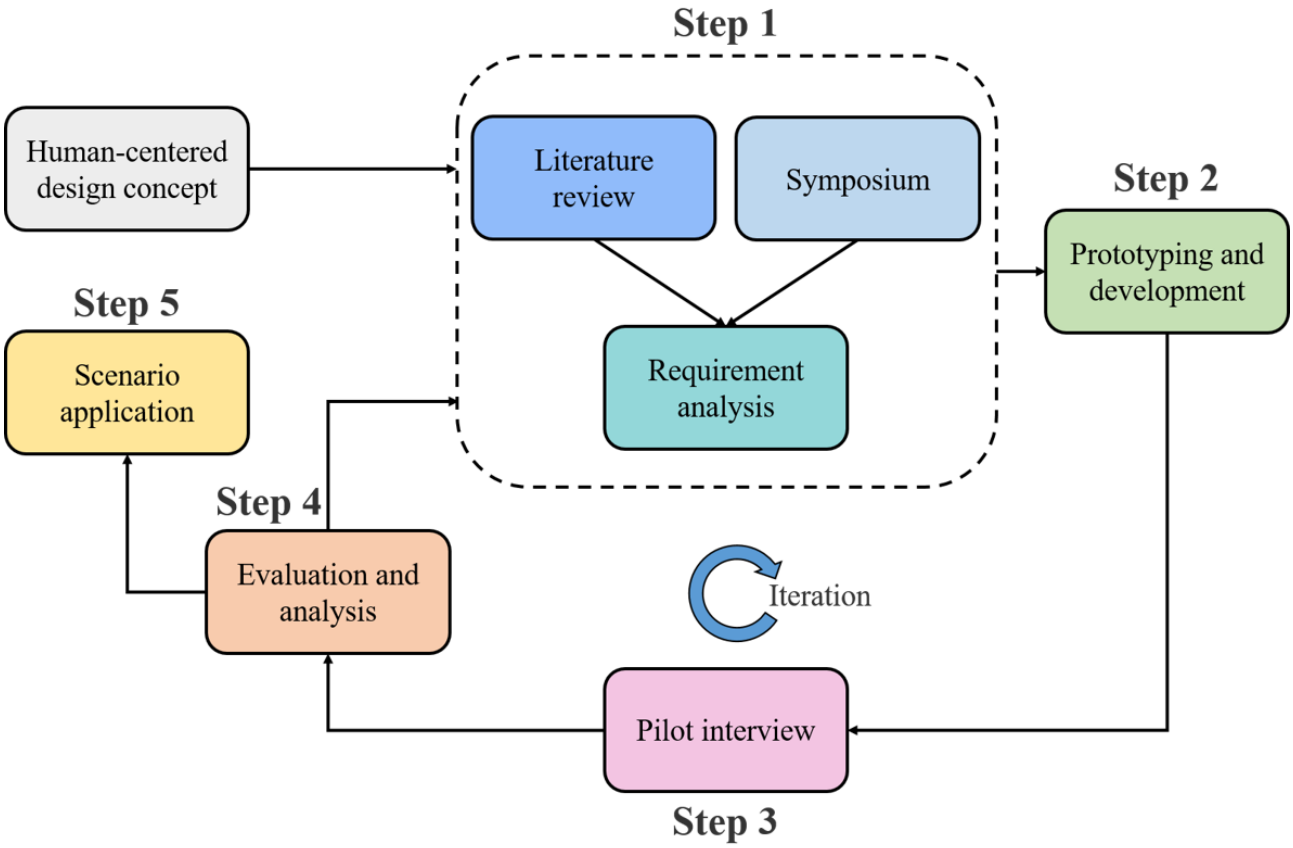
This research was reviewed and approved by the Biomedical Ethics Review Committee of Taiyuan University of Technology (20230118). The patients and participants provided their written informed consent to participate in this study. We provided US $10 to eligible older adults as compensation for participation.

Experimental Design

Design Overview

To design simple and effective “drawing and dragging” tasks for assessing hand motor function, we applied a human-centered design approach [24], as shown in Figure 1. Specifically, the approach included (1) a literature review and symposium to analyze the requirements of patients with MCI, (2) prototype design and development, (3) pilot interviews, (4) system evaluation and analysis, and (5) scenario application. In step 1, we surveyed approximately 100 previous studies and performed a detailed analysis of relevant studies. In addition, we conducted symposiums, mainly involving communication between patients with MCI, healthy older adults, caregivers, neurologists, nurses, and engineers regarding what to do and how to interact. These efforts were mainly aimed at extracting user requirements and applying them to our cognitive system. On the basis of these findings, the engineers designed and developed a task prototype in step 2. Then, in step 3, we conducted pilot interviews to better understand the different preferences and demands of patients with MCI. In other words, the interviews were semistructured exchanges, and the experimental prototype was tested among healthy older adults, patients with MCI, and caregivers to refine our task further. These first 3 steps constitute an iterative design circle to continuously improve and adapt the task to the requirements of patients with MCI. In step 4, we evaluated the extent to which the designed task could effectively differentiate between patients with MCI and HC participants and how well the user experience was. Step 5 is the scenario application, that is, the extension to home, community, and clinical settings. In the experimental design phase, the main focus is on steps 1, 2, and 3 and the iterations between them, while the evaluation and analysis (step 4) are described in the *Results* section.

Figure 1. The human-centered design approach is applied to the process of cognitive tasks, and the first 3 steps form an iterative design circle to improve and adapt the tasks continuously.



Stakeholders

The stakeholders engaged in the experimental design encompass our development team (2 neurologists, 2 nurses, and 3 engineers), patients with MCI, healthy older adults, and caregivers. Neurologists and nurses are from our partner hospitals and have been working for 10 years on diagnosing and treating neurological diseases. The engineers comprised 1 PhD and 2 MS students specializing in human-computer interaction. In addition, we recruited 49 patients with MCI, 49 healthy older adults, and 4 caregivers. They were mainly from the community, nursing home, and hospital. All older adults completed the Montreal Cognitive Assessment (MoCA) test [25], the Clinical Dementia Rating (CDR) test [26], and an informed consent form and underwent comprehensive medical evaluations conducted by experienced neurologists (including detailed medical history, systematic physical examination, and imaging studies). Furthermore, neurologists conducted clinical interviews with patients or their caregivers and collected self-reports to assess and diagnose MCI comprehensively. Among the 49 patients with MCI, all had MoCA scores <26 and CDR scores of 0.5. Brain magnetic resonance imaging or computed tomography scans revealed no other structural abnormalities associated with cognitive impairment. We provided the participants with a notebook and a pen

Prototype 1: Design and Iteration

Using clinically validated paper-based drawing tasks as screening elements, we conducted a comprehensive review of clinical MCI screening tools and invited 4 (8%) of the 49 patients with MCI and 4 (8%) of the 49 healthy older adults to participate in a symposium. We focused on the following points:

1. Which of the existing MCI diagnostic methods in the clinical setting involve hand motor function?
2. Does the screening tool address integrated cognitive abilities (ie, memory, attention, visuospatial, motor planning, and executive functions)?
3. Does the application of screening tools extend to hospitals, nursing homes, or households?
4. The psychology of older adults, considering their preferences and cognitive limitations

The symposium results identified the CDT and RCFT as the primary tasks for the cognitive assessment system. Subsequently, in step 2 (Figure 1), the engineers designed and developed prototype 1 based on these tasks. Prototype 1 consisted of a tablet-based dCDT and a digital RCFT. The dCDT required participants to complete three tasks in the clock drawing area: (1) draw a clock face, (2) write all the numbers in the correct position, and (3) use the pointers to indicate “10 minutes past 11-o’clock.” Similarly, in the digital RCFT, participants were prompted to copy the Rey-Osterrieth Complex Figure [ROCF] on the left side within the copy area. Next, to assess the feasibility of the 2 drawing tasks in prototype 1 among older adults, we conducted interviews (step 3; Figure 1) with 6 older adults (patients with MCI: 3/49, 6%; healthy older adults: 3/49, 6%) from the community. The interviews involved engaging the older adults in the tests and soliciting their feedback and

opinions. Some common dialogues from the study participants are as follows:

> *Experimenter: Could you share your thoughts on these tests?*
>
> *Patient with MCI: What time do the pointers indicate? I forgot.*
>
> *Patient with MCI: Do others have their minute pointers pointing at the number '2'? Am I the only one with the minute pointer pointing at the number '10'?*
>
> *Patient with MCI: This figure (RCFT) is too complex to complete, and I prefer not to continue.*
>
> *Healthy individual: Am I performing as well as others?*
>
> *Healthy individual: Does it matter that the number of positions are painted at different intervals?*
>
> *Healthy individual: RCFT is too time-consuming, and my eyes are tired.*

Prototype 2: Design and Iteration

Using the pilot study (step 3; Figure 1), we found that older adults were somewhat confused by the rules of dCDT (Figure 2A) but were generally receptive and engaged in its implementation. However, the digital RCFT was less than ideal. Participants thought that the ROCF was too complex and challenging to complete at first glance. In addition, it typically took ≥10 minutes to complete, which was not user-friendly to older adults. Hence, in the second symposium, we modified the original figure to a relatively simplified version while retaining its main framework. The modification work was collaboratively undertaken by 2 neurologists from our affiliated hospital, each with >10 years of experience, to ensure a balanced representation of global and local components. While the original RCFT comprises 18 components (4 global and 14 local components) [27], our simplified version of the RCFT consists of 5 global components and 4 local components, as illustrated in Figure 3.

In the global components section, we maintained the large rectangle with horizontal, vertical, and diagonal crossings; the large triangle (only the position has changed); and the 4 horizontal lines in the upper left panel, while the rest of the regions or local components have been simplified. Regarding the local components section, we made the following modifications: (1) squares replaced the diamonds, (2) the 3-point circle was moved to the bottom, and (3) three triangles staggered up and down instead of 5 parallel lines. We staggered the triangles up and down to allow participants to focus more on local components. Finally, some overlapping lines and detail parts outside the outline were removed in the simplified version. After the development, a raw copy of sRCFT is shown in Figure 2B.

Figure 2. Interactive interface of drawing and dragging tasks. (A) digital Clock Drawing Test (dCDT; single task), (B) simplified Rey-Osterrieth Complex Figure Test (sRCFT; raw copy), (C) dCDT (dual task), (D) sRCFT (delayed recall), (E) clock "drag and drop", (F) sRCFT "point and line".

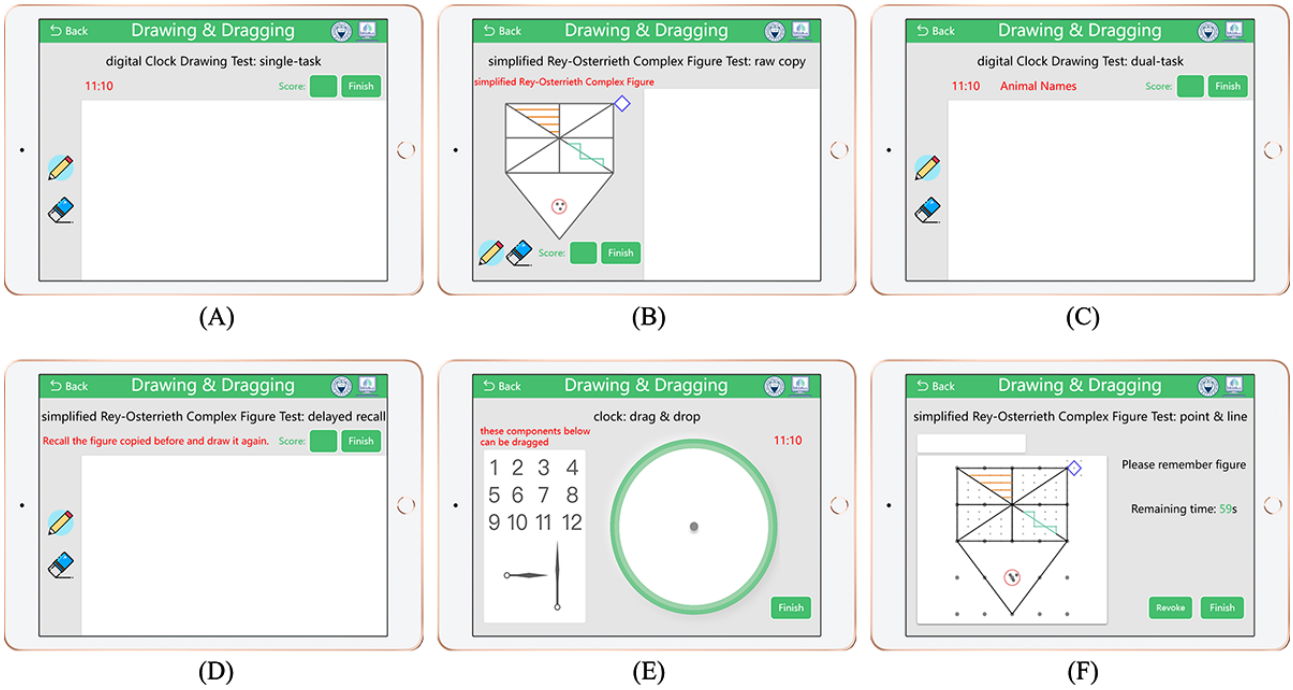
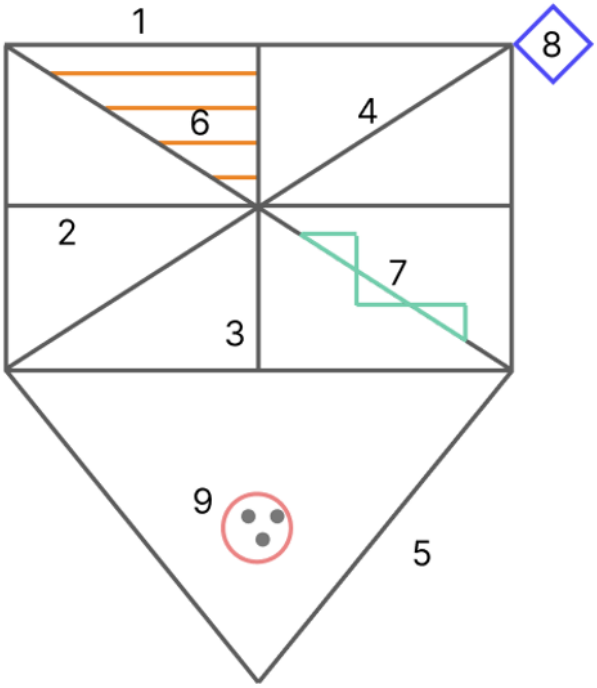


Figure 3. The simplified Rey-Osterrieth Complex Figure consists of 5 global and 4 local components.



Global	1	Large rectangle
	2	Horizontal line
	3	Vertical line
	4	Diagonal cross
	5	Triangle
Local	6	4 parallel lines
	7	3 triangles
	8	Small square
	9	A smiley face

Subsequently, we conducted another pilot interview to investigate the performance of the sRCFT compared to the original RCFT on cognitive screening. We recorded the distribution of original RCFT and sRCFT scores for 24 (49%) out of 49 healthy older adults and 24 (49%) out of 49 patients with MCI in the raw copy and delayed recall (waiting 20 minutes to recall and draw the figure) tasks. The normality test for the score variables revealed that none of the groups met normality. Therefore, we used the median (IQR) for statistical description and the nonparametric Mann-Whitney *U* test for comparison between groups. As depicted in Table 1, the HC group scored significantly higher than the MCI group on the delayed recall task, while no significant difference was observed between the 2 groups in the raw copy task. Moreover, we examined the correlation between the raw copy and delayed recall scores of the original RCFT and sRCFT using Spearman correlation analysis. A positive correlation was noted between the original RCFT and sRCFT score (raw copy: r=0.812; P<.001 and delayed recall: r=0.816; P<.001; Figure 4).

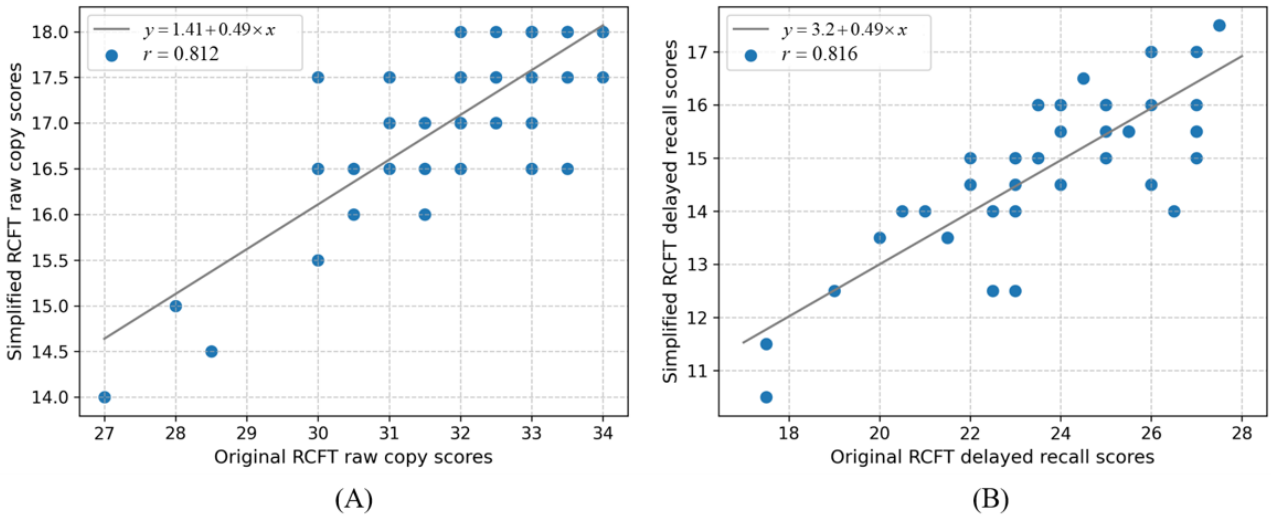
Table 1. Raw copy and delayed recall scores of healthy older adults and patients with mild cognitive impairment on the original and simplified Rey-Osterrieth Complex Figure Test (RCFT).

	HC[a] group (n=24)	MCI[b] group (n=24)	*P* value	*z* score
Original RCFT: raw copy (IQR 0-36)	32.5 (31.5-33.5)	32 (31-33)	.18	−1.350
Simplified RCFT: raw copy (IQR 0-18)	17.25 (16.5-17.5)	17 (16.5-18)	.94	−0.074
Original RCFT: delayed recall (IQR 0-36)	25.25 (24-27)	22.75 (21-24.5)	<.001	−3.639
Simplified RCFT: delayed recall (IQR 0-18)	16 (15-16.875)	14 (13.5-15)	<.001	−4.208

[a]HC: healthy control.

[b]MCI: mild cognitive impairment.

Figure 4. Correlation between raw copy and delayed recall scores on original and simplified RCFT: (A) raw copy and (B) delayed recall. RCFT: Rey-Osterrieth Complex Figure Test.



Prototype 3: Design and Iteration

As is well known, MCI encompasses changes in cognitive performance, including declines in short-term memory, working memory, logical thinking, verbal expression, spatial cognition, and executive function [24]. To comprehensively assess these aspects, we enhanced the drawing task during the third symposium, which involved adding 3 dual tasks to the dCDT and introducing a 10-minute delayed recall task in the sRCFT, as depicted in Figures 2C and 2D. In this prototype, 3 language function–related tasks include series 1 (low load: count backward from 100), animal names (medium load: name as many animals as possible), and series 3 (high load: subtract 3 from 100 and give the result) [28,29]. We combined the 3 language function–related tasks with the dCDT, resulting in 3 dual tasks. In the dual task, participants were required to perform a single dCDT task while completing a task related to language function. Series 1 was applied to test essential language functions, whereas series 2 and 3 focused on memory and logical thinking skills, respectively. Finally, according to the Schulman criteria, the dCDT score ranges from 1 point (ie, a perfectly accomplished CDT) to 6 points (ie, severe impairment and no identifiable clock) [30].

Performance in the delayed recall condition helps examiners assess visuospatial memory in declarative memory associated with the hippocampus and related areas of the right temporal lobe [31,32]. However, in the traditional delayed recall task, participants were asked to recall the previous figure 30 minutes after completing the RCFT raw copy task. Long waits are unwelcome and unpleasant for older adults. Therefore, we shortened the delayed recall time to 10 minutes [33] and adjusted the drawing sequence. The participants' performance in sRCFT will be scored in accuracy and location [34]. The score weights of the 9 components are equal, namely 2 (accurately drawn and correctly located); 1 (accurately drawn and incorrectly located or inaccurately drawn and correctly located); 0.5 (inaccurately drawn and incorrectly placed but identifiable); or 0 (inaccurately drawn, incorrectly located, and unidentifiable). Therefore, the possible range of raw scores is 0 to 18. A lower score indicates a more severe visual perception or construction function impairment.

In addition, during the nursing home interviews, 4 (8%) out of 49 healthy older adults mentioned that they were not accustomed to using the Apple Pencil for drawing and suggested incorporating a practice mode before initiating the test. Following discussions with the development engineers, we promptly integrated their feedback and added a practice mode, allowing free drawing without a time limit. At the same time, we assisted in teaching during the practice process to ensure that the participants could use the Apple Pencil in their experiments.

Prototype 4: Design and Iteration

To meet the demands for cognitive screening for older adults alone at home, we designed "drag and drop" and "point and line" tasks to indirectly assess their cognitive status by capturing subtle changes in hand motor function. Participating experimental designers included 4 (8%) out of 49 patients with MCI, 4 (8%) out of 49 healthy older adults, 2 (50%) out of 4 caregivers, and all our team members.

Similar to drawing, drag and drop is used to assess hand motor function and detect potential cognitive decline in patients with MCI [35,36]. So, we designed a clock "drag and drop" task, as shown in Figure 2E. The clock "drag and drop" task requires participants to drag or drop numbers and pointers from the drag area to the inside of the clock face, then rotate the pointers to point "10 minutes past 11." On the other hand, concerning sRCFT, we transformed the drawing task into a "point and line" task, as shown in Figure 2F. First, the participants were asked to memorize the simplified ROCF (including his position, belt, and all details) for 60 seconds. Then, the participant began recalling, connecting dots, and drawing lines to reproduce the target pattern. Specifically, the clock "drag and drop" task was scored on a scale of 0 to 3 (ie, numbers position, numbers order, and pointers indication; a high score indicates a perfect clock).

The sRCFT "point and line" task has the same scoring rules as the sRCFT drawing task (ie, a score between 0 and 18). In addition, during the pilot interviews, 4 (8%) out of 49 patients with MCI and 4 (8%) out of 49 healthy older adults from our partner hospitals completed the tests.

Experimental Participants and Procedure

Before initiating the study, we conducted a power analysis to estimate the sample size required to detect significant differences between the MCI and HC groups. This analysis considered an expected effect size of 0.3, a significance level of .05, and a statistical power of 0.8, requiring at least 82 participants per group [37]. Similar research in the field also informed our sample size, which typically used participant numbers within this range for comparable outcomes [38-41].

We rerecruited 207 participants from the geriatrics and neurology department research clinics at our collaborating hospitals who had not previously been involved in the design or development of the experiment. Participants were identified using a purposive sampling method [42], with the process being meticulously overseen by experienced neurologists. The inclusion criteria were that participants must (1) have normal hearing and vision or corrected to normal, (2) be aged >65 years, (3) have completed the MMSE test, (4) have completed the MoCA test, (5) have completed the CDR test, (6) be capable of moderate exercise without physical disabilities, (7) have no severe depressive symptoms or other mental illnesses, (8) be capable of using smart devices (eg, smartphones and tablets). Neurologists contacted potential participants during their clinic visits and explained the study's purpose, related procedures, and the possible impact of the research findings. Once potential participants expressed interest, neurologists conducted comprehensive medical evaluations, including detailed medical history collection; physical examinations; brain imaging (magnetic resonance imaging or computed tomography scans); and cognitive function assessments (using the MMSE, MoCA, and CDR scales). Of 207 participants, the MCI group comprised 108 (52.2%) participants who scored <26 on the MoCA and had a CDR score of 0.5, while the HC group included 99 (47.8%) healthy older adults without symptoms of cognitive decline. Brain imaging scans revealed no structural abnormalities causing cognitive impairment. Furthermore, all patients with MCI met the criteria proposed by the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association [43]. We also administered the habitual hand questionnaire [44], which consisted of 13 items, to all participants. The MCI and healthy groups were matched for age, gender, hand preference, education, average sleep duration (in general), exercise habit (regularly engaging in exercise or infrequently), and years of smart device use. Table 2 summarizes the clinical and demographic information for all participants.

Table 2. Clinical and demographic characteristics (n=207).

	MCI[a] (n=108)	HC[b] (n=99)	*P* value
Age (y), mean (SD)	71.34 (4.48)	70.11 (4.00)	.13
Gender, n (%)			.81
Woman	64 (59.2)	57 (57.6)	
Man	44 (40.7)	42 (42.4)	
Hand preference, n (%)			.43
Left	11 (10.2)	7 (7.1)	
Right	97 (89.8)	92 (92.9)	
Education years, mean (SD)	6.21 (3.28)	6.81 (2.89)	.31
Hours of sleep, mean (SD)	5.93 (1.20)	6.28 (1.06)	.11
Exercise habit, n (%)			.63
Yes	52 (48.1)	51 (51.5)	
No	56 (51.9)	48 (48.5)	
Smart device use years, mean (SD)	5.52 (2.54)	5.87 (2.52)	.47
MoCA[c], mean (SD)	24.09 (1.13)	27.04 (1.21)	<.001
MMSE[d], mean (SD)	25.31 (1.13)	28.00 (0.92)	<.001

[a]MCI: mild cognitive impairment.

[b]HC: healthy control.

[c]MoCA: Montreal Cognitive Assessment.

[d]MMSE: Mini-Mental State Examination.

We implemented digital "drawing and dragging" tasks on the Lenovo Qitian M530-A154 (AMD Ryzen 7 PRO 2700/16 GB). Our system introduction video can be found in Multimedia Appendix 1. Before the experiment began, participants could draw freely with an Apple Pencil until they were prepared to collect data. After the practice mode, given the sRCFT (delayed

recall) task rules, we clarified the execution sequence of digital drawing tasks, namely (1) sRCFT raw copy, (2) dCDT single task, (3) dCDT dual task, (4) sRCFT delayed recall, (5) clock "drag and drop," and (6) sRCFT "point and line." All tests were run on an iPad 2019 tablet (seventh generation, 3 GB/128 GB, with 10.2 inches and 2160×1620 touch screen), and Apple Pencil was configured for drawing and lining. The experimenter is engaged in the entire data collection process, clarifying the experiment's rules and procedures for the participants and documenting their feedback. At the beginning of the experiment, the experimenter explained the whole process to the participants and encouraged them to draw freely to familiarize themselves with the use of Apple Pencil. Once participants felt prepared, the test commenced.

Structured Questionnaire and Unstructured Interview

A structured questionnaire was used to investigate users' experiences of the overall cognitive screening process. The questionnaire is rated on a scale of 1 to 5, with higher scores indicating better performance in this area. The questionnaire consisted of four questions.

- Question 1: are the tasks easy to understand and interact with?
- Question 2: do the tasks evoke self-awareness and inspire them?
- Question 3: are the tasks are interesting?
- Question 4: would the tasks be used consistently in future daily life?

The unstructured interview was designed to gather user feedback on the digital cognitive tests, aiming to enhance the system further. The discussion centered on two primary themes:

(1) what are your thoughts on the "drawing and dragging" tasks?
(2) What aspects do you believe require optimization?

Data Processing and Feature Extraction

Tablet-based "drawing and dragging" tasks can assess the hand motor function and cognitive abilities of older adults. The data obtained from these tests were used for feature extraction, serving as a digital biomarker to distinguish individuals with MCI from those who were healthy. Therefore, we conducted data cleaning procedures before extracting the features, which involved removing outliers and ensuring data consistency. Next, we extracted 5 features related to hand motor function, including time, stroke, frequency, score, and sequence. All these features were extracted from the drawing or dragging tasks participated by older adults. A comprehensive overview of the details of these 5 categories of features can be found in Multimedia Appendix 2.

Statistical Analysis

Statistical analysis was conducted using SPSS software (IBM Corp) to analyze tablet device demographic characteristics and cognitive data. Age, years of education, sleep duration, years of smart device use, MoCA scores, and MMSE scores were described using means (SDs), while gender, hand preference, and exercise habits were presented as percentages. Furthermore, we conducted the Kolmogorov-Smirnov test to assess the normal distribution of all variables. For those variables conforming to a normal distribution, we applied 2-tailed *t* tests (for continuous variables) and chi-square tests (for categorical variables) to determine the significance of intergroup differences. For nonnormally distributed variables, we used the nonparametric Mann-Whitney *U* test to assess intergroup differences and estimated CIs using the Hodges-Lehmann estimator. In addition, logistic regression analysis was performed to evaluate the diagnostic value of the chosen variables in discriminating between healthy older adults and patients with MCI. The statistical significance level for all tests was set at $P<.05$.

Results

Analysis of Hand Motor Function Features

Table 3 exhibits the statistical comparison results between the MCI and HC groups concerning hand motor function features. In terms of MCI detection, the findings reveal discrepancies in the features selected across different tasks, indicating variations in the number of valid features among them. For instance, in the dCDT tasks (ie, dCDT single task, dCDT series 1, dCDT animal names, and dCDT series 3), the number of features with significant ($P<.05$) was 7, 7, 2, and 8, respectively. Furthermore, our analysis based on the number of selected features indicates that time- and score-based features outperformed stroke- and frequency-based features. Notably, features with significant ($P<.001$) were predominantly observed in time- and score-based features, while they were rare in stroke- and frequency-based features (only observed in sRCFT raw copy and clock drag and drop). Particularly, significant differences between the 2 groups were observed regarding time in the air, time on the surface, and time being dragged. Compared to the MCI group, the HC group exhibited significantly shorter times for switching between strokes and quicker drawing or dragging speeds. Similarly, the *P* values of the score features were mostly <.05, except for the sRCFT task (raw copy), which yielded the preferred features for subsequent logistic regression model construction. However, the dCDT animal names outcome was deemed unsatisfactory. During data collection, it was observed that most participants (151/207, 72.9%) were more accustomed to using the names of the 12 Chinese zodiac animals (Chinese folk culture), which may coincide with the clock numbers (1-12).

Table 3. Statistical comparisons were conducted for 4 categories of features between healthy older adults and patients with mild cognitive impairment.

Task	Time, *P* values							Stroke, *P* values						Frequency, *P* values			Score, *P* values
	1[a]	2[b]	3[c]	4[d]	5[e]	6[f]	7[g]	8[h]	9[i]	10[j]	11[k]	12[l]	13[m]	14[n]	15[o]	16[p]	17[q]
T1[r]	.233	.525	<.001	.152	.061	<.001	<.001	.033	.121	.261	.239	.014	.077	.237	.024	.274	<.001
T2[s]	.730	.235	.013	.265	.451	<.001	.256	.017	.065	.057	.632	.155	.011	.489	.045	.006	<.001
T3[t]	.625	.156	.071	.059	.077	.022	.366	.641	.095	.115	.541	.514	.513	.093	.336	.532	.019
T4[u]	.254	.785	<.001	.082	.084	<.001	.514	.012	.025	.074	.365	.224	0.002	.258	.009	.008	<.001
T5[v]	.299	—[w]	<.001	—	—	<.001	—	—	.054	—	—	.088	—	—	<.001	—	.569
T6[x]	<.001	—	<.001	—	—	<.001	—	—	.006	—	—	.067	—	—	.399	—	<.001
T7[y]	.058	<.001	—	<.001	—	—	—	—	—	—	—	—	—	<.001	—	—	.029
T8[z]	<.001	—	—		.457	—	<.001	—	—	—	—	—	—	—	—	.004	.006

[a]1: thinking time for the first stroke.
[b]2: circle painting time or umber drag time.
[c]3: number painting time or figure painting time.
[d]4: pointer painting time or pointer drag time.
[e]5: circle unpainted time or global drawing time.
[f]6: number unpainted time or figure unpainted time.
[g]7: pointer unpainted time or local drawing time.
[h]8: circle total stroke.
[i]9: number total stroke or figure total stroke.
[j]10: pointer total stroke.
[k]11: circle pen-up stroke.
[l]12: number pen-up stroke or figure pen-up stroke.
[m]13: pointer pen-up stroke.
[n]14: circle painting frequency or hand drag frequency.
[o]15: number painting frequency or figure painting frequency.
[p]16: pointer painting frequency or figure drawing frequency.
[q]17: score.
[r]T1: digital Clock Drawing Test single task.
[s]T2: digital Clock Drawing Test dual task (series 1).
[t]T3: digital Clock Drawing Test dual task (animal names).
[u]T4: digital Clock Drawing Test dual task (series 3).
[v]T5: simplified Rey-Osterrieth Complex Figure Test raw copy.
[w]Not applicable.
[x]T6: simplified Rey-Osterrieth Complex Figure Test delayed recall.
[y]T7: clock drag and drop.
[z]T8: sRCFT point and line.

Analysis of Drawing Sequences

We converted the drawing of the dCDT and sRCFT into pseudocolor images encoded by a series of colors in the order of strokes (Multimedia Appendix 3). In the dCDT single task, patients with MCI tended to rotate the tablet when writing numbers (making some numbers look upside down) and drawing lines to determine the 12 o'clock direction (MCI group: numbers 2 and 5 in Multimedia Appendix 3). Furthermore, upon analyzing the dCDT dual task, we noted that healthy older adults may exhibit errors in number positioning and slight gaps between numbers in series 1s due to increased cognitive load (HC group: numbers 2, 3, and 4 in Multimedia Appendix 3). As anticipated, both groups encountered challenges with the relatively complex series 3s. Notably, even some healthy older adults displayed significant gaps between numbers. However, in the spatial organization of the sRCFT, there was no discernible difference between the 2 groups (ie, global-first or local-first approach or top-first and bottom-second strategies). Particularly, healthy older adults exhibited good recall of figures, albeit with minor omissions in some details. Conversely, patients with MCI showed poor recall of figures, indicating deficits in visual memory and spatial construction abilities.

Diagnostic Value of "Drawing and Dragging" Tasks

To further explore the diagnostic value of the designed tasks in distinguishing patients with MCI from HC participants, we used a forward stepwise inclusion method in which the features extracted from the task were entered into a logistic regression model, where the diagnostic group (HC vs MCI) was considered as the dependent variable, and the task-extracted features were used as the independent variables, as shown in Table 4.

Table 4. Digital drawing and dragging tasks to construct logistic regression models with selected variables and their performance.

Model and selected variable	β	OR[a] (95% CI)	*P* value
dCDT[b]: single task			
Score	1.430	4.180 (2.070-8.441)	<.001
Number unpainted time	0.178	1.195 (1.075-1.327)	.001
Pointer unpainted time	0.326	1.385 (1.166-1.646)	<.001
dCDT: dual task			
Number painting time	0.415	1.515 (1.228-1.869)	<.001
Number unpainted time	0.092	1.097 (1.036-1.161)	.001
Number painting frequency	0.343	1.409 (1.080-1.837)	.01
sRCFT[c]: raw copy			
Figure painting time	0.141	1.152 (1.071-1.238)	<.001
Figure unpainted time	0.168	1.183 (1.085-1.290)	<.001
sRCFT: delayed recall			
Score	−0.689	0.502 (0.339-0.744)	.001
Figure painting time	0.149	1.161 (1.058-1.274)	.002
Figure unpainted time	0.096	1.101 (1.028-1.178)	.006
C lock: drag and drop			
Number drag time	0.135	1.145 (1.066-1.230)	<.001
Pointer drag time	0.260	1.297 (1.147-1.468)	<.001
Hand drag frequency	0.487	1.627 (1.268-2.087)	<.001
sRCFT: point and line			
Thinking time for the first stroke	−0.856	0.425 (0.259-0.696)	.001
Local drawing time	0.220	1.246 (1.032-1.506)	.02
Local components score	0.118	1.126 (1.063-1.192)	<.001

[a]OR: odds ratio.

[b]dCDT: digital Clock Drawing Test.

[c]sRCFT: simplified Rey-Osterrieth Complex Figure Test.

Furthermore, we used 4 metrics to measure the classification performance of the models, including accuracy, sensitivity, specificity, and area under the curve, as shown in Table 5. Receiver operating characteristic curves for logistic regression models were plotted, as depicted in Figure 5. Notably, the sRCFT delayed recall reached 88.4% (183/207) of the highest detection accuracy, while its specificity results were also the best (90/99, 91%). Meanwhile, the area under the curve indicating the authenticity of the detection method is the highest in sRCFT delayed recall, signifying superior performance. In the dCDT dual task, 88% (95/108) of all predicted patients with MCI were actual patients with MCI. In contrast, the sRCFT raw copy yielded lower classification results, with an accuracy of 82.1% (170/207). Our interpretation of these findings is that relatively complex tasks (eg, dCDT dual task and sRCFT delayed recall) are associated with integrated cognitive abilities, including executive functioning, language comprehension and expression, information extraction, and spatial visualization, making them more effective in detecting cognitive deficits.

Table 5. Diagnostic value of digital drawing and home dragging tasks in healthy older adults and patients with mild cognitive impairment.

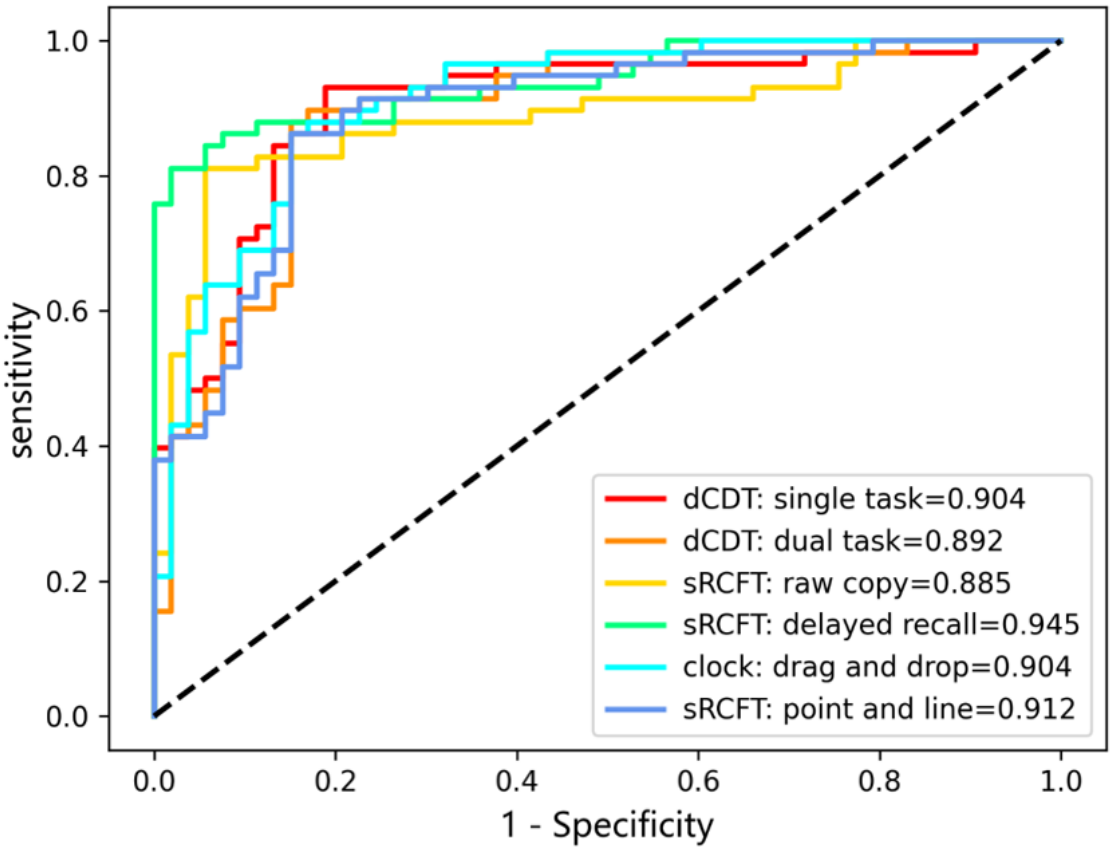
Model	Accuracy	Sensitivity	Specificity	AUC[a] (95% CI)	*P* value
dCDT[b]: single task	0.845	0.861	0.828	0.904 (0.846-0.962)	<.001
dCDT: dual task	0.855	0.880	0.828	0.892 (0.830-0.954)	<.001
sRCFT[c]: raw copy	0.816	0.843	0.788	0.885 (0.820-0.950)	<.001
sRCFT: delayed recall	0.884	0.861	0.909	0.945 (0.903-0.986)	<.001
clock: drag and drop	0.845	0.861	0.828	0.904 (0.846-0.962)	<.001
sRCFT: point and line	0.855	0.861	0.848	0.912 (0.859-0.965)	<.001

[a]AUC: area under the curve.

[b]dCDT: digital Clock Drawing Test.

[c]sRCFT: simplified Rey-Osterrieth Complex Figure Test.

Figure 5. Receiver operating characteristic curves of our cognitive system. dCDT: digital Clock Drawing Test; sRCFT: simplified Rey-Osterrieth Complex Figure Test.



Comparison of Our Cognitive System With Existing Studies

We compared 4 aspects of sensitivity, specificity, administration time, and self-administration with digital cognitive tests identified in recent years to identify MCI and dementia, as shown in Table 6. For specificity and sensitivity, we averaged the results of the 4 drawing tasks and the 2 dragging tasks. The findings suggest that our cognitive system is equivalent to the best-performing Vigilance and Memory Test [45] in accurately discriminating patients with MCI. Similarly, in identifying healthy older adults, the performance on both types of tasks surpassed that observed in most recent studies, except for slightly lower performance than the screening system based on ROCF [46]. Although the total time for the digital drawing tasks was approximately 15 minutes, each task ranged from 3 to 5 minutes, aligning with the administration time (typically between <5 and 35 minutes) observed in most studies. Moreover, in the home dragging task, the time was generally shorter, and participants could achieve self-administration, a feature rarely observed in previous studies.

Table 6. Our cognitive system compared to existing mild cognitive impairment detection studies.

Test name	Study and year	Sensitivity	Specificity	Administration time (min)	Self- administration
Digital Tree Drawing Test	Robens et al [15], 2019	0.560	0.830	<5	—[a]
FACE[b]-memory	Alegret et al [41], 2020	0.734	0.721	30	Yes
Vigilance and Memory Test	Fung and Lam [45], 2020	0.861	0.753	15	No
Screening System based on ROCF[c]	Cheah et al [46], 2019	0.756	0.864	35	No
Digital Clock Drawing Test	Müller et al [47], 2019	0.854	0.775	<5	—
Smart Aging Serious Game	Cabinio et al [48], 2020	0.844	0.755	—	—
Virtual supermarket	Eraslan Boz et al [49], 2020	0.630	0.720	25	Yes
Digital drawing task (average)	This paper	0.861	0.838	15	No
Home dragging task (average)	This paper	0.861	0.838	<5	Yes

[a]Not applicable.

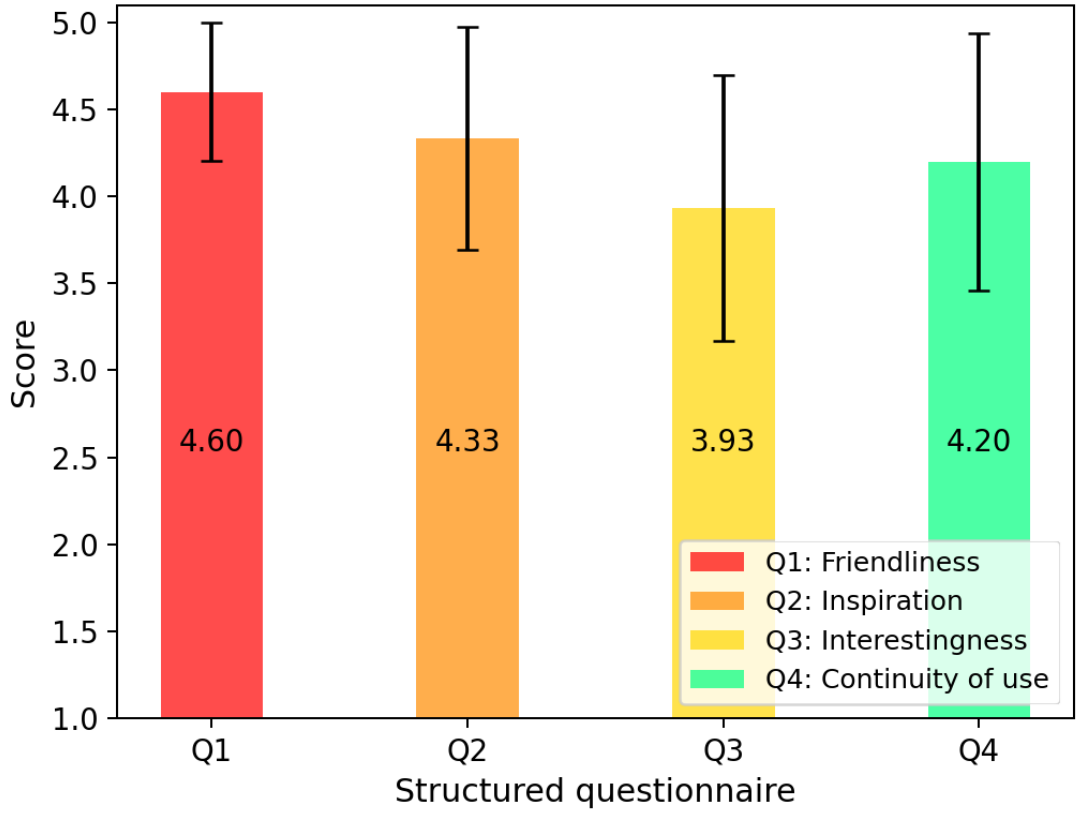
[b]FACE: Face-Name Associative Memory Exam.

[c]ROCF: Rey-Osterrieth Complex Figure.

User Experience of Cognitive Screening Process

Figure 6 illustrates the results of the questionnaire experiment, where the error bars represent SD. All 207 participants rated the screening process highly across all aspects. The mean score for question 1 (4.60) was the highest among all questions, suggesting that our system was user-friendly, enabling most older adults to complete each test. However, the score for question 3 (3.93) was the lowest of all 4 questions, possibly due to the system's limited interaction, resulting in tasks being perceived as less engaging. Furthermore, most older adults (179/207, 86.5%) reported that our system engaged their cognition and helped identify cognitive deficits during self-screening (question 2: 4.33), expressing an interest in continued use in the future (question 4: 4.20).

Figure 6. User experience results with the structure questionnaire. Q: question.



Similarly, we gleaned valuable insights from unstructured interviews. All participants concurred that the screening tool is user-friendly and practical. Furthermore, older adults offered valuable suggestions for enhancing the screening tool: (1)

integrating social features to bolster user engagement and (2) providing instructional videos for each test to aid participants in comprehending the task requirements and procedures more effectively.

Discussion

Design Advantages Behind "Drawing and Dragging" Tasks

First, we digitized the traditional drawing test and assessed cognitive performance using a digital pen and tablet. The screening system recorded task-related data (eg, time, score, stroke, frequency, and sequence) and automatically uploaded these data to a remote cloud server where they could be analyzed and visualized through statistical analysis for easy access by neurologists and patients. Second, we captured the hand motor function features of each module (ie, clock face, number, and pointer) in dCDT to further explore the specific cognitive deficits of patients with MCI, which are rarely mentioned in the existing literature. In addition, the sRCFT was well received by older adults in practical testing settings and was easier to administer. Third, we introduced 3 tasks related to language function in the dCDT, expanding the assessment capabilities of the drawing system and increasing its potential application in detecting various cognitive impairments. Fourth, each task was time efficient, taking only 3 to 5 minutes. Older adults were willing to engage with and accept this, effectively avoiding the negative emotional impact of prolonged testing. Fifth, the screening system provides drawing and dragging interaction, which maintains the traditional drawing test requirements while meeting patients' needs for self-perception screening. In other words, the system can be applied to hospitals and clinics to facilitate physicians to understand patients' conditions in a timely and effective manner and can also meet the self-assessment of the cognitive level of older adults at home and provide an early warning means.

Detection Accuracy and Interpretability of Features

As for the cognitive assessment results, the 4 drawing and 2 dragging tasks performed well in discriminating abilities (average accuracy of 85.2%), comparable to the results of alternative state-of-the-art methods [47,50,51]. In addition, statistical hypothesis testing was conducted on the 5 categories of critical features acquired during the cognitive assessment. The analysis results provide three insights: (1) time features are the most prominent in cognitive assessment; (2) key features vary in different tasks; (3) most older adults strive for perfection in drawing or dragging tasks and ignore the features of movement kinematics, such as long-term stagnation in the air, multiple drawings to make the figure symmetrical, and slow drawing or dragging speed.

Regarding the digital pen stroke data analysis, we compared differences between the 2 groups in the dCDT. Typically, healthy older adults prioritized drawing the numbers 12, 6, 3, and 9 to divide the clock face area accurately and make the clock more perfect, partly reflecting their ability to generate adequate decision-making strategies [52].

In contrast, patients with MCI were accustomed to rotating the tablet to write numbers in sequence, counterclockwise to write numbers from 12, unable to indicate "10 minutes past 11 o'clock" and refine the scale between numbers 11 and 12. These behavioral patterns reveal deficiencies in patients with MCI's abilities related to time perception, numerical comprehension, and the organization and planning required for executing cognitive tasks [30,53]. On the other hand, when comparing drawing sequences between healthy older adults and patients with MCI in the sRCFT, most participants followed similar patterns: (1) from top to bottom, (2) from global to local, and (3) checking and filling in missing information. Some patients used unconventional methods, such as mirror drawing or completing the global structure last, possibly due to personal painting habits. Notably, patients with MCI scored lower and omitted many local features in delayed recall, suggesting a link between the loss of detailed features and visual-spatial dysfunction or deficits in visual-spatial working memory [54,55]. Furthermore, our study primarily focused on extracting local features, which may offer better identification of patients with MCI.

Experience of "Drawing and Dragging" Tasks

In terms of user experience, 207 participants rated the ease of use positively, with an average score of 4.60 (SD 0.398). In contrast, the average score for task enjoyment was 3.93 (SD 0.765), which is less desirable. The "drawing and dragging" tasks are designed as a serious game to maintain rigor while assessing participants' perceptions through teaching and training methods. However, this approach may be less conducive to generating enjoyment and may challenge user interest, leading to a perception of dullness and tedium. Moving forward, our focus will be on enhancing plot design and incentives and making it resemble a real game by incorporating interactivity and feedback mechanisms.

Limitations and Future Work

For our study, it is important to acknowledge several potential limitations and proposed solutions. First, our sample size was relatively small, comprising only 207 participants. To address this, we intend to conduct a longitudinal study to validate the efficacy of the developed MCI screening tool, aiming to recruit a larger and more representative sample. Second, considering the visual condition of older adults, the "drawing and dragging" tasks were performed on a 10.2-inch touchscreen tablet. Therefore, whether this study's experimental results can be extrapolated to smaller-screen smartphones remains to be determined. We will validate these smartphone tasks to fully evaluate their applicability, considering the user experience on different devices.

Conclusions

We present an MCI detection system based on the digital "drawing and dragging" by assessing the movement kinetics of older adults during various tasks. The interactive system comprises 4 digital drawing tasks and 2 home dragging tasks. We report how these 6 cognitive tasks are designed, optimized, and evaluated through research on traditional clinical screening tools, discussions with related neurologists, and user experience

feedback. Then, specific parameters were combined for different tasks, and different mixed models were constructed. The experimental outcomes demonstrate the efficacy of the proposed model in distinguishing patients with MCI from healthy older adults, garnering positive reception from multiple users. In a broader context, this study advances novel approaches for nuanced feature extraction of movement kinetics in individuals with MCI, offering tangible support for developing cognitive impairment detection systems.

Acknowledgments

This work was supported by the Central Guidance for Local Science and Technology Development Funds (YDZJSX2022C004) and National Natural Science Foundation Collaboration (U21A20469). In addition, the authors would like to thank all the volunteers for their participation in their experiment.

Authors' Contributions

AL contributed to paper writing, experimental design, and data analysis, while JL contributed to experimental design and system development. JC and WW collected and organized the data. SC proofread the paper for formatting and grammar. JZ and YQ were responsible for experimental design and system optimization.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Experiment introduction video.
[MP4 File (MP4 Video), 159981 KB - mhealth_v12i1e48777_app1.mp4]

Multimedia Appendix 2
Detailed description of features.
[PDF File (Adobe PDF File), 252 KB - mhealth_v12i1e48777_app2.pdf]

Multimedia Appendix 3
Visualization of drawing sequences.
[PDF File (Adobe PDF File), 1111 KB - mhealth_v12i1e48777_app3.pdf]

References

1. Yates JA, Clare L, Woods RT. What is the relationship between health, mood, and mild cognitive impairment? J Alzheimers Dis 2016 Dec 06;55(3):1183-1193. [doi: 10.3233/jad-160611] [Medline: 27792011]
2. Petersen RC, Lopez O, Armstrong MJ, Getchius TS, Ganguli M, Gloss D, et al. Practice guideline update summary: mild cognitive impairment: report of the guideline development, dissemination, and implementation subcommittee of the American Academy of Neurology. Neurology 2018 Jan 16;90(3):126-135 [FREE Full text] [doi: 10.1212/WNL.0000000000004826] [Medline: 29282327]
3. Petersen RC, Roberts RO, Knopman DS, Boeve BF, Geda YE, Ivnik RJ, et al. Mild cognitive impairment: ten years later. Arch Neurol 2009 Dec 01;66(12):1447-1455 [FREE Full text] [doi: 10.1001/archneurol.2009.266] [Medline: 20008648]
4. Mitchell AJ, Shiri-Feshki M. Rate of progression of mild cognitive impairment to dementia--meta-analysis of 41 robust inception cohort studies. Acta Psychiatr Scand 2009 Apr 03;119(4):252-265. [doi: 10.1111/j.1600-0447.2008.01326.x] [Medline: 19236314]
5. Kinsella GJ, Mullaly E, Rand E, Ong B, Burton C, Price S, et al. Early intervention for mild cognitive impairment: a randomised controlled trial. J Neurol Neurosurg Psychiatry 2009 Jul 29;80(7):730-736. [doi: 10.1136/jnnp.2008.148346] [Medline: 19332424]
6. Livingston G, Sommerlad A, Orgeta V, Costafreda SG, Huntley J, Ames D, et al. Dementia prevention, intervention, and care. Lancet 2017 Dec;390(10113):2673-2734. [doi: 10.1016/s0140-6736(17)31363-6]
7. Gosztolya G, Vincze V, Tóth L, Pákáski M, Kálmán J, Hoffmann I. Identifying mild cognitive impairment and mild Alzheimer's disease based on spontaneous speech using ASR and linguistic features. Comput Speech Lang 2019 Jan;53:181-197. [doi: 10.1016/j.csl.2018.07.007]
8. Storey JE, Rowland JT, Basic D, Conforti DA. Accuracy of the clock drawing test for detecting dementia in a multicultural sample of elderly Australian patients. Int Psychogeriatr 2002 Sep 10;14(3):259-271. [doi: 10.1017/s1041610202008463] [Medline: 12475087]

9. Schmitt AL, Livingston RB, Galusha JM, Davis KM. Correlation between the bicycle drawing task and neuropsychological variables in patients referred for a dementia evaluation. Percept Mot Skills 2009 Dec 01;109(3):721-730. [doi: 10.2466/pms.109.3.721-730] [Medline: 20178271]
10. Müller S, Preische O, Heymann P, Elbing U, Laske C. Increased diagnostic accuracy of digital vs. conventional clock drawing test for discrimination of patients in the early course of Alzheimer's disease from cognitively healthy individuals. Front Aging Neurosci 2017 Apr 11;9:101 [FREE Full text] [doi: 10.3389/fnagi.2017.00101] [Medline: 28443019]
11. Dion C, Arias F, Amini S, Davis R, Penney D, Libon DJ, et al. Cognitive correlates of digital clock drawing metrics in older adults with and without mild cognitive impairment. J Alzheimers Dis 2020;75(1):73-83 [FREE Full text] [doi: 10.3233/JAD-191089] [Medline: 32250300]
12. Piers RJ, Devlin KN, Ning B, Liu Y, Wasserman B, Massaro JM, et al. Age and graphomotor decision making assessed with the digital clock drawing test: the Framingham heart study. J Alzheimers Dis 2017;60(4):1611-1620 [FREE Full text] [doi: 10.3233/JAD-170444] [Medline: 29036819]
13. Kim KW, Lee SY, Choi J, Chin J, Lee BH, Na DL, et al. A comprehensive evaluation of the process of copying a complex figure in early- and late-onset Alzheimer disease: a quantitative analysis of digital pen data. J Med Internet Res 2020 Aug 12;22(8):e18136 [FREE Full text] [doi: 10.2196/18136] [Medline: 32491988]
14. Müller S, Preische O, Heymann P, Elbing U, Laske C. Diagnostic value of a tablet-based drawing task for discrimination of patients in the early course of Alzheimer's disease from healthy individuals. J Alzheimers Dis 2017;55(4):1463-1469. [doi: 10.3233/JAD-160921] [Medline: 27858717]
15. Robens S, Heymann P, Gienger R, Hett A, Müller S, Laske C, et al. The digital tree drawing test for screening of early dementia: an explorative study comparing healthy controls, patients with mild cognitive impairment, and patients with early dementia of the Alzheimer type. J Alzheimers Dis 2019;68(4):1561-1574. [doi: 10.3233/JAD-181029] [Medline: 30909229]
16. Whitson HE, Potter GG, Feld JA, Plassman BL, Reynolds K, Sloane R, et al. Dual-task gait and Alzheimer's disease genetic risk in cognitively normal adults: a pilot study. J Alzheimers Dis 2018;64(4):1137-1148 [FREE Full text] [doi: 10.3233/JAD-180016] [Medline: 30010120]
17. Ansai JH, Aurichio TR, Rebelatto JR. Relationship between dual task walking, cognition, and depression in oldest old people. Int Psychogeriatr 2016 Jan;28(1):31-38. [doi: 10.1017/S1041610215000915] [Medline: 26132488]
18. Montero-Odasso MM, Sarquis-Adamson Y, Speechley M, Borrie MJ, Hachinski VC, Wells J, et al. Association of dual-task gait with incident dementia in mild cognitive impairment: results from the gait and brain study. JAMA Neurol 2017 Jul 01;74(7):857-865 [FREE Full text] [doi: 10.1001/jamaneurol.2017.0643] [Medline: 28505243]
19. Aoki K, Ngo TT, Mitsugami I, Okura F, Niwa M, Makihara Y, et al. Early detection of lower MMSE scores in elderly based on dual-task gait. IEEE Access 2019;7:40085-40094. [doi: 10.1109/access.2019.2906908]
20. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975 Nov;12(3):189-198. [doi: 10.1016/0022-3956(75)90026-6] [Medline: 1202204]
21. Ali N, Liu J, Tian H, Pan W, Tang Y, Zhong Q, et al. A novel dual-task paradigm with story recall shows significant differences in the gait kinematics in older adults with cognitive impairment: a cross-sectional study. Front Aging Neurosci 2022 Dec 15;14:992873 [FREE Full text] [doi: 10.3389/fnagi.2022.992873] [Medline: 36589542]
22. McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA work group under the auspices of department of health and human services task force on Alzheimer's disease. Neurology 1984 Jul;34(7):939-944. [doi: 10.1212/wnl.34.7.939] [Medline: 6610841]
23. Crowe SF. The differential contribution of mental tracking, cognitive flexibility, visual search, and motor speed to performance on parts A and B of the Trail Making test. J Clin Psychol 1998 Aug;54(5):585-591. [doi: 10.1002/(sici)1097-4679(199808)54:5<585::aid-jclp4>3.0.co;2-k] [Medline: 9696108]
24. Giacomin J. What is human centred design? Des J 2015 Apr 28;17(4):606-623. [doi: 10.2752/175630614x14056185480186]
25. Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, et al. The Montreal cognitive assessment, MoCA: a brief screening tool for mild cognitive impairment. J Am Geriatr Soc 2005 Apr 30;53(4):695-699. [doi: 10.1111/j.1532-5415.2005.53221.x] [Medline: 15817019]
26. Morris JC. The Clinical Dementia Rating (CDR): current version and scoring rules. Neurology 1993 Nov;43(11):2412-2414. [doi: 10.1212/wnl.43.11.2412-a] [Medline: 8232972]
27. Meyers JE, Meyers KR. Rey complex figure test and recognition trial professional manual. In: Psychological Assessment Resources. Florida: Odessa; 1995.
28. Beauchet O, Dubost V, Aminian K, Gonthier R, Kressig RW. Dual-task-related gait changes in the elderly: does the type of cognitive task matter? J Mot Behav 2005 Jul;37(4):259-264. [Medline: 15967751]
29. Ehsani H, Mohler MJ, O'Connor K, Zamrini E, Tirambulo C, Toosizadeh N. The association between cognition and dual-tasking among older adults: the effect of motor function type and cognition task difficulty. Clin Interv Aging 2019;14:659-669 [FREE Full text] [doi: 10.2147/CIA.S198697] [Medline: 31040655]
30. Shulman KI. Clock-drawing: is it the ideal cognitive screening test? Int J Geriatr Psychiatry 2000 Jun;15(6):548-561. [doi: 10.1002/1099-1166(200006)15:6<548::aid-gps242>3.0.co;2-u] [Medline: 10861923]

31. Lezak MD, Howieson DB, Loring DW, Tranel D. Neuropsychological Assessment. Oxford, UK: Oxford University Press; 2004.
32. Milner B. Psychological aspects of focal epilepsy and its neurosurgical management. Adv Neurol 1975;8:299-321. [Medline: 804234]
33. Salvadori E, Dieci F, Caffarra P, Pantoni L. Qualitative evaluation of the immediate copy of the Rey-Osterrieth Complex Figure: comparison between vascular and degenerative MCI patients. Arch Clin Neuropsychol 2019 Feb 01;34(1):14-23. [doi: 10.1093/arclin/acy010] [Medline: 29420698]
34. Ghaderyan P, Abbasi A, Saber S. A new algorithm for kinematic analysis of handwriting data; towards a reliable handwriting-based tool for early detection of Alzheimer's disease. Expert Syst Appl 2018 Dec;114:428-440. [doi: 10.1016/j.eswa.2018.07.052]
35. Carment L, Abdellatif A, Lafuente-Lafuente C, Pariel S, Maier MA, Belmin J, et al. Manual dexterity and aging: a pilot study disentangling sensorimotor from cognitive decline. Front Neurol 2018 Oct 29;9:910 [FREE Full text] [doi: 10.3389/fneur.2018.00910] [Medline: 30420830]
36. Tian F, Fan X, Fan J, Zhu Y, Gao J, Wang D, et al. What can gestures tell?: detecting motor impairment in early Parkinson's from common touch gestural interactions. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 2019 Presented at: CHI '19; May 4-9, 2019; Scotland, UK p. 1-14 URL: https://dl.acm.org/doi/10.1145/3290605.3300313 [doi: 10.1145/3290605.3300313]
37. Cohen J. A power primer. Psychol Bull 1992 Jul;112(1):155-159. [doi: 10.1037//0033-2909.112.1.155] [Medline: 19565683]
38. Chan JY, Wong A, Yiu B, Mok H, Lam P, Kwan P, et al. Electronic cognitive screen technology for screening older adults with dementia and mild cognitive impairment in a community setting: development and validation study. J Med Internet Res 2020 Dec 18;22(12):e17332 [FREE Full text] [doi: 10.2196/17332] [Medline: 33337341]
39. Zhang W, Zheng X, Tang Z, Wang H, Li R, Xie Z, et al. Combination of paper and electronic trail making tests for automatic analysis of cognitive impairment: development and validation study. J Med Internet Res 2023 Jun 09;25:e42637 [FREE Full text] [doi: 10.2196/42637] [Medline: 37294606]
40. Skirrow C, Meszaros M, Meepegama U, Lenain R, Papp KV, Weston J, et al. Validation of a remote and fully automated story recall task to assess for early cognitive impairment in older adults: longitudinal case-control observational study. JMIR Aging 2022 Sep 30;5(3):e37090 [FREE Full text] [doi: 10.2196/37090] [Medline: 36178715]
41. Alegret M, Muñoz N, Roberto N, Rentz DM, Valero S, Gil S, et al. A computerized version of the short form of the Face-Name Associative Memory Exam (FACEmemory®) for the early detection of Alzheimer's disease. Alzheimers Res Ther 2020 Mar 16;12(1):25 [FREE Full text] [doi: 10.1186/s13195-020-00594-6] [Medline: 32178724]
42. Polit DF, Beck CT. Nursing Research: Generating and Assessing Evidence for Nursing Practice. Philadelphia, PA: Lippincott Williams & Wilkins; 2008.
43. McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR, Kawas CH, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011 May 22;7(3):263-269 [FREE Full text] [doi: 10.1016/j.jalz.2011.03.005] [Medline: 21514250]
44. Chapman LJ, Chapman JP. The measurement of handedness. Brain Cogn 1987 Apr;6(2):175-183. [doi: 10.1016/0278-2626(87)90118-7] [Medline: 3593557]
45. Fung AW, Lam LC. Validation of a computerized Hong Kong - vigilance and memory test (HK-VMT) to detect early cognitive impairment in healthy older adults. Aging Ment Health 2020 Jan 01;24(1):186-192. [doi: 10.1080/13607863.2018.1523878] [Medline: 30270640]
46. Cheah WT, Chang WD, Hwang JJ, Hong SY, Fu LC, Chang YL. A screening system for mild cognitive impairment based on neuropsychological drawing test and neural network. In: Proceedings of the 2019 IEEE International Conference on Systems, Man and Cybernetics. 2019 Presented at: SMC '19; October 6-9, 2019; Bari, Italy p. 3543-3548 URL: https://ieeexplore.ieee.org/abstract/document/8913880 [doi: 10.1109/smc.2019.8913880]
47. Müller S, Herde L, Preische O, Zeller A, Heymann P, Robens S, et al. Diagnostic value of digital clock drawing test in comparison with CERAD neuropsychological battery total score for discrimination of patients in the early course of Alzheimer's disease from healthy individuals. Sci Rep 2019 Mar 05;9(1):3543 [FREE Full text] [doi: 10.1038/s41598-019-40010-0] [Medline: 30837580]
48. Cabinio M, Rossetto F, Isernia S, Saibene FL, Di Cesare M, Borgnis F, et al. The use of a virtual reality platform for the assessment of the memory decline and the hippocampal neural injury in subjects with mild cognitive impairment: the validity of Smart Aging Serious Game (SASG). J Clin Med 2020 May 06;9(5):1355 [FREE Full text] [doi: 10.3390/jcm9051355] [Medline: 32384591]
49. Eraslan Boz H, Limoncu H, Zygouris S, Tsolaki M, Giakoumis D, Votis K, et al. A new tool to assess amnestic mild cognitive impairment in Turkish older adults: virtual supermarket (VSM). Neuropsychol Dev Cogn B Aging Neuropsychol Cogn 2020 Sep 04;27(5):639-653. [doi: 10.1080/13825585.2019.1663146] [Medline: 31482749]
50. Zhang Y, Chen Y, Yu H, Lv Z, Yang X, Hu C, et al. What can "drag and drop" tell? Detecting mild cognitive impairment by hand motor function assessment under dual-task paradigm. Int J Hum Comput Stud 2021 Jan;145:102547. [doi: 10.1016/j.ijhcs.2020.102547]

51. Li A, Xue C, Wu R, Wu W, Zhao J, Qiang Y. Unearthing subtle cognitive variations: a digital screening tool for detecting and monitoring mild cognitive impairment. Int J Hum Comput Interact 2024 Apr 04:1-21. [doi: 10.1080/10447318.2024.2327179]
52. Renier M, Gnoato F, Tessari A, Formilan M, Busonera F, Albanese P, et al. A correlational study between signature, writing abilities and decision-making capacity among people with initial cognitive impairment. Aging Clin Exp Res 2016 Jun 2;28(3):505-511. [doi: 10.1007/s40520-016-0549-y] [Medline: 26936371]
53. Parsey CM, Schmitter-Edgecombe M. Quantitative and qualitative analyses of the clock drawing test in mild cognitive impairment and Alzheimer disease: evaluation of a modified scoring system. J Geriatr Psychiatry Neurol 2011 Jun 05;24(2):108-118. [doi: 10.1177/0891988711402349] [Medline: 21546651]
54. Casarotti A, Papagno C, Zarino B. Modified Taylor Complex Figure: normative data from 290 adults. J Neuropsychol 2014 Sep 06;8(2):186-198. [doi: 10.1111/jnp.12019] [Medline: 23647550]
55. Piau A, Wild K, Mattek N, Kaye J. Current state of digital biomarker technologies for real-life, home-based monitoring of cognitive function for mild cognitive impairment to mild Alzheimer disease and implications for clinical care: systematic review. J Med Internet Res 2019 Aug 30;21(8):e12785 [FREE Full text] [doi: 10.2196/12785] [Medline: 31471958]

Abbreviations

AD: Alzheimer dementia
CDR: Clinical Dementia Rating
CDT: Clock Drawing Test
dCDT: digital Clock Drawing Test
HC: healthy control
MCI: mild cognitive impairment
MMSE: Mini-Mental State Examination
MoCA: Montreal Cognitive Assessment
RCFT: Rey-Osterrieth Complex Figure Test
ROCF: Rey-Osterrieth Complex Figure
sRCFT: simplified Rey-Osterrieth Complex Figure Test

Edited by L Buis; submitted 06.05.23; peer-reviewed by S Cheng, L Brown, L Yao, N Xiao, A Guaita; comments to author 23.02.24; revised version received 07.03.24; accepted 22.04.24; published 26.06.24.

Please cite as:
Li A, Li J, Chai J, Wu W, Chaudhary S, Zhao J, Qiang Y
Detection of Mild Cognitive Impairment Through Hand Motor Function Under Digital Cognitive Test: Mixed Methods Study
JMIR Mhealth Uhealth 2024;12:e48777
URL: https://mhealth.jmir.org/2024/1/e48777
doi:10.2196/48777
PMID:

Review

Using mHealth Technologies for Case Finding in Tuberculosis and Other Infectious Diseases in Africa: Systematic Review

Don Lawrence Mudzengi[1,2], BSc, MSc; Herbert Chomutare[3], BSc, MPH; Jeniffer Nagudi[1], BPharm, MSc; Thobani Ntshiqa[1,2], BCur, MPH, MBA; J Lucian Davis[4], MD; Salome Charalambous[1,2], MSc, MD, PhD; Kavindhran Velen[1], BSc, MSc, PhD

[1]The Aurum Institute, Johannesburg, South Africa

[2]School of Public Health, University of the Witwatersrand, Johannesburg, South Africa

[3]Ministry of Health LISTEN Project, Mbabane, Swaziland

[4]Department of Epidemiology of Microbial Diseases, Yale School of Public Health, Connecticut, CT, United States

Corresponding Author:
Don Lawrence Mudzengi, BSc, MSc
The Aurum Institute
29 Queens Road
Parktown
Johannesburg, 2193
South Africa
Phone: 27 833697946
Email: dmudzengi@auruminstitute.org

Abstract

Background: Mobile health (mHealth) technologies are increasingly used in contact tracing and case finding, enhancing and replacing traditional methods for managing infectious diseases such as Ebola, tuberculosis, COVID-19, and HIV. However, the variations in their development approaches, implementation scopes, and effectiveness introduce uncertainty regarding their potential to improve public health outcomes.

Objective: We conducted this systematic review to explore how mHealth technologies are developed, implemented, and evaluated. We aimed to deepen our understanding of mHealth's role in contact tracing, enhancing both the implementation and overall health outcomes.

Methods: We searched and reviewed studies conducted in Africa focusing on tuberculosis, Ebola, HIV, and COVID-19 and published between 1990 and 2023 using the PubMed, Scopus, Web of Science, and Google Scholar databases. We followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines to review, synthesize, and report the findings from articles that met our criteria.

Results: We identified 11,943 articles, but only 19 (0.16%) met our criteria, revealing a large gap in technologies specifically aimed at case finding and contact tracing of infectious diseases. These technologies addressed a broad spectrum of diseases, with a predominant focus on Ebola and tuberculosis. The type of technologies used ranged from mobile data collection platforms and smartphone apps to advanced geographic information systems (GISs) and bidirectional communication systems. Technologies deployed in programmatic settings, often developed using design thinking frameworks, were backed by significant funding and often deployed at a large scale but frequently lacked rigorous evaluations. In contrast, technologies used in research settings, although providing more detailed evaluation of both technical performance and health outcomes, were constrained by scale and insufficient funding. These challenges not only prevented these technologies from being tested on a wider scale but also hindered their ability to provide actionable and generalizable insights that could inform public health policies effectively.

Conclusions: Overall, this review underscored a need for organized development approaches and comprehensive evaluations. A significant gap exists between the expansive deployment of mHealth technologies in programmatic settings, which are typically well funded and rigorously developed, and the more robust evaluations necessary to ascertain their effectiveness. Future research should consider integrating the robust evaluations often found in research settings with the scale and developmental rigor of programmatic implementations. By embedding advanced research methodologies within programmatic frameworks at the design thinking stage, mHealth technologies can potentially become technically viable and effectively meet specific contact tracing health outcomes to inform policy effectively.

(JMIR Mhealth Uhealth 2024;12:e53211) doi:10.2196/53211

KEYWORDS

mobile health; mHealth; design thinking; tuberculosis; Ebola; HIV; COVID-19; infectious diseases; contact tracing; mobile phone

Introduction

Background

Mobile health (mHealth) technologies have become increasingly popular tools for facilitating data collection and delivery of health services worldwide [1,2]. The emergence of COVID-19 in 2020 increased interest in using mHealth technologies for contact tracing due to their ability to achieve case finding goals without the need for physical contact with an infected person [1,2]. Before COVID-19, mHealth technologies had already been used in Africa for contact tracing of infectious diseases such as Ebola [3-7], tuberculosis [8-11], and HIV, albeit with limited success. Compared to paper-based systems, the most apparent advantages of mHealth include its ability to reduce repetitive tasks and errors, systematic delivery of services, and improved monitoring due to efficient data processing in databases. Therefore, at face value, mHealth technologies have the potential to overcome challenges encountered when using traditional paper-based systems for contact tracing, thereby improving outcomes.

Contact tracing is a strategy for actively and systematically screening for symptoms among individuals exposed to someone with a transmissible disease to determine whether they require further diagnostic evaluation [12-14]. A key advantage of contact tracing is that, in principle, it reduces the time from when an individual falls ill with an infectious disease to when they are diagnosed, preventing further transmission to healthy persons. A contact, defined as any person living with or someone in a social circle who has regular contact with an individual with a transmissible disease, is at the highest risk of contracting the same disease due to their proximity to the infected person. Therefore, contacts are defined as a high-risk priority group for contact tracing [12]. The World Health Organization has long recommended and supported contact tracing for tuberculosis and Ebola [15,16] and promotes contact tracing as a critical intervention for tuberculosis control in high-burden countries [16].

Tuberculosis contact tracing can be seen as a cascade of activities that begins with finding an individual with the disease of interest, referred to as the index patient, and collecting information about their close contacts [12]. These activities are followed by a household visit to screen the enumerated persons for tuberculosis symptoms and may include collecting sputum samples from symptomatic contacts for laboratory testing—anyone testing positive for tuberculosis is referred to health facilities for linkage to care [12]. During household visits, contact tracing also serves as a pathway for accessing household contacts eligible for and initiating tuberculosis preventive therapy [14]. There are variations to contact tracing, with recent modalities using portable chest x-rays for identifying individuals eligible for tuberculosis testing [17,18] and oral swabs as an alternative to sputum samples [19,20]. In South Africa, contact tracing has now evolved from testing only symptomatic persons to universal testing of all household contacts and, in the process, initiating tuberculosis preventive therapy in eligible contacts [21].

The process for conducting contact tracing is rigorous. Each step in the contact tracing cascade requires documentation to enable contact follow-up, communication of results, linkage to care, and program monitoring and evaluation. Contact tracing programs rely significantly on efficient data collection to inform decision-making and patient management—in the absence of this, the process loses its cost-effectiveness and may become unattractive to national tuberculosis programs. Within Africa, countries have relied on inefficient paper-based data collection that overburdens outreach workers responsible for tracing household contacts but also results in poor data quality due to inevitable human error and inadequate accountability due to the manual processes required to collate field data [12,22-25]. These systems could be improved or overhauled by introducing mHealth technologies to optimize the documentation of activities in each step of the cascade.

Objectives

Despite the potential of mHealth to overcome traditional paper-based system challenges, there is evidence suggesting that improvements in contact tracing outcomes using mHealth technologies remain insufficient and “largely unproven” [24,26-28]. To address the evidence gaps, this systematic review assessed technologies on tuberculosis, Ebola, HIV, and COVID-19. While our focus remains predominantly on tuberculosis, we included these additional diseases due to similarities in contact tracing methods, enhancing the wider applicability of our findings. We aimed to synthesize information on the development, implementation, and evaluation of these technologies to better guide future work and enhancements. We hypothesized that synthesizing this existing information on the technologies will reveal valuable insights into the continuum of mHealth apps, their benefits, and limitations, thereby shaping improvements in future contact tracing efforts.

Methods

Literature Search

We conducted a systematic review of mHealth technologies used for contact tracing for selected infectious diseases and reported the results using the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. We conducted an iterative search of all studies published in or translated into English from January 1, 1990, to December 31, 2023, in electronic databases, including PubMed (MEDLINE), Scopus, Web of Science, and Google Scholar (Multimedia Appendix 1). The year 1990 was chosen because it represents the earliest period during which the literature suggests that digital phones were introduced and available for use in health [29,30]. In addition, the 2023 cutoff allowed for the inclusion

of technologies that may have been used during the COVID-19 pandemic. The following search terms were applied within the selected databases: *tuberculosis*, *mHealth* or *mobile health*, *telehealth* or *telemedicine* and *case finding* or *contact tracing*/*investigation* or *tuberculosis screening* or *COVID-19* or *Ebola* or *HIV*. We used Medical Subject Heading (MeSH) terms to search in PubMed. We used subject headings and keywords in databases that do not use MeSH terms, such as Scopus, Web of Science, and Google Scholar. The search terms are presented in Multimedia Appendix 1. The Boolean operators "AND," "OR," and "NOT" were used to join words in all databases to improve the accuracy and relevance of the literature.

Inclusion and Exclusion Criteria

The following study inclusion criteria were applied: (1) studies conducted in Africa and published between 1990 and 2023 (2) that included an mHealth technology either as an intervention or part of procedures for contact tracing (3) used to screen for tuberculosis, COVID-19, Ebola, and HIV or find and screen contacts of people infected with these diseases.

The following studies were excluded: (1) mHealth modeling studies; (2) mHealth protocols and proposals; and (3) mHealth systematic reviews, commentaries, and scoping reviews.

Screening of the Literature, Extraction, and Analysis

The initial database search was conducted by DLM, and all references were uploaded to a reference management software library (EndNote version 20; Clarivate Analytics) for abstract and title screening. In total, 2 reviewers (DLM and HC) performed the initial review of all articles according to the inclusion and exclusion criteria. Where there was disagreement on the classification of a reference, a third reviewer (JN) conducted an additional review and confirmed the final classification.

A web-based data extraction tool was developed on Microsoft Office 365 Forms (Microsoft Corp; Multimedia Appendix 2). Extracted data elements included the study title, year of publication, country, location where the technology was implemented (community, facility, or both), study design, target disease, and type of technology used. The same form also contained sections to capture qualitative data about how the technologies were developed and their implementation processes, including challenges and the outcomes to measure effectiveness.

We summarized and synthesized the systematic review results using the Joanna Briggs Institute approach [31], and the data were presented in tables and narrative text. The themes used in the analysis were predetermined: development, implementation, and outcomes. The development theme described the steps taken in the development of the technologies. The implementation theme described how the technologies were deployed, what and how they collected contact tracing data, and challenges and successes. The final theme focused on the contact tracing outcomes measured when mHealth technologies were used.

Quality Assessments and Risk of Bias

The included studies were assessed for risk of bias using tools appropriate for each study type. The steps followed in assessing risk are detailed in Multimedia Appendix 3 [6,8,27,32-48]. The Cochrane Risk of Bias 2 tool was used for randomized clinical trials [49]. Technologies used in programmatic settings and pretest-posttest studies were assessed using the Quality Assessment Tool for Before-After (pretest-posttest) Studies With No Control Group developed by the US National Heart, Lung, and Blood Institute [50]. Finally, cross-sectional studies were assessed using the Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies also from the National Heart, Lung, and Blood Institute [50].

Definitions

The World Health Organization broadly defines eHealth as "the use of information and communications technology (ICT) in support of health and health-related fields, including health care services, health surveillance, health literature, health education, knowledge and research" [51].

mHealth is a branch of eHealth that refers to mobile wireless technologies to support public health objectives [51,52]. mHealth technologies are mainly used on portable devices such as phones and tablets [52,53] and allow for the ubiquitous provision of services such as public health surveillance, sharing of clinical information, data collection, health behavior communication, and the use of mobile technologies to reach health goals [54].

"Programmatic pretest-posttest" describes the use of technologies in programmatic settings to deliver an interventions to large populations targeting specific outputs and outcomes.

"In-house software platform" refers to bespoke technologies without using existing platforms. For example, many public health projects build forms on platforms such as Open Data Kit (ODK), Epi Info, and REDCap (Research Electronic Data Capture; Vanderbilt University), and in this review, these were not considered in-house software platforms.

"The effectiveness of an mHealth technology used in programmatic settings" is defined in the context of a project or program designed to implement an intervention with a target and without the use of research principles. Effectiveness in this sense is the achievement of a set target, such as the number of people screened using an mHealth technology.

"Effectiveness in a research project of an mHealth technology" is defined as the improvement of a contact tracing outcome before and after, such as the number of people diagnosed when screened using an mHealth technology.

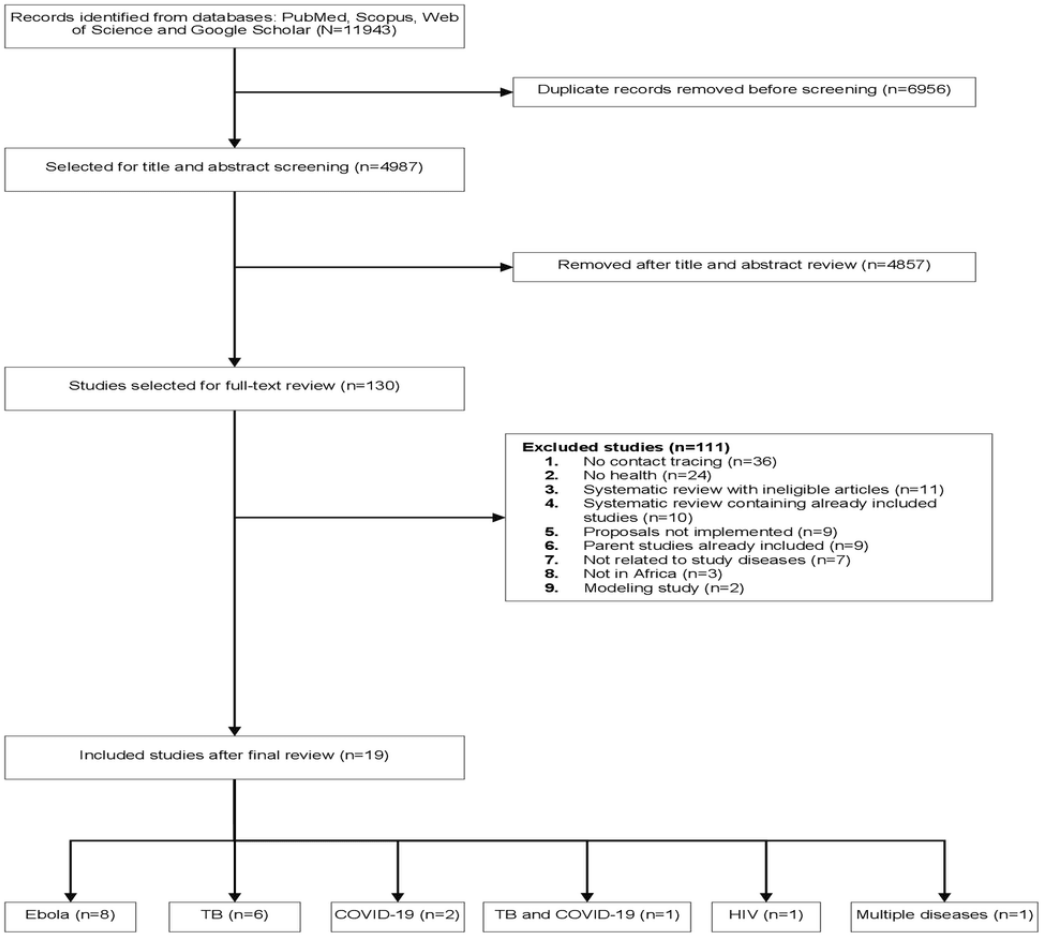
Results

Literature Search

Figure 1 provides a PRISMA flowchart for the literature search and review process. The initial literature search yielded 11,943 articles. After removing 58.24% (6956/11,943) of duplicates, 41.76% (4987/11,943) of the articles remained for screening. A review of titles and abstracts led to the exclusion of 97.39%

(4857/4987) of the articles, resulting in 130 articles being selected for full-text review. Only 19 studies were retained after the full-text review [6,8,27,32-47]. Of these 19 studies, 8 (42%) were on Ebola [32-39], 6 (32%) were on tuberculosis [6,27,40-43], 2 (11%) were on COVID-19 [44,45], 1 (5%) was on HIV [8], 1 (5%) was on tuberculosis and COVID-19 [46], and 1 (5%) was on multiple diseases in humans and animals [47]. No study needed translation.

Figure 1. The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram for the systematic review. TB: tuberculosis.



Description of the Studies

Table 1 provides an overview of the key characteristics of the 19 studies that formed the basis of this systematic review. The studies covered a range of diseases, including Ebola, tuberculosis, COVID-19, HIV, and multiple diseases in humans and animals. In total, 74% (14/19) of the included studies were about technologies used for contact tracing of Ebola and tuberculosis. Ebola had the most studies with 42% (8/19), whereas tuberculosis had 32% (6/19). One other technology was used for both tuberculosis and COVID-19. The remaining 4 studies discussed technologies for the contact tracing of HIV (n=1, 25%), COVID-19 (n=2, 50%), and multiple diseases among humans and animals (n=1, 25%). Multimedia Appendix 4 [6,8,27,32-47,55-75] provides a detailed summary of each technology.

Table 1. Description of all retained studies.

Disease and target users	Technology type	Study	Country	Study design	Location
Ebola					
Health workers	Smartphone app, GIS[a], and BI[b] dashboard	Tom-Aba et al [34], 2015	Nigeria	Programmatic pretest-posttest	Community
Health workers	Smartphone app and BI dashboard	Sacks et al [32], 2015	Guinea	Programmatic pretest-posttest	Community
Contacts	SMS text messaging and phone calls	Jia and Mohamed [33], 2015	Sierra Leone	Programmatic pretest-posttest	Community
Health workers	Smartphone app	Adeoye et al [37], 2017	Nigeria	Programmatic pretest-posttest	Community
Contacts	Phone calls	Alpren et al [36], 2017	Sierra Leone	Programmatic pretest-posttest	Community
Health workers	Mobile tracking	Wolfe et al [35], 2017	Liberia	Programmatic pretest-posttest	Community
Health workers	Smartphone app	Danquah et al [38], 2019	Sierra Leone	Cross-sectional	Community
Health workers	Smartphone app	Whitesell et al [39], 2021	DRC[c]	Programmatic pretest-posttest	Community
Tuberculosis					
Health workers	Smartphone app and GIS	Chisunkha et al [6], 2016	Malawi	Cross-sectional	Community
Health workers	Smartphone app	Ha et al [40], 2016	Botswana	Pretest-posttest	Community
Contacts	USSD[d]	Diaz and Moturi [42], 2019	Tanzania	Programmatic pretest-posttest	Community
Health workers	Smartphone app	Davis et al [27], 2019	Uganda	Trials	Used or intended for use in both the facility and community
Health workers	Smartphone app	Szkwarko et al [41], 2021	Kenya	Cross-sectional	Facility
Both	Smartphone app	URC[e] [43], 2020	South Africa	Programmatic pretest-posttest	Used or intended for use in both the facility and community
COVID-19					
Contacts	Web application and USSD	Owoyemi et al [45], 2021	Nigeria	Programmatic pretest-posttest	Community
Contacts	Smartphone app	Mugenyi et al [44], 2021	Uganda	Pretest-posttest	Community
Tuberculosis and COVID-19					
Contacts	USSD and WhatsApp	Praekelt.org [46], 2021	South Africa	Programmatic pretest-posttest	Community
HIV					
Health workers	Smartphone app	Rajput et al [8], 2012	Kenya	Programmatic pretest-posttest	Community
Multiple diseases					
Health workers	Smartphone app	Karimuribo et al [47], 2017	Tanzania	Programmatic pretest-posttest	Community

[a]GIS: geographic information system.

[b]BI: business intelligence.

[c]DRC: Democratic Republic of the Congo.

[d]USSD: unstructured supplementary service data.

[e]URC: University Research Co.

The type of mHealth technology varied across the selected studies. Exactly 42% (5/19) used communications platforms with SMS and phone calls accounting for 11% (2/19) [33,36] and unstructured supplementary service data or WhatsApp used in 16% (3/19) [42,45,46]. Smartphone apps were the most common, appearing in 68% (13/19) of the studies [6,8,27,32,34,37-41,43,44,47]. Among these smartphone apps, 25% (3/12) incorporated geographic information systems [6,34] or a business intelligence dashboard [32,34]. Only one out of the 19 (5%) used a mobile tracking system managed through cellphone tower technologies [35].

Most of the technologies (16/19, 84%) [6,8,32-40,42,44-47] were designed for use in the community during outreach visits, whereas 11% (2/19) were used in both the community [27,43] and the household and only 5% (1/19) were designed for use in a health facility [41]. Of all 19 technologies, 11 (58%) were designed for use by outreach workers to capture data during contact tracing [6,8,27,32,34,37-41,43,47], whereas 7 (37%) were designed for self-screening disease symptoms by the contacts [33,35,36,42,44-46] and only 1 (5%) could be used by both patients and outreach workers [43]. In contact-targeted technologies, contacts interacted with the system by sending messages, responding to built-in prompts, and making or receiving phone calls for contact tracing, whereas outreach workers used the technologies to facilitate data collection. Only 5% (1/19) of the studies used a technology to facilitate linkage to care of household contacts, a distal outcome from contact tracing [27].

Regarding study designs, 68% (13/19) of the studies that implemented technologies in programmatic settings used pretest-posttest designs. In these cases, the technology served as an intervention aimed at achieving specific contact tracing or case finding objectives within broad population programs, often without the technology of stringent research methodologies. The remaining 6 studies used various designs: 3 (50%) were pretest-posttest studies, 2 (33%) were cross-sectional studies, and 1 (17%) was a randomized trial.

The Uses of Technologies in the Included Studies

The uses of the technologies in the included studies are also presented in Table 1. As predetermined in the inclusion criteria, all technologies were designed to assist in contact tracing and infectious disease case finding. However, some were primarily designed for disease surveillance incorporating contact-tracing functionality [41,45,47,76]. The predominant contact tracing modality in the technologies involved identifying index patients, finding their contacts, and collecting data using mobile technologies instead of paper.

Other studies used unconventional contact tracing methods that did not necessarily require initial contact with index patients or arrangements with household contacts. For instance, in a study conducted in Malawi by Chisunkha et al [6], Google Earth, a publicly available software, was used to identify and visit households for recruitment into a chronic airway disease and tuberculosis trial even in the absence of previous knowledge of index patients. This approach involved the use of global positioning technology to locate households within a disease hot spot area on a map. Subsequently, data collectors were dispatched to these identified locations to recruit participants for the trial. While the system was not exclusively designed as a contact tracing project but rather as a tool to facilitate the location of participants within a study, the authors proposed that it could also serve contact tracing purposes in remote settings. This approach allowed for the swift identification of households before actual visits, thereby reducing costs and enhancing efficiency [6]. In Monrovia, Liberia, Wolfe et al [35] described an Ebola contact tracing system in which the government issued subpoenas to cell phone companies forcing them to provide locations of contacts that outreach workers could not locate. The authors reported that these subpoenas assisted in successfully locating 29 missing contacts [35]. In Sierra Leone, Alpren et al [36] described a repurposed call center, also for Ebola, in which the public acted as primary informants to authorities to report details and locations of known or suspected patients with Ebola (live alerts) and deaths (death alerts). The study showed >10,000 weekly alerts at the peak of the Ebola outbreak in October 2014. Cumulatively, between 2014 and 2016, the call center received 248,789 death alerts and 95,136 live alerts [36]. Finally, Mugenyi et al [44] developed and implemented the Wetaase app in a small pilot study to monitor the symptoms and movement of household members during the COVID-19 outbreak in Uganda. Unique to this technology was that household members downloaded and self-reported daily symptoms through Wetaase, and if symptomatic, they were tested for COVID-19 by workers visiting the household. Only 101 participants were enrolled, and out of an expected daily report of 8949 in 90 days, the app achieved 6617 reports, a use rate of 78%. Of these 6617 reports, only 57 (0.8%) self-reported COVID-19 symptoms, and no cases were diagnosed.

Development of mHealth Technologies for Contact Tracing

Type of Platform for Development

In total, 84% (16/19) of the technologies in the included studies were digital technologies developed to work on mobile phones. Of these 16 technologies, 11 (69%) were developed by customizing existing platforms such as ODK and CommCare, while 6 (38%) were developed as in-house software platforms. Of the 11 customized technologies, ODK was used to develop 5 (45%), and CommCare was used to develop 3 (27%) [27,32,38], while Unstructured Supplementary Service Data was used to develop 2 (18%) technologies and only 1 (9%) was developed using KoboToolbox. The remaining 16% (3/19) of the technologies used existing cellular network infrastructure such as SMS text messaging and phone calls. Of these 3 technologies that used existing cellular network infrastructure, 2 (67%) used phone calls, while 1 (33%) used cell phone tower technology.

Scope and Scale of Implementation

The scope and implementation scale of the technologies in the included studies varied depending on the intended use, that is,

in programmatic or research settings. Technologies used in programmatic settings of contact tracing activities had a larger scope aimed at targeting large sections of the population to meet contact tracing targets of the program. Large-scope technologies in the selected studies included the Surveillance Outbreak Response Management and Analysis System (SORMAS) in Nigeria, the Academic Model Providing Access to Healthcare (AMPATH) in Kenya, Tanzania's Tambua tuberculosis and Afyadata, tuberculosis HealthCheck in South Africa, and the Ebola contact tracing app in Guinea. For example, AMPATH in Kenya was developed to facilitate screening among >2 million individuals in 3 years [8]. Similarly, HealthCheck in South Africa was initially designed for population-wide self-screening of COVID-19 and expanded to tuberculosis at a national level [46]. In Tanzania, Afyadata, which veterinary specialists conceptualized, had a large scope for screening and reporting diseases in animals and humans [47,55]. Among humans, Afyadata has been used for finding diseases such as Ebola [47,77,78], COVID-19 [56,79], cholera [55,80,81], brucellosis [82,83], and an impetigo-like outbreak [47,55,84]. In contrast, technologies developed for research studies had a narrow scope, mainly built to demonstrate the value of mHealth in research settings or as research data collection tools, and had no evidence of uptake beyond the research. In Botswana, Ha et al [40] demonstrated that mHealth was better than paper by developing a mobile technology using ODK, but only 376 contacts were screened before the technology was probably retired. Chisunkha et al [6] developed a mobile technology using ODK to enumerate households for a tuberculosis contact tracing trial in Malawi and a small population. We did not find any further information about the tenure of this mHealth technology after the enumeration of the households. In Kenya, Szkwarko et al [41] developed PPTBMAPP to facilitate tuberculosis screening of 276 children in a health facility and demonstrated its superiority over paper, but no further use is documented. Also in Uganda, a randomized trial used an mHealth technology developed using CommCare as a data collection and screening tool and then evaluated the effectiveness of SMS text messaging to complete tuberculosis evaluation within 2 weeks. This technology screened only 919 contacts in the intervention and control arms [27].

Funding for Developing mHealth Technologies

Large-scale technologies developed for programmatic settings often received substantial funding for extensive development and implementation processes. For instance, SORMAS received €850,000 (US $929,509) from the European Union [85], Afyadata obtained US $450,000 from the Skoll Global Threats Fund [55], and the CommCare app for HIV screening was supported by a US $74.9–million US Agency for International Development–AMPATH program [86]. Tambua tuberculosis received government backing through Tanzania's National Tuberculosis and Leprosy Programme, which facilitated support and acceptance from end users [42].

In contrast, many research projects included in this review did not report their funding sources. This lack of reported funding highlights the resource disparities between technologies used in programmatic settings and research studies. Research projects often operated with limited financial resources, which impacted the scope and scale of their technology development processes. For example, in Botswana, Ha et al [40] evaluated a pragmatic mHealth technology for tuberculosis contact tracing. While the technology demonstrated promising outcomes compared to traditional paper-based methods, its application was limited to a pilot study across 7 urban and semiurban health facilities, and there is no record of its broader implementation, underscoring a challenge of limited scope in assessing the technology's full potential and scalability [40].

Participatory Development Processes as an Indicator of Successful Implementation

In total, 21% (4/19) of the technologies—SORMAS [57,76], Afyadata [47,55], AMPATH [8], and Tambua tuberculosis [42]—documented the use of design thinking frameworks in their development processes. Design thinking frameworks increase the likelihood of producing technologies that are fit for the intended purposes by using participatory processes to capture all the requirements and what outcomes the technology must deliver and how and also address the context in which the technology will be used [57,58]. Design thinking also involves extensive consultation with relevant stakeholders such as public health experts, government policy makers, health workers, and information technologists.

For example, Afyadata was intended to be a surveillance technology for use among humans and animals to monitor common diseases in the 2 fields. Professionals from human and animal health, as well as software developers, extensively collaborated to elicit project needs and goals. Relevant government officials were also involved and endorsed the technology throughout all its development and implementation stages. As a result, Afyadata is now widely used for surveillance purposes in Tanzania and has been used in the surveillance of cholera cases, monitoring of hygiene and sanitation practices, early detection of Ebola in neighboring Democratic Republic of the Congo, and surveillance of diseases occurring among animals. Afyadata was also adopted for the screening of COVID-19 in neighboring Mozambique.

Similarly, Nigeria's SORMAS followed a design thinking framework as a response to its precursor, the Sense Follow-up app. The Sense Follow-up app was an Android mobile technology rapidly developed in an emergency to support data collection during home visits following contacts of patients with Ebola for 21 days to document people who had been in contact with index patients with Ebola and support data collection during the first Ebola outbreak in Nigeria in 2014 [34,87]. However, the development was not sufficiently consultative because of time constraints imposed by the quick response time required to control Ebola. It was later deemed to have failed to meet the needs of the Ebola control team, specifically that it did not sufficiently support the bidirectional exchange of information; did not address case finding [57]; and had complex data manageability because of its modular architecture with separate systems for data storage, functions, and format and its interface [34,87]. The same team that developed the Sense Follow-up app then teamed up with philanthropists and public health and IT experts in designing SORMAS following a design thinking approach. The design thinking framework guided the SORMAS

development team to identify system, user, and technical requirements and addressed all the shortcomings of the Sense Follow-up app. Although there is limited evidence of its evaluation, SORMAS is now a widely used technology for contact tracing various diseases in Nigeria. In addition, because it is easily customizable, SORMAS was central in response to COVID-19.

Similar to Afyadata and SORMAS, the technology for HIV case finding in the AMPATH program with clear objectives was able to screen thousands as had been originally planned [8]. It was designed to be reliable in resource-constrained settings, scalable to allow for screening of >2 million people in 3 years, and open-source software; easily integrate with other systems and devices; and have GPS capabilities. As a result, public health experts, software technologists, and government stakeholders contributed to successfully developing a technology to screen >1 million HIV contacts.

However, in Guinea, the Ebola contact tracing technology was developed using participatory processes and design thinking but lacked government policy maker buy-in, which led to suboptimal implementation, and most of the intended results were not achieved [32]. Tuberculosis HealthCheck and COVID-19 alert technology in South Africa did not have documented participatory development, but government support led to their nationwide rollout [46].

All technologies used in research studies were developed by research teams and did not have any documented use of design thinking frameworks or participatory development. These technologies were used to reach research objectives, and their tenure did not extend beyond the project. However, the research team in Uganda that implemented the trial by Davis et al [27] conducted a post hoc analysis of the trial using an implementation science framework, the Consolidated Framework for Implementation Research, to understand the reasons for the poor performance of the mHealth technology on their main outcome [59]. While not explicitly mentioning design thinking as a framework to follow, the researchers conceded that the lack of consultation with local stakeholders was a critical gap in development that may have deprived them of a robust outer setting, which is a critical indicator for successful implementation even in smaller research studies.

Effectiveness of the mHealth Technologies and Reporting of Outcomes

Table 2 provides an overview of outcome reporting and technology development in the included studies. Technologies used in programmatic settings with a pretest-posttest setup were designed to reach specific targets and mainly reported outputs without comparisons, whereas research studies designed to answer a specific question reported both outputs and outcomes. Hence, effectiveness in the 2 approaches was defined differently. In technologies used for programmatic settings, effectiveness was defined as rolling out the technology to a large population and producing output. In research studies, effectiveness was defined as improving outcomes.

Table 2. Outcomes and app development.

Study title (design)	Output or outcomes	Development and evaluation

"Home-based tuberculosis contact investigation in Uganda: a household randomised trial" [27] (randomized study)	• Primary outcome: completion of tuberculosis evaluation within 14 days of enrollment among contacts requiring additional evaluation for tuberculosis
	• Secondary outcomes:
	  • Tuberculosis diagnosis and treatment initiation
	  • HIV diagnosis and linkage to HIV care
		The app was developed by public health professionals with experience in implementing and evaluating mHealth[a] technologies and by Dimagi CommCare specialists.

"Use of a mobile technology for Ebola contact tracing and monitoring in Northern Sierra Leone: a proof-of-concept study" [38] (cross-sectional study)	• Number of Ebola contacts identified and traced per arm • Time to find contact • Ebola cases detected per arm • Implementation feasibility	Developed by an information technologist in the United States working with the study team in Sierra Leone.
"Implementation of digital technology solutions for a lung health trial in rural Malawi" [6] (cross-sectional study)	• Number of households located on the map	No app development; Google Earth was used to identify households.
"Using a mobile technology to Improve paediatric presumptive Tuberculosis identification in western Kenya" [41] (cross-sectional study)	• Proportion of children recorded in the presumptive tuberculosis register • Proportion of children testing positive for tuberculosis	The app was developed in Bangladesh and customized for use in this study.
"Evaluation of a Mobile Health Approach to Tuberculosis Contact Tracing in Botswana" [40] (pretest-posttest study)	• Number of contacts screened using paper vs the app • Time to complete screening using paper vs the app • Missing and illogical data between paper and the app • User satisfaction with paper vs the app	Researchers partnered with an IT firm to produce the contact tracing app.
"Feasibility of using a mobile app to monitor and report COVID-19-related symptoms and people's movements in Uganda" [44] (pretest-posttest study)	• Household members using the technology • Household members reporting symptoms • Household members referred for testing • Household members diagnosed with COVID-19	Developed by a technology designer in consultation with officials from the Ministry of Health. No further details of the development process are given.
"The 117-call alert system in Sierra Leone: from rapid Ebola notification to routine death reporting" [36] (programmatic pretest-posttest)	• Number of live Ebola alerts • Number of death alerts • Confirmed new cases from alerts	No app development, and a call center was repurposed to report suspected Ebola cases.
"A Smartphone App (AfyaData) for Innovative One Health Disease Surveillance from Community to National Levels in Africa: Intervention in Disease Surveillance" [47] (programmatic pretest-posttest)	• Number of disease cases reported among humans and animals	Designed through a collaboration between public health and ICT[b] specialists and government personnel. They developed a theory of change to guide development and implementation.
"Ebola virus disease contact tracing activities, lessons learned and best practices during the Duport Road outbreak in Monrovia, Liberia, November 2015" [35] (programmatic pretest-posttest)	• The total number of contacts traced • Total households found • Total high- and low-risk contacts found • Total symptomatic contacts • Contact deaths recorded	No app development. The program used subpoenas to cell phone companies to provide location details of suspected Ebola contacts.
"Introduction of Mobile Health Tools to Support Ebola Surveillance and Contact Tracing in Guinea" [32] (programmatic pretest-posttest)	• Total number of Ebola contacts monitored on CommCare • IT and implementation challenges	An IT team developed it in the United States with local public health specialists in Guinea, local UN[c] partners, and the government.

Study title (design)	Output or outcomes	Development and evaluation
"USAID/South Africa Tuberculosis South Africa Project (TBSAP) Midterm Evaluation Report" [43] (programmatic pretest-posttest)	• Not reported	Not reported
"Development and implementation of the Ebola Exposure Window Calculator: A tool for Ebola virus disease outbreak field investigations" [39] (programmatic pretest-posttest)	• Exposure window for contact tracing to take place	Not reported
"Innovative Technological Approach to Ebola Virus Disease Outbreak Response in Nigeria Using the Open Data Kit and Form Hub Technology" [34] (programmatic pretest-posttest)	• Number of users downloading the app • Number of active users of the app	CDC[d], WHO[e], and DRC[f] field teams and Johns Hopkins teams developed the technology, and the government endorsed its use.
"Evaluation of an Android-based mHealth system for population surveillance in developing countries" [8] (programmatic pretest-posttest)	• Number of households visited for home-based counselling and testing and data collected using the app • Time to complete screening using the app vs PDA • App usability • Cost of development and implementation	Members of the AMPATH[g] project team developed the app in collaboration with staff from the Ministry of Health in Kenya. They had clear objectives for their scope and what they wanted the app to achieve.
"Evaluating the use of cell phone messaging for community Ebola syndromic surveillance in high-risk settings in Southern Sierra Leone" [33] (programmatic pretest-posttest)	• Suspected Ebola cases • Confirmed Ebola cases • Reported deaths	No app development
"Implementing Surveillance and Outbreak Response Management and Analysis System (SORMAS) for Public Health in West Africa- Lessons Learnt and Future Direction" [37] (programmatic pretest-posttest)	• Implementation outcomes from piloting the technology	Design thinking workshops were held in Nigeria and Germany to assess software requirements. Nigeria Centre for Disease Control and Prevention, Port Health, and other stakeholders were also involved in the design thinking methodology. Contact tracing requirements for other diseases, not just Ebola, were also considered in the design process.
"Using mHealth to self-screen and promote Tuberculosis awareness in Tanzania" [42] (programmatic pretest-posttest)	• Number of people screened using the technology • Number of presumptive cases screened using the app	Designed by the government and other implementors using the Ministry of Health mHealth platforms in a participatory process.
"TB HealthCheck puts tuberculosis self-screening in everyone's hands ahead of World TB Day" [46] (programmatic pretest-posttest)	• Not clearly stated	Not clearly stated
"Mobile health approaches to disease surveillance in Africa; Wellvis COVID triage tool" [45] (programmatic pretest-posttest)	• Risk of COVID-19 after self-screenings	The development process is not described in detail.

[a]mHealth: mobile health.

[b]ICT: information and communications technology.

[c]UN: United Nations.

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

[e]WHO: World Health Organization.

[f]DRC: Democratic Republic of the Congo.

[g]AMPATH: Academic Model Providing Access to Healthcare.

Of all the included studies, only the trial from Uganda randomly assigned participants to different groups, comparing outcomes objectively. The trial had primary outcomes of contacts completing tuberculosis evaluation within 14 days and secondary outcomes of treatment initiation and linking patients to care [27]. Contacts were assigned to the control, and SMS text messaging–facilitated interventions were at a household level. The trial showed that the mHealth technology had no effect on the primary outcome.

In addition, although they did not randomize participants or calculate sample size as in the randomized trial, the cross-sectional surveys and pretest-posttest studies included in this systematic review attempted to evaluate the performance of the mHealth technologies by reporting outcomes such as proportions of contacts diagnosed, and some compared performance between groups. For example, a study in Sierra Leone compared the number of Ebola contacts identified and traced between the Ebola contact tracing app and a paper-based data collection system using a conveniently selected sample [38]. The researchers also evaluated completion, the proportion of cases detected, and implementation feasibility between the 2 arms. Similarly, the study by Ha et al [40] also compared outcomes in proportions before and after implementation and in contrast to a paper-based system. This pretest-posttest study compared the number of contacts screened, time to complete an evaluation, and data quality between using a contact tracing app and paper-based systems [40]. Other cross-sectional studies only reported outcomes in one conveniently selected group [6,41], and one pretest-posttest study only reported outputs [44].

Programmatic setting studies predominantly reported immediate outputs or implementation experiences of the mHealth technologies. SORMAS, AMPATH, and Afyadata, which were the largest and most comprehensively developed technologies, only reported absolute numbers of the people screened and implementation experiences rather than actual outcomes to determine the performance of the mHealth technology. In addition, Tambua tuberculosis in Tanzania reported immediate outputs on tuberculosis screening but did not have an evaluation done on implementation outcomes. In South Africa, ConnecTB and tuberculosis HealthCheck, both used to facilitate tuberculosis screening, did not provide sufficient details on the outputs or evaluation of the technologies. Nevertheless, these technologies in programmatic settings reached large populations for contact tracing.

Discussion

Principal Findings

We reviewed the continuum of mHealth technologies used for contact tracing and case finding. We synthesized this information to improve understanding of mHealth's value in contact tracing and inform how future projects can develop or implement it efficiently to improve contact tracing or implementation outcomes. Only 19 studies met the criteria and were included in the review. The technologies were developed by either customizing existing platforms or creating original software. Some technologies used to trace contacts did not require software development but used existing cellular network infrastructure. Technologies developed using design thinking frameworks with participatory activities had a higher likelihood of implementation fidelity. However, the effectiveness of these technologies was not sufficiently evaluated, and the outcomes of most technologies were only reported in small research studies.

Our search yielded a greater number of articles on studies using mHealth technologies for contact tracing than those obtained in previous reviews, likely reflecting the growing adoption of such technologies during the search period. In addition, our iterative search strategy, which featured relaxed criteria and encompassed a variety of infectious diseases, may have played a significant role in identifying this increased volume of articles. The review has also presented the entire pathway of using these technologies from the development stage until deployment and evaluation of outcomes. Our findings also confirm that mHealth outperforms traditional paper-based contact tracing regarding the timeliness of screening, data accuracy, and streamlining of the screening process. However, little evidence exists of mHealth's impact or incremental value on contact tracing outcomes compared to paper because of underreporting. Despite this challenge, our study underscores the potential of using validated design frameworks during the developmental phase, which is likely to enhance the overall effectiveness of such solutions. Future public health projects that intend to develop and implement mHealth technologies can consider these insights as a guide into some of the prerequisites for implementing useful technologies that can meaningfully improve outcomes.

Our systematic review suggests that a more systematic approach to development using design thinking frameworks with participatory development, including buy-in from policy makers, could help technologies in achieving the intended case finding outputs and outcomes when compared to an approach that does not use these frameworks. For example, SORMAS was developed in Nigeria following design thinking and participatory processes involving public health experts, information technologists, and government officials. SORMAS was responding to an earlier mHealth app, Sense Follow-up, which had failed at the implementation stage due to not meeting user requirements because of poor development. Similarly, Afyadata in Tanzania and AMPATH in Kenya both involved extensive consultative processes, increasing their outputs. Contrastingly, even a collaboratively built technology may fail at implementation without government backing. For example, in Sierra Leone, Sacks et al [32] reported poor outputs from the Ebola contact tracing program, partly due to the lack of government officials' commitment to supporting the program. In South Africa, ConnecTB did not document any design thinking frameworks or participatory development and was discontinued due to technical challenges despite having the policy makers' support. None of the research studies documented the use of a design thinking framework, which may have affected their results, most of which were suboptimal, with the technology failing to show an impact on case finding. A post hoc analysis of the trial by Davis et al [27] also recommended using design thinking frameworks even in research studies to ensure that they are impactful [59].

In addition to using established frameworks in the development stage, financial resources were a significant factor in determining the extent of development and implementation. For example, among the technologies used in programmatic settings, SORMAS received an initial funding of €850,000 (US $929,509) from the European Union [85], Afyadata received US $450,000 from the Skoll Global Threats Fund [55], and the CommCare app for HIV screening was developed under a US $74.9–million US Agency for International Development–AMPATH program [86]. In addition, Tambua

tuberculosis was government backed through Tanzania's National Tuberculosis and Leprosy Programme [42]. Government involvement also assisted in obtaining buy-in from the ultimate end users of the technologies. However, compared to studies implementing technologies in programmatic settings, most research studies included in this review did not have the financial resources to execute the extensive development processes required for a technology to succeed. However, the dilemma is that well-funded programs tend to develop and implement large and sustainable technologies but often overlook comprehensive evaluations and, hence, fail to inform any future work meaningfully. Research studies, on the other hand, while rigorously evaluating their technologies, undergo a suboptimal development process due to limited funding. Consequently, despite restricted resources, research studies tend to use more robust evaluation methods in their assessments. For example, in Botswana, Ha et al [40] evaluated a pragmatic mHealth technology for tuberculosis contact tracing, but it was only implemented in a few settings, possibly due to limited funding, and the researchers could not conduct a large study to fully evaluate the program. Therefore, nesting evaluations in large programs as implementation research processes may be a way to improve evidence of mHealth technologies' utility for contact tracing.

The primary focus of contact tracing programs is on improving case finding. However, there is also an expectation in some studies that improved case finding should also trickle down to the betterment of other outcomes down the tuberculosis care and treatment cascade, such as treatment initiation and completion [88,89]. However, using current evidence, it can be argued for mHealth technologies that, unless initially planned and incorporated in the design process, the desired contact tracing outcomes further down the cascade are inappropriate to determine the success of a contact tracing program and the effectiveness of mHealth technologies. Only 1 study in this review, a randomized trial, reported treatment initiation further down the contact tracing cascade [24,27], and the rest focused on outputs immediately after screening, such as numbers screened and testing positive. When using mHealth, immediate implementation outcomes such as technology development, implementation fidelity, feasibility, and acceptability could more appropriately measure effectiveness. This is because mHealth technologies can only measure what they are designed for. For instance, all technologies used in programmatic setting studies were designed to find and screen contacts but did not have modules developed for linkage to care, which would have allowed outcomes further down the cascade to be measured. Although this narrow focus on immediate outputs may limit the tenure of the technologies because they cannot demonstrate any value for the desired contact tracing outcomes, these outcomes are not determined only by finding the people with the disease. Instead, additional steps are required to optimize care linkage and treatment adherence. Small research studies with a limited scope cannot include all the steps from finding contacts, linking them to care, and showing a reduction in disease incidence, but without these steps, they may be too expensive and less likely to be adopted. Therefore, as discussed in the previous section, a participatory developmental process must be used to include such capabilities for mHealth and, where possible, develop the technology within a large well-funded program.

A cost analysis of the mHealth technology for contact tracing in Uganda found that this technology was underused because of its limited scope despite substantial investment in the development stages and initial implementation [90]. The authors suggested expanding the scope by increasing the volume of contacts served or expanding the use of the technology to the later stages of the contact tracing cascade. From this review, most technologies used to support Ebola contact tracing became defunct after the outbreak subsided, thus ultimately costing health programs because they were not used long enough to absorb or justify the costs of their development. Only SORMAS, which had an expanded scope beyond a single disease and activities beyond tracing, survived and continues to receive funding. Afyadata in Tanzania also has a large scope of contact tracing and surveillance activities in humans and animals and continues to be valuable to both health systems. Therefore, health programs should consider using technologies for contact tracing and case finding more broadly to capitalize on synergies and bring down costs to make them more sustainable. This will expand their scope and include multiple diseases or higher volumes of patients or expand the use of mHealth to all stages of the cascade, increasing the likelihood of the interventions' sustainability. Investment cases for future mHealth technologies may be supported if there is evidence of their cost-effectiveness, and this may be necessitated by exploiting economies of scale and scope. Health economics studies and models have also shown that unit costs of interventions may decrease and be optimized for cost-effectiveness by implementing at the appropriate scale, within the right scope [91,92], and without compromising quality by exceeding these bounds [93].

Limitations and Strengths of the Study

The primary limitation of our study stems from the context in which most of the reviewed mHealth technologies were deployed. Predominantly used within programmatic settings rather than through rigorous formal research, these technologies often lacked comprehensive outcome evaluations. Consequently, without these evaluations, our review faces challenges in definitively assessing the impact of mHealth technologies on enhancing contact tracing efforts. Despite this, the deployment of these technologies in well-funded programmatic settings offers valuable insights into the foundational requirements for successful contact tracing technologies, demonstrating how strategic funding can facilitate robust development and implementation.

In addition, our review faces limitations due to the unavailability of some key information, such as tenure of technology use beyond the initial studies, especially for technologies evaluated in research settings. The scarcity of detailed data in both primary literature and supplementary documents makes it difficult to assess whether the technologies were continued or discontinued after the research. This gap underscores the vital need for comprehensive documentation and rigorous reporting standards in the mHealth domain.

Recommendations and Conclusions

The essential ingredients for developing functional and impactful mHealth technologies, whether for contact tracing or other health care interventions, include using a participatory design thinking framework, securing adequate funding, and establishing a clear plan for evaluating the implementation outcomes of the technology.

Many projects, especially research projects, often face challenges of limited funds to fully develop and implement their technologies. In such cases, an alternative approach can be integrating the development of mHealth technologies within larger technologies used in programmatic setting projects. This integration can provide the necessary resources and infrastructure to sufficiently develop and efficiently evaluate the technology. However, such ideal scenarios may not always be feasible, and the use of participatory design thinking frameworks becomes a prerequisite for developing an effective mHealth technology. This approach emphasizes collaboration with end users, stakeholders, and experts to ensure that the technology aligns with user needs, is inclusive, and considers the context-specific requirements.

Authors' Contributions

The conception of this work was presented by SC and KV as part of DLM's PhD studies. Manuscript reviews were diligently conducted by HC and JN. JLD provided invaluable senior scientific guidance based on his extensive experience in contact tracing. TN collaborated with DLM to structure the systematic review chapters and review the written content drawing insights from their involvement in the Community and Universal Testing for tuberculosis project—a contact tracing cluster randomized trial—in which both are researchers dedicated to the development of a mobile health technology to support tuberculosis contact tracing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search terms.
[PDF File (Adobe PDF File), 116 KB - mhealth_v12i1e53211_app1.pdf]

Multimedia Appendix 2
Systematic review data extraction tool.
[PDF File (Adobe PDF File), 231 KB - mhealth_v12i1e53211_app2.pdf]

Multimedia Appendix 3
Quality assessment of the included studies.
[DOCX File , 70 KB - mhealth_v12i1e53211_app3.docx]

Multimedia Appendix 4
Detailed explanation of each article in the systematic review.
[DOCX File , 127 KB - mhealth_v12i1e53211_app4.docx]

Multimedia Appendix 5
PRISMA checklist.
[PDF File (Adobe PDF File), 124 KB - mhealth_v12i1e53211_app5.pdf]

References

1. Asadzadeh A, Kalankesh LR. A scope of mobile health solutions in COVID-19 pandemics. Inform Med Unlocked 2021;23:100558 [FREE Full text] [doi: 10.1016/j.imu.2021.100558] [Medline: 33842688]
2. Tilahun B, Gashu KD, Mekonnen ZA, Endehabtu BF, Angaw DA. Mapping the role of digital health technologies in prevention and control of COVID-19 pandemic: review of the literature. Yearb Med Inform 2021 Aug;30(1):26-37 [FREE Full text] [doi: 10.1055/s-0041-1726505] [Medline: 34479378]
3. Brinkel J, Krämer A, Krumkamp R, May J, Fobil J. Mobile phone-based mHealth approaches for public health surveillance in sub-Saharan Africa: a systematic review. Int J Environ Res Public Health 2014 Nov 12;11(11):11559-11582 [FREE Full text] [doi: 10.3390/ijerph111111559] [Medline: 25396767]
4. Falzon D, Raviglione M, Bel EH, Gratziou C, Bettcher D, Migliori GB. The role of eHealth and mHealth in tuberculosis and tobacco control: a WHO/ERS consultation. Eur Respir J 2015 Aug;46(2):307-311 [FREE Full text] [doi: 10.1183/09031936.00043315] [Medline: 26232477]

5. Falzon D, Timimi H, Kurosinski P, Migliori GB, Van Gemert W, Denkinger C, et al. Digital health for the End TB Strategy: developing priority products and making them work. Eur Respir J 2016 Jul;48(1):29-45 [FREE Full text] [doi: 10.1183/13993003.00424-2016] [Medline: 27230443]
6. Chisunkha B, Banda H, Thomson R, Squire SB, Mortimer K. Implementation of digital technology solutions for a lung health trial in rural Malawi. Eur Respir J 2016 Jun;47(6):1876-1879 [FREE Full text] [doi: 10.1183/13993003.00045-2016] [Medline: 27076597]
7. Babirye D, Shete PB, Farr K, Nalugwa T, Ojok C, Nantale M, et al. Feasibility of a short message service (SMS) intervention to deliver tuberculosis testing results in peri-urban and rural Uganda. J Clin Tuberc Other Mycobact Dis 2019 Aug;16:100110 [FREE Full text] [doi: 10.1016/j.jctube.2019.100110] [Medline: 31720434]
8. Rajput ZA, Mbugua S, Amadi D, Chepngeno V, Saleem JJ, Anokwa Y, et al. Evaluation of an android-based mHealth system for population surveillance in developing countries. J Am Med Inform Assoc 2012;19(4):655-659 [FREE Full text] [doi: 10.1136/amiajnl-2011-000476] [Medline: 22366295]
9. Cherutich P, Golden M, Betz B, Wamuti B, Ng'ang'a A, Maingi P, et al. Surveillance of HIV assisted partner services using routine health information systems in Kenya. BMC Med Inform Decis Mak 2016 Jul 20;16:97 [FREE Full text] [doi: 10.1186/s12911-016-0337-9] [Medline: 27439397]
10. Coleman J, Bohlin KC, Thorson A, Black V, Mechael P, Mangxaba J, et al. Effectiveness of an SMS-based maternal mHealth intervention to improve clinical outcomes of HIV-positive pregnant women. AIDS Care 2017 Jul;29(7):890-897. [doi: 10.1080/09540121.2017.1280126] [Medline: 28107805]
11. Abdulrahman S, Ganasegeran K. Chapter 11 - m-Health in public health practice: a constellation of current evidence. In: Jude HD, Balas VE, editors. Telemedicine Technologies: Big Data, Deep Learning, Robotics, Mobile and Remote Applications for Global Healthcare. Cambridge, MA: Academic Press; 2019:171-182.
12. Fox GJ, Barry SE, Britton WJ, Marks GB. Contact investigation for tuberculosis: a systematic review and meta-analysis. Eur Respir J 2013 Jan;41(1):140-156 [FREE Full text] [doi: 10.1183/09031936.00070812] [Medline: 22936710]
13. Baxter S, Goyder E, Chambers D, Johnson M, Preston L, Booth A. Interventions to improve contact tracing for tuberculosis in specific groups and in wider populations: an evidence synthesis. Health Serv Deliv Res 2017 Jan;5(1):1-102. [doi: 10.3310/hsdr05010] [Medline: 28121090]
14. Velen K, Shingde RV, Ho J, Fox GJ. The effectiveness of contact investigation among contacts of tuberculosis patients: a systematic review and meta-analysis. Eur Respir J 2021 Dec;58(6):2100266 [FREE Full text] [doi: 10.1183/13993003.00266-2021] [Medline: 34016621]
15. Contact tracing during an outbreak of Ebola virus disease. World Health Organization Regional Office for Africa. 2014. URL: https://www.icao.int/safety/CAPSCA/PublishingImages/Pages/Ebola/contact-tracing-during-outbreak-of-ebola.pdf [accessed 2024-07-26]
16. Recommendations for investigating contacts of persons with infectious tuberculosis in low- and middle-income countries. World Health Organization. 2012 Apr 11. URL: https://www.who.int/publications/i/item/9789241504492 [accessed 2024-07-26]
17. Breuninger M, van Ginneken B, Philipsen RH, Mhimbira F, Hella JJ, Lwilla F, et al. Diagnostic accuracy of computer-aided detection of pulmonary tuberculosis in chest radiographs: a validation study from sub-Saharan Africa. PLoS One 2014 Sep 5;9(9):e106381 [FREE Full text] [doi: 10.1371/journal.pone.0106381] [Medline: 25192172]
18. Murphy K, Habib SS, Zaidi SM, Khowaja S, Khan A, Melendez J, et al. Computer aided detection of tuberculosis on chest radiographs: an evaluation of the CAD4TB v6 system. Sci Rep 2020 Mar 26;10(1):5492 [FREE Full text] [doi: 10.1038/s41598-020-62148-y] [Medline: 32218458]
19. Cox H, Workman L, Bateman L, Franckling-Smith Z, Prins M, Luiz J, et al. Oral swab specimens tested with Xpert MTB/RIF ultra assay for diagnosis of pulmonary tuberculosis in children: a diagnostic accuracy study. Clin Infect Dis 2022 Dec 19;75(12):2145-2152 [FREE Full text] [doi: 10.1093/cid/ciac332] [Medline: 35579497]
20. Kang YA, Koo B, Kim OH, Park JH, Kim HC, Lee HJ, et al. Gene-based diagnosis of tuberculosis from oral swabs with a new generation pathogen enrichment technique. Microbiol Spectr 2022 Jun 29;10(3):e0020722 [FREE Full text] [doi: 10.1128/spectrum.00207-22] [Medline: 35587206]
21. National guidelines on the treatment of tuberculosis infection. National Department of Health, Republic of South Africa. 2023. URL: https://knowledgehub.health.gov.za/elibrary/national-guidelines-treatment-tuberculosis-infection [accessed 2024-04-19]
22. Thind D, Charalambous S, Tongman A, Churchyard G, Grant AD. An evaluation of 'Ribolola': a household tuberculosis contact tracing programme in North West Province, South Africa. Int J Tuberc Lung Dis 2012 Dec;16(12):1643-1648. [doi: 10.5588/ijtld.12.0074] [Medline: 23131263]
23. Deery CB, Hanrahan CF, Selibas K, Bassett J, Sanne I, Van Rie A. A home tracing program for contacts of people with tuberculosis or HIV and patients lost to care. Int J Tuberc Lung Dis 2014 May;18(5):534-540 [FREE Full text] [doi: 10.5588/ijtld.13.0587] [Medline: 24903789]
24. Armstrong-Hough M, Turimumahoro P, Meyer AJ, Ochom E, Babirye D, Ayakaka I, et al. Drop-out from the tuberculosis contact investigation cascade in a routine public health setting in urban Uganda: a prospective, multi-center study. PLoS One 2017 Nov 6;12(11):e0187145 [FREE Full text] [doi: 10.1371/journal.pone.0187145] [Medline: 29108007]

25. Chetty-Makkan CM, deSanto D, Lessells R, Charalambous S, Velen K, Makgopa S, et al. Exploring the promise and reality of ward-based primary healthcare outreach teams conducting TB household contact tracing in three districts of South Africa. PLoS One 2021 Aug 13;16(8):e0256033 [FREE Full text] [doi: 10.1371/journal.pone.0256033] [Medline: 34388206]
26. Meyer AJ, Babirye D, Armstrong-Hough M, Mark D, Ayakaka I, Katamba A, et al. Text messages sent to household tuberculosis contacts in Kampala, Uganda: process evaluation. JMIR Mhealth Uhealth 2018 Nov 20;6(11):e10239 [FREE Full text] [doi: 10.2196/10239] [Medline: 30459147]
27. Davis JL, Turimumahoro P, Meyer AJ, Ayakaka I, Ochom E, Ggita J, et al. Home-based tuberculosis contact investigation in Uganda: a household randomised trial. ERJ Open Res 2019 Jul;5(3):00112-2019 [FREE Full text] [doi: 10.1183/23120541.00112-2019] [Medline: 31367636]
28. Anglemyer A, Moore TH, Parker L, Chambers T, Grady A, Chiu K, et al. Digital contact tracing technologies in epidemics: a rapid review. Cochrane Database Syst Rev 2020 Aug 18;8(8):CD013699 [FREE Full text] [doi: 10.1002/14651858.CD013699] [Medline: 33502000]
29. Skiba DJ. Intellectual property issues in the digital health care world. Nurs Adm Q 1997;21(3):11-20. [Medline: 9215002]
30. Fiordelli M, Diviani N, Schulz PJ. Mapping mHealth research: a decade of evolution. J Med Internet Res 2013 May 21;15(5):e95 [FREE Full text] [doi: 10.2196/jmir.2430] [Medline: 23697600]
31. Santos WM, Secoli SR, Püschel VA. The Joanna Briggs Institute approach for systematic reviews. Rev Lat Am Enfermagem 2018 Nov 14;26:e3074 [FREE Full text] [doi: 10.1590/1518-8345.2885.3074] [Medline: 30462787]
32. Sacks JA, Zehe E, Redick C, Bah A, Cowger K, Camara M, et al. Introduction of mobile health tools to support Ebola surveillance and contact tracing in Guinea. Glob Health Sci Pract 2015 Nov 12;3(4):646-659 [FREE Full text] [doi: 10.9745/GHSP-D-15-00207] [Medline: 26681710]
33. Jia K, Mohamed K. Evaluating the use of cell phone messaging for community Ebola syndromic surveillance in high risked settings in Southern Sierra Leone. Afr Health Sci 2015 Sep;15(3):797-802 [FREE Full text] [doi: 10.4314/ahs.v15i3.13] [Medline: 26957967]
34. Tom-Aba D, Olaleye A, Olayinka AT, Nguku P, Waziri N, Adewuyi P, et al. Innovative technological approach to Ebola virus disease outbreak response in Nigeria using the open data kit and form hub technology. PLoS One 2015 Jun 26;10(6):e0131000 [FREE Full text] [doi: 10.1371/journal.pone.0131000] [Medline: 26115402]
35. Wolfe CM, Hamblion EL, Schulte J, Williams P, Koryon A, Enders J, et al. Ebola virus disease contact tracing activities, lessons learned and best practices during the Duport Road outbreak in Monrovia, Liberia, November 2015. PLoS Negl Trop Dis 2017 Jun;11(6):e0005597 [FREE Full text] [doi: 10.1371/journal.pntd.0005597] [Medline: 28575034]
36. Alpren C, Jalloh MF, Kaiser R, Diop M, Kargbo S, Castle E, et al. The 117 call alert system in Sierra Leone: from rapid Ebola notification to routine death reporting. BMJ Glob Health 2017;2(3):e000392 [FREE Full text] [doi: 10.1136/bmjgh-2017-000392] [Medline: 28948044]
37. Adeoye OO, Tom-Aba D, Ameh CA, Ojo OE, Ilori EA, Gidado SO, et al. Implementing surveillance and outbreak response management and analysis system (SORMAS) for public health in west Africa- lessons learnt and future direction. Intl J Tropical Dis Health 2017 Jan 10;22(2):1-17. [doi: 10.9734/IJTDH/2017/31584]
38. Danquah LO, Hasham N, MacFarlane M, Conteh FE, Momoh F, Tedesco AA, et al. Use of a mobile application for Ebola contact tracing and monitoring in northern Sierra Leone: a proof-of-concept study. BMC Infect Dis 2019 Sep 18;19(1):810 [FREE Full text] [doi: 10.1186/s12879-019-4354-z] [Medline: 31533659]
39. Whitesell A, Bustamante ND, Stewart M, Freeman J, Dismer AM, Alarcon W, et al. Development and implementation of the Ebola exposure window calculator: a tool for Ebola virus disease outbreak field investigations. PLoS One 2021 Aug 5;16(8):e0255631 [FREE Full text] [doi: 10.1371/journal.pone.0255631] [Medline: 34352008]
40. Ha YP, Tesfalul MA, Littman-Quinn R, Antwi C, Green RS, Mapila TO, et al. Evaluation of a mobile health approach to tuberculosis contact tracing in Botswana. J Health Commun 2016 Oct 26;21(10):1115-1121 [FREE Full text] [doi: 10.1080/10810730.2016.1222035] [Medline: 27668973]
41. Szkwarko D, Amisi JA, Peterson D, Burudi S, Angala P, Carter EJ. Using a mobile application to improve pediatric presumptive TB identification in western Kenya. Int J Tuberc Lung Dis 2021 Jun 01;25(6):468-474. [doi: 10.5588/ijtld.20.0890] [Medline: 34049609]
42. Diaz N, Moturi E. Using mHealth to self-screen and promote TB awareness in Tanzania. Challenge TB. 2019. URL: https://www.challengetb.org/publications/tools/briefs/CTB_Tanzania_mHealth.pdf [accessed 2024-07-26]
43. USAID/South Africa: Tuberculosis South Africa Project (TBSAP) midterm evaluation report. United States Agency for International Development. 2020 Oct 15. URL: https://pdf.usaid.gov/pdf_docs/PA00XB1Q.pdf [accessed 2024-07-26]
44. Mugenyi L, Nsubuga RN, Wanyana I, Muttamba W, Tumwesigye NM, Nsubuga SH. Feasibility of using a mobile app to monitor and report COVID-19 related symptoms and people's movements in Uganda. PLoS One 2021;16(11):e0260269 [FREE Full text] [doi: 10.1371/journal.pone.0260269] [Medline: 34797878]
45. Owoyemi A, Ikpe R, Toye M, Rewane A, Abdullateef M, Obaseki E, et al. Mobile health approaches to disease surveillance in Africa; Wellvis COVID triage tool. Digit Health 2021 Feb 20;7:2055207621996876 [FREE Full text] [doi: 10.1177/2055207621996876] [Medline: 33680485]

46. TB HealthCheck puts self-screening in everyone's hands ahead of World TB Day. Rising Sun. 2021 Mar 24. URL: https://risingsunnewspapers.co.za/lnn/1151916/tb-healthcheck-puts-self-screening-in-everyones-hands-ahead-of-world-tb-day/ [accessed 2024-07-26]
47. Karimuribo ED, Mutagahywa E, Sindato C, Mboera L, Mwabukusi M, Kariuki Njenga M, et al. A smartphone app (AfyaData) for innovative one health disease surveillance from community to national levels in Africa: intervention in disease surveillance. JMIR Public Health Surveill 2017 Dec 18;3(4):e94 [FREE Full text] [doi: 10.2196/publichealth.7373] [Medline: 29254916]
48. Kangbai JB. Social network analysis and modeling of cellphone-based syndromic surveillance data for Ebola in Sierra Leone. Asian Pac J Trop Med 2016 Sep;9(9):851-855 [FREE Full text] [doi: 10.1016/j.apjtm.2016.07.005] [Medline: 27633297]
49. Sterne JA, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019 Aug 28;366:l4898 [FREE Full text] [doi: 10.1136/bmj.l4898] [Medline: 31462531]
50. Study quality assessment tools. National Institutes of Health National Heart, Lung, and Blood Institute. URL: https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools [accessed 2022-05-15]
51. WHO guideline: recommendations on digital interventions for health system strengthening. World Health Organization. 2019. URL: https://iris.who.int/bitstream/handle/10665/311941/9789241550505-eng.pdf?ua=1 [accessed 2024-07-26]
52. mHealth: use of mobile wireless technologies for public health. World Health Organization. 2016 May 27. URL: https://apps.who.int/gb/ebwha/pdf_files/EB139/B139_8-en.pdf [accessed 2024-07-26]
53. Khandpur RS. Telemedicine: Technology and Applications (mHealth, TeleHealth and eHealth). Delhi, India: PHI Learning; 2017.
54. Crico C, Renzi C, Graf N, Buyx A, Kondylakis H, Koumakis L, et al. mHealth and telemedicine apps: in search of a common regulation. Ecancermedicalscience 2018 Jul 11;12:853 [FREE Full text] [doi: 10.3332/ecancer.2018.853] [Medline: 30079115]
55. The DODRES project: revolutionizing human and animal disease surveillance in Tanzania. Ending Pandemic. 2017. URL: https://endingpandemics.org/wp-content/uploads/2017/04/EP_DODRES_CaseStudy.pdf [accessed 2022-06-20]
56. Divi N, Smolinski M. EpiHacks, a process for technologists and health experts to cocreate optimal solutions for disease prevention and control: user-centered design approach. J Med Internet Res 2021 Dec 15;23(12):e34286 [FREE Full text] [doi: 10.2196/34286] [Medline: 34807832]
57. Fähnrich C, Denecke K, Adeoye OO, Benzler J, Claus H, Kirchner G, et al. Surveillance and Outbreak Response Management System (SORMAS) to support the control of the Ebola virus disease outbreak in West Africa. Euro Surveill 2015 Mar 26;20(12):21071 [FREE Full text] [doi: 10.2807/1560-7917.es2015.20.12.21071] [Medline: 25846493]
58. Grainger C. A software for disease surveillance and outbreak response: insights from implementing SORMAS in Nigeria and Ghana. Federal Ministry for Economic Cooperation and Development, Germany. 2020 Mar. URL: https://health.bmz.de/studies/a-software-for-disease-surveillance-and-outbreak-response/ [accessed 2024-07-26]
59. Meyer AJ, Armstrong-Hough M, Babirye D, Mark D, Turimumahoro P, Ayakaka I, et al. Implementing mHealth interventions in a resource-constrained setting: case study from Uganda. JMIR Mhealth Uhealth 2020 Jul 13;8(7):e19552 [FREE Full text] [doi: 10.2196/19552] [Medline: 32673262]
60. Gaythorpe K, Morris A, Imai N, Stewart M, Freeman J, Choi M. Chainchecker: an application to visualise and explore transmission chains for Ebola virus disease. PLoS One 2021 Feb 19;16(2):e0247002 [FREE Full text] [doi: 10.1371/journal.pone.0247002] [Medline: 33606709]
61. Sormas in Nigeria: adapting a fully integrated surveillance system to track COVID-19. Exemplars in Global Health. URL: https://www.exemplars.health/emerging-topics/epidemic-preparedness-and-response/digital-health-tools/sormas-nigeria [accessed 2024-07-26]
62. Wools-Kaloustian KK, Sidle JE, Selke HM, Vedanthan R, Kemboi EK, Boit LJ, et al. A model for extending antiretroviral care beyond the rural health centre. J Int AIDS Soc 2009 Sep 29;12:22 [FREE Full text] [doi: 10.1186/1758-2652-12-22] [Medline: 19788755]
63. Anokwa Y, Ribeka N, Parikh T, Borriello G, Were MC. Design of a phone-based clinical decision support system for resource-limited settings. In: Proceedings of the Fifth International Conference on Information and Communication Technologies and Development. 2012 Presented at: ICTD '12; March 12-15, 2012; Atlanta, GA. [doi: 10.1145/2160673.2160676]
64. Were M, Kariuki J, Chepng'eno V, Wandabwa M, Ndege S, Braitstein P, et al. Leapfrogging paper-based records using handheld technology: experience from Western Kenya. Stud Health Technol Inform 2010;160(Pt 1):525-529. [Medline: 20841742]
65. Gates B. Meet the heroes in Africa fighting back against this pandemic—and working to prevent the next one. The Gates Notes. 2021 Dec 7. URL: https://www.gatesnotes.com/These-African-heroes-are-fighting-back-against-the-pandemic [accessed 2022-06-20]
66. Littman-Quinn R, Chandra A, Schwartz A, Fadlelmola F, Ghose S, Luberti A, et al. mHealth applications for telemedicine and public health intervention in Botswana. In: Proceeding of the IST-Africa 2011 Conference Proceedings. 2011 Presented at: IST-Africa 2011; May 11-13, 2011; Gaborone, Botswana.

67. Mosweunyane G, Nkgau T, Makhura O, Seipone T. USSD system for TB contact tracing: server, database and security issues. In: Proceedings of the 9th European Conference on IS Management and Evaluation. 2015 Presented at: ECIME 2015; September 21-22, 2015; Bristol, UK. [doi: 10.2316/p.2014.815-026]
68. Ncube B, Mars M, Scott RE. The need for a telemedicine strategy for Botswana? A scoping review and situational assessment. BMC Health Serv Res 2020 Aug 26;20(1):794 [FREE Full text] [doi: 10.1186/s12913-020-05653-0] [Medline: 32843017]
69. Banda HT, Mortimer K, Bello GA, Mbera GB, Namakhoma I, Thomson R, et al. Informal health provider and practical approach to lung health interventions to improve the detection of chronic airways disease and tuberculosis at primary care level in Malawi: study protocol for a randomised controlled trial. Trials 2015 Dec 17;16(1):576 [FREE Full text] [doi: 10.1186/s13063-015-1068-4] [Medline: 26679768]
70. Brunette W. Open data kit 2: building mobile application frameworks for disconnected data management. University of Washington. URL: https://digital.lib.washington.edu/researchworks/handle/1773/46432 [accessed 2024-07-26]
71. White EB, Meyer AJ, Ggita JM, Babirye D, Mark D, Ayakaka I, et al. Feasibility, acceptability, and adoption of digital fingerprinting during contact investigation for tuberculosis in Kampala, Uganda: a parallel-convergent mixed-methods analysis. J Med Internet Res 2018 Nov 15;20(11):e11541 [FREE Full text] [doi: 10.2196/11541] [Medline: 30442637]
72. HealthConnect in South Africa: a chatbot tool for pandemic response. Exemplars in Global Health. URL: https://www.exemplars.health/emerging-topics/epidemic-preparedness-and-response/digital-health-tools/healthconnect-in-south-africa#:~:text=The%20program%20scaled%20up%20quickly,countries%20including%20Uganda%20and%20Nigeria [accessed 2024-07-31]
73. Completion assessment executive summary: HealthConnect. Global Innovation Fund. 2022 Dec. URL: https://www.globalinnovation.fund/assets/uploads/PDF-Documents/Completion-Assessment/HealthConnect-Investment-Assessment-Executive-Summary.pdf [accessed 2024-07-31]
74. National department of health and Clinton health access initiative. North West Department of Education, Republic of South Africa. 2023 Mar 10. URL: https://desd.nwpg.gov.za/?news=national-department-of-health-and-clinton-health-access-initiative [accessed 2024-07-31]
75. Mkhize highlights the way forward for SA's TB response. South African Online Portal. 2021 Feb 5. URL: https://sacoronavirus.co.za/2021/02/05/mkhize-highlights-the-way-forward-for-sas-tb-response/ [accessed 2024-07-31]
76. Tom-Aba D, Silenou BC, Doerrbecker J, Fourie C, Leitner C, Wahnschaffe M, et al. The surveillance outbreak response management and analysis system (SORMAS): digital health global goods maturity assessment. JMIR Public Health Surveill 2020 Apr 29;6(2):e15860 [FREE Full text] [doi: 10.2196/15860] [Medline: 32347809]
77. Tanzania's new digital App that can track Ebola and other epidemics. SACIDS Foundation for One Health. 2018 May 25. URL: https://sacids.org/news/tanzanias-new-digital-app-can-track-ebola-epidemics/ [accessed 2024-03-04]
78. Buguzi S. Tanzanians on digital mission in Ebola-hit Congo. The Citizen. 2018 Oct 22. URL: https://www.thecitizen.co.tz/tanzania/magazines/health-/tanzanians-on-digital-mission-in-ebola-hit-congo-2659320 [accessed 2024-03-04]
79. AfyaData App in Mozambique. SACIDS Foundation for One Health YouTube page. 2021 Jul 19. URL: https://www.youtube.com/watch?v=yUMM99Gqbfc [accessed 2024-03-04]
80. TechnoHealth surveillance newsletter, volume 1, number 8. Southern African Centre for Infectious Disease Surveillance. 2016 Oct. URL: https://endingpandemics.org/wp-content/uploads/2016/10/TECHNO_SURV_OCT_2016.pdf [accessed 2024-07-26]
81. Enhancing One Health surveillance systems. Open Access Government. 2019 Sep 8. URL: https://www.openaccessgovernment.org/one-health-surveillance-systems/62354/ [accessed 2024-03-04]
82. Ntirandekura JB, Matemba LE, Kimera SI, Muma JB, Karimuribo ED. Brucellosis and its associated risk factors to humans and domestic ruminants in Kagera Ecosystem, Tanzania. Afr Health Sci 2021 Jun 02;21(2):523-530 [FREE Full text] [doi: 10.4314/ahs.v21i2.6] [Medline: 34795704]
83. Mligo BJ, Sindato C, Yapi RB, Mwabukusi M, Mathew C, Mkupasi EM, et al. Effect of awareness training to frontline health workers and the use of e-based technology on reporting of brucellosis cases in selected pastoral communities, Tanzania: a quasi-experimental study. One Health Outlook 2023 Oct 11;5(1):13 [FREE Full text] [doi: 10.1186/s42522-023-00084-3] [Medline: 37817289]
84. Mustafa UK, Kreppel KS, Brinkel J, Sauli E. Digital technologies to enhance infectious disease surveillance in Tanzania: a scoping review. Healthcare (Basel) 2023 Feb 06;11(4):470 [FREE Full text] [doi: 10.3390/healthcare11040470] [Medline: 36833004]
85. Disease surveillance and outbreak response technology in West Africa. European Commission. URL: https://international-partnerships.ec.europa.eu/news-and-events/stories/disease-surveillance-and-outbreak-response-technology-west-africa_en [accessed 2024-04-19]
86. Academic model providing access to healthcare. United States Agency for International Development. 2016. URL: https://2012-2017.usaid.gov/sites/default/files/documents/1860/AMPATH.pdf [accessed 2024-07-26]
87. Sense Followup: optimising the workflow of contact tracers in the West African Ebola epidemic. eHealth Africa. URL: https://ehealthafrica.github.io/case-studies/sense-followup.html [accessed 2024-07-31]

88. Ayles HM, Sismanidis C, Beyers N, Hayes RJ, Godfrey-Faussett P. ZAMSTAR, The Zambia South Africa TB and HIV Reduction Study: design of a 2 x 2 factorial community randomized trial. Trials 2008 Nov 07;9:63 [FREE Full text] [doi: 10.1186/1745-6215-9-63] [Medline: 18992133]
89. Martinson NA, Lebina L, Webb EL, Ratsela A, Varavia E, Kinghorn A, et al. Household contact tracing with intensified tuberculosis and human immunodeficiency virus screening in South Africa: a cluster-randomized trial. Clin Infect Dis 2022 Sep 14;75(5):849-856 [FREE Full text] [doi: 10.1093/cid/ciab1047] [Medline: 34950944]
90. Turimumahoro P, Tucker A, Gupta AJ, Tampi RP, Babirye D, Ochom E, et al. A cost analysis of implementing mobile health facilitated tuberculosis contact investigation in a low-income setting. PLoS One 2022 Apr 1;17(4):e0265033 [FREE Full text] [doi: 10.1371/journal.pone.0265033] [Medline: 35363783]
91. Guinness L, Kumaranayake L, Hanson K. A cost function for HIV prevention services: is there a 'u' - shape? Cost Eff Resour Alloc 2007 Nov 05;5(1):13 [FREE Full text] [doi: 10.1186/1478-7547-5-13] [Medline: 17983475]
92. Bernet PM, Singh S. Economies of scale in the production of public health services: an analysis of local health districts in Florida. Am J Public Health 2015 Apr;105 Suppl 2(Suppl 2):S260-S267. [doi: 10.2105/AJPH.2014.302350] [Medline: 25689207]
93. Bousmah MA, Iwuji C, Okesola N, Orne-Gliemann J, Pillay D, Dabis F, et al. Costs and economies of scale in repeated home-based HIV counselling and testing: evidence from the ANRS 12249 treatment as prevention trial in South Africa. Soc Sci Med 2022 Jul;305:115068 [FREE Full text] [doi: 10.1016/j.socscimed.2022.115068] [Medline: 35665689]

Abbreviations

AMPATH: Academic Model Providing Access to Healthcare
MeSH: Medical Subject Heading
mHealth: mobile health
ODK: Open Data Kit
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
REDCap: Research Electronic Data Capture
SORMAS: Surveillance Outbreak Response Management and Analysis System

Edited by L Buis; submitted 03.10.23; peer-reviewed by D Tom-Aba, Y Ha, J Li; comments to author 27.11.23; revised version received 04.03.24; accepted 05.06.24; published 26.08.24.

Please cite as:
Mudzengi DL, Chomutare H, Nagudi J, Ntshiqa T, Davis JL, Charalambous S, Velen K
Using mHealth Technologies for Case Finding in Tuberculosis and Other Infectious Diseases in Africa: Systematic Review
JMIR Mhealth Uhealth 2024;12:e53211
URL: https://mhealth.jmir.org/2024/1/e53211
doi:10.2196/53211
PMID:

Efficacy of a WeChat-Based, Multidisciplinary, Full-Course Nutritional Management Program on the Nutritional Status of Patients With Ovarian Cancer Undergoing Chemotherapy: Randomized Controlled Trial

Xiaojuan Tian[1*], BS; Yan Liu[1*], BS; Jiahua Zhang[1*], MSc; Lixiao Yang[1], MSc; Linyao Feng[1], BS; Aidong Qi[1], BS; Hanjiazi Liu[1], BS; Pengju Liu[2], BS; Ying Li[3], BS

[1]Ward 2, Gynaecological Oncology Centre, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences, Beijing, China

[2]Department of Nutrition, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences, Beijing, China

[3]Ward 1, Department of Obstetrics, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences, No.1 Shuaifuyuan, Dongcheng District, Beijing, China

[*]these authors contributed equally

Corresponding Author:
Ying Li, BS
Ward 1, Department of Obstetrics, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences, , No.1 Shuaifuyuan, Dongcheng District, Beijing, , China

Abstract

Background: As the most malignant type of cancer in the female reproductive system, ovarian cancer (OC) has become the second leading cause of death among Chinese women. Chemotherapy is the main treatment for patients with OC, and its numerous adverse effects can easily lead to malnutrition. It is difficult to centrally manage patients with OC in the intervals between chemotherapy. The use of WeChat, an effective mobile tool, in chronic disease management has been highlighted.

Objective: This study aimed to implement a continuous follow-up strategy and health monitoring based on the WeChat platform for patients with OC undergoing chemotherapy to ensure that each phase of chemotherapy was delivered on schedule and to improve the survival rate of patients with OC.

Methods: Participants were recruited and randomly assigned to either the WeChat-based nutrition intervention group or the usual care group. A self-administered general information questionnaire was used at enrollment to obtain basic information about the patients. The Patient-Generated Subjective Global Assessment (PG-SGA) Scale was used to investigate the nutritional status of the patients at 3 time points (T0=before the first admission to the hospital for chemotherapy, T1=2 weeks after the first chemotherapy, and T6=2 weeks after the sixth chemotherapy). The blood indices of patients were investigated through the inhospital health care system at 3 times（T0=before the first admission to the hospital for chemotherapy, T1=2 weeks after the first chemotherapy, and T6=2 weeks after the sixth chemotherapy). Patients in the intervention group were introduced to the nutrition applet, invited to join the nutrition management group chat, and allowed to consult on nutritional issues in private chats with nutrition management team members. Linear mixed models were used to analyze changes in each nutritional indicator in the 2 groups, with their baseline measurements as covariates; with group, time, and group-time interactions considered as fixed effects; and with patients considered as random effects.

Results: A total of 96 patients with OC undergoing chemotherapy were recruited into the study. Distribution was based on a 1:1 ratio, with 48 patients each in the nutrition intervention group and the usual care group. The attrition rate after the first chemotherapy session was 18.75%. The mixed linear model revealed that the group-based effect and the group-time interaction effect on PG-SGA scores were significant ($F_{38,38}$=4.763, P=.03; $F_{37,37}$=6.368, P=.01), whereas the time-based effect on PG-SGA scores was not ($F_{38,38}$=0.377; P=.54). The findings indicated that the group-based effect, the time-based effect, and the group-time interaction effect on nutrition-inflammation composite indices were significant ($F_{38,38}$=7.653, P=.006; $F_{38,38}$=13.309, P<.001; $F_{37,37}$=92.304, P<.001; $F_{37,38}$=110.675, P<.001; $F_{38,38}$=10.379, P=.002; and $F_{37,37}$=5.289, P=.02).

Conclusions: This study provided evidence that a WeChat-based, multidisciplinary, full-course nutritional management program can significantly improve the nutritional status of patients with OC during chemotherapy.

Trial Registration: ClinicalTrials.gov NCT06379191; https://clinicaltrials.gov/study/NCT06379191

(JMIR Mhealth Uhealth 2024;12:e56475) doi:10.2196/56475

KEYWORDS

WeChat; nutrition management; ovarian cancer; chemotherapy; mobile health

Introduction

Ovarian cancer (OC) is the most malignant type of tumor in the female reproductive system, with a poor prognosis [1]. According to the latest statistics, there are 196,000 estimated OC cases, 45,000 estimated new cases, and 29,000 estimated OC deaths in China, making this disease become the second leading cause of death among Chinese women [2].

Approximately 90% of patients with OC receive chemotherapy. The carboplatin-paclitaxel combination as the first-line chemotherapy regimen for OC has shown considerable efficacy over the past 30 years [3]. Unfortunately, one of the major adverse effects of chemotherapy is malnutrition. Malnutrition is defined as a nutritional condition in which deficiencies of energy, protein, and other nutrients have measurable adverse effects on tissue or body form, function, and clinical outcomes [4]. The most common adverse effects of chemotherapy, including anorexia, altered taste and smell, food aversions, nausea and vomiting, mucositis, xerostomia, constipation, diarrhea, and early satiety, negatively affect nutritional status [5,6]. Malnutrition has been shown to be one of the leading causes of death in patients with OC [7-9]. It not only severely diminishes the efficacy of treatment but also leads to increased complications, decreased quality of life, prolonged hospitalization, increased health care costs, and shorter survival time [10]. Malnutrition may occur at any time during the perichemotherapy period, making it necessary to manage nutrition throughout the entire course of chemotherapy for patients with OC.

Surprisingly, there is less information available on nutritional management for patients with OC undergoing chemotherapy. Nutritional interventions in previous studies recommended to increase protein and calorie intake [11,12] or to modify the structure of the diet by adopting a Mediterranean diet [11]. The form of intervention was mainly counseled with a dietitian every 3 weeks through telephone [11]. The previous studies lacked real-time monitoring and timely guidance for patients, and there has been a lag in the nutritional management. In addition, they mentioned only guidance on energy intake and dietary patterns and lacked targeted nutritional guidance for chemotherapy-related adverse effects. The guideline provided more detailed dietary guidance for patients with OC undergoing chemotherapy, such as not using restrictive energy intake diets [13,14] and taking oral nutritional supplements according to the actual situation [15,16], which can help improve the nutritional knowledge of patients with OC. However, it did not mention in what way (eg, online or offline), and at what frequency and intensity, the intervention for patients treated with chemotherapy for OC is more conducive to their nutritional management. In addition, the difficulty of centrally managing patients recuperating at home between chemotherapy treatments should also be considered. Personalized nutritional guidance and education are the preferred methods of nutritional intervention for patients with OC treated with chemotherapy and should be carried out throughout the consultation [13]. Therefore, the construction of personalized nutritional management programs and the development of innovative telemedicine interventions for patients with OC undergoing chemotherapy have become urgent issues.

WeChat, a very popular social application in China, has more than 1 billion monthly active users. It is easy to operate and offers multiple functions such as text and voice messaging, free voice and video calls, group chats, subscriptions to public accounts and applets, and so forth. WeChat has been demonstrated to be an effective and more cost-efficient technological tool for chronic disease management [17,18]. Currently, the applications of WeChat in patients with cancer mainly focus on discharge follow-up [19], symptom management [20,21], cancer prevention [22], and intervention of psychological problems [23]. The effect of WeChat application in the nutritional management of patients with cancer has not been explored. Therefore, the aim of this study was to implement a continuous follow-up strategy and health monitoring based on a WeChat platform for patients with OC undergoing chemotherapy to ensure that each phase of chemotherapy was delivered on schedule and to improve the survival rate of patients with OC.

Methods

Study Design

We conducted a parallel, 2-armed, open-label, randomized controlled trial to examine the efficacy of the WeChat-based nutrition intervention compared with usual nutritional care and to explore preliminary effects of the intervention. The study was conducted between February 2023 and October 2023 at Peking Union Medical College Hospital, Chinese Academy of Medical Sciences in Beijing, China.

Participants

Patients were eligible if they were aged 18 years or older with pathologically confirmed OC [24]. The chemotherapy regimen was paclitaxel in combination with carboplatin. Patients had to have normal cognitive ability and proficiency in the use of WeChat. Patients were excluded if they had a malignant tumor of another system, a serious illness, or failure of vital organs (eg, the heart, lungs, liver, and kidneys) or if they were receiving enteral or parenteral nutritional support. All patients provided written informed consent (Multimedia Appendix 1).

Randomization and Allocation

Patients were randomly assigned (1:1) to receive the WeChat-based remote nutrition intervention or usual nutritional care. Randomization was stratified by disease stage (stages I and II vs stages III and IV) and surgical procedure (laparoscopy vs laparotomy) to balance treatment assignment. The random allocation sequence was generated by the study designers (XJT and YL, who were not involved in patient enrollment) by means of a table of random numbers, and the allocation was hidden by the envelope.

Study Procedure

Intervention Group

Establishment of Nutrition Management Team

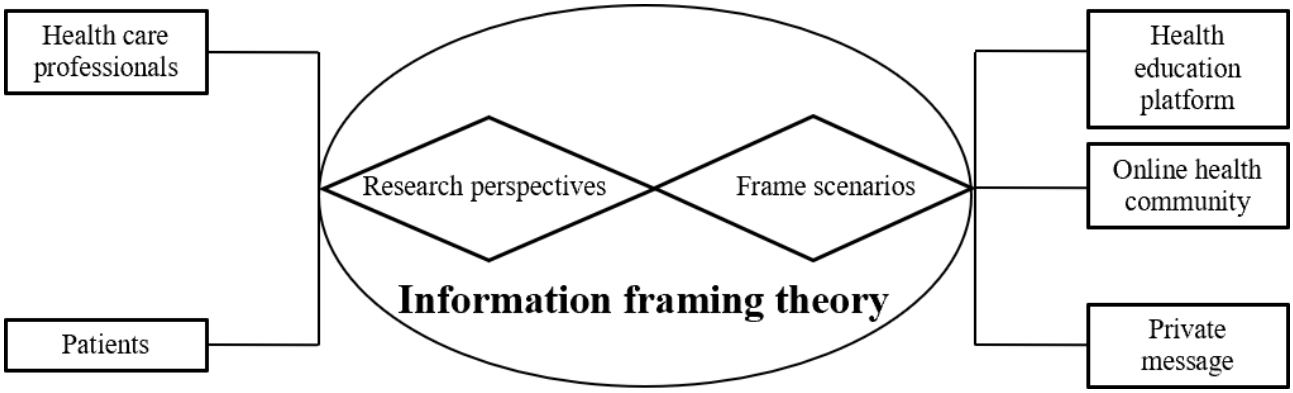
The nutrition management team consisted of a nursing administrator (bachelor's degree, associate nurse practitioner), 2 gynecologic oncologists (PhD, attending physician), 2 oncology specialist nurses (bachelor's degree, charge nurse practitioner), 1 nutrition specialist nurse (bachelor's degree, charge nurse practitioner), 1 dietitian (PhD, associate physician practitioner), and 2 master's degree students in nursing.

Design of the WeChat Platform

Overview

The construction of the WeChat platform relied on information framing theory [25], which emphasized the impact of information on human health behavioral decision-making and health outcomes. It focused the interaction process between people and information on the manifestation of information itself, while reinforcing the dominance of information in both research perspectives and research contexts. The WeChat platform constructed in this study considered the perspectives of both health care professionals and patients and set up 3 different scenarios based on the different needs of the 2 subjects: a health education platform (WeChat applet), a web-based health community (WeChat group chat), and a private question-and-answer session (WeChat private message), with the aim of realizing the maximum benefits of nutritional management. Specific relationships are displayed in Figure 1.

Figure 1. Information framing theory diagram.

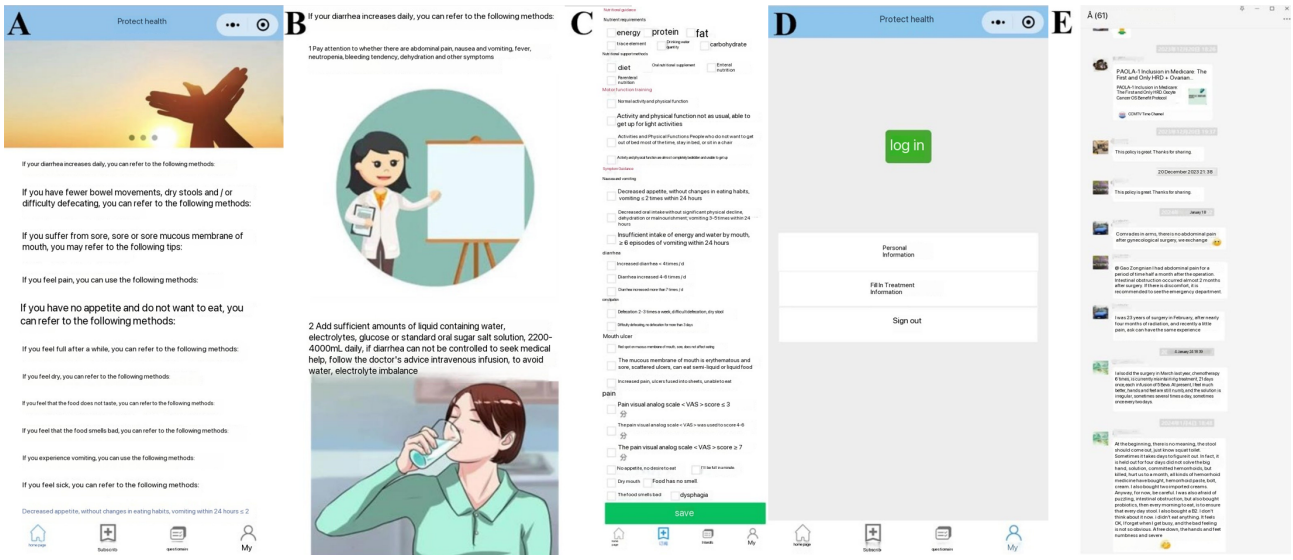


WeChat Applet

With regard to the development of the nutrition WeChat applet "Good Nutrition," it consisted of 2 ports: the computer side and the WeChat side. The computer side was operated and managed by health professionals. The WeChat terminal could be operated by patients with OC undergoing chemotherapy. Patient privacy and data security were protected through access control and permission control, patient data transfer and anonymization, redundant storage, and data backup. The applet was set up with 4 modules: health education column, subscription content column, questionnaire column, and personal information column (Figure 2A).

The content of the health education column was compiled based on a previously constructed multidisciplinary full-course nutritional management program (Table S1 in Multimedia Appendix 2) for patients with OC during chemotherapy, which mainly included chemotherapy symptom management (eg, diarrhea, nausea and vomiting, mouth ulcers), nutritional risk management (based on the Patient-Generated Subjective Global Assessment [PG-SGA] score), and physical activity management. This section provided comprehensive management measures based on symptom severity rating, nutritional risk level, and activity level. In addition, we used a combination of text and images in the presentation of the content, making it detailed and easy to understand (Figure 2B).

Figure 2. WeChat applet interface settings. (A) The applet was set up with 4 modules, namely, health education column, subscription content column, questionnaire column, and personal information column. (B) The content of the health education column. (C) The subscription content section summarized the 3 main sections of instructions: nutritional guidance, motor function training, and symptom guidance. (D) The personal information section was simple to set up and included basic patient information, treatment information, and an option to log out. (E) A group chat for patient-directed learning and encouraging patients to ask questions, share personal experiences, and discuss lifestyle topics.



The subscription content section summarized the 3 main sections of instructions: nutritional guidance (nutrient requirements and nutritional support methods), motor function training, and symptom guidance (nausea, vomiting, diarrhea, constipation, mouth ulcers, and pain). On the one hand, patients could subscribe to the content they were interested in according to their own situation, and the system would remind them to check whether the subscription content has been updated. On the other hand, members of the nutrition team could obtain the subscription status of the patients from the back end, so as to have a better understanding of the adverse effects of the individuals during chemotherapy and thus give more targeted guidance (Figure 2C).

Patients were briefly quizzed on a regular basis in the form of a questionnaire, and open-ended questions were included to ask whether they had any questions about the nutritional management process.

With regard to the log-in interface, the personal information section was simple to set up and included basic patient information, treatment information, and an option to log out (Figure 2D).

Group Chat

Throughout the intervention phase, we sent out evidence-based nutrition papers taken from WeChat applets in group chats for patient-directed learning and encouraged patients to ask questions, shared personal experiences, and discussed lifestyle topics. We also critically analyzed the public papers shared by patients to help them judge the quality of nutritional information (Figure 2E).

Private Message

Patients can always choose to have a private chat with a member of the nutrition management team via WeChat. For example, a patient with an aggravated chemotherapy reaction may seek help in a private chat, and our doctors would give some professional advice. In addition, for common problems, we also posted answers in the group for other patients' reference.

Implementation Procedure

Detailed and clear interventions are shown in Table 1.

Table . Interventions.

Time of intervention	Objectives of intervention	Interventions
1. The first admission for chemotherapy	To enable patients to master the nutrition management WeChat platform (WeChat applet, WeChat group chat, WeChat private message) function and operation use method	WeChat applet: 1. Searched for the "Good Nutrition" WeChat applet 2. Added it to My Applet 3. Opened the notification permission 4. Logged in the Applet 5. Introduced to the main functional sections Group chat: 1. Invited to join the group chat 2. Informed that throughout the intervention phase, evidence-based scientific papers would be sent to the group chat. Asking questions, sharing personal experiences, and discussing lifestyles were permitted. Group chat could help identify questionable nutrition information. Answer questions via private message: 1. Allowed private chats with nutrition team members if in doubt.
2. During each session of chemotherapy	Enable patients to understand in advance the possible adverse reactions related to chemotherapy and the countermeasures	Explained the complications associated with chemotherapy, and given a chemotherapy care booklet and a leaflet on diet and nutrition. Before discharge, the patients were given the nutritional guidebook again.
3. Each chemotherapy interval	Real-time monitoring of patients' chemotherapy adverse reactions and nutritional status, and timely resolution of patients' dilemmas in coping with chemotherapy-related adverse reactions	WeChat applet: 1. Updated the health education content every 2 months. 2. Checked the top 3 health education topics in the subscription content weekly, and sent relevant papers to the group chat for discussion and analysis. 3. Pushed the nutritional knowledge questionnaire weekly, according to the patients' answer situation in the group chat for nutritional guidance. Group chat: 1. Summarized nutritional knowledge questions once a week and answered centrally. 2. Summarized and corrected the misconceptions shared by patients during the day every night. Answer questions via private message: 1. One nutrition team member would be responsible for answering questions via private message and replying promptly after seeing the message every day.

Control Group

For patients in the usual care group, upon admission, patients were provided with admission counseling, given an explanation about the complications associated with chemotherapy, and given a chemotherapy care booklet and a leaflet on diet and nutrition. Before discharge, the patients were given the nutritional guidebook again. Responsible nurses made 1 telephone follow-up visit between chemotherapy sessions.

Data Collection and Outcomes

Self-reported sociodemographic characteristics (eg, age, BMI, marital status, education level, etc) and clinical disease features (eg, disease stage, surgical procedure, etc) were collected at baseline.

The primary outcome was assessed by the PG-SGA Scale to evaluate the nutritional status of patients at 3 time points (T0=before the first admission to the hospital for chemotherapy, T1=2 weeks after the first chemotherapy, and T6=2 weeks after the sixth chemotherapy). The PG-SGA was used as a prognostic tool developed specifically for patients with cancer to evaluate

the nutritional status [26]. It consisted of 2 subscales: the patient self-assessment scale and the medical staff assessment scale. The former integrated short-term weight loss, food intake (including amount eaten, type of food, manner of eating, etc), symptoms affecting eating, and activity or physical functioning. The latter included medical history, metabolic stress, and physical examination provided by medical staff [27]. Each item in the PG-SGA Scale has a score range of 0 - 4. The more severe the symptoms in relation to malnutrition, the higher the assigned value. The total score was compared and analyzed in three groups: (1) well nourished (0 - 1 point), (2) moderately or suspected of being malnourished (2 - 8 point), and (3) severely malnourished (≥9 point) [28]. A total score between 0 and 35 quantitatively informs the severity of malnutrition and types of intervention needed: 0 - 1 point indicates no need for any intervention, 2 - 3 points suggest education needs for the patients and the family, 4 - 8 points indicate the need for a referral to a dietitian, and a score of 9 or more recommends an action of critical nutritional intervention. The PG-SGA Scale had high validity (sensitivity 100%, specificity 88%) [29]. It had been accepted by the Oncology Nutrition Dietetic Practice Group of the American Dietetic Association as the standard for nutrition assessment for patients with cancer [26,30,31].

The secondary outcome was to compare the blood test indices of the 2 groups at the same 3 time points, including nutrition-related blood indices, such as total protein (g/L), albumin (g/L), prealbumin (g/L), and hemoglobin (g/L); inflammation-related blood indices, such as leukocytes (10^9/L), lymphocytes (10^9/L), neutrophils (10^9/L), and platelets (10^9/L); and nutrition-inflammation composite indices, such as the prognostic nutritional index (PNI) and systemic immunoinflammatory index (SII). The PNI reflected the chronic inflammation, immune status, and nutritional status of patients with cancer and can be used to predict the risk of postchemotherapy complications in patients with cancer. The PNI is calculated as "albumin (g/L) + 5 × lymphocyte (10^9/L)" [32]. The SII is a new scoring system proposed by Chinese scholars in 2014 to assess the functional status of the immune system in patients with tumor [33]. The SII is calculated as "platelet (10^9/L) × neutrophil (10^9/L)]/lymphocyte (10^9/L)," and the composite score combining the 3 indexes can better reflect the balance of the body's inflammatory response and immune status. The nutrition-related blood indices (total protein, albumin, prealbumin, and hemoglobin), inflammation-related blood indices (leukocytes, lymphocytes, neutrophils, and platelets), and nutrition-inflammation composite indices (PNI and SII) were collected from hospital electronic medical records.

Patient enrollment and baseline information collection were done by the member of the group responsible for recruitment, who was aware only of the patient inclusion and exclusion criteria and was not aware of the patient grouping. The assessment of the outcome indicators—PG-SGA and blood test indices—was done by the member responsible for efficacy assessment, who was also unaware of the patient subgroups. This study was open-labeled.

Statistical Analysis

Based on the primary outcome of PG-SGA score, a target sample size of 78 participants was estimated to provide 90% power with an α value of .05 to detect a difference between control and intervention groups. This calculation was calculated using PASS 15.0 (NCSS LLC), referring to data from previous study [34], taking into account a 20% dropout rate. Baseline characteristics were presented as medians (IQRs) or percentages and were compared using the chi-square test or the Mann-Whitney test, as appropriate. A linear mixed model was used to analyze these score changes, with their baseline measurements used as covariates. Group, time, and group-time interaction were treated as fixed effects, while the patient was treated as a random effect. Missing data due to loss to follow-up were assumed to be missing at random. The estimated within- and between-group differences were reported with their respective 95% CIs. A 2-sided *P* value of <.05 was considered statistically significant. All data analyses were performed using SPSS (version 25.0; IBM Corp) on an intention-to-treat basis.

Ethical Considerations

This study was approved by the Medical Ethics Committee of Peking Union Medical College Hospital, Chinese Academy of Medical Sciences (ZS-3411).

Results

Participant Flow and Baseline Characteristics

A total of 120 patients were notified of the study before they were admitted to the hospital for the first chemotherapy treatment, of which 18 patients did not fulfill the inclusion criteria and 6 patients refused to participate in the study. Therefore, 96 patients were involved in the random allocation, that is, 48 patients each in the intervention and control groups. During the follow-up, there was sample dropout in both groups: 1 (1/48, 2.1%) patient and 5 (5/48, 10.4%) patients in both groups who refused to continue participating in the study, in addition to 8 (8/48, 16.67%) patients and 4 (4/48, 8.3%) patients in both groups who went to other hospitals for chemotherapy. The flow diagram is shown in Figure 3. There were no statistically significant differences in demographic and clinical characteristics between the intervention and control groups, as shown in Table 2.

Figure 3. Flowchart of recruitment.

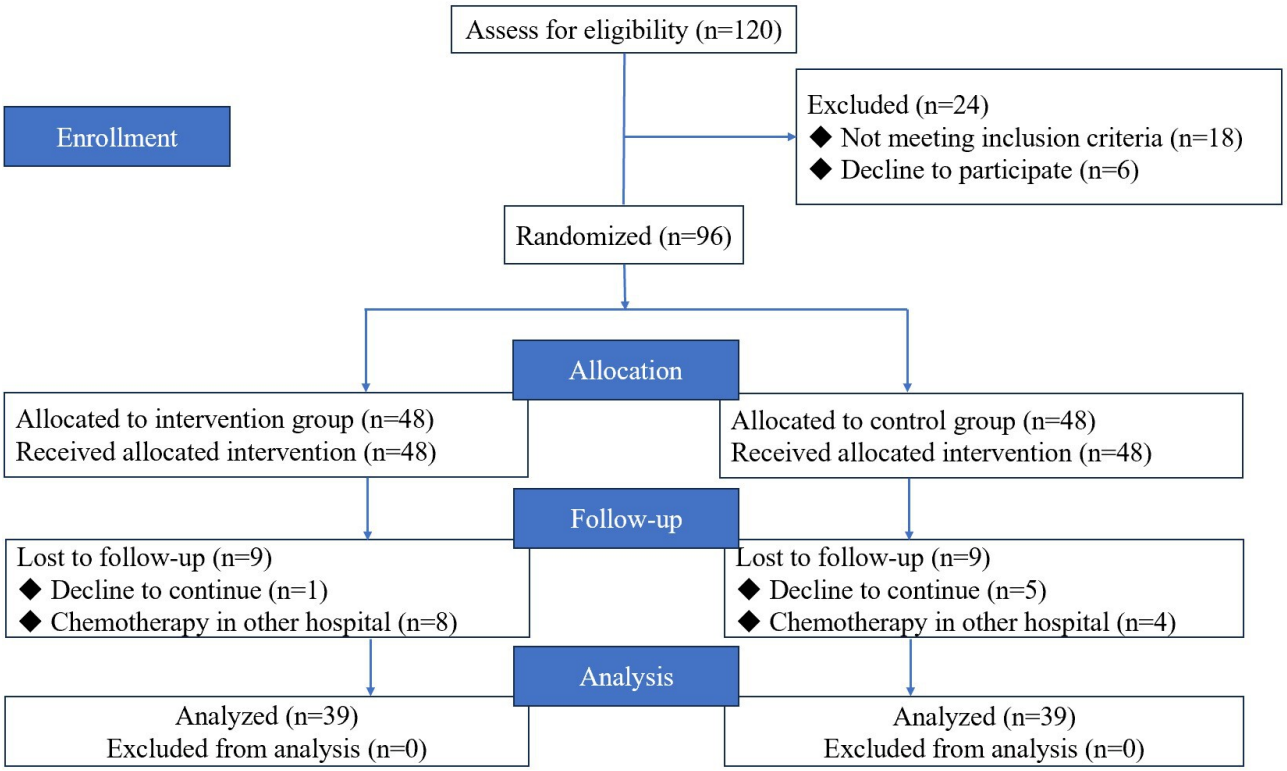


Table . Demographic and clinical characteristics.

Variables	Intervention group (n=39)	Control group (n=39)	Chi-square (*df*)/*Z*	*P* value
Age (years), median (IQR)	55.00 (49.00-60.00)	56.00 (48.00-65.00)	−0.50[a]	.62
BMI (kg/m^2), median (IQR)	22.48 (20.45-24.40)	22.14 (19.78-24.50)	−0.71[a]	.48
Marital status, n (%)			0.000 (38)[b]	>.99
Married	37 (94.9)	37 (94.9)	N/A[c]	N/A
Unmarried, divorced, or widowed	2 (5.1)	2 (5.1)	N/A	N/A
Education level, n (%)			0.56 (38)[b]	.76
Junior high school and below	13 (33.3)	10 (25.6)	N/A	N/A
High school	11 (28.2)	12 (30.8)	N/A	N/A
College degree and above	15 (38.5)	17 (43.6)	N/A	N/A
Number of children, n (%)			0.47 (38)[b]	>.99
0	1 (2.6)	2 (5.1)	N/A	N/A
1	25 (64.1)	25 (64.1)	N/A	N/A
≥2	13 (33.3)	12 (30.8)	N/A	N/A
Family monthly income per capita (¥[d]), n (%)			1.39 (38)[b]	.71
≤2000	7 (18.0)	7 (18.0)	N/A	N/A
2001-4000	9 (23.1)	9 (23.1)	N/A	N/A
4001-6000	10 (25.6)	14 (35.8)	N/A	N/A
>6000	13 (33.3)	9 (23.1)	N/A	N/A
Professional status, n (%)			0.83 (38)[b]	.36
Active employee	19 (48.7)	24 (61.54)	N/A	N/A
Inactive employee	20 (51.3)	15 (38.46)	N/A	N/A
Type of medical insurance, n (%)			1.46 (38)[b]	.74
Free medical care	5 (12.8)	2 (5.1)	N/A	N/A
Beijing Medical Insurance	8 (20.5)	8 (20.5)	N/A	N/A
Remote medical insurance	20 (51.3)	22 (56.4)	N/A	N/A
New rural cooperative medical insurance	6 (15.4)	7 (18.0)	N/A	N/A
Residence, n (%)			0.11 (38)[b]	.75
Town	33 (84.6)	34 (87.2)	N/A	N/A
Rural	6 (15.4)	5 (12.8)	N/A	N/A
Stage, n (%)			0.87 (38)[b]	.87
I	5 (12.8)	5 (12.8)	N/A	N/A
II	15 (38.5)	17 (43.6)	N/A	N/A
III	17 (43.6)	13 (33.3)	N/A	N/A
IV	2 (5.1)	4 (10.3)	N/A	N/A
Surgical procedure, n (%)			0.05 (38)[b]	.82
Laparoscopy	19 (48.7)	18 (46.1)	N/A	N/A
Laparotomy	20 (51.3)	21 (53.9)	N/A	N/A

[a]*Z* value.

[b]Chi-square value.

[c]N/A: not applicable.

[d]A currency exchange rate of ¥1=US $0.14 is applicable.

Nutritional Status—PG-SGA Score

Figure 4 demonstrates the PG-SGA scores of the intervention and control groups at T0, T1, and T6. Among the 3 time points, there was a statistically significant difference in PG-SGA scores between the 2 groups only at T6. By longitudinal comparison, we found that the difference of PG-SGA scores was statistically significant in the intervention group between T0 and T1 (P<.05), as well as in the control group, with scores lower than the T0 period. However, comparing T0 and T6, only T6 in the control group had a lower score than T0 (P<.05). The mixed linear model revealed that the group-based effect on PG-SGA scores was significant (F_{47}=4.763; P=.03), whereas the time-based effect on PG-SGA scores was not (F_{47}=0.377; P=.54). The group-time interaction effect on PG-SGA scores was statistically significant (F_{46}=6.368; P=.01) (Table 3).

Figure 4. Changes in PG-SGA scores at T0, T1, and T6 between intervention and control groups. T0=prechemotherapy; T1=after first chemotherapy; and T6=after sixth chemotherapy. *There was a significant difference between the intervention and control groups (P<.05). PG-SGA: Patient-Generated Subjective Global Assessment.

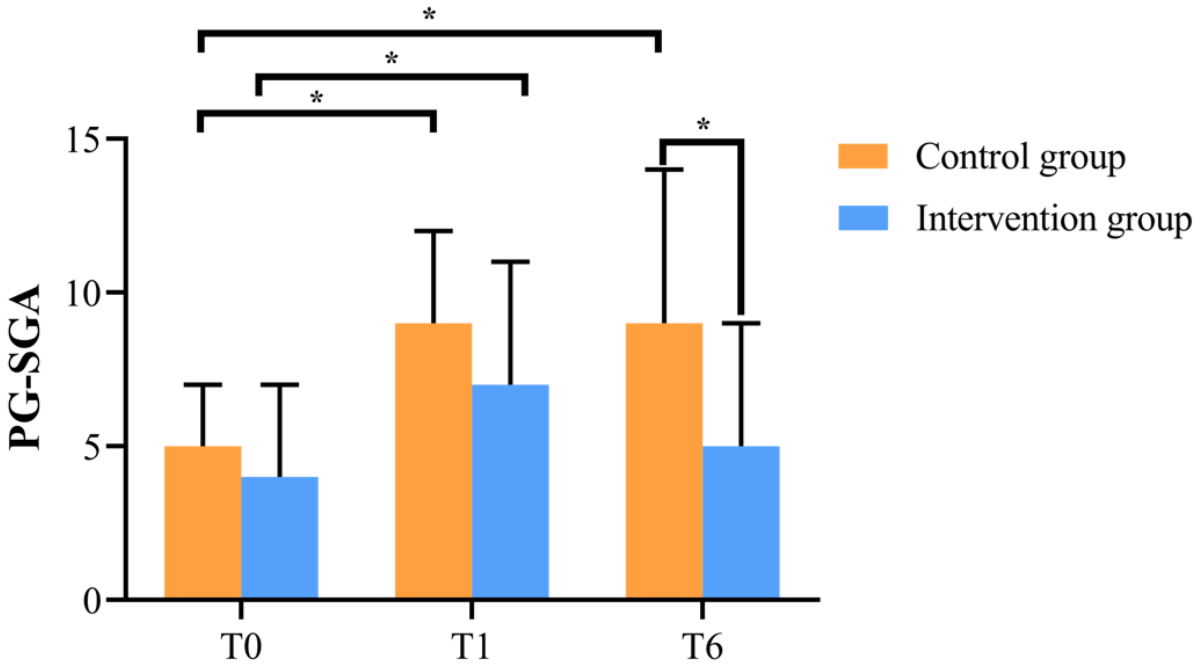


Table . Changes within groups and comparisons between groups regarding the Patient-Generated Subjective Global Assessment (PG-SGA) scores: linear mixed model analysis[a].

Groups	T0 (n=39)	T1 (n=39)		T6 (n=39)	
	Score, median (IQR)	Score, median (IQR)	Change from T0, median difference (95% CI)	Score, median (IQR)	Change from T0, median difference (95% CI)
Intervention	4.00 (3.00 to 7.00)	7.00 (4.00 to 11.00)	−3.00 (−4.00 to 0.00)	5.00 (3.00 to 9.00)	−1.00 (−2.00 to 1.00)
Control	5.00 (2.00 to 7.00)	9.00 (2.00 to 12.00)	−4.00 (−5.00 to 0.00)	9.00 (4.00 to 14.00)	−4.00 (−6.00 to −2.00)
Change between groups, median difference (95% CI)	1.00 (−1.00 to 2.00)	2.00 (−2.00 to 2.00)	N/A[b]	4.00 (1.00 to 5.00)	N/A

[a]Linear mixed model was used for the analysis of changes within group and comparisons between groups of the PG-SGA scores, with T0 measurement of the PG-SGA scores as covariate; group, time, and group × time interaction as fixed effects; and patient as random effect. PG-SGA: (group) F_{47}=4.763, P=.03; (time) F_{47}=0.377, P=.54; (group × time interaction) F_{46}=6.368, P=.01.

[b]N/A: not applicable.

Compound Immune-Nutritional and Compound Immune-Inflammatory Indicators in Blood

Among the 3 time points, there was a statistically significant difference in the PNI and SII between the 2 groups only at T6. By longitudinal comparison, there was a statistically significant difference in the comparison between all 3 time points for the PNI (Figure 5A) and SII (Figure 5B) in the intervention group, as well as in the control group (P<.05) (Figure 5). The mixed linear model revealed that the group-based effect on nutrition-inflammation composite indices was significant (F_{47}=7.653, P=.006; F_{47}=13.309, P<.001), as well as the time-based effect on nutrition-inflammation composite indices (F_{47}=92.304, P<.001; F_{47}=110.675, P<.001). The group-time interaction effect on nutrition-inflammation composite indices was statistically significant (F_{46}=10.379, P=.002; F_{46}=5.289, P=.02) (Table 4).

Figure 5. Changes in compound immune-nutritional and compound immune-inflammatory indicators in blood at T0, T1, and T6 between intervention and control groups. The line graphs show the differences and changes in prognostic nutritional index (**A**) and systemic immunoinflammatory index (**B**) between the intervention and control groups at T0, T1, and T6. *There was a significant difference between the intervention and control groups (P<.05). With regard to "a," "b," and "c," different letters at the 2 time points indicate that there is a significant difference in the blood indices of the intervention or control group at the 2 time points (P<.05).

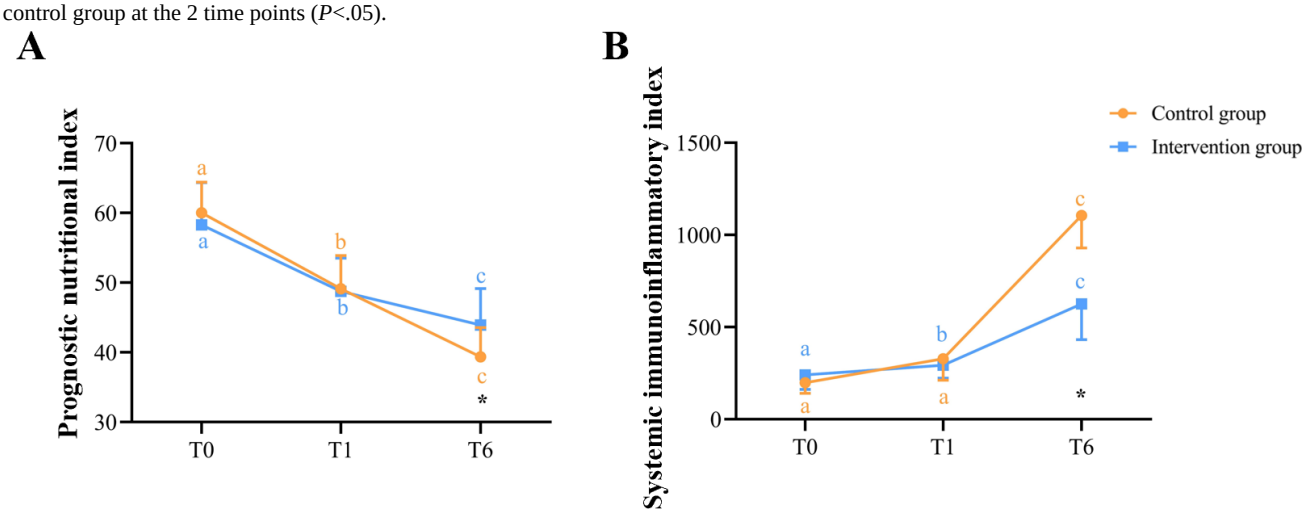


Table . Changes within groups and comparisons between groups regarding the nutrition-inflammation composite indices: linear mixed model analysis[a].

Groups	T0 (n=39)	T1 (n=39)		T6 (n=39)	
	Score, median (IQR)	Score, median (IQR)	Change from T0, median difference (95% CI)	Score, median (IQR)	Change from T0, median difference (95% CI)
PNI [b]					
Intervention	57.95 (53.50 to 61.00)	48.90 (44.55 to 52.80)	9.05 (6.550 to 11.50)	44.10 (41.20 to 47.35)	13.85 (11.45 to 16.60)
Control	60.40 (56.70 to 63.05)	48.5 (46.2 to 51.25)	11.9 (9.20 to 13.30)	39.85 (36.80 to 42.65)	20.55 (18.80 to 22.55)
Change between groups, median difference (95% CI)	−2.45 (−0.15 to 4.80)	0.4 (−2.20 to 2.35)	N/A[c]	−4.25 (−6.70 to −2.25)	N/A
SII [d]					
Intervention	241.07 (161.70 to 306.38)	293.6834 (223.17 to 382.31)	−52.612 (−113.46 to −16.37)	625.37 (432.02 to 1273.21)	−384.30 (−679.19 to −280.35)
Control	197.6443 (140.86 to 327.68)	328.31 (212.17 to 552.86)	−130.66 (−206.08 to −58.14)	1106.15 (929.42 to 1672.49)	−908.51 (−1026.72 to −785.50)
Change between groups, median difference (95% CI)	−43.4271 (−72.70 to 27.57)	34.6246 (−31.67 to 127.02)	N/A	480.7798 (213.29 to 624.26)	N/A

[a]Linear mixed model was used for the analysis of changes within group and comparisons between groups of the nutrition-inflammation composite indices, with T0 measurement of the nutrition-inflammation composite indices as covariate; group, time, and group × time interaction as fixed effects; and patient as random effect.

[b]PNI (prognostic nutritional index): (group) F_{47}=7.653, P=.006; (time) F_{47}=92.304, P<.001; (group × time interaction) F_{46}=10.379, P=.002.

[c]N/A: not applicable.

[d]SII (systemic immunoinflammatory index): (group) F_{47}=13.309, P<.001; (time) F_{47}=110.675, P<.001; (group × time interaction) F_{46}=5.289, P=.02.

Nutrition-Related and Inflammation-Related Blood Indices

Changes in blood indices related to nutrition and inflammation at T0, T1, and T6 are shown in Figures 6 and 7, respectively. Figure 6 showed the differences and changes in total protein (Figure 6A), albumin (Figure 6B), prealbumin (Figure 6C), and hemoglobin (Figure 6D) between the intervention and control groups at T0, T1, and T6. Figure 7 showed the differences and changes in white blood cells (Figure 7A), lymphocytes (Figure 7B), neutrophils (Figure 7C), and platelets (Figure 7D) between the intervention and control groups at T0, T1, and T6. The mixed linear model displayed that the group-based effect on total protein (F_{47}=19.712; P<.001), prealbumin (F_{47}=10.029; P=.002), hemoglobin (F_{47}=6.225; P=.01), lymphocytes (F_{47}=3.921; P=.05), and neutrophils (F_{47}=72.058; P<.001) was significant. The time-based effect on total protein (F_{47}=75.642; P<.001), albumin (F_{47}=24.496; P<.001), prealbumin (F_{47}=5.054; P=.03), hemoglobin (F_{47}=27.744; P<.001), leukocytes (F_{47}=29.950; P<.001), lymphocytes (F_{47}=76.220; P<.001), neutrophils (F_{47}=182.218; P<.001), and platelets (F_{47}=9.350; P=.003) was statistically significant. The group-time interaction effect was statistically significant on total protein (F_{46}=5.626; P=.02), albumin (F_{46}=7.346; P=.007), hemoglobin (F_{46}=6.225; P=.01), leukocytes (F_{46}=3.328; P=.07), and neutrophils (F_{46}=45.555; P<.001). Detailed information is given in Table S2 in Multimedia Appendix 2.

Figure 6. Changes in nutritional indicators in blood at T0, T1, and T6 between intervention and control groups. The line graphs show the differences and changes in total protein (**A**), albumin (**B**), prealbumin (**C**), and hemoglobin (**D**) between the intervention and control groups at T0, T1, and T6. *There was a significant difference between the intervention and control groups (P<.05). With regard to "a, "b," and "c," different letters at the 2 time points indicate that there is a significant difference in the blood indices of the intervention or control group at the 2 time points, whereas the same letter indicates that there was no significant difference (P>.05).

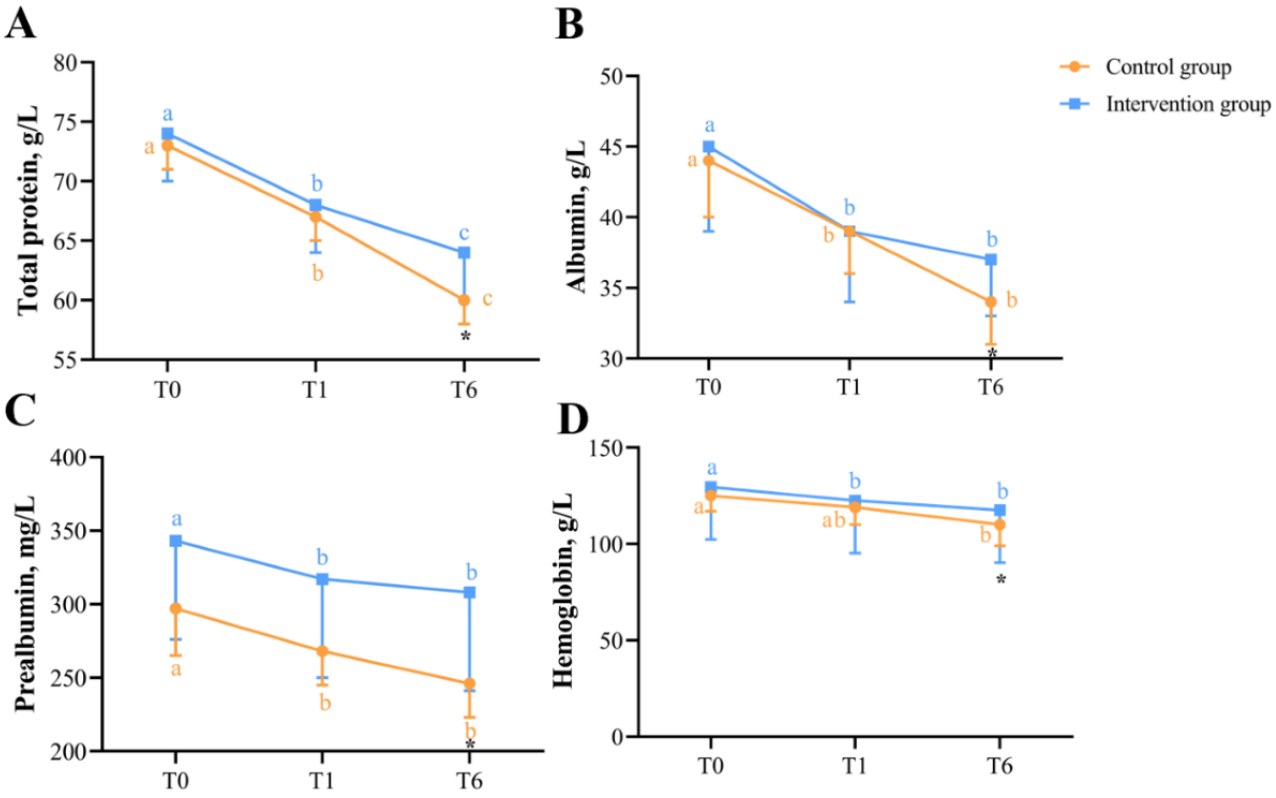
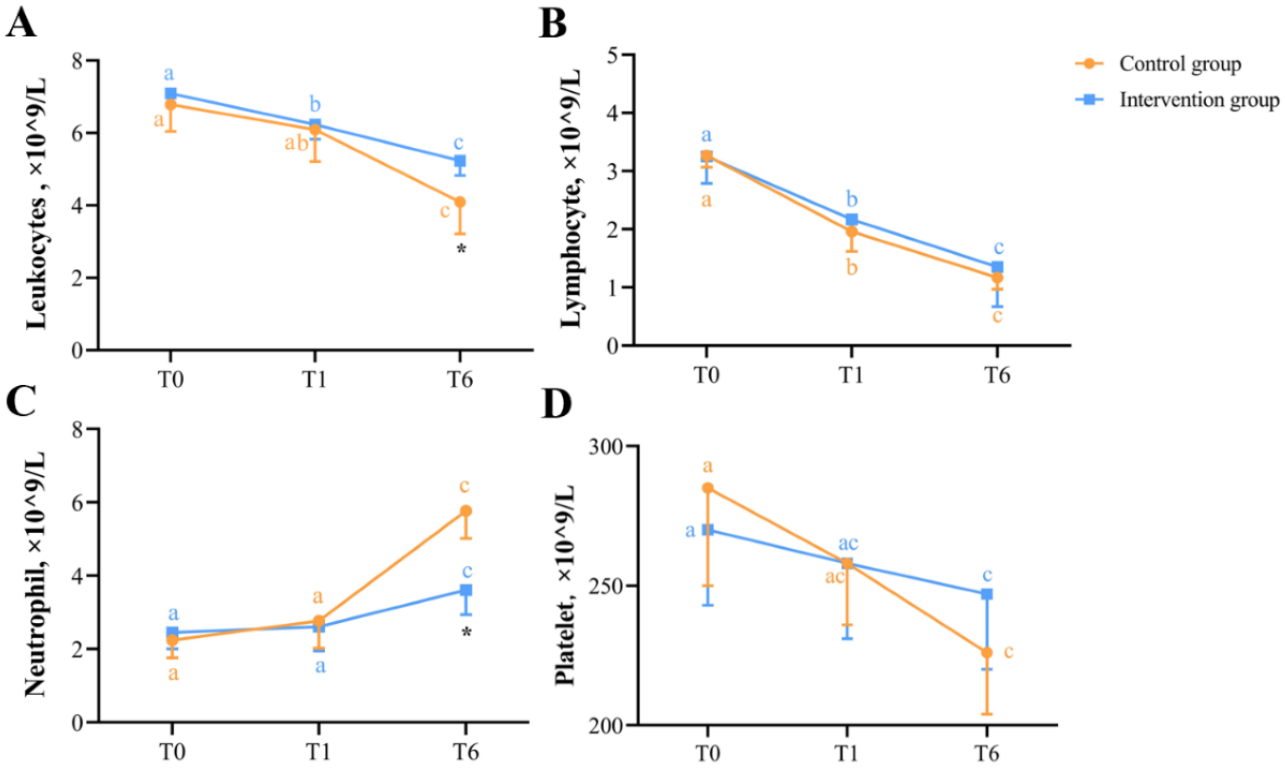


Figure 7. Changes in inflammatory parameters in blood at T0, T1, and T6 between intervention and control groups. The line graphs show the differences and changes in white blood cells (**A**), lymphocytes (**B**), neutrophils (**C**), and platelets (**D**) between the intervention and control groups at T0, T1, and T6. *There was a significant difference between the intervention and control groups (P<.05). With regard to "a," "b," and "c," different letters at the 2 time points indicate that there is a significant difference in the blood indices of the intervention or control group at the 2 time points (P<.05).



Discussion

Principal Findings

Owing to the difficulty of uniformly managing patients with OC undergoing chemotherapy during the chemotherapy intervals, remote and easily accessible alternatives are needed to improve the nutritional status of patients with OC undergoing chemotherapy. This study of patients with OC undergoing chemotherapy analyzed the effect of a full-course nutritional management program based on the WeChat platform on nutritional status. The results obtained showed that the interventions led to significant improvements in nutritional status in terms of reducing PG-SGA scores, increasing nutrition-related blood indices, and decreasing inflammation-related blood indices.

Comparison With Previous Work

Patients in both the intervention and control groups showed significant increases in the PG-SGA scores after the first chemotherapy compared with baseline, demonstrating that the nutritional status was markedly impaired in patients undergoing chemotherapy. This finding was consistent with those of other related studies [35-37]. The 2 main reasons are analyzed as follows. On the one hand, gastrointestinal symptoms such as nausea and vomiting caused by the paclitaxel-combined carboplatin chemotherapy regimen are the most pronounced and severe after the first chemotherapy session [38], which limits the patient's nutrient intake and absorption, resulting in a loss of body weight and muscle mass, and an elevated PG-SGA score. On the other hand, patients may also develop a cluster of peripheral neurological symptoms after the first chemotherapy session, affecting their motor function [39].

There was no significant difference in PG-SGA scores between the 2 groups after the first chemotherapy treatment, suggesting that the short-term benefits of nutritional intervention strategies have not yet been demonstrated. In addition to the adverse effects of chemotherapy, some patients may experience the following symptoms after their first chemotherapy treatment: anxiety, depression, pain, sexual dysfunction, fatigue, personality changes, and other symptoms [40]. Their normal adaptive mechanisms are challenged and their future planning is compromised, leading to a reduction in their disease self-efficacy [41].

At the sixth chemotherapy follow-up visits, the PG-SGA scores of the intervention group were not significantly different from those at baseline, indicating that the program helped patients gradually return to their prechemotherapy nutritional status during the sixth course of chemotherapy. Meanwhile, the PG-SGA scores of the intervention group remained significantly lower than those of the control group at sixth chemotherapy. This result was consistent with the study by Wang et al [42]. Similarly, we found significant differences between the 2 groups after the sixth session of chemotherapy for both indicators, the PNI and SII. After the sixth chemotherapy session, the intervention group had a higher PNI level and a lower SII level than the control group, which suggested that the nutritional intervention program during chemotherapy may further improve the long-term nutritional prognosis of patients after chemotherapy. The WeChat platform was divided into 3 main interventions: WeChat applet, WeChat group chat, and private

message. First of all, the health education content of the health education module in the WeChat applet was a comprehensive, detailed, and authoritative holistic nutritional intervention program during chemotherapy for patients to review and learn through literature review and expert correspondence, so as to enhance their nutritional literacy. The section was presented in the form of text combined with pictures to easily stimulate patients' interest in reading. The knowledge subscription section was for patients to subscribe to the content according to their own interests and symptoms after chemotherapy, so that patients could read and learn repeatedly anytime and anywhere. The questionnaire test was to regularly check the learning effect of patients, so as to provide targeted intervention. The joint intervention of the 3 sections improved patients' nutritional literacy during chemotherapy, thus realizing the improvement of nutritional status. Compared with Keum et al [43], who intervened in patients' nutritional status through a smartphone app, the WeChat applet in our study was more convenient and economical as it did not require downloads. Second, the WeChat group chat broke through the limitations of time and space to provide a platform for patients to communicate anytime, anywhere [44]. This platform enabled patients to share nutritional experiences, released negative emotions, and gained encouragement from each other [45]. In addition, nutritional team members would correct incorrect nutritional concepts and knowledge in the group chat in a timely manner. Relevant discussions in the group chat could be retained in the WeChat group for a long period of time, which facilitated repeated learning. Peer support from health care professionals and patients in the WeChat group chat also enhanced patients' trust and reliance on the nutrition intervention program [46]. The private massage provided a platform for patients to consult and took timely and effective countermeasures when they encountered nutritional problems [47].

Limitations and Strengths

There are some limitations in our study. First, to participate in the WeChat-based nutrition intervention program, patients must have mobile devices with web access, which suggested that the patients were younger or have supportive family members. Second, the sample size was small. A barrier to recruiting more participants is that most patients choose to go home for follow-up chemotherapy due to their financial situation, and it is difficult to access the nutritional status of patients. A larger sample size could enhance the knowledge gleaned about the efficacy of the intervention. In addition, our study captured only patients' self-reported objective physical outcomes such as weight and specific types of chemotherapy-related symptoms and did not assess patients' psychological perceptions of their nutritional status. In-depth qualitative interviews may be able to compensate for the bias in outcomes. Finally, the study was conducted in a public general tertiary care hospital, so it is necessary to test whether the good results observed can be generalized to other health care institutions.

This study also has some strengths. First, the study site was in a large general tertiary hospital that received patients from all over the country, so the study samples were representative of patients with OC undergoing chemotherapy with malnutrition conditions occurring in different regions of China. In addition, the nutritional assessment tool in the study has been internationally validated, thus allowing for replication and comparison with other studies. Another advantage was that the trial was conducted through a WeChat applet, which was easy to use and did not require downloading an additional application. In addition, the study included only those patients undergoing chemotherapy for OC, so the patient sample was more homogeneous than studies that included patients undergoing chemotherapy for gynecologic oncology.

Conclusions

In this study, a normative WeChat-based nutritional management program lasting for 6 chemotherapy intervals significantly improved nutritional status in patients with OC undergoing chemotherapy compared with conventional controls. Improving personalized nutritional intervention strategies in standard nutritional management protocols for patients with OC and introducing the WeChat platform as a mobile intervention model can enhance patients' nutritional status during chemotherapy and improve their quality of life.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Informed consent form for this study.
[DOC File, 55 KB - mhealth_v12i1e56475_app1.doc]

Multimedia Appendix 2
Nutritional program and nutrition-related and inflammation-related blood indices.
[ZIP File, 43 KB - mhealth_v12i1e56475_app2.zip]

Checklist 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File, 9431 KB - mhealth_v12i1e56475_app3.pdf]

References

1. Wang Z, Guo E, Yang B, et al. Trends and age-period-cohort effects on mortality of the three major gynecologic cancers in China from 1990 to 2019: cervical, ovarian and uterine cancer. Gynecol Oncol 2021 Nov;163(2):358-363. [doi: 10.1016/j.ygyno.2021.08.029] [Medline: 34507827]
2. Feng J, Xu L, Chen Y, Lin R, Li H, He H. Trends in incidence and mortality for ovarian cancer in China from 1990 to 2019 and its forecasted levels in 30 years. J Ovarian Res 2023 Jul 14;16(1):139. [doi: 10.1186/s13048-023-01233-y] [Medline: 37452315]
3. Kim A, Ueda Y, Naka T, Enomoto T. Therapeutic strategies in epithelial ovarian cancer. J Exp Clin Cancer Res 2012 Feb 13;31(1):14. [doi: 10.1186/1756-9966-31-14] [Medline: 22330607]
4. Elia M. Defining, recognizing, and reporting malnutrition. Int J Low Extrem Wounds 2017 Dec;16(4):230-237. [doi: 10.1177/1534734617733902] [Medline: 29145755]
5. Van Cutsem E, Arends J. The causes and consequences of cancer-associated malnutrition. Eur J Oncol Nurs 2005;9(Suppl 2):S51-S63. [doi: 10.1016/j.ejon.2005.09.007] [Medline: 16437758]
6. Hertlein L, Kirschenhofer A, Fürst S, et al. Malnutrition and clinical outcome in gynecologic patients. Eur J Obstet Gynecol Reprod Biol 2014 Mar;174:137-140. [doi: 10.1016/j.ejogrb.2013.12.028] [Medline: 24485666]
7. Rinninella E, Fagotti A, Cintoni M, et al. Nutritional interventions to improve clinical outcomes in ovarian cancer: a systematic review of randomized controlled trials. Nutrients 2019 Jun 21;11(6):1404. [doi: 10.3390/nu11061404] [Medline: 31234395]
8. Uccella S, Mele MC, Quagliozzi L, et al. Assessment of preoperative nutritional status using BIA-derived phase angle (PhA) in patients with advanced ovarian cancer: correlation with the extent of cytoreduction and complications. Gynecol Oncol 2018 May;149(2):263-269. [doi: 10.1016/j.ygyno.2018.03.044] [Medline: 29550182]
9. Laky B, Janda M, Cleghorn G, Obermair A. Comparison of different nutritional assessments and body-composition measurements in detecting malnutrition among gynecologic cancer patients. Am J Clin Nutr 2008 Jun;87(6):1678-1685. [doi: 10.1093/ajcn/87.6.1678] [Medline: 18541556]
10. Meijers JMM, Halfens RJG, Wilson L, Schols JMGA. Estimating the costs associated with malnutrition in Dutch nursing homes. Clin Nutr 2012 Feb;31(1):65-68. [doi: 10.1016/j.clnu.2011.08.009] [Medline: 21880401]
11. Maurer T, Belau MH, von Grundherr J, et al. Randomised controlled trial testing the feasibility of an exercise and nutrition intervention for patients with ovarian cancer during and after first-line chemotherapy (BENITA-study). BMJ Open 2022 Feb 23;12(2):e054091. [doi: 10.1136/bmjopen-2021-054091] [Medline: 35197344]
12. Stelten S, Hoedjes M, Kenter GG, et al. Rationale and study protocol of the Physical Activity and Dietary intervention in women with OVArian cancer (PADOVA) study: a randomised controlled trial to evaluate effectiveness of a tailored exercise and dietary intervention on body composition, physical function and fatigue in women with ovarian cancer undergoing chemotherapy. BMJ Open 2020 Nov 4;10(11):e036854. [doi: 10.1136/bmjopen-2020-036854] [Medline: 33148722]
13. Muscaritoli M, Arends J, Bachmann P, et al. ESPEN practical guideline: clinical nutrition in cancer. Clin Nutr 2021 May;40(5):2898-2913. [doi: 10.1016/j.clnu.2021.02.005] [Medline: 33946039]
14. Rock CL, Thomson CA, Sullivan KR, et al. American Cancer Society nutrition and physical activity guideline for cancer survivors. CA Cancer J Clin 2022 May;72(3):230-262. [doi: 10.3322/caac.21719] [Medline: 35294043]
15. Roila F, Molassiotis A, Herrstedt J, et al. 2016 MASCC and ESMO guideline update for the prevention of chemotherapy- and radiotherapy-induced nausea and vomiting and of nausea and vomiting in advanced cancer patients. Ann Oncol 2016 Sep;27(Suppl 5):v119-v133. [doi: 10.1093/annonc/mdw270] [Medline: 27664248]
16. Alderman B, Allan L, Amano K, et al. Multinational Association of Supportive Care in Cancer (MASCC) expert opinion/guidance on the use of clinically assisted nutrition in patients with advanced cancer. Support Care Cancer 2022 Apr;30(4):2983-2992. [doi: 10.1007/s00520-021-06613-y] [Medline: 34665311]
17. Xia J, Hu S, Xu J, Hao H, Yin C, Xu D. The correlation between glucose fluctuation from self-monitored blood glucose and the major adverse cardiac events in diabetic patients with acute coronary syndrome during a 6-month follow-up by WeChat application. Clin Chem Lab Med 2018 Nov 27;56(12):2119-2124. [doi: 10.1515/cclm-2018-0220] [Medline: 30016270]
18. He C, Wu S, Zhao Y, et al. Social media-promoted weight loss among an occupational population: cohort study using a WeChat mobile phone app-based campaign. J Med Internet Res 2017 Oct 23;19(10):e357. [doi: 10.2196/jmir.7861] [Medline: 29061555]
19. Lyu KX, Zhao J, Wang B, et al. Smartphone application WeChat for clinical follow-up of discharged patients with head and neck tumors: a randomized controlled trial. Chin Med J (Engl) 2016 Dec 5;129(23):2816-2823. [doi: 10.4103/0366-6999.194635] [Medline: 27900995]
20. Qiao S, Tang L, Zhang W, et al. Nurse-led follow-up to outpatients with cancer pain treated with opioids at home-telephone calls plus WeChat versus telephone calls only: a quasi-experimental study. Patient Prefer Adherence 2019;13:923-931. [doi: 10.2147/PPA.S203900] [Medline: 31239650]
21. Ni X, Lou Y, Hu W, et al. Development of mobile health-based self-management support for patients with lung cancer: a stepwise approach. Nurs Open 2022 May;9(3):1612-1624. [doi: 10.1002/nop2.1185] [Medline: 35128821]

22. Chen L, Yang X, Fu L, Liu X, Yuan C. Using the extended parallel process model to examine the nature and impact of breast cancer prevention information on mobile-based social media: content analysis. JMIR Mhealth Uhealth 2019 Jun 24;7(6):e13987. [doi: 10.2196/13987] [Medline: 31237239]
23. Wang J, Howell D, Shen N, et al. mHealth supportive care intervention for parents of children with acute lymphoblastic leukemia: quasi-experimental pre- and postdesign study. JMIR Mhealth Uhealth 2018 Nov 19;6(11):e195. [doi: 10.2196/mhealth.9981] [Medline: 30455166]
24. Orr B, Edwards RP. Diagnosis and treatment of ovarian cancer. Hematol Oncol Clin North Am 2018 Dec;32(6):943-964. [doi: 10.1016/j.hoc.2018.07.010] [Medline: 30390767]
25. Gallagher KM, Updegraff JA. Health message framing effects on attitudes, intentions, and behavior: a meta-analytic review. Ann Behav Med 2012 Feb;43(1):101-116. [doi: 10.1007/s12160-011-9308-7] [Medline: 21993844]
26. Bauer J, Capra S, Ferguson M. Use of the scored Patient-Generated Subjective Global Assessment (PG-SGA) as a nutrition assessment tool in patients with cancer. Eur J Clin Nutr 2002 Aug;56(8):779-785. [doi: 10.1038/sj.ejcn.1601412] [Medline: 12122555]
27. Jager-Wittenaar H, Ottery FD. Assessing nutritional status in cancer: role of the Patient-Generated Subjective Global Assessment. Curr Opin Clin Nutr Metab Care 2017 Sep;20(5):322-329. [doi: 10.1097/MCO.0000000000000389] [Medline: 28562490]
28. Isenring E, Bauer J, Capra S. The scored Patient-generated Subjective Global Assessment (PG-SGA) and its association with quality of life in ambulatory patients receiving radiotherapy. Eur J Clin Nutr 2003 Feb;57(2):305-309. [doi: 10.1038/sj.ejcn.1601552] [Medline: 12571664]
29. Marshall S, Young A, Bauer J, Isenring E. Malnutrition in geriatric rehabilitation: prevalence, patient outcomes, and criterion validity of the Scored Patient-Generated Subjective Global Assessment and the Mini Nutritional Assessment. J Acad Nutr Diet 2016 May;116(5):785-794. [doi: 10.1016/j.jand.2015.06.013] [Medline: 26212447]
30. Ottery FD. Rethinking nutritional support of the cancer patient: the new field of nutritional oncology. Semin Oncol 1994 Dec;21(6):770-778. [Medline: 7992092]
31. Kwang AY, Kandiah M. Objective and subjective nutritional assessment of patients with cancer in palliative care. Am J Hosp Palliat Care 2010 Mar;27(2):117-126. [doi: 10.1177/1049909109353900] [Medline: 19959839]
32. Li J, Zhu N, Wang C, et al. Preoperative albumin-to-globulin ratio and prognostic nutritional index predict the prognosis of colorectal cancer: a retrospective study. Sci Rep 2023;13(1):17272. [doi: 10.1038/s41598-023-43391-5]
33. Hu B, Yang XR, Xu Y, et al. Systemic immune-inflammation index predicts prognosis of patients after curative resection for hepatocellular carcinoma. Clin Cancer Res 2014 Dec 1;20(23):6212-6222. [doi: 10.1158/1078-0432.CCR-14-0442] [Medline: 25271081]
34. Zeng X, Huang X, Wang P, et al. The application of the PDCA cycle in the nutritional management of patients with nasopharyngeal carcinoma. Support Care Cancer 2023 Apr 10;31(5):251. [doi: 10.1007/s00520-023-07724-4] [Medline: 37036536]
35. Qin N, Jiang G, Zhang X, Sun D, Liu M. The effect of nutrition intervention with oral nutritional supplements on ovarian cancer patients undergoing chemotherapy. Front Nutr 2021;8:685967. [doi: 10.3389/fnut.2021.685967] [Medline: 34249995]
36. Um MH, Choi MY, Lee SM, Lee IJ, Lee CG, Park YK. Intensive nutritional counseling improves PG-SGA scores and nutritional symptoms during and after radiotherapy in Korean cancer patients. Support Care Cancer 2014 Nov;22(11):2997-3005. [doi: 10.1007/s00520-014-2304-2] [Medline: 24906838]
37. Miao J, Wang L, Ong EHW, et al. Effects of induction chemotherapy on nutrition status in locally advanced nasopharyngeal carcinoma: a multicentre prospective study. J Cachexia Sarcopenia Muscle 2023 Apr;14(2):815-825. [doi: 10.1002/jcsm.13196] [Medline: 36872457]
38. Huang J, Gu L, Zhang L, Lu X, Zhuang W, Yang Y. Symptom clusters in ovarian cancer patients with chemotherapy after surgery: a longitudinal survey. Cancer Nurs 2016;39(2):106-116. [doi: 10.1097/NCC.0000000000000252] [Medline: 25837811]
39. Kolb NA, Smith AG, Singleton JR, et al. The association of chemotherapy-induced peripheral neuropathy symptoms and the risk of falling. JAMA Neurol 2016 Jul 1;73(7):860-866. [doi: 10.1001/jamaneurol.2016.0383] [Medline: 27183099]
40. Fu L, Feng X, Jin Y, et al. Symptom clusters and quality of life in gastric cancer patients receiving chemotherapy. J Pain Symptom Manage 2022 Feb;63(2):230-243. [doi: 10.1016/j.jpainsymman.2021.09.003] [Medline: 34537311]
41. Wu X, Xu H, Zhang X, et al. Self-efficacy, hope as mediators between positive coping and resilience among patients with gastric cancer before the first chemotherapy. Cancer Nurs 2021;44(1):79-85. [doi: 10.1097/NCC.0000000000000753] [Medline: 31743154]
42. Wang X, Zeng H, Li L, et al. Personalized nutrition intervention improves nutritional status and quality of life of colorectal cancer survivors in the community: a randomized controlled trial. Nutrition 2022;103-104:111835. [doi: 10.1016/j.nut.2022.111835] [Medline: 36183482]
43. Keum J, Chung MJ, Kim Y, et al. Usefulness of smartphone apps for improving nutritional status of pancreatic cancer patients: randomized controlled trial. JMIR Mhealth Uhealth 2021 Aug 31;9(8):e21088. [doi: 10.2196/21088] [Medline: 34463630]

44. Xu M, Yang X, Liu L, Dai Y, Xu M, Lin S. Effect of the WeChat platform health management and refined continuous nursing model on life quality of patients with acute myocardial infarction after PCI. J Healthc Eng 2021;2021:5034269. [doi: 10.1155/2021/5034269] [Medline: 34880976]
45. Luo J, Dong X, Hu J. Effect of nursing intervention via a chatting tool on the rehabilitation of patients after total hip arthroplasty. J Orthop Surg Res 2019 Dec 9;14(1):417. [doi: 10.1186/s13018-019-1483-4] [Medline: 31818330]
46. Duan Y, Li S, Su Q, Xu S, Lu G. Influence of exercise prescription intervention based on WeChat on glycolipid metabolism and fitness of suboptimal-health teachers. Medicine (Baltimore) 2024 May 24;103(21):e38167. [doi: 10.1097/MD.0000000000038167] [Medline: 38788028]
47. Yang B, Liu JF, Xie WP, Cao H, Chen Q. The effects of WeChat follow-up management to improve the parents' mental status and the quality of life of premature newborns with patent ductus arteriosus. J Cardiothorac Surg 2021 Aug 21;16(1):235. [doi: 10.1186/s13019-021-01617-5] [Medline: 34419135]

Abbreviations

OC: ovarian cancer
PG-SGA: Patient-Generated Subjective Global Assessment
PNI: prognostic nutritional index
SII: systemic immunoinflammatory index

Edited by L Buis; submitted 17.01.24; peer-reviewed by A Hassan, J Liu, Y Wang; revised version received 15.06.24; accepted 26.08.24; published 04.11.24.

Please cite as:
Tian X, Liu Y, Zhang J, Yang L, Feng L, Qi A, Liu H, Liu P, Li Y
Efficacy of a WeChat-Based, Multidisciplinary, Full-Course Nutritional Management Program on the Nutritional Status of Patients With Ovarian Cancer Undergoing Chemotherapy: Randomized Controlled Trial
JMIR Mhealth Uhealth 2024;12:e56475
URL: https://mhealth.jmir.org/2024/1/e56475
doi:10.2196/56475

Engagement in mHealth-Prompted Self-Measured Blood Pressure Monitoring Among Participants Recruited From a Safety-Net Emergency Department: Secondary Analysis of the Reach Out Trial

Lesli E Skolarus[1], MS, MD; Chun Chieh Lin[2], MBA, PhD; Sonali Mishra[3], PhD; William Meurer[4,5], MS, MD; Mackenzie Dinh[4], MS; Candace Whitfield[4], BS; Ran Bi[2], MS; Devin Brown[5], MS, MD; Rockefeller Oteng[4], MD; Lorraine R Buis[6], PhD; Kelley Kidwell[7], PhD

[1]Department of Neurology, Stroke and Vascular Neurology, Northwestern University, Chicago, IL, United States

[2]Department of Neurology, The Ohio State University, Columbus, OH, United States

[3]Department of Internal Medicine-Cardiology, University of Michigan, Ann Arbor, MI, United States

[4]Department of Emergency Medicine, University of Michigan, Ann Arbor, MI, United States

[5]Department of Neurology, University of Michigan, Ann Arbor, MI, United States

[6]Department of Family Medicine, University of Michigan, Ann Arbor, MI, United States

[7]Department of Biostatistics, University of Michigan, Ann Arbor, MI, United States

Corresponding Author:
Lesli E Skolarus, MS, MD
Department of Neurology, Stroke and Vascular Neurology, Northwestern University, , Chicago, IL, , United States

Abstract

Background: Hypertension, a key modifiable risk factor for cardiovascular disease, is more prevalent among Black and low-income individuals. To address this health disparity, leveraging safety-net emergency departments for scalable mobile health (mHealth) interventions, specifically using text messaging for self-measured blood pressure (SMBP) monitoring, presents a promising strategy. This study investigates patterns of engagement, associated factors, and the impact of engagement on lowering blood pressure (BP) in an underserved population.

Objective: We aimed to identify patterns of engagement with prompted SMBP monitoring with feedback, factors associated with engagement, and the association of engagement with lowered BP.

Methods: This is a secondary analysis of data from Reach Out, an mHealth, factorial trial among 488 hypertensive patients recruited from a safety-net emergency department in Flint, Michigan. Reach Out participants were randomized to weekly or daily text message prompts to measure their BP and text in their responses. Engagement was defined as a BP response to the prompt. The k-means clustering algorithm and visualization were used to determine the pattern of SMBP engagement by SMBP prompt frequency—weekly or daily. BP was remotely measured at 12 months. For each prompt frequency group, logistic regression models were used to assess the univariate association of demographics, access to care, and comorbidities with high engagement. We then used linear mixed-effects models to explore the association between engagement and systolic BP at 12 months, estimated using average marginal effects.

Results: For both SMBP prompt groups, the optimal number of engagement clusters was 2, which we defined as high and low engagement. Of the 241 weekly participants, 189 (78.4%) were low (response rate: mean 20%, SD 23.4) engagers, and 52 (21.6%) were high (response rate: mean 86%, SD 14.7) engagers. Of the 247 daily participants, 221 (89.5%) were low engagers (response rate: mean 9%, SD 12.2), and 26 (10.5%) were high (response rate: mean 67%, SD 8.7) engagers. Among weekly participants, those who were older (>65 years of age), attended some college (vs no college), married or lived with someone, had Medicare (vs Medicaid), were under the care of a primary care doctor, and took antihypertensive medication in the last 6 months had higher odds of high engagement. Participants who lacked transportation to appointments had lower odds of high engagement. In both prompt frequency groups, participants who were high engagers had a greater decline in BP compared to low engagers.

Conclusions: Participants randomized to weekly SMBP monitoring prompts responded more frequently overall and were more likely to be classed as high engagers compared to participants who received daily prompts. High engagement was associated with a larger decrease in BP. New strategies to encourage engagement are needed for participants with lower access to care.

Trial Registration: ClinicalTrials.gov NCT03422718; https://clinicaltrials.gov/study/NCT03422718

International Registered Report Identifier (IRRID): RR2-10.1186/s13063-020-04340-z

(JMIR Mhealth Uhealth 2024;12:e54946) doi:10.2196/54946

KEYWORDS

hypertension; self-measured blood pressure; mobile health; blood pressure; emergency; blood pressure monitoring; risk factor; cardiovascular; cardiovascular disease; utilization; feedback; care; systolic blood pressure; emergency department; mHealth; health disparities; engagement

Introduction

Hypertension is the most important modifiable risk factor for cardiovascular disease and is more prevalent among Black and low-income people [1-3]. Given the ubiquity and inequities of hypertension, scalable approaches are needed to identify and treat Americans with hypertension.

One scalable approach may be by partnering with safety-net emergency departments (EDs)—hospitals where over 25% of patients are Medicaid recipients—to reach a medically underserved population to initiate mobile health (mHealth) interventions. There are over 136 million ED visits annually [4], and Black Americans and those with low incomes frequently use the ED [5].

mHealth is readily scalable as nearly all American adults (96%) have a mobile phone and over 80% use it for text messaging [6]. Text messaging is a leading form of communication partly because of its cost, ease, and low technical threshold. mHealth hypertension interventions have primarily targeted medication adherence, self-management (including diet and physical activity), and self-measured blood pressure (SMBP) monitoring. SMBP monitoring involves the regular measurement of BP by a patient outside the clinical setting. The outcomes of these mHealth hypertension interventions have had mixed results [7-10]. SMBP reduces BP, in part by promoting the habit of medication adherence and a healthy lifestyle, but little is known about SMBP in the Black and low-income population [11-15]. Additionally, more work is needed to understand engagement with mHealth text messaging interventions and how engagement relates to clinical outcomes.

"Engagement" has been conceptualized differently across fields and studies, but generally refers to users' use of and experience with the system [16]. Although engagement with an intervention is distinct from the performance of health behavior, as users may abandon the use of an intervention while continuing to perform the target behavior [17,18], engagement is associated with intervention efficacy in multiple health domains [19]. Engagement with an intervention may describe interaction with either push- or pull-based interventions [20]. The former describes a user responding to an intervention, for instance, by responding to a text message; the second describes the user calling on the intervention in response to their own needs, for instance, by opening a dashboard to review their past BP measurements. Although push interventions have the potential to increase users' performance of critical health behaviors, such as SMBP, they run the risk of burdening or annoying users, leading to disengagement [20]. Despite work to categorize and evaluate the efficacy of various engagement strategies for mHealth interventions [21], dropout rates remain high [22]. More work is needed to understand how users interact with mHealth interventions, their target health behaviors, and what factors contribute to engagement or disengagement with the intervention.

In this study, we examine engagement with a push intervention in which participants were asked to text SMBP measurements to the research team in response to a text prompt. We identify patterns of engagement with SMBP monitoring, factors associated with engagement, and the association of engagement with lowered BP.

Methods

Ethical Considerations

This study was approved by the University of Michigan Institutional Review Board (HUM00138470) and the ED site Institutional Review Board (1199877). All participants provided written informed consent. The original informed consent allowed for deidentified use in additional analyses and research studies. Participants of the Reach Out trial were given an automated BP cuff, US $20 at enrollment, US $25 after the completion of a 6-month follow-up visit, and US $30 after the completion of a 12-month follow-up visit. If needed, transportation was provided to follow-up visit(s).

Design

This is a nonprespecified secondary analysis of data from the Reach Out Trial. Reach Out was a randomized, controlled, 2×2×2 factorial design mHealth clinical trial to reduce BP among hypertensive safety-net ED patients seeking care for conditions likely to be discharged from the ED (identifier NCT03422718 on ClinicalTrials.gov) [23]. Enrollment occurred from March 2019 to March 2020. Participants were randomized to (1) prompted SMBP monitoring (daily vs weekly) with feedback, (2) tailored healthy behavior text messaging (daily vs none), and (3) facilitated primary care provider (PCP) scheduling and transportation (yes vs no) for 12 months. There were minimal differences across the intervention arms [24]. Among participants, BP declined over the 12-month intervention period. There was no difference in change in systolic BP among the 3 mHealth components [24].

Prompted SMBP Monitoring With Feedback

Participants were randomized to receive daily or weekly automated text prompts to take their BP and text the results to the study team. All participants were given a BP cuff at the time of enrollment. Sample prompts included the following: "This is your reminder to take your BP. REPLY with your BP to Reach Out!" Participants randomized to weekly prompted SMBP monitoring were sent up to 2 follow-up reminders over a 24-hour period if they remained unresponsive. Participants randomized to daily prompted SMBP monitoring were not sent follow-up reminders. All participants received an automated confirmation

text for each BP texted to the study team. Each week, all participants received a tailored feedback message comparing the participant's recent BP to goal BP, along with general encouragement. For example, if a participant reported an SMBP with systolic BP higher than the threshold but diastolic BP lower than the threshold, the following text was sent: "Your BP is 150/75. Your top number is above the normal range, but your bottom number is normal. Meds, eating healthy, and exercise can lower that top number!" Participants also received a monthly text that contained a graph of their self-reported BPs, with tailored interpretation. In this context, engagement was defined as a BP response to an SMBP prompt.

Covariates and BP

Covariates were chosen based on prior research suggesting an association with engagement in SMBP monitoring [13]. Age, race (self-reported Black vs non-Black), sex, education (no college vs any college), insurance type (eg, Medicaid, Medicare, private, uninsured, multiple, or other types of insurance), relationship status (living with someone or married vs living alone), and employment status were self-reported. We also queried access to hypertension care, including the presence of a PCP [25], diagnosis of hypertension, prior hypertension medication in the last 6 months, and inability to attend medical appointments due to lack of transportation (yes vs no). Finally, we queried medical comorbidities (eg, stroke, congestive heart failure, myocardial infarction, and kidney disease) [25].

Baseline BP measurements were the mean of the median of the remote systolic BPs received during the 3-week eligibility phase. All BP outcomes were intended to be measured in person. However, outcomes were assessed remotely for some participants at 6 months and for all participants at 12 months due to COVID-19 research restrictions. In-person BP assessments were conducted by trained research team members using an OMRON 7 Series Upper Arm BP Monitor and following standard procedures [26]. If performed remotely, participants were asked to provide 3 BP measurements, each 1 minute apart. Participants communicated these to the study team through phone calls or text messages. Participants were also asked to send a picture of their BP cuff on their arms to confirm the correct orientation.

Analysis

We used the k-means clustering algorithm and visualization to determine the pattern of SMBP monitoring engagement by SMBP prompt frequency. For both prompted SMBP frequency groups (weekly and daily), the optimal number of clusters of engagement via the elbow method was 2, which we defined as high and low engagement (Multimedia Appendices 1-4) [27]. For each daily and weekly prompted SMBP frequency group, we used descriptive statistics to describe engagement type and assessed the univariate association of demographics, access to care, and comorbidities with high engagement using logistic regression models. Finally, we sought to determine whether engagement was associated with a difference in systolic BP at 12 months using a linear mixed-effects model with 12-month systolic BP as the outcome variable and baseline BP, engagement type (high vs low), time measured in days from randomization to 1 year, and the interaction of time and engagement type as fixed effects, with a random participant effect. Systolic BP was estimated using average marginal effects from the fully adjusted model. K-mode clustering was performed using Python (version 3.11.2; Python Software Foundation), and all other analyses were conducted using SAS (version 9.4; SAS Institute).

Results

Participants

A total of 488 participants were randomized into the intervention; 241 (49.4%) were randomized to weekly monitoring and 247 (50.6%) were randomized to daily monitoring. Within this safety-net ED, of the 241 randomized participants, 117 (48.6%) used Medicaid. Demographics, access to care, and comorbidities for each prompted SMBP frequency group are included in Tables 1 and 2.

Table . Association of engagement among participants who received weekly blood pressure prompts.

Characteristics		Weekly total, n (%)	Low engager, n (%)	High engager, n (%)	Univariate analysis in predicting high engagers	
					Odds ratio (95% CI)	*P* value
Total		241 (100)	189 (78.4)	52 (21.6)	—[a]	—
Demographics						
	Older than 65 years	19 (7.9)	8 (4.2)	11 (21.2)	6.1 (2.3 - 16.0)	<.001
	Women	148 (61.4)	117 (61.9)	31 (59.6)	0.9 (0.5 - 1.7)	.76
	Non-Hispanic Black people	132 (54.8)	105 (55.6)	27 (51.9)	0.9 (0.5 - 1.6)	.64
	Married or living with someone	60 (24.9)	40 (21.2)	20 (38.5)	2.3 (1.2 - 4.5)	.01
	Not employed	123 (51.0)	94 (49.7)	29 (55.8)	1.3 (0.7 - 2.4)	.44
Education						
	Some high school education, high school graduate, or trade school	118 (49.0)	102 (54.0)	16 (30.8)	1	—
	Some college education or college graduate	123 (54.8)	87 (46.0)	36 (69.2)	2.6 (1.4 - 5.1)	.004
Access to care						
	Under the care of a primary care doctor	195 (80.9)	146 (77.3)	49 (94.2)	4.8 (1.4 - 16.2)	.01
	Previous diagnosis of hypertension	190 (78.8)	146 (77.3)	44 (84.6)	1.6 (0.7 - 3.7)	.25
	Prior medication for hypertension in the last 6 months	136 (56.4)	96 (50.8)	40 (76.9)	3.2 (1.6 - 6.5)	.001
	Lack of transportation for medical appointments	49 (20.3)	45 (23.8)	4 (7.7)	0.3 (0.1 - 0.8)	.02
Health insurance						
	Medicaid	117 (48.6)	101 (53.4)	16 (30.8)	1	—
	Private	50 (20.8)	38 (20.1)	12 (23.1)	2.0 (0.9 - 4.6)	.11
	Medicare	22 (9.1)	10 (5.3)	12 (23.1)	7.6 (2.8 - 20.4)	<.001
	Other insurance	5 (2.1)	4 (2.1)	1 (1.9)	1.6 (0.2 - 15.0)	.69
	Uninsured	19 (7.9)	15 (7.9)	4 (7.7)	1.7 (0.5 - 5.7)	.40
	Multiple insurances	28 (11.6)	21 (11.1)	7 (13.5)	2.1 (0.8 - 5.8)	.15
Comorbidities						
	Stroke or transient ischemic attack	17 (7.1)	13 (6.9)	4 (7.7)	1.1 (0.4 - 3.6)	.84
	Congestive heart failure	12 (5.0)	7 (3.7)	5 (9.6)	2.8 (0.8 - 9.1)	.09
	Myocardial infarction	13 (5.4)	8 (4.2)	5 (9.6)	2.4 (0.8 - 7.7)	.14
	Kidney disease	16 (6.6)	10 (5.3)	6 (11.5)	2.3 (0.8 - 6.8)	.12
Intervention components						
	Healthy behavior texts	121 (50.2)	100 (52.9)	21 (40.4)	0.6 (0.3 - 1.1)	.11

Characteristics		Weekly total, n (%)	Low engager, n (%)	High engager, n (%)	Univariate analysis in predicting high engagers	
					Odds ratio (95% CI)	*P* value
	Primary care provider–facilitated scheduling and transportation	119 (49.4)	94 (49.7)	25 (48.1)	0.9 (0.5 - 1.7)	.83

[a]Not applicable.

Table . Association of engagement among participants who received daily blood pressure prompts.

Characteristics		Daily total, n (%)	Low engager, n (%)	High engager, n (%)	Univariate analysis in predicting high engagers	
					Odds ratio (95% CI)	*P* value
Total		247 (100)	221 (89.5)	26 (10.5)	—[a]	—
Demographics						
	Older than 65 years	18 (7.3)	15 (6.8)	3 (11.5)	1.8 (0.5 - 6.7)	.38
	Women	151 (61.1)	132 (59.7)	19 (73.1)	1.8 (0.7 - 4.5)	.19
	Non-Hispanic Black people	130 (52.6)	120 (54.3)	10 (38.5)	0.5 (0.2 - 1.2)	.13
	Married or living with someone	65 (26.3)	54 (24.4)	11 (42.3)	2.3 (1.0 - 5.2)	.06
	Not employed	119 (48.2)	105 (47.5)	14 (53.9)	1.3 (0.6 - 2.9)	.54
Education						
	Some high school education, high school graduate, or trade school	122 (49.4)	113 (51.1)	9 (34.6)	1	—
	Some college education or college graduate	125 (50.6)	108 (48.9)	17 (65.4)	2.0 (0.9 - 4.6)	.12
Access to care						
	Under the care of a primary care doctor	185 (74.9)	163 (73.8)	22 (84.6)	2.0 (0.7 - 5.9)	.23
	Previous diagnosis of hypertension	195 (79.0)	173 (78.3)	22 (84.6)	1.5 (0.5 - 4.6)	.46
	Prior medication for hypertension in the last 6 months	137 (55.5)	119 (53.9)	18 (69.2)	1.9 (0.8 - 4.6)	.14
	Lack of transportation for medical appointments	52 (21.1)	49 (22.2)	3 (11.5)	0.5 (0.1 - 1.6)	.22
Health insurance						
	Medicaid	127 (51.4)	119 (53.9)	8 (30.8)	1	—
	Private	48 (19.4)	41 (18.6)	7 (26.9)	2.5 (0.9 - 7.4)	.09
	Medicare	23 (9.3)	21 (9.5)	2 (7.7)	1.4 (0.3 - 7.1)	.67
	Other insurance	3 (1.2)	3 (1.4)	0 (0)	0.0 (<0->1000)	.99
	Uninsured	29 (11.8)	25 (11.3)	4 (15.4)	2.4 (0.7 - 8.5)	.18
	Multiple insurances	17 (6.9)	12 (5.4)	5 (19.2)	6.2 (1.8 - 22.0)	.005
Comorbidities						
	Stroke or transient ischemic attack	19 (7.7)	16 (7.2)	3 (11.5)	1.7 (0.5 - 6.2)	.40
	Congestive heart failure	11 (4.5)	11 (5.0)	0 (0)	<0.01 (<0.01->999.99)	.97
	Myocardial infarction	18 (7.3)	15 (6.8)	3 (11.5)	1.8 (0.5 - 6.7)	.38
	Kidney disease	9 (3.6)	9 (4.1)	0 (0)	<0.01 (<0.01->999.99)	.98
Intervention components						

Characteristics		Daily total, n (%)	Low engager, n (%)	High engager, n (%)	Univariate analysis in predicting high engagers	
					Odds ratio (95% CI)	*P* value
	Healthy behavior texts	120 (48.6)	108 (48.9)	12 (46.2)	0.9 (0.4 - 2.0)	.79
	Primary care provider–facilitated scheduling and transportation	123 (49.8)	109 (49.3)	14 (53.9)	1.2 (0.5 - 2.7)	.66

[a]Not applicable.

Weekly SMBP

Of the 241 participants randomized to weekly SMBP prompts, 189 (78.4%) were classed as low engagers, and 52 (21.6%) were classed as high engagers (Table 1). Low engagers had a mean response rate of 20% (SD 23.4), while high engagers had a mean response rate of 86% (SD 8.7). Participants who were low engagers decreased their engagement early in the trial (Multimedia Appendix 3). Participants older than 65 years (odds ratio [OR] 6.1, 95% CI 2.3 - 16.0; *P*<.001), married or living with someone compared to living alone (OR 2.3, 95% CI 1.2 - 4.5; *P*=.01), and those who attended some college (OR 2.6, 95% CI 1.4 - 5.1; *P*=.004) compared to those without college attendance were more likely to be high engagers. Participants with more access to care, including participants with Medicare (OR 7.6, 95% CI 2.8 - 20.4; *P*<.001) compared to participants with Medicaid, those with the care of a primary care doctor (OR 4.8, 95% CI 1.4 - 16.2; *P*=.01), and those taking antihypertensive medication in the last 6 months (OR 3.2, 95% CI 1.6 - 6.5; *P*=.001) had higher odds of high engagement. Participants who lacked transportation to appointments (OR 0.3, 95% CI 0.1 - 0.8; *P*=.02) had lower odds of high engagement. There was no association between engagement with PCP facilitation and healthy behavior text messaging. Participants who were high engagers had a greater decline in BP (−8.1 mm Hg, 95% CI −12.6 to −3.6), compared to low engagers (−2.4 mm Hg, 95% CI −6.6 to 1.9), and lower BP at 12 months (high engagers: 128.9 mm Hg, 95% CI 125.7 - 132.1 vs low engagers: 136.8 mm Hg, 95% CI 133.2 - 140.4; Figure 1 and Table 3).

Figure 1. Estimated blood pressure (BP) and change in BP by engagement and time among participants who received BP prompts. SBP: systolic blood pressure.

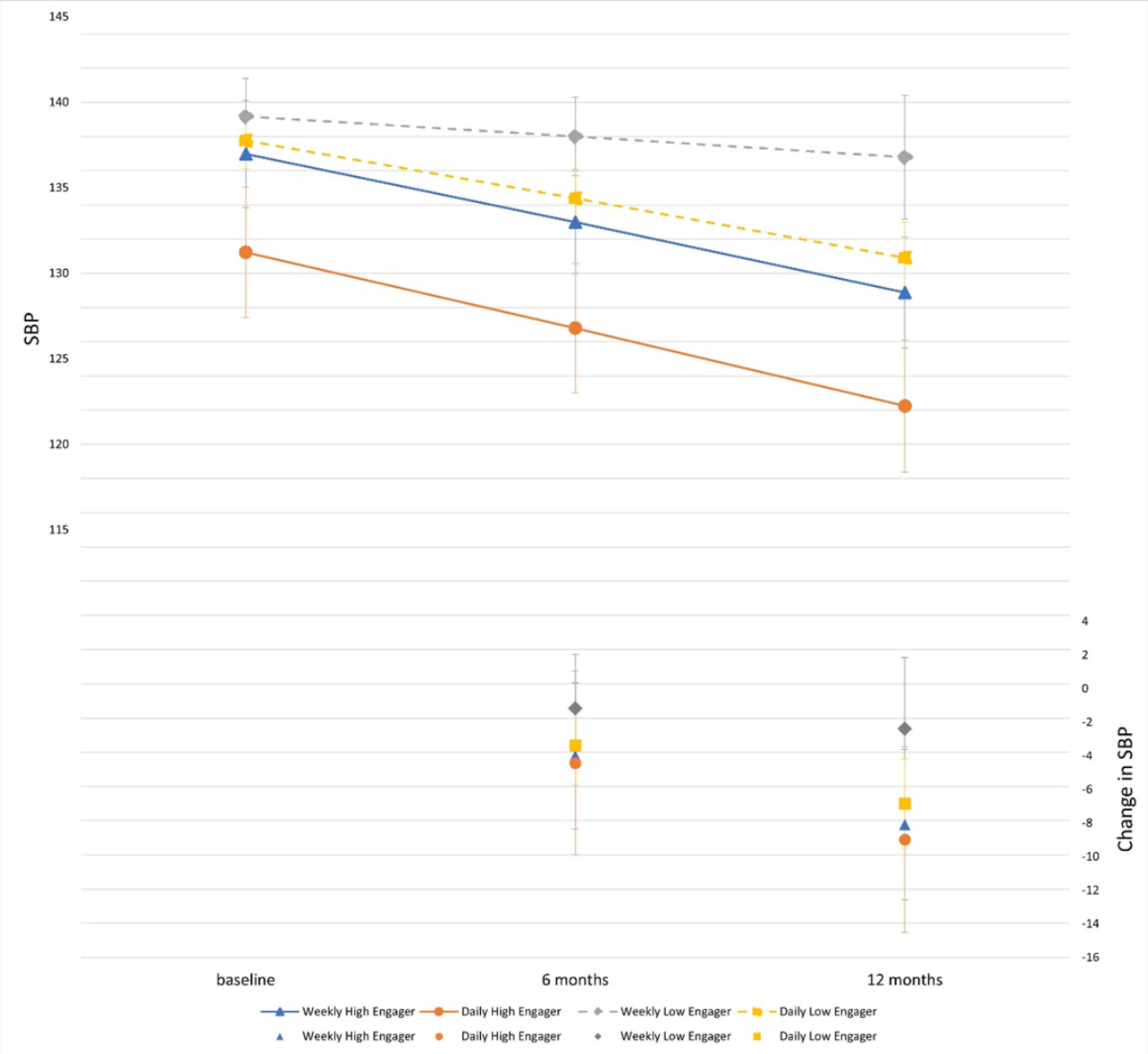


Table . Estimated blood pressure by engagement and time among participants who received blood pressure prompts.

SBP[a] prompts	High engager					Low engager				
	BL[b] SBP (mm Hg; 95% CI)	6-month SBP (mm Hg; 95% CI)	Change from BL to 6-month SBP (95% CI)	12-month SBP (mm Hg; 95% CI)	Change from BL to 12-month SBP (95% CI)	BL SBP (mm Hg; 95% CI)	6-month SBP (mm Hg; 95% CI)	Change from BL to 6-month SBP (95% CI)	12-month SBP (mm Hg; 95% CI)	Change from BL to 12-month SBP (95% CI)
Weekly	137.0 (133.8 to 140.1)	133.0 (130.0 to 136.0)	−4.0 (−8.3 to 0.4)	128.9 (125.7 to 132.1)	−8.1 (−12.6 to −3.6)	139.2 (136.9 to 141.4)	138.0 (135.7 to 140.3)	−1.2 (−4.4 to 2.02)	136.8 (133.2 to 140.4)	−2.4 (−6.6 to 1.9)
Daily	131.2 (127.4 to 135.1)	126.8 (123.0 to 130.6)	−4.4 (−9.9 to 1.04)	122.2 (118.4 to 126.1)	−9.0 (−14.5 to −3.5)	137.8 (136.1 to 139.4)	134.4 (132.7 to 136.1)	−3.4 (−5.7 to −1.03)	130.9 (128.8 to 133.0)	−6.8 (−9.5 to −4.2)

[a]SBP: systolic blood pressure.

[b]BL: baseline.

Daily SMBP

Of the 247 participants randomized to daily SMBP prompts, 221 (89.5%) were classed as low engagers and 26 (10.5%) as high engagers (Table 2). Low engagers had a response rate of 9% (SD 12.2), while high engagers had a mean response rate of 67% (SD 8.7). Participants who were low engagers decreased their engagement early in the trial (Multimedia Appendix 4). Participants who were married or living with someone had a trend toward high engagement (OR 2.3 95% CI 1.0 - 5.2; *P*=.06), and participants with multiple insurance types were more likely to have high engagement (OR 6.2, 95% CI 1.8 - 22.0; *P*=.005) compared to Medicaid alone. Participants who were high engagers had a greater decline in BP (−9.0 mm Hg, 95% CI −14.5 to −3.5) compared to low engagers (−6.8 mm Hg, 95% CI −9.5 to −4.2) and lower BP at 12 months (high engagers: 122.2 mm Hg, 95% CI 118.4 - 12.1 vs low engagers 130.9 mm Hg, 95% CI 128.8 - 133.0; Figure 1 and Table 3).

Discussion

In this secondary analysis of a mHealth clinical trial among participants recruited from a safety-net ED, we identified 2 distinct patterns of SMBP engagement.

High and Low Engagement: Interpretation and Implications

In both daily and weekly reminder conditions, participants could be clustered into high and low engagers. A larger proportion of participants randomized to weekly SMBP monitoring prompts were high engagers and had greater engagement overall compared to participants who received daily prompts. All engagement groups had a decline in BP, even low engagers in the daily condition who exhibited a very low response rate.

In both daily and weekly conditions, many of the factors associated with high engagement were factors also associated with better health outcomes in general: high engagers generally had attended college, had treated their hypertension with an antihypertensive medication in the last 6 months, had Medicare, and had more access to care (ie, they had reported a PCP at enrollment or reported having transportation to medical appointments). Perhaps due to these factors, high engagers had slightly lower baseline systolic BP than low engagers in both conditions (Table 3). High engagers also exhibited a greater decrease in BP than low engagers. This may be related to their baseline characteristics, which could promote a habit of medication adherence and a practice of a healthy lifestyle (through SMBP) [11-15]. Another possible explanation for this finding, which would require further study, is that high engagers learned more about what impacted their BP from frequent monitoring. In this case, researchers should explore strategies for promoting engagement to turn low engagers into high engagers.

We found that engagement levels declined quickly and did not recover, suggesting re-engagement attempts should begin early on. Re-engagement strategies could take the form of direct outreach from the study team. However, for the sake of scalability, researchers should also explore techniques for re-engaging participants through the intervention itself. Although there is not yet a good evidence base for re-engagement strategies for mHealth interventions, existing research suggests that when users re-engage, it is for the same health motivations that brought them to the system in the first place [28]. Work examining the adoption and re-adoption of wearables has also found that the visual appeal of new devices is also associated with re-engagement [29]. These findings suggest that reminding users of their health goals and prompting them to monitor their BP could be an effective re-engagement technique. Although researchers have little control over the general visual appeal of text messages, future interventions could explore whether varying the form of reminders to include more graphical elements, for instance, by sending images as text messages, could promote engagement or re-engagement.

Another possible explanation of our findings is that high engagers were more likely to engage in SMBP because they were better able to act on their BP. For instance, high engagers may have been able to talk to their PCP about their recorded measurements [30]. By contrast, low engagers, who overall had less access to care, may have been less able to consult clinicians about what additional management was needed, leading to discouragement and disengagement. This interpretation suggests that SMBP alone is unable to mitigate the impact of decreased access to care. Additional strategies, such as connecting participants to primary care—which was initiated in the Reach Out trial but ended due to COVID-19, the addition of community-health worker support, or activating social networks may be needed.

Prompt Frequency, Engagement, and BP

Participants in the weekly condition exhibited higher engagement, and higher engagement was associated with a greater decrease in BP. The lower engagement observed in the daily condition could be due to habituation, as observed in other mHealth contexts [17,31]—users may simply have gotten used to the notifications and begun ignoring them. Lower engagement could signal frustration or dissatisfaction with the intervention, impacting real-world use. Other studies had similar findings—higher engagement was linked to improved outcomes [32,33]. More work is needed to evaluate the reasons for higher engagement. To get better insight into the effect of each message, future work could analyze not just the response rate but also the rate at which messages are read and the rate at which particular behaviors associated with hypertension management, such as medication taking, are performed. Such granular data could enable researchers to get more insight into how different doses of the intervention function. Future works could also focus on the efficacy of prompt frequency alongside habituation to better inform how to optimize engagement.

Limitations and Conclusions

Our study has limitations. This was a nonprespecified secondary analysis of clinical trial data. Reach Out was a single-center trial; thus, results may not be generalizable to other safety-net ED populations, particularly those with non–English-speaking populations. Finally, we cannot exclude that a small subset of participants completed SMBP monitoring but did not text in their BPs, and as a result, we may be overestimating the

proportion of low engagers. If high engagers were misclassified as low engagers, this should bias our findings toward the null.

In conclusion, about 16% of safety-net ED participants were highly engaged in prompted SMBP monitoring with feedback. Participants in the weekly condition overall exhibited higher engagement compared to daily prompts. BP decreased among all participants, but those with higher engagement had a greater decline in BP overall, supporting weekly rather than daily BP prompts. New strategies to encourage engagement are needed for participants who were not taking antihypertensive medication and had lower access to care.

Acknowledgments

This study was funded by the National Institutes of Health, National Institutes of Minority Health and Health Disparities (R01 MD011516).

Data Availability

The data supporting this study's findings are available from the corresponding author upon reasonable request.

Authors' Contributions

LES, DB, WM, MD, KK, and LRB contributed to the study conception and design. Data collection was carried out by MD and CW. Analysis and interpretation of results were conducted by KK, CCL, LES, WM, DB, and RB. LES and CCL prepared the draft for the manuscript. RO, SM, DB, WM, MD, KK, LRB, CW, and RB were in charge of the manuscript's critical revision. All authors reviewed the results and the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Weekly self-measured blood pressure engagement of optimal K.
[PNG File, 15 KB - mhealth_v12i1e54946_app1.png]

Multimedia Appendix 2
Daily self-measured blood pressure engagement of optimal K.
[PNG File, 14 KB - mhealth_v12i1e54946_app2.png]

Multimedia Appendix 3
Frequency of self-measured blood pressure (SMBP) response to weekly SMBP prompts clustered by engagement.
[PNG File, 58 KB - mhealth_v12i1e54946_app3.png]

Multimedia Appendix 4
Frequency of self-measured blood pressure (SMBP) response to daily SMBP prompts clustered by engagement.
[PNG File, 137 KB - mhealth_v12i1e54946_app4.png]

References

1. Turnbull F. Effects of different blood-pressure-lowering regimens on major cardiovascular events: results of prospectively-designed overviews of randomised trials. Lancet 2003 Nov 8;362(9395):1527-1535. [doi: 10.1016/s0140-6736(03)14739-3] [Medline: 14615107]
2. Cheng S, Claggett B, Correia AW, et al. Temporal trends in the population attributable risk for cardiovascular disease: the Atherosclerosis Risk in Communities Study. Circulation 2014 Sep 2;130(10):820-828. [doi: 10.1161/CIRCULATIONAHA.113.008506] [Medline: 25210095]
3. SPRINT Research Group, Wright JT, Williamson JD, et al. A randomized trial of intensive versus standard blood-pressure control. N Engl J Med 2015 Nov 26;373(22):2103-2116. [doi: 10.1056/NEJMoa1511939] [Medline: 26551272]
4. Weiss AJ, Jiang HJ. Most frequent reasons for emergency department visits, 2018. : Agency for Healthcare Research and Quality; 2021 URL: https://www.hcup-us.ahrq.gov/reports/statbriefs/sb286-ED-Frequent-Conditions-2018.pdf [accessed 2024-05-08]
5. Gindi RM, Cohen RA, Kirzinger WK. Emergency Room Use Among Adults Aged 18-64: Early Release of Estimates From the National Health Interview Survey, January-June 2011: National Center for Health Statistics; 2012.
6. Cell phone activities 2013. Pew Research Center. 2013. URL: http://www.pewinternet.org/2013/09/19/cell-phone-activities-2013/ [accessed 2024-05-08]

7. Mehta SJ, Volpp KG, Troxel AB, et al. Electronic pill bottles or bidirectional text messaging to improve hypertension medication adherence (way 2 text): a randomized clinical trial. J Gen Intern Med 2019 Nov;34(11):2397-2404. [doi: 10.1007/s11606-019-05241-x] [Medline: 31396815]
8. Buis L, Hirzel L, Dawood RM, et al. Text messaging to improve hypertension medication adherence in African Americans from primary care and emergency department settings: results from two randomized feasibility studies. JMIR mHealth uHealth 2017 Feb 1;5(2):e9. [doi: 10.2196/mhealth.6630] [Medline: 28148474]
9. Li R, Liang N, Bu F, Hesketh T. The effectiveness of self-management of hypertension in adults using mobile health: systematic review and meta-analysis. JMIR mHealth uHealth 2020 Mar 27;8(3):e17776. [doi: 10.2196/17776] [Medline: 32217503]
10. Tam HL, Wong EML, Cheung K, Chung SF. Effectiveness of text messaging interventions on blood pressure control among patients with hypertension: systematic review of randomized controlled trials. JMIR mHealth uHealth 2021 Sep 22;9(9):e24527. [doi: 10.2196/24527] [Medline: 34550078]
11. Ostchega Y, Zhang G, Kit BK, Nwankwo T. Factors associated with home blood pressure monitoring among US adults: National Health and Nutrition Examination Survey, 2011-2014. Am J Hypertens 2017 Nov 1;30(11):1126-1132. [doi: 10.1093/ajh/hpx101] [Medline: 28633432]
12. Doane J, Buu J, Penrod MJ, Bischoff M, Conroy MB, Stults B. Measuring and managing blood pressure in a primary care setting: a pragmatic implementation study. J Am Board Fam Med 2018;31(3):375-388. [doi: 10.3122/jabfm.2018.03.170450] [Medline: 29743221]
13. Ostchega Y, Berman L, Hughes JP, Chen TC, Chiappa MM. Home blood pressure monitoring and hypertension status among US adults: the National Health and Nutrition Examination Survey (NHANES), 2009-2010. Am J Hypertens 2013 Sep;26(9):1086-1092. [doi: 10.1093/ajh/hpt054] [Medline: 23604493]
14. Conn VS, Ruppar TM, Chase JA, Enriquez M, Cooper PS. Interventions to improve medication adherence in hypertensive patients: systematic review and meta-analysis. Curr Hypertens Rep 2015 Dec;17(12):94. [doi: 10.1007/s11906-015-0606-5] [Medline: 26560139]
15. Khoong EC, Olazo K, Rivadeneira NA, et al. Mobile health strategies for blood pressure self-management in urban populations with digital barriers: systematic review and meta-analyses. NPJ Digit Med 2021 Jul 22;4(1):114. [doi: 10.1038/s41746-021-00486-5] [Medline: 34294852]
16. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267. [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
17. Park J, Lee U. Understanding disengagement in just-in-time mobile health interventions. Proc ACM Interact Mob Wearable Ubiquitous Technol 2023 Jun 12;7(2):1-27. [doi: 10.1145/3596240]
18. Yardley L, Spring BJ, Riper H, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
19. Schoeppe S, Alley S, Van Lippevelde W, et al. Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: a systematic review. Int J Behav Nutr Phys Act 2016 Dec 7;13(1):127. [doi: 10.1186/s12966-016-0454-y] [Medline: 27927218]
20. Walton A, Nahum-Shani I, Crosby L, Klasnja P, Murphy S. Optimizing digital integrated care via micro-randomized trials. Clin Pharmacol Ther 2018 Jul;104(1):53-58. [doi: 10.1002/cpt.1079] [Medline: 29604043]
21. Wagner B, Liu E, Shaw SD, et al. ewrapper: engagement strategies in mHealth. Proc ACM Int Conf Ubiquitous Comput 2017 Sep;2017:790-798. [doi: 10.1145/3123024.3125612] [Medline: 29362728]
22. Meyerowitz-Katz G, Ravi S, Arnolda L, Feng X, Maberly G, Astell-Burt T. Rates of attrition and dropout in app-based interventions for chronic disease: systematic review and meta-analysis. J Med Internet Res 2020 Sep 29;22(9):e20283. [doi: 10.2196/20283] [Medline: 32990635]
23. Meurer WJ, Dinh M, Kidwell KM, et al. Reach out behavioral intervention for hypertension initiated in the emergency department connecting multiple health systems: study protocol for a randomized control trial. Trials 2020 Jun 3;21(1):456. [doi: 10.1186/s13063-020-04340-z] [Medline: 32493502]
24. Skolarus LE, Dinh M, Kidwell KM, et al. Reach out emergency department: a randomized factorial trial to determine the optimal mobile health components to reduce blood pressure. Circ Cardiovasc Qual Outcomes 2023 May;16(5):e009606. [doi: 10.1161/CIRCOUTCOMES.122.009606] [Medline: 37192282]
25. Price RJ, Brenner AB, Lin CC, Burke JF, Skolarus LE. Two million stroke survivors utilize medical visit companions: the other person in the room? Neurology 2019 Nov 12;93(20):899-901. [doi: 10.1212/WNL.0000000000008465] [Medline: 31619484]
26. Blood pressure devices. US Blood Pressure Validated Device Listing. 2022. URL: https://www.validatebp.org/ [accessed 2024-05-08]
27. Thorndike RL. Who belongs in the family? Psychometrika 1953 Dec;18(4):267-276. [doi: 10.1007/BF02289263]
28. Lin Z, Althoff T, Leskovec J. I'll be back: on the multiple lives of users of a mobile activity tracking application. Proc Int World Wide Web Conf 2018 Apr;2018:1501-1511. [doi: 10.1145/3178876.3186062] [Medline: 29780978]

29. Burns K, Nicholas R, Beatson A, Chamorro-Koc M, Blackler A, Gottlieb U. Identifying mobile health engagement stages: interviews and observations for developing brief message content. J Med Internet Res 2020 Sep 22;22(9):e15307. [doi: 10.2196/15307] [Medline: 32960181]
30. Zolnierek KBH, Dimatteo MR. Physician communication and patient adherence to treatment: a meta-analysis. Med Care 2009 Aug;47(8):826-834. [doi: 10.1097/MLR.0b013e31819a5acc] [Medline: 19584762]
31. Klasnja P, Smith S, Seewald NJ, et al. Efficacy of contextually tailored suggestions for physical activity: a micro-randomized optimization trial of HeartSteps. Ann Behav Med 2019 May 3;53(6):573-582. [doi: 10.1093/abm/kay067] [Medline: 30192907]
32. Gazit T, Gutman M, Beatty AL. Assessment of hypertension control among adults participating in a mobile technology blood pressure self-management program. JAMA Netw Open 2021 Oct 1;4(10):e2127008. [doi: 10.1001/jamanetworkopen.2021.27008] [Medline: 34652447]
33. Kaplan AL, Cohen ER, Zimlichman E. Improving patient engagement in self-measured blood pressure monitoring using a mobile health technology. Health Inf Sci Syst 2017 Dec;5(1):4. [doi: 10.1007/s13755-017-0026-9] [Medline: 29081974]

Abbreviations

BP: blood pressure
ED: emergency department
mHealth: mobile health
OR: odds ratio
PCP: primary care provider
SMBP: self-measured blood pressure

Edited by A Mavragani; submitted 28.11.23; peer-reviewed by R Newton, R Zhuang; revised version received 05.04.24; accepted 09.04.24; published 12.06.24.

Please cite as:
Skolarus LE, Lin CC, Mishra S, Meurer W, Dinh M, Whitfield C, Bi R, Brown D, Oteng R, Buis LR, Kidwell K
Engagement in mHealth-Prompted Self-Measured Blood Pressure Monitoring Among Participants Recruited From a Safety-Net Emergency Department: Secondary Analysis of the Reach Out Trial
JMIR Mhealth Uhealth 2024;12:e54946
URL: https://mhealth.jmir.org/2024/1/e54946
doi:10.2196/54946

Original Paper

Development of a Health Behavioral Digital Intervention for Patients With Hypertension Based on an Intelligent Health Promotion System and WeChat: Randomized Controlled Trial

Ting Sun[1,2,3], MSN; Xuejie Xu[3], BSN; Zenghui Ding[2], PhD; Hui Xie[3], PhD; Linlin Ma[3]; Jing Zhang[3]; Yuxin Xia[3], BSN; Guoli Zhang[3], BSN; Zuchang Ma[2], PhD

[1]Graduate School, University of Science and Technology of China, Hefei, China

[2]Hefei Institutes of Physical Science, Chinese Academy of Sciences, Hefei, China

[3]School of Nursing, Bengbu Medical University, Bengbu, China

Corresponding Author:
Zuchang Ma, PhD
Hefei Institutes of Physical Science
Chinese Academy of Sciences
#350 Shushan Lake Road
Shushan District
Hefei, 230031
China
Phone: 86 13956984669
Email: ZCMa121@126.com

Abstract

Background: The effectiveness of timely medication, physical activity (PA), a healthy diet, and blood pressure (BP) monitoring for promoting health outcomes and behavioral changes among patients with hypertension is supported by a substantial amount of literature, with “adherence” playing a pivotal role. Nevertheless, there is a lack of consistent evidence regarding whether digital interventions can improve adherence to healthy behaviors among individuals with hypertension.

Objective: The aim was to develop a health behavioral digital intervention for hypertensive patients (HBDIHP) based on an intelligent health promotion system and WeChat following the behavior change wheel (BCW) theory and digital micro-intervention care (DMIC) model and assess its efficacy in controlling BP and improving healthy behavior adherence.

Methods: A 2-arm, randomized trial design was used. We randomly assigned 68 individuals aged >60 years with hypertension in a 1:1 ratio to either the control or experimental group. The digital intervention was established through the following steps: (1) developing digital health education materials focused on adherence to exercise prescriptions, Dietary Approaches to Stop Hypertension (DASH), prescribed medication, and monitoring of BP; (2) using the BCW theory to select behavior change techniques; (3) constructing the intervention's logic following the guidelines of the DMIC model; (4) creating an intervention manual including the aforementioned elements. Prior to the experiment, participants underwent physical examinations at the community health service center's intelligent health cabin and received intelligent personalized health recommendations. The experimental group underwent a 12-week behavior intervention via WeChat, while the control group received routine health education and a self-management manual. The primary outcomes included BP and adherence indicators. Data analysis was performed using SPSS, with independent sample *t* tests, chi-square tests, paired *t* tests, and McNemar tests. A *P* value <.05 was considered statistically significant.

Results: The final analysis included 54 participants with a mean age of 67.24 (SD 4.19) years (n=23 experimental group, n=31 control group). The experimental group had improvements in systolic BP (–7.36 mm Hg, *P*=.002), exercise time (856.35 metabolic equivalent [MET]-min/week, *P*<.001), medication adherence (0.56, *P*=.001), BP monitoring frequency (*P*=.02), and learning performance (3.23, *P*<.001). Both groups experienced weight reduction (experimental: 1.2 kg, *P*=.002; control: 1.11 kg, *P*=.009) after the intervention. The diet types and quantities for both groups (*P*<.001) as well as the subendocardial viability ratio (0.16, *P*=.01) showed significant improvement. However, there were no statistically significant changes in other health outcomes.

Conclusions: The observations suggest our program may have enhanced specific health outcomes and adherence to health behaviors in older adults with hypertension. However, a longer-term, larger-scale trial is necessary to validate the effectiveness.

Trial Registration: Chinese Clinical Trial Registry ChiCTR2200062643; https://www.chictr.org.cn/showprojEN.html?proj=172782

International Registered Report Identifier (IRRID): RR2-10.2196/46883

(JMIR Mhealth Uhealth 2024;12:e53006) doi:10.2196/53006

KEYWORDS

adherence; hypertension; health behavior; mHealth; digital health

Introduction

Hypertension and Health Behavior Interventions

Hypertension, as a high-prevalence chronic disease, has become an important risk factor for many diseases (eg, stroke, renal disease) and a major contributor to the global burden of disease [1,2]. Approximately one-third of older adults with hypertension fail to achieve their blood pressure (BP) control goals [3]. The reasons for the low rate of hypertension control are related to high-risk lifestyles such as poor dietary habits and low levels of physical activity (PA).

Relevant studies have shown that adherence to recommended health behaviors can significantly reduce systolic blood pressure (SBP) by an average of 4.0 mm Hg to 5.6 mm Hg and diastolic blood pressure (DBP) by an average of 4.1 mm Hg to 5.3 mm Hg in individuals with hypertension [4]. Engaging in a wide range of exercise training can lead to average reductions of 4.08 mm Hg to 8.24 mm Hg in SBP and 2.5 mm Hg to 4.0 mm Hg in DBP [5]. Regular BP monitoring behaviors result in average reductions of 2.53 mm Hg to 4.7 mm Hg in SBP and 1.45 mm Hg to 2.4 mm Hg in DBP [6,7]. Adherence to Dietary Approaches to Stop Hypertension (DASH) not only reduces SBP and DBP by approximately 5.5 mm Hg and 3 mm Hg, respectively, but also reduces the chance of developing hypertension by 26% [8,9]. High medication adherence, combined with comprehensive interventions like diet and exercise management, leads to improved BP control [10,11]. It is evident that comprehensive health behavior interventions can achieve effective BP control, with adherence playing a pivotal role.

In recent years, several studies exploring intelligent health promotion systems that incorporate advanced technologies like artificial intelligence, wearable devices, and mobile communication have consistently shown their efficacy in managing chronic diseases [12-14]. In addition, numerous studies support that independent mobile health (mHealth) apps play a vital role in community-based patient management [10,12,14-16]. WeChat, China's predominant social communication mobile app, serves as one platform for mHealth interventions. It boasts a staggering daily user count of up to 902 million people and over 1 billion monthly active users spanning all age groups [17]. Several studies have indicated that mHealth-based interventions can enhance health outcomes, quality of life, and self-care among patients with chronic diseases [18,19]. BP monitoring adherence (BPMA), dietary habits, and self-efficacy behaviors of patients with hypertension have been somewhat improved by these interventions based on mHealth [20-22]. However, other studies have shown no significant change or little improvement in BP, medicine, exercise, and DASH adherence [20,23-25].

Theoretical Framework

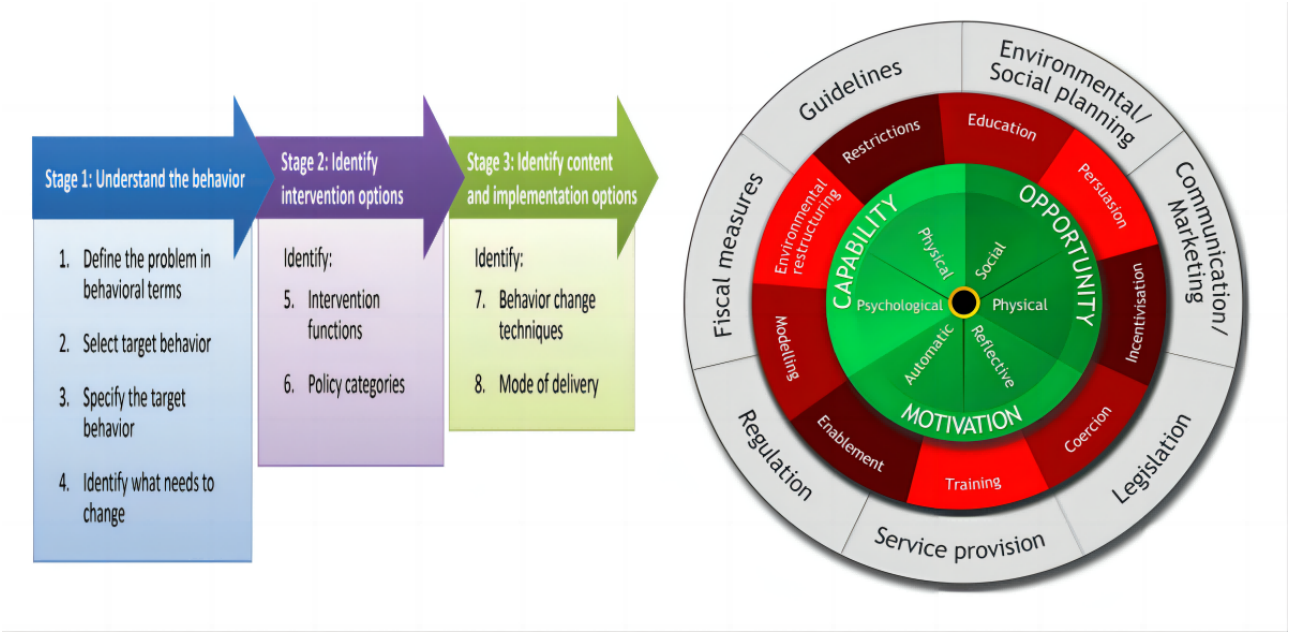
The behavior change wheel (BCW), which integrates 19 relevant theoretical frameworks for behavior change, was first proposed by Michie et al in 2011 [26]. As Figure 1 illustrates, it consists of 3 tiers. The inner tier is the Capability, Opportunity, Motivation-Behavior (COM-B) model, which is used to identify barriers to intervening in the target behavior. The second tier comprises the following 9 intervention categories intended to tackle identified behavioral obstacles: education, persuasion, motivation, coercion, training, restriction, environmental restructuring, demonstration, and empowerment. The outermost tier encompasses 7 policy categories (eg, regulation and legislation) that aid in the implementation of macrolevel interventions [27]. This theory consists of 3 steps: understanding the behavior, identifying intervention options, and identifying content and implementing the options (ie, behavior change techniques [BCTs]). Michie et al [28], along with other scholars, developed “The behavior change technique taxonomy of 93 hierarchically clustered techniques,” which contains 93 BCTs and provides names, definitions, and examples. In the final step, researchers can select the necessary BCTs from this taxonomy list.

Scholars have applied this theory to community health promotion, health care management, and nursing care, resulting in positive outcomes [29-32]. Additionally, some researchers have extended its application to digital and mHealth interventions [33]. The effectiveness of this lies in its ability to aid interveners in the systematic and scientific identification of intervention functions and specific BCTs for behavior change problems. Despite this, the theory does not provide much insight into the components of the intervention. The BCW can be supplemented by the digital micro-intervention care (DMIC) model [25,34].

In 2020, the DMIC model proposed by Baumel and colleagues [34] provided a reference paradigm. This theoretical model promotes shorter, more focused interventions, known as micro-interventions. They can be highly focused on implementation in people's daily lives to help intervention recipients achieve desired short-term goals (the basis for achieving long-term goals). The DMIC comprises the following 3 core concepts: events (in-the-moment attempts at change or impact toward the overall target of the intervention), decision rules (guiding which events are deployed and when), and proximal assessments (assessing the impact of the event). Events are similar to specific BCTs, while decision rules deploy events in a meaningful way based on time, user status, or environmental

information, allowing interveners to dynamically adjust the content of micro-interventions. Overall outcome assessment, proximal event outcome assessment, and assessment of user participation in the micro-intervention are 3 types of impact assessment for digital micro- interventions. Based on the results of the overall and proximal assessments and the continuous recording of the user engagement experience (ie, measuring the quality of attention, engagement, and immersion during the use of the program), it is possible to identify individuals with low levels of engagement and later search for the causes, modify the intervention decision rules, and re-engage the user. This theoretical framework advocates for interventions that aim to achieve specific objectives through in-the-moment intervention elements. These elements may not be directly tied to the attainment of a broader clinical goal [34]. Each intervention in every event represents an immediate attempt to modify or influence the overall goal of the intervention. This implies that, in order to achieve a clinical objective, interventions should be broken down into numerous small steps and regulated through proximal assessments, ultimately leading to the attainment of the overall outcome.

Figure 1. Behavior change wheel.



Goal of This Study

Therefore, our objective was to develop a health behavioral digital intervention for hypertensive patients (HBDIHP) based on BCW and DMIC and assess the effectiveness of this program in 2 groups after 3 months of intervention. This program involved exercise, diet, BP monitoring, and medication adherence intervention strategies. Consequently, we aimed to assess the effectiveness of this approach for enhancing outcomes for older adults with hypertension.

Methods

Development of HBDIHP

Community-Oriented Intelligent Health Promotion System

The Intelligent Health Promotion System is a cloud platform–based system that leverages health sign data and health questionnaire responses to generate intelligent health reports, personalized exercise prescriptions, and other tailored health recommendations. It also tracks and monitors individual health data. This system comprises the following 3 main layers: perception layer, decision layer, and application layer (see Figure 2).

This system is installed in the Intelligent Health Cabin at community health service centers (see Figure 3). It currently supports a range of connected instruments, including cardiovascular function monitors, arteriosclerosis detectors, body composition monitors, bone densitometers, and physical fitness detectors. After completing assessments with these instruments, the system sends those collected data to its central cloud platform database. Participants need to continue to fill out various questionnaires, such as chronic disease history questionnaires, medication profiles, and family medical history surveys. The system then uses both instrument data and questionnaire responses to activate the intelligent decision-making and inference engine, which generates comprehensive reports. These reports provide a comprehensive evaluation of an individual's health status and offer personalized, evidence-based recommendations, namely personalized intervention plans.

Figure 2. Intelligent health promotion system architecture diagram.

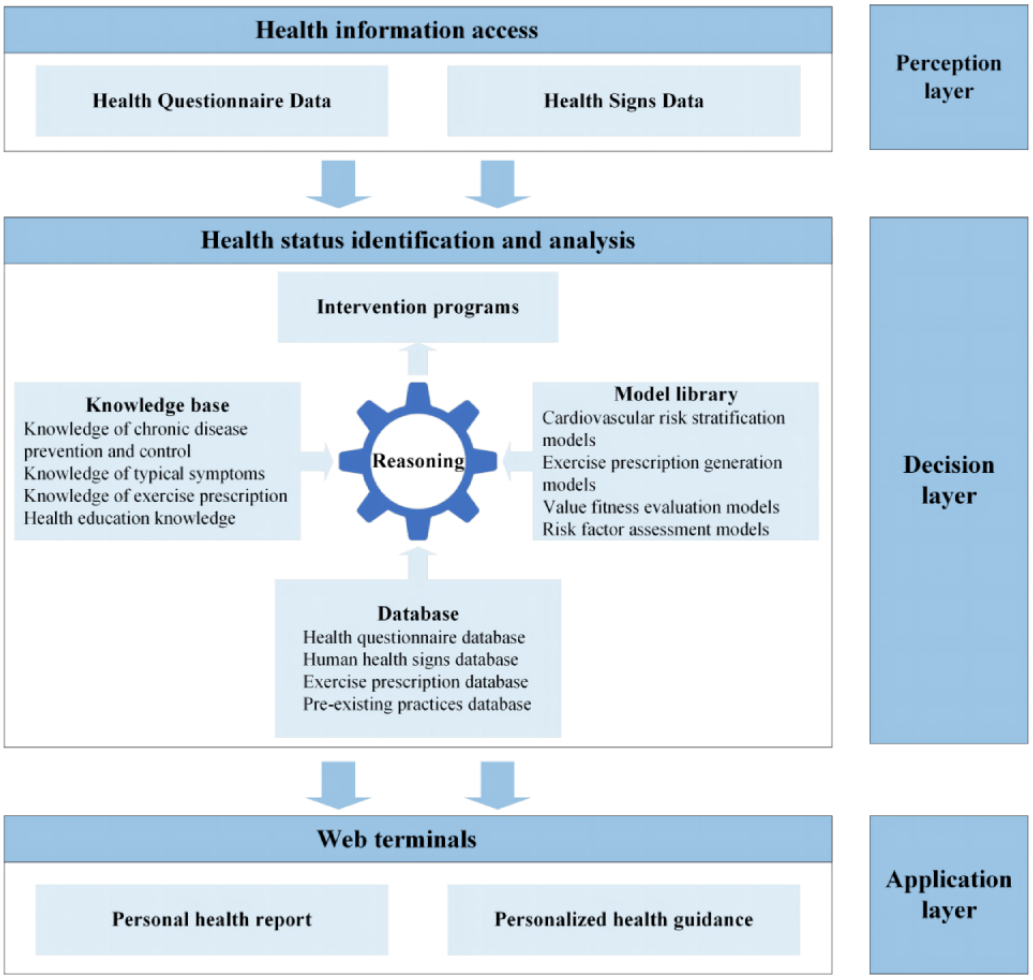


Figure 3. Intelligent Health Cabin.

Assessment of Health Outcomes and Risk Prediction

The comprehensive report provides an overview of participants' health, identifying health issues in areas such as the cardiovascular system, lipid metabolism, musculoskeletal system, lifestyle, and physical fitness, while explaining the meaning of abnormal indicators. In the cardiovascular assessment, the system not only evaluates participants' cardiovascular systems based on instrument results but also predicts participants' heart age and the risk of cardiovascular disease using a machine learning model.

Personalized Health Advice

The comprehensive report offers personalized health advice to each resident, which includes exercise prescriptions, dietary recommendations, and suggestions for behavior correction. The exercise prescription, based on the design principles of the American College of Sports Medicine guidelines (frequency, intensity, type, time, volume, and progression [FITT-VP])

combined with the Transtheoretical Model, delineates intervention plans tailored to the health care stage, exercise habit formation stage, scientific fitness stage, and exercise habit maintenance stage. After being processed by the intelligent decision module, the system generates exercise prescriptions customized for individual residents (see Figure 4), considering the exclusion of exercise contraindications. These prescriptions encompass exercise recommendations, principles, weekly plans, exercise correction, precautions, and exercise guidance (see Figure 5) [13].

A dietary guidance and behavior correction database was constructed using technical strategies such as expert systems and knowledge graphs. In practice, the system provides dietary recommendations and behavior correction suggestions (eg, health advice for sedentary individuals to change their unhealthy habits) based on diet-related questionnaires, medical history collection, and physical examination results (see Figure 6).

Figure 4. Exercise prescription generation flowchart.

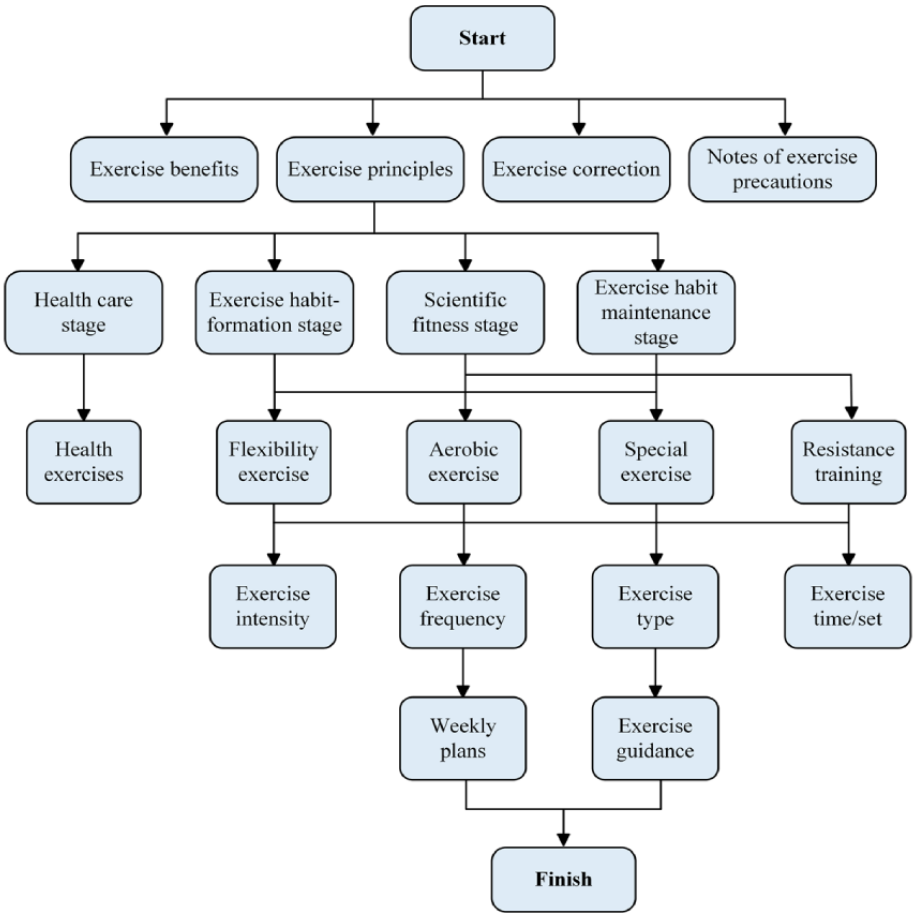


Figure 5. Personalized exercise prescription.

Figure 6. Dietary recommendations and behavior correction suggestions.

Development of the Health Behavioral Digital Intervention

A multidisciplinary working group for digital health intervention strategies was established during the entire intervention program design process. This group included 3 patients with hypertension, 1 clinical expert in hypertension, 1 management expert, and 1 behavioral psychology expert. Group-focused interviews were carried out at every stage of development. At this stage, experts evaluated intervention strategies using their professional knowledge, while patients with hypertension assessed the acceptability and utility of the interventions from their perspective. After finalizing the intervention scheme, experts who were not involved in the development were invited to assess its structural validity.

Defining Health Behavior Management Targets and Health Education Content

The identification of the 4 key health behavior management targets and associated health education content for hypertension (see Table 1) was based on existing literature and previous research findings.

Table 1. Health education targets and content.

Health behaviors	Targets	Health education content
Physical activity (exercise)	Recommended exercise levels according to exercise prescription	Relevant knowledge and exercise skills for personalized intelligent exercise prescriptions
Healthy diet	Choose types and quantities of food according to the DASH[a] principles	Knowledge related to DASH and techniques for assessing simplified DASH grading diet index score
Taking medication	Adherence to prescribed medication	Knowledge of medications
Monitoring of blood pressure	The recommended frequency of home blood pressure monitoring according to clinical guidelines	Knowledge and skills in blood pressure monitoring

[a]DASH: Dietary Approaches to Stop Hypertension.

Development of Digital Health Education Materials

Patient health education needs for the 4 targets were gathered through focus group interviews. Subsequently, evidence-based principles were used to develop textual materials related to exercise, diet, medication taking, and BP monitoring. These materials underwent content validity assessments by experts before being transformed into health education videos. For diet, in addition to creating health education videos based on the DASH guidelines, the research team developed a simplified DASH grading diet index score. The scoring system included items for evaluating daily meals, such as grains, vegetables, fruits, protein sources, cooking oil, and compliance with recommended food types and quantities. Items 1 through 7 represented positive scoring criteria for each recommended food category and quantity, while items 8 through 10 incurred deductions. The aim was to educate patients on recommended dietary behaviors through the acquisition of DASH knowledge and proficiency in using the simplified DASH grading diet index score. For exercise, in addition to providing general knowledge about exercise, the research team augmented the guidance materials (eg, videos on how to perform recommended exercise types) without altering the existing elements of the intelligent exercise prescription. Medication adherence and BP monitoring were primarily addressed through instructional videos that imparted relevant knowledge and skills. All relevant videos were accompanied by textual materials that were ultimately compiled into the “hypertension self-management manual.” See Figure 7 for an overview.

Figure 7. Hypertension self-management manual.

Development of Digital Intervention Scheme Based on BCW and DMIC

We used the systematic workflow of the BCW to identify appropriate intervention categories and suitable BCTs. Details of this process have been published in other papers [35]. Based on the identified BCTs, we developed corresponding textual content, which, along with the previously mentioned digital health education materials, collectively form the foundational elements of the DMIC theory. Subsequently, additional elements of the DMIC theory, namely proximal assessment indicators and decision rules, were determined based on literature review and expert opinions.

First, the BCW theory was used to identify barriers related to capability, opportunity, and motivation affecting adherence to health behaviors by patients with hypertension. Intervention categories were chosen to address these barriers, including methods like education, persuasion, and incentivization. Figure 8 illustrates the process of developing the mHealth intervention scheme for improving adherence.

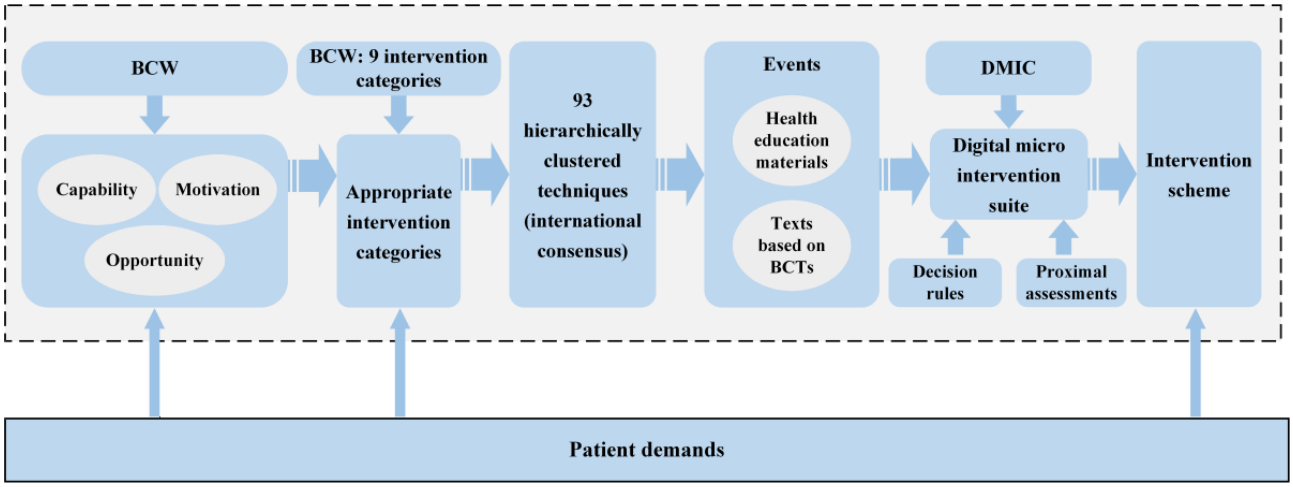
Second, BCTs, such as feedback on behavior, prompts, self-monitoring of behavior, and verbal persuasion about capability, were selected from the BCT taxonomy, which had already been coded and organized by researchers like Michie et al [28]. Furthermore, based on the selected BCT, specific textual content was prepared (eg, BCT: focus on past success; text: “You have successfully quit smoking in the past, and we believe you can also develop a scientific exercise habit! Keep it up!”).

Third, following the DMIC model, we classified the textual content corresponding to BCTs and digital health education materials as “events.” Subsequently, we established decision rules that determined when and in what order interventions for these events should take place. Additionally, we established proximal assessment indicators (eg, knowledge level) for tailoring intervention strategies to individual patients and their corresponding events (eg, continue learning if qualified, relearn if not). Events, proximal assessment, and decision rules were integrated into intervention units organized by chronological stages, including assessing and preparing, committing and planning, and reinforcing behavioral habits. Notably, health education was primarily implemented during the assessing and preparing phase (first week).

Fourth, the intervention scheme was validated by an expert panel through 2 rounds of Delphi surveys. The panel of 15 experts included 6 nutrition experts, 2 clinical cardiologists, 3 clinical nurses and nursing teachers, and 4 exercise experts. The content validity index (CVI) was calculated and assessed using the item-level CVI (I-CVI) and a 4-point scale, respectively. The first round of assessments using the I-CVI ranged from 0.6 to 0.8. After adjusting the content based on the feedback and comments from the experts, the I-CVI reached 1 in the second round.

Finally, all intervention logic and guidance content were compiled into an intervention manual. This manual included daily intervention tasks (eg, questionnaires, text based on BCTs, and the order of videos to be sent), communication guidelines, and personalized guidance strategies (eg, after assessing the extent of the patient's knowledge and skills, the approach to providing personalized guidance can be provided).

Figure 8. The process of shaping the mobile health (mHealth) intervention scheme for adherence. BCT: behavior change technique; BCW: behavior change wheel; DMIC: digital micro-intervention care.

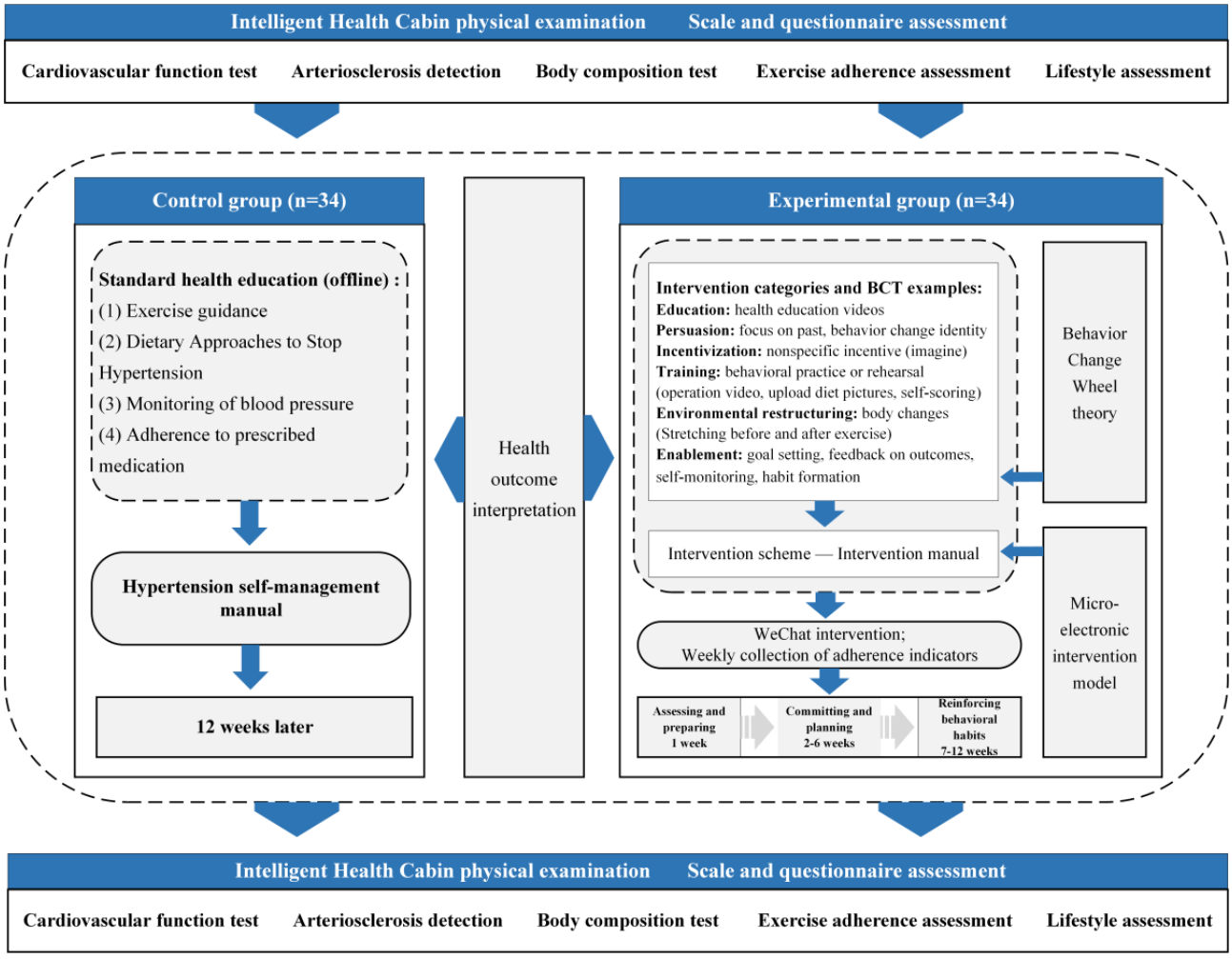


Study Design

Trial Design and Setting

The study protocol was previously published [35]. This was a randomized controlled trial. The experimental group received the health behavioral digital intervention based on an intelligent health promotion system and WeChat for 12 weeks, while the control group received routine health services and was provided with a "Hypertension Self-Management Manual" to guide daily health behaviors (refer to Figure 9). The trial was conducted at 2 community health centers, both located in Anhui Province: Sanxiao Kou Community Health Service Center and Dongfeng Community Health Service Center.

Figure 9. Trial design.



Patients

The staff at the community health service centers recruited eligible patients with hypertension within their jurisdiction through phone calls or verbal invitations using a convenience sampling method. The inclusion criteria for participants were as follows: (1) diagnosed with primary hypertension or currently taking antihypertensive medication; (2) aged >60 years; and (3) proficient in using smartphones and the WeChat application. The exclusion criteria were as follows: (1) patients with hypertension undergoing nonpharmacological treatment; (2) those with diabetes, kidney disease, or other conditions requiring special dietary and exercise considerations; (3) individuals participating in or having participated in other health management projects; and (4) those unable to measure BP in a home environment.

Sample Size

The sample size for this study was determined based on the effect size. According to a previous meta-analysis comparing the BP-lowering effects of mHealth interventions and other traditional methods used by patients with hypertension, the mHealth experimental group demonstrated a significant reduction in BP, with an effect size of 0.7 [36]. The sample size for this study was calculated using GPower 3.1, assuming α=.05 and β=.2 and accounting for a 20% dropout rate, resulting in a final sample size of 68 participants.

Randomization, Allocation, and Blinding

Prior to randomization, a researcher unfamiliar with the experimental design used SPSS (version 23; IBM Corp) to generate 68 random numbers. Subsequently, using the visual binning method, these numbers were divided into 2 groups (experimental group and control group). The paper slips containing the labeled random numbers were placed into sealed envelopes, and another researcher unaware of the experimental design was responsible for assigning participants to either the experimental group or control group. This researcher sequentially opened the envelopes and assigned participants to the experimental group or control group based on the numbers. Additionally, during the data collection phases before and after the experiment, we trained and employed 2 nursing graduate students unfamiliar with the study design. The data collection for the experimental group intervention process was conducted by 5 nursing undergraduates who were also unfamiliar with the experimental design.

Outcomes

Primary Outcomes

Before and 12 weeks after the intervention, we used a cardiovascular function monitoring device (BX-CFTI-100, Intelligent Machine Institute) to measure participants' BP. Measurement details have been published [13]. Through WeChat, 5 undergraduate students unfamiliar with the experimental design collected adherence indicators from the experimental group participants weekly. These indicators included exercise adherence, calculated as the ratio of actual weekly exercise time meeting the prescribed intensity to the total prescribed weekly exercise time; dietary adherence, assessed using the weekly average score of the simplified DASH grading diet index score (see protocol paper for details [35]); medication adherence, assessed using the weekly average score of the Modified Morisky Scale (Chinese version-MMS-8, Certificate Number: 8538-1877-1559-6025-5310) [37-39]; and BPMA calculated as the ratio of the actual weekly frequency of BP monitoring to the total recommended weekly frequency.

Secondary Outcomes

Secondary outcomes were assessed before and after the intervention. Heart rate and subendocardial viability ratio (SEVR) were measured using the cardiovascular function monitor; brachial-ankle pulse wave velocity was measured using an arterial stiffness monitor (BX-AS-100, Intelligent Machine Institute); weight was measured using a body composition analyzer (BX-BCA-100, Intelligent Machine Institute); and lifestyle was assessed through an online questionnaire administered to participants pre- and postintervention. Participants were questioned about health-related behaviors, including the frequency of smoking, drinking, diet, and PA. For smoking, the lifelong smoking quantity of participants was calculated based on the quantity and weekly frequency of cigarettes smoked. For those who had quit smoking, this calculation also included their smoking history before quitting. The alcohol content in one bottle of the most popular alcoholic beverages in Anhui Province is as follows: beer (500 mL, 3.2% alcohol): 17.5 g; white liquor (450 mL, 42% alcohol): 210 g; and wine (750 mL, 13.5%-14% alcohol): 97.5 g. Daily alcohol consumption was calculated using these values. Missing values for smoking and drinking data were set to zero. Participants were asked about the types, duration (in minutes), and frequency (per week) of PA they engaged in. PA time was determined in minutes per metabolic equivalent (MET) per day (min/MET/day) based on activity codes and MET intensities in the "Compendium of Physical Activities." Weekly exercise time (MET-min/week) and weekly PA time (MET-min/week) were calculated, with missing values set to the median [40].

Statistical Analysis

Data analysis was carried out using SPSS, with independent sample *t* tests and chi-square tests used to evaluate the significance of differences between the 2 groups. Paired *t* tests and McNemar tests were used to compare differences within the same group before and after the 3-month intervention. A *P* value <.05 was considered statistically significant.

Ethical Considerations

The study was approved by the Ethics Committee of Bengbu Medical College in June 2022 under approval number 2022-103, and the study began after informed consent was obtained from patients with hypertension and all participants allowed their data to be used anonymously. An independent data manager conducted weekly checks on the database to ensure its integrity and security. The implementation of data lockup aims to prevent any postmodification. All exported data must undergo anonymization by the data manager before statistical analysis can be conducted to safeguard the participants' information. Each participant who underwent the intervention and data collection at the community health center was given a gift of daily necessities valued at approximately US $6.87.

Results

Participants

Between September 5, 2022, and September 19, 2022, a total of 68 patients with hypertension were recruited through phone or verbal invitations from the 2 community health centers in Anhui Province. Participants were re-invited to the community health service center for health data and information collection between December 20, 2022, and January 5, 2023. A total of 54 participants (30 women and 24 men; mean age 67.24, SD 4.19 years) were included in the final analysis: 23 in the experimental group and 31 in the control group. Exclusions were due to various reasons, including hospitalization for illness (2 individuals), inability to complete the postintervention assessment (3 individuals), health conditions deteriorating to the point of hindering PA (4 individuals), voluntary withdrawal from the study (4 individuals), and a change in antihypertensive medication (1 individual). Other patients did not change their hypertension medication during the intervention.

Baseline Data

No statistically significant differences in health outcomes (eg, SBP, SEVR), adherence indicators (eg, exercise time, PA time, medication adherence), and learning performance were observed between the experimental group and control group at baseline.

Intervention Effect

Table 2 displays the baseline and posttest results for both groups regarding health outcomes, adherence indicators, and learning performance. Significant changes were observed in SBP (–7.36 mm Hg, P=.05), SEVR (0.16, P=.01), exercise time (856.35 MET-min/week, P=.03), medication adherence (0.56, P=.02), BP monitoring frequency (P=.046), and learning performance (3.23, P<.001) in the intergroup comparison after 12 weeks. The PA time increased for the experimental group in the before-and-after comparisons (P=.045). Both groups experienced a reduction in weight after the intervention (experimental: 1.2 kg, P=.002; control: 1.11 kg, P=.009). Furthermore, the Cohen d values reflecting effect size were greater than 0.5 for all variables except PA time, indicating at least an intermediate effect size. Among these variables, the health outcomes of SEVR, recommended diet types (eg, meeting recommendations occasionally, sometimes, often), recommended diet quantities (eg, meeting recommendations occasionally, sometimes, often), BP monitoring frequency (eg, measure daily, measure 1-3 times a week, measure whenever remember), and learning performance had Cohen d values >0.8, suggesting a large effect size.

Table 2. Effects of the health behavioral digital intervention for hypertensive patients (HBDIHP) effects (N=54).

Measure	Control group				Experimental group				Cohen *d*	*P* value[a]
	Baseline	12 weeks	Cohen *d*	*P* value[a]	Baseline	12 weeks	Cohen *d*	*P* value[a]		
Health outcomes, mean (SD)										
SBP[b] (mm Hg)	136.94 (18.44)	133.10 (15.02)	0.549	.10	135.43 (17.48)	125.74 (14.76)	1.457	.002	0.636	.05
DBP[c] (mm Hg)	76.03 (9.20)	75.58 (6.94)	0.129	.73	78.39 (8.81)	75.96 (6.38)	0.286	.19	0.516	.84
Heart rate (bpm)	74.58 (10.58)	74.58 (9.45)	0.149	≥.99	77.09 (9.18)	73.91 (9.84)	0.548	.08	0.52	.35
SEVR[d]	1.11 (0.26)	1.03 (0.15)	0.487	.25	1.12 (0.21)	1.19 (0.25)	0.736	.13	0.806	.01
baPWV[e] (m/s)	16.87 (2.48)	16.67 (2.22)	0.21	.64	16.23 (1.82)	16.51 (1.64)	0.4	.51	0.523	.78
Weight (kg)	69.54 (9.93)	68.34 (9.52)	0.308	.009	71.48 (9.92)	70.37 (9.47)	0.721	.002	0.56	.45
Adherence indicators, mean (SD)										
Exercise time (MET[f]-min/week)	1263.3 (775.50)	1823.74 (1208.06)	1.143	.009	1578.78 (709.95)	2680.09 (1604.09)	1.554	<.001	0.616	.03
Physical activity time (MET-min/week)	2462.09 (1405.61)	2861.03 (1484.81)	0.688	.15	3166.82 (2277.17)	3742.13 (2138.65)	0.649	.045	0.492	.09
Medication adherence score	6.99 (1.02)	7.09 (1.00)	0.247	.46	6.77 (1.22)	7.65 (0.49)	1.584	.001	0.68	.02
Recommended diet types, n (%)									1.318	<.001
Occasionally	—[g]	21 (68)	—	—	—	3 (13)	—	—		
Sometimes	—	8 (26)	—	—	—	6 (26)	—	—		
Often	—	2 (7)	—	—	—	14 (61)	—	—		
Recommended diet quantity, n (%)									1.18	<.001
Occasionally	—	21 (68)	—	—	—	3 (13)	—	—		
Sometimes	—	5 (16)	—	—	—	5 (22)	—	—		
Often	—	5 (16)	—	—	—	15 (65)	—	—		
Blood pressure monitoring frequency, n (%)			1.298	<.001			0.654	.02	1.318	.046
Measure daily	7 (23)	10 (32)			0	12 (52)				
Measure 1-3 times a week	16 (52)	14 (45)			18 (78)	9 (39)				
Measure whenever remember	8 (26)	7 (23)			5 (22)	2 (9)				
Learning performance, mean (SD)	9.32 (1.79)	9.55 (1.77)	0.323	.15	9.91 (1.51)	12.78 (2.04)	1.617	<.001	0.887	<.001

[a]12-week intergroup *P* value.

[b]SBP: systolic blood pressure.

[c]DBP: diastolic blood pressure.

[d]SEVR: subendocardial viability ratio.

[e]baPWV: brachial-ankle pulse wave velocity.

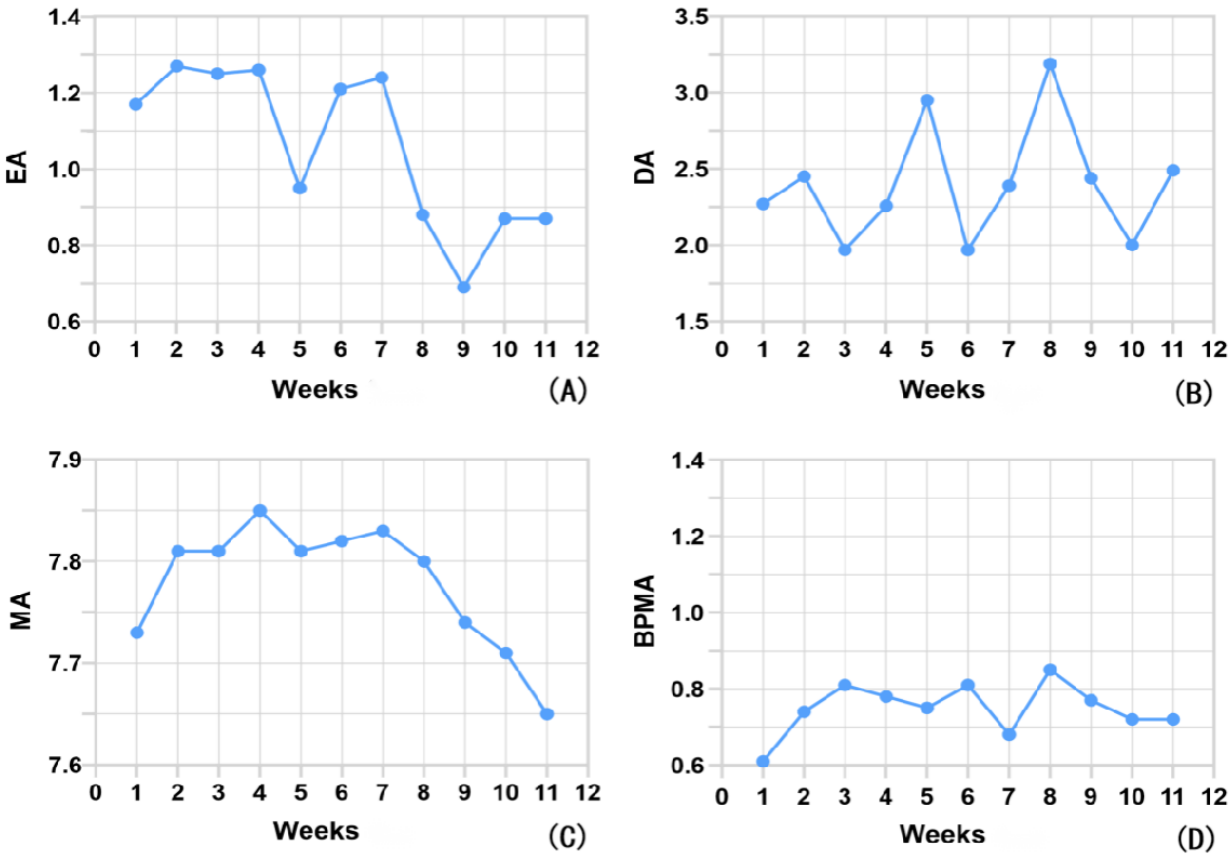
[f]MET: metabolic equivalent.

[g]Not applicable.

Changes in Weekly Adherence Indicators in the Experimental Group

The experimental group collected weekly average adherence indicators from the first week after the assessing and preparing phase (the second week of the project) through the eleventh week (the twelfth week of the project), as depicted in the curve shown in Figure 10. From weeks 1 to 4, the exercise adherence curve indicated that this phase was when the intervention participants were most actively engaged in PA. They exceeded the prescribed exercise volume (>1). By week 5, the participants showed lower exercise adherence, followed by 2 weeks of rebound. Subsequently, exercise adherence sharply declined, reaching a stable state in weeks 10 and 11. The dietary adherence curve indicators exhibited the most noticeable fluctuations, reaching peaks in weeks 4 and 8 and valleys in weeks 6 and 10. Medication adherence gradually increased from week 1 to week 4, experienced fluctuations, and then steadily declined starting at week 7. Among the 4 indicators, the BPMA curve had smaller fluctuations. It steadily increased from week 1 to week 3, reached a low point in week 7, and remained relatively stable at 0.72 thereafter.

Figure 10. Average adherence indicators by week in the intervention: (A) exercise adherence (EA), (B) dietary adherence (DA), (C) medication adherence (MA), (D) blood pressure monitoring adherence (BPMA).



Discussion

Health Outcomes

We observed varying degrees of positive impacts on health outcomes for patients with hypertension through this health behavioral digital intervention. The intervention demonstrated a significant effect on SBP control. It may also be effective at improving SEVR (with statistical differences observed between the 2 groups after the intervention). However, there is insufficient evidence to conclude that a WeChat-based digital intervention is more effective than a conventional intervention for weight improvement. Based on the current results, both interventions appear to have a positive effect on weight reduction. No statistically significant differences were found in other health outcome indicators.

The improvements in BP after our program's intervention were similar to previously published findings. Using methods such as text messaging, electronic reminders, and sharing health education links, both SBP and DBP can be reduced [20-22]. However, our project only demonstrated a significant reduction in SBP and not DBP, consistent with the results found by some scholars [41-43]. The reduction in BP may have resulted from the intervention measures we designed based on BCW and the DMIC model, which enhanced patients' knowledge and adherence to healthy behaviors. Several behaviors interacted dynamically and influenced each other, such as increased disease knowledge leading to regular BP monitoring at home [8]. Regular attention to BP can assist individuals with better managing their condition [44], ultimately enhancing BP control. Both the DASH diet and exercise, either alone or in combination, can relax the smooth muscle of blood vessel walls to some extent, promoting blood circulation and consequently

lowering BP [45,46]. Older adults with decreased vascular elasticity may have a tendency for elevated SBP and decreased DBP (as seen in our study with high mean SBP and normal mean DBP) [47]. They may be more sensitive to reductions in SBP. Another possible reason for the reduction in SBP, but not DBP, could be attributed to variations in study populations, interventions, age groups, and medication usage. Regardless, the results of this study were generally consistent with the antihypertensive effects reported in the guidelines [48]. Additionally, the exercise prescription aligned with the guideline recommendations of engaging in at least 30 minutes of moderate-intensity dynamic aerobic exercise per week (such as walking, jogging, cycling, or swimming) and a minimum of 2 to 3 days of resistance training per week. Simultaneously, the formulation of personalized behavior change strategies has further increased patient adherence to the exercise prescription.

SEVR, a reliable indicator of myocardial oxygen supply and demand [49], may be more closely associated with exercise. Our previous study demonstrated an improvement in this indicator after an exercise intervention [13], and other researchers have also observed significant enhancements in SEVR across different age groups (18-80 years) following exercise interventions [50]. Regarding weight, there was a reduction observed in both groups of patients with hypertension when comparing pre- and postintervention data. This reduction aligns with the results of our previous 1-arm before-and-after study [13]. However, the WeChat-based intervention did not demonstrate superiority, possibly due to the limited sample size with this experiment. There were no statistically significant differences observed in heart rate and brachial-ankle pulse wave velocity. This finding contradicted our 1-year study results, which may be attributed to the small sample size and the relatively short duration of the study. Some scholars pointed out that exercise training to improve atherosclerosis requires higher exercise intensity and longer duration, while the participants in this study were older adults with chronic diseases and the exercise prescriptions mostly provided lower exercise intensity [51].

Adherence Indicators

This program enhanced the health knowledge of community-based patients with hypertension and fostered compliance with medication, BP monitoring, exercise, and dietary guidelines, which was consistent with the findings in some previous literature [20-22]. The improvement in knowledge highlighted the efficacy of educating patients with hypertension, serving as the cornerstone of healthy behavior change.

The improvement in medication adherence in our research supports the findings of other research [52-54]. This study indicated that the Morisky scores before and after the intervention (6.77-7.65) were within the moderate range, with slightly better outcomes compared with the results reported by Morawski et al [55] (same assessment tool: 12 weeks, 6-6.3). Medication education and adherence reminders based on video, text messages, and other mobile app functions were some of the most common interventions for medication adherence in cardiovascular diseases [56]. These were key BCTs in our project. Foreman et al [57] suggested that medication text reminders reinforcing medication adherence can lead to higher oral medication adherence among patients with hypertension. Information reminders also help patients maintain higher compliance over time. However, long-term, high-frequency reminders may result in response fatigue in patients [58]. This may explain why our research showed that medication adherence was highest during weeks 2 to 6 (commitment and planning stage) but significantly declined after week 7 (behavioral habit consolidation stage). This highlights the need for researchers to adopt additional behavior change strategies to address patient fatigue during this phase. The effectiveness of electronic reminders and health education in improving patient compliance with BP monitoring has also been confirmed [53]. Based on our research findings, the curve for BPMA showed the least fluctuation. This suggests that BPMA is relatively stable for patients. It is crucial to emphasize the importance of BP monitoring to patients and establish corresponding health behaviors, particularly during the initial phase of the intervention. Additionally, monitoring BP is often the first health behavior action adopted by patients, and it plays a pivotal role in the transition to other health behaviors [44]. Therefore, it should receive special attention during the initial stages.

Most studies have consistently demonstrated the impact of various exercise types and intensities on BP improvement. Research involving mHealth or telehealth interventions through smartphones for patients with coronary heart disease or hypertension has also reported their positive effects on exercise compliance [22,59], which were consistent with our results. Our research revealed a significant rise and reduction in the exercise prescription adherence curve. This indicator was highest during weeks 2 to 6, but it significantly declined starting from week 7. This suggests that emphasis should be placed on maintaining exercise habits. The effectiveness of DASH in improving BP is unquestionable. However, the results of a systematic review indicated only weak evidence supporting the use of smartphone apps to enhance DASH dietary adherence and reduce BP [60]. From our study's perspective, there was significant fluctuation in DASH dietary adherence. This variability may be attributed to the complexity of dietary management compared with other health behaviors, making it difficult to implement and provide feedback [61]. This poses a greater challenge for researchers in designing interventions to improve dietary compliance among patients with hypertension.

In this study, considering the decline in participant adherence indicators after a certain period of intervention, effective strategies may help alleviate participant adherence fatigue, thereby sustaining and enhancing patient engagement. For exercise adherence, research indicates that the effectiveness of interventions is not necessarily correlated with longer intervention periods or higher frequencies. Therefore, tailoring interventions to individual preferences, using different proven therapeutic intervention types for specific target populations, maintaining intervention frequencies above once per week, and ensuring a moderate planned duration may be crucial factors in promoting intervention adherence [62]. Additionally, involving professionals from different disciplines (such as psychologists, doctors, and nurses), having professionals supervise the

implementation of intervention plans, and actively engaging in social interactions with staff and other participants have proven effective in enhancing participant adherence and increasing engagement [62]. Regular home visits [63], actively implementing strategies to increase participant self-efficacy [64], and incorporating gamification elements into interventions [65] are also effective strategies to address participant compliance fatigue. During the intervention implementation process, dynamically assessing and understanding the drivers and barriers of adherence based on the participant's stage and providing personalized decision support and motivation can effectively enhance participant adherence and engagement [66,67].

Notably, there is a lack of research on behavior change interventions for hypertension based on the BCW and DMIC. Therefore, a more intricate experimental design and thorough investigation are required to understand the precise mechanisms underlying the effectiveness of this project, including which components and specific BCTs are effective.

Limitations

This study has several limitations. First, we were unable to implement blinding for the personnel involved in the hypertensive health behavior interventions. To mitigate this, we established standardized intervention procedures, provided intervention strategies tailored to different patient types, edited intervention guidance language based on BCTs, incorporated the aforementioned content into the intervention manual, and conducted training for all intervention personnel. Second, the generalizability of our trial's results may be limited for populations of patients with hypertension residing elsewhere, as they may possess sociodemographic and comorbidity characteristics distinct from those of our study participants. Third, some measurement indicators relied on patient self-report, which could potentially affect the credibility of the results. Fourth, considering the workload, this study did not assess changes in compliance indicators in the control group nor did it collect more comprehensive dietary-related information. This requires correction in future experiments of app-based hypertension health behavior interventions.

Conclusions

The observations suggest that our program may have improved specific health outcomes and adherence to health behaviors in older adults with hypertension. In terms of health outcomes, participants observed significant improvements in SBP, SEVR, and weight. Moreover, there were noteworthy changes in adherence indicators, such as exercise duration, medication adherence, PA duration, frequency of BP monitoring, and learning performance. However, due to our small sample size and short intervention duration, a larger sample size and longer randomized controlled trial are needed to validate the intervention's effects, explore its mechanisms, and identify the specific design elements that are effective. Additionally, among the 4 adherence behaviors, dietary adherence is the most susceptible to external influences, and more BCTs targeting dietary adherence should be considered in intervention design.

Acknowledgments

This work was supported by the Demonstration of Integrated Application of Community Science and Fitness in the Central Region (grant number 2022YFC2010200), Anhui Provincial Department of Science and Technology (grant number 202103a07020004), and Anhui Provincial Department of Education (grant numbers SK2021A0432, 2022lhpysfjd063, and gxyq2022040). The funding source had no role in the study design, data collection, management, data analysis, data interpretation, report writing, or the decision to submit for publication. TS extends gratitude to the scale developer, Mr Donald E Morisky, for generously granting the license to utilize the MMAS-8 (Morisky Medication Adherence Scale 8-item) for this study (Certificate Number: 8538-1877-1559-6025-5310). The MMAS-4 Scale, content, name, and trademarks are protected by US copyright and trademark laws. Permission for use of the scale and its coding is required. A license agreement is available from MMAR, LLC, www.moriskyscale.com. The authors would also like to thank Huaqing Liu and Lihua Zhou for preparing the simplified Dietary Approaches to Stop Hypertension (DASH) grading diet index score.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Consort E-health V1.6.
[PDF File (Adobe PDF File), 3003 KB - mhealth_v12i1e53006_app1.pdf]

References

1. NCD Risk Factor Collaboration (NCD-RisC). Worldwide trends in hypertension prevalence and progress in treatment and control from 1990 to 2019: a pooled analysis of 1201 population-representative studies with 104 million participants. Lancet 2021 Sep 11;398(10304):957-980 [FREE Full text] [doi: 10.1016/S0140-6736(21)01330-1] [Medline: 34450083]
2. The Writing Committee of the Report on Cardiovascular Health and Diseases in China. Report on Cardiovascular Health and Diseases in China 2022: an Updated Summary. Biomed Environ Sci 2023 Aug 20;36(8):669-701 [FREE Full text] [doi: 10.3967/bes2023.106] [Medline: 37711081]

3. Hamrahian SM. Medication non-adherence: a major cause of resistant hypertension. Curr Cardiol Rep 2020 Sep 10;22(11):133. [doi: 10.1007/s11886-020-01400-3] [Medline: 32910342]
4. Wu J, Napoleone J, Linke S, Noble M, Turken M, Rakotz M, et al. Long-term results of a digital hypertension self-management program: retrospective cohort study. JMIR Cardio 2023 Aug 24;7:e43489 [FREE Full text] [doi: 10.2196/43489] [Medline: 37463311]
5. Edwards JJ, Deenmamode AHP, Griffiths M, Arnold O, Cooper NJ, Wiles JD, et al. Exercise training and resting blood pressure: a large-scale pairwise and network meta-analysis of randomised controlled trials. Br J Sports Med 2023 Oct;57(20):1317-1326. [doi: 10.1136/bjsports-2022-106503] [Medline: 37491419]
6. Etminani K, Tao Engström A, Göransson C, Sant'Anna A, Nowaczyk S. How behavior change strategies are used to design digital interventions to improve medication adherence and blood pressure among patients with hypertension: systematic review. J Med Internet Res 2020 Apr 09;22(4):e17201 [FREE Full text] [doi: 10.2196/17201] [Medline: 32271148]
7. Ho JK, Carnagarin R, Matthews VB, Schlaich MP. Self-monitoring of blood pressure to guide titration of antihypertensive medication-a new era in hypertension management? Cardiovasc Diagn Ther 2019 Feb;9(1):94-99 [FREE Full text] [doi: 10.21037/cdt.2018.08.01] [Medline: 30881886]
8. Rahimlou M, Grau N, Banaie-Jahromi N, Taheri M, Khosravi A, Mavrommatis Y, et al. Association of adherence to the dietary approach to stop hypertension and Mediterranean diets with blood pressure in a non-hypertensive population: Results from Isfahan Salt Study (ISS). Nutr Metab Cardiovasc Dis 2022 Jan;32(1):109-116. [doi: 10.1016/j.numecd.2021.09.029] [Medline: 34893410]
9. Verma N, Rastogi S, Chia Y, Siddique S, Turana Y, Cheng H, et al. Non-pharmacological management of hypertension. J Clin Hypertens (Greenwich) 2021 Jul;23(7):1275-1283 [FREE Full text] [doi: 10.1111/jch.14236] [Medline: 33738923]
10. Li ZC, He LH. The application of "hospital-community-family" three-dimensional information management in the follow-up of hypertension in community. Chin Primary Health Care 2018;32(7):4-6 [FREE Full text]
11. Wang S. Effect evaluation on comprehensive intervention of hypertension under family physician model. China Med Pharm 2018;8(22):254-257.
12. Xu J, Wei Y, Ye J. Informatization of exercise prescription and its application. Sport Sci Res 2023;44(1):45-48+54.
13. Sun T, Xu Y, Xie H, Ma Z, Wang Y. Intelligent personalized exercise prescription based on an eHealth promotion system to improve health outcomes of middle-aged and older adult community dwellers: pretest-posttest study. J Med Internet Res 2021 May 24;23(5):e28221 [FREE Full text] [doi: 10.2196/28221] [Medline: 34028359]
14. Ning Z, Li S, Ma S. Development of smart body assessment and health promotion system for middle-aged and elderly people. Proceedings of the Third Academic Exchange Conference on National Fitness and Scientific Exercise 2021:69-70. [doi: 10.26914/c.cnkihy.2021.048574]
15. Xiong S, Berkhouse H, Schooler M, Pu W, Sun A, Gong E, et al. Effectiveness of mHealth interventions in improving medication adherence among people with hypertension: a systematic review. Curr Hypertens Rep 2018 Aug 07;20(10):86. [doi: 10.1007/s11906-018-0886-7] [Medline: 30088110]
16. Han H, Guo W, Lu Y, Wang M. Effect of mobile applications on blood pressure control and their development in China: a systematic review and meta-analysis. Public Health 2020 Aug;185:356-363. [doi: 10.1016/j.puhe.2020.05.024] [Medline: 32738577]
17. Yang P, Zhu X, Chen Z. Research advances in hypertension management based on mobile healthcare. Chin J Hypertens 2020;28(5):415-419.
18. Behaviour change: digital and mobile health interventions. NICE. 2020 Oct 7. URL: https://www.nice.org.uk/guidance/NG183 [accessed 2023-09-17]
19. Jamaladin H, van de Belt TH, Luijpers LC, de Graaff FR, Bredie SJ, Roeleveld N, et al. Mobile apps for blood pressure monitoring: systematic search in app stores and content analysis. JMIR Mhealth Uhealth 2018 Nov 14;6(11):e187 [FREE Full text] [doi: 10.2196/mhealth.9888] [Medline: 30429116]
20. Bingham JM, Black M, Anderson EJ, Li Y, Toselli N, Fox S, et al. Impact of telehealth interventions on medication adherence for patients with type 2 diabetes, hypertension, and/or dyslipidemia: a systematic review. Ann Pharmacother 2021 May;55(5):637-649. [doi: 10.1177/1060028020950726] [Medline: 32815400]
21. Thom SJM, Sivakumar B, Ayodele T, Tan MC, Brown JM, Arcand J. Impact of mHealth interventions on supporting dietary adherence in cardiovascular disease: a systematic review. J Nutr Educ Behav 2023 Jun;55(6):419-436. [doi: 10.1016/j.jneb.2023.03.004] [Medline: 37097263]
22. Bozorgi A, Hosseini H, Eftekhar H, Majdzadeh R, Yoonessi A, Ramezankhani A, et al. The effect of the mobile "blood pressure management application" on hypertension self-management enhancement: a randomized controlled trial. Trials 2021 Jun 24;22(1):413 [FREE Full text] [doi: 10.1186/s13063-021-05270-0] [Medline: 34167566]
23. Alzahrani S, Bin Muammar MF, Bin Muammar AF, Alolah A, Almutawa M. The adoption and acceptance of mHealth interventions for self-management of hypertension among adult patients: a systematic review. Cureus 2022 Nov;14(11):e31584 [FREE Full text] [doi: 10.7759/cureus.31584] [Medline: 36540533]
24. Li R, Liang N, Bu F, Hesketh T. The effectiveness of self-management of hypertension in adults using mobile health: systematic review and meta-analysis. JMIR Mhealth Uhealth 2020 Mar 27;8(3):e17776 [FREE Full text] [doi: 10.2196/17776] [Medline: 32217503]

25. Koetsier LW. The development and implementation of integrated care for childhood overweight and obesity in the Netherlands PhD-Thesis - Research and graduation internal. Vrije Universiteit Amsterdam. 2023 Jun 22. URL: https://research.vu.nl/en/publications/the-development-and-implementation-of-integrated-care-for-childho [accessed 2024-03-15]
26. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011 Apr 23;6:42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
27. Cai L, Shan Y, Du L, Li Y, Zhou Y, Jiang T. Overview and practice of the behavior change wheel theory in foreign contexts. Nurs J Chin People Liberation Army 2019;36(7):59-62.
28. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95 [FREE Full text] [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
29. Barber CEH, Spencer N, Bansback N, Zimmermann GL, Li LC, Richards DP, et al. Development of an implementation strategy for patient decision aids in rheumatoid arthritis through application of the behavior change wheel. ACR Open Rheumatol 2021 May;3(5):312-323 [FREE Full text] [doi: 10.1002/acr2.11250] [Medline: 33793083]
30. Daoud N, Jung YE, Sheikh Muhammad A, Weinstein R, Qaadny A, Ghattas F, et al. Facilitators and barriers to smoking cessation among minority men using the behavioral-ecological model and Behavior Change Wheel: A concept mapping study. PLoS One 2018;13(10):e0204657 [FREE Full text] [doi: 10.1371/journal.pone.0204657] [Medline: 30356254]
31. Bai I, Isenor JE, Reeve E, Whelan AM, Martin-Misener R, Burgess S, et al. Using the behavior change wheel to link published deprescribing strategies to identified local primary healthcare needs. Res Social Adm Pharm 2022 Aug;18(8):3350-3357. [doi: 10.1016/j.sapharm.2021.12.001] [Medline: 34895842]
32. Oliwa JN, Nzinga J, Masini E, van Hensbroek MB, Jones C, English M, et al. Improving case detection of tuberculosis in hospitalised Kenyan children-employing the behaviour change wheel to aid intervention design and implementation. Implement Sci 2020 Nov 25;15(1):102 [FREE Full text] [doi: 10.1186/s13012-020-01061-4] [Medline: 33239055]
33. Gomez R, Habib A, Maidment DW, Ferguson MA. Smartphone-connected hearing aids enable and empower self-management of hearing loss: a qualitative interview study underpinned by the behavior change wheel. Ear Hear 2022;43(3):921-932. [doi: 10.1097/AUD.0000000000001143] [Medline: 34653030]
34. Baumel A, Fleming T, Schueller SM. Digital micro interventions for behavioral and mental health gains: core components and conceptualization of digital micro intervention care. J Med Internet Res 2020 Oct 29;22(10):e20631 [FREE Full text] [doi: 10.2196/20631] [Medline: 33118946]
35. Sun T, Zhao H, Ding Z, Xie H, Ma L, Zhang Y, et al. Evaluating a WeChat-based health behavioral digital intervention for patients with hypertension: protocol for a randomized controlled trial. JMIR Res Protoc 2023 Sep 12;12:e46883 [FREE Full text] [doi: 10.2196/46883] [Medline: 37698909]
36. Mao Y, Lin W, Wen J, Chen G. Impact and efficacy of mobile health intervention in the management of diabetes and hypertension: a systematic review and meta-analysis. BMJ Open Diabetes Res Care 2020 Sep;8(1):1 [FREE Full text] [doi: 10.1136/bmjdrc-2020-001225] [Medline: 32988849]
37. Morisky DE, Ang A, Krousel-Wood M, Ward HJ. Predictive validity of a medication adherence measure in an outpatient setting. J Clin Hypertens (Greenwich) 2008 May;10(5):348-354 [FREE Full text] [doi: 10.1111/j.1751-7176.2008.07572.x] [Medline: 18453793]
38. Berlowitz DR, Foy CG, Kazis LE, Bolin LP, Conroy MB, Fitzpatrick P, SPRINT Research Group. Effect of intensive blood-pressure treatment on patient-reported outcomes. N Engl J Med 2017 Aug 24;377(8):733-744 [FREE Full text] [doi: 10.1056/NEJMoa1611179] [Medline: 28834483]
39. Bress AP, Bellows BK, King JB, Hess R, Beddhu S, Zhang Z, SPRINT Research Group. Cost-effectiveness of intensive versus standard blood-pressure control. N Engl J Med 2017 Aug 24;377(8):745-755 [FREE Full text] [doi: 10.1056/NEJMsa1616035] [Medline: 28834469]
40. Ainsworth BE, Haskell WL, Whitt MC, Irwin ML, Swartz AM, Strath SJ, et al. Compendium of physical activities: an update of activity codes and MET intensities. Med Sci Sports Exerc 2000 Sep;32(9 Suppl):S498-S504. [doi: 10.1097/00005768-200009001-00009] [Medline: 10993420]
41. Zhou L, He L, Kong Y, Lai Y, Dong J, Ma C. Effectiveness of mHealth interventions for improving hypertension control in uncontrolled hypertensive patients: A meta-analysis of randomized controlled trials. J Clin Hypertens (Greenwich) 2023 Jul 06;25(7):591-600 [FREE Full text] [doi: 10.1111/jch.14690] [Medline: 37409556]
42. Aovare P, Abdulai K, Laar A, van der Linden EL, Moens N, Richard E, et al. Assessing the effectiveness of mHealth interventions for diabetes and hypertension management in Africa: systematic review and meta-analysis. JMIR Mhealth Uhealth 2023 Aug 29;11:e43742-e43742 [FREE Full text] [doi: 10.2196/43742] [Medline: 37646291]
43. Zha P, Qureshi R, Porter S, Chao Y, Pacquiao D, Chase S, et al. Utilizing a mobile health intervention to manage hypertension in an underserved community. West J Nurs Res 2020 Mar;42(3):201-209. [doi: 10.1177/0193945919847937] [Medline: 31057081]
44. Wu D, Huyan X, She Y, Hu J, Duan H, Deng N. Exploring and characterizing patient multibehavior engagement trails and patient behavior preference patterns in pathway-based mHealth hypertension self-management: analysis of use data. JMIR Mhealth Uhealth 2022 Feb 03;10(2):e33189 [FREE Full text] [doi: 10.2196/33189] [Medline: 35113032]

45. Nystoriak MA, Bhatnagar A. Cardiovascular effects and benefits of exercise. Front Cardiovasc Med 2018;5:135 [FREE Full text] [doi: 10.3389/fcvm.2018.00135] [Medline: 30324108]
46. Akhlaghi M. Dietary Approaches to Stop Hypertension (DASH): potential mechanisms of action against risk factors of the metabolic syndrome. Nutr. Res. Rev 2019 Jul 30;33(1):1-18. [doi: 10.1017/s0954422419000155]
47. Laurent S, Boutouyrie P. Arterial stiffness and hypertension in the elderly. Front Cardiovasc Med 2020 Oct 29;7:544302 [FREE Full text] [doi: 10.3389/fcvm.2020.544302] [Medline: 33330638]
48. Williams B, Mancia G, Spiering W, Agabiti Rosei E, Azizi M, Burnier M, ESC Scientific Document Group. 2018 ESC/ESH Guidelines for the management of arterial hypertension. Eur Heart J 2018 Sep 01;39(33):3021-3104 [FREE Full text] [doi: 10.1093/eurheartj/ehy339] [Medline: 30165516]
49. Murakami T, Takeda A, Takei K, Tateno S, Kawasoe Y, Niwa K. The cardiac blood supply-workload balance in children. Heart Vessels 2015 Sep 3;30(5):626-631 [FREE Full text] [doi: 10.1007/s00380-014-0537-0] [Medline: 24989969]
50. Radhakrishnan J, Swaminathan N, Pereira NM, Henderson K, Brodie DA. Acute changes in arterial stiffness following exercise in people with metabolic syndrome. Diabetes Metab Syndr 2017;11(4):237-243. [doi: 10.1016/j.dsx.2016.08.013] [Medline: 27575048]
51. Park J, Choi Y, Myoenzono K, Yoshikawa T, Tagawa K, Isobe T, et al. Effects of aerobic exercise training on the arterial stiffness and intramyocellular or extramyocellular lipid in overweight and obese men. Clin Exp Hypertens 2020 May 18;42(4):302-308. [doi: 10.1080/10641963.2019.1649686] [Medline: 31392903]
52. Xu H, Long H. The effect of smartphone app-based interventions for patients with hypertension: systematic review and meta-analysis. JMIR Mhealth Uhealth 2020 Oct 19;8(10):e21759 [FREE Full text] [doi: 10.2196/21759] [Medline: 33074161]
53. Chandler J, Sox L, Kellam K, Feder L, Nemeth L, Treiber F. Impact of a culturally tailored mHealth medication regimen self-management program upon blood pressure among hypertensive Hispanic adults. Int J Environ Res Public Health 2019 Apr 06;16(7):1 [FREE Full text] [doi: 10.3390/ijerph16071226] [Medline: 30959858]
54. Zare S, Rezaee R, Aslani A, Shirdeli M, Kojuri J. Moving toward community based telehealth services using mhealth for hypertensive patients. Int J Technol Assess Health Care 2019 Sep 24;35(5):379-383. [doi: 10.1017/s0266462319000655]
55. Morawski K, Ghazinouri R, Krumme A, Lauffenburger JC, Lu Z, Durfee E, et al. Association of a smartphone application with medication adherence and blood pressure control: the MedISAFE-BP randomized clinical trial. JAMA Intern Med 2018 Jun 01;178(6):802-809 [FREE Full text] [doi: 10.1001/jamainternmed.2018.0447] [Medline: 29710289]
56. Yeung DL, Alvarez KS, Quinones ME, Clark CA, Oliver GH, Alvarez CA, et al. Low-health literacy flashcards and mobile video reinforcement to improve medication adherence in patients on oral diabetes, heart failure, and hypertension medications. J Am Pharm Assoc (2003) 2017;57(1):30-37 [FREE Full text] [doi: 10.1016/j.japh.2016.08.012] [Medline: 27816544]
57. Foreman KF, Stockl KM, Le LB, Fisk E, Shah SM, Lew HC, et al. Impact of a text messaging pilot program on patient medication adherence. Clin Ther 2012 May;34(5):1084-1091. [doi: 10.1016/j.clinthera.2012.04.007] [Medline: 22554973]
58. Finitsis DJ, Pellowski JA, Johnson BT. Text message intervention designs to promote adherence to antiretroviral therapy (ART): a meta-analysis of randomized controlled trials. PLoS One 2014;9(2):e88166 [FREE Full text] [doi: 10.1371/journal.pone.0088166] [Medline: 24505411]
59. Song Y, Ren C, Liu P, Tao L, Zhao W, Gao W. Effect of smartphone-based telemonitored exercise rehabilitation among patients with coronary heart disease. J Cardiovasc Transl Res 2020 Aug;13(4):659-667 [FREE Full text] [doi: 10.1007/s12265-019-09938-6] [Medline: 31820334]
60. Alnooh G, Alessa T, Hawley M, de Witte L. The use of Dietary Approaches to Stop Hypertension (DASH) mobile apps for supporting a healthy diet and controlling hypertension in adults: systematic review. JMIR Cardio 2022 Nov 02;6(2):e35876 [FREE Full text] [doi: 10.2196/35876] [Medline: 36322108]
61. Hever J, Cronise R. Plant-based nutrition for healthcare professionals: implementing diet as a primary modality in the prevention and treatment of chronic disease. J Geriatr Cardiol 2017 May;14(5):355-368 [FREE Full text] [doi: 10.11909/j.issn.1671-5411.2017.05.012] [Medline: 28630615]
62. Collado-Mateo D, Lavín-Pérez AM, Peñacoba C, Del Coso J, Leyton-Román M, Luque-Casado A, et al. Key factors associated with adherence to physical exercise in patients with chronic diseases and older adults: an umbrella review. Int J Environ Res Public Health 2021 Feb 19;18(4):2023 [FREE Full text] [doi: 10.3390/ijerph18042023] [Medline: 33669679]
63. Zhang L, Chen C, Zheng C, Zhao M, Jiang J. Implementing home visit to improve treatment compliance of community hypertension patients. Chinese Nursing Management 2015;15(2):186-190 [FREE Full text] [doi: 10.5005/jp/books/12577_7]
64. Sun SY, Sun GZ, Gao M, Yu TX, Liu SX, Tang ZJ, et al. Mediating effect of self-efficacy between self-regulatory fatique and exerciseadherence in patients with chronic heart failure. Journal of Nursing 2023;30(22):63-67 [FREE Full text] [doi: 10.16460/j.issn1008-9969.2023.22.063]
65. Ouyang P, Wang J, Ali U. The impact of gamification on the patient's engagement in the online health community. Aslib J Inf Manage 2022 Apr 15;74(6):1196-1213. [doi: 10.1108/ajim-10-2021-0314]
66. Zhu J, Cai W, Tao Z, Ji C, Yuan P, Chen L. Analysis of rehabilitation exercise compliance and its influencing factors in young and middle-aged patients with hemorrhagic stroke. Chin J Nurs 2023;58(5):580-586 [FREE Full text] [doi: 10.3761/j.issn.0254-1769.2023.05.010]

67. Bonetti L, Tolotti A, Anderson G, Nania T, Vignaduzzo C, Sari D, et al. Nursing interventions to promote patient engagement in cancer care: a systematic review. Int J Nurs Stud 2022 Sep;133:104289. [doi: 10.1016/j.ijnurstu.2022.104289] [Medline: 35751947]

Abbreviations

BCT: behavior change technique
BCW: behavior change wheel
BP: blood pressure
BPMA: blood pressure monitoring adherence
COM-B: Capability, Opportunity, Motivation-Behavior
CVI: content validity index
DA: dietary adherence
DASH: Dietary Approaches to Stop Hypertension
DBP: diastolic blood pressure
DMIC: digital micro-intervention care
FITT-VP: frequency, intensity, type, time, volume, and progression
HBDIHP: health behavioral digital intervention for hypertensive patients
I-CVI: item-level content validity index
MET: metabolic equivalent
mHealth: mobile health
PA: physical activity
SBP: systolic blood pressure
SEVR: subendocardial viability ratio

Edited by L Buis; submitted 22.09.23; peer-reviewed by R Cui, E Aigbokhan, M Putilina, Q Lv, X Yang; comments to author 23.10.23; revised version received 22.12.23; accepted 26.02.24; published 05.04.24.

Please cite as:
Sun T, Xu X, Ding Z, Xie H, Ma L, Zhang J, Xia Y, Zhang G, Ma Z
Development of a Health Behavioral Digital Intervention for Patients With Hypertension Based on an Intelligent Health Promotion System and WeChat: Randomized Controlled Trial
JMIR Mhealth Uhealth 2024;12:e53006
URL: https://mhealth.jmir.org/2024/1/e53006
doi:10.2196/53006
PMID:38578692

Original Paper

Effects of the "AI-TA" Mobile App With Intelligent Design on Psychological and Related Symptoms of Young Survivors of Breast Cancer: Randomized Controlled Trial

Lulu Jiang[1], MMed; Jiehui Xu[2], RN, MSN; Yanwei Wu[2], BS, RN; Yanyan Liu[1], PhD; Xiyi Wang[1*], PhD; Yun Hu[1*], PhD

[1]School of Nursing, Shanghai Jiao Tong University, Shanghai, China

[2]Renji Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai, China

[*]these authors contributed equally

Corresponding Author:
Yun Hu, PhD
School of Nursing, Shanghai Jiao Tong University
No.277 Chongqing South Roud, Huangpu District
Shanghai, 200025
China
Phone: 86 021 63546078
Email: huyunsy@shsmu.edu.cn

Abstract

Background: Young women often face substantial psychological challenges in the initial years following cancer diagnosis, leading to a comparatively lower quality of life than older survivors. While mobile apps have emerged as potential interventions, their effectiveness remains inconclusive due to the diversity in intervention types and variation in follow-up periods. Furthermore, there is a particular dearth of evidence regarding the efficacy of these apps' intelligent features in addressing psychological distress with these apps.

Objective: This study aims to evaluate the effectiveness of a mobile app with intelligent design called "*AI-TA*" on cancer-related psychological health and ongoing symptoms with a randomized controlled design.

Methods: Women aged 18 to 45 years diagnosed with breast cancer were randomly assigned to the intervention or control group. The intervention was *AI-TA*, which included 2-way web-based follow-up every 2 weeks. Both intention-to-treat (ITT) and per-protocol (PP) analyses employed repeated measurement analysis of variance. The participants' background features, primary outcomes (psychological distress and frequency, self-efficacy, and social support), and secondary outcomes (quality of life) were measured using multiple instruments at 3 time points (baseline, 1-month intervention, and 3-month intervention).

Results: A total of 124 participants were randomly allocated to the control group (n=62, 50%) or intervention group (n=62, 50%). In total, 92.7% (115/124) of the participants completed the intervention. Significant improvements in psychological symptoms (Memorial Symptom Assessment Scale-Short Form) were observed in the ITT group from baseline to 1-month intervention relative to the control group (ITT vs control: 1.17 vs 1.23; P<.001), which persisted at 3-month follow-up (ITT vs control: 0.68 vs 0.91; P<.001). Both the ITT and PP groups exhibited greater improvements in self-efficacy (Cancer Behavior Inventory-Brief Version) than the control group at 1-month (ITT vs PP vs control: 82.83 vs 77.12 vs 65.35; P<.001) and 3-month intervention (ITT vs PP vs control: 92.83 vs 89.30 vs 85.65; P<.001). However, the change in social support (Social Support Rating Scale) did not increase significantly until 3-month intervention (ITT vs control: 50.09 vs 45.10; P=.002) (PP vs control: 49.78 vs 45.10; P<.001). All groups also experienced beneficial effects on quality of life (Functional Assessment of Cancer Therapy-Breast), which persisted at 3-month follow-up (P<.001).

Conclusions: The intelligent mobile app *AI-TA* incorporating intelligent design shows promise for reducing psychological and cancer-related symptoms among young survivors of breast cancer.

Trial Registration: Chinese Clinical Trial Registry ChiCTR2200058823; https://www.chictr.org.cn/showproj.html?proj=151195

(JMIR Mhealth Uhealth 2024;12:e50783) doi:10.2196/50783

KEYWORDS

mobile app; artificial intelligence; interactivity; breast cancer; psychological symptoms; self-efficacy; social support; quality of life

Introduction

Background

Breast cancer is a significant health concern for women globally, particularly in China, where the incidence and mortality rates have been steadily rising, accounting for 12.2% and 9.6%, respectively, of the total cases in the world [1,2]. In 2020, alone, approximately 416,371 women were newly diagnosed with breast cancer [3]. Moreover, the peak prevalence of breast cancer among Chinese women occurring between 45 and 55 years, which is younger than that of their Western counterparts [2-4].

Life after breast cancer, especially for younger survivors, often entails adverse psychological consequences [5]. Young survivors of breast cancer have greater psychologic morbidity than older women and age-matched women with no cancer history; this includes elevated levels of psychological distress and frequency of persistent disease for at least 2 years after diagnosis [6]. A substantial proportion of younger women experience long-term iatrogenic effects, including fatigue, persistent pain, lymphedema, and infertility, all of which may negatively affect psychological health [7,8]. Other cancer-related symptoms, such as psychosocial maladjustment, have also been reported during both cancer treatment and rehabilitation [9]. A lack of confidence and preparedness to cope with cancer can intensify survivors' distress, hinder their reintegration into society, reduce their self-efficacy, and cause significant impairment in quality of life [10,11]. Understanding the dynamic demands of young survivors of breast cancer is crucial for providing targeted and culturally sensitive support. Our previous research on young survivors of breast cancer indicated that psychological support is desired early in diagnosis, and there is more focus on information provided during treatment [12]. Thus, recognizing the unique characteristics of young survivors of breast cancer is vital for delivering tailored and comprehensive psychosocial care.

Research indicates that young survivors of breast cancer have more complex and dynamic needs and face challenges related to cultural norms, psychological disturbances, and a decreased quality of survivorship [13]. Web-based programs that leverage the accessibility, availability, and cost-effectiveness of the internet have been widely used in breast cancer interventions [14,15]. However, most programs aimed at improving well-being in survivors of cancer are not tailored to the specific functions, components, or characteristics of the target population. Information is often generalized and looped and does not accurately align resources with individual needs, rendering them ineffective for many patients. This means that there is no coordinated, personalized, or supportive care; rather, there is only a 1-way relationship between programs and patients.

Incorporating artificial intelligence (AI) into interventions offers a promising avenue to address these challenges. AI, as a major component of the internet, can enhance technical interventions and interactions through the use of a sophisticated blend of human-computer and human-human techniques [16]. Specifically, AI algorithms can analyze user input and provide tailored advice or support based on the user's history and preferences. This personalized interaction increases user engagement and satisfaction, effectively bridging psychological gaps and facilitating a deeper understanding of user needs. At the humanistic level, AI can significantly enhance communication and collaboration. For example, AI-driven platforms can facilitate social support networks or groups, connecting individuals with similar interests or experiences. This approach is particularly useful in therapeutic contexts or web intervention. In health care, AI can analyze patient data in real time to provide up-to-date, personalized health recommendations or alerts, empowering users to access relevant information and engage in dynamic dialogues [17]. Building on previous studies and feedback from interventionists, young survivors of breast cancer, and health care professionals, we developed an intelligent interactive mobile app called "*AI-TA*" (a WeChat Mini Program) guided by a person-centered care (PCC) framework [12,18,19]. The PCC emphasizes collaborative partnerships between patients and health care providers [20]. It has been shown to enhance patients' convictions to engage in desired activities and take responsibility for disease management and clinical outcomes [21-23]. Informed by our pilot study results, we have made necessary adjustments to the intervention strategy and module design [24]. In this study, we expand the sample size to further enhance the effectiveness and generalizability of the findings, hypothesizing that users of *AI-TA* will experience significant improvements in psychological symptoms.

Objective

The purpose of this study is to comprehensively assess the impact of an innovative mobile app, "AI-TA," which features intelligent design elements, on the psychological health and ongoing symptoms experienced by young survivors of breast cancer.

Methods

Study Design

This study was designed as a multicenter, 3-month, parallel group, single-blind, 2-arm randomized controlled trial conducted in 3 university-affiliated hospitals from January 2022 to December 2022. This study investigated the effectiveness of "*AI-TA*" on psychological and related symptoms of young survivors of breast cancer from baseline (T0) to 2 follow-up points (1 month [T1] and 3 months [T2]; Multimedia Appendix 1). The trial was approved by the Chinese Clinical Trial Registry (ChiCTR2200058823).

Recruitment

Participants were recruited through convenience sampling, aligning with findings from prior studies [25-27], and

considering the menopausal age of women. The inclusion and exclusion criteria for the participants are provided below in Textbox 1.

Textbox 1. Inclusion and exclusion criteria.

Inclusion criteria

- Chinese females
- Aged 18-45 years
- Diagnosed with stage 1-3 breast cancer
- Able to access the internet using computer or mobile devices
- Able to read and write in Chinese (traditional or simplified)
- Provided informed consent

Exclusion criteria

- Chronic or acute physical conditions that significantly impair daily functioning or require intensive medical care and supervision that could detract from intervention participation or measurement of outcomes
- Serious cognitive or communication barriers (including but not limited to medical diagnoses of advanced dementia, severe aphasia, or other neurological conditions significantly impairing understanding or expression)
- Recurrent or metastatic breast cancer
- Concurrent involvement in other studies

Randomization

To ensure the quality of the entire study and prevent selection and information bias, our researchers received unified training and were divided into different group roles: (1) recruiter: 2 breast clinical nurses will strictly recruit young survivors of breast cancer according to the inclusion and exclusion criteria and record recruitment information; (2) an independent master candidate randomly assigned participants to 2-armed parallel groups at a 1:1 ratio via a computer-generated digital sequence; (3) intervener: an experienced researcher conducted the interventions, with another researcher serving as the intervention companion, who was responsible for supervision and evaluation; and (4) data collectors: collect and analyze all participants' data and the results of the intervention by double checking. During the process, blinding was applied to the recruiters and data collectors.

Procedure

The participants in the control and intervention groups received oral and written instructions on how to use the *AI-TA*, which combines a mobile app with fortnightly web-based follow-up. Each participant used her own WeChat ID to register and log in with a unique or random number for access to *AI-TA* and was told not to discuss the research with other patients so that no identifying information would be linked to them and to reduce contamination. At first enrollment, they were required to complete and return electronic questionnaires, which included sociodemographics, cancer-related characteristics, and psychological and accompanying symptoms in *AI-TA* after informed consent was obtained. The initial baseline evaluation of symptoms was classified at T0. Data collection and assessments of outcomes took place over 2 time points in the follow-up period. T1 assessment took place at 1 month after allocation (intermediate period of intervention), and T2 assessment took place at 3 months after allocation (end point of the intervention). Follow-up assessments were collected via *AI-TA*–assisted self-reported surveys. Each result was saved and available for participants to view their own result at any time. In addition, all participants were awarded CNY 100 (US $13.80) upon completing all the assessments. In addition, there was a questionnaire to assess participants' interaction of and satisfaction with the *AI-TA* mobile app program for further improvements in the following research (Multimedia Appendix 2).

Intervention

AI-TA consisted of several modules designed to support young survivors of breast cancer in various aspects of their survivorship. The mobile app stored reliable resources uploaded by health care professionals occasionally from time to time and covered psychological counseling, coping effectiveness, symptom management, social security, etc in text, image, and animation formats. It also allowed participants to synchronously save their log-in, comments, likes, history, duration, and traces. In addition, the health care professionals invited breast clinical experts to hold salon lectures, focusing on common problems in treatments involving diet guidance, functional exercise, tube maintenance, and other guidance. Question and answer sessions were incorporated, and recorded videos were made available for review. Visual representations of *AI-TA* are shown in Figure 1. Additional details relating to the intervention construction can be found in Figure 2.

Figure 1. Visual representations of AI-TA.

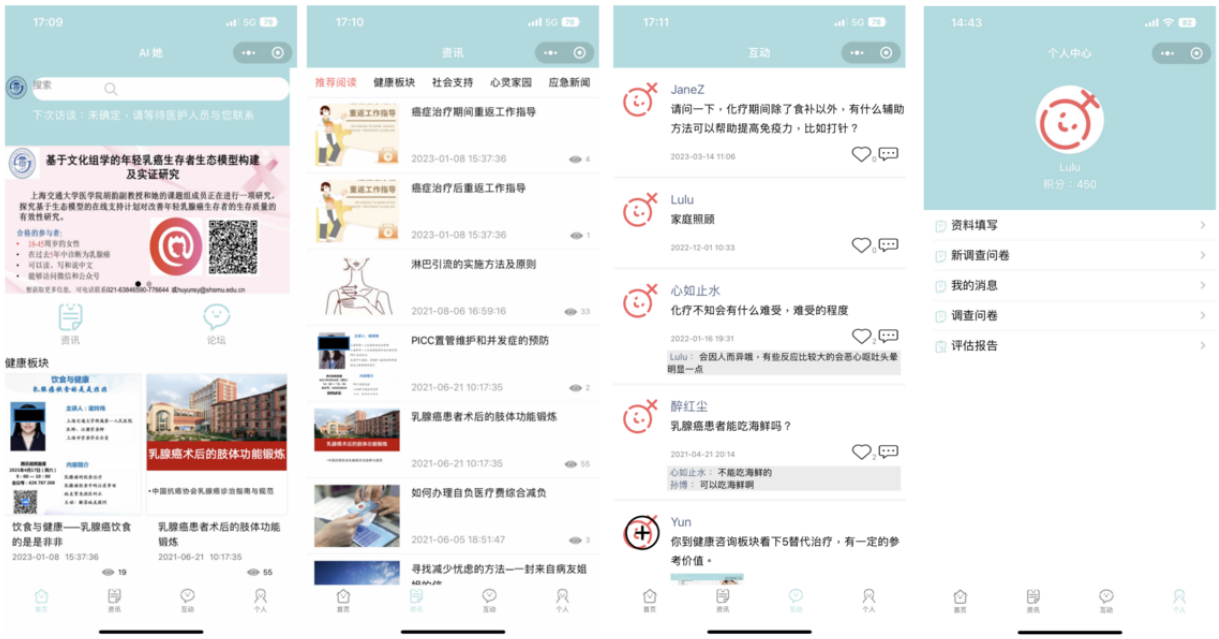
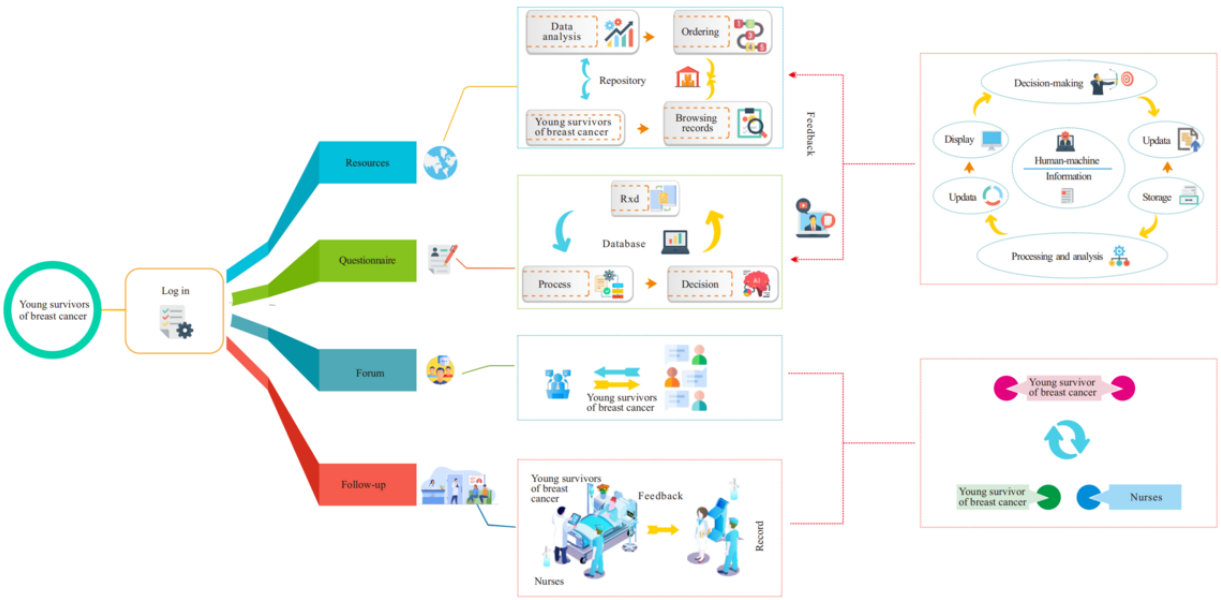


Figure 2. Frame diagram for the intervention design.



AI-Driven System

An AI-driven system is central to our intervention, using text extraction techniques and behavioral data analysis to provide personalized recommendations. This innovative approach significantly enhances user engagement and retention by tailoring content to individual preferences and interests. The following functions were undertaken: (1) personalized content delivery (the AI system analyzes responses from questionnaires to prioritize issues and deliver tailored content; for instance, if young survivors of breast cancer reported low physical activity and poor sleep quality, keywords such as “physical,” “exercise,” “activity,” and “sleep” were used to recommend relevant articles), (2) symptom tracking and management (the system regularly tracks survivors’ symptoms and allows health care professionals to update and tailor content based on their browsing preferences, enhancing the relevance and effectiveness of the information provided), (3) data monitoring and evaluation

(AI algorithms can calculate and monitor assessment progress as well as synchronous storage in the server backend), (4) social support network (AI can enable the formation of connections in forums with independent YBCS; it fosters a supportive web-based environment through interactive question and answer sessions, experience sharing, and emotional support), and (5) privacy and security (adhering to health care data regulations, the AI system uses encryption measures to ensure the utmost privacy and security of user data; AI-driven system is a vital link that enables the exchange and feedback of information in peer-to-peer interactions between young survivors of breast cancer and health care professionals and enhances the interactive experience). This dynamic approach ensures that the intervention remains relevant, engaging, and supportive of the unique needs of survivors.

2-Way Web-Based Follow-Up

Furthermore, the intervention program included fortnightly web-based follow-up using 2-way communication through private messages or calls. This approach encouraged narration from young survivors of breast cancer and aimed to establish a partnership using PCC communication skills such as open-ended questions, reflections, and summaries. In the initial conversation, health care professionals focused on listening to survivors' narratives about daily life events and customs (diet, motion, pressure, hobbies, relationships, and sharing) to build trust relationships. The subsequent step entailed the anticipation and cocreation of a health plan jointly based on their feedback through discussion and agreement, including goals, resources, and needs. The contents regarding what participants had talked about, how they felt, what goals they had, and what they wanted to accomplish will be the points for the forthcoming conversations to consider. During the 3-month intervention, participants were also free to get in touch with the health care professionals during office hours. Each follow-up was recorded and uploaded to the platform.

General Information Support

The control group was granted access to general information on the mobile app, with all modules available in *AI-TA* except for forums and intelligent recommendations. This meant that they could not participate in the forum and obtain recommendations provided by the system based on questionnaire results. In addition, young survivors of breast cancer in the control group also had no follow-up conversations.

Measures and Instruments

A comprehensive set of questions was used to assess participants' sociodemographic and health characteristics, including age, height, weight, habitation, educational attainment, marital status, employment, income, offspring, parent, cancer stage (stages 1-3), cancer type, diagnosis time, and treatment.

Primary Outcome Measures

Psychological Distress and Frequency

The Memorial Symptom Assessment Scale-Short Form (MSAS-SF) was used to assess the frequency and severity of psychological symptoms during the past 7 days [28]. The distress level of each symptom was rated on a 5-point Likert scale (0="not at all," 1="a little bit," 2="somewhat," 3="quite a bit," and 4="very much"). If the symptom was not present, a value of 0 was assigned. The frequency of psychological symptoms is rated from 1 to 4 (1="rarely" to 4="almost constantly").

Self-Efficacy

The Cancer Behavior Inventory-Brief Version (CBI-B) developed by Heitzmann et al [29] was adopted to rate self-efficacy for coping with cancer. It is used to assess four factors: (1) maintaining independence and positive attitude, (2) participating in medical care, (3) coping and stress management, and (4) managing affect. There are 12 items in total (rated on a 9-point Likert scale, ranging from 1="not at all confident" to 9="totally confident"); a score≤36 is considered low, a score between 37 and 72 is considered moderate, and a score between 73 and 108 is considered high.

Social Support

The Social Support Rating Scale (SSRS) is a 10-item questionnaire developed by Xiao [30] for measuring social support, including objective social support, subjective social support, and use of social support. A higher score indicated more social support. The SSRS has been widely used and has shown acceptable reliability and validity in the cancer population. An SSRS score≤22 is considered poor social support, a score between 23 and 44 is considered moderate social support, and a score between 45 and 66 indicates adequate social support.

Secondary Outcome Measures

The Functional Assessment of Cancer Therapy-Breast (FACT-B) translated and adapted by Wan et al [31] was used to evaluate the quality of life of patients with breast cancer. The 5 dimensions included physical well-being, social or family well-being, emotional well-being, functional well-being, and additional concerns (cancer type–specific questions). A total of 36 items were scored on a 5-point Likert scale (0="not at all," 1="a little bit," 2="somewhat," 3="quite a bit," and 4="very much"). Among them, 19 items were scored in a reverse manner. Higher scores represent better quality of life.

Sample Size

This study used G*Power (version 3.1; HHU) to calculate the necessary sample size. On the basis of a similarly designed study, mobile app support reduced the psychological symptoms among survivors of breast cancer with an effect size of 0.77 [32]. Considering the conservative estimate and the variability of previous pilot research [24] as well as the statistical power [33], we estimated that 66 participants were needed to compare between-group differences and present a large effect size (d=0.8) in the primary outcome after intervention, with an α level of .05 (2-sided test), 80% statistical power, 1:1 allocation rate, and 20% attrition rate. Thus, a final sample of 124, with 62 (50%) individuals in each group, was adequate.

Ethical Considerations

The trial complied with the ethical guidelines of the Declaration of Helsinki and was approved by the ethics committee of Public Health and Nursing Research, School of Medicine, Shanghai

Jiao Tong University (SJTUPN-201803). All participants provided electronic informed consent before enrollment in the study. All data and information were anonymized according to the established guidelines, and a password-protected document containing participants' personal information was stored on secure servers.

Statistical Analysis

Data analysis was performed using SPSS (version 26.0; IBM Corp). Descriptive and comparative statistics were used to characterize the study groups (eg, percentage or mean and SD). A total of 2-sample *t* tests (2-tailed) and chi-square or Fisher exact tests were used where appropriate, and these tests assessed demographic variable differences between the intervention and control groups. Before performing the *t* test, the continuous variables were checked for normality using the Shapiro-Wilk test, and all the data were revealed to be normal ($P>.05$). To confirm the improvements in psychological symptoms, the baseline and postintervention results of the dependent variables were analyzed using the paired *t* tests, whereas the 2-sample *t* tests were used to detect differences between the intervention and control groups at each time point. To estimate the effects of the intervention on the outcomes over time, a linear mixed effect model for repeated measurements was performed. The main effects of group, time, and group×time interaction effects were examined. The significance level was set at $P<.05$ (2 sided).

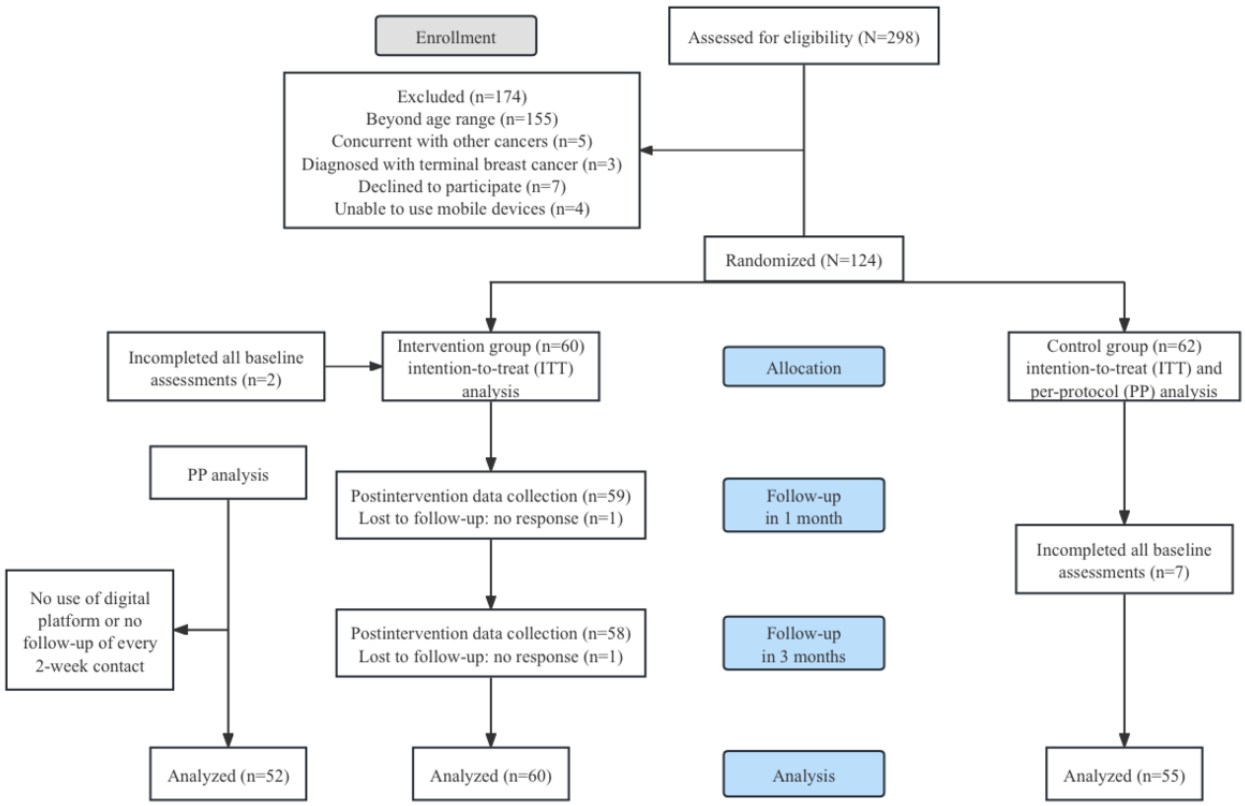
For this study, both intention-to-treat (ITT) and per-protocol (PP) analyses were conducted. The primary analysis used an ITT approach, which can reflect the results of all participants randomly assigned to receive intervention; missing fields were imputed with the expectation-maximization algorithm. Post hoc sensitivity analyses for missing data were performed to ensure the integrity and reliability of the trial outcomes (Multimedia Appendix 3 [24]). The PP group analysis included participants who fully followed the intervention protocol. The primary end points for evaluating the efficacy of *AI-TA* were MSAS-SF, CBI-B, and SSRS to assess psychological symptoms (distress and frequency), self-efficacy, and social support, respectively, at T2. Secondary end point was FACT-B measures of quality of life.

Results

Participants

Data were collected through questionnaires at T0, T1, and T2. Approximately 7.3% (9/124) of participants (2/62, 3% in the intervention group and 7/62, 11% in the control group) did not complete all baseline assessment at T0. At T1 and T2, 1.7% (2/115) of participants did not return their questionnaires despite being reminded. A flowchart of the study participants is given in Figure 3.

Figure 3. A flowchart of the study participants.



Overview

Table 1 provides an overview of the sociodemographic and health characteristics of young survivors of breast cancer at T0. In this study, the participants had an average age of 40.21 (SD 4.24) years, a mean height of 161.16 (SD 4.24) cm, and a mean weight of 55.9 (SD 8.56) kg. The average time since diagnosis

was 1.42 (SD 0.3) years. Approximately 89.6% (103/115) participants were living in urban areas and about 90.4% (104/115) were married. Approximately 93.9% (108/115) had children, and 81.7% (94/115) of the participants' parents were in good health. A total of >50% were employed, and their monthly income was >CNY 10,000 (US $1379.44). In addition, participants tended to be highly educated; approximately 50.4% (58/115) were college graduates and some had graduate degrees. Almost 94.8% (109/115) had invasive breast cancer, and approximately 65.2% (75/115) were diagnosed with stage 2 or stage 3 breast cancer. A total of 11.3% (13/115) only underwent surgical treatment, 42.6% (49/115) only received adjuvant treatment, and 46.1% (53/115) had both. No significant differences were observed in the demographic characteristics between the intervention and control groups, except that the time since diagnosis of participants in the control group was significantly longer than that in the intervention group (P=.03). However, this difference did not remain when comparing the control and PP groups (P=.18).

Table 1. Participant characteristics at baseline (N=115).

Characteristics	Control group (n=55)	ITT[a] group (n=60)	ITT group, *P* value	PP[b] group (n=52)	PP group, *P* value
Age (y), mean (SD)	40.02 (4.32)	40.40 (4.16)	.68	40.66 (4.34)	.27
Height (cm), mean (SD)	160.55 (4.36)	161.76 (4.12)	.19	161.71 (4.43)	.32
Weight (kg), mean (SD)	55.37 (7.51)	56.43 (9.61)	.57	56.60 (9.56)	.86
Diagnosis (y), mean (SD)	1.53 (0.34)	1.31 (0.26)	.03	1.50 (0.29)	.18
Residence, n (%)			.62		.69
Urban	48 (87.3)	55 (91.7)		50 (96.2)	
Rural	7 (12.7)	5 (8.3)		2 (3.8)	
Marital status, n (%)			.74		.09
Married	51 (92.7)	53 (88.3)		48 (92.4)	
Single	3 (5.5)	5 (8.3)		2 (3.8)	
Other	1 (1.8)	2 (3.4)		2 (3.8)	
Have a child, n (%)			.62		.63
Yes	52 (94.5)	56 (93.3)		49 (94.2)	
No	3 (5.5)	4 (6.7)		3 (5.8)	
Parent, n (%)			.99		.44
Both	46 (83.6)	48 (80)		44 (84.5)	
Either	7 (12.7)	11 (18.3)		8 (15.5)	
Neither	2 (3.7)	1 (1.7)		0 (0)	
Work status, n (%)			.95		.55
Unemployed	19 (34.5)	18 (30)		14 (26.9)	
Employed	36 (63.5)	42 (70)		38 (73.1)	
Education level, n (%)			.60		.36
Less than or equal to junior college	24 (43.6)	33 (55)		26 (50)	
College	26 (47.3)	23 (38.3)		23 (44.2)	
Postgraduate	5 (9.1)	4 (6.7)		3 (5.8)	
Monthly income (CNY ¥), n (%)			.57		.90
<5000 (US $689.72)	11 (20)	11 (18.3)		8 (15.5)	
5000-10,000 (US $689.72-1379.44)	15 (27.3)	17 (28.3)		15 (28.8)	
>10,000 (US $1379.44)	29 (52.7)	32 (53.4)		29 (55.7)	
Type of breast cancer, n (%)			.45		.89
Invasive	51 (92.7)	58 (96.7)		50 (96.2)	
Noninvasive	4 (7.3)	2 (3.3)		2 (3.8)	
Stage of breast cancer, n (%)			.13		.07
1	19 (34.5)	21 (35)		14 (26.9)	
2	30 (54.5)	28 (46.7)		28 (53.8)	
3	6 (11)	11 (8.3)		10 (19.3)	
Therapy for breast cancer, n (%)			.34		.66
Operation	5 (9.1)	8 (13.3)		3 (5.8)	
Adjuvant therapy[c]	26 (47.3)	23 (38.4)		20 (38.5)	
Both	24 (43.6)	29 (48.3)		29 (55.7)	

[a]ITT: intention-to-treat.

[b]PP: per-protocol.

[c]Includes radiotherapy, chemotherapy, endocrine therapy, etc.

The Effects on Primary Outcomes

Table 2 presents the overall test results of the intervention effect on psychological symptoms through repeated measures analysis. There were statistically significant group effects, time effects, and group×time interaction effects on the changes in the MSAS-SF and CBI-B scores. Significant time effects ($F_{2,108}$=236.123; P<.001) and group×time interaction effects ($F_{2,108}$=36.639; P<.001) were found in the ITT analysis for the SSRS score, but only a significant time effect ($F_{1.767,95.423}$=231.187; P<.001) was found in the PP analysis.

As demonstrated in Table 3, there were no significant differences in any of the variables between groups at T0, which supported successful randomization. At T2, psychological distress (ITT vs PP vs control: 0.40 vs 0.41 vs 0.93; P<.001) was significantly different among the groups. In addition, each group exhibited significant differences (ITT vs PP vs control: 92.83 vs 89.30 vs 85.65; P<.001) in self-efficacy; however, compared with those in the control group, the ITT group (P=.002) and PP group (P<.001) did not show significant increases in social support scores until T2.

Table 2. Repeated measures analysis of variance results for total scores.

Variable	ITT[a]			PP[b]		
	Group effect, *F* value (*P* value)	Time effect, *F* value (*P* value)	Group×time effect, *F* value (*P* value)	Group effect, *F* value (*P* value)	Time effect, *F* value (*P* value)	Group×time effect, *F* value (*P* value)
MSAS-SF[c]	14.118 (<.001)	75.718 (<.001)	2.219 (<.001)	13.916 (<.001)	75.007 (<.001)	2.222 (<.001)
CBI-B[d]	7.956 (.007)	526.864 (<.001)	23.850 (<.001)	7.838 (.007)	529.502 (<.001)	23.558 (<.001)
SSRS[e]	0.099 (.75)	236.123 (<.001)	36.639 (<.001)	0.123 (.19)	231.187 (<.001)	37.928 (.16)
FACT-B[f]	6.081 (.02)	275.261 (<.001)	36.978 (<.001)	2.571 (.02)	325.216 (<.001)	45.457 (<.001)

[a]ITT: intention-to-treat; n=60 in the intervention group and n=55 in the control group.

[b]PP: per-protocol; n=52 in the intervention group and n=55 in the control group.

[c]MSAS-SF: Memorial Symptom Assessment Scale-Short Form.

[d]CBI-B: Cancer Behavior Inventory-Brief Version.

[e]SSRS: Social Support Rating Scale.

[f]FACT-B: Functional Assessment of Cancer Therapy-Breast.

Table 3. Changes in psychological and related symptoms over time between each group.

Variable	Control group (n=55), mean (SD)	ITT[a] group (n=60)		PP[b] group (n=52)	
		Values, mean (SD)	*P* value[c]	Values, mean (SD)	*P* value[c]
MSAS-SF[d]					
T0	1.51 (0.50)	1.56 (0.47)	.25	1.59 (0.43)	.20
T1	1.23 (0.37)[e]	1.17 (0.46)[e]	<.001	1.07 (0.50)[e]	.37
T2	0.91 (0.22)[e]	0.68 (0.21)[e,f]	<.001	0.88 (0.20)[e]	.05
Psychological frequency					
T0	1.52 (0.29)	1.43 (0.36)	.67	1.50 (0.25)	.49
T1	1.45 (0.14)	1.41 (0.19)	.006	1.36 (0.17)	<.001
T2	0.95 (0.21)[e,f]	0.83 (0.13)	<.001	0.77 (0.20)[e,f]	<.001
Psychological distress					
T0	1.59 (0.14)	1.48 (0.11)	.38	1.37 (0.18)	.75
T1	1.22 (0.12)[e]	0.79 (0.25)[e]	<.001	0.80 (0.13)[e]	<.001
T2	0.93 (0.17)[e,f]	0.40 (0.16)[e,f]	<.001	0.41 (0.19)[e,f]	<.001
Global distress					
T0	1.46 (0.26)	1.30 (0.17)	.07	1.31 (0.22)	.92
T1	1.23 (0.11)[e]	0. 94 (0.23)[e]	<.001	0.95 (0.14)[e]	<.001
T2	1.03 (0.13)[e,f]	0.41 (0.20)[e,f]	<.001	0.54 (0.11)[e,f]	<.001
Physical distress					
T0	1.08 (0.15)	1.16 (0.10)	.30	1.13 (0.17)	.41
T1	0.75 (0.14)[e]	0.56 (0.25)[e]	.002	0.61 (0.15)[e]	.03
T2	0.49 (0.15)[e,f]	0.32 (0.12)[e,f]	<.001	0.39 (0.11)[e,f]	.01
CBI-B[g]					
T0	37.07 (3.50)	36.13 (6.72)	.08	36.62 (3.81)	.43
T1	65.35 (3.06)[e]	82.83 (5.70)[e]	<.001	77.12 (4.27)[e]	<.001
T2	85.65 (2.79)[e,f]	92.83 (3.04)[e,f]	<.001	89.30 (5.18)[e,f]	<.001
Maintaining independence and a positive attitude					
T0	5.27 (1.56)	5.39 (1.88)	.79	5.28 (1.63)	.22
T1	6.00 (1.41)[e]	6.98 (1.88)[e]	.04	6.97 (1.88)[e]	.02
T2	6.48 (1.35)[e,f]	7.84 (1.48)[e,f]	.001	7.85 (1.47)[e,f]	.001
Participating in medical care					
T0	6.32 (1.55)	6.27 (2.06)	.57	6.07 (1.74)	.55
T1	6.89 (1.31)[e]	7.28 (1.56)[e]	.01	7.27 (1.56)[e]	.03
T2	7.77 (1.19)[e,f]	7.78 (1.63)[e]	.03	7.80 (1.35)[e,f]	.05
Coping and stress management					
T0	4.91 (1.67)	5.07 (1.69)	.91	5.14 (1.16)	.89
T1	5.76 (1.24)[e]	6.91 (1.55)[e]	.001	6.88 (1.56)[e]	.04
T2	6.79 (1.18)[e,f]	7.74 (1.27)[e,f]	<.001	7.70 (1.26)[e,f]	<.001
Managing affect					

Variable	Control group (n=55), mean (SD)	ITT[a] group (n=60)		PP[b] group (n=52)	
		Values, mean (SD)	*P* value[c]	Values, mean (SD)	*P* value[c]
T0	5.55 (0.95)	5.40 (1.81)	.14	5.39 (1.85)	.13
T1	5.60 (1.67)[e]	6.62 (1.57)[e]	<.001	6.62 (1.58)[e]	<.001
T2	6.76 (0.59)[e,f]	7.62 (1.47)[e,f]	<.001	7.63 (1.45)[e,f]	<.001
SSRS[h]					
T0	39.33 (7.39)	38.04 (8.19)	.07	39.47 (7.24)	.50
T1	42.31 (6.68)[e]	42.67 (7.94)[e]	.73	42.45 (7.80)[e]	.74
T2	45.10 (6.44)[e,f]	50.09 (4.95)[e,f]	.002	49.78 (5.09)[e,f]	<.001
Objective social support					
T0	10.24 (3.58)	9.53 (3.08)	.26	9.70 (3.12)	.53
T1	11.62 (3.24)[e]	11.76 (3.51)[e]	.62	11.82 (3.32)[e]	.78
T2	12.69 (3.19)[e,f]	14.66 (2.63)[e,f]	.02	14.88 (2.57)[e,f]	.01
Subjective social support					
T0	21.62 (4.19)	21.72 (4.99)	.36	21.33 (4.66)	.42
T1	22.48 (3.87)[e]	22.62 (4.88)[e]	.80	23.52 (4.42)[e]	.61
T2	23.4 (3.71)[e,f]	25.28 (2.74)[e,f]	.01	25.61 (2.62)[e,f]	.03
Use of social support					
T0	7.48 (2.03)	7.57 (2.03)	.31	7.36 (2.04)	.13
T1	8.21 (1.8)[e]	8.29 (2.22)[e]	.40	8.39 (2.21)[e]	.24
T2	9 (1.75)[e,f]	10.16 (1.33)[e,f]	.007	10.09 (1.38)[e,f]	.04

[a]ITT: intention-to-treat.

[b]PP: per-protocol.

[c]*P* value of between-group differences.

[d]MSAS-SF: Memorial Symptom Assessment Scale-Short Form.

[e]Compared with T0, *P*<.05.

[f]Compared with T1, *P*<.05.

[g]CBI-B: Cancer Behavior Inventory-Brief Version.

[h]SSRS: Social Support Rating Scale.

The Effects on Secondary Outcomes

First, the analysis highlighted a statistically significant group effect (*P*=.02), time effect (*P*<.001), and group×time interaction effect (*P*<.001) for quality of life in both the ITT and PP groups (Table 2).

The results at T1 (ITT vs PP vs control: 106.68 vs 105.73 vs 100.33; *P*<.05) and T2 (ITT vs PP vs control: 124.47 vs 126.04 vs 113.50; *P*<.001) indicated that there was significant improvement in overall quality of life. Notably, significant differences between and within all groups were found in functional well-being and additional concerns both in T1 and T2. Compared with those in the control group, the ITT and PP groups did not show significant increases in physical, social or family, and emotional well-being until T2 (*P*<.05; Table 4).

Table 4. Changes in quality of life over time between each group.

Variable	Control group (n=55), mean (SD)	ITT[a] group (n=60)		PP[b] group (n=52)	
		Values, mean (SD)	*P* value[c]	Values, mean (SD)	*P* value[c]
FACT-B [d]					
T0	92.48 (15.15)	90.02 (12.44)	.52	91.64 (12.85)	.30
T1	100.33 (12.32)[e]	106.68 (12.49)[e]	.01	105.73 (12.54)[e]	.02
T2	113.50 (11.20)[e,f]	124.47 (9.14)[e,f]	<.001	126.04 (10.69)[e,f]	<.001
Physical well-being					
T0	20.5 (4.27)	20.04 (5.05)	.65	19.73 (4.99)	.75
T1	21.17 (3.66)[e]	21.61 (4.44)[e]	.45	22.8 (4.41)[e]	.44
T2	22 (3.57)[e,f]	24.91 (3.00)[e,f]	<.001	24.48 (3.73)[e,f]	<.001
Social or family well-being					
T0	19.86 (5.07)	20.11 (7.93)	.09	19.48 (5.35)	.07
T1	20.52 (3.76)[e]	21.31 (4.37)[e]	.39	21.10 (4.51)[e]	.39
T2	21.45 (3.47)[e,f]	24.89 (2.88)[e,f]	<.001	24.73 (2.24)[e,f]	<.001
Emotional well-being					
T0	12.88 (4.27)	12.87 (4.28)	.98	12.48 (4.51)	.84
T1	15.45 (3.01)[e]	15.98 (2.57)[e]	.39	18.36 (4.78)[e]	.40
T2	17.24 (2.99)[e,f]	18.50 (2.57)[e,f]	<.001	22.48 (3.51)[e,f]	.02
Functional well-being					
T0	12.57 (5.11)	12.89 (6.75)	.81	12.14 (5.06)	.48
T1	15.55 (3.92)[e]	17.86 (4.72)[e]	.008	17.91 (4.21)[e]	.008
T2	18.29 (3.63)[e,f]	23.85 (3.84)[e,f]	<.001	22.14 (3.08)[e,f]	<.001
Additional concerns (cancer type–specific questions)					
T0	26.67 (3.69)	26.91 (5.85)	.57	26.48 (4.11)	.48
T1	27.64 (3.35)[e]	29.92 (4.78)[e]	.02	28.00 (3.66)[e]	.02
T2	28.52 (3.38)[e,f]	32.31 (3.29)[e,f]	<.001	31.17 (3.52)[e,f]	<.001

[a]ITT: intention-to-treat.

[b]PP: per-protocol.

[c]*P* value of between-group differences.

[d]FACT-B: Functional Assessment of Cancer Therapy-Breast.

[e]Compared with T0, *P*<.05.

[f]Compared with T1, *P*<.05.

Discussion

Principal Findings

This randomized controlled trial examined the effectiveness of an internet-enabled, mobile, intelligent interactive intervention for young survivors of breast cancer over a 3-month period. The findings demonstrated the benefits of using a mobile app and engaging in 2-way web-based follow-up. Significant improvements were observed in psychological symptoms, including distress and frequency, indicating the positive impact of the intervention. In addition, there was a noticeable trend toward improvement in quality-of-life outcomes, with both the ITT and PP analyses showing consistent overall outcomes.

We observed that the MSAS-SF score decreased from moderate to mild in all groups, and psychological distress also significantly decreased by 1.08 in the ITT group and 0.96 in the PP group from baseline; these findings were more pronounced than those of American survivors of breast cancer [32] and survivors of lung cancer [34]. In addition, a significant reduction between groups in the frequency of psychological problems was found (eg, sadness, worry, irritability, and nervousness), with 0.83 in the ITT group and 0.77 in the PP

group at T2. Clinical evidence has confirmed that survivors of cancer who approach people considered isolated and marginalized with stigmatized conditions and underserved populations to confide negative emotions and relieve themselves may become stuck in a psychological trap [35]. *AI-TA* adopted multiple approaches to alleviate psychological distress and reduce the frequency of psychological symptoms. For example, individuals could connect and offer spiritual support to each other in a private forum, allowing young survivors of breast cancer to openly share experience and advice, thereby overcoming the hesitation often caused by traditional cultural norms. Web-based follow-up enables continuous care between health care professionals and survivors, providing targeted educational contents such as information about stress relief, emotional management, and other relevant information [36]. By integrating the assessments from AI-driven system with qualitative insights from interviews, the follow-up can be more accurate to address the nuanced needs of young survivors of breast cancer, ensuring that the care provided is both relevant and effective. In this study, young survivors of breast cancer had a lower frequency of psychological symptoms at baseline and therefore had little margin for improvement at T1 and T2 in the ITT group; conversely, the PP group demonstrated notable enhancements. This disparity may relate to the duration or intensity of the intervention [37]. In addition, the greater increase in physical and global distress in the intervention group identified the necessity of intelligent interactive support for young survivors of breast cancer.

The findings reported in this study align with the literature, indicating that the intervention program *AI-TA* had a positive effect on increasing self-efficacy levels, as reported previously [32]. In particular, the ITT and PP groups had already reached a high level of self-efficacy at T1, which was faster than the control group. Among them, coping and stress management showed a borderline significant trend between groups at T1, and it became better at T2. This is likely because these survivors lacked motivation and familiarity with *AI-TA* at first, resulting in insufficient in-depth effects on young survivors of breast cancer [38]. Although the intervention group exhibited more favorable changes in several self-efficacy variables, all the groups experienced positive changes, which indicated that the general information support also played a certain role in promoting young survivors of breast cancer.

In this study, the changes in symptoms among survivors of breast cancer were similar to changes in their levels of self-efficacy. Several explanations have been proposed to account for the relationship between syndrome and self-efficacy: patients with cancer with high self-efficacy have a high level of health beliefs, which may promote the recovery from symptoms; in contrast, those with low self-efficacy are prone to negative emotions such as anxiety and depression, which are not conducive to recovery from psychological and physical symptoms [39,40]. AI-driven system and fortnightly web-based follow-up encouraged young survivors of breast cancer to actively engage with the provided content, enhancing their participation and initiative. This interactive process improved their confidence and self-efficacy in managing their symptoms, as the AI continually adapts to their evolving needs and responses.

Furthermore, while all groups' total social support scores were sustained and reached adequacy by T2, no significant effect was observed across any dimension at T1. This initial absence can be attributed to the challenges faced by young survivors of breast cancer. Frequently undergoing treatments and grappling with severe side effects, young survivors likely found themselves with limited energy to engage with the AI-driven tool (*AI-TA*) or to communicate effectively with health care professionals [41]. Besides, research has identified several factors that affect the perception of social support of survivors of cancer. Specifically, young patients with a collectivist orientation who value in-group solidarity and interdependence may feel alienated from or resist joining groups perceived as outside their usual social circles in a short term [42]. Initially, *AI-TA* may not show a significant impact. The novelty of such apps and their integration into survivors' lives requires time to manifest tangible benefits. However, as these AI-driven systems evolve to more accurately assess and respond to daily symptoms, their potential to significantly enhance health management and symptom control for young survivors with breast cancer grows. Over time, continued engagement with *AI-TA* is likely to foster social support, deeper understanding of the disease, and overall better well-being for them. The gradual accumulation of these positive effects underscores the promise of long-term interventions to bolster survivors' outcomes.

This work revealed that the *AI-TA* mobile app has been shown to work effectively in improving quality of life. The findings in physical, social or family, and emotional well-being of these survivors did not increase until T2. Previous literature has also been published in the field of internet-based or computer-based interventions for survivors of breast cancer, and the results indicate that internet support has no significant impact on quality of life in recently diagnosed survivors of breast cancer [43]. Particularly in the early postoperative period and before and after chemotherapy, the recovery of physical well-being and role function was slower [44]. In addition to confronting the challenges of disease itself, young survivors of breast cancer often experience negative emotions associated with work, childbirth, support, and other pressures as well as feelings about being abandoned by the medical system [45]. The integration of web-based follow-up through AI-driven system fosters continuous interaction and support, which is crucial for these survivors managing sensitive and often underdiscussed topics such as sexual health. These not only allow real-time monitoring and assistance but also facilitate a space for them to seek guidance and share experiences securely and comfortably.

Limitations

There are several limitations to this research that weaken the generalizability of these findings and warrant further investigation. Firstly, because of the relatively small population size and heterogeneity of treatment, a small study may not detect significant effects on outcomes related to the whole psychological symptom. Secondly, it is possible that young survivors of breast cancer with different types of cancer would react differently to the content of this intervention. Furthermore,

the duration of this study was short; thus, in the future, long-term interventions could be carried out to detect differences between groups.

Conclusions

The mobile app *AI-TA* demonstrated significant benefits in addressing the psychological health needs of young survivors of breast cancer during their survivorship journey. The consistent duration, intelligent support, and ease of interaction and web-based follow-up facilitated through digital platforms contributed to the success of the intervention. Specifically, AI-driven features such as personalized content delivery based on user feedback, symptom tracking and management, and interactive support networks have proven crucial for enhancing self-efficacy and social support among these survivors. Emphasis should be placed on optimizing the frequency of interaction and content delivery during an intervention to sustain user engagement without inducing fatigue. The observed effect size on psychological and related symptoms warrants further exploration, prompting future research to expand and investigate the efficacy of such AI-driven interventions in larger trials and across diverse populations over extended periods.

Acknowledgments

The authors would like to thank the involved clinical specialists, nurses, and other technicians for their assistance and active guidance throughout the study. This study was supported by the National Natural Science Foundation of China (71804112), the Nursing Talent Team Construction Project of Shanghai Jiao Tong University, the School of Medicine (SHSMU2021-21), and the Shanghai Sailing Program (21YF1422400).

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) form V1.6.
[PDF File (Adobe PDF File), 949 KB - mhealth_v12i1e50783_app1.pdf]

Multimedia Appendix 2
Questionnaire to assess participants' interaction and satisfaction with the AI-TA mobile app program.
[DOCX File , 26 KB - mhealth_v12i1e50783_app2.docx]

Multimedia Appendix 3
Post hoc sensitivity analyses for missing data.
[DOCX File , 25 KB - mhealth_v12i1e50783_app3.docx]

References

1. Siegel RL, Miller KD, Wagle NS, Jemal A. Cancer statistics, 2023. CA Cancer J Clin 2023 Jan;73(1):17-48 [FREE Full text] [doi: 10.3322/caac.21763] [Medline: 36633525]
2. Fan L, Strasser-Weippl K, Li JJ, St Louis J, Finkelstein DM, Yu KD, et al. Breast cancer in China. Lancet Oncol 2014 Jun;15(7):e279-e289. [doi: 10.1016/S1470-2045(13)70567-9] [Medline: 24872111]
3. Population factsheets. International Agency for Research on Cancer. URL: https://gco.iarc.fr/today/en/fact-sheets-populations [accessed 2024-05-22]
4. Li T, Mello-Thoms C, Brennan PC. Descriptive epidemiology of breast cancer in China: incidence, mortality, survival and prevalence. Breast Cancer Res Treat 2016 Oct;159(3):395-406. [doi: 10.1007/s10549-016-3947-0] [Medline: 27562585]
5. Cai T, Huang Y, Huang Q, Xia H, Yuan C. Symptom trajectories in patients with breast cancer: an integrative review. Int J Nurs Sci 2021 Dec 21;9(1):120-128 [FREE Full text] [doi: 10.1016/j.ijnss.2021.12.011] [Medline: 35079613]
6. Bower JE, Partridge AH, Wolff AC, Thorner ED, Irwin MR, Joffe H, et al. Targeting depressive symptoms in younger breast cancer survivors: the pathways to wellness randomized controlled trial of mindfulness meditation and survivorship education. J Clin Oncol 2021 Nov 01;39(31):3473-3484 [FREE Full text] [doi: 10.1200/JCO.21.00279] [Medline: 34406839]
7. Nakamura Z, Deal A, Nyrop K, Chen YT, Quillen LJ, Brenizer T, et al. Serial assessment of depression and anxiety by patients and providers in women receiving chemotherapy for early breast cancer. Oncologist 2021 Feb;26(2):147-156 [FREE Full text] [doi: 10.1002/onco.13528] [Medline: 32946156]
8. Miaja M, Platas A, Martinez-Cannon BA. Psychological impact of alterations in sexuality, fertility, and body image in young breast cancer patients and their partners. Rev Invest Clin 2017;69(4):204-209 [FREE Full text] [doi: 10.24875/ric.17002279] [Medline: 28776605]

9. Zhang Y, Zhang X, Li N, He H, Chen J, Zhu M, et al. Factors associated with psychosocial adjustment in newly diagnosed young to middle-aged women with breast cancer: a cross-sectional study. Eur J Oncol Nurs 2023 Aug;65:102357. [doi: 10.1016/j.ejon.2023.102357] [Medline: 37321131]
10. Hu Y, Xu J, Wang X, Shi Y, Chen M, Im EO. Socio-ecological environmental characteristics of young Chinese breast cancer survivors. Oncol Nurs Forum 2021 Sep 01;48(5):481-490. [doi: 10.1188/21.ONF.481-490] [Medline: 34411083]
11. Schmidt ME, Scherer S, Wiskemann J, Steindorf K. Return to work after breast cancer: the role of treatment-related side effects and potential impact on quality of life. Eur J Cancer Care (Engl) 2019 Jul;28(4):e13051. [doi: 10.1111/ecc.13051] [Medline: 31033073]
12. Jiang L, Wang X, Hu Y. Research of comprehensive needs of young breast cancer survivors based on user portrait. Nurs J Chin People Liberation Army 2022;39(6):26-30. [doi: 10.3969/j.issn.1008-9993.2022.06.007]
13. Chen SQ, Sun N, Ge W, Su JE, Li QR. The development process of self-acceptance among Chinese women with breast cancer. Jpn J Nurs Sci 2020 Apr;17(2):e12308. [doi: 10.1111/jjns.12308] [Medline: 31845525]
14. Kane K, Kennedy F, Absolom KL, Harley C, Velikova G. Quality of life support in advanced cancer-web and technological interventions: systematic review and narrative synthesis. BMJ Support Palliat Care 2023 Dec 07;13(e2):e221-e234. [doi: 10.1136/bmjspcare-2020-002820] [Medline: 33903259]
15. Sotirova MB, McCaughan EM, Ramsey L, Flannagan C, Kerr DP, O'Connor SR, et al. Acceptability of online exercise-based interventions after breast cancer surgery: systematic review and narrative synthesis. J Cancer Surviv 2021 Apr;15(2):281-310 [FREE Full text] [doi: 10.1007/s11764-020-00931-6] [Medline: 32930924]
16. Jiang S. Functional interactivity in social media: an examination of Chinese health care organizations' microblog profiles. Health Promot Int 2019 Feb 01;34(1):38-46. [doi: 10.1093/heapro/dax056] [Medline: 28973346]
17. Weidener L, Fischer M. Role of ethics in developing AI-based applications in medicine: insights from expert interviews and discussion of implications. JMIR AI 2024 Jan 12;3:e51204. [doi: 10.2196/51204]
18. Xie YD, Wang X, Huang Y. A comparative study of medical APP interface design based on emotional design principles. Med Inform 2022;35(2):27-30. [doi: 10.3969/j.issn.1006-1959.2022.02.007]
19. Hu Y, Cheng C, Chee W, Im EO. Issues in internet-based support for chinese-american breast cancer survivors. Inform Health Soc Care 2020;45(2):204-216. [doi: 10.1080/17538157.2019.1624967] [Medline: 31282245]
20. Britten N, Ekman I, Naldemirci Ö, Javinger M, Hedman H, Wolf A. Learning from Gothenburg model of person centred healthcare. BMJ 2020 Sep 01;370:m2738. [doi: 10.1136/bmj.m2738] [Medline: 32873594]
21. Pirhonen L, Olofsson EH, Fors A, Ekman I, Bolin K. Effects of person-centred care on health outcomes-a randomized controlled trial in patients with acute coronary syndrome. Health Policy 2017 Feb;121(2):169-179. [doi: 10.1016/j.healthpol.2016.12.003] [Medline: 28062091]
22. Wolf A, Fors A, Ulin K, Thorn J, Swedberg K, Ekman I. An eHealth diary and symptom-tracking tool combined with person-centered care for improving self-efficacy after a diagnosis of acute coronary syndrome: a substudy of a randomized controlled trial. J Med Internet Res 2016 Feb 23;18(2):e40 [FREE Full text] [doi: 10.2196/jmir.4890] [Medline: 26907584]
23. Fors A, Blanck E, Ali L, Ekberg-Jansson A, Fu M, Lindström Kjellberg I, et al. Effects of a person-centred telephone-support in patients with chronic obstructive pulmonary disease and/or chronic heart failure - a randomized controlled trial. PLoS One 2018 Aug 31;13(8):e0203031 [FREE Full text] [doi: 10.1371/journal.pone.0203031] [Medline: 30169539]
24. Jiang L, Wang X, Xu J, Wu Y, Hu Y. Construction of an intelligent interactive nursing information support system and its application in young breast cancer survivors. Chin J Nurs 2023;58(6):654-661. [doi: 10.3761/j.issn.0254-1769.2023.06.002]
25. Rosenberg SM, Dominici LS, Gelber S, Poorvu PD, Ruddy KJ, Wong JS, et al. Association of breast cancer surgery with quality of life and psychosocial well-being in young breast cancer survivors. JAMA Surg 2020 Nov 01;155(11):1035-1042 [FREE Full text] [doi: 10.1001/jamasurg.2020.3325] [Medline: 32936216]
26. Kim K, Park H. Factors affecting anxiety and depression in young breast cancer survivors undergoing radiotherapy. Eur J Oncol Nurs 2021 Feb;50:101898. [doi: 10.1016/j.ejon.2021.101898] [Medline: 33465702]
27. Xu J, Wang X, Chen M, Shi Y, Hu Y. Family interaction among young Chinese breast cancer survivors. BMC Fam Pract 2021 Jun 21;22(1):122 [FREE Full text] [doi: 10.1186/s12875-021-01476-y] [Medline: 34154539]
28. Fu L, Hu Y, Lu Z, Zhou Y, Zhang X, Chang VT, et al. Validation of the simplified Chinese version of the memorial symptom assessment scale-short form among cancer patients. J Pain Symptom Manage 2018 Jul;56(1):113-121 [FREE Full text] [doi: 10.1016/j.jpainsymman.2018.03.024] [Medline: 29627566]
29. Heitzmann CA, Merluzzi TV, Jean-Pierre P, Roscoe JA, Kirsh KL, Passik SD. Assessing self-efficacy for coping with cancer: development and psychometric analysis of the brief version of the Cancer Behavior Inventory (CBI-B). Psychooncology 2011 Mar;20(3):302-312. [doi: 10.1002/pon.1735] [Medline: 20878830]
30. Xiao SY. The theoretical basis and research application of social support rating scale. J Clin Psychiatry 1994;4:98-100 [FREE Full text]
31. Wan C, Zhang D, Tang X, Zhang W, Li W, Ren H, et al. Introduction on measurement scale of quality of life for patients with breast cancer: Chinese version of FACT-B. China Cancer 2002;11(6):318-320. [doi: 10.3969/j.issn.1004-0242.2002.06.004]

32. Im EO, Kim S, Yang YL, Chee W. The efficacy of a technology-based information and coaching/support program on pain and symptoms in Asian American survivors of breast cancer. Cancer 2020 Feb 01;126(3):670-680 [FREE Full text] [doi: 10.1002/cncr.32579] [Medline: 31714598]
33. Cohen J. Statistical power analysis. Curr Dir Psychol Sci 2016 Jun;1(3):98-101. [doi: 10.1111/1467-8721.ep10768783]
34. Hung HY, Wu LM, Chen KP. Determinants of quality of life in lung cancer patients. J Nurs Scholarsh 2018 May;50(3):257-264. [doi: 10.1111/jnu.12376] [Medline: 29569334]
35. Wang JH, Adams I, Huang E, Ashing-Giwa K, Gomez SL, Allen L. Physical distress and cancer care experiences among Chinese-American and non-Hispanic White breast cancer survivors. Gynecol Oncol 2012 Mar;124(3):383-388 [FREE Full text] [doi: 10.1016/j.ygyno.2011.11.029] [Medline: 22115854]
36. Ko NY, Fikre TG, Buck AK, Restrepo E, Warner ET. Breast cancer survivorship experiences among Black women. Cancer 2023 Sep;129(S19):3087-3101. [doi: 10.1002/cncr.34634] [Medline: 37691522]
37. Tripepi G, Chesnaye NC, Dekker FW, Zoccali C, Jager KJ. Intention to treat and per protocol analysis in clinical trials. Nephrology (Carlton) 2020 Jul;25(7):513-517. [doi: 10.1111/nep.13709] [Medline: 32147926]
38. Jakob R, Harperink S, Rudolf AM, Fleisch E, Haug S, Mair JL, et al. Factors influencing adherence to mHealth apps for prevention or management of noncommunicable diseases: systematic review. J Med Internet Res 2022 May 25;24(5):e35371 [FREE Full text] [doi: 10.2196/35371] [Medline: 35612886]
39. Yuan L, Yuan L. Effectiveness of nursing Intervention on anxiety, psychology and self-efficacy among elderly patients with acute coronary syndrome after percutaneous coronary intervention: an observational cohort study. Medicine (Baltimore) 2021 Aug 20;100(33):e26899 [FREE Full text] [doi: 10.1097/MD.0000000000026899] [Medline: 34414946]
40. Kurt S, Altan Sarikaya N. Correlation of self-efficacy and symptom control in cancer patients. Support Care Cancer 2022 Jul;30(7):5849-5857. [doi: 10.1007/s00520-022-06972-0] [Medline: 35364732]
41. Zhu J, Ebert L, Liu X, Wei D, Chan SW. Mobile breast cancer e-support program for Chinese women with breast cancer undergoing chemotherapy (part 2): multicenter randomized controlled trial. JMIR Mhealth Uhealth 2018 Apr 30;6(4):e104 [FREE Full text] [doi: 10.2196/mhealth.9438] [Medline: 29712622]
42. To C, Leslie LM, Torelli CJ, Stoner JL. Culture and social hierarchy: collectivism as a driver of the relationship between power and status. Organ Behav Hum Decis Process 2020 Mar;157:159-176. [doi: 10.1016/j.obhdp.2019.12.006]
43. Loiselle CG, Edgar L, Batist G, Lu J, Lauzier S. The impact of a multimedia informational intervention on psychosocial adjustment among individuals with newly diagnosed breast or prostate cancer: a feasibility study. Patient Educ Couns 2010 Jul;80(1):48-55. [doi: 10.1016/j.pec.2009.09.026] [Medline: 19854604]
44. Zhou K, Wang W, Zhao W, Li L, Zhang M, Guo P, et al. Benefits of a WeChat-based multimodal nursing program on early rehabilitation in postoperative women with breast cancer: a clinical randomized controlled trial. Int J Nurs Stud 2020 Jun;106:103565. [doi: 10.1016/j.ijnurstu.2020.103565] [Medline: 32272281]
45. Matthews H, Semper H. 'Dropped from the system': the experiences and challenges of long-term breast cancer survivors. J Adv Nurs 2017 Jun;73(6):1355-1365. [doi: 10.1111/jan.13237] [Medline: 27987220]

Abbreviations

AI: artificial intelligence
CBI-B: Cancer Behavior Inventory-Brief Version
FACT-B: Functional Assessment of Cancer Therapy-Breast
ITT: intention-to-treat
MSAS-SF: Memorial Symptom Assessment Scale-Short Form
PCC: person-centered care
PP: per-protocol
SSRS: Social Support Rating Scale

Edited by L Buis, S Ma; submitted 23.07.23; peer-reviewed by D Chao, S Rennick-Egglestone; comments to author 09.01.24; revised version received 28.01.24; accepted 06.05.24; published 04.06.24.

Please cite as:
Jiang L, Xu J, Wu Y, Liu Y, Wang X, Hu Y
Effects of the "AI-TA" Mobile App With Intelligent Design on Psychological and Related Symptoms of Young Survivors of Breast Cancer: Randomized Controlled Trial
JMIR Mhealth Uhealth 2024;12:e50783
URL: https://mhealth.jmir.org/2024/1/e50783
doi:10.2196/50783
PMID:38833298

Original Paper

Feasibility and Preliminary Effects of a Social Media–Based Peer-Group Mobile Messaging Smoking Cessation Intervention Among Chinese Immigrants who Smoke: Pilot Randomized Controlled Trial

Nan Jiang[1], PhD; Ariel Zhao[1], BS; Erin S Rogers[1], MPH, DrPH; Ana Paula Cupertino[2], PhD; Xiaoquan Zhao[3], PhD; Francisco Cartujano-Barrera[2], MD; Katherine Siu[1], MS; Scott E Sherman[1,4], MPH, MD

[1]Department of Population Health, New York University Grossman School of Medicine, New York, NY, United States

[2]Department of Public Health Sciences, University of Rochester Medical Center, Rochester, NY, United States

[3]Department of Communication, College of Humanities and Social Sciences, George Mason University, Fairfax, VA, United States

[4]Department of Medicine, VA New York Harbor Healthcare System, New York, NY, United States

Corresponding Author:
Nan Jiang, PhD
Department of Population Health
New York University Grossman School of Medicine
17th Fl
180 Madison Ave
New York, NY, 10016
United States
Phone: 1 646 501 3553
Email: Nan.Jiang@nyulangone.org

Abstract

Background: Chinese immigrants experience significant disparities in tobacco use. Culturally adapted tobacco treatments targeting this population are sparse and the use is low. The low use of these treatment programs is attributed to their exclusive focus on individuals who are ready to quit and the wide range of barriers that Chinese immigrants face to access these programs. To support Chinese immigrant smokers at all levels of readiness to quit and address their access barriers, we developed the WeChat Quit Coach, a culturally and linguistically appropriate WeChat (Tencent Holdings Limited)–based peer group mobile messaging smoking cessation intervention.

Objective: This study aims to assess the feasibility, acceptability, and preliminary effects of WeChat Quit Coach.

Methods: We enrolled a total of 60 Chinese immigrant smokers in 2022 in New York City for a pilot randomized controlled trial (RCT) and a single-arm pilot test. The first 40 participants were randomized to either the intervention arm (WeChat Quit Coach) or the control arm (self-help print material) using 1:1 block randomization stratified by sex. WeChat Quit Coach lasted 6 weeks, featuring small peer groups moderated by a coach, daily text messages with text questions, and chat-based instant messaging support from the coach in response to peer questions. The next 20 participants were enrolled in the single-arm pilot test to further assess intervention feasibility and acceptability. All 60 participants were offered a 4-week supply of complimentary nicotine replacement therapy. Surveys were administered at baseline and 6 weeks, with participants in the pilot RCT completing an additional survey at 6 months and biochemical verification of abstinence at both follow-ups.

Results: Of 74 individuals screened, 68 (92%) were eligible and 60 (88%) were enrolled. The majority of participants, with a mean age of 42.5 (SD 13.8) years, were male (49/60, 82%) and not ready to quit, with 70% (42/60) in the precontemplation or contemplation stage at the time of enrollment. The pilot RCT had follow-up rates of 98% (39/40) at 6 weeks and 93% (37/40) at 6 months, while the single-arm test achieved 100% follow-up at 6 weeks. On average, participants responded to daily text questions for 25.1 days over the 42-day intervention period and 23% (9/40) used the chat-based instant messaging support. Most participants were satisfied with WeChat Quit Coach (36/39, 92%) and would recommend it to others (32/39, 82%). At 6 months, self-reported 7-day point prevalence abstinence rates were 25% (5/20) in the intervention arm and 15% (3/20) in the control arm, with biochemically verified abstinence rates of 25% (5/20) and 5% (1/20), respectively.

Conclusions: WeChat Quit Coach was feasible and well-received by Chinese immigrants who smoke and produced promising effects on abstinence. Large trials are warranted to assess its efficacy in promoting abstinence in this underserved population.

Trial Registration: ClinicalTrials.gov NCT05130788; https://clinicaltrials.gov/study/NCT05130788

(JMIR Mhealth Uhealth 2024;12:e59496) doi:10.2196/59496

KEYWORDS

smoking cessation; tobacco; mHealth; social media; Chinese American; immigrant; smoking; smoker; mobile messaging; randomized controlled trial; tobacco use; feasibility; acceptability; nicotine replacement therapy

Introduction

The prevalence of smoking has declined significantly in the United States over the past 50 years [1]. However, smoking rates remain disproportionately high in socioeconomically disadvantaged populations, including Chinese immigrants (foreign-born Chinese Americans). In New York City (NYC), the city with the largest Chinese immigrant population in the United States, 28% of Chinese American men smoke compared with 18% of the total NYC men [2]. Chinese immigrants, accounting for 68% of the NYC Chinese American population, are more likely to smoke than US-born Chinese Americans (New York City Department of Health and Mental Hygiene, unpublished data, 2018).

Culturally adapted tobacco treatments for Chinese immigrants are sparse [3,4] and the use is low. For example, the National Asian Smokers' Quitline (ASQ) engages only about 2000 Asian American callers nationwide annually (the number of Chinese-speaking callers is unknown) [5]. The low use of tobacco treatment programs is attributed to 2 major reasons. First, current tobacco treatment programs are designed to offer cessation support to individuals who are ready to quit smoking (ie, they plan to quit within a month). However, only 6%-33% of Chinese Americans who smoke are ready to quit [6-8]. This is largely due to the limited knowledge about the harms of smoking, strong attachment to traditional Chinese prosmoking norms, lack of behavioral capability for quitting, limited social support to facilitate cessation, and low self-efficacy [9-13]. Current tobacco treatment programs often provide no support to move the vast majority of individuals not ready to quit to the stage of being ready to quit. Second, Chinese immigrants have limited access to evidence-based tobacco treatments. Access barriers include limited awareness of treatment resources, skepticism about treatment effects, and time and economic constraints [9-11,14,15]. Less than 20% of Chinese American smokers who attempt to quit use pharmacotherapy or behavioral interventions [7,16]. Research is needed to explore intervention strategies that can engage a broad group of Chinese immigrant smokers, including those not ready to quit, and address the wide range of barriers to cessation.

Over the past decade, social media has been increasingly studied as a tool for tobacco treatment. Social media can reach large populations, allow users to access at their own time of convenience with low or no cost, and may alleviate access barriers to tobacco treatment. A total of 2 systematic reviews have provided evidence supporting the feasibility and acceptability of social media–based smoking cessation interventions [17,18]. In addition, 2 randomized controlled trials (RCTs) conducted among English-speaking smokers compared social media–based tobacco treatment with other cessation support [19,20]. Both interventions feature peer group messaging support, with mixed findings, for example, a Facebook (Meta)–based intervention resulted in abstinence rates comparable to the control at 3, 6, and 12 months [19], whereas a Twitter-based intervention yielded a higher abstinence rate than the control on day 60 [20].

A total of 3 RCTs among Chinese-speaking smokers in Hong Kong [21] and mainland China [22,23] compared WhatsApp (Meta)–based or WeChat (Tencent Holdings Limited)–based interventions with usual cessation care or no tobacco treatment, all using a one-on-one intervention model. A WhatsApp-based intervention led to higher abstinence than the control at 6 months (8% vs 5% biochemically validated 7-day point prevalence abstinence) [21]. A WeChat-based intervention also produced higher abstinence than the control at 26 weeks (12% vs 3% biochemically validated continuous abstinence) [23], but the trial only included individuals ready to quit. Another trial of a WeChat-based intervention did not report any abstinence outcomes [22]. Thus far, the impact of using social media to engage and treat smokers, particularly those not ready to quit, is still unclear.

To reduce tobacco use among Chinese immigrants, our multidisciplinary team developed a culturally and linguistically appropriate social media intervention named WeChat Quit Coach for Chinese immigrants who smoke, across all levels of readiness to quit. Informed by the social cognitive theory and the socioecological model, WeChat Quit Coach addresses multilevel barriers to cessation [24]. The intervention features small, private peer groups moderated by a coach, daily WeChat text messages with text questions, and chat-based instant messaging support from the coach responding to peer questions. By using WeChat, the most widely used social media platform among Chinese with approximately 1.3 billion monthly active users worldwide [25], WeChat Quit Coach holds the potential to reach a large population of Chinese immigrants. In our previous study, we found that 94% of Chinese immigrants who smoke in NYC use WeChat, and that WeChat is used more frequently than other platforms, including Facebook, Twitter, WhatsApp, Line (Line Corporation), and short messaging [10]. This study sought to assess the feasibility, acceptability, and preliminary effects of WeChat Quit Coach among Chinese immigrants who smoke.

Methods

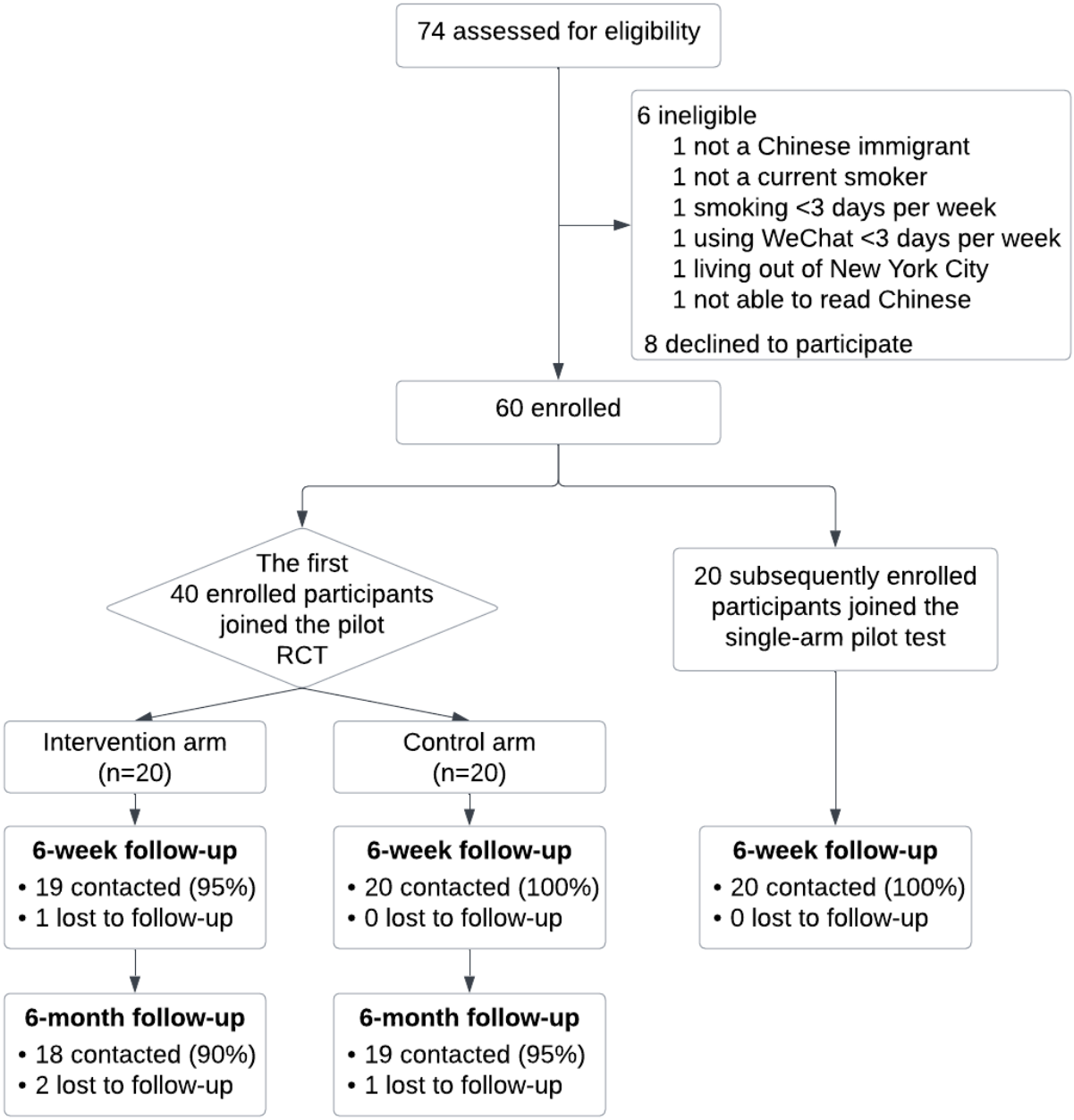
Ethics Approval

The study was approved by the Institutional Review Board of New York University Grossman School of Medicine (i20-01959) and preregistered at ClinicalTrials.gov (NCT05130788).

Study Design

As shown in Figure 1, we conducted an open-label, 2-arm pilot RCT and a single-arm pilot test among 60 participants. The first 40 participants were enrolled in the pilot RCT. To obtain more data on feasibility and acceptability, we enrolled an additional 20 participants in a single-arm pilot test.

Figure 1. Study schema. RCT: randomized controlled trial.



Participants in the pilot RCT were randomized in a 1:1 ratio to the intervention (6-week WeChat Quit Coach) or control arm (self-help print material). Both arms received a 4-week supply of nicotine patches and lozenges by request. Research staff conducted in-person baseline assessment and follow-up phone assessments at 6 weeks and 6 months postintervention initiation and performed biochemical validation through exhaled carbon monoxide (CO) tests at both time points. Participants in the single-arm pilot test received the same treatment as the intervention arm (6-week WeChat Quit Coach and nicotine replacement therapy [NRT]) and completed an in-person baseline survey and a phone survey at 6 weeks. Participants received a US $20 gift card for each survey completed and an additional US $20 for completing each biochemical validation test, irrespective of the test result.

Participants and Recruitment

Participants were recruited from February to December 2022, primarily in 3 NYC communities that have high concentrations of Chinese immigrants, including Flushing (Queens, NY), Sunset Park (Brooklyn, NY), and Chinatown (Manhattan, NY). Inclusion criteria included (1) self-identified as a Chinese immigrant; (2) age ≥18 years; (3) had smoked at least 100 cigarettes in their lifetime; (4) smoked ≥3 days per week; (5) used WeChat ≥3 days per week; (6) able to read Chinese; (7) lived in NYC; and (8) was somewhat interested in quitting, which was assessed by a question, "Which statement best describes your intention to quit? (A) I don't want to quit at all; (B) I may quit at some point, but not within the next 6 months; (C) I plan to quit within the next 6 months; (D) I plan to quit within the next 30 days; (E) I am trying to quit." People who chose an answer other than "A" were considered to be somewhat interested in quitting. Exclusion criteria included (1) current participation in other tobacco treatments, (2) pregnancy or breastfeeding, and (3) inability to connect with research staff through WeChat.

In collaboration with community-based organizations (CBOs), we disseminated study flyers at community events (eg, CBOs' workshops, health fairs, and immigrant resource fairs) and posted flyers in CBOs' offices and on their WeChat accounts. The flyer contained information about the study purpose, a study WeChat QR code, and a study phone number. In addition, CBOs and study participants referred potential participants to the research staff.

Potential participants contacted our research staff by scanning the study WeChat QR code or calling the study number. The research staff conducted eligibility screenings over the phone and arranged an in-person study visit for eligible individuals. At the visit, the research staff administered the written consent process and a paper-and-pencil baseline survey. The first 40 participants were enrolled in the pilot RCT. The subsequent 20 participants were enrolled in the single-arm pilot test.

Randomization

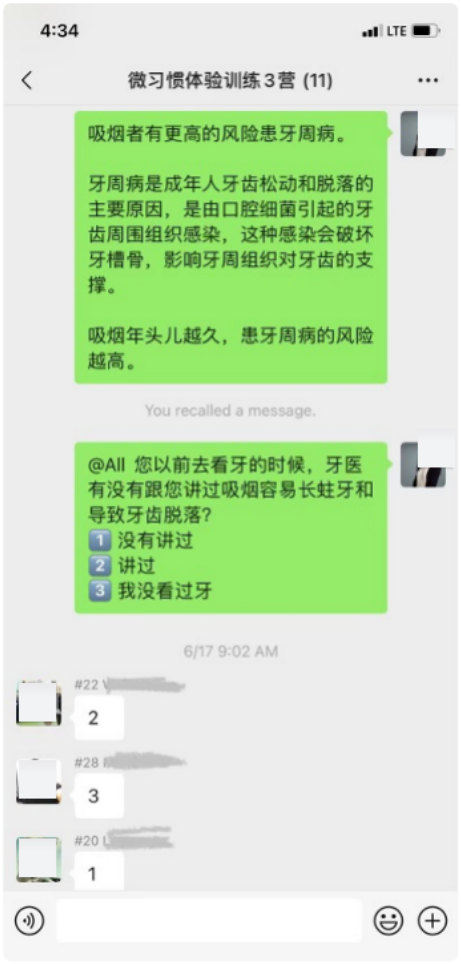
A total of 40 participants were randomized to the intervention or control arm (n=20 per arm) using block randomization stratified by sex (male or female). A randomization module was created and uploaded into REDCap (Research Electronic Data Capture). The research staff performed the allocation. Participants and research staff were unblinded to the assignment.

Intervention Arm

The research staff created a WeChat peer group every other month. Each group comprised of participants newly enrolled during the 2-month period who were either randomized to the intervention arm or enrolled in the single-arm pilot test (4-10 participants per group), along with a coach (NJ) and a research assistant. During the 6-week intervention period, the coach sent a WeChat text message to the group every day at 9 AM. The messages aimed to (1) enhance motivation to quit by building awareness about the health effects of smoking, quitting methods (eg, quitting preparation and relapse prevention) and cognitive or behavioral tips (eg, coping and refusal strategies), and available Chinese-language tobacco treatment programs (eg, ASQ and local smoking cessation programs); (2) challenge social norms that perpetuate smoking by denormalizing cigarette sharing culture and highlighting the fact that most Chinese immigrants do not smoke; (3) improve self-efficacy through motivational contents and tips for handling slips; and (4) encourage NRT use by addressing misconceptions. Details about the development process of WeChat Quit Coach are available elsewhere [24].

Following the daily WeChat text message, a text question was sent to the WeChat group each day to promote engagement (eg, "What is the longest period you've stayed abstinent in previous quit attempts?"). Participants were encouraged to respond daily and ask their own questions either in the group or directly to the coach, who responded within 24 hours. Figure 2 shows a screenshot of the intervention. If a participant did not respond for 3 consecutive days, a reminder message was sent to him or her through WeChat by the research assistant. Each participant received up to 3 reminders during the intervention. Participants could comment on others' responses and withdraw from the intervention at any time. The research assistant moderated peer interactions and monitored engagement.

Figure 2. A screenshot of the WeChat Quit Coach intervention.



Smokers have higher risks of developing gum diseases compared to nonsmokers.

Gum diseases, infections in the gums that affect the bone structure supporting teeth, are the primary cause of tooth loss among adults.

The longer one smoke, the greater the risk of developing gum diseases.

Question: Have you ever been told by a dentist that smoking increases the risk of tooth decay or loss?

1. No
2. Yes
3. Never visited a dentist

Control Arm

Participants allocated to the control arm received a leaflet with information about pharmacotherapy and behavioral interventions for smoking cessation, including the ASQ and local Chinese-language tobacco treatment programs.

NRT Supply

All 60 participants were offered a 4-week supply of free nicotine patches and lozenges. The participants could request NRT medications by texting the study WeChat account or calling the study number, with research staff screening for contraindications over the phone. If no contraindication was reported, participants were recommended a combination of nicotine patches and lozenges with doses as per package instructions. Participants could choose the form (eg, patch, lozenge, or both) and request

a lower dose. NRT medications and an instruction sheet (Chinese version) were then delivered by mail or in person.

Key Measures

Primary outcomes included feasibility, engagement, and acceptability. Feasibility was measured by eligibility rate (defined as the proportion of individuals screened who were eligible), enrollment rate (the proportion of eligible individuals who enrolled), follow-up rate (the proportion of participants who completed follow-up assessment at each time point), and the number of withdrawals. Engagement and acceptability were assessed at 6 weeks among participants from the intervention arm and the single-arm pilot test. Engagement was defined as the number of days participants responded to text questions and the number of participants who posted their own questions for instant messaging support. Acceptability was assessed through Likert scale questions regarding participants' perceptions about the intervention (eg, the timing of text messages and length of the intervention), helpfulness of intervention components (ie, text messages, text questions, and chat-based instant messaging support), overall satisfaction, and willingness to recommend it to others.

Secondary outcomes, assessed at 6 weeks and 6 months in the pilot RCT, included self-reported and biochemically verified 7-day point prevalence abstinence (yes or no). Participants reporting abstinence in the past 7 days were invited to participate in an exhaled CO test administered by research staff. A CO concentration of ≤6 parts per million indicated biochemically verified abstinence [26]. Additional outcomes encompassed quit attempts (yes or no; defined as no smoking for at least 24 hours because of trying to quit), change in smoking knowledge score, progression to a more advanced stage of change (yes or no), NRT use (yes or no), and use of other tobacco treatment (yes or no). Smoking knowledge was assessed using a 5-item measure adapted from the Global Adult Tobacco Survey [27], "Based on what you know, does smoking cigarettes cause (1) respiratory diseases, (2) lung cancer, (3) coronary heart disease, (4) stroke, (5) diabetes?" (0="no" or "not sure"; 1="yes"). The score of smoking knowledge was the sum of the points (ranging between 0 and 5). The stage of change was assessed by a question "Which statement best describes your plan about quitting?" with answer options of 1="I am not interested in quitting at all" (precontemplation), 2="I may quit in the future, but not in the next 6 months" (precontemplation), 3="I plan to quit within the next 6 months" (contemplation), 4="I plan to quit within the next 30 days" (preparation), and 5="I am trying to quit now" (action). Participants reporting a more advanced stage of change at follow-up, compared with baseline, were deemed to have transitioned to a more advanced stage.

Other measures included sociodemographic characteristics (eg, age, gender, education, birth country, years of residence in the United States, marital and employment status, and household income level), smoking behaviors (eg, age of smoking initiation, number of smoking days per week, cigarette consumption per day, time to first cigarette in the morning, and current use of other tobacco products), quitting experience (eg, quit attempts in the past 12 months, reasons for past quit attempts, and quitting methods ever used), and smoke-free home rule ("Which statement best describes the rule about smoking inside your home? (1) Smoking is not allowed anywhere in my home [complete rule]; (2) Smoking is allowed in some places or at some times in my home [partial rule]; (3) Smoking is allowed anywhere in my home [no rule]; (4) There are no rules about smoking in my home [no rule]" [28]).

Statistical Analysis

We performed descriptive statistics to summarize the variables of interest. Secondary outcomes were assessed using an intention-to-treat approach, with missing data treated as no change compared with baseline. To compare the differences between intervention and control arms, effect size estimates were computed, including odds ratios for categorical variables (eg, self-reported and biochemically validated abstinence) and Hedges *g* for the continuous variable (ie, smoking knowledge score) and their 95% CIs. For variables containing zero count cells, odds ratios and 95% CIs were not reported. All data analyses were performed using Stata 17 (StataCorp LLC).

Results

Feasibility

Of the 74 potential participants screened, 68 (92%) were eligible, and 60 (88%) enrolled, with no withdrawals. For the pilot RCT, follow-up rates were 98% (39/40 participants) at 6 weeks and 93% (37/40) at 6 months. The single-arm test achieved 100% follow-up at 6 weeks.

Sample Characteristics

On average, participants were 42.5 (SD 13.8) years old and had resided in the United States for 13.1 (SD 7.7) years (Table 1). Most were male (49/60, 82%), born in mainland China (58/60, 97%), and employed full-time (41/60, 68%). Out of 60, 15 participants (25%) had a middle school education (comparable to US 9th grade) or lower, and 28 participants (47%) reported an annual household income of US $55,000 or less.

Table 1. Characteristics of study participants at baseline (N=60).

Characteristics	Pilot RCT[a]		Single-arm pilot test (n=20)	Total (N=60)
	Intervention (n=20)	Control (n=20)		
Age (years), mean (SD)	45.5 (13.5)	42.5 (14.8)	39.6 (13.2)	42.5 (13.8)
Male, n (%)	14 (70)	15 (75)	20 (100)	49 (82)
Years of residence in the United States, mean (SD)	13.5 (7.7)	12.5 (5.1)	13.3 (9.9)	13.1 (7.7)
Place of birth, n (%)				
Mainland China	19 (95)	20 (100)	19 (95)	58 (97)
Other	1 (5)	0 (0)	1 (5)	2 (3)
Education, n (%)				
Middle school or less	5 (25)	5 (25)	5 (25)	15 (25)
High school or vocational high school	5 (25)	7 (35)	8 (40)	20 (33)
Some college, no degree or associate degree	4 (20)	2 (10)	2 (10)	8 (13)
Bachelor’s or advanced degree	6 (30)	6 (30)	5 (25)	17 (28)
Marital status, n (%)				
Single, never married	5 (25)	2 (10)	9 (45)	16 (27)
Married, living with a spouse	9 (45)	13 (65)	11 (55)	33 (55)
Married, living apart with spouse	1 (5)	2 (10)	0 (0)	3 (5)
Divorced	5 (25)	3 (15)	0 (0)	8 (13)
Employment status, n (%)				
Full-time employed	14 (70)	11 (55)	16 (80)	41 (68)
Part-time employed	1 (5)	5 (25)	2 (10)	8 (13)
Other	5 (25)	4 (20)	2 (10)	11 (18)
Annual household income level, n (%)				
Less than US $25,000	6 (30)	5 (25)	2 (10)	13 (22)
US $25,000-US $55,000	5 (25)	5 (25)	5 (25)	15 (25)
More than US $55,000	6 (30)	5 (25)	2 (10)	13 (22)
Not reported or not sure	3 (15)	5 (25)	11 (55)	19 (32)
Age of smoking initiation (years), mean (SD)	19.4 (8.4)	17.2 (4.1)	17 (4.3)	17.9 (5.9)
Current smoking status, n (%)				
Daily smoker	16 (80)	15 (75)	16 (80)	47 (78)
Nondaily smoker	4 (20)	5 (25)	4 (20)	13 (22)
Cigarette consumption per day, mean (SD)	15.5 (10)	9.4 (6.4)	11.2 (6)	12.0 (8)
Time to first cigarette in the morning, n (%)				
≤5 minutes	9 (45)	2 (10)	3 (15)	14 (23)
6-30 minutes	2 (10)	6 (30)	6 (30)	14 (23)
31-60 minutes	2 (10)	1 (5)	4 (20)	7 (12)
>60 minutes	6 (30)	5 (25)	5 (25)	16 (27)
No pattern or not sure	1 (5)	6 (30)	2 (20)	9 (15)
Current e-cigarette use, n (%)	5 (25)	8 (40)	10 (50)	23 (38)
Smoke-free home rule, n (%)				
Complete rule	9 (45)	10 (50)	13 (65)	32 (53)
Partial rule	3 (15)	6 (30)	2 (10)	11 (18)

Characteristics	Pilot RCT[a]		Single-arm pilot test (n=20)	Total (N=60)
	Intervention (n=20)	Control (n=20)		
No rule	8 (40)	4 (20)	5 (25)	17 (28)
Smoking knowledge score, mean (SD)	2.1 (1.6)	1.9 (1.3)	2.3 (1.8)	2.1 (1.6)
Quit attempts in the past 12 months, n (%)	10 (50)	5 (25)	8 (40)	23 (38)
Ever use of evidence-based tobacco treatment [b] , n (%)	3 (15)	3 (15)	5 (25)	11 (18)
Used tobacco treatment clinics or programs	0 (0)	0 (0)	1 (5)	1 (2)
Consulted a family doctor	0 (0)	1 (5)	0 (0)	1 (2)
Nicotine replacement therapy	3 (15)	3 (15)	4 (20)	10 (17)
Other smoking cessation medications	0 (0)	0 (0)	0 (0)	0 (0)
Called a quitline	0 (0)	0 (0)	0 (0)	0 (0)
Reasons for past quit attempts[b], n (%)				
Concern for my health	14 (70)	13 (65)	13 (65)	40 (67)
Concern for other's health	4 (20)	2 (10)	4 (20)	10 (17)
Family or roommate against smoking	5 (25)	1 (5)	4 (20)	10 (17)
Smoking is not allowed in my home	1 (5)	0 (0)	1 (5)	2 (3)
Smoking is not allowed in my workplace	3 (15)	1 (5)	2 (10)	6 (10)
To set a good example for my child	3 (15)	3 (15)	1 (5)	7 (12)
To save money	1 (5)	1 (5)	1 (5)	3 (5)
Other reasons	0 (0)	1 (5)	3 (15)	4 (7)
Stage of change, n (%)				
Precontemplation	11 (55)	9 (45)	4 (20)	24 (40)
Contemplation	4 (20)	5 (25)	9 (45)	18 (30)
Preparation	3 (15)	2 (10)	5 (25)	10 (17)
Action	2 (10)	4 (20)	2 (20)	8 (13)

[a]RCT: randomized controlled trial.

[b]Multiple responses, do not add up to 100%.

In total, 47 out of 60 participants (78%) reported daily smoking. On average, participants started smoking at 17.9 (SD 5.9) years old, smoked 12 (SD 8) cigarettes per day, and had a smoking knowledge score of 2.1 (SD 1.6). Participants were predominantly not ready to quit (42/60, 70% in precontemplation or contemplation stage). While 23 out of 60 participants (38%) reported current e-cigarette use, no participants reported current use of hookah, cigars, or smokeless tobacco (data not shown). Out of 60, 32 participants (53%) reported having a complete smoke-free home rule, 23 participants (38%) reported past 12-month quit attempts, and 11 participants (18%) had used evidence-based tobacco treatment, including 10 participants (17%) used NRT, 1 participant (2%) consulted a doctor, and 1 participant (2%) sought help from a smoking cessation program. None reported ever calling a quitline.

Engagement and Acceptability

On average, participants responded to WeChat text questions on 25.1 (SD 11) days out of 42 days (median 26.5, IQR 19-33 days). Reportedly, 9 out of 40 participants (23%) sent their own questions to receive chat-based instant messaging support from the coach. In total, 3 out of 40 participants (8%) interacted with peers. One of the participants commented on another's responses to a text question, while 2 participants each posted a message to communicate with the entire peer group.

Of the 39 participants who completed the 6-week follow-up assessment, 36 participants (92%) were satisfied or very satisfied with the intervention (Table 2). Participants predominantly agreed that WeChat Quit Coach enhanced their motivation to quit (35/39, 90%), increased awareness about how to quit (34/39, 87%), and promoted confidence in quitting (32/39, 82%). Most (32/39, 82%) would recommend the intervention to others, while 8 out of 39 participants (21%) indicated the intervention duration was too short. When asked about perceptions about the intervention, 34 out of 39 participants (87%) agreed that text messages were helpful, 33 participants (85%) indicated that text questions were helpful, and 25 participants (64%) found it enjoyable to answer text questions. Of the 9 who sent their own

questions for instant messaging support, 7 participants (78%) reported such support was helpful.

Table 2. Acceptability of WeChat Quit Coach among participants (n=39).

Perceptions	Participants, n (%)
Daily text messages	
Timing of text messages	
Too early	5 (13)
Too late	0 (0)
Just right	21 (54)
Does not matter	13 (33)
Length of text messages	
Appropriate	29 (74)
Too long	5 (13)
Other	5 (13)
"Do you find text messages helpful?"	
Helpful or very helpful	34 (87)
Neither helpful nor unhelpful	4 (10)
Unhelpful or very unhelpful	1 (3)
"In general, the messages are easy to understand."	
Agree or strongly agree	36 (92)
Neither agree nor disagree	3 (8)
Disagree or strongly disagree	0 (0)
Ever showed or sent text messages to others	10 (26)
Daily text questions	
"Do you enjoy responding to the text questions?"	
Enjoy or very much enjoy	25 (64)
Neutral	11 (28)
Do not enjoy	3 (8)
"Do you find text questions helpful?"	
Helpful or very helpful	33 (85)
Neither helpful nor unhelpful	4 (10)
Unhelpful or very unhelpful	2 (5)
Chat-based instant messaging support[a]	
"Do you find the instant messaging support helpful?"	
Helpful or very helpful	7 (78)
Neither helpful nor unhelpful	2 (22)
Unhelpful or very unhelpful	0 (0)
Overall intervention	
Satisfaction with the intervention	
Satisfied or very satisfied	36 (92)
Neither satisfied nor dissatisfied	3 (78)
Dissatisfied or very dissatisfied	0 (0)
Agree or strongly agree with the statements	
"It makes me want to quit smoking."	35 (90)
"It is helpful for quitting smoking."	32 (82)
"It makes me feel more confident that I can quit."	32 (82)

Perceptions	Participants, n (%)
"It makes me feel I knew how to quit."	34 (87)
"It makes me want to try again if a quit attempt is unsuccessful."	30 (77)
Length of the intervention (6 weeks)	
Appropriate	30 (77)
Too short	8 (21)
Too long	1 (2)
Would recommend the intervention to others	
Definitely yes	13 (33)
Probably yes	19 (49)
Probably no	5 (13)
Definitely no	2 (5)

[a]Among participants who sent in their own questions to receive chat-based instant messaging support (n=9).

Cessation Outcomes

At 6 weeks, 20% (4/20) of intervention participants reported 7-day point prevalence abstinence and were validated biochemically (Table 3). None of the control participants reported abstinence. At 6 months, 25% (5/20) of intervention participants reported abstinence and were validated, while 15% (3/20) of control participants reported abstinence and only 5% (1/20) were validated.

Table 3. Intention-to-treat analyses of preliminary effects of WeChat Quit Coach (n=40).

	6 weeks			6 months		
	Intervention arm	Control arm	Effect size estimates[a], OR[b] or Hedges *g* (95% CI)	Intervention arm	Control arm	Effect size estimates[a], OR or Hedges *g* (95% CI)
Self-reported 7-day point prevalence abstinence, n (%)	4 (20)	0 (0)	—[c]	5 (25)	3 (15)	OR 1.89 (0.38-9.27)
Biochemically verified 7-day point prevalence abstinence, n (%)	4 (20)	0 (0)	—	5 (25)	1 (5)	OR 6.33 (0.67-60.16)
Quit attempts, n (%)	11 (55)	7 (35)	OR 2.27 (0.54-9.82)	15 (75)	13 (65)	OR 0.62 (0.16-2.43)
Change in smoking knowledge score, mean (SD)	1.6 (1.6)	−0.3 (1.0)	Hedges *g* 1.10 (0.45-1.79)	—	—	—
Transition to a more advanced stage of change[d], n (%)	8 (40)	0 (0)	—	7 (35)	5 (25)	OR 1.62 (0.41-6.34)
Use of NRT[e], n (%)	11 (55)	14 (70)	OR 0.52 (0.12-2.3)	11 (55)	14 (70)	OR 0.52 (0.12-2.3)
Use of other tobacco treatment programs, n (%)	0 (0)	0 (0)	—	0 (0)	0 (0)	—

[a]For categorical variables, we computed the odds ratio and its 95% CI; for the continuous variable, we computed the Hedges *g* and its 95% CI.

[b]OR: odds ratio.

[c]Not applicable.

[d]Compared with baseline.

[e]NRT: nicotine replacement therapy.

Compared with baseline, intervention participants showed an average increase of 1.6 (SD 1.6) points in smoking knowledge at 6 weeks, while the score remained unchanged (mean -0.3, SD 1 points) for control participants. At both follow-up time

points, more intervention participants reported quit attempts and progressed to a more advanced stage of change, while more control participants requested NRT during the study period than intervention participants (17/20, 85% vs 16/20, 80%; data not shown in tables) and reported NRT use at both follow-ups (14/20, 70% vs 11/20, 55%).

Discussion

This is the first study of a social media–based smoking cessation intervention targeting Chinese immigrants. The sociodemographic characteristics of our participants correspond to the overall population traits of Chinese immigrants in NYC, including low income and low education [29]. The study demonstrates the feasibility and acceptability of a WeChat-based smoking cessation intervention and shows promising early efficacy on biochemically verified abstinence at 6 months.

It was found that 70% (42/60) of the participants were in the precontemplation or contemplation stage at enrollment. It suggests that the intervention is attractive to participants, including those not determined to quit in the near term. WeChat Quit Coach targets individuals at all levels of readiness to quit and is implemented through a platform deeply integrated into the daily lives of Chinese immigrants. The intervention has the potential to extend the reach of tobacco treatment within this population.

WeChat Quit Coach created favorable user experiences, as indicated by high levels of acceptability and engagement. Participants were overwhelmingly satisfied with the intervention. The interaction feature (daily text questions) was deemed enjoyable and helpful overall, resulting in a high engagement level compared with other social media peer group cessation interventions [20,30]. It is encouraging that participants post their own questions to receive instant messaging support, resulting in a usage rate (23%) comparable to that of a WhatsApp-based one-on-one cessation intervention for smokers in Hong Kong, which also features instant messaging support (17%) [21]. The results suggest that our intervention generally aligns with the needs of Chinese immigrant smokers. In total, 8 of the 39 participants (21%) reported the intervention was too short, highlighting opportunities for further refinement.

WeChat Quit Coach appears to be promising in promoting abstinence, with a biochemically verified abstinence rate of 25% at 6 months, which is higher than those reported in previous trials of social media interventions for Chinese-speaking smokers in Hong Kong (8% vs 5% biochemically validated 7-day point prevalence abstinence at 6 months [21]) and China (12% vs 3% biochemically validated continuous abstinence at 26 weeks [23]). Other outcomes, including quit attempts and smoking knowledge, were also higher in the intervention arm than control. Larger trials are needed to evaluate the impact of such intervention on abstinence among Chinese immigrants.

A noteworthy finding is that 55% (11/20) of intervention participants and 70% (14/20) of control participants reported NRT use. The high use of NRT could be attributed to the availability of complimentary NRT and the convenient process to request NRT. Chinese immigrants have limited awareness and hold widespread misconceptions about NRT despite its over-the-counter availability [10,14]. Moreover, Chinese immigrants have a high poverty rate compared with immigrants overall in the United States [31]. These factors contribute to the underuse of NRT. Our study indicates that Chinese immigrant smokers are receptive to using NRT when access barriers (eg, cost) are minimized. Our intervention might be more successful in promoting NRT use if cues to try NRT were added to our message library.

This study has several strengths. It is the first to explore a social media–based smoking cessation intervention for Chinese immigrants, a disadvantaged population with high smoking rates. It tests an innovative intervention using a culturally appropriate platform. The study fills gaps in research on tobacco-related disparities and the application of mobile technology for tobacco treatment. It also benefits from the inclusion of both individuals who are ready to quit and those not ready to quit, high follow-up rates, biochemically validated abstinence, and provision of NRT. This study has several limitations. First, as a pilot study aiming to test the feasibility, acceptability, and preliminary effects of our intervention, it was not adequately powered to assess treatment efficacy on abstinence. Second, participants were recruited from NYC, thus limiting the generalizability of our findings to other geographic regions.

This study supports the feasibility of a WeChat-based smoking cessation intervention for recruiting Chinese immigrant smokers across different levels of readiness to quit. The high levels of engagement, acceptability, and promising abstinence outcomes suggest that the intervention may be viable for this population. Larger trials are warranted to determine its efficacy.

Acknowledgments

We sincerely thank Professor Brian D Elbel (New York University Grossman School of Medicine) for the crucial support in securing approval from the Institutional Review Board for this research project. This work was supported by the National Institute on Minority Health and Health Disparities at the National Institutes of Health (K01MD014165) and New York University Grossman School of Medicine's Department of Population Health (Global Health Pilot Grant Funding Award). The funding organizations played no role in study design, data collection and analysis, manuscript writing, or the decision to submit the paper for publication.

Data Availability

The data sets generated during and/or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

All authors contributed to the writing-review and editing. NJ, APC, and SES performed conceptualization. AZ and NJ performed and contributed to the investigation, data curation, and writing the original draft. KS and NJ performed formal analysis. SES and NJ performed the methodology and supervision. NJ handled the funding acquisition, project administration, validation, and visualization, and managed resources and software.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) Checklist.
[PDF File (Adobe PDF File), 2987 KB - mhealth_v12i1e59496_app1.pdf]

References

1. The Health Consequences of Smoking: 50 Years of Progress. A Report of the Surgeon General. Atlanta, GA: Centers for Disease Control and Prevention; 2014.
2. Smoking among New York City men and boys. New York City Department of Health and Mental Hygiene. 2018 Jun. URL: https://www1.nyc.gov/assets/doh/downloads/pdf/epi/databrief103.pdf [accessed 2019-03-02]
3. Chen M, Tang H. Review of smoking cessation research among Asian Americans: the state of the research. Nicotine Tob Res 2007;9 Suppl 3:485-493. [doi: 10.1080/14622200701587052] [Medline: 17978977]
4. Liu JJ, Wabnitz C, Davidson E, Bhopal RS, White M, Johnson MR, et al. Smoking cessation interventions for ethnic minority groups--a systematic review of adapted interventions. Prev Med 2013;57(6):765-775. [doi: 10.1016/j.ypmed.2013.09.014] [Medline: 24076130]
5. Chen C, Anderson CM, Babb SD, Frank R, Wong S, Kuiper NM, et al. Evaluation of the Asian smokers' quitline: a centralized service for a dispersed population. Am J Prev Med 2021 Mar;60(3 Suppl 2):S154-S162 [FREE Full text] [doi: 10.1016/j.amepre.2020.01.033] [Medline: 33663703]
6. Ma GX, Tan Y, Toubbeh J, Su X. Differences in stages of change of smoking behavior among current smokers of four Asian American subgroups. Addict Behav 2003;28(8):1431-1439. [doi: 10.1016/s0306-4603(03)00071-6] [Medline: 14512065]
7. Fu S, Ma G, Tu X, Siu P, Metlay J. Cigarette smoking among Chinese Americans and the influence of linguistic acculturation. Nicotine Tob Res 2003;5(6):803-811. [doi: 10.1080/14622200310001614566] [Medline: 14668064]
8. Shelley D, Fahs M, Scheinmann R, Swain S, Qu J, Burton D. Acculturation and tobacco use among Chinese Americans. Am J Public Health 2004;94(2):300-307. [doi: 10.2105/ajph.94.2.300] [Medline: 14759946]
9. Spence W, Zhu L. Perceptions of smoking cessation among Glasgow's Chinese community. Tob Prev Cessat 2017;3:127 [FREE Full text] [doi: 10.18332/tpc/77942] [Medline: 32432201]
10. Jiang N, Zhang Y, Qian X, Thorpe L, Trinh-Shevrin C, Shelley D. Chinese immigrant smokers' access barriers to tobacco cessation services and experience using social media and text messaging. Tob Prev Cessat 2020;6:52 [FREE Full text] [doi: 10.18332/tpc/125942] [Medline: 33083680]
11. Spigner C, Yip M, Huang B, Tu SP. Chinese and Vietnamese adult male smokers' perspectives regarding facilitators of tobacco cessation behavior. Asian Pac J Cancer Prev 2007;8(3):429-435 [FREE Full text] [Medline: 18159982]
12. Ma G, Shive S, Ma X, Toubbeh J, Tan Y, Lan Y, et al. Social influences on cigarette smoking among mainland Chinese and Chinese Americans: a comparative study. Am J Health Stud 2013;28(1):12-20 [FREE Full text] [Medline: 24511220]
13. Kwong K, Ferketich AK, Wewers ME, Shek A, Tsang T, Tso A. Development and evaluation of a physician-led smoking cessation intervention for low-income Chinese Americans. J Smok Cessat 2012;4(2):92-98. [doi: 10.1375/jsc.4.2.92]
14. Tsang IK, Tsoh JY, Wong C, Le K, Cheng JW, Nguyen AN, et al. Understanding and use of nicotine replacement therapy and nonpharmacologic smoking cessation strategies among Chinese and Vietnamese smokers and their families. Prev Chronic Dis 2014;11:E26 [FREE Full text] [doi: 10.5888/pcd11.130299] [Medline: 24556252]
15. Katigbak C, Maglalang DD, Chao Y, Au H, Liang W, Zuo S. Cultural perspectives on tobacco use and cessation among Chinese American immigrants: a community-engaged qualitative study. J Transcult Nurs 2018;30(4):350-358. [doi: 10.1177/1043659618817583]
16. Yu ESH, Chen EH, Kim KK, Abdulrahim S. Smoking among Chinese Americans: behavior, knowledge, and beliefs. Am J Public Health 2002;92(6):1007-1012. [doi: 10.2105/ajph.92.6.1007] [Medline: 12036797]

17. Naslund JA, Kim SJ, Aschbrenner KA, McCulloch LJ, Brunette MF, Dallery J, et al. Systematic review of social media interventions for smoking cessation. Addict Behav 2017;73:81-93 [FREE Full text] [doi: 10.1016/j.addbeh.2017.05.002] [Medline: 28499259]
18. Luo T, Li M, Williams D, Phillippi S, Yu Q, Kantrow S, et al. Using social media for smoking cessation interventions: a systematic review. Perspect Public Health 2021;141(1):50-63. [doi: 10.1177/1757913920906845] [Medline: 32077368]
19. Ramo DE, Thrul J, Delucchi KL, Hall S, Ling PM, Belohlavek A, et al. A randomized controlled evaluation of the tobacco status project, a Facebook intervention for young adults. Addiction 2018;113:1683-1695 [FREE Full text] [doi: 10.1111/add.14245] [Medline: 29797621]
20. Pechmann C, Delucchi K, Lakon CM, Prochaska JJ. Randomised controlled trial evaluation of Tweet2Quit: a social network quit-smoking intervention. Tob Control 2017;26(2):188-194 [FREE Full text] [doi: 10.1136/tobaccocontrol-2015-052768] [Medline: 26928205]
21. Wang MP, Luk TT, Wu Y, Li WH, Cheung DY, Kwong AC, et al. Chat-based instant messaging support integrated with brief interventions for smoking cessation: a community-based, pragmatic, cluster-randomised controlled trial. Lancet Digit Health 2019;1(4):e183-e192 [FREE Full text] [doi: 10.1016/S2589-7500(19)30082-2] [Medline: 33323188]
22. Luo T, Li MS, Williams D, Fritz J, Phillippi S, Yu Q, et al. Implementation of a WeChat-based smoking cessation program for chinese smokers. Int J Environ Res Public Health 2021 Oct 25;18(21):11189 [FREE Full text] [doi: 10.3390/ijerph182111189] [Medline: 34769707]
23. Tang J, Yang J, Liu Y, Liu X, Li L, Sun Y, et al. Efficacy of WeChat-based online smoking cessation intervention ('WeChat WeQuit') in China: a randomised controlled trial. EClinicalMedicine 2023;60:102009 [FREE Full text] [doi: 10.1016/j.eclinm.2023.102009] [Medline: 37251625]
24. Jiang N, Rogers ES, Cupertino P, Zhao X, Cartujano-Barrera F, Lyu JC, et al. Development of a WeChat-based mobile messaging smoking cessation intervention for Chinese immigrant smokers: qualitative interview study. JMIR Form Res 2022;6(6):e36091 [FREE Full text] [doi: 10.2196/36091] [Medline: 35771603]
25. Number of monthly active WeChat users from 3rd quarter 2013 to 3rd quarter 2023 (in millions). Statista. URL: https://www.statista.com/statistics/255778/number-of-active-wechat-messenger-accounts/ [accessed 2023-12-16]
26. Benowitz N, Bernert J, Foulds J, Hecht S, Jacob P, Jarvis MJ, et al. Biochemical verification of tobacco use and abstinence: 2019 update. Nicotine Tob Res 2020;22(7):1086-1097 [FREE Full text] [doi: 10.1093/ntr/ntz132] [Medline: 31570931]
27. Noncommunicable disease surveillance, monitoring and reporting: global adult tobacco survey. WHO. 2020. URL: https://www.who.int/teams/noncommunicable-diseases/surveillance/systems-tools/global-adult-tobacco-survey/questionnaire [accessed 2022-03-31]
28. Kegler MC, Bundy L, Haardörfer R, Escoffery C, Berg C, Yembra D, et al. A minimal intervention to promote smoke-free homes among 2-1-1 callers: a randomized controlled trial. Am J Public Health 2015;105(3):530-537. [doi: 10.2105/AJPH.2014.302260] [Medline: 25602863]
29. Wang A. Chinese in NYC: a profile. Asian American Federation. 2022. URL: https://www.aafederation.org/research/chinese-in-nyc-a-profile/ [accessed 2023-04-26]
30. Ramo DE, Thrul J, Chavez K, Delucchi KL, Prochaska JJ. Feasibility and quit rates of the tobacco status project: a Facebook smoking cessation intervention for young adults. J Med Internet Res 2015;17(12):e291 [FREE Full text] [doi: 10.2196/jmir.5209] [Medline: 26721211]
31. Rosenbloom R, Batalova J. Chinese immigrants in the United States. Migration Policy Institute. 2023. URL: https://www.migrationpolicy.org/article/chinese-immigrants-united-states [accessed 2023-03-18]

Abbreviations

ASQ: Asian Smokers' Quitline
CBO: community-based organization
CO: carbon monoxide
NRT: nicotine replacement therapy
NYC: New York City
RCT: randomized controlled trial
REDCap: Research Electronic Data Capture

Edited by L Buis; submitted 13.04.24; peer-reviewed by TT Luk, H Meng, YS Wu; comments to author 09.05.24; revised version received 31.05.24; accepted 05.06.24; published 22.07.24.

Please cite as:
Jiang N, Zhao A, Rogers ES, Cupertino AP, Zhao X, Cartujano-Barrera F, Siu K, Sherman SE
Feasibility and Preliminary Effects of a Social Media–Based Peer-Group Mobile Messaging Smoking Cessation Intervention Among Chinese Immigrants who Smoke: Pilot Randomized Controlled Trial
JMIR Mhealth Uhealth 2024;12:e59496
URL: https://mhealth.jmir.org/2024/1/e59496
doi:10.2196/59496
PMID:39037756

Original Paper

A Chatbot-Delivered Stress Management Coaching for Students (MISHA App): Pilot Randomized Controlled Trial

Sandra Ulrich[1*], MSc; Natascha Lienhard[1*], MSc; Hansjörg Künzli[1], LicPhil; Tobias Kowatsch[2,3,4], PhD

[1]School of Applied Psychology, Zurich University of Applied Sciences, Zurich, Switzerland

[2]Institute for Implementation Science in Health Care, University of Zurich, Zurich, Switzerland

[3]School of Medicine, University of St. Gallen, St.Gallen, Switzerland

[4]Centre for Digital Health Interventions, Department of Management, Technology and Economics, ETH Zurich, Zurich, Switzerland

[*]these authors contributed equally

Corresponding Author:
Sandra Ulrich, MSc
School of Applied Psychology
Zurich University of Applied Sciences
Pfingstweidstrasse 96
Zurich, 8005
Switzerland
Phone: 41 58 934 ext 8451
Email: sandra.ulrich@zhaw.ch

Abstract

Background: Globally, students face increasing mental health challenges, including elevated stress levels and declining well-being, leading to academic performance issues and mental health disorders. However, due to stigma and symptom underestimation, students rarely seek effective stress management solutions. Conversational agents in the health sector have shown promise in reducing stress, depression, and anxiety. Nevertheless, research on their effectiveness for students with stress remains limited.

Objective: This study aims to develop a conversational agent–delivered stress management coaching intervention for students called MISHA and to evaluate its effectiveness, engagement, and acceptance.

Methods: In an unblinded randomized controlled trial, Swiss students experiencing stress were recruited on the web. Using a 1:1 randomization ratio, participants (N=140) were allocated to either the intervention or waitlist control group. Treatment effectiveness on changes in the primary outcome, that is, perceived stress, and secondary outcomes, including depression, anxiety, psychosomatic symptoms, and active coping, were self-assessed and evaluated using ANOVA for repeated measure and general estimating equations.

Results: The per-protocol analysis revealed evidence for improvement of stress, depression, and somatic symptoms with medium effect sizes (Cohen d=−0.36 to Cohen d=−0.60), while anxiety and active coping did not change (Cohen d=−0.29 and Cohen d=0.13). In the intention-to-treat analysis, similar results were found, indicating reduced stress (β estimate=−0.13, 95% CI −0.20 to −0.05; P<.001), depressive symptoms (β estimate=−0.23, 95% CI −0.38 to −0.08; P=.003), and psychosomatic symptoms (β estimate=−0.16, 95% CI −0.27 to −0.06; P=.003), while anxiety and active coping did not change. Overall, 60% (42/70) of the participants in the intervention group completed the coaching by completing the postintervention survey. They particularly appreciated the quality, quantity, credibility, and visual representation of information. While individual customization was rated the lowest, the target group fitting was perceived as high.

Conclusions: Findings indicate that MISHA is feasible, acceptable, and effective in reducing perceived stress among students in Switzerland. Future research is needed with different populations, for example, in students with high stress levels or compared to active controls.

Trial Registration: German Clinical Trials Register DRKS 00030004; https://drks.de/search/en/trial/DRKS00030004

(JMIR Mhealth Uhealth 2024;12:e54945) doi:10.2196/54945

KEYWORDS

conversational agent; mobile health; mHealth; smartphone; stress management; lifestyle; behavior change; coaching; mobile phone

Introduction

Background

Stress is rapidly becoming a major issue affecting adults in high-income countries, especially during periods of uncertainty and worry. Chronic stress is closely related to mental illnesses such as anxiety disorders and depression, leading to various symptoms such as sleep disturbances, pain, dizziness, cardiovascular and digestive problems, as well as fatigue [1,2]. Younger individuals, particularly students [3-7], are experiencing a decline in mental health on a global scale [8,9]. Studies indicate that approximately 11% of students experience impairments such as anxiety, depression, exhaustion, and burnout-like symptoms [1,10]. Furthermore, a high level of stress can have a negative impact on academic performance, resulting in changes in study direction, prolonged studies, and even dropout [11,12].

Students encounter distinct challenges during their academic journey, including the need to assimilate a substantial amount of content, effectively manage their time, cope with performance expectations, and handle examination pressure [13]. In addition, the developmentally sensitive period associated with this age group, combined with the academic environment, can contribute to increased stress levels [6]. Furthermore, compared to previous generations, today's students appear to exhibit lower stress tolerance and inadequate stress coping mechanisms, which further exacerbate the situation [1,14,15]. Notably, a recent study by Ehrentreich et al [16] reported that stress levels among students have increased by nearly 40% due to the impact of the COVID-19 pandemic.

To prevent students from experiencing chronic stress and its long-term effects, the implementation of appropriate prevention programs is crucial. These programs aim to promote students' self-management and stress management skills, including learning and time management techniques, to help them effectively cope with stress and to counteract increasing stress levels in the target group [10,17]. Studies have demonstrated the positive impact of interventions such as behavioral therapy–based approaches, relaxation and mindfulness exercises, psychoeducation, and time and study management strategies in reducing stress among students [10,18,19]. Typically, evidence-based stress management programs combine psychoeducational sessions with relaxation exercises [20-22]. Importantly, stress management programs should be specifically tailored to the needs of students. By considering the target group's real-life context, these programs facilitate the transfer of acquired skills into everyday life [23].

Despite the importance of stress management programs for students, successful uptake remains challenging [24]. Unfortunately, individuals experiencing stress often do not make use of stress management techniques for several reasons. These include the fear of being stigmatized [25], underestimation of the impact of stress, limited availability of therapy options, and high cost, particularly for young people in education [26,27].

Low-threshold, mobile health (mHealth) interventions such as smartphone apps could potentially bridge this gap. A meta-analysis by Weisel et al [28] highlighted the advantages of apps, including location and time independence, reduced stigmatization, and low costs [29]. Initial evidence suggests that smartphone apps can effectively reduce perceived stress, distress, depression, and anxiety and improve quality of life, psychological health, well-being, and self-regulation among student populations [30-32]. However, reported disadvantages of digital interventions, such as low adherence, legal concerns, lack of therapist relationship, and arbitrary scheduling, may diminish their effectiveness [29,33].

Conversational agents (CAs), commonly known as chatbots, are designed to simulate humanlike conversations and are increasingly used in clinical and nonclinical settings [34-36]. Initial findings demonstrate the feasibility, acceptance, and effectiveness of CAs in various health domains [37,38], including promoting physical activity [39]; managing pain [40]; reducing substance abuse [41,42]; improving depression, distress, and stress [43]; enhancing general wellness and pain [44]; and facilitating self-adherence and psychoeducation [38]. Although large language model (LLM)–based CAs have recently gained increasing attention [45], they are still subject to basic research in computer science because of several severe shortcomings, such as hallucinations and nonconscious bias, among others [46]. Therefore, LLM-based CAs are not yet appropriate for safe and ethical delivery of several-week health interventions [47]. Hence, we decided to implement an established, safe, and transparent approach to using CAs and used a rule-based CA [39,40,48-51].

Studies investigating the effectiveness of stress management interventions delivered by a CA specifically tailored to the needs of students are still lacking. While recent studies have explored interventions such as Stressbot, developed with Meta's Messenger (Meta Platforms, Inc) and CA Atena, accessible via Telegram messaging app (developed by the Digital Health Lab at Fondazione Bruno Kessler FBK research center), their focus has been limited to short-term outcomes or specific topics. For instance, while Stressbot aimed to reinforce coping self-efficacy, its intervention period was only 7 days [52]. Similarly, CA Aetna's positive psychology and cognitive behavioral approaches with a tailored focus on the unique needs of the COVID-19 pandemic rather than the life context of students led to inconclusive outcomes regarding anxiety and stress reduction [53]. Furthermore, a previous study evaluating an artificial intelligence (AI)–based chatbot that provided self-help interventions for students to reduce depression lacked detailed descriptions of evidence-based intervention designs, leaving uncertainty about the elements implemented [54]. However, evidence-based design is vital in developing CA-based coaching intervention programs [34] and stress management interventions

for specific groups such as students [23]. To our knowledge, there is no study describing the development and evaluation of the effectiveness of a CA-delivered stress management coaching program lasting several weeks and adapted to the specific context of students in their everyday lives.

Consequently, we have developed an evidence-based, scalable, and CA-delivered stress management coaching intervention for students called MISHA. It combines the following components: (1) providing psychoeducation about stress, mindfulness, and relaxation; (2) fostering participant motivation for self-reflection on stress and stress reactions; and (3) guiding participants in the regular practice of mindfulness and relaxation techniques. This comprehensive approach addresses key aspects of stress management, including knowledge acquisition, self-reflection, and practical application of mindfulness and relaxation techniques [19,55]. By focusing on these evidence-based intervention components, MISHA aims to empower students with effective tools and strategies to reduce stress and its long-term effects.

Objectives

The goal of this pilot study was twofold: (1) to develop a scalable, evidence-based coaching intervention specifically designed for students and delivered via a CA and (2) to assess the coaching intervention's effectiveness, engagement, and acceptance.

Methods

Intervention

App Development

MISHA was developed in collaboration with the ETH Zurich using the open-source software platform MobileCoach [56], designed for rule-based digital health interventions [48,57-59]. MISHA features a chat-based interface with multimedia elements and regular notifications to engage users. The app includes a chat channel, an audio library with relaxation exercises, psychoeducational illustrations, and frequently asked questions (Figure 1). Communications takes place via predefined but dynamic answer options or by providing free-text input. Study participants were provided with access to a beta version of the MISHA app for Android (Google LLC) devices through Firebase [60] and for iOS (Apple Inc) devices through TestFlight [61].

Figure 1. Screenshots of the MISHA app (coach selection, chat interface, reminder, and audio library). Translation from German to English, screenshot Select coach: "Choose a coach"; screenshot Chat with coach: "Effective time management can support you and prevent or reduce stress. Shall we discuss this?", "Yes, I'm interested.", "Great, you're on board. Today, we'll focus on reflecting on your personal thought and behavior patterns related to time management. Remember, time management is primarily self-management.", "Really?", "Perhaps you've experienced this yourself or observed it in others…"; screenshot Reminder: "Have you relaxed today? See you tomorrow", "Dear Isabelle, tomorrow I'll show you a relaxation exercise"; screenshot Audio library: "Progressive Relaxation - Introduction (long)", "Progressive Relaxation - Brief", "Progressive Relaxation - Extended", "Seated Meditation", "Footprints in the Snow", "Waterfall".

Coaching Concept of MISHA

The intervention concept for MISHA draws inspiration from an effective face-to-face prevention program [62], adapting its content and topics to suit a CA-delivered approach. MISHA's chat messages and notifications are aligned with the health action process approach (HAPA) model, emphasizing both motivational and volitional processes in behavior change [63].

MISHA integrates evidence-based strategies from cognitive behavioral therapy (CBT), mindfulness, and psychoeducation to provide information about stressors and coping techniques [55,64]. The stress management program includes fundamental elements derived from CBT, such as cognitive restructuring, identification, evaluation, and modification of maladaptive thought patterns [65]. In addition, techniques such as behavioral activation and activity monitoring from CBT were applied to

directly support the participants in their desired goals in a collaborative approach. For further details on CBTs and session elements, refer to Multimedia Appendix 1. The overall aim is to empower participants to reflect on their daily stressors and effectively manage their stress with new coping techniques.

Coaching Content

MISHA offers a consecutive 12-session coaching program based on the stress management manual by Kaluza [20]. Sessions cover psychoeducation on stress, relaxation techniques, and student-specific topics such as examination anxiety. Topics are personalized, for example, setting goals, individual appointments with the CA, or selecting a CA. Participants can schedule sessions every 2 to 4 days, completing the program in 24 to 54 days (refer to Multimedia Appendix 1 for an overview of sessions and a detailed description of the content). Throughout the coaching, participants receive personalized feedback on the progression of the coaching, motivational reminders, and reminders in case of inactivity (refer to Multimedia Appendix 2 for detailed information on reminders). Personalization on an individual level is essential in promoting trust, engagement, adherence, and effectiveness to digital health interventions [66,67].

Study Design and Procedure

We conducted an unblinded, 2-armed, pilot randomized controlled trial in a population of university students in Switzerland. Study participants were allocated either to a 4-week to 7-week coaching intervention or to a 40-day waitlist control group. This research project was registered at the German Clinical Trials Register accredited by the World Health Organization (DRKS00030004). The trial was conducted following CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) guidelines. No significant content changes were made to the coaching intervention during the study period.

After downloading the MISHA app, participants were greeted and provided with information about the study procedure and coaching program. They were explicitly informed that the app does not serve as a substitute for psychotherapy and were given guidance on where to seek further help if needed. Study information was displayed within the app. To proceed, participants had to provide electronic informed consent by confirming that they had read and understood the study information. Subsequently, inclusion criteria were checked, and participants were directed to the baseline self-assessment at preintervention (time point 1; T1) using the app's in-built LimeSurvey platform (LimeSurvey Project). The MobileCoach software automatically randomized participants into either the intervention or the waitlist control group by a 1:1 allocation using random numbers (0 to 1), with numbers <0.5 assigned to the intervention group. Participants from the intervention group started the coaching program immediately. Upon program completion (1) by working through all the modules or (2) after 54 intervention days, participants were directed to the postintervention survey (time point 2; T2) before moving to the final goodbye session. During the intervention, further self-reported outcomes (eg, goal achievement) and use data (eg, total minutes spent on in-app relaxation) were gathered.

Participants from the waitlist control group received short weekly chat messages from MISHA, informing them about the remaining duration of their wait and encouraging them to continue their participation in the study. After 40 days of waiting, they were presented with the postintervention survey (T2) and given the opportunity to participate in the coaching program.

There was no human involvement throughout the study; however, participants had the option to contact the study team via email if they encountered technical issues or encountered problems with app download.

Ethical Considerations

The Cantonal Ethics Committee of Zurich (KEK-ZH, BASEC-Nr. Req-2020-01038) reviewed the research project and confirmed that the study did not fall within the scope of the Human Research Act. All participants gave informed electronic consent by selecting a checkbox before enrolling in the study and were informed about their right to opt out at any time. Their data were deidentified. Participants who completed the postintervention survey had the opportunity to win a voucher worth CHF 200 (US $224.73). In addition, students of applied psychology at Zurich University of Applied Sciences had the opportunity to earn 5 test person hours.

Recruitment

From October 6, 2021, to the end of October 2021, flyers were distributed via email to students at the University of Zurich, the Zurich University of Teacher Education, University of Applied Sciences Northwestern Switzerland School of Education, the University of Teacher Education in Special Needs Zurich, and the Zurich University of Applied Sciences. In addition, the flyer was posted on Facebook (Meta Platforms, Inc) and LinkedIn (Microsoft Corp). The app could be downloaded via flyer by following a web link. Eligibility was determined within the MISHA app by self-report and included the following: (1) being aged ≥18 years; (2) possession of and basic knowledge in the use of a smartphone; (3) sufficient knowledge of the German language; and (4) being a student at a Swiss university, university of applied sciences, university of teacher education, or college of higher education.

Outcomes

Primary Outcome

To measure the effectiveness of the program, we assessed perceived stress at preintervention (T1) and postintervention (T2) time points using the German version of the Perceived Stress Scale, a self-report questionnaire consisting of 10 items [68]. Participants rated their responses on a scale ranging from 0 (never) to 5 (very often).

Secondary Outcomes

We measured secondary outcomes, including depression, anxiety, somatic symptoms, and active coping, at preintervention and postintervention time points by self-report. Multimedia Appendix 3 presents all outcomes and time points.

Depression, Anxiety, and Somatic Symptoms

We used the Patient Health Questionnaire Somatic, Anxiety, and Depressive Symptom Scales [69] to detect depressive symptoms, anxiety, and somatic symptoms, which consists of the Patient Health Questionnaire-9 (PHQ-9), Generalized Anxiety Disorder-7, and the Patient Health Questionnaire-15. The PHQ-9 is a 9-item questionnaire assessing depressive symptoms [70]. Participants rate the frequency of each symptom over the past 2 weeks, ranging from 0 (not at all) to 3 (nearly every day). The Generalized Anxiety Disorder-7 is a 7-item questionnaire that measures anxiety symptoms [71]. Participants rate the frequency of each symptom over the past 2 weeks, ranging from 0 (not at all) to 3 (nearly every day). The Patient Health Questionnaire-15 is a 15-item questionnaire measuring psychosomatic symptoms [72]. Participants rate the severity of each symptom over the previous 4 weeks, ranging from 0 (not bothered at all) to 2 (bothered a lot). For this study, items 14 (trouble with sleeping) and 15 (ie, low energy or tiredness) were collected in the PHQ-9 (similar in both questionnaires) but had to be converted according to the manual [73]. By combining these individual components, the PHQ Somatic, Anxiety, and Depressive Symptoms Scales provide a comprehensive assessment of depressive symptoms, anxiety, and somatic symptoms.

Active Coping

According to the HAPA model [74], we evaluated participants' engagement in stress management activities by asking them to rate how often they had actively taken steps to reduce stress in the past 5 days. The question was assessed on a rating scale ranging from 1 (never) to 4 (regularly). This allowed us to understand the participants' level of proactive involvement in managing their stress.

Predictor: Self-Efficacy Expectancy

Various health behavior change models, including the HAPA model [74], consider self-efficacy expectancy to be a key aspect of health behavior change. However, research findings on the impact on stress interventions are mixed [75-77]. To address this, we assessed self-efficacy expectancy using the General Self-Efficacy Scale [78]. Before the intervention, participants rated their agreement with statements on their ability to handle tasks effectively on a 4-point Likert scale ranging from 1 (not at all true) to 4 (exactly true). The total score of the General Self-Efficacy Scale ranges from 10 to 40, with higher scores indicating higher self-efficacy.

Exploratory

Working Alliance

To assess the interaction between participants and MISHA, we used the German version of the Working Alliance Inventory-Short Revised [79] after the intervention. This self-report questionnaire comprises 12 items that capture the quality of the therapeutic relationship and collaboration between participants and the CA via 3 dimensions: goal, task, and bond. Responses were rated with an adapted scale from 1 (I do not agree at all) to 6 (I completely agree) after the intervention.

Subjective Stress Expertise and Goal Achievement

Throughout the coaching period, we assessed participants' goal achievement 3 times (sessions 1, 6, and 11) using a scale of 1 to 10, where 1 referred to the goal as clearly not achieved and 10 referred to the goal as fully achieved. We further measured participants' stress expertise 3 times (sessions 2, 5, and 13) using a similar scale, ranging from 1 (no idea how stress manifests itself in me) to 10 (I know exactly how I react when under stress).

Engagement and Acceptance

The extent to which a participant has to engage with the intervention to derive the maximum benefits is termed intended use [80]. For MISHA, we defined intended use for participants as completing the postintervention assessment, regardless of completing all sessions. This definition was based on the fact that participants may have varied goals and desired outcomes, leading to differences in their use of MISHA's features, including frequency and duration [81,82]. It also implies that participants do not necessarily need to interact with all available intervention components. Furthermore, participants might discontinue using the intervention upon achieving their personal goals, indicating that nonuse is not due to loss of interest [83,84]. In addition, we ground this approach on the self-determination theory, where autonomy by providing choice is essential [85].

To assess participants' engagement in the coaching program, we analyzed use data from the intervention group by calculating the ratio of replied conversational turns based on the number of SMS text messages sent by MISHA in relation to SMS text messages replied by participants. Furthermore, we tracked the number of sessions completed by participants and the number of reminders sent to participants in cases of inactivity (ie, if participants stopped interacting during a session). In addition, we tracked the number of minutes of audio files played by participants throughout the intervention.

We evaluated the feasibility and acceptance of MISHA using the user version of the Mobile App Rating Scale (uMARS) [86] after the intervention. The uMARS is a validated questionnaire that assesses the dimensions of engagement, functionality, esthetics, information, perceived quality, and perceived impact. All subscales use a 5-point Likert scale ranging from 1 to 5, where higher scores indicate a more favorable judgment. In this study, 19 items were translated from English to German to assess engagement (eg, entertainment, interest, customization, interactivity, and target group of the app), information (eg, quality of information, quantity of information, visual information, and credibility of source), perceived quality (eg, recommendation, use, payment, and overall rating), and perceived impact (eg, awareness, knowledge, attitudes, behavior change, seeking help, and intention to change). In addition to the uMARS, participants had the opportunity to provide feedback in free text prompted by the following questions: "What did you like most about the MISHA app?" and "What would you improve in the MISHA app?"

Sample Size Calculation

The sample size was estimated for a generalized estimating equation (GEE) based on a repeated-measure (within-between

interaction) ANOVA. A small to medium time by group interaction effect size (Cohen *f*=0.15) for the primary outcome perceived stress due to prior results [87] was expected. The G*Power (Heinrich-Heine-Universität Düsseldorf) analysis [88] revealed that a sample size of 90 participants would be sufficient with a power of 0.80 and a correlation of *r*=0.5 between measurements. Owing to the high percentage of dropouts observed in earlier studies, the target sample was increased to 180 participants [89].

Data Analysis

Descriptive statistics, independent 2-tailed *t* tests, and chi-square tests were conducted to analyze baseline differences in demographics and outcomes between the intervention and control groups.

In our analysis, we examined the effectiveness of the intervention by assessing changes in the primary outcome perceived stress scores over time within each group (intervention and control) and comparing these changes between groups. We first conducted a per-protocol (PP) analysis, including only participants who completed both surveys. This was done using a repeated-measure ANOVA with perceived stress as the dependent variable, time as the within-subject factor, and group as the between-group factor. Secondary outcomes, including depression, anxiety, psychosomatic symptoms, and active coping, were analyzed accordingly.

In compliance with the CONSORT (Consolidated Standards of Reporting Trials) guidelines, we also conducted an intention-to-treat (ITT) analysis wherein all randomized participants were included, regardless of their adherence to the coaching intervention. This analysis was performed using GEE. In model 1, we conducted an unadjusted evaluation with time (T1 and T2), group (intervention and control), and treatment (group by time interaction) as independent variables, with perceived stress as the dependent variable. The incorporation of time allows the examination of the dependent variable stress over different time points, the incorporation of group allows for comparison of stress between groups, and the interaction between group and time allows for an examination of whether the changes in outcomes over time differ between the intervention and control groups. In model 2, we did an adjusted analysis with the inclusion of the covariate general self-efficacy for the primary outcome perceived stress. The same independent variables were considered as in model 1. Secondary outcomes were evaluated accordingly. A log link function, gamma distribution, and unstructured covariance structure were applied. This modeling approach provided the best fit with the outcomes and allowed us to avoid restrictions on the covariance structure. To reduce the impact of influential observations and outlier effects, we used a robust estimator, which is consistent with standard procedures when using GEE.

Using GEE [90] offered several advantages. First, it allowed us to consider the correlations between the measurement times in longitudinal data, which is important for analyzing repeated measures. In addition, GEE allowed us to include incomplete data sets using an estimating equation to handle missing data. GEEs use all available data and estimate missing outcome values under the assumption of missing completely at random (MCAR). To assess the assumption of MCAR, we conducted the Little MCAR test. Calculations of between-group effect sizes (Cohen *d*) were based on the pooled SD and labeled as small (Cohen *d*=0.2), medium (Cohen *d*=0.5), and large (Cohen *d*=0.8). Furthermore, we explored the potential relation of working alliance and perceived impact on treatment outcomes using a correlation. All statistical analyses were performed using SPSS software (version 28; IBM Corp). We applied qualitative content analysis [91,92] using thematic maps [93] to answer the open-ended questions.

Results

Demographics and Baseline Scores

In total, 230 individuals downloaded the app. Of the 230 individuals, 148 (64.3%) were assessed for eligibility and completed the baseline survey. Before randomization, of the 148 participants, 8 (3.5%) discontinued using the app and 140 (60.9%) were randomized into intervention (70/140, 50%) and waitlist control (70/140, 50%) groups. The complete participant flow is depicted in Figure 2.

Participants had a mean age of 26.71 (SD 6.29) years. While 23.6% (33/140) of the participants identified as men, 73.6% (103/140) as women, and 2.1% (3/140) as nonbinary, 0.7% (1/140) declined to provide information about their gender (Table 1). Regarding relationship status, 59.3% (83/140) of the participants reported being married or in a relationship, while 40.7% (57/140) were single. Regarding educational background, most participants (90/140, 64.3%) had an apprenticeship or vocational or high-school diploma. A substantial proportion of the participants (37/140, 26.4%) had a university degree at the bachelor level or higher vocational education or training, while 8.6% (12/140) had other qualifications. Regarding their field of study, most participants (131/140, 93.6%) were studying at a university of applied sciences or university, while 5% (7/140) were studying at other institutions. The participants had a degree in (applied) psychology (124/140, 88.6%), social sciences (6/140, 4.4%), or other fields (7/140, 5%). There were no differences between groups for any of the outcomes at baseline.

Figure 2. Study flowchart. ITT: intention-to-treat; PP: per-protocol; T1: time point 1.

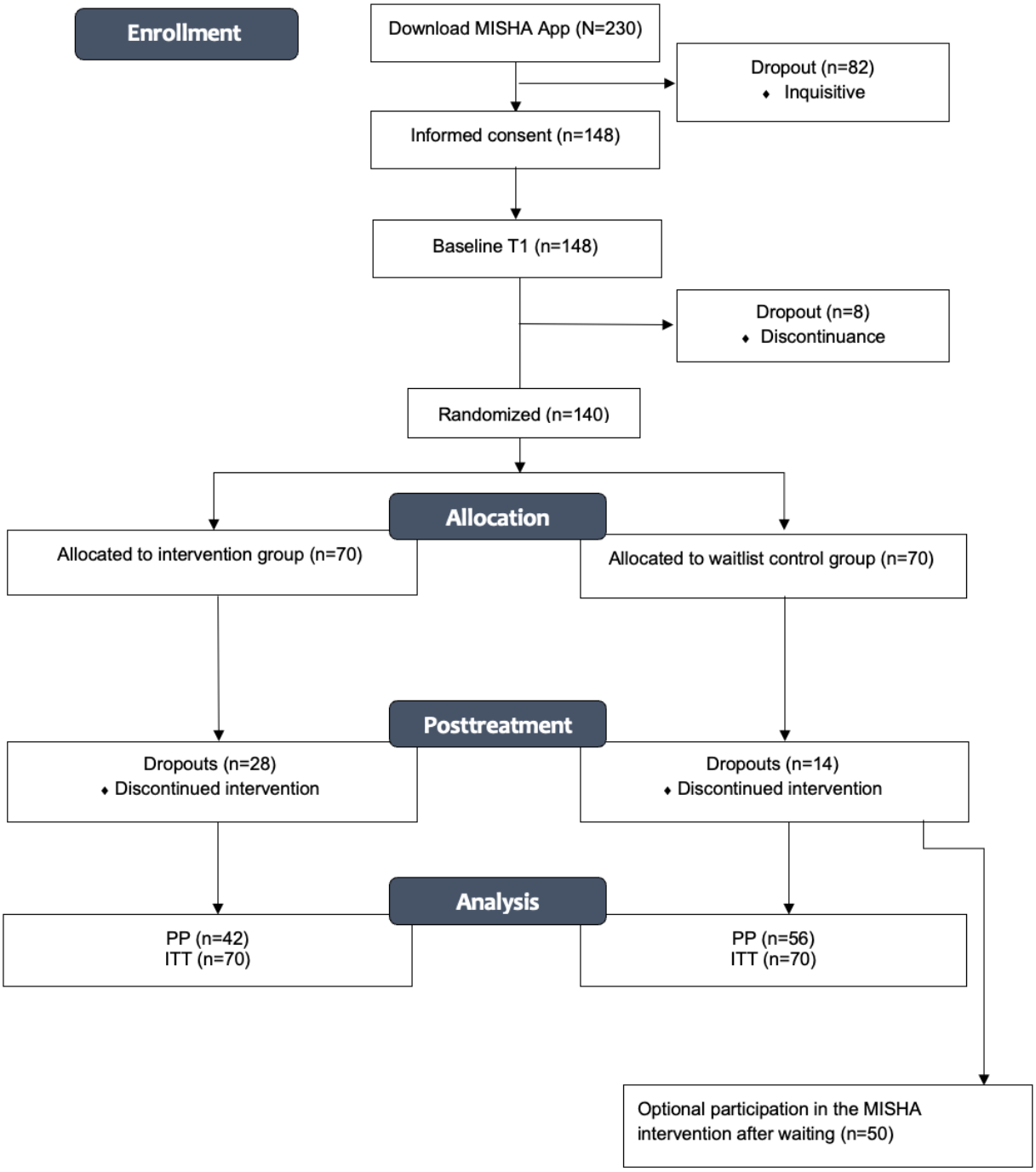


Table 1. Sample description at baseline (n=140).

Outcome	Control group (n=70)	Intervention group (n=70)	*P* value[a]
Age (y), mean (SD)	26.21 (5.56)	27.22 (6.96)	*.75*
Gender, n (%)			*.78*
Man	17 (24.3)	16 (22.9)	
Woman	52 (74.3)	51 (72.8)	
Nonbinary	1 (1.4)	2 (2.9)	
Not specified	0 (0)	1 (1.4)	
Highest education, n (%)			*.78*
Apprenticeship, vocational training, or high-school diploma	47 (67.1)	43 (61.4)	
Higher vocational education and training	6 (8.6)	7 (10)	
Degree at BSc[b] level	17 (24.3)	20 (28.6)	
Relationship status, n (%)			*.86*
Single	29 (41.4)	28 (40)	
Married or in relationship	41 (58.6)	42 (60)	
Study institute, n (%)			*.39*
University of Applied Science	67 (95.7)	64 (91.5)	
University and Swiss Federal Institute of Technology ETH	3 (4.3)	4 (5.7)	
University of Education	0 (0)	1 (1.4)	
Others	0 (0)	1 (1.4)	
Study subject, n (%)			*.33*
Applied psychology	63 (92.6)	60 (87.2)	
Social Work	0 (0)	2 (2.9)	
Information or technology	1 (1.5)	0 (0)	
Economics and business	1 (1.5)	1 (1.4)	
Pedagogy	0 (0.0)	1 (1.4)	
Natural and earth sciences	0 (0)	1 (1.4)	
Social sciences	3 (4.4)	3 (4.3)	
Other	0 (0)	1 (1.4)	
Outcomes, mean (SD)			
Perceived stress (PSS-10[c])	28.79 (5.27)	28.4 (5.45)	.67
Depression (PHQ-9[d])	8.16 (4.57)	7.83 (4.16)	.66
Anxiety (GAD-7[e])	6.84 (4.05)	6.69 (3.77)	.81
Psychosomatic symptoms (PHQ-15[f])	9.26 (4.09)	8.87 (4.39)	.59
Self-efficacy (GSES[g])	29.09 (3.36)	29.21 (2.86)	.81
Active coping (HAPA[h])	2.43 (0.79)	2.29 (0.85)	.31

[a]Baseline group comparison between intervention group and waitlist control group with *t* test or chi-square test. Italicized values are statistically significant.

[b]BSc: Bachelor of Science.

[c]PSS-10: Perceived Stress Scale-10.

[d]PHQ-9: Patient Health Questionnaire-9.

[e]GAD-7: Generalized Anxiety Disorder-7.

[f]PHQ-15: Patient Health Questionnaire-15.

[g]GSES: General Self-Efficacy Scale.

[h]HAPA: health action process approach.

Effectiveness

To evaluate the effectiveness of the intervention and to take missing data into account, a PP analysis of the time by group interaction was conducted followed by an ITT analysis. For the PP analysis (Table 2), we found evidence of a treatment effect (group by time interaction) from pre- to postintervention time points between the intervention and control groups for stress (P=.001; Cohen d=−0.60), depressive symptoms (P=.003; Cohen d=−0.50), and psychosomatic symptoms (P=.010; Cohen d=−0.36) but not for anxiety and active coping behavior.

In the ITT analysis for the unadjusted model (model 1), we found evidence of a treatment effect (group by time interaction) from pre- to postintervention time points between the intervention and control groups for stress (P<.001), depressive symptoms (P=.003), and psychosomatic symptoms (P=.003). No treatment effect was found for anxiety (P=.13) and active coping (P=.09).

After adjusting for the covariate self-efficacy expectancy (model 2), we found evidence of treatment effect sizes similar to model 1 (Table 3). Furthermore, there was evidence for an effect of self-efficacy expectancy on perceived stress (P<.001), depression (P<.001), anxiety (P<.001), and psychosomatic symptoms (P<.001) but not on active coping.

Table 2. Preintervention and postintervention means, results of the per-protocol (PP) repeated-measure ANOVA analysis, and between-group effect sizes (Cohen *d*) of primary and secondary outcomes (n=98).

Measure	Preintervention, mean (SD)	Postintervention, mean (SD)	Between-group effect sizes (intervention group vs waitlist control group after the intervention)			
			Cohen *d*[a] (95% CI[b])	Partial η^2	ANOVA	
					F test (*df*)	*P* value
Primary outcome						
Perceived stress (PSS-10[c])						
Intervention (n=42)	28.41 (5.53)	24.24 (5.93)	−0.60 (−1.01 to −0.19)	0.10	10.69 (1, 96)	.001
Control (n=56)	28.36 (4.93)	27.61 (5.38)	—[d]	—	—	—
Secondary outcomes						
Depression (PHQ-9[e])						
Intervention (n=42)	7.90 (4.24)	5.95 (3.45)	−0.50 (−0.91 to −0.10)	0.09	9.29 (1, 96)	.003
Control (n=56)	7.86 (4.13)	7.86 (4.02)	—	—	—	—
Anxiety (GAD-7[f])						
Intervention (n=42)	6.52 (3.69)	5.62 (3.22)	−0.29 (−0.69 to 0.11)	0.03	3.18 (1, 96)	.08
Control (n=56)	6.41 (3.32)	6.59 (3.47)	—	—	—	—
Somatic symptoms (PHQ-15[g])						
Intervention (n=42)	9.19 (4.81)	7.50 (3.78)	−0.36 (−0.76 to −0.04)	0.07	6.92 (1, 96)	.01
Control (n=56)	9.07 (3.89)	9.00 (4.43)	—	—	—	—
Active coping (HAPA[h])						
Intervention (n=42)	2.21 (0.87)	2.67 (0.75)	0.13 (−0.27 to 0.53)	0.04	3.60 (1, 96)	.06
Control (n=56)	2.45 (0.81)	2.57 (0.78)	—	—	—	—

[a]Cohen *d* values based on means and the pooled SD of the PP analysis.

[b]95% CI of Cohen *d* (between groups, after the intervention).

[c]PSS-10: Perceived Stress Scale-10.

[d]Not applicable.

[e]PHQ-9: Patient Health Questionnaire-9.

[f]GAD-7: Generalized Anxiety Disorder-7.

[g]PHQ-15: Patient Health Questionnaire-15.

[h]HAPA: health action process approach.

Table 3. Results of the outcome intention-to-treat analysis (model 1), including self-efficacy as covariate (model 2), using generalized estimating equations.

Outcome	Model 1[a]		Model 2[b]	
	β estimate (SE; 95% CI)	*P* value	β estimate (SE; 95% CI)	*P* value
Perceived stress (PSS-10[c])				
Intercept	3.36 (—[d])	—	4.18 (—)	—
Time[e]	−0.03 (0.02; −0.08 to 0.05)	.17	−0.04 (0.02; −0.08 to 0.01)	.12
Group[f]	−0.13 (0.03; −0.05 to 0.08)	.69	−0.01 (0.03; −0.07 to 0.05)	.75
Treatment[g]	−0.13 (0.04; −0.20 to −0.05)	<.001	−0.12 (0.04; −0.19 to −0.04)	.001
Self-efficacy	—	—	−0.03 (0.01; −0.04 to −0.02)	<.001
Depression (PHQ-9[h])				
Intercept	2.22 (—)	—	3.98 (—)	—
Time	−0.01 (0.05; −0.11 to −0.09)	.83	−0.20 (0.05; −0.12 to 0.08)	.69
Group	−0.04 (0.08; −0.20 to 0.12)	.65	−0.01 (0.08; −0.16 to 0.14)	.87
Treatment	−0.23 (0.08; −0.38 to −0.08)	.003	−0.21 (0.07; −0.35 to −0.06)	.006
Self-efficacy	—	—	−0.06 (0.01; −0.08 to −0.04)	<.001
Anxiety (GAD-7[i])				
Intercept	2.06 (—)	—	3.71 (—)	—
Time	−0.00 (0.06; −0.12 to 0.12)	.99	−0.00 (0.06; −0.12 to 0.11)	.94
Group	−0.02 (0.08; −0.18 to 0.14)	.81	−0.01 (0.08; −0.17 to 0.14)	.91
Treatment	−0.14 (0.09; −0.31 to 0.04)	.13	−0.11 (0.09; −0.28 to 0.06)	.22
Psychosomatic symptoms (PHQ-15[j])				
Intercept	2.33 (—)	—	3.90 (—)	—
Time	−0.01 (0.04; −0.08 to 0.61)	.77	−0.01 (0.04; −0.08 to 0.06)	.78
Group	−0.04 (0.07; −0.18 to 0.11)	.60	−0.03 (0.07; −0.17 to 0.11)	.68
Treatment	−0.16 (0.06; −0.27 to −0.06)	.003	−0.15 (0.06; −0.26 to −0.04)	.007
Self-efficacy	—	—	−0.06 (0.01; −0.08 to −0.03)	<.001
Active coping (HAPA[k])				
Intercept	0.89 (—)	—	0.69 (—)	—
Time	0.05 (0.05; −0.03 to 0.14)	.23	0.05 (0.05; −0.04 to 0.14)	.24
Group	−0.06 (0.06; −0.17 to 0.05)	.28	−0.06 (0.06; −0.17 to 0.05)	.26
Treatment	0.11 (0.07; −0.02 to 0.25)	.09	0.12 (0.07; −0.02 to 0.25)	.09
Self-efficacy	—	—	0.01 (0.01; −0.01 to 0.02)	.39

[a]Model 1: unadjusted model (without covariate).

[b]Model 2: adjusted model for general self-efficacy expectancy.

[c]PSS-10: Perceived Stress Scale-10.

[d]Not applicable.

[e]Time effect represents the rate of improvement for both the intervention and waitlist control groups.

[f]Group effect represents intervention or waitlist control group.

[g]Treatment effect is represented by group and time interaction.

[h]PHQ-9: Patient Health Questionnaire-9.

[i]GAD-7: Generalized Anxiety Disorder-7.

[j]PHQ-15: Patient Health Questionnaire-15.

[k]HAPA: health action process approach.

Exploratory

Regarding the working alliance, participants in the intervention group reported a mean working alliance score of 4.23 (SD 0.89) after the intervention. When exploring the potential influence of the working alliance on changes in outcomes from pre- to postintervention time points, we did not find evidence for correlations on any of the outcomes (Pearson correlation *r* ranging from −0.021 to 0.223). The participants rated their subjective stress expertise and goal achievement throughout the coaching program (3 times). For goal achievement, we observed a significant increase from the first to the third measurement with a large effect size (Cohen *d*=−1.07). Table 4 provides further details.

Table 4. Means for subscales bond, task, and goal of working alliance and results of a paired *t* test for stress expertise and goal achievement.

	Start of the intervention, mean (SD)	End of the intervention, mean (SD)	*t* test (*df*)	*P* value[a]	Cohen *d* (95% CI)
WAI-SR[b](n=42)					
Total	—[c]	4.23 (0.89)	—	—	—
Bond	—	4.20 (1.01)	—	—	—
Task	—	4.18 (0.82)	—	—	—
Goal	—	4.30 (0.84)	—	—	—
Stress expertise (n=45)	7.51 (1.47)	7.64 (1.60)	0.47 (44)	.64	−0.07 (−0.36 to 0.22)
Goal achievement (n=24)	3.88 (2.54)	6.71 (2.14)	−5.24 (23)	<.001	−1.07 (−1.57 to −0.56)

[a]Within group comparison: start of intervention versus end of intervention.

[b]WAI-SR: Working Alliance Inventory with Likert scale ranging from 1 to 7.

[c]Not applicable.

Engagement and Acceptance

In the intervention group, 60% (42/70) of the participants finished the coaching program by completing the postintervention survey (completers) and used the intervention as intended. Although the Little test indicated that values were MCAR for perceived stress (χ^2_1=0.5; *P*=.47), depression (χ^2_1=0.2; *P*=.63), anxiety (χ^2_1=2.0; *P*=.16), psychosomatic symptoms (χ^2_1=0.6; *P*=.80), and active coping (χ^2_1=0.1; *P*=.82), we conducted a dropout analysis due to the potential risk of differential attrition, particularly with significantly higher dropouts observed in the intervention group [94]. The analysis revealed no significant differences in outcomes (eg, stress and depression) or demographics (ie, gender and age) between completers and dropouts.

Overall, 45% (19/42) of the completers worked through all 13 sessions, played a mean of 86.52 (SD 120.54) minutes of relaxation audios, and received a mean of 115.88 (SD 5.06) reminders; Table 5 provides further information. On average, MISHA sent 400 (SD 205.61) SMS text messages and participants answered a mean of 297.54 (SD 169.80) SMS text messages, resulting in an average engagement ratio of 74.3%.

The participants in the intervention group (42/70, 60%) rated the subscale information highest, with a mean of 4.26 (SD 0.46), followed by engagement (mean 3.42, SD 0.70), perceived impact (mean 3.35, SD 0.87), and subjective quality of the app (mean 2.99, SD 0.87). Regarding engagement, individual customization was rated lowest with a mean of 2.71 (SD 0.84), while the target group fit was perceived as high (mean 3.95, SD 0.90). The participants liked the visual information of the CA and rated it high regarding correctness, clarity, and logic (mean 4.45, SD 0.55). Only a few participants (2/42, 2%) showed a high willingness to pay for the app (mean 2.10, SD 0.91) or anticipated high future use (mean 2.98, SD 1.05). The recommendation of the app to others was good, with a mean of 3.43 (SD 1.19) within the subjective app quality scale.

Table 5. Indicators of engagement: intended use, session completion, relaxation, and reminders.

Indicators of engagement	Values, n (%)	Relaxation applied, mean (SD)	Reminders received, mean (SD)[a]
Completers (intended use[b])	42 (100)	86.52 (120.54)	15.88 (5.06)
Completed all sessions	19 (45)	97.11 (71.43)	14.89 (5.75)
Stopped interacting after session 12	3 (7)	299.33 (377.33)	17.67 (1.15)
Stopped interacting after session 11	3 (7)	29.00 (26.91)	17.67 (0.58)
Stopped interacting after session 10	1 (2)	12.00 (0.0)	27.00 (0.0)
Stopped interacting after session 9	2 (5)	46.00 (26.87)	21.00 (1.41)
Stopped interacting after session 8	4 (10)	53.50 (42.45)	16.00 (3.92)
Stopped interacting after session 7	3 (7)	89.33 (82.25)	18.00 (1.73)
Stopped interacting after session 6	3 (7)	56.00 (73.53)	17.00 (1.0)
Stopped interacting after session 5	1 (2)	26.00 (0.0)	14.00 (0.0)
Stopped interacting after session 4	1 (2)	16.00 (0.0)	9.00 (0.0)
Stopped interacting after session 3	2 (5)	4.00 (4.24)	8.50 (2.12)
Stopped interacting after session 2	0	—[c]	—
Stopped interacting after session 1	0	—	—

[a]Reminders in case of inactivity during sessions.

[b]Intended use is defined by completing the postintervention survey, regardless of number of sessions that were completed.

[c]Not applicable.

Qualitative Feedback

The participants in the intervention group had the opportunity to provide free-text responses regarding their positive feedback on the CA intervention (Figure 3) and suggestions for improvement (Figure 4). The number of responses is displayed within the circles.

Figure 3. Thematic map of positive participant feedback.

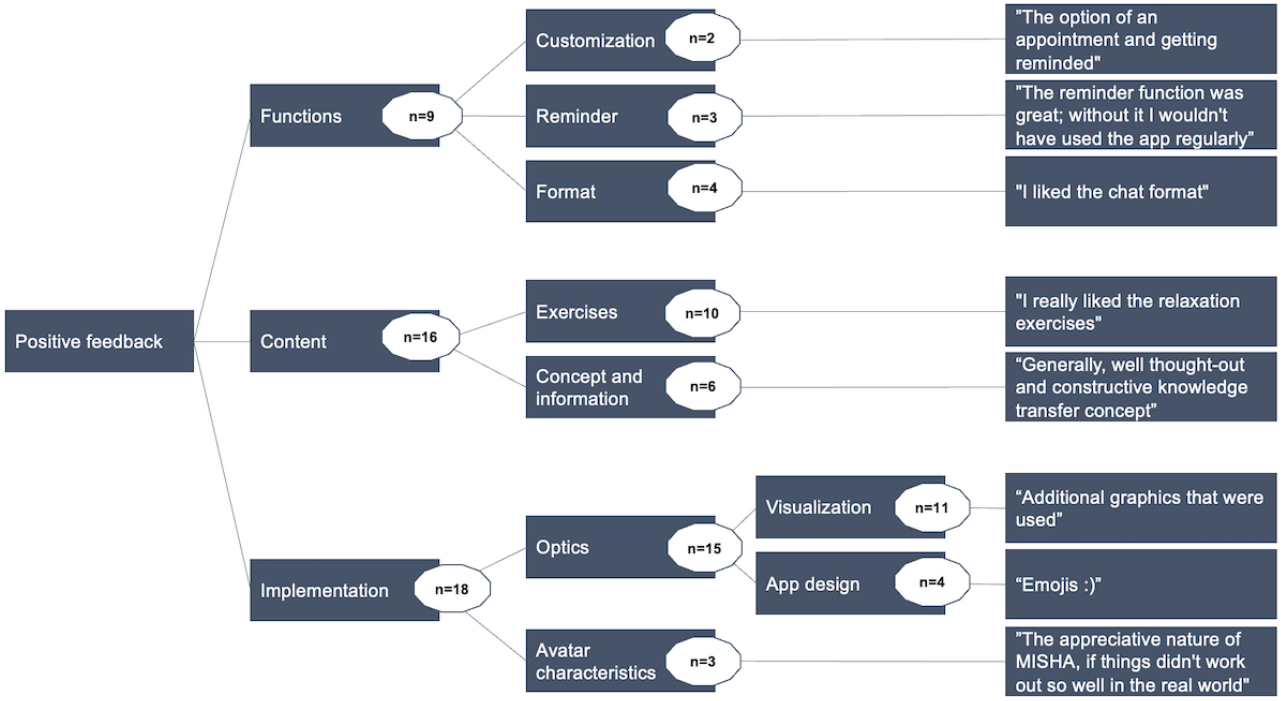
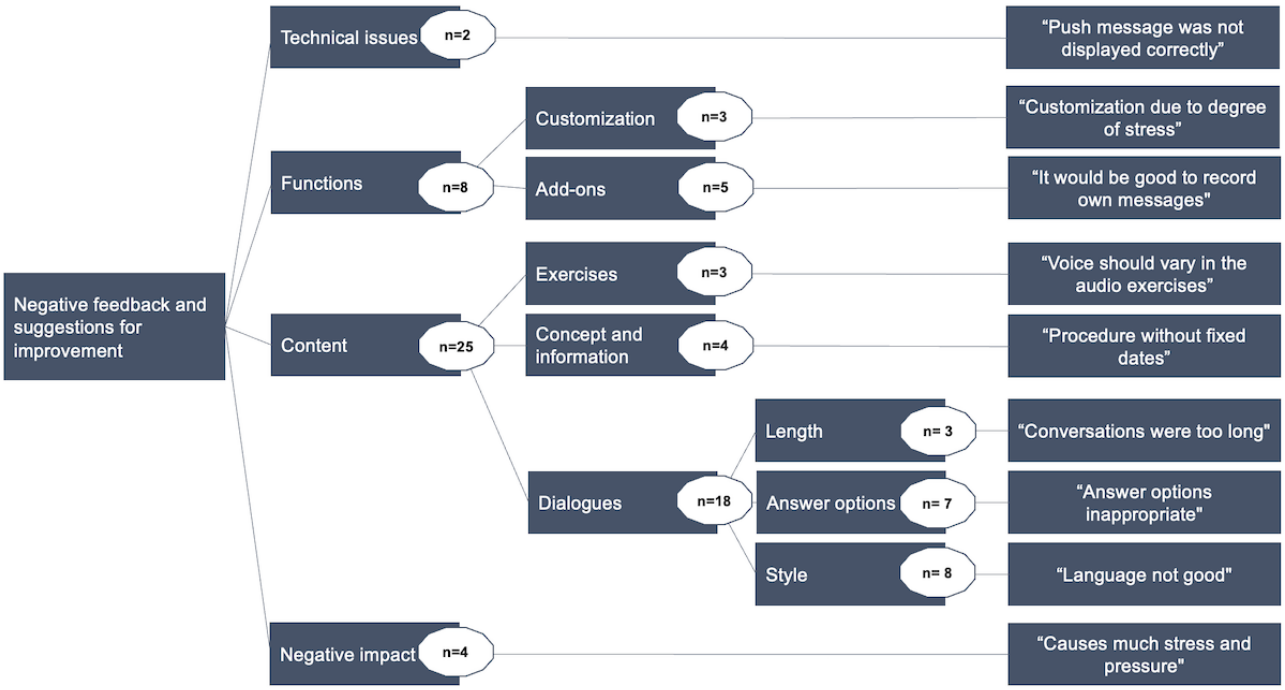


Figure 4. Thematic map of negative participant feedback.



Discussion

Principal Findings

This study aimed to describe the development and evaluation of the effectiveness of the MISHA app, a rule-based, CA-delivered stress management coaching intervention specifically tailored to the living environment of students. We described the MISHA app's evidence-based design and systematic evaluation. In both the PP and ITT analyses, we found evidence of decreased stress levels among participants in the intervention group compared to those in the control group, with a medium to large between-group effect (PP: Cohen d=−0.60). In addition, we observed evidence of a reduction in depressive symptoms with a medium to large effect (Cohen d=−0.50) as well as in psychosomatic symptoms with a small to medium effect (Cohen d=−0.36), while anxiety and active coping did not change. In the ITT analyses, a weak relation was found between self-efficacy and perceived stress, depression, anxiety, and psychosomatic symptoms, while the treatment effect persisted for stress, depression, and psychosomatic symptoms.

Our findings are consistent with other studies evaluating CA effectiveness in nonclinical populations. For instance, a study on CA Shim [95] among young adults with stress, despite a small sample size, reported stress reduction and improved psychological well-being, mirroring our results. Another study by Maciejewski and Smoktunowicz [52] assessed Stressbot, a 7-day messenger CA intervention aimed at enhancing coping self-efficacy among university students. Initial results showed reduced stress levels and improved self-efficacy postintervention. A large single-arm study evaluated Viki, an instant-messenger platform-based intervention [96], and found reduced stress and depressive symptoms. However, unlike our study, they reported a significant decrease in anxiety. In our study, the concurrent COVID-19 pandemic situation or upcoming examinations may have triggered increased uncertainties and fears. In a study involving CA Atena [53], the overall reduction in anxiety and stress levels may not have been substantial; however, the intervention showed promise in supporting individuals with high stress levels during the COVID-19 pandemic. Another study evaluated an AI-driven CA with the aim of reducing depression in university students by reflecting on their emotions, thoughts, and behavior [54]. The authors found reduced levels of depression and anxiety in the intervention group.

With its strong focus on goal setting, a crucial element in coaching [97], and being based on a behavior change model [74], the MISHA coaching intervention appears to effectively help students manage their stress. Toward the end of the coaching program, participants significantly rated their goal achievement higher with a large effect (Cohen d=−1.07), indicating the intervention's effectiveness in this regard. However, some participants expressed a desire for customization options, particularly regarding stress levels.

Regarding evidence from mHealth interventions for students, a study by Yang et al [30] found positive effects on stress and overall well-being in a 30-day app-based intervention on stress management through mindfulness meditation among medical students. A systematic review confirmed that digital interventions for the enhancement of mental well-being among college students can be effective in improving depression, anxiety, and mental well-being [98].

Given the mixed findings regarding the impact of self-efficacy expectancy on stress interventions targeting students [75-77], we explored whether self-efficacy was related to perceived stress. We found only a weak relation, while the treatment effectiveness remained unchanged. Therefore, in this study,

self-efficacy does not seem to have influenced the treatment's effectiveness in reducing stress.

In line with other studies [30,99,100], participants formed a working alliance with CA MISHA. Qualitative analyses revealed participants' appreciation for MISHA's supportive nature, especially during challenging moments. Most participants enjoyed interacting with MISHA, found the information provided appropriate, and expressed increased intention to change their behavior related to stress. Some desired additional features (eg, voice recording), found answer options or language style to be inappropriate, and disliked the lengthy dialogues. The various exercises, reminders, and visualizations were perceived as positive, and the constructive knowledge transfer was appreciated. In summary, it appears that a CA could be a well-accepted medium for stress prevention measures among students.

Limitations

This study has several limitations. First, despite statistically significant findings, it is essential to recognize that the absolute improvement in perceived stress, depressive symptoms, and psychosomatic symptoms was small. However, these improvements may still hold clinical relevance, and students experiencing even slight relief from perceived stress can benefit from CA-based coaching. Medium effect sizes indicate practical significance but may not always translate into substantial clinical change, and results should be interpreted with caution and in light of the context. Furthermore, all participants were self-selected, which limits the generalizability of our findings and introduces the potential for self-selection bias. Participants may have a particular interest in the subject and, therefore, cannot be considered a representative sample. It is important to note that their preexisting characteristics may differ from those in the broader population, and caution should be exercised when generalizing these findings to a wider context. Furthermore, this study is based on a convenience sample and should not be considered representative of all students. In particular, our sample, with most studying psychology (123/140, 87.9%) and predominantly woman participants (103/140, 73.6%), does not accurately reflect the student population in Switzerland, which shows an approximately even gender distribution (53% woman) [101]. Therefore, questions remain regarding the accessibility of the intervention to individuals who may not have an interest in psychological content and whether men and women can be equally reached by a mindfulness-focused chatbot such as MISHA.

Second, regarding engagement, we have analyzed use data from the intervention group, including completion rates, session completion, SMS text message response rate, reminders, and use of media player for relaxation. These objective measures offer valuable insights into participants' interactions with the coaching program and help ensure the robustness of our findings. However, it is difficult to measure how devoted participants were when using the app. To date, there is no consensus on measuring engagement in digital interventions [81]. According to Perski et al [102], engagement can be defined as a multidimensional construct that can be measured using self-reported outcomes, use data, or even psychophysical parameters. Future research should assess participants' time and motivation for offline engagement with exercises, while considering aspects of attention, interest, and emotions. Furthermore, in-depth use data should be gathered to assess the association between engagement, effectiveness, and optimal intended use.

Third, in this study, participants established a working alliance with the CA. However, it is important to acknowledge that CAs lack humanlike empathy or emotions [103]. They may struggle to understand the nuances of human language and lack the emotional intelligence and personal experience of a human, even if they can express empathy-like utterances. A recent study demonstrated that human-AI collaboration outperformed human-to-human collaboration, leading to a 19.6% increase in empathy in peer-to-peer text-based mental health support conversations [104]. While AI can mimic empathy and generate appropriate responses in text-based conversations, it is important to remember that these are still artificial constructs.

Fourth, various technical limitations need to be listed. At the beginning of the intervention, there were technical difficulties related to the audio files of the relaxation exercises. Some exercises could not be played. In addition, several participants indicated that the app was not updating properly; however, this issue was resolved within a few days. Furthermore, there was a 2-day interruption at the beginning because a technical adjustment had to be made to ensure that the system could recognize completed sessions. It remains unclear whether these technical issues led to more dropouts, frustration, or nonuse of the exercises. Notably, the recording of the minutes of listened audio files did not function flawlessly. While audio minutes were measured, they must be interpreted with great caution due to uncertainty in measurement. In addition, if the display of push notifications on the mobile phone was not set as the default, some SMS text messages were displayed without text. The number of people for whom this was the case and whether it negatively affected adherence cannot be conclusively determined. Any reported bugs in MISHA were addressed by a member of the study team within a 24-hour timeframe. There were no reported instances of server downtime.

Fifth, it is important to recognize the potential for improvements to enhance interaction in MISHA. The nature of the current CA is rule based: while allowing for evidence-based program development, the flexibility of interaction is limited by predefined answer options. While participants appreciated various aspects such as visualization, reminders, or exercises, personalized input via text input was missing, and some answer options were perceived as inappropriate. AI-based technology such as LLMs or natural language processing could be considered to improve text processing in MISHA. Natural language processing and LLM enable the CA to interpret user inputs more dynamically with increased natural interactions [105,106]. AI-based CAs are increasingly applied in health care to provide education and disease management. The literature on AI-based CAs indicates high overall performance and satisfactory user experience, high engagement, and positive health-related outcomes [107]. However, to date, CA interventions in the field of mental health are almost entirely rule based [108]. Ethical considerations concerning AI

technology should be addressed to mitigate potential misjudgments and risks. Research highlights the critical issue of inadequate transparency in data input and algorithms, undermining the reliability and validity of results [107,109]. Currently, both rule-based and LLM-based CAs are suitable for administering script-based interventions such as CBT elements, including psychoeducation, goal setting, or reflective tasks. While in the future, LLM-based interventions may be able to deliver more complex interventions in the field of psychology, it is crucial to consider the potential risk and limitations of implementing these technologies [110].

Sixth, it is important to acknowledge the possibility of a digital placebo effect [111]. In an unblinded trial, participants might attribute their improvements to the mere use of an mHealth intervention rather than its specific components. Expectations and engagement could introduce positive bias into the outcomes. Future research should carefully plan control conditions, which might include active control groups or sham interventions [111].

Conclusions

This paper outlines the evidence-based development of MISHA, a scalable coaching intervention specifically designed for students in their everyday life. The results of this study confirmed that CA-based coaching can be successfully delivered and is effective in reducing stress in students. It could not be confirmed that self-efficacy is related to the treatment effect. The establishment of a strong working alliance between participants and the CA, along with their perceived goal achievement, further reinforces the potential effectiveness of this intervention. Future research should involve students from diverse academic backgrounds, analyze effectiveness over time, incorporate active control groups, and improve user interaction. Overall, providing psychoeducation on stress, coupled with relaxation techniques, seems to empower students with effective tools and strategies for stress reduction.

Acknowledgments

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflicts of Interest

TK is affiliated with the Centre for Digital Health Interventions, a joint initiative of the Institute for Implementation Science in Health Care, University of Zurich; the Department of Management, Technology, and Economics at ETH Zurich; and the Institute of Technology Management and School of Medicine at the University of St Gallen. Centre for Digital Health Interventions is funded in part by CSS, a Swiss health insurer; Mavie Next, an Austrian health insurer; and MTIP, a Swiss digital health investor. Furthermore, TK is a cofounder of Pathmate Technologies, a university spin-off company that creates and delivers digital clinical pathways. However, neither CSS nor Pathmate Technologies was involved in this research.

Multimedia Appendix 1
Overview coaching content.
[PDF File (Adobe PDF File), 250 KB - mhealth_v12i1e54945_app1.pdf]

Multimedia Appendix 2
Reminder escalation.
[PDF File (Adobe PDF File), 168 KB - mhealth_v12i1e54945_app2.pdf]

Multimedia Appendix 3
Outcome and time points.
[PDF File (Adobe PDF File), 120 KB - mhealth_v12i1e54945_app3.pdf]

Multimedia Appendix 4
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1213 KB - mhealth_v12i1e54945_app4.pdf]

References

1. Beiter R, Nash R, McCrady M, Rhoades D, Linscomb M, Clarahan M, et al. The prevalence and correlates of depression, anxiety, and stress in a sample of college students. J Affect Disord 2015 Mar 01;173:90-96. [doi: 10.1016/j.jad.2014.10.054] [Medline: 25462401]
2. Löwe B, Spitzer RL, Zipfel S, Herzog W. PHQ-D gesundheitsfragebogen für patienten. Pfizer. 2002. URL: https://www.klinikum.uni-heidelberg.de/fileadmin/Psychosomatische_Klinik/download/PHQ_Kurzanleitung1.pdf [accessed 2022-09-03]
3. Stress in America: money, inflation, war pile on to nation stuck in COVID-19 survival mode. American Psychological Association. URL: https://www.apa.org/news/press/releases/stress/2022/march-2022-survival-mode [accessed 2022-08-08]

4. Gesundheit der studierenden an den schweizer hochschulen: 1 gesundheitszustand der studierenden. Bundesamt für Statistik. 2018. URL: https://www.swissstats.bfs.admin.ch/collection/ch.admin.bfs.swissstat.de.issue181518601600/article/issue181518601600-05 [accessed 2023-12-17]
5. Hofmann FH, Sperth M, Holm-Hadulla RM. Psychische belastungen und probleme studierender. Psychotherapeut 2017 Aug 28;62(5):395-402. [doi: 10.1007/s00278-017-0224-6]
6. Pogrebtsova E, Craig J, Chris A, O'Shea D, González-Morales MG. Exploring daily affective changes in university students with a mindful positive reappraisal intervention: a daily diary randomized controlled trial. Stress Health 2018 Mar 17;34(1):46-58. [doi: 10.1002/smi.2759] [Medline: 28512859]
7. Shaffique S, Farooq SS, Anwar H, Asif HM, Akram M, Jung SK. Meta-analysis of prevalence of depression, anxiety and stress among university students. RADS J Biol Res Appl Sci 2020 Jun;11(1):27-32. [doi: 10.37962/jbas.v11i1.308]
8. Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition. Washington, DC: American Psychiatric Association; 2013.
9. Auerbach RP, Mortier P, Bruffaerts R, Alonso J, Benjet C, Cuijpers P, et al. WHO World Mental Health Surveys International College Student Project: prevalence and distribution of mental disorders. J Abnorm Psychol 2018 Oct;127(7):623-638 [FREE Full text] [doi: 10.1037/abn0000362] [Medline: 30211576]
10. Gusy B, Lesener T, Wolter C. Burnout bei studierenden. PiD 2018 Sep 03;19(03):90-94. [doi: 10.1055/a-0556-2588]
11. Eicher V, Staerklé C, Clémence A. I want to quit education: a longitudinal study of stress and optimism as predictors of school dropout intention. J Adolesc 2014 Oct 15;37(7):1021-1030. [doi: 10.1016/j.adolescence.2014.07.007] [Medline: 25128662]
12. Middendorff E, Apolinarski B, Becker K, Bornkessel P, Brandt T, Heissenberg S, et al. Die wirtschaftliche und soziale lage der studierenden in Deutschland 2016. Bundesministerium für Bildung und Forschung. 2017. URL: https://www.bmbf.de/SharedDocs/Publikationen/de/bmbf/4/31338_21_Sozialerhebung_2016_Zusammenfassung.pdf?__blob=publicationFile&v=3 [accessed 2022-09-08]
13. Kumaraswamy N. Academic stress, anxiety and depression among college students - a brief review. Int Rev Soc Sci Humanit 2013;1(5):135-143.
14. Bland HW, Melton BF, Welle P, Bigham L. Stress tolerance: new challenges for millennial college students. Coll Stud J 2012;46(2):362-375. [doi: 10.1037/t39417-000]
15. Mackenzie S, Wiegel JR, Mundt M, Brown D, Saewyc E, Heiligenstein E, et al. Depression and suicide ideation among students accessing campus health care. Am J Orthopsychiatry 2011 Jan;81(1):101-107 [FREE Full text] [doi: 10.1111/j.1939-0025.2010.01077.x] [Medline: 21219281]
16. Ehrentreich S, Metzner L, Deraneck S, Blavutskaya Z, Tschupke S, Hasseler M. Einflüsse der coronapandemie auf gesundheitsbezogene verhaltensweisen und belastungen von studierenden [Article in German]. Präv Gesundheitsf 2021 Aug 23;17(3):364-369. [doi: 10.1007/s11553-021-00893-2]
17. Ackermann E, Schumann W. Die Uni ist kein Ponyhof: zur psychosozialen situation von studierenden. Praev Gesundheitsf 2010 Apr 22;5(3):231-237. [doi: 10.1007/s11553-010-0234-5]
18. Mak WW, Tong AC, Yip SY, Lui WW, Chio FH, Chan AT, et al. Efficacy and moderation of mobile app-based programs for mindfulness-based training, self-compassion training, and cognitive behavioral psychoeducation on mental health: randomized controlled noninferiority trial. JMIR Ment Health 2018 Oct 11;5(4):e60 [FREE Full text] [doi: 10.2196/mental.8597] [Medline: 30309837]
19. Yusufov M, Nicoloro-SantaBarbara J, Grey NE, Moyer A, Lobel M. Meta-analytic evaluation of stress reduction interventions for undergraduate and graduate students. Int J Stress Manag 2019 May;26(2):132-145. [doi: 10.1037/str0000099]
20. Kaluza G. Stressbewältigung: Trainingsmanual Zur Psychologischen Gesundheitsförderung. Berlin, Heidelberg: Springer; 2018.
21. Meichenbaum D. Intervention bei Stress: Anwendung und Wirkung des Stressimpfungstrainings. Berne, Switzerland: Hogrefe AG; 2012.
22. Reschke K, Schröder H. Optimistisch den Stress meistern: Ein Programm für Gesundheitsförderung, Therapie und Rehabilitation. Tübingen, Germany: dgvt-Verlag; 2010.
23. Kaluza G. Psychologische gesundheitsförderung und prävention im erwachsenenalter. Zeitschrift für Gesundheitspsychologie 2006 Oct;14(4):171-196. [doi: 10.1026/0943-8149.14.4.171]
24. Marsh CN, Wilcoxon SA. Underutilization of mental health services among college students: an examination of system-related barriers. J Coll Stud Psychother 2015 Jul 07;29(3):227-243. [doi: 10.1080/87568225.2015.1045783]
25. Song X, Anderson T, Himawan L, McClintock A, Jiang Y, McCarrick S. An investigation of a cultural help-seeking model for professional psychological services with U.S. and Chinese samples. J Cross Cult Psychol 2019 Oct 14;50(9):1027-1049. [doi: 10.1177/0022022119878506]
26. Figueroa CA, Aguilera A. The need for a mental health technology revolution in the COVID-19 pandemic. Front Psychiatry 2020 Jun 3;11:523 [FREE Full text] [doi: 10.3389/fpsyt.2020.00523] [Medline: 32581891]
27. Hunt J, Eisenberg D. Mental health problems and help-seeking behavior among college students. J Adolesc Health 2010 Jan;46(1):3-10. [doi: 10.1016/j.jadohealth.2009.08.008] [Medline: 20123251]

28. Weisel KK, Fuhrmann LM, Berking M, Baumeister H, Cuijpers P, Ebert DD. Standalone smartphone apps for mental health-a systematic review and meta-analysis. NPJ Digit Med 2019 Dec 2;2(1):118 [FREE Full text] [doi: 10.1038/s41746-019-0188-8] [Medline: 31815193]
29. Laux G. Online-/internet-programme zur psychotherapie bei depression - eine zwischenbilanz. J Neurologie Neurochirurgie Psychiatrie 2017;18(1):16-24.
30. Yang E, Schamber E, Meyer RM, Gold JI. Happier healers: randomized controlled trial of mobile mindfulness for stress management. J Altern Complement Med 2018 May;24(5):505-513. [doi: 10.1089/acm.2015.0301] [Medline: 29420050]
31. Sun S, Lin D, Goldberg S, Shen Z, Chen P, Qiao S, et al. A mindfulness-based mobile health (mHealth) intervention among psychologically distressed university students in quarantine during the COVID-19 pandemic: a randomized controlled trial. J Couns Psychol 2022 Mar;69(2):157-171 [FREE Full text] [doi: 10.1037/cou0000568] [Medline: 34264696]
32. Schulte-Frankenfeld PM, Trautwein FM. App-based mindfulness meditation reduces perceived stress and improves self-regulation in working university students: a randomised controlled trial. Appl Psychol Health Well Being 2022 Nov;14(4):1151-1171 [FREE Full text] [doi: 10.1111/aphw.12328] [Medline: 34962055]
33. Linardon J, Fuller-Tyszkiewicz M. Attrition and adherence in smartphone-delivered interventions for mental health problems: a systematic and meta-analytic review. J Consult Clin Psychol 2020 Jan;88(1):1-13. [doi: 10.1037/ccp0000459] [Medline: 31697093]
34. Dhinagaran DA, Martinengo L, Ho MH, Joty S, Kowatsch T, Atun R, et al. Designing, developing, evaluating, and implementing a smartphone-delivered, rule-based conversational agent (DISCOVER): development of a conceptual framework. JMIR Mhealth Uhealth 2022 Oct 04;10(10):e38740 [FREE Full text] [doi: 10.2196/38740] [Medline: 36194462]
35. Tudor Car L, Dhinagaran DA, Kyaw BM, Kowatsch T, Joty S, Theng YL, et al. Conversational agents in health care: scoping review and conceptual analysis. J Med Internet Res 2020 Aug 07;22(8):e17158 [FREE Full text] [doi: 10.2196/17158] [Medline: 32763886]
36. He Y, Yang L, Qian C, Li T, Su Z, Zhang Q, et al. Conversational agent interventions for mental health problems: systematic review and meta-analysis of randomized controlled trials. J Med Internet Res 2023 Apr 28;25:e43862 [FREE Full text] [doi: 10.2196/43862] [Medline: 37115595]
37. Bird T, Mansell W, Wright J, Gaffney H, Tai S. Manage your life online: a web-based randomized controlled trial evaluating the effectiveness of a problem-solving intervention in a student sample. Behav Cogn Psychother 2018 Sep 25;46(5):570-582 [FREE Full text] [doi: 10.1017/S1352465817000820] [Medline: 29366432]
38. Vaidyam AN, Wisniewski H, Halamka JD, Kashavan MS, Torous JB. Chatbots and conversational agents in mental health: a review of the psychiatric landscape. Can J Psychiatry 2019 Jul;64(7):456-464 [FREE Full text] [doi: 10.1177/0706743719828977] [Medline: 30897957]
39. Kramer JN, Künzler F, Mishra V, Smith SN, Kotz D, Scholz U, et al. Which components of a smartphone walking app help users to reach personalized step goals? results from an optimization trial. Ann Behav Med 2020 Jun 12;54(7):518-528 [FREE Full text] [doi: 10.1093/abm/kaaa002] [Medline: 32182353]
40. Hauser-Ulrich S, Künzli H, Meier-Peterhans D, Kowatsch T. A smartphone-based health care chatbot to promote self-management of chronic pain (SELMA): pilot randomized controlled trial. JMIR Mhealth Uhealth 2020 Apr 03;8(4):e15806. [doi: 10.2196/15806] [Medline: 32242820]
41. Prochaska JJ, Vogel EA, Chieng A, Kendra M, Baiocchi M, Pajarito S, et al. A therapeutic relational agent for reducing problematic substance use (Woebot): development and usability study. J Med Internet Res 2021 Mar 23;23(3):e24850 [FREE Full text] [doi: 10.2196/24850] [Medline: 33755028]
42. Haug S, Paz Castro R, Scholz U, Kowatsch T, Schaub MP, Radtke T. Assessment of the efficacy of a mobile phone-delivered just-in-time planning intervention to reduce alcohol use in adolescents: randomized controlled crossover trial. JMIR Mhealth Uhealth 2020 May 26;8(5):e16937 [FREE Full text] [doi: 10.2196/16937] [Medline: 32452818]
43. Abd-Alrazaq AA, Rababeh A, Alajlani M, Bewick BM, Househ M. Effectiveness and safety of using chatbots to improve mental health: systematic review and meta-analysis. J Med Internet Res 2020 Jul 13;22(7):e16021 [FREE Full text] [doi: 10.2196/16021] [Medline: 32673216]
44. Ma T, Sharifi H, Chattopadhyay D. Virtual humans in health-related interventions: a meta-analysis. In: Proceedings of the Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems. 2019 Presented at: CHI EA '19; May 4-9, 2019; Glasgow, Scotland. [doi: 10.1145/3290607.3312853]
45. Gilbert S, Harvey H, Melvin T, Vollebregt E, Wicks P. Large language model AI chatbots require approval as medical devices. Nat Med 2023 Oct 30;29(10):2396-2398. [doi: 10.1038/s41591-023-02412-6] [Medline: 37391665]
46. Hastings J. Preventing harm from non-conscious bias in medical generative AI. Lancet Digit Health 2024 Jan;6(1):e2-e3 [FREE Full text] [doi: 10.1016/S2589-7500(23)00246-7] [Medline: 38123253]
47. Goldberg CB, Adams L, Blumenthal D, Brennan PF, Brown N, Butte AJ, et al. To do no harm — and the most good — with AI in health care. NEJM AI 2024 Feb 22;1(3). [doi: 10.1056/AIp2400036]
48. Ollier J, Neff S, Dworschak C, Sejdiji A, Santhanam P, Keller R, et al. Elena+ care for COVID-19, a pandemic lifestyle care intervention: intervention design and study protocol. Front Public Health 2021 Oct 21;9:625640 [FREE Full text] [doi: 10.3389/fpubh.2021.625640] [Medline: 34746067]

49. Stieger M, Flückiger C, Rüegger D, Kowatsch T, Roberts BW, Allemand M. Changing personality traits with the help of a digital personality change intervention. Proc Natl Acad Sci U S A 2021 Mar 23;118(8):e2017548118 [FREE Full text] [doi: 10.1073/pnas.2017548118] [Medline: 33558417]
50. Castro O, Mair JL, Salamanca-Sanabria A, Alattas A, Keller R, Zheng S, et al. Development of "LvL UP 1.0": a smartphone-based, conversational agent-delivered holistic lifestyle intervention for the prevention of non-communicable diseases and common mental disorders. Front Digit Health 2023 May 10;5:1039171 [FREE Full text] [doi: 10.3389/fdgth.2023.1039171] [Medline: 37234382]
51. Ulrich S, Gantenbein AR, Zuber V, Von Wyl A, Kowatsch T, Künzli H. Development and evaluation of a smartphone-based chatbot coach to facilitate a balanced lifestyle in individuals with headaches (BalanceUP App): randomized controlled trial. J Med Internet Res 2024 Jan 24;26:e50132 [FREE Full text] [doi: 10.2196/50132] [Medline: 38265863]
52. Maciejewski J, Smoktunowicz E. Low-effort internet intervention to reduce students' stress delivered with Meta's Messenger chatbot (Stressbot): a randomized controlled trial. Internet Interv 2023 Sep;33:100653 [FREE Full text] [doi: 10.1016/j.invent.2023.100653] [Medline: 37575678]
53. Gabrielli S, Rizzi S, Bassi G, Carbone S, Maimone R, Marchesoni M, et al. Engagement and effectiveness of a healthy-coping intervention via chatbot for university students during the COVID-19 pandemic: mixed methods proof-of-concept study. JMIR Mhealth Uhealth 2021 May 28;9(5):e27965 [FREE Full text] [doi: 10.2196/27965] [Medline: 33950849]
54. Liu H, Peng H, Song X, Xu C, Zhang M. Using AI chatbots to provide self-help depression interventions for university students: a randomized trial of effectiveness. Internet Interv 2022 Mar;27:100495 [FREE Full text] [doi: 10.1016/j.invent.2022.100495] [Medline: 35059305]
55. Regehr C, Glancy D, Pitts A. Interventions to reduce stress in university students: a review and meta-analysis. J Affect Disord 2013 May 15;148(1):1-11. [doi: 10.1016/j.jad.2012.11.026] [Medline: 23246209]
56. MobileCoach homepage. MobileCoach. URL: https://www.mobile-coach.eu/ [accessed 2021-11-27]
57. Haug S, Paz Castro R, Kowatsch T, Filler A, Dey M, Schaub MP. Efficacy of a web- and text messaging-based intervention to reduce problem drinking in adolescents: results of a cluster-randomized controlled trial. J Consult Clin Psychol 2017 Mar;85(2):147-159 [FREE Full text] [doi: 10.1037/ccp0000138] [Medline: 27606700]
58. Haug S, Paz Castro R, Kowatsch T, Filler A, Schaub MP. Efficacy of a technology-based, integrated smoking cessation and alcohol intervention for smoking cessation in adolescents: results of a cluster-randomised controlled trial. J Subst Abuse Treat 2017 Nov;82:55-66. [doi: 10.1016/j.jsat.2017.09.008] [Medline: 29021116]
59. Stieger M, Eck M, Rüegger D, Kowatsch T, Flückiger C, Allemand M. Who wants to become more conscientious, more extraverted, or less neurotic with the help of a digital intervention? J Res Personal 2020 Aug;87:103983. [doi: 10.1016/j.jrp.2020.103983]
60. Firebase homepage. Firebase. URL: https://firebase.google.com/?hl=de [accessed 2021-11-27]
61. Testing apps with TestFlight. Apple. URL: https://testflight.apple.com/join/oAvL0b5N [accessed 2021-11-27]
62. Seidl MH, Limberger MF, Ebner-Priemer UW. Entwicklung und evaluierung eines stressbewältigungsprogramms für studierende im hochschulsetting. Zeitschrift für Gesundheitspsychologie 2016 Jan;24(1):29-40. [doi: 10.1026/0943-8149/a000154]
63. Schwarzer R. Psychologie des Gesundheitsverhaltens: Einführung in die Gesundheitspsychologie. Göttingen, Germany: Hogrefe Verlag GmbH & Company KG; 2004.
64. Shah LB, Klainin-Yobas P, Torres S, Kannusamy P. Efficacy of psychoeducation and relaxation interventions on stress-related variables in people with mental disorders: a literature review. Arch Psychiatr Nurs 2014 Apr;28(2):94-101. [doi: 10.1016/j.apnu.2013.11.004] [Medline: 24673782]
65. Wenzel A. Basic strategies of cognitive behavioral therapy. Psychiatr Clin North Am 2017 Dec;40(4):597-609. [doi: 10.1016/j.psc.2017.07.001] [Medline: 29080588]
66. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
67. Karekla M, Kasinopoulos O, Neto DD, Ebert DD, van Daele T, Nordgreen T, et al. Best practices and recommendations for digital interventions to improve engagement and adherence in chronic illness sufferers. Eur Psychol 2019 Jan;24(1):49-67. [doi: 10.1027/1016-9040/a000349]
68. Klein EM, Brähler E, Dreier M, Reinecke L, Müller KW, Schmutzer G, et al. The German version of the Perceived Stress Scale - psychometric characteristics in a representative German community sample. BMC Psychiatry 2016 May 23;16:159 [FREE Full text] [doi: 10.1186/s12888-016-0875-9] [Medline: 27216151]
69. Spitzer RL, Williams JB, Kroenke K. Patient Health Questionnaire--SADS (PHQ-SADS). APA PsycTests. URL: https://psycnet.apa.org/doiLanding?doi=10.1037%2Ft06164-000 [accessed 2024-05-30]
70. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
71. Löwe B, Decker O, Müller S, Brähler E, Schellberg D, Herzog W, et al. Validation and standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the general population. Med Care 2008 Mar;46(3):266-274. [doi: 10.1097/MLR.0b013e318160d093] [Medline: 18388841]

72. Kroenke K, Spitzer RL, Williams JB. The PHQ-15: validity of a new measure for evaluating the severity of somatic symptoms. Psychosom Med 2002;64(2):258-266. [doi: 10.1097/00006842-200203000-00008] [Medline: 11914441]
73. Spitzer RL, Williams JB, Kroenke K. Instruction manual: instructions for patient health questionnaire (PHQ) and GAD-7 measures. Primary Care Collaborative. URL: https://www.pcpcc.org/sites/default/files/resources/instructions.pdf [accessed 2023-02-07]
74. Schwarzer R. Modeling health behavior change: how to predict and modify the adoption and maintenance of health behaviors. Appl Psychol 2008 Jan 30;57(1):1-29. [doi: 10.1111/j.1464-0597.2007.00325.x]
75. Brenninkmeijer V, Lagerveld SE, Blonk RW, Schaufeli WB, Wijngaards-de Meij LD. Predicting the effectiveness of work-focused CBT for common mental disorders: the influence of baseline self-efficacy, depression and anxiety. J Occup Rehabil 2019 Mar;29(1):31-41. [doi: 10.1007/s10926-018-9760-3] [Medline: 29450678]
76. Gençoğlu C, Şahin E, Topkaya N. General self-efficacy and forgiveness of self, others, and situations as predictors of depression, anxiety, and stress in university students. Educ Sci Theory Pract 2018;18(3):605-626. [doi: 10.12738/estp.2018.3.0128]
77. Büttner TR, Dlugosch GE. Stress im studium: die rolle der selbstwirksamkeitserwartung und der achtsamkeit im stresserleben von studierenden. Präv Gesundheitsf 2013 Jan 18;8:106-111. [doi: 10.1007/s11553-012-0369-7]
78. Schwarzer R. Self-Efficacy: Thought Control Of Action. Milton Park, UK: Taylor & Francis; 1992.
79. Munder T, Wilmers F, Leonhart R, Linster HW, Barth J. Working Alliance Inventory-Short Revised (WAI-SR): psychometric properties in outpatients and inpatients. Clin Psychol Psychother 2010;17(3):231-239. [doi: 10.1002/cpp.658] [Medline: 20013760]
80. Kelders SM, Kok RN, Ossebaard HC, van Gemert-Pijnen JE. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012 Nov 14;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
81. Sieverink F, Kelders SM, van Gemert-Pijnen JE. Clarifying the concept of adherence to eHealth technology: systematic review on when usage becomes adherence. J Med Internet Res 2017 Dec 06;19(12):e402 [FREE Full text] [doi: 10.2196/jmir.8578] [Medline: 29212630]
82. Moller AC, Merchant G, Conroy DE, West R, Hekler E, Kugler KC, et al. Applying and advancing behavior change theories and techniques in the context of a digital health revolution: proposals for more effectively realizing untapped potential. J Behav Med 2017 Mar;40(1):85-98 [FREE Full text] [doi: 10.1007/s10865-016-9818-7] [Medline: 28058516]
83. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
84. Christensen H, Mackinnon A. The law of attrition revisited. J Med Internet Res 2006 Sep 29;8(3):e20; author reply e21 [FREE Full text] [doi: 10.2196/jmir.8.3.e20] [Medline: 17032636]
85. Ryan RM, Deci EL. Self-Determination Theory: Basic Psychological Needs in Motivation, Development, and Wellness. New York, NY: Guilford Publications; 2017.
86. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and validation of the user version of the mobile application rating scale (uMARS). JMIR Mhealth Uhealth 2016 Jun 10;4(2):e72 [FREE Full text] [doi: 10.2196/mhealth.5849] [Medline: 27287964]
87. Linardon J, Cuijpers P, Carlbring P, Messer M, Fuller-Tyszkiewicz M. The efficacy of app-supported smartphone interventions for mental health problems: a meta-analysis of randomized controlled trials. World Psychiatry 2019 Oct;18(3):325-336 [FREE Full text] [doi: 10.1002/wps.20673] [Medline: 31496095]
88. Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007 May;39:175-191. [doi: 10.3758/BF03193146]
89. Torous J, Lipschitz J, Ng M, Firth J. Dropout rates in clinical trials of smartphone apps for depressive symptoms: a systematic review and meta-analysis. J Affect Disord 2020 Mar 15;263:413-419. [doi: 10.1016/j.jad.2019.11.167] [Medline: 31969272]
90. Zeger SL, Liang KY, Albert PS. Models for longitudinal data: a generalized estimating equation approach. Biometrics 1988 Dec;44(4):1049-1060. [doi: 10.2307/2531734]
91. Mayring P. Qualitative content analysis. Forum Qual Soc Res 2000;1 [FREE Full text] [doi: 10.17169/fqs-1.2.1089]
92. Mayring P. Qualitative inhaltsanalyse. In: Mey G, Mruck K, editors. Handbuch Qualitative Forschung in der Psychologie. Wiesbaden, Germany: Springer; 2020.
93. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
94. Goldberg SB, Bolt DM, Davidson RJ. Data missing not at random in mobile health research: assessment of the problem and a case for sensitivity analyses. J Med Internet Res 2021 Jun 15;23(6):e26749 [FREE Full text] [doi: 10.2196/26749] [Medline: 34128810]
95. Ly KH, Ly AM, Andersson G. A fully automated conversational agent for promoting mental well-being: a pilot RCT using mixed methods. Internet Interv 2017 Oct 10;10:39-46 [FREE Full text] [doi: 10.1016/j.invent.2017.10.002] [Medline: 30135751]

96. Daley K, Hungerbuehler I, Cavanagh K, Claro HG, Swinton PA, Kapps M. Preliminary evaluation of the engagement and effectiveness of a mental health chatbot. Front Digit Health 2020 Nov 30;2:576361 [FREE Full text] [doi: 10.3389/fdgth.2020.576361] [Medline: 34713049]
97. Clutterbuck D, Spence G. Working with goals in coaching. In: Bachkirova T, Spence G, Drake D, editors. SAGE Handbook of Coaching. Thousand Oaks, CA: SAGE Publications; 2016:218-237.
98. Lattie EG, Adkins EC, Winquist N, Stiles-Shields C, Wafford QE, Graham AK. Digital mental health interventions for depression, anxiety, and enhancement of psychological well-being among college students: systematic review. J Med Internet Res 2019 Jul 22;21(7):e12869 [FREE Full text] [doi: 10.2196/12869] [Medline: 31333198]
99. Bickmore TW, Mitchell SE, Jack BW, Paasche-Orlow MK, Pfeifer LM, Odonnell J. Response to a relational agent by hospital patients with depressive symptoms. Interact Comput 2010 Jul 01;22(4):289-298 [FREE Full text] [doi: 10.1016/j.intcom.2009.12.001] [Medline: 20628581]
100. Heim E, Rötger A, Lorenz N, Maercker A. Working alliance with an avatar: how far can we go with internet interventions? Internet Interv 2018 Mar;11:41-46 [FREE Full text] [doi: 10.1016/j.invent.2018.01.005] [Medline: 30135758]
101. Verteilung der Studierenden an Fachhochschulen in der Schweiz nach Geschlecht von 2011/2012 bis 2022/2023. Statista. URL: https://de.statista.com/statistik/daten/studie/306922/umfrage/verteilung-der-studierenden-an-fachhochschulen-in-der-schweiz-nach-geschlecht/ [accessed 2023-02-08]
102. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
103. Carlbring P, Hadjistavropoulos H, Kleiboer A, Andersson G. A new era in internet interventions: the advent of Chat-GPT and AI-assisted therapist guidance. Internet Interv 2023 Apr;32:100621 [FREE Full text] [doi: 10.1016/j.invent.2023.100621] [Medline: 37273936]
104. Sharma A, Lin IW, Miner AS, Atkins DC, Althoff T. Human–AI collaboration enables more empathic conversations in text-based peer-to-peer mental health support. Nat Mach Intell 2023 Jan 23;5(1):46-57. [doi: 10.1038/s42256-022-00593-2]
105. Montenegro JL, da Costa CA, da Rosa Righi R. Survey of conversational agents in health. Expert Syst Appl 2019 Sep;129:56-67 [FREE Full text] [doi: 10.1016/j.eswa.2019.03.054]
106. Suta P, Mongkolnam P, Chan JH, Lan X, Wu B. An overview of machine learning in chatbots. Int J Mech Eng Robot Res 2020;9(4):502-510. [doi: 10.18178/ijmerr.9.4.502-510]
107. Schachner T, Keller R, V Wangenheim F. Artificial intelligence-based conversational agents for chronic conditions: systematic literature review. J Med Internet Res 2020 Sep 14;22(9):e20701 [FREE Full text] [doi: 10.2196/20701] [Medline: 32924957]
108. Lim SM, Shiau CW, Cheng LJ, Lau Y. Chatbot-delivered psychotherapy for adults with depressive and anxiety symptoms: a systematic review and meta-regression. Behav Ther 2022 Mar;53(2):334-347. [doi: 10.1016/j.beth.2021.09.007] [Medline: 35227408]
109. Aggarwal A, Tam CC, Wu D, Li X, Qiao S. Artificial intelligence-based chatbots for promoting health behavioral changes: systematic review. J Med Internet Res 2023 Mar 24;25:e40789 [FREE Full text] [doi: 10.2196/40789] [Medline: 36826990]
110. Stade E, Stirman SW, Ungar LH, Boland CL, Schwartz HA, Yaden DB, et al. Large language models could change the future of behavioral healthcare: a proposal for responsible development and evaluation. PsyArXiv Preprints Preprint posted online on April 25, 2023 [FREE Full text] [doi: 10.31234/osf.io/cuzvr]
111. Torous J, Firth J. The digital placebo effect: mobile mental health meets clinical psychiatry. Lancet Psychiatry 2016 Mar;3(2):100-102. [doi: 10.1016/S2215-0366(15)00565-9] [Medline: 26851322]

Abbreviations

AI: artificial intelligence
CA: conversational agent
CBT: cognitive behavioral therapy
CONSORT: Consolidated Standards of Reporting Trials
CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
GEE: generalized estimating equation
HAPA: health action process approach
ITT: intention-to-treat
LLM: large language model
MCAR: missing completely at random
mHealth: mobile Health
PHQ-9: Patient Health Questionnaire-9
PP: per-protocol
T1: time point 1

T2: time point 2
uMARS: user version of the Mobile App Rating Scale

Edited by M Sobolev; submitted 04.12.23; peer-reviewed by E Mitsea, W Wei, I Liu, J Greist; comments to author 20.02.24; revised version received 05.04.24; accepted 03.05.24; published 26.06.24.

Please cite as:
Ulrich S, Lienhard N, Künzli H, Kowatsch T
A Chatbot-Delivered Stress Management Coaching for Students (MISHA App): Pilot Randomized Controlled Trial
JMIR Mhealth Uhealth 2024;12:e54945
URL: https://mhealth.jmir.org/2024/1/e54945
doi:10.2196/54945
PMID:38922677

Original Paper

Conversational Chatbot for Cigarette Smoking Cessation: Results From the 11-Step User-Centered Design Development Process and Randomized Controlled Trial

Jonathan B Bricker[1,2], PhD; Brianna Sullivan[1], MS; Kristin Mull[1], MS; Margarita Santiago-Torres[1], PhD; Juan M Lavista Ferres[3], PhD

[1]Division of Public Health Sciences, Fred Hutch Cancer Center, Seattle, WA, United States

[2]Department of Psychology, University of Washington, Seattle, WA, United States

[3]AI for Good Lab, Microsoft Corporation, Redmond, WA, United States

Corresponding Author:
Jonathan B Bricker, PhD
Division of Public Health Sciences
Fred Hutch Cancer Center
1100 Fairview Avenue N
Seattle, WA, 98109
United States
Phone: 1 206 667 5074
Email: jbricker@fredhutch.org

Abstract

Background: Conversational chatbots are an emerging digital intervention for smoking cessation. No studies have reported on the entire development process of a cessation chatbot.

Objective: We aim to report results of the user-centered design development process and randomized controlled trial for a novel and comprehensive quit smoking conversational chatbot called *QuitBot*.

Methods: The 4 years of formative research for developing QuitBot followed an 11-step process: (1) specifying a conceptual model; (2) conducting content analysis of existing interventions (63 hours of intervention transcripts); (3) assessing user needs; (4) developing the chat's persona ("personality"); (5) prototyping content and persona; (6) developing full functionality; (7) programming the QuitBot; (8) conducting a diary study; (9) conducting a pilot randomized controlled trial (RCT); (10) reviewing results of the RCT; and (11) adding a free-form question and answer (QnA) function, based on user feedback from pilot RCT results. The process of adding a QnA function itself involved a three-step process: (1) generating QnA pairs, (2) fine-tuning large language models (LLMs) on QnA pairs, and (3) evaluating the LLM outputs.

Results: We developed a quit smoking program spanning 42 days of 2- to 3-minute conversations covering topics ranging from motivations to quit, setting a quit date, choosing Food and Drug Administration–approved cessation medications, coping with triggers, and recovering from lapses and relapses. In a pilot RCT with 96% three-month outcome data retention, QuitBot demonstrated high user engagement and promising cessation rates compared to the National Cancer Institute's SmokefreeTXT text messaging program, particularly among those who viewed all 42 days of program content: 30-day, complete-case, point prevalence abstinence rates at 3-month follow-up were 63% (39/62) for QuitBot versus 38.5% (45/117) for SmokefreeTXT (odds ratio 2.58, 95% CI 1.34-4.99; *P*=.005). However, Facebook Messenger intermittently blocked participants' access to QuitBot, so we transitioned from Facebook Messenger to a stand-alone smartphone app as the communication channel. Participants' frustration with QuitBot's inability to answer their open-ended questions led to us develop a core conversational feature, enabling users to ask open-ended questions about quitting cigarette smoking and for the QuitBot to respond with accurate and professional answers. To support this functionality, we developed a library of 11,000 QnA pairs on topics associated with quitting cigarette smoking. Model testing results showed that Microsoft's Azure-based QnA maker effectively handled questions that matched our library of 11,000 QnA pairs. A fine-tuned, contextualized GPT-3.5 (OpenAI) responds to questions that are not within our library of QnA pairs.

Conclusions: The development process yielded the first LLM-based quit smoking program delivered as a conversational chatbot. Iterative testing led to significant enhancements, including improvements to the delivery channel. A pivotal addition was the inclusion of a core LLM–supported conversational feature allowing users to ask open-ended questions.

Trial Registration: ClinicalTrials.gov NCT03585231; https://clinicaltrials.gov/study/NCT03585231

(JMIR Mhealth Uhealth 2024;12:e57318) doi:10.2196/57318

KEYWORDS

chatbot; conversational agent; conversational agents; digital therapeutics; smoking cessation; development; develop; design; smoking; smoke; smokers; quit; quitting; cessation; chatbots; large language model; LLM; LLMs; large language models; addict; addiction; addictions; mobile phone

Introduction

Background

Cigarette smoking accounts for 8 million premature deaths and 25% of all cancer deaths annually [1,2]. Despite advancements in government policies, antismoking campaigns, and shifting societal norms, existing smoking cessation interventions continue to have limited treatment engagement and cessation rates [3-9]. While this is a problem for the general population of people who smoke, the issue is particularly pronounced in communities considered marginalized, synonymous with groups considered vulnerable or disadvantaged, which are segments of society facing systemic disadvantages and barriers in accessing resources and opportunities. Populations considered marginalized, marked by factors such as racial or ethnic minority status, sexual or gender identity differences, low education and income levels, higher unemployment rates, or an increased prevalence of mental illness, encounter discrimination, social exclusion, and limited influence in decision-making processes.

Challenges in treatment engagement and cessation efficacy across all communities of people who smoke are compounded by a scarcity of trained clinicians and significant barriers, including cost and lack of insurance, hindering access to existing clinician-delivered interventions [10-14]. Given that 1.3 billion people in the world smoke cigarettes, with 70% of them wanting to quit, it would be impractical to have enough trained clinicians to help people quit smoking [15,16]. Indeed, only 5% of cessation attempts are aided by a health professional [17]. Consequently, there is an enormous need for high-impact, cost-effective, population-level interventions for smoking cessation.

A well-documented finding from research on clinician-delivered treatments has emphasized the significance of therapeutic conversations as powerful drivers of patient engagement [18-21]. Therapeutic conversations, which form a social-emotional bond with the user, have predicted treatment engagement and, subsequently, health outcomes across various treatments and settings [12,18,22]. A new technology provides an opportunity to leverage engaging therapeutic conversations. Advances in machine learning, large language models (LLMs), and cloud computing are now making it possible to create and widely disseminate conversational chatbots for behavior change coaching.

Unlike the chatbots used in customer service contexts, conversational chatbots for behavior change coaching are designed to form long-term social-emotional connections with users, even as they are made aware that chatbots are merely computer software that use language to communicate with users [23,24]. Conversational chatbots for coaching are designed to be supportive and empathic, offering reflective listening, personalized responses, and timely advice aligned with the user's individual needs [25,26]. In the context of cessation, conversational chatbots can enhance engagement through an informal therapeutic conversational style tailored to users' unique barriers to quitting smoking [27]. Conversational chatbots only require a text response to operate, making them ideal for all individuals who smoke, including those with low technology literacy [26]. Overall, conversational chatbots offer a cost-effective communication platform, accessible at any time, and have the potential for high population-level reach, making them a valuable tool in smoking cessation interventions.

To date, research on conversational chatbots for smoking cessation is scarce. Existing literature revealed a limited number of empirical studies, often exhibiting low methodological quality [28]. There is a notable paucity of randomized controlled trials (RCTs) focusing on conversational chatbots for smoking cessation, and while promising results have emerged, they have been limited by low quit rates [29]. Several conversational chatbots for smoking cessation in the public domain include Florence (World Health Organization) [30], Bella (Solutions4Health) [31], and Alex AI (Alex Therapeutics) [32]. However, we are not aware of publications on their efficacy, with only the Florence app having reported user's receptivity results [33]. Critical to creating useful and engaging conversational chatbots is following a user-centered design development process [34]. Similar to most chatbots, the development of the chatbots listed above lacks context for how they were designed and any user-centered design that involved conducting a needs assessment or including user feedback during the development process [28,35]. The few studies that have provided development details only describe early design phases, such as coding 30 quit coaching calls for prototype development, without empirical efficacy data [28,35,36]. In sum, the literature on chatbots for smoking cessation offers only partial accounts on how they were developed or report on initial stages of development.

Objective

To address these gaps, this paper describes the comprehensive 4-year, 11-step user-centered design development process for a novel quit smoking conversational chatbot named "QuitBot." This single report aims to summarize the entirety of the QuitBot development process.

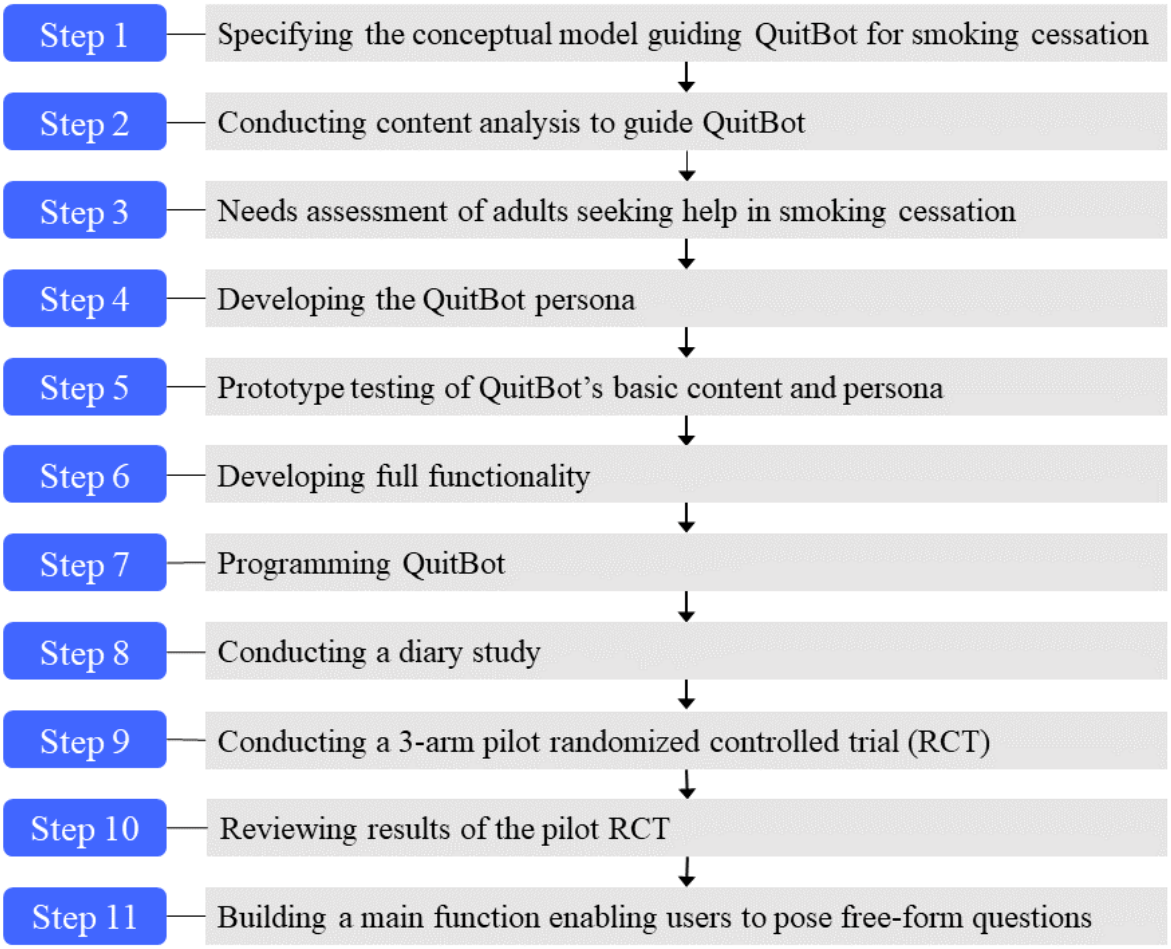
Methods

Overview of the Formative Research Process

The 4 years of formative research for developing QuitBot followed an 11-step process, consistent with a user-centered design framework (Figure 1) [37,38].

The steps were as follows: (1) specifying a conceptual model to guide the QuitBot intervention targets; (2) conducting content analysis of existing smoking cessation interventions to guide the QuitBot coaching conversations; (3) conducting a needs assessment to determine what an adult seeking help in quitting smoking would need from a cessation chatbot; (4) developing the QuitBot persona, or personality of the chatbot, to shape the user's experience of and bond with the QuitBot chatbot; (5) prototyping QuitBot's basic content and persona; (6) developing the full functionality of the QuitBot; (7) programming the QuitBot; (8) conducting a diary study for user feedback on their interactions with QuitBot and its design and content; (9) conducting a pilot RCT to test QuitBot for smoking cessation; (10) reviewing results of the pilot RCT; and (11) adding a free-form question and answer (QnA) function, based on user feedback from pilot RCT results. The process of adding the QnA function itself involved a three-step process: (1) generating QnA pairs, (2) fine-tuning LLMs on the QnA pairs, and (3) evaluating the LLM model outputs.

Figure 1. Overview of QuitBot's formative research process.



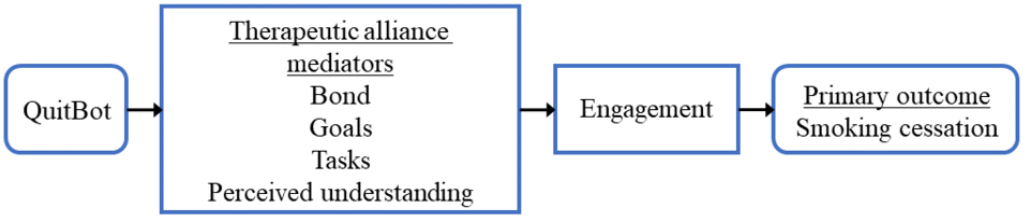
Step 1: Specifying the Conceptual Model Guiding QuitBot for Smoking Cessation

The conceptual model guiding the development of QuitBot for smoking cessation (Figure 2) focuses on impacting user engagement through 4 therapeutic alliance processes. The four processes are as follows: (1) *bond* with QuitBot, (2) agreement on smoking cessation *goal*, (3) agreement on *tasks* for achieving smoking cessation goal, and (4) *perception that QuitBot understands* user's current needs [18].

These working alliance processes have predicted smoking cessation [39] and quit attempts [40] and have mediated the impact of human therapist–delivered smoking cessation interventions [40]. QuitBot uses various strategies to establish a therapeutic alliance, including expressing empathy for the user's struggles [41,42], engaging in social dialogue [43,44], using metarelational communication (ie, discuss the relationship) [45], and expressing happiness while interacting with the user [42]. Language constructs such as inclusive pronouns [41], politeness strategies [46], and the use of greetings and farewells rituals [47] contribute to the creation of this alliance as well. Compared to a technology that did not use these verbal behaviors, a conversational chatbot for physical activity increased these therapeutic alliance processes, which in turn was predictive of higher engagement with the chatbot [48].

Agreement on smoking cessation goal starts by collaboratively setting a quit date (eg, "Have you thought about a specific day you would like to quit? Generally, I recommend about 14 days away."). QuitBot enhances perceived understanding by promptly addressing the user's immediate needs (eg, "You say you are tempted by friends who smoke. Here's a tip that might help..."). In addition, self-disclosure [49] is used to foster perceived understanding, generating various positive outcomes, especially when the listener responds with support and validation [50]. A chatbot that used self-disclosure increased the user's perception that the chatbot understood their needs, which in turn predicted more positive mood [51].

Figure 2. Conceptual model of QuitBot for smoking cessation.



Step 2: Conducting Content Analysis to Guide QuitBot

The content analysis aimed to establish a natural flow of coaching conversations for QuitBot, aligned with US Clinical Practice Guidelines for smoking cessation [52]. In the initial phase of the content analysis, we interviewed a panel of experts, including 3 smoking cessation counselors, a smoking cessation master trainer, and a tobacco cessation scientist from our team. This panel consisted of 4 women and 1 man, with 20% (1/5) from racial and ethnic minority backgrounds. Among them, 40% (2/5) held a PhD in clinical psychology, while 60% (3/5) had master's degrees in counseling or social work. Collectively, they had a wealth of experience ranging from 3 to 20 years, with an average of 8 (SD 4.6) years, in developing and delivering smoking cessation interventions. Deductive coding of these interviews and expert consensus iteratively lead to the formulation of common themes, domain-specific responses, and anticipated user interactions that QuitBot should address. We identified common conversation topics about smoking cessation, including triggers to smoke (ie, physical, emotional, and situational triggers), motivations to quit, and barriers to quitting. Interviews also highlighted the importance of QuitBot's persona to be sensitive and empathetic to the user and to express that their concerns are being heard.

Guided by this expert consensus on conversation topics, the second phase was to extract the content and flow of smoking interventions as they naturally occur in actual interactions between cessation counselors and patients. To achieve this, we conducted semantic analysis of verbatim manually transcribed intervention conversation transcripts obtained from our telephone counseling intervention trial, randomly selected among those who did and did not quit smoking (R01 DA038411) [53]. A total of 159 call transcripts (equating to 63 h and 23 min) from 117 unique participants were randomly selected, constituting a 7.8% (159/2038) sample from each of the 5 sessions (with an average session duration of 22.9 min) of an efficacious behavioral intervention for smoking cessation with a 25% thirty-day point prevalence abstinence (PPA) rate at the 12-month follow-up [53]. These sessions covered various topics, including motivations to quit, triggers to smoke, barriers to quitting, setting a quit date, developing a quit plan, education and proper use of Food and Drug Administration (FDA)–approved medications for quitting smoking, coping skills for dealing with urges, enlisting social support, and strategies for avoiding external cues to smoke. Participants had a mean age of 47.4 (SD 12.7) years, with 43.6% (51/117) male participants and 21.4% (25/117) from racial and ethnic minority backgrounds.

Transcripts underwent deductive coding using a predefined codebook to identify common conversation topics related to smoking cessation, such as triggers to smoke (ie, physical, emotional, and situational triggers), motivations to quit, and barriers to quitting. Using a supervised machine learning approach, these topics formed the basis of QuitBot's entity extraction, wherein elements of the unstructured transcript text were coded into predefined categories. Subsequently, we determined intent classifications, which involved discerning the meaning of the user's text. Finally, we mapped out the natural conversational flow of both the chatbot and the range of verbal responses and comments that users might provide in response to the chatbot. The entity extraction, intent classifications, and conversational mapping were conducted using the LUIS conversational artificial intelligence (AI) program [54].

Step 3: Needs Assessment of Adults Seeking Help in Smoking Cessation

Assessing the needs of adults seeking help in smoking cessation interventions shapes what the user should be able to do with a chatbot. To assess user needs, we first analyzed the results of the content analysis phase. Subsequently, we conducted interviews with 5 adults who had participated in our human-delivered smoking cessation interventions within the past year (including 2 who quit and 3 who did not quit) [55]. Participants had a mean age of 46.1 (SD 10.4) years, with 40% (2/5) female participants and 40% (2/5) from racial and ethnic minority backgrounds. The interviews queried participants about their personal background and smoking history, expectations for a smoking cessation program, experiences with a human cessation coach, perceptions regarding setting, keeping and changing quit dates, coping skills for urges to smoke, and attitudes toward and expectations of what a chatbot could do for helping them quit smoking. Semistructured interviews were conducted in person at the lead author's user experience (UX) Research HABIT laboratory. The deductive thematic analysis organized the user's responses by grouping them into themes,

reviewing the themes, and then interpreting them [56-58]. The themes of the key user needs identified were (1) a coach who can make a personal connection, (2) on-demand help with urges, and (3) skills for preparation to quit and preventing relapse.

Step 4: Developing the QuitBot "Persona"

The user's bond with the chatbot is impacted by its persona [48]. On the basis of interviews with smoking cessation coaches and our master trainer, we created the persona to foster a strong bond with users. Presented to the user as a computer program (eg, "I'm a bot designed to help you live smoke free"), elements of the QuitBot persona included expressions of empathy [41,42], social dialogue [43,44], metarelational communication (ie, discuss the relationship) [45], and expressing happiness to see the user [42]. In addition, specific language constructs, including inclusive pronouns [41], politeness strategies [46], and greetings and farewells rituals, were integrated to enhance the UX and promote a respectful dialogue [47]. Finally, we established 11 core values for the persona, serve as guiding principles for QuitBot's behavior throughout conversations.

Step 5: Prototype Testing of QuitBot's Basic Content and Persona

The prototyping testing of QuitBot's basic content and persona aimed to assess users' initial responses to basic smoking cessation conversations between the user and the persona. Stimuli were built using botmock [59,60] to develop the chat dialogue, which was then integrated into Facebook Messenger (FM; Meta Platforms, Inc) using Chatfuel [61]. Participants had a guided initial chat conversation introducing the chatbot and program goals, querying about triggers for smoking, and setting a quit date. Subsequently, they interacted with QuitBot for a second conversation, focusing on tracking triggers to smoke. For both conversations, a UX researcher frequently paused to prompt participants to think-aloud their experiences with QuitBot. Real-time interactions between the user and QuitBot were facilitated by a UX researcher using the Chatfuel program [61]. To evaluate this process, 75-minute individual interviews were conducted with 8 adults interested in quitting smoking. Four were chosen because they thought a chatbot could be helpful for quitting smoking, while the remaining 4 were selected because they were unsure or skeptical that a chatbot would be helpful for quitting smoking. The mean age of the participants was 42 (SD 11.1) years, with 38% (3/8) male participants, 50% (4/8) female participants, and 12% (1/8) transgender participants. In addition, 38% (3/8) of the participants had high school education or less and 25% (2/8) reported being from racial or ethnic minority backgrounds.

Semistructured Interviews

Semistructured interviews were conducted in person at the lead author's UX Research HABIT laboratory. A deductive thematic analysis method was used to organize user responses into themes, review those themes, and then interpret them [56-58]. Despite initial skepticism from half of participants (4/8, 50%) regarding the usefulness of interacting with a digital coach, the results showed a notable shift in the interest in QuitBot by the end of the interview: 100% (8/8) reported that a chatbot such as QuitBot would be valuable for helping someone quit smoking, with 88% (7/8) expressing willingness to try this chatbot for quitting. In addition, all participants (8/8, 100%) found QuitBot easy to use, noting its conversational tone as "encouraging," "polite," and "reassuring." They deemed the length and speed of onboarding conversations appropriate and felt comfortable providing conversational responses. Participants expressed surprise at the "humanness" of QuitBot's avatar, noting its informal, reassuring, accessible, and easy-to-talk-to demeanor.

When discussing whether the avatar should have a gender or a name, there was consensus among participants in favor of a female persona, with the name "Ellen" deemed appropriate (interestingly, one of the initial participants spontaneously suggested "How about a woman named 'Ellen'?"). Later participants concurred with this choice when asked by the UX researcher.

Participants also expressed a desire for more actionable suggestions and to open and close each conversation with a specific plan of action. In response, we added a plan outlining what to anticipate, letting them know that the avatar would initiate a chat the following day and introduce a new quitting smoking skill in the subsequent conversation. Some participants wanted additional time to decide on a quit date, prompting us to include a dialogue indicating that they postponed setting a quit date until they felt ready. In addition, participants suggested visualizing their progress in quitting smoking, such as through a graph. In response, we added a progress chart displaying the number of cigarettes smoked over time. Overall, participants described feeling "captivated" by the content and expressed eagerness to learn more.

Step 6: Developing Full Functionality

Building upon the prototype as the foundation, we applied insights from the earlier steps to develop a full program consisting of 42 days of 2- to 3-minute focused conversations. These conversations were distributed over several phases of treatment: a prequit phase (14 days of content), quit day (1 day of content), and postquit phase (27 days of content). There are also conversations for those not ready to quit smoking by day 14 (6 days of content) and conversations for those who have relapsed (3 days of content). The content, described in Textboxes 1 and 2, follows US Clinical Practice Guidelines for cessation interventions [62]. The program content was presented as a continuous conversation, built on user input from prior conversations. This ensures a personalized and adaptive approach based on the user's stated motivations to quit, triggers to smoke, and number of cigarettes smoked. QuitBot was proactive and provided daily prompts to start a structured text conversation with Ellen at the user's preferred time, such as "Hi Alex, are you free to chat?" Users also had the flexibility to reach out to Ellen at any time for on-demand help with urges, inspiration, mood, and slips.

Textbox 1. QuitBot and SmokefreeTXT (SFT): phases and corresponding content.

Phase (number of days) and content of both SFT and QuitBot

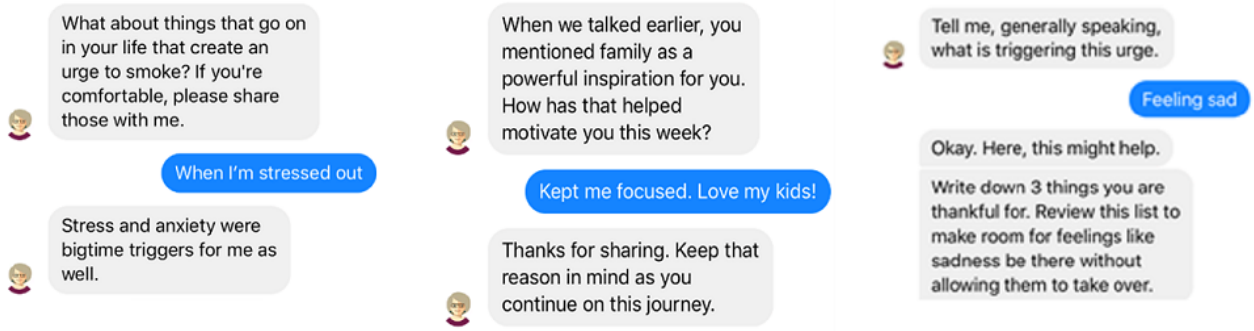
- Prequit day (14 days)
  - Triggers to smoke, motivations for quitting, setting a quit date, Food and Drug Administration–approved medication information, skills to be aware of and cope with urges, and cessation progress tracking
- Quit day (1 day)
  - Encouragement and smoking status check-in
- Postquit (28 days)
  - Withdrawal symptoms education, slips and relapse prevention, managing mood, managing cravings, and cessation progress tracking
- Not ready or quit date >14 days
  - Reviews motivations for quitting and cessation progress tracking
- Anytime help
  - Skills to cope with urges, mood, and slips

Textbox 2. Content communication.

How content is communicated: SmokefreeTXT (SFT) and QuitBot

- SFT: sends texts of the content, answers to daily check-ins (eg, number of cigarettes smoked today), get 1-2 text responses, answers to entering anytime help keywords (eg, “CRAVE”), and get 1 text response
- QuitBot: digital coach sends the user a greeting to start a 2- to 3-minute conversation, presents content in a dialogue with the user via engagement features described in Figure 3 (eg, tailored responses and empathy), and answers to entering anytime help keywords (eg, “CRAVE”) initiate a dialogue

Figure 3. Representative functionalities of QuitBot include (A) determining triggers, (B) maintaining motivation, and (C) providing anytime urge help.



Step 7: Programming QuitBot

We initially sought a development architecture with the flexibility to interact with QuitBot on any major consumer communication channel (eg, as a stand-alone app, FM, and Slack). Such flexibility adapts to current consumer trends in communication technology use, making QuitBot available for use on the channels with current high population-level reach. To determine which communication channel would be used for interacting with QuitBot, we conducted a web-based survey of 100 US adults who smoke, asking them which of these channels they would prefer for a chatbot: stand-alone app, WhatsApp (Meta Platforms, Inc), FM (Meta Platforms, Inc), Skype (Skype, Inc), or Slack (Slack Technologies, LLC). The majority of respondents (74/100, 74%) preferred FM, citing its familiarity, ubiquity, and ease of use. FM is an instant messaging service for online chats. At the time of the study (ie, 2019), there were >133 million FM users in the United States (1.3 billion globally) and FM hosts >300,000 chatbots, with 27% of them for health care (eg, exercise) [63-65]. Following these findings, we hosted QuitBot on FM.

Therefore, we custom built an architecture using the Microsoft Bot Framework that uses Microsoft Azure for the cloud computing and Microsoft Language Understanding (LUIS) platform for the natural language understanding of the QuitBot guide *Ellen*. The preference for natural language understanding over an if-then decision-based conversation flow was made to ensure a more natural and open-ended interaction, allowing a broad range of responses and better conveying that the user is being heard and understood. QuitBot's LUIS allows it to understand common text shorthand. Users can respond freely

or simply select from a menu of responses. If QuitBot does not understand a free response, it will say so and ask the participant to rephrase the response. QuitBot was written in the programming language of Node.js [66].

Step 8: Conducting a Diary Study

We conducted a diary study to obtain ongoing feedback on users' interactions with QuitBot, its design, and content. In user-centered design research, a diary study of 2 weeks with 6 to 12 participants is recommended to obtain this initial feedback [67,68]. Accordingly, we conducted a single-arm 14-day diary study of the program with 9 adults who were smoking at least daily (all smoked ≥30 cigarettes/d), were interested in quitting smoking, and recruited from around the United States via Facebook advertisements. Four were chosen because they were skeptical about chatbots being able to help someone quit smoking, while the remaining 5 were neutral about them. Participant demographics were as follows: mean age 40.4 (SD 13.4) years, 11% (1/9) from racial and ethnic minority backgrounds, 44% (4/9) female, and 67% (6/9) had less than a bachelor's degree.

All 9 participants completed the following: (1) the 60-minute video-based orientation focusing on how to use QuitBot and complete the daily diary entries; (2) 14 evening diary entries (15 min each) about their daily interactions with QuitBot, its design, and content; (3) on day 7, a midpoint 15-minute video call with a member of our user research team to review their impressions to date; and (4) a 60-minute video call exit interview with a member of our user research team. A PhD-level UX researcher with >20 years of experience conducted the interviews. (Example questions from the exit interview are as follows: "Which parts of the app did you find the most helpful? Why?") Semistructured interviews were conducted in person at the lead author's UX Research HABIT laboratory. The deductive thematic analysis organized the user's responses by grouping them into themes, reviewing the themes, and then interpreting them [56-58].

The results showed that, although the focus was on usability, by day 14, three participants quit smoking and the remaining 6 participants reduced to 3 or 4 cigarettes per day. Ratings for usefulness ("Overall, how useful was the QuitBot app for helping you quit smoking?"), satisfaction ("Overall, how satisfied were you with the QuitBot?"), and likelihood of recommending QuitBot ("To what extent would you recommend QuitBot to someone who would like to quit smoking?") were all high: 4.33, 4.67, and 4.88, respectively, on a 0 (not all) to 5 (extremely) scale. All 9 users felt highly supported by Ellen and liked her persona. They liked the skills training for coping with smoking urges and lapses. Their feedback yielded minor content edits and fixes of technical bugs. Representative functionalities of QuitBot are shown in Figure 3.

Diversity and Inclusion in UX Design

The diversity of race, gender, age, and educational characteristics of users who participated in our UX design studies influenced the design in many ways, including Ellen's persona design (eg, men and women both preferred a female persona), Ellen's stories of people who have quit smoking (eg, they were diverse in age, gender, race, and education), use of language (eg, fifth-grade reading level, informal, and respectful), and user interface (eg, larger response buttons and larger font size [69]).

Step 9: Conducting a 3-Arm Pilot RCT

The favorable feedback from the diary study led us to conduct a 3-arm parallel pilot RCT comparing QuitBot (n=200) to the SmokefreeTXT (SFT; n=149) intervention and to a QuitBot delayed access control group (n=55). Following expert recommendations for pilot RCT design [70,71], the feasibility outcomes were the study's primary focus to inform the further development of QuitBot and design of a future full-scale trial of QuitBot. As this pilot RCT was the first time QuitBot was tested and no prior RCTs had been reported on any quit smoking chatbot, estimated effect sizes were unknown. Instead, the sample sizes were based on comparable sample sizes from prior pilot studies we had conducted in our laboratory [72,73]. Participants were recruited nationwide and were randomized to the intervention arm using randomly permuted blocks of size 2, 4, and 6, stratified by biological sex (male vs female), heaviness of smoking index score (≤4 vs >4), and percent confidence in being smoke-free in 12 months (≤70% vs >70%). The study was double-blinded, with both interventions called "QuitBot."

Ethical Considerations

All study procedures were approved by the Fred Hutch Cancer Center Institutional Review Board (8659/RG1001766). The clinical trial protocol was approved by the Fred Hutch Scientific Review Committee (FHIRB008659), and the trial was registered on ClinicalTrials.gov (NCT03585231). There were no deviations to the registered protocol. All study participants provided informed consent, and data were deidentified for privacy and confidentiality.

Eligibility Criteria for the Pilot RCT

The inclusion criteria were as follows: (1) age ≥18 years; (2) having smoked at least 1 cigarette a day for at least the past 12 months; (3) wanting to quit cigarette smoking within the next 14 days; (4) if concurrently using any other nicotine or tobacco products, wanting to quit using them within the next 14 days; (5) being interested in learning skills to quit smoking; (6) being willing to be randomly assigned to either condition; (7) residing in the United States; (8) having daily access to their own smartphone; (9) having both SMS text messaging and FM on their smartphone (criteria 8 and 9 were required to receive each interventions' content); (10) being willing and able to read in English; and (11) not using other smoking cessation interventions. Individuals deemed ineligible to participate were directed to the smokefree.gov website and the 800-QUIT-NOW number for access to their state's quitline resources.

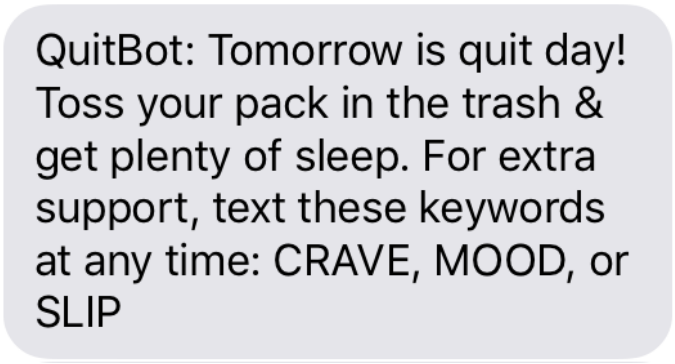
SFT Comparison Condition

For the past 20 years, mobile phone–delivered SMS text messaging interventions have been a prominent technology for delivering smoking cessation interventions [74-78]. Each year, SMS text messaging smoking cessation interventions are reaching >300,000 US adults who smoke and 6 million adults who smoke worldwide [79,80]. SFT's 42-day program was

developed by the National Cancer Institute (NCI). SFT is the most widely accessible SMS text messaging program in the United States. SFT is nonproprietary and free to the public, thereby providing maximal accessibility and replicability. Daily messages are sent about the importance of quitting smoking, setting a quit date, preparing to quit, quitting, and maintaining abstinence. Daily messages check in about quit status. Three keywords can be proactively sent by users to receive help anytime: "CRAVE" (on how to cope with urges), "MOOD" (on how to cope with moods triggering smoking), and "SLIP" (on how to cope with lapses). Participants do not need to respond to or otherwise engage with SFT messages to complete the SFT program. Refer to Figure 4 for sample messages.

NCI's SFT contractor (ICF International [81]) provided us with the full content of SFT so that we could internally host a secured private version for research. In both SFT and QuitBot, participants receive 2 prompts per day (3 on the quit day). Comparisons between QuitBot and SFT are shown in Textboxes 1 and 2.

Figure 4. Sample SmokefreeTXT text message.



QuitBot Delayed Access Comparison Condition

To explore the unique impact of QuitBot on smoking cessation, considering that some participants might quit smoking without intervention, we introduced a delayed access comparison condition. In this condition, 55 participants received delayed access to QuitBot after completing the 3-month follow-up survey. The delayed access comparison condition served the ethical purpose of providing participants access to a treatment (as opposed to no treatment at all).

Outcome Measures

Outcome data were collected through an encrypted web-based survey. Participants not completing the web-based survey were sequentially offered the survey via phone, mailed survey, and postcard. The primary feasibility outcomes were (1) sufficient accrual of the planned number of study participants, (2) balanced demographic and smoking characteristics at baseline between study arms, and (3) retention of the primary 30-day PPA smoking outcome at the 3-month follow-up. Intervention engagement was assessed based on comparing the active treatment study arms on the number of times and number of days participants interacted with their assigned intervention. All interactions with the participants' assigned interventions were objectively logged using an internally hosted secure server. The primary smoking cessation outcome was 30-day PPA, based on compete-case analysis, and 7-day complete-case PPA was secondary.

Statistical Analysis for the Pilot RCT

The feasibility of the pilot RCT was assessed based on sufficient accrual, balanced randomization, and adequate follow-up data retention rates that did not differ between arms. Baseline characteristics were compared between the 3 study arms using ANOVA for continuous variables and Fisher exact tests for categorical variables and were summarized with the "arsenal" package in R (version 4.2.3; R Foundation for Statistical Computing) [82,83]. We used generalized linear models to assess differences between study arms in the number of days participants used their intervention.

We used negative binomial models, implemented with the R package "MASS" [84], to compare treatment arms on total number interactions because the data were heavily right-skewed. Logistic regression models were used to test the effect of the treatment arm on binary smoking cessation outcomes. On the basis of evidence from meta-analyses of SMS text messaging trials [85], all outcome models were adjusted for the 3 factors used in stratified randomization: biological sex (male vs female), heaviness of smoking index score (≤4 vs >4), and percent confidence in being smoke-free in 12 months (≤70% vs >70%). Wald tests for pairwise comparisons of each outcome between study arms were adjusted for multiple comparisons with the Holm procedure [86]. Statistical tests were considered significant at $\alpha<.05$. Deductive thematic analysis organized participants' comments about QuitBot by grouping them into themes, reviewing the themes, and then interpreting them [56-58].

Results

Step 10: Main Results of the Pilot RCT

Recruitment Was Successful

On the basis of our successful methods for national recruitment [87], we developed and tailored Facebook advertisements with ongoing monitoring and adjustment of recruitment yield. These efforts resulted in screening 2954 participants, with 1380 eligible, 583 consenting, and 418 randomized between September 2018 and June 2019. After the completion of study participation, 14 participants were found to be cases of fraud, duplicate participants, or in the same household as another

participant, leading to a total of 404 participants included in analyses.

Randomization

The 3 stratification conditions were balanced at baseline on all measured characteristics (all *P* values >.05). As shown in Table 1, participants were on average 36 years old, 70% (283/404) were women, 28.9% (116/401) reported being from racial or ethnic minority backgrounds, 52.7% (213/404) were unemployed, 83.9% (339/404) had no college degree, 71.5% (289/404) smoked more than one-half pack daily, and 59.9% (242/404) had high cigarette dependence (Fagerström Test for Cigarette Dependence scores of ≥6). The characteristics of this FM sample are very similar to those of other digital health intervention trials [85,88,90].

Table 1. Baseline participant characteristics by study arm.

Characteristic	Total (n=404)	SmokefreeTXT (n=149)	Delayed (n=55)	QuitBot (n=200)	*P* value
Baseline characteristic					
Age (y), mean (SD)	36.0 (10.4)	36.2 (11.2)	35.6 (9.6)	35.9 (9.9)	.92
Gender, n (%)					
Woman	283 (70)	103 (69.1)	39 (70.9)	141 (70.5)	.95
Man	121 (30)	46 (30.1)	16 (29.1)	59 (29.5)	
Race, n (%)					.52
Asian	2 (0.5)	0 (0)	0 (0.0)	2 (1)	
Black or African American	51 (12.6)	21 (14.1)	9 (16.4)	21 (10.5)	
Native American or Alaska Native	12 (3)	4 (2.7)	0 (0)	8 (4)	
Native Hawaiian or Pacific Islander	1 (0.2)	0 (0)	0 (0)	1 (0.5)	
White	296 (73.3)	110 (73.8)	40 (72.7)	146 (73)	
Multiple races	31 (7.7)	13 (8.7)	4 (7.3)	14 (7)	
Unknown race	11 (2.7)	1 (0.7)	2 (3.6)	8 (4)	
Hispanic ethnicity, n (%)	28 (6.9)	7 (4.7)	6 (10.9)	15 (7.5)	.27
Minority race or ethnicity (n=401), n (%)	116 (28.9)	42 (28.2)	17 (30.9)	57 (28.9)	.93
Married, n (%)	104 (25.7)	32 (21.5)	16 (29.1)	56 (28)	.32
Employed, n (%)	191 (47.3)	80 (53.7)	24 (43.6)	87 (43.5)	.14
No college degree, n (%)	339 (83.9)	126 (84.6)	47 (85.5)	166 (83)	.87
Heavy alcohol use (n=395), n (%)	47 (11.9)	18 (12.5)	6 (11.3)	23 (11.6)	.96
Positive depression screening results (n=402), n (%)	223 (55.5)	91 (61.5)	28 (50.9)	104 (52.3)	.17
Smoking behavior					
FTCD[a] score, mean (SD)	5.7 (2.0)	5.5 (2.0)	6.1 (2.2)	5.7 (2.0)	.17
High nicotine dependence, n (%)	242 (59.9)	88 (59.1)	36 (65.5)	118 (59)	.66
Smokes more than one-half pack per day, n (%)	289 (71.5)	98 (65.8)	42 (76.4)	149 (74.5)	.14
Smokes >1 pack per day, n (%)	66 (16.3)	20 (13.4)	14 (25.5)	32 (16)	.11
First cigarette within 5 minutes of waking, n (%)	205 (50.7)	75 (50.3)	34 (61.8)	96 (48)	.19
Smoked for ≥10 years, n (%)	317 (78.5)	112 (75.2)	44 (80)	161 (80.5)	.46
Used e-cigarettes at least once in the past month, n (%)	122 (30.2)	42 (28.2)	16 (29.1)	64 (32)	.73
Quit attempts in the past 12 months (n=377), mean (SD)	1.6 (4.7)	1.6 (3.3)	1.1 (3.2)	1.7 (5.8)	.68
At least 1 quit attempt in the past 12 months (n=377), n (%)	145 (38.5)	51 (37.8)	16 (30.2)	78 (41.3)	.33
Confidence to quit smoking, mean (SD)	64.1 (27.0)	62.6 (27.0)	72.2 (27.3)	62.9 (26.8)	.05
Friend and partner smoking					
Close friends who smoke, mean (SD)	2.8 (1.7)	2.8 (1.7)	2.7 (1.6)	2.8 (1.8)	.97
Number of adults in home who smoke, mean (SD)	1.5 (0.9)	1.4 (0.9)	1.7 (1.1)	1.5 (0.8)	.19
Living with partner who smokes, n (%)	145 (35.9)	51 (34.2)	24 (43.6)	70 (35)	.43

[a]FTCD: Fagerström Test for Cigarette Dependence.

The 3-Month Follow-Up Rates Were High

To maximize outcome data completion, we followed our team's successful protocol [87]: 4 sequential survey modalities (first web, followed by phone, mail, and postcard). As agreed in the informed consent, participants received US $25 for submitting their responses and received an additional US $10 bonus for completing the web survey within 24 hours. The achieved *outcome survey completion rate of 96%* provided confidence in the follow-up survey methods. The data retention did not

differ between study arms (*P*=.54). Given the limitations of the pilot budget, cessation data were self-reported.

Engagement and Cessation Results Were Promising for QuitBot

The number of times participants interacted with their assigned intervention was 1.3 times greater in QuitBot as compared to SFT (incidence rate ratio 1.33, 95% CI 1.04-1.70; *P*=.02; Table 2). Participants used their assigned intervention 11 days longer in the QuitBot arm than in the SFT arm (point estimate 11.5, 95% CI 4.9-18.1; *P*=.001). QuitBot's intervention completion results are substantial when considering that each day's content involved a 2- to 3-minute conversation. (By contrast, SFT participants did not need to respond to or otherwise engage at all with their messages to complete their program; daily SFT text messages were sent automatically.) Participant engagement was limited by QuitBot's inability to answer participants' open-ended questions (see the *Representative QuitBot Comments* section). Therefore, cessation results are reported for all participants and for participants who completed their assigned intervention.

For all participants, the 30-day PPA rates at 3-month follow-up were 31.1% (59/190) for QuitBot versus 34.7% (50/144) for SFT (QuitBot vs SFT: odds ratio [OR] 0.81, 95% CI 0.50-1.29; *P*=.36; Table 3) versus 7% (4/54) for delayed treatment (QuitBot vs delayed: OR 5.97, 95% CI 2.04-17.45; *P*=.002). For those who completed their assigned intervention (ie, viewed all 42 days of planned content), the 30-day, complete-case, PPA rates at 3-month follow-up were 63% (39/62) for QuitBot versus 38.5% (45/117) for SFT (QuitBot vs SFT: OR 2.58, 95% CI 1.34-4.99; *P*=.005). The pattern of results was highly similar for the outcome of 7-day, complete-case, PPA rates at 3-month follow-up, albeit with higher abstinence rates in each study arm.

Table 2. Comparison of QuitBot and SmokefreeTXT (SFT) interventions on 3-month engagement outcomes.

Study engagement outcome	SFT (n=149), mean (SD; median)	QuitBot (n=200), mean (SD; median)	QuitBot vs SFT			
			IRR[a] (95% CI)	*P* value	PE[b] (95% CI)	*P* value
Number of times interacted (n=266)	24.2 (25.8; 15)	32.9 (29.0; 25)	1.33 (1.04-1.70)	.02	—[c]	—
Days from randomization to last input	44.1 (22.7; 54)	55.7 (36.0; 70)	—	—	11.5 (4.9-18.1)	<.001

[a]IRR: incidence rate ratio.

[b]PE: point estimate.

[c]Not applicable.

Table 3. Comparison of QuitBot and SmokefreeTXT (SFT) interventions and delayed intervention on 3-month cessation outcomes.

Study outcome	SFT (n=149), n (%)	Delayed[a] (n=55), n (%)	QuitBot (n=200), n (%)	QuitBot vs SFT		QuitBot vs delayed[a]	
				OR[b] (95% CI)	*P* value	OR (95% CI)	*P* value
30-day cigarette abstinence among all participants (n=388)	50 (35)	4 (7)	59 (31)	0.81 (0.50-1.29)	.36	5.97 (2.04-17.45)	.002
30-day cigarette abstinence among program completers (n=179)	45 (38)	—[c]	39 (63)	2.58 (1.34-4.99)	.005	—	—
7-day cigarette abstinence among all participants (n=388)	76 (53)	5 (9)	91 (48)	0.79 (0.51-1.22)	.28	10.08 (3.79-26.80)	<.001
7-day cigarette abstinence among program completers (n=179)	70 (60)	—	50 (81)	2.63 (1.24-5.55)	.01	—	—

[a]Three-month delay in receiving QuitBot.

[b]OR: odds ratio.

[c]Not applicable.

Representative QuitBot Comments

Comments from QuitBot arm trial participants reflected a strong overall bond with the chatbot's persona:

> *I loved Ellen. She was always there when I needed her.*
>
> *Ellen was always there for me when I had a craving.*
>
> *I love how engaged she was, I could really quit with her there to talk to.*
>
> *She made me feel like I was not alone.*
>
> *She was there without making me feel ashamed.*
>
> *She was kind, nonjudgmental.*
>
> *She held me accountable.*
>
> *Felt like a friend encouraging me.*

Conversely, participants were frustrated by QuitBot's inability to respond to their specific questions about quitting smoking:

> *I could not ask questions and get real answers back.*

I could not ask it real live questions.

I wanted to write my own questions.

Can't ask any question.

Not being able to respond to my questions.

I wish you could talk to her...without it being a constant couple of options.

I didn't like how it selected responses.

The fact that you cannot ask a question and [it] has no idea what you are saying unless you select one of the options.

Main Conclusions From the Pilot RCT

Our main conclusions were as follows: (1) the intervention demonstrated potential for rigorous testing based on sufficient accrual, balanced randomization, and high retention rates; (2) overall, there was a strong engagement with QuitBot; and (3) promising quit rates were observed, particularly among participants who completed the content of their assigned intervention. The effectiveness of QuitBot was evident, as quit rates in the delayed condition group were significantly lower (59/190, 31.1% vs 4/54, 7%; *P*=.002), indicating a net percentage point increase in smoking cessation of 24%. Therefore, it is highly unlikely that effects of QuitBot were merely due to the passage of time or baseline motivation to quit smoking (ie, few participants quit without offering intervention).

Challenges were also identified, potentially impacting participant engagement and quit rates. Specifically, QuitBot's inability to respond to participants' own questions about quitting smoking led to a significant level of frustration. While the participant can answer questions asked by the QuitBot (eg, "Tell me what is triggering your urge."), the reverse was not possible: participants could not ask QuitBot their own questions. A QuitBot feature that allows participants to ask free-form questions would be needed to address this limitation.

Technical Limitations of the FM Platform

While FM was the preferred communication platform from our survey results, Facebook introduced changes that would limit participants' engagement with QuitBot as well as our own access to user data: (1) Facebook made policy changes that revoked access permissions to proactively outreach (eg, to invite participant to check in or start a conversation), effectively removing our ability to proactively contact users (restricting that ability to news-related apps only); and (2) Facebook made platform changes that restricted our ability to access demographic information of users, inhibiting data collection. Facebook's changes raised concerns about the feasibility of QuitBot's conversational functionality and data collection. This critical limitation could be addressed by transitioning to a stand-alone smartphone app communication platform, enhancing accessibility and control for both participants and the development and research teams.

Step 11: Building a Main Function Enabling Users to Pose Free-Form Questions About Smoking

Overview

The goal of this specific QuitBot refinement was to build a main function of QuitBot that would enable users to pose free-form questions about quitting cigarette smoking and for the QuitBot to respond with accurate, concise, professional, and nonrepetitive answers. This was an iterative 3-step process, which is detailed in Table 4.

Table 4. Steps, sources, and results of QuitBot's question and answer (QnA) iterative development process.

Step	Source (year)	Results
1. Generate QnA pairs	• Alexander Street therapy transcripts (2020) • National Cancer Institute call center transcripts (2020) • HABIT laboratory cessation counseling intervention transcripts (2020) • HABIT laboratory digital intervention content (2020) • HABIT laboratory clinical team generates QnA (2020-2021) • Prolific survey of adults who Smoke (2021)	• 11,000 smoking QnA pairs • 8223 chitchat QnA pairs
2. Training LLM[a] models on QnA pairs	• Azure application programming interface (2020-2023) • DialoGPT (2021) • ParlAI (2021) • Davinci GPT-3 (2021) • Curie GPT-3 (2021) • Ada GPT-3 (2021) • Contextualized GPT-3.5 (2022) • GPT-4.0 (2023)	• Models with higher self-scored confidence about answers provided: processed by Azure application programming interface • Models with lower self-scored confidence about answers provided: handled by Curie GPT-3 (2021) and replaced by contextualized GPT-3.5 (2022)
3. Evaluating LLM outputs	• Automated evaluation: pertinence and grammaticalness (2020-2023) • Manual evaluation: accuracy and tone (2021-2023) • 14-day user resting (2021)	• Identified the answers that were repetitive, incorrect, or had impersonal or nonprofessional tone

[a]LLM: large language model.

Step 11.1. Generating QnA Pairs

The first step was to develop a knowledge base of QnA pairs focused on the topics of quitting cigarette smoking. Smoking cessation clinical data sources included the Alexander Street data sets of counseling transcripts [90], NCI call center transcripts of smoking cessation coaching conversations [91], and transcripts of counseling sessions from our Fred Hutch Cancer Center research laboratory's prior smoking cessation intervention trials [53]. The categories of smoking cessation questions spanned a wide range, including medications to aid smoking cessation, the role of vaping e-cigarettes in quitting smoking, health consequences of smoking on self and others, staying motivated to quit smoking, triggers to smoke, barriers to quitting smoking, tips for managing cravings and withdrawal, and relapse prevention and recovery.

For generating a diversity of QnA sources, the knowledge base was broad, drawn from Alexander Street transcripts of therapy sessions [90], NCI call center transcripts of web-based Live Chats [91], Fred Hutch Cancer Center's HABIT laboratory cessation counseling intervention transcripts [53], written clinical content from HABIT laboratory digital smoking cessation interventions [92,93], and manual generation of questions and answers by HABIT laboratory clinical team members. We created a sequence-to-sequence (seq2seq) training model and processed cleaned transcripts, generating >8000 QnA pairs specifically focused on the topic of quitting cigarette smoking. As detailed in step 3, our evaluations led us to revisit step 1. In this iteration, we generated 2000 new questions posed by 32 adults interested in quitting cigarette smoking in a Prolific web-based survey. Then, our HABIT laboratory clinical team members generated answers to those questions. We also generated 1000 additional new questions from NCI call transcripts and transcripts related to our laboratory's intervention research. This iterative process yielded >11,000 QnA pairs.

Step 11.2. Training LLMs on QnA Pairs

The second step was to use these QnA pairs to train a series of LLMs. The first LLM deployed was an Azure application programming interface customized for the chatbot's 6-level chat architecture:

1. Motivations to quit smoking, triggers to smoke, and frequency of smoking
2. Dispatcher dialogue flow, to interpret and route utterances to the correct target, either Azure's 8000 QnA library or a custom AI model
3. QnA-maker-based "caring" personality chitchat to express empathy and provide supportive interactions with the user
4. QnA-based custom "help" to respond to trigger words such as suicide
5. Custom transformer-based model to handle chitchat QnA scenarios in dialogues that do not align with any specific intent
6. Custom transformer-based model designed to address smoking-related QnA scenarios in dialogues that do not align with any specific intent

In 2021, we tested 4 neural network models: a preexisting DialoGPT "Chitchat" model, 2 smoking content–trained DialoGPT models (1 trained on our 8000 QnA pairs and 1 on NCI chatlogs), and 1 smoking content–trained ParlAI model. We then accessed OpenAI's GPT-3 and conducted a comparative analysis of its model performance against our then best-performing model DialoGPT. GPT-3 outperformed the specially trained DialoGPT. Subsequently, to further improve GPT-3 performance, we compared the following model options: (1) Davinci, (2) Curie, and (3) fine-tuned Curie, concluding that the fine-tuned Curie model provided the best performance. Upon the release of GPT-3.5 (March 2022), we compared its performance with the fine-tuned GPT-3 Curie model [94]. This evaluation involved batch testing of questions that were not present in the QnA library, followed by a ranked review of answers (Textbox 3). GPT-3.5 provided better or comparable answer quality but required additional contextualizing parameters to match fine-tuned performance tailored for smoking-specific content.

Therefore, we used GPT-3.5 to expand the Azure QnA library by generating variations of existing questions and answers. This procedure expanded the Azure QnA library's capacity to accommodate a wider range of user question formulations, providing QuitBot with greater control over the quality of answers delivered to users.

Textbox 3. Comparisons of large language model answers to 3 sample questions on quitting smoking.

Sample question 1: "What is the quickest way to stop smoking?"

- Davinci: "The quickest way to stop smoking is to not start!"
- Curie: "Tapering down is the best way to stop smoking. Quitting smoking is a process. Make a quit plan and notice what triggers you to smoke."
- Fine-tuned Curie: "Our program was created to give you a plan that will work for you. We start by making a quit plan. Write down what inspires you to quit smoking. Choose a quit date and start cutting back on cigarettes."
- GPT-3.5: "The quickest way to stop smoking is to quit cold turkey, which means stopping smoking abruptly and without any aids such as nicotine replacement therapy or medication. However, quitting cold turkey can be difficult and the withdrawal symptoms can be challenging. We found that it works best to follow the program steps: make a quit plan and notice what triggers you to smoke. Then use quit smoking medication to help with withdrawal and use our tools to help urges pass as you quit gradually. Either way, it's important to find a method that works for you and stick with it."

Sample question 2: "I smoke when I am lonely. What should I do?"

- Davinci: "You are addicted to the nicotine in cigarettes. The only way to truly stop smoking is through nicotine replacements or prescription medications."
- Curie: "It's hard to feel lonely. Notice what thoughts and feelings are there and try not smoking them away."
- Fine-tuned Curie: "The first step is to make a quit plan. Write down what inspires you to quit smoking. Choose a quit date and start cutting back on cigarettes. Be aware of what triggers you to smoke."
- GPT-3.5: "It's hard to feel lonely. Notice what thoughts and feelings are there and try not smoking them away. Trying to be a friend to yourself in that moment and do something that would be nourishing to you."

Sample question 3: "Aren't e-cigarettes like the nicotine inhaler?"

- Davinci: "No, because unlike nicotine inhalers, e-cigarettes are used as simultaneous consumption of nicotine and air."
- Curie: "No. While e-cigarettes/vaping don't have the harmful contaminants of tobacco smoke they are not safe."
- Fine-tuned Curie: "They appear similar but are actually very different. The content of the nicotine inhaler is regulated by the FDA and when used stays mostly in throat and mouth. E-cigarettes are less harmful than regular cigarettes however, they are still addictive and contain noxious chemicals which can damage the heart and blood vessels."
- GPT-3.5: "Although they may look similar, the content of nicotine inhalers and e-cigarettes are actually different. The nicotine inhalers are regulated by the FDA and the nicotine content stays mostly in the throat and mouth. The e-cigarettes are not regulated, and the nicotine content can vary greatly, sometimes having more nicotine than a cigarette. They can also have other harmful substances besides nicotine. Because it is heated into a mist, the vapor goes into the lungs."

Step 11.3. Evaluating LLM Outputs

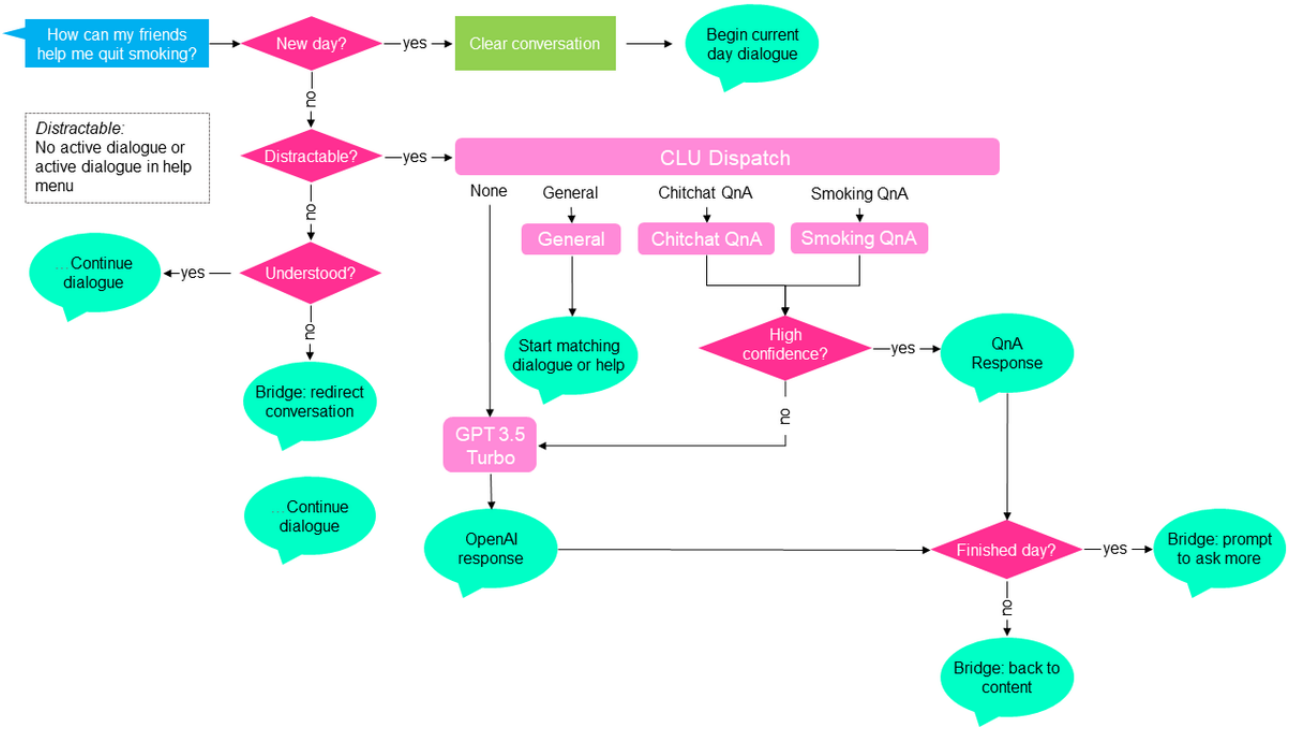
The third step in the development of the QuitBot's QnA feature involved evaluating the results of the LLMs using both automated and manual evaluation methods as well as user testing from adults who wanted to quit smoking. The automated evaluation used a scoring system of entropy (pertinence) and perplexity (grammaticalness) scores to rate the quality of answers generated by the LLMs. In addition, we implemented a filtering and scoring system to enhance the QuitBot's ability to communicate when it fails to comprehend a question. In such cases, the QuitBot asks users to be more specific if the confidence score for an answer falls below a certain threshold. The manual evaluation was conducted by trained raters in our HABIT laboratory, who hand scored the results of the models on measures of answer *accuracy* (yes or no), *repetitiveness* (yes or no), and *tone* (acceptable or needs improvement). Answers requiring improvement underwent manual revisions and were included into future iterations of model testing.

After retraining the model, in 2022, adults who smoke (n=14) were recruited from nationally placed Facebook advertisements to provide feedback on QuitBot prototype over a period of 14 days. Participant demographics were as follows: mean age 44.6 (SD 9.9) years, 43% (6/14) from minority race and ethnicity backgrounds, 64% (9/14) female, and 50% (7/14) had less than a bachelor's degree. A PhD-level UX researcher with 20 years of experience conducted the interviews. Deductive thematic analysis organized the user's responses by grouping them into themes, reviewing the themes, and then interpreting them [56-58].

Participants were highly receptive to the structured clinical conversations and noted that the free-form chats required additional fine-tuning to enhance comprehension of the users' questions. On the basis of this feedback, we determined the final organization of the chatbot architecture, combining 2 different models (Figure 5):

1. Microsoft's Azure QnA maker to handle questions matching our library of 11,000 QnA pairs that are based on scientific and clinical expertise. QnA maker uses Microsoft Language Understanding framework (LUIS) to leverage transformer models for responding to structured questions, with vetted answers stated in a professional manner.
2. Fine-tuned GPT-3.5 Turbo model to respond to questions that are not within our library of QnA pairs. Answers accurately with human-like variability, with different wording variations each time. GPT-3.5 is a backup to our QnA library.

Figure 5. QuitBot's architecture for handling free-form questions about quitting smoking. CLU: conversational language understanding; QnA: question and answer.



From the perspective of computer science, the QuitBot's chatbot's program runs on a *finite state machine* [95], which is a model of a system that runs on a limited set of modes. Depending on which mode the program is in, the QuitBot will behave in one manner or another. Ellen's state machine tracks what it is currently doing and combines interactions with the user to determine the next state. As the user can say anything, a hierarchy of possible intentions ranked by importance to the current state is used to decide the response to the user and the next state of the conversation.

A daily welcome exchange is an example. In this example, it is the start of the day and Ellen is in the "welcome" state. The user opens the app and says, "Hi!" To handle this prompt, the user's text goes through several steps in the finite state machine, illustrated in Figure 5: (1) categorize the user's intent; (2) determine if the intent is relevant to the current state of the conversation; (3) accordingly, move the current state of the conversation; and (4) formulate a response. In this case, "Hi!" is interpreted as a greeting intent, which is relevant to the current conversation. Ellen moves the state of conversation to "daily check-in" and responds with a greeting of her own, "Good morning. Thanks for checking in." When formulating a response, the user's intent determines which AI model will be used. General banter goes to a "chitchat" model powered by our Azure QnA library, smoking questions go to the Azure QnA model specifically trained on smoking questions, and unknown or low confidence intents go to GPT-3.5 Turbo. Finally, prewritten responses from our Azure QnA library that fit into the current conversation are used for specific scenarios.

Final Version of QuitBot

The final version of QuitBot is a stand-alone app that features (1) a personal coach (named "Ellen") who supports the user; (2) a series of 42 days of 2- to 3-minute structured clinical conversations with Ellen, guiding the user through distinct stages of quitting smoking; and (3) the ability for users to pose any free-form question related to quitting smoking. The structured conversations provide the valuable function of a clear step-by-step program for staying motivated, learning about one's triggers to smoke, setting a quit date, and staying smoke-free. Complementing the structured conversations, the free-form question feature provides users the freedom to ask their own questions, the option to address unique clinical needs, and the opportunity to follow-up on the content provided in the structured conversations. The combination of both structured and free-form conversation features is intended to balance their main strengths and limitations: the structured clinical format offers a guided program on quitting smoking, albeit with limited user question flexibility, while the open-ended format provides freedom but may encounter instances of not fully understanding the user's questions to give them clear guidance, despite the positive performance of the QnA feature thus far. Representative screenshots of QuitBot are provided in Multimedia Appendix 1.

Discussion

Principal Findings

This paper described the research group's 4-year process of developing a conversational chatbot for cigarette smoking cessation ("QuitBot"). The user-centered development process yielded a comprehensive quit smoking program that follows a series of 42 days of 2- to 3-minute structured clinical conversations. The program content covers topics ranging from motivations to quit, setting a quit date, choosing FDA-approved medications, identifying and coping with a wide range of triggers to smoke, and recovering from lapses or relapses. The program content, which follows the US Clinical Practice

Guidelines for smoking cessation, is presented as a continuous conversation, built on user input from prior conversations. QuitBot is available for both proactive and on-demand assistance at any time. Users can continue to interact with QuitBot after completing the 42 days of conversations.

Pilot RCT testing of QuitBot showed that the intervention had high user engagement and promising cessation rates, especially among participants who completed their assigned intervention. However, Facebook made policy changes that revoked access permissions to proactively outreach (eg, to invite participant to check in or start a conversation), effectively removing our ability to proactively contact users (restricting that ability to news-related apps only). This limitation was addressed by changing the FM communication platform to a stand-alone smartphone app that is fully accessible and controllable by the development and research teams.

The primary feedback from users was their frustration that the QuitBot could not respond to their own questions about quitting smoking. Therefore, we created the core conversational feature that would allow users to ask free-form and open-ended questions about quitting cigarette smoking and for the QuitBot to respond with accurate, concise, professional, and nonrepetitive answers. We developed a library of 11,000 QnA pairs on the topic of quitting cigarette smoking. The results of our model testing showed that Microsoft's Azure-based QnA maker could handle any question that matched our library of 11,000 QnA pairs. In contrast, a fine-tuned, contextualized GPT-3.5 could answer new questions that were not within our library of QnA pairs.

QuitBot has several key limitations that might present a challenge for users who expect fast responses to their questions. QuitBot was designed for users to wait until the end of the 2- to 3-minute structured clinical conversations before they can ask free-form questions. This design element was necessary to prevent breaking the logic of each of the structured conversations and thereby going off on tangents without an ability to return to the structured conversation. We address this design element by asking the user to hold onto their questions until the end of the structured conversation at various times throughout the program. To date, this message appears to have been effective at training the user to wait until the end of the structured conversation to ask free-form questions. The second major limitation is the response time latency for free-form questions when the GPT servers are running at capacity. While usually the response latency is only a few seconds, we have observed some instances where it can take up to 30 seconds. To address this potential delay, we inform users that it may take a moment to answer their questions and appreciate their patience.

By contrast, this study has numerous strengths that have the potential to advance clinical intervention development research and practice to aid smoking cessation. Most importantly, this study illustrates the value of following a methodical, user-centered design framework in the development of technology interventions. The framework has yielded a chatbot with a comprehensive step-by-step clinical program for quitting smoking and possesses a broad knowledge base on the topic of quitting smoking. QuitBot allows users to ask free-form and open-ended questions about quitting smoking, with answers informed by a broad set of clinical experience and scientific research. This technical capability has been afforded by the LLMs that underlie the state-of-the-art versions of Azure QnA Maker and GPT. The result is that users can obtain accurate and informative answers to their questions, which would otherwise be difficult to glean and evaluate from other accessible digital resources such as internet searches.

By contrast, prior reports of chatbots only address certain aspects of the quit smoking process, such as providing reflections on the pros and cons of smoking or helping ambivalent adults who smoke in contemplating a quit attempt. Only 8% of participants rated such chatbots as helpful [36]. Similarly, early iterations of QuitBot, which relied on a forced-choice answer format, left participants wishing responses tailored to their individual needs [96]. Likewise, a study of 6 users of a tablet-based chatbot, aimed at encouraging them to contemplate quitting and set a quit date, was limited by a forced-choice answer format [97]. In the only prior RCT of a smoking cessation chatbot, responses were confined to preset scripts and had an outcome data retention rate of only 45% [98].

Although users are informed that QuitBot is only a computer program, the supportive and conversational tone of the messages has the potential to lead to a long-term social-emotional connection. Indeed, the interim trial result of a mean of 72 days from first to last use is longer than we have observed in rule-based SMS text messaging interventions for smoking cessation (which typically last about 7 days) [85,99,100] and longer than typical human clinician–delivered interventions, such as telephone quit coaching (which typically last about 7 days) [101,102]. The length of intervention engagement is a strong predictor of treatment success [103,104], so these initial results on QuitBot's engagement certainly appear promising.

While we developed and tested QuitBot in the United States for an English-speaking audience, the program could be tailored to other nationalities and languages across the world. As a health behavioral change platform, QuitBot has the potential to be adapted to other behavior changes, including alcohol and drug use, dietary change, and physical activity.

Lessons Learned

QuitBot was developed in the midst of rapid changes in LLM technology, during what is arguably one of the most rapidly transformative periods of AI history (2020 to 2023) [105,106]. Thus, the most important lesson we learned was the value of investing the time in continuing to iterate and improve on our free-form QnA feature as new LLMs were continuously being released. Indeed, when we started to develop the free-form QnA feature in 2020, LLM capabilities were primitive by the current (January 2024) standards. Despite being based on 345 *million* parameters, we learned that DialoGPT was limited in its ability to determine the intent of our questions. By the time GPT-3.5 was released (based on 175 *billion* parameters), the performance of the free-form QnA feature was far superior, which in turn allowed us to improve the quality of answers provided by Azure QnA.

The second most important lesson we learned is the challenge of training an LLM model for a specific clinical domain. The popular press has provided ample examples of LLMs such as GPT providing very detailed answers to questions in a wide variety of topics [105,107,108]. While much has been written about the tendency for LLMs to "hallucinate" (ie, providing confident-sounding answers that are factually wrong or fabricated) [109], the more common problem we encountered in our development process was providing an extensive knowledge base to address highly specific questions within a clinical domain. In our experience, off-the-shelf LLMs are like dilettantes: they possess broad knowledge but lack depth in a particular subject. From this project, we glean that this characteristic holds particularly true when the subject matter requires clinical expertise and familiarity with scientific literature within a specific clinical domain. Overcoming this challenge required multiple iterations to build a knowledge base grounded in empirically supported best practices for smoking cessation. The responses needed to be accurate and clinically sensitive, suggesting that a similar knowledge-building process will be essential for developing chatbots in any other clinical domain.

Conclusions

The development process yielded a comprehensive, fully developed, quit smoking program delivered through a conversational chatbot. Iterative testing led to improvements in the delivery platform, and a core LLM–supported conversational feature was integrated, enabling users to pose open-ended questions about quitting cigarette smoking. Our next step is testing QuitBot's efficacy for smoking cessation in a full-scale RCT.

Acknowledgments

The authors gratefully acknowledge the National Cancer Institute for providing their call center transcripts; James Jacoby and Tara Ward at Moby Inc for providing QuitBot development expertise and technical project management; Jessica Harris, Sarah Hwang, Julie Packard, Victoria Sanborn, and Christeine Terry at the Fred Hutch HABIT laboratory for supporting the QnA development; Daniella Kim at Halibut Flats for leading the user testing; Andrew Pham and Adam Rhine for their design and programming expertise in the early phase of developing QuitBot; the research staff at OpenAI for providing access to and technical support for fine-tuning GPT; and the research staff, including Anusua Trivedi, Marci Strong, and Thomas Roca, at Microsoft's AI for Good for providing technical expertise in large language models. An interview about QuitBot with the lead author and Trevor Noah is available on YouTube [110].

Conflicts of Interest

None declared.

Multimedia Appendix 1
Representative screenshots of QuitBot.
[DOCX File , 1231 KB - mhealth_v12i1e57318_app1.docx]

Multimedia Appendix 2
CONSORT eHealth checklist.
[PDF File (Adobe PDF File), 95 KB - mhealth_v12i1e57318_app2.pdf]

References

1. Jha P. Avoidable global cancer deaths and total deaths from smoking. Nat Rev Cancer 2009 Sep;9(9):655-664. [doi: 10.1038/nrc2703] [Medline: 19693096]
2. Safiri S, Nejadghaderi SA, Abdollahi M, Carson-Chahhoud K, Kaufman JS, Bragazzi NL, et al. Global, regional, and national burden of cancers attributable to tobacco smoking in 204 countries and territories, 1990-2019. Cancer Med 2022 Jul;11(13):2662-2678 [FREE Full text] [doi: 10.1002/cam4.4647] [Medline: 35621231]
3. Smoking cessation: a report of the surgeon general. U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health. 2020. URL: https://www.hhs.gov/sites/default/files/2020-cessation-sgr-full-report.pdf [accessed 2024-07-12]
4. Patnode CD, Henderson JT, Coppola EL, Melnikow J, Durbin S, Thomas RG. Interventions for tobacco cessation in adults, including pregnant persons: updated evidence report and systematic review for the US preventive services task force. JAMA 2021 Jan 19;325(3):280-298. [doi: 10.1001/jama.2020.23541] [Medline: 33464342]
5. Littman D, Sherman SE, Troxel AB, Stevens ER. Behavioral economics and tobacco control: current practices and future opportunities. Int J Environ Res Public Health 2022 Jul 04;19(13):8174 [FREE Full text] [doi: 10.3390/ijerph19138174] [Medline: 35805833]
6. Choi K, Jones JT, Ruybal AL, McNeel TS, Duarte DA, Webb Hooper MW. Trends in education-related smoking disparities among U.S. Black or African American and White adults: intersections of race, sex, and region. Nicotine Tob Res 2023 Mar 22;25(4):718-728 [FREE Full text] [doi: 10.1093/ntr/ntac238] [Medline: 36239224]

7. Ho JY, Elo IT. The contribution of smoking to black-white differences in U.S. mortality. Demography 2013 Apr;50(2):545-568 [FREE Full text] [doi: 10.1007/s13524-012-0159-z] [Medline: 23086667]
8. Kcomt L, Evans-Polce RJ, Engstrom CW, West BT, McCabe SE. Discrimination, sexual orientation discrimination, and severity of tobacco use disorder in the United States: results from the national epidemiologic survey on alcohol and related conditions-III. Nicotine Tob Res 2021 May 24;23(6):920-930 [FREE Full text] [doi: 10.1093/ntr/ntaa197] [Medline: 32996575]
9. Arrazola RA, Griffin T, Lunsford NB, Kittner D, Bammeke P, Courtney-Long EA, et al. US cigarette smoking disparities by race and ethnicity - keep going and going!. Prev Chronic Dis 2023 Jun 01;20:E45 [FREE Full text] [doi: 10.5888/pcd20.220375] [Medline: 37262328]
10. Girvalaki C, Mechili E, Papadakis S, Nikitara K, Demin A, Trofor A, et al. Current practices and perceived barriers to tobacco-treatment delivery among healthcare professionals from 15 European countries. The EPACTT Plus project. Tob Prev Cessat 2020 Jan 21;6:6 [FREE Full text] [doi: 10.18332/tpc/115033] [Medline: 32548343]
11. Sakka S, Al-Shatanawi TN, Bataineh DZ, Haddad W, Al Tamimi S, AL Salamat H, et al. Knowledge, attitude, practice and perceived barriers towards smoking cessation services among community pharmacists. Pharm Pract 2022 Mar 29;20(1):2637. [doi: 10.18549/pharmpract.2022.1.2637]
12. Julius RJ, Novitsky MAJ, Dubin WR. Medication adherence: a review of the literature and implications for clinical practice. J Psychiatr Pract 2009 Jan;15(1):34-44. [doi: 10.1097/01.pra.0000344917.43780.77] [Medline: 19182563]
13. Twyman L, Bonevski B, Paul C, Bryant J. Perceived barriers to smoking cessation in selected vulnerable groups: a systematic review of the qualitative and quantitative literature. BMJ Open 2014 Dec 22;4(12):e006414 [FREE Full text] [doi: 10.1136/bmjopen-2014-006414] [Medline: 25534212]
14. Pipe AL, Evans W, Papadakis S. Smoking cessation: health system challenges and opportunities. Tob Control 2022 Mar;31(2):340-347. [doi: 10.1136/tobaccocontrol-2021-056575] [Medline: 35241609]
15. Smoking cessation—the role of healthcare professionals and health systems. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/tobacco/sgr/2020-smoking-cessation/fact-sheets/healthcare-professionals-health-systems/index.html [accessed 2024-07-12]
16. WHO report on the global tobacco epidemic, 2023: protect people from tobacco smoke. World Health Organization. 2023 Jul 31. URL: https://www.who.int/publications/i/item/9789240077164 [accessed 2024-07-12]
17. Babb S, Malarcher A, Schauer G, Asman K, Jamal A. Quitting smoking among adults - United States, 2000-2015. MMWR Morb Mortal Wkly Rep 2017 Jan 06;65(52):1457-1464 [FREE Full text] [doi: 10.15585/mmwr.mm6552a1] [Medline: 28056007]
18. Horvath AO, Greenberg LS. The Working Alliance: Theory, Research, and Practice. Hoboken, NJ: Wiley; Apr 1994.
19. Meier PS, Barrowclough C, Donmall MC. The role of the therapeutic alliance in the treatment of substance misuse: a critical review of the literature. Addiction 2005 Mar;100(3):304-316 [FREE Full text] [doi: 10.1111/j.1360-0443.2004.00935.x] [Medline: 15733244]
20. Becker MH, Rosenstock IM. Compliance with medical advice. In: Steptoe A, Matthews A, editors. Health Care and Human Behavior. London, UK: Academic Press; 1984.
21. Boardman T, Catley D, Grobe JE, Little TD, Ahluwalia JS. Using motivational interviewing with smokers: do therapist behaviors relate to engagement and therapeutic alliance? J Subst Abuse Treat 2006 Dec;31(4):329-339. [doi: 10.1016/j.jsat.2006.05.006] [Medline: 17084786]
22. Joe GW, Simpson DD, Dansereau DF, Rowan-Szal GA. Relationships between counseling rapport and drug abuse treatment outcomes. Psychiatr Serv 2001 Sep;52(9):1223-1229. [doi: 10.1176/appi.ps.52.9.1223] [Medline: 11533397]
23. Gardiner PM, McCue KD, Negash LM, Cheng T, White LF, Yinusa-Nyahkoon L, et al. Engaging women with an embodied conversational agent to deliver mindfulness and lifestyle recommendations: a feasibility randomized control trial. Patient Educ Couns 2017 Sep;100(9):1720-1729 [FREE Full text] [doi: 10.1016/j.pec.2017.04.015] [Medline: 28495391]
24. Provoost S, Lau HM, Ruwaard J, Riper H. Embodied conversational agents in clinical psychology: a scoping review. J Med Internet Res 2017 May 09;19(5):e151 [FREE Full text] [doi: 10.2196/jmir.6553] [Medline: 28487267]
25. Bickmore TW, Utami D, Matsuyama R, Paasche-Orlow MK. Improving access to online health information with conversational agents: a randomized controlled experiment. J Med Internet Res 2016 Jan 04;18(1):e1 [FREE Full text] [doi: 10.2196/jmir.5239] [Medline: 26728964]
26. King AC, Campero I, Sheats JL, Castro Sweet CM, Garcia D, Chazaro A, et al. Testing the comparative effects of physical activity advice by humans vs. computers in underserved populations: the COMPASS trial design, methods, and baseline characteristics. Contemp Clin Trials 2017 Oct;61:115-125 [FREE Full text] [doi: 10.1016/j.cct.2017.07.020] [Medline: 28739541]
27. Ayers JW, Poliak A, Dredze M, Leas EC, Zhu Z, Kelley JB, et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. JAMA Intern Med 2023 Jun 01;183(6):589-596. [doi: 10.1001/jamainternmed.2023.1838] [Medline: 37115527]
28. Whittaker R, Dobson R, Garner K. Chatbots for smoking cessation: scoping review. J Med Internet Res 2022 Sep 26;24(9):e35556 [FREE Full text] [doi: 10.2196/35556] [Medline: 36095295]

29. Bendotti H, Lawler S, Chan GC, Gartner C, Ireland D, Marshall HM. Conversational artificial intelligence interventions to support smoking cessation: a systematic review and meta-analysis. Digit Health 2023 Nov 03;9:20552076231211634 [FREE Full text] [doi: 10.1177/20552076231211634] [Medline: 37928336]
30. AI for quitting tobacco initiative. World Health Organization. 2020 Jul 13. URL: https://www.who.int/news-room/feature-stories/detail/ai-for-quitting-tobacco-initiative [accessed 2024-07-12]
31. Quit with Bella homepage. Quit With Bella. URL: https://www.quitwithbella.com/ [accessed 2024-07-12]
32. Alex - quit smoking. Alex Therapeutics. URL: https://www.alexquitsmoking.com/ [accessed 2024-07-12]
33. Loveys K, Antoni M, Donkin L, Sagar M, Broadbent E. Comparing the feasibility and acceptability of a virtual human, teletherapy, and an e-manual in delivering a stress management intervention to distressed adult women: pilot study. JMIR Form Res 2023 Feb 09;7:e42390 [FREE Full text] [doi: 10.2196/42390] [Medline: 36757790]
34. de Vito Dabbs A, Myers B, Mc Curry KR, Dunbar-Jacob J, Hawkins R, Begey A, et al. User-centered design and interactive health technologies for patients. Comput Inform Nurs 2009;27(3):175-183 [FREE Full text] [doi: 10.1097/NCN.0b013e31819f7c7c] [Medline: 19411947]
35. Bendotti H, Ireland D, Lawler S, Oates D, Gartner C, Marshall HM. Introducing quin: the design and development of a prototype chatbot to support smoking cessation. Nicotine Tob Res 2024 Apr 22;26(5):612-620. [doi: 10.1093/ntr/ntad217] [Medline: 37936253]
36. Almusharraf F, Rose J, Selby P. Engaging unmotivated smokers to move toward quitting: design of motivational interviewing-based chatbot through iterative interactions. J Med Internet Res 2020 Nov 03;22(11):e20251 [FREE Full text] [doi: 10.2196/20251] [Medline: 33141095]
37. ISO 9241-210:2019 Ergonomics of human-system interaction: part 210: human-centred design for interactive systems. International Organization for Standardization. 2019. URL: https://www.iso.org/standard/77520.html [accessed 2024-07-12]
38. ISO 13407:1999(en) Human-centred design processes for interactive systems. International Organization for Standardization. URL: https://www.iso.org/obp/ui/#iso:std:iso:13407:ed-1:v1:en [accessed 2024-06-28]
39. Warlick C, Richter KP, Catley D, Gajewski BJ, Martin LE, Mussulman LM. Two brief valid measures of therapeutic alliance in counseling for tobacco dependence. J Subst Abuse Treat 2018 Mar;86:60-64 [FREE Full text] [doi: 10.1016/j.jsat.2017.12.010] [Medline: 29415852]
40. Klemperer EM, Hughes JR, Callas PW, Solomon LJ. Working alliance and empathy as mediators of brief telephone counseling for cigarette smokers who are not ready to quit. Psychol Addict Behav 2017 Feb;31(1):130-135 [FREE Full text] [doi: 10.1037/adb0000243] [Medline: 28165273]
41. Havens L. Making Contact: Uses of Language in Psychotherapy. Cambridge, MA: Harvard University Press; 1986.
42. Okun BF, Kantrowitz RE. Effective Helping: Interviewing and Counseling Techniques. Boston, MA: Cengage Learning; 1976.
43. Laver J. Communicative functions of phatic communion. In: Organization of Behavior in Face-to-Face Interaction. Berlin, Germany: De Gruyter Mouton; 1975.
44. Dainton M, Stafford L. Routine maintenance behaviors: a comparison of relationship type, partner similarity and sex differences. J Soc Pers Relatsh 1993;10(2):255-271. [doi: 10.1177/026540759301000206]
45. Stafford L, Canary DJ. Maintenance strategies and romantic relationship type, gender and relational characteristics. J Soc Pers Relatsh 1991;8(2):217-242. [doi: 10.1177/0265407591082004]
46. Brown P, Levinson SC. Politeness: Some Universals in Language Usage. Cambridge, MA: Cambridge University Press; 1987.
47. Laver J. Linguistic routines and politeness in greeting and parting. In: Conversational Routine. Berlin, Germany: De Gruyter Mouton; 1981.
48. Bickmore T, Gruber A, Picard R. Establishing the computer-patient working alliance in automated health behavior change interventions. Patient Educ Couns 2005 Oct;59(1):21-30. [doi: 10.1016/j.pec.2004.09.008] [Medline: 16198215]
49. Archer RL. Self-disclosure. In: Wegner DM, Vallacher RR, editors. The Self in Social Psychology. New York, NY: Oxford University Press; 1980:183-205.
50. Shenk CE, Fruzzetti AE. The impact of validating and invalidating responses on emotional reactivity. J Soc Clin Psychol 2011 Feb 22;30(2):163-183. [doi: 10.1521/jscp.2011.30.2.163]
51. Ho A, Hancock J, Miner AS. Psychological, relational, and emotional effects of self-disclosure after conversations with a chatbot. J Commun 2018 Aug;68(4):712-733 [FREE Full text] [doi: 10.1093/joc/jqy026] [Medline: 30100620]
52. Clinical Practice Guideline Treating Tobacco Use and Dependence 2008 Update Panel, Liaisons, and Staff. A clinical practice guideline for treating tobacco use and dependence: 2008 update. A U.S. Public Health Service report. Am J Prev Med 2008 Aug;35(2):158-176 [FREE Full text] [doi: 10.1016/j.amepre.2008.04.009] [Medline: 18617085]
53. Bricker JB, Sullivan BM, Mull KE, Torres AJ, Carpenter KM. Full-scale randomized trial comparing acceptance and commitment therapy telephone-delivered coaching with standard telephone-delivered coaching among Medicare/uninsured Quitline callers. Nicotine Tob Res 2022 Oct 17;24(10):1556-1566 [FREE Full text] [doi: 10.1093/ntr/ntac052] [Medline: 35196381]
54. What is language understanding (LUIS)? Microsoft. 2024 Jan 19. URL: https://learn.microsoft.com/en-us/azure/ai-services/luis/what-is-luis [accessed 2024-07-12]

55. Bricker JB, Mull KE, Santiago-Torres M, Miao Z, Perski O, Di C. Smoking cessation smartphone app use over time: predicting 12-month cessation outcomes in a 2-arm randomized trial. J Med Internet Res 2022 Aug 18;24(8):e39208 [FREE Full text] [doi: 10.2196/39208] [Medline: 35831180]
56. Sharp H, Preece J, Rogers Y. Interaction Design: Beyond Human-Computer Interaction. Hoboken, NJ: John Wiley & Sons; 2019.
57. Human Centered Design Toolkit, Second Edition. San Francisco, CA: IDEO; 2011.
58. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
59. Botmock. crunchbase. URL: https://www.crunchbase.com/organization/botmock [accessed 2024-06-27]
60. Botmock. Wayback Machine. URL: https://web.archive.org/web/20210914233413/https://botmock.com/ [accessed 2024-07-15]
61. Chatfuel homepage. Chatfuel. URL: https://chatfuel.com/ [accessed 2024-07-12]
62. Fiore M. Treating Tobacco Use and Dependence: 2008 Update. Rockville, MD: U.S. Department of Health and Human Services, Public Health Service; 2008.
63. Kraus R. Facebook is really proud of its 300,000 business bots, despite claiming it will put 'people first'. Mashable. 2018 May 1. URL: https://mashable.com/article/messenger-bots-f8-2018 [accessed 2024-07-12]
64. Number of Facebook Messenger users in the United States from 2018 to 2022. Statista. URL: https://www.statista.com/statistics/558283/number-of-fb-messenger-users-usa/ [accessed 2024-07-12]
65. Acceptance of artificial intelligence chatbots by customers worldwide, as of 2017, by service. Statista. URL: https://www.statista.com/statistics/717098/worldwide-customer-chatbot-acceptance-by-industry/ [accessed 2024-07-12]
66. Node.js® - run JavaScript everywhere. OpenJS Foundation. URL: https://nodejs.org/en [accessed 2024-02-12]
67. Guest G, Bunce A, Johnson L. How many interviews are enough?: an experiment with data saturation and variability. Field Methods 2006;18(1):59-82. [doi: 10.1177/1525822x05279903]
68. Daae J, Boks C. A classification of user research methods for design for sustainable behaviour. J Clean Prod 2015 Nov 01;106:680-689 [FREE Full text] [doi: 10.1016/j.jclepro.2014.04.056]
69. Portz JD, Bayliss EA, Bull S, Boxer RS, Bekelman DB, Gleason K, et al. Using the technology acceptance model to explore user experience, intent to use, and use behavior of a patient portal among older adults with multiple chronic conditions: descriptive qualitative study. J Med Internet Res 2019 Apr 08;21(4):e11604 [FREE Full text] [doi: 10.2196/11604] [Medline: 30958272]
70. Leon AC, Davis LL, Kraemer HC. The role and interpretation of pilot studies in clinical research. J Psychiatr Res 2011 May;45(5):626-629 [FREE Full text] [doi: 10.1016/j.jpsychires.2010.10.008] [Medline: 21035130]
71. Rounsaville BJ, Carroll KM, Onken LS. A stage model of behavioral therapies research: getting started and moving on from stage I. Clin Psychol Sci Pract 2001;8(2):133-142. [doi: 10.1093/clipsy.8.2.133]
72. Bricker J, Wyszynski C, Comstock B, Heffner JL. Pilot randomized controlled trial of web-based acceptance and commitment therapy for smoking cessation. Nicotine Tob Res 2013 Oct;15(10):1756-1764 [FREE Full text] [doi: 10.1093/ntr/ntt056] [Medline: 23703730]
73. Bricker JB, Mull KE, Kientz JA, Vilardaga R, Mercer LD, Akioka KJ, et al. Randomized, controlled pilot trial of a smartphone app for smoking cessation using acceptance and commitment therapy. Drug Alcohol Depend 2014 Oct 01;143:87-94 [FREE Full text] [doi: 10.1016/j.drugalcdep.2014.07.006] [Medline: 25085225]
74. Boland VC, Stockings EA, Mattick RP, McRobbie H, Brown J, Courtney RJ. The methodological quality and effectiveness of technology-based smoking cessation interventions for disadvantaged groups: a systematic review and meta-analysis. Nicotine Tob Res 2018 Feb 07;20(3):276-285. [doi: 10.1093/ntr/ntw391] [Medline: 28034998]
75. Chen YF, Madan J, Welton N, Yahaya I, Aveyard P, Bauld L, et al. Effectiveness and cost-effectiveness of computer and other electronic aids for smoking cessation: a systematic review and network meta-analysis. Health Technol Assess 2012;16(38):1-205, iii-v [FREE Full text] [doi: 10.3310/hta16380] [Medline: 23046909]
76. Graham AL, Amato MS. Twelve million smokers look online for smoking cessation help annually: health information national trends survey data, 2005-2017. Nicotine Tob Res 2019 Jan 04;21(2):249-252 [FREE Full text] [doi: 10.1093/ntr/nty043] [Medline: 29660037]
77. Hall AK, Cole-Lewis H, Bernhardt JM. Mobile text messaging for health: a systematic review of reviews. Annu Rev Public Health 2015 Mar 18;36:393-415 [FREE Full text] [doi: 10.1146/annurev-publhealth-031914-122855] [Medline: 25785892]
78. Rodgers A, Corbett T, Bramley D, Riddell T, Wills M, Lin RB, et al. Do u smoke after txt? results of a randomised trial of smoking cessation using mobile phone text messaging. Tob Control 2005 Aug;14(4):255-261 [FREE Full text] [doi: 10.1136/tc.2005.011577] [Medline: 16046689]
79. Cheung YT, Chan CH, Lai CK, Chan WF, Wang MP, Li HC, et al. Using WhatsApp and Facebook online social groups for smoking relapse prevention for recent quitters: a pilot pragmatic cluster randomized controlled trial. J Med Internet Res 2015 Oct 22;17(10):e238 [FREE Full text] [doi: 10.2196/jmir.4829] [Medline: 26494159]
80. Durmaz S, Ergin I, Durusoy R, Hassoy H, Caliskan A, Okyay P. WhatsApp embedded in routine service delivery for smoking cessation: effects on abstinence rates in a randomized controlled study. BMC Public Health 2019 Apr 08;19(1):387 [FREE Full text] [doi: 10.1186/s12889-019-6727-z] [Medline: 30961557]

81. ICF International homepage. ICF International. URL: https://www.icf.com/ [accessed 2024-07-12]
82. Heinzen E, Sinnwell J, Atkinson E, Gunderson T, Dougherty G, Votruba P, et al. arsenal: an arsenal of 'R' functions for large-scale statistical summaries. The Comprehensive R Archive Network. 2021. URL: https://cran.r-project.org/web/packages/arsenal/index.html [accessed 2024-07-12]
83. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing. 2019. URL: https://www.R-project.org/ [accessed 2024-04-16]
84. Venables WN, Ripley BD. Modern Applied Statistics with S. Berlin, Germany: Springer; 2002.
85. Whittaker R, McRobbie H, Bullen C, Rodgers A, Gu Y. Mobile phone-based interventions for smoking cessation. Cochrane Database Syst Rev 2016 Apr 10;4(4):CD006611 [FREE Full text] [doi: 10.1002/14651858.CD006611.pub4] [Medline: 27060875]
86. Holm S. A simple sequentially rejective multiple test procedure. Scand J Stat 1979;6(2):65-70.
87. Watson NL, Mull KE, Heffner JL, McClure JB, Bricker JB. Participant recruitment and retention in remote eHealth intervention trials: methods and lessons learned from a large randomized controlled trial of two web-based smoking interventions. J Med Internet Res 2018 Aug 24;20(8):e10351 [FREE Full text] [doi: 10.2196/10351] [Medline: 30143479]
88. Scott-Sheldon LA, Lantini R, Jennings EG, Thind H, Rosen RK, Salmoirago-Blotcher E, et al. Text messaging-based interventions for smoking cessation: a systematic review and meta-analysis. JMIR Mhealth Uhealth 2016 May 20;4(2):e49 [FREE Full text] [doi: 10.2196/mhealth.5436] [Medline: 27207211]
89. Taylor GM, Dalili MN, Semwal M, Civljak M, Sheikh A, Car J. Internet-based interventions for smoking cessation. Cochrane Database Syst Rev 2017 Sep 04;9(9):CD007078 [FREE Full text] [doi: 10.1002/14651858.CD007078.pub5] [Medline: 28869775]
90. Counseling and psychotherapy transcripts: volume I. Alexander Street. URL: https://search.alexanderstreet.com/psyc [accessed 2023-11-21]
91. Get live, online assistance from NCI's Cancer Information Service. National Institutes of Health National Cancer Institute. URL: https://livehelp.cancer.gov/app/chat/chat_launch [accessed 2023-11-21]
92. Bricker JB, Mull KE, McClure JB, Watson NL, Heffner JL. Improving quit rates of web-delivered interventions for smoking cessation: full-scale randomized trial of WebQuit.org versus Smokefree.gov. Addiction 2018 May;113(5):914-923 [FREE Full text] [doi: 10.1111/add.14127] [Medline: 29235186]
93. Bricker JB, Watson NL, Mull KE, Sullivan BM, Heffner JL. Efficacy of smartphone applications for smoking cessation: a randomized clinical trial. JAMA Intern Med 2020 Nov 01;180(11):1472-1480 [FREE Full text] [doi: 10.1001/jamainternmed.2020.4055] [Medline: 32955554]
94. Models. OpenAI Platform. URL: https://platform.openai.com/docs/models [accessed 2024-02-12]
95. Wang J. Formal Methods in Computer Science. Boca Raton, FL: CRC Press; 2019.
96. Alphonse A, Stewart K, Brown J, Perski O. Exploring users' experiences with a quick-response chatbot within a popular smoking cessation smartphone app: semistructured interview study. JMIR Form Res 2022 Jul 07;6(7):e36869 [FREE Full text] [doi: 10.2196/36869] [Medline: 35797093]
97. Abdullah AS, Gaehde S, Bickmore T. A tablet based embodied conversational agent to promote smoking cessation among veterans: a feasibility study. J Epidemiol Glob Health 2018 Dec;8(3-4):225-230 [FREE Full text] [doi: 10.2991/j.jegh.2018.08.104] [Medline: 30864768]
98. Olano-Espinosa E, Avila-Tomas JF, Minue-Lorenzo C, Matilla-Pardo B, Serrano Serrano ME, Martinez-Suberviola FJ, et al. Effectiveness of a conversational chatbot (Dejal@bot) for the adult population to quit smoking: pragmatic, multicenter, controlled, randomized clinical trial in primary care. JMIR Mhealth Uhealth 2022 Jun 27;10(6):e34273 [FREE Full text] [doi: 10.2196/34273] [Medline: 35759328]
99. Christofferson DE, Hertzberg JS, Beckham JC, Dennis PA, Hamlett-Berry K. Engagement and abstinence among users of a smoking cessation text message program for veterans. Addict Behav 2016 Nov;62:47-53 [FREE Full text] [doi: 10.1016/j.addbeh.2016.06.016] [Medline: 27318948]
100. Coa KI, Wiseman KP, Higgins B, Augustson E. Associations between engagement and outcomes in the SmokefreeTXT program: a growth mixture modeling analysis. Nicotine Tob Res 2019 Apr 17;21(5):663-669 [FREE Full text] [doi: 10.1093/ntr/nty073] [Medline: 29668984]
101. Piñeiro B, Wetter DW, Vidrine DJ, Hoover DS, Frank-Pearce SG, Nguyen N, et al. Quitline treatment dose predicts cessation outcomes among safety net patients linked with treatment via Ask-Advise-Connect. Prev Med Rep 2019 Jan 17;13:262-267 [FREE Full text] [doi: 10.1016/j.pmedr.2019.01.009] [Medline: 30723660]
102. Matkin W, Ordóñez-Mena JM, Hartmann-Boyce J. Telephone counselling for smoking cessation. Cochrane Database Syst Rev 2019 May 02;5(5):CD002850 [FREE Full text] [doi: 10.1002/14651858.CD002850.pub4] [Medline: 31045250]
103. Heminger CL, Boal AL, Zumer M, Abroms LC. Text2Quit: an analysis of participant engagement in the mobile smoking cessation program. Am J Drug Alcohol Abuse 2016 Jul;42(4):450-458. [doi: 10.3109/00952990.2016.1149591] [Medline: 27120396]
104. Zeng EY, Heffner JL, Copeland WK, Mull KE, Bricker JB. Get with the program: adherence to a smartphone app for smoking cessation. Addict Behav 2016 Dec;63:120-124 [FREE Full text] [doi: 10.1016/j.addbeh.2016.07.007] [Medline: 27454354]

105. Toloka Team. The history, timeline, and future of LLMs. Toloka. 2023 Jan 26. URL: https://toloka.ai/blog/history-of-llms/ [accessed 2024-07-12]
106. Tucci L. A guide to artificial intelligence in the enterprise. TechTarget. 2024 Jan 25. URL: https://www.techtarget.com/searchenterpriseai/Ultimate-guide-to-artificial-intelligence-in-the-enterprise [accessed 2024-07-12]
107. Bushwick S. What the new GPT-4 AI can do. Scientific American. 2023 Mar 16. URL: https://www.scientificamerican.com/article/what-the-new-gpt-4-ai-can-do/ [accessed 2024-07-12]
108. Roose K. GPT-4 is exciting and scary. The New York Times. 2023 Mar 15. URL: https://www.nytimes.com/2023/03/15/technology/gpt-4-artificial-intelligence-openai.html [accessed 2024-07-12]
109. Metz C, Collins K. 10 ways GPT-4 is impressive but still flawed. The New York Times. 2023 Mar 14. URL: https://www.nytimes.com/2023/03/14/technology/openai-new-gpt4.html [accessed 2024-07-12]
110. How can large language models help combat addiction | The Prompt with Trevor Noah. YouTube. URL: https://www.youtube.com/watch?v=6bTQfWsWC7g [accessed 2024-07-12]

Abbreviations

AI: artificial intelligence
FDA: Food and Drug Administration
FM: Facebook Messenger
LLM: large language model
NCI: National Cancer Institute
OR: odds ratio
PPA: point prevalence abstinence
RCT: randomized controlled trial
SFT: SmokefreeTXT
UX: user experience

Edited by L Buis; submitted 12.02.24; peer-reviewed by H Bendotti, S Macha, GK Gupta; comments to author 28.02.24; revised version received 18.04.24; accepted 18.06.24; published 23.07.24.

Please cite as:
Bricker JB, Sullivan B, Mull K, Santiago-Torres M, Lavista Ferres JM
Conversational Chatbot for Cigarette Smoking Cessation: Results From the 11-Step User-Centered Design Development Process and Randomized Controlled Trial
JMIR Mhealth Uhealth 2024;12:e57318
URL: https://mhealth.jmir.org/2024/1/e57318
doi:10.2196/57318
PMID:38913882

Original Paper

Initial Evaluation of Acceptability, Engagement, and Effectiveness of the MO App to Provide Tailored and Comprehensive Support for Smoking Cessation: Development and Usability Study

Shuo Zhou[1], PhD; Paul Brunetta[2], MD; Joshva Silvasstar[3], MS; Greg Feldman[4], MA; Nicolas Oromi[5], MBA; Sheana Bull[3], MPH, PhD

[1]Department of Communication Studies, School of Communication and System Health Lab, Hong Kong Baptist University, Hong Kong, China (Hong Kong)

[2]Fontana Tobacco Treatment Center, University of California San Francisco, San Francisco, CA, United States

[3]Department of Community and Behavioral Health, mHealth Impact Lab, Colorado School of Public Health, University of Colorado, Aurora, CO, United States

[4]Mobile Applications for Connected Health, Lakewood, CO, United States

[5]Ingenious Agency, Montevideo, Uruguay

Corresponding Author:
Shuo Zhou, PhD
Department of Communication Studies
School of Communication and System Health Lab
Hong Kong Baptist University
913, No. 5 Hereford Rd, KL
Hong Kong
China (Hong Kong)
Phone: 852 34118230
Email: shuozhou@hkbu.edu.hk

Abstract

Background: Despite the growing availability of smoking cessation apps, low engagement and cessation rates have remained a significant challenge. To address this issue, we used a user-centered design to iteratively develop a mobile app (MO) to provide comprehensive, tailored, and evidence-based content to support smokers in their quitting journey.

Objective: This study examined the acceptability, use, and preliminary efficacy of the MO app for smoking cessation. Specifically, we sought to understand smokers' preferred features, engagement, and satisfaction with MO; identify concerns in using the app and ways to improve the app; and evaluate its smoking cessation outcomes.

Methods: Through 3 cohorts, we recruited 10, 12, and 85 adult smokers who attempted to quit smoking to pilot-test the MO app between December 2019 and July 2022. Participants were instructed to complete a baseline survey, interact with the app for 6 weeks, and fill in a postsurvey at week 6. Participants in cohort 3 completed an additional postsurvey at week 12. Participants' app use was tracked and analyzed. The primary outcome measures were participants' 7-day point prevalence abstinence at 6 and 12 weeks.

Results: Participants reported high levels of satisfaction with the MO app across all 3 cohorts, rating it between 4.40 and 4.76 on a scale of 5 for acceptability. Users engaged with app activities for an average of 89 to 159 times over 35 days. The most liked features of the app included "quit plan," "tracking," "reminders and notifications," "MOtalks," and "motivational quotes." The 7-day point prevalence abstinence rate of the modified intention to treat population in cohort 3 was 58% at 6 weeks and 52% at 12 weeks. Those who interacted more frequently with app features and engaged with more diverse activities were more likely to maintain abstinence at weeks 6 and 12. For each additional time logged into the app, the odds of staying abstinent at week 12 increased by 5% (odds ratio [OR] 1.05, 95% CI 1.01-1.08). Participants who earned >5000 points during app use also had higher odds of quitting at both 6 weeks (OR 3.12, 95% CI 1.25-7.75) and 12 weeks (OR 4.65, 95% CI 1.83-11.76), compared with those who earned <5000 points.

Conclusions: Our study demonstrated that MO is a feasible mobile phone app with high acceptability and usability and can effectively deliver smoking cessation support to individuals who want to quit. Implications for developing and evaluating mobile phone apps for smoking cessation are discussed.

(JMIR Mhealth Uhealth 2024;12:e55239) doi:10.2196/55239

KEYWORDS

smoking cessation; tobacco; mobile phone app; mHealth; mobile health; iterative design; feasibility; acceptability; engagement; efficacy; mobile phone

Introduction

Background

Tobacco use remains the leading cause of preventable death in the United States and globally, with nearly half a million people dying each year in the United States alone due to tobacco-related diseases. Over a billion people worldwide still smoke [1], which highlights the enormous global burden of tobacco use. Recent evidence indicates that quitting before the age of 40 years can reduce the risk of premature death by 90%. This underscores the importance of effective tobacco cessation interventions. Cessation at any age can provide major disease risk reduction benefits. Despite the increasing awareness of the adverse health effects of smoking and the clear benefits of quitting, cessation rates remain low, particularly among highly nicotine-dependent individuals who face a high risk of relapse [2]. Nicotine is one of the most addictive substances known in medicine, which explains the major challenges of successful quitting and contributes to the persistently low cessation rates. Thus, there is a critical need for accessible and clinically validated smoking cessation interventions that can effectively address the barriers and challenges associated with nicotine addiction.

Compared with spontaneous unmedicated cessation rates (about 2%/year), nicotine cessation therapies have documented success rates of approximately 20% at 1 year. Motivated smokers who receive professional counseling and evidence–based cessation medications can have a success rate as high as 35% at 1 year [2,3]. However, traditional counseling interventions have limited reach, and the multiple available medications for cessation can be confusing for people seeking help. Patients face various barriers to accessing face-to-face tobacco cessation services, including financial challenges, language issues, traveling difficulties, and low-time commitment for in-person counseling [4,5]. The social stigma for free telephone-based counseling [6], lower-income patients' distrust of health care systems [7], and low awareness of the health risks of smoking further exacerbate these challenges for smokers with low socioeconomic status (SES). Hence, there is an urgent need to disseminate high-quality cessation treatments to the widest possible target audiences, including those in remote locations, those without health insurance, and those who cannot or will not receive counseling in person. Innovative strategies that deliver evidence–based cessation interventions to smokers attempting to quit could further enhance the success rate of quitting and prevent tobacco-related diseases while improving both physical and psychological well-being. Therefore, exploring novel approaches to promote smoking cessation is a critical public health priority.

Mobile Apps for Smoking Cessation

The ubiquity and widespread availability of smartphones provide an opportunity to extend the reach of smoking cessation professional counseling and evidence-based treatment beyond clinics. As of 2021, 85% of the adults in the United States own a smartphone, which covers low-income populations who are known to have higher smoking rates. In 2017, mobile health (mHealth) apps were downloaded 3.7 billion times worldwide [8], and this number increased dramatically during and after the COVID-19 pandemic. The high demand for mHealth apps represents a promising opportunity to leverage mobile technologies to facilitate smoking cessation and reduce health disparities in quitting.

Increasing evidence has shown the positive outcomes of mobile phone apps for smoking cessation [9-15]. According to a review, the cessation rate of people using mobile apps ranges from 13% to 24%, higher than the average cessation rate of using SMS text messaging alone (10%) [16]. Combining smartphone-based interventions with pharmacotherapy yields greater smoking abstinence rates than using pharmacotherapy alone [17]. However, limited and mixed evidence exists for the long-term effectiveness and retention rates of mobile phone apps. Very few of the popular smoking cessation apps (4% of the top 50 apps) are evidence based and adhere to the US Public Health Service's Clinical Practice Guidelines for Treating Tobacco Use and Dependence [18-21]. Currently, available apps lack tailored features and integration of multiple evidence–based approaches into the app design, such as a comprehensive review of medication options. In addition, existing apps have limited support throughout the different phases of smoking cessation, including preparation to quit, maintenance of cessation, and relapse prevention and support.

We developed a comprehensive and integrative smoking cessation mobile phone app called "MO"—an abbreviation for "mobile" and "motivation"—to address these limitations. The aim of MO is to provide user-centered, tailored cessation support using evidence-based approaches, integrating cognitive-behavioral therapy, peer support, and medication information. MO provides learning materials regarding how to quit smoking in different formats (eg, texts, audios, videos, and quizzes) and explains in easy-to-understand language the intensely addictive nature of nicotine. The app offers cessation and relapse support throughout the "prequit and planning," "quit and maintenance," and "relapse prevention" phases of smoking cessation and helps the user remain motivated toward quitting. Existing smoking cessation apps often have low user engagement and retention rates. MO seeks to address this by consistently motivating users to quit and encouraging long-term engagement to increase the efficacy of quitting.

Development of MO Smoking Cessation App

The MO app was developed with guidance from the US Public Health Service's Clinical Practice Guidelines for Treating Tobacco Use and Dependence, using the 5 *A*'s model (ask, advise, assess, assist, and arrange) and more than 80

evidence–based behavioral modification and motivation techniques. The app was designed to guide users step-by-step throughout their quitting process.

In the prequit or planning phase, there are 3 main functions: creating an individualized profile, tracking smoking behaviors, and advising on tasks and activities to prepare for quitting. The app's first function (setting up profile) assesses factors, such as level of nicotine addiction, health conditions, use of medications, quit date, motivations, and reasons for quitting, gathered from user profile information to make tailored recommendations based on proprietary algorithms and using tenets of motivational interviewing. The second function (tracking behaviors) is to help users track smoking behaviors (eg, concurrent activities, locations of smoking, time of the day, and the number of cigarettes smoked), identify triggers, and guide them through the process of choosing substitutive activities or distractions to replace smoking. The third function (task advising and learning) is to prepare users to stop smoking through education and specific activities, such as obtaining medications, modifying their environment, and informing family and friends to support their quit effort. General advice and strategies on how to stop smoking are available in text, audio, and video formats for users to learn at their own pace. Each day before the quit day, users can use this function to go through a specific theme preparing them to quit smoking. Cessation medications are described in detail and consistent with existing guidelines, yet none are specifically recommended, and the user is encouraged to make the decision about medication use with a health care provider.

In the quit or maintenance phase, the app also has 3 main functions. The first is the daily check-in, which prompts the users to track their moods, urges, smoking behavior, use of medications, and interactions with their social network. The second function is the "urge" button, which provides immediate support and strategies for managing urges to smoke. The third function delivers educational content through video clips and quizzes based on user motivations and data gathered in the prequit phase. The smoker will guide themselves through content that matches their highest interest and view new modules or review old modules to reinforce knowledge about smoking cessation.

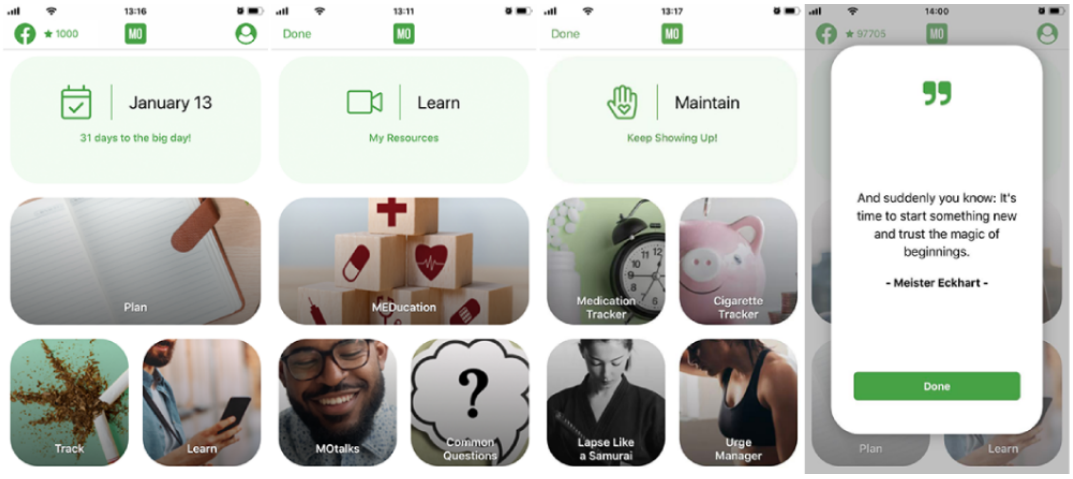
The app also includes functions to manage lapses (a momentary slip), relapses (return to smoking that may last for several days before resuming the cessation program), and collapses (a major reduction in motivation and return to smoking), which are often neglected by smoking cessation apps. The MO app anticipates these events, supports users through these inevitably difficult moments, and keeps users engaged and motivated. Specific functions for this phase include educational videos and audio segments helping users determine what happened, reestablish motivations, and how to prevent it the next time.

The app incorporates several features to encourage long-term user engagement. Every time a user opens the app, an uplifting quote will pop up to boost smokers' motivation to quit. Gamification features, including a point system and badges, help users celebrate their success in the quitting process. This feature visualizes users' progress and motivates them to stay abstinent. As each user completes a module, they will receive certain points (250, 500, or 1000) for the activity they engaged in. These points are designed to recognize the actions users take toward the quitting goal. Modules that are reviewed can be repeated to accumulate additional points. This "nudge" feature assigns more points to activities that have scientific evidence suggesting a positive role in cessation. Points are shown on the home screen, which may help users feel invested in their progress and remind them of their previous efforts if a lapse or relapse occurs after the quit date. Existing evidence has consistently suggested that financial incentives can improve smoking cessation outcomes [22]. As a potential reward, users who obtained >10,000 points will enter a lottery for a US $50 gift card in addition to the study compensation.

Informed by text-to-quit data suggesting social group aggregation may improve outcomes [23], the MO app has an internal chat feature that allows daily posting to the quitter's support network. This is a secure, customizable platform uniquely created for the app. The chat function allows for secure data management and professional moderation, with the goal to provide a supportive, understanding community breaking the negative effects of isolation and motivating users through positive social contagion. In addition, the story booth feature allows users to record and share brief video clips about their cessation journey and to view other people's stories.

The main features of the MO app are illustrated in Figure 1. The development of MO involved a web-based asynchronous focus group with end users to evaluate and provide suggestions for the initial prototypes of the app. Participants were asked to post comments to a web-based discussion board as to the viability, content usability, and attractiveness of each app element and the ability for each to engage the intended audience. We incorporated participants' comments and developed the beta version of the MO smoking cessation app.

Figure 1. Interface and main features of the MO smoking cessation app.



Research Questions and Hypotheses

This study used an iterative process to evaluate and improve the feasibility and usability of the beta version of the app. A series of studies were conducted to address the following research questions (RQs) and hypotheses related to the app's acceptability, user engagement, and preliminary effectiveness on smoking cessation.

- RQ1: Are participants satisfied with using MO for smoking cessation?
- RQ2: What are the most liked features of the app?
- RQ3: What are the least liked features of the app?
- RQ4: What are the suggestions for improving the app?
- RQ5: What are the 7-day point prevalence abstinence rates at 6 and 12 weeks after using the MO app?
- Hypothesis 1: Participants who engage with a greater number of app activities are more likely to self-report quitting at the 6 and 12-week follow-ups.
- Hypothesis 2: Participants who earned more points are more likely to self-report quitting at the 6 and 12-week follow-ups.

Methods

Overview of the Study Design

Studies were conducted in 3 participant cohorts to examine the RQs and hypotheses mentioned earlier. In cohorts 1 and 2, we examined acceptability and usability of the app among a group of end users to ensure the various features work as intended, are perceived acceptable by the intended users, and can be easily used. Features and interfaces of the app will be refined based on participants' use experience and feedback. In cohort 3, we additionally assessed the preliminary effectiveness of the MO app on smoking cessation outcomes.

Participant Recruitment

We intended to recruit 10, 10, and 80 participants into cohorts 1, 2, and 3, respectively, during the following 3 periods: December 19, 2019, to March 19, 2020; August 10 to December 10, 2020; and April 7 to July 26, 2022. Eligible participants included (1) persons aged between 18 and 59 years; (2) who were current smokers consuming at least 5 cigarettes daily for at least the past 12 months, as consistent with cessation trials [24]; (3) self-reported a relatively high level of commitment to quit smoking of at least 7 on a scale of 0 to 10; (4) who were interested in learning skills to quit smoking; (5) resided in the United States; (6) had at least daily access to their own iPhone with internet access; (7) were able to read English; and (8) were not using other smoking cessation interventions (including apps, medications, or other intervention studies). We excluded participants who were younger than 18 years or older than 59 years as they may have had different cessation motivations; who had health contraindications to nicotine patch use; who were actively taking medication for depression, anxiety, or quitting smoking; who used illicit drugs or marijuana; or who failed to provide contact information.

Participants were recruited using web-based advertising through Craigslist (Craig Newmark) posts, Facebook (Meta), and Google (Google LLC) advertisements, and a university listserv for research participant recruitment. People who saw the advertisements could also refer the study information to their smoking relatives or friends. Interested participants clicked the enrollment link in the advertisements, which directed them to a short screening survey. Once eligibility was established, participants were shown a short video containing the informed consent materials.

Eligible participants who consented to participate were enrolled in the study. They were asked to complete a baseline survey to document their smoking history, use of nicotine replacement, and other health behaviors. We documented factors known to correlate with smoking cessation, such as social support, attitudes and norms related to smoking, and intentions to quit.

Participants were instructed to download MO and interact with the app and its various features for 6 weeks. A 10-minute walkthrough video was sent to every participant to introduce the main features of the app. The walkthrough video is available in Multimedia Appendix 1. A postsurvey was developed to examine app acceptability and usability and was sent to participants at 6 weeks (day 42) and at 12 weeks (day 84, only for cohort 3). Participants that returned the 6-week and 12-week questionnaires were part of a completer's analysis of efficacy. Participants not returning the questionnaire were imputed as nonresponders and having returned to smoking.

Measures

Acceptability

Acceptability refers to the degree to which an intervention is deemed acceptable by its intended users. In cohorts 1 and 2, we measured users' acceptability of MO using 3 items: "I enjoyed using the MO smoking cessation app," "The MO smoking cessation app was useful to me," and "The MO smoking cessation app was easy to navigate" on a 5-point Likert scale (1=strongly disagree and 5=strongly agree). In cohort 3, we additionally asked 6 questions adapted from previous studies on app-based smoking cessation interventions [25] to have a more comprehensive evaluation of users' satisfaction with and perceived usefulness of the MO app for smoking cessation.

App Engagement

We tracked participants' interaction with and navigation through the MO app through Firebase Performance Monitoring (Google LLC) for 12 weeks. Specifically, we recorded the type of activities users engaged in within the app, whether they completed each module or task, and the frequency of engagement. We also asked users to identify their most liked and disliked app features, explain why they liked or disliked certain features, and provide suggestions for improving the app through open-ended survey questions.

Effectiveness

In cohort 3, smoking cessation outcomes were additionally examined. Participants were asked to self-report their smoking status at 6 and 12 weeks through 2 postenrollment surveys, including the 7-day point prevalence abstinence rate as the primary outcome variable, number of cigarettes smoked, and frequency of smoking in the past month. Self-report 7-day point prevalence abstinence, which is one of the commonly used measures for evaluating users' smoking cessation outcome [26], asks whether participants maintain smoke-free in the past 7 days.

Analytic Approach

The intent-to-treat (ITT) group typically includes all participants enrolled in the study who were assigned to receive a treatment. The modified ITT (mITT) is a subset of the ITT sample that removed participants who did not receive or start the treatment. In our study, participants who completed the screening test and successfully enrolled were considered the ITT group, and those who downloaded and used the MO app were considered the mITT group. The mITT group is more appropriate for the app engagement and preliminary efficacy analyses because they are the true participants who became exposed to and received the smoking cessation treatment, removing participants who did not download, open, or interact with the MO app at all. Because acceptability is a subjective evaluation of participants' satisfaction with the app, we were only able to analyze data input by participants who completed the postsurvey. Those who dropped out of the study were not included in the acceptability analysis.

To examine acceptability (RQ1), we calculated participants' average level of satisfaction with the MO smoking cessation app. We anticipated that at least 80% of the users will be "satisfied" or "very satisfied" with the app. To address RQs on usability (RQs 2-4), we qualitatively analyzed participants' responses to open-ended questions—to identify themes around (1) difficulties in using or navigating the app, (2) liked features, (3) disliked features, and (4) suggestions for improving the app. We also evaluated engagement by calculating multiple aspects of user experience with the app, including the number of times the app was opened and the frequency and types of features used.

To explore RQ 5, we assessed the percentage of participants remaining abstinent from cigarette smoking and use of other nicotine or tobacco products in the previous 7 days at 6 and 12 weeks after using the app. All participants who completed the baseline survey, downloaded the app, and answered the questionnaire were included in this completer analysis. Participants that were lost to follow-up at the 6th and 12th weeks were imputed to be smoking.

We also examined the dose response relationship between quit status and greater use of the app in general (hypothesis 1) and more points earned (hypothesis 2) using logistic regressions to test the proposed hypotheses. We conducted exploratory analyses to assess the relationship between use of specific features of the app and the participant's quit status.

Ethical Considerations

The study was reviewed and approved by the Colorado Multiple Institutional Review Board (protocol number 14-2225). Participants were asked to view the informed consent video. Contact information for research staff was included in the video in case participants had any questions during the consent process. The material covered in the video was made available in a printable document so the participant could download and print it if desired. Participation in this study was completely voluntary. Participants may decline to participate or withdraw from the study at any time without penalty. Participants in cohorts 1 and 2 received a US $20 gift card incentive for participation. For cohort 3, participants received a US $40 gift card incentive for participation, as 2 follow-up surveys (at 6 and 12 weeks) were collected. All information was kept confidential, and the data were deidentified. Study procedures were in accordance with institutional ethical standards for conducting human subjects research and with the Helsinki Declaration.

Results

Participant Recruitment and Demographics

We enrolled 15, 19, and 132 eligible participants (the ITT groups) into the 3 cohorts. Among them, 9 (60%), 12 (63%), and 85 (64.4%) participants (the mITT group) downloaded and used the MO app. All statistics were calculated based on this mITT sample. The retention rates at 6-week follow-up were 90%, 83%, and 66%, respectively. For cohort 3, the retention rate at 12 weeks was 55%. Participants' basic demographic characteristics, smoking history, level of commitment to quit smoking, and sources of recruitment were summarized in Table 1.

Table 1. Demographics of participants across the 3 cohorts.

Characteristic	Cohort 1 (n=9), n (%)	Cohort 2 (n=12), n (%)	Cohort 3 (n=85), n (%)
Sex			
Male	5 (56)	3 (25)	66 (78)
Female	4 (44)	9 (75)	19 (22)
Age (years)			
18-30	4 (44)	2 (17)	34 (40)
31-40	4 (44)	5 (42)	42 (49)
41-50	1 (11)	2 (17)	7 (8)
51-59	0 (0)	3 (25)	2 (2)
Ethnicity			
Hispanic	3 (33)	0 (0)	7 (8)
Non-Hispanic	6 (67)	12 (100)	78 (92)
Race			
Black or African American	2 (22)	5 (42)	54 (64)
White	6 (67)	6 (50)	30 (35)
Other	1 (11)	1 (8)	1 (1)
Age started smoking (years)			
13-17	4 (44)	7 (58)	12 (14)
18-21	2 (22)	4 (33)	39 (46)
22-30	2 (22)	1 (8)	30 (36)
31-40	1 (1)	0 (0)	2 (2)
41-50	0 (0)	0 (0)	1 (1)
Smoking history (years)			
1-5	3 (33)	4 (33)	35 (41)
5-10	4 (44)	4 (33)	33 (39)
10-20	2 (22)	4 (33)	13 (15)
>20	0 (0)	0 (0)	4 (5)
Attempted to quit smoking in the past year			
Yes	3 (33)	6 (50)	61 (72)
No	1 (11)	6 (50)	22 (26)
Not sure	5 (56)	0 (0)	2 (2)
Previous use of nicotine replacement			
Yes	—[a]	7 (58)	20 (24)
Have used gum[b]	—	4 (33)	16 (19)
Have used skin patches[b]	—	6 (50)	9 (11)
Have used lozenge[b]	—	2 (17)	9 (11)
Have used inhaler[b]	—	1 (8)	13 (15)
No	—	5 (42)	65 (77)
Parental smoking in childhood			
Yes	—	8 (67)	56 (66)
No	—	4 (33)	25 (39)
Not sure	—	0 (0)	4 (5)

Characteristic	Cohort 1 (n=9), n (%)	Cohort 2 (n=12), n (%)	Cohort 3 (n=85), n (%)
Be often around people who smoke (eg, family member, friend, or colleague)			
Yes	—	9 (75)	78 (92)
No	—	3 (25)	7 (8)
Often in social situations where others are smoking			
Yes	—	8 (67)	80 (94)
No	—	4 (33)	4 (5)
Not sure	—	0 (0)	1 (1)
Commitment to quit smoking (0-10 scale)			
7	2 (22)	3 (25)	10 (12)
8	0 (0)	1 (8)	13 (15)
9	2 (22)	4 (33)	17 (20)
10	5 (56)	4 (33)	45 (53)
Sources of knowing this study			
Facebook	8 (89)	5 (42)	48 (56)
Google	0 (0)	5 (42)	24 (28)
Craigslist	0 (0)	0 (0)	8 (9)
Friends	1 (11)	0 (0)	4 (5)
Listserv	0 (0)	2 (17)	1 (1)

[a]Not available.

[b]These questions were only asked among participants who indicated previous use of nicotine replacement.

Acceptability

Overall, participants had high levels of satisfaction with using the MO app. At 6 weeks, participants from the 3 cohorts reported 4.67, 4.40, and 4.76 out of 5 for app satisfaction. Table 2 shows details regarding participants' acceptability of the app.

Table 2. Acceptability of the MO app across the 3 cohorts.

Statements	Cohort 1: 6 weeks, mean (SD)	Cohort 2: 6 weeks, mean (SD)	Cohort 3: 6 weeks, mean (SD)	Cohort 3: 12 weeks, mean (SD)
Satisfaction				
I enjoyed using the MO smoking cessation app	4.63 (0.70)	4.60 (0.66)	4.86 (0.35)	4.87 (0.34)
The MO smoking cessation app was useful to me	4.75 (0.43)	4.30 (0.64)	4.71 (0.49)	4.62 (0.80)
The MO smoking cessation app was easy to navigate	4.63 (0.48)	4.30 (0.64)	4.70 (0.66)	4.62 (0.71)
Satisfaction with the app for cessation treatment				
I can rely on the app to provide guidance that will help me to quit smoking and stay quit.	—[a]	—	4.59 (0.53)	4.70 (0.55)
I feel that the app provides smoking cessation treatment that is personalized to my specific needs.	—	—	4.46 (0.71)	4.57 (0.58)
I believe that the app will help me to quit smoking and stay quit.	—	—	4.54 (0.66)	4.55 (0.72)
The app knows how to help me to quit smoking.	—	—	4.61 (0.53)	4.70 (0.47)
I believe I can depend on the app.	—	—	4.46 (0.60)	4.47 (0.88)
I find the app to be annoying.	—	—	1.31 (0.57)	1.48 (0.91)

[a]Not available.

App Engagement

On the basis of the tracked data, Table 3 summarizes participants' general use of the app and engagement with specific features. Due to technical issues, we were not able to track 3 participants' app use in cohort 3. Therefore, data analysis on app engagement in cohort 3 was based on 82 (96%) of 85 participants. In general, participants perceived almost all the features to be useful. The top 5 most liked features of the app included "quit plan," "tracking," "reminders and notifications," "MOtalks," and "motivational quotes." MOtalks are short video or audio lectures of a specific theme related to smoking cessation. Participants explained that the "quit plan" feature "helped with learning about the habit and situations that trigger me to smoke and how to make a plan to deal with it," and the "tracking" feature "showed me some patterns I wasn't aware of." Participants liked the "reminders and notifications" because "the daily reminders came as pop-up notifications to enter your smoking status for the day. With that it makes you feel responsible and like you are making progress." Participants also enjoyed the "quotes" because they were "uplifting" and "motivational." These findings were consistent with users' actual engagement with various features, with "login" as the most frequently used function, followed by "quit plan" and "medication tracker." The least liked feature of the app was the "comments." Because this feature was not frequently used, participants mentioned that "I could not see any comments or messages there."

Table 3. Use of the MO app.

	Cohort 1	Cohort 2	Cohort 3
General use of the app, mean (SD)			
Duration of app use (days)	37.56 (29.79)	35.33 (36.56)	46.31 (41.52)
Total activities engaged (times)	89.22 (76.62)	159.60 (213.97)	158.89 (236.71)
Unique activities engaged (types)	37.11 (23.85)	79.42 (76.45)	67.66 (80.83)
Log-in times	17.56 (21.38)	13.75 (28.57)	9.98 (15.79)
Points earned	20,430 (25916.20)	26,933.75 (33859.66)	28,764.48 (40,157.13)
Engagement with specific app features (most frequently engaged features), mean score			
Quit plan	8.00	6.46	12.47
Medication tracker	4.89	3.00	3.56
MOtalks: meet your hosts	0.89	4.5	2.25
Why quit	1.11	0.83	2.20
MOtalks: denial stories	0.22	2.33	2.12
MOtalks: triggers	0.33	1.00	1.85
Quit date setting	1.11	2.75	1.74

Suggestions for Improvement

The following themes emerged from thematic analysis of participants' responses to open-ended questions regarding things to be improved for the app. The first identified theme was "availability." Participants recommended "enabling it in all devices—that is, iOS and Android." The second theme was "staying logged in." Participants mentioned that "there should be an option to always stay logged in since no one else uses my phone but me. The ability to stay logged in after closing the app for a while will save me the time of having to log in with my details again or entering my phone passcode since my phone is virtually always on me." "Stability" was the third theme, as participants suggested to "make the app more stable and working perfectly to avoid it crashing." The last theme was "removing visual smoking triggers." As participants put it, "the app is perfect as it is, but taking out pictures of cigarettes and someone smoking would help a lot." In addition to the general themes, participants also pointed out that certain features could be improved. For example, for the tracking function, participants could not see all the days they smoked even within that month. Participants also think the "common questions" feature "needs more development and more options to pick from."

Preliminary Effectiveness

We performed a series of logistic regression analyses to examine whether participants with greater app engagement are more likely to self-report quitting at the 6 and 12-week follow-ups (hypothesis 1). Results showed a significant effect of total activities engaged on both 6-week cessation status (Wald χ^2_1=4.41; *P*=.04) and 12-week cessation status (Wald χ^2_1=7.93; *P*=.005). The odds of staying abstinent at week 6 and week 12 increased by 0.3% (odds ratio [OR] 1.003, 95% CI 1.00-1.01) and 0.4% (OR 1.004, 95% CI 1.001-1.01), respectively, for every additional app activity engaged.

The types of activities participants engaged in also significantly predicted the odds of quitting at 6 weeks (Wald χ^2_1=4.19; *P*=.04) and 12 weeks (Wald χ^2_1=7.97; *P*=.005). The odds of staying abstinent at week 6 and week 12 increased by 0.6% (OR 1.006, 95% CI 1.00-1.01) and 0.9% (OR 1.009, 95% CI 1.003-1.02), respectively, for every additional type of activity engaged.

Use frequency, indicated by "login times," did not significantly influence smoking cessation rate at 6 weeks (Wald χ^2_1=1.85; *P*=.17, OR 1.02, 95% CI 0.99 to 1.06) but predicted the odds of quitting at 12 weeks (Wald χ^2_1=5.25; *P*=.02). The odds of maintaining abstinence from smoking at week 12 increased 1.05 times for each additional log-in (OR 1.05, 95% CI 1.01-1.08).

To examine whether earning more points is associated with a higher likelihood of self-reported quitting (hypothesis 2), we recoded participants' points earned into 4 dummy coded categories: >5000 points, >10,000 points, >50,000 points, and >100,000 points. Among them, only ">5000 points" significantly predicted the odds of quitting at 6 weeks (Wald χ^2_1=5.96; *P*=.02) and 12 weeks (Wald χ^2_1=10.41; *P*=.001). Participants who earned >5000 points had higher odds of quitting at 6 weeks (OR 3.12, 95% CI 1.25-7.75) and 12 weeks (OR 4.65, 95% CI 1.83-11.76). Changes in other smoking-related outcomes from baseline, 6 weeks, to 12 weeks in cohort 3 were summarized in Table 4.

Table 4. Smoking-related outcomes in cohort 3 (N=85).

Characteristic	Baseline upon installation, n (%)	6 weeks, n (%)	12 weeks, n (%)
Number of participants completed the survey			
Completed	85 (100)	56 (66)	47 (55)
Missed	0 (0)	29 (34)	38 (45)
Smoking status			
Active smoker	85 (100)	8 (9)	5 (6)
Quit	0 (0)	42 (49)	40 (47)
Not sure	0 (0)	6 (7)	2 (3)
Abstinent from smoking in the past 7 days			
Yes	—[a]	49 (58)	44 (52)
No	—	7 (8)	3 (4)
Smoking in the past 30 days			
None of the days	—	25 (29)	31 (36)
A few days	—	24 (28)	12 (14)
Some days	—	5 (6)	3 (4)
Most days	—	2 (26)	1 (1)
Cigarettes smoked per day (on average)			
0 cigarette/d	0 (0)	48 (56)	42 (49)
1 cigarette/d	14 (16)	2 (26)	1 (1)
2-4 cigarettes/d	28 (33)	4 (5)	4 (5)
5-10 cigarettes/d	24 (28)	2 (2)	0 (0)
11-20 cigarettes/d	11 (13)	0 (0)	0 (0)
21-30 cigarettes/d	8 (9)	0 (0)	0 (0)
FTND score[b] (0-10 scale)			
Very low (0-2)	7 (8)	6 (7)	3 (4)
Low (3-4)	34 (40)	1 (1)	1 (1)
Moderate (5)	17 (20)	1 (1)	1 (1)
High (6-7)	18 (21)	0 (0)	0 (0)
Very high (8-10)	8 (9)	0 (0)	0 (0)

[a]Not available.

[b]FTND: Fagerstrom Test for Nicotine Dependence. This score was only assessed and analyzed among active smokers.

We also performed a series of post hoc analyses to explore the effects of specific app features on cessation outcomes. In the prequit and planning phase, engaging with "why track," "what influences are around me," "quit story," "meducation vaping," or "meducation nicotine addiction" predicted a greater likelihood of quitting at both week 6 and week 12. Completing reasons for "why quit" and "MOtalks triggers" did not affect the 6-week cessation status but was significantly associated with

participants' 12-week cessation status. Completing "about me" was not linked with cessation status at either time point. Detailed results are reported in Table 5.

Table 5. Effects of engaging with specific app features on cessation outcomes.

	Wald chi-square	*P* value	OR[a] (95% CI)
Quit at week 6			
About me	0.19	.67	1.31 (0.39-4.94)
Why quit	2.39	.12	2.05 (0.83-5.07)
Why track	4.35	.04	2.75 (1.06-7.12)
Quit story	4.53	.03	2.71 (1.08-6.76)
What influences are around me	5.05	.03	3.19 (1.16-8.75)
MOtalks triggers	3.39	.07	3.15 (0.93-10.69)
Meducation vaping	3.94	.05	2.93 (1.01-8.47)
Meducation nicotine addiction	5.12	.02	3.63 (1.19-11.12)
Medication tracker	3.57	.06	2.66 (0.96-7.36)
MOtalks medication	0.82	.37	1.72 (0.53-5.58)
MOtalks breathing	2.61	.11	2.42 (0.83-7.07)
MOtalks urges	0.64	.42	1.57 (0.52-4.76)
MOtalks stress	4.09	.04	3.50 (1.04-11.79)
MOtalks tools	0.82	.37	1.72 (0.53-5.58)
Quit at week 12			
About me	3.05	.08	3.10 (0.87-11.06)
Why quit	5.79	.02	3.22 (1.24-8.36)
Why track	6.39	.01	3.33 (1.31-8.48)
Quit story	7.67	.006	3.66 (1.46-9.16)
What influences are around me	6.40	.01	3.50 (1.33-9.24)
MOtalks triggers	6.94	.008	5.20 (1.52-17.74)
Meducation vaping	4.41	.04	2.94 (1.07-8.02)
Meducation nicotine addiction	5.47	.02	3.45 (1.22-9.73)
Medication tracker	8.83	.003	4.76 (1.70-13.31)
MOtalks medication	2.91	.09	2.79 (0.86-9.05)
MOtalks breathing	6.64	.01	4.13 (1.40-12.15)
MOtalks urges	2.80	.09	2.58 (0.85-7.83)
MOtalks stress	8.01	.005	5.83 (1.72-19.78)
MOtalks tools	2.91	.09	2.79 (0.86-9.05)

[a]OR: odds ratio.

In the quit and maintenance phase, we found a marginally significant effect of using the medication tracking feature on the 6-week cessation status (Wald χ^2_1=3.57; *P*=.06; OR 2.66, 95% CI 0.96-7.36), and a significant effect on the 12-week cessation status (Wald χ^2_1=8.83; *P*=.003; OR 4.76, 95% CI 1.70-13.31). Similarly, participants who completed "MOtalks breathing" had a greater chance of quitting at the 12 weeks than those who did not. Engaging "MOtalks medication" was not associated with cessation outcomes.

In the relapse prevention phase, learning from "MOtalks stress" was significantly linked with a higher likelihood of quitting at 6 and 12 weeks. However, engaging with "MOtalks urges" or "MOtalks tools" did not predict cessation status.

Discussion

Overview

In this study, we used a systematic approach to develop and pilot-test the MO mobile phone app for smoking cessation. This process involved iterative engagement with users, incorporating

their feedback to make improvements and updates to the app. We used a combination of qualitative and quantitative methods, including closed and open-ended questions and app use tracking, to gather user feedback and refine the app's features.

In cohort 1, we identified and fixed the issue of frequent crashes and enhanced the point reward system by adding badges. We also tailored the progress reports and reminders based on users' stages of quitting. On the basis of user data analysis in cohort 2, we further refined the MO app by introducing a new "chat" feature, which served as an internal support group to replace a prior support group feature that required participants to exit the app to engage it. Discussions in this chat were moderated by a researcher on an intermittent basis. In addition, we implemented a new reward mechanism where users achieving the intermediate stage (5 stages in total: beginner, fast starter, intermediate, master, and expert) of 20,000 points were entered into a lottery for a US $50 gift card. These updated features aimed to promote engagement and motivation among participants to quit smoking. The performance of the app steadily increased from cohort 1 to 3, with cohort 3 reporting the highest level of acceptability and engagement, as indicated by the duration of active app use and total points earned. The "quit plan" feature was the most frequently used among cohort 3 participants. Participants also frequently interacted with and positively rated the MO talks. The most popular topics of MO talks were "meet your host," "denial stories," and "breathing."

Principal Findings

Results demonstrated that the MO app was promising in assisting smokers to quit. The 7-day point prevalence abstinence rates of the mITT population were 58% (49/85) at week 6 and 52% (44/85) at week 12, while the 30-day point prevalence abstinence rates were 29% (25/85) at week 6 and 36% (31/85) at week 12. These cessation rates were higher than the 7-day point prevalence abstinence rates reported for other smoking cessation apps, which typically ranged from 13% to 35% at the 12-week follow-up [9,16,24]. We also found that participants who engaged more frequently with the app's features, interacted with a greater variety of features, and earned more points were more likely to maintain abstinence from smoking. Participants that did not complete the study questionnaire were imputed to return to smoking. This conservative approach may not capture any benefits, such as decreased cigarette use for a period. Furthermore, no adverse events or safety and privacy concerns were reported during the study period. Despite minor technical issues, such as unexpected crashes, all features of the app functioned as intended, indicating its feasibility for everyday use by smokers.

The significant association between app engagement and cessation outcome found in this pilot study underscores the critical role of an engaging system in driving and sustaining abstinence. Previous research has also shown that improved abstinence rates were linked with increased app engagement, indicated by a greater number of app openings and adherence to app features (ie, completing required tasks) [16,27,28]. In a pilot randomized clinical trial, the number of interactions with app features completely mediated cessation outcomes [29]. Our study contributes to existing literature by identifying the types of activities users engaged in and the points earned through interactions with app features as additional forms of engagement that significantly predict smoking cessation outcomes. We found that users who received >5000 points were more likely to remain abstinence at weeks 6 and 12. However, users who accumulated >10,000 points did not show significantly different cessation outcomes compared with those with fewer points. This suggests that earning 5000 points represents a critical threshold for the app's effectiveness. The finding that additional points beyond this threshold did not enhance cessation outcomes does not necessarily indicate a ceiling effect but rather reflects the complex and recursive nature of smoking cessation. Users who found quitting relatively easy may not need to earn many points and can quit at an early stage. In contrast, users who accumulated >10,000 points might indicate 2 possibilities: on one hand, they were highly motivated to quit and thus engaged more with the app; and by contrast, they faced greater challenges in quitting, requiring prolonged app engagement. While both scenarios result in greater points earned, the former is associated with more favorable quitting outcomes, whereas the latter is linked to continued smoking or an extended period required to quit.

Our data shed light on several strategies for designing engaging health apps. First, smoking cessation apps should provide users with the flexibility to create their own quitting plans, including setting quit dates, identifying reasons to quit, establishing quitting goals, and learning about medication options. These features were the most frequently used in the MO app. Second, the app should enhance users' efficacy in quitting smoking and equip them with skills to cope with relapses. Features that facilitated the tracking of smoking habits, sent reminders, and provided educational materials on strategies for managing urges were rated as the most liked features by participants. These design considerations align with the self-determination theory [30,31], which suggests that behaviors are driven by the internal needs of autonomy (ie, taking control of the quitting process), competence (ie, having necessary skills to achieve the cessation goal), and relatedness (ie, social connections and sense of belonging). Features related to "autonomy" and "competence" were well received by MO users. While MO also addressed the "relatedness" need by establishing a secure peer-support community within the app through chat and story sharing functions, these features were not completely used by participants. This finding is consistent with previous research that showed low levels of engagement and perceived usefulness for smoking cessation apps that relied on social network members [32]. One potential reason for such limited engagement is the challenge participants faced in building trust with other users within a relatively short period. However, the chat and story booth features may play a more important role for users in the later stage of smoking cessation, especially when they experience relapses. In addition to sharing experiences with smoking cessation peers, individuals may also need support and encouragement from family members or close friends. While receiving social support has been positively associated with changes in smoking cessation stages [33], meta-analyses yielded mixed results regarding the effects of interventions that enhance partner and family support on cessation rates [34,35]. Therefore, stronger empirical evidence is needed to determine the

effectiveness of integrating family-based support into the design of smoking cessation apps.

In response to open-ended questions soliciting suggestions for increasing app engagement, participants emphasized the importance of reducing barriers to app use and making it as accessible as possible. Recommendations included streamlining the log-in process and ensuring compatibility with both iPhone and Android devices. Enabling logins with biometrics, such as using fingerprints or facial recognition, would further enhance the convenience and security of the log-in process. Participants were also inspired by the motivational talks and uplifting quotes provided by the app, as these elements helped build confidence and foster positive beliefs about quitting smoking. On the basis of this finding, it might be helpful to further refine the "story booth" function, including providing prompts, instructions, or incentives to encourage former successful quitters who used the app to record their experience, allowing different forms of story sharing (eg, textual, audio, or video stories), and helping current users find quitters sharing similar backgrounds with them so that they are more likely to learn from and become motivated by peers' quitting journey. Furthermore, the point system and financial incentives were found to be useful in increasing participants' motivation to remain engaged with the MO app. The financial incentives may be particularly attractive to and useful in assisting low-SES smokers in quitting. These reward mechanisms could be reinforced in future iterations and adapted to nonstudy users. For example, the point system could be used to unlock advanced features (eg, personalized motivational quotes, an artificial intelligence-powered smoking cessation chatbot). In addition, enabling users to share their achievements on social media or with significant others could further enhance engagement and motivation. While it may not be feasible to provide large-scale monetary rewards to public users, alternative financial incentives could be offered in the form of free web-based consultation with professionals, free nicotine cessation medications, or coupons for healthy products.

Given the proliferation of numerous cessation apps with varying quality, clinical research data are essential to help smokers make informed choices and enhance their chances of success. Although there has been an increase in clinical research on mobile phone-based smoking cessation interventions in recent years, we are still in the early stages of accumulating rigorous clinical research evidence to establish best practices for smoking cessation apps. Health care professionals require credible data to confidently recommend mHealth platforms to their patients. Our findings were based on user experience and cessation outcomes of a good mix of males and females as well as smokers of different ethnicities and races, representing a diverse smoking population. By identifying users' level of app engagement (eg, 5000 points) associated with cessation, the app platform can serve as a tool for health care professionals to encourage patients to reach and surpass this activity level. Moreover, this mobile-based system can be easily integrated into clinical practices, such as short advice or counseling sessions, complementing clinical support for smoking cessation and assisting in tracking patients' adherence to cessation treatment.

Limitations

The study had several limitations. First, being a pilot evaluation study, the sample size was small, and the representativeness of the sample cannot be guaranteed. Therefore, the findings may not be generalizable to the entire population of smokers intending to quit. Although recruitment through web-based advertisements proved to be feasible and efficient in recruiting eligible participants, it is important to note that the participants in our study exhibited high levels of technology literacy and motivation to quit smoking. Participants who joined this research were young, with more than 85.8% (91/106) of participants aged younger than 40 years, and most of them had a smoking history between 1 and 10 years. We also set an eligibility criterion requiring participants to have a high intention to quit (ie, scored at least a 7 on a scale of 0-10 on a self-assessed commitment scale). The unique characteristics of the sample, including the younger age, relatively short smoking history, and strong motivation to quit, could be reasons for the high smoking cessation rate of using the MO app. In addition, the study was conducted during the COVID-19 pandemic, which presented a unique opportunity for smokers to quit due to social distancing and quarantine policies that significantly reduced social and environmental smoking triggers. This health crisis also made staying healthy a more salient goal in people's minds, and therefore, participants may have been more motivated to quit smoking. By contrast, social isolation, anxiety, and depression increased during the pandemic, and self-reported cessation attempts and cessation-related medications declined at the beginning of the pandemic [36]. The unique characteristics of participants, social contexts, and the special period should be considered when interpreting this study's findings.

Second, the follow-up period in this study was relatively short (6 and 12 weeks) for testing the app's engagement and effectiveness. Smoking cessation is a long-term process involving potential setbacks and relapses. In future studies, we plan to follow-up with app users for 6 months and 1 year to examine the apps' long-term cessation effects and the dynamic relationship between app engagement and cessation outcomes at different stages of quitting.

Third, this study assessed cessation outcomes through self-reported data from web-based surveys, without biochemical verification of cotinine levels or exhaled carbon monoxide monitoring. Obtaining biochemically verified cessation status was not feasible during the pandemic period, and the Society for Research on Nicotine and Tobacco Subcommittee on Biochemical Verification has suggested that biochemical confirmation may not be necessary in certain studies [37]. Considering cost and other constraints, self-reporting can be considered an acceptable standard method in cessation trials and clinical practice [38]. However, future studies incorporating biochemically verified data would be valuable to check whether the high cessation rate after using the MO app can be replicated.

Finally, this study did not have time-stamped historical logs of user engagement with each app feature, so we were unable to demonstrate users' dynamic engagement with specific activities over time. To improve the user-centered design and usability of the app, it is critical to understand how users interact with

the app initially, whether they can quickly learn and explore the various functions, and how users' initial, instant responses to the app differ from adaptive responses several weeks later. It is useful for future studies to track users' first-week use and compare it with use at 1 month.

One advantage of the MO app is that it incorporates the medication tracking function to facilitate smokers with medication use and compliance of cessation medications in everyday settings. Due to the small sample size, we were only able to analyze the effect of using the medication tracking function on cessation outcomes and were underpowered to have more detailed descriptions about how they used this function, what kinds of medications they used, and the corresponding effects. Further research should fill this gap by providing empirical evidence on how using different cessation medications and adherence rates, facilitated by smoking cessation apps, are associated with cessation outcomes.

Our next step is to conduct a larger-scale randomized clinical trial that compares MO with other standard smoking cessation apps to further explore its effectiveness on short-term and long-term quitting outcomes. It is also worth investigating whether the app is accessible to and acceptable to low-SES smokers, which is critical to help reduce disparities in smoking cessation. While mHealth apps have enhanced the accessibility of health interventions to the public, it is unclear whether disparities exist in the use of mobile technologies for smoking cessation. Considering that the low-SES population comprises most current smokers, understanding the effects of MO on facilitating smoking cessation among this group is necessary.

Conclusions

MO is a feasible mobile phone app that effectively provides smoking cessation support to individuals who desire to quit. The study demonstrated high acceptability, usability, and potential efficacy of the app in enhancing cessation outcomes. Further research, including a randomized controlled trial, is needed to rigorously evaluate long-term app engagement and effects on smoking cessation.

Acknowledgments

The authors want to thank Dawn Repola for her invaluable contributions and support throughout the app development and research process. We'd also like to acknowledge the insightful comments and contributions of Lisa Kroon, PharmD and Suzanne Harris, RN in the early development phase of the platform.

Conflicts of Interest

GF is an employee and owner, and PB is an owner and co-founder of Mobile Applications for Connected Health (MACH), a digital health LLC. This study was funded by a grant from the Fontana Tobacco Treatment Center at the University of California, San Francisco, and supported by Mobile Applications for Connected Health (MACH) LLC. The funders were not involved in the process of data collection, analyses, or interpretation.

Multimedia Appendix 1
A walkthrough video explaining the main features of the MO smoking cessation app.
[MP4 File (MP4 Video), 19261 KB - mhealth_v12i1e55239_app1.mp4]

References

1. Tobacco. World Health Organization. 2023 Jul 31. URL: https://www.who.int/news-room/fact-sheets/detail/tobacco [accessed 2023-06-20]
2. Burke MV, Ebbert JO, Schroeder DR, McFadden DD, Hays JT. Treatment outcomes from a specialist model for treating tobacco use disorder in a medical center. Medicine (Baltimore) 2015 Nov;94(44):e1903 [FREE Full text] [doi: 10.1097/MD.0000000000001903] [Medline: 26554789]
3. Stead LF, Lancaster T. Behavioural interventions as adjuncts to pharmacotherapy for smoking cessation. Cochrane Database Syst Rev 2012 Dec 12;12:CD009670. [doi: 10.1002/14651858.CD009670.pub2] [Medline: 23235680]
4. Cartujano-Barrera F, Lee D'Abundo M, Arana-Chicas E, Chock S, Valera P, Kamen CS, et al. Barriers and facilitators of smoking cessation among Latinos living with HIV: perspectives from key leaders of community-based organizations and clinics. Int J Environ Res Public Health 2021 Mar 26;18(7):3437 [FREE Full text] [doi: 10.3390/ijerph18073437] [Medline: 33810269]
5. Rodgers-Melnick SN, Zanotti K, Lee RT, Webb Hooper M. Demographic and clinical predictors of engaging in tobacco cessation counseling at a comprehensive cancer center. JCO Oncol Pract 2022 May;18(5):e721-e730. [doi: 10.1200/OP.21.00458] [Medline: 34995104]
6. Sheffer CE, Brackman SL, Cottoms N, Olsen M. Understanding the barriers to use of free, proactive telephone counseling for tobacco dependence. Qual Health Res 2011 Aug;21(8):1075-1085 [FREE Full text] [doi: 10.1177/1049732311404248] [Medline: 21464470]

7. Zhou S, Li Y, Levinson AH. A cross-sectional national survey to explore the relationship between smoking and political abstention: evidence of social mistrust as a mediator. SSM Popul Health 2021 Jul 01;15:100856 [FREE Full text] [doi: 10.1016/j.ssmph.2021.100856] [Medline: 34277923]
8. Mobile fact sheet. Pew Research Center. 2024 Jan 31. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2024-01-31]
9. Marler JD, Fujii CA, Utley MT, Balbierz DJ, Galanko JA, Utley DS. Outcomes of a comprehensive mobile smoking cessation program with nicotine replacement therapy in adult smokers: pilot randomized controlled trial. JMIR Mhealth Uhealth 2022 Nov 24;10(11):e41658 [FREE Full text] [doi: 10.2196/41658] [Medline: 36257323]
10. Bricker JB, Watson NL, Mull KE, Sullivan BM, Heffner JL. Efficacy of smartphone applications for smoking cessation: a randomized clinical trial. JAMA Intern Med 2020 Nov 01;180(11):1472-1480 [FREE Full text] [doi: 10.1001/jamainternmed.2020.4055] [Medline: 32955554]
11. Iacoviello BM, Steinerman JR, Klein DB, Silver TL, Berger AG, Luo SX, et al. Clickotine, a personalized smartphone app for smoking cessation: initial evaluation. JMIR Mhealth Uhealth 2017 Apr 25;5(4):e56 [FREE Full text] [doi: 10.2196/mhealth.7226] [Medline: 28442453]
12. Webb J, Peerbux S, Smittenaar P, Siddiqui S, Sherwani Y, Ahmed M, et al. Preliminary outcomes of a digital therapeutic intervention for smoking cessation in adult smokers: randomized controlled trial. JMIR Ment Health 2020 Oct 06;7(10):e22833 [FREE Full text] [doi: 10.2196/22833] [Medline: 33021488]
13. Huff JW, Green K. Demonstration of active sodium transport across the isolated rabbit corneal endothelium. Curr Eye Res 1981;1(2):113-114. [doi: 10.3109/02713688109001735] [Medline: 7297096]
14. Masaki K, Tateno H, Nomura A, Muto T, Suzuki S, Satake K, et al. A randomized controlled trial of a smoking cessation smartphone application with a carbon monoxide checker. NPJ Digit Med 2020 Mar 12;3:35 [FREE Full text] [doi: 10.1038/s41746-020-0243-5] [Medline: 32195370]
15. Kato A, Tanigawa T, Satake K, Nomura A. Efficacy of the Ascure smoking cessation program: retrospective study. JMIR Mhealth Uhealth 2020 May 14;8(5):e17270 [FREE Full text] [doi: 10.2196/17270] [Medline: 32406856]
16. Regmi K, Kassim N, Ahmad N, Tuah NA. Effectiveness of mobile apps for smoking cessation: a review. Tob Prev Cessat 2017 Apr 12;3:12 [FREE Full text] [doi: 10.18332/tpc/70088] [Medline: 32432186]
17. Guo YQ, Chen Y, Dabbs AD, Wu Y. The effectiveness of smartphone app-based interventions for assisting smoking cessation: systematic review and meta-analysis. J Med Internet Res 2023 Apr 20;25:e43242 [FREE Full text] [doi: 10.2196/43242] [Medline: 37079352]
18. Rajani NB, Weth D, Mastellos N, Filippidis FT. Adherence of popular smoking cessation mobile applications to evidence-based guidelines. BMC Public Health 2019 Jun 13;19(1):743 [FREE Full text] [doi: 10.1186/s12889-019-7084-7] [Medline: 31196062]
19. Abroms LC, Lee Westmaas J, Bontemps-Jones J, Ramani R, Mellerson J. A content analysis of popular smartphone apps for smoking cessation. Am J Prev Med 2013 Dec;45(6):732-736 [FREE Full text] [doi: 10.1016/j.amepre.2013.07.008] [Medline: 24237915]
20. Hoeppner BB, Hoeppner SS, Seaboyer L, Schick MR, Wu GW, Bergman BG, et al. How smart are smartphone apps for smoking cessation? A content analysis. Nicotine Tob Res 2016 May;18(5):1025-1031 [FREE Full text] [doi: 10.1093/ntr/ntv117] [Medline: 26045249]
21. Haskins BL, Lesperance D, Gibbons P, Boudreaux ED. A systematic review of smartphone applications for smoking cessation. Transl Behav Med 2017 Jun;7(2):292-299 [FREE Full text] [doi: 10.1007/s13142-017-0492-2] [Medline: 28527027]
22. Notley C, Gentry S, Livingstone-Banks J, Bauld L, Perera R, Hartmann-Boyce J. Incentives for smoking cessation. Cochrane Database Syst Rev 2019 Jul 17;7(7):CD004307 [FREE Full text] [doi: 10.1002/14651858.CD004307.pub6] [Medline: 31313293]
23. Bock BC, Heron KE, Jennings EG, Magee JC, Morrow KM. User preferences for a text message-based smoking cessation intervention. Health Educ Behav 2013 Apr;40(2):152-159. [doi: 10.1177/1090198112463020] [Medline: 23086555]
24. Taylor GM, Dalili MN, Semwal M, Civljak M, Sheikh A, Car J. Internet-based interventions for smoking cessation. Cochrane Database Syst Rev 2017 Sep 04;9(9):CD007078 [FREE Full text] [doi: 10.1002/14651858.CD007078.pub5] [Medline: 28869775]
25. Hébert ET, Ra CK, Alexander AC, Helt A, Moisiuc R, Kendzor DE, et al. A mobile just-in-time adaptive intervention for smoking cessation: pilot randomized controlled trial. J Med Internet Res 2020 Mar 09;22(3):e16907 [FREE Full text] [doi: 10.2196/16907] [Medline: 32149716]
26. Cheung KL, de Ruijter D, Hiligsmann M, Elfeddali I, Hoving C, Evers SM, et al. Exploring consensus on how to measure smoking cessation. A Delphi study. BMC Public Health 2017 Nov 21;17(1):890 [FREE Full text] [doi: 10.1186/s12889-017-4902-7] [Medline: 29162043]
27. Ubhi HK, Michie S, Kotz D, Wong WC, West R. A mobile app to aid smoking cessation: preliminary evaluation of SmokeFree28. J Med Internet Res 2015 Jan 16;17(1):e17 [FREE Full text] [doi: 10.2196/jmir.3479] [Medline: 25596170]

28. Zeng EY, Heffner JL, Copeland WK, Mull KE, Bricker JB. Get with the program: adherence to a smartphone app for smoking cessation. Addict Behav 2016 Dec;63:120-124 [FREE Full text] [doi: 10.1016/j.addbeh.2016.07.007] [Medline: 27454354]
29. Browne J, Halverson TF, Vilardaga R. Engagement with a digital therapeutic for smoking cessation designed for persons with psychiatric illness fully mediates smoking outcomes in a pilot randomized controlled trial. Transl Behav Med 2021 Sep 15;11(9):1717-1725 [FREE Full text] [doi: 10.1093/tbm/ibab100] [Medline: 34347865]
30. Deci EL, Ryan RM. Self-determination theory. In: Van Lange PA, Kruglanski AW, Higgins ET, editors. Handbook of Theories of Social Psychology. Thousand Oaks, CA: SAGE Publications; 2012:416-436.
31. Deci EL, Olafsen AH, Ryan RM. Self-determination theory in work organizations: the state of a science. Annu Rev Organ Psychol Organ Behav 2017 Mar;4:19-43. [doi: 10.1146/annurev-orgpsych-032516-113108]
32. Schwaninger P, Berli C, Scholz U, Lüscher J. Effectiveness of a dyadic buddy app for smoking cessation: randomized controlled trial. J Med Internet Res 2021 Sep 09;23(9):e27162 [FREE Full text] [doi: 10.2196/27162] [Medline: 34499045]
33. Wagner J, Burg M, Sirois B. Social support and the transtheoretical model: relationship of social support to smoking cessation stage, decisional balance, process use, and temptation. Addict Behav 2004 Jul;29(5):1039-1043. [doi: 10.1016/j.addbeh.2004.02.058] [Medline: 15219355]
34. Hubbard G, Gorely T, Ozakinci G, Polson R, Forbat L. A systematic review and narrative summary of family-based smoking cessation interventions to help adults quit smoking. BMC Fam Pract 2016 Jun 24;17:73 [FREE Full text] [doi: 10.1186/s12875-016-0457-4] [Medline: 27342987]
35. Park EW, Tudiver F, Schultz JK, Campbell T. Does enhancing partner support and interaction improve smoking cessation? A meta-analysis. Ann Fam Med 2004;2(2):170-174 [FREE Full text] [doi: 10.1370/afm.64] [Medline: 15083859]
36. Bandi P, Asare S, Majmundar A, Xue Z, Han X, Westmaas JL, et al. Changes in smoking cessation-related behaviors among US adults during the COVID-19 pandemic. JAMA Netw Open 2022 Aug 01;5(8):e2225149 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.25149] [Medline: 35913738]
37. Benowitz NL, Bernert JT, Foulds J, Hecht SS, Jacob P, Jarvis MJ, et al. Biochemical verification of tobacco use and abstinence: 2019 update. Nicotine Tob Res 2020 Jun 12;22(7):1086-1097 [FREE Full text] [doi: 10.1093/ntr/ntz132] [Medline: 31570931]
38. Hutton HE, Wilson LM, Apelberg BJ, Tang EA, Odelola O, Bass EB, et al. A systematic review of randomized controlled trials: web-based interventions for smoking cessation among adolescents, college students, and adults. Nicotine Tob Res 2011 Apr;13(4):227-238. [doi: 10.1093/ntr/ntq252] [Medline: 21350042]

Abbreviations

ITT: intent-to-treat
mHealth: mobile health
mITT: modified intent-to-treat
OR: odds ratio
RQ: research question
SES: socioeconomic status

Edited by L Buis; submitted 06.12.23; peer-reviewed by H Tateno, R Marshall; comments to author 26.03.24; revised version received 31.05.24; accepted 26.08.24; published 29.10.24.

Please cite as:
Zhou S, Brunetta P, Silvasstar J, Feldman G, Oromi N, Bull S
Initial Evaluation of Acceptability, Engagement, and Effectiveness of the MO App to Provide Tailored and Comprehensive Support for Smoking Cessation: Development and Usability Study
JMIR Mhealth Uhealth 2024;12:e55239
URL: https://mhealth.jmir.org/2024/1/e55239
doi:10.2196/55239
PMID:

Original Paper

Efficacy of the mHealth App Intellect in Improving Subclinical Obsessive-Compulsive Disorder in University Students: Randomized Controlled Trial With a 4-Week Follow-Up

Madeline Lee Yoon Li[1], BSocSci; Stephanie Lee Si Min[1], PhD; Oliver Sündermann[1], PhD

National University of Singapore, Singapore, Singapore

Corresponding Author:
Oliver Sündermann, PhD
National University of Singapore
9 Arts Link
#13-02
Singapore, 117570
Singapore
Phone: 65 80435120 ext 5120
Email: psymos@nus.edu.sg

Abstract

Background: Obsessive-compulsive disorder (OCD) is the third most prevalent mental health disorder in Singapore, with a high degree of burden and large treatment gaps. Self-guided programs on mobile apps are accessible and affordable interventions, with the potential to address subclinical OCD before symptoms escalate.

Objective: This randomized controlled trial aimed to examine the efficacy of a self-guided OCD program on the mobile health (mHealth) app Intellect in improving subclinical OCD and maladaptive perfectionism (MP) as a potential moderator of this predicted relationship.

Methods: University students (N=225) were randomly assigned to an 8-day, self-guided app program on OCD (intervention group) or cooperation (active control). Self-reported measures were obtained at baseline, after the program, and at a 4-week follow-up. The primary outcome measure was OCD symptom severity (Obsessive Compulsive Inventory–Revised [OCI-R]). Baseline MP was assessed as a potential moderator. Depression, anxiety, and stress (Depression Anxiety and Stress Scales-21) were controlled for during statistical analyses.

Results: The final sample included 192 participants. The intervention group reported significantly lower OCI-R scores compared with the active control group after the intervention (partial eta-squared [η_p^2]=0.031; *P*=.02) and at 4-week follow-up (η_p^2=0.021; *P*=.044). A significant, weak positive correlation was found between MP and OCI-R levels at baseline (*r*=0.28; *P*<.001). MP was not found to moderate the relationship between condition and OCI-R scores at postintervention (*P*=.70) and at 4-week follow-up (*P*=.88).

Conclusions: This study provides evidence that the self-guided OCD program on the Intellect app is effective in reducing subclinical OCD among university students in Singapore. Future studies should include longer follow-up durations and study MP as a moderator in a broader spectrum of OCD symptom severity.

Trial Registration: ClinicalTrials.gov NCT06202677; https://clinicaltrials.gov/study/NCT06202677

(JMIR Mhealth Uhealth 2024;12:e63316) doi:10.2196/63316

KEYWORDS

mobile health app; self-guided interventions; obsessive-compulsive disorder; cognitive behavioral therapy; maladaptive perfectionism; randomized controlled trial; behavioral; efficacy; mHealth app; university students; Singapore; symptoms

Introduction

Background

Obsessive-compulsive disorder (OCD) is one of the most common mental health disorders, with an estimated lifetime prevalence of 3.6% in Singapore [1] and from 2% to 3% globally [2]. OCD is characterized by the presence of recurrent unwanted thoughts, urges, or images (obsessions, compulsions, or both). Common contents of obsessions include preventing harm from mistakes or contamination; concern over order and incompleteness; and intrusive violent, religious, or sexual thoughts [3]. Typically, obsessions are ego-dystonic and inconsistent with one's fundamental beliefs and personality. Hence, obsessions evoke significant anxiety or distress, which may drive the individual to carry out repetitive and excessive behaviors or mental acts (compulsions) that aim to neutralize the discomfort or avert the threat of a feared event.

OCD is associated with serious disability and decreased quality of life [4], leading to significant impairment in areas of family relations, social interactions, education, and work [5]. However, OCD-related impairment is not restricted to individuals who meet the full diagnostic criteria. OCD symptoms exist on a spectrum that varies in frequency and intensity across clinical and nonclinical populations [6]. Compared with the general population, individuals with subclinical OCD symptoms reported significantly lower life satisfaction and quality of life, and greater health care use [7]. In nonclinical populations, prevalence rates of intrusive thoughts have ranged from 74% to 99.4%, with individuals also found to carry out compulsive behaviors, such as checking, washing, or thought suppression [8,9]. Hence, recent research has emphasized early detection and intervention for individuals at risk of developing OCD [10].

The cognitive-behavioral model of OCD posits that obsessional distress stems from an individual's appraisal of normal intrusive thoughts, images, and urges as extremely significant or threatening. Additionally, the individual appraises that the occurrence of the intrusion indicates their potential responsibility for causing harm [11]. These appraisals trigger distress and anxiety, which culminate in the decision to perform compulsions. Consequently, the short-term relief following compulsions negatively reinforces the behavior, leading to an increased behavioral tendency for compulsions over time [12]. As clinical obsessions develop from intrusive thoughts that most people experience [13], OCD symptoms in nonclinical samples are often investigated to uncover findings that can be extended to clinical populations.

The first-line psychotherapeutic treatment for OCD is cognitive behavioral therapy (CBT) with exposure and response prevention (ERP) [14]. ERP involves gradually confronting obsessions while refraining from compulsions, breaking the cycle of compulsions by teaching individuals to tolerate distress without engaging in counterproductive behaviors. ERP facilitates inhibitory learning, helping individuals form nonthreat associations that inhibit threat responses [15]. Extensive research has supported the superior efficacy of CBT in reducing OCD symptoms compared with placebo, relaxation therapy, and anxiety management [16,17]. The efficacy of CBT has been similarly demonstrated across countries, OCD severity, and treatment settings [18]. However, up to 50% of individuals drop out of treatment prematurely and approximately 25% refuse treatment [19]. Barriers to CBT include cost, limited availability of qualified practitioners, societal stigma, and difficulties attending sessions [20,21]. Hence, only a minority of individuals with OCD access evidence-based treatment [22].

Computerized internet-based CBT (ICBT) can reduce these barriers and overcome the limitations of face-to-face psychological interventions. Digital interventions are accessible from any location with an internet connection, offer a broad reach, are cost-effective, and are available 24/7 [23]. ICBT has demonstrated comparable effectiveness to face-to-face CBT for OCD [24], and self-guided ICBT has been suggested as an alternative to remote clinical support [25]. Self-guided ICBT has proven superior in reducing OCD symptom severity levels compared with waitlist and active control conditions [23,26]. Thus, self-guided digital interventions present a promising option for addressing the treatment gap and facilitating the delivery of evidence-based care.

While mobile health (mHealth) apps have gained traction in mental interventions for anxiety, stress, and depression [27], few studies are available on the efficacy of self-guided, mHealth app–based OCD programs that use CBT with ERP. Two randomized controlled trials (RCTs) examined the effectiveness of GG Relationships, Doubt, and Obsessions. However, as GG Relationships, Doubt, and Obsessions was designed to target symptoms of relationship OCD, the study samples were more selective, requiring all participants to have romantic relationship experience [28,29]. An RCT and an open pilot trial investigated the effectiveness of the apps OCfree and nOCD for OCD symptoms and found statistically significant reductions in OCD symptom severity at postintervention [30,31]. However, these apps were used alongside in-person therapy. Hence, it remains unclear whether self-guided mHealth apps for OCD are effective without therapist support. Moreover, previous studies focused on clinical OCD samples, potentially limiting generalizability to those with mild or subclinical symptoms who might benefit from low-intensity, self-guided treatments [32]. The authors also highlighted limitations, such as small sample sizes (N≤50), the lack of a control group not undergoing CBT, and the necessity for replication in future RCTs. Other limitations include the absence of follow-up assessments in 2 out of 4 studies and only 1 Asian sample being studied [29,31]. Notable gaps remain regarding the effectiveness of self-guided mHealth interventions for subclinical OCD populations.

Moreover, variables affecting the efficacy of self-guided mHealth app programs for OCD are underexplored. Only Gamoran and Doron [33] examined moderators in a sample of 46,955 users of OCD.app. Initial OCD severity, age, gender, trait mood, and app use data were identified as predictors of change. However, no psychological variables were analyzed as potential outcome moderators, leaving a gap in OCD treatment research [34]. A potential moderator of OCD outcomes from mHealth app programs is maladaptive perfectionism (MP). Perfectionism refers to one's tendency to demand and pursue unrealistically high standards of performance, accompanied by excessively critical self-evaluations [35]. Frost et al [36]

proposed a 6D measure of perfectionism, which involves being overly concerned about making mistakes (CM), setting and pursuing high personal standards, perceiving high parental expectations (PE) and parental criticism (PC), doubting the quality of one's actions (DA), and preferring order and organization. As a multidimensional construct, perfectionism comprises adaptive and maladaptive aspects. While perfectionistic individuals seek excellence which may promote well-being [37], they may experience dysfunction when excessively fixating on mistakes and meeting external standards. MP is commonly observed in individuals with OCD [38]. It is a risk factor for OCD development and has been viewed as an essential predisposing trait [39]. Additionally, MP has been shown to impede OCD treatment [40]. As moderators can help to identify individuals who may face a poor prognosis and assist in matching individuals with appropriate treatments, it would be useful to examine how baseline MP levels can affect intervention outcomes for self-guided CBT-based mHealth app programs for OCD.

This Study

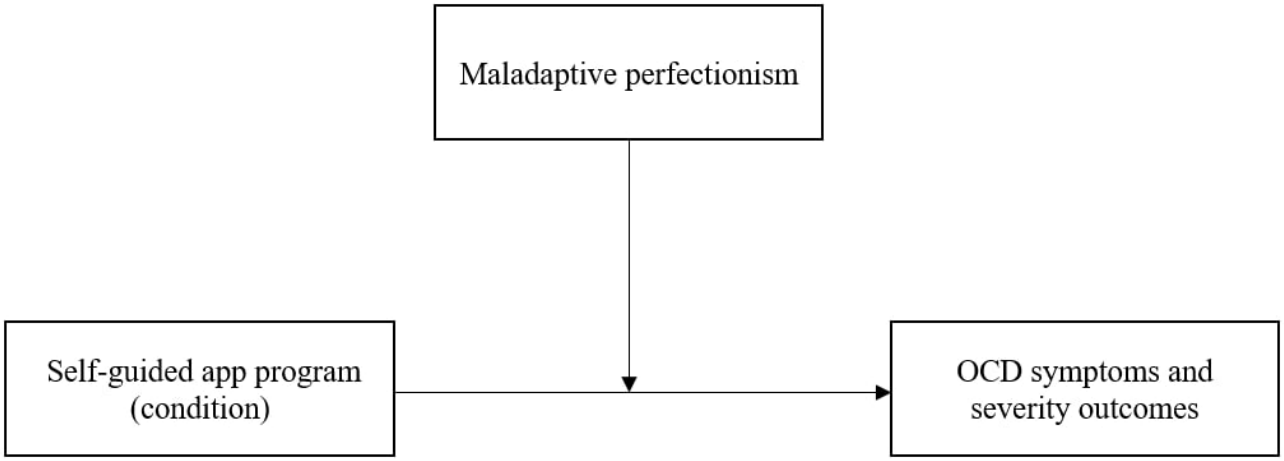
Using an RCT design, we evaluated the efficacy of a self-guided program on an mHealth app in reducing subclinical OCD symptoms and severity among university students from Singapore. Additionally, we examined the association between MP and OCD symptom severity and investigated MP as a moderator of intervention outcomes. Although the mean age of onset for OCD is around 20 years, intervening in university students with subclinical symptoms can still prevent further progression. Early intervention, even at this stage, is crucial for reducing the risk of symptom escalation and long-term impairment.

Given that OCD ranks among the top 3 most prevalent mental health conditions in Singapore, with a lifetime prevalence that is approximately 0.4% to 1.6% higher than the global average, it is crucial to identify accessible and effective OCD interventions. In particular, Singaporeans aged 18 to 34 years have the highest lifetime prevalence rates of OCD, at 6.7% [1]. Yet, OCD remains poorly recognized and has the second-largest treatment gap out of at least 7 mental disorders studied in Singapore [41]. Given the typical onset of OCD in adolescence and early adulthood and its likely chronic course without appropriate intervention [42], it is critical for this population to access evidence-based interventions before symptoms escalate [1]. Young adults have responded positively to mHealth apps, citing their appeal due to flexibility, interactivity, and accessibility [43]. Furthermore, as a population with high smartphone use and common obstacles to accessing traditional mental health services, mHealth apps offer considerable promise to engage Singapore's young adults [44,45].

Given Singapore's cultural context, MP is also a relevant variable to examine. Extensive research has indicated that Asian populations generally exhibit significantly higher levels of perfectionism compared with African and White populations [46,47], which may be due to cross-cultural differences in parenting. Asian parents were found to set higher expectations for their children than European American parents [48] and provide conditional love based on their children's achievements [49]. Additionally, Asian parents typically adopt more controlling and directive parenting styles compared with European parents [50], which is associated with the development of MP [51]. Thus, individuals who grew up in such demanding environments may develop perfectionism as they internalize these standards [52]. In a sample of 302 Singaporean children, a large proportion was found to exhibit high levels of MP [53]. This finding may be attributed to parental intrusiveness and the general emphasis on academic achievement in Singapore, which have been identified as factors that predict increased perfectionistic strivings and concerns [54,55]. Hence, as findings on OCD and MP may be culturally influenced, these variables are pertinent to study in Singapore for the development of effective interventions.

The hypotheses of the study are in 3 parts. First, we expected participants in the OCD program to report significant reductions in OCD outcomes compared with the active control group at postintervention and 4-week follow-up. Second, we hypothesized that baseline OCD symptom severity and MP will be positively correlated. Third, we predicted that participants with the lowest levels of MP would experience the greatest reduction in OCD symptoms and severity at postintervention and 4-week follow-up (Figure 1).

Figure 1. Predicted relationship with maladaptive perfectionism as a moderator of the direct effect of the self-guided app program on OCD symptom and severity outcomes. OCD: obsessive-compulsive disorder.



Methods

Design

This RCT used a 2×3 mixed factorial design with condition (intervention vs active control) as the between-subjects factor and time of assessment (baseline vs postintervention vs 4-week follow-up) as the within-subjects factor. The intervention group participated in an 8-day, self-guided OCD program, and the active control group participated in an 8-day, self-guided cooperation program. The dependent variable measured was OCD symptom severity. Additionally, general psychological distress was measured to be controlled for during analyses, and MP was measured for correlation and moderation analyses.

Ethical Considerations

The study's ethics was approved by the National University of Singapore's (NUS) Institutional Review Board (NUS-IRB-2023-444). The study was also preregistered with ClinicalTrials.gov (registration NCT06202677). All study procedures adhered to the ethical standards outlined by the Declaration of Helsinki (as revised in 2000). Participants were provided with a detailed Participant Information Sheet, which explained the study's objectives, procedures, potential risks, and benefits. Written informed consent was obtained from all participants prior to their enrollment in the study, and they were explicitly informed of their right to withdraw at any point without penalty.

Power Calculation

A meta-analysis examining low-intensity, technology-delivered CBT programs for OCD found small effect size reductions in OCD symptoms posttreatment [56]. Using Cohen *f*=0.10, α=0.05, and power=0.80, G*Power 3.1 revealed a minimum number of 164 participants. Attrition rates in similar mHealth app–based studies conducted among NUS students have ranged from 1% to 44.9% [57-59]. Additionally, accounting for 10% of invalid data typically encountered in web-based research studies [60], this study aimed to recruit a minimum of 206 participants.

Recruitment and Study Participants

The sample of 225 participants had a mean age of 21.68 (SD 1.71) years and consisted of 171 female individuals, 52 male individuals, 1 nonbinary individual, and 1 individual of other gender. Recruitment was carried out by word of mouth and on the university recruitment platforms. Participants were reimbursed with either course credits or monetary cash (see the *Procedure* section below). The inclusion criteria required participants to be a Singapore citizen or permanent resident, be able to read and understand English, be a student from the NUS between the ages of 18 and 30 years, present with mild or subclinical OCD symptoms, and own a mobile phone to download the Intellect app for use in the study. Participants were also required to refrain from using external mental health services during their participation in the study. Individuals with moderate to high OCD symptom severity, as determined by their OCI-R (Obsessive Compulsive Inventory–Revised) scores, were excluded and directed to appropriate resources for the necessary treatment

Materials

Intellect is an mHealth app that can be downloaded from the Apple App Store and the Google Play Store. It hosts numerous self-guided programs that have been reviewed by clinical psychologists and cognitive behavioral therapists. The app adheres to the highest standards of data security and privacy, using zero-knowledge encryption to ensure user data confidentiality. Various self-guided programs on the app have been validated in previous RCTs using samples of Singaporean university students: "Stress," "Improving your Body Image," and "Anxiety & Worry" [57-59]. In this study, participants used Intellect to access either the "Obsessive Compulsive Disorder" or "Cooperation" learning program, depending on their assigned condition.

Learning Programs

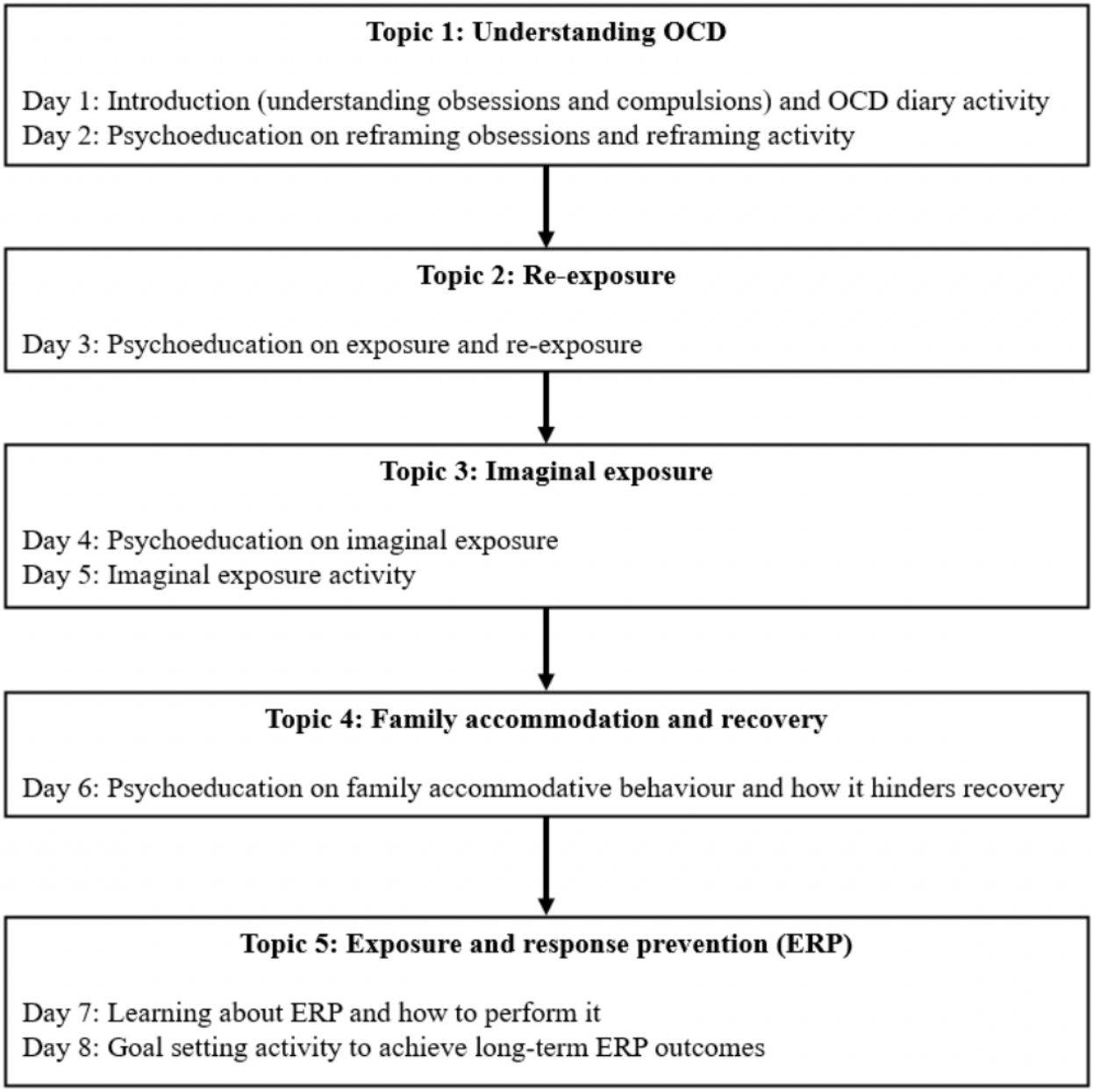
OCD Program (Intervention Condition)

This 8-day program used principles of CBT with ERP for OCD. Each daily session took approximately 5 minutes to complete. The program commenced with psychoeducation on obsessions and compulsions, and how to identify them. Subsequently, through a series of content learning and daily exercises that increase in difficulty, participants were guided to conduct exposure exercises while tolerating distress and refraining from compulsions. For example, in Topic 2 (Re-exposure), participants were asked to face their feared stimuli (eg, touching a "contaminated" surface) and refrain from performing compulsive behaviors (eg, excessive hand-washing) while observing and tolerating the discomfort that arose. This program structure aimed to enable participants to actively practice using the skills that were taught, while not overwhelming them. During the final program session, participants were guided to set goals for continual progress. Due to the challenging nature of ERP, the program also provided advice on concerns that are frequently encountered during ERP. An overview and timeline of the program are presented in Table 1 and Figure 2.

Table 1. Overview of obsessive-compulsive disorder (OCD) program.

Topics	Content
Topic 1: Understanding OCD	• Identify personal obsessions and compulsions • Identify when/where obsessions and compulsions arise • Understand how the importance placed on obsessions causes distress • Understand that obsessions are not based on reality • Develop skills to reframe obsessions
Topic 2: Re-exposure	• Introduce the concept of re-exposure • Understand the function of exposure and re-exposure exercises • Develop skills to carry out re-exposure • Develop skills to tolerate obsessional distress
Topic 3: Imaginal exposure	• Introduce the concept of imaginal exposure • Develop skills to carry out imaginal exposure
Topic 4: Family accommodation and recovery	• Introduce the concept of family accommodation • Understand how family accommodative behaviors can hinder recovery • Develop skills to identify family accommodative behaviors
Topic 5: Exposure and response prevention (ERP)	• Introduce the concept of ERP • Understand the importance of prioritizing long-term OCD reduction goals over short-term discomfort • Develop skills to confront obsessions and resist compulsions • Develop goal-setting skills for long-term ERP progress

Figure 2. Timeline of OCD program. ERP: exposure and response prevention; OCD: obsessive-compulsive disorder.



Cooperation Program (Active Control Condition)

Participants in the active control group were allocated to an 8-day, self-guided program on cooperation, which was housed within the same Intellect app but offered a different module unrelated to OCD content. Having an active control group ensured that participants in both groups shared as many experiences as possible. This controlled for nonspecific factors and attentional effects, which could impact the study's results [61]. The program aimed to improve participants' collaborative skills and interpersonal wellness through a series of content learning and practice exercises, including activities reflective exercises on communication skills. To ensure that participants in both conditions spent a comparable amount of time and effort, the approximate duration for completing each daily session was matched to be 5 minutes.

Procedure

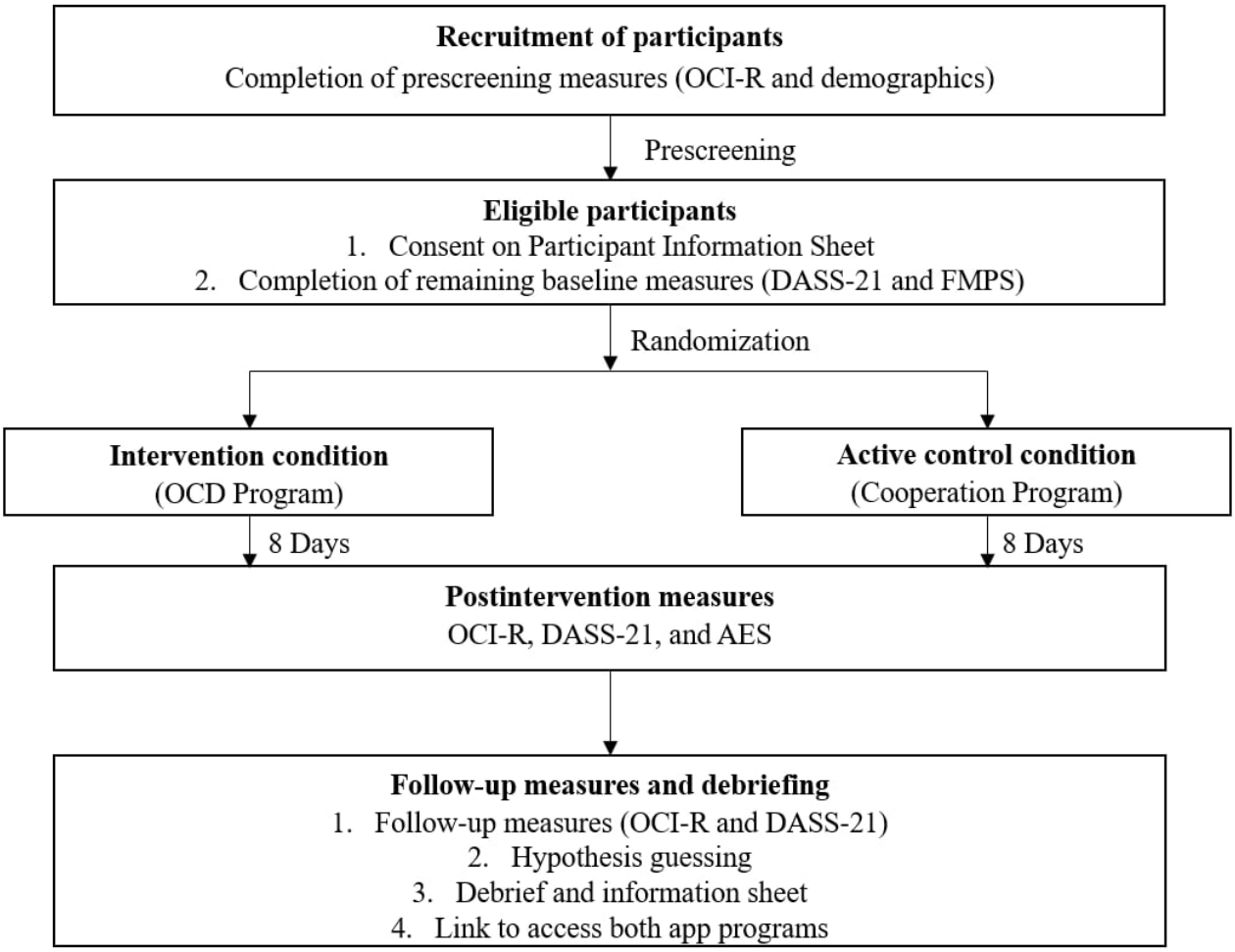
Interested individuals signed up for the study by responding to the web-based advertisement via a link to the web-based survey hosted on Qualtrics. First, they completed a prescreening test comprising measures on OCD and demographics and were screened for eligibility according to the inclusion criteria. Eligible participants read the Participation Information Sheet, and after providing informed consent, completed measures on psychological distress, and perfectionism.

Thereafter, participants were randomly allocated to the intervention or active control condition, using computer-generated random numbers in a 1:1 ratio. To minimize potential confounding effects, participants were instructed not to use any other mHealth apps for the study duration. Additionally, to reduce demand characteristics, participants were not informed of the conditions' functions or the study's objective to evaluate the efficacy of the OCD program. Instead,

participants were informed that the study examined well-being and the learning effectiveness of mental health–related topics using a smartphone app. Participants were then instructed to download the Intellect app, create an account, and access the program of their allocated condition. Participants in the intervention group underwent 8 days of the OCD program, while participants in the active control group underwent 8 days of the cooperation program. Participation in each program took approximately 5 minutes per day. To encourage study adherence, participants were sent a daily reminder to complete app activities via WhatsApp Messenger or Telegram Messenger. To verify program completion, participants sent a screenshot of their learning logs to the investigator.

After verification, participants were sent a postintervention survey link to complete the measures on OCD, psychological distress, and the App Engagement Scale (AES). Four weeks after program completion, participants were sent a follow-up survey link to complete measures on OCD and psychological distress. To further control for demand effects and confounders, participants were asked to guess the study's hypothesis and indicate whether they had undergone mental health treatment during the past 4 weeks. Thereafter, they were debriefed on the study's objectives and hypotheses. Participants who participated in the cooperation program were given access to the OCD program and vice versa. Participants were reimbursed with either SGD 10 (US $7.43) or 2 course credits after completion of the postintervention survey, and an additional SGD 2.50 (US $1.90) or 0.5 course credits after the follow-up survey. Figure 3 presents the study procedure.

Figure 3. Study flowchart. AES: App Engagement Scale; DASS-21: Depression Anxiety Stress Scale-21; FMPS: Frost Multidimensional Perfectionism Scale; OCD: obsessive-compulsive disorder; OCI-R: Obsessive Compulsive Inventory–Revised.



Measures

OCI-R Questionnaire

OCI-R [62] is an 18-item questionnaire that assesses OCD symptoms across 6 subscales: washing, checking, neutralizing, obsessing, ordering, and hoarding. Items are rated on a 5-point Likert scale (0=not at all; 4=extremely). OCD symptom severity can be derived by summing all the items to obtain a total score. Total OCI-R scores correspond to mild (OCI-R=0-15), moderate (OCI-R=16-27), and severe (OCI-R=28-72) OCD symptom severity [63]. In this study, the OCI-R was used as a measure of OCD symptom severity at baseline, postintervention, and 4-week follow-up. Participants with scores of 21 and above were screened out due to a likely presence of clinically significant OCD [62]. Subclinical OCD refers to above-average levels of symptoms that do not meet the clinical threshold but can still cause distress and functional impairment. Reducing even these subclinical symptoms is important to prevent potential escalation to more severe forms of OCD. The OCI-R

showed acceptable internal consistency, with Cronbach α of 0.68.

Frost Multidimensional Perfectionism Scale

The Frost Multidimensional Perfectionism Scale (FMPS) [36] is a 35-item scale that assesses perfectionism across 6 subscales: CM, personal standards, PE, PC, DA, and preferring order and organization. Items are rated on a 5-point Likert scale (1=Strongly disagree; 5=Strongly agree). The FMPS has demonstrated good internal consistency and convergent validity with other measures of perfectionism within nonclinical and clinical samples [36]. In this study, the MP dimension (FMPS–Maladaptive Perfectionism [FMPS-MP]) was obtained by summing scores on the subscales of CM, DA, PE, and PC [64]. Perfectionism was assessed as a moderator of OCD symptom outcomes but was not directly addressed in the intervention content. The FMPS and FMPS-MP showed excellent internal consistency with Cronbach α of 0.91 and 0.92, respectively.

Depression Anxiety Stress Scale-21

The Depression Anxiety Stress Scale-21 (DASS-21) [65] is a set of 3 scales assessing depression, anxiety, and stress, and it is commonly used in student samples [66]. Each subscale contains 7 items, which are rated on a 4-point Likert scale (0=Did not apply to me at all; 3=Applied to me very much or most of the time). In this study, the DASS-21 total score was computed by summing the scores of all items and used as a measure of general psychological distress [65]. Additionally, as depression, stress, and anxiety have been identified in the pathogenesis and maintenance of OCD [67-69], baseline total DASS-21 scores were controlled for during analyses to disentangle the effect of the intervention on OCD symptoms from depression, stress, and anxiety. Internal consistency of the DASS-21 was excellent, with Cronbach α of 0.91.

AES Instrument

The AES [70] is a 7-item scale that assesses app engagement. Each item is rated on a 5-point Likert scale (1=definitely disagree; 5=definitely agree). The total score is derived by summing all the items. Internal consistency of the AES is good, with Cronbach α of 0.88.

Analytic Approach

Incomplete responses and participants who indicated that they received psychotherapy during the study's duration were excluded. Invalid data due to careless or insufficient effort responding were then sequentially identified in 2 steps and removed using a multiple hurdles approach [60]. First, response times of less than 500 seconds on the baseline survey, less than 215 seconds on the postintervention survey, and less than 180 seconds on the follow-up survey suggested that participants may have rushed through the questionnaires and app activities. Long string analysis and attention checks were carried out to confirm if these participants' responses were invalid. Responses were identified and removed if the same option was selected across the entire scale of any measure or responses to attention-check items were incorrect. A total of 32 participants were excluded for failing these data quality checks, including response times indicating rushed responses, long string analysis, and attention check failures, as their data were deemed unreliable for analysis.

Finally, data were screened for univariate outliers that fell beyond 3 SDs from the mean. Winsorization was conducted to address these outliers by substituting them with the closest nonoutlying value [71]. Subsequently, all analyses were conducted with and without the outliers. As the results were consistent across both datasets, only results from the initial dataset were reported.

SPSS (version 26.0, IBM Corp) was used for all statistical analyses. Independent 2-tailed *t* tests, Mann-Whitney *U* tests, and chi-square tests were conducted to analyze differences between the intervention and active control groups on the baseline and demographic variables. While the initial intention-to-treat (ITT) analysis was conducted on the full dataset of 115 and 109 participants, 32 participants were excluded after failing data quality checks. Thus, the final numbers analyzed were 94 and 98, reflecting a modified per-protocol analysis after quality checks.

Missing data were assessed for attrition biases using independent 2-tailed *t* tests, chi-square tests, and Mann-Whitney *U* tests to ensure there were no significant differences between participants who completed all assessments and those who did not. As all missing data were verified to be missing at random, they were addressed using ITT analyses [72], whereby missing data on the postintervention and follow-up assessment were substituted with data from the last completed questionnaire. Missing AES data were replaced using the mean score.

Preliminary Analyses

Main Analyses

Analysis of covariance (ANCOVA) examined if changes in OCD symptom severity at postintervention and follow-up were significantly different between the intervention and active control groups. ANCOVA is recommended for inferential testing intervention effects [73], as controlling for baseline scores ensures that results at postintervention and follow-up are attributable to the intervention [74]. ANCOVA was conducted on postintervention and follow-up OCI-R scores, with baseline OCI-R and DASS-21 scores as the covariates. The α level was set at $P<.05$. Cohen [75] guidelines for eta square were used to interpret effect sizes (η_p^2), whereby 0.01 to 0.05 indicates a small effect, 0.06 to 0.13 indicates a moderate effect, and 0.14 and above indicates a large effect.

Correlation and Moderation Analyses

Pearson correlation was conducted to test if baseline OCI-R scores were correlated with FMPS-MP scores. Due to nonnormality of baseline OCI-R scores, the bias-corrected and accelerated method (BCa) with 1000 bootstrap iterations was used to generate 95% CIs for Pearson correlation coefficients [76]. Pearson correlation analysis with BCa is robust to nonnormality and retains type I error control by generating more accurate CIs [77,78].

Subsequently, Hayes PROCESS (version 4.2, IBM Corp, macro model 1) was used for moderation analyses. A total of 2 models

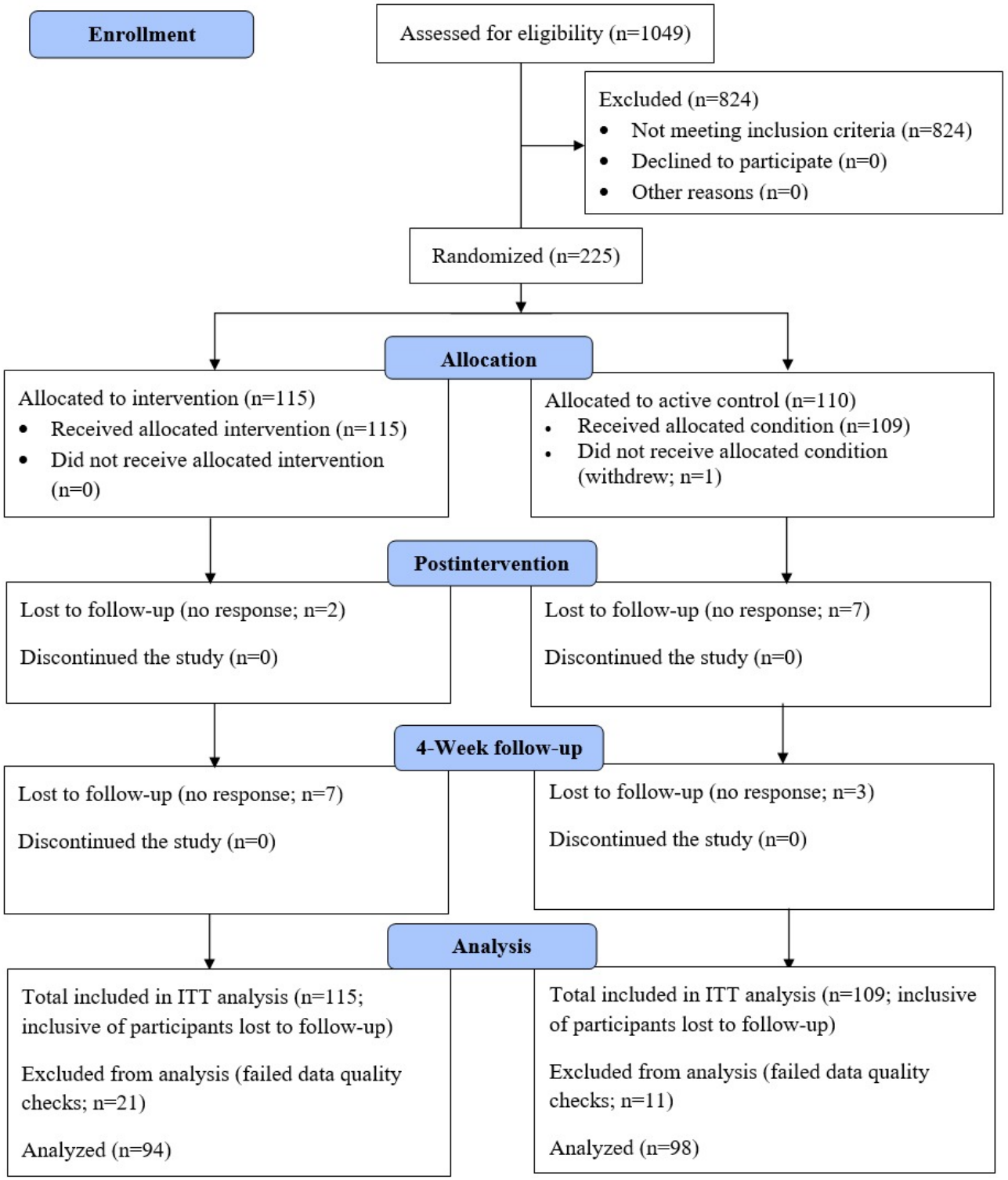
were analyzed. Postintervention and follow-up OCI-R scores were used as the dependent variables in the first and second models, respectively. Each model ran a simple moderation analysis to investigate the moderating effects of MP on the relationship between condition and OCI-R scores at postintervention or follow-up. Baseline OCI-R and DASS-21 scores were entered as the covariate for all analyses. The 95% CIs were used, with 5000 bootstrap iterations. The interaction term between condition and FMPS-MP was interpreted, whereby $P<.05$ indicates that MP moderates OCI-R scores at postintervention or follow-up.

Results

Participant Characteristics

A total of 1049 responses were screened for eligibility. Of these, 225 participants met the inclusion criteria, completed the baseline survey, and were randomly assigned to either the intervention or active control condition. One participant withdrew from the study before completing the postintervention assessment, resulting in 224 participants who completed the intervention. Among these, 215 (95.56%) participants completed the postintervention assessment, and 204 (90.67%) participants completed the follow-up assessment. Additionally, 32 participants were excluded from the analyses after failing data quality checks, leading to a final sample of 192 participants included in the statistical analyses (Figure 4).

Figure 4. CONSORT (Consolidated Standards of Reporting Trials) flow diagram. ITT: Intention to treat.



At baseline, 111 participants exhibited mild OCD symptom levels, and 81 participants exhibited moderate OCD symptom levels. Psychological distress was measured using the DASS-21, with average scores at baseline (mean 11.45, SD 8.66), postintervention (mean 9.93, SD 8.37), and follow-up (mean 11.59, SD 9.46), indicating that the participants were generally experiencing mild distress throughout the study.

Preliminary Analyses

The intervention and active control groups did not differ significantly on baseline demographic and outcome variables (Tables 2 and 3).

Table 2. Descriptive statistics for demographic variable of gender by condition.

Demographic variable	Full sample (n=192), n (%)	Intervention condition (n=94), n (%)	Active control condition (n=98), n (%)	*P* value
Gender				.47
Women	147 (76.6)	74 (78.7)	73 (74.5)	
Men	43 (22.4)	19 (20.2)	24 (24.5)	
Nonbinary	1 (0.5)	1 (1.1)	0 (0.0)	
Other	1 (0.5)	0 (0.0)	1 (1.0)	

Table 3. Descriptive statistics for baseline demographic and outcome variables by condition (n=192).

Variable	Intervention condition (n=94), mean (SD)	Active control condition (n=98), mean (SD)	*P* value
FMPS-MP[a]	61.21 (14.56)	57.93 (12.60)	.10
Age (years)	21.69 (1.46)	21.74 (1.82)	.82
OCI-R[b]	12.83 (5.41)	13.01 (5.18)	.86
DASS-21[c]	11.83 (9.32)	11.09 (8.00)	.79

[a]FMPS-MP: Frost Multidimensional Perfectionism Scale–Maladaptive Perfectionism.

[b]OCI-R: Obsessive Compulsive Inventory–Revised.

[c]DASS-21: Depression Anxiety Stress Scale-21.

For participants lost to postintervention and follow-up, Mann-Whitney *U* tests, chi-square tests, and independent 2-tailed *t* tests revealed no attrition-related biases across demographic and outcome variables. ITT analysis was conducted to address participants' missing data by bringing forward scores from their last completed survey. Mean substitution was conducted to address missing AES scores.

Independent 2-tailed *t* tests indicated that participants in both conditions enjoyed using the app, with no significant differences (*P*=.81) found between the intervention group (mean 25.33, SD 3.90) and the active control group (mean 25.47, SD 4.01) on the AES.

Participants were directly asked after the study if they believed they could guess the hypothesis. Mann-Whitney *U* tests indicated that participants who correctly guessed the study's hypothesis did not differ significantly from participants who incorrectly guessed the hypothesis on all measures at postintervention and follow-up.

Main Analyses

The assumptions for ANCOVA were met, except for the assumption for normality. However, due to the robustness of ANCOVA against deviations from normality when sample sizes are larger than 100 [77,78], analysis using ANCOVA proceeded. At both postintervention and follow-up, the intervention group reported significantly lower OCI-R scores compared with the active control group, with a small effect size (Table 4).

Table 4. Means, SDs, univariate *F* values, and effect sizes for OCI-R[a] scores at postintervention and 4-week follow-up.

	Intervention, mean (SD)	Control, mean (SD)	*F* test (*df*)	*P* value	ES[b]
Baseline	12.83 (5.41)	13.01 (5.18)	—[c]	—	—
Postintervention	11.03 (7.40)	13.04 (6.53)	5.79[d] (1)	.02[e]	.030
Follow-up	10.53 (7.36)	12.37 (7.46)	4.11[d] (1)	—	—

[a]OCI-R: Obsessive Compulsive Inventory–Revised.

[b]Effect size of 0.01=small, 0.06=moderate, 0.14=large [75].

[c]Not applicable.

[d]*P*<.05.

[e]Significant *P* value at .05.

Correlation and Moderation Analyses

Results from Pearson correlation analysis revealed that OCI-R and FMPS-MP were significantly correlated (*P*<.001) with a weak positive correlation (*r*=.28), and the BCa bootstrap (95% CI 0.14-0.41) did not include zero.

Moderation analyses indicated a significant effect in the overall moderation model with OCI-R scores as the dependent variable

and FMPS-MP as the moderator. At postintervention, the model showed R^2=0.38, $F_{5,186}$=22.57, and P<.001, and at follow-up, it showed R^2=0.29, $F_{5,186}$=15.23, and P<.001. However, the interaction term between the condition and FMPS-MP was not statistically significant at postintervention and follow-up. Results are displayed in Table 5.

Table 5. Standardized coefficients, SEs, t values, *P* values, and 95% CIs for moderation analyses.

Model and variable	OCI-R[a]				
	β	SE	*P* value	*t*	95% CI
1 (T1[b]-T2[c])					
Constant	4.85	5.93	.41	0.82	–6.85 to 16.56
Condition	–3.49	3.69	.34	–0.95	–10.77 to 3.78
FMPS-MP[d]	0.01	0.10	.95	0.07	–0.19 to 0.21
Condition×FMPS-MP	0.02	0.06	.70	0.39	–0.10 to 0.14
DASS-21[e]	0.08	0.05	.12	1.55	–0.02 to 0.18
OCI-R	0.70	0.08	<.001[f]	8.19	0.53 to 0.86
2 (T1-T3[g])					
Constant	–0.51	6.72	.94	–0.077	–13.77 to 12.74
Condition	–1.65	4.18	.69	–0.39	–9.88 to 6.59
FMPS-MP	0.14	0.11	.23	1.21	–0.09 to 0.36
Condition×FMPS-MP	–0.01	0.07	.88	–0.16	–0.15 to 0.12
DASS-21	0.16	0.06	.01[h]	2.79	0.05 to 0.28
OCI-R	0.29	0.10	<.001[f]	4.26	0.22 to 0.60

[a]OCI-R: Obsessive Compulsive Inventory–Revised.

[b]T1: preintervention.

[c]T2: postintervention.

[d]FMPS-MP: Frost Multidimensional Perfectionism Scale–Maladaptive Perfectionism.

[e]DASS-21: Depression Anxiety Stress Scale-21.

[f]*P* <.001.

[g]T3: follow-up.

[h]*P*<.05.

Discussion

Principal Findings

This study evaluated the efficacy of an 8-day, self-guided OCD program on an mHealth app compared with active control in reducing subclinical OCD symptoms and severity in a sample of Asian university students. The study further investigated the correlation between MP and OCD symptomatology, and MP as a factor that might help identify individuals who will benefit most from the self-guided program. The study addresses important gaps in the literature regarding the dearth of RCTs studying evidence-based mHealth app programs for OCD and the factors that influence the efficacy of such programs.

Our hypotheses regarding the efficacy of the app-based OCD program were supported. The intervention group reported significant reductions in OCD symptom severity compared with the active control group at both postintervention and follow-up. Consistent with past studies on longer ICBT or in-person interventions, our findings revealed that an 8-day, self-guided program using principles of CBT with ERP can reduce OCD symptoms and severity [16,23,79]. The small effect size reductions for OCD symptom severity in our OCD program are also comparable with a 12-week self-guided ICBT program for OCD at postintervention [80]. Taken together, these findings suggest that OCD gains can be obtained in a shorter duration through self-guided mHealth app programs. However, as there is limited research regarding the optimal duration of self-guided interventions, future research should examine how the efficacy of self-guided, low-intensity interventions for OCD might vary with program duration. While longer interventions may offer more comprehensive content and exercises, shorter interventions may offer increased convenience and improve program completion rates.

Our study also extended the findings from guided, 6-week-long, app-based program by Hwang et al [31], demonstrating that even without therapist support, app-based CBT can be beneficial for individuals with OCD symptoms. However, compared with the study by Hwang et al [31], which used a sample with clinical OCD, our study's subclinical sample did not yield large effect

size reductions for OCD. Despite our study's overall small effect sizes, this result is larger or comparable to effect sizes found after using the self-guided OCD.app for approximately 8.52 days. Specifically, Gamoran and Doron [33] found null effect sizes for participants with mild OCD symptom levels, and small effect sizes for participants with moderate OCD symptom levels. Research from previous RCTs has suggested that mHealth app use leads to greater decreases in symptoms for participants with moderate to severe OCD symptom levels [81], compared with those with less severe OCD symptom levels [29]. As such, because our study's sample comprised participants with both mild and moderate OCD symptom severity, it is possible that analyzing their data collectively diluted the effect size. Moreover, participants with OCI-R scores near zero were included in our study. Such participants would expectedly have a limited reduction in OCD symptom severity after the intervention. As participants in both groups generally reported low OCI-R scores at baseline, it is possible that a floor effect limited intervention effects. Hence, it would be fruitful for future RCTs to investigate how effect sizes may vary across different levels of OCD symptom severity after using self-guided app-based OCD programs. Additionally, the efficacy of such programs should be tested in samples with clinically significant levels of OCD symptom severity to better understand their clinical utility.

Overall, these findings provide preliminary support for the use of CBT with ERP on mHealth apps to reduce OCD symptom severity in Asian university students. As such, our findings align with the National Institute for Health and Care Excellence [32] stepped-care approach for managing OCD-related symptoms using low-intensity treatments. Additionally, the 4-week follow-up effects were encouraging as they suggest that the app-based OCD program can produce sustained effects, addressing the lack of follow-up data in mHealth app research [82]. As ICBT for OCD has been found to produce treatment effects maintained after 2 years [83], future mHealth app studies for OCD may seek to administer longer follow-up assessments to evaluate the long-term maintenance of treatment gains.

Our hypothesis that MP predicts OCD symptom severity was also supported. This finding corroborated with prior research that MP is highly prevalent in individuals with OCD [38]. Based on the cognitive-behavioral model of OCD, exhibiting higher levels of MP increases the risk of OCD as individuals are prone to beliefs that there is a perfect method to performing any task (eg, "I must order my books perfectly parallel to each other"). Additionally, individuals with high MP often have an inflated sense of responsibility over obsessions, anticipating that small mistakes and failures will have catastrophic consequences which they are responsible for preventing [84]. Consequently, obsessive distress and anxiety drive the individual to perform compulsions to dispel the threat of their obsessions [85]. However, the strength of the correlation found between MP and OCD symptoms in our study was weaker than that of prior research [86]. This may be due to a range restriction of OCI-R scores. Given that participants with clinically significant levels of OCD were excluded from our sample, our study's participants represent a narrower segment of the OCD severity spectrum than the general, broader OCD population. By limiting the variability in OCD symptom severity, the true strength of the correlation between MP and OCD symptom severity may have been obscured [87].

Furthermore, our hypothesis that MP would moderate the decrease of OCD symptoms at postintervention and follow-up was not supported. This result was unexpected, given that MP has been found to impede OCD treatment [40,88]. The weak correlation between MP and OCD symptom severity may have consequently led to and explained the lack of moderation effect of MP on the decrease of OCD symptoms at postintervention and follow-up. Another possible explanation for the lack of significant moderation effect could be due to the sample's high mean MP and small spread (mean 59.46, SD 13.72) [86,89]. As such, there might have been insufficient variability in MP scores needed to detect a moderating effect. Future studies may consider further examining MP as a moderator in samples with a wider range of OCD symptom severity.

Strengths and Limitations

The RCT design enabled the establishment of causal conclusions from the study's intervention and results [90]. Additionally, using an active control group controlled for attentional influences and nonspecific variables. Our study also had an attrition rate of 9.33%, which is much lower than participant dropout rates of up to 44.44% in similar studies on app-based OCD programs [81]. Furthermore, no attrition-related biases were detected in participants who discontinued the study. Altogether, these factors strengthened the overall validity of our results.

However, our study had several limitations. First, the low attrition and relatively high compliance observed may be partially attributable to the daily messaging by the study team, which is unlikely to be replicable in real-world settings. Next, data regarding how long participants spent applying skills from the programs in their daily lives were not collected. CBT research has found that greater application of CBT skills improves treatment outcomes [91], hence it is probable that participants who practiced the skills taught in our programs more frequently would have made increased gains. Moreover, there was no objective assessment of participant engagement and effort expended on the intervention programs as the time participants spent using the app was not tracked or controlled for. However, as poor engagement with mHealth apps has been associated with reduced gains [92], measures were taken to encourage adherence and ensure that app activities were fully completed. Participants received daily reminders to complete the app activities, and the app was programmed such that participants could only proceed to the next session if all preceding sessions were completed. Crucially, all participants were verified to have completed their assigned program. Given that no significant differences were found on the AES, both groups likely shared similar levels of engagement and time spent on the programs. Regardless, future studies should control for the duration of mHealth app use and investigate how the duration of app engagement may impact treatment outcomes.

Next, self-report measures are susceptible to individual biases, which could have contributed to the study's results. Social desirability bias is a concern because participant blinding was

not fully feasible, given that the study's hypothesis might be guessed based on the content of the allocated program. However, participants who correctly guessed the study's hypothesis did not report significantly different results on OCD outcomes compared with incorrect participants, which gives us some confidence that the effect of social desirability bias was limited. Nonetheless, our results should be interpreted with caution. Retrospective recall biases may also be present as the OCI-R required participants to report their experiences in the past month. To strengthen the accuracy of participants' self-reports on treatment outcomes, future studies may use daily journaling methods which can minimize memory distortion from prolonged recall periods [93].

Finally, as participants were recruited from one university and comprised mostly of female students, the external validity of our study may be limited. Future studies may replicate our findings with a more diverse sample of young adults.

In sum, the Intellect app offers a valuable solution for early intervention in subclinical OCD, helping prevent the escalation of symptoms into more severe forms. Its self-guided, CBT-based structure provides accessible support without the need for therapist involvement, addressing gaps in traditional mental health services. By offering a scalable and cost-effective option, the app enables individuals to manage their mental health more effectively, improving quality of life and reducing the risk of future clinical impairment.

Conclusion

Overall, our study provides preliminary support for the efficacy of a self-guided mHealth app program using principles of CBT with ERP in improving OCD symptom severity in young adults, producing gains maintained after 4 weeks. While MP was found to predict OCD symptom severity, MP did not moderate treatment outcomes. Identifying moderators can optimize outcome and treatment delivery on OCD-related mHealth apps. Hence, the impact of MP on treatment outcomes should be further studied in a broader spectrum of OCD symptom severity. Given the potential for cost-effective and accessible mHealth app programs to reach young adults and overcome traditional treatment barriers, future studies should aim to improve the efficacy of these programs.

Conflicts of Interest

The study was partly funded by Intellect Pte Ltd. OS holds the position of Vice President of Clinical at Intellect Pte Ltd; he also holds equity in Intellect.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.2).
[PDF File (Adobe PDF File), 784 KB - mhealth_v12i1e63316_app1.pdf]

References

1. Subramaniam M, Abdin E, Vaingankar J, Shafie S, Chang S, Seow E, et al. Obsessive-compulsive disorder in Singapore: prevalence, comorbidity, quality of life and social support. Ann Acad Med Singapore 2020;49(1):15-25. [doi: 10.47102/annals-acadmedsg.2019185]
2. Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the national comorbidity survey replication. Arch Gen Psychiatry 2005;62(6):593-602. [doi: 10.1001/archpsyc.62.6.593] [Medline: 15939837]
3. Abramowitz JS, Deacon BJ, Olatunji BO, Wheaton MG, Berman NC, Losardo D, et al. Assessment of obsessive-compulsive symptom dimensions: development and evaluation of the dimensional obsessive-compulsive scale. Psychol Assess 2010;22(1):180-198. [doi: 10.1037/a0018260] [Medline: 20230164]
4. Bobes J, González MP, Bascarán MT, Arango C, Sáiz PA, Bousoño M. Quality of life and disability in patients with obsessive-compulsive disorder. Eur Psychiatry 2001;16(4):239-245. [doi: 10.1016/s0924-9338(01)00571-5] [Medline: 11418275]
5. Bystritsky A, Liberman RP, Hwang S, Wallace CJ, Vapnik T, Maindment K, et al. Social functioning and quality of life comparisons between obsessive-compulsive and schizophrenic disorders. Depression Anxiety 2001;14(4):214-218. [doi: 10.1002/da.1069] [Medline: 11754128]
6. Olatunji BO, Williams BJ, Haslam N, Abramowitz JS, Tolin DF. The latent structure of obsessive-compulsive symptoms: a taxometric study. Depression Anxiety 2008;25(11):956-968. [doi: 10.1002/da.20387] [Medline: 17943983]
7. Grabe HJ, Meyer CH, Hapke U, Rumpf H, Freyberger H, Dilling H, et al. Prevalence, quality of life and psychosocial function in obsessive-compulsive disorder and subclinical obsessive-compulsive disorder in northern Germany. Eur Arch Psychiatry Clin Neurosci 2000;250(5):262-268. [doi: 10.1007/s004060070017] [Medline: 11097170]
8. Langlois F, Freeston MH, Ladouceur R. Differences and similarities between obsessive intrusive thoughts and worry in a non-clinical population: study 2. Behav Res Ther 2000;38(2):175-189. [doi: 10.1016/s0005-7967(99)00028-5] [Medline: 10661002]
9. Belloch A, Morillo C, Giménez A. Effects of suppressing neutral and obsession-like thoughts in normal subjects: beyond frequency. Behav Res Ther 2004;42(7):841-857. [doi: 10.1016/j.brat.2003.07.007] [Medline: 15149902]

10. Fineberg NA, Dell'Osso B, Albert U, Maina G, Geller D, Carmi L, et al. Early intervention for obsessive compulsive disorder: an expert consensus statement. Eur Neuropsychopharmacol 2019;29(4):549-565 [FREE Full text] [doi: 10.1016/j.euroneuro.2019.02.002] [Medline: 30773387]
11. Salkovskis PM. Cognitive-behavioural factors and the persistence of intrusive thoughts in obsessional problems. Behav Res Ther 1989;27(6):677-683. [doi: 10.1016/0005-7967(89)90152-6] [Medline: 2610662]
12. Taylor S, Abramowitz JS, McKay D. Cognitive-behavioral models of obsessive-compulsive disorder. In: Antony MM, Purdon C, Summerfeldt LJ, editors. Psychological Treatment of Obsessive-Compulsive Disorder: Fundamentals and beyond. Washington, DC: American Psychological Association; 2007.
13. Morillo C, Belloch A, García-Soriano G. Clinical obsessions in obsessive-compulsive patients and obsession-relevant intrusive thoughts in non-clinical, depressed and anxious subjects: where are the differences? Behav Res Ther 2007;45(6):1319-1333. [doi: 10.1016/j.brat.2006.11.005] [Medline: 17208197]
14. Öst LG, Havnen A, Hansen B, Kvale G. Cognitive behavioral treatments of obsessive-compulsive disorder. a systematic review and meta-analysis of studies published 1993-2014. Clin Psychol Rev 2015;40:156-169. [doi: 10.1016/j.cpr.2015.06.003] [Medline: 26117062]
15. Jacoby RJ, Abramowitz JS. Inhibitory learning approaches to exposure therapy: a critical review and translation to obsessive-compulsive disorder. Clin Psychol Rev 2016;49:28-40. [doi: 10.1016/j.cpr.2016.07.001] [Medline: 27521505]
16. Lindsay M, Crino R, Andrews G. Controlled trial of exposure and response prevention in obsessive-compulsive disorder. Br J Psychiatry 1997;171(2):135-139. [doi: 10.1192/bjp.171.2.135] [Medline: 9337948]
17. Reid JE, Laws KR, Drummond L, Vismara M, Grancini B, Mpavaenda D, et al. Cognitive behavioural therapy with exposure and response prevention in the treatment of obsessive-compulsive disorder: a systematic review and meta-analysis of randomised controlled trials. Compr Psychiatry 2021;106:152223 [FREE Full text] [doi: 10.1016/j.comppsych.2021.152223] [Medline: 33618297]
18. Ferrando C, Selai C. A systematic review and meta-analysis on the effectiveness of exposure and response prevention therapy in the treatment of obsessive-compulsive disorder. J Obsessive Compuls Relat Disord 2021;31:100684. [doi: 10.1016/j.jocrd.2021.100684]
19. Maltby N, Tolin DF. A brief motivational intervention for treatment-refusing OCD patients. Cogn Behav Ther 2005;34(3):176-184. [doi: 10.1080/16506070510043741] [Medline: 16195056]
20. Mancebo MC, Eisen JL, Sibrava NJ, Dyck IR, Rasmussen SA. Patient utilization of cognitive-behavioral therapy for OCD. Behav Ther 2011;42(3):399-412 [FREE Full text] [doi: 10.1016/j.beth.2010.10.002] [Medline: 21658523]
21. Wootton BM, Dear BF, Johnston L, Terides MD, Titov N. Self-guided internet administered treatment for obsessive-compulsive disorder: results from two open trials. J Obsessive Compuls Relat Disord 2014;3(2):102-108. [doi: 10.1016/j.jocrd.2014.03.001]
22. Kohn R, Saxena S, Levav I, Saraceno B. The treatment gap in mental health care. Bull World Health Organ 2004;82(11):858-866 [FREE Full text] [Medline: 15640922]
23. Wootton BM, Karin E, Titov N, Dear BF. Self-guided internet-delivered cognitive behavior therapy (ICBT) for obsessive-compulsive symptoms: a randomized controlled trial. J Anxiety Disord 2019;66:102111. [doi: 10.1016/j.janxdis.2019.102111] [Medline: 31301476]
24. Wootton BM. Remote cognitive-behavior therapy for obsessive-compulsive symptoms: a meta-analysis. Clin Psychol Rev 2016;43:103-113. [doi: 10.1016/j.cpr.2015.10.001] [Medline: 26494179]
25. Lundström L, Flygare O, Andersson E, Enander J, Bottai M, Ivanov VZ, et al. Effect of internet-based vs face-to-face cognitive behavioral therapy for adults with obsessive-compulsive disorder: a randomized clinical trial. JAMA Netw Open 2022;5(3):e221967 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.1967] [Medline: 35285923]
26. Rees CS, Anderson RA, Kane RT, Finlay-Jones AL. Online obsessive-compulsive disorder treatment: preliminary results of the "OCD? not me!" self-guided internet-based cognitive behavioral therapy program for young people. JMIR Ment Health 2016;3(3):e29 [FREE Full text] [doi: 10.2196/mental.5363] [Medline: 27381977]
27. Lecomte T, Potvin S, Corbière M, Guay S, Samson C, Cloutier B, et al. Mobile apps for mental health issues: meta-review of meta-analyses. JMIR Mhealth Uhealth 2020;8(5):e17458 [FREE Full text] [doi: 10.2196/17458] [Medline: 32348289]
28. Cerea S, Ghisi M, Bottesi G, Carraro E, Broggio D, Doron G. Reaching reliable change using short, daily, cognitive training exercises delivered on a mobile application: the case of relationship obsessive compulsive disorder (ROCD) symptoms and cognitions in a subclinical cohort. J Affect Disord 2020;276:775-787. [doi: 10.1016/j.jad.2020.07.043] [Medline: 32738662]
29. Roncero M, Belloch A, Doron G. A novel approach to challenging OCD related beliefs using a mobile-app: an exploratory study. J Behav Ther Exp Psychiatry 2018;59:157-160. [doi: 10.1016/j.jbtep.2018.01.008] [Medline: 29425951]
30. Gershkovich M, Middleton R, Hezel DM, Grimaldi S, Renna M, Basaraba C, et al. Integrating exposure and response prevention with a mobile app to treat obsessive-compulsive disorder: feasibility, acceptability, and preliminary effects. Behav Ther 2021;52(2):394-405. [doi: 10.1016/j.beth.2020.05.001] [Medline: 33622508]
31. Hwang H, Bae S, Hong JS, Han DH. Comparing effectiveness between a mobile app program and traditional cognitive behavior therapy in obsessive-compulsive disorder: evaluation study. JMIR Ment Health 2021;8(1):e23778. [doi: 10.2196/23778] [Medline: 33464208]

32. National Institute for Clinical Excellence. Obsessive-Compulsive Disorder. In: Core Interventions in the Treatment of Obsessive-Compulsive Disorder and Body Dysmorphic Disorder. Leicester, United Kingdom: British Psychological Society; 2006.
33. Gamoran A, Doron G. J Obsessive Compuls Relat Disord 2023;36:100782. [doi: 10.1016/j.jocrd.2023.100782]
34. Knopp J, Knowles S, Bee P, Lovell K, Bower P. A systematic review of predictors and moderators of response to psychological therapies in OCD: do we have enough empirical evidence to target treatment? Clin Psychol Rev 2013;33(8):1067-1081 [FREE Full text] [doi: 10.1016/j.cpr.2013.08.008] [Medline: 24077387]
35. Frost RO, Steketee G. Perfectionism in obsessive-compulsive disorder patients. Behav Res Ther 1997;35(4):291-296 [FREE Full text] [doi: 10.1016/s0005-7967(96)00108-8] [Medline: 9134783]
36. Frost RO, Marten P, Lahart C, Rosenblate R. The dimensions of perfectionism. Cogn Ther Res 1990;14(5):449-468. [doi: 10.1007/Bf01172967]
37. Martinelli M, Chasson GS, Wetterneck CT, Hart JM, Björgvinsson T. Perfectionism dimensions as predictors of symptom dimensions of obsessive-compulsive disorder. Bull Menninger Clin 2014;78(2):140-159. [doi: 10.1521/bumc.2014.78.2.140] [Medline: 24870847]
38. Pinto A, Mancebo MC, Eisen JL, Pagano ME, Rasmussen SA. The brown longitudinal obsessive compulsive study: clinical features and symptoms of the sample at intake. J Clin Psychiatry 2006;67(5):703-711 [FREE Full text] [doi: 10.4088/jcp.v67n0503] [Medline: 16841619]
39. Rhéaume J, Freeston MH, Dugas MJ, Letarte H, Ladouceur R. Perfectionism, responsibility and obsessive-compulsive symptoms. Behav Res Ther 1995;33(7):785-794. [doi: 10.1016/0005-7967(95)00017-r] [Medline: 7677716]
40. Pinto A, Liebowitz MR, Foa EB, Simpson HB. Obsessive compulsive personality disorder as a predictor of exposure and ritual prevention outcome for obsessive compulsive disorder. Behav Res Ther 2011;49(8):453-458 [FREE Full text] [doi: 10.1016/j.brat.2011.04.004] [Medline: 21600563]
41. Subramaniam M, Abdin E, Vaingankar JA, Shafie S, Chua HC, Tan WM, et al. Minding the treatment gap: results of the singapore mental health study. Soc Psychiatry Psychiatr Epidemiol 2020;55(11):1415-1424 [FREE Full text] [doi: 10.1007/s00127-019-01748-0] [Medline: 31317246]
42. Fineberg NA, Hengartner MP, Bergbaum C, Gale T, Rössler W, Angst J. Remission of obsessive-compulsive disorders and syndromes; evidence from a prospective community cohort study over 30 years. Int J Psychiatry Clin Pract 2013;17(3):179-187. [doi: 10.3109/13651501.2013.777744] [Medline: 23428237]
43. Agapie E, Chang K, Patrachari S, Neary M, Schueller SM. Understanding mental health apps for youth: focus group study with latinx youth. JMIR Form Res 2022;6(10):e40726 [FREE Full text] [doi: 10.2196/40726] [Medline: 36256835]
44. Nguyen NC, Bosch OJH, Ong FY, Seah JS, Succu A, Nguyen TV, et al. A systemic approach to understand smartphone usage in Singapore. Syst Res Behav Sci 2015;33(3):360-380. [doi: 10.1002/sres.2348]
45. Salamanca-Sanabria A, Jabir AI, Lin X, Alattas A, Kocaballi AB, Lee J, et al. Exploring the perceptions of mHealth interventions for the prevention of common mental disorders in university students in Singapore: qualitative study. J Med Internet Res 2023;25:e44542 [FREE Full text] [doi: 10.2196/44542] [Medline: 36939808]
46. Castro JR, Rice KG. Perfectionism and ethnicity: implications for depressive symptoms and self-reported academic achievement. Cultur Divers Ethnic Minor Psychol 2003;9(1):64-78. [doi: 10.1037//1099-9809.9.1.64]
47. Perera MJ, Chang EC. Ethnic variations between Asian and European Americans in interpersonal sources of socially prescribed perfectionism: it's not just about parents!. Asian Am J Psychol 2015;6(1):31-37. [doi: 10.1037/a0036175]
48. Oishi S, Sullivan HW. The mediating role of parental expectations in culture and well‐being. J Pers 2005;73(5):1267-1294. [doi: 10.1111/j.1467-6494.2005.00349.x] [Medline: 16138873]
49. Flett GL, Hewitt PL. Perfectionism: theory, research, and treatment. In: Flett GL, Hewitt PL, editors. Perfectionism and Malad Justment: An Overview of Theoretical, Definitional, and Treatment. Washington, DC: American Psychological Association; 2002.
50. Rudy D, Halgunseth LC. Psychological control, maternal emotion and cognition, and child outcomes in individualist and collectivist groups. J Emotional Abuse 2005;5(4):237-264. [doi: 10.1300/j135v05n04_04]
51. Kawamura KY, Frost RO, Harmatz MG. The relationship of perceived parenting styles to perfectionism. Pers Individ Differ 2002;32(2):317-327. [doi: 10.1016/s0191-8869(01)00026-5]
52. Wang L, Wong YJ, Chung YB. Family perfectionism, shame, and mental health among Asian American and Asian international emerging adults: mediating and moderating relationships. Asian Am J Psychol 2018;9(2):117-126. [doi: 10.1037/aap0000098]
53. Hong RY, Lee SSM, Chng RY, Zhou Y, Tsai F, Tan SH. Developmental trajectories of maladaptive perfectionism in middle childhood. J Pers 2017;85(3):409-422. [doi: 10.1111/jopy.12249] [Medline: 26919690]
54. Damian LE, Stoeber J, Negru-Subtirica O, Băban A. On the development of perfectionism: the longitudinal role of academic achievement and academic efficacy. J Pers 2017;85(4):565-577. [doi: 10.1111/jopy.12261] [Medline: 27237456]
55. Luo W, Hogan DJ, Yeung AS, Sheng YZ, Aye KM. Attributional beliefs of Singapore students: relations to self-construal, competence and achievement goals. Educ Psychol 2013;34(2):154-170. [doi: 10.1080/01443410.2013.785056]

56. Hoppen LM, Kuck N, Bürkner PC, Karin E, Wootton BM, Buhlmann U. Low intensity technology-delivered cognitive behavioral therapy for obsessive-compulsive disorder: a meta-analysis. BMC Psychiatry 2021;21(1):322 [FREE Full text] [doi: 10.1186/s12888-021-03272-5] [Medline: 34193113]
57. Kosasih FR, Yee VTS, Toh SHY, Sündermann O. Efficacy of intellect's self-guided anxiety and worry mobile health programme: a randomized controlled trial with an active control and a 2-week follow-up. PLOS Digit Health 2023;2(5):e0000095 [FREE Full text] [doi: 10.1371/journal.pdig.0000095] [Medline: 37224139]
58. Ong WY, Sündermann O. Efficacy of the mental health app “Intellect” to improve body image and self-compassion in young adults: a randomized controlled trial with a 4-week follow-up. JMIR Mhealth Uhealth 2022;10(11):e41800. [doi: 10.2196/41800]
59. Toh SHY, Tan JHY, Kosasih FR, Sündermann O. Efficacy of the mental health app intellect to reduce stress: randomized controlled trial with a 1-month follow-up. JMIR Form Res 2022;6(12):e40723 [FREE Full text] [doi: 10.2196/40723] [Medline: 36515984]
60. Curran PG. Methods for the detection of carelessly invalid responses in survey data. J Exp Soc Psychol 2016;66:4-19. [doi: 10.1016/j.jesp.2015.07.006]
61. Fernández-López R, Riquelme-Gallego B, Bueno-Cavanillas A, Khan KS. Influence of placebo effect in mental disorders research: a systematic review and meta-analysis. Eur J Clin Invest 2022;52(7):e13762 [FREE Full text] [doi: 10.1111/eci.13762] [Medline: 35224726]
62. Foa EB, Huppert JD, Leiberg S, Langner R, Kichic R, Hajcak G, et al. The obsessive-complusive inventory: development and validation of a short version. Psychol Assess 2002;14(4):485-495. [doi: 10.1037//1040-3590.14.4.485]
63. Abramovitch A, Abramowitz JS, Riemann BC, McKay D. Severity benchmarks and contemporary clinical norms for the obsessive-compulsive inventory-revised (OCI-R). J Obsessive Compuls Relat Disord 2020;27:100557. [doi: 10.1016/j.jocrd.2020.100557]
64. Frost RO, Heimberg RG, Holt CS, Mattia JI, Neubauer AL. A comparison of two measures of perfectionism. Pers Individ Differ 1993;14(1):119-126. [doi: 10.1016/0191-8869(93)90181-2]
65. Lovibond PF, Lovibond SH. The structure of negative emotional states: comparison of the depression anxiety stress scales (DASS) with the beck depression and anxiety inventories. Behav Res Ther 1995;33(3):335-343. [doi: 10.1016/0005-7967(94)00075-U]
66. Laranjeira C, Querido A, Sousa P, Dixe MA. Assessment and psychometric properties of the 21-item depression anxiety stress scale (DASS-21) among portuguese higher education students during the COVID-19 pandemic. Eur J Investig Health Psychol Educ 2023;13(11):2546-2560 [FREE Full text] [doi: 10.3390/ejihpe13110177] [Medline: 37998067]
67. Adams TG, Kelmendi B, Brake CA, Gruner P, Badour CL, Pittenger C. The role of stress in the pathogenesis and maintenance of obsessive-compulsive disorder. Chronic Stress (Thousand Oaks) 2018;2:247054701875804 [FREE Full text] [doi: 10.1177/2470547018758043] [Medline: 29527593]
68. Diniz JB, Miguel EC, de Oliveira AR, Reimer AE, Brandão ML, de Mathis MA, et al. Outlining new frontiers for the comprehension of obsessive-compulsive disorder: a review of its relationship with fear and anxiety. Revista Brasileira de Psiquiatria 2012;34:S81-S103. [doi: 10.1016/s1516-4446(12)70056-9]
69. Shaw AM, Carbonella JY, Arditte Hall KA, Timpano KR. Obsessive-compulsive and depressive symptoms: the role of depressive cognitive styles. J Psychol 2017;151(6):532-546. [doi: 10.1080/00223980.2017.1372341] [Medline: 29077549]
70. Bakker D, Rickard N. Engagement in mobile phone app for self-monitoring of emotional wellbeing predicts changes in mental health: MoodPrism. J Affect Disord 2018;227:432-442. [doi: 10.1016/j.jad.2017.11.016] [Medline: 29154165]
71. Weichle T, Hynes DM, Durazo-Arvizu R, Tarlov E, Zhang Q. Impact of alternative approaches to assess outlying and influential observations on health care costs. Springerplus 2013;2(1):614 [FREE Full text] [doi: 10.1186/2193-1801-2-614] [Medline: 24303338]
72. McCoy E. Understanding the intention-to-treat principle in randomized controlled trials. West J Emerg Med 2017;18(6):1075-1078. [doi: 10.5811/westjem.2017.8.35985] [Medline: 29085540]
73. Mistry EA, Yeatts SD, Khatri P, Mistry AM, Detry M, Viele K, et al. National Institutes of Health Stroke Scale as an outcome in stroke research: value of ANCOVA over analyzing change from baseline. Stroke 2022;53(4). [doi: 10.1161/strokeaha.121.034859]
74. Vickers AJ, Altman DG. Statistics notes: analysing controlled trials with baseline and follow up measurements. BMJ 2001;323(7321):1123-1124 [FREE Full text] [doi: 10.1136/bmj.323.7321.1123] [Medline: 11701584]
75. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd ed. Mahwah, NJ: Lawrence Erlbaum Associates; 1988.
76. Field A. Discovering Statistics Using IBM SPSS Statistics. 5th ed. Thousand Oaks, CA: SAGE Publications; 2018.
77. Bishara AJ, Hittner JB. Testing the significance of a correlation with nonnormal data: comparison of pearson, spearman, transformation, and resampling approaches. Psychol Methods 2012;17(3):399-417. [doi: 10.1037/a0028087] [Medline: 22563845]
78. Lee WC, Rodgers JL. Bootstrapping correlation coefficients using univariate and bivariate sampling. Psychol Methods 1998;3(1):91-103. [doi: 10.1037//1082-989x.3.1.91]

79. Mahoney AE, Mackenzie A, Williams AD, Smith J, Andrews G. Internet cognitive behavioural treatment for obsessive compulsive disorder: a randomised controlled trial. Behav Res Ther 2014;63:99-106. [doi: 10.1016/j.brat.2014.09.012] [Medline: 25461784]
80. WATERNAUX CM. Asymptotic distribution of the sample roots for a nonnormal population. Biometrika 1976;63(3):639-645. [doi: 10.1093/biomet/63.3.639]
81. Akin-Sari B, Inozu M, Haciomeroglu AB, Cekci BC, Uzumcu E, Doron G. Cognitive training via a mobile application to reduce obsessive-compulsive-related distress and cognitions during the COVID-19 outbreaks: a randomized controlled trial using a subclinical cohort. Behav Ther 2022;53(5):776-792 [FREE Full text] [doi: 10.1016/j.beth.2021.12.008] [Medline: 35987538]
82. Linardon J, Cuijpers P, Carlbring P, Messer M, Fuller-Tyszkiewicz M. The efficacy of app-supported smartphone interventions for mental health problems: a meta-analysis of randomized controlled trials. World Psychiatry 2019;18(3):325-336 [FREE Full text] [doi: 10.1002/wps.20673] [Medline: 31496095]
83. Andersson E, Steneby S, Karlsson K, Ljótsson B, Hedman E, Enander J, et al. Long-term efficacy of Internet-based cognitive behavior therapy for obsessive–compulsive disorder with or without booster: a randomized controlled trial. Psychol. Med 2014;44(13):2877-2887. [doi: 10.1017/s0033291714000543]
84. Freeston MH, Rhéaume J, Ladouceur R. Correcting faulty appraisals of obsessional thoughts. Behav Res Ther 1996;34(5-6):433-446. [doi: 10.1016/0005-7967(95)00076-3] [Medline: 8687365]
85. Frost RO, Novara C, Rhéaume J. Cognitive approaches to obsessions and compulsions. In: Frost RO, Steketee G, editors. Perfectionism in Obsessive Compulsive Disorder. Oxford, United Kingdom: Pergamon/Elsevier Science Inc; 2002:91-105.
86. Wu KD, Cortesi GT. Relations between perfectionism and obsessive-compulsive symptoms: examination of specificity among the dimensions. J Anxiety Disord 2009;23(3):393-400. [doi: 10.1016/j.janxdis.2008.11.006] [Medline: 19110399]
87. Sackett PR, Yang H. Correction for range restriction: an expanded typology. J Appl Psychol 2000;85(1):112-118. [doi: 10.1037//0021-9010.85.1.112]
88. Pinto A, Dargani N, Wheaton MG, Cervoni C, Rees CS, Egan SJ. Perfectionism in obsessive-compulsive disorder and related disorders: what should treating clinicians know? J Obsessive Compuls Relat Disord 2017;12:102-108. [doi: 10.1016/j.jocrd.2017.01.001]
89. Welch HA, Agras WS, Lock J, Halmi KA. Perfectionism, anorexia nervosa, and family treatment: how perfectionism changes throughout treatment and predicts outcomes. Int J Eat Disord 2020;53(12):2055-2060. [doi: 10.1002/eat.23396] [Medline: 33094868]
90. Hariton E, Locascio JJ. Randomised controlled trials - the gold standard for effectiveness research: study design: randomised controlled trials. BJOG 2018;125(13):1716 [FREE Full text] [doi: 10.1111/1471-0528.15199] [Medline: 29916205]
91. Camacho KS, Page AC, Hooke GR. An exploration of the relationships between patient application of CBT skills and therapeutic outcomes during a two-week CBT treatment. Psychother Res 2021;31(6):778-788. [doi: 10.1080/10503307.2020.1845414] [Medline: 33186064]
92. Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. Evid Based Ment Health 2018;21(3):116-119 [FREE Full text] [doi: 10.1136/eb-2018-102891] [Medline: 29871870]
93. Te Braak P, van Tienoven TP, Minnen J, Glorieux I. Data quality and recall bias in time-diary research: the effects of prolonged recall periods in self-administered online time-use surveys. Sociol Methodol 2022;53(1):115-138. [doi: 10.1177/00811750221126499]

Abbreviations

AES: App Engagement Scale
ANCOVA: analysis of covariance
BCa: bias-corrected and accelerated method
CBT: cognitive behavioral therapy
CM: concerned about making mistakes
DA: doubting the quality of one's actions
DASS-21: Depression Anxiety Stress Scale-21
ERP: exposure and response prevention
FMPS-MP: Frost Multidimensional Perfectionism Scale
FMPS-MP: Frost Multidimensional Perfectionism Scale–Maladaptive Perfectionism
ICBT: internet-based CBT
ITT: intention-to-treat
mHealth: mobile health
MP: maladaptive perfectionism
NUS: National University of Singapore
OCD: obsessive-compulsive disorder

OCI-R: Obsessive Compulsive Inventory–Revised
PC: parental criticism
PE: parental expectations
RCT: randomized controlled trial

Edited by D Cox; submitted 19.06.24; peer-reviewed by E Choi, S Arnaez; comments to author 18.08.24; revised version received 18.09.24; accepted 12.11.24; published 16.12.24.

Please cite as:
Lee Yoon Li M, Lee Si Min S, Sündermann O
Efficacy of the mHealth App Intellect in Improving Subclinical Obsessive-Compulsive Disorder in University Students: Randomized Controlled Trial With a 4-Week Follow-Up
JMIR Mhealth Uhealth 2024;12:e63316
URL: https://mhealth.jmir.org/2024/1/e63316
doi:10.2196/63316
PMID:

Original Paper

The Roles of Trust in Government and Sense of Community in the COVID-19 Contact Tracing Privacy Calculus: Mixed Method Study Using a 2-Wave Survey and In-Depth Interviews

Hyunjin Kang[1], PhD; Jeong Kyu Lee[2], PhD; Edmund WJ Lee[1], PhD; Cindy Toh[3], BA

[1]Wee Kim Wee School of Communication and Information, Nanyang Technological University, Singapore, Singapore

[2]Department of Health and Exercise Science, University of Oklahoma, Norman, OK, United States

[3]Department of Anthropology, Columbia University, New York, NY, United States

Corresponding Author:
Hyunjin Kang, PhD
Wee Kim Wee School of Communication and Information
Nanyang Technological University
31 Nanyang Link
Singapore, 637718
Singapore
Phone: 65 69083431
Email: hjkang@ntu.edu.sg

Abstract

Background: Contact tracing technology has been adopted in many countries to aid in identifying, evaluating, and handling individuals who have had contact with those infected with COVID-19. Singapore was among the countries that actively implemented the government-led contact tracing program known as TraceTogether. Despite the benefits the contact tracing program could provide to individuals and the community, privacy issues were a significant barrier to individuals' acceptance of the program.

Objective: Building on the privacy calculus model, this study investigates how the perceptions of the 2 key groups (ie, government and community members) involved in the digital contact tracing factor into individuals' privacy calculus of digital contact tracing.

Methods: Using a mixed method approach, we conducted (1) a 2-wave survey (n=674) and (2) in-depth interviews (n=12) with TraceTogether users in Singapore. Using structural equation modeling, this study investigated how trust in the government and the sense of community exhibited by individuals during the early stage of implementation (time 1) predicted privacy concerns, perceived benefits, and future use intentions, measured after the program was fully implemented (time 2). Expanding on the survey results, this study conducted one-on-one interviews to gain in-depth insights into the privacy considerations involved in digital contact tracing.

Results: The results from the survey showed that trust in the government increased perceived benefits while decreasing privacy concerns regarding the use of TraceTogether. Furthermore, individuals who felt a connection to community members by participating in the program (ie, the sense of community) were more inclined to believe in its benefits. The sense of community also played a moderating role in the influence of government trust on perceived benefits. Follow-up in-depth interviews highlighted that having a sense of control over information and transparency in the government's data management were crucial factors in privacy considerations. The interviews also highlighted surveillance as the most prevalent aspect of privacy concerns regarding TraceTogether use. In addition, our findings revealed that trust in the government, particularly the perceived transparency of government actions, was most strongly associated with concerns regarding the secondary use of data.

Conclusions: Using a mixed method approach involving a 2-wave survey and in-depth interview data, we expanded our understanding of privacy decisions and the privacy calculus in the context of digital contact tracing. The opposite influences of privacy concerns and perceived benefit on use intention suggest that the privacy calculus in TraceTogether might be viewed as a rational process of weighing between privacy risks and use benefits to make an uptake decision. However, our study demonstrated that existing perceptions toward the provider and the government in the contact tracing context, as well as the perception of the community triggered by TraceTogether use, may bias user appraisals of privacy risks and the benefits of contact tracing.

(JMIR Mhealth Uhealth 2024;12:e48986) doi:10.2196/48986

KEYWORDS

COVID-19; contact tracing technology; privacy calculus; trust in government; sense of community; mixed method; mobile phone

Introduction

Background

Contact tracing technology has been implemented in many countries to help identify, assess, and manage individuals exposed to persons infected with COVID-19 [1]. The main aims of a digital contact tracing system are not only to notify those who may have contracted the virus unknowingly but also to provide corresponding information and instructions from public health officials. Hence, some of contact tracing applications are designed to track the geolocations and mobility patterns of individuals using GPS and location-based solutions [2,3]. Digital contact tracing identifies individuals' close contacts with people confirmed or suspected of having COVID-19 through Bluetooth technology [4] or even allows authorities to make decisions as to whether certain individuals are certified as healthy enough to enter public spaces [5]. The use and implementation of contact tracing apps during COVID-19 brings concerns of privacy intrusion and surveillance to the forefront, as the attempt to digitally map and trace the spread of the virus is conducted through apps that collect some of the most personal and sensitive data. Thus, privacy concerns were identified as a significant barrier to adopting the contact tracing system, significantly influencing the intention to participate in the program [6,7].

Applying the privacy calculus model, that is, the extent to which individuals are willing to disclose personal information after weighing the risks and benefits, this study examines how individuals' perceptions of two important key entities—government and community members—involved in digital contact tracing play roles in their privacy calculus for adopting contact tracing. First, most contact tracing systems, including those in Singapore, Ireland, and France, were introduced and managed by the national government [6]. Derived from the social exchange concept, prior research on privacy suggests that trust in a service provider is a significant determinant of privacy calculus on the web [8,9]. Considering the managerial roles played by the government for contact tracing measures and the extensive personal information collected and analyzed for this effort, we postulated that individuals' trust in the government would be the key influencer of their privacy calculus in digital contact tracing.

Community members are another key entity involved in the privacy calculus in digital contact tracing. The active participation of community members in the contact tracing program is crucial for the success of preventive efforts. However, digital contact tracing implies that the participants are connected through a web-based system for tracking, which may trigger privacy risks among users. Therefore, the sense of community triggered by the contact tracing measure may influence the relationship between privacy concerns and benefits, shaping one's intention to participate in contact tracing efforts.

This study focuses on understanding privacy calculus in the context of using TraceTogether, a nationwide contact tracing program implemented in Singapore during the COVID-19 pandemic. The Government Technology Agency of Singapore, also known as GovTech, launched TraceTogether on March 20, 2020, followed by the launch of a token on June 28 to increase accessibility for those without a smartphone. In late December 2020, it became mandatory to check using TraceTogether before entering various places, including restaurants, workplaces, schools, places of worship, and shopping malls. As COVID-19 prevention measures eased, TraceTogether check-ins were no longer required in most settings from April 26, 2022.

Objective

This study investigates how trust in government and the sense of community during the early stage of contact tracing influence privacy calculus and intention to use the technology during the later stage when contact tracing becomes compulsory. To this end, a 2-wave survey and a follow-up in-depth interview with those who participated in the contact tracing program in Singapore were conducted. Building on the survey results, in-depth interviews were conducted to gain insights into privacy calculus in digital contact tracings, as well as the ways and reasons why trust in government and sense of community impact privacy calculus in contact tracing for COVID-19 prevention.

Literature Review

Theoretical Framework: Privacy Calculus Theory

Overview

Privacy concerns about contact tracing programs during the COVID-19 pandemic have been raised because of the potential risks of extensive data collection, surveillance, and the possibility of unauthorized access [6,7], necessitating a careful balance between public health benefits and safeguarding individuals' private information. Thus, privacy calculus theory has been used as one of the most robust theories within the field of communication [10] to unpack the underlying theoretical mechanisms in explaining the use of contact tracing apps through the lens of privacy. Having its theoretical roots and origins in economics and social exchange theory, individuals are assumed to be rational in making decisions and weighing trade-offs, and these decisions can be mathematically modeled as a function of the interaction between the benefits and risks of adopting a behavior [11]. For these reasons, privacy calculus theory offers a useful framework for investigating the influence of competing positive and negative beliefs on individuals' willingness to engage with technologies by capturing the extent of trade-offs individuals are willing to make in relation to privacy [12]. For instance, despite knowing that companies such as Google or Waze could identify individuals' precise geocoordinates, individuals who commute by driving may be willing to disclose this otherwise sensitive information (where they are and their mobility patterns) in exchange for the benefit of getting to their desired locations.

Privacy Concerns About Digital Contact Tracing

At the heart of privacy calculus theory is the act of self-disclosure of personal data, which is a product of two opposing forces: privacy concerns and the perceived benefits of information disclosure. *Privacy concerns* can be defined as individuals' fears and apprehensions over the possible loss of privacy, as well as how their personal data could be used or abused [13,14]. Privacy concerns are multidimensional by nature, for example, [15,16]. Privacy concerns encompass 3 dimensions: *perceived surveillance*, *perceived intrusion*, and *secondary use of information* [16]. At a fundamental level, privacy concerns are the worries that people are being watched. *Perceived surveillance* encapsulates this notion and refers to individuals' perceptions of the act of data collection pertaining to their lives—whether legitimate or illegal [17]. In the digital age, this could take the form of devices and technological platforms that capture individuals' personal activities by recording their behaviors [18]. For instance, people living in Florida expressed concerns over the Florida Department of Health's Healthy Together contact tracing app in collecting data such as contact lists, phone numbers, and medical conditions [19]. Relatedly, *perceived intrusion* is another dimension of privacy concerns and refers to the extent to which others are able to make independent decisions about possessing or soliciting information. For instance, in China, the government implemented a QR code system for contact tracing and assigned colors, such as green, yellow, or red, depending on the health status of individuals and the risk of exposure by triangulating data from public transport and health care systems [20], which determined whether they had access to public places [21]. The third dimension of privacy concerns is the secondary use of information, defined as the use of personal data collected from individuals for purposes other than what it was intended for [22]. As a large volume of data were collected by contact tracing apps, there were fears over how the government would use (or abuse) potentially sensitive data [20,21].

In the context of contact tracing app use, a recent study showed that when people are concerned about the extent of surveillance and intrusion of the implementation of contact tracing apps, this may negatively influence their intention to use, as fears of public health surveillance could be transformed into routine monitoring of populations extending beyond the purpose of infectious disease management [23]. This is consistent with prior evidence of privacy concerns and different facets of technology use or behaviors to engage in privacy protection behaviors [13,24]. In Singapore, after the government announced that police would have access to TraceTogether data for criminal investigations and after correcting perceptions that the data were only collected for contact tracing and no other purpose, some members of the public expressed concerns and disappointment and indicated their intention to use it less [25].

Perceived Benefits

In this study, perceived benefits were conceptualized as the health benefits that individuals would receive by using contact tracing apps, such as protecting them from COVID-19, making them more informed, or improving the overall public health capacity in managing and keeping infections low [12]. Hence, the use of contact tracing during the COVID-19 pandemic addresses a common problem by providing a crucial tool for communal health management, where the benefits extend beyond individual interests.

Although the loss of personal privacy is a major concern for different populations during the use of COVID-19 contact tracing apps, some have recognized the potential benefits of digital surveillance. First, advocates of "big data systems" would argue that tracking and mapping the spread of infectious diseases through digital means (eg, electronic health records) is more efficient than the traditional means of calling and asking people to remember their close contacts during the day and would reduce the burden on health care workers [26]. The data collected could enable public health agencies to quickly map potential hotspots in real time [27] and take decisive actions to curb the spread of COVID-19. Second, the use of contact tracing technology may help users feel empowered. Research has shown that individuals may gain psychological relief and peace of mind by using contact tracing apps as they provide a sense of empowerment and certainty—having data and knowledge about whether they were in close proximity with the infected, as they coped with the ambiguity of the pandemic by monitoring and interacting with their own health data [19]. A study conducted in Sweden [7] found that the benefits of contact tracing apps were positively associated with willingness to use them. They found that the perceived prosocial usefulness of contact tracing apps was positively associated with willingness to use contact tracing apps. Research has also underscored the prevalence of the privacy paradox that, despite concerns over privacy loss, people are willing to accept and use apps as long as they are perceived to be beneficial in lowering infection rates [7].

Trust in Government and Sense of Community as Antecedents of Privacy Calculus in TraceTogether

Overview

Communication privacy management theory, built upon the privacy calculus model, uses the metaphor of "boundaries" to explain the motivation to reveal or withhold information [28]. According to this theory, boundary openness or closure for information flow is largely dependent on how individuals perceive institutional privacy assurances. In other words, the cognitive processes of risk control assessment and privacy concerns are context specific. Thus, privacy-relevant beliefs are "better related to individuals' own information experiences and social contexts rather than regarded as a global consequence of technology use per se" [16]. This approach aligns with the notion that privacy is a dynamic process [29,30]. This study also investigated the antecedents that shape privacy calculus in the context of digital contact tracing.

Specifically, we posit that the service provider (ie, government) of the tracing system and other community members who use the device to collectively fight against COVID-19 would be the key entities that would influence individual users' intentions to use a contact tracing device. This study examines how (1) trust in government and (2) the sense of community triggered by the use of contact tracing technology are associated with one's privacy calculus regarding digital contact tracing, thereby shaping one's intention to adopt digital tracing technology.

Trust in Government

Social exchange theory suggests that users disclose personal information to gain intangible benefits from a relationship, given that these perceived benefits outweigh the perceived risks [31]. The benefits of a social exchange might differ from an economic exchange in that the former may rely on existing social ties and involve intangible values, including emotions and social power [31]. Similar to other contact tracing programs, TraceTogether is government-initiated. Therefore, user acceptance and adoption are deeply linked to Singaporeans' trust and confidence in their government [26].

Trust is established when individuals or groups harboring positive perceptions of one another enable the relationship to achieve anticipated results [32]. When an individual places trust in another person, group, or organization, they find themselves liberated from the burden of anxiety and obligation to constantly monitor the actions of the other party [33]. Thus, it can be expected that people who use the contact tracing system will perceive social exchange as more reliable if they trust the government. Supporting this notion, prior studies on web-based privacy have demonstrated that trust in web-based service providers or vendors significantly influences privacy management decisions by decreasing privacy concerns [8,9]. A study in Singapore [34] found that political trust could mediate Singaporeans' privacy concerns, and those who exhibited trust in their government tended to have more positive attitudes toward digital contact tracing technology. This finding is echoed in multiple studies on contact tracing apps in countries including France [35], Japan [36], Germany [37], and the United Kingdom [38].

Trust in service providers can also be critical to increasing users' perceived disclosure benefits [39]. Building on the concept of privacy calculus [40], one study [41] argues that gaining users' willingness to disclose personal information requires the exchange to be based on a fair social contract. In other words, users find the benefits more attractive when they perceive the provider as trustworthy, leading to a positive bias in their privacy calculus. Taken together, we hypothesized that the effects of trust in the government on privacy calculus in TraceTogether would be as follows:

> *H1: Trust in government increases the perceived benefit of using contact tracing apps.*
>
> *H2: Trust in government will decrease privacy concerns when using contact tracing apps.*

Sense of Community Triggered by Contact Tracing

The effectiveness of the contact tracing program for infectious disease prevention largely depends on the participation rate of individuals in the community. Moreover, recognizing the collective action nature of contact tracing underscores its significance in navigating the social dilemma posed by the virus, emphasizing the shared responsibility in curbing the spread for the greater well-being of the community. To enhance the uptake rate of TraceTogether, one way in which the government has promoted the application to Singaporeans is by framing digital contact tracing as a community effort and responsibility. The TraceTogether interface included various elements designed to encourage feelings of closeness and connectedness with other users. For example, the TraceTogether app indicates the number of activated devices nearby and shows how many Bluetooth signals are exchanged among them. The language used in an official promotional video for TraceTogether also reinforced this, depicting the tool as being for "'community-driven' contact tracing suggesting that this is a 'grassroots' mechanism of fighting the virus" [42]. Therefore, group membership and collective benefit have been leveraged to encourage user adoption and acceptance of TraceTogether.

However, the contact tracing program connects those in the vicinity via Bluetooth technology to estimate the proximity and duration of encounters, which may also trigger privacy concerns. Communication privacy management theory [28] suggests that privacy management is not just about deciding to disclose or withdraw personal information with others (personal boundaries). When one's personal information is shared, those who can access it become co-owners of the information, suggesting that privacy management also involves comanaging collective boundaries. Thus, the disruption of the synchronized management of collective privacy boundaries among co-owners is a significant source of privacy risk perception. Although individual participants in the contact tracing program cannot access others' information directly, contact tracing implies that people encountered may be notified of one's infection. Moreover, signaling connectivity with others through interface cues may trigger privacy risk heuristics, heightening privacy concerns among users [43].

Collectively, we expect that the perception of "the greater good" for the improved safety of loved ones and the community is a significant factor influencing the privacy calculus of being part of the contact tracing program [44,45]. This study hypothesizes a paradoxical function of the sense of community such that the sense of community triggered by digital contact tracing will enhance perceived benefits while simultaneously presenting privacy concerns at the same time:

> *H3: The sense of community triggered by contact tracing apps increases the perceived benefits of contact tracing apps.*
>
> *H4: The sense of community triggered by contact tracing apps will increase the privacy concerns of contact tracing apps.*

This study also examined the interplay between one's perception of the government and the community on privacy calculus in digital contact tracing. Although trust in government has developed over time, a sense of community can be promoted through various design and marketing elements of digital contact tracing. As discussed above, TraceTogether was promoted to Singaporeans by framing contact tracing as a community effort and responsibility [46]. The user interface of the application was also designed to imbue users with a sense of community through various design elements. Thus, the study attempts to test whether one's trust in government factors in the privacy calculus in contact tracing would vary as a function of the larger community that they perceive through participation in the collective contact tracing effort. However, given the absence of literature documenting the moderating role of sense of

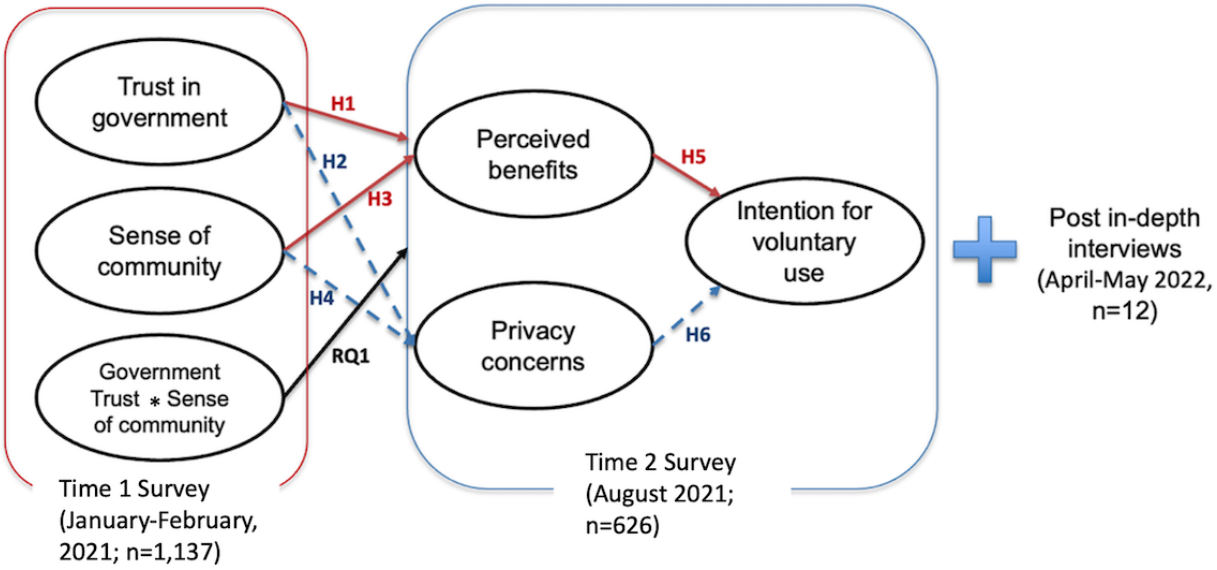
community in the relationship between trust in government and privacy calculus, this study investigates the interplay between sense of community and trust in government, proposing the following research question:

> *RQ1. Will the sense of community generated by contact tracing apps moderate the influence of trust in government on privacy concerns and perceived benefits of using contact tracing apps?*

Finally, based on the basic tenet of privacy calculus theory discussed earlier, we also hypothesize that privacy concerns and perceived benefits will have opposite associations with the intention to use TraceTogether. Given that the use of a contact tracing system itself implies the disclosure of personal information, we tested the intention to use TraceTogether as the outcome of the privacy calculus. Checking through TraceTogether was mandatory to enter most venues, including shopping malls, offices, workplaces, and schools, at the time of the second wave of data collection when use intention had been measured. Therefore, the measurement was designed to assess their intention to use TraceTogether voluntarily until the pandemic ended, rather than merely assessing their intention to use the device for a certain duration. The study model and methodology are illustrated in Figure 1.

> *H5: Perceived benefits are positively associated with the intention to actively participate in contact tracing programs.*
>
> *H6: Privacy concerns are negatively associated with the intention to actively participate in the contact tracing programs.*

Figure 1. Summary of the study model and methodology. H: hypothesis; RQ: research question.



Study 1: Two–Wave Longitudinal Survey

Methods

Recruitment

To test our hypotheses and research questions, we used a mixed method approach involving a 2-wave survey and in-depth interviews. The first survey was conducted from January to February 2021 with panel members provided by a local research company. English-speaking Singaporean citizens or permanent residents aged 21 years or above participated in the survey (N=1198). The quota sampling procedure was employed to ensure that the study sample represented the proportion of the ethnic structure of the national population [47]. The first study was conducted in an omnibus survey format to understand the various perceptions and behaviors related to COVID-19 and contact tracing practices. Among all participants who completed the survey, 94.91% (n=1137) responded that they had downloaded the TraceTogether mobile app or collected the token.

A follow-up second-wave survey was carried out in August 2021, approximately 6 months after the first survey. This survey assessed privacy calculus factors: privacy concerns, perceived benefits, and adoption intention. A second-wave survey invitation was sent to those who completed the first survey. In total, 674 participants completed the second-wave survey (response rate: 59.3%). Data from those who either downloaded the TraceTogether mobile app or received the token and who completed both studies at the point of the first wave of data collection were included in the study sample (n=626; see Table 1 for the descriptive statistics of the study sample). Our sample comprised 88.7% (n=555) Chinese, 4.3% (n=27) Malay, and 7% (n=44) Indian individuals. In comparison to the ethnic structure of the national population in Singapore, where Chinese people constitute 74%, Malay people 13%, and Indian people 9% [48], our sample overrepresented the Chinese group and underrepresented the Malay group. Moreover, the sample included only a small portion of participants aged older than 65 years (n=25; 4%), thereby underrepresenting the older population. During the first wave of data collection (January-February 2021), the TraceTogether app or token could be used for checking at selected venues where people were likely to be in close contact for prolonged periods or where high human traffic is expected. These venues included schools, educational institutes, shopping malls, restaurants, and workplaces. People can still choose to scan the QR code

displayed at the entrance instead of using TraceTogether. From June 2021, the use of TraceTogether became mandatory when entering these venues.

Table 1. Descriptive statistics.

Variables	Values (N=626)
Gender, n (%)	
Male	332 (53)
Female	294 (47)
Age group (y), n (%)	
21-34	130 (20.8)
35-49	281 (44.9)
50-64	190 (30.3)
65 or older	25 (4)
Mean age (y), mean (SD)	44.9 (11)
Ethnicity, n (%)	
Chinese	555 (88.7)
Malay	27 (4.3)
Indian	44 (7)
Education level, n (%)	
Secondary and below	13 (2)
Preuniversity	270 (43.1)
University degree or above	343 (54.8)
Yearly household income (Singaporean $[a]), n (%)	
<2000	37 (5.9)
2000-6000	185 (29.6)
6000-10,000	241 (38.5)
≥10,000	163 (26)
Self-assessed health status, n (%)	
Poor	18 (3)
Fair	163 (26)
Good	293 (46.8)
Very good	126 (20.1)
Excellent	26 (4)
TraceTogether adoption	
Have downloaded TraceTogether app (Yes)	Wave 1: 494 (78.9); wave 2: 586 (93.6)
Have collected TraceTogether token (Yes)	Wave 1: 452 (72.2); wave 2: 549 (87.7)

[a]Singaporean $1=US $0.75.

Measurement

The two exogenous variables (government trust, sense of community) and control variables (demographic information and TraceTogether use) were measured in the first-wave survey. Privacy calculus factors—privacy concerns and perceived benefits—and TraceTogether use intention were measured in the second-wave survey.

Trust in government was measured using 3 items adapted from Wu et al [49]. The *sense of community* was measured using 5 items adapted from 2 studies [50,51]. The sense of community captured the extent to which users felt a sense of belonging and emotional ties with other TraceTogether users. *Privacy concerns* were assessed using 10-item measures which captured 3 aspects of privacy concerns, namely perceived surveillance, perceived intrusion, and secondary use of personal information [16]. The *perceived benefits* of using TraceTogether were measured with 5 items that assessed the extent to which the participant perceived the health and information benefits they received from using TraceTogether [52]. Finally, *use intention* was

measured using three items. Checking TraceTogether was mandatory to enter most public venues. Therefore, the measurement assessed their intention to use TraceTogether voluntarily until the pandemic ended, rather than asking about their intention to use the device within a certain period. All items used a 7-point Likert scale (1=strongly disagree to 7=strongly agree) unless otherwise indicated. Table 2 lists the measurement items and their factor loadings.

Table 2. Measurement items and factor loadings.

Variable and items	Values, mean (SD)	Factor loading, β
Trust in government	4.80 (1.53)	
Even if not monitored, I would trust the government to do the right thing.		.91
I trust the government to protect my personal information.		.96
I believe that the government is trustworthy.		.95
Sense of community (sentences complete the phrase "When I use TraceTogether..."	4.51 (1.62)	
It makes me feel like I am part of a community.		.89
I feel a high sense of attachment with other users.		.94
I feel an emotional connection with other users.		.93
It reminds me of the people around me.		.90
It makes me feel a sense of belonging.		.94
Privacy concerns	4.71 (1.50)	
I believe that my location is being monitored in real time when using TraceTogether.		.83
I am concerned that TraceTogether is collecting too much information about me.		.88
I am concerned that TraceTogether may monitor my activities.		.91
I am concerned that my activities are being monitored.		.91
I feel that as a result of using TraceTogether, others know more about me than I am comfortable with.		.88
I believe that as a result of using TraceTogether, information about me that I consider private is now more readily available to others than I prefer.		.90
I feel that as a result of using TraceTogether, information about me is out there that, if used, will invade my privacy.		.90
I am concerned that TraceTogether may use my personal information for other purposes without notifying me or getting my authorization.		.90
I am concerned that TraceTogether may use my information for other purposes.		.87
I am concerned that TraceTogether may share my personal information with other entities without getting my authorization.		.87
Perceived benefits	5 (1.39)	
Using TraceTogether would improve my access to my health information related to COVID-19.		.86
Using TraceTogether would improve my access to my health information related to COVID-19.		.87
Using TraceTogether would improve my ability to manage my health.		.93
Using the COVID app would improve the quality of health care.		.91
I would manage my health more effectively using TraceTogether.		.90
TraceTogether use intention	5.23 (1.40)	
I will use TraceTogether even if TraceTogether becomes optional when entering public venues.		.86
I am willing to use TraceTogether until the COVID pandemic ends.		.74
I plan to use TraceTogether even if I am not monitored.		.86

Ethical Considerations

The study procedure and questionnaire were reviewed and approved by Nanyang Technological University Institutional Review Board (IRB-2022-213). A consent form was displayed on the first page of the survey, and the participants were required to provide their consent to complete the study.

Results

Measurement Model

Confirmatory factor analysis was conducted using Mplus (version 8.8; Muthen & Muthen) statistical software with a maximum likelihood estimation method. The confirmatory factor analysis results indicated a reasonable fit to data: χ^2_{286}=937.9 (*P*<.001), root mean square error of approximation=0.06 (90% CI 0.056-0.065), comparative fit index=0.97, and Tucker-Lewis index=0.96). The indicators reflected their respective latent variables, as evidenced by the high factor loadings. The magnitude of all factor loadings and Cronbach α were equal to or greater than 0.74 and 0.86, respectively. The measurement model showed robust convergent and discriminant validity: composite reliability and average variance extracted (AVE) were greater than 0.70 and 0.50, respectively. The square roots of the AVEs of all observed variables were greater than the intercorrelations between variables. All factors were significantly correlated with each other. See Table 3 for composite reliability, AVE, and Cronbach α, and see Table 4 for the correlation matrix.

Table 3. Validity and reliability.

	CR[a]	AVE[b]	Cronbach α
Government trust	0.958	0.885	0.958
Sense of community	0.964	0.844	0.964
Privacy concerns	0.973	0.783	0.974
Perceived benefits	0.952	0.798	0.952
Behavior intention	0.862	0.676	0.859

[a]CR: composite reliability.

[b]AVE: average variance extracted.

Table 4. Correlation matrix (Pearson r).

	Government trust	Sense of community	Privacy concerns	Perceived benefits
Sense of community				
r	0.63	—[a]	—	—
P value	<.001	—	—	—
Privacy concerns				
r	−0.27	−0.17	—	—
P value	<.001	<.001	—	—
Perceived benefits				
r	0.48	0.54	−0.17	—
P value	<.001	<.001	<.001	—
Behavior intention				
r	0.44	0.47	−0.21	0.63
P value	<.001	<.001	<.001	<.001

[a]Not applicable.

Model Testing

The structural equation modeling (SEM) result without the interaction term in the model revealed a good fit: χ^2_{562}= 1382.9 (*P*<.001), root mean square error of approximation=0.048 (90% CI 0.045-0.052), comparative fit index=0.957, and Tucker-Lewis index=0.952. As these model fit indices have not been developed for the SEM models assessing latent interaction effects, we used a log-likelihood (−2 LL) ratio test to compare the SEM model where the interaction was not estimated (parsimonious model) with the model where the interaction was estimated (more complex model) [53]. The test was statistically significant, meaning that the model without the interaction term represented a significant loss of fit relative to the more complex model. Thus, we conclude that the model with the interaction term is well-fitted. In addition, the model with the interaction term had lower values of the 2 fit indices, Akaike information criteria and Bayesian information criteria, in comparison to the model without the interaction term.

Our analysis revealed that trust in the government significantly reduced privacy concerns of TraceTogether (β=−.39; *P*<.001), whereas sense of community did not have a significant impact on privacy concerns (β=.02; *P*=.86). However, both government trust and sense of community significantly increased the perceived benefits of using TraceTogether (government trust: β=0.28; *P*<.001; sense of community: β=.54; *P*<.001). H1, H2

and H3 were supported, while H 4 was not supported. Regarding RQ1, the results indicated that trust in the government and sense of community interacted only with perceived benefits (β=−.14; *P*=.004). The results are summarized in Figure 2. The interaction pattern (Figure 3) shows that the triggering sense of community has a larger impact on those who display lower trust in the government than on those with high trust. This finding suggests that triggering a sense of community can buffer the negative impact of skepticism in the government on their perceptions of the benefits they can obtain using TraceTogether.

Figure 2. Summary of structural equation model results. *P<.05, **P<.01, ***P<.001. Age, gender, ethnicity, education, income, and use frequency of TraceTogether use measured at time 1 were controlled.

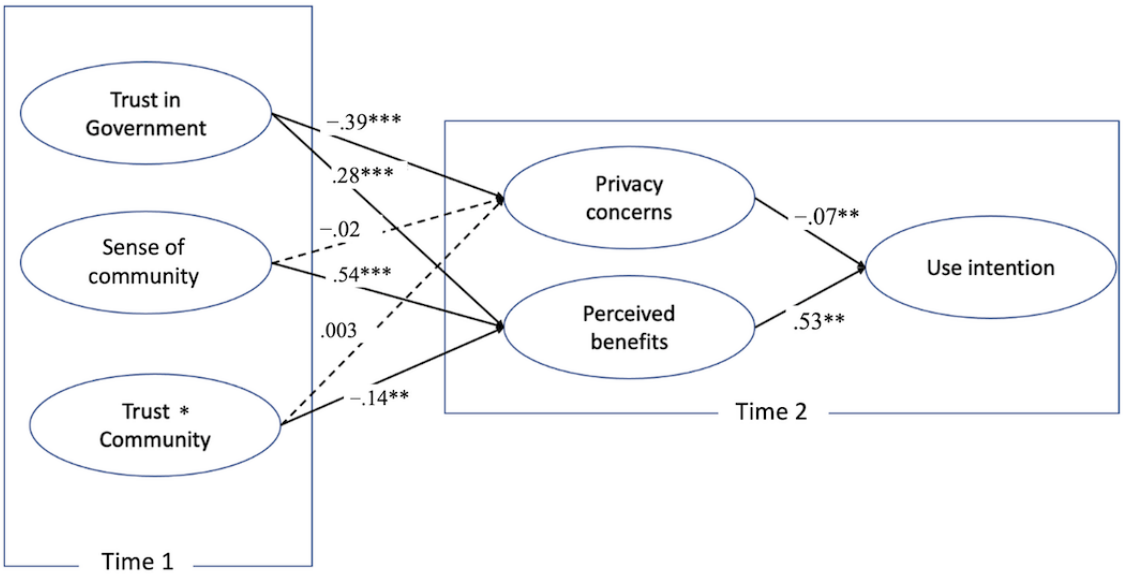
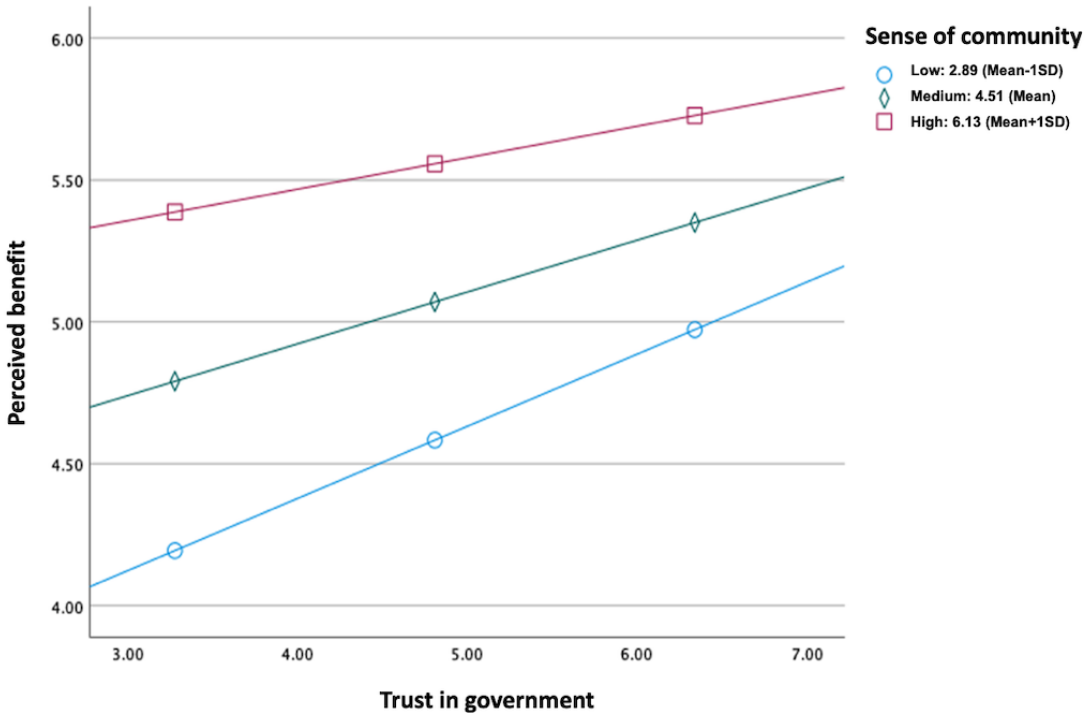


Figure 3. Interaction between the trust in government and sense of community on perceived benefits.



Finally, supporting the privacy calculus hypothesis, privacy concerns decreased TraceTogether use intention (β=−.07; *P*=.04) whereas perceived benefit increased use intention (β=.53; *P*<.001). Hence, H5 and H6 were supported.

Study 2: After In-Depth Interviews

Overview

The survey results illustrated the effects of trust in the government and a sense of community on TraceTogether privacy calculus and, ultimately, TraceTogether users' intentions. However, with the survey data, we could not identify which aspects of privacy concerns were factored in the privacy

calculus, influencing their intention to use TraceTogether. In addition, given the significant roles of trust in government and sense of community in TraceTogether and privacy calculus identified by the survey results, we conducted in-depth interviews to further understand the underlying reasons for these effects. The purpose of this study was two-fold: 1) to understand the reasons underlying the effects of government and sense of community on the privacy calculus of TraceTogether use, and 2) to obtain a deeper understanding of different aspects of privacy concerns among TraceTogether users and the implications for their privacy calculus.

Methods

Participants and Procedure

To obtain additional insights into the privacy calculus of TraceTogether, we conducted in-depth interviews with 12 Singaporeans between April 13 and May 11, 2022. We recruited Singaporeans (mean age 38, SD 13.28; 7 female participants, 5 male participants; see Table 5) who used TraceTogether during the COVID-19 pandemic. Among them, 7 were Chinese, 3 Malay, and 2 Indian. A trained research assistant conducted the interviews using a web-based conference program (ie, Zoom) in English. Participants were given a link to the consent form before starting the interview, which began once their consent was received. Each interview followed a preconstructed list of questions to understand their perceptions and use of TraceTogether. After completing the 4 interviews, the TraceTogether check-in was no longer used, starting from April 26. Subsequently, for the remaining 8 interviews, the participants were asked to answer questions reflecting their prior experiences with TraceTogether. Interview recruitment was stopped when we reached a state of theoretical saturation [54].

Table 5. Demographics of in-depth interview participants.

ID	Gender	Ethnicity	Age (y)	Occupation
1	Male	Indian	24	Student
2	Male	Chinese	23	Student
3	Female	Chinese	23	Student
4	Female	Chinese	49	Manager
5	Male	Chinese	56	Teacher
6	Female	Chinese	49	Teacher
7	Female	Malay	50	Teacher
8	Female	Malay	31	Teacher
9	Male	Malay	31	Clinician
10	Male	Chinese	51	Worker in the Singapore armed forces
11	Female	Chinese	49	Administrator
12	Female	Indian	24	Social media and project manager

Data Analysis

We used the 3-step protocol [55], which provides a structured framework for qualitative researchers to systematically analyze and derive meaningful interpretations from their data, involving 3 stages of data analysis: open-coding, axial and hierarchical coding, and theoretical interpretation, outlined as follows: First, 2 research assistants, including the interviewer, reviewed and coded the transcribed interviews independently. They read the transcribed interviews line by line and assigned the initial codes according to the themes of the original conversations (eg, concerns about using TraceTogether and adoption barriers). The second step was the axial and hierarchical coding stage; 2 researchers identified the second-level interpretive themes based on the theoretical framework and hypotheses (eg, surveillance, intrusion, and secondary use of data) by analyzing and grouping the initial coding performed by the 2 research assistants. Prior literature on web-based privacy calculus and privacy issues of digital contact tracing guided this coding stage. Finally, the researchers discussed how to interpret the themes that surfaced in the second stage in order to answer the research questions.

Ethical Considerations

Participants were given a link to the consent form before starting the interview, which began once their consent was received. The consent form, interview questions, and protocols were reviewed and approved by the primary investigator's institute (IRB-2022-213).

Results

Overview

In a scenario where policies on and implementation of a contact tracing system were rolled out urgently and comprehensively, privacy became a significant barrier to adopting the program. Our study revealed that trust in the government serves as a social lubricant, effectively mitigating privacy concerns arising from the inherent risks associated with the tracing program, thereby facilitating smoother implementation of the program. Our interview results indicated that trust in the government translated into the belief that the government would possess adequate security measures to safeguard citizens' privacy and prevent the misuse of their information. However, considering other community members while participating in the contact tracing program does not necessarily alleviate privacy concerns. The

interviews also highlighted surveillance as the most prevalent aspect of privacy concerns regarding TraceTogether use. In addition, our findings revealed that trust in the government, particularly the perceived transparency of government actions, was most related to concerns regarding the secondary use of data. The detailed results, accompanied by key supporting quotes, are outlined in the following sections.

The Role of Trust in Government in TraceTogether Privacy Calculus

The interview responses also suggested a significant influence of trust in the government on both perceived benefits and privacy concerns. Consistent with the findings of our web-based survey, the results indicated that individuals who trust the government were more likely to perceive TraceTogether as beneficial while exhibiting lower privacy concerns. For example, participant 7 (female, age 50 y) did not understand why the app should track user locations; therefore, she felt that she did not want to use it. However, she explained that she downloaded the app because she trusted the government, stating, “I trust that my government knows what they’re doing to the people. I trust that it is for the good of us.” Conversely, participant 12 (female, age 24 y), who indicated distrust in the government, did not perceive any additional benefits of using TraceTogether beyond using QR codes for checking-in, stating, “I feel like using the QR code would still work.” Furthermore, she voiced several privacy concerns, remarking, “What if this thing is tracking me 24/7? What if I am at home has my constant location at home, and my address is being transmitted?” She also referenced the public backlash at the start of 2021, when Foreign Minister Vivian Balakrishnan revealed that the police could access TraceTogether data for criminal investigations. This announcement contradicted his statement in June of the previous year that TraceTogether data would only be used for contact tracing [56]. She mentioned that she felt betrayed by government. Regarding the same incident, however, participant 5 (male, age 56 y) felt indifferent, stating: “There’s nothing to hide, so I am not concerned. Those who are criminals, they should be afraid. Good for them. They are better afraid.” His indifference was his strong trust in the government. He added, “I think the government is smart enough not to use it against themselves, because if it blows out of proportion, I tell you, the government will pay a high price. I don’t think they try to do that.”

The Role of Sense of Community in TraceTogether Privacy Calculus

Our survey data revealed that a sense of community when using TraceTogether significantly increased the perceived benefit but did not influence their privacy concerns. The interview respondents also understood that collective efforts were required to make the digital contact tracing system successful. For example, participant 1 (male, age 24 y) mentioned, “If enough people download it and contact tracing becomes easier, then we can, in a sense, fight this disease and end it earlier instead of letting it prolong.” The responses suggest that protecting the community is a strong motivation for adopting a digital contact tracing system. However, consistent with the survey results, the interview responses indicated that a sense of community did not necessarily alleviate concerns about privacy invasion through the contact tracing system. Participant 5 (male, age 56 y) stated, “It’s for the greater good. If it’s going to protect the country from this virus, it would be good. It won’t last forever. These two years, you have had your privacy invaded, but subsequently, you are fine.”

Dissecting the Multiple Aspects of Privacy Concerns in TraceTogether

As discussed earlier, privacy concerns are not a single-faceted concept, but encompass multiple aspects, including perceived surveillance, perceived intrusion, and secondary use of data [16]. Through the postinterviews, we attempted to understand the different aspects of privacy concerns among TraceTogether users.

Surveillance stood out as the most prominent privacy concern, as mentioned by 7 of 12 respondents. Respondents explained that lack of control contributes to their fear and anxiety about being watched, even if they feel that they have nothing to hide. Some respondents preferred using a TraceTogether token to the mobile app to reclaim a sense of control over where and when their locations were traced. For example, participant 7 (female, age 50 y) mentioned, “I am not afraid because I don’t have anything to hide and that I am doing anything wrong or I am a criminal... It’s just that... I don’t like to know that my personal particulars are being shared.” Participant 4 (female, age 49 y) also shared, “I can put it [token] at home...They cannot even trace me to home. They can trace the data on my phone.” These deliberate choices can be understood as small acts of resistance to mandated surveillance.

The respondents also shared that the digital contact tracing system, in fact, collected more information beyond what was necessary, such as how participant 4 (female, age 49 y) expressed her concerns. “Government is going to trace me where I go, what I do, who I meet. I feel very invaded by my privacy.” From the responses, we discovered that perceived intrusion becomes prominent because they do not know what information is to be tracked and used and the purpose of tracking clearly. Participant 2 (male, age 23 y) said, “I never know what they use the information for... They want to check whether we stay at home. It’s still not right because it’s an invasion of privacy.” Participant 4 (female, age 49 y) also raised concerns that TraceTogether would gather information beyond what has been told: “[if] I go to the shopping mall, and then they start to want to know how much you spend and say how much you spend in that shopping mall or what exact activities are being conducted.”

In addition, several respondents cited distrust in the government as a factor that increased privacy concerns, specifically with regard to the secondary use of data. Regarding the government’s announcement that the police could use TraceTogether data for criminal investigations, participant 3 (female, age 23 y) expressed her concerns, saying, “I thought it was a bit annoying because there was a lack of transparency about whatever they were going to use the app. If they wanted to do it for this stuff, they should have said it earlier and not let it be like, ‘Oops, I said it as an accident.’” The government’s transparency was revealed as an important factor in determining users’ privacy concerns about the secondary use of data. Transparency on how contact tracing apps work can not only affect user confidence,

but also determine ethical considerations. Participant 2 (male, age 23 y) shared his point, stating, "If they ask for your consent and you decide to give it (tracked information), then sure. If they do it on their own, then that's not okay." Participant 1 (male, age 24 y) shared a similar point, explaining: "It's just that I'm perfectly fine with them in using this app for in terms of national security or intelligence or what, but what I want them to do is be upfront and open."

Discussion

Theoretical Implications

Using a mixed method approach involving a 2-wave survey and in-depth interview data, we expanded our understanding of privacy decisions and calculus in the context of digital contact tracing. Our study has several notable findings. First, we found that privacy concerns had a marginal negative association with use intention, whereas perceived benefits had a strong positive association, consistent with previous research [57,58]. The opposite influences of privacy concerns and perceived benefit on use intention suggest that the privacy calculus in TraceTogether might be viewed as a rational process of weighing privacy risks and use benefits to make an uptake decision. Second, although our findings on the relationships between trust in government and sense of community and privacy calculus are largely consistent with existing research [59], our study demonstrated that their existing perceptions toward the provider and the government in the contact tracing context, as well as the perception of the community triggered by TraceTogether use, may bias user appraisals of privacy risks and benefits of contact tracing. Finally, our research extends the existing knowledge of privacy calculus by delineating how the sense of community moderated the relationship between perceived benefit and trust in government, such that individuals' sense of community elevated perceived benefit, even for those with low trust in government. The study results provides several important theoretical implications: First, the study illuminated how people's trust in the system provider plays a role in privacy calculus in digital contact tracing systems. The digital contact tracing system, including TraceTogether, is a special form of a web-based system built and managed by government organizations during an unprecedented pandemic. Digital contact tracing entails delegating the right to track and process location, contact, and health-related information to the government's discretion. Therefore, we posit that trust in government is a powerful factor in determining social exchange decisions.

As expected, trust in the government significantly influenced privacy calculus at the later stage when the use of TraceTogether became mandatory for entering most venues and workplaces. The results support prior research, which suggests the importance of trust in the government for the successful implementation of the contact tracing program for infectious disease prevention (eg, [32-36]). Trust in the government becomes pivotal in government-citizen relationships, especially during the implementation of new policies. This is largely because trust can reduce the transaction costs in relationships [60]. Our interview responses offer deeper insights into how trust in the government lowers transaction costs, specifically addressing privacy risks in the context of participating in the contact tracing program. We found that one's privacy calculus can be biased according to the level of trust displayed by individuals, and an interviewee showing distrust in the government tends to exaggerate the privacy risk of TraceTogether while doubting its effectiveness without specific evidence. Conversely, another interviewee, who started to use TraceTogether without knowing its purpose of tracking one's location, shared that she downloaded the app simply because she trusted the government, and thus trusted that the app was good for them.

Our results also revealed that the sense of community elevated by TraceTogether use positively influenced users' benefit appraisal, even though it did not reduce privacy concerns. This finding is in line with research [61], which found that in largely collectivistic societies (eg, China), there is a greater acceptance of contact tracing apps compared with more individualistic societies (eg, the United States, Germany), and people are generally willing to accept these technologies, especially if they are considered effective. In our study, the results might be because Singaporeans accept the spirit of the "greater good" underlying the contact tracing measure in the pandemic situation in the community, supporting the extant literature. A prior study compared Singapore and Switzerland and attributed the former's higher rates of acceptance and adoption to its norms that prioritize the "interests of the community" [46]. Interestingly, privacy concerns regarding the use of TraceTogether persisted even when a sense of community was fostered. The follow-up interviews revealed these findings. For instance, our interviews indicated that considering fellow community members while participating in the contact tracing program motivated individuals to endure the discomfort of sacrificing privacy for the greater good rather than directly alleviating the concerns themselves. This aligns with prior studies on the role of sense of community in coping behaviors; they showed that a stronger sense of community motivates individuals to focus on problem-solving coping behaviors [62,63]. By extending these findings to our study, it can be inferred that strengthening the sense of community compels individuals to concentrate on addressing the larger community issue while accepting potential risks associated with participating in the program to tackle this issue.

Our postinterview responses provided an in-depth understanding of privacy calculus in TraceTogether use by dissecting different aspects of privacy concerns [16] that TraceTogether users exhibited. The interviewees responded that they felt they were being surveyed through TraceTogether, which often required excessive personal information. However, interviewees were generally willing to bear surveillance and intrusion for the greater good of the community and their own safety. However, the results indicated that having a sense of control over information is an important factor in the privacy calculus in TraceTogether. Most interviewees also called for greater transparency in data collection and secondary use of data. They actively sought strategies to better control their personal information. For example, some participants indicated that they prefer to use the token rather than the app so that they can take the token only when TraceTogether checking-in is required.

Hence, our study results support the notion that privacy management is a dynamic process involving "*selective control* of access to the self or to one's group" [64].

Another notable finding of our study was how the sense of community could elevate the benefits of contact tracing despite low trust in the government. This is important considering that norms-based approaches [65] are effective in modifying health behaviors in diverse contexts. Existing research has identified several moderating factors that influence users' intention to use contact tracing apps. For instance, demographic factors, such as gender, racial and ethnic identity, and education level, are significant moderators of contact tracing app design on people's intention to install apps [66]. Separately, another study found that self-efficacy was a significant moderator of trust and intention to use contact tracing apps [67]. Although these findings are important for technology developers, they view technology use and acceptance through a largely technocentric and individualistic approach, even though public health scholars [68] have argued for the need to consider societal factors as driving factors of adoption, as evidenced by research [69] on how norms could drive app use, even among marginalized communities.

Practical Implications

Our study offers valuable insights into the policy and design considerations concerning digital contact tracing programs. The interview results revealed that, although Singaporeans are willing to accept a certain level of surveillance for preventive measures via contact tracing, issues arise when data use lacks transparency, potentially causing individuals to reconsider their positive intentions to participate in the greater good. Moreover, because digital contact tracing was made mandatory, they also felt that they had no control over their information, which may overturn their privacy calculus, and thus opted out from participating in the program if it became optional. The results suggest that for the successful implementation of the contact tracing program, it is crucial to implement policy measures to ensure clear and transparent communication regarding the use of data collected from contact tracing practices. In addition, providing as many control options as possible would encourage users to maintain active participation in the contact tracing program.

The moderating effect of a sense of community prompted by contact tracing has valuable practical implications. The interaction result indicated that the negative impact of distrust in government on contact tracing program participation can be overcome by promoting the "social good" nature of the contact tracing program via campaign or application design. According to our study results, the campaign for TraceTogether that framed contact tracing as a community effort and responsibility, with the campaign slogan being "Protect your community," might have enhanced community participation, especially among those with lower trust in the government. If, unfortunately, another pandemic emerges in the future, and a similar program is needed, the contact tracing program can actively incorporate social- and community-related features. For example, in the early stage of implementation, the application can be promoted through social networks, allowing users to invite and encourage their family and friends to opt for the program.

Additionally, the contact tracing app interface can be designed to cue a sense of community when using it. Illustrations of the family and community for the application interface design, as TraceTogether has done, would be an effective strategy to enhance participation in the program. TraceTogether also displays the number of Bluetooth signals shared by others. However, such information can be framed as emphasizing collective efforts for the greater good of the community. For example, information can be shown through a visual illustration of people forming networks.

Limitation and Future Research Direction

This study had some limitations. Although the longitudinal survey design allowed us to establish the temporal sequencing of the variables of interest, the dependent variable (ie, behavioral intention) was assessed in the follow-up survey without obtaining the baseline measurement. Hence, we cannot fully leverage the longitudinal nature of the data to establish the direction of the association or causal relationships between the variables. Future studies may use longitudinal data with repeated measures of the variables (two or more time points) to address them. Additionally, the data provide limited potential for generalizability. Survey participants with a higher socioeconomic status (eg, education) were overrepresented. In addition, the use of self-reported data in surveys may pose the risk of social desirability and recall biases. Finally, we acknowledge that the mandatory use of TraceTogether between waves 1 and 2 of our survey might have an impact on the use intention of participants, as existing research [70] on a Singapore sample showed high adoption due to the mandatory use of the app for entering public venues. To circumvent this, we modified our items such that they captured participants' volitional use in contexts when TraceTogether became voluntary, rather than capturing actual use behavior as the dependent variable due to social desirability bias. Despite the above limitations, we are confident that this study has presented valid evidence of the relationships among trust, privacy concerns, perceived benefits, and TraceTogether use intentions.

Conflicts of Interest

None declared.

References

1. Coronavirus disease (COVID-19): contact tracing. World Health Organization. URL: https://www.who.int/news-room/questions-and-answers/item/coronavirus-disease-covid-19-contact-tracing [accessed 2022-08-31]

2. González-Cabañas J, Cuevas Á, Cuevas R, Maier M. Digital contact tracing: large-scale geolocation data as an alternative to bluetooth-based apps failure. Electronics 2021 May 05;10(9):1093. [doi: 10.3390/electronics10091093]
3. Min-Allah N, Alahmed BA, Albreek EM, Alghamdi LS, Alawad DA, Alharbi AS, et al. A survey of COVID-19 contact-tracing apps. Comput Biol Med 2021 Oct;137:104787 [FREE Full text] [doi: 10.1016/j.compbiomed.2021.104787] [Medline: 34482197]
4. Kamel Boulos MN, Geraghty EM. Geographical tracking and mapping of coronavirus disease COVID-19/severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) epidemic and associated events around the world: how 21st century GIS technologies are supporting the global fight against outbreaks and epidemics. Int J Health Geogr 2020 Mar 11;19(1):8 [FREE Full text] [doi: 10.1186/s12942-020-00202-8] [Medline: 32160889]
5. Mozur P, Zhong R, Krolik A. In coronavirus fight, China gives citizens a color code, with red flags. The New York Times. 2020 Mar 1. URL: https://www.seattletimes.com/business/questions-over-smart-phone-app-that-chinas-using-in-coronavirus-fight/ [accessed 2024-01-29]
6. Fox G, van der Werff L, Rosati P, Takako Endo P, Lynn T. Examining the determinants of acceptance and use of mobile contact tracing applications in Brazil: an extended privacy calculus perspective. J Assoc Inf Sci Technol 2021 Nov 10;73(7):944-967. [doi: 10.1002/asi.24602]
7. Carlsson Hauff J, Nilsson J. Individual costs and societal benefits: the privacy calculus of contact-tracing apps. JCM 2021 Sep 07;40(2):171-180. [doi: 10.1108/jcm-03-2021-4559]
8. Taddei S, Contena B. Privacy, trust and control: which relationships with online self-disclosure? Comput Human Behav 2013 May;29(3):821-826. [doi: 10.1016/j.chb.2012.11.022]
9. Zimmer JC, Arsal RE, Al-Marzouq M, Grover V. Investigating online information disclosure: effects of information relevance, trust and risk. Inf Manag 2010 Mar;47(2):115-123. [doi: 10.1016/j.im.2009.12.003]
10. Dinev T, Hart P. An extended privacy calculus model for e-commerce transactions. Inf Syst Res 2006 Mar;17(1):61-80. [doi: 10.1287/isre.1060.0080]
11. Kang H, Oh J. Communication privacy management for smart speaker use: integrating the role of privacy self-efficacy and the multidimensional view. New Media Soc 2021 Jun 30;25(5):1153-1175. [doi: 10.1177/14614448211026611]
12. Fox G, Clohessy T, van der Werff L, Rosati P, Lynn T. Exploring the competing influences of privacy concerns and positive beliefs on citizen acceptance of contact tracing mobile applications. Comput Human Behav 2021 Aug;121:106806. [doi: 10.1016/j.chb.2021.106806]
13. Ho SS, Lwin MO, Yee AZ, Lee EW. Understanding factors associated with Singaporean adolescents' intention to adopt privacy protection behavior using an extended theory of planned behavior. Cyberpsychol Behav Soc Netw 2017 Sep;20(9):572-579. [doi: 10.1089/cyber.2017.0061] [Medline: 28876967]
14. Krasnova H, Veltri NF, Günther O. Self-disclosure and privacy calculus on social networking sites: the role of culture. Bus Inf Syst Eng 2012 Apr 26;4(3):127-135. [doi: 10.1007/s12599-012-0216-6]
15. Kwee-Meier ST, Bützler JE, Schlick C. Development and validation of a technology acceptance model for safety-enhancing, wearable locating systems. Behav Inf Technol 2016 Feb 25;35(5):394-409. [doi: 10.1080/0144929x.2016.1141986]
16. Xu H, Gupta S, Rosson MB, Carroll JM. Measuring mobile users' concerns for information privacy. In: Proceedings of the 2012 International Conference on Information Systems. 2012 Presented at: ICIS '12; December 16-19, 2012; Orlando, FL p. 2278-2293.
17. Malhotra NK, Kim SS, Agarwal J. Internet users' information privacy concerns (IUIPC): the construct, the scale, and a causal model. Inf Syst Res 2004 Dec;15(4):336-355. [doi: 10.1287/isre.1040.0032]
18. Solove DJ. A taxonomy of privacy. Univ PA Law Rev 2006;154(3):477. [doi: 10.2307/40041279]
19. Baik J(, Jang E. Where horizontal and vertical surveillances meet: sense-making of US COVID-19 contact-tracing apps during a health crisis. Mob Media Commun 2022 Feb 11;10(3):468-486. [doi: 10.1177/20501579221078674]
20. Zhang X. Decoding China's COVID-19 health code apps: the legal challenges. Healthcare (Basel) 2022 Aug 05;10(8):1479 [FREE Full text] [doi: 10.3390/healthcare10081479] [Medline: 36011136]
21. Cong W. From pandemic control to data-driven governance: the case of China’s health code. Front Polit Sci 2021 Apr 14;3:627959. [doi: 10.3389/fpos.2021.627959]
22. Smith HJ, Milberg SJ, Burke SJ. Information privacy: measuring individuals' concerns about organizational practices. MIS Q 1996 Jun;20(2):167. [doi: 10.2307/249477]
23. Samuel G, Roberts SL, Fiske A, Lucivero F, McLennan S, Phillips A, et al. COVID-19 contact tracing apps: UK public perceptions. Crit Public Health 2022 Jan 01;32(1):31-43 [FREE Full text] [doi: 10.1080/09581596.2021.1909707] [Medline: 35221546]
24. Mohamed N, Ahmad IH. Information privacy concerns, antecedents and privacy measure use in social networking sites: evidence from Malaysia. Comput Human Behav 2012 Nov;28(6):2366-2375. [doi: 10.1016/j.chb.2012.07.008]
25. Elangvan N, Tan YL. Some TraceTogether users upset with Govt’s revelation on police access to data, say they’ll use it less. Today. 2021. URL: https://www.todayonline.com/singapore/woman-pleads-guilty-fail-report-rape-unconscious-friend-open-bank-accounts-scammers-2354456 [accessed 2024-01-29]

26. Zhou X, Lee EW, Wang X, Lin L, Xuan Z, Wu D, et al. Infectious diseases prevention and control using an integrated health big data system in China. BMC Infect Dis 2022 Apr 06;22(1):344 [FREE Full text] [doi: 10.1186/s12879-022-07316-3] [Medline: 35387590]
27. Lee EW, Yee AZ. Toward data sense-making in digital health communication research: why theory matters in the age of big data. Front Commun 2020 Feb 27;5 [FREE Full text] [doi: 10.3389/fcomm.2020.00011]
28. Petronio S, Altman I. Boundaries of Privacy: Dialectics of Disclosure. New York, NY: State University of New York Press; 2002.
29. Ahern S, Eckles D, Good N, King S, Naaman M, Nair R. Over-exposed?: privacy patterns and considerations in online and mobile photo sharing. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2007 Presented at: CHI '07; April 28-May 3, 2007; San Jose, CA p. 357-366 URL: https://dl.acm.org/doi/10.1145/1240624.1240683 [doi: 10.1145/1240624.1240683]
30. Joinson A, Reips UD, Buchanan T, Schofield CB. Privacy, trust, and self-disclosure online. Hum Comput Interact 2010 Jan;25(1):1-24 [FREE Full text] [doi: 10.1080/07370020903586662]
31. Chellappa RK, Sin RG. Personalization versus privacy: an empirical examination of the online consumer's dilemma. Inf Technol Manage 2005 Apr;6(2-3):181-202. [doi: 10.1007/s10799-005-5879-y]
32. Wheeless LR, Grotz J. The measurement of trust and its relationship to self-disclosure. Hum Commun Res 1977 Mar;3(3):250-257 [FREE Full text] [doi: 10.1111/j.1468-2958.1977.tb00523.x]
33. Levi M, Stoker L. Political trust and trustworthiness. Annu Rev Polit Sci (Palo Alto) 2000 Jun;3(1):475-507. [doi: 10.1146/annurev.polisci.3.1.475]
34. Wang Y, Ngien A, Ahmed S. Nationwide adoption of a digital contact tracing app: examining the role of privacy concern, political trust, and technology literacy. Commun Stud 2022 Jul 01;73(4):364-379. [doi: 10.1080/10510974.2022.2094982]
35. Guillon M, Kergall P. Attitudes and opinions on quarantine and support for a contact-tracing application in France during the COVID-19 outbreak. Public Health 2020 Nov;188:21-31 [FREE Full text] [doi: 10.1016/j.puhe.2020.08.026] [Medline: 33059232]
36. He Y, Yatsuya H, Ota A, Tabuchi T. The association of public trust with the utilization of digital contact tracing for COVID-19 in Japan. Public Health Pract (Oxf) 2022 Dec;4:100279 [FREE Full text] [doi: 10.1016/j.puhip.2022.100279] [Medline: 35719273]
37. Kozyreva A, Lorenz-Spreen P, Lewandowsky S, Garrett PM, Herzog SM, Pachur T, et al. Psychological factors shaping public responses to COVID-19 digital contact tracing technologies in Germany. Sci Rep 2021 Sep 21;11(1):18716 [FREE Full text] [doi: 10.1038/s41598-021-98249-5] [Medline: 34548550]
38. Lewandowsky S, Dennis S, Perfors A, Kashima Y, White JP, Garrett P, et al. Public acceptance of privacy-encroaching policies to address the COVID-19 pandemic in the United Kingdom. PLoS One 2021 Jan 22;16(1):e0245740 [FREE Full text] [doi: 10.1371/journal.pone.0245740] [Medline: 33481877]
39. Metzger MJ. Effects of site, vendor, and consumer characteristics on web site trust and disclosure. Commun Res 2016 Jun 30;33(3):155-179. [doi: 10.1177/0093650206287076]
40. Laufer RS, Wolfe M. Privacy as a concept and a social issue: a multidimensional developmental theory. J Soc Issues 2010 Apr 14;33(3):22-42. [doi: 10.1111/j.1540-4560.1977.tb01880.x]
41. Culnan MJ, Bies RJ. Consumer privacy: balancing economic and justice considerations. J Soc Issues 2003 Apr 29;59(2):323-342. [doi: 10.1111/1540-4560.00067]
42. Stevens H, Haines MB. TraceTogether: pandemic response, democracy, and technology. East Asian Sci Technol Soc 2020 Nov 23;14(3):523-532. [doi: 10.1215/18752160-8698301]
43. Sundar S, Kang H, Wu M, Go E, Zhang B. Unlocking the privacy paradox: do cognitive heuristics hold the key? In: Proceedings of the CHI '13 Extended Abstracts on Human Factors in Computing Systems. 2013 Presented at: CHI EA '13; April 27-May 2, 2013; Paris, France p. 811-816 URL: https://dl.acm.org/doi/10.1145/2468356.2468501 [doi: 10.1145/2468356.2468501]
44. Redmiles EM. User concerns and tradeoffs in technology-facilitated COVID-19 response. Digit Gov Res Pract 2020 Nov 17;2(1):1-12 [FREE Full text] [doi: 10.1145/3428093]
45. Williams SN, Armitage CJ, Tampe T, Dienes K. Public attitudes towards COVID-19 contact tracing apps: a UK-based focus group study. Health Expect 2021 Apr 12;24(2):377-385 [FREE Full text] [doi: 10.1111/hex.13179] [Medline: 33434404]
46. Geber S, Ho SS. Examining the cultural dimension of contact-tracing app adoption during the COVID-19 pandemic: a cross-country study in Singapore and Switzerland. Inf Commun Soc 2022 Jun 08;26(11):2229-2249. [doi: 10.1080/1369118x.2022.2082880]
47. Population in brief 2020. Strategy Group. 2020. URL: https://www.strategygroup.gov.sg/files/media-centre/publications/population-in-brief-2020.pdf [accessed 2022-08-31]
48. Demographics of Singapore in 2023 - structure, labor, regional trends. OOSGA. URL: https://oosga.com/demographics/sgp/#:~:text=Executive%20Summary,%2C%20and%20Indians%20(9%25) [accessed 2024-01-29]
49. Wu KW, Huang SY, Yen DC, Popova I. The effect of online privacy policy on consumer privacy concern and trust. Comput Human Behav 2012 May;28(3):889-897. [doi: 10.1016/j.chb.2011.12.008]

50. McMillan DW, Chavis DM. Sense of community: a definition and theory. J Community Psychol 1986;14(1):6-23 [FREE Full text] [doi: 10.1002/1520-6629(198601)14:1<6::aid-jcop2290140103>3.0.co;2-i]
51. Kalyanaraman S, Sundar SS. The psychological appeal of personalized content in web portals: does customization affect attitudes and behavior? J Commun 2006 Mar;56(1):110-132. [doi: 10.1111/j.1460-2466.2006.00006.x]
52. Li H, Gupta A, Zhang J, Sarathy R. Examining the decision to use standalone personal health record systems as a trust-enabled fair social contract. Decis. Support Syst 2014 Jan;57:376-386. [doi: 10.1016/j.dss.2012.10.043]
53. Maslowsky J, Jager J, Hemken D. Estimating and interpreting latent variable interactions: a tutorial for applying the latent moderated structural equations method. Int J Behav Dev 2015 Jan 01;39(1):87-96 [FREE Full text] [doi: 10.1177/0165025414552301] [Medline: 26478643]
54. Morse JM. The significance of saturation. Qual Health Res 2016 Jul 01;5(2):147-149. [doi: 10.1177/104973239500500201]
55. Tracy SJ. Qualitative Research Methods: Collecting Evidence, Crafting Analysis, Communicating Impact. 2nd edition. Hoboken, NJ: John Wiley & Sons; 2013.
56. Chee K. Vivian Balakrishnan says he 'deeply regrets' mistake on TraceTogether data. The Straits Times. 2021 Feb 2. URL: https://www.straitstimes.com/singapore/vivian-balakrishnan-says-he-deeply-regrets-mistake-on-tracetogether-data-first-realised-it [accessed 2024-01-29]
57. Walrave M, Waeterloos C, Ponnet K. Adoption of a contact tracing app for containing COVID-19: a health belief model approach. JMIR Public Health Surveill 2020 Sep 01;6(3):e20572 [FREE Full text] [doi: 10.2196/20572] [Medline: 32755882]
58. Shahidi N, Tossan V, Bourliataux-Lajoinie S, Cacho-Elizondo S. Behavioural intention to use a contact tracing application: the case of StopCovid in France. J Retail Consum Serv 2022 Sep;68:102998 [FREE Full text] [doi: 10.1016/j.jretconser.2022.102998]
59. Akintoye S, Ogoh G, Krokida Z, Nnadi J, Eke D. Understanding the perceptions of UK COVID-19 contact tracing app in the BAME community in Leicester. J Inf Commun Ethics Soc 2021 Dec 07;19(4):521-536. [doi: 10.1108/jices-06-2021-0071]
60. Fukuyama F. Trust: The Social Virtues and the Creation of Prosperity. Florence, MA: Free Press; 1996.
61. Kostka G, Habich-Sobiegalla S. In times of crisis: public perceptions toward COVID-19 contact tracing apps in China, Germany, and the United States. New Media Soc 2022 Apr 03:146144482210832. [doi: 10.1177/14614448221083285]
62. Bachrach KM, Zautra AJ. Coping with a community stressor: the threat of a hazardous waste facility. J Health Soc Behav 1985 Jun;26(2):127. [doi: 10.2307/2136602]
63. Wandersman A, Giamartino GA. Community and individual difference characteristics as influences on initial participation. American J of Comm Psychol 1980 Apr;8(2):217-228. [doi: 10.1007/bf00912661]
64. Altman I. The Environment and Social Behavior: Privacy, Personal Space, Territory, Crowding. Pacific Grove, CA: Brooks/Cole Publishing Company; 1975.
65. Rimal RN. Modeling the relationship between descriptive norms and behaviors: a test and extension of the theory of normative social behavior (TNSB). Health Commun 2008;23(2):103-116. [doi: 10.1080/10410230801967791] [Medline: 18443998]
66. Li T, Cobb C, Yang J, Baviskar S, Agarwal Y, Li B, et al. What makes people install a COVID-19 contact-tracing app? Understanding the influence of app design and individual difference on contact-tracing app adoption intention. Pervasive Mob Comput 2021 Aug;75:101439 [FREE Full text] [doi: 10.1016/j.pmcj.2021.101439] [Medline: 36569467]
67. Ukpabi D, Olaleye S, Karjaluoto H. Factors influencing tourists' intention to use COVID-19 contact tracing app. In: Wörndl W, Koo C, Stienmetz JL, editors. Information and Communication Technologies in Tourism 2021. Cham, Switzerland: Springer; 2021:504-516.
68. Bao H, Lee EW. Examining theoretical frameworks and antecedents of health apps and wearables use: a scoping review. Health Commun 2023 Nov 15:1-11. [doi: 10.1080/10410236.2023.2283655] [Medline: 37968803]
69. Bao H, Lee EW. Examining the antecedents and health outcomes of health apps and wearables use: an integration of the technology acceptance model and communication inequality. Behav Inf Technol 2023 Feb 23:1-22 [FREE Full text] [doi: 10.1080/0144929x.2023.2183062]
70. Chow BW, Lim YD, Poh RC, Ko A, Hong GH, Zou SW, et al. Use of a digital contact tracing system in Singapore to mitigate COVID-19 spread. BMC Public Health 2023 Nov 16;23(1):2253 [FREE Full text] [doi: 10.1186/s12889-023-17150-0] [Medline: 37974135]

Abbreviations

AVE: average variance extracted
SEM: structural equation modeling

Edited by L Buis; submitted 14.05.23; peer-reviewed by R Rahimi, M Dzandu, L Kojan; comments to author 16.10.23; revised version received 12.12.23; accepted 25.01.24; published 07.03.24.

Please cite as:
Kang H, Lee JK, Lee EWJ, Toh C
The Roles of Trust in Government and Sense of Community in the COVID-19 Contact Tracing Privacy Calculus: Mixed Method Study Using a 2-Wave Survey and In-Depth Interviews
JMIR Mhealth Uhealth 2024;12:e48986
URL: https://mhealth.jmir.org/2024/1/e48986
doi:10.2196/48986
PMID:38451602

Original Paper

Medical Information Protection in Internet Hospital Apps in China: Scale Development and Content Analysis

Jiayi Jiang[1], PhD; Zexing Zheng[1], PhD

Law School, Central South University, Changsha, China

Corresponding Author:
Zexing Zheng, PhD
Law School
Central South University
Number 932 Lushan South Road
Yuelu District
Changsha, 410083
China
Phone: 86 18718479720
Email: xing986@126.com

Abstract

Background: Hospital apps are increasingly being adopted in many countries, especially since the start of the COVID-19 pandemic. Web-based hospitals can provide valuable medical services and enhanced accessibility. However, increasing concerns about personal information (PI) and strict legal compliance requirements necessitate privacy assessments for these platforms. Guided by the theory of contextual integrity, this study investigates the regulatory compliance of privacy policies for internet hospital apps in the mainland of China.

Objective: In this paper, we aim to evaluate the regulatory compliance of privacy policies of internet hospital apps in the mainland of China and offer recommendations for improvement.

Methods: We obtained 59 internet hospital apps on November 7, 2023, and reviewed 52 privacy policies available between November 8 and 23, 2023. We developed a 3-level indicator scale based on the information processing activities, as stipulated in relevant regulations. The scale comprised 7 level-1 indicators, 26 level-2 indicators, and 70 level-3 indicators.

Results: The mean compliance score of the 52 assessed apps was 73/100 (SD 22.4%), revealing a varied spectrum of compliance. Sensitive PI protection compliance (mean 73.9%, SD 24.2%) lagged behind general PI protection (mean 90.4%, SD 14.7%), with only 12 apps requiring separate consent for processing sensitive PI (mean 73.9%, SD 24.2%). Although most apps (n=41, 79%) committed to supervising subcontractors, only a quarter (n=13, 25%) required users' explicit consent for subcontracting activities. Concerning PI storage security (mean 71.2%, SD 29.3%) and incident management (mean 71.8%, SD 36.6%), half of the assessed apps (n=27, 52%) committed to bear corresponding legal responsibility, whereas fewer than half (n=24, 46%) specified the security level obtained. Most privacy policies stated the PI retention period (n=40, 77%) and instances of PI deletion or anonymization (n=41, 79%), but fewer (n=20, 38.5%) committed to prompt third-party PI deletion. Most apps delineated various individual rights, but only a fraction addressed the rights to obtain copies (n=22, 42%) or to refuse advertisement based on automated decision-making (n=13, 25%). Significant deficiencies remained in regular compliance audits (mean 11.5%, SD 37.8%), impact assessments (mean 13.5%, SD 15.2%), and PI officer disclosure (mean 48.1%, SD 49.3%).

Conclusions: Our analysis revealed both strengths and significant shortcomings in the compliance of internet hospital apps' privacy policies with relevant regulations. As China continues to implement internet hospital apps, it should ensure the informed consent of users for PI processing activities, enhance compliance levels of relevant privacy policies, and fortify PI protection enforcement across the information processing stages.

(JMIR Mhealth Uhealth 2024;12:e55061) doi:10.2196/55061

KEYWORDS

hospital apps; privacy policy; personal information protection; policy evaluation; content analysis

Introduction

Background

The emergence and rapid expansion of hospital apps represents a significant milestone in the evolution of global health care services [1,2], especially during the COVID-19 pandemic [3-7]. These digital platforms provide a range of medical services, from digital consultations [8,9] to telemedicine [10] and digital care management [6,11]. Their growing use reflects a trend toward digital health solutions as enhanced, accessible, and cost-efficient health care services [12].

However, the rise of hospital apps has been accompanied by substantial concerns regarding patient privacy and data security [13-15], as with other mobile health (mHealth) applications [16,17]. The apps' extensive collection of personal health and medical information, as well as the sensitive nature of that data, suggest a need for comprehensive, rigorously enforced regulations to prevent unauthorized access, misuse, and disclosure. In regions like the United States [18,19] and the European Union, similar digital health initiatives have been developed that focus on interoperability, patient-centricity, and adherence to strict data protection regulations, such as the HIPAA (Health Insurance Portability and Accountability Act) in the United States [20,21] and the General Data Protection Regulation in the EU [22,23]. Striking a balance between leveraging the benefits of digital health services and ensuring the confidentiality and integrity of patient information remains an ongoing challenge in the industry.

In China, the response to the evolution of digital health care has been swift, aided by the prevalence of mobile internet [24] and the development of mHealth services [25-27], as evidenced by the 2014 launch of the country's first officially approved internet hospital in Guangdong province [28,29]. This milestone, coupled with the enactment of several "internet plus healthcare" policies [30], has led to a surge in digital hospital apps, bringing the terms "internet diagnosis" and "internet hospitals" into the national health care context [31-33]. "Internet diagnosis" encompasses medical services provided in digital form by registered doctors, including consultations for certain common and chronic diseases and "internet plus" family doctor services [34]. Hospital apps are divided into two categories: (1) digital extensions of traditional hospitals and (2) stand-alone entities operated by internet enterprises [34]. The former involves local doctors and patients, whereas the latter combines the resources of various medical institutions to expand service to patients across different locations.

Internet hospital apps offer digital consultation, appointment scheduling, diagnosis and treatment of common and chronic diseases, and medical guidance, as well as prescription and delivery of medications and other treatments [32,35]. These apps have significantly enhanced health care by addressing the disparities in resource distribution and access across the mainland of China's large population [26,31,36] and improved overall patient experiences by enhancing communication, transparency, and efficiency [37]. The COVID-19 pandemic further underscored the efficacy of digital health care providers, which facilitated crucial health care services for prevention and control in the pandemic's early stages [38-42].

However, the existing application of these apps presents significant challenges to patient information protection [43,44]. First, sensitive personal information (PI) generated during medical visits, such as biometric and health data, is vulnerable to unauthorized sharing and cyberattacks, which can lead to privacy breaches [45,46]. Second, the complexity of integrating and applying health data weakens individuals' control over their health information once it transforms into big health care data [47-49]. Finally, the difficulty of implementing and upholding informed consent is compounded by the lack of unified industry standards and the realities of "algorithmic black boxes," which often leave individual patients in a relatively disadvantaged position [50,51].

To manage these issues, China has established a regulatory framework to protect PI. Since 2017, the Information Security Technology-Personal Information Specification (PI Specification) has been adopted as a voluntary standard for PI protection practice by all kinds of enterprises in information processing activities [52-54]. In addition, the Personal Information Protection Law (PIPL), guided by the Chinese Civil Code [55] and effective starting November 11, 2021, serves as the nation's first comprehensive national PI legislation. The PIPL specifies the rights of individuals and the obligations of PI processors [56,57]. The Chinese government has also made a specific commitment to protect personal health information and prohibits illegal processing, trade, or disclosure of personal health information in article 92 of the Law on the Promotion of Basic Medical Care, Hygiene, and Health, enacted on June 1, 2020.

Internet hospital apps represent a critical intersection of PI and digital technology, which underscores the urgent need for scrutiny of these providers' privacy policies within a framework that balances self-regulation and governmental oversight [58,59]. Privacy policies delineate how PI processors collect, use, disclose, and manage a customer or client's data [60]. They are also the primary grounds for the transparent data processing requirements mandated by privacy-related regulations [61]. Drawing inspiration from contextual integrity (CI) theory, we also investigated how the privacy policies articulated and adhered to the norms of information flow [62]. In keeping with Nissenbaum's [63] assertion that privacy "is preserved when informational norms are respected and violated when informational norms are breached," we set the basis of evaluation with a focus on the norms and values that govern appropriate flows of PI.

Previous research on privacy compliance of mHealth apps in different countries has identified gaps between rules for privacy protection and the apps' implementations in various aspects, such as lack of complete privacy policies, lack of informed consent, and insufficient protection of sensitive data [64-73]. Such investigations have also raised concerns about internet hospital apps' uneven design quality and the challenges in minimizing users' cognitive load while ensuring information security [74,75]. However, these studies have not thoroughly

examined web-based hospitals' compliance with China's comprehensive legal framework for PI protection.

This study uses a legal framework to assess the compliance of internet hospital apps' privacy policies with China's PI-related regulations. The Methods section elaborates on the collection and selection of sample apps, describes the development of an evaluation scale based on relevant policy documents, and outlines the procedures for app assessment and scoring. In the Results section, we present the compliance scores of sample apps. The Discussion section contextualizes these results within the broader landscape of mHealth app privacy compliance, underscoring the importance of legal compliance in the evolving digital health landscape.

Objective

In this study, we aimed to (1) collect the privacy policies of internet hospital apps developed for users in the mainland of China, (2) develop a scale based on the provisions stipulated in the PIPL, PI Specification, and rules of the hospitals, (3) assess the compliance of the privacy policies within the regulatory framework of PI protection, and (4) offer recommendations for improving the legal compliance of internet hospital apps' privacy policies to enhance PI protection in the evolving landscape of mHealth innovation. This study contributes to the global discussion on balanced policies for PI protection in digital health initiatives in the postpandemic era and provides insights for policymakers, hospital app providers, and users across different countries while highlighting the importance of improving legal compliance and strengthening enforcement.

Methods

Study Design

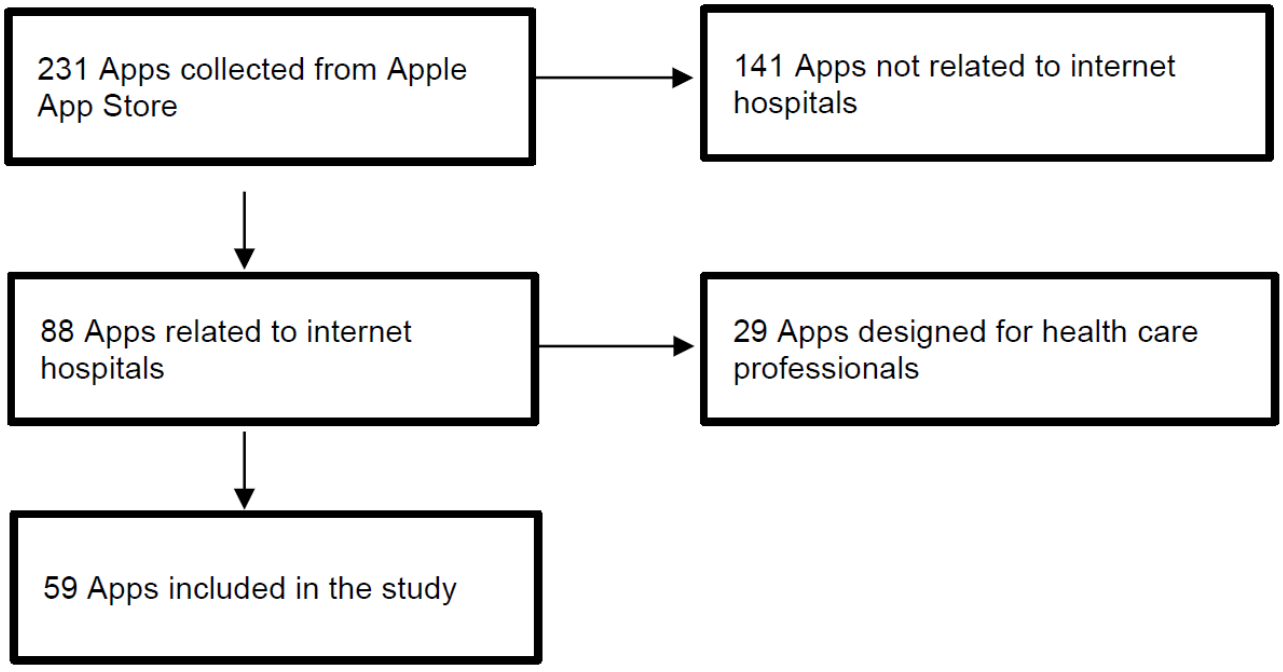
We conducted a content analysis of the privacy policies of internet hospital apps available in the Apple App Store in the mainland of China and evaluated their compliance with the PIPL, PI Specification, and hospital app rules. Drawing from CI theory, we considered the adherence of internet hospital apps' privacy policies to PI norms as essential to PI protection.

App Selection and Inclusion Criteria

In this study, we focused on the privacy policies of internet hospital apps available in the Apple App Store and tailored for the market of Chinese mainland. To identify relevant apps, we used the keyword "internet hospital" (*hu lian wang yi yuan* in Chinese) to search on Diandian (*Dian Shu Ju* in Chinese), a prominent mobile data analytics platform in China. We conducted the search on November 7, 2023.

The apps included in the sample fell under the following definitions: (1) apps or platforms specifically developed to provide a range of hospital app services, and (2) apps intended for use by the general population rather than health care professionals. Excluded apps fit the following: (1) apps designed for health care professionals managing internal hospital operations, and (2) apps with scope or functionality unrelated to hospital app services, such as those dedicated to health insurance, maintaining a healthy lifestyle, or health education and popular science. The initial search resulted in a total of 231 apps, out of which 59 met the inclusion criteria and were included in the final analysis upon review (Figure 1). We obtained and reviewed the full text of corresponding privacy policies as text files or screenshots from the sample apps between November 8 and 23, 2023.

Figure 1. Filtering procedure for the selection and inclusion of hospital apps.



Development of the Compliance Evaluation Scale

Overview

We systematically developed a compliance evaluation scale to assess the privacy policies of internet hospital apps against the PIPL, PI Specification (GB/*t* 35273-2020), and the Rules for Regulation of Internet Diagnosis and Treatment Management (for Trial Implementation). This process entailed the following sections.

Comprehensive Review

First, we obtained and meticulously reviewed the full text of the PIPL, PI Specification, and the aforementioned rules to understand the comprehensive regulatory framework governing PI protection in hospital apps.

Indicator Development

Based on the information processing activities delineated in these policy documents, we identified level-1 evaluation indicators encompassing critical processing stages such as PI collection and usage; PI storage and protection; PI sharing, transfer, disclosure, and transmission; PI deletion; individual rights; and PI processor duties. In addition, we introduced "general attributes" as an additional level-1 indicator to evaluate the overall transparency and ongoing maintenance efforts.

Indicator Elaboration

We then translated the specific chapters and clauses of these policy documents into a more granular set of 26 level-2 indicators and 70 level-3 indicators, which provided a detailed framework for our evaluation. Each level-2 indicator represented a crucial component within the respective PI processing stage—our level-1 indicators. For example, in the stage of PI collection and usage, we followed different rules for general and sensitive PI as stipulated in the PIPL, and further developed 2 level-2 indicators to evaluate the collection and usage of general PI and sensitive PI. Moreover, we established level-3 indicators to assess the specific compliance requirements as detailed under each level-2 indicator. For example, concerning the collection and usage of sensitive PI, we identified 7 level-3 indicators according to the PIPL, which included: highlighting sensitive PI, stating the specific purpose, clarifying the sufficient necessity, implementing stringent protective measures, communicating the implications of processing sensitive PI, obtaining separate explicit consent for processing sensitive PI, and requiring explicit consent for processing PI of minors.

Operational Definitions and Examples

To ensure clarity and consistency in our assessment and support the application of our evaluation criteria, we have included brief explanations, example sentences, and references to the relevant provisions of the policy documents for all the level-3 indicators in Multimedia Appendix 1.

Scoring and Evaluation Procedure

We adopted a binary scoring system for level-3 indicators, awarding a score of 1 for privacy policies that adequately addressed a given indicator and 0 for those that did not. This allowed us to calculate the compliance rate for each level-3 indicator based on the proportion of policies scoring "1" from our app sample. We then calculated the scoring rate for each level-2 indicator as the arithmetic mean of the scoring rates for its associated level-3 indicators. Similarly, we determined the compliance rate for the level-1 indicators as the mean of the scoring rates of the corresponding level-2 indicators, which reflects the overall compliance of each app in specific stages of the information processing activities. The overall compliance of each app's privacy policy was quantified by aggregating the scores of all level-3 indicators and converting this total into a percentage to denote the app's compliance level.

To ensure the reliability of our evaluation, 2 independent raters (JJ and ZZ) were engaged to assess the privacy policies of all 59 internet hospital apps between November 8 and 24, 2023. To gauge interrater reliability, both raters independently evaluated a randomly selected subset of 20 apps (34% of the total), achieving an intraclass correlation coefficient of 0.986 (P<.001), indicating nearly perfect agreement. Following this assessment, the raters convened to discuss score discrepancies in their initial evaluations. After this, the raters divided the remainder of the apps equally and applied the unified standards to ensure scoring consistency.

Results

Sample Collection

We accessed 59 internet hospital apps available in the Apple App Store for Chinese mainland users by registering as users with our own identity documents and mobile phone numbers. We obtained the full text of 52 privacy policies. A small but significant percentage of apps (7/59, 12%) altogether lacked a separate privacy policy, a fundamental requirement for safeguarding PI. This absence is a critical oversight and represents direct noncompliance with established PI protection laws, suggesting an urgent need for these apps to develop and implement comprehensive privacy policies.

Compliance Evaluation

The overall compliance landscape among the 52 assessed privacy policies was mixed. The mean compliance score of all policies was 73 of a possible 100 (SD 22.4%). Moreover, 36 apps (69%) surpassed the mean score, whereas 16 apps (31%) fell below.

The evaluation results for level-1 and level-2 indicators are listed in Figure 2 and Table 1. Level-1 indicators were ranked by score from highest to lowest, as follows: general attributes (mean 92.1%, SD 16.5%); PI collection and usage (mean 81.5%, SD 17.9%); PI sharing, transfer, disclosure, and transmission (mean 75%, SD 25.2%); PI storage and protection (mean 71.5%, SD 30.7%); individual rights (mean 68.4%, SD 31.5%); PI deletion (mean 64.7%, SD 34.8%); and PI processor duties (mean 59.4%, SD 28.4%). The names and evaluation results for each app are listed in Multimedia Appendix 2.

The privacy policies' general attributes (mean 92.1%, SD 16.5%) scored high, indicating effective efforts in transparency and maintenance. For level-2 indicators, PI processors and services recorded an impressive compliance rate of 95.2% (SD 20.2%), indicating a majority of the privacy policies effectively identified the parties responsible for processing PI and providing

services. Policy transparency was a standout area, with a perfect score of 100% (SD 0%) reflecting the apps' commitment to clear and open communication with users. Policy maintenance also emerged as a strong suit, scoring 84.6% (SD 30.3%). This suggests a significant proportion of apps were proactive in updating their privacy policies, a vital aspect of best practices following the implementation of the PIPL. Specifically, 25 apps updated their privacy policies after the PIPL came into force. However, a concerning 12 apps failed to mention either the effective or updated date of their policies, whereas 15 updated their privacy policies before the PIPL came into effect.

Regulations for a description of the collection and usage of general PI had an average compliance rate of 90.4% (SD 14.7%). This indicates the majority of internet hospital apps were conscientious in describing how general PI is collected and used within their service functions. Our evaluation found all the reviewed privacy policies specified the purpose and methods of collecting and using PI, demonstrating a high level of transparency. Additionally, a substantial 90% (n=47) of apps provided a list of the types of PI collected, while 83% (n=43) of the policies clarified the consequences of not providing PI. In terms of differentiating between essential and nonessential PI for services, compliance stood at 69% (n=36). Although significant clarity was currently provided, an opportunity still remained for apps to enhance user understanding of the purpose and optional nature of PI collection.

Meanwhile, the scoring rate of collection and usage of sensitive PI was lower (mean 73.9%, SD 24.2%). We observed strong compliance rates for describing specific purposes (n=48, 92%), protective measures (n=46, 88.5%), implications (n=43, 83%), and necessity (n=41, 79%) of processing sensitive PI. Most assessed apps required explicit consent for processing minors' PI (n=43, 83%). However, the requirement to obtain separate explicit consent for processing sensitive PI revealed a significant gap, with only 23% (n=12) of apps complying.

In the PI storage and protection stage (mean 71.5%, SD 30.7), the scoring of level-2 indicators varied slightly. The compliance rate of storage security was 71.2% (SD 29.3%). Most apps explained potential security risks (n=46, 88.5%) and organizational management measures. Fewer than half (n=24, 46%) outlined the compulsory level of technical security measures. As for security incidents (mean 71.8%, SD 36.6%), although a significant portion of apps committed to notifying users (n=43, 83%) and reporting security incidents (n=42, 81%), just over half of PI processors (n=27, 52%) committed to assuming legal responsibility in the event of such an incident.

In the stage of PI sharing, transfer, disclosure, and transmission (mean 75%, SD 25.2%), the scoring rate of level-2 indicators varied substantially. For public disclosure (mean 93.3%, SD 24.1%), we observed high compliance in specifying conditions for potential public PI disclosure (n=49, 94%) and requiring separate consent for such practices (n=48, 92%). These rates indicated a high degree of transparency and respect for user consent for public disclosure. As for the compliance rate of PI sharing and transfer (mean 77.5%, SD 30%), most privacy policies introduced information about PI recipients (n=37, 71%), the types of PI transferred (n=38, 73%), and the safety precautions adopted in advance (n=37, 71%). In addition, most apps explained the purposes (n=44, 85%) and methods (n=44, 85%) of PI transfer, described the rules governing PI transfer during specific events (n=37, 71%), and required separate consent for sharing or transferring PI (n=44, 85%). Regarding cross-border transmission (mean 71.2%, SD 43.1%), most apps specified PI storage locations (n=39, 75%), whereas fewer mentioned compliance with relevant cross-border transmission laws (n=35, 67%). However, the compliance rate of subcontracting PI processing was low (mean 51.9%, SD 27.7%). Although most apps committed to supervising the subcontracted PI processing activities (n=41, 79%), only a quarter (n=13, 25%) required separate consent for these activities.

Figure 2. Compliance evaluation scores of internet hospital apps for level-1 indicators. PI: personal information.

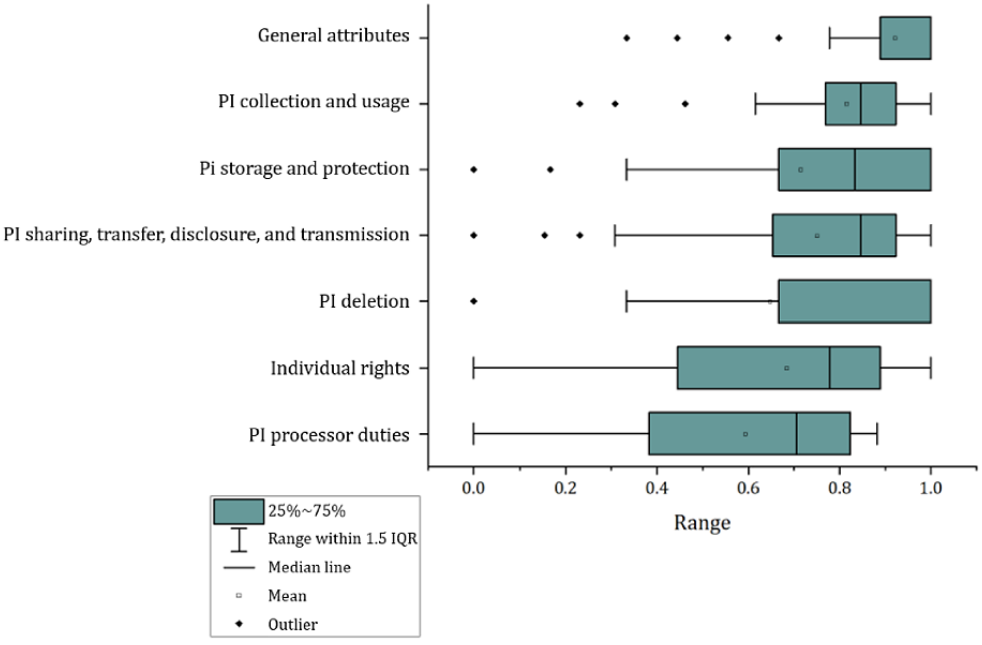


Table 1. Compliance evaluation scores of internet hospital apps for level-2 indicators.

Evaluation results on level-2 indicators	Compliance rate (%), mean (SD)
PI[a] processors and service	95.2 (20)
Policy transparency	100 (0)
Policy maintenance	84.6 (30)
Collection and use of general PI in service functions	90.4 (15)
Collection and use of sensitive PI in service functions	73.9 (24)
Storage security	71.2 (29)
Security incidents	71.8 (37)
Subcontracting of PI processing	51.9 (28)
PI sharing and transfer	77.5 (30)
Public disclosure	93.3 (24)
Cross-border transmission	71.2 (43)
Retention period	76.9 (42)
Deletion and cessation	58.7 (38)
Inquiry of PI	80.8 (39)
Obtain copies of PI	42.3 (49)
Correction of PI	80.8 (39)
Deletion of PI	80.8 (39)
Explanation regarding PI processing	82.7 (38)
Consent withdrawal	51.9 (34)
Deregistration	76.9 (42)
Consent exception scenarios	67.3 (46)
PI protection officer disclosure	48.1 (49)
Compliance audits	11.5 (38)
Impact assessment procedures	13.5 (15)
Request management	72.9 (36)
Complaint management	64.7 (40)

[a]PI: personal information.

In the stage of PI deletion (mean 64.7%, SD 34.8%), most privacy policies stated the PI retention period (mean 76.9%, SD 42.1%). In contrast, the scoring rate for deletion and cessation was lower (mean 58.7%, SD 37.6%). Although most apps committed to PI deletion or anonymization after a retention period (n=41, 79%), only 20 apps (38.5%) required third parties to delete PI or cease processing after the same period.

Concerning individual rights (mean 68.4%, SD 31.5%), most apps explained individuals' various rights effectively, including the rights to inquire about (n=42, 81%), correct (n=42, 81%), and delete PI (n=42, 81%); cancel the account (n=40, 77%); withdraw or modify consent (n=41, 79%); and request an explanation of the privacy policy (n=43, 83%). However, only 22 apps recognized the right of users to obtain copies of their PI (n=22, 42%) and only 13 explained the right to refuse business marketing using automated decision-making. A majority of apps (n=35, 67%) listed exceptions for obtaining consent as provided by applicable laws or administrative regulations.

Concerning PI processor duties, we found a compliance rate of 59.3% (SD 28.4%). Fewer than half of the apps appointed a PI officer and disclosed their information in their privacy policies (n=25, 48%). A quarter of the apps (n=13, 25%) presented impact assessment procedures, whereas 11.5% of apps (n=6) engaged in compliance audits. Many apps provided methods for individuals to inquire about (n=41, 79%), correct (n=41, 79%), and delete PI (n=41, 79%); clarify PI processing rules (n=43, 83%); cancel their account (n=41, 79%); withdraw or modify consent (n=40, 79%); and understand limits on the use of automated decision-making (n=29, 56%). However, fewer than half of all the studied apps provided methods to obtain copies of PI (n=21, 40%). Although many apps provided a means for lodging complaints by disclosing contact information (n=37, 71%), fewer committed to addressing these complaints within the stated time limits or explained the methods of dispute resolution (n=32, 61.5%).

Discussion

Principal Findings

We developed the evaluation scale to align with the characteristics of internet hospital apps, drawing from the essential CI parameters of context, attributes, actors, and transmission principles. We set the context in the realms of internet diagnosis and hospitals and categorized the attributes of PI, emphasizing the distinction between sensitive and general PI processing activities. Actors encompassed both app users and PI processors, including third-party entities outlined in the privacy policies. We translated the transmission principles of lawfulness, legitimacy, necessity, good faith, minimal impact, openness, and transparency into indicators that aligned with PI-related regulations.

Our review of 52 privacy policies from internet hospital apps in the mainland of China reveals a varied spectrum of compliance. The compliance score of the apps' privacy policies varied (mean 73%, SD 22.4%), with some apps demonstrating robust compliance, whereas others fell short. This suggests a need for enhanced regulatory oversight and standardized practices. We also identified variations in legal compliance across different stages of the information processing activities, as shown in Figure 2 and Table 1. This underscores the varying application of PI-related regulations in digital hospital apps, raising concerns about users' potential exposure to privacy risks.

First, our analysis indicates a notable gap between compliance rates for sensitive PI protection (mean 73.9%, SD 24.2%) and general PI protection (mean 90.4%, SD 14.7%), raising significant concerns regarding the provision of stringent safeguards for sensitive PI [76,77]. This gap is especially concerning given the PIPL (specifically section 2, chapter II) mandates special protection for sensitive PI. The inadequate compliance in this area also potentially diminishes users' awareness and understanding of the risks associated with the processing of their sensitive PI. Article 28 of the PIPL stipulates that PI processors may only process sensitive PI with a specified purpose, sufficient necessity, and stringent protective measures. Alarmingly, the practice of seeking explicit consent for processing sensitive PI—a fundamental requirement for lawful processing and respecting user rights—is not as widespread as it should be, implying a pervasive reliance on blanket consent strategies among digital hospital apps. These findings also suggest privacy policies often fail to provide the necessary clarity for users to understand the distinctions between various types of PI and the specific reasons for their processing. Enhancing privacy policies to offer more detailed explanations would not only align with the PIPL's mandate but also elevate the standard of user empowerment, enabling individuals to make informed decisions about their PI.

Second, the fact that many apps did not fully elucidate the role of third-party subcontractors or the conditions of PI sharing, transferring, or deletion in privacy policies may hinder users' understanding of the destinations and protections of their PI, which could consequently affect their trust and the integrity of their informed consent [78]. The lack of detailed disclosure about PI processors (including involved third parties) and protocols for PI sharing and transfer, particularly in critical scenarios like mergers or acquisitions, underscores a disconnect between regulatory intentions and the operational realities of data governance within these digital platforms. In addition, the apps' handling of PI deletion remains challenging and becomes more complex when third-party subcontracting activities are involved [79,80]. It is particularly problematic when privacy policies do not clearly communicate how these third parties are managed or if they are held to the same rigorous standards of PI protection as the primary PI processors.

Third, we found users' rights to inquire about, correct, and delete personal data, along with other user-centric controls, were generally recognized within the apps' privacy policies. The rights of individuals to manage their PI are paramount in the domain of digital health [81,82]. However, a deeper look into the specifics of these policies uncovers a gap in the acknowledgment of users' right to obtain copies of their own PI, a provision stipulated in article 45 of the PIPL. More concerning is the fact that only a quarter of the apps addressed the right to refuse business marketing through automated decision-making, even though article 24 of the PIPL calls for transparency, fairness, and the right to receive an explanation and be able to opt out of such marketing. A lack of explanation of these crucial rights might inadvertently hinder app users from fully realizing their entitlements under PI protection norms.

Finally, our findings reveal inadequacies in how internet hospital apps execute PI-related responsibilities, even though the roles and responsibilities of PI processors are central to the protection of PI [83]. The absence of clear methods for users to obtain copies of their PI or comprehensive explanations of automated decision-making processes stands in stark contrast to the protective intent of the PIPL. Moreover, the relatively low scores of PI protection officer disclosure, compliance audits, and impact assessment procedures suggest a concerning lapse in institutional oversight. Such critical mechanisms are essential for the proactive identification of vulnerabilities and agile adaptation to emergent technological threats.

Recommendations

Since the creation of China's first internet hospital, the nation's government has shown commendable support for the industry in its policy making [25,32,84,85]. China's "internet plus" policy paves the way for a promising future for internet hospital apps beyond their role in the prevention and control of the COVID-19 pandemic [41,86]. Constructing health and medical big data requires the aggregation and integration of personal health care information, so it is essential to address PI risks posed by big data technology. The public-interest nature of health and medical information in areas like infectious disease control, medical research, and public safety further underscores the importance of the reasonable use and sufficient protection of PI [87,88]. However, the rapidly growing sector of internet diagnosis and hospitals still grapples with gaps in patient information protection [43,89], necessitating a balanced approach that judiciously considers both the advantages of processing PI and the inherent challenges associated with PI protection [90].

There is a pressing need to standardize obtaining informed consent in internet hospital services. The prevalent absence of

explicit consent, particularly in subcontracting processes, raises significant privacy concerns, ranging from unauthorized data collection to inadequate user disclosure and excessive data harvesting [91-94]. Individuals often find their control over their own health information reduced, especially as it becomes integrated into big data [48,95]. Drawing on CI's focus on the principles of actors and transmission, it is essential to adopt a dynamic consent model to reinforce granular control over PI. Implementing robust privacy impact assessments and creating transparent platforms for sharing privacy policies can further enhance public trust [50].

Improving the compliance of privacy policies and their enforcement mechanisms requires adherence to CI principles across the information processing stages. This includes ensuring clarity in the collection and use of PI, enhancing protection for PI storage, and promoting transparency in the sharing, transfer, and deletion of PI. Emphasizing the internet diagnosis and hospitals' context and the attributes of PI can help in distinguishing between sensitive and general PI, ensuring PI processing activities are aligned with users' needs and rights. Expanding acknowledgment of individual rights concerning users' PI, a core aspect of CI, involves not only recognizing users' rights to access, correct, and delete their PI, but also ensuring they are adequately informed about the purposes and methods of PI processing. This can be achieved through regular audits, impact assessments, and the appointment of PI protection officers, which ensure internet hospital apps not only comply with legal standards but also respect ethical digital health practices.

Limitations

This study, while comprehensive in scope, encountered several limitations common in privacy policy analyses. First, our methodology primarily relied on content analysis of privacy policies, which may not fully capture the actual practices and implementation effectiveness of these policies. There is often a gap between what is stated in policy documents and how those policies are executed. Thus, the findings may not accurately reflect the on-the-ground application of apps' privacy standards. Second, this study was confined to the examination of publicly available privacy policies, without delving into the apps' technical backend and data-handling processes. This limitation means we could not assess the real-world effectiveness of the stated privacy measures or the security of the apps' data management systems. Future researchers could benefit from incorporating technical audits, user-experience studies, and automated analysis, which could provide a more holistic and dynamic view of privacy protection in internet hospital apps.

Conclusions

Our comprehensive evaluation of privacy policies from 52 internet hospital apps in the mainland of China highlights a landscape marked by varied compliance with relevant regulations. Despite some apps demonstrating adherence to legal standards, notable gaps persist, especially in protecting sensitive PI, obtaining informed consent, and clearly delineating individual rights. Inspired by CI theory, in this study, we underscore the urgent need for enhanced regulatory oversight, standardized privacy practices, and a commitment to user empowerment through transparent, comprehensive privacy policies. Addressing these challenges is critical, not only for protecting PI but also for fostering trust and facilitating the sustainable growth of digital health care services in China and other countries.

Acknowledgments

The research is funded by the Social Science Fund Projects of Hunan Province (23YBQ003 and 23YBQ011).

Conflicts of Interest

None declared.

Multimedia Appendix 1
This chart provides the 3-level indicator scale developed to evaluate the compliance status of the sample apps.
[XLSX File (Microsoft Excel File), 26 KB - mhealth_v12i1e55061_app1.xlsx]

Multimedia Appendix 2
This chart provides a list of 59 internet hospital apps that met the inclusion criteria.
[XLSX File (Microsoft Excel File), 18 KB - mhealth_v12i1e55061_app2.xlsx]

References

1. Mosa ASM, Yoo I, Sheets L. A systematic review of healthcare applications for smartphones. BMC Med Inform Decis Mak 2012;12:67 [FREE Full text] [doi: 10.1186/1472-6947-12-67] [Medline: 22781312]
2. Weinstein RS, Lopez AM, Joseph BA, Erps KA, Holcomb M, Barker GP, et al. Telemedicine, telehealth, and mobile health applications that work: opportunities and barriers. Am J Med 2014;127(3):183-187. [doi: 10.1016/j.amjmed.2013.09.032] [Medline: 24384059]
3. Lee SM, Lee D. Opportunities and challenges for contactless healthcare services in the post-COVID-19 era. Technol Forecast Soc Change 2021;167:120712 [FREE Full text] [doi: 10.1016/j.techfore.2021.120712] [Medline: 33654330]

4. Kondylakis H, Katehakis DG, Kouroubali A, Logothetidis F, Triantafyllidis A, Kalamaras I, et al. COVID-19 mobile apps: a systematic review of the literature. J Med Internet Res 2020;22(12):e23170 [FREE Full text] [doi: 10.2196/23170] [Medline: 33197234]
5. Bokolo AJ. Application of telemedicine and eHealth technology for clinical services in response to COVID-19 pandemic. Health Technol (Berl) 2021;11(2):359-366 [FREE Full text] [doi: 10.1007/s12553-020-00516-4] [Medline: 33469474]
6. Sharma S, Kumari B, Ali A, Yadav RK, Sharma AK, Sharma KK, et al. Mobile technology: a tool for healthcare and a boon in pandemic. J Family Med Prim Care 2022;11(1):37-43 [FREE Full text] [doi: 10.4103/jfmpc.jfmpc_1114_21] [Medline: 35309626]
7. Sitammagari K, Murphy S, Kowalkowski M, Chou SH, Sullivan M, Taylor S, et al. Insights from rapid deployment of a "Virtual Hospital" as standard care during the COVID-19 pandemic. Ann Intern Med 2021;174(2):192-199 [FREE Full text] [doi: 10.7326/M20-4076] [Medline: 33175567]
8. Greenhalgh T, Vijayaraghavan S, Wherton J, Shaw S, Byrne E, Campbell-Richards D, et al. Virtual online consultations: advantages and limitations (VOCAL) study. BMJ Open 2016;6(1):e009388 [FREE Full text] [doi: 10.1136/bmjopen-2015-009388] [Medline: 26826147]
9. Car J, Koh GCH, Foong PS, Wang CJ. Video consultations in primary and specialist care during the covid-19 pandemic and beyond. BMJ 2020;371:m3945. [doi: 10.1136/bmj.m3945] [Medline: 33082127]
10. Leite H, Hodgkinson IR, Gruber T. New development: 'healing at a distance'—telemedicine and COVID-19. Public Money Manage 2020;40(6):483-485. [doi: 10.1080/09540962.2020.1748855]
11. Bokolo Anthony Jnr. Use of telemedicine and virtual care for remote treatment in response to COVID-19 pandemic. J Med Syst 2020;44(7):132 [FREE Full text] [doi: 10.1007/s10916-020-01596-5] [Medline: 32542571]
12. Haleem A, Javaid M, Singh RP, Suman R. Telemedicine for healthcare: capabilities, features, barriers, and applications. Sens Int 2021;2:100117 [FREE Full text] [doi: 10.1016/j.sintl.2021.100117] [Medline: 34806053]
13. Rowland SP, Fitzgerald JE, Holme T, Powell J, McGregor A. What is the clinical value of mHealth for patients? NPJ Digit Med 2020;3(1):4 [FREE Full text] [doi: 10.1038/s41746-019-0206-x] [Medline: 31970289]
14. Lupton D, Jutel A. 'It's like having a physician in your pocket!' a critical analysis of self-diagnosis smartphone apps. Soc Sci Med 2015;133:128-135. [doi: 10.1016/j.socscimed.2015.04.004] [Medline: 25864149]
15. Flaherty JL. Digital diagnosis: privacy and the regulation of mobile phone health applications. Am J Law Med 2014;40(4):416-441. [Medline: 27530051]
16. Tangari G, Ikram M, Ijaz K, Kaafar MA, Berkovsky S. Mobile health and privacy: cross sectional study. BMJ 2021;373:n1248 [FREE Full text] [doi: 10.1136/bmj.n1248] [Medline: 34135009]
17. Martínez-Pérez B, de LTI, López-Coronado M. Privacy and security in mobile health apps: a review and recommendations. J Med Syst 2015;39(1):181. [doi: 10.1007/s10916-014-0181-3] [Medline: 25486895]
18. Yu SWY, Hill C, Ricks ML, Bennet J, Oriol NE. The scope and impact of mobile health clinics in the United States: a literature review. Int J Equity Health 2017;16(1):178 [FREE Full text] [doi: 10.1186/s12939-017-0671-2] [Medline: 28982362]
19. Kichloo A, Albosta M, Dettloff K, Wani F, El-Amir Z, Singh J, et al. Telemedicine, the current COVID-19 pandemic and the future: a narrative review and perspectives moving forward in the USA. Fam Med Community Health 2020;8(3):e000530 [FREE Full text] [doi: 10.1136/fmch-2020-000530] [Medline: 32816942]
20. Glenn T, Monteith S. Privacy in the digital world: medical and health data outside of HIPAA protections. Curr Psychiatry Rep 2014;16(11):494. [doi: 10.1007/s11920-014-0494-4] [Medline: 25218603]
21. Soumitra SB, Hyunmin K, Oluwaseyi OI, Kumar N, Bhatt J, Wyant DK, et al. Privacy and security issues in mobile health: current research and future directions. Health Policy Technol 2017;6(2):188-191 [FREE Full text] [doi: 10.1016/j.hlpt.2017.01.004]
22. Yuan B, Li J. The policy effect of the General Data Protection Regulation (GDPR) on the digital public health sector in the European Union: an empirical investigation. Int J Environ Res Public Health 2019;16(6):1070 [FREE Full text] [doi: 10.3390/ijerph16061070] [Medline: 30934648]
23. Rumbold JMM, Pierscionek B. The effect of the general data protection regulation on medical research. J Med Internet Res 2017;19(2):e47 [FREE Full text] [doi: 10.2196/jmir.7108] [Medline: 28235748]
24. The 52nd statistical report on China's internet development. China Internet Network Information Center. 2023. URL: https://www.cnnic.com.cn/IDR/ReportDownloads/202311/P020231121355042476714.pdf [accessed 2024-05-29]
25. Hsu J, Liu D, Yu YM, Zhao HT, Chen ZR, Li J, et al. The top Chinese mobile health apps: a systematic investigation. J Med Internet Res 2016;18(8):e222 [FREE Full text] [doi: 10.2196/jmir.5955] [Medline: 27573724]
26. Lv Q, Jiang Y, Qi J, Zhang Y, Zhang X, Fang L, et al. Using mobile apps for health management: a new health care mode in China. JMIR mHealth uHealth 2019;7(6):e10299 [FREE Full text] [doi: 10.2196/10299] [Medline: 31162131]
27. Xie Z, Nacioglu A, Or C. Prevalence, demographic correlates, and perceived impacts of mobile health app use amongst Chinese adults: cross-sectional survey study. JMIR mHealth uHealth 2018;6(4):e103 [FREE Full text] [doi: 10.2196/mhealth.9002] [Medline: 29699971]
28. Tu J, Wang C, Wu S. The internet hospital: an emerging innovation in China. Lancet Glob Health 2015;3(8):e445-e446 [FREE Full text] [doi: 10.1016/S2214-109X(15)00042-X] [Medline: 26187488]

29. Wu D, Hesketh T, Shu H, Lian W, Tang W, Tian J. Description of an online hospital platform, China. Bull World Health Organ 2019;97(8):578-579 [FREE Full text] [doi: 10.2471/BLT.18.226936] [Medline: 31384077]
30. Yang F, Shu H, Zhang X. Understanding "internet plus healthcare" in China: policy text analysis. J Med Internet Res 2021;23(7):e23779 [FREE Full text] [doi: 10.2196/23779] [Medline: 34309581]
31. Wang Y, Liu Y, Shi Y, Yu Y, Yang J. User perceptions of virtual hospital apps in China: systematic search. JMIR mHealth uHealth 2020;8(8):e19487 [FREE Full text] [doi: 10.2196/19487] [Medline: 32687480]
32. Han Y, Lie RK, Guo R. The internet hospital as a Telehealth Model in China: systematic search and content analysis. J Med Internet Res 2020;22(7):e17995 [FREE Full text] [doi: 10.2196/17995] [Medline: 32723721]
33. Wang H, Liang L, Du C, Wu Y. Implementation of online hospitals and factors influencing the adoption of mobile medical services in China: cross-sectional survey study. JMIR mHealth uHealth 2021;9(2):e25960 [FREE Full text] [doi: 10.2196/25960] [Medline: 33444155]
34. National HC, National AFTCM. Circular on issuing Administrative Measures for Internet Diagnosis (for Trial Implementation) and other policy documents. Guo Wei Yi Fa (2018) no25. Beijing. 2018 Jul 17. URL: https://www.gov.cn/gongbao/content/2019/content_5358684.htm [accessed 2024-05-30]
35. Xie X, Zhou W, Lin L, Fan S, Lin F, Wang L, et al. Internet hospitals in China: cross-sectional survey. J Med Internet Res 2017;19(7):e239 [FREE Full text] [doi: 10.2196/jmir.7854] [Medline: 28676472]
36. Cui F, Ma Q, He X, Zhai Y, Zhao J, Chen B, et al. Implementation and application of telemedicine in China: cross-sectional study. JMIR mHealth uHealth 2020;8(10):e18426 [FREE Full text] [doi: 10.2196/18426] [Medline: 33095175]
37. Lu C, Hu Y, Xie J, Fu Q, Leigh I, Governor S, et al. The use of mobile health applications to improve patient experience: cross-sectional study in Chinese public hospitals. JMIR mHealth uHealth 2018;6(5):e126 [FREE Full text] [doi: 10.2196/mhealth.9145] [Medline: 29792290]
38. Yan A, Zou Y, Mirchandani DA. How hospitals in mainland China responded to the outbreak of COVID-19 using information technology-enabled services: an analysis of hospital news webpages. J Am Med Inform Assoc 2020;27(7):991-999 [FREE Full text] [doi: 10.1093/jamia/ocaa064] [Medline: 32311036]
39. Liu J, Liu S, Zheng T, Bi Y. Physicians' perspectives of telemedicine during the COVID-19 pandemic in China: qualitative survey study. JMIR Med Inform 2021;9(6):e26463 [FREE Full text] [doi: 10.2196/26463] [Medline: 33945493]
40. Xu X, Cai Y, Wu S, Guo J, Yang L, Lan J, et al. Assessment of internet hospitals in China during the COVID-19 pandemic: national cross-sectional data analysis study. J Med Internet Res 2021;23(1):e21825 [FREE Full text] [doi: 10.2196/21825] [Medline: 33417586]
41. Gong K, Xu Z, Cai Z, Chen Y, Wang Z. Internet hospitals help prevent and control the epidemic of COVID-19 in China: multicenter user profiling study. J Med Internet Res 2020;22(4):e18908 [FREE Full text] [doi: 10.2196/18908] [Medline: 32250962]
42. Wu J, Wang J, Nicholas S, Maitland E, Fan Q. Application of big data technology for COVID-19 prevention and control in China: lessons and recommendations. J Med Internet Res 2020;22(10):e21980 [FREE Full text] [doi: 10.2196/21980] [Medline: 33001836]
43. Liu M. Dilemmas and legislative countermeasures in the development of the internet medical service in China: commentary on measures for the administration of internet diagnosis and treatment (for trial implementation) and two other documents. J Southwest Univ (Soc Sci Ed) 2022;48(02):37-48. [doi: 10.13718/j.cnki.xdsk.2022.02.004]
44. Wang X, Li F. The status and regulation of the development of internet hospitals in China. Health Econ Res 2020;37(11):23-25. [doi: 10.14055/j.cnki.33-1056/f.2020.11.022]
45. Li L, Ding H. Risk and prevention and control strategies of privacy leakage of patients under the context of 'internet+ healthcare' in China. Med Soc 2023;36(01):57-63. [doi: 10.13723/j.yxysh.2023.01.011]
46. Chen Y. Research on healthcare data sharing and personal information protection. J Intell 2023;42(05):192-199.
47. Price WN, Cohen IG. Privacy in the age of medical big data. Nat Med 2019;25(1):37-43 [FREE Full text] [doi: 10.1038/s41591-018-0272-7] [Medline: 30617331]
48. Abouelmehdi K, Beni-Hessane A, Khaloufi H. Big healthcare data: preserving security and privacy. J Big Data 2018;5(1):1. [doi: 10.1186/s40537-017-0110-7]
49. Zhang L, Wang H, Li Q, Zhao MH, Zhan QM. Big data and medical research in China. BMJ 2018;360:j5910 [FREE Full text] [doi: 10.1136/bmj.j5910] [Medline: 29437562]
50. Yin M. Conflict and reconciliation in big data construction and privacy protection in health care. Med Soc 2023;36(02):125-131. [doi: 10.13723/j.yxysh.2023.02.023]
51. Sun Z, Liu X, Tian Z. Discussing the security protection of health care big data. Chin Health Serv Manage 2021;38(07):518-520, 525.
52. Creemers R. China's emerging data protection framework. J Cybersecurity 2022;8(1):tyac011. [doi: 10.1093/cybsec/tyac011]
53. Berti R. Data protection law: a comparison of the latest legal developments in China and European Union. Eur J Privacy Law Technol 2020;1:34 [FREE Full text]
54. Standardization Administration of China. Information security technology—Personal information security specification (GB/t 35273-2020). Standardization Administration of China. 2020 Oct 01. URL: https://openstd.samr.gov.cn/bzgk/gb/newGbInfo?hcno=4568F276E0F8346EB0FBA097AA0CE05E [accessed 2024-05-30]

55. Cui S, Qi P. The legal construction of personal information protection and privacy under the Chinese Civil Code. Comput Law Secur Rev 2021;41:105560 [FREE Full text] [doi: 10.1016/j.clsr.2021.105560]
56. Liu Y, Huang L, Yan W, Wang X, Zhang R. Privacy in AI and the IoT: the privacy concerns of smart speaker users and the personal information protection law in China. Telecommun Policy 2022;46(7):102334 [FREE Full text] [doi: 10.1016/j.telpol.2022.102334]
57. Calzada I. Citizens' data privacy in China: the state of the art of the personal Information Protection Law (PIPL). Smart Cities 2022;5(3):1129-1150. [doi: 10.3390/smartcities5030057]
58. Bowie NE, Jamal K. Privacy rights on the internet: self-regulation or government regulation? Bus Ethics Q 2006;16(3):323-342. [doi: 10.5840/beq200616340]
59. Tang Z, Hu Y, Smith MD. Gaining trust through online privacy protection: self-regulation, mandatory standards, or caveat emptor. J Manage Inf Syst 2008;24(4):153-173. [doi: 10.2753/mis0742-1222240406]
60. Karjoth G, Schunter M. A privacy policy model for enterprises. : IEEE; 2002 Presented at: Proceedings 15th IEEE Computer Security Foundations Workshop. CSFW-15; June 24–26 2002; Cape Breton, NS, Canada p. 271-281. [doi: 10.1109/csfw.2002.1021821]
61. Flavián C, Guinalíu M. Consumer trust, perceived security and privacy policy: three basic elements of loyalty to a web site. Ind Manage Data Syst 2006;106(5):601-620. [doi: 10.1108/02635570610666403] [Medline: 18657047]
62. Shaffer G. Applying a contextual integrity framework to privacy policies for smart technologies. J Inf Policy 2021;11:222-265. [doi: 10.5325/jinfopoli.11.2021.0222]
63. Nissenbaum H. Privacy as contextual integrity. Wash L Rev 2004;79:119 [FREE Full text]
64. Fan M, Yu L, Chen S. An Empirical Evaluation of GDPR Compliance Violations in Android mHealth Apps. Coimbra, Portugal: IEEE; 2020 Presented at: 2020 IEEE 31st International Symposium on Software Reliability Engineering (ISSRE); October 12-15 2020; Coimbra, Portugal p. 253-264. [doi: 10.1109/issre5003.2020.00032]
65. Wang M, Jiang Z. The defining approaches and practical paradox of sensitive data: an investigation of data protection laws in 92 Countries and Regions and 200 data breaches in the world. Int J Commun 2017;11:3286-3305. [doi: 10.1201/9781138069848-29]
66. Braghin C, Cimato S, Libera AD. Are mHealth apps secure? A case study. : IEEE; 2018 Presented at: 2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC); July 23-27 2018; Tokyo, Japan p. 335-340. [doi: 10.1109/compsac.2018.10253]
67. Alfawzan N, Christen M, Spitale G, Biller-Andorno N. Privacy, data sharing, and data security policies of women's mHealth apps: scoping review and content analysis. JMIR mHealth uHealth 2022;10(5):e33735 [FREE Full text] [doi: 10.2196/33735] [Medline: 35522465]
68. Benjumea J, Ropero J, Rivera-Romero O, Dorronzoro-Zubiete E, Carrasco A. Privacy assessment in mobile health apps: scoping review. JMIR mHealth uHealth 2020;8(7):e18868 [FREE Full text] [doi: 10.2196/18868] [Medline: 32459640]
69. Hutton L, Price BA, Kelly R, McCormick C, Bandara AK, Hatzakis T, et al. Assessing the privacy of mHealth apps for self-tracking: heuristic evaluation approach. JMIR mHealth uHealth 2018;6(10):e185 [FREE Full text] [doi: 10.2196/mhealth.9217] [Medline: 30348623]
70. Sunyaev A, Dehling T, Taylor PL, Mandl KD. Availability and quality of mobile health app privacy policies. J Am Med Inform Assoc 2015;22(e1):e28-e33 [FREE Full text] [doi: 10.1136/amiajnl-2013-002605] [Medline: 25147247]
71. Galvin HK, DeMuro PR. Developments in privacy and data ownership in mobile health technologies, 2016-2019. Yearb Med Inform 2020;29(1):32-43 [FREE Full text] [doi: 10.1055/s-0040-1701987] [Medline: 32823298]
72. Jiang J, Zheng Z. Personal information protection and privacy policy compliance of health code apps in China: scale development and Content analysis. JMIR mHealth uHealth 2023;11:e48714 [FREE Full text] [doi: 10.2196/48714] [Medline: 37990813]
73. Ni Z, Wang Y, Qian Y. Privacy policy compliance of chronic disease management apps in China: scale development and content evaluation. JMIR mHealth uHealth 2021;9(1):e23409 [FREE Full text] [doi: 10.2196/23409] [Medline: 33507159]
74. Zhang M, Shen J, Liu H, Yan W. Privacy policy evaluation of internet hospital apps in China from the perspective of cognitive load and content compliance. J Mod Inf 2023;43(03):110-122.
75. Liu Q, Liu H, Qin Z, Ma C. Research on users' privacy protection policy of m-health application based on content analysis. Chin Hosp 2019;23(09):20-23. [doi: 10.19660/j.issn.1671-0592.2019.09.06]
76. Li Y, Saxunová D. A perspective on categorizing personal and sensitive data and the analysis of practical protection regulations. Procedia Comput Sci 2020;170:1110-1115 [FREE Full text]
77. Rakers J. Managing professional personal sensitive information. : Association for Computing Machinery; 2010 Presented at: SIGUCCS '10: Proceedings of the 38th annual ACM SIGUCCS fall conference: navigation and discovery; October 24 2010; New York, NY, United States URL: https://doi.org/10.1145/1878335.1878340
78. Tschider C. The consent myth: improving choice for patients of the future. Wash UL Rev 2018;96:1505 [FREE Full text]
79. Garg S, Goldwasser S, Vasudevan P. Formalizing Data Deletion in the Context of the Right to Be Forgotten. Cham. Switzerland: Springer International Publishing; 2020:373-402.
80. Shu J, Zhang Y, Li J, Li B, Gu D. Why data deletion fails? A study on deletion flaws and data remanence in android systems. ACM Trans. Embed. Comput. Syst 2017;16(2):1-22. [doi: 10.1145/3007211]

81. Nissenbaum H. Protecting Privacy in an Information Age: The Problem of Privacy in Public. Routledge: Taylor & Francis Group; 2017:141-178.
82. Nurgalieva L, O'Callaghan D, Doherty G. Security and privacy of mHealth applications: a scoping review. 2020 Presented at: IEEE Access; June 04 2020; NA p. 104247-104268. [doi: 10.1109/access.2020.2999934]
83. Wolters PTJ. The security of personal data under the GDPR: a harmonized duty or a shared responsibility? Int Data Priv Law 2017;7(3):165-178. [doi: 10.1093/idpl/ipx008]
84. Lai Y, Chen S, Li M, Ung COL, Hu H. Policy interventions, development trends, and service innovations of internet hospitals in China: documentary analysis and qualitative interview study. J Med Internet Res 2021;23(7):e22330 [FREE Full text] [doi: 10.2196/22330] [Medline: 34283025]
85. Zhang H, Zhang H, Wang X, Yang Z, Zhao Y. Analysis of requirements for developing an mHealth-based health management platform. JMIR mHealth uHealth 2017;5(8):e117 [FREE Full text] [doi: 10.2196/mhealth.5890] [Medline: 28778840]
86. Li Y, Hu H, Rozanova L, Fabre G. COVID-19 and internet hospital development in China. Epidemiologia (Basel) 2022;3(2):269-284 [FREE Full text] [doi: 10.3390/epidemiologia3020021] [Medline: 36417257]
87. Yang Z, Cao G, Luo X. Legal protection and fair use of patient health and medical information. Med Soc 2022;35(02):140-144. [doi: 10.13723/j.yxysh.2022.02.027]
88. Sun L, Buijsen M. Mobile health in China: does it meet availability, accessibility, acceptability and quality standards? Health policy and technology. Health Policy Technol 2022;11(3):100660 [FREE Full text] [doi: 10.1016/j.hlpt.2022.100660]
89. Ye G, Chen J, Zhang S, Yuan H. Research on the current situation, problems and countermeasures of internet medical treatment of traditional Chinese medicine. Chin Hosp 2023;27(11):34-39. [doi: 10.19660/j.issn.1671-0592.2023.11.08]
90. Xu L. Internet Hospital: Challenges and Opportunities in China. Switzerland: Springer International Publishing; 2016:85-90.
91. Ministry of industry and information technology reported 31 Apps (SDKs) with infringment on users rights and interest. Information and Communication Bureau. URL: https://www.cnii.com.cn/tx/202307/t20230710_485484.html [accessed 2023-10-18]
92. Ministry of industry and information technology reported 56 apps (SDKs) with infringment on users rights and interest. Information and Communication Bureau. 2023. URL: https://www.tjcac.gov.cn/xxh/dzzw/202305/t20230508_6233305.html [accessed 2023-10-18]
93. Ministry of industry and information technology reported 46 apps (SDKs) with infringment on users rights and interest. Information and Communication Bureau. 2023. URL: https://www.cnii.com.cn/tx/202302/t20230208_445308.html [accessed 2023-10-18]
94. Ministry of industry and information technology reported 55 apps (SDKs) with infringment on users rights and interest. Information and Communication Bureau. 2023. URL: http://news.china.com.cn/2023-03/22/content_85183464.htm?f=pad&a=true [accessed 2023-10-18]
95. Abouelmehdi K, Beni-Hssane A, Khaloufi H, Saadi M. Big data security and privacy in healthcare: a review. Procedia Comput Sci 2017;113:73-80 [FREE Full text] [doi: 10.1016/j.procs.2017.08.292]

Abbreviations

CI: contextual integrity
HIPAA: Health Insurance Portability and Accountability Act
mHealth: mobile health
PI: personal information
PIPL: Personal Information Protection Law

Edited by L Buis; submitted 01.12.23; peer-reviewed by J Shi, Y Lee, S Liu; comments to author 17.01.24; revised version received 23.03.24; accepted 22.05.24; published 21.06.24.

Please cite as:
Jiang J, Zheng Z
Medical Information Protection in Internet Hospital Apps in China: Scale Development and Content Analysis
JMIR Mhealth Uhealth 2024;12:e55061
URL: https://mhealth.jmir.org/2024/1/e55061
doi:10.2196/55061
PMID:38904994

Nutrition-Related Mobile Apps in the Spanish App Stores: Quality and Content Analysis

Cesar I Fernandez-Lazaro[1,2], MPH, PharmD, PhD; Gema Santamaría[3], MSc; Annika Fernandez Milano[4], MPH; Maria I Martin-Vergel[5], MSc; Diego Fernandez-Lazaro[6,7], PharmD, PhD

[1]Department of Preventive Medicine and Public Health, School of Medicine, University of Navarra, Pamplona, Spain

[2]IdiSNA, Navarra Institute for Health Research, Pamplona, Spain

[3]Departamento de Anatomía y Radiología, Facultad de Ciencias de la Salud, Campus de Soria, Universidad de Valladolid, Soria, Spain

[4]IQVIA Commercial GmbH & Co OHG, Frankfurt am Main, Germany

[5]Departamento de Ciencias Farmacéuticas y de la Salud, Facultad de Farmacia, Universidad CEU San Pablo, Madrid, Spain

[6]Grupo de Investigación Reconocido "Neurobiología," Facultad de Medicina, Universidad de Valladolid, Valladolid, Spain

[7]Departamento de Biología Celular, Genética, Histología y Farmacología, Facultad de Ciencias de la Salud, Campus de Soria, Universidad de Valladolid, Soria, Spain

Corresponding Author:
Cesar I Fernandez-Lazaro, MPH, PharmD, PhD
Department of Preventive Medicine and Public Health, School of Medicine, University of Navarra, , Pamplona, , Spain

Abstract

Background: Mobile apps represent accessible and cost-effective tools to improve nutrition and prevent chronic diseases. However, most of these apps have been characterized as having limited functionality, raising concerns about their effectiveness, acceptability, and efficacy.

Objective: The aims of the study were to assess the quality of popular nutrition-related app platforms in Spain and to describe their characteristics and functionalities.

Methods: We screened apps providing information on dietary advice, food advice, and nutritional content in the Apple App Store and Google Play Store in Spain from March 2 to March 16, 2024. Apps with a star rating of ≥4 (of 5 stars), those available in Spanish, those that were free of charge, those last updated after January 2022, those with >500 reviews, and those with >500,000 downloads were included. The quality of apps was assessed using the user version of the Mobile App Rating Scale (uMARS). General characteristics and nutritional, health, and market-related functionalities of the nutrition-related apps were described. Correlations among total and uMARS sections, star ratings, and number of reviews and downloads were evaluated.

Results: Among the 1460 apps identified in the search, 42 apps met the criteria. The majority of these (n=20, 48%) aimed at recording and analyzing food intake, followed by those providing nutritional plans or diets (n=9, 21%), advising on healthy habits (n=7, 17%), and offering recipes (n=6, 14%). The most prevalent nutritional functionalities offered were recording and monitoring body measurements (n=30, 71%), food tracking (n=26, 62%), and dietary analysis (n=25, 60%), whereas nutrition education was less common (n=16, 38%). Among market-related functionalities, advertisements were the most common among the study apps (n=30, 71%), followed by the option of sharing on social media (n=29, 69%) and customizable reminders (n=26, 62%). Sharing the recorded information in the app with health professionals was infrequent (n=1, 2%). The mean (SD) total uMARS score (maximum 5 points) was 3.78 (0.35), while the mean (SD) uMARS scores for functionality, aesthetics, engagement, and information were 4.21 (0.38), 3.94 (0.54), 3.51 (0.46), and 3.48 (0.44), respectively. Lower mean scores were observed for the subjective quality (mean 2.65, SD 0.56) and perceived impact (mean 3.06, SD 0.67). Moderate to strong positive significant correlations were mostly observed between total uMARS and section-specific uMARS scores, while the correlations between the uMARS section scores were mostly moderate positive. Total uMARS scores were very weakly correlated with user rating, number of reviews, and number of downloads.

Conclusions: The quality of popular nutrition-related app platforms in Spain was acceptable, with observed remarkable differences between sections. The majority of the apps were appealing due to their user-friendly interfaces. Only a few apps, however, provided dietary structure analysis or nutritional education. Further research is needed to assess the long-term impact of these apps on users.

(JMIR Mhealth Uhealth 2024;12:e52424) doi:10.2196/52424

KEYWORDS

mobile apps; mHealth; mobile health; app; nutritional; nutrition; dietary; eating; diet; food; lifestyle; Spain; Spanish; chronic diseases; chronic; review; quality; MARS; Mobile App Rating Scale; uMARS; user version of the Mobile App Rating Scale; assessment; mobile phone

Introduction

Noncommunicable diseases—commonly known as chronic diseases—are responsible for more than 60% of the global disease burden, representing a public health concern [1]. Their high prevalence and mortality rates constitute a threat to individuals, families, health care systems, and governments due to their health and economic impacts [2,3]. Mounting and robust scientific evidence suggests that nutrition, a potentially modifiable lifestyle factor, is linked to chronic diseases, opening up new opportunities for their prevention [4,5]. For instance, a recent umbrella review of meta-analyses, which included 116 primary prospective cohort studies, revealed inverse associations between healthy dietary patterns and risk of chronic diseases, including type 2 diabetes, certain types of cancer, and cardiovascular disease, among others. Conversely, positive associations were found between unhealthy dietary patterns and the risk of chronic diseases [6]. Additionally, extensive research on the Mediterranean diet, one of the most studied and well-known dietary patterns globally, has consistently demonstrated its positive impact on cardiovascular health. This includes a decreased risk in developing lifestyle diseases, such as obesity, hypertension, metabolic syndrome, and dyslipidemia, as well as being linked to lower rates of diabetes, glycemic control, and an age-related cognitive decline [7,8].

In the search for accessible and cost-effective interventions to improve nutrition and prevent chronic diseases, mobile health (mHealth) apps have emerged as an extension of traditional approaches [9]. mHealth apps may provide continuous monitoring, health care provider communication and support, reminders, patient education, and patient involvement strategies that facilitate chronic disease management [10,11]. Moreover, mHealth apps have proven to be effective in achieving healthier lifestyles through behavioral change by helping to increase physical activity levels and improve dietary patterns [12-14]. These relatively new digital health methods stand out for their low cost, innovative functionalities, and high usability [15]. Moreover, the use of these apps has remarkably increased in recent years [16-18]. Currently, 6.9 billion users—86.1% of the world population—own a smartphone [19] with access to more than 350,000 apps in top app stores in 2021 (with approximately 90,000 new mHealth apps added in 2020) [15]. Among these apps, approximately 11% were categorized as diet and nutrition-related apps.

The thousands of mHealth apps available in top stores enable smartphone users to choose the app that best meets their needs. However, over 50% of mHealth apps have limited functionality [20], and there are several concerns about their effectiveness, acceptability, and efficacy [21,22], indicating room for improvement [23]. Moreover, reports suggest that most users stop using mHealth apps after installation or a few interactions [24,25] due to a lack of desired features or the apps not being easy to use [26].

Spain is among the European countries with the highest smartphone penetration rates, with over 47 million people (97% of the population) owning a smartphone. The use of mHealth apps is also continuously increasing, particularly the nutrition-related apps that account for 34% of the mHealth apps market [27,28]. No studies, however, have been conducted to evaluate the quality and functionality of nutrition apps in the Spanish app stores. Therefore, the aims of this study were to evaluate the quality of popular nutrition-related app platforms in Spain and describe the different characteristics and functionalities of these apps.

Methods

Search Strategy

The iOS (Spanish Apple App Store) and Android (Spanish Google Play Store) platforms were used to search for nutrition-related apps from March 2 to March 16, 2024. The search on the Android platform was conducted via laptop, using the Google Chrome browser in an incognito window while logged out of the Google account to prevent any influence of user preferences. The search on the iOS platform was conducted on a mobile device, an iPhone 11 Pro. Of note, the displayed apps may differ from those in other countries due to the search having been carried out in Spanish app stores. The following Spanish search terms were used: “Nutrición” (nutrition), “Dieta” (diet), and “Dieta saludable” (healthy diet). All results for each keyword on both platforms were reviewed.

Screening and Selection Criteria

The selection process of the apps was conducted following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) [29]. All apps were initially independently screened and analyzed by 2 study researchers (GS and MIM-V) based on the description and screenshots on their download page, and disagreements between them were resolved by a third reviewer (CIF-L). We based the selection of apps on the following criteria: (1) a minimum rating of 4 of 5 stars; (2) available in the Spanish language; (3) freemium (free of charge); (4) last update after January 2022; (5) more than 500 reviews; (6) more than 500,000 downloads; and (7) dietary advice (eg, diet plan or meal planning), food advice (eg, cooking process or characteristics of food), or nutritional information of food or drinks (eg, nutrient content or recommended intake). The following exclusion criteria were applied: (1) apps initially free of charge but with a monthly subscription later on, (2) apps used as a food diary without dietary analysis, (3) apps requiring external equipment (scales and smartwatches) for use, (4) apps designed for use by health care professionals, (5) restaurant apps for making reservations or displaying menus, (6) apps for taking photos or videos of food, (7) educational food-related apps for children, (8) apps that allow offline payments for health services, (9) apps that exclusively monitor weight and physical activity or display recipes (without including nutritional

information), and (10) apps targeted at specific populations such as pregnant women or individuals diagnosed with diabetes.

After identifying the nutrition-related apps that met all the inclusion criteria and none of the exclusion criteria, duplicates within and between the platforms were removed before each researcher downloaded and explored the remaining apps. Following a more in-depth analysis, apps were examined by the same study researchers and selected for further analysis based on the abovementioned criteria.

Data Extraction

The following general characteristics were extracted from the nutrition-related apps that met these criteria: number of downloads, number of reviews, number of stars, date of last update, and a brief description of the main goal of the apps. Since the Apple App Store does not provide the number of downloads, this information and the rest of the characteristics related to the apps were extracted from the Google Play Store when the apps were available on both platforms. In addition, the nutritional, health, and market-related functionalities of these apps were extracted by the study researchers after a literature review and discussions with experts and were recorded in a Microsoft Excel spreadsheet (Microsoft Corp).

Quality Assessment Using the User Version of the Mobile App Rating Scale

The user version of the Mobile App Rating Scale (uMARS) was used to evaluate the quality of the apps [30]. The uMARS is an adaptation of the Mobile App Rating Scale (MARS) developed for end users, while the MARS is used by professionals who require training and expertise in mHealth in order to perform the assessment [31]. The uMARS is composed of 4 objective sections, each containing multiple items: engagement, functionality, aesthetics, and information. The engagement section (5 items) assesses how entertaining, interesting, customizable, and interactive the app is and whether it fits the target group. The functionality section (4 items) tests the app's performance, ease of use, navigation, and gestural design. The aesthetics section (3 items) evaluates the layout, graphic design, and overall visual appeal. Finally, the information section (4 items) focuses on the quality and quantity of the information, the visual information, and the credibility of the sources. All items are assessed on a 5-point scale (1=inadequate, 2=poor, 3=acceptable, 4=good, and 5=excellent). In addition, the uMARS includes the sections subjective quality (4 items) and perceived impact (6 items), which evaluate whether the app may increase awareness, change attitudes, and promote changes toward healthier behaviors. The same 2 study researchers (GS and MIM-V) independently assessed the quality of the nutrition-related apps using the uMARS. The researchers evaluated the apps from the perspective of users rather than experts.

Statistical Analysis

Descriptive statistics, including frequency and proportions, were used to summarize the general characteristics and the nutritional, health, and market-related functionalities of the selected apps. The overall and section-specific uMARS scores, along with the uMARS scores by the main aim of the apps, were described by their maxima, minima, means, medians, and IQRs and were graphically represented by boxplots. Pearson and Spearman correlation coefficients for normally and nonnormally distributed data, respectively, were determined between app characteristics and total and section-specific uMARS scores. The distribution of the data was checked with the Shapiro-Wilk test. All analyses were performed with Stata (version 16.0; StataCorp LLC), with a 2-tailed level of statistical significance set at $P \leq .05$.

Ethical Considerations

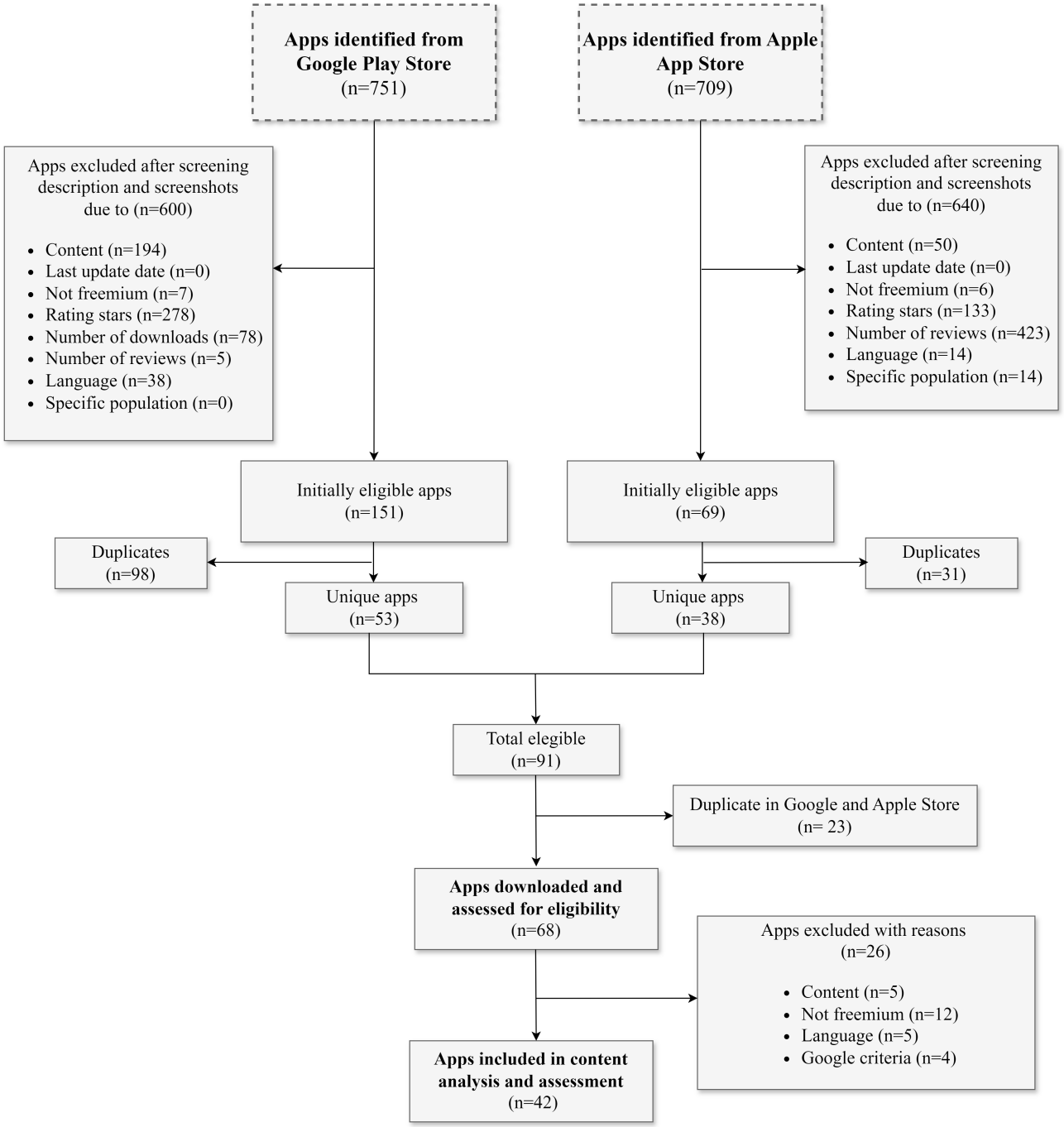
The research conducted in this study did not require the participation of human participants, and no personal data were referenced or collected. The data used in the study were obtained solely from publicly available sources on app stores, and as a result, ethics approval was not needed. This rationale aligns with the institutional policies of the research location [32].

Results

Search Results

Our search yielded a total of 1460 (n=751 from Google Play Store and n=709 from Apple App Store) apps (Figure 1). Of these, 1240 were excluded for not meeting the inclusion criteria (Google Play Store [n=600]: content: n=194, not freemium: n=7, rating stars: n=278, number of downloads: n=78, number of reviews: n=5, and language: n=38; and Apple App Store [n=640]: content: n=50, not freemium: n=6, rating stars: n=133, number of reviews: n=423, language: n=14, and targeted specific population, n=14). After removing duplicates on the same platform (Google Play Store: n=98 and Apple App Store: n=31) and between the Google Play Store and Apple App Store (n=23), the remaining 68 apps were downloaded and further assessed for eligibility. Of these, 26 apps were then excluded (content: n=5, not freemium: n=12, language: n=5, and Google criteria: n=4), resulting in a total of 42 nutrition-related apps that were included in the study for an in-depth analysis.

Figure 1. Flowchart of the selection of nutrition-related apps for the in-depth analysis and quality assessment.



General Characteristics of the Nutrition-Related Apps

The general characteristics of the 42 apps included in the study are displayed in Table 1. A large proportion of the apps (n=21, 50%) were exclusively available on the Android platform, and only a small proportion was exclusively available on the iOS platform (n=5, 12%). The remaining apps (n=16, 38%) were available on both platforms. A great proportion of the selected apps had a rating of 4.6 to 4.8 stars (n=18, 43%), at least 50,000 reviews (n=17, 41%), 1.0 to 4.9 million downloads (n=18, 43%), and offered a premium version with additional paid features (n=38, 91%). Furthermore, the primary aim of many apps was to record and analyze food intake (n=20, 48%), followed by providing nutritional plans or diets (n=9, 21%), advising on healthy habits (n=7, 17%), and offering recipes (n=6, 14%).

Table . General characteristics of the nutrition-related apps included in the study (N=42).

Characteristics	Values, n (%)
Platform	
Android	21 (50)
iOS	5 (12)
Android and iOS	16 (38)
User star rating	
4.0 - 4.2	6 (14)
4.3 - 4.5	16 (38)
4.6 - 4.8	18 (43)
4.9 - 5.0	2 (5)
Number of reviews (in thousands)	
<5.0	9 (21)
5.0 - 14.9	10 (24)
15.0 - 49.9	6 (14)
≥50.0	17 (41)
Number of downloads (in millions)	
Missing	5 (12)
<1.0	4 (10)
1.0 - 4.9	18 (43)
5.0 - 9.9	5 (12)
10.0 - 49.9	7 (17)
≥50.0	3 (7)
Premium version	
Yes	38 (91)
No	4 (9)
Main aim	
Nutritional plan or diet	9 (21)
Recipes	6 (14)
Recording and analyzing food intake	20 (48)
Advice on healthy habits	7 (17)

Nutritional and Health Functionalities

The nutritional and health functionalities of the apps included in the study are described in Table 2. More than half of the apps offered a search function for food and nutritional information (n=23, 55%), specifically on energy content (n=23, 55%) and macronutrients (n=21, 50%). A lower proportion of apps informed about fiber (n=9, 21%), micronutrients (n=6, 14%), and food additives (n=2, 5%). Around two-thirds of the apps included a function to record food intake (n=26, 62%), of which quantification by weight (n=21, 50%) and portion size (n=19, 45%) were the most common. Recipes and dietary plans were provided by 48% (n=20) of the apps, and around one-third of these were customizable to the user's preferences, intolerances, or lifestyle (n=14, 33%). The majority of the apps also offered the recording and monitoring of body measurements (n=30, 71%), such as weight (n=30, 71%) and BMI (n=14, 33%). In addition to nutritional functionalities, a considerable number of apps provided health functionalities related to physical activity (n=25, 60%), with the most common functionalities being type (n=23, 55%) and duration of physical activity (n=22, 52%). Among the apps that offered dietary analysis and suggestions (n=25, 60%), most provided recommendations on water and calorie intake (n=24, 57%) and energy analyses (n=23, 55%), followed by dietary structure analyses (n=20, 48%). The fewest apps provided nutritional education (n=16, 38%).

Table . Nutritional and health functionalities of the nutrition-related apps included in the study (N=42).

Functionalities	Values, n (%)
Searching food and providing nutritional information	23 (55)
Energy	23 (55)
Macronutrients	21 (50)
Micronutrients	6 (14)
Food additives	2 (5)
Fiber	9 (21)
Recording food intake	26 (62)
Quantified by weight	21 (50)
Quantified by portion	19 (45)
Favorite foods	17 (41)
Water consumption	18 (43)
Providing recipes and dietary plans	20 (48)
Customizable according to preferences, intolerances, or lifestyle	14 (33)
With specific amount of food	12 (29)
Shopping list	12 (29)
Recording and monitoring body measurements	30 (71)
Weight	30 (71)
Waist circumference	11 (26)
BMI	14 (33)
Recording physical activity	25 (60)
Type of physical activity	23 (55)
Duration of physical activity	22 (52)
Burned calories	20 (48)
Pedometer	7 (17)
Dietary analysis and suggestions	25 (60)
Energy analysis	23 (55)
Total energy	23 (55)
Energy balance	12 (29)
Energy ratio of meals	5 (12)
Dietary structure analysis	20 (48)
Food groups	1 (2)
Distribution of macronutrients	19 (45)
Monitoring of micronutrient intake and advising on required minimum	1 (2)
Target weight and BMI	11 (26)
Water and calorie intake recommendations	24 (57)
Nutritional education	16 (38)
Independent education module	14 (33)
Minimum micronutrient requirements, their function, and sources	3 (7)

Market-Related Functionalities

The market-related functionalities of the included apps are described in Table 3. Advertisements were the most common market-related functionality among the selected apps (n=30, 71%), followed by the option of sharing on social media (n=29, 69%) and customizable reminders (n=26, 62%). A total of 43% (n=18) of the apps provided challenges and incentives such as

points, badges, or rankings to enhance users' motivation. A more novel function, sharing the information recorded in the app with health professionals, was only provided by 1 (2%) app, while communication with other users was available in 13 (31%) apps. A total of 21 (50%) apps included in the study offered a method to recognize foods through intelligent recognition technology; specifically, 17 (41%) apps offered barcode or QR code recognition, and 14 (33%) offered photo recognition.

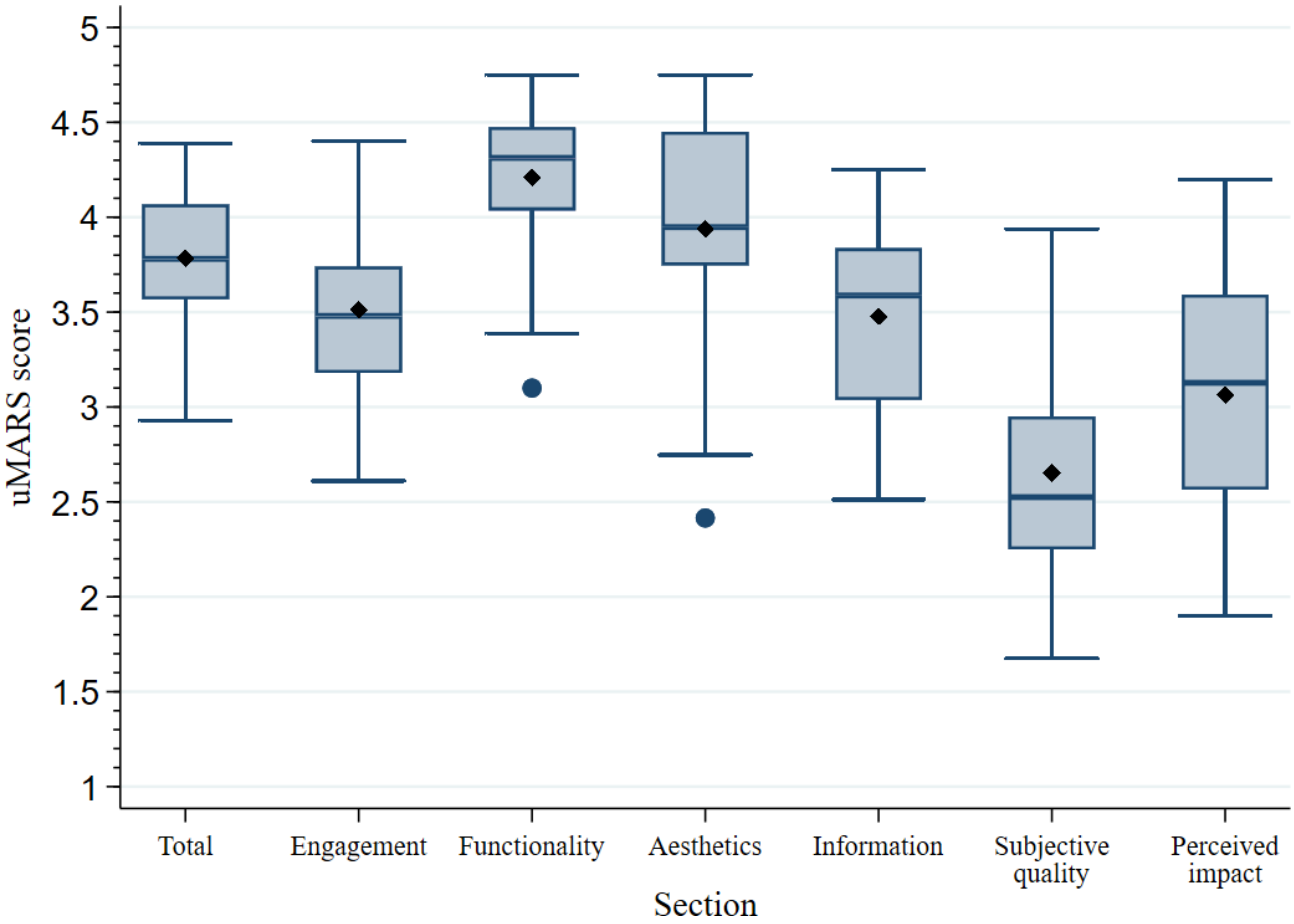
Table . Market-related functionalities of the nutrition-related apps included in the study (N=42).

Functionalities		Values, n (%)
Communicating with other users		13 (31)
Sharing on social media		29 (69)
Receiving invitations to challenges and incentives		18 (43)
Receiving reminders (push notifications)		26 (62)
Receiving advertisement		30 (71)
Buying products		3 (7)
Sharing with health professionals		1 (2)
Intelligent recognition technology		21 (50)
	Barcode or QR code	17 (41)
	Photo	14 (33)

Assessment of the Quality Content

The quality content of the selected nutrition-related apps was assessed using the uMARS scale. The results of the assessment are displayed in Figure 2. A total of 12 (29%) nutrition-related apps scored a mean higher than 4 of 5 points on the uMARS scale. The app with the highest mean total score was 4.39 (SD 0.32), and the lowest mean total score was 2.93 (SD 0.86). The mean total score for all the apps was 3.78 (SD 0.35), and the total median score was 3.78 (IQR 3.57-4.07). Regarding uMARS section-specific scores, functionality was the section with the highest scores, with a mean of 4.21 (SD 0.38), followed by aesthetics (mean 3.94, SD 0.54), engagement (mean 3.51, SD 0.46), and information (mean 3.48, SD 0.44). There was some variability in the scores for each section of the uMARS, with functionality scores ranging from 3.10 to 4.75, aesthetics scores from 2.42 to 4.75, engagement scores from 2.61 to 4.40, and information scores from 2.51 to 4.25.

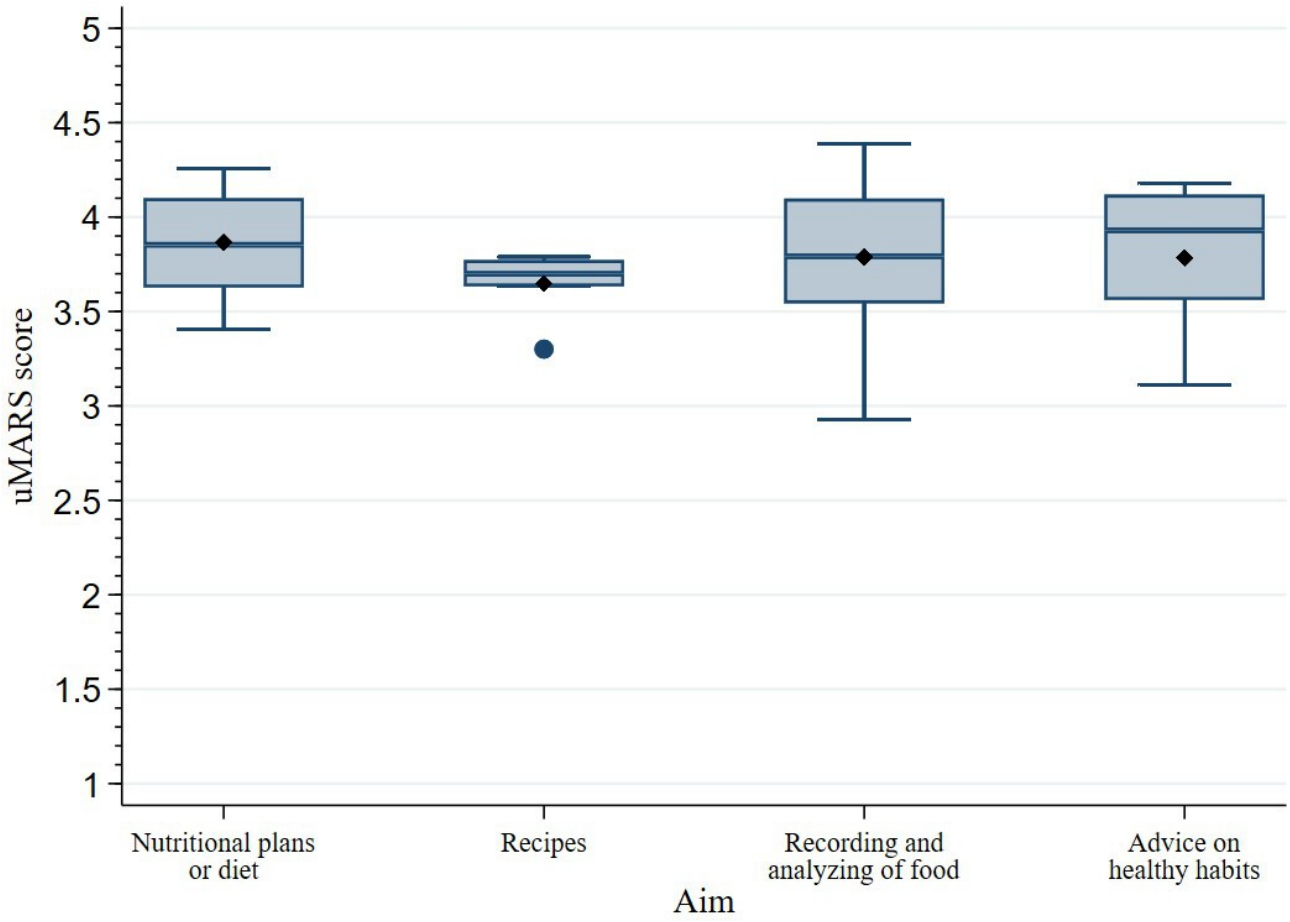
Figure 2. Quality assessment of the nutrition-related apps of the study (N=42). The box plots represent the mean total and section-specific scores of the user version of the Mobile App Rating Scale (uMARS) assessment scores. The line within the boxes represents the median score, the diamond within the boxes represents the mean score, the dot outside the boxes represents an outlier score, and the whiskers represent the maximum and minimum values (except in the presence of outliers).



The sections subjective quality and perceived impact of the uMARS were additionally assessed. The mean scores for the subjective quality and perceived impact sections were lower at 2.65 (SD 0.56) and 3.06 (SD 0.67), respectively, than those of the previously described app quality sections (engagement, functionality, aesthetics, and information). With the exception of aesthetics, a wider range of scores was observed for the perceived impact (1.90 to 4.20) and subjective quality (1.67 to 3.94). Details of the quality content assessment can be found in Multimedia Appendix 1.

The results of the quality content assessment, according to the main aim of the selected nutrition-related apps, are displayed in Figure 3. The apps whose main aim was to provide nutritional plans or diets were the best rated, with a total mean score of 3.87 (SD 0.33). They were followed by the apps that recorded and analyzed food intake (mean 3.79, SD 0.39), the apps that mainly provided advice on healthy habits (mean 3.78, SD 0.37), and those that primarily offered recipes (mean 3.65, SD 0.18). App-specific scores can be found in Multimedia Appendix 2.

Figure 3. Quality assessment according to the aim of the nutrition-related apps of the study (N=42). The box plots represent the mean total score of the user version of the Mobile App Rating Scale (uMARS) assessment by the aim of the apps. The line within the boxes represents the median score, the diamond within the boxes represents the mean score, the dot outside the boxes represents an outlier score, and the whiskers represent the maximum and minimum values (except in the presence of outliers).



Relationship Between App Characteristics, Total uMARS Scores, and Section-Specific uMARS Scores

Pearson and Spearman correlation coefficients between app characteristics and total and section-specific uMARS scores are shown in Table 4. Overall, strong positive significant correlations were observed between total uMARS and section-specific uMARS scores, while the correlations between the uMARS section scores were mostly moderate positive. Notably, total uMARS scores were weakly correlated with user rating, number of reviews, and number of downloads.

Table . Correlations between app characteristics and total and section-specific user version of the Mobile App Rating Scale (uMARS) scores of the nutrition-related apps included in the study (N=42).

Characteristics	uMARS total	User star rating	Number of reviews	Number of downloads[a]	uMARS section-specific				
					Engagement	Functionality	Aesthetics	Information	Subjective quality
User star rating									
Value	0.256[b]								
P value	.13								
Number of reviews									
Value	0.196[b]	0.031[b]							
P value	.25	.85							
Number of downloads [a]									
Value	0.135[b]	0.113[b]	0.860[b]						
P value	.43	.51	<.001						
uMARS engagement									
Value	0.791[c]	0.221[b]	0.191[b]	0.151[b]					
P value	<.001	.19	.26	.37					
uMARS functionality									
Value	0.767[b]	0.158[b]	0.116[b]	0.165[b]	0.353[b]				
P value	<.001	.35	.50	.33	.03				
uMARS aesthetics									
Value	0.743[b]	0.218[b]	0.324[b]	0.260[b]	0.444[b]	0.613[b]			
P value	<.001	.19	.05	.12	.006	<.001			
uMARS information									
Value	0.708[c]	0.123[b]	0.003[b]	0.098[b]	0.618[c]	0.451[b]	0.290[b]		
P value	<.001	.47	.99	.57	<.001	.005	.08		
uMARS subjective quality									
Value	0.757[c]	0.220[b]	0.143[b]	0.203[b]	0.788[c]	0.592[b]	0.346[b]	0.675[c]	
P value	<.001	.19	.40	.23	<.001	<.001	.03	<.001	
uMARS perceived impact									
Value	0.735[c]	0.147[b]	0.178[b]	0.138[b]	0.810[c]	0.497[b]	0.413[b]	0.778[c]	0.763[c]
P value	<.001	.39	.29	.42	<.001	.002	.01	<.001	<.001

[a]Information on the number of downloads was missing for 5 apps.

[b]Correlation assessed using Spearman rank correlation.

[c]Correlation assessed using Pearson rank correlation.

Discussion

Principal Results

To the authors' knowledge, this is the first study conducted with apps from the Spanish app stores (on the iOS and Android platforms) aimed at evaluating the quality of the most popular and freemium nutrition-related apps and at describing their characteristics and functionalities. The study search identified 42 apps that met the inclusion criteria, most of which were aimed at recording and analyzing food intake. The overall quality of the apps included in the study was acceptable; however, remarkable differences between the section-specific quality scores of the uMARS were revealed, with functionality having received the highest rating of the 4 uMARS sections.

The most common functions that the apps provided were recording of body measurements and food intake, searching for food and nutritional information, and conducting dietary analysis. Monitoring and search functions were also found to

be the most common strategies in nutrition-related apps to elicit behavioral change, as indicated elsewhere [33,34]. Only a few apps included in this study offered nutritional education or recipes and dietary plan functionalities, for which a higher development burden could be the reason. The elaboration of these functions is costly and time-consuming, as it requires close collaboration with health care professionals [33].

Comparison With Prior Work

The mean total uMARS score in our study (3.78, SD 0.35) was slightly higher compared to similar nutrition-related app studies conducted in China (3.5) [33] and Korea (2.9) [34]. Unlike our study, the Chinese study [33] included apps regardless of ratings and downloads, while the Korean study [34] used the MARS (opposed to the uMARS used in our study), which requires an assessment by experts rather than by users. Experts may be more inclined to evaluate the quality of the apps more rigorously. Thus, these methodological differences may explain the assessment disparities between the previous studies [33,34] and our study.

In our study, the information section of the uMARS scored the lowest among the 4 main sections, consistent with the study by Martinon et al [35], in which only one-third of the apps offered scientific evidence. This raises concerns about the validity of the information. The information conveyed through apps should be based on empirical evidence, and its accuracy should be rigorously verified, as misinformation can negatively affect users' well-being and health goals. A specialized grading scale designed to evaluate nutritional content could be beneficial in assessing the accuracy of information more effectively [33]. Previous studies have reported overestimations [36] or underestimations [37] of energy intake, while others have found close alignment with validated reference methods [38,39]. The variability in energy intake and nutrient measurements across apps may be explained by the lack of alternative serving sizes offered by the apps [40] and the use of different databases based on distinct nutritional reference guides [36,38,39]. Moreover, self-reported methods of nutritional assessment are subject to a degree of measurement bias, such as underestimation and overestimation of dietary intake, which poses an additional challenge for accuracy calculations [41]. An image-based assessment of the food type and portion could counteract this bias [42]. Along with water and calorie intake recommendations and dietary structure analysis, energy analysis has emerged as one of the most common functionalities offered by dietary analysis apps in our study. In another study [43], energy analysis and calorie intake recommendations were even more prevalent among the apps examined. A possible explanation for this dominance could be the ease of implementing energy calculations within an app [33] and users' interest in tracking energy intake to pursue weight loss goals, as self-monitoring is a crucial element of behavioral weight loss [44,45]. However, the assessment of food group composition or dietary patterns may provide a more comprehensive approach to disease prevention or treatment [46]. In our study and similar research, only a few of the included apps that provided dietary structure analysis actually assessed food group composition [33,39].

Despite the majority of the apps including reminders to inform users about unachieved objectives, and nearly half of the apps offering challenges and incentives, the engagement section scored the second lowest of the 4 main sections, similar to elsewhere [33]. Ongoing motivation is fundamental to continuously engage individuals in successfully achieving app goals and sustaining healthy habits in the long term [47,48]. The inclusion of customizable reminders [49], and in particular gamification, such as receiving digital rewards once a challenge has been successfully completed, may increase the user's motivation through positive, playful experiences [40]. In addition, users seek social support through preferably close-knit communities and continuous guidance by health care professionals. Due to time constraints [47] and a lack of nutritional training [50], physicians often struggle to provide support to their patients regarding nutritional changes. Nutrition-related and other mHealth apps may facilitate the establishment of multidisciplinary health care teams that include dietitians, who can provide nutritional counseling and long-term motivational support to patients. The importance of involving dietitians in promoting dietary lifestyle changes has been studied elsewhere [51,52]. The app environment simplifies the collaboration of different professional groups within the health care system, which would not be feasible otherwise due to location and time constraints.

The functionality and aesthetics sections scored the highest and second highest, respectively. Previous studies have reported functionality as the top uMARS section [33,35,53,54]. When designing apps, it is fundamental to ensure that they work seamlessly and intuitively, as apps that are easy to navigate have been found to reduce usability barriers and motivate people who are less familiar with technology [55]. Regarding aesthetics, the inclusion of visual elements, such as pictures, food icons, or charts, has been previously described as enhancing the user experience [33,34]. For instance, Li et al [33] found that apps with higher aesthetic ratings performed better in overall uMARS scores than those with lower ratings.

The other 2 additional sections of the uMARS, namely, "subjective quality" and "perceived impact," were rated the lowest, which is consistent with findings from other studies [34,35,54]. The low subjective score in relation to the overall uMARS score implies that despite good ratings for engagement, functionality, aesthetics, and information quality, the raters would not recommend the app [35].

Limitations

The results of the study should be interpreted in light of the following limitations. First, this study only included nutrition-related apps in the Spanish language from the app stores in Spain, excluding poorly rated apps (<4 stars), those not free of charge, those with fewer than 500 reviews, and those with fewer than 500,000 downloads. Therefore, this may limit the generalizability of the findings to freemium and the most popular nutrition-related apps in the Spanish app store. Second, since the apps were evaluated from the perspective of users, the assessments could vary from those made by nutrition experts or a joint evaluation by both groups. Third, despite the uMARS [30] having been used in other nutrition-related assessment

studies [33] and being considered a reliable measure of app quality with similar psychometric characteristics and results to the standard version [30,31], the use of questionnaires specifically designed for the assessment of nutrition-related apps such as the App Quality Evaluation may provide more accurate results [56].

Conclusions

This study provides a comprehensive overview of the most demanded nutrition-related apps in the Spanish app market. We found that the majority of the apps were appealing due to their user-friendly interfaces, potentially attracting a larger user base and enhancing adherence. However, few apps provided dietary structure analysis or nutritional education, hindering users' ability to follow a well-balanced diet. Additionally, much of the information provided within the apps raises concerns about its validity. To ensure evidence-based content, collaboration between app developers and nutrition experts is crucial during the app's design phase. Multidisciplinary health care teams, including dietitians, should support patients through mHealth apps to enhance patients' long-term dietary lifestyle changes. Moreover, the findings of this study can help users choose suitable apps and support app developers in the development or refinement of nutrition-related apps. Further research is needed to assess the long-term impact of these apps on users.

Acknowledgments

The authors express their gratitude to Carmen Garcia Mulas for her exemplary dedication as a research assistant. Furthermore, special thanks are extended to the Ministry of Science and Innovation, the Government of Spain, and the European Union for their generous funding of this project. Open Access funding was provided by the Ministry of Science and Innovation, the Government of Spain, and the European Union (project PID2022-141358OBI00).

Authors' Contributions

CIF-L, GS, and DF-L contributed to the conception and design of the study. GS and MIM-V conducted the search and performed the quality assessment. CIF-L and AFM conducted the statistical analysis, interpreted the results, and drafted the first version of the manuscript. GS, MIM-V, and DF-L contributed to writing and editing the manuscript. CIF-L had primary responsibility for the final content of the manuscript. CIF-L resolved disagreements between the 2 reviewers (GS and MIM-V). All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The user version of the Mobile App Rating Scale total and section-specific scores of the nutrition-related apps (N=42).
[DOCX File, 31 KB - mhealth_v12i1e52424_app1.docx]

Multimedia Appendix 2
Results of the total and section-specific quality assessment scores of the user version of the Mobile App Rating Scale assessment for the nutrition-related apps in this study (N=42).
[DOCX File, 48 KB - mhealth_v12i1e52424_app2.docx]

References

1. Roser M, Ritchie H, Spooner F. Burden of disease. Our World in Data. 2021. URL: https://ourworldindata.org/burden-of-disease [accessed 2023-06-17]
2. Noncommunicable diseases. World Health Organization. 2022. URL: https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases [accessed 2023-05-13]
3. About global NCDs. Division of Global Health Protection, Global Health, Centers for Disease Control and Prevention. 2021. URL: https://www.cdc.gov/globalhealth/healthprotection/ncd/global-ncd-overview.html [accessed 2023-06-17]
4. GBD 2017 Diet Collaborators. Health effects of dietary risks in 195 countries, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2019 May 11;393(10184):1958-1972. [doi: 10.1016/S0140-6736(19)30041-8] [Medline: 30954305]
5. Schulze MB, Martínez-González MA, Fung TT, Lichtenstein AH, Forouhi NG. Food based dietary patterns and chronic disease prevention. BMJ 2018 Jun 13;361:k2396. [doi: 10.1136/bmj.k2396] [Medline: 29898951]
6. Jayedi A, Soltani S, Abdolshahi A, Shab-Bidar S. Healthy and unhealthy dietary patterns and the risk of chronic disease: an umbrella review of meta-analyses of prospective cohort studies. Br J Nutr 2020 Dec 14;124(11):1133-1144. [doi: 10.1017/S0007114520002330] [Medline: 32600500]

7. Galbete C, Schwingshackl L, Schwedhelm C, Boeing H, Schulze MB. Evaluating Mediterranean diet and risk of chronic disease in cohort studies: an umbrella review of meta-analyses. Eur J Epidemiol 2018 Oct;33(10):909-931. [doi: 10.1007/s10654-018-0427-3] [Medline: 30030684]
8. Guasch-Ferré M, Willett WC. The Mediterranean diet and health: a comprehensive overview. J Intern Med 2021 Sep;290(3):549-566. [doi: 10.1111/joim.13333] [Medline: 34423871]
9. Oldenburg B, Taylor CB, O'Neil A, Cocker F, Cameron LD. Using new technologies to improve the prevention and management of chronic conditions in populations. Annu Rev Public Health 2015 Mar 18;36:483-505. [doi: 10.1146/annurev-publhealth-031914-122848] [Medline: 25581147]
10. Yi JY, Kim Y, Cho YM, Kim H. Self-management of chronic conditions using mHealth interventions in Korea: a systematic review. Healthc Inform Res 2018 Jul;24(3):187-197. [doi: 10.4258/hir.2018.24.3.187] [Medline: 30109152]
11. Lewis J, Ray P, Liaw ST. Recent worldwide developments in eHealth and mHealth to more effectively manage cancer and other chronic diseases—a systematic review. Yearb Med Inform 2016 Nov 10;25(1):93-108. [doi: 10.15265/IY-2016-020] [Medline: 27830236]
12. Schoeppe S, Alley S, Van Lippevelde W, et al. Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: a systematic review. Int J Behav Nutr Phys Act 2016 Dec 7;13(1):127. [doi: 10.1186/s12966-016-0454-y] [Medline: 27927218]
13. Pradal-Cano L, Lozano-Ruiz C, Pereyra-Rodríguez JJ, et al. Using mobile applications to increase physical activity: a systematic review. Int J Environ Res Public Health 2020 Nov 7;17(21):8238. [doi: 10.3390/ijerph17218238] [Medline: 33171871]
14. Scarry A, Rice J, O'Connor EM, Tierney AC. Usage of mobile applications or mobile health technology to improve diet quality in adults. Nutrients 2022 Jun 12;14(12):2437. [doi: 10.3390/nu14122437] [Medline: 35745167]
15. Baig MM, GholamHosseini H, Connolly MJ. Mobile healthcare applications: system design review, critical issues and challenges. Australas Phys Eng Sci Med 2015 Mar;38(1):23-38. [doi: 10.1007/s13246-014-0315-4] [Medline: 25476753]
16. Digital health trends 2021: innovation, evidence, regulation, and adoption. IQVIA. 2021. URL: https://www.iqvia.com/insights/the-iqvia-institute/reports/digital-health-trends-2021 [accessed 2023-07-07]
17. mHealth—statistics and facts. Statista. 2023. URL: https://www.statista.com/topics/2263/mhealth/ [accessed 2023-06-17]
18. Market research report: mHealth apps market size, share & COVID-19 impact analysis, 2021-2028. Fortune Business Insights. 2021. URL: https://www.fortunebusinessinsights.com/mhealth-apps-market-102020 [accessed 2023-07-03]
19. How many smartphones are in the world? BankMyCell. 2018. URL: https://www.bankmycell.com/blog/how-many-phones-are-in-the-world [accessed 2023-07-05]
20. Patient adoption of mHealth, use, evidence and remaining barriers to mainstream acceptance. IMS Institute for Healthcare Informatics. 2015. URL: https://www.iqvia.com/-/media/iqvia/pdfs/institute-reports/patient-adoption-of-mhealth.pdf [accessed 2023-06-25]
21. Larsen ME, Huckvale K, Nicholas J, et al. Using science to sell apps: evaluation of mental health app store quality claims. NPJ Digit Med 2019;2:18. [doi: 10.1038/s41746-019-0093-1] [Medline: 31304366]
22. Marcolino MS, Oliveira JAQ, D'Agostino M, Ribeiro AL, Alkmim MBM, Novillo-Ortiz D. The impact of mHealth interventions: systematic review of systematic reviews. JMIR Mhealth Uhealth 2018 Jan 17;6(1):e23. [doi: 10.2196/mhealth.8873] [Medline: 29343463]
23. Singh K, Drouin K, Newmark LP, et al. Many mobile health apps target high-need, high-cost populations, but gaps remain. Health Aff (Millwood) 2016 Dec 1;35(12):2310-2318. [doi: 10.1377/hlthaff.2016.0578] [Medline: 27920321]
24. Kayyali R, Peletidi A, Ismail M, Hashim Z, Bandeira P, Bonnah J. Awareness and use of mHealth apps: a study from England. Pharmacy (Basel) 2017 Jun 14;5(2):33. [doi: 10.3390/pharmacy5020033] [Medline: 28970445]
25. Nearly 1 in 4 people abandon mobile apps after only one use. TechCrunch. 2016. URL: https://techcrunch.com/2016/05/31/nearly-1-in-4-people-abandon-mobile-apps-after-only-one-use/?guccounter=1 [accessed 2023-07-25]
26. Mustafa AS, Ali N, Dhillon JS, Alkawsi G, Baashar Y. User engagement and abandonment of mHealth: a cross-sectional survey. Healthcare (Basel) 2022 Jan 24;10(2):221. [doi: 10.3390/healthcare10020221] [Medline: 35206837]
27. Usage of health apps in Spain 2022. Statista. 2024. URL: https://www.statista.com/forecasts/1010582/popular-health-app-types-in-spain [accessed 2024-08-26]
28. Spain: smartphone penetration rate 2019-2028. Statista. 2024. URL: https://www.statista.com/statistics/568268/predicted-smartphone-user-penetration-rate-in-spain/ [accessed 2024-08-26]
29. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:71. [doi: 10.1136/bmj.n71] [Medline: 33782057]
30. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and validation of the user version of the Mobile Application Rating Scale (uMARS). JMIR Mhealth Uhealth 2016 Jun 10;4(2):e72. [doi: 10.2196/mhealth.5849] [Medline: 27287964]
31. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile App Rating Scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27. [doi: 10.2196/mhealth.3422] [Medline: 25760773]
32. Comité Ético de Investigación de la Universidad de Valladolid. UVainvestiga. 2024. URL: https://investiga.uva.es/buenas-practicas/comites-de-etica/comite-etico-de-investigacion [accessed 2024-07-19]

33. Li Y, Ding J, Wang Y, Tang C, Zhang P. Nutrition-related mobile apps in the China app store: assessment of functionality and quality. JMIR Mhealth Uhealth 2019 Jul 30;7(7):e13261. [doi: 10.2196/13261] [Medline: 31364606]
34. Choi J, Chung C, Woo H. Diet-related mobile apps to promote healthy eating and proper nutrition: a content analysis and quality assessment. Int J Environ Res Public Health 2021 Mar 28;18(7):3496. [doi: 10.3390/ijerph18073496] [Medline: 33800531]
35. Martinon P, Saliasi I, Bourgeois D, et al. Nutrition-related mobile apps in the French app stores: assessment of functionality and quality. JMIR Mhealth Uhealth 2022 Mar 14;10(3):e35879. [doi: 10.2196/35879] [Medline: 35285817]
36. Bzikowska-Jura A, Sobieraj P, Raciborski F. Low comparability of nutrition-related mobile apps against the Polish reference method—a validity study. Nutrients 2021 Aug 20;13(8):2868. [doi: 10.3390/nu13082868] [Medline: 34445026]
37. Griffiths C, Harnack L, Pereira MA. Assessment of the accuracy of nutrient calculations of five popular nutrition tracking applications. Public Health Nutr 2018 Jun;21(8):1495-1502. [doi: 10.1017/S1368980018000393] [Medline: 29534771]
38. Ferrara G, Kim J, Lin S, Hua J, Seto E. A focused review of smartphone diet-tracking apps: usability, functionality, coherence with behavior change theory, and comparative validity of nutrient intake and energy estimates. JMIR Mhealth Uhealth 2019 May 17;7(5):e9232. [doi: 10.2196/mhealth.9232] [Medline: 31102369]
39. Fallaize R, Zenun Franco R, Pasang J, Hwang F, Lovegrove JA. Popular nutrition-related mobile apps: an agreement assessment against a UK reference method. JMIR Mhealth Uhealth 2019 Feb 20;7(2):e9838. [doi: 10.2196/mhealth.9838] [Medline: 30785409]
40. Chen J, Cade JE, Allman-Farinelli M. The most popular smartphone apps for weight loss: a quality assessment. JMIR Mhealth Uhealth 2015 Dec 16;3(4):e104. [doi: 10.2196/mhealth.4334] [Medline: 26678569]
41. Ravelli MN, Schoeller DA. Traditional self-reported dietary instruments are prone to inaccuracies and new approaches are needed. Front Nutr 2020;7:90. [doi: 10.3389/fnut.2020.00090] [Medline: 32719809]
42. Lee CD, Chae J, Schap TE, et al. Comparison of known food weights with image-based portion-size automated estimation and adolescents' self-reported portion size. J Diabetes Sci Technol 2012 Mar 1;6(2):428-434. [doi: 10.1177/193229681200600231] [Medline: 22538157]
43. Franco RZ, Fallaize R, Lovegrove JA, Hwang F. Popular nutrition-related mobile apps: a feature assessment. JMIR Mhealth Uhealth 2016 Aug 1;4(3):e85. [doi: 10.2196/mhealth.5846] [Medline: 27480144]
44. Hartmann-Boyce J, Johns DJ, Jebb SA, Aveyard P, Behavioural Weight Management Review Group. Effect of behavioural techniques and delivery mode on effectiveness of weight management: systematic review, meta-analysis and meta-regression. Obes Rev 2014 Jul;15(7):598-609. [doi: 10.1111/obr.12165] [Medline: 24636238]
45. Burke LE, Wang J, Sevick MA. Self-monitoring in weight loss: a systematic review of the literature. J Am Diet Assoc 2011 Jan;111(1):92-102. [doi: 10.1016/j.jada.2010.10.008] [Medline: 21185970]
46. Hu FB. Dietary pattern analysis: a new direction in nutritional epidemiology. Curr Opin Lipidol 2002 Feb;13(1):3-9. [doi: 10.1097/00041433-200202000-00002] [Medline: 11790957]
47. Solbrig L, Jones R, Kavanagh D, May J, Parkin T, Andrade J. People trying to lose weight dislike calorie counting apps and want motivational support to help them achieve their goals. Internet Interv 2017 Mar;7:23-31. [doi: 10.1016/j.invent.2016.12.003] [Medline: 28286739]
48. Fernández-Lázaro CI, Ruiz-Canela M, Martínez-González M. Deep dive to the secrets of the PREDIMED trial. Curr Opin Lipidol 2021 Feb 1;32(1):62-69. [doi: 10.1097/MOL.0000000000000731] [Medline: 33315620]
49. Prestwich A, Perugini M, Hurling R. Can the effects of implementation intentions on exercise be enhanced using text messages? Psychol Health 2009 Jul;24(6):677-687. [doi: 10.1080/08870440802040715] [Medline: 20205020]
50. Shrimpton R, Blaney S. The issue of nutrition training in health professions. Public Health Nutr 2018 Sep;21(13):2526-2527. [doi: 10.1017/S136898001800126X] [Medline: 29766834]
51. Howatson A, Wall CR, Turner-Benny P. The contribution of dietitians to the primary health care workforce. J Prim Health Care 2015 Dec 1;7(4):324-332. [doi: 10.1071/hc15324] [Medline: 26668838]
52. Bednarczuk B, Czekajło-Kozłowska A. Role of nutritional support provided by qualified dietitians in the prevention and treatment of non-communicable diseases. Rocz Panstw Zakl Hig 2019;70(3):235-241. [doi: 10.32394/rpzh.2019.0080] [Medline: 31515982]
53. Mandracchia F, Llauradó E, Tarro L, Valls RM, Solà R. Mobile phone apps for food allergies or intolerances in app stores: systematic search and quality assessment using the Mobile App Rating Scale (MARS). JMIR Mhealth Uhealth 2020 Sep 16;8(9):e18339. [doi: 10.2196/18339] [Medline: 32936078]
54. Bardus M, Ali A, Demachkieh F, Hamadeh G. Assessing the quality of mobile phone apps for weight management: user-centered study with employees from a Lebanese university. JMIR Mhealth Uhealth 2019 Jan 23;7(1):e9836. [doi: 10.2196/mhealth.9836] [Medline: 30672742]
55. Boulos MNK, Wheeler S, Tavares C, Jones R. How smartphones are changing the face of mobile and participatory healthcare: an overview, with example from eCAALYX. Biomed Eng Online 2011 Apr 5;10:24. [doi: 10.1186/1475-925X-10-24] [Medline: 21466669]
56. DiFilippo KN, Huang W, Chapman-Novakofski KM. A new tool for nutrition app quality evaluation (AQEL): development, validation, and reliability testing. JMIR Mhealth Uhealth 2017 Oct 27;5(10):e163. [doi: 10.2196/mhealth.7441] [Medline: 29079554]

Abbreviations

MARS: Mobile App Rating Scale
mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
uMARS: user version of the Mobile App Rating Scale

Edited by L Buis; submitted 04.09.23; peer-reviewed by JP Chart Pascual, J Camargo; revised version received 28.04.24; accepted 05.06.24; published 04.10.24.

Please cite as:
Fernandez-Lazaro CI, Santamaría G, Fernandez Milano A, Martin-Vergel MI, Fernandez-Lazaro D
Nutrition-Related Mobile Apps in the Spanish App Stores: Quality and Content Analysis
JMIR Mhealth Uhealth 2024;12:e52424
URL: https://mhealth.jmir.org/2024/1/e52424
doi:10.2196/52424

Smartphone Apps for Pulmonary Hypertension: Systematic Search and Content Evaluation

Nerea Báez Gutiérrez[1*], PharmD; Héctor Rodríguez Ramallo[2*], PharmD; Elva María Mendoza-Zambrano[3], MD; Berenice Brown Arreola[3], MD; Bernardo Santos Ramos[2], PharmD; Laila Abdel-kader Martín[2], PhD; Remedios Otero Candelera[3], PhD

[1]Pharmacy Department, Hospital Universitario Virgen Macarena, Seville, Spain

[2]Pharmacy Department, Hospital Universitario Virgen del Rocio, Avenida Manuel Siurot, S/n, Seville, Spain

[3]Instituto de Biomedicina de Sevilla, Universidad de Sevilla, Hospital Universitario Virgen del Rocío, Seville, Spain

[*]these authors contributed equally

Corresponding Author:
Héctor Rodríguez Ramallo, PharmD
Pharmacy Department, Hospital Universitario Virgen del Rocio, , Avenida Manuel Siurot, S/n, Seville, , Spain

Abstract

Background: Pulmonary hypertension (PH) is a chronic and complex condition, requiring consistent management and education. The widespread use of smartphones has opened possibilities for mobile health apps to support both patients and health care professionals in monitoring and managing PH more effectively.

Objective: This study aimed to identify and assess the quality of free smartphone apps for PH targeted at either patients or health care professionals.

Methods: A systematic search was conducted on freely available apps for patients with PH and health care professionals, accessed from a Spanish IP address, on Android (Google Play) and iOS (App Store) platforms. Searches were performed in October 2022 and 2023. Apps were independently analyzed by two reviewers, focusing on general characteristics. Quality assessment was based on the Mobile Application Rating Scale (MARS) framework, and Mann-Whitney *U* tests compared mean MARS scores against specific variables.

Results: In the overall study, 21 apps were identified. In the 2022 search, 19 apps were listed (9 iOS, 7 Android, 3 available on both platforms). In the subsequent 2023 search, 16 apps were identified (6 Android, 7 iOS, 3 available on both platforms). Of those identified in 2022, 14 remained available in 2023, with only 7 updated since 2022. In addition, 12 apps targeted patients or the general population, while 9 targeted health care professionals; none involved patients in the development or design. Conversely, 13 apps involving health care professionals were identified. There were 10 apps that received pharmaceutical industry funding. The primary goal for 81% (17/21) of the apps was to disseminate general information about PH. The overall mean MARS quality was acceptable in 2022 and 2023, with mean ratings of 3.1 (SD 0.6) and 3.3 (SD 0.5), respectively. The functionality category achieved the highest scores in both years, indicating ease of use and intuitive navigation. In contrast, the subjective quality domain consistently received the lowest ratings in the MARS assessment across both years. None of the apps underwent clinical testing themselves; however, 2 incorporated tools or algorithms derived from trials. The overall quality of iOS apps statistically outperformed that of Android apps in both years (P<.05). Furthermore, the involvement of health care professionals in app development was associated with enhanced quality, a trend observed in both years (P=.003 for both years).

Conclusions: This review of mobile health apps for PH reveals their emergent development stage, with generally acceptable quality but lacking refinement. It highlights the critical role of health care professionals in app development, as they contribute significantly to quality and reliability. Despite this, a notable stagnation in app quality and functionality improvement over 2 years points to a need for continuous innovation and clinical validation for effective clinical integration. This research advocates for future app developers to actively engage with health care professionals, integrate patient insights, and mandate rigorous clinical validation for PH management.

(JMIR Mhealth Uhealth 2024;12:e57289) doi:10.2196/57289

KEYWORDS

pulmonary hypertension; mobile apps; smartphone; eHealth; mHealth; app; hypertension; chronic condition; mobile health app; monitoring; systematic search; app development; clinical validation; evaluation; pulmonary

Introduction

Pulmonary hypertension (PH) is a pathophysiological disorder encompassing various clinical conditions, often associated with cardiovascular and respiratory diseases [1]. Effective management of PH requires a comprehensive, multidisciplinary approach within specialized centers, ensuring enhanced patient care through collaborative diagnosis, treatment, and continuous quality monitoring [1,2]. Regular communication between patients, their informal caregivers, and the health care professionals involved, including those external to the primary treatment team, is vital for effective disease management [3,4]. This communication enhances patient adherence to the prescribed therapy and ensures meticulous monitoring [5,6]. In this context, mobile health (mHealth) technologies, particularly smartphone apps, present a novel opportunity to enhance disease management. These apps have become pivotal resources for either health care professionals or patients in managing and monitoring diseases such as PH [7,8].

Although the integration of mHealth technologies could expand patient care in PH, it raises significant questions regarding the willingness of patients and health care professionals to invest financially in these digital health tools. Research exploring the willingness to pay for health management apps among these groups is limited [9-12]. Findings suggest a general reluctance among patients to spend money on such apps [9,11,13], likely influenced by the widespread availability of free alternatives that offer similar functionalities. This reluctance underscores the importance of focusing on these no-cost options, reflecting the economic realities and preferences of most patients and health care providers.

Assessing the quality of these free health apps is crucial in order to identify deficiencies and opportunities for improvement and address concerns regarding the accuracy of medical information, user privacy, and accessibility—particularly for populations with limited technology access [14,15]. This analysis can guide the development of future apps and provide valuable insights to health care professionals and patients, aiding them in the selection of more effective and safer apps for health management. However, it is important to acknowledge that the specific application of these technologies in PH may still need to be explored [16].

Currently, our knowledge about the number and quality of free apps for caregivers and patients with PH, as well as the variation in their availability and quality across years, is limited. Therefore, this cross-sectional study aims to conduct a thorough evaluation of commercially available free PH apps, targeted at either patients or health care professionals, for 2 consecutive years. This analysis will focus on the characteristics and overall quality of these apps, addressing the identified gaps in evidence and providing insights into their development and evolution.

Methods

Study Design

This study was designed to evaluate the quality and availability of free smartphone apps for PH targeted at either patients or health care professionals over 2 consecutive years, 2022 and 2023. The decision to conduct the study across these 2 years was based on the dynamic nature of mHealth technologies, where rapid advancements and frequent updates are common. By analyzing the apps over this period, we aimed to capture changes, improvements, and trends that might occur within a relatively short time frame, providing a comprehensive view of the current state and evolution of PH-related mHealth apps.

Search Strategy

A systematic search of mHealth apps specifically relating to PH was conducted using a Spanish IP address on the iOS (App Store) and Android (Google Play) platforms. The initial search was performed on October 5, 2022, with a follow-up search on October 10, 2023, to compare the evolution of apps over 1 year. The search terms used were “pulmonary hypertension,” “pulmonary,” “lung,” “SEPAR” (Sociedad Española de Patología Respiratoria), “European Respiratory Society,” “ERS,” “European Society of Cardiology,” “ESC,” and, when applicable, their equivalents in the Spanish language.

The following inclusion criteria were used for the selection of apps:

- Apps targeted at patients with PH
- Apps targeted at health care professionals providing care for patients with PH
- Apps that included general information about PH

The following exclusion criteria were used:

- Apps that were not freely available
- Apps that were not available in English and/or Spanish
- Apps providing information exclusively about human anatomy
- Apps providing exclusively quizzes or trivia about a disease
- Apps that included inappropriate content (eg, horoscopes, astrology)
- Apps aimed exclusively at fundraising
- Apps with nonfunctional links

The apps had to comply with at least 1 inclusion criteria and none of the exclusion criteria to be included in the data extraction phase.

The search was executed on mobile devices, reflecting a real user’s approach to finding and downloading apps. This method ensures the selection of apps that are directly accessible and functional for end users, aligning with the typical user experience on smartphones and tablets.

The names and descriptions of the apps from the search in the Google Play Store and the Apple App Store were screened against predefined selection criteria. The apps that met the inclusion criteria were downloaded for further screening using a smartphone (Android version 13) for Google Play Store apps. For the apps downloaded from the App Store, an iPad Pro 2020 (Apple Inc) running iOs iPadOS 15.1 was used. For apps available on both platforms, the iOS version was selected for evaluation due to the relatively smaller number of apps on this platform, a method consistent with established methodologies in previous studies [17-19]. This approach not only streamlined the analysis within the confines of our resource constraints but

also was supported by literature indicating minimal variability between the platforms in terms of usability and overall quality scores [20-23].

Data Extraction

Data were obtained from the online app descriptions (including app characteristics and the narrative text) found on the App Store and Google Play Store by 2 independent researchers. Data were extracted and entered into a structured Microsoft Excel 2021 database (Microsoft Corp).

Variables collected for each app were the following: name, developer/owner name, type of developer/owner (eg, commercial, scientific societies, patient associations), platform (Android or iOS), language (English and/or Spanish), app store category (ie, education, health and fitness, medicine), date of publication (year), date of the last update (year), participation of health care professionals in the design and/or development of the apps (yes/no), and participation of patients in the design/development of the apps.

For the purposes of this study, an update was defined as any new version of an app released by its developers, which may include improvements in functionality, user interface changes, additional features, bug fixes, or any other modifications aimed at enhancing the app.

The involvement of health care professionals was acknowledged if specified in the app's description or if health organizations, such as scientific societies and hospitals, developed the app. Patient participation was noted when explicitly mentioned in the app's description.

Technical aspects of the apps were also documented, including features such as social media sharing capabilities (eg, Facebook, Twitter), the requirement of a login, and the necessity of web access for functionality. Additionally, information was collected on whether the apps had received funding or sponsorship from the pharmaceutical industry, and on their potential commercial interests, including the presence of advertisements or options for purchasing services.

The purposes of the apps were classified into the following categories: assessment (ie, apps providing clinical scales, classifying adverse event severity, interpretation for laboratory findings), general information (ie, apps providing information about PH, medication, adverse events), and monitoring of clinical parameters (ie, apps providing a register of laboratory parameters, symptoms). Apps aimed at patients or caregivers were also classified into the following categories: register of patient activities (ie, apps providing calendars for patients to add appointments and treatment administration) and contact with health care professionals.

A descriptive analysis was developed, with continuous and discrete variables presented as mean (SD) and n (%), respectively. The means of continuous variables were compared using the Mann-Whitney *U* test. Results with a *P* value <.05 were considered statistically significant. The data generated were analyzed using IBM SPSS Statistics (version 26.0; IBM Corp).

App Quality Evaluation

To evaluate app quality, the Mobile Application Rating Scale (MARS) was used [24]. The MARS is a validated system to assess health apps, which has been described as the most comprehensive for evaluating technical information and capabilities of apps [24-26]. This tool has been widely used to evaluate health apps designed for many diseases [17,27-29].

The MARS comprises 23 items across 5 subscales:

- Engagement: evaluates the entertainment, interest, customization, interactivity, and adequacy of the target group.
- Functionality: assesses the performance, ease of use, navigation, and gestural design of the app.
- Aesthetics: examines the layout, graphics, and visual appeal of the app.
- Information: assesses the accuracy of the description, establishment of goals, quality and quantity of information, visual information, credibility of the source, and evidence-based development of the app.
- Subjective quality: determines willingness to recommend the app, times the apps will be used, willingness to pay, and overall rating of the app.

Each criterion was evaluated from 1 to 5 (1=inadequate; 2=poor; 3=acceptable; 4=good; 5=excellent). A mean score of the 5 subscales was calculated to describe overall quality.

Mean scores of the engagement, functionality, aesthetics, and information quality objective subscales were calculated, and an overall mean app quality total score was determined. Mean scores were used instead of total scores to accommodate items rated as "Not applicable" and align with familiar star rating formats. The subjective quality items were scored separately or averaged into a mean subjective quality score. This separation ensured that the inherently subjective nature of these measures, based solely on the evaluators' personal opinions and without any predefined criteria, did not unduly influence the overall objective quality assessment. This approach provides a clearer distinction between subjective perceptions and objective functionality and content quality [24].

Before the app search was carried out, reviewers read and became acquainted with the MARS tool. All authors then discussed each rating criterion to achieve a consensus on how to apply it. The first app included was evaluated concurrently by all reviewers to ensure a common understanding and application of the MARS tool. Subsequently, each app included in the study was independently evaluated by 2 reviewers. The scores of each item from these independent evaluations were then compared. In cases where the initial assessments aligned, the average of these scores was calculated to establish the overall mean app quality total score. In instances of significant discrepancy between the 2 primary reviewers' scores—defined as a difference greater than 1 point—where a consensus could not be reached through discussion, a third reviewer was consulted. This third reviewer provided an independent assessment, and the overall mean app quality score was determined by averaging the 2 closest scores among the 3 reviewers.

The MARS was utilized to assess the quality of the apps identified in searches conducted in October 2022 and 2023. For apps identified in both searches, a separate evaluation was conducted after each search if there had been an update to the app. This approach was adopted to examine the changes implemented and to determine whether there was a variation in the app's quality.

Ethical Considerations

This study did not involve human participants or personal data collection. According to local policies, ethics approval was not required for this research.

Results

Description

In the scope of this study, searches conducted in October 2022 and October 2023 resulted in the cumulative identification of 21 apps. The initial search in 2022 identified 19 apps, from a total of 835 apps screened (Figure 1). Of these 19 apps, only 7 received version updates during 2022.

The subsequent search in 2023 identified 16 apps, of which 14 had been initially identified in the 2022 search and remained available. Only 7 of these apps had received version updates in 2023, as compared to their status after the first search in 2022. Moreover, the 2023 search revealed the introduction of 2 new apps unavailable in the preceding year. Table 1 provides detailed general characteristics of the apps included in this analysis.

Almost half of the assessed apps received funding from the pharmaceutical industry (10/21, 48%), and 4 (19%) were developed by the pharmaceutical industry. This influence was observed mostly in the iOS platform apps.

The primary objective of most of the evaluated apps (17/21; 81%) was to provide general information about PH, as well as other pulmonary pathologies in some instances (n=8; 38%). Of the apps targeted at health care professionals, 10 allowed patient assessments through the use of calculators, facilitating the evaluation/classification of patients. There were 2 apps (6 min. test and Pulmonary hypertension) that provided information and tools to assess pediatric patients with PH. None of the apps facilitated the recording of clinical parameters for patient monitoring. Regarding the apps intended for patient use, none permitted users to register activities or contact health care professionals.

For the apps found in the 2022 search, health care professionals contributed to the development or design of 13 of 19 apps (68%). By the 2023 search, health care professionals were involved in 11 of 16 apps (69%). Regarding patient engagement in the development or design process of these apps, such details were not explicitly stated in the descriptions of any evaluated apps. However, the apps "phaware" and "phaware: Aware That I'm Rare," developed by Phaware Global Association (a collective of patients, caregivers, and medical professionals), suggest a potential for patient involvement in designing these apps. Table 2 describes the main characteristics of the assessed apps.

Figure 1. PRISMA flow diagram illustrating the initial search conducted in 2022. PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

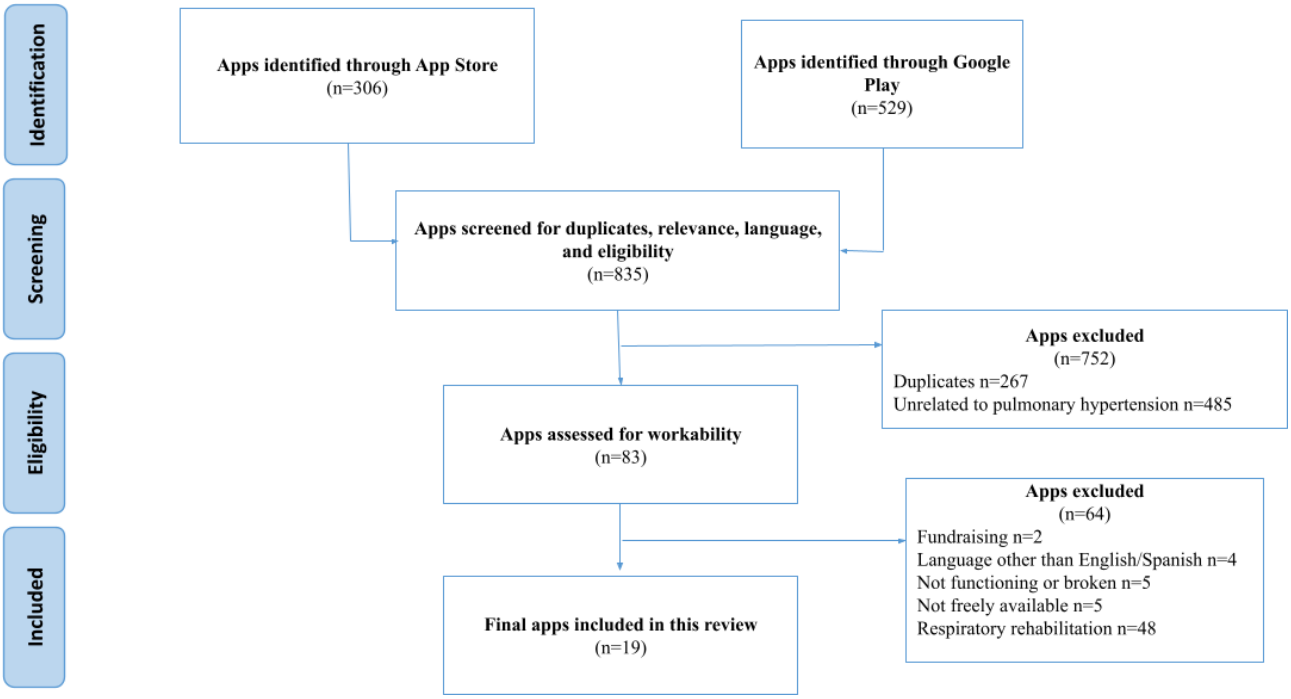


Table . General characteristics of apps.

App		Year			Platform		Language	
Name	Developer	Search	Publication	Last update	iOS	Android	English	Spanish
6 min.test	PH Austria - Lunghochdruck	2022/23	2020	2020	X		X	X
6 Minute Walker	Appricode	2022/23	2021	2021		X	X	X
ATS journals App	American Thoracic Society Inc	2022/23	2013	2021	X		X	
Detect PAH	Janssen EMEA	2022	2022	2022	X		X	
Dx3D	Merck Sharp & Dohme Corp.	2022/23	2021	2023	X		X	
Echoright/Echoright Pro[a]	Janssen EMEA	2022/23	2019	2023	X	X	X	
ePulmonology Review	DKBmed LLC	2022/23	2018	2020	X	X	X	
ESC Pocket Guidelines	ESC European	2022/23	2016	2023	X	X	X	
Hipertensión pulmonar	Goodstore	2022	2021	2021		X		X
Las enfermedades respiratorias	Devo-DreamTeam	2022/23	2020	2023		X	X	X
Lung Diseases and Treatment	StudySpring	2023	2023	2023		X	X	
MSD PH art	Merck Sharp & Dohme LLC	2022/23	2022	2023	X		X	
Phaware: Aware That I'm Rare	Phaware global association	2022/23	2016	2023	X		X	
Phaware	Phaware global association	2022/23	2014	2023	X		X	
Pulmonary & Diseases	Medico_guide	2022	2018	2021		X	X	
Pulmonary hypertension	Corey Chartan	2022/23	2018	2018	X		X	
Pulmonary hypertension support	MyHealthTeams	2022	2018	2019	X		X	
Respiratory Diseases and Treatments	Xstream Apps	2022	2020	2020		X	X	
Respiratory Diseases Treatment	Medi Science	2023	2023	2023		X	X	
Respiratory Disease & Treatment	Patrikat Softech	2022/23	2018	2022		X	X	
Respiratory diseases & Treatment	KIFEDHA APP	2022/23	2020	2021		X	X	

[a]Echoright Pro is the new version (2023) of the Echoright app (2022).

Table . Characteristics of the assessed apps.

Characteristics	2022 search (n=19), n (%)	2023 search (n=16), n (%)	Total (n=21), n (%)
Target audience			
Health professional	9 (47)	8 (50)	9 (43)
Patients/general population	10 (53)	8 (50)	12 (57)
Contain advertisements	7 (37)	6 (38)	9 (43)
Industry-funded [a]	10 (53)	8 (50)	10 (48)
Type of developer			
Commercial/private companies	9 (47)	7 (44)	11 (52)
Scientific/patient society/non-governmental organization	6 (32)	6 (38)	6 (29)
Pharmaceutical industry	4 (21)	3 (19)	4 (19)
Age group			
All audiences	5 (26)	5 (31)	7 (44)
>4 years	4 (21)	4 (25)	4 (19)
>12 years	5 (26)	4 (25)	5 (24)
>17 years	5 (26)	3 (19)	5 (24)
Technical aspects of apps			
Allows sharing (Facebook, Twitter, etc)	13 (68)	10 (62)	15 (71)
Requires login	8 (42)	6 (38)	8 (38)
Needs web access to function	9 (48)	7 (44)	9 (43)

[a]Apps that have received funding or sponsorship from the pharmaceutical industry, irrespective of whether the developer is the pharmaceutical industry or not.

Assessment of App Quality

The overall MARS app quality was considered acceptable in both 2022 and 2023, with mean scores of 3.1 (SD 0.6) and 3.3 (SD 0.5), respectively. Table 3 presents the collected MARS overall quality scores for reference.

The quality assessment of the apps, as detailed in Multimedia Appendices 1 and 2, presents the mean and standard deviations for each MARS item score for the 2022 and 2023 searches, respectively. The engagement scores exhibited significant variation across the apps. Notably, 58% (11/19) of the apps in 2022 and 56% (9/16) in 2023 scored below 3, with the lowest ratings observed in customization and interactivity for both years. In contrast, the functionality category received the highest ratings, achieving a consistent score of 3.8 in both 2022 and 2023. A notable 53% (n=10) of apps in 2022 and 69% (n=11) in 2023 were evaluated with scores of 4 or higher, reflecting commendable performance characterized by ease of use and intuitive navigation. Aesthetically, all evaluated apps received favorable ratings, attributed to their well-designed layouts, graphical design, and visual appeal.

In terms of informational content, most of the apps offered clear descriptions of their features in the app store. However, a significant portion lacked specific objectives, with 47% (n=9) in 2022 and 75% (n=12) in 2023 falling into this category. Evaluations generally favored the quality of the information provided over the quantity, indicating a preference for well-articulated and purposeful content rather than a larger volume of less specific information.

In terms of subjective quality, this domain was the lowest rated in the MARS assessment for both years, with a score of 1.9. Approximately 63% (n=12) of the apps in 2022 and 50% (n=8) in 2023 received a score of 2 or lower, indicating that they were not highly recommendable. Furthermore, 42% (n=8) of the apps in 2022 and 44% (n=7) in 2023 achieved a 3-point score for frequency of use. Notably, when responding to the MARS subjective quality item “Would you pay for this app?” none of the evaluators found the apps worthy of purchase.

Table . The Mobile Application Rating Scale overall quality scores.

App name and year	Engagement, mean (SD)	Functionality, mean (SD)	Aesthetics, mean (SD)	Information, mean (SD)	Subjective quality, mean (SD)	Overall score, mean (SD)
6 min.test						
2022/23[a]	2.4 (0.5)	4.0 (0)	3.3 (0.6)	3.2 (0.8)	1.8 (0.5)	3.2 (0.7)
6 Minute Walker						
2022/23[a]	2.6 (1.1)	4.0 (0)	3.0 (0)	2.4 (0.9)	2.0 (0.8)	3.0 (0.7)
ATS journals APP						
2022/23[a]	3.2 (1.1)	4.0 (0)	3.3 (0.6)	4.2 (0.4)	2.5 (1.0)	3.7 (0.5)
Detect PAH						
2022	2.4 (1.1)	3.8 (0.5)	3.3 (0.6)	2.8 (0.8)	1.5 (0.6)	3.1 (0.6)
Dx3D						
2022	3 (1.4)	3.8 (0.5)	3.7 (0.6)	3.7 (0.5)	2.3 (1)	3.52 (0.3)
2023	3.2 (1.1)	4.0 (0)	4.0 (0)	3.7 (0.5)	2.3 (1)	3.7 (0.4)
Echoright/Echoright Pro [b]						
2022	3.2 (1.3)	4.0 (0)	4.0 (0)	3.7 (0.5)	2.3 (1)	3.72 (0.4)
2023	3.2 (1.3)	4.0 (0)	4.0 (0)	3.8 (1)	2.3 (1)	3.8 (0.4)
ePulmonology Review						
2022/23[a]	2.8 (1.3)	4.0 (0)	4.0 (0)	3.4 (0.5)	2.3 (1)	3.55 (0.6)
ESC Pocket Guidelines						
2022	3.2 (1.3)	4.3 (0.5)	3.0 (0)	3.0 (0)	2 (0.8)	3.4 (0.6)
2023	3.4 (1.1)	4.3 (0.5)	4.0 (0)	3.6 (0.9)	2.5 (1)	3.8 (0.4)
Hipertensión pulmonar						
2022	1.8 (0.8)	3.5 (0.6)	2.0 (0)	2.2 (0.8)	1.3 (0.5)	2.4 (0.8)
Las enfermedades respiratorias						
2022	2.6 (0.5)	3.8 (0.5)	2.7 (0.6)	3.2 (1.3)	1.5 (0.6)	3.1 (0.5)
2023	2.8 (0.4)	3.8 (0.5)	3.0 (0)	2.8 (1.1)	1.5 (0.6)	3.1 (0.5)
Lung Diseases and Treatment						
2023	2.2 (1.1)	4.0 (0)	3.7 (0.6)	3.0 (1.2)	1.5 (0.6)	3.2 (0.8)
MSD PH art						
2022	3.2 (1.3)	4.0 (0)	4.0 (0)	3.6 (0.5)	3 (1.4)	3.7 (0.4)
2023	3.2 (0.8)	4.3 (0.5)	4.0 (0)	3.5 (1.3)	2.5 (1)	3.7 (0.5)
phaware						
2022/23[a]	2.8 (0.8)	3.3 (0.5)	3.0 (0)	4.0 (1.4)	2.0 (0.8)	3.3 (0.5)
phaware: Aware That I´m Rare						
2022/23[a]	3.2 (0.4)	4.0 (0)	3.7 (0.6)	4.0 (0)	2.3 (1)	3.7 (0.4)
Pulmonary hypertension						
2022/23[a]	3 (1.4)	4.8 (0.5)	3.7 (0.6)	3.2 (0.8)	2.3 (1)	3.7 (0.8)
Pulmonary hypertension support						
2022	3.4 (0.9)	3.8 (0.5)	3.7 (0.6)	3.6 (0.5)	2.5 (1)	3.6 (0.1)
Pulmonary & Diseases						
2022	1.6 (0.5)	3.5 (1)	1.7 (0.6)	1.7 (0.8)	1.0 (0)	2.1 (0.9)
Respiratory Diseases and Treatments						

App name and year	Engagement, mean (SD)	Functionality, mean (SD)	Aesthetics, mean (SD)	Information, mean (SD)	Subjective quality, mean (SD)	Overall score, mean (SD)
2022	2.2 (1.1)	4.0 (0)	2.7 (0.6)	2.7 (1)	1.5 (0.6)	2.9 (0.8)
Respiratory Diseases Treatment						
2023	1.6 (0.5)	3.5 (0.6)	3.3 (0.6)	2.4 (0.9)	1.0 (0)	2.7 (0.9)
Respiratory Disease & Treatment						
2022	1.2 (0.4)	2.5 (0.6)	1.7 (0.6)	1.8 (0.8)	1 (0)	1.8 (0.5)
2023	1.8 (0.8)	2.8 (0.5)	2.7 (0.6)	2.6 (0.4)	1.3 (0.5)	2.5 (0.4)
Respiratory diseases & Treatment						
2022/23[a]	1.2 (0.4)	3.0 (0)	2.0 (0)	1.8 (0.8)	1.0 (0)	2.0 (0.7)
Total, mean (SD)						
2022	2.58 (0.7)	3.8 (0.5)	3.1 (0.8)	3.1 (0.8)	1.9 (0.6)	3.1 (0.6)
2023	2.7 (0.7)	3.8 (0.5)	3.4 (0.6)	3.2 (0.7)	1.9 (0.5)	3.3 (0.5)

[a]The app remained available in both years; however, a reevaluation was not conducted due to the absence of any updates since the initial assessment.
[b]Echoright Pro is the new version (2023) of the Echoright app (2022).

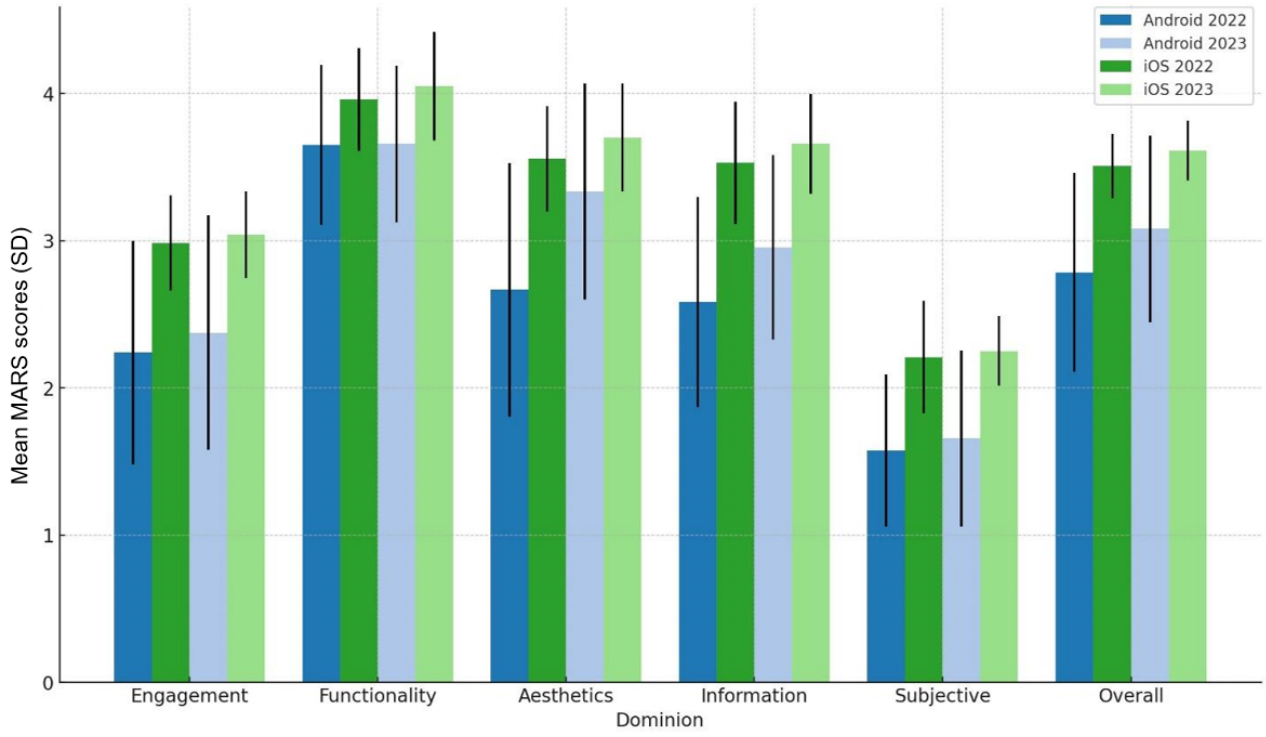
Evolution of Apps Across Two Years

The mean overall MARS score exhibited a slight improvement from 2022 to 2023 (+0.2). However, this variation could be considered not statistically significant, suggesting overall quality remained stable across the evaluated years.

Multimedia Appendix 2 details the differences in scores based on the evaluated platform. Specifically, iOS apps consistently received higher ratings compared to Android apps in terms of the overall mean MARS score for both years, with a difference of +0.7 (*P*=.008) in 2022 and +0.5 (*P*=.03) in 2023. The 2022 search revealed statistically significant differences in the domains of engagement (+0.8, *P*=.02), aesthetics (+0.9, *P*=.01), information (+0.9, *P*=.005), and subjective quality (+0.6, *P*=.008). In contrast, the 2023 search found significant differences only in the information (+0.8, *P*=.01) and subjective quality (+0.6, *P*=.03) domains. Figure 2 illustrates these variations in each MARS domain across the 2 years (2022 - 2023).

Regarding the involvement of health care professionals, apps that incorporated their input in design or development showed higher overall MARS ratings in both the 2022 (+1.1, *P*<.001) and 2023 (+0.9, *P*=.002) searches. Enhanced ratings were observed across all 5 domains in both searches, as detailed in Multimedia Appendix 3.

Figure 2. Quality differences between platforms across the 2-year evaluations.



Discussion

Main Results

Health apps have the potential to become a standard component of care for individuals with chronic conditions. These apps have a demonstrated beneficial impact on patients' health, including enhanced adherence to prescribed medications, improved quality of life, and reduced reliance on health care services [16,30-32]. To our knowledge, this is the first study that systematically evaluated and compared the quality of free mHealth apps aimed at patients with PH and health care professionals over 2 years using a cross-sectional approach, with the aim of examining the progression in the quality and features of PH apps.

In this systematic evaluation, 21 unique PH apps were identified, including 19 in 2022 and 16 in 2023. The overall mean app quality score improved from 3.1 (SD 0.6) in 2022 to 3.3 (SD 0.5) in 2023. According to the MARS 5-point scale, a score above 3, which denotes "acceptable" quality, generally indicates that an app is functional and offers a satisfactory user experience but may still lack certain advanced features or innovative design elements that could elevate its rating to "excellent." However, it is important to note that the MARS framework does not provide explicit guidelines on interpreting the overall mean app quality total score, leaving some ambiguity in assessing app quality [24,25]. Therefore, we acknowledge that for health care apps, especially for complex conditions like PH, more stringent quality standards for "acceptable" quality could be necessary, such as a cutoff score of 3.5 (indicating above-average quality), as defined by Terhorst et al, which may better reflect the standards needed for effective health care management apps [33].

Throughout the 2-year study, from the 19 apps initially identified in 2022, only 14 remained available in 2023. Notably, these persisting apps scored above the median on the MARS, suggesting that higher quality may have contributed to their continued availability in the market. The decrease in availability of the remaining apps calls for further investigation. Potential reasons for this decline could include low overall quality, poor user engagement, lack of economic viability, regulatory challenges, obsolescence in content or technology, organizational changes, or negative performance and reviews [34-36]. In contrast, most apps that remained available over the 2-year period showed durability, although with a relatively low frequency of updates. The minimal changes observed, even after updates, indicate a lack of significant advancement over the year. This might indicate a continuing interest in the digital management of PH but also points to a potential lack of advancement, a significant issue in the fast-evolving mHealth sector [37]. The limited updates could result from funding constraints or inadequate user feedback, both critical for driving innovation in this area.

Simultaneously, the involvement of health care professionals in creating mHealth apps emerged as a crucial factor influencing both the quality and the accuracy of the information provided [17,28,38]. A systematic analysis focusing on the contribution of experts and their compliance with medical evidence in mHealth apps revealed a notable trend: many health apps lack health care professional input and are not consistently grounded in evidence-based information [38]. This is noteworthy considering that the involvement of health care experts in the app development process does not automatically guarantee the inclusion of evidence-based content [38,39]. However, in the context of PH, our review found that more than two-thirds of PH-related apps were developed with the active involvement of health care professionals, correlating with higher app quality scores (2022: 13/19, 68%; 2023: 11/16, 69%). These findings suggest that health care professionals are a key part in developing high-quality PH-related apps.

Moreover, specific app functionalities exhibit significant potential for health care professionals managing PH. Key features such as pulmonary arterial pressure monitoring for patients with implanted devices and easy access to up-to-date PH-specific treatment guidelines are distinctly advantageous [40,41]. Within the scope of this review, 2 apps provided access to PH-specific treatment guidelines ("ESC pocket guidelines" and "ATS journals APP"), and 10 apps included risk assessment tools. However, a critical issue is the infrequent updates; only half of the apps from the year 2022 were updated that year, and just 43% of those available in 2023 had received recent updates, underscoring a significant gap in maintaining current information.

Mobile apps targeted at patients with PH could profoundly impact health outcomes by enhancing self-management, monitoring, and communication [16,30,42]. Such apps could aid patients with PH through various functionalities like symptom tracking, ensuring medication adherence, enabling remote monitoring, offering patient education, creating support networks, facilitating telemedicine, and providing mental health support [43-45]. However, our findings reveal that the majority of apps lack direct patient engagement in their design, resulting in the omission of key features like direct contact with health care professionals and patient activity registration. Notably, only 2 apps, "Pulmonary Hypertension Support" and "phaware," provided community access, with the latter doing so through external links. This gap highlights the need for greater patient involvement in app development to ensure the tools meet the actual needs of end users, potentially through user-centered design approaches [46,47]. The apparent mismatch between app functionalities and patient needs suggests an area ripe for research, aimed at aligning app development more closely with patient priorities [48-50].

A significant finding from our systematic evaluation is the absence of proven clinical efficacy among PH apps, highlighted by the fact that none underwent formal testing in clinical trials for their efficacy in PH management—a stark contrast to chronic conditions like heart failure or chronic obstructive pulmonary disease, where mHealth interventions have demonstrated clinical benefits [51-54]. Notably, apps like "Detect PAH" by Janssen EMEA, which includes a clinically validated diagnostic algorithm [55-57], and the "6 min.test" app by PH Austria, featuring a pediatric-adapted 6-minute walking distance test [58-60], showed promise but lacked validation in broader clinical settings. This lack of broader clinical validation is consistent with findings from other studies assessing health apps [17,28,61]. Further research is crucial to comprehensively

evaluate these apps prior to their integration into standard PH management, considering the potential for eHealth to add to the treatment burden, especially for vulnerable patients [62,63].

Additional Considerations

Beyond the limitations identified in this systematic evaluation, free applications for patients with PH and caregivers should also focus on enhancing interoperability and ensuring robust data privacy.

Enhancing interoperability with existing health care systems is essential to maximize the utility of free health apps. Interoperability ensures seamless communication with other health information systems, which is crucial for real-time updates and integration of the latest treatment guidelines and diagnostic tools. This connectivity keeps apps relevant and effective for both patients and health care providers [64]. By adhering to established health data standards such as Health Level 7 [65] and Fast Healthcare Interoperability Resources [66], these free apps can improve their adaptability and usefulness, facilitating better clinical decision-making and improving patient outcomes [51-54].

However, free apps often pose significant privacy and security risks. According to a report by Timeero [67], free apps tend to share more user data compared to paid apps, increasing the risk of data breaches and unauthorized access. The report indicates that free apps share, on average, 7 times more data points than their paid counterparts, often monetizing user data by selling it to third parties. This practice raises substantial concerns, especially given the sensitive nature of health data. App developers need to implement robust data protection measures to safeguard patient information. This involves encryption of data both at rest and in transit, secure user authentication mechanisms, and regular security audits to prevent unauthorized access and data breaches [68]. Additionally, compliance with regulatory requirements such as the General Data Protection Regulation [69] in Europe or the Health Insurance Portability and Accountability Act [70] in the United States is critical. For free apps, which may lack the financial backing of paid apps, ensuring these privacy and security measures can be challenging but is necessary to gain user trust and ensure safe use in a health care setting.

Limitations

This systematic evaluation is subject to limitations. The evaluation process utilizes the MARS framework [24,26], which introduces a degree of subjectivity in app categorization. However, the dual-reviewer methodology applied in this study likely reduces the impact of this subjectivity. In addition, the MARS framework, while being a widely recognized tool for app quality assessment [17,27,28,71], does not account for crucial factors such as privacy, security, and update frequency, which are critical for comprehensive health software quality assessment [72,73].

In addition, our perspective as health care professionals may influence our interpretation of patient-oriented apps' relevance and utility for patients, potentially overlooking user-specific needs and preferences [49,50]. Furthermore, our analysis was primarily based on app store descriptions and publicly available information, which limits our ability to verify the extent or nature of health care professional or patient involvement in the development of these apps. Additionally, the methodology involved systematic searches conducted via mobile devices in order to simulate the experience of app users. This approach potentially introduces bias based on app store search algorithms and the variable availability of apps. Finally, the review was confined to apps available on Google Play and the App Store. Although these are the major platforms covering most market-available apps [74], other platforms such as Windows Phone and Blackberry Market were not included due to the unavailability of devices for the reviewers, which may have led to the omission of certain PH-related apps.

Conclusions

This systematic evaluation of PH-related mHealth apps targeted at patients and health care professionals has provided critical insights into the current state and potential of these tools. Our analysis identified a relatively small cohort of apps, with an average quality score deemed acceptable, yet revealing a considerable gap in the development and refinement of these digital resources. Furthermore, the involvement of health care professionals in the design and development of mHealth apps was found to be a significant factor in enhancing the quality and reliability of the information these tools provide. Despite this positive impact, our study observed a stagnation in the evolution of app quality and features over 2 years, highlighting a critical need for continuous innovation and improvement. Additionally, our findings emphasize the lack of clinical validation for most of these apps, which is a pivotal step that should not be overlooked if these tools are to be integrated into standard patient care.

In conclusion, while the landscape of mHealth apps for PH shows promise, it remains underdeveloped. Future efforts should focus on leveraging the expertise of health care professionals, integrating patient perspectives, and ensuring that apps undergo rigorous clinical validation. Such measures would significantly enhance the utility, relevance, and effectiveness of mHealth tools, potentially transforming them into indispensable components of PH management. This evolution requires a concerted effort from stakeholders to realize the full potential of mHealth in supporting the complex needs of patients with PH.

Conflicts of Interest

None declared.

Multimedia Appendix 1

The Mobile Application Rating Scale—engagement, functionality, aesthetics, and information domains (2022 search).
[DOCX File, 25 KB - mhealth_v12i1e57289_app1.docx]

Multimedia Appendix 2
Differences between platform scores.
[DOCX File, 7 KB - mhealth_v12i1e57289_app2.docx]

Multimedia Appendix 3
Differences between health care professionals' participation in the apps' development.
[DOCX File, 7 KB - mhealth_v12i1e57289_app3.docx]

Multimedia Appendix 4
The Mobile Application Rating Scale—engagement, functionality, aesthetics, and information domains (2023 search).
[DOCX File, 21 KB - mhealth_v12i1e57289_app4.docx]

References

1. Humbert M, Kovacs G, Hoeper MM, et al. 2022 ESC/ERS Guidelines for the diagnosis and treatment of pulmonary hypertension. Eur Heart J 2022 Oct 11;43(38):3618-3731. [doi: 10.1093/eurheartj/ehac237] [Medline: 36017548]
2. Barberà JA, Escribano P, Morales P, et al. Standards of care in pulmonary hypertension. Consensus statement of the Spanish Society of Pulmonology and Thoracic Surgery (SEPAR) and the Spanish Society of Cardiology (SEC) [Article in Spanish]. Rev Esp Cardiol 2008 Feb;61(2):170-184. [doi: 10.1016/S1885-5857(08)60093-6] [Medline: 18364186]
3. Wu RC, Lo V, Rossos P, et al. Improving hospital care and collaborative communications for the 21st century: key recommendations for general internal medicine. Interact J Med Res 2012 Sep 24;1(2):e9. [doi: 10.2196/ijmr.2022] [Medline: 23612055]
4. Makoul G. Essential elements of communication in medical encounters: the Kalamazoo consensus statement. Acad Med 2001 Apr;76(4):390-393. [doi: 10.1097/00001888-200104000-00021] [Medline: 11299158]
5. Dong TX, Zhu Q, Wang ST, et al. Diagnostic and prognostic value of echocardiography in pulmonary hypertension: an umbrella review of systematic reviews and meta-analyses. BMC Pulm Med 2023 Jul 10;23(1):253. [doi: 10.1186/s12890-023-02552-y] [Medline: 37430308]
6. SEC Working Group for the 2022 ESC/ERS guidelines for the diagnosis and treatment of pulmonary hypertension and SEC Guidelines Committee. Comments on the 2022 ESC/ERS guidelines for the diagnosis and treatment of pulmonary hypertension. Rev Esp Cardiol (Eng Ed) 2023 May;76(5):294-300. [doi: 10.1016/j.rec.2022.11.001]
7. Salvi D, Poffley E, Tarassenko L, Orchard E. App-based versus standard six-minute walk test in pulmonary hypertension: mixed methods study. JMIR mHealth uHealth 2021 Jun 7;9(6):e22748. [doi: 10.2196/22748] [Medline: 34096876]
8. Salvi D, Poffley E, Orchard E, Tarassenko L. The mobile-based 6-minute walk test: usability study and algorithm development and validation. JMIR mHealth uHealth 2020 Jan 3;8(1):e13756. [doi: 10.2196/13756] [Medline: 31899457]
9. Collado-Borrell R, Escudero-Vilaplana V, Calles A, et al. Oncology patient interest in the use of new technologies to manage their disease: cross-sectional survey. J Med Internet Res 2018 Oct 23;20(10):e11006. [doi: 10.2196/11006] [Medline: 30355554]
10. Sun L, Jiao C, Wang Y, Xiao Q, Zhang Y, Wu Y. A survey on the willingness to use physical activity smartphone applications (apps) in patients with chronic diseases. Stud Health Technol Inform 2016;225:1032-1033. [Medline: 27332468]
11. Jezrawi R, Balakumar S, Masud R, et al. Patient and physician perspectives on the use and outcome measures of mHealth apps: exploratory survey and focus group study. Dig Health 2022;8:20552076221102773. [doi: 10.1177/20552076221102773] [Medline: 35646382]
12. Uncovska M, Freitag B, Meister S, Fehring L. Patient acceptance of prescribed and fully reimbursed mHealth apps in Germany: an UTAUT2-based online survey study. J Med Syst 2023 Jan 27;47(1):14. [doi: 10.1007/s10916-023-01910-x] [Medline: 36705853]
13. Yang K, Li Y, Qi H. Determinants of and willingness to use and pay for digital health technologies among the urban elderly in Hangzhou, China. Risk Manag Healthc Policy 2023;16:463-478. [doi: 10.2147/RMHP.S393767] [Medline: 37007299]
14. Houser SH, Flite CA, Foster SL. Privacy and security risk factors related to telehealth services - a systematic review. Perspect Health Inf Manag 2023;20(1):1f. [Medline: 37215337]
15. Miró J, Llorens-Vernet P. Assessing the quality of mobile health-related apps: interrater reliability study of two guides. JMIR mHealth uHealth 2021 Apr 19;9(4):e26471. [doi: 10.2196/26471] [Medline: 33871376]
16. Gonzalez-Garcia MC, Fatehi F, Varnfield M, et al. Use of eHealth in the management of pulmonary arterial hypertension: review of the literature. BMJ Health Care Inform 2020 Sep;27(3):e100176. [doi: 10.1136/bmjhci-2020-100176] [Medline: 32928780]

17. Báez Gutiérrez N, Rodríguez Ramallo H, Fernández González M, Abdel-Kader Martín L. Smartphone apps for patients with hematologic malignancies: systematic review and evaluation of content. JMIR mHealth uHealth 2022 Sep 20;10(9):e35851. [doi: 10.2196/35851] [Medline: 36125860]
18. Devan H, Farmery D, Peebles L, Grainger R. Evaluation of self-management support functions in apps for people with persistent pain: systematic review. JMIR mHealth uHealth 2019 Feb 12;7(2):e13080. [doi: 10.2196/13080] [Medline: 30747715]
19. Bunova A, Wiemker V, Gornyi B, Ferreira-Borges C, Neufeld M. Russian-language mobile apps for reducing alcohol use: systematic search and evaluation. JMIR mHealth uHealth 2022 Jan 10;10(1):e31058. [doi: 10.2196/31058] [Medline: 35006083]
20. Gerner M, Vuillerme N, Aubourg T, et al. Review and analysis of German mobile apps for inflammatory bowel disease management using the Mobile Application Rating Scale: systematic search in app stores and content analysis. JMIR mHealth uHealth 2022 May 3;10(5):e31102. [doi: 10.2196/31102] [Medline: 35503246]
21. McKay F, Chan L, Cerio R, et al. Assessing the quality and behavior change potential of vaping cessation apps: systematic search and assessment. JMIR mHealth uHealth 2024 Mar 15;12:e55177. [doi: 10.2196/55177] [Medline: 38532616]
22. Toïgo M, Marc J, Hayot M, Moulis L, Carbonnel F. Quality assessment of smartphone medication management apps in France: systematic search. JMIR mHealth uHealth 2024 Mar 18;12:e54866. [doi: 10.2196/54866] [Medline: 38498042]
23. Kim BY, Sharafoddini A, Tran N, Wen EY, Lee J. Consumer mobile apps for potential drug-drug interaction check: systematic review and content analysis using the Mobile App Rating Scale (MARS). JMIR mHealth uHealth 2018 Mar 28;6(3):e74. [doi: 10.2196/mhealth.8613] [Medline: 29592848]
24. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile App Rating Scale: a new tool for assessing the quality of health mobile apps. JMIR mHealth uHealth 2015 Mar 11;3(1):e27. [doi: 10.2196/mhealth.3422] [Medline: 25760773]
25. Terhorst Y, Philippi P, Sander LB, et al. Validation of the Mobile Application Rating Scale (MARS). PLoS ONE 2020;15(11):e0241480. [doi: 10.1371/journal.pone.0241480] [Medline: 33137123]
26. Dawson RM, Felder TM, Donevant SB, et al. What makes a good health “app”? Identifying the strengths and limitations of existing mobile application evaluation tools. Nurs Inq 2020 Apr;27(2):e12333. [doi: 10.1111/nin.12333] [Medline: 31854055]
27. Masterson Creber RM, Maurer MS, Reading M, Hiraldo G, Hickey KT, Iribarren S. Review and analysis of existing mobile phone apps to support heart failure symptom monitoring and self-care management using the Mobile Application Rating Scale (MARS). JMIR mHealth uHealth 2016 Jun 14;4(2):e74. [doi: 10.2196/mhealth.5882] [Medline: 27302310]
28. Kwan YH, Ong WJ, Xiong M, et al. Evaluation of mobile apps targeted at patients with spondyloarthritis for disease monitoring: systematic app search. JMIR mHealth uHealth 2019 Oct 28;7(10):e14753. [doi: 10.2196/14753] [Medline: 31661080]
29. Ramsey RR, Caromody JK, Voorhees SE, et al. A systematic evaluation of asthma management apps examining behavior change techniques. J Allergy Clin Immunol Pract 2019;7(8):2583-2591. [doi: 10.1016/j.jaip.2019.03.041] [Medline: 30954644]
30. Collado-Borrell R, Escudero-Vilaplana V, Narrillos-Moraza Á, Villanueva-Bueno C, Herranz-Alonso A, Sanjurjo-Sáez M. Patient-reported outcomes and mobile applications. A review of their impact on patients’ health outcomes. Farm Hosp 2022 Apr 27;46(3):173-181. [Medline: 36183211]
31. Casey L, Clough B, Green H, et al. Computer-based interactive health communications for people with chronic disease. SHTT 2014 Apr;2:29. [doi: 10.2147/SHTT.S42684]
32. Debon R, Coleone JD, Bellei EA, De Marchi ACB. Mobile health applications for chronic diseases: a systematic review of features for lifestyle improvement. Diabetes Metab Syndr Clin Res Rev 2019 Jul;13(4):2507-2512. [doi: 10.1016/j.dsx.2019.07.016]
33. Terhorst Y, Rathner EM, Baumeister H, Sander L. «Hilfe aus dem App-Store?»: Eine systematische Übersichtsarbeit und Evaluation von Apps zur Anwendung bei Depressionen [Article in German]. Verh Ther 2018 Feb 1;28(2):101-112. [doi: 10.1159/000481692]
34. van Velthoven M, Powell J. Do health apps need endorsement? Challenges for giving advice about which health apps are safe and effective to use. D Health 2017;3:2055207617701342. [doi: 10.1177/2055207617701342] [Medline: 29942588]
35. Pagliari C. Commercial health apps: in the user’s interest? BMJ 2019 Mar 21;364:l1280. [doi: 10.1136/bmj.l1280] [Medline: 30898789]
36. Mulder T. Health apps, their privacy policies and the GDPR. Eur J Law Technol 2019;10(1) [FREE Full text]
37. Zakerabasali S, Ayyoubzadeh SM, Baniasadi T, Yazdani A, Abhari S. Mobile health technology and healthcare providers: systemic barriers to adoption. Healthc Inform Res 2021 Oct;27(4):267-278. [doi: 10.4258/hir.2021.27.4.267] [Medline: 34788907]
38. Subhi Y, Bube SH, Rolskov Bojsen S, Skou Thomsen AS, Konge L. Expert involvement and adherence to medical evidence in medical mobile phone apps: a systematic review. JMIR mHealth uHealth 2015 Jul 27;3(3):e79. [doi: 10.2196/mhealth.4169] [Medline: 26215371]

39. Dittrich F, Albrecht UV, Scherer J, et al. Development of open backend structures for health care professionals to improve participation in app developments: pilot usability study of a medical app. JMIR Form Res 2023 Apr 13;7:e42224. [doi: 10.2196/42224] [Medline: 37052998]
40. McBeath KCC, Angermann CE, Cowie MR. Digital technologies to support better outcome and experience of care in patients with heart failure. Curr Heart Fail Rep 2022 Jun;19(3):75-108. [doi: 10.1007/s11897-022-00548-z] [Medline: 35486314]
41. Wu S, Hoang HB, Yang JZ, et al. Drug-drug interactions in the management of patients with pulmonary arterial hypertension. Chest 2022 Dec;162(6):1360-1372. [doi: 10.1016/j.chest.2022.06.042] [Medline: 35841932]
42. Stubbs H, Brewis M, Church C, Johnson M. Towards telemedicine in pulmonary hypertension: assessing the feasibility of remote quality of life and exercise capacity assessment. Pulm Circ 2022 Oct;12(4):e12144. [doi: 10.1002/pul2.12144] [Medline: 36381292]
43. Liu S, Li J, Wan DY, et al. Effectiveness of eHealth self-management interventions in patients with heart failure: systematic review and meta-analysis. J Med Internet Res 2022 Sep 26;24(9):e38697. [doi: 10.2196/38697] [Medline: 36155484]
44. Pérez Peñate G, Ochoa Parra N, Domingo Morera JA, et al. Evaluation of a digital health system (PAHcareTM) for routine care of patients with pulmonary arterial hypertension: the CBS-PAH study protocol. Front Public Health 2022;10:954487. [doi: 10.3389/fpubh.2022.954487] [Medline: 36568758]
45. Fitzpatrick PJ. Improving health literacy using the power of digital communications to achieve better health outcomes for patients and practitioners. Front Digit Health 2023;5:1264780. [doi: 10.3389/fdgth.2023.1264780] [Medline: 38046643]
46. Chudyk AM, Ragheb S, Kent D, et al. Patient engagement in the design of a mobile health app that supports enhanced recovery protocols for cardiac surgery: development study. JMIR Perioper Med 2021 Nov 30;4(2):e26597. [doi: 10.2196/26597] [Medline: 34851299]
47. Adus S, Macklin J, Pinto A. Exploring patient perspectives on how they can and should be engaged in the development of artificial intelligence (AI) applications in health care. BMC Health Serv Res 2023 Oct 26;23(1):1163. [doi: 10.1186/s12913-023-10098-2] [Medline: 37884940]
48. Tarricone R, Cucciniello M, Armeni P, et al. Mobile health divide between clinicians and patients in cancer care: results from a cross-sectional international survey. JMIR mHealth uHealth 2019 Sep 6;7(9):e13584. [doi: 10.2196/13584] [Medline: 31493318]
49. Nair SC, Hassen WA, Sreedharan J, Qawasmeh K, Ibrahim H. Comparison of perceptions of unmet supportive care needs between cancer patients and their oncologists. Supp Care Cancer 2019 Nov;27(11):4229-4235. [doi: 10.1007/s00520-019-04705-4] [Medline: 30847701]
50. Liu P, Li X, Zhang XM. Healthcare professionals' and patients' assessments of listed mobile health apps in China: a qualitative study. Front Public Health 2023;11:1220160. [doi: 10.3389/fpubh.2023.1220160] [Medline: 37780445]
51. Ding X, Wen Y, Tian Z, et al. Effect of e-health intervention on disease management in patients with chronic heart failure: a meta-analysis. Front Cardiovasc Med 2022;9:1053765. [doi: 10.3389/fcvm.2022.1053765] [Medline: 36824289]
52. Kitsiou S, Vatani H, Paré G, et al. Effectiveness of mobile health technology interventions for patients with heart failure: systematic review and meta-analysis. Can J Cardiol 2021 Aug;37(8):1248-1259. [doi: 10.1016/j.cjca.2021.02.015] [Medline: 33667616]
53. Kooij L, Vos PJE, Dijkstra A, van Harten WH. Effectiveness of a mobile health and self-management app for high-risk patients with chronic obstructive pulmonary disease in daily clinical practice: mixed methods evaluation study. JMIR mHealth uHealth 2021 Feb 4;9(2):e21977. [doi: 10.2196/21977] [Medline: 33538699]
54. Shaw G, Whelan ME, Armitage LC, Roberts N, Farmer AJ. Are COPD self-management mobile applications effective? A systematic review and meta-analysis. NPJ Prim Care Respir Med 2020 Apr 1;30(1):11. [doi: 10.1038/s41533-020-0167-1] [Medline: 32238810]
55. Coghlan JG, Denton CP, Grünig E, et al. Evidence-based detection of pulmonary arterial hypertension in systemic sclerosis: the DETECT study. Ann Rheum Dis 2014 Jul;73(7):1340-1349. [doi: 10.1136/annrheumdis-2013-203301] [Medline: 23687283]
56. Young A, Moles VM, Jaafar S, et al. Performance of the DETECT algorithm for pulmonary hypertension screening in a systemic sclerosis cohort. Arthritis Rheumatol 2021 Sep;73(9):1731-1737. [doi: 10.1002/art.41732] [Medline: 33760392]
57. Mihai C, Antic M, Dobrota R, et al. Factors associated with disease progression in early-diagnosed pulmonary arterial hypertension associated with systemic sclerosis: longitudinal data from the DETECT cohort. Ann Rheum Dis 2018 Jan;77(1):128-132. [doi: 10.1136/annrheumdis-2017-211480] [Medline: 29061853]
58. Saraff V, Schneider J, Colleselli V, et al. Sex-, age-, and height-specific reference curves for the 6-min walk test in healthy children and adolescents. Eur J Pediatr 2015 Jun;174(6):837-840. [doi: 10.1007/s00431-014-2454-8] [Medline: 25491900]
59. Geiger R, Strasak A, Treml B, et al. Six-minute walk test in children and adolescents. J Pediatr 2007 Apr;150(4):395-399. [doi: 10.1016/j.jpeds.2006.12.052] [Medline: 17382117]
60. Rauchenzauner M, Schneider J, Colleselli V, et al. Comparing modalities of conducting the six-minute walk test in healthy children and adolescents. Minerva Pediatr 2019 Jun;71(3):229-234. [doi: 10.23736/S0026-4946.18.05185-X] [Medline: 31017381]

61. Connor K, Brady RRW, de Beaux A, Tulloh B. Contemporary hernia smartphone applications (apps). Hernia 2014 Aug;18(4):557-561. [doi: 10.1007/s10029-013-1130-7] [Medline: 23801277]
62. Mair FS, Montori VM, May CR. Digital transformation could increase the burden of treatment on patients. BMJ 2021 Nov 25;375:2909. [doi: 10.1136/bmj.n2909] [Medline: 34824093]
63. Mair FS, May CR. Thinking about the burden of treatment. BMJ 2014 Nov 10;349:g6680. [doi: 10.1136/bmj.g6680] [Medline: 25385748]
64. New European interoperability framework: promoting seamless services and data flows for European public administrations. European Commission. 2017. URL: https://ec.europa.eu/isa2/sites/default/files/eif_brochure_final.pdf [accessed 2024-10-16]
65. Introduction to HL7 standards. HL7 International. URL: https://www.hl7.org/implement/standards/ [accessed 2024-10-16]
66. Fast healthcare interoperability resources. eCQI Resource Center. URL: https://ecqi.healthit.gov/fhir?qt-tabs_fhir=about [accessed 2024-10-16]
67. Data security concerns: free vs. paid apps. Timeero. URL: https://timeero.com/post/free-vs-paid-apps#toc-how-can-timeero-help-you- [accessed 2024-10-16]
68. Arora S, Yttri J, Nilse W. Privacy and security in mobile health (mHealth) research. Alcohol Res 2014;36(1):143-152. [Medline: 26259009]
69. Radley-Gardner O, Beale H, Zimmermann R, editors. Fundamental Texts On European Private Law, 2nd edition: Hart Publishing Ltd; 2016. [doi: 10.5040/9781782258674]
70. Modifications to the HIPAA privacy, security, enforcement, and breach notification rules under the Health Information Technology for Economic and Clinical Health Act and the Genetic Information Nondiscrimination Act; other modifications to the HIPAA rules. Federal Register. URL: https://www.federalregister.gov/documents/2013/01/25/2013-01073/modifications-to-the-hipaa-privacy-security-enforcement-and-breach-notification-rules-under-the [accessed 2024-10-16]
71. Tinschert P, Jakob R, Barata F, Kramer JN, Kowatsch T. The potential of mobile apps for improving asthma self-management: a review of publicly available and well-adopted asthma apps. JMIR mHealth uHealth 2017 Aug 2;5(8):e113. [doi: 10.2196/mhealth.7177] [Medline: 28768606]
72. Lewis TL, Wyatt JC. mHealth and mobile medical apps: a framework to assess risk and promote safer use. J Med Internet Res 2014 Sep 15;16(9):e210. [doi: 10.2196/jmir.3133] [Medline: 25223398]
73. Aljedaani B, Babar MA. Challenges with developing secure mobile health applications: systematic review. JMIR mHealth uHealth 2021 Jun 21;9(6):e15654. [doi: 10.2196/15654] [Medline: 34152277]
74. Biggest app stores in the world 2022. Statista. URL: https://www.statista.com/statistics/276623/number-of-apps-available-in-leading-app-stores/ [accessed 2024-10-16]

Abbreviations

MARS: Mobile Application Rating Scale
mHealth: mobile health
PH: pulmonary hypertension

Edited by L Buis; submitted 11.02.24; peer-reviewed by A Hassan, H Stubbs, R Chen; revised version received 01.07.24; accepted 30.08.24; published 30.10.24.

Please cite as:
Báez Gutiérrez N, Rodríguez Ramallo H, Mendoza-Zambrano EM, Brown Arreola B, Santos Ramos B, Abdel-kader Martín L, Otero Candelera R
Smartphone Apps for Pulmonary Hypertension: Systematic Search and Content Evaluation
JMIR Mhealth Uhealth 2024;12:e57289
URL: https://mhealth.jmir.org/2024/1/e57289
doi:10.2196/57289

Mobile Apps for Common Noncommunicable Disease Management: Systematic Search in App Stores and Evaluation Using the Mobile App Rating Scale

Khang Jin Cheah[1*], MSc; Zahara Abdul Manaf[1*], PhD; Arimi Fitri Mat Ludin[1*], PhD; Nurul Huda Razalli[1*], PhD; Norfilza Mohd Mokhtar[2*], PhD; Sawal Hamid Md Ali[3*], PhD

[1]Center for Healthy Ageing and Wellness, Faculty of Health Sciences, Universiti Kebangsaan Malaysia, Kuala Lumpur, Malaysia

[2]Faculty of Medicine, Universiti Kebangsaan Malaysia, Kuala Lumpur, Malaysia

[3]Electronic and Systems Engineering, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, Selangor, Malaysia

[*]all authors contributed equally

Corresponding Author:
Zahara Abdul Manaf, PhD
Center for Healthy Ageing and Wellness, Faculty of Health Sciences, Universiti Kebangsaan Malaysia, , Kuala Lumpur, , Malaysia

Abstract

Background: The success of mobile apps in improving the lifestyle of patients with noncommunicable diseases through self-management interventions is contingent upon the emerging growth in this field. While users of mobile health (mHealth) apps continue to grow in number, little is known about the quality of available apps that provide self-management for common noncommunicable diseases such as diabetes, hypertension, and obesity.

Objective: We aimed to investigate the availability, characteristics, and quality of mHealth apps for common noncommunicable disease health management that included dietary aspects (based on the developer's description), as well as their features for promoting health outcomes and self-monitoring.

Methods: A systematic search of English-language apps on the Google Play Store (Google LLC) and Apple App Store (Apple Inc) was conducted between August 7, 2022, and September 13, 2022. The search terms used included *weight management*, *obesity*, *diabetes*, *hypertension*, *cardiovascular diseases*, *stroke*, and *diet*. The selected mHealth apps' titles and content were screened based on the description that was provided. Apps that were not designed with self-management features were excluded. We analyzed the mHealth apps by category and whether they involved health care professionals, were based on scientific testing, and had self-monitoring features. A validated and multidimensional tool, the Mobile App Rating Scale (MARS), was used to evaluate each mHealth app's quality based on a 5-point Likert scale from 1 (inadequate) to 5 (excellent).

Results: Overall, 42 apps were identified. Diabetes-specific mHealth apps accounted for 7% (n=3) of the market, hypertension apps for 12% (n=5), and general noncommunicable disease management apps for 21% (n=9). About 38% (n=16) of the apps were for managing chronic diseases, while 74% (n=31) were for weight management. Self-management features such as weight tracking, BMI calculators, diet tracking, and fluid intake tracking were seen in 86% (n=36) of the apps. Most mHealth apps (n=37, 88%) did not indicate whether there was involvement of health professionals in app development. Additionally, none of the apps reported scientific evidence demonstrating their efficacy in managing health. The overall mean MARS score was 3.2 of 5, with a range of 2.0 to 4.1. Functionality was the best-rated category (mean score 3.9, SD 0.5), followed by aesthetics (mean score 3.2, SD 0.9), information (mean score 3.1, SD 0.7), and engagement (mean score 2.9, SD 0.6).

Conclusions: The quality of mHealth apps for managing chronic diseases was heterogeneous, with roughly half of them falling short of acceptable standards for both quality and content. The majority of apps contained scant information about scientific evidence and the developer's history. To increase user confidence and accomplish desired health outcomes, mHealth apps should be optimized with the help of health care professionals. Future studies on mHealth content analysis should focus on other diseases as well.

(JMIR Mhealth Uhealth 2024;12:e49055) doi:10.2196/49055

KEYWORDS

mHealth apps; mobile health; health apps; chronic diseases; self-management; app quality; apps; app; application; applications; quality; MARS; Mobile App Rating Scale; mHealth; chronic; review methods; review methodology; review of apps

Introduction

Globally, noncommunicable diseases (NCDs) account for 74% of all deaths yearly (41 million people), which highlights the global health burden [1]. In Malaysia, the latest National Health and Morbidity Survey reported that two-thirds of the Malaysian population have at least 1 of 3 common NCDs, namely diabetes, hypertension, or hypercholesterolemia, and about half of the population (50.1%) are overweight or obese [2]. There are barriers to care access, delivery, and self-management for the management of NCDs, such as being unable to visit a clinic in a timely manner and long consultation waiting times [3], prompting policy makers to improve the health care system. The introduction of mobile health (mHealth) technology presents an opportunity for patient self-monitoring, helping health care providers personalize the management of patients [4] and increase cost-effectiveness throughout health management [5].

According to a definition by the World Health Organization, mHealth is the use of mobile and wireless devices such as mobile phones, tablets, and personal digital assistants to support health care management [6]. In the rapidly growing mHealth app market, the presence of these apps could facilitate the health care management system. A recent review of the literature revealed that researchers have recognized that mHealth could be an effective tool in chronic disease management [4] and improve patients' self-management behavior [7]. A growing body of research demonstrates the health benefits of mHealth interventions for patients with NCDs in terms of enhancing patient self-monitoring and health outcomes in NCDs such as type 2 diabetes [8], obesity [9], and cardiovascular diseases [10].

Self-management is crucial in the daily management of chronic diseases to improve quality of life and reduce management costs [11]. However, poor self-management among patients with chronic disease has been observed [12]. Technologies such as mHealth have the potential capacity to empower patients requiring support in their self-management efforts. A review by Cruz-Ramos et al [10] demonstrated that many mHealth apps for cardiovascular diseases support self-management features such as medical advice, reminders, and self-monitoring notifications [10]. Moreover, research has also found that self-monitoring of weight and dietary intake is associated with positive outcomes for weight loss [13]. As the key to person-centered care, mHealth apps enhance self-management for chronic conditions by providing personalized goal setting, active reminders, social interaction, and support [14]. Hence, mHealth apps could help build decision-support systems that bridge the gap between self-management and conventional health care management.

As of 2022, there were nearly 2.67 million mobile apps available on the Google Play Store. Of these, more than 130, 000 apps were health care or health and fitness apps [15]. The number of mHealth apps available on the Google Play Store and Apple App Store continues to grow [16-18]. Globally, it was estimated that 6.6 billion individuals own a smartphone, and the number is expected to grow to 7.7 billion by the year 2027 [19], which allows mHealth technology to be more accessible to individuals. However, caution must be taken regarding this growth, as the evidence related to its efficacy and benefits for chronic health disease management is not well identified.

Previous research on the perception and usability of mHealth apps for NCD management has demonstrated a growing interest in user-centric health-tracking mobile apps among the population with chronic illness [20,21]. An empirical study conducted to predict patients' intentions to continue using mHealth services as part of self-managing their chronic conditions revealed that the participants had high intentions to continue the use of mHealth services [22]. The use of mHealth apps is highly encouraged, as it has been linked to higher rates of health-promoting behavior among people with chronic medical conditions [23]. Self-management using mHealth apps could be a part of health management, as people living with NCDs have the autonomy to take responsibility for their health. Relevant content analysis studies have been carried out in different geographical areas [24-26], but limited studies have been conducted among Southeast Asian countries. As such, there is an emerging need to bridge this research gap by initiating more content analysis studies in Malaysia to contribute to more holistic development of mHealth apps.

This paper provides a review of the current landscape of mHealth apps, with an emphasis on common chronic disease management. Understanding trends in mHealth apps and their relevant features will benefit users in developing informed decisions, as well as help health care providers improve the quality of mHealth. To make better-informed decisions, the reliability of currently existing mHealth apps should be explored. This study aims to describe mHealth apps available in conventional app stores for common NCDs, determine their health categories, explore the features they focus on, identify neglected areas, and evaluate their quality using the Mobile App Rating Scale (MARS) assessment tool.

Methods

This review involved a systematic search of apps available in mobile phone app stores. The protocol adhered to the 5-step framework outlined by Arksey and O'Malley [27], which includes (1) identifying the research question; (2) identifying relevant apps; (3) selecting apps; (4) charting the data; and (5) collating, summarizing, and reporting the results.

Identifying the Research Question

This review aimed to answer the following questions: "What are the available mHealth apps in the Malaysian market for common NCDs?" "What are the app features available in the mHealth apps for NCDs in Malaysia?" and "What is the quality of the mHealth apps for NCDs in Malaysia?"

Identifying Relevant Apps

The search was conducted from July 7, 2022, to August 14, 2022, on the Apple App Store (Apple Inc) and Google Play Store (Google LLC) using the search terms *weight loss*, *obesity*, *diabetes*, *hypertension and cardiovascular diseases*, *stroke*, *weight management*, and *diet*. The search terms were identified using appropriate Medical Subject Headings terms as well as the free text of keywords. The selected NCDs were chosen

because they are among the most prevalent chronic diseases in Malaysia and around the world.

Selecting Apps

To be eligible for inclusion in this review, the title and content of the identified mHealth apps were screened and filtered based on the descriptions provided by the app developer. An app met the inclusion criteria if it (1) was developed in English; (2) had self-monitoring feature(s); and (3) was developed for chronic disease management, namely obesity, type 2 diabetes, hypertension, and cardiovascular disease. An app was excluded based on the following criteria: it was a heart rate tracker, exercise tracker, or game; or it was a medicine-delivery, appointment-based, recipe-sharing, or research study app.

Charting the Data

The data were extracted based on feature categories, which included (1) type of health management, (2) number of downloads, (3) country of the developers, and (4) app features. A data extraction table with the mHealth apps' basic information, such as country of origin, name of app developer, and number of downloads, is shown in Table 1. The details of the mHealth apps' features and characteristics are outlined in Multimedia Appendix 1.

Table . Overview of selected mHealth apps.

Name of app	Number of downloads	Availability of in-app purchase	Cost of in-app purchase (US $)	Country of app developer	App developer	Operating system
MyFitness Pal: Calorie Counter	100,000,000	Yes	0.76-0.70	US	MyFitnessPal, Inc	Android
Health & fitness tracker with calorie counter	5,000,000	Yes	0.65-2.90	India	DROID INFINITY	Android
Lifesum: Healthy eating & diet	10,000,000	Yes	3.95-50.30	Sweden	Lifesum	Android
Withings Health Mate	1,000,000	No	0.00	France	Withings	Android
Fitbit	50,000,000	Yes	7.20-290.70	US	Fitbit LLC	Android
HealthifyMe	10,000,000	Yes	7.65-72.10	Singapore	HealthifyMe	Android
Noom: Weight loss	10,000,000	Yes	0.76-142.10	US	Noom Inc	Android
Personal Health Monitor	100,000	No	0.00	Ukraine	Extrawest	Android
Life Extend: Healthy Habits	100,000	Yes	3.10-203.30	US	LifeOmic	Android
Qardio Heart Health	100,000	No	0.00	Canada	Qardio Inc	Android
FitTrack My-Health: Track Scale	100,000	Yes	9.61-92.90	US	Fittrack	Android
Calorie counter by lose it!	10,000,000	Yes	4.25-161.75	US	FitNow Inc	Android
One Drop: Better Health Today	1,000,000	Yes	20.10-20.35	US	One Drop	Android
Calorie Counter - MyNetDiary	1,000,000	Yes	3.75-60.10	US	MyNetDiary.com	Android
Healthi: Personal Weight Loss	500,000	Yes	1.30-52.45	US	Sunshine Health Studios	Android
Health Diet Foods Fitness Help	500,000	Yes	2.20-21.90	India	RecoveryBull.com	Android
Unimeal: Healthy Diet&Workouts	100,000	Yes	5.50-95.10	Cyprus	Uniwell	Android
Health Click Away	10,000	Yes	3.10-21.85	US	HealthClickAway	Android
Possible-Nutrition Weight Loss	1,000,000	Yes	0.00	India	Truweight Wellness	Android
Health Club-Home workouts& Fitness-calorie tracker	50,000	No	0.00	__[a]	Health Club Group	Android
Smart Diet Planner weight loss	100,000	Yes	9.2-28.40	India	Appneurons Technologies Private Limited	Android
Heart Care Health & Diet Tips	10,000	Yes	2.2-21.90	India	RecoveryBull.com	Android
Calorie Counter + (Nutracheck)	1,000,000	Yes	1.85-36.10	England	Nutracheck	Android
Health Mate - Calorie Counter & Weight Loss App	500,000	Yes	0.90-5.69	India	PIXEL BYTES	Android
Doctor2u- One Stop Healthcare	500,000	Yes	2.20	Malaysia	BP Healthcare Group	Android

Name of app	Number of downloads	Availability of in-app purchase	Cost of in-app purchase (US $)	Country of app developer	App developer	Operating system
BookDoc- Go Active Get reward	500,000	Yes	1.75-17.50	Malaysia	BookDoc	Android
Health Pal - Fitness, Weight loss coach, Pedometer	1,000,000	Yes	1.65	India	Digit Grove	Android
Creda- manage chronic condition	100,000	Yes	1.75-18.40	US	KnowYourMeds Inc	Android
Mhealth	—	No	0.00	—	mutifun LLC	iOS
my Mhealth	—	No	0.00	England	my mhealth	iOS
Zero: Fasting & Health Tracker	—	Yes	10.10-69.95	US	Zero Longevity Science Inc	iOS
Lose Weight at Home in 30 Days	—	Yes	5.25-54.65	Hong Kong	ABISHKKING LIMITED	iOS
BodyFast Intermittent Fasting	—	Yes	3.95-61.2	Germany	BodyFast GmbH	iOS
BetterMe: Health Coaching	—	Yes	4.60-36.70	Cyprus	BetterMe Limited	iOS
Weight Loss Running by Slimkit	—	Yes	8.75-38.25	UK	MONTIBUS LTD	iOS
My Diet Coach - Weight Loss	—	Yes	2.20-8.75	US	Easy Tiger Apps LLC	iOS
Fitness Coach & Diet: FitCoach	—	Yes	9.20-57.95	Cyprus	A.L. AMAZING APPS LIMITED	iOS
Argus: Calorie Counter & Step	—	Yes	9.20-27.35	US	Azumio Inc	iOS
Speedoc - Care Comes to You	—	No	0.00	Singapore	Speedoc	iOS
Glucose Buddy Diabetes Tracker	—	Yes	3.70-54.65	US	Azumio Inc	iOS
DOC2US - Trusted Online Doctor	—	No	0.00	Malaysia	Doc2Us	iOS
Foodvisor - Nutrition & Diet	—	Yes	17.9-80.90	France	Foodvisor	iOS

[a]Not available (information was not found in app stores).

Collating, Summarizing, and Reporting the Results

After obtaining the screening results, we performed a descriptive analysis, comparison, and functionality assessment based on the information provided by the app developers. In addition, we analyzed the mHealth apps' category, as well as whether health care experts were involved, whether they were based on scientific testing, and whether they had self-monitoring, based on the description provided by the app providers. Additional information, such as star ratings and the presence of a privacy policy, was also tabulated. A star rating offers a quick overview of an app's overall user satisfaction, making it a valuable component of quality assessment. The presence of a privacy policy reveals how an app manages user data, ensuring the protection of users' private health information, which is crucial for informed decision-making.

Quality Assessment of mHealth Apps

The quality of the mHealth apps was assessed using the validated MARS evaluation tool, which has demonstrated excellent internal consistency (α=.90) [28]. The MARS has 4 sections for objective evaluation: engagement (eg, the level of entertainment provided, interactivity, and appropriateness of app content), functionality (eg, app performance, ease of use, and navigation), aesthetics (eg, layout, graphics quality, and overall visual appeal), and information (eg, accuracy of the app description, source of information, and quality of information). The subjective quality evaluation section of the MARS subscale has 4 items. However, we excluded this section in this study as the aim was to assess the apps' quality objectively. Each item was evaluated using a Likert scale with a score range from 1 (inadequate) to 5 (excellent). The overall quality score was calculated based on the mean scores for each of the 4 sections. A mean score of 3 was considered the minimum acceptable score, whereas a score greater than 4 of 5 was preferable. Before

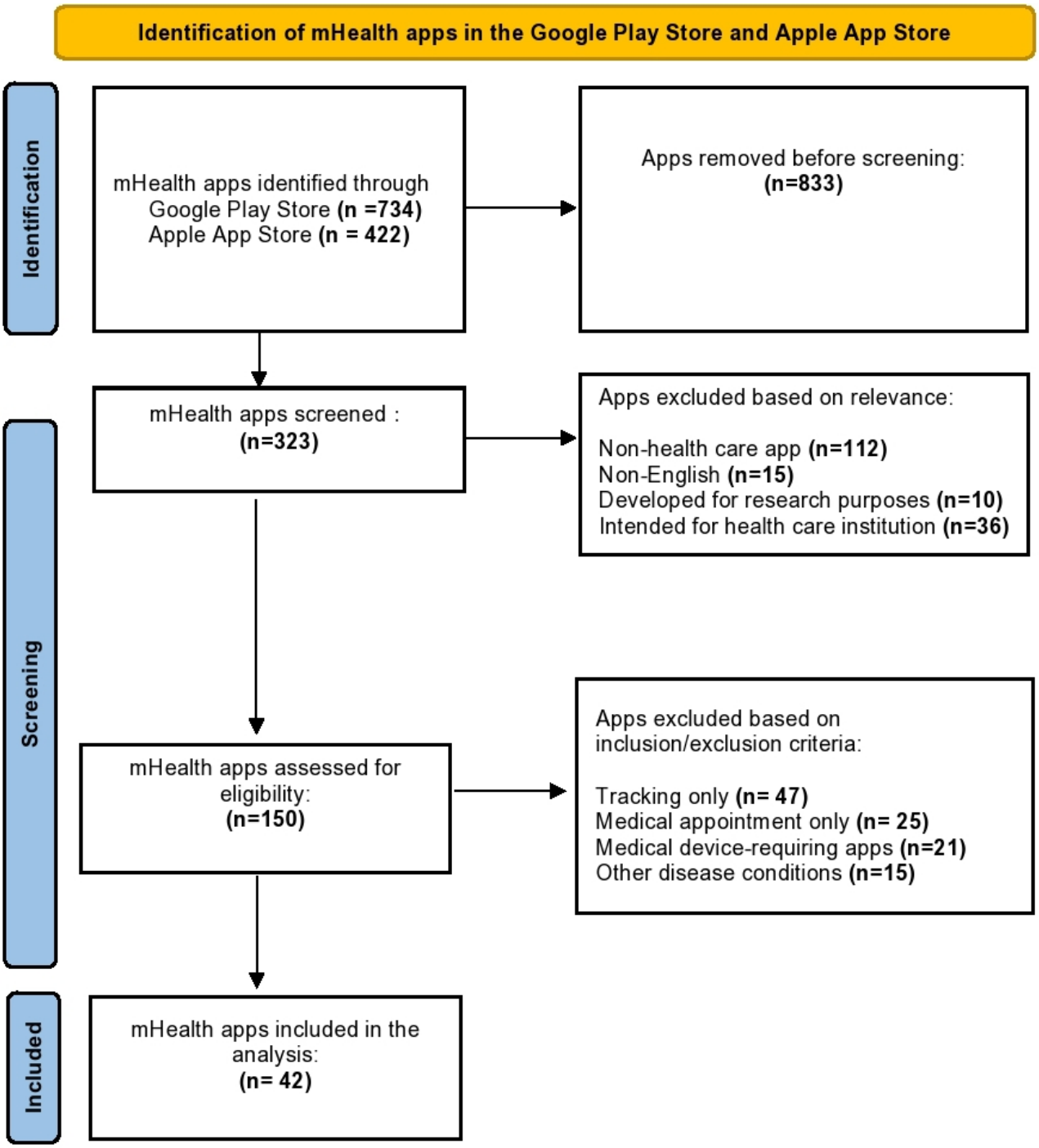
the evaluation, 2 authors independently used each of the apps and conducted the quality assessment in agreement with each other; disagreements were resolved through discussion with a third author.

Results

Search Results

Our search found a total of 1156 apps through keyword retrieval from the Apple App Store and Google Play Store (Figure 1). After removing duplicates, 323 apps were screened; 150 apps met the inclusion criteria as apps focusing on selected chronic disease management (iOS: n=103; Android: n=47) and were included for eligibility assessment. Among them, 42 were included in this study for analysis.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram for the systematic search and selection of mobile health (mHealth) apps.



Sample Distribution and App Characteristics

A total of 42 apps (n=28 Android and n=14 iOS apps) were included in this review. Among these, 31 (74%) were for weight management while 11 (21%) were dedicated to chronic disease health management. The chronic diseases commonly targeted by the apps included cardiovascular disease (n=1, 2%), type 2 diabetes (n=3, 7%) and hypertension (n=5, 12%). However, there were 9 apps (21%) that did not specify which chronic diseases they targeted. Most of the apps received recent updates, in the year 2022. The general characteristics of the mHealth apps are described in Table 2. The download count is exclusive to Android apps, as this information was not obtainable for iOS apps.

Table . Sample distribution of chronic disease management apps (n=42).

Category		Apps, n (%)
Weight management		31 (74)
Type 2 diabetes mellitus		3 (7)
Hypertension		5 (12)
Cardiovascular disease		1 (2)
General NCDs[a]		9 (21)
Year updated		
	2020	1 (2)
	2021	5 (12)
	2022	36 (86)
Downloads (Android only), n		
	10,000-99,999	3 (7)
	100,000-999,999	12 (29)
	≥1,000,000	13 (31)
Cost to download		
	Free	42 (100)

[a]NCD: noncommunicable disease (this included hypertension, type 2 diabetes mellitus, and cardiovascular diseases).

As shown in Table 1, about 81% (n=34) of the apps provided in-app purchases for live-chat subscriptions, ad removal, and premium feature subscriptions ranging from weekly to yearly, among others. The price for each purchase was between US $0.66 to US $203.26. All the apps included in this review can be freely downloaded by users. Most of the apps (n=15, 36%) were developed in the United States, followed by India (n=7, 17%), Malaysia (n=3, 7%) and Singapore (n=2, 5%).

Star Ratings and Privacy Policies of the Included mHealth Apps

Two apps on the Google Play Store and 6 apps on the Apple App Store received no reviews. The number of users (more than 1 million) who provided ratings for Android apps was significantly greater than the number for iOS apps. Most apps received a user rating of more than 4 stars of a total of 5. In this study, all the apps had a privacy policy. Table 3 shows the star ratings and privacy policies of the included apps.

Table . Star ratings and privacy policies of the included mobile health apps.

	Android[a]	iOS[b]
Star rating (1-5 stars), mean (SD)	4.3 (0.3)	4.6 (0.3)
Privacy policy, n (%)	28 (100)	14 (100)

[a]Apps without ratings: n=2.

[b]Apps without ratings: n=6.

App Feature Assessment

The most common feature that was available on the mHealth apps was self-monitoring (n=35, 83%), which allows users to track their body weight, food intake, fluid intake, and other health indicators. Approximately 16 of 42 apps provided online consultation or personalized feedback via the app as well as goal-setting features (n=35, 83%). Meanwhile, there were no evidence-based apps that used scientific testing or multidisciplinary team involvement in app development, based on the descriptions provided by the developers. About 62% of apps (n=26) introduced at least 1 health care professional (eg, health coach, nutritionist, or dietitian) in the health management of the app. The overview of the functionality characteristics of the apps is presented in Table 4.

Table . Overview of functionality assessment of selected mobile health apps (n=42).

Components		Apps, n (%)	
		Yes	No
Multidisciplinary team involvement in app development		0 (0)	42 (100)
Health care professional involvement in app health management		26 (62)	16 (38)
Self-monitoring features			
	Overall	35 (83)	7 (17)
	Weight tracker (eg, BMI)	35 (83)	7 (17)
	Diet tracker	29 (69)	13 (31)
	Water intake tracker	16 (38)	26 (62)
	Step count tracker	10 (24)	32 (76)
	Exercise tracker	13 (31)	29 (69)
	Personalised feedback (eg, chat with doctor, nutritionist, health coach)	16 (38)	26 (62)
Goal setting			
	Overall	35 (83)	7 (17)
	Weight	35 (83)	7 (17)
	Nutrient intake	16 (38)	26 (62)
	Steps activity	9 (21)	33 (79)
	Exercise	5 (12)	37 (88)
Medical condition monitoring			
	Overall	13 (31)	29 (69)
	Blood pressure	9 (21)	33 (79)
	Glucose	8 (19)	34 (81)
	Heart rate	7 (17)	35 (83)
Social support		15 (36)	27 (64)
Evidence-based testing		0 (0)	42 (100)

App Quality Assessment

The average MARS score among the 42 apps was 3.2 of 5, with a range between 2.0 and 4.1. Of the 4 MARS domains, functionality scored the highest (3.9/5), followed by aesthetics (3.2/5), information (3.1/5), and engagement (2.9/5). There was a large gap in the scores of each subdomain, with the engagement score ranging from 1.6 to 4.0, functionality score ranging from 3.0 to 5.0, aesthetics score ranging from 1.0 to 5.0, and information score ranging from 1.7 to 5.0. The MARS functionality score had the smallest range, and the information score had the largest range. Table 5 shows the MARS subdomain ratings and the total mean score.

Table . App quality rating scores using the Mobile App Rating Scale (n=42).

Objective quality rating	Mean score (SD)	Minimum to maximum score
Engagement	2.9 (0.6)	1.6-4.0
Functionality	3.9 (0.5)	3.0-5.0
Aesthetics	3.2 (0.9)	1.0-5.0
Information	3.1 (0.7)	1.7-5.0
Total quality rating	3.2 (0.5)	2.0-4.1

Discussion

Principal Findings

Most NCD management apps in Malaysia lack scientific evidence of efficacy, do not involve multidisciplinary teams, and require significant feature improvements. This study provides a snapshot of current mHealth apps for common NCDs in 2022 that were primarily commercially based. It revealed that the mHealth apps varied in terms of features, functionality, and disease management. The most common category that the mHealth apps currently focus on is self-management of weight, followed by hypertension and type 2 diabetes. Some apps need to be purchased to access extended features, health management information, and medical advice.

Self-monitoring or self-management are crucial components of long-term chronic disease management. Self-management of chronic diseases has been reported to be related to positive health outcomes among patients [29]. Most of the mHealth apps (n=35, 83%) from this study incorporated self-monitoring features that allow the user to monitor their weight, blood pressure, blood glucose level, dietary intake, and fluid intake. The most common self-monitoring features found in the apps were weight trackers, food intake trackers, and step trackers. As technology has advanced, self-management through mHealth apps has been the subject of investigation. Systematic reviews in Korea [14] and the United States [12] examined their impact, shedding light on their potential role in achieving desirable health outcomes through self-management. This suggests that self-monitoring features in mHealth apps may be instrumental in achieving desirable health outcomes.

Some of the apps (n=15, 36%) incorporated social support features to allow users to engage with other users as well as health professionals. These social features could be key for users to continue engaging with the mHealth apps to improve their health. Social support has been shown to improve patients' health and well-being, and this also applies to online social support networks [30]. Only 38% (n=16) of the apps offered communication with health care professionals via instant messaging or robotic automated message chat functions. These features are beneficial in the self-management of NCDs, as they allow the user to communicate with a health care professional and receive immediate feedback. There is evidence that 2-way communication between patients and health care professionals can improve health outcomes [31]; therefore, app developers should prioritize the inclusion of communication features and health care professionals in app management.

This study reveals a lack of evidence supporting the use of mHealth apps, with none of the apps reporting scientific evidence to indicate the effectiveness of their health management. Based on the star ratings, we cannot deny that mHealth apps could potentially help users improve their health outcomes. Given the fact that most apps on the Apple App Store (iOS) and Google Play Store (Android) did not provide evidence-based testing to prove their effectiveness [32], the app developers were able to make false or misleading claims about their apps. Moreover, health care providers are less likely to feel skeptical of the role of mHealth apps in health care management if the app is supported by research as clinical evidence [33].

The multidisciplinary team approach is a treatment domain that optimizes the health of patients with chronic diseases [34]. Promoting a multidisciplinary team approach is crucial for coordinating the health care system and aiding patients in self-management [35]. However, our research revealed that the majority of mHealth apps did not incorporate a multidisciplinary team to support health management. In this study, only 16% (n=7) of the apps indicated the participation of medical doctors, specialists, nurses, or dietitians. Our findings emphasize the need for multidisciplinary team involvement in health management by using reliable and high-quality mHealth apps. mHealth apps have the potential to function as proactive disease self-management tools [36], meeting diverse needs through collaborative efforts with multidisciplinary teams within the realm of mHealth technology.

The apps in the app stores generally had high star ratings, ranging from 4.3 to 4.6 of 5 stars, which may suggest good user satisfaction. However, it is important to note that user ratings may not always accurately reflect app quality. Therefore, the MARS was used in this study. The overall MARS score of the included apps was 3.2 of 5, which is considered acceptable. Considering the maximum and minimum MARS scores, substantial variability was observed across domains, including aesthetics, which had a range from 1.0 to 5.0. This variance may indicate a significant diversity in app quality, with some evidence suggesting the presence of low-quality apps currently available in the market. However, lower MARS scores could potentially be attributed to reviews of freely available app features.

The review and analysis of mobile app quality for common chronic diseases is crucial for future mHealth app development, as poor app quality can limit their effectiveness in health management [37]. Research on the quality of mHealth apps in Southeast Asia is limited, and the existing studies have primarily focused on COVID-19–related apps rather than those related to chronic disease management [38,39]. Therefore, the authors compared the MARS results with health care app studies from other countries. Our findings align with an assessment of nutrition-related mHealth apps in Korea, where the majority received an average rating, with a mean score of 3.28 of 5 [40]. In contrast, health apps designed for behavioral change in Denmark achieved a slightly higher average quality score, with an average MARS score of 3.48 of 5 [26]. Consistent with previous studies, our study revealed that functionality was best rated (3.9/5), whereas engagement consistently received the lowest score, with the mean score being 2.9 of 5 [41,42]. This indicates that the app developers focused on the functionality of the apps as an essential element in delivering outstanding experiences to the users. On the other hand, some studies discovered that information scored the lowest or fell into the low-to-moderate category [43-45]. This discrepancy highlights that the mHealth apps in the market need special attention to provide more advanced and effective features and capabilities for health management. Importantly, our results indicate that the key areas for improvement in mHealth apps are engagement and information. App developers should prioritize these

domains, focusing on customization, interactive information delivery, and the integration of prompts such as feedback and reminders. Additionally, enhancing sharing functionality and offering more evidence-based content, engaging visuals, and data-driven information for users would improve the overall mHealth app quality.

From a health management perspective, mHealth interventions offer a significant opportunity to facilitate the monitoring of chronic conditions and improve self-management skills [14,24,26]. Evaluating the quality of mHealth apps can help us identify their positive impact on health and behavior outcomes among people with chronic diseases. According to Ryan and Sawin [46], who described an individual and family self-management theory, successful self-management should involve three key components: (1) individual competence, (2) individual motivation, and (3) social factors. mHealth apps have the potential to bridge these factors together to facilitate health management. Additionally, mobile technology promises to enable real-time remote monitoring systems and prompt feedback systems to improve health management [12]. App developers and users should be alert, as our findings show that most apps available in the market are lacking such components.

Strengths and Limitations

The strengths of this study included that the findings were derived and evaluated from the clinical point of view, the study used a validated tool to determine the quality of mHealth apps, and the results can publicly provide review data to users as well as further direction for mHealth app development. This study has several identified limitations. First, we conducted a comprehensive review solely of the apps' free features, excluding premium features due to budget constraints. Second, at the time the study was completed, it is possible that new mHealth apps or updated features had been released, which we were unable to consider. Third, the review was limited to specific common NCDs, and we could not provide an overview of other diseases. Fourth, our focus was on English-language apps available for download in commonly used mobile app stores, which might limit the generalizability of our findings. According to the world's largest ranking of countries and regions by English skills, Malaysia ranks among the top 3 in the English Proficiency Index among Asian countries [47]. Nevertheless, we acknowledge that the focus on English-language apps could result in some percentage of Malaysians being left out. Importantly, app developers and researchers are increasingly recognizing the need to cater to diverse linguistic needs within mHealth apps to address the cultural and language diversity of countries like Malaysia [48,49].

Future Direction for mHealth App Development

The findings of this study reveal that the current mHealth app market for managing NCDs in Malaysia is still in its nascent stages and is marked by a shortage of high-quality mHealth apps. In contrast to a previous study conducted in Malaysia in 2017, which predominantly featured informational apps [50], the market is currently undergoing a shift. It is now pivoting toward health management apps. This changing trend underscores the potential for mHealth technology to serve as a cornerstone in the management of chronic diseases in the future. For instance, the results of a prior study that specifically examined hypertension indicated that health care apps could serve as valuable additions to conventional treatment methods [51]. To effectively address NCDs, mHealth apps should incorporate self-monitoring capabilities, such as health tracking, goal setting, and personalized feedback. Additionally, app developers should emphasize elevating the overall quality of their apps by incorporating a variety of perspectives, including input from relevant health professionals and the integration of scientific evidence. It is essential to include this information in the app description to establish trust among users.

Conclusion

A search for mHealth apps for common NCDs available in the Google Play Store and the Apple App Store revealed that most apps focused on weight management, followed by hypertension and type 2 diabetes mellitus. Self-monitoring features such as weight trackers, diet trackers, and step trackers are the core functions of current mHealth apps. This review also highlights the current market's lack of evidence-based mHealth apps designed specifically for the self-management of chronic diseases. The lack of multidisciplinary teams in app development and health management was observed in the app stores. Evidently, the quality of mHealth apps currently available in the market should undergo ongoing assessment and enhancement to optimize their benefits for users in the realm of health management. Reviews of these apps can offer valuable insights to researchers, health care providers, and app developers, aiding them in delivering high-quality apps for effective health management. App developers and public health authorities should prioritize the development of evidence-based mHealth apps to enhance the mHealth ecosystem for users. Future studies on mHealth content analysis and app evaluation should encompass a broader spectrum of diseases, aiming for a more comprehensive approach that benefits diverse populations.

Authors' Contributions

KJC and ZAM contributed to conceptualization, methodology, and formal analysis; KJC contributed to writing (original draft preparation); KJC, ZAM, AFML, NHR, SHMA, and NMM contributed to writing (review and editing); ZAM, AFML, and NHR supervised the project. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Assessment of the functionality of selected mobile health apps.
[DOCX File, 18 KB - mhealth_v12i1e49055_app1.docx]

References

1. Noncommunicable diseases. World Health Organization. 2021. URL: https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases [accessed 2022-09-19]
2. National Health and Morbidity Survey. Non-Communicable Diseases, Healthcare Demand, and Health Literacy—Key Findings: Institute for Public Health, Ministry of Health, Malaysia; 2019.
3. Foo KM, Sundram M, Legido-Quigley H. Facilitators and barriers of managing patients with multiple chronic conditions in the community: a qualitative study. BMC Public Health 2020 Feb 27;20(1):273. [doi: 10.1186/s12889-020-8375-8] [Medline: 32106838]
4. Fan K, Zhao Y. Mobile health technology: a novel tool in chronic disease management. Intell Med 2022 Feb;2(1):41-47. [doi: 10.1016/j.imed.2021.06.003]
5. Rinaldi G, Hijazi A, Haghparast-Bidgoli H. Cost and cost-effectiveness of mHealth interventions for the prevention and control of type 2 diabetes mellitus: a protocol for a systematic review. BMJ Open 2019 Apr 11;9(4):e027490. [doi: 10.1136/bmjopen-2018-027490] [Medline: 30975686]
6. Allegrante JP, Wells MT, Peterson JC. Interventions to support behavioral self-management of chronic diseases. Annu Rev Public Health 2019 Apr 1;40:127-146. [doi: 10.1146/annurev-publhealth-040218-044008] [Medline: 30601717]
7. Peñarrieta MI, Flores-Barrios F, Gutiérrez-Gómez T, Piñones-Martínez S, Resendiz-Gonzalez E, Quintero-Valle LM. Self-management and family support in chronic diseases. J Nurs Educ Pract 2015 Mar;5(11):73-80. [doi: 10.5430/jnep.v5n11p73]
8. Bene BA, O'Connor S, Mastellos N, Majeed A, Fadahunsi KP, O'Donoghue J. Impact of mobile health applications on self-management in patients with type 2 diabetes mellitus: protocol of a systematic review. BMJ Open 2019 Jun 25;9(6):e025714. [doi: 10.1136/bmjopen-2018-025714] [Medline: 31243029]
9. Eisenhauer CM, Brito F, Kupzyk K, et al. Mobile health assisted self-monitoring is acceptable for supporting weight loss in rural men: a pragmatic randomized controlled feasibility trial. BMC Public Health 2021 Aug 18;21(1):1568. [doi: 10.1186/s12889-021-11618-7] [Medline: 34407782]
10. Cruz-Ramos NA, Alor-Hernández G, Colombo-Mendoza LO, Sánchez-Cervantes JL, Rodríguez-Mazahua L, Guarneros-Nolasco LR. mHealth apps for self-management of cardiovascular diseases: a scoping review. Healthcare 2022;10(2):322. [doi: 10.3390/healthcare10020322]
11. WHO Global Observatory for eHealth. Atlas eHealth country profiles: based on the findings of the second global survey on eHealth. World Health Organization. 2011. URL: https://apps.who.int/iris/handle/10665/44502 [accessed 2022-09-19]
12. Wang H, Ho AF, Wiener RC, Sambamoorthi U. The association of mobile health applications with self-management behaviors among adults with chronic conditions in the United States. Int J Environ Res Public Health 2021 Sep 30;18(19):10351. [doi: 10.3390/ijerph181910351] [Medline: 34639651]
13. Eastman A, Dixon BN, Ross KM. Associations between effort, importance, and self-monitoring during and after a 12-week behavioural weight management program. Obes Sci Pract 2020 Oct;6(5):447-453. [doi: 10.1002/osp4.431] [Medline: 33082986]
14. Yi JY, Kim Y, Cho YM, Kim H. Self-management of chronic conditions using mHealth interventions in Korea: a systematic review. Healthc Inform Res 2018 Jul;24(3):187-197. [doi: 10.4258/hir.2018.24.3.187] [Medline: 30109152]
15. Number of Android apps on Google Play. AppBrain. 2022. URL: https://www.appbrain.com/stats/number-of-android-apps [accessed 2022-09-19]
16. Number of mHealth apps available in the Apple App Store from 1st quarter 2015 to 2nd quarter 2022. Statista. 2022. URL: https://www.statista.com/statistics/779910/health-apps-available-ios-worldwide/ [accessed 2022-09-19]
17. Number of mHealth apps available in the Google Play Store from 1st quarter 2015 to 2nd quarter 2022. Statista. 2022. URL: https://www.statista.com/statistics/779919/health-apps-available-google-play-worldwide/ [accessed 2022-09-19]
18. Kalgotra P, Raja U, Sharda R. Growth in the development of health and fitness mobile Apps amid COVID-19 pandemic. Digit Health 2022;8(3):20552076221129070. [doi: 10.1177/20552076221129070] [Medline: 36211794]
19. Number of smartphone subscriptions worldwide from 2016 to 2021, with forecasts from 2022 to 2027. Statista. 2022. URL: https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/ [accessed 2022-09-19]
20. Birkhoff SD, Smeltzer SC. Perceptions of smartphone user-centered mobile health tracking apps across various chronic illness populations: an integrative review. J Nurs Scholarsh 2017 Jul;49(4):371-378. [doi: 10.1111/jnu.12298] [Medline: 28605151]
21. Lee M, Kang D, Yoon J, et al. The difference in knowledge and attitudes of using mobile health applications between actual user and non-user among adults aged 50 and older. PLoS One 2020;15(10):e0241350. [doi: 10.1371/journal.pone.0241350] [Medline: 33108792]

22. Song T, Deng N, Cui T, et al. Measuring success of patients' continuous use of mobile health services for self-management of chronic conditions: model development and validation. J Med Internet Res 2021 Jul 13;23(7):e26670. [doi: 10.2196/26670] [Medline: 34255685]
23. Mahmood A, Kedia S, Wyant DK, Ahn S, Bhuyan SS. Use of mobile health applications for health-promoting behavior among individuals with chronic medical conditions. Digit Health 2019;5:2055207619882181. [doi: 10.1177/2055207619882181] [Medline: 31656632]
24. Gong E, Zhang Z, Jin X, et al. Quality, functionality, and features of Chinese mobile apps for diabetes self-management: systematic search and evaluation of mobile apps. JMIR Mhealth Uhealth 2020 Apr 7;8(4):e14836. [doi: 10.2196/14836] [Medline: 32255432]
25. Ming LC, Untong N, Aliudin NA, et al. Mobile health apps on COVID-19 launched in the early days of the pandemic: content analysis and review. JMIR Mhealth Uhealth 2020 Sep 16;8(9):e19796. [doi: 10.2196/19796] [Medline: 32609622]
26. Bricca A, Pellegrini A, Zangger G, Ahler J, Jäger M, Skou ST. The quality of health apps and their potential to promote behavior change in patients with a chronic condition or multimorbidity: systematic search in App Store and Google Play. JMIR Mhealth Uhealth 2022 Feb 4;10(2):e33168. [doi: 10.2196/33168] [Medline: 35119367]
27. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
28. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile App Rating Scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27. [doi: 10.2196/mhealth.3422] [Medline: 25760773]
29. Panagioti M, Richardson G, Small N, et al. Self-management support interventions to reduce health care utilisation without compromising outcomes: a systematic review and meta-analysis. BMC Health Serv Res 2014 Aug 27;14:356. [doi: 10.1186/1472-6963-14-356] [Medline: 25164529]
30. Wright K. Social networks, interpersonal social support, and health outcomes: a health communication perspective. Front Commun 2016;1:10. [doi: 10.3389/fcomm.2016.00010]
31. Donevant SB, Estrada RD, Culley JM, Habing B, Adams SA. Exploring app features with outcomes in mHealth studies involving chronic respiratory diseases, diabetes, and hypertension: a targeted exploration of the literature. J Am Med Inform Assoc 2018 Oct 1;25(10):1407-1418. [doi: 10.1093/jamia/ocy104] [Medline: 30137383]
32. Powell AC, Landman AB, Bates DW. In search of a few good apps. JAMA 2014 May 14;311(18):1851-1852. [doi: 10.1001/jama.2014.2564] [Medline: 24664278]
33. Kelly MP, Heath I, Howick J, Greenhalgh T. The importance of values in evidence-based medicine. BMC Med Ethics 2015 Oct 12;16(1):69. [doi: 10.1186/s12910-015-0063-3] [Medline: 26459219]
34. Di Serafino F, Pascucci D, Sassano M, et al. Systematic review on multidisciplinarity and management of multimorbid chronic patients in hospital. Eur J Public Health 2020 Sep 1;30(Supplement_5):ckaa166.501. [doi: 10.1093/eurpub/ckaa166.501]
35. Keddy AC, Packer TL, Audulv Å, et al. The Team Assessment of Self-Management Support (TASMS): a new approach to uncovering how teams support people with chronic conditions. Healthc Manage Forum 2021 Jan;34(1):43-48. [doi: 10.1177/0840470420942262] [Medline: 32729329]
36. Grundy Q. A review of the quality and impact of mobile health apps. Annu Rev Public Health 2022 Apr 5;43(1):117-134. [doi: 10.1146/annurev-publhealth-052020-103738] [Medline: 34910582]
37. Aungst T, Seed S, Gobin N, Jung R. The good, the bad, and the poorly designed: the mobile app stores are not a user-friendly experience for health and medical purposes. Digit Health 2022;8:20552076221090038. [doi: 10.1177/20552076221090038] [Medline: 35401997]
38. Lee B, Ibrahim SA, Zhang T. Mobile apps leveraged in the COVID-19 pandemic in East and South-East Asia: review and content analysis. JMIR Mhealth Uhealth 2021 Nov 11;9(11):e32093. [doi: 10.2196/32093] [Medline: 34748515]
39. Gonzales A, Custodio R, Lapitan MC, Ladia MA. Mobile applications in the Philippines during the COVID-19 pandemic: systematic search, use case mapping, and quality assessment using the Mobile App Rating Scale (MARS). BMC Digit Health 2023;1(1):8. [doi: 10.1186/s44247-023-00007-2] [Medline: 38014368]
40. Lee HJ, Seong SH, Chung H, Lee YJ, Kim JH. Systematic search and qualitative evaluation of dietary supplement mobile applications: using the Mobile Application Rating Scale (MARS). Korean J Clin Pharm 2023 Mar 31;33(1):51-61. [doi: 10.24304/kjcp.2023.33.1.51]
41. Masterson Creber RM, Maurer MS, Reading M, Hiraldo G, Hickey KT, Iribarren S. Review and analysis of existing mobile phone apps to support heart failure symptom monitoring and self-care management using the Mobile Application Rating Scale (MARS). JMIR Mhealth Uhealth 2016 Jun 14;4(2):e74. [doi: 10.2196/mhealth.5882] [Medline: 27302310]
42. Siddiqui NR, Hodges SJ, Sharif MO. Orthodontic apps: an assessment of quality (using the Mobile App Rating Scale (MARS)) and behaviour change techniques (BCTs). Prog Orthod 2021 Sep 13;22(1):25. [doi: 10.1186/s40510-021-00373-5] [Medline: 34514529]
43. Chavez S, Fedele D, Guo Y, et al. Mobile apps for the management of diabetes. Diabetes Care 2017 Oct;40(10):e145-e146. [doi: 10.2337/dc17-0853] [Medline: 28774944]

44. Salazar A, de Sola H, Failde I, Moral-Munoz JA. Measuring the quality of mobile apps for the management of pain: systematic search and evaluation using the Mobile App Rating Scale. JMIR Mhealth Uhealth 2018 Oct 25;6(10):e10718. [doi: 10.2196/10718] [Medline: 30361196]
45. McKay FH, Wright A, Shill J, Stephens H, Uccellini M. Using health and well-being apps for behavior change: a systematic search and rating of apps. JMIR Mhealth Uhealth 2019 Jul 4;7(7):e11926. [doi: 10.2196/11926] [Medline: 31274112]
46. Ryan P, Sawin KJ. The Individual and Family Self-Management Theory: background and perspectives on context, process, and outcomes. Nurs Outlook 2009;57(4):217-225. [doi: 10.1016/j.outlook.2008.10.004] [Medline: 19631064]
47. Malaysia | EF English proficiency index. Education First. 2023. URL: https://www.ef.com/wwen/epi/regions/asia/malaysia/ [accessed 2023-11-10]
48. Nicol V. Mobile health (mHealth) language considerations. My Language Connection. 2022. URL: https://www.mylanguageconnection.com/mobile-health-mhealth-language-considerations/ [accessed 2023-11-10]
49. Nguyen HL, Tran K, Doan PLN, Nguyen T. Demand for mobile health in developing countries during COVID-19: Vietnamese's perspectives from different age groups and health conditions. Patient Prefer Adherence 2022;16:265-284. [doi: 10.2147/PPA.S348790] [Medline: 35140459]
50. Seman SAA, Ramayah T. Are we ready to app? A study on mHealth apps, its future, and trends in Malaysia context. In: Pelet JE, editor. Mobile Platforms, Design, and Apps for Social Commerce: IGI Global; 2017. [doi: 10.4018/978-1-5225-2469-4.ch005]
51. Agnihothri S, Cui L, Rajan B, Banerjee A, Ramanujan R. Mobile health application usage and quality of care at a hypertension clinic: an observational cohort study. J Hypertens 2021 Nov 1;39(11):2265-2271. [doi: 10.1097/HJH.0000000000002909] [Medline: 34074997]

Abbreviations

MARS: Mobile App Rating Scale
mHealth: mobile health
NCD: noncommunicable disease
NHMS: National Health and Morbidity Survey
WHO: World Health Organization

Edited by L Buis; submitted 19.05.23; peer-reviewed by C Cheng, P Rasche; revised version received 12.11.23; accepted 15.11.23; published 12.03.24.

Please cite as:
Cheah KJ, Abdul Manaf Z, Fitri Mat Ludin A, Razalli NH, Mohd Mokhtar N, Md Ali SH
Mobile Apps for Common Noncommunicable Disease Management: Systematic Search in App Stores and Evaluation Using the Mobile App Rating Scale
JMIR Mhealth Uhealth 2024;12:e49055
URL: https://mhealth.jmir.org/2024/1/e49055
doi:10.2196/49055

Assessing the Quality and Behavior Change Potential of Vaping Cessation Apps: Systematic Search and Assessment

Fiona McKay[1,2], BSc, MPH, PhD; Lilian Chan[3], MBBS, MPH, PhD; Rebecca Cerio[3], MA; Sandra Rickards[3], PhD; Phillipa Hastings[3], BHS, MPH; Kate Reakes[3], BE, MPH, Grad Cert, MBA, MPA; Tracey O'Brien[3], DCH, BSc, LLM, MBA, MBChB; Matthew Dunn[1,2], BA, PhD

[1]Instutite for Health Transformation, Faculty of Health, Deakin University, Geelong, Victoria, Australia

[2]The School of Health and Social Development, Faculty of Health, Deakin University, Geelong, Victoria, Australia

[3]Cancer Institute of New South Wales, St. Leonards, New South Wales, Australia

Corresponding Author:
Fiona McKay, BSc, MPH, PhD
Instutite for Health Transformation, Faculty of Health, Deakin University, , Geelong, Victoria, , Australia

Abstract

Background: An increasing number of people are using vapes (e-cigarettes), and with growing evidence of associated harms, there is a need for acceptable cessation support and interventions. Smartphone apps for health and well-being have increased in popularity and use. Limited published literature assesses the potential of apps to support vaping cessation.

Objective: A systematic search of vaping cessation apps currently available in Australia for iOS and Android platforms was conducted. Apps were assessed against established health app assessment tools for quality and behavior change potential.

Methods: A systematic search through the Australian Apple iTunes and Google Play stores was conducted using the search terms “vape”; “vaping”; “e-cigarette”; and “cessation,” “quit,” or “quitting” in May 2023. Only apps that encouraged the cessation of vaping were included. App descriptions were reviewed to determine if they were relevant for inclusion in this study, and relevant apps were downloaded onto the appropriate mobile device for review. The Mobile App Rating Scale (MARS) was used to rate the quality (engagement, functionality, aesthetics, and information) of the apps using an overall score out of 5. The App Behavior Change Scale (ABACUS) was used to assess the behavior change potential of each app using a score out of 21.

Results: An initial search of the app stores yielded 220 Android apps and 124 iOS apps. Screening against the inclusion criteria left 20 iOS apps and 10 Android apps for review. Six apps were available on both operating systems, and these were downloaded, reviewed, and reported separately for each operating system. The average MARS score for all apps assessed in this review was 3.1 (SD 0.41) out of 5. The reviewed apps overall performed well for the MARS elements relating to functionality, such as ease of use and navigation, but had the lowest scores for information-related elements, such as credibility. The number of ABACUS behavior change features per app ranged from 0 to 19 out of 21, with a mean of 8.9 (SD 4.51). The apps commonly included information-related features, such as requesting baseline information. The least common behavior change features were those relating to goal-setting, such as asking about the user’s willingness for behavior change and providing feedback on current actions in comparison to future goals.

Conclusions: The identified vaping cessation apps had moderate levels of quality and some behavior change components. Future vaping cessation apps could benefit from including more features that are known to support behavior change, such as goal-setting, to improve the potential benefit of these apps to support people to stop vaping. As guidelines for vaping cessation continue to be established, future apps need to reference these in their development.

(JMIR Mhealth Uhealth 2024;12:e55177) doi:10.2196/55177

KEYWORDS

e-cigarettes; quit vaping apps; health apps; behavior change apps; behavior application; behavior; app; application; vaping; smoking; review; vapes; cessation; support; smartphone app; well-being; vape; mobile device; vaping cessation

Introduction

The use of e-cigarettes, also known as vaping, is a growing health issue, particularly among young people [1,2]. There is increasing concern regarding potential links between vaping and lung, heart, and brain damage as e-cigarettes often contain cancer-causing agents, toxins, heavy metals, and very fine particles that can cause adverse health effects [3-5]. There is also concern that vaping is increasing the prevalence of nicotine addiction, and people who vape are three times as likely as those who do not take up tobacco smoking [3]. Young adults and teenagers disproportionally comprise the vaping population [6].

In Australia, 19.8% of people aged 18-24 years used e-cigarettes, compared with 8.9% in the general adult population [7].

Due to the increasing awareness of the adverse effects of vaping, there is a substantial number of people interested in quitting vaping [8]. In this context, the Australian Federal Government has recently legislated significant reforms to limit the accessibility of e-cigarettes [9]. The addictive nature of nicotine in e-cigarettes means that it can be difficult for people to cease vaping, highlighting the need for developing high-quality, easily accessible, and evidence-based cessation supports. There is currently limited evidence on vaping cessation interventions [10]. Research on tobacco smoking cessation has found that mobile phone–based interventions are effective and acceptable smoking cessation support among young adults [10,11]. As nearly all (99%) people aged 18-24 years in Australia have a smartphone [12], this is a potential medium with which to explore vaping cessation support. Preliminary research demonstrates that smartphone apps would be acceptable or preferred as a vaping cessation tool for people aged 14-25 years [13], and a survey among US high school students found that of those seeking vaping cessation support, 18% had used a mobile app or SMS text messaging support [14].

The use of mobile phones to assist with behavior change (also known as mobile health) has become increasingly common, and apps have been developed for a range of health issues, including smoking cessation [15,16], alcohol reduction or cessation [17,18], and increasing physical activity [19]. There has been some research into the effectiveness of health behavior change apps, and several reviews have identified the need for these apps to have greater integration of behavior change theories and techniques [20-22].

Given the recency of vaping as a health issue, there has been scarce research on the use of apps for vaping cessation. A review in 2020 found that most vaping-related apps available on Google Play promoted vaping (they provided instructions on creating e-liquids, finding stores that sell vaping products, etc), and only 3% supported vaping cessation [23]. A subsequent Canadian review identified only 8 apps that were available on both Android and iOS platforms that had been created for vaping cessation [24]. The review assessed the apps for quality and content, and concluded that there were a limited number of apps available for vaping cessation and highlighted a need for more evidence-based practice in the development of future apps.

While the use of apps for self-management of health issues is increasing, assessment of their ability to support people in changing their health behavior is in its nascent stages, and the rapid pace of change in the app market presents a challenge to thoroughly evaluate their effectiveness in changing health behaviors. Assessment tools that appraise apps based on the inclusion of evidence-based behavior change techniques, such as goal-setting and self-monitoring, are a practical way of rapidly assessing the potential of apps to promote health behavior change [25].

As practitioners are increasingly seeking ways to support people to quit vaping, there is a need for information on both the quality and behavior change potential of vaping cessation support apps. This study aims to assess vaping cessation apps available in Australia by conducting a systematic search and using established health app assessment tools to gauge app quality and behavior change potential.

Methods

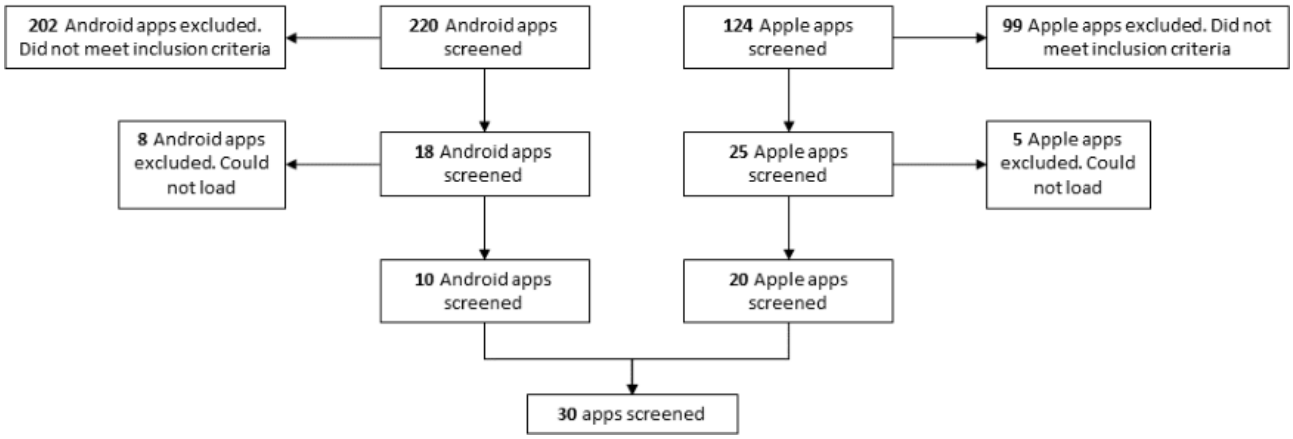
Ethical Considerations

Ethical approval was not required as all data were available in the public domain, and no human participants were involved with this study [26].

Sample Selection and Inclusion Criteria

The Australian Apple iTunes and Google Play stores were searched in May 2023 to identify vaping cessation or quitting apps. The Australian app store was searched as the research is based in Australia, and the research team was interested in the apps available in the Australian context. The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guideline [27] was modified and adapted to guide this review (Figure 1). All apps that were designed to encourage or promote vaping cessation that were available in Australia were included for analysis. Apps that were designed for smoking cessation but contained some vaping cessation components were also included. Apps that sought to promote behavior change but were not specifically related to vaping or e-cigarette cessation were excluded. Search terms were developed to include any vaping apps: “vape”; “vaping”; “e-cigarette”; and “cessation,” “quit,” or “quitting.”

Figure 1. App screening process.



All apps available for download in Australian app stores containing any of the above keywords in either the title or description were downloaded for analysis. App titles and descriptions were read to determine inclusion in the review. The inclusion criteria included apps that were created for vaping cessation and those in the English language. Any apps promoting vaping or e-cigarettes were excluded. While previous studies of apps have inclusion criteria such as an average user rating or recent update [21], given the small number of apps and the recency of vaping cessation as an issue, broad inclusion criteria were used to include all available apps.

All apps were downloaded for use on an iPhone or Samsung Galaxy Android phone. Where an app was available on both Google Play (Android) and Apple iTunes (iOS), the app was downloaded on each device and rated separately on each operating system because features and functionality may vary across the same app when developed for different operating systems.

The title and description of downloaded apps were recorded in an Excel (Microsoft Corporation) spreadsheet and rated. This study used two scales to rate the apps. The first was the Mobile App Rating Scale (MARS) [28] for quality, and the second was the App Behavior Change Scale (ABACUS) [25] to determine the potential for behavior change. The MARS and the ABACUS show good internal consistency and interrater reliability (MARS: α=.90, intraclass correlation coefficient 0.79 [28]; ABACUS: α=.93, intraclass correlation coefficient 0.91 [25]).

Quality Review and Rating (MARS)

The quality of each app was rated using the MARS [28]. This rating scale examines 19 app elements across four domains: engagement, functionality, aesthetics, and information. Each element was scored using questions on a 5-point ordinal scale. A detailed description of each element and its scoring criteria are provided in the MARS tool [27]. Based on the scores of each element, an aggregated score is calculated for each domain, and an overall app quality mean score (out of 5) is calculated. All apps were reviewed and scored by two authors (FM and MD) who have experience in evaluating health-promoting apps, including those that address addiction and behavior change. Where quality scores differed, the reviewers considered the app together, sought consensus, and determined a final score. The subjective quality rating of the MARS was not included in the analysis due to the subjective nature of this score, in line with commentary from Stoyanov et al [28]. The MARS assessment tool also includes identification of the strategies used by the app (eg, monitoring/tracking, goal-setting, information/education, meditation/mindfulness, and cognitive behavioral therapy) and affiliations of the app (eg, commercial, government, nongovernment organization, or university). These two sections do not contribute to the overall MARS score but are useful to understand the nature and context of the app.

Each app was downloaded and, consistent with other studies [28-30], used for approximately 10 minutes to allow the reviewers to familiarize themselves with the functionality of the app and user experience. Reviewers attempted to use all parts of each app. During the testing, the reviewers noted if the app crashed or if its functions were not accessible. Apps that were not able to be opened or that crashed were removed from the analysis.

Behavior Change Potential Review and Rating (ABACUS)

The ABACUS [25] comprises 21 items and was used to examine the apps' potential to support behavior change concerning goal-setting, action-planning, barrier identification, self-monitoring, and feedback. Each app was reviewed and scored by two authors (FM and MD). Each app was first explored by the reviewer to gain familiarity with the app and the interface. Reviewers used all app functions including images, cartoons, videos, record-keeping, calendars, and reminders. A total score out of 21 was calculated by summing the item scores. Where behavior change scores differed, reviewers considered the app together, sought consensus, and determined a final score.

Results

Overview of Search

The initial search of the app stores using the specified search terms yielded 220 Android apps and 124 iOS apps (Figure 1). These apps were screened through a review of the title and description; 18 Android apps and 25 iOS apps met the inclusion criteria and were included for assessment and downloaded onto the appropriate device. Eight Android apps and 5 iOS apps could not be downloaded or were no longer available and were

excluded from the sample, leaving 10 Android apps and 20 iOS apps included in this review.

Overview of Included Apps

Most apps that met the inclusion criteria were designed for Apple iOS (n=20), with a smaller number available for the Android operating system (n=10); a minority were available for both operating systems (n=6; see Multimedia Appendix 1 for names of apps). Six apps were created by a commercial organization, 2 by a university, and the remainder (n=22) had an unknown affiliation or developer. Most apps were available for free (n=27); however, of those available for free, most (n=21) had some form of in-app purchase to "enhance" the app experience.

Of the included apps, most used the strategies of allowing the user to monitor/track information (n=25), providing the user with information/education (n=15), providing strategies or tips (n=12), or allowing goal-setting (n=13). A smaller number of apps provided features related to meditation (n=9), cognitive behavioral therapy (n=4), or relaxation (n=2) strategies.

MARS Assessment for Quality

The average overall MARS score of all reviewed apps was 3.1 (SD 0.41). Individual MARS element mean scores ranged from 4 for the elements "ease of use" and "navigation" to 0 for the element "evidence base," as no app included any information about scientific trialing or testing of the app. iOS apps and Android apps were of similar quality, both with an overall mean MARS score of 3.1. Across all apps, the functionality domain score was the highest, while the information domain score was the lowest (Table 1).

Table . Mean app quality calculated using the Mobile Application Rating Scale.

Domain		iOS apps, mean (SD)	Android apps, mean (SD)	Total, mean (SD)
Engagement score		2.7 (0.7)	2.6 (0.8)	2.6 (0.7)
	Entertainment	2.4 (0.8)	2.5 (0.9)	2.4 (0.8)
	Interest	2.4 (0.9)	2.7 (1.0)	2.5 (0.9)
	Customization	2.6 (0.8)	2.4 (0.9)	2.6 (0.8)
	Interactivity	2.8 (0.7)	2.2 (1.0)	2.6 (0.9)
	Target group	3.2 (0.6)	3.1 (0.7)	3.1 (0.6)
Functionality score		3.9 (0.2)	3.9 (0.5)	3.9 (0.4)
	Performance	3.7 (0.5)	3.7 (0.9)	3.7 (0.6)
	Ease of use	3.9 (0.2)	4.0 (0.5)	4.0 (0.6)
	Navigation	4.0 (0.2)	4.1 (0.4)	4.0 (0.3)
	Gestural design	3.9 (0.2)	4.0 (0.6)	3.9 (0.4)
Aesthetics score		3.4 (0.5)	3.6 (0.5)	3.4 (0.5)
	Layout	3.8 (0.3)	3.8 (0.7)	3.8 (0.5)
	Graphics	3.5 (0.5)	3.8 (0.8)	3.6 (0.6)
	Visual appeal	2.8(0.9)	3.3 (0.5)	3.0 (0.8)
Information score		2.7 (0.5)	2.5 (0.7)	2.6 (0.5)
	Accuracy of app description	3.1 (0.6)	2.9 (0.6)	3.1 (0.7)
	Goals	2.8 (0.5)	2.3 (1.0)	2.6 (0.7)
	Quality of information	3.1 (0.7)	3.0 (1.1)	3.1 (0.8)
	Quantity of information	3.1 (0.8)	3.0 (1.2)	3.1 (0.9)
	Visual information	2.4 (0.7)	3.1 (1.0)	2.6 (0.9)
	Credibility	1.6 (0.7)	1.6 (0.6)	1.6 (0.7)
	Evidence base	0 (0.0)	0 (0.0)	0 (0.0)
Total objective score		3.1 (0.5)	3.1 (0.4)	3.1 (0.5)

When considering specific MARS elements across all apps, low mean scores were reported for evidence bases (0, as no apps had this element), credibility (1.6, SD 0.67), and entertainment (2.4, SD 0.84). The highest mean scores were obtained for ease of use (4.0, SD 0.33), navigation (4.0, SD 0.29), and gestural design (3.9, SD 0.41). Over half (n=17, 57%) of all apps were rated with an overall MARS score >3.0.

ABACUS Assessment for Potential to Support Behavior Change

In assessing the apps for behavior change potential, the number of ABACUS behavior change features per app ranged from 0 to 19 out of 21, with a mean of 8.9 (SD 4.51). On average, iOS apps had more behavior change features (9.5, SD 4.6) than Android apps (7.8, SD 4.4). The most common behavior change feature of the 30 apps was the request for baseline information (n=23, 77%), followed by the ability to self-monitor behavior (n=22, 73%) and the ability to personalize or customize the app (n=20, 67%). The least common features were the ability to understand the difference between current action and future goals (n=6, 20%), asking about willingness for behavior change (n=4, 13%), and the ability to export data from the app (n=3, 10%). Table 2 shows the frequencies of the 21 behavior change features evaluated in the apps.

Table . Behavioral change features identified.

Behavior change feature		iOS apps (n=20), n (%)	Android apps (n=10), n (%)	Total apps (N=30), n (%)
Knowledge and information				
	Ability to customize and personalize features	14 (70)	6 (60)	20 (67)
	Consistency with national guidelines or created with expertise	10 (50)	3 (30)	13 (43)
	Request for baseline information	16 (80)	7 (70)	23 (77)
	Instruction on how to perform the behavior	8 (40)	6 (60)	14 (47)
	Information about the consequences of continuing or discontinuing behavior	11 (55)	6 (60)	17 (57)
Goals and planning				
	Request for willingness for behavior change	3 (15)	1 (10)	4 (13)
	Setting of goals	8 (40)	4 (40)	12 (40)
	Ability to review goals, update, and change when necessary	6 (30)	3 (30)	9 (30)
Feedback and monitoring				
	Ability to quickly and easily understand the difference between current action and future goals	5 (25)	1 (10)	6 (20)
	Ability to allow the user to easily self-monitor behavior	16 (80)	6 (60)	22 (73)
	Ability to share behaviors with others or allow for social comparison	10 (50)	3 (30)	13 (43)
	Ability to give the user feedback—either from a person or automatically	8 (40)	3 (30)	11 (37)
	Ability to export data from app	1 (5)	2 (20)	3 (10)
	Material or social reward or incentive	8 (40)	3 (30)	11 (37)
	General encouragement	11 (55)	5 (50)	16 (53)
	Reminders or prompts or cues for activity	13 (65)	4 (40)	17 (57)
	App encourages positive habit formation	8 (40)	2 (20)	10 (33)
	App allows or encourages practice or rehearsal in addition to daily activities	10 (50)	5 (50)	15 (50)
	Opportunity to plan for barriers	8 (40)	2 (20)	10 (33)
	Assistance with or suggest restructuring the physical or social environment	8 (40)	2 (20)	10 (33)

Behavior change feature		iOS apps (n=20), n (%)	Android apps (n=10), n (%)	Total apps (N=30), n (%)
	Assistance with distraction or avoidance	7 (35)	4 (40)	11 (37)

Overall Assessment of Individual Apps

The top 5 apps according to their ABACUS and MARS scores are shown in Table 3. While the app with the highest ABACUS score did not have the highest MARS score, the top 5 apps for each operating system were the same, albeit in a different order (see Multimedia Appendix 1 for the ABACUS and MARS scores of all 30 apps included in this review).

Table . Top apps by App Behavior Change Scale (ABACUS) and Mobile App Rating Scale (MARS) scores.

App name	Operating system	Developer, affiliation	MARS score (out of 5)	ABACUS score (out of 21)	Subjective quality	Costs (up front, in-app purchases)[a]
Quit smoking. Stop vaping app	iOS	Elena Minina, unknown	3.4	19	4	Free up front, no in-app purchases
Quit vaping for good	iOS	Quit Vaping LLC, unknown	3.6	18	4	Free up front, no in-app purchases
Quit Tracker: Stop Smoking	Android	despDev, unknown	3.7	15	2.5	Free up front, Aus $4.99 per feature
QuitSure Quit Smoking Smartly	iOS	Instaquit.org, commercial	3.4	15	3	Free up front, varies Aus $9.99-$39.99
QuitSure Quit Smoking Smartly	Android	QuitSure, commercial	3.8	14	3	Free up front, varies Aus $0.49-$99.99
Kwit Quit smoking for good	iOS	KWIT, unknown	3.8	14	4	Free up front, varies Aus $6.49-$119.99

[a]A currency exchange rate of Aus $1=US $0.66 is applicable.

Discussion

Principal Findings

This review identified 30 vaping cessation apps available in app stores and assessed their quality and potential to support behavior change. The 30 apps were created for either iOS (n=20) or Android (n=10), with 6 being available on both operating systems. On average, the vaping cessation apps performed best in functionality features such as navigation and ease of use, and commonly had behavior change features such as allowing users to self-monitor their behaviors. Areas of deficiency were related to specific behavior change strategies such as comprehensive goal-setting, a lack of evidence of trials or testing, and a lack of transparency in the source of the app and its information.

Behavior Change Potential

This review goes beyond the existing research [24] by including an assessment of behavior change potential through the use of the ABACUS tool. The mean ABACUS score identified in this review of vaping cessation apps of 8.9 out of 21 is comparable to the mean ABACUS score of apps focused on other health behaviors, such as 7.6 for physical activity apps, 8.0 for apps to reduce alcohol consumption, and 8.7 for apps to improve mental well-being [21]. Notably, the mean ABACUS score for tobacco cessation apps has been reported as 10.2 [21]. As vaping cessation has some similarities with tobacco smoking cessation [31], app developers could look toward smoking cessation apps to identify strategies and features that could be adapted for vaping cessation apps.

The findings of this assessment indicate that there is potential for developers to create vaping cessation apps that include more features known to support behavior change. Goal-setting is an important feature of behavior change interventions [32], but only 12 of the 30 apps included a goal-setting feature. App developers could consider including more comprehensive goal-setting features, such as adding elements that ask the user about their willingness or readiness for behavior change and providing the user with feedback on how their current actions compare to their future goal, such as in the form of a visual graphic. Future apps could also consider the inclusion of additional features that draw on behavior change strategies, such as meditation, cognitive behavioral therapy, and relaxation, particularly as evidence shows that a significant proportion of young people vape to relieve stress or anxiety [33].

Quality of Apps

The overall quality of vaping cessation apps as assessed by the MARS was moderate, scoring 3.1 out of 5. A review of apps related to other health behaviors identified mean scores ranging from 2.71 for healthy eating apps to 3.26 for apps that aimed to improve mental well-being [21]. One of the areas where vaping cessation apps scored poorly was in credibility, with a mean score of 1.6 compared with a mean of 2.11 that has been identified for health behavior apps overall [21]. It would be

valuable for future vaping apps to provide more transparency about their information, development, and funding source. One major challenge is that information and guidelines on vaping cessation are in their infancy and may vary across different countries or jurisdictions. As health organizations begin to publish clinical guidelines for vaping cessation [34,35], it would be beneficial for app developers to consider these in the app development process.

The other key area where apps could be strengthened to provide more potential for behavior change is in the establishment of an evidence base. No reviewed app included evidence of being trialed or tested by the developer or an external party. This is a challenge for health apps broadly as the app landscape changes rapidly, and because app development generally outpaces research and knowledge translation, there is often little time for rigorous research. In practice, developing an evidence base for vaping cessation apps may need to occur in parallel to the promotion and general uptake of these apps, as the need for these apps is already present among people who want to stop vaping, particularly among young people for whom digital cessation support is acceptable [14]. Trialing and evaluation of existing and new vaping cessation apps will allow for continual development and improvement over time, and future apps could benefit from partnerships between app developers, researchers, and health behavior change experts.

Considerations for Clinicians and Practitioners

Individuals may seek out assistance and advice from their health care providers when considering vaping cessation. While the treatment approaches for tobacco cessation are long established, there is little guidance relating to vaping cessation [36] nor are there evidence-based clinical guidelines on cessation of dual use (the use of both vapes and tobacco cigarettes) [37]. The Canadian Paediatric Society recently provided preliminary clinical guidance on e-cigarette cessation for young people [38]. These guidelines suggest that behavioral therapy, either in combination with or without pharmacotherapy, should be considered when supporting young people in ceasing vaping or e-cigarette use [38]. This is consistent with Australian guidelines [35] and guidelines from the American Heart Association and American College of Cardiology [39] that suggest that practitioners engage in a range of strategies when advising young people on vaping cessation. A recent randomized controlled trial suggested that tailored mobile interventions could be considered an effective tool to support vaping cessation [40]. As such, in their discussions about vaping cessation, practitioners could encourage people interested in vaping cessation to engage in behavior strategies, such as those apps included in this review.

Considerations for App Developers

Future vaping cessation apps must be developed with an understanding of their target audience and in collaboration with users and clinicians. Users from different age groups may require different features to support their vaping cessation efforts, particularly as they may have different patterns and motivations for vaping [35]. The names and descriptions of some apps included in this review suggest that they were initially developed for tobacco smoking cessation and subsequently adapted for vaping cessation. In Australia, 10.7% of those aged 14 to 17 years are dual users [7], and there are similarities in barriers to and motivations for quitting for both behaviors [31]. Therefore, apps that provide dual support for both health behaviors could be beneficial for some audiences. However, it is important for app developers to create apps with this dual purpose in mind, rather than simply adding vaping cessation to existing smoking cessation app titles.

Finally, it is important to consider the barriers to the uptake of vaping cessation apps. While most of the apps identified in this review were free, many had in-app purchases to access additional features. If a user is unwilling or cannot afford to pay for the additional features, the effectiveness of the app may be compromised. The cost must not be a barrier for people to access health apps [41] and to ensure that people who may not have the ability to pay can access the full level of support for vaping cessation.

Limitations

While there are important findings presented here, this study presents an analysis of apps available at a point in time (May 2023) and only those available in the Australian app stores—app stores in other countries may have different apps. Although this research provides a reference point for further research into the quality of vaping cessation apps, given the fast-moving nature of this field, app developers may modify and update the features of these apps. While the development of the ABACUS assessment tool included reference to tobacco cessation evidence, the tool has only been studied for validity and reliability for physical activity apps [25]. There are perhaps differences in behavior change techniques relevant to vaping cessation in comparison to physical activity. It is also important to note the potential limitation of reviewing apps that were available on both iOS and Android platforms separately. While this may potentially duplicate the identification of certain features in the pool of available apps, this approach has been designed with careful consideration that recognizes that app features and functionality may vary when developed for different operating systems. Unlike other reviews [21], we did not include the app store star rating in our analysis. Many of the apps identified in the app stores were relatively new and as a result had a small number of reviews, making this an unreliable indicator of quality or usability.

Finally, it is important to note that with these data, we are unable to draw firm conclusions relating to long-term behavior change or provide clinical recommendations. It is also known that apps are sometimes used only for a short duration [42]. This is an area that needs more research attention as we strive to create apps that will be able to assist in improving population health at a low cost.

Conclusion

This review of vaping cessation apps found that while apps that are currently available performed reasonably in terms of quality, this review suggests that there is room for improvement, particularly in including features that support behavior change. There is a growing interest and need for effective apps to support people to stop vaping, and future vaping cessation apps could

be improved by including specific features known to support behavior change, such as goal-setting, meditation, and relaxation activities, and observing the growing body of clinical guidelines for vaping cessation.

Acknowledgments

This research is funded by the Cancer Institute of New South Wales as part of their vaping control program.

Conflicts of Interest

None declared.

Multimedia Appendix 1
App Behavior Change Scale and Mobile App Rating Scale scores for included apps.
[DOCX File, 18 KB - mhealth_v12i1e55177_app1.docx]

Checklist 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOCX File, 32 KB - mhealth_v12i1e55177_app2.docx]

References

1. Hammond D, Rynard VL, Reid JL. Changes in prevalence of vaping among youths in the United States, Canada, and England from 2017 to 2019. JAMA Pediatr 2020 Aug 1;174(8):797-800. [doi: 10.1001/jamapediatrics.2020.0901] [Medline: 32364581]
2. Miech R, Johnston L, O'Malley PM, Bachman JG, Patrick ME. Trends in adolescent vaping, 2017-2019. N Engl J Med 2019 Oct 10;381(15):1490-1491. [doi: 10.1056/NEJMc1910739] [Medline: 31532955]
3. Banks E, Yazidjoglou A, Brown S, et al. Electronic cigarettes and health outcomes: umbrella and systematic review of the global evidence. Med J Aust 2023 Apr 3;218(6):267-275. [doi: 10.5694/mja2.51890] [Medline: 36939271]
4. National Academies of Sciences, Engineering, and Medicine, Health and Medicine Division, Board on Population Health and Public Health Practice, Committee on the Review of the Health Effects of Electronic Nicotine Delivery Systems. In: Eaton DL, Kwan LY, Stratton K, editors. Public Health Consequences of E-Cigarettes: National Academies Press (US); 2018. [Medline: 29894118]
5. Scientific Committee on Health, Environmental and Emerging Risks. Opinion on electronic cigarettes. European Commission: Public Health. 2021 Apr 16. URL: https://health.ec.europa.eu/system/files/2022-08/scheer_o_017.pdf [accessed 2023-10-26]
6. Reid JL, Rynard VL, Czoli CD, Hammond D. Who is using e-cigarettes in Canada? Nationally representative data on the prevalence of e-cigarette use among Canadians. Prev Med 2015 Dec;81:180-183. [doi: 10.1016/j.ypmed.2015.08.019] [Medline: 26348453]
7. Wakefield M, Haynes A, Tabbakh T, Scollo M, Durkin S. Current vaping and current smoking in the Australian population aged 14+ years: February 2018-March 2023. Australian Government: Department of Health and Aged Care. 2023 May. URL: https://www.health.gov.au/sites/default/files/2023-06/current-vaping-and-smoking-in-the-australian-population-aged-14-years-or-older-february-2018-to-march-2023.pdf [accessed 2023-11-07]
8. Cuccia AF, Patel M, Amato MS, Stephens DK, Yoon SN, Vallone DM. Quitting e-cigarettes: quit attempts and quit intentions among youth and young adults. Prev Med Rep 2021 Jan 5;21:101287. [doi: 10.1016/j.pmedr.2020.101287] [Medline: 33489721]
9. Butler M. Taking action on smoking and vaping. Australian Government: Department of Health and Aged Care. 2023 May 2. URL: https://www.health.gov.au/ministers/the-hon-mark-butler-mp/media/taking-action-on-smoking-and-vaping [accessed 2023-10-20]
10. Berg CJ, Krishnan N, Graham AL, Abroms LC. A synthesis of the literature to inform vaping cessation interventions for young adults. Addictive Behaviors 2021 Aug;119:106898. [doi: 10.1016/j.addbeh.2021.106898] [Medline: 33894483]
11. Gowarty MA, Longacre MR, Vilardaga R, Kung NJ, Gaughan-Maher AE, Brunette MF. Usability and acceptability of two smartphone apps for smoking cessation among young adults with serious mental illness: mixed methods study. JMIR Ment Health 2021 Jul 7;8(7):e26873. [doi: 10.2196/26873] [Medline: 34255699]
12. Smartphone ownership in Australia as of July 2019, by age. Statista. 2019. URL: https://www.statista.com/statistics/730101/australia-smartphone-ownership-by-age/ [accessed 2023-10-20]
13. Huma ZE, Struik L, Bottorff JL, Hasan MK. Preferences for mobile-supported e-cigarette cessation interventions among young adults: qualitative descriptive study. JMIR Form Res 2022 Apr 1;6(4):e33640. [doi: 10.2196/33640] [Medline: 35363140]

14. Dai HD, Hanh P, Guenzel N, Morgan M, Kerns E, Winickoff JP. Adoption of vaping cessation methods by US adolescent e-cigarette users. Pediatrics 2023 Nov 1;152(5):e2023062948. [doi: 10.1542/peds.2023-062948] [Medline: 37781732]
15. Bricker JB, Mull KE, Santiago-Torres M, Miao Z, Perski O, Di C. Smoking cessation smartphone app use over time: predicting 12-month cessation outcomes in a 2-arm randomized trial. J Med Internet Res 2022 Aug 18;24(8):e39208. [doi: 10.2196/39208] [Medline: 35831180]
16. Rajani NB, Weth D, Mastellos N, Filippidis FT. Adherence of popular smoking cessation mobile applications to evidence-based guidelines. BMC Public Health 2019 Jun 13;19(1):743. [doi: 10.1186/s12889-019-7084-7] [Medline: 31196062]
17. Farren C, Farrell A, Hagerty A, McHugh C. A 6-month randomized trial of a smartphone application, ucontroldrink, in aiding recovery in alcohol use disorder. Eur Addict Res 2022;28(2):122-133. [doi: 10.1159/000519945] [Medline: 34802002]
18. Leightley D, Williamson C, Rona RJ, et al. Evaluating the efficacy of the drinks:ration mobile app to reduce alcohol consumption in a help-seeking military veteran population: randomized controlled trial. JMIR Mhealth Uhealth 2022 Jun 20;10(6):e38991. [doi: 10.2196/38991] [Medline: 35724966]
19. Damaševičius R, Kim J, Dourado VZ. Editorial: use of smartphone applications to increase physical activity and fitness. Front Public Health 2021 Jan 4;9:713306. [doi: 10.3389/fpubh.2021.713306] [Medline: 35059367]
20. Brown HM, Bucher T, Collins CE, Rollo ME. A review of pregnancy iPhone apps assessing their quality, inclusion of behaviour change techniques, and nutrition information. Matern Child Nutr 2019 Jul;15(3):e12768. [doi: 10.1111/mcn.12768] [Medline: 30569549]
21. McKay FH, Wright A, Shill J, Stephens H, Uccellini M. Using health and well-being apps for behavior change: a systematic search and rating of apps. JMIR Mhealth Uhealth 2019 Jul 4;7(7):e11926. [doi: 10.2196/11926] [Medline: 31274112]
22. Milne-Ives M, Lam C, De Cock C, Van Velthoven MH, Meinert E. Mobile apps for health behavior change in physical activity, diet, drug and alcohol use, and mental health: systematic review. JMIR Mhealth Uhealth 2020 Mar 18;8(3):e17046. [doi: 10.2196/17046] [Medline: 32186518]
23. Meacham MC, Vogel EA, Thrul J. Vaping-related mobile apps available in the Google Play store after the Apple ban: content review. J Med Internet Res 2020 Nov 13;22(11):e20009. [doi: 10.2196/20009] [Medline: 33185565]
24. Sanchez S, Kundu A, Limanto E, Selby P, Baskerville NB, Chaiton M. Smartphone apps for vaping cessation: quality assessment and content analysis. JMIR Mhealth Uhealth 2022 Mar 28;10(3):e31309. [doi: 10.2196/31309] [Medline: 35343904]
25. McKay FH, Slykerman S, Dunn M. The app behavior change scale: creation of a scale to assess the potential of apps to promote behavior change. JMIR Mhealth Uhealth 2019 Jan 25;7(1):e11130. [doi: 10.2196/11130] [Medline: 30681967]
26. National statement on ethical conduct in human research 2023. National Health and Medical Research Council. URL: https://www.nhmrc.gov.au/about-us/publications/national-statement-ethical-conduct-human-research-2023 [accessed 2024-03-08]
27. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Syst Rev 2021 Mar 29;10(1):89. [doi: 10.1186/s13643-021-01626-4] [Medline: 33781348]
28. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile App Rating Scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27. [doi: 10.2196/mhealth.3422] [Medline: 25760773]
29. Knitza J, Tascilar K, Messner EM, et al. German mobile apps in rheumatology: review and analysis using the Mobile Application Rating Scale (MARS). JMIR Mhealth Uhealth 2019 Aug 5;7(8):e14991. [doi: 10.2196/14991] [Medline: 31381501]
30. Lull C, von Ahnen JA, Gross G, et al. German mobile apps for patients with psoriasis: systematic search and evaluation. JMIR Mhealth Uhealth 2022 May 26;10(5):e34017. [doi: 10.2196/34017] [Medline: 35617014]
31. Sanchez S, Kaufman P, Pelletier H, et al. Is vaping cessation like smoking cessation? A qualitative study exploring the responses of youth and young adults who vape e-cigarettes. Addict Behav 2021 Feb;113:106687. [doi: 10.1016/j.addbeh.2020.106687] [Medline: 33045643]
32. Prestwich A, Moore S, Kotze A, Budworth L, Lawton R, Kellar I. How can smoking cessation be induced before surgery? A systematic review and meta-analysis of behavior change techniques and other intervention characteristics. Front Psychol 2017 Jun 7;8:915. [doi: 10.3389/fpsyg.2017.00915] [Medline: 28638356]
33. Donaldson CD, Stupplebeen DA, Fecho CL, Ta T, Zhang X, Williams RJ. Nicotine vaping for relaxation and coping: race/ethnicity differences and social connectedness mechanisms. Addict Behav 2022 Sep;132:107365. [doi: 10.1016/j.addbeh.2022.107365] [Medline: 35605411]
34. Guide to support young people to quit e-cigarettes. NSW Health. 2023 Jul. URL: https://www.health.nsw.gov.au/tobacco/Publications/e-cigarette-young-people-guide.pdf [accessed 2023-10-24]
35. Wahhad M, Milne BM. Clinician's guide to supporting adolescents and young adults quit vapes. The Sydney Children's Hospitals Network. URL: https://www.health.nsw.gov.au/tobacco/Publications/e-cigarette-young-people-guide.pdf [accessed 2023-10-24]

36. Zawertailo L, Kouzoukas E, Fougere C, Dragonetti R, Veldhuizen S, Selby P. Clinical guidance for e-cigarette (vaping) cessation: results from a modified Delphi panel approach. Prev Med Rep 2023 Aug 19;35:102372. [doi: 10.1016/j.pmedr.2023.102372] [Medline: 37654515]
37. Kundu A, Kouzoukas E, Zawertailo L, et al. Scoping review of guidance on cessation interventions for electronic cigarettes and dual electronic and combustible cigarettes use. CMAJ Open 2023 Apr 18;11(2):E336-E344. [doi: 10.9778/cmajo.20210325] [Medline: 37072138]
38. Chadi N, Vyver E, Bélanger RE. Protecting children and adolescents against the risks of vaping. Paediatr Child Health 2021 Sep 17;26(6):358-374. [doi: 10.1093/pch/pxab037] [Medline: 34552676]
39. Dias KJ, Ferreira G, Martin KJ, Pignataro RM. Defining the role of the physical therapist in addressing vaping and smoking cessation. Cardiopulm Phys Ther J 2022 Oct;33(4):140-150. [doi: 10.1097/CPT.0000000000000199]
40. Graham AL, Amato MS, Cha S, Jacobs MA, Bottcher MM, Papandonatos GD. Effectiveness of a vaping cessation text message program among young adult e-cigarette users: a randomized clinical trial. JAMA Intern Med 2021 Jul 1;181(7):923-930. [doi: 10.1001/jamainternmed.2021.1793] [Medline: 33999133]
41. Peng W, Kanthawala S, Yuan S, Hussain SA. A qualitative study of user perceptions of mobile health apps. BMC Public Health 2016 Nov 14;16(1):1158. [doi: 10.1186/s12889-016-3808-0] [Medline: 27842533]
42. Hamamura T, Suganuma S, Ueda M, Mearns J, Shimoyama H. Standalone effects of a cognitive behavioral intervention using a mobile phone app on psychological distress and alcohol consumption among Japanese workers: pilot nonrandomized controlled trial. JMIR Ment Health 2018 Mar 22;5(1):e24. [doi: 10.2196/mental.8984] [Medline: 29567634]

Abbreviations

ABACUS: App Behavior Change Scale
MARS: Mobile App Rating Scale
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by L Buis; submitted 04.12.23; peer-reviewed by C Berg, G Gross; revised version received 14.02.24; accepted 20.02.24; published 15.03.24.

Please cite as:
McKay F, Chan L, Cerio R, Rickards S, Hastings P, Reakes K, O'Brien T, Dunn M
Assessing the Quality and Behavior Change Potential of Vaping Cessation Apps: Systematic Search and Assessment
JMIR Mhealth Uhealth 2024;12:e55177
URL: https://mhealth.jmir.org/2024/1/e55177
doi:10.2196/55177

Patient-Centered Chronic Wound Care Mobile Apps: Systematic Identification, Analysis, and Assessment

Tassilo Dege[1*], MSc, MD; Bernadette Glatzel[1*]; Vanessa Borst[2], MSc; Franziska Grän[1], MD; Simon Goller[1], MD; Caroline Glatzel[1], MD; Matthias Goebeler[1], MD; Astrid Schmieder[1], MD

[1]Department of Dermatology, Venereology and Allergology, University Hospital Würzburg, Würzburg, Germany

[2]Department of Computer Science, University of Würzburg, Würzburg, Germany

[*]these authors contributed equally

Corresponding Author:
Tassilo Dege, MSc, MD
Department of Dermatology, Venereology and Allergology, University Hospital Würzburg, , Würzburg, , Germany

Abstract

Background: The prevalence of chronic wounds is predicted to increase within the aging populations in industrialized countries. Patients experience significant distress due to pain, wound secretions, and the resulting immobilization. As the number of wounds continues to rise, their adequate care becomes increasingly costly in terms of health care resources worldwide. eHealth support systems are being increasingly integrated into patient care. However, to date, no systematic analysis of such apps for chronic wounds has been published.

Objective: The aims of this study were to systematically identify and subjectively assess publicly available German- or English-language mobile apps for patients with chronic wounds, with quality assessments performed by both patients and physicians.

Methods: Two reviewers independently conducted a systematic search and assessment of German- or English-language mobile apps for patients with chronic wounds that were available in the Google Play Store and Apple App Store from April 2022 to May 2022. In total, 3 apps met the inclusion and exclusion criteria and were reviewed independently by 10 physicians using the German Mobile App Rating Scale (MARS) and the System Usability Scale (SUS). The app with the highest mean MARS score was subsequently reviewed by 11 patients with chronic wounds using the German user version of the MARS (uMARS) and the SUS. Additionally, Affinity for Technology Interaction (ATI) scale scores were collected from both patients and physicians.

Results: This study assessed mobile apps for patients with chronic wounds that were selected from a pool of 118 identified apps. Of the 73 apps available in both app stores, 10 were patient oriented. After excluding apps with advertisements or costs, 3 apps were evaluated by 10 physicians. Mean MARS scores ranged from 2.64 (SD 0.65) to 3.88 (SD 0.65) out of 5, and mean SUS scores ranged from 50.75 (SD 27) to 80.5 (SD 17.7) out of 100. *WUND APP* received the highest mean MARS score (mean 3.88, SD 0.65 out of 5) among physicians. Hence, it was subsequently assessed by 11 patients and achieved a similar rating (uMARS score: mean 3.89, SD 0.4 out of 5). Technical affinity, as measured with the ATI scale, was slightly lower in patients (score: mean 3.62, SD 1.35 out of 6) compared to physicians (score: mean 3.88, SD 1.03 out 6).

Conclusions: The quality ratings from physicians and patients were comparable and indicated mediocre app quality. Technical affinity, as assessed by using the ATI scale, was slightly lower for patients. Adequate apps for patients with chronic wounds remain limited, emphasizing the need for improved app development to meet patient needs. The ATI scale proved valuable for assessing technical affinity among different user groups.

(JMIR Mhealth Uhealth 2024;12:e51592) doi:10.2196/51592

KEYWORDS

chronic wounds; chronic leg ulcers; mobile applications; evaluation; mental health; Mobile Application Rating Scale; System Usability Scale; affinity for technology interaction; ATI; teledermatology; disease management; health app; skin; eHealth; telemedicine; mHealth; mobile health; app; apps; applications; quality; rating; wound; wounds; chronic; ulcer; ulcers; sore; sores; dermatology; chronic wound

Introduction

Background

Aging societies in industrialized nations are experiencing an increasing prevalence of chronic wounds, resulting in growing challenges in patient care. The high costs of therapy and personnel often limit proper treatment. Chronic wounds significantly impact patients' quality of life, requiring intensive therapy multiple times per week, along with regular medical checkups. Moreover, they have substantial economic implications, including hospitalizations, personnel costs (eg, wound care nurses, home health care services, physicians, wound managers, day clinics, etc), material costs, and management (eg, medical transportation). Purwins et al [1] identified hospitalizations, nursing staff, and material costs as the main contributors to the overall expenses.

Mobile health (mHealth) apps hold promise for bridging gaps in health care. However, a lack of evidence exists for the effectiveness of available mHealth apps [2], and high-quality trials are needed to examine their effects [3]. So far, only reviews of the use of mHealth apps without systematic searches and evaluations have been performed [4]. To address this gap in evidence for patients with chronic wounds, a systematic analysis is necessary to gather reliable data in this area.

A study by Svendsen et al [5] demonstrated that patient-centered smartphone apps can significantly improve treatment adherence in cohorts of patients with psoriasis and patients with rheumatic diseases. Participants expressed favorable views toward medical apps, indicating a willingness to use such apps if available.

To subjectively assess app quality, the Mobile App Rating Scale (MARS) [6] was developed, which evaluates engagement, functionality, aesthetics, and information. Additionally, the System Usability Scale (SUS) is a 10-item questionnaire for assessing the usability of a system. It has been effectively applied to evaluate mobile apps related to dementia, depression, pediatric obesity, and smoking cessation [7]. Moreover, the Affinity for Technology Interaction (ATI) scale [8] provides an easy and reliable means to quantify an individual's technology affinity. An aim of this study was to conduct a focused analysis on the core functionalities and core features of the included apps; therefore, we excluded apps with advertisements in order to prevent disruptions in users' experiences and advertisements' effects on usability [9]. Advertisements can be confounding variables and make it difficult to compare apps' performance or users' experiences. Advertisements can frustrate or annoy users and reduce the clarity and intuitiveness of an app's interface [10]. We also excluded paid apps to prevent bias toward users with financial means, maintain analysis inclusivity, and ensure that preferences align with intrinsic app quality rather than financial considerations, thereby enhancing the analysis' fairness and validity.

Aim of This Study

The primary objectives of this study were to identify and evaluate publicly available smartphone apps designed for patients with chronic wounds. The assessment aimed to provide subjective quality ratings for these apps, while also collecting data on the technical affinity of this specific patient group. To date, a systematic review and assessment of smartphone apps tailored for patients with chronic wounds has not been conducted.

Methods

Ethical Considerations

This study was conducted in accordance with the Declaration of Helsinki, and ethical approval was waived by the local ethics committee of the University of Würzburg.

App Selection

A systematic search of the German Google Play Store and the Apple App Store was performed from April 2022 to May 2022. The search terms used were "wound," "pressure ulcer," "ulcus," "Wunde," "Wunddokumentation," and "wound documentation." Two independent reviewers searched each app store. The inclusion criteria were that apps had to (1) be available in both app stores, (2) be available in English or German, and (3) be specifically designed for patients. Apps that were not available free of charge and apps that contained advertisements were excluded.

The following information, when available in the app stores and on the associated app websites, was collected: app name, target group (eg, patients and medical personnel), cost, platform, advertisements, features, and search term used to identify the app.

Evaluation of App Quality

The MARS [11] was developed for professional raters to evaluate mobile apps, and it is a validated and reliable scale. The user version of the MARS (uMARS) [12] was designed for users to evaluate the quality of mHealth apps. Both scales are based on a 5-point Likert scale for the following four sections: "Engagement," "Aesthetics," "Functionality," and "Information." Additionally, there is a "Subjective" section. Studies using the MARS have already been performed for several chronic diseases and apps related to breast cancer. The uMARS [12] has been broadly applied to evaluate apps for rheumatic diseases, weight loss, nutrition tracking, and menstrual tracking. The SUS is a simple, 10-item attitude Likert scale that gives an overall view of subjective assessments of usability [13].

Prior to the evaluation, suitable apps were selected based on the inclusion and exclusion criteria, resulting in a total of 3 apps. The quality of these three apps was then evaluated by 10 physicians using the MARS and SUS. Before the assessment, the physicians first watched a short training video that explained the MARS and were then asked to use the apps for more than 10 minutes.

Finally, the best app—the one with the highest mean MARS score among the physicians—was evaluated by 11 patients with chronic wounds using the uMARS and SUS.

Evaluation of Technical Affinity

Many studies that evaluate the quality of mobile apps lack information on cohorts' technical affinity, which is necessary

to assess and interpret the results. Therefore, the ATI scale [8] was used to gather information on physicians' and patients' technical affinity. The ATI scale was designed to quantify a tendency to actively engage in intensive technology interaction or a tendency to avoid technology interaction. In both the patient group and the physician group, ATI scale scores were collected. A Pearson correlation was used to correlate the ages of patients and ATI scale scores.

Comparative Analysis of Patients' and Physicians' Data

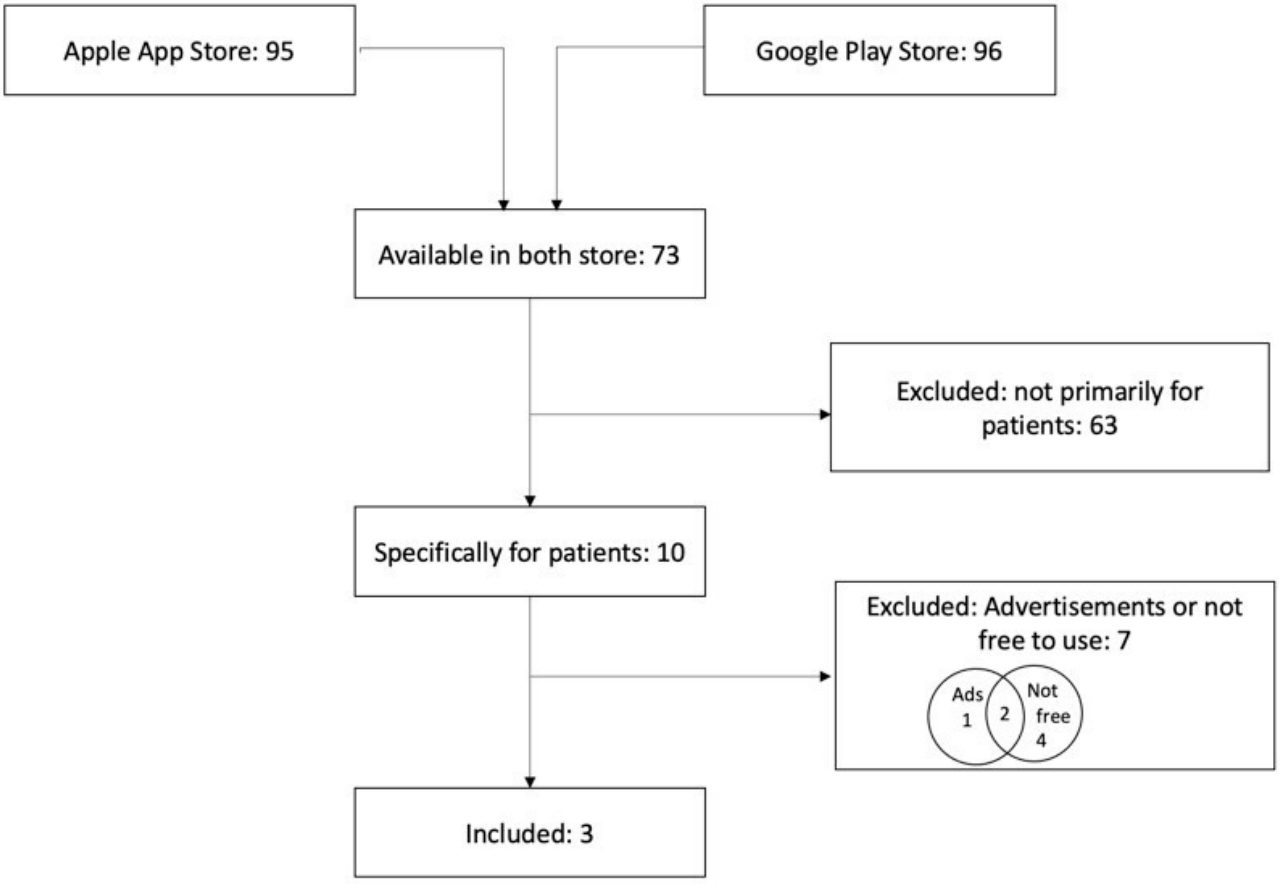
The MARS results represented the physicians' evaluations, as stated in the *Evaluation of App Quality* section, whereas uMARS results corresponded to the patients' evaluations. As a next step, after normalization, the Mann-Whitney *U* test was used to analyze whether there was a significant difference between MARS and uMARS scores and between patients' and physicians' SUS and ATI scale median scores. A comparison of the five subcategories within the MARS was performed. *P* values of <.05 were considered significant. The data analysis was performed with SPSS 23 (IBM Corp).

Results

App Selection

A total of 118 apps were identified—95 in the Apple App Store and 96 in the Google Play Store—of which 73 were available in both app stores (Figure 1). Of these 73 apps, 10 were specifically designed for patients. Of these 10 apps, 1 contained advertising, 4 were not free of charge, and 3 met both exclusion criteria, resulting in a total of 7 apps that were excluded from further analysis.

Figure 1. Flowchart illustrating the screening process for identifying suitable mobile apps.



Evaluation of App Quality

A total of 3 apps—*WoundEducation*, *APD Skin Monitoring*, and *WUND APP*—met all inclusion and exclusion criteria and were evaluated by 10 physicians. *WoundEducation* provides an overview of different wound types and treatment options. *APD Skin Monitoring* includes an automated area calculation function to quantify wound area. *WUND APP* offers advice and information on chronic wounds and contains a diary function to track patient-related outcomes, such as pain or the level of wound secretion.

The mean MARS scores (and SDs) of the physicians, including subcategory scores and SUS scores, are provided in Table 1.

Table . Evaluation of *WUND APP*, *APD Skin Monitoring*, and *WoundEducation* by 10 physicians using the Mobile App Rating Scale (MARS) and System Usability Scale (SUS).

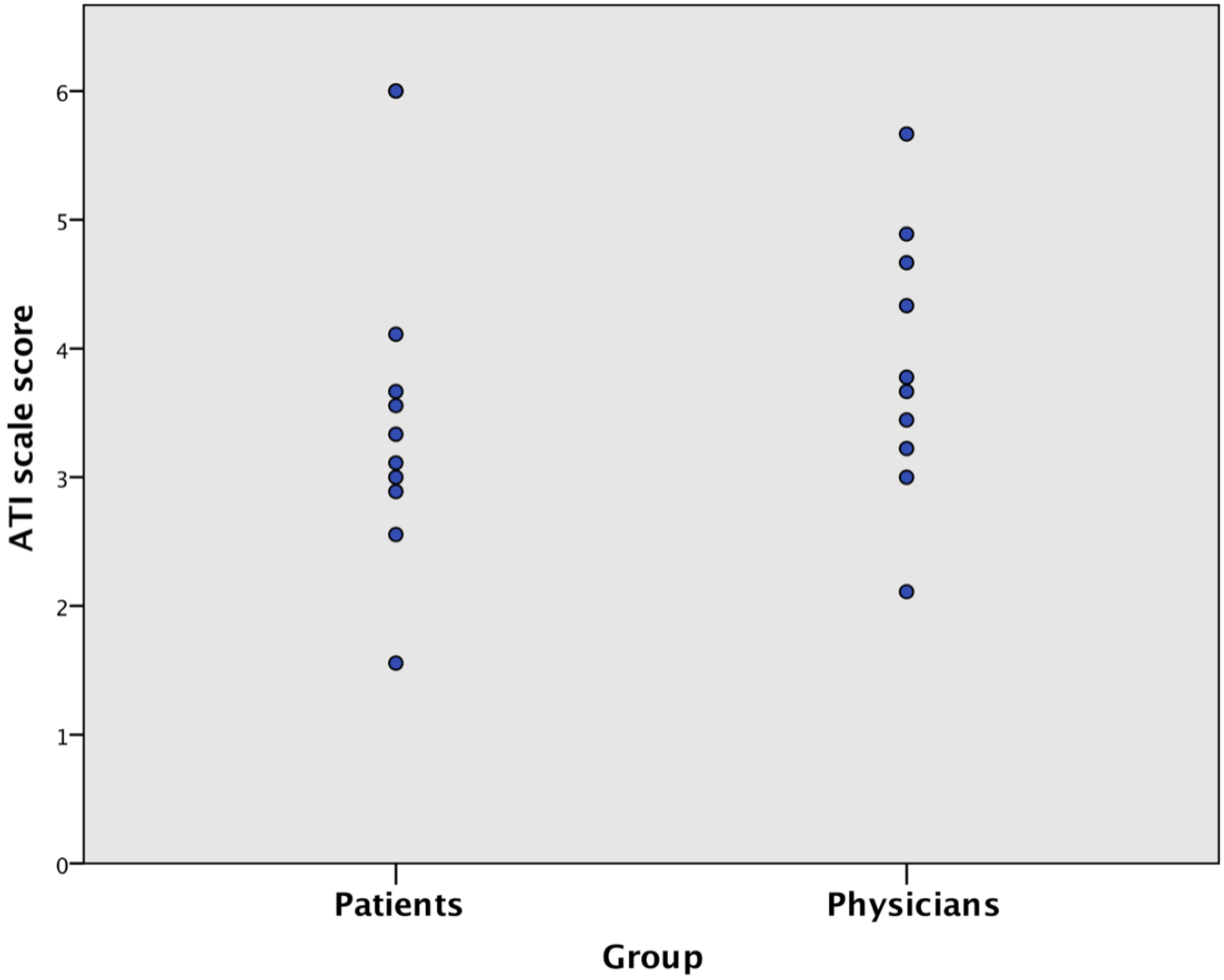
Mobile app name	MARS score, mean (SD)	Engagement score, mean (SD)	Functionality score, mean (SD)	Aesthetics score, mean (SD)	Information score, mean (SD)	Psychotherapy score, mean (SD)	SUS score (%), mean (SD)
WUND APP	3.88 (0.65)	3.36 (0.89)	4.38 (0.66)	4.13 (0.76)	3.67 (0.56)	2.98 (0.36)	80.5 (17.7)
WoundEducation	3.01 (0.5)	2.14 (0.44)	3.93 (0.99)	2.73 (0.75)	3.25 (0.58)	2.57 (0.46)	72.75 (18.5)
APD Skin Monitoring	2.64 (0.65)	2.36 (0.56)	2.65 (0.83)	2.77 (0.69)	2.79 (0.82)	2.47 (0.63)	50.75 (27)

WUND APP had the highest mean MARS score (mean 3.88, SD 0.65 out of 5) and was subsequently analyzed by 11 patients with chronic wounds. *WUND APP* had a similar mean uMARS score (mean 3.89, SD 0.4 out of 5) when analyzed by the patients.

Evaluation of Technical Affinity

The ATI scale scores ranged from 1.56 to 6 out of 6 for patients and from 2.11 to 5.67 out of 6 for physicians (Figure 2). An ATI scale score of >3 indicates average technology affinity, and a score of >4 indicates high technology affinity [8]. Mean ATI scale scores were slightly lower for patients (mean 3.62, SD 1.35 out of 6) than those for physicians (mean 3.88, SD 1.03 out of 6), but the difference was not statistically significant (P=.43). Patients ranged in age from 28 to 70 years; 7 were male, and 4 were female. Physicians ranged in age from 26 to 43 years; 4 were male, and 6 were female. The Pearson correlation coefficient between ATI scale scores and the ages of patients was –0.706 (P=.02), indicating a strong negative correlation.

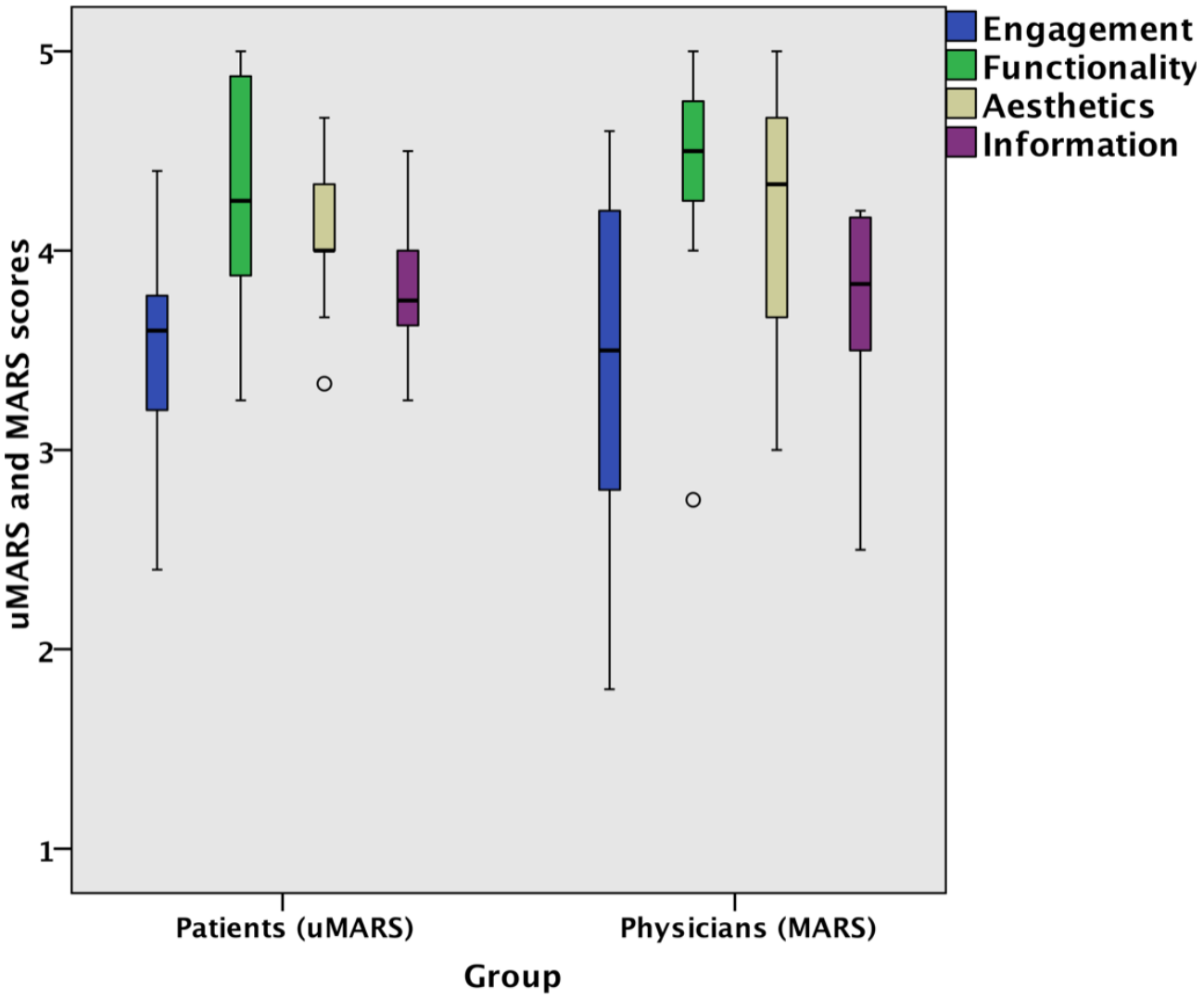
Figure 2. Scatter plot depicting ATI scale scores in patients and physicians. ATI: Affinity for Technology Interaction.



Comparative Analysis of Patients' and Physicians' Data

MARS and uMARS scores were almost identical. The subcategory "Information" was rated worse by physicians, while the subcategory "Functionality" was rated worse by patients (Figure 3). For the subcategories "Aesthetics" and "Engagement," no differences between physicians and patients were found. No statistically significant difference (Mann-Whitney *U* test) was found between MARS and uMARS scores (P=.76) or between patients' and physicians' SUS scores (P=.39; Multimedia Appendix 1).

Figure 3. Comparison of MARS and uMARS subscale scores ("Engagement," "Functionality," "Aesthetics," and "Information" scores). Physicians (professional raters) used the MARS, and patients used the uMARS. MARS: Mobile App Rating Scale; uMARS: user version of the Mobile App Rating Scale.



Discussion

Principal Findings

This study represents the first systematic examination that aimed to identify and assess smartphone apps specifically designed for patients with chronic wounds. The quality of these apps was evaluated by independent professional reviewers (physicians) and patients using well-validated scoring systems—the MARS, uMARS, and SUS. Using subjective measures in the assessment of eHealth is relevant and essential, and it is crucial to recognize this fact. Agreement among the assessors, even for subjective measures, is a good sign of the reliability of the assessment [14].

The overall findings indicate that the apps received moderate ratings. Among all apps evaluated, *WUND APP* achieved the highest mean MARS score (mean 3.88, SD 0.65) when assessed by physicians. Similarly, it received a mean uMARS score of 3.89 (SD 0.4) when evaluated by patients. Given that the preselection was exclusively conducted by physicians, patients' assessments might have differed.

WUND APP includes a diary function for documenting wound photos and patient-related outcomes, such as pain or wound secretion. Furthermore, it has a reminder function for necessities such as physician appointments and dressing changes. Its user interface is clear, and the app is user-friendly. Consequently, physicians assigned high scores for the subcategories "Engagement" (mean 3.36, SD 0.89), "Functionality" (mean 4.38, SD 0.66), and "Aesthetics" (mean 4.13, SD 0.76). Additionally, the app provides information on the causes, diagnostics, and treatments of the different wound types,

resulting in a mean MARS score of 3.67 (SD 0.56) for the subcategory "Information."

WoundEducation presents information in a straightforward structure that resembles an article with embedded links, delivers medical information, and illustrates various wound types through examples. The app received high ratings for overall functionality (mean 3.93, SD 0.99) and information provision (mean 3.25, SD 0.58). However, it demonstrated limited interactivity, as reflected by having the lowest score for the "Engagement" subcategory (mean 2.14, SD 0.44). Additionally, its aesthetics (mean 2.73, SD 0.75) were rated the lowest among the three apps.

APD Skin Monitoring [15] uses a coin as a reference for calculating the area of a wound, proving particularly useful in assessing wound progression and healing status regardless of whether the wounds are irregularly shaped. In addition, the coloration of wounds can be analyzed and tracked over time. However, the app received the lowest ratings for the subcategories "Functionality" (mean 2.65, SD 0.83) and "Engagement" (mean 2.36, SD 0.56).

Our findings revealed that a limited number of wound apps were specifically tailored for patients, accounting for only 14% (10/73) of all wound apps that were available in both app stores. Overall, both physicians and patients rated *WUND APP* similarly; however, although not statistically significant, it is noteworthy that differences emerged in 2 subcategories ("Information": P=.97; "Functionality": P=.56). Physicians rated the information content lower, which could be attributed to their expert medical knowledge. As professionals, they may have had higher expectations regarding the app's information content and may have found certain aspects less informative. In a recent study, it was shown that patients trusted recommendations and reviews from medical organizations and health care professionals when selecting apps [16]; their motivation to continue using apps was driven by features that supported goal setting and tracking, data sharing, decision-making, and empowerment.

On the other hand, patients rated the functionality of *WUND APP* lower, which was possibly due to their limited involvement in the app's development process. A systemic review reported that health care professionals were engaged in the development process for only 35% of the 7 analyzed apps that were specifically designed for patients with rheumatoid arthritis [17]. Another systematic review revealed that patients were only engaged in the development process for 15% of the 32 analyzed apps that were designed for individuals with rheumatic and musculoskeletal diseases [18]. The inclusion of patients in the development process of future wound apps could help to ensure that the apps meet their specific needs and preferences. Our study highlights the scarcity of wound apps designed explicitly for patients and the importance of involving patients in the app development process. Tailoring apps to meet patients' specific requirements and involving them in the design process would likely result in improved app functionality and overall user satisfaction.

To the best of our knowledge, this study represents the first systematic study to collect data on the technical affinity of patients with wounds. Surprisingly, no statistically significant difference in technical affinity was observed between wound care patients and physicians, even with the inclusion of patients aged up to 70 years (P=.43). Nevertheless, future studies and app development projects should aim to include older patients and comprehensively assess and address their specific needs. The low adoption and use of mHealth apps among older patients are frequently attributed to inadequate designs [19].

In 2019, Germany introduced a digital health app (DiGA) directory that includes scientifically validated apps. Physicians can prescribe DiGAs in a manner similar to how they prescribe medications [20]. However, to date, no DiGAs specifically tailored for patients with chronic wounds are available in the directory. It is worth noting that physicians with a higher technical affinity and those who are female hold significantly more positive attitudes toward DiGAs [21].

The absence of wound care apps highlights the unmet potential for innovative digital solutions to address the needs of patients with chronic wounds. The inclusion of validated and effective wound care apps in the DiGA directory could significantly improve patient outcomes and health care management in this specific area. It is imperative for future app development initiatives to focus on developing and validating apps that cater to the unique requirements of patients with chronic wounds, to provide them with accessible and effective digital health care resources.

Multiple attempts have been made to integrate mHealth apps into wound care [22], including apps for wound care measurements [23], wound care dressing decision support systems [24], and home-based self-management systems [25]. An Australian study assessed an artificial intelligence app for wound assessment, involving 166 patients in the standard group and 124 in the intervention group. The intervention group demonstrated significantly improved wound documentation, along with positive outcomes such as enhanced patient adherence, efficient digital care provision, and substantial reductions in wound size [26].

A cultural shift toward greater technology affinity has been accelerated by the COVID-19 pandemic [27]. Patients are increasingly using mobile apps when they perceive clear benefits to their use, such as reducing social contact during the COVID-19 pandemic. For patients with chronic wounds, these benefits may include using an app as a diary, while for others, apps may offer the advantage of saving time and transportation costs through telemedicine services. Telemedical approaches could significantly alter and improve the current wound care landscape, bearing the potential to improve the efficiency, accuracy, and accessibility of both diagnoses and treatments. Telemedicine in chronic wound management was shown to be noninferior to conventional standard care in a systematic review and meta-analysis [28]. The earlier diagnosis of complications, such as wound infections, can reduce the use of antibiotics and lower health care costs by preventing hospital stays.

Limitations

Due to the limited number of apps that met our strict inclusion and exclusion criteria, only a small subset could be analyzed in

this study. The selection of apps was conducted within a relatively short time frame, which might have further restricted the available options for evaluation.

The nonsignificant differences between physicians' and patients' ratings in this study may be attributed to the small sample of only 11 patients and 10 physicians, which resulted in limited statistical power. Therefore, we emphasize the importance of cautious interpretation and the consideration of larger sample sizes for future research.

Another factor that influenced this study's results was the inclusion of patients who agreed to evaluate *WUND APP*. This approach potentially introduced selection bias, as our patient sample may not represent the true technical affinity of all patients with chronic wounds. The overall technical affinity of a larger and more diverse patient population might be lower than what was observed in this study. Future studies with larger patient cohorts might provide a more comprehensive understanding of the technical affinity and usability experience of patients with chronic wounds who use mHealth apps. Further, the assessment relied on a qualitative survey—a method that may be susceptible to various biases. However, to truly ascertain the effectiveness of *WUND APP*, a randomized controlled trial is needed.

Conclusions

Patient involvement is crucial in app development. By involving all stakeholders, including physicians, wound care experts, and patients, throughout the development process, apps can be tailored to meet the specific needs and preferences of the end users, resulting in increased user satisfaction and improved health outcomes.

The validated ATI scale proved to be a valuable tool for evaluating an individual's technical affinity. In future studies and app evaluations, technical affinity should be determined to generalize outcomes to specific patient and consumer cohorts.

Acknowledgments

TD is supported by the Clinician Scientist Program (project number: Z-2/CSP-25) of the Interdisciplinary Center of Clinical Research (IZKF) of the Medical Faculty, University of Würzburg.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Patient and physician characteristics (age and sex) and the Affinity for Technology Interaction scale, System Usability Scale, Mobile App Rating Scale (MARS), and user version of the MARS scores (including subcategory scores) of all 11 patients and 10 physicians who participated.
[DOCX File, 19 KB - mhealth_v12i1e51592_app1.docx]

References

1. Purwins S, Herberger K, Debus ES, et al. Cost-of-illness of chronic leg ulcers in Germany. Int Wound J 2010 Apr;7(2):97-102. [doi: 10.1111/j.1742-481X.2010.00660.x] [Medline: 20529149]
2. Buijink AWG, Visser BJ, Marshall L. Medical apps for smartphones: lack of evidence undermines quality and safety. Evid Based Med 2013 Jun;18(3):90-92. [doi: 10.1136/eb-2012-100885] [Medline: 22923708]
3. Zhao J, Freeman B, Li M. Can mobile phone apps influence people's health behavior change? an evidence review. J Med Internet Res 2016 Oct 31;18(11):e287. [doi: 10.2196/jmir.5692] [Medline: 27806926]
4. Koepp J, Baron MV, Martins PRH, et al. The quality of mobile apps used for the identification of pressure ulcers in adults: systematic survey and review of apps in app stores. JMIR Mhealth Uhealth 2020 Jun 16;8(6):e14266. [doi: 10.2196/14266] [Medline: 32470916]
5. Svendsen MT, Andersen F, Andersen KH, et al. A smartphone application supporting patients with psoriasis improves adherence to topical treatment: a randomized controlled trial. Br J Dermatol 2018 Nov;179(5):1062-1071. [doi: 10.1111/bjd.16667] [Medline: 29654699]
6. Messner EM, Terhorst Y, Barke A, et al. The German version of the Mobile App Rating Scale (MARS-G): development and validation study. JMIR Mhealth Uhealth 2020 Mar 27;8(3):e14479. [doi: 10.2196/14479] [Medline: 32217504]
7. Hyzy M, Bond R, Mulvenna M, et al. System Usability Scale benchmarking for digital health apps: meta-analysis. JMIR Mhealth Uhealth 2022 Aug 18;10(8):e37290. [doi: 10.2196/37290] [Medline: 35980732]
8. Franke T, Attig C, Wessel D. A personal resource for technology interaction: development and validation of the Affinity for Technology Interaction (ATI) scale. Int J Hum Comput Interact 2019;35(6):456-467. [doi: 10.1080/10447318.2018.1456150]
9. Haggag O, Grundy J, Abdelrazek M, Haggag S. A large scale analysis of mHealth app user reviews. Empir Softw Eng 2022;27(7):196. [doi: 10.1007/s10664-022-10222-6] [Medline: 36246486]

10. Lobo EH, Abdelrazek M, Frølich A, et al. Detecting user experience issues from mHealth apps that support stroke caregiver needs: an analysis of user reviews. Front Public Health 2023 May 25;11:1027667. [doi: 10.3389/fpubh.2023.1027667] [Medline: 37304084]
11. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile App Rating Scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27. [doi: 10.2196/mhealth.3422] [Medline: 25760773]
12. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and validation of the user version of the Mobile Application Rating Scale (uMARS). JMIR Mhealth Uhealth 2016 Jun 10;4(2):e72. [doi: 10.2196/mhealth.5849] [Medline: 27287964]
13. Brooke J. SUS: a quick and dirty usability scale. In: Jordan PW, Thomas, B, McClelland IL, Weerdmeester B, editors. Usability Evaluation In Industry: CRC Press; 1996. [doi: 10.1201/9781498710411]
14. Powell R, Brown K, Davies M, et al. The value of the modern vascular surgeon to the health care system: a report from the Society for Vascular Surgery Valuation Work Group. J Vasc Surg 2021 Feb;73(2):359-371.e3. [doi: 10.1016/j.jvs.2020.05.056] [Medline: 32585182]
15. Wu W, Yong KYW, Federico MAJ, Gan SKE. The APD Skin Monitoring app for wound monitoring: image processing, area plot, and colour histogram. Scientific Phone Apps and Mobile Devices 2019 May 28. [doi: 10.30943/2019/28052019]
16. Jezrawi R, Balakumar S, Masud R, et al. Patient and physician perspectives on the use and outcome measures of mHealth apps: exploratory survey and focus group study. Digit Health 2022 May 24;8:20552076221102773. [doi: 10.1177/20552076221102773] [Medline: 35646382]
17. Luo D, Wang P, Lu F, Elias J, Sparks JA, Lee YC. Mobile apps for individuals with rheumatoid arthritis: a systematic review. J Clin Rheumatol 2019 Apr;25(3):133-141. [doi: 10.1097/RHU.0000000000000800] [Medline: 29933327]
18. Najm A, Gossec L, Weill C, Benoist D, Berenbaum F, Nikiphorou E. Mobile health apps for self-management of rheumatic and musculoskeletal diseases: systematic literature review. JMIR Mhealth Uhealth 2019 Nov 26;7(11):e14730. [doi: 10.2196/14730] [Medline: 31769758]
19. Liu N, Yin J, Tan SSL, Ngiam KY, Teo HH. Mobile health applications for older adults: a systematic review of interface and persuasive feature design. J Am Med Inform Assoc 2021 Oct 12;28(11):2483-2501. [doi: 10.1093/jamia/ocab151] [Medline: 34472601]
20. Sauermann S, Herzberg J, Burkert S, Habetha S. DiGA - a chance for the German healthcare system. J Eur CME 2021 Dec 23;11(1):2014047. [doi: 10.1080/21614083.2021.2014047] [Medline: 34992948]
21. Dahlhausen F, Zinner M, Bieske L, Ehlers JP, Boehme P, Fehring L. Physicians' attitudes toward prescribable mHealth apps and implications for adoption in Germany: mixed methods study. JMIR Mhealth Uhealth 2021 Nov 23;9(11):e33012. [doi: 10.2196/33012] [Medline: 34817385]
22. Friesen MR, Hamel C, McLeod RD. A mHealth application for chronic wound care: findings of a user trial. Int J Environ Res Public Health 2013 Nov 19;10(11):6199-6214. [doi: 10.3390/ijerph10116199] [Medline: 24256739]
23. Biagioni RB, Carvalho BV, Manzioni R, Matielo MF, Neto FCB, Sacilotto R. Smartphone application for wound area measurement in clinical practice. J Vasc Surg Cases Innov Tech 2021 Feb 26;7(2):258-261. [doi: 10.1016/j.jvscit.2021.02.008] [Medline: 33997567]
24. Jordan S, McSwiggan J, Parker J, Halas GA, Friesen M. An mHealth app for decision-making support in wound dressing selection (wounds): protocol for a user-centered feasibility study. JMIR Res Protoc 2018 Apr 24;7(4):e108. [doi: 10.2196/resprot.9116] [Medline: 29691213]
25. Zhang J, Mihai C, Tüshaus L, Scebba G, Distler O, Karlen W. Wound image quality from a mobile health tool for home-based chronic wound management with real-time quality feedback: randomized feasibility study. JMIR Mhealth Uhealth 2021 Jul 30;9(7):e26149. [doi: 10.2196/26149] [Medline: 34328440]
26. Barakat-Johnson M, Jones A, Burger M, et al. Reshaping wound care: evaluation of an artificial intelligence app to improve wound assessment and management amid the COVID-19 pandemic. Int Wound J 2022 Oct;19(6):1561-1577. [doi: 10.1111/iwj.13755] [Medline: 35212459]
27. Rettinger L, Klupper C, Werner F, Putz P. Changing attitudes towards teletherapy in Austrian therapists during the COVID-19 pandemic. J Telemed Telecare 2023 Jun;29(5):406-414. [doi: 10.1177/1357633X20986038] [Medline: 33430678]
28. Chen L, Cheng L, Gao W, Chen D, Wang C, Ran X. Telemedicine in chronic wound management: systematic review and meta-analysis. JMIR Mhealth Uhealth 2020 Jun 25;8(6):e15574. [doi: 10.2196/15574] [Medline: 32584259]

Abbreviations

ATI: Affinity for Technology Interaction
DiGA: digital health app
MARS: Mobile App Rating Scale
mHealth: mobile health
SUS: System Usability Scale
uMARS: user version of the Mobile App Rating Scale

Edited by L Buis; submitted 04.08.23; peer-reviewed by C Jacob, M Zhou; revised version received 17.12.23; accepted 25.01.24; published 21.03.24.

Please cite as:
Dege T, Glatzel B, Borst V, Grän F, Goller S, Glatzel C, Goebeler M, Schmieder A
Patient-Centered Chronic Wound Care Mobile Apps: Systematic Identification, Analysis, and Assessment
JMIR Mhealth Uhealth 2024;12:e51592
URL: https://mhealth.jmir.org/2024/1/e51592
doi:10.2196/51592

Reliability Issues of Mobile Nutrition Apps for Cardiovascular Disease Prevention: Comparative Study

Dang Khanh Ngan Ho[1], PhD; Wan-Chun Chiu[1,2], PhD; Jing-Wen Kao[1], MSc; Hsiang-Tung Tseng[1], MSc; Cheng-Yu Lin[1], BSc; Pin-Hsiang Huang[1], BSc; Yu-Ren Fang[1], BSc; Kuei-Hung Chen[1], BSc; Ting-Ying Su[1], BSc; Chia-Hui Yang[1], BSc; Chih-Yuan Yao[3], PhD; Hsiu-Yueh Su[1,4], MSc; Pin-Hui Wei[2], MSc; Jung-Su Chang[1,5,6,7,8], PhD

[1]School of Nutrition and Health Sciences, College of Nutrition, Taipei Medical University, 250 Wuxing Street, Xinyi District, Taipei, Taiwan

[2]Department of Nutrition, Wan Fang Hospital, Taipei Medical University, Taipei, Taiwan

[3]Department of Computer Science and Information Engineering, National Taiwan University of Science and Technology, Taipei, Taiwan

[4]Department of Dietetics, Taipei Medical University Hospital, Taipei, Taiwan

[5]Graduate Institute of Metabolism and Obesity Sciences, College of Nutrition, Taipei Medical University, Taipei, Taiwan

[6]Nutrition Research Center, Taipei Medical University Hospital, Taipei, Taiwan

[7]Chinese Taipei Society for the Study of Obesity (CTSSO), Taipei, Taiwan

[8]TMU Research Center for Digestive Medicine, Taipei Medical University, Taipei, Taiwan

Corresponding Author:
Jung-Su Chang, PhD
School of Nutrition and Health Sciences, College of Nutrition, Taipei Medical University, , 250 Wuxing Street, Xinyi District, Taipei, , Taiwan

Abstract

Background: Controlling saturated fat and cholesterol intake is important for the prevention of cardiovascular diseases. Although the use of mobile diet-tracking apps has been increasing, the reliability of nutrition apps in tracking saturated fats and cholesterol across different nations remains underexplored.

Objective: This study aimed to examine the reliability and consistency of nutrition apps focusing on saturated fat and cholesterol intake across different national contexts. The study focused on 3 key concerns: data omission, inconsistency (variability) of saturated fat and cholesterol values within an app, and the reliability of commercial apps across different national contexts.

Methods: Nutrient data from 4 consumer-grade apps (COFIT, MyFitnessPal-Chinese, MyFitnessPal-English, and LoseIt!) and an academic app (Formosa FoodApp) were compared against 2 national reference databases (US Department of Agriculture [USDA]–Food and Nutrient Database for Dietary Studies [FNDDS] and Taiwan Food Composition Database [FCD]). Percentages of missing nutrients were recorded, and coefficients of variation were used to compute data inconsistencies. One-way ANOVAs were used to examine differences among apps, and paired 2-tailed *t* tests were used to compare the apps to national reference data. The reliability across different national contexts was investigated by comparing the Chinese and English versions of MyFitnessPal with the USDA-FNDDS and Taiwan FCD.

Results: Across the 5 apps, 836 food codes from 42 items were analyzed. Four apps, including COFIT, MyFitnessPal-Chinese, MyFitnessPal-English, and LoseIt!, significantly underestimated saturated fats, with errors ranging from −13.8% to −40.3% (all $P<.05$). All apps underestimated cholesterol, with errors ranging from −26.3% to −60.3% (all $P<.05$). COFIT omitted 47% of saturated fat data, and MyFitnessPal-Chinese missed 62% of cholesterol data. The coefficients of variation of beef, chicken, and seafood ranged from 78% to 145%, from 74% to 112%, and from 97% to 124% across MyFitnessPal-Chinese, MyFitnessPal-English, and LoseIt!, respectively, indicating a high variability in saturated fats across different food groups. Similarly, cholesterol variability was consistently high in dairy (71%-118%) and prepackaged foods (84%-118%) across all selected apps. When examining the reliability of MyFitnessPal across different national contexts, errors in MyFitnessPal were consistent across different national FCDs (USDA-FNDSS and Taiwan FCD). Regardless of the FCDs used as a reference, these errors persisted to be statistically significant, indicating that the app's core database is the source of the problems rather than just mismatches or variances in external FCDs.

Conclusions: The findings reveal substantial inaccuracies and inconsistencies in diet-tracking apps' reporting of saturated fats and cholesterol. These issues raise concerns for the effectiveness of using consumer-grade nutrition apps in cardiovascular disease prevention across different national contexts and within the apps themselves.

(JMIR Mhealth Uhealth 2024;12:e54509) doi:10.2196/54509

KEYWORDS

mobile apps; mHealth; dietary assessment; validity; cardiovascular disease prevention; app; apps; applications; application; nutrition; cardiovascular; nutrients; fitness; diet; mobile health

Introduction

Cardiovascular diseases (CVDs) are leading causes of global mortality and major contributors to disability [1]. Excessive intake of saturated fats and cholesterol was significantly associated with increased risks of CVDs and all-cause mortality [2-4]. The American Heart Association has emphasized that reducing saturated fat intake in favor of unsaturated fats can significantly diminish CVD risks [5]. Managing these nutrients is crucial for early prevention efforts [6,7]. In this context, the advent of mobile diet-tracking apps may offer a modern solution for monitoring dietary saturated fats and cholesterol, with benefits such as accessibility, cost-effectiveness, and efficiency over traditional dietary assessment methods [8-10].

Khazen et al [11] has categorized diet-tracking apps into 2 main types: academic and commercial. Academic apps are developed with research-based validation but are often limited by their geographical scope and reliance on local food composition databases (FCDs). Some examples of academic apps are Australia's Electronic Dietary Intake Assessment [12], Canada's Keenoa [13], and Taiwan's Formosa FoodApp [14]. On the other hand, commercial apps (such as MyFitnessPal, FatSecret, and Lose It!) are known for their extensive, international FCDs and are designed to cater to multilingual users [11,15,16]. They include a function that allows users to add new food products not available in the existing FCD, characterizing them as consumer-grade apps with a consumer-oriented approach. Language availability is another factor differentiating academic and consumer-grade apps [11]. Academic apps typically support only 1 or 2 languages and are tailored for specific research within certain populations, thus limiting their use in global studies [11,17]. In contrast, commercial apps, with their support for multiple languages [18], are aimed at a global audience, positioning them as versatile or "universal" tools accessible to users across the world.

The primary concerns with consumer-grade apps are the quality and reliability of their FCDs to estimate energy and nutrient intake [11,19]. Regarding the reliability, academic apps have been in agreement with conventional self-report dietary assessment methods such as 24-hour dietary recall [14,20,21]. In contrast, the accuracy of consumer-grade apps varies, generally showing a trend of underestimating nutritional values, with variability in accuracy across different commercial apps [19,22]. This is likely due to the varying quality of FCDs among commercial apps, which tend to suffer from missing or redundant data [19,22]. Although many studies have emphasized the disparities in dietary intake of total energy and macronutrients between data from commercial apps and national food databases [10,22,23], very few have delved into the variability in saturated fat and cholesterol values. While many commercial apps derive data from nutrition labels with mandated disclosures for saturated fats and cholesterol, these aspects in commercial nutrition apps remain insufficiently evaluated.

Currently, the reliability of the consumer-grade apps in tracking nutrients across different national standards remains underexplored. This study aimed to examine the reliability of mobile nutrition apps by focusing on saturated fat and cholesterol intake and the consistency of commercial apps across different national contexts. This critical evaluation focused on 3 key concerns: omission of saturated fat and cholesterol data, data consistency within an app, and the reliability of commercial apps across different national contexts.

Methods

App Selection

We selected widely used apps from 2 distinct regions, Taiwan and the United States, to represent linguistic and cultural differences. Specifically, we chose the academic app Formosa FoodApp [14] and the commercial app COFIT [24] given their widespread clinical and educational acceptance in Taiwan. In addition, we also explored 2 commercial US-based apps—MyFitnessPal and Lose It!—based on their broad acknowledgment, dominant presence in both clinical settings and among consumers, and citations in earlier literature [10,15,19]. The primary sources for nutrient data included the Taiwan FCD, US Department of Agriculture (USDA)–Food and Nutrient Database for Dietary Studies (FNDDS), and US food manufacturers and restaurants. Notably, MyFitnessPal and LoseIt! make user-generated food entries available to all users (Table 1).

Table . Description of the mobile app food composition databases (FCDs) in the study.

Characteristics	Formosa FoodApp	COFIT	MyFitnessPal-Chinese	MyFitnessPal-English	LoseIt!
Type of FCD	Academic	Commercial	Commercial	Commercial	Commercial
Manufacturer and country	Professor Susan Chang's Lab (Taipei Medical University), Taiwan	Cofit Healthcare, Taiwan	Francisco Partners, United States	Francisco Partners, United States	FitNow, United States
FCD language	Chinese	Chinese	Chinese (added by users)	English	English
FCD sources	Taiwan FCD, USDA[a]-FNDSS[b], Vietnam FCD, Indonesia FCD, food manufacturers, and restaurants	Taiwan FCD, food manufacturers, and restaurants	Taiwan FCD and users	USDA-FNDSS, food manufacturers, restaurants, and users	USDA-FNDSS, food manufacturers, restaurants, and users
User-added function	No	No	Yes	Yes	Yes

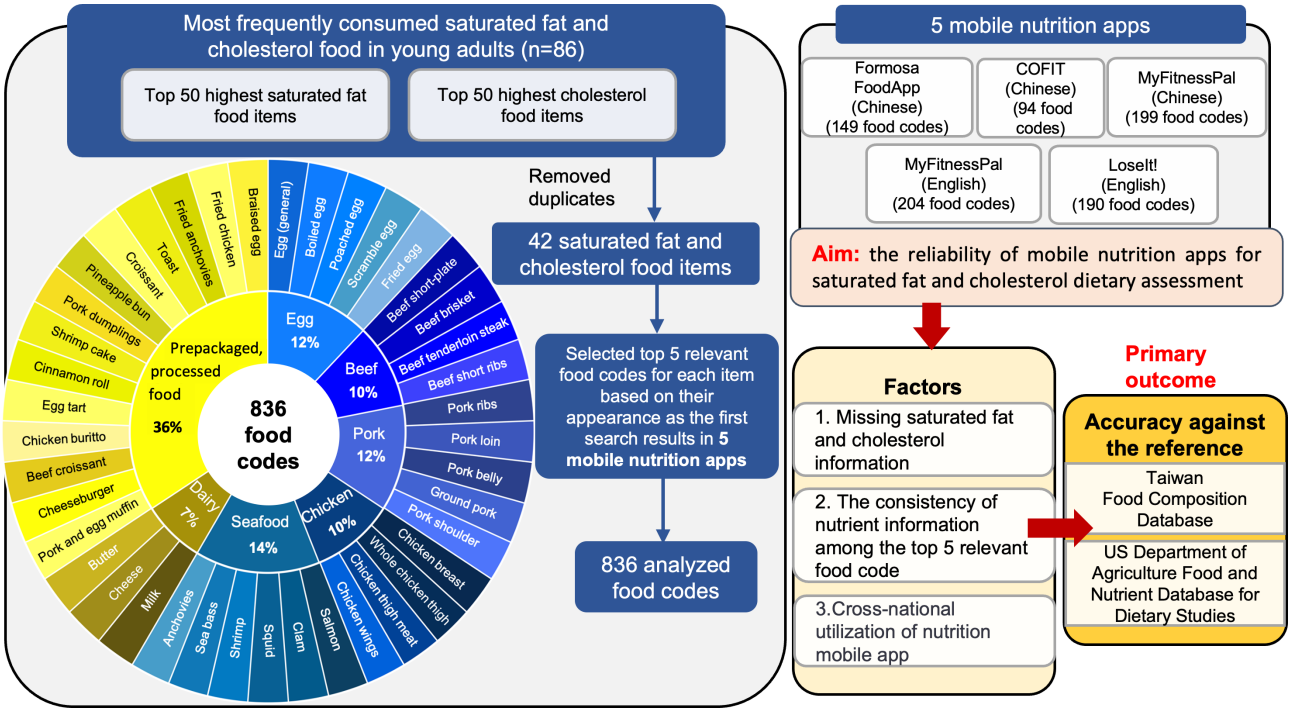
[a]USDA: US Department of Agriculture.

[b]FNDSS: Food and Nutrient Database for Dietary Studies.

Food Items and Code Selection

Figure 1 shows the flowchart of the food item and food code selection process. Briefly, data from our Formosa FoodApp validation study were used for the analyses [14], which involved 86 healthy adults aged 19-26 years who tracked their daily diet using the Formosa FoodApp. To identify food items that contributed to high cholesterol and saturated fat intake, a nutrition-trained investigator ranked frequently consumed, saturated fat– and cholesterol-containing items from our prior study. The initial assessment generated 2 lists of 50 food items each. After eliminating duplicates and similar entries, we finalized a list comprising 42 unique food items.

Figure 1. Flowchart of the selection process for food items and food codes. Food items included are those that contribute the most to cholesterol and saturated fat intake levels, ranked based on findings from our previous Formosa FoodApp validation study [14].



To mitigate potential biases from initially using data from the Formosa FoodApp, which may predominantly reflect Taiwanese dietary preferences, we verified the availability of these 42 food items across all apps selected for this study. The comprehensive list of these food items is provided in Multimedia Appendix 1 for reference. We further grouped these 42 food items into broader categories—such as eggs, beef, pork, chicken, dairy, seafood, and prepackaged and processed foods—to enhance the representativeness of our analysis across diverse national diets.

Food items were entered using the search feature for the free versions of Formosa FoodApp, COFIT, MyFitnessPal, and Lose It!. Searches were performed across Formosa FoodApp and COFIT in Chinese and LoseIt! in English. To explore MyFitnessPal's reliability across different national contexts,

searches were conducted in both Chinese and English, and we compared the errors in nutrient data from its English and Chinese versions. To ensure a clear distinction between MyFitnessPal in Chinese and English, we refer to them as MyFitnessPal-Chinese and MyFitnessPal-English, respectively, throughout the manuscript. Given the vast array of food codes that appeared in the search results and to make the analysis more manageable and representative, we selected the top 5 relevant food codes that appeared first in the search results for each item as representative subsets. Energy contents and nutrients including total carbohydrate, protein, total fat, saturated fat, and cholesterol values were extracted from the apps (Figure 1). The selection process underwent a review and confirmation by a second investigator. Portion sizes for the selected food codes were standardized at 100 g across all databases. Data from all selected apps and reference databases were documented for each food code and entered into a Microsoft Excel dataset.

Data and Statistical Analysis

We calculated the mean percentage error for energy, total carbohydrate, protein, total fat, saturated fat, and cholesterol values as the difference between the nutrient intake values from the apps and the corresponding values from the reference databases, expressed as a percentage error. This error, calculated for each food code in each app, is given by the following:

$$\text{Percentage error (\%)} = \frac{\text{App nutrient value} - \text{Reference nutrient value}}{\text{Reference nutrient value}} \times 100$$

We focused on 2 key components of variability for saturated fats and cholesterol: missing nutrient data and inconsistencies of nutrient data for the same food item across 5 codes. Missing nutrient information for each food item was determined as the ratio of missing nutrient information to the total selected food codes for that item, presented as a percentage:

$$\text{Percentage of missing data (\%)} = \frac{\text{Number of missing nutrient information in the selected food codes}}{\text{Total number of selected food codes}} \times 100$$

We calculated coefficients of variation (CVs) to address nutrient data inconsistencies across 5 food codes for the same item within an app. CVs were obtained by taking the square root of the average reporting variance within a food item, divided by the mean of the items (app values over food codes). A CV was calculated for each food item, and then mean CVs were determined based on the food group and app. Our analysis involved paired 2-tailed *t* tests for app-to-reference comparisons (paired within food codes) and a 1-way ANOVA for app-to-app variations. A linear regression analysis was used to assess the factor associated with the percentage of underestimated error of nutrients compared to reference databases. To assess the reliability of a commercial app universally used in different national contexts, we conducted a cross-national analysis, comparing errors in both MyFitnessPal-Chinese and MyFitnessPal-English against the USDA-FNDDS and Taiwan FCD reference databases. All analyses were performed in SPSS (version 23; IBM Corp) and GraphPad Prism 8 (GraphPad Software), with *P* values of <.05 considered statistically significant.

Ethical Considerations

This study conducted a secondary analysis of data previously collected in an earlier study [14], which received approval from the Taipei Medical University Institutional Review Board (N202101046). Written informed consent was obtained from all participants. The data used in this analysis were deidentified.

Results

Characteristics of Mobile Nutrition Apps

Among the selected nutrition apps, Formosa FoodApp is an academic mobile nutrition app, while COFIT, MyFitnessPal-Chinese, MyFitnessPal-English, and LoseIt! are all commercial apps (Table 1). Figure 1 details the distribution of 42 food items across 7 main food categories. The dominant food groups were prepackaged processed foods (15 items), seafood (6 items), and pork and egg (5 items each).

Reliability of Mobile Nutrition Apps Across Different National Standards

Figure 2A presents the mean percentage errors of app nutrient data against national reference databases. Most apps showed no differences in energy or macronutrients compared to their respective national references, with the exception of LoseIt!, which underestimated protein by −1.3% and fats by −22.4% compared to the USDA-FDNSS. COFIT, MyFitnessPal (in both languages), and LoseIt! underestimated saturated fats with errors ranging from −16.3% to −40.3% (all $P<.05$). All apps underestimated cholesterol, with errors ranging from −26.3% to −60.3% (all $P<.05$).

Figure 2B displays the mean percentage errors in MyFitnessPal-Chinese and MyFitnessPal-English using the USDA-FNDSS and Taiwan FCD. The Chinese version underestimated saturated fats by around −40% (−43.2% for the Taiwan FCD and −39.4% for the USDA-FNDSS) and cholesterol by −60% (−58.2% for the USDA-FNDSS and −60.3% for the Taiwan FCD). In contrast, the English version showed −10% errors for saturated fat (−16.8% for the USDA-FNDSS and −4.3% for the Taiwan FCD) and −40% for cholesterol (−40.4% for the USDA-FNDSS and −38.8% for the Taiwan FCD). Notably, these discrepancies, significant at $P<.01$ for both nutrients, remained unchanged regardless of the reference database used.

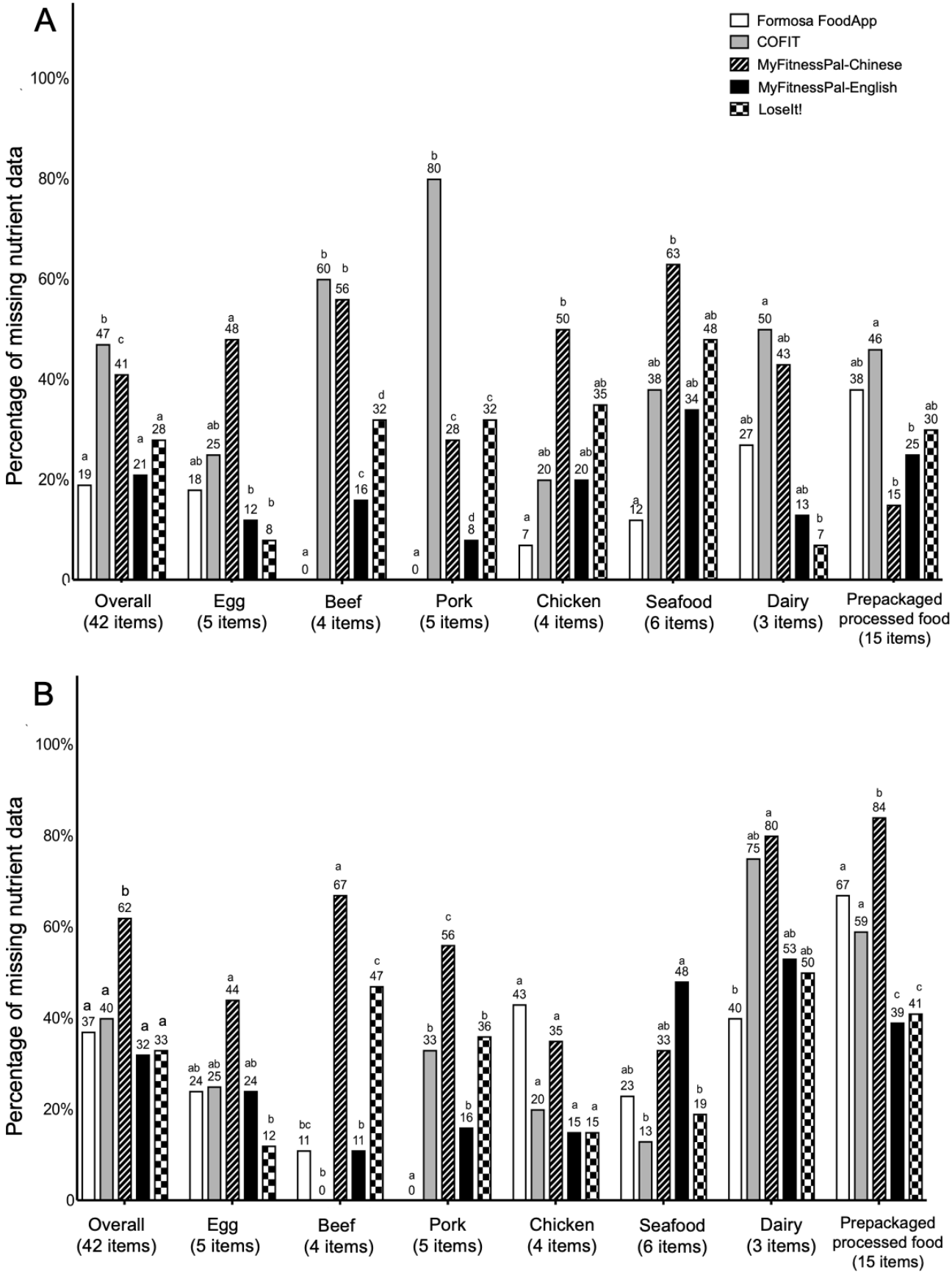
Figure 2. (A) Heat map of mean percentage errors of nutrients across mobile apps compared to reference databases. Different letters (a, b, ab, c, bc, and abc) indicate statistically significant differences in mean percentage errors among groups based on 1-way ANOVA ($P<.05$). a: statistically different from b, bc, and c; b: statistically different from a, ab, and c; ab: not significantly different from a or b; c: statistically different from a, ab, and b; bc: not significantly different from b or c, but different from a; abc: not significantly different from a, b, or c. Asterisks (*) denote the level of statistical significance in paired 2-tailed *t* tests between the app and the reference database: *$P<.05$, **$P<.01$, ***$P<.001$, and ****$P<.0001$. (B) Cross-national comparison of percentage errors in saturated fats and cholesterol between the English and Chinese platforms of MyFitnessPal. **$P<.01$ by a 1-way ANOVA indicating differences in percentage errors among groups. FCD: Food Composition Database; FNDDS: Food and Nutrient Database for Dietary Studies; USDA: US Department of Agriculture.

Omissions of Saturated Fats and Cholesterol

Omission rates for different nutrients considerably varied. Total fats exhibited the lowest omission range, spanning from 0% to 21%. In comparison, saturated fats and cholesterol showed higher rates of omission, ranging from 19% to 47% and from 21% to 62%, respectively (data not shown). COFIT showed the highest missing information for saturated fats at 47% (Figure 3A), while MyFitnessPal-Chinese recorded the highest omission at 62% for cholesterol (Figure 3B).

For saturated fats, both COFIT and MyFitnessPal-Chinese showed substantial omissions, especially for beef (60% and 56%, respectively), pork (80% and 28%, respectively), and dairy (50% and 43%, respectively). For cholesterol, there were significant differences in omission rates between MyFitnessPal-Chinese and MyFitnessPal-English (62% vs 32%; $P<.05$), notably for beef (67% vs 11%), pork (56% vs 16%), dairy (80% vs 53%), and prepackaged processed foods (84% vs 39%).

Figure 3. Omission rates (percentage of missing nutrient information) for (A) saturated fats and (B) cholesterol across the mobile nutrition apps. Different letters (a, b, c, d, ab, and bc) indicate a significant *P* value ($P<.05$) for Fisher test comparing differences in omission rates between the apps among the mobile app food databases. a: statistically different from b, bc, c, and d; b: statistically different from a, ab, c, and d; ab: not significantly different from a or b; c: statistically different from a, ab, b, and d; d: statistically different from a, b, and c; bc: not significantly different from b or c, but different from a and d.

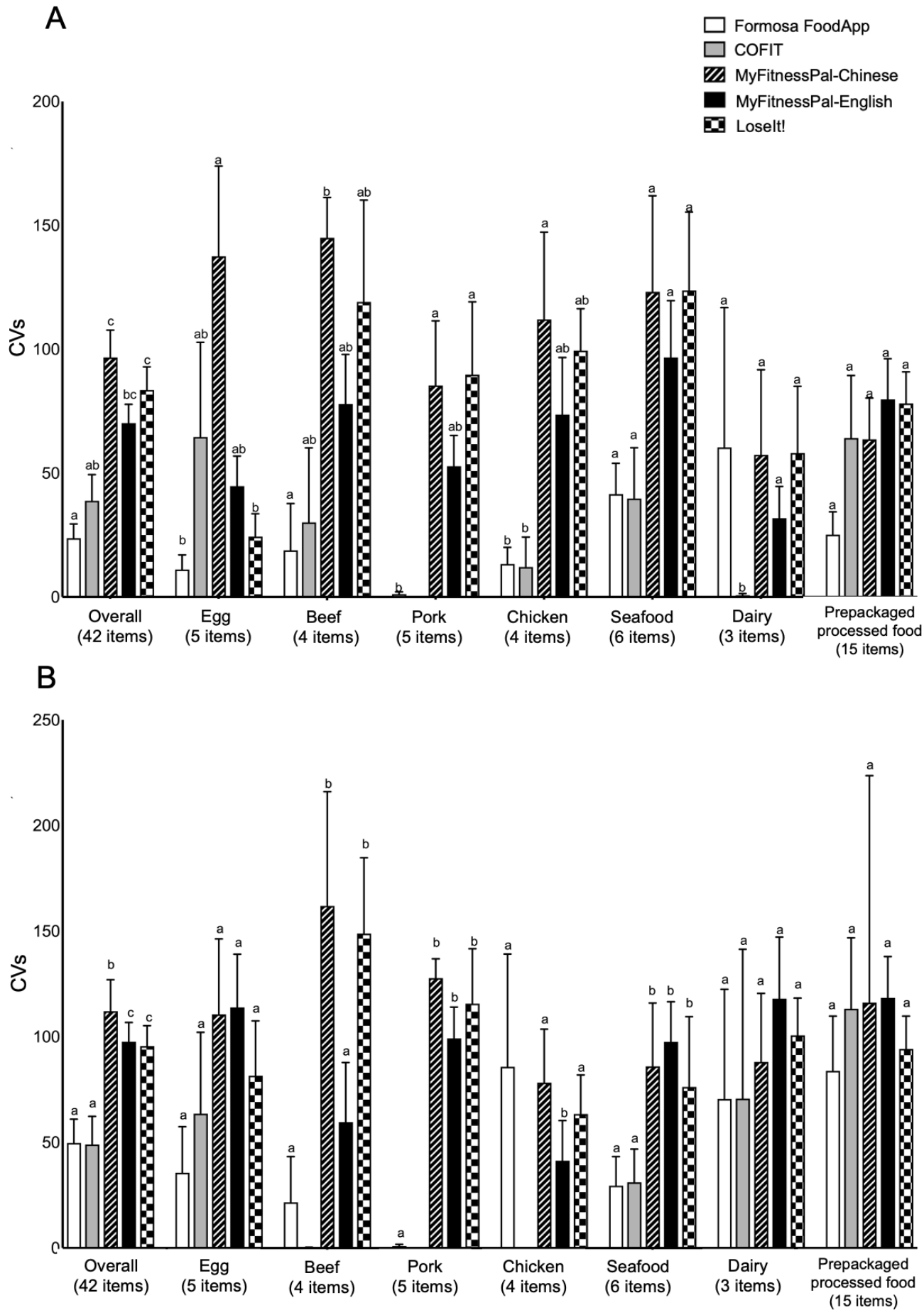


Variability Extents of Saturated Fat and Cholesterol

We calculated CVs to address saturated fat and cholesterol data consistencies, by assessing 5 food codes for identical items within a given app (Figure 4). For saturated fats, the lowest mean CVs were observed in Formosa FoodApp (23.8%) and COFIT (38.9%). The highest mean CVs were observed in MyFitnessPal-Chinese (96.7%) and LoseIt! (83.6%). A similar trend was also observed for cholesterol.

Saturated fat variability was high for beef (78%-145%), chicken (74%-112%), and seafood (97%-124%) among MyFitnessPal-Chinese, MyFitnessPal-English, and LoseIt!, respectively. Cholesterol variability was high for dairy (71%-118%) and prepackaged foods across all apps (84%-118%; Figure 4).

Figure 4. Mean coefficient of variations (CVs) for (A) saturated fats and (B) cholesterol, calculated as the percentage ratio of the SD to the mean, for 42 food items grouped by food groups across the apps. Different letters (a, b, ab, and c) indicate a significant *P* value (*P*<.05) for 1-way ANOVA comparing differences in the CVs among the apps within food groups. a: statistically different from b and c; b: statistically different from a, ab, and c; ab: not significantly different from a or b; c: statistically different from a, ab, and b.



Discussion

Principal Findings

To the best of our knowledge, this is the first study to critically examine the reliability of mobile nutrition apps' reported data on saturated fats and cholesterol against 2 different national FCDs. We found significant errors up to −40.3% for saturated fats and −60.3% for cholesterol, suggesting that these apps may provide unreliable data when assessing diets, particularly in the context of CVD prevention. These findings align with the findings by Shinozaki and Murakami [25] on MyFitnessPal's nutrient underestimation of saturated fats and cholesterol (85.1% and 97.6%, respectively). Similarly, Siniarski et al [16] also reported errors up to 57.3% in mobile app predictions for saturated fats compared to Polish reference data. Furthermore, our results demonstrate that errors in MyFitnessPal's nutrient reporting are consistent across different national FCDs (USDA-FNDDS and Taiwan FCD). Such errors remained statistically significant, regardless of the FCDs used as a reference. This problem suggests that the errors are not merely due to mismatches or variations in external FCDs but are rooted in the app's core database.

Data inconsistencies within consumer-grade apps, evidenced by significant omissions and variability, compromise their reliability. For instance, saturated fat variability was notably high across different food groups, with beef ranging from 78% to 145%, chicken from 74% to 112%, and seafood from 97% to 124% among MyFitnessPal-Chinese, MyFitnessPal-English, and LoseIt!, respectively. Similarly, cholesterol variability was consistently high in dairy (71%-118%) and prepackaged foods (84%-118%) across all apps. COFIT lacked almost 50% of its saturated fat data, and MyFitnessPal-Chinese was missing over 60% of its cholesterol information. This is concerning as food labeling regulations require clear disclosure of these nutrients [6], yet many entries incorrectly list these values as zero, deviating from standard guidelines [26]. This issue is even more pressing in the context of prepackaged foods. Previous research from our team emphasized this challenge, revealing that data omissions related to prepackaged foods significantly impact the reliability of mobile dietary assessments [14]. This emphasizes the urgency for databases to stay updated with new market introductions.

The lack of transparency in food sourcing within consumer-grade apps, particularly when combined with the extensive volume of FCDs [11,27], can significantly confuse users during food logging [28,29]. As consumer-grade apps allow for a vast array of food items to be listed—often without clear sourcing information, users face the daunting task of navigating through numerous results to find the most accurate match for their consumed foods. The entries frequently do not specify their nutritional data sources. Our data showed that only a fraction of the analyzed food codes transparently disclosed their sources. Specifically, just 51.5% (n=105) of MyFitnessPal-English's 204 food codes and 43.7% (n=87) of its Chinese counterpart's 199 food codes provided clear sourcing information. Meanwhile, LoseIt! transparently sourced only 37.4% (n=71) of its 190 codes. Formosa FoodApp led with 78.2% (n=115) for its 147 codes, while COFIT lagged at 61% (n=57) for its 94 codes. This lack of clarity can lead to confusion and inaccuracies, underscoring the need for improved transparency in food sourcing to facilitate more precise dietary tracking.

Our analysis also delved into the complexities of maintaining accuracy within multilingual universal apps, using MyFitnessPal as a case study. Although errors in MyFitnessPal's nutrient reporting are consistent across different national FCDs (USDA-FNDDS and Taiwan FCD), it is important to note that the larger errors observed in MyFitnessPal-Chinese, as opposed to the English version, highlight another challenge for universal apps: ensuring uniform accuracy across different languages and, by extension, cultural contexts. Several factors may contribute to the discrepancies observed in MyFitnessPal. First, the integration of multiple FCDs combined with user-generated content without stringent quality control can introduce variability in nutrient data [27]. Second, nutritional composition can also vary by region due to differences in food production, processing, and preparation practices, further complicating accurate nutrient estimation [30]. Lastly, user-generated content can exacerbate inaccuracies, as this content may not undergo rigorous verification processes, leading to inconsistencies in nutrient reporting [27]. Together, these factors highlight the complexities of providing accurate, universal dietary tracking tools and the critical need for app developers to address these challenges to enhance the reliability and global applicability of their platforms.

This study has highlighted significant errors and inconsistencies in various consumer-grade apps. Academic apps such as Formosa FoodApp, which show fewer reporting errors and greater transparency in nutrient sourcing, are crucial for health professionals who require precise and validated data to make evidence-based dietary recommendations, particularly for CVD prevention. However, these apps encounter financial and regional limitations that restrict their usability. As academic apps are usually funded by research grants or academic institutions for specific purposes, their development and maintenance are often constrained, limiting updates and expansion. Moreover, many academic apps are tailored for specific regional diets, such as the Taiwanese focus of Formosa FoodApp, which narrows their international relevance. Additionally, these apps often lack engaging user interfaces and features that are essential for widespread user adoption [14]. Innovative funding models supporting extensive food databases and multilingual capabilities are needed to make these tools globally accessible and competitive with commercial apps. Khazen et al [11] suggested that an ideal dietary tracking tool would combine the features of commercial apps with the accuracy of academic apps, increasing their utility in global research and public health initiatives.

Our study contributed to the current literature by critically examining how accurately mobile apps can estimate the intake of saturated fats and cholesterol in different national contexts. Additionally, by investigating the consistency of data within and across the apps, we have highlighted several challenges these apps face in providing reliable nutritional information in the context of CVD prevention. Nonetheless, our methodology has inherent limitations. The data from a young demographic

may introduce bias, potentially limiting the applicability of findings across all age groups. By focusing on only the top 5 food codes per item, we might not have captured the full extent of the data variability within each app. This strategy, although deliberate, possibly overlooked broader inaccuracies or consistencies in less frequently selected food items. Our cross-national examination predominantly relied on MyFitnessPal's performance, potentially narrowing the applicability of our findings to other multilingual apps. Furthermore, the discrepancy in the number of selected food codes across apps—with MyFitnessPal at 204 and COFIT at 94, for instance—introduces bias when comparing results across platforms. Additionally, we did not factor in user behaviors such as app update frequencies, which can intermittently influence data accuracy. The evolving nature of these apps, with ongoing food item additions and updates by developers and users alike, suggests that database accuracy might shift over time. As the landscape of mobile health apps is dynamic, future studies would benefit from broader sampling and recognizing of these continual database evolutions, ensuring a more encompassing perspective on app reliability.

Conclusions

Our study reveals substantial inaccuracies and inconsistencies in mobile nutrition apps' reporting of saturated fats and cholesterol, highlighting challenges in ensuring data reliability across different national contexts and within the apps themselves. As digital tools are increasingly incorporated in CVD prevention and care, rigorous assessments and continuous refinements are crucial to ensure that they serve as dependable resources for both professionals and the public.

Acknowledgments

The authors would like to thank Trung Vinh Ly and Chi Cuong Hoang for developing the Formosa FoodApp. Finally, we thank all students and research staff at the School of Nutrition and Health Sciences, Taipei Medical University, for supporting our undertaking of this work. All persons named in the *Acknowledgments* section gave permission to be named. JSC was supported by grants from the National Science and Technology Council, Taiwan (MOST 111-2410-H-038-019-MY2 and MOST 111-2320-B-038-030-MY3).

Data Availability

The datasets generated during and/or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

DKNH and JSC designed the study. JWK, HTT, CYL, PHH, YRF, KHC, TYS, CHY, and CYY contributed to development of the Formosa FoodApp. WCC, HYS, PHW, JSC, and DKNH acquired or interpreted the data. DKNH conducted the statistical analysis. JSC and DKNH drafted the manuscript. JSC supervised the study and obtained funding. All authors read and agreed to the published version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Description of mobile app databases and analyzed food items in the study.
[DOCX File, 18 KB - mhealth_v12i1e54509_app1.docx]

References

1. Roth GA, Mensah GA, Johnson CO, et al. Global burden of cardiovascular diseases and risk factors, 1990-2019: update from the GBD 2019 study. J Am Coll Cardiol 2020 Dec 22;76(25):2982-3021. [doi: 10.1016/j.jacc.2020.11.010] [Medline: 33309175]
2. Zhong VW, van Horn L, Cornelis MC, et al. Associations of dietary cholesterol or egg consumption with incident cardiovascular disease and mortality. JAMA 2019 Mar 19;321(11):1081-1095. [doi: 10.1001/jama.2019.1572] [Medline: 30874756]
3. Kim Y, Je Y, Giovannucci EL. Association between dietary fat intake and mortality from all-causes, cardiovascular disease, and cancer: a systematic review and meta-analysis of prospective cohort studies. Clin Nutr 2021 Mar;40(3):1060-1070. [doi: 10.1016/j.clnu.2020.07.007] [Medline: 32723506]
4. Zhao B, Gan L, Graubard BI, Männistö S, Albanes D, Huang J. Associations of dietary cholesterol, serum cholesterol, and egg consumption with overall and cause-specific mortality: systematic review and updated meta-analysis. Circulation 2022 May 17;145(20):1506-1520. [doi: 10.1161/CIRCULATIONAHA.121.057642] [Medline: 35360933]
5. Sacks FM, Lichtenstein AH, Wu JHY, et al. Dietary fats and cardiovascular disease: a presidential advisory from the American Heart Association. Circulation 2017 Jul 18;136(3):e1-e23. [doi: 10.1161/CIR.0000000000000510] [Medline: 28620111]

6. Gooding HC, Ning H, Gillman MW, et al. Application of a lifestyle-based tool to estimate premature cardiovascular disease events in young adults: the Coronary Artery Risk Development in Young Adults (CARDIA) study. JAMA Intern Med 2017 Sep 1;177(9):1354-1360. [doi: 10.1001/jamainternmed.2017.2922] [Medline: 28715555]
7. Gillman MW. Primordial prevention of cardiovascular disease. Circulation 2015 Feb 17;131(7):599-601. [doi: 10.1161/CIRCULATIONAHA.115.014849] [Medline: 25605661]
8. Jospe MR, Fairbairn KA, Green P, Perry TL. Diet app use by sports dietitians: a survey in five countries. JMIR Mhealth Uhealth 2015 Jan 22;3(1):e7. [doi: 10.2196/mhealth.3345] [Medline: 25616274]
9. Ambrosini GL, Hurworth M, Giglia R, Trapp G, Strauss P. Feasibility of a commercial smartphone application for dietary assessment in epidemiological research and comparison with 24-h dietary recalls. Nutr J 2018 Jan 9;17(1):5. [doi: 10.1186/s12937-018-0315-4] [Medline: 29316930]
10. Lin AW, Morgan N, Ward D, et al. Comparative validity of mostly unprocessed and minimally processed food items differs among popular commercial nutrition apps compared with a research food database. J Acad Nutr Diet 2022 Apr;122(4):825-832.e1. [doi: 10.1016/j.jand.2021.10.015] [Medline: 34662722]
11. Khazen W, Jeanne JF, Demaretz L, Schäfer F, Fagherazzi G. Rethinking the use of mobile apps for dietary assessment in medical research. J Med Internet Res 2020 Jun 18;22(6):e15619. [doi: 10.2196/15619] [Medline: 32554383]
12. Rangan AM, O'Connor S, Giannelli V, et al. Electronic Dietary Intake Assessment (e-DIA): comparison of a mobile phone digital entry app for dietary data collection with 24-hour dietary recalls. JMIR Mhealth Uhealth 2015 Oct 27;3(4):e98. [doi: 10.2196/mhealth.4613] [Medline: 26508282]
13. Ji Y, Plourde H, Bouzo V, Kilgour RD, Cohen TR. Validity and usability of a smartphone image-based dietary assessment app compared to 3-day food diaries in assessing dietary intake among Canadian adults: randomized controlled trial. JMIR Mhealth Uhealth 2020 Sep 9;8(9):e16953. [doi: 10.2196/16953] [Medline: 32902389]
14. Ho DKN, Chiu WC, Kao JW, et al. Mitigating errors in mobile-based dietary assessments: effects of a data modification process on the validity of an image-assisted food and nutrition app. Nutrition 2023 Dec;116:112212. [doi: 10.1016/j.nut.2023.112212] [Medline: 37776838]
15. Ferrara G, Kim J, Lin S, Hua J, Seto E. A focused review of smartphone diet-tracking apps: usability, functionality, coherence with behavior change theory, and comparative validity of nutrient intake and energy estimates. JMIR Mhealth Uhealth 2019 May 17;7(5):e9232. [doi: 10.2196/mhealth.9232] [Medline: 31102369]
16. Siniarski A, Sobieraj P, Samel-Kowalik P, Sińska B, Milewska M, Bzikowska-Jura A. Nutrition-related mobile applications - should they be used for dietary prevention and treatment of cardiovascular diseases? Nutr Metab Cardiovasc Dis 2022 Nov;32(11):2505-2514. [doi: 10.1016/j.numecd.2022.07.010] [Medline: 36064687]
17. Eldridge AL, Piernas C, Illner AK, et al. Evaluation of new technology-based tools for dietary intake assessment-an ILSI Europe Dietary Intake and Exposure Task Force evaluation. Nutrients 2018 Dec 28;11(1):55. [doi: 10.3390/nu11010055] [Medline: 30597864]
18. Maringer M, Wisse-Voorwinden N, Veer PV, Geelen A. Food identification by barcode scanning in the Netherlands: a quality assessment of labelled food product databases underlying popular nutrition applications. Pub Health Nutr 2019 May;22(7):1215-1222. [doi: 10.1017/S136898001800157X] [Medline: 29962361]
19. Griffiths C, Harnack L, Pereira MA. Assessment of the accuracy of nutrient calculations of five popular nutrition tracking applications. Public Health Nutr 2018 Jun;21(8):1495-1502. [doi: 10.1017/S1368980018000393] [Medline: 29534771]
20. Moyen A, Rappaport AI, Fleurent-Grégoire C, Tessier AJ, Brazeau AS, Chevalier S. Relative validation of an artificial intelligence-enhanced, image-assisted mobile app for dietary assessment in adults: randomized crossover study. J Med Internet Res 2022 Nov 21;24(11):e40449. [doi: 10.2196/40449] [Medline: 36409539]
21. Carter MC, Burley VJ, Nykjaer C, Cade JE. 'My Meal Mate' (MMM): validation of the diet measures captured on a smartphone application to facilitate weight loss. Br J Nutr 2013 Feb 14;109(3):539-546. [doi: 10.1017/S0007114512001353] [Medline: 22717334]
22. Fallaize R, Zenun Franco R, Pasang J, Hwang F, Lovegrove JA. Popular nutrition-related mobile apps: an agreement assessment against a UK reference method. JMIR Mhealth Uhealth 2019 Feb 20;7(2):e9838. [doi: 10.2196/mhealth.9838] [Medline: 30785409]
23. Bzikowska-Jura A, Sobieraj P, Raciborski F. Low comparability of nutrition-related mobile apps against the Polish reference method-a validity study. Nutrients 2021 Aug 20;13(8):2868. [doi: 10.3390/nu13082868] [Medline: 34445026]
24. Wang JS, Hsieh RH, Tung YT, Chen YH, Yang C, Chen YC. Evaluation of a technological image-based dietary assessment tool for children during pubertal growth: a pilot study. Nutrients 2019 Oct 20;11(10):2527. [doi: 10.3390/nu11102527] [Medline: 31635141]
25. Shinozaki N, Murakami K. Evaluation of the ability of diet-tracking mobile applications to estimate energy and nutrient intake in Japan. Nutrients 2020 Oct 29;12(11):3327. [doi: 10.3390/nu12113327] [Medline: 33138088]
26. Westrich BJ, Buzzard IM, Gatewood LC, McGovern PG. Accuracy and efficiency of estimating nutrient values in commercial food products using mathematical optimization. J Food Compos Anal 1994 Dec;7(4):223-239. [doi: 10.1006/jfca.1994.1026]
27. Chen J, Berkman W, Bardouh M, Ng CYK, Allman-Farinelli M. The use of a food logging app in the naturalistic setting fails to provide accurate measurements of nutrients and poses usability challenges. Nutrition 2019 Jan;57:208-216. [doi: 10.1016/j.nut.2018.05.003] [Medline: 30184514]

28. Lieffers JRL, Arocha JF, Grindrod K, Hanning RM. Experiences and perceptions of adults accessing publicly available nutrition behavior-change mobile apps for weight management. J Acad Nutr Diet 2018 Feb;118(2):229-239.e3. [doi: 10.1016/j.jand.2017.04.015] [Medline: 28625662]
29. Cordeiro F, Epstein DA, Thomaz E, et al. Barriers and negative nudges: exploring challenges in food journaling. Proc SIGCHI Conf Hum Factor Comput Syst 2015 Apr;2015:1159-1162. [doi: 10.1145/2702123.2702155] [Medline: 26894233]
30. Ross J, Gao J. Overcoming the language barrier in mobile user interface design: a case study on a mobile health app. arXiv. Preprint posted online on May 16, 2016. [doi: 10.48550/arXiv.1605.04693]

Abbreviations

CV: coefficient of variation
CVD: cardiovascular disease
FCD: food composition database
FNDDS: Food and Nutrient Database for Dietary Studies
USDA: US Department of Agriculture

Edited by L Buis; submitted 12.11.23; peer-reviewed by YC Chen, N Skinner, R Ciptaningtyas; revised version received 14.05.24; accepted 24.06.24; published 04.09.24.

Please cite as:
Ho DKN, Chiu WC, Kao JW, Tseng HT, Lin CY, Huang PH, Fang YR, Chen KH, Su TY, Yang CH, Yao CY, Su HY, Wei PH, Chang JS
Reliability Issues of Mobile Nutrition Apps for Cardiovascular Disease Prevention: Comparative Study
JMIR Mhealth Uhealth 2024;12:e54509
URL: https://mhealth.jmir.org/2024/1/e54509
doi:10.2196/54509

“Internet+Nursing Service” Mobile Apps in China App Stores: Functionality and Quality Assessment Study

Shuo Yuan[1,2*], MSN; Min Liu[1*], MSN; Yuqi Peng[1], MSN; Jinrui Hu[1], MSN; Bingyan Li[1], MSN; Xia Ding[3], BPS; Lunfang Xie[1], PhD

[1]School of Nursing, Anhui Medical University, Hefei, China

[2]Department of Cardiology II, Anhui No.2 Provincial People’s Hospital, Hefei, China

[3]Nursing Department, Anhui No.2 Provincial People’s Hospital, Hefei, China

[*]these authors contributed equally

Corresponding Author:
Lunfang Xie, PhD
School of Nursing, Anhui Medical University, , Hefei, , China

Abstract

Background: As the Chinese society ages and the concern for health and quality of life grows, the demand for care services in China is increasing. The widespread use of internet technology has greatly improved the convenience and efficiency of web-based services. As a result, the Chinese government has been implementing “Internet+Nursing Services” since 2019, with mobile apps being the primary tools for users to access these services. The quality of these apps is closely related to user experience and the smooth use of services.

Objective: This study aims to evaluate the functionality, services, and quality of “Internet+Nursing Service” apps; identify weaknesses; and provide suggestions for improving service programs and the research, development, improvement, and maintenance of similar apps.

Methods: In December 2022, two researchers searched for “Internet+Nursing Service” apps by applying the search criteria on the Kuchuan mobile app monitoring platform. After identifying the apps to be included based on ranking criteria, they collected information such as the app developer, app size, version number, number of downloads, user ratings, and number and names of services. Afterward, 5 trained researchers independently evaluated the quality of the apps by using the Chinese version of the user version of the Mobile App Rating Scale (uMARS-C). The total uMARS-C score was based on the average of the five evaluators’ ratings.

Results: A total of 17 “Internet+Nursing Service” apps were included. Among these, 12 (71%) had been downloaded more than 10,000 times, 11 (65%) had user ratings of 4 or higher, the median app size was 62.67 (range 22.71 - 103; IQR 37.51-73.47) MB, 16 (94%) apps provided surgical wound dressing change services, 4 (24%) covered first-tier cities, and only 1 (6%) covered fourth-tier cities. The median total uMARS-C score was 3.88 (range 1.92-4.92; IQR 3.71-4.05), which did not correlate with app store user ratings (r=0.003; P=.99). The quality of most apps (11/17, 65%) was average. Most apps (12/17, 71%) were rated as “good” or above (≥4 points) in terms of information quality, layout, graphics, performance, and ease of use; however, the vast majority of apps were rated as “fair” or even “poor” (<4 points) in terms of credibility (14/17, 82%) and demand (16/17, 94%).

Conclusions: “Internet+Nursing Service” apps need to broaden their service coverage, increase service variety, and further optimize their service structure. The overall quality of these apps is generally poor. App developers should collaborate with medical professionals and communicate with target users before launching their products to ensure accurate content, complete functionality, and good operation that meets user needs.

(JMIR Mhealth Uhealth 2024;12:e52169) doi:10.2196/52169

KEYWORDS

mobile phone; mobile applications; home care services; telemedicine; nursing services; China; apps

Introduction

As IT continues to develop, the number of internet users worldwide has reached 4.95 billion in 2022, including 1.067 billion internet users in China, with an internet penetration rate of 75.6%. The proportion of mobile phone users accessing the internet has reached 99.6% [1]. Mobile apps continue to grow steadily, with people using them for communication, shopping, office work, and socializing. The field of internet health care is becoming increasingly standardized, and it was the fastest-growing field in terms of user scale in 2022. As of December 2022, the number of internet health care users in

China reached 363 million—an increase of 64.66 million when compared to December 2021—accounting for 34% of all internet users [1]. Mobile health (mHealth) apps are the main media for mHealth and the primary channels for users to obtain medical support and health information. In 2019, 65.9% of internet users chose medical and health apps for medical consultations when feeling unwell [2].

People are increasingly inclined to use health care apps to manage their health due to these apps' cost-effectiveness, convenience, and speed in accessing health information [3-5] and ability to provide evidence-based health information for better health guidance [6]. There are several popular health care apps in China. According to a report on Statista, as of December 2022, the most popular medical app in China was Ping An Good Doctor—a health care platform owned by the Ping An Insurance Group—with almost 23 million monthly active users [7]. The mobile portal provides real-time medical consultations, web-based appointment booking services, and a health-related discussion forum.

As people become more health conscious, the demand for health care services is increasing. To alleviate the pressure on offline medical institutions and meet the needs of the public, China launched "Internet+Nursing Services" [8]. These services are provided by registered nurses from fixed medical institutions and operate on a web-based app and offline service model. They are designed to serve discharged patients or special populations with medical conditions and limited mobility [9]. Since 2019, "Internet+Nursing Service" organizations in China's provinces and regions have been actively providing services to countless older adults living at home, pregnant women, infants, and young children who have difficulty with leaving their homes. These services have received unanimous positive reviews [10,11]. The "Internet+Nursing Service" model integrates nursing services with internet technology, using internet IT to break the spatial limitations of traditional medical treatment. It can meet the multilevel nursing care needs of service users, allow for personalized and continuous care, alleviate social problems that result in difficulties with visiting a physician, improve the quality of life of service users, and broaden the channels of communication between nurses and patients [12].

Patients place orders through an app, and managers dispatch orders based on various factors, such as web-based nurses' qualifications, professionalism, and distance. Platform nurses receive orders within a specified period of time and travel to patients' homes to provide services, including routine nursing operations (eg, intramuscular injections, intravenous injections, urinary catheterization, gastric tube insertion, and blood sample collection), as well as specialty care (eg, peripherally inserted central venous catheter medication exchange, wound stoma care, and neonatal examinations). By downloading high-quality "Internet+Nursing Service" mobile apps, users can quickly access the care services they need without leaving their homes. However, low-quality apps may not only affect the user experience but also make it impossible for users to obtain the care services they need.

As people become increasingly dependent on smartphones and apps, they are also becoming more concerned about the quality of apps. Users not only expect apps to function properly but also have high expectations for their aesthetics, security, and personalization settings [13]. App quality issues can affect the user experience, determine whether users continue to use the app, and even lead to economic and property losses for both the users and the app development departments. mHealth apps are a special type of app, and studies have shown that factors such as usability, navigability, accuracy of information, and security all affect the user's experience with and evaluation of an mHealth app. If an mHealth app is of low quality, users may doubt its usefulness and effectiveness, which can even lead to users obtaining low-quality health services and generating erroneous health management concepts that affect their health beliefs and behaviors [6,14,15].

In recent years, various software development organizations have been paying more attention to improving app quality, but there are still deficiencies [16]. The quality of an app is mainly judged by checking user ratings and reviews on app stores. However, the actual quality of an app can be unclear, and it is impossible to know whether an app's functions are comprehensive or whether its content is scientific based on the data displayed on app store pages. Therefore, it is necessary to objectively evaluate the quality of medical and health apps to identify shortcomings in their development; promote their continuous improvement; improve their applicability, usage experience, and compliance (eg, compliance with data protection laws); and allow users to use apps that are of high quality, are reasonably designed, and are safe to use. There are various methods for evaluating the quality of medical apps, which are also known as *mHealth apps*. According to a study by Stoyanov et al [17], the quality of mHealth apps is evaluated based on different categories, including engagement, functionality, aesthetics, information quality, and subjective quality [18]. There are various methods for evaluating mHealth services, such as the use of questionnaires, the conduction of interviews, and observation [18]. A systematic review by Nouri et al [19] identified the following seven main classes of assessment criteria for mHealth apps: design, information and content, usability, functionality, ethical issues, security and privacy, and user-perceived value. These criteria can be used to assess the quality of a medical app.

Most studies on "Internet+Nursing Services" focus on service effects [20-22], the establishment of service quality index systems [23], risk management strategies [24,25], and the demand for services from various groups [26-28]. However, there is little attention paid to the quality of "Internet+Nursing Service" apps and a lack of studies that use evaluation tools to objectively evaluate these apps' quality and functions. The aim of this study was to review the "Internet+Nursing Service" mobile apps that are available on China's app stores and evaluate their quality.

Methods

Selection of the "Internet+Nursing Service" Apps

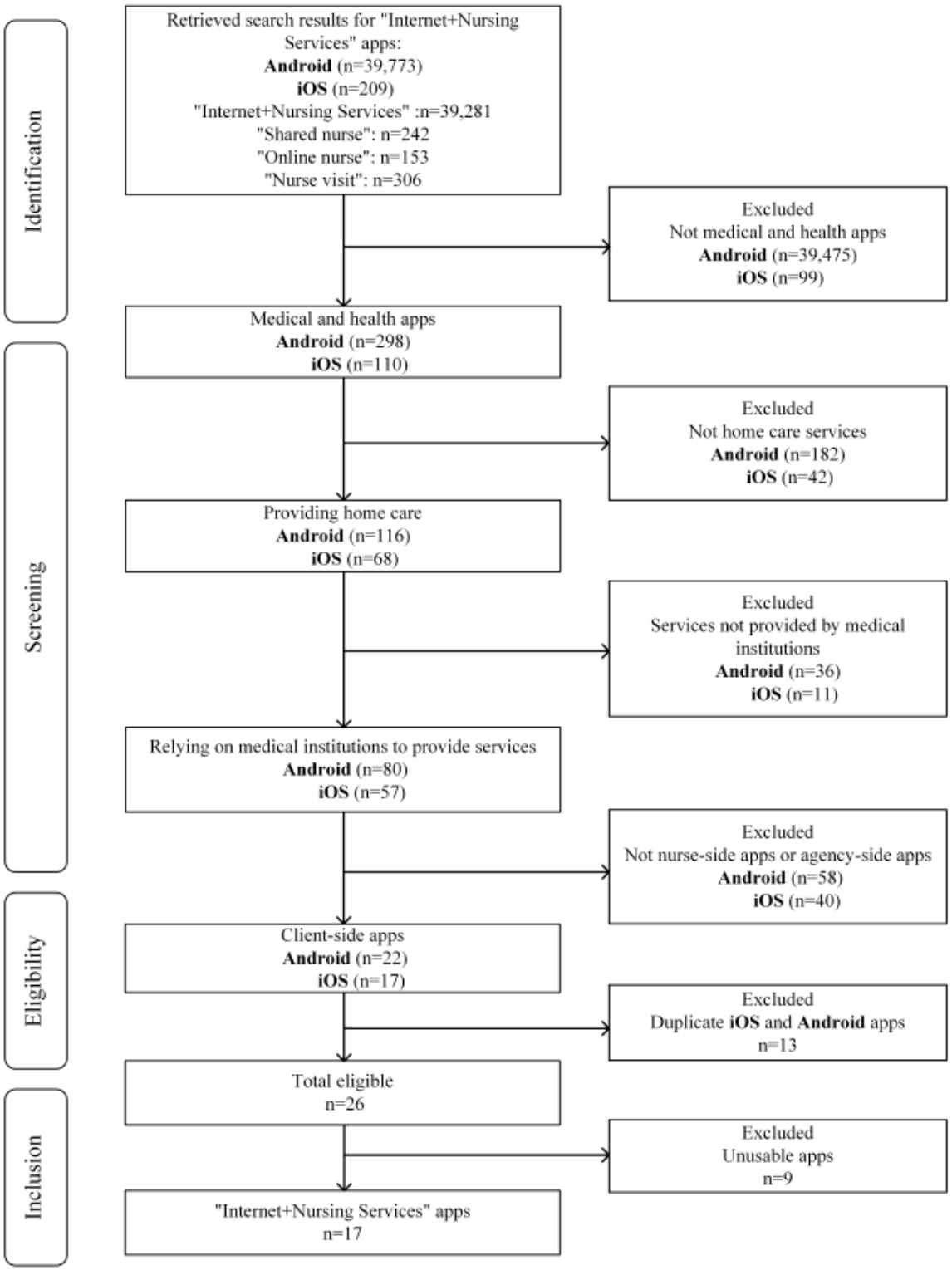
We used the Kuchuan mobile app monitoring platform (Beijing Kuchuan Technology Co.) to monitor data from the iOS and Android app stores. This platform provides real-time information

about mobile app developers, the latest versions of apps, and the number of app downloads. Two researchers searched for apps that were available as of December 1, 2022, using the keywords "Internet+Nursing Service," "Home Nursing," "Nurses at Home," "online nurse," and "shared nurse."

The inclusion criteria for the apps were (1) apps with content that includes home nursing services, (2) apps categorized as health care apps, (3) apps in Chinese, (4) free apps, and (5) functional apps. The exclusion criteria were (1) non–user-side apps, (2) old versions of the same app, (3) apps with different names but the same content, and (4) duplicate apps.

Two researchers independently screened the apps based on their names, profiles, and display images. They then discussed their findings to finalize the list of evaluated apps (Figure 1).

Figure 1. Flowchart of the selection of apps.



Selection of a Standardized Rating Scale for Mobile Apps

We used the Chinese version of the user version of the Mobile App Rating Scale (uMARS-C), and we obtained authorization from the authors of the uMARS-C [29]. Adapted from the Mobile App Rating Scale (MARS), the user version of the MARS (uMARS) has been used for assessing a wide variety of apps, including apps for mental health [30], rheumatology patient management [31], cancer risk assessment [32], and hospital registration [33]. The uMARS-C includes 14 objective items that are rated on a 5-point Likert scale and divided into the following three dimensions: engagement, functionality, and information. Dimension scores are calculated by dividing the total entry score by the number of entries, and the uMARS-C total score is calculated by dividing the total dimension score by the number of dimensions. According to the rating scale, a uMARS-C total score of 1 indicates poor quality, a score of 2 indicates inadequate quality, a score of 3 indicates fair quality, a score of 4 indicates good quality, and a score of 5 indicates excellent quality.

The uMARS-C has good reliability and validity, with a Cronbach α coefficient of 0.890 and dimension Cronbach α coefficients ranging from 0.853 to 0.895. The test-retest reliability value is 0.967, the item content validity index ranges from 0.78 to 1.00, and the scale content validity index/average is 0.969.

Process of Evaluating "Internet+Nursing Service" Apps

A total of 5 researchers—2 nurses with more than 5 years of experience in their roles, 2 graduate nursing students, and 1 professional internet engineer with more than 7 years of experience in their role—assessed the quality of the apps. Before the evaluation, we made sure that each researcher properly understood and was familiar with the uMARS-C. To ensure their understanding of the scale, they downloaded and assessed 2 mHealth apps that were not included in this study. When there was a difference of more than 2 points in dimension scores or total scores, they discussed until reaching a consensus. The researchers then downloaded the apps that were included in the final analysis onto iOS and Android smartphones. After downloading the apps, they used each app for at least 10 minutes and independently evaluated the ease of use, performance, security, and settings of each app, using the uMARS-C. Basic app information was collected from the app store download page, including the app developer, app size (in MB), version number, number of downloads, user ratings (ranging from 0 to 5), and number and names of services, among others. Some app download data were missing because the iOS app store did not provide these data. The researchers also graded the service coverage cities based on the categorized statistical service items in the *Beijing Internet Home Care Service Item Catalog* (2022 edition) [34] and the city class divisions in the *2022 China's City Business Attractiveness Ranking* [35].

Statistical Analysis

We used EpiData 3.1 (EpiData Association) for data entry and SPSS 24.0 (IBM Corp) for statistical analysis. Nonnormally distributed measurement data were expressed as medians and quartiles, while count data were expressed as numbers and percentages. The uMARS-C dimension scores for each app were averaged across the five raters, and the final scores were calculated by using the scale's formula.

Ethical Considerations

This study did not involve human subjects, clinical trials, and vulnerable groups and was therefore exempt from ethical approval.

Results

Characteristics of Selected Apps

Our search found a total of 39,982 apps (iOS: n=209; Android: n=39,773). After initial screening based on the inclusion and exclusion criteria, we downloaded 26 apps. After using them, we excluded 8 apps that were not working properly, leaving a total of 17 apps, which were included in this study (Table 1 and Multimedia Appendix 1). Of these 17 apps, 4 (24%) were released by medical institutions and 13 (76%) were released by corporations (Table 1). Further, 12 (71%) apps had been downloaded more than 10,000 times, with Champion Nurse having the highest number of downloads (n=37,321,776). The median app size was 62.67 (range 22.71-103; IQR 37.51-73.47) MB. App store user ratings ranged from 2.8 to 5.0, with 14 (82%) apps being rated 3.0 or higher and 11 (65%) apps being rated 4.0 or higher. In terms of service coverage, 4 (24%) of the "Internet+Nursing Service" apps covered first-tier cities, including Beijing; 8 (47%) covered new first-tier cities; 4 (24%) covered second-tier cities; and 1 (6%) covered fourth-tier cities.

Table . Characteristics of the "Internet+Nursing Service" apps.

Characteristic		Apps (N=17), n (%)
Download count [a]		
	0-9999	2 (12)
	10,000-99,999	7 (41)
	100,000-999,999	3 (18)
	1,000,000-9,999,999	1 (6)
	≥10,000,000	1 (6)
Platform		
	iOS and Android	13 (76)
	iOS	3 (18)
	Android	1 (6)
User rating (number of stars)		
	0-2.9	3 (18)
	3.0-3.9	3 (18)
	4.0-4.9	5 (29)
	5.0	6 (35)
Developer		
	Individual developer	8 (47)
	Corporation	9 (53)
Number of services provided		
	1-10	2 (12)
	11-20	9 (53)
	21-30	6 (35)
Service coverage city class		
	First-tier cities	4 (24)
	New first-tier cities	8 (47)
	Second-tier cities	4 (24)
	Fourth-tier cities	1 (6)

[a]Download count for Android apps only.

Categories of Nursing Services Provided by Apps

The 17 apps provided at-home nursing services, including intravenous injection, intramuscular injection, and nebulized inhalation services, among others. These services were classified as *Health Assessment and Guidance*, *Clinical Nursing*, *Maternal and Infant Nursing*, *TCM* (traditional Chinese medicine) *Nursing*, *Specialty Nursing*, *Hospice*, and *Rehabilitation Nursing* services per the categories in the *Beijing Internet Home Care Service Item Catalog* [34]. Any services that were not in the catalog were classified as *uncategorized items*.

Of the 17 apps, 16 (94%) provided surgical wound dressing change services (*Clinical Nursing* category), 15 (88%) provided services for the maintenance of peripherally inserted central catheters (*Specialty Nursing* category), 14 (82%) provided maternal and infant nursing services (*Maternal and Infant Nursing* category), and 10 (59%) provided TCM nursing services (*TCM Nursing* category) and disease rehabilitation guidance (*Rehabilitation Nursing* category). Further, 4 (24%) apps provided hospice care services (*Hospice* category), 3 (18%) provided health assessments and guidance (*Health Assessment and Guidance* category), and only 1 (6%) app provided gastrointestinal decompression, previsit physical examination, conjunctival capsule irrigation, T-tube drainage care, and family room services (uncategorized items).

Quality of the "Internet+Nursing Service" Apps

In our study, the Cronbach α coefficient of the uMARS-C was 0.871, and the dimension Cronbach α coefficients ranged from 0.761 to 0.811. Based on the uMARS-C scores, of the 17 apps, 1 (6%) was rated as "poor," 11 (65%) were rated as "fair," 5 (29%) were rated as "good," and none were rated as "insufficient" or "excellent." The median total score for the "Internet+Nursing Service" apps was 3.88 (range 1.92-4.92; IQR 3.71-4.05; Figure 2), with Champion Nurse having the

highest score (4.92) and Health WuHan having the lowest score (1.92). There was no significant correlation between app store user ratings and total uMARS-C scores (r=0.003; P=.99). The median information dimension score was 3.97 (range 1.86-5.00; IQR 3.57-4.29), the median functionality dimension score was 3.95 (range 1.25-5.00; IQR 3.25-4.50), and the median engagement dimension score was 3.80 (range 1.00-5.00; IQR 3.33-4.33; Figure 3). A heat map comparing the scores of each entry in the uMARS-C for the 17 "Internet+Nursing Service" apps showed that Health WuHan scored below 3 points in most entries (11/14, 79%; Multimedia Appendix 2). Further, 14 (82%) apps scored below 4.0 points in the *Credibility* entry, indicating an average or poor level; 16 (94%) scored below 4.0 points in the *Demand* entry; 14 (82%) scored at a good or above level in the *Quality of Information* entry; and 12 (71%) scored at a good or above level in the *Layout*, *Graphics*, *Performance*, and *Ease of Use* entries.

Figure 2. uMARS-C overall scores of the "Internet+Nursing Service" apps (N=17). The bottom and top edges of the boxes represent the first and third quartiles, respectively; the lines within the boxes represent the medians; the ends of the bottom and top whiskers represent the minimum and maximum values, respectively; and the circles represent outliers. uMARS-C: Chinese version of the user version of the Mobile App Rating Scale.

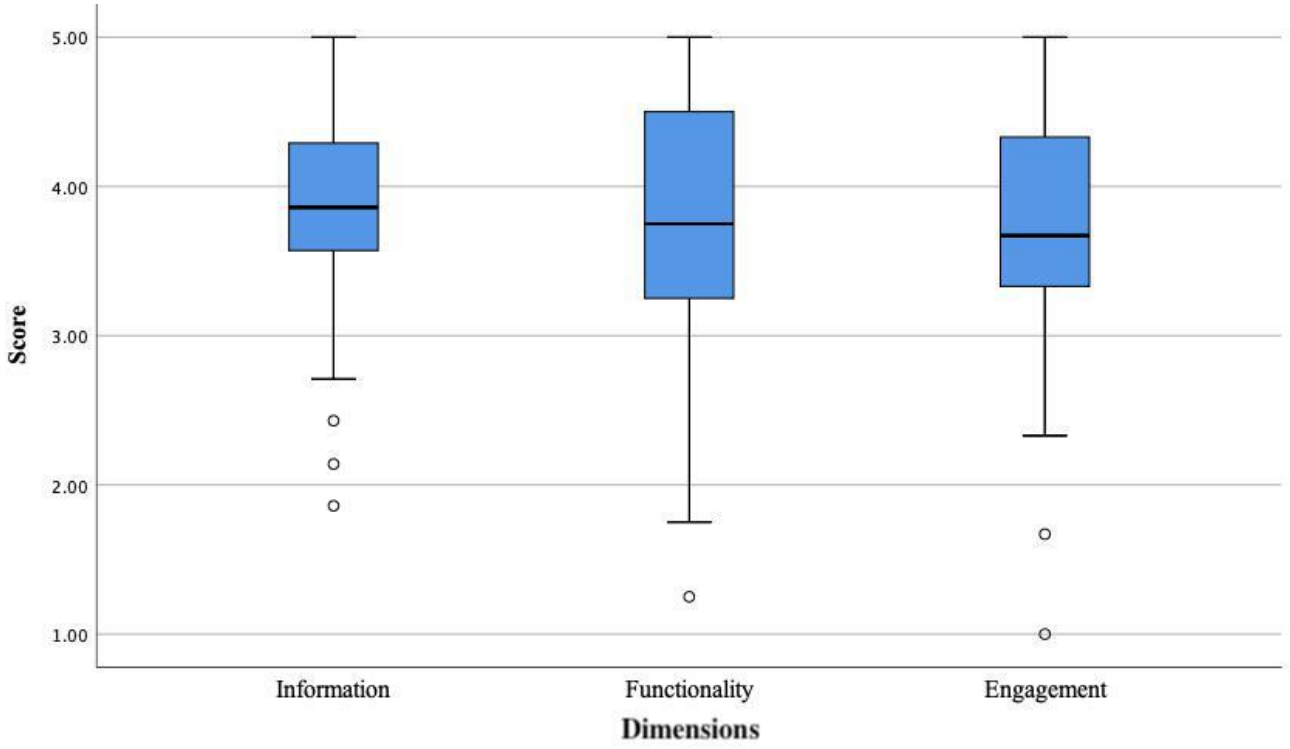
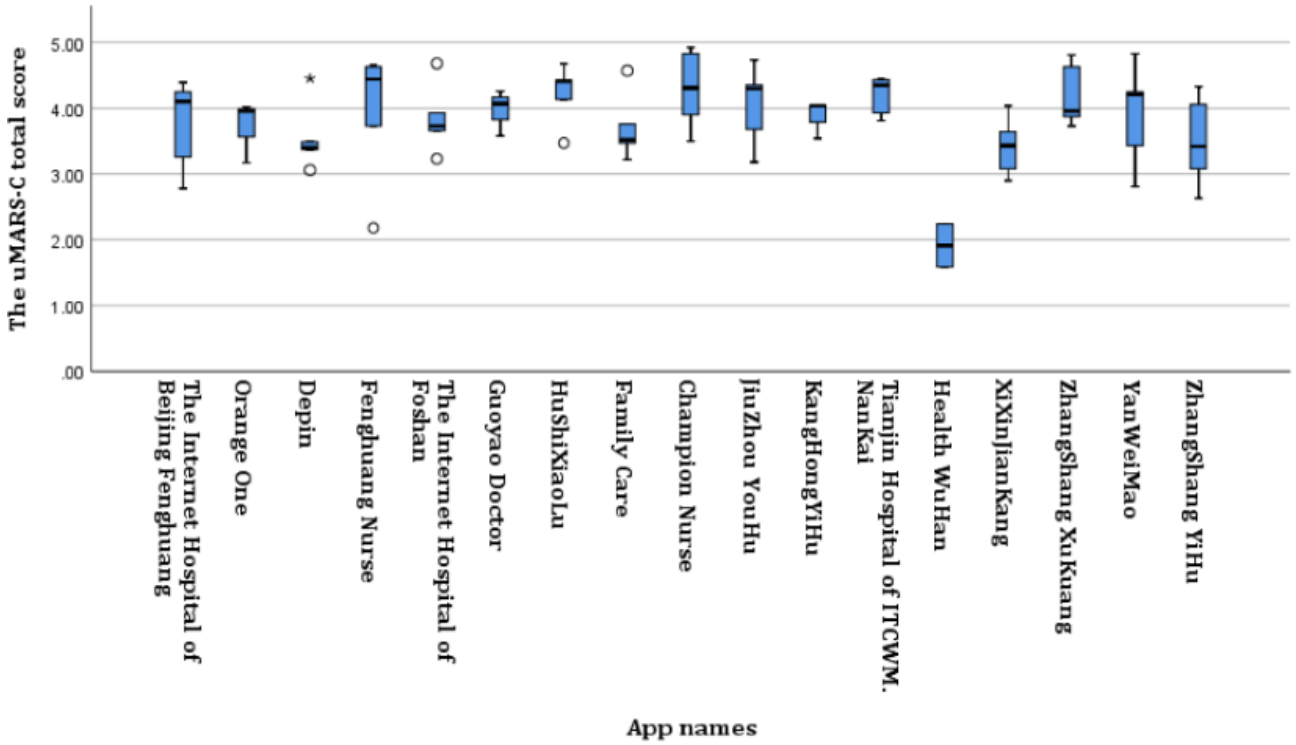


Figure 3. uMARS-C dimension scores of "Internet+Nursing Service" apps (N=17). The bottom and top edges of the boxes represent the first and third quartiles, respectively; the lines within the boxes represent the medians; the ends of the bottom and top whiskers represent the minimum and maximum values, respectively; the circles represent outliers; and the asterisk represents an extreme outlier. uMARS-C: Chinese version of the user version of the Mobile App Rating Scale.



Discussion

Principal Results

As internet technology continues to develop, all industries are integrating the "Internet+" model to promote innovation and development [1]. However, unlike services in other industries, "Internet+Nursing Services," as web-based health care services, are characterized by the high risks and high professionalism of the medical industry, as well as the special risks associated with mHealth [25].

There are several risks associated with "Internet+Nursing Service" apps. One of the risks is that technology barriers can prevent some patients from accessing telehealth services. These barriers can include a lack of access to the internet, a lack of access to the necessary devices, or difficulty with using the technology. Another risk is that there may be issues with insurance coverage for telehealth services, as well as regulatory obstacles that can limit the use of telehealth [36]. Therefore, it is necessary to strengthen the supervision of "Internet+Nursing Service" apps and strictly control all aspects. "Internet+Nursing Services" are human-centered services and aim to improve health. Therefore, "Internet+Nursing Service" app design should fully consider user characteristics and needs. In this study, we downloaded and used "Internet+Nursing Service" apps, evaluated them from a user perspective, and rated them objectively based on our usage experience.

Most apps (16/17, 94%) provided services in first-tier cities, new first-tier cities, or second-tier cities, with only 1 app providing "Internet+Nursing Services" in a fourth-tier city. The "Internet+Nursing Service" scope does not yet cover remote areas and townships. Data from China's seventh national census show that the rural population consists of about 509.79 million people, accounting for 36.11% of the total population [37]. However, 80% of medical resources are concentrated in medium- and large-sized medical institutions in medium- and large-sized cities, leading to an imbalance between the demand for care and the supply of care resources for the grassroots population [38].

Providing high-quality nursing resources and services is key to improving the health and quality of life of people at the grassroots level. Medical institutions at all levels should actively promote the distribution of medical resources to enhance access to basic medical and public health services. Therefore, "Internet+Nursing Service" apps should integrate medical resources and promote a 3-tier "hospital-community-family" linkage to bring professional nursing services into the homes of grassroots people. This would encourage more medical institutions to provide home care, expand the scope of services, make full use of medical resources, and address the imbalance between the supply of and demand for medical resources.

Our study found that "Internet+Nursing Service" apps provide a limited number of service programs—mostly routine care programs—with few special care programs, such as psychological care, hospice care, and child care programs. Only 4 of the 17 apps provided hospice care services, and none provided psychological care services, despite the high demand for these programs. In one study, it was found that 92.3% of the older adult population in urban and rural areas needed psychological comfort [39], and in another study, 10.32% of housebound older adults believed that hospice care should be carried out [40]. Further, as the concept of childbearing changes,

people are pursuing more scientific and specialized childcare, and child health care and nursing have become more emphasized. Research has shown that providing nutritional guidance, growth and development guidance, and child psychological care to families of preterm infants through "Internet+Nursing Service" platforms could promote the growth and intellectual development of preterm infants [41]. Medical institutions in each region should have an in-depth understanding of the needs of service users and the characteristics of different groups of people. They should gradually expand the list of "Internet+Nursing Services" by taking into account the actual situations of medical institutions and the local medical resources to optimize the structure of service items and meet the needs of service users.

The total quality scores of the 17 apps ranged from 1.92 to 4.92, with 1 (6%) app rated as "poor" and 11 (65%) rated as "fair," indicating that the overall quality of "Internet+Nursing Service" apps was not good. This may be related to the fact that app development engineers do not fully understand the medical industry. Most "Internet+Nursing Service" apps (9/17, 53%) were developed by corporations, and the developers may not have taken into account the specificity of medical software before development. They also may not have communicated well with medical staff during app development or understood the content and characteristics of "Internet+Nursing Service" apps.

The process of target user evaluation not only strengthens the interactions between users and the software but also identifies weak points (ie, from user feedback) that developers may have missed [42]. Our researchers found that some apps had problems, such as crashes and the inability to log in during use, which affected the user experience. Therefore, research and development organizations need to conduct premarket research and postmaintenance work to ensure the smooth operation of platforms.

From the heat map (Multimedia Appendix 2), it can be seen that most apps (12/17, 71%) scored at a good or above level in terms of the *Layout*, *Graphics*, *Performance*, and *Ease of Use* entries. This indicates that most app development teams pay more attention to the visual effects, ease of use, and smoothness of their apps. However, of the 17 apps, 14 (82%) had average scores of less than 4.0 in the *Credibility* entry, and 16 (94%) had average scores of less than 4.0 in the *Demand* entry, indicating average or even poor performance in these areas. During the evaluation, our researchers found that some app development teams did not clearly label the source or publisher information when publishing health science articles or videos. This may cause users to doubt the authenticity, authority, and reliability of the articles when reading them, affecting their ability to build health knowledge and manage their own health [43]. It is important to publish health science articles with scientific evidence and to clarify the sources, authors, time of publication, and applicable populations of standardized content to increase the credibility of information. In terms of meeting user needs, we found that most apps provided information from health science literature in text form, with content mostly focused on introducing services, medical institutions, and health care experts. We also noted that after content was published on an app, the content was not updated for a long time. Some apps also only had 2 to 3 articles and could not meet user needs. Therefore, in addition to not meeting user needs in terms of service programs, there are also deficiencies in providing up-to-date and relevant health information.

App developers can improve the credibility of health science articles and videos by taking several steps. First, they should ensure that the information provided is based on scientific evidence and comes from reputable sources. This can be done by clearly labeling the sources, authors, time of publication, and applicable populations of the content. Second, they should regularly update the content to ensure that it is current and relevant. Third, they should provide references or links to the original sources of information, so that users can verify the accuracy of the information. By taking these steps, app developers can increase the credibility of their health science articles and videos and help users build their health knowledge and manage their own health.

Health science popularization should aim to provide basic concepts and knowledge in the field of health, with a focus on healthy lifestyles and behaviors. Health science content should be regularly updated to keep up with social hot spots, seasonal changes, and the occurrence of epidemics to provide users with the most up-to-date and relevant information. To achieve this, app development teams should conduct market research to understand the needs and characteristics of their users, including users' cultural levels and reading habits. They should also keep track of social hot spots, seasonal changes, and other factors to provide relevant health policies, basic medical knowledge, diet and exercise guidance, psychological guidance, and knowledge regarding disease prevention or first aid in daily life. In addition to providing information in graphic form, app developers can also use video and audio formats to present health science information from multiple angles, dimensions, and levels to meet the needs of users at different levels. By providing accurate and scientific health information in a timely manner, app developers can help users improve their health knowledge and quality of life.

Our study found no correlation between uMARS-C ratings and app store ratings, suggesting that app store ratings do not reflect the quality of apps. This may have been due to the small number of app store ratings, differences in app store rating mechanisms, developer marketing strategies, or users' preferences (eg, app favoritism among users). If the quality of apps is judged solely based on app store ratings, users may download low-quality apps, thereby affecting their usage experience and even causing them to distrust care services. Additionally, health care professionals may be unable to accurately recommend high-quality apps to patients or their families. This further demonstrates the necessity of evaluating the quality of "Internet+Nursing Service" apps by using objective rating scales.

Limitations

This study has several limitations. First, we only searched for "Internet+Nursing Service" apps that were updated until December 1, 2022, and did not consistently track the uploads and downloads of related apps. Second, we only used iPhone,

Huawei, and Xiaomi phones to download and evaluate apps and did not use other systems, such as Meizu, Samsung, and Windows phones. During app development, the development team may modify app functions for different systems due to differences in system algorithms, resulting in differences in app functions. Therefore, future research should take system differences into account and conduct more comprehensive quality evaluations of apps for different systems. Third, there were only 5 researchers in this study; all were under the age of 30 years and had a high level of e-literacy, which may have introduced bias in entries, such as the *Ease of Use* entry, during the evaluation process. However, older people, who are the main target of "Internet+Nursing Services," have varying levels of e-literacy and may have different understandings and judgments of an app's ease of use. Therefore, future studies may consider including evaluators with different backgrounds and health literacy levels.

Conclusions

In this study, we used the uMARS-C to evaluate the quality of "Internet+Nursing Service" apps. We found that the service coverage of these apps was concentrated in first-tier cities (eg, Beijing and Shanghai), new first-tier cities, and some second-tier cities, with a limited number of service items and a need to optimize the structure of service items. The quality evaluation results showed that the quality of apps was not good, especially in terms of information credibility and meeting users' needs. Further, the scale scores did not correlate with app store scores. Therefore, "Internet+Nursing Service" app development teams need to pay attention to improving the quality of their apps. Before releasing an "Internet+Nursing Service" app, they should fully understand the needs of their target users, as well as the characteristics of this type of app, and communicate with relevant professionals in the field. They should also orient the release of health knowledge in the app toward user needs and improve the credibility and readability of content. After releasing an app, it is necessary to maintain and update it to ensure its normal operation and the timely updating of health education content.

Acknowledgments

We would like to thank three other researchers (Guanghan Zhang, Ying Liu, and Liping Fan) for their participation. This work was supported by the Key Project of Natural Science Research in Colleges and Universities of Anhui Province (grant KJ2021A0255).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Details of the 17 "Internet+Nursing Service" apps.
[DOC File, 23 KB - mhealth_v12i1e52169_app1.doc]

Multimedia Appendix 2
Heat map of the average scores for each item and app. The colors range from blue (worst score) to white (best score).
[DOCX File, 1693 KB - mhealth_v12i1e52169_app2.docx]

References

1. The 51st China statistical report on internet development [Article in Chinese]. China Internet Network Information Center. 2023. URL: https://www.cnnic.net.cn/n4/2023/0302/c199-10755.html [accessed 2023-03-02]
2. Ying F, Xing-Hong JI. Research on user satisfaction based on medical and health app [Article in Chinese]. Computer Engineering & Software 2020 Nov 30;41(10):38-41. [doi: 10.3969/j.issn.1003-6970.2020.10.010]
3. West JH, Belvedere LM, Andreasen R, Frandsen C, Hall PC, Crookston BT. Controlling your "app"etite: how diet and nutrition-related mobile apps lead to behavior change. JMIR Mhealth Uhealth 2017 Jul 10;5(7):e95. [doi: 10.2196/mhealth.7410] [Medline: 28694241]
4. Sieverdes JC, Treiber F, Jenkins C, Hermayer K. Improving diabetes management with mobile health technology. Am J Med Sci 2013 Apr;345(4):289-295. [doi: 10.1097/MAJ.0b013e3182896cee] [Medline: 23531961]
5. LeBeau K, Huey LG, Hart M. Assessing the quality of mobile apps used by occupational therapists: evaluation using the user version of the Mobile Application Rating Scale. JMIR Mhealth Uhealth 2019 May 1;7(5):e13019. [doi: 10.2196/13019] [Medline: 31066712]
6. Pagoto S, Schneider K, Jojic M, DeBiasse M, Mann D. Evidence-based strategies in weight-loss mobile apps. Am J Prev Med 2013 Nov;45(5):576-582. [doi: 10.1016/j.amepre.2013.04.025] [Medline: 24139770]
7. Thomala LL. Leading medical apps by monthly active users in China 2023. Statista. 2023. URL: https://www.statista.com/statistics/1043739/china-leading-healthcare-Apps-monthly-active-users [accessed 2023-12-12]

8. Policy interpretation on the pilot work of "internet + nursing service" [Article in Chinese]. National Health Commission of the People's Republic of China. 2019. URL: https://www.gov.cn/zhengce/zhengceku/2019-10/08/content_5436955.htm [accessed 2019-01-22]
9. Dianrui Y, Mei S. Research progress of "internet + nursing service" in the health management of diabetic patients [Article in Chinese]. Nursing Practice and Research 2022 Sep;19(18):2751-2755 [FREE Full text]
10. Yan W, Liu L, Huang WZ, et al. Study on the application of the internet + nursing service in family rehabilitation of common bone and joint diseases in the elderly. Eur Rev Med Pharmacol Sci 2022 Sep;26(18):6444-6450. [doi: 10.26355/eurrev_202209_29743] [Medline: 36196694]
11. Li X. Cost-effectiveness analysis of "internet +nursing service" model for discharged patients with cholelithiasis with T tube. China National Knowledge Infrastructure. 2021. URL: https://kns.cnki.net/kcms/detail/detail.aspx?FileName=1022431015.nh&DbName=CMFD2022 [accessed 2024-02-09]
12. Shuhong Z. Informatization construction and risk analysis of internet + nursing service model [Article in Chinese]. Ke JI Fang 2021 Jun(18):79-81. [doi: 10.19392/j.cnki.1671-7341.202118037]
13. Dongge R. Research on intention of using fitness app continuously based on SOR theory [Article in Chinese]. China National Knowledge Infrastructure. URL: https://kns.cnki.net/kcms2/article/ [accessed 2024-02-10]
14. Chan KL, Chen M. Effects of social media and mobile health apps on pregnancy care: meta-analysis. JMIR Mhealth Uhealth 2019 Jan 30;7(1):e11836. [doi: 10.2196/11836] [Medline: 30698533]
15. Ploderer B, Brown R, Seng LSD, Lazzarini PA, van Netten JJ. Promoting self-care of diabetic foot ulcers through a mobile phone app: user-centered design and evaluation. JMIR Diabetes 2018 Oct 10;3(4):e10105. [doi: 10.2196/10105] [Medline: 30305266]
16. Gareth S. The importance of software quality for the digital world [Article in Chinese]. Software and Integrated Circuit 2022(4):16-18. [doi: 10.19609/j.cnki.cn10-1339/tn.2022.04.019]
17. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile App Rating Scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27. [doi: 10.2196/mhealth.3422] [Medline: 25760773]
18. Hajesmaeel-Gohari S, Khordastan F, Fatehi F, Samzadeh H, Bahaadinbeigy K. The most used questionnaires for evaluating satisfaction, usability, acceptance, and quality outcomes of mobile health. BMC Med Inform Decis Mak 2022 Jan 27;22(1):22. [doi: 10.1186/s12911-022-01764-2] [Medline: 35081953]
19. Nouri R, Kalhori SRN, Ghazisaeedi M, Marchand G, Yasini M. Criteria for assessing the quality of mHealth apps: a systematic review. J Am Med Inform Assoc 2018 Aug 1;25(8):1089-1098. [doi: 10.1093/jamia/ocy050] [Medline: 29788283]
20. Jingjing W, Dongfu Q. Research on the application of "internet + nursing service" platform in a certain area [Article in Chinese]. Journal of Nanjing Medical University 2022 Apr;22(2):180-185. [doi: 10.7655/NYDXBSS20220214]
21. Zegen T, Wenxia Y, Haiyan L, Yunjing Y, Lixia W, Xinping L. The application effect of "internet+nursing service" in the continuity of indwelling catheter care [Article in Chinese]. Chinese Journal of Clinical Research 2021 Dec 20;34(12):1736-1738. [doi: 10.13429/j.cnki.cjcr.2021.12.035]
22. Liying C, Jing L, Jiegang Z. Development status quo and influencing factors of "internet+nursing service" for medical institutions in Luoyang city [Article in Chinese]. Chin Nurs Res 2022 Aug 10;36(15):2801-2804 [FREE Full text]
23. Zhifang R, Xueli G, Yanling W, Lili Z, Qian X. Construction of the quality evaluation index system of "internet+ nursing service" based on Donabedian theory and SERVQUAL model. Chinese Nursing Management 2022 Mar 15;22(3):391-396. [doi: 10.3969/j.issn.1672-1756.2022.03.014]
24. Pan-Pan H, Hong C, Ying-Hong Z, Yan W, Lian-Sheng Z. Construction of "internet + nursing service" risk management evaluation index system based on 4R-model crisis management theory. Journal of Nursing 2022 May;29(9):16-20. [doi: 10.16460/j.issn1008-9969.2022.09.016]
25. Hui L, Feng L. Studying on the risk control strategy of internet + nursing services based on SARF theory. Chinese Health Service Management 2022 Jun 15;39(6):477-480 [FREE Full text]
26. Fang L, Zixia Z, Yan Z, Yuehua Z. Investigation on the influencing factors of chronic disease patients' awareness and acceptance of "internet + nursing services" in Dongtai city [Article in Chinese]. Chinese and Foreign Medical Research 2022 Mar 5;20(7):98-101. [doi: 10.14033/j.cnki.cfmr.2022.07.024]
27. Wenrong L, Haimei G, Xiaoling W, Qin Z, Yunyun H, Lihui L. A qualitative study on the attitude, experience and needs of family caregivers towards internet + nursing services [Article in Chinese]. Journal of Nursing 2022 Aug 25;29(16):75-78. [doi: 10.16460/j.issn1008-9969.2022.16.075]
28. Nanxuan Z, Yinxin W. Research on residents' cognition and willingness of "Internet plus Nursing Service": a case study of Fujian Province [Article in Chinese]. Journal of Fujian Medical University(Social Science Edition) 2021 Dec 15;22(6):24-29 [FREE Full text]
29. Chenlin Z, Xuemei L, Xiaoyan L, Huixin W. Reliability and validity of the Chinese version of the user version of Mobile Application Rating Scale. Chinese Journal of Modern Nursing 2020;26(19):2608-2612. [doi: 10.3760/cma.j.cn115682-20190806-02814]

30. Lambrecht A, Vuillerme N, Raab C, et al. Quality of a supporting mobile app for rheumatic patients: patient-based assessment using the user version of the Mobile Application Scale (uMARS). Front Med (Lausanne) 2021 Jul 22;8:715345. [doi: 10.3389/fmed.2021.715345] [Medline: 34368202]
31. Li Y, Ding J, Wang Y, Tang C, Zhang P. Nutrition-related mobile apps in the China app store: assessment of functionality and quality. JMIR Mhealth Uhealth 2019 Jul 30;7(7):e13261. [doi: 10.2196/13261] [Medline: 31364606]
32. Adam A, Hellig JC, Perera M, Bolton D, Lawrentschuk N. 'Prostate Cancer Risk Calculator' mobile applications (apps): a systematic review and scoring using the validated user version of the Mobile Application Rating Scale (uMARS). World J Urol 2018 Apr;36(4):565-573. [doi: 10.1007/s00345-017-2150-1] [Medline: 29222595]
33. Qingyuan H, Diehua H, Ye W. Function and quality assessment of app appointment registry in hospital using uMARS [Article in Chinese]. Chinese Journal of Medical Library and Information Science 2018 Jul 15;27(7):14-18. [doi: 10.3969/j.issn.1671-3982.2018.07.003]
34. Beijing Internet Home Care Service Item Catalog (2022 edition) [Article in Chinese]. Beijing Municipal Health Commission. 2022. URL: https://wjw.beijing.gov.cn/zwgk_20040/ylws/202210/t20221011_2833382.html [accessed 2024-02-09]
35. 2022 China's City Business Attractiveness Ranking [Article in Chinese]. YiCai. URL: https://www.yicai.com/topic/101425010/ [accessed 2024-02-09]
36. Deering M. What are the pros and cons of telehealth nursing? NurseJournal. 2022. URL: https://nursejournal.org/careers/telehealth-nurse/pros-and-cons/ [accessed 2022-09-14]
37. Major figures on 2020 population census of China [Article in Chinese]. National Bureau of Statistics. 2021 Nov 26. URL: https://www.stats.gov.cn/sj/pcsj/rkpc/d7c/ [accessed 2024-02-09]
38. 2022 China health care statistics yearbook [Article in Chinese]. National Health Commission of People's Republic of China. 2023 May 17. URL: http://www.nhc.gov.cn/mohwsbwstjxxzx/tjtjnj/202305/6ef68aac6bd14c1eb9375e01a0faa1fb.shtml [accessed 2024-02-09]
39. Huanhuan W, Jing Z. Current demand for "internet+" home nursing service by elderly patients in an urban and rural area and analysis of influencing factors [Article in Chinese]. Shanxi Medical Journal 2022 Nov 15;51(21):2418-2421 [FREE Full text]
40. Shuyu F, Xulan L, Hua L, Dan Z. Study on the demand and influencing factors of the elderly living at home for "internet+nursing services" [Article in Chinese]. Modern Nurse 2023 Feb 1;30(2):148-152. [doi: 10.19793/j.cnki.1006-6411.2023.06.040]
41. Cuiwei C, Dan L, Shuchan O, Lianhao D, Zhiyao T. Effect of continuous nursing based on internet+ on the growth and development and quality of life of premature infants [Article in Chinese]. Journal of Shenyang Medical College 2022 Jul 15;24(4):396-399. [doi: 10.16753/j.cnki.1008-2344.2022.04.014]
42. Koepp J, Baron MV, Martins PRH, et al. The quality of mobile apps used for the identification of pressure ulcers in adults: systematic survey and review of apps in app stores. JMIR Mhealth Uhealth 2020 Jun 16;8(6):e14266. [doi: 10.2196/14266] [Medline: 32470916]
43. Guidelines on the establishment of a sound mechanism for the publication and dissemination of health science knowledge across the media [Article in Chinese]. National Health Commission of the People's Republic of China. 2022 May 31. URL: http://www.nhc.gov.cn/xcs/s3581/202205/1c67c12c86b44fd2afb8e424a2477091.shtml [accessed 2022-07-01]

Abbreviations

MARS: Mobile App Rating Scale
mHealth: mobile health
TCM: traditional Chinese medicine
uMARS: user version of the Mobile App Rating Scale
uMARS-C: Chinese version of the user version of the Mobile App Rating Scale

Edited by L Buis; submitted 28.08.23; peer-reviewed by E Prochaska, F Carrouel; revised version received 07.01.24; accepted 11.01.24; published 16.02.24.

Please cite as:
Yuan S, Liu M, Peng Y, Hu J, Li B, Ding X, Xie L
"Internet+Nursing Service" Mobile Apps in China App Stores: Functionality and Quality Assessment Study
JMIR Mhealth Uhealth 2024;12:e52169
URL: https://mhealth.jmir.org/2024/1/e52169
doi:10.2196/52169

Review

Functionality and Quality of Asthma mHealth Apps and Their Consistency With International Guidelines: Structured Search and Evaluation

Billy Robinson[1], BMedSci, BMBS, MMed; Eleni Proimos[2], BMedSc, MD; Daniel Zou[3], BBioMedSci, MD; Enying Gong[4], BA, MPH, PhD; Brian Oldenburg[5,6], PhD; Katharine See[1], MBBS

[1]Department of Respiratory Medicine, Northern Health, Epping, Australia

[2]Faculty of Medicine, Nursing and Health Sciences, Monash University, Melbourne, Australia

[3]Faculty of Medicine, Dentistry and Health Sciences, University of Melbourne, Melbourne, Australia

[4]School of Population Medicine and Public Health, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

[5]Baker Heart and Diabetes Institute, Melbourne, Australia

[6]Academic and Research Collaborative in Health, LaTrobe University, Melbourne, Australia

Corresponding Author:
Billy Robinson, BMedSci, BMBS, MMed
Department of Respiratory Medicine
Northern Health
185 Cooper St
Epping, 3076
Australia
Phone: 61 38405 8000
Email: billymed1994@gmail.com

Abstract

Background: Asthma is a chronic respiratory disorder requiring long-term pharmacotherapy and judicious patient self-management. Few studies have systematically evaluated asthma mobile health (mHealth) apps for quality and functionality; however, none have systematically assessed these apps for their content alignment with international best practice guidelines.

Objective: This review aims to conduct a systematic search and evaluation of current mHealth apps in the Australian marketplace for their functionality, quality, and consistency with best practice guidelines.

Methods: The most recent Global Initiative for Asthma (GINA) guidelines were reviewed to identify key recommendations that could be feasibly incorporated into an mHealth app. We developed a checklist based on these recommendations and a modified version of a previously developed framework. App stores were reviewed to identify potential mHealth apps based on predefined criteria. Evaluation of suitable apps included the assessment of technical information, an app quality assessment using the validated Mobile App Rating Scale (MARS) framework, and an app functionality assessment using the Intercontinental Medical Statistics Institute for Health Informatics (IMS) Functionality Scoring System. Finally, the mHealth apps were assessed for their content alignment with the GINA guidelines using the checklist we developed.

Results: Of the 422 apps initially identified, 53 were suitable for further analysis based on inclusion and exclusion criteria. The mean number of behavioral change techniques for a single app was 3.26 (SD 2.27). The mean MARS score for all the reviewed apps was 3.05 (SD 0.54). Of 53 apps, 27 (51%) achieved a total MARS score of ≥3. On average, the reviewed apps achieved 5.1 (SD 2.79) functionalities on the 11-point IMS functionality scale. The median number of functionalities identified was 5 (IQR 2-7). Overall, 10 (22%) of the 45 apps with reviewer consensus in this domain provided general knowledge regarding asthma. Of 53 apps, skill training in peak flow meters, inhaler devices, recognizing or responding to exacerbations, and nonpharmacological asthma management were identified in 8 (17%), 12 (25%), 11 (28%), and 14 (31%) apps, respectively; 19 (37%) apps could track or record “asthma symptoms,” which was the most commonly recorded metric. The most frequently identified prompt was for taking preventive medications, available in 9 (20%) apps. Five (10%) apps provided an area for patients to store or enter their asthma action plan.

Conclusions: This study used a unique checklist developed based on the GINA guidelines to evaluate the content alignment of asthma apps. Good-quality asthma apps aligned with international best practice asthma guidelines are lacking. Future app

development should target the currently lacking key features identified in this study, including the use of asthma action plans and the deployment of behavioral change techniques to engage and re-engage with users. This study has implications for clinicians navigating the ever-expanding mHealth app market for chronic diseases.

Trial Registration: PROSPERO CRD42021269894; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=269894

International Registered Report Identifier (IRRID): RR2-10.2196/33103

(JMIR Mhealth Uhealth 2024;12:e47295) doi:10.2196/47295

KEYWORDS

asthma; mobile health; mHealth; app; mobile; chronic disease; systematic review; smartphone; review methodology; respiratory; compliance; guideline; guidelines; review of apps; evaluation; quality; best practices; apps; mobile phone

Introduction

Background

Asthma is a chronic respiratory disorder that is clinically defined as a combination of typical episodic respiratory symptoms, such as wheezing, shortness of breath, cough, chest tightness, and significant variable reversible airflow limitation [1]. When the frequency or severity of these symptoms increases compared with the baseline respiratory status, it represents an "asthma exacerbation" or "flare-up" [2]. Judicious self-monitoring and management of regular asthma medications, symptoms, and exacerbations are key to allowing patients with asthma to live with a high quality of life and prevent hospitalizations or death [3]. The Global Initiative for Asthma (GINA) regularly releases updated best practice asthma guidelines based on reviews of scientific literature by an international panel of experts [4]. Many local asthma management guidelines have been derived from these international guidelines. In addition to pharmacotherapy, the guidelines advise that patient education on medication adherence, exacerbation recognition, and management is key to self-management [4].

Asthma is a significant chronic health issue worldwide, affecting 1% to 18% of the global population [4]. Australia is no exception, with asthma affecting millions and accounting for 34% of Australia's respiratory disease burden and 2.5% of the total disease burden [5]. Asthma leads to numerous emergency department visits and urgent health care visits [5]. Furthermore, those living with asthma report a poorer quality of life and are less likely to rate their health status as excellent or very good [5]. When observing the total cost that asthma has on the Australian health system, it is evident that hospital-related costs outweigh non–hospital-related costs (Aus $205 million/year [approximately US $150 million] vs Aus $163 million/year [approximately US $120 million]) [5]. Theoretically, reducing exacerbations would reduce the requirement for hospitalizations; unplanned primary care presentations; and indirect costs, such as work absenteeism, and thus assist in reducing these costs.

With the increasing availability of smartphones, mobile health (mHealth) apps have become accessible to a large percentage of the population and represent a potential medium through which patients can improve their ability to self-manage asthma. Deloitte's recent review of Australia's telecommunication status found that 89% of the Australian population uses smartphones [6]. These apps are already available for download and use; however, it is imperative that a review of their quality, functionality, and alignment with evidence-based best practices is conducted to inform both users and health professionals. These apps represent an opportunity for new ways to empower patients to track asthma symptoms, learn about their condition, and undertake practical self-management strategies. The established Mobile App Rating Scale (MARS) is generally used to assess the usability and overall quality of mHealth apps [7,8]. Although systematic evaluations of asthma mobile apps have been conducted in the past, many of these studies did not assess the apps' functionality or quality using a validated rating scale, such as MARS [9-11]. Furthermore, to our knowledge, none of these prior evaluations assessed all available apps systematically for the presence and quality of information they provide compared with available best practice management guidelines, such as the GINA guidelines [9-11].

This systematic search and evaluation review assessed the functionality and quality of free and paid asthma mHealth apps targeted toward adults with asthma available from the Apple App Store (iOS) and Google Play Store (Android), as well as their consistency with recommendations from the GINA guidelines, making it the first review of its kind.

Objective

The objective of this review was to conduct a systematic search of available English-language mHealth apps targeted toward adults with asthma in Australia, to evaluate their overall quality and functionality and to assess the consistency and quality of the content and information they provide in alignment with current best practice guidelines for asthma management.

Methods

Overview

The GINA guidelines were reviewed by 2 medical professionals to identify and establish a consensus of key recommendations from the guidelines that could feasibly be incorporated into an app for asthma management. The mobile apps in the selected app stores were identified and screened based on the selection criteria. Finally, we assessed the quality, functionality, and alignment of the apps with the guidelines identified in the first step of the screened mHealth apps. An in-depth description of the research protocol was published the previous year [12].

Study Setting

Given the primary residences of the researchers involved in this review, this study was conducted by medical practitioners,

medical students, and digital health researchers using apps available in Australia. The mHealth apps presented in English on Australian mobile app stores were assessed. Some mHealth apps identified in this review may not be available in regions outside Australia. Similarly, apps available in other regions may not be available in Australia. However, given that most of the apps identified in this review are also available in other English-speaking regions such as the United Kingdom and the United States, the results are largely generalizable to these regions. Given that the researchers were adult medical practitioners who did not manage pediatric patients, only those mHealth apps targeted toward adults with asthma were evaluated.

Wherever possible, the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines for systematic reviews were followed [13]. Given that this was a review of mobile apps instead of journal articles, some items in the PRISMA checklist were not relevant to this review. The checklist is shown in Multimedia Appendix 1.

Review of the GINA Guidelines and Checklist Creation

To assess the usability and overall quality of the app, we used the established MARS [7,8]. A review of the available literature using the CINAHL, MEDLINE, Embase, and PubMed databases revealed that 1 research group had developed an asthma app assessment framework yet to be derived and validated into an instrument [14]. For the reasons outlined in our published research protocol, we decided to combine aspects of the framework by Guan et al [12] with our own checklist derived from the GINA guidelines. Two reviewers, BR and KS, independently assessed the 2020 GINA guidelines for identifiable recommendations that could be incorporated into an mHealth app. Following this, the reviewers examined each other's identified recommendations to see whether a consensus could be reached on the recommendations from the GINA guidelines that could be incorporated into an mHealth app. The identified recommendations from each author and those where a consensus was reached, which represent the final identified recommendations, are shown in Table 1.

A final checklist modified from the framework by Guan et al [14] (Table 1) was developed to include recommendations we identified from the above process while excluding the information we gathered through the MARS framework. To determine app consistency with the GINA guidelines, participants were assessed for the presence or absence of features identified through this process. This is further discussed in subsequent sections.

Table 1. Recommendations identified from the Global Initiative for Asthma guidelines that could be incorporated into a mobile health app.

Reviewer 1	Reviewer 2	Consensus reached
Assess symptom control (eg, ACQ[a])	Support for assessing symptom control for a 4-week period	Support for assessing symptom control for a 4-week period considering the frequency of asthma symptoms, night waking because of asthma, frequency of SABA[b] use, and any activity limitation because of asthma; uses recognized screening, symptom control or numerical asthma control tools, and tracks peak flow measurement
Ability to self-track symptoms with or without peak flow	—[c]	Encourages patients to track symptoms and peak flow measurements
Risk factors for future exacerbations	Helps users identify the future risk of exacerbations	Helps users identify the risk of future exacerbations
Screens for comorbidities and education regarding managing them	Screens for comorbidities and assists patients with managing them	Screens for relevant comorbidities and educates patients on the management of these comorbidities
Inhaler technique with or without video	Provides education on appropriate inhaler techniques	Provides education on appropriate inhaler techniques with visual aids
Ability to record action plan	Provides an area for patients to keep and refer to their written action plan	Provides an area for patients to keep and refer to their written action plan
Reminder to engage with primary care	Reminds users to see their HCP[d] for management and review of asthma	Provides reminders to users to see their HCP for management and review of asthma
—	Specifically provides suggestion to see an HCP if a patient is using only a SABA.	Specifically provides suggestion to see an HCP if a patient is using a SABA alone
Medication adherence	Prompts users to adhere to controller medications even when symptoms are infrequent	Prompts users to adhere to controller medications even when symptoms are infrequent
General asthma education	Provides knowledge on general asthma information, management of asthma, modifiable risk factors and strategies to address them, and when to see an HCP	Provides knowledge on general asthma information, management of asthma, modifiable risk factors and strategies to address them, when to see an HCP, and identification and management of comorbidities
Help with activating action plan	Provides advice on when to refer to a patient's asthma action plan based on self-monitoring of symptoms or PEF[e]	Provides advice on when to refer to a patient's asthma action plan based on self-monitoring of symptoms or PEF
—	Prompts patient to see the primary HCP if features of asthma exacerbation (symptoms and SABA use) are identified using the app	Prompts patient to see the primary HCP if features of asthma exacerbation (symptoms and SABA use) are identified using the app

[a]ACQ: Asthma Control Questionnaire.

[b]SABA: short-acting β-agonist.

[c]Recommendation identified by one reviewer but not the other.

[d]HCP: health care provider.

[e]PEF: peak expiratory flow.

Identification, Screening, and Selection of Mobile Apps for Review

This review included both free and paid apps from the 2 most popular app stores in Australia across the iOS and Android operating systems: the Apple App Store (Apple) and Google Play Store (Google). Our published protocol outlines the steps taken for quality assurance [12]. Our approach for identifying these apps followed the approach used in similar studies [9-11]. Before commencing the initial search for apps, the reviewers ensured that the operating systems on the chosen smartphones were up to date. Each reviewer used different phones to assess the apps, but all updated the Android operating system (OS) to the Android 11 OS (Google) when reviewing apps from the Google Play Store. In the search bar of each store, we input the term *asthma*. Two reviewers (BR and DZ) independently searched both app stores on August 10, 2021, from Melbourne, Australia. After obtaining the results for this search term, each reviewer stored the information on an Excel (Microsoft Corp) spreadsheet (Multimedia Appendix 2). The reviewers then compared their results to ensure that they captured all the available apps.

For further evaluation of all the apps identified above, BR and DZ individually reviewed the app title, description, and attached photos and determined whether the app met all inclusion criteria and no exclusion criteria. An identified asthma app was included in the evaluation stage if all the following applied: (1) its primary role was related to asthma, (2) it was targeted to those with asthma, (3) it could be run on mobile phones, and (4) it was in English. Apps were excluded if any of the following

applied: (1) they were not primarily related to asthma, (2) they were primarily targeted toward health care professionals (as stated in the product description), (3) they were not in English, and (4) they were targeted toward pediatric patients. This information was entered into an Excel (Microsoft Corp) spreadsheet for record-keeping (Multimedia Appendix 3). For further evaluation, all apps identified as meeting the above criteria were downloaded by a third reviewer (EP) who identified apps that did not install or function properly after downloading, eliminating them from the review. Finally, the last round of screening was conducted by the reviewer EP. In this round, duplications (ie, apps available on both stores), inaccessible apps, and "lite" version apps, where a pro version was available, were removed from the review. This process was comparable with similar reviews that examined the quality of mobile apps for diabetes self-management [15].

App Evaluation and Data Extraction

Reviewer Training

A day-long training session was conducted before the initial data extraction. This training session was similar to the one performed by Gong et al [15] for their diabetes app review [15]. A step-by-step reference guide was created by the primary researchers to inform reviewers regarding how to complete the various frameworks and checklists involved in the study. This is provided in Multimedia Appendix 4 [3,4,16,17].

App Evaluation and Data Extraction Overview

An internet database was established on Qualtrics (Qualtrics International Inc) for data extraction. A total of 3 reviewers were involved in data collection. A web-based random team generator was used for all apps identified for further evaluation during the screening process so that each app was randomly allocated to 2 assessors (BR, DZ, or EP). The 2 assessors assigned to the app independently reviewed the in-store app description, downloaded the app, and used it for a minimum of 20 minutes to become familiar with all its functions [12]. The reviewers subsequently conducted the evaluations and entered the data into the Qualtrics database. Each reviewer performed this process individually without communicating their results to one another. There are 4 key aspects of the app evaluation and data extraction process, as summarized in the checklist provided in Multimedia Appendix 5 and the step-by-step guide to data collection in Multimedia Appendix 4.

Technical Information About the App

The first step in data collection involved gathering basic technical information about the app. The decision of which technical information to include was based on the MARS checklist and previous app review studies [7,9,15]. This was derived from publicly available information in the in-store app descriptions and in-app information sections. If required, the app developer's website was used. The technical information collected included the app name, date of release, date of update, developer, developer affiliations, price, rating, number of ratings, platform or platforms, size of the app, and number of downloads. A checklist for this section is provided in Multimedia Appendix 5.

App Quality Assessment

The app quality assessment was completed using the MARS tool to objectively determine the quality of the apps selected. This scale has 4 separate domains that are assessed to evaluate mobile app quality. These domains are engagement, functionality, esthetics, and information quality [7]. A total of 19 items, each with a 5-point scale regarding quality in the 4 domains mentioned above, make up the MARS score [7]. This framework is presented in Multimedia Appendix 5. Reviewers completed this tool by entering the information into the Qualtrics (Qualtrics International Inc) checklist for each app. Once this was completed, the mean score for that domain and the overall MARS score were calculated for every app. Following the objective MARS section, there are several subjective questions to evaluate user satisfaction and the perceived impact of the app on the user's knowledge, attitude, motivation to change, likelihood of change, and awareness of the importance of changing their asthma self-management [7]. These questions were answered by reviewers based on their experience using the app and their knowledge gained through the training sessions and clinical practice. Once the data were collected, the mean total MARS value and SD were calculated for each app.

The MARS tool assesses the presence or absence of 19 behavioral change techniques (BCTs). Although 93 types of BCTs are known, only the techniques outlined in MARS were assessed. This approach aligns with previous research, with a scope comparable with our review [7,10,15]. These techniques are outlined in Multimedia Appendix 4. To capture the presence of BCTs, an app was considered to have a BCT present, even if only 1 reviewer identified it. The median number of BCTs and the corresponding IQR of the apps was calculated using Stata (StataCorp) statistical software.

App Functionality

App functionality refers to what the app can do for a user and is an important marker of whether an app offers much benefit to users and the overall quality of the app. Although the MARS framework examines the overall quality of a mobile app, it focuses on the performance, ease of use, gestural design navigation, and navigation of the app [7]. Therefore, the Intercontinental Medical Statistics Institute for Health Informatics (IMS) Functionality Scoring System, henceforth known as the IMS functionality score, was used. This score was developed by the above institute and is based on 7 functionality criteria and 4 subcategories in the *record* functionality criterion. The IMS functionality score focuses on the scope of functions, including the ability of the app to inform, instruct, record, display, guide, remind or alert, and communicate information [18]. Each app was assessed for having or not having these functions and then given a total functionality score between 0 and 11 [18]. To capture the presence of all functionalities, an app was considered to have a functionality present even if one reviewer identified it. The mean, median, and IQR were calculated. The reviewers assessed each app for these functions and entered the data into Qualtrics (Qualtrics International Inc) database. This scoring system is presented in Multimedia Appendix 5.

Presence of App Features Consistent With Asthma Guidelines

As discussed above, key recommendations that could feasibly be incorporated into an asthma mHealth app were identified from the GINA guidelines. These recommendations, summarized in Table 1, were used to develop a more extensive checklist provided in Multimedia Appendix 5. The main functions of the app that we were interested in assessing in our checklist included asthma information, self-management skill training (including peak flow use, inhaler technique, and nonpharmacological strategies), monitoring of asthma symptoms, risk evaluation, and prompting (medication reminders, referring to action plan reminders, and suggestions for seeking health advice). Each of the selected apps was assessed using this checklist, and the data were entered into the Qualtrics (Qualtrics International Inc) database. To ensure consistency, an app was only assessed for the presence of the above function if both reviewers reached a consensus that the said feature was present. For single-reviewed those apps, the sole reviewer's analysis was used to determine whether the app did or did not have the examined feature.

Quality Assurance, Data Management, and Data Analysis

Training was provided to all the researchers, and a handbook for reviewers was provided to ensure the quality of this research. Selected apps were allocated to reviewers using a web-based random allocation software, and 2 different major app databases were searched to reduce selection bias. The apps were independently reviewed by 2 reviewers to reduce the likelihood of bias affecting the results. A protocol was published to reduce publication bias and enhance the transparency of this study [12]. During app evaluation, all data were entered into either an Excel (Microsoft Corp) spreadsheet during the screening process or into the web-based Qualtrics (Qualtrics International Inc) database. These were stored on a cloud-based system that only the researcher team could access. Once the evaluation was completed, all data were downloaded for subsequent analysis. This analysis comprised a descriptive analysis and calculation of the mean and SD or median and IQR.

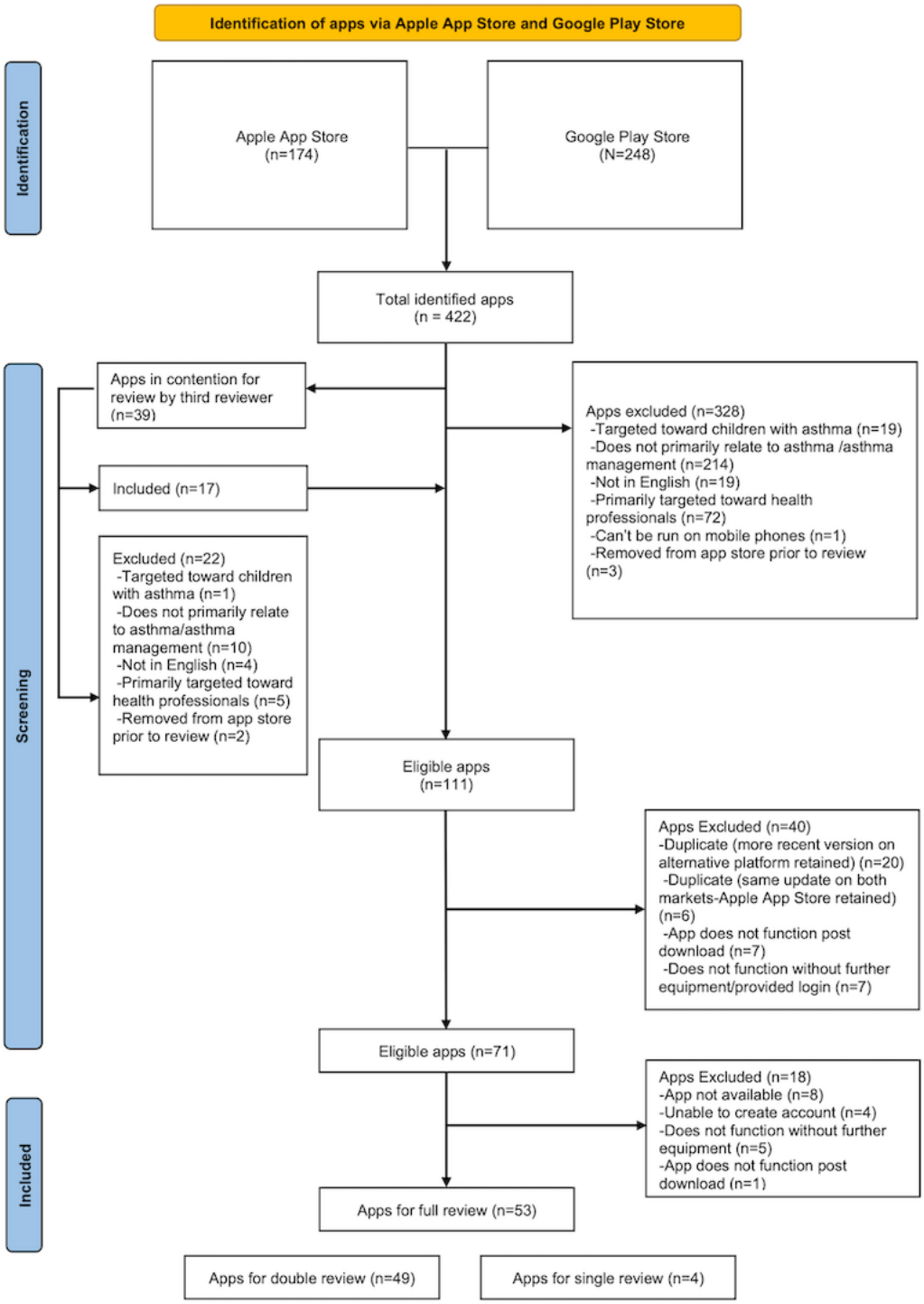
All data analyses were performed using Stata statistical software version 14 (StataCorp). Visual figures were generated using Excel (version 16; Office 365; Microsoft Corp).

Results

Identification, Screening, and Selection of Mobile Apps for Review

The process and results of identifying, screening, and selecting mobile apps are shown in Figure 1. A total of 174 unique apps from the Apple Store and 248 unique apps from the Google Play Store were identified. These 422 apps were assessed by 2 reviewers (BR and DZ). In total, 94 apps met all inclusion criteria and no exclusion criteria, although there was a discrepancy between the reviewers' opinions regarding the eligibility of 39 apps. A third reviewer (EP) identified 17 of these 39 apps as suitable for further review, which resulted in a total of 111 apps suitable for further assessment. A total of 40 apps were removed for reasons outlined in Figure 1. When a more recently updated app was available on 1 platform compared with the other, the older version was excluded from the review. When duplicate apps were available on both platforms and had been updated on the same date, the app from the Apple App Store was retained, whereas the app from the Google Play Store was excluded from the review. This was done to ensure consistency between reviewers and prevent skewing of results by assessing the same app twice. This resulted in 71 apps that were suitable for a complete assessment. Given the delay between the identification of apps and analysis, 18 of the above 71 apps were no longer assessable for the reasons outlined in Figure 1. A total of 4 apps were downloaded by 1 reviewer but not the other before they were removed from the market. This resulted in 49 apps assessed by 2 reviewers and 4 apps assessed by 1 reviewer (53 apps and 102 total reviews).

Figure 1. App screening process and results based on PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines.



Technical Information About the App

The technical information for each reviewed app can be found in Multimedia Appendix 2. Of the 53 apps assessed, 29 (55%) were from the Apple App Store, and 24 (45%) were from the Google Play Store. A total of 19 (36%) apps downloaded from the Apple App Store were also available on Google Play Store. As outlined above, apps available on both marketplaces were only downloaded from the Apple App Store and assessed on the iOS platform. The apps' last date of update ranged from February 2016 to April 2022. A total of 26 (49%) apps were updated from January 2020. The mean app size was 46.33 MB, and the median app size was 27 MB (IQR 9.2 MB-47.38 MB). App developers were primarily technical companies (28 apps), health care or pharmaceutical companies (4 apps), or a combination of both (4 apps). Six apps were created by private individuals, 3 were created from research or clinical institutions and the remaining 8 were created from developers from a variety of other backgrounds. The number of app downloads ranged from 10 to >10,000. Of 53 apps, 24 (45%) apps had a published user rating, and the median rating was 4 out of 5 (IQR 2.9-4.9). The number of people who provided a rating ranged from 0 to 523, with a median of 2.5 ratings per app (IQR 1-20). Of 53 apps, 42 (79%) apps were completely free, 5 (9%) apps required

users to pay to download, and 6 (11%) of the above free apps had the in-app ability to upgrade for a cost.

Presence of Behavior Change Techniques

The total number of each BCT identified across the 53 apps assessed is demonstrated in Table 2. The most frequent BCT observed was *self-monitoring or tracking*, which was identified in 38 (72%) of the 53 apps. The next commonly identified BCTs were *information or education*, seen in 33 (62%) apps, and *advice, tips, strategies, or skills training*, seen in 26 (49%) apps. The average number of BCTs in a single app from those reviewed was 3.26 (SD 2.27). The median number of BCTs for the apps reviewed was 3 (IQR 1-4).

Table 2. Assessed behavioral change techniques and the number of apps found to use these techniques (n=53).

Behavioral change technique	Apps with this technique, n (%)
Information or education	33 (62)
Self-monitoring or tracking	38 (72)
Advice or tips or strategies or skills training	26 (49)
Assessment	17 (32)
Feedback	12 (23)
Self-efficacy	9 (17)
Model or demonstrate behavior	3 (6)
Rewards and self-rewards	6 (11)
Goal setting	5 (9)
Provide social support	3 (6)
Perceived risks	1 (2)
Model or demonstrate behavior	3 (6)
Action planning	11 (21)
Motivation	2 (4)
Motivational readiness	2 (4)
Mindfulness or meditation	1 (2)
Perceived benefit	1 (2)

App Quality (MARS)

The mean MARS score for all reviewed apps was 3.05 (SD 0.54)**.** Of the 53 apps, 27 (51%) achieved a total MARS score of ≥3. A score of 3 on the MARS tool correlates to an "acceptable" quality app, <3 is inadequate or poor quality, and >3 represents a good or exceptional app [7]. Functionality was the highest rated MARS category with a mean score of 3.85 (SD 0.52), followed by esthetics with a mean score of 3.21 (SD 0.6). The information category had an average score of 2.78 (SD 0.83), and engagement had a mean score of 2.77 (SD 0.59). Table 3 shows the mean score for each of the 19 items on the MARS tool. Notably, the apps reviewed had higher scores in the gestural design, app description accuracy, and ease of use domains and lower scores in the evidence base, credibility, and entertainment domains. The mean scores for the quality and quantity of information were 3.21 (SD 1.95) and 2.68 (SD 1.63), respectively.

The final component of MARS allows reviewers to complete a subjective assessment of their opinions on the app. For double-reviewed apps, the mean value for the level of agreement for each domain was first calculated. Therefore, the number of apps will not be a whole number. This is summarized in Table 4 and demonstrates that there were few apps that reviewers would recommend to others, use >2 times in a 12-month period, or pay for. Only 8% (n=4) of the apps were rated >3 stars by the reviewers. For >50% of the apps, reviewers either disagreed or strongly disagreed that the app would impact the user's knowledge, attitudes, and intentions to change or change the rate of asthma exacerbations. For <20% of the apps, reviewers either strongly agreed or agreed that the app would improve the user's knowledge, attitudes, awareness, or intention to change behaviors to improve asthma self-management. However, reviewers either strongly agreed or agreed that 29% (n=15) and 27% (n=14) of apps would encourage users to seek further help in asthma management and reduce asthma exacerbations, respectively.

Table 3. Mean score for each category in the Mobile App Rating Scale (MARS) tool for the 53 assessed apps. Each category is assessed on a 5-point scale.

MARS category	Mean score for category
Entertainment	2
Interest	2
Customization	2
Interactivity	2
Target group	3
Performance	3
Ease of use	3
Navigation	3
Gestural design	4
Layout	3
Graphics	3
Visual appeal	2
Accuracy of app description	4
Goals	3
Quality of information	3
Quantity of information	2
Visual information	2
Credibility	1
Evidence base	1

Table 4. Results from the subjective assessment section of the Mobile App Rating Scale framework (n=53).

	Apps, n (%)
Would you recommend this app to people who might benefit from it?	
Not at all, I would not recommend this app to anyone.	12 (23)
There are very few people I would recommend this app to.	18 (35)
Maybe, there are several people whom I would recommend it to.	18 (35)
There are many people I would recommend this app to.	4 (8)
Definitely, I would recommend this app to everyone.	0 (0)
How many times do you think you would use this app in the next 12 months?	
0	4(8)
1-2	28 (52)
3-10	7 (13)
11-50	14 (27)
>50	0 (0)
Would you pay for this app?	
No	34 (64)
Yes	18 (35)
Maybe	1 (2)
What is your overall star rating for the app?	
1 (one of the worst apps I have used)	10 (19)
2	20 (38)
3 (average)	18 (35)
4	3 (6)
5 (one of the best apps I have used)	1 (2)
Strongly disagree or disagree that the app will improve	
Awareness	40 (75)
Knowledge	37 (69)
Attitudes	32 (60)
Intention to change	32 (60)
Help seeking	27 (50)
Behavior change (reduce asthma exacerbations)	29 (54)
Strongly agree or agree that the app will improve	
Awareness	8 (15)
Knowledge	10 (19)
Attitudes	7 (13)
Intention to change	6 (12)
Help seeking	15 (29)
Behavior change (reduce asthma exacerbations)	14 (27)

IMS Functionality Score

Out of a potential 11 functionalities, an average IMS functionality score of 5.1 (SD 2.79) was achieved by the reviewed apps. The median number of functionalities identified was 5 (IQR 2-7). A total of 3 apps had 11 functionalities, although most apps had the ability to capture user-entered data (n=37, 70%), provide information in a variety of formats (n=35, 66%), and provide instructions to the user (n=33, 62%). Out of 53 apps, 7 (13%) apps had the ability to send alerts or propose behavioral interventions based on the collected data. The total number of apps that met these IMS functionality criteria is summarized in Table 5.

Table 5. Total number of apps meeting each of the Intercontinental Medical Statistics Institute for Health Informatics (IMS) functionality criteria (n=53).

IMS functionality	Apps containing this functionality, n (%)
Inform	35 (66)
Instruct	33 (62)
Record	37 (70)
Display	30 (57)
Guide	11 (21)
Remind or alert	27 (51)
Communicate	17 (32)
Collect data	37 (70)
Share data	26 (49)
Evaluate data	12 (23)
Intervene	7 (13)

Presence of App Features Consistent With Asthma Guidelines

Knowledge

Table 6 outlines the results of evaluating the presence of information or knowledge identified as important based on the GINA guidelines. This details the number of apps where reviewer consensus was achieved, the number of apps that provided knowledge or did not provide knowledge in the subcategories, whether knowledge was individualized, and whether knowledge was based on evidence.

Table 6. Number of apps providing knowledge in asthma in the various relevant domains identified as important based on the Global Initiative for Asthma guidelines (n=53).

	Reviewer consensus achieved, n (%)	Apps that provided knowledge in this domain, n (%)	Apps that did not provide knowledge in this domain, n (%)	Apps that provided individualized knowledge, n (%)	Apps that provided evidence-based knowledge, n (%)
General asthma knowledge	45 (85)	10 (22)	35 (78)	0 (0)	7 (16)
Asthma medications	46 (87)	13 (28)	33 (72)	0 (0)	12 (26)
Exacerbation management	44 (83)	14 (32)	30 (68)	0 (0)	12 (27)
Asthma risk factors and triggers	46 (87)	11 (24)	35 (76)	0 (0)	11 (24)

Skill Training

The number of apps that provide skills training in peak flowmeter use, inhaler device use, recognizing and responding to asthma exacerbations, and nonpharmacological management to reduce asthma exacerbations are summarized in Table 7.

Table 7. Number of apps which provide the specific skill training in the areas identified as important in the Global Initiative for Asthma guidelines (n=53).

	Apps with reviewer consensus, n (%)	Apps that provide skill training, n (%)	Apps which provide personalized skill training, n (%)
App provides general skill training in peak flowmeter use	46 (87)	8 (17)	0 (0)
Describes why and when to use peak flowmeter	0 (0)	5 (11)	0 (0)
Describes operational criteria for peak flowmeter	0 (0)	2 (4)	0 (0)
Demonstrates the use of peak flow meter through photos or videos	0 (0)	5 (11)	0 (0)
App provides general skill training in inhaler device use	49 (93)	12 (25)	0 (0)
Describes how to use a spacer	0 (0)	0 (0)	0 (0)
Demonstrates how to use a spacer through videos or photos	0 (0)	0 (0)	0 (0)
Demonstrates how to care for a spacer	0 (0)	0 (0)	0 (0)
Describes how to use common inhaler devices	0 (0)	0 (0)	0 (0)
Demonstrates how to use common inhaler devices through videos or photos	0 (0)	0 (0)	0 (0)
App provides general skill training in recognizing and responding to exacerbations	39 (74)	11 (28)	3 (8)
Encourages patients to monitor for signs of asthma exacerbation	0 (0)	9 (23)	1 (3)
Provide an area for asthma action plan	0 (0)	5 (13)	2 (5)
Specifically guides patients to use their asthma action plan	0 (0)	3 (8)	2 (5)
Provide information on how to use an asthma action plan	0 (0)	7 (18)	1 (3)
Prompts patient to see health care provider when required	0 (0)	7 (18)	0 (0)
App provides general skill training in nonpharmacological management strategies to reduce asthma exacerbations	45 (85)	14 (31)	1 (2)
Helps identify triggers that make symptoms worse	0 (0)	9 (20)	1 (2)
Advises avoidance of environmental smoke exposure	0 (0)	13 (29)	0 (0)
Advises avoidance of medications that can worsen asthma	0 (0)	2 (4)	0 (0)
Advises avoidance of occupation exposures	0 (0)	1 (2)	0 (0)
Advises on the avoidance of allergen exposure	0 (0)	13 (29)	0 (0)
Advises on avoidance of indoor or outdoor pollution	0 (0)	12 (27)	1 (2)
Advises on avoidance of emotional stress	0 (0)	4 (9)	0 (0)
Advises on regular moderate physical activity	0 (0)	6 (13)	1 (2)

App's Ability to Track and Display Health Information

Table 8 demonstrates the results of the assessment of whether apps had the ability to track and display different aspects of a user's key asthma information. All information is only related to apps where reviewer consensus was achieved or those that were reviewed by a single researcher. This demonstrates that most apps did not support tracking of all relevant asthma data, and for those that did, manual data input was the predominant entry method.

Table 8. Number of apps which allowed tracking and displaying of the specified asthma information identified as important from the Global Initiative for Asthma guidelines and by what means this information could be input into the app (n=53).

App's ability to track and display users' asthma information	Apps with reviewer consensus, n (%)	Apps that allow tracking and recording of data, n (%)	Apps with manual data input, n (%)	Apps that allow data entry through external sensors or devices, n (%)	Apps with the ability to create tables or graphs demonstrating trends or analysis of entered data, n (%)
Asthma symptoms	51 (96)	19 (37)	19 (37)	1 (2)	14 (27)
Night waking because of asthma	49 (92)	4 (8)	4 (8)	0 (0)	2 (4)
Activity limitation because of asthma	47 (89)	3 (6)	3 (6)	0 (0)	3 (6)
Peak flow meter values	50 (94)	20 (40)	17 (34)	4 (8)	20 (40)
SABA[a] use	46 (87)	14 (30)	14 (30)	3 (7)	10 (21)
Preventive medication adherence	48 (91)	11 (23)	11 (23)	3 (6)	8 (17)

[a]SABA: short-acting β-agonist.

App's Ability to Provide Prompts or Reminders

Table 9 demonstrates the results of the assessment of whether the apps could provide reminders or prompts on the areas deemed relevant from the review of the GINA guidelines, as outlined in the Methods section. Overall, there were few apps that provided reminders or prompts to users, with only 9 (20%) out of 53 apps providing a reminder to use a person's daily preventive medication. Few apps prompted users to assess the severity and frequency of their asthma symptoms or to seek health advice based on the data they entered.

Table 9. Number of apps that provided reminders and prompts on the specified asthma features chosen based on Global Initiative for Asthma guidelines (n=53).

App provides reminders or prompts on asthma features	Apps with reviewer consensus, n (%)	Apps that provide reminders or prompts, n (%)	Apps where reminders or prompts can be individualized, n (%)
Assessing asthma symptoms for the last month	50 (94)	2 (4)	1 (2)
Appointment with physicians	47 (89)	1 (2)	1 (2)
Performing peak flow test	50 (94)	5 (10)	4 (8)
Preventive medication adherence	45 (85)	9 (20)	8 (18)
Checking the date of expiry and dosage of inhalers	51 (96)	0 (0)	0 (0)
Warning of changing health data (ie, very frequent exacerbations)	53 (100)	0 (0)	0 (0)
Seeking urgent health advice based on the health data the user inputs into the app	52 (98)	2 (4)	2 (4)

Other App Information

Table 10 summarizes the other important features assessed during this review. Only 1 app allowed the user to make an appointment with a physician. Of import, only 5 (10%) of the apps were identified as having an area where users could keep a record of their asthma action plan. Most of these (4 apps) allowed the user to type in their action plan, and 1 app allowed users to upload a photo of their action plan. Only 1 app was identified using recognized asthma screening tools to assess patient's current asthma symptoms and severity.

Table 10. Summary of number of apps containing further features and content identified as important (n=53).

	Reviewer consensus, n (%)	Apps with this feature, n (%)	Apps without this feature, n (%)
Make appointment with physicians	53 (100)	1 (2)	52 (98)
For web-based consultation	0 (0)	0 (0)	53 (100)
For face-to-face consultation	0 (0)	1 (2)	52 (98)
App provides an area for patients to keep record of their asthma action plan	52 (98)	5 (10)	47 (90)
Can type in action plan	0 (0)	4 (8)	48 (92)
Can upload a photo of action plan	0 (0)	1 (2)	51 (98)
The app includes social forums or blogs that promote peer-support and communication among asthma patients	50 (94)	1 (2)	49 (98)
The users could send recorded data to others	44 (83)	11 (25)	33 (75)
To physicians	0 (0)	11 (25)	33 (75)
To each other	0 (0)	9 (21)	33 (75)
The app could help the users to evaluate the risk of having future asthma exacerbations	44 (83)	2 (5)	42 (96)
Using a validated scoring system	0 (0)	0 (0)	44 (100)
Yes, but not using a validated scoring system	0 (0)	2 (5)	42 (96)
The app could guide the users to find out the closest pharmacy, hospital, or clinic	51 (96)	0 (0)	51 (96)
The app uses recognized screening, symptom control and numerical asthma control tools	46 (87)	1 (2)	45 (98)
Yes, all 3	0 (0)	0 (0)	46 (100)
Yes, screening tool	0 (0)	0 (0)	46 (100)
Yes, symptom control tool	0 (0)	0 (0)	46 (100)
Yes, numerical asthma control tools	0 (0)	1 (2)	45 (98)
The app allows users to connect to wearable technology	53 (100)	8 (15)	45 (85)
Smartwatch	0 (0)	0 (0)	53 (100)
Activity sensor (eg, Fitbit)	0 (0)	0 (0)	53 (100)
Smart peak flowmeter	0 (0)	4 (8)	49 (93)
Handheld spirometer	0 (0)	3 (6)	50 (94)
Oximeter	0 (0)	2 (4)	51 (96)
Other	0 (0)	6 (11)	47 (89)

Discussion

Principal Findings

This review aimed to evaluate the quality and functionality of asthma apps and their consistency with international best practice guidelines. We conducted a comprehensive review of 422 asthma apps available on the Australian App Store and Google Play Store, of which we selected 53 apps that met our inclusion and exclusion criteria for detailed analysis. The most common reason for app exclusion was that they were not primarily related to asthma. Most apps were developed by technical companies rather than health care facilities and clinical or research institutes. This lack of involvement of practicing experts in the field of health care is concerning, as these apps are primarily related to the management of chronic health conditions. Apps that involved a clinician during the designing phase demonstrated a greater ability to facilitate behavioral change than those that did not [19]. We believe that in the future, asthma mHealth apps should be developed in consultation with health care professionals and organizations to ensure that they meet an appropriate standard.

Our review revealed that most asthma apps do not use key BCTs that can promote behavioral changes through feedback, goal setting, and rewards. Instead, the most commonly used BCTs were self-monitoring or tracking, education, and advice. This is consistent with other studies and demonstrates a potential area where asthma apps could be improved in the future [19,20]. Functionalities describe what an app can do for its user and are an important marker of whether an app offers much benefit to users. Our review demonstrated that although basic functionalities, such as informing, reading, collecting, instructing, and displaying, were prominent, more complex functionalities were lacking. The ability to evaluate and intervene based on the app information entered is not present

in most apps. Creating apps with this functionality could, for example, guide patients to see their health care provider based on the data they are entering, such as excessive short-acting β-agonist use or asthma symptoms. This feature was demonstrated in the Asthmahub app from NHS Wales, one of the higher-rated apps on assessment.

Our evaluation using the MARS tool showed that the asthma mHealth apps performed poorly in the information and engagement domains compared with the functionality and esthetics categories. These findings are consistent with previous asthma app reviews that showed poorer results in these categories [9]. Only half of the apps achieved an acceptable standard, and even fewer achieved a dual rating average >4, indicating a "good-quality" app. These apps were Kiss myAsthama, Asthmahub, and AioCare Patient. Acceptable or good ratings for subcategories such as app credibility, evidence base, and entertainment were particularly lacking in the apps we assessed. In contrast to this lack of evidence-based content identified through app assessment with the MARS framework, assessment with our checklist found that the knowledge presented in apps was largely based on evidence. "Evidence-based" in the MARS checklist refers to whether an app has undergone a clinical trial. By contrast, in our checklist we refer to "evidence-based" as that the information presented in the app was factual and in alignment with our clinical knowledge of asthma and the GINA guidelines. This explains the contrasting results between the MARS framework and our checklist for "evidence-based" knowledge. Although information was often factual and based on guidelines, the apps had not undergone clinical trials. The subjective star rating provided by reviewers was rarely >3, whereas the mean user rating was 3.56 out of 5, indicating a discrepancy between the perspectives of the reviewers and the user. This discrepancy in user ratings and reviewers' perspectives has been demonstrated in similar app reviews in the past [9]. We propose that this could be because of our health care background bias when assessing the apps, even though we were attempting to assess apps from the perspective of a patient with asthma, or because we were approaching the review with a critical lens following an objective app assessment, a different mindset from the usual user.

Our review demonstrated that asthma mHealth apps do not contain key features consistent with international best practice guidelines for asthma management. Few apps contained important information regarding asthma and asthma management, and even when provided, they were not individualized to the user. Personalization is a key part in the management of asthma, with an individual's triggers, baseline respiratory function, and inhaler device being common things we assess for, educate on, and consider when managing asthma in the community. Furthermore, personalized therapeutic management is a key aspect of asthma management, as outlined in the GINA guidelines [4]. There is also an expectation from previous research that personalized technological interventions may lead to better health outcomes, although this has not yet been specifically demonstrated for asthma mHealth apps [21]. The NHS Wales AsthmaHub app provides one solution to this problem by starting with the creation of a profile in which the user answers questions regarding asthma triggers, control, and medications. This allows the app to provide information targeted toward the user. Skill training in a field can be achieved by apps through written and visual information to teach users how to use peak flow meters, inhalers, and nonpharmacological strategies to manage asthma and asthma exacerbation. These are skills that the GINA guidelines advocate patients to learn and become proficient in, with the support of their health care provider [4]. Training in the use of inhaler devices, spacers, and peak flow meters, all of which in our experience, patients can inadvertently misuse, is lacking in most apps and is something that we feel is key to be included in an asthma mHealth app. Although not essential for all inhaler types, such as dry-powder inhalers, spacers are considered an essential part of using metered-dose inhaler preventer and reliever medications. Therefore, we believe it is reasonable that a high-quality asthma mHealth app should include advice to the users regarding the use of spacers, particularly given the prevalence of salbutamol or albuterol metered-dose inhaler use among those with asthma. No apps could individualize this training to the user, and given the magnitude of devices on the market and the difference in how they work, this represents a deficiency. In total, 40% (n=20) of the apps could record and display peak flow meter values and asthma symptoms. Peak flow is representative of worsening airway obstruction or asthma. If this value is reducing, it can indicate that a person's asthma may be on the precipice of or in the middle of an exacerbation, and as such, it can be a valuable metric for patients and clinicians to monitor. Other key metrics that could be tracked were lacking, such as the amount of short-acting β-agonists and adherence to preventive medications. It is important for these metrics to be recorded, as excessive SABA use can indicate an asthma exacerbation or poorly controlled asthma and should prompt a review by a health care professional. Subtler signs of poor asthma control, such as activity limitation or night waking because of asthma, were rarely recorded and represented missed opportunities. Prompting and reminding users to do something is a basic functionality that is largely underutilized by apps. Chronic asthma management involves remembering to consume daily inhaler medications, assessing symptoms, and regular interactions with a health care provider [4]. All these tasks lend themselves to a reminder from an app to assist with asthma management, and their absence from most apps is a missed opportunity.

Asthma action plans are a key aspect of contemporary patient self-management of asthma [4]. The ability to quickly refer to this plan on a user's digital device should be easy to include; however, most apps lack this feature. We see this as a significant deficiency and missed opportunity. Few apps use validated asthma symptom scoring systems for users to assess their symptoms and risk of exacerbation. Wearable technology and external sensors are a growing medium through which data related to a patient's asthma status can be captured, yet only few apps use this technology. When external sensors are used, they are often smart spirometers (such as in AioCare patient) or peak flow meters, both of which carry extra costs and may not be palatable to all users. No apps were found to be linked to smartwatches or activity sensors, the use of which is becoming more prominent [22]. With the availability of pulse

oximetry in smartwatches, this may be a new method for asthma apps to collect crucial data in the future.

Overall, we determined the quality of apps to be average at best, with many lacking features consistent with international best practice guidelines. Three apps achieved a MARS rating of >4. These were Kissmyasthma, Asthmahub, and AioCare patient. Their alignment of these apps with international best practice guidelines was mixed, with Aiocare patients not having many of the functions deemed important by the guidelines. The Kissmyasthma and Asthmahub apps had great amounts of information related to asthma, consistent with the guidelines. Asthmahub stood alone in having these features and many features to support skill training and track and record information; however, its prompting or reminding functions were minimal. Notably, both these apps that ranked higher in MARS and our checklist were developed by health services or medical research centers (NHS Wales, the University of Sydney, The University of Melbourne, and the Woolcock Institute of Medical Research). All 11 functionalities from the IMS scale were identified in Asthmahub and Aiocare patients, and Kissmyasthma had 6 functionalities, which were above the mean identified. A total of 6 to 8 BCTs were identified in these apps, which, although greater than the median number of BCTs, still did not include several potential BCTs.

Comparison With Previous Work

Prior studies that have conducted these assessments have primarily focused on evaluating the quality and functionality of apps using the MARS framework [9-11]. From a review of the literature for the past 5 years, only 2 prior studies were found to have conducted some sort of assessment of the alignment of apps with asthma self-management principles. Both these studies only looked at free apps, eliminating a number of apps from the review [11,23]. Data collection for both reviews occurred years ago [11,23]. In the rapidly developing marketplace of mobile apps, a number of new apps have been released during this time. Our review examined both free and paid apps, and provided an updated assessment, given that our data collection took place in 2022. Furthermore, Househ et al [23] did not assess apps from the Apple App Store, focusing only on the Google Play Store, and therefore, did not fully assess the breadth of available English-language apps in the marketplace. These authors evaluated whether apps included or did not include information consistent with GINA guidelines as per a checklist created by 1 author [23]. This was limited to asthma information and education and did not include further features, such as the ability of an app to track information, provide asthma skill training, or personalize information. This review also did not examine the overall app quality using the validated MARS framework [23].

Our review benefits from having 2 independent clinicians review the guidelines to establish all GINA self-management recommendations that can be feasibly incorporated into an mHealth app and review app quality using the MARS framework. Furthermore, we examined not only the presence of information, but also the presence or absence of the ability to track asthma symptoms and provide reminders and skill training, as well as all features derived from the GINA guidelines provided in Multimedia Appendix 5. As part of their app review, Tan et al [11] established a framework for assessing the alignment of mHealth apps with the theoretical principles of self-management of allergic rhinitis or asthma [11]. A total of 6 asthma self-management principles were identified based on a literature review and author consensus [11]. Our review has taken this a step further, specifically deriving self-management principles from the international best practice GINA guidelines and creating a more extensive checklist based on these principles. The inclusion of paid apps, the creation of an asthma self-management principle checklist derived from international best practice guidelines, and the up-to-date nature makes our study unique.

Limitations

This study has several limitations. First, we limited it to English-language apps available from 2 app stores in Australia. Although these stores make up a significant portion of the market, other stores do exist, such as the Blackberry store, which were not included in the review. It is also expected that some apps available on the Australian marketplace may not be available on international marketplaces, whereas some apps available internationally may not be available on the Australian marketplace, and thus not included in the review. Although the operating systems for phones used by reviewers were updated on the same day to ensure that the same OS was on each smartphone, the phones themselves were different models. This may have affected the user or reviewer experience to an unknown degree, reducing standardization in this study. A total of 4 apps were reviewed by 1 researcher only as they were removed from the market or were not available to the second reviewer by the time they tried to assess it. This may skew some results, although previous studies have only single-reviewed apps; therefore, the fact that the most apps in this study are double-reviewed is a strength of this research.

Furthermore, we only assessed the presence or absence of those BCTs embedded in the MARS checklist. This limited the number of BCTs that we assessed; however, reviewing the presence or absence of all 93 BCTs was outside the scope of this review and our protocol and requires further research.

This evaluation was researcher-based, and thus does not reflect the real assessment from a patient's perspective. Future research should also include people with asthma to determine their responses to the quality and functionality of these apps.

Conclusions

Currently, there is a lack of high-quality asthma apps aligned with international best practice asthma guidelines. Most apps do not provide patients with important asthma information, skills training in key aspects of asthma management, recording and tracking relevant data, or reminders regarding asthma control. The lack of interaction with smart technology or use of asthma action plans are significant flaws that merit attention in future apps. Few BCTs or in-depth functionalities have been deployed to engage and re-engage with users or generate meaningful behavioral modifications. Again, we see that app designers are typically skilled in providing an esthetically pleasing functional app but lack skills in providing engagement and information, as assessed in the MARS tool. Future app

development should target the key features identified in this study as currently lacking. Furthermore, not only are high-quality asthma mHealth apps lacking, there are minimal robust clinical trials examining the use of these apps and their effect on patient outcomes. Further research in this area will be valuable to determine the true clinical utility of these apps.

A wide spectrum of technological quality, accuracy, and breadth of health information was seen among available apps. Given this spectrum of poor-to high-quality apps in an unregulated market, it is unlikely that future guidelines or health professionals will be able to make generic recommendations to patients regarding asthma mHealth app use and instead will need to make targeted recommendations about specific apps. Guidelines that incorporate reviews such as this review that identify apps known to be of high quality, such as Asthmahub, AioCare patient, and Kissmyasthma, will be an important future resource for general clinicians navigating the ever-expanding mHealth app market for chronic diseases.

Acknowledgments

Funding was received from the Northern Health Respiratory Department (Victoria, Australia) for the costs associated with submitting the protocol and manuscript and purchasing mobile health apps. No funding or sponsorship was received from the mobile app developers or third parties.

Authors' Contributions

BO and KS conceived this study. BR developed the study protocol by adapting a protocol previously created by EG for a diabetes mobile health app review project. KS and BR reviewed the Global Initiative for Asthma guidelines to establish key recommendations that could be feasibly incorporated into a mobile health app. BO and KS provided valuable guidance and advice for this research. BR, DZ, and EP reviewed the apps. BR was the primary author of this manuscript. EG and BR also provided training to all the reviewers of the project. All authors contributed to the refinement of the study protocol and approval of the final manuscript. The Respiratory Research Team at the Northern Hospital assisted in providing general advice to the authors.

Conflicts of Interest

The Northern Health Respiratory Department is providing funding for this study. The head of this department, KS, is a key author of this paper. The department does not have competing or fiduciary interests that could be affected by the results of this study. KS is not responsible for directly reviewing the apps included in the study or for the data analysis.

Multimedia Appendix 1
Completed PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[PDF File (Adobe PDF File), 769 KB - mhealth_v12i1e47295_app1.pdf]

Multimedia Appendix 2
App technical information.
[XLSX File (Microsoft Excel File), 23 KB - mhealth_v12i1e47295_app2.xlsx]

Multimedia Appendix 3
Data from initial app review process to identify apps that met inclusion but not exclusion criteria.
[XLSX File (Microsoft Excel File), 265 KB - mhealth_v12i1e47295_app3.xlsx]

Multimedia Appendix 4
Stepwise guide on the key steps of data extraction and evaluation for reviewers to follow.
[DOCX File , 161 KB - mhealth_v12i1e47295_app4.docx]

Multimedia Appendix 5
Data extraction forms (technical information, Mobile App Rating Scale, Intercontinental Medical Statistics functionality score, and asthma assessment checklist).
[DOCX File , 47 KB - mhealth_v12i1e47295_app5.docx]

References

1. Diagnosing asthma in adults. National Asthma Council. URL: https://www.asthmahandbook.org.au/diagnosis/adults [accessed 2021-01-04]
2. Reddel HK, Taylor DR, Bateman ED, Boulet LP, Boushey HA, Busse WW, et al. An official American Thoracic Society/European Respiratory Society statement: asthma control and exacerbations: standardizing endpoints for clinical

asthma trials and clinical practice. Am J Respir Crit Care Med 2009 Jul 01;180(1):59-99. [doi: 10.1164/rccm.200801-060st] [Medline: 19535666]
3. Pinnock H. Supported self-management for asthma. Breathe (Sheff) 2015 Jun;11(2):98-109 [FREE Full text] [doi: 10.1183/20734735.015614] [Medline: 26306110]
4. Global Strategy for Asthma Management and Prevention. Global Initiative for Asthma. URL: https://ginasthma.org/wp-content/uploads/2020/04/GINA-2020-full-report_-final-_wms.pdf [accessed 2022-01-05]
5. Chronic respiratory conditions. Australian Institute of Health and Welfare, Government of Australia. URL: https://www.aihw.gov.au/reports/chronic-respiratory-conditions/chronic-respiratory-conditions/contents/about [accessed 2022-01-04]
6. Mobile nation 2019: the 5G future. Deloitte. 2019. URL: https://www2.deloitte.com/content/dam/Deloitte/au/Documents/Economics/deloitte-au-economics-mobile-nation-2019-080419.pdf [accessed 2022-01-04]
7. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
8. Terhorst Y, Philippi P, Sander LB, Schultchen D, Paganini S, Bardus M, et al. Validation of the mobile application rating scale (MARS). PLoS One 2020 Nov 2;15(11):e0241480 [FREE Full text] [doi: 10.1371/journal.pone.0241480] [Medline: 33137123]
9. Tinschert P, Jakob R, Barata F, Kramer JN, Kowatsch T. The potential of mobile apps for improving asthma self-management: a review of publicly available and well-adopted asthma apps. JMIR Mhealth Uhealth 2017 Aug 02;5(8):e113 [FREE Full text] [doi: 10.2196/mhealth.7177] [Medline: 28768606]
10. Ramsey RR, Caromody JK, Voorhees SE, Warning A, Cushing CC, Guilbert TW, et al. A systematic evaluation of asthma management apps examining behavior change techniques. J Allergy Clin Immunol Pract 2019 Nov;7(8):2583-2591 [FREE Full text] [doi: 10.1016/j.jaip.2019.03.041] [Medline: 30954644]
11. Tan R, Cvetkovski B, Kritikos V, O'Hehir RE, Lourenço O, Bousquet J, et al. Identifying an effective mobile health application for the self-management of allergic rhinitis and asthma in Australia. J Asthma 2020 Oct 24;57(10):1128-1139. [doi: 10.1080/02770903.2019.1640728] [Medline: 31274044]
12. Robinson B, Gong E, Oldenburg B, See K. Functionality and quality of asthma mHealth apps and their consistency with international guidelines: protocol for a systematic search and evaluation of mobile apps. JMIR Res Protoc 2022 Feb 09;11(2):e33103. [doi: 10.2196/33103] [Medline: 35138257]
13. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Syst Rev 2021 Mar 29;10(1):89 [FREE Full text] [doi: 10.1186/s13643-021-01626-4] [Medline: 33781348]
14. Guan Z, Sun L, Xiao Q, Wang Y. Constructing an assessment framework for the quality of asthma smartphone applications. BMC Med Inform Decis Mak 2019 Oct 15;19(1):192. [doi: 10.1186/s12911-019-0923-8] [Medline: 31615493]
15. Gong E, Zhang Z, Jin X, Liu Y, Zhong L, Wu Y, et al. Quality, functionality, and features of Chinese mobile apps for diabetes self-management: systematic search and evaluation of mobile apps. JMIR Mhealth Uhealth 2020 Apr 07;8(4):e14836. [doi: 10.2196/14836] [Medline: 32255432]
16. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95 [FREE Full text] [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
17. Papi A, Brightling C, Pedersen SE, Reddel HK. Asthma. Lancet 2018 Feb 24;391(10122):783-800. [doi: 10.1016/S0140-6736(17)33311-1] [Medline: 29273246]
18. Aitken M, Gauntlett C. Patient apps for improved healthcare: from novelty to mainstream. IMS Institute for Healthcare Informatics. URL: http://ignacioriesgo.es/wp-content/uploads/2014/03/iihi_patient_apps_report_editora_39_2_1.pdf [accessed 2022-01-04]
19. Fitzgerald M, McClelland T. What makes a mobile app successful in supporting health behaviour change? Health Educ J 2017;76(3):373-381 [FREE Full text] [doi: 10.1177/0017896916681179]
20. Edwards EA, Lumsden J, Rivas C, Steed L, Edwards LA, Thiyagarajan A, et al. Gamification for health promotion: systematic review of behaviour change techniques in smartphone apps. BMJ Open 2016 Oct 04;6(10):e012447 [FREE Full text] [doi: 10.1136/bmjopen-2016-012447] [Medline: 27707829]
21. Lustria ML, Noar SM, Cortese J, Van Stee SK, Glueckauf RL, Lee J. A meta-analysis of web-delivered tailored health behavior change interventions. J Health Commun 2013;18(9):1039-1069. [doi: 10.1080/10810730.2013.768727] [Medline: 23750972]
22. Corbett P, Schwager E. Digital consumer trends- touch-less, virtual, health-conscious and eco-wise. Deloitte. 2021. URL: https://www2.deloitte.com/au/en/blog/technology-media-telecommunications-blog/2021/digital-consumer-trends-touch-less-healthier-wiser.html [accessed 2021-12-09]
23. Househ M, Hossain N, Jamal A, Zakaria N, Elmetwally A, Alsalamah M, et al. A cross-sectional content analysis of android applications for asthma. Health Informatics J 2017 Jun 03;23(2):83-95 [FREE Full text] [doi: 10.1177/1460458215627289] [Medline: 26940137]

Abbreviations

BCT: behavioral change technique
GINA: Global Initiative for Asthma
IMS: Intercontinental Medical Statistics Institute for Health Informatics
MARS: Mobile App Rating Scale
mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by L Buis; submitted 14.03.23; peer-reviewed by Y Zhang, R Almeida, JA Fonseca, PM Matricardi; comments to author 18.07.23; revised version received 21.10.23; accepted 21.11.23; published 10.01.24.

Please cite as:
Robinson B, Proimos E, Zou D, Gong E, Oldenburg B, See K
Functionality and Quality of Asthma mHealth Apps and Their Consistency With International Guidelines: Structured Search and Evaluation
JMIR Mhealth Uhealth 2024;12:e47295
URL: https://mhealth.jmir.org/2024/1/e47295
doi:10.2196/47295
PMID:38198204

Review

Developing a Comprehensive List of Criteria to Evaluate the Characteristics and Quality of eHealth Smartphone Apps: Systematic Review

Janette Ribaut[1,2], PhD, RN; Annette DeVito Dabbs[3,4], PhD, RN; Fabienne Dobbels[1,5], PhD; Alexandra Teynor[6], PhD, Dipl-Ing; Elisabeth Veronica Mess[6], MA; Theresa Hoffmann[1,7], MSc, RN; Sabina De Geest[1,5], PhD, RN

[1]Institute of Nursing Science, Department Public Health, University of Basel, Basel, Switzerland

[2]Department of Hematology, University Hospital Basel, Basel, Switzerland

[3]School of Nursing, Department of Acute & Tertiary Care, University of Pittsburgh, Pittsburgh, PA, United States

[4]Clinical Translational Science Institute, University of Pittsburgh, Pittsburgh, PA, United States

[5]Academic Center for Nursing and Midwifery, Department of Public Health and Primary Care, KU Leuven, Leuven, Belgium

[6]Department of Computer Science, University of Applied Sciences, Augsburg, Germany

[7]Department Pflege und Betreuung, Genossenschaft Alterszentrum Kreuzlingen, Kreuzlingen, Switzerland

Corresponding Author:
Sabina De Geest, PhD, RN
Institute of Nursing Science
Department Public Health
University of Basel
Bernoullistrasse 28
Basel, 4056
Switzerland
Phone: 41 61 207 09 16
Fax: 41 61 207 09 55
Email: sabina.degeest@unibas.ch

Abstract

Background: The field of eHealth is growing rapidly and chaotically. Health care professionals need guidance on reviewing and assessing health-related smartphone apps to propose appropriate ones to their patients. However, to date, no framework or evaluation tool fulfills this purpose.

Objective: Before developing a tool to help health care professionals assess and recommend apps to their patients, we aimed to create an overview of published criteria to describe and evaluate health apps.

Methods: We conducted a systematic review to identify existing criteria for eHealth smartphone app evaluation. Relevant databases and trial registers were queried for articles. Articles were included that (1) described tools, guidelines, dimensions, or criteria to evaluate apps, (2) were available in full text, and (3) were written in English, French, German, Italian, Portuguese, or Spanish. We proposed a conceptual framework for app evaluation based on the dimensions reported in the selected articles. This was revised iteratively in discussion rounds with international stakeholders. The conceptual framework was used to synthesize the reported evaluation criteria. The list of criteria was discussed and refined by the research team.

Results: Screening of 1258 articles yielded 128 (10.17%) that met the inclusion criteria. Of these 128 articles, 30 (23.4%) reported the use of self-developed criteria and described their development processes incompletely. Although 43 evaluation instruments were used only once, 6 were used in multiple studies. Most articles (83/128, 64.8%) did not report following theoretical guidelines; those that did noted 37 theoretical frameworks. On the basis of the selected articles, we proposed a conceptual framework to explore 6 app evaluation dimensions: *context*, *stakeholder involvement*, *features and requirements*, *development processes*, *implementation*, and *evaluation*. After standardizing the definitions, we identified 205 distinct criteria. Through consensus, the research team relabeled 12 of these and added 11 more—mainly related to ethical, legal, and social aspects—resulting in 216 evaluation criteria. No criteria had to be moved between dimensions.

Conclusions: This study provides a comprehensive overview of criteria currently used in clinical practice to describe and evaluate apps. This is necessary as no reviewed criteria sets were inclusive, and none included consistent definitions and terminology.

Although the resulting overview is impractical for use in clinical practice in its current form, it confirms the need to craft it into a purpose-built, theory-driven tool. Therefore, in a subsequent step, based on our current criteria set, we plan to construct an app evaluation tool with 2 parts: a short section (including 1-3 questions/dimension) to quickly disqualify clearly unsuitable apps and a longer one to investigate more likely candidates in closer detail. We will use a Delphi consensus-building process and develop a user manual to prepare for this undertaking.

Trial Registration: PROSPERO International Prospective Register of Systematic Reviews CRD42021227064; https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42021227064

(JMIR Mhealth Uhealth 2024;12:e48625) doi:10.2196/48625

KEYWORDS

telemedicine; smartphone; mobile apps; program evaluation; decision-making; systematic review; mobile phone

Introduction

Background

eHealth, that is, "the use of information and communication technology for health" [1], can support the delivery of interventions for self-management support and behavior change in patients with acute and chronic illnesses [2,3]. According to the World Health Organization (WHO) [4], self-care health interventions can be classified into individual agency (eg, promoting awareness about self-care), health information seeking (eg, education for informed health decision-making), social and community support (eg, peer mentorship and counseling), personal health tracking (eg, home-based records for health and diagnostic data), self-diagnosis of health conditions (eg, self-testing), self-management of health (eg, self-medication or treatment), individuals' links to their health systems (eg, individuals sharing data with health care professionals [HCPs]), and individuals' financial outlays for health (eg, expenses for prescription medicines). However, a recent evaluation of self-care interventions delivered via eHealth apps noted that only 20% of the 100 included apps used evidence-based information, whereas experienced HCPs considered only 32% to be useful and deemed 52% to be misleading and 11% to be dangerous [5].

Searching for a common characteristic linking the most effective apps, several systematic reviews and meta-analyses have found that those developed on firm theoretical foundations are more likely to be effective [6,7]. However, a systematic review of health-promoting smartphone apps found that only 55.6% of the included 27 studies described a theoretical basis for their smartphone app development [8]. A 2018 review found that only 8 (1.2%) of 681 smartphone apps to support medication adherence had documented evidence of their effectiveness. Such evidence is health care systems' main consideration regarding certification and reimbursement [9]. Furthermore, although a user-centered design (sometimes also called human-centered design) and the involvement of patients and HCPs in the development of smartphone apps is known to provide insights into end users' needs and helps ensure both relevant, reliable content and high quality [9,10], only 84 (12.3%) of the apps in this review had been developed in collaboration with HCPs. None reported patient involvement in their development processes [9].

Owing to the increasing availability of eHealth smartphone apps, it is vitally important and increasingly challenging for HCPs to identify, evaluate, and recommend relevant, trustworthy, and high-quality eHealth smartphone apps [11-14]. One tempting way for HCPs to form a first idea of an eHealth smartphone app's quality is the star ratings and written reviews it receives on an app store [15]. However, this information is often subjective and distorted by individuals, comes from unverified or fraudulent sources, or provides no insights on an eHealth app's quality [15-18]. HCPs also face a lack of reliable guidance on evaluating eHealth smartphone apps' applicability to clinical practice [13,19,20]. Therefore, many are now struggling to describe and evaluate eHealth smartphone apps. A guideline regarding their characteristics and quality using standardized methods that will allow HCPs to propose reliable apps to their patients is needed [21,22].

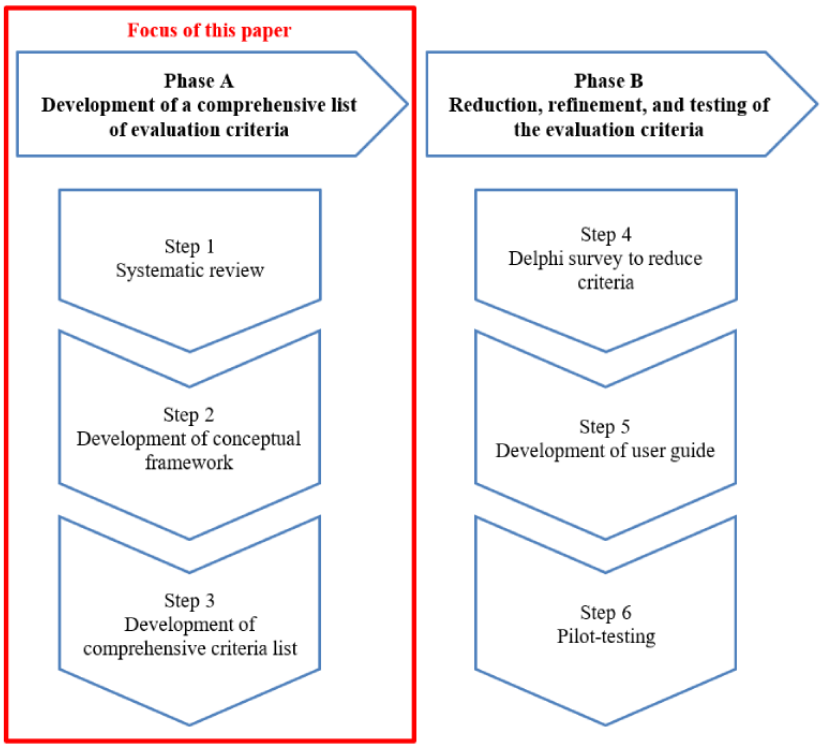
Previous efforts to evaluate apps have generally focused on guidance for researchers [23-25]. Although the criteria were often overly complex or tailored to specific health areas, they also tended to be incomplete. Their underlying theories, scientific rationales, and development processes have rarely been described [20,25-27]. Furthermore, their unsuitable foci, nontransparent development processes, complexity, and often excessive time demands make them a poor fit for clinical practice. Finally, the existing instruments used a variety of criteria that only partially overlapped [20,23-27]. A clear description and evaluation of an app is important as, in rapidly evolving fields, even small changes or improvements to an app can have significant impacts on its use and usefulness [28]. To date, evaluation tools to help clinicians describe and evaluate eHealth apps, allowing them to recommend high-quality apps to their patients and share their thoughts using common terminology, are lacking [29,30].

Objectives

Therefore, the aim of this study was to obtain an overview of the evaluation criteria used in the literature. This process was conducted in three steps: (1) conducting a systematic review to identify existing criteria for evaluating eHealth apps, (2) developing a conceptual framework for the evaluation of eHealth smartphone apps, and (3) developing a comprehensive list of criteria for describing and evaluating eHealth smartphone apps. This was the foundational phase 1 of an overarching project to develop and pilot-test a theory-based tool to help HCPs evaluate the characteristics and quality of eHealth smartphone apps in a practical and standardized way (Figure 1). Phase 2 will involve

narrowing down, refining, and testing the evaluation criteria via 3 further steps: conducting a Delphi survey to narrow down the criteria (step 4), developing a user guide for the processes of description and evaluation (step 5), and pilot-testing the user guide and processes with HCPs (step 6). This paper focuses on reporting on the foundational phase 1 and outlining the proposed steps for phase 2.

Figure 1. Overview of the 2 phases and steps in the development of the eHealth smartphone app evaluation tool (the focus of this paper is framed on the left side).



Methods

Design

We used a 3-step descriptive, iterative, and developmental approach in phase 1. We first conducted a systematic review, then iteratively developed a new conceptual framework, and finished by compiling a comprehensive list of criteria for the evaluation of eHealth apps. The methodology for each of these steps is described in detail in the following sections. As this study did not deal with human participants or identifiable data, no ethics approval was needed.

Step 1: Systematic Review

To identify existing criteria for evaluating eHealth apps, we conducted a systematic review complying with the *Cochrane Handbook for Systematic Reviews of Interventions* [31]. The manuscript was written following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [32].

Protocol and Registration

Our review was registered in PROSPERO (CRD42021227064 [33]). No other protocol has been published.

Eligibility Criteria

Our systematic review included studies on any health condition fulfilling the following inclusion criteria: all had to be primary studies or reviews (1) explicitly describing tools, guidelines, dimensions, instruments, criteria or items, or development processes for tools to evaluate eHealth smartphone apps; (2) clearly describing the evaluation of interventions delivered via eHealth smartphone apps according to predefined (self-developed or existing) criteria; (3) having full-text versions available; and (4) being available in English, French, German, Italian, Portuguese, or Spanish. We incorporated criteria encompassing a broad spectrum of health care areas, including health promotion, prevention, and both physical and mental health. Recognizing the interconnected nature of physical and mental health and the diverse purposes and user groups for which eHealth is used, we integrated a wide array of objectives and stakeholders into our evaluation. Although these areas may exhibit distinct characteristics, it is conceivable that there are fundamental criteria that could be consistently applied in evaluating eHealth apps across different domains. These fundamental criteria may encompass aspects such as user-friendliness, data security, privacy, and usability, forming a shared foundation for evaluation to ensure that essential quality aspects are addressed. In summary, our systematic review encompasses these comprehensive topics to provide a thorough evaluation of eHealth quality criteria, which are applicable to diverse health care needs. Our aim was to encourage consistency and standardization in the evaluation process. Articles were excluded if they (1) described criteria to evaluate interventions delivered via eHealth websites and videos, among other media, but not smartphone apps; and (2) were study protocols, conference abstracts, editorials, or letters to the editor.

Information Sources

We queried the MEDLINE (OvidSP), CENTRAL (via Cochrane), CINAHL (EBSCOhost), and Web of Science databases. Supplementary searches were conducted on trial registries (ClinicalTrials.gov and WHO trial registry) and the reference lists of the included papers. The search was conducted on December 5, 2022. No time restrictions were imposed.

Search Strategy

We developed our MEDLINE search string based on the terms used in articles on (partially) similar topics [34-41] combined with key Medical Subject Heading and free-text terms (see the search strategy in Multimedia Appendix 1). For the other databases, we adapted the search string accordingly. We combined thematic blocks with various keywords related to *eHealth*, *smartphone*, *application*, *evaluation*, and *tool*. No filters were applied.

Selection Process

All identified titles and abstracts were independently screened for relevance by 2 reviewers (JR and TH). The full texts were assessed by the same reviewers using the criteria described previously. The reasons for full-text exclusion were reported. In one case of disagreement, an independent third researcher (SDG) contributed to help reach a consensus.

Data Collection Process and Data Items

In total, 2 reviewers (JR and TH) independently extracted the data and cross-checked their results. We extracted information on the author, year, country, research question or study aim, design, operating system, population or specific condition, main intended intervention purpose, name of the tool, and framework or theoretical guidance. The intended purpose of each eHealth app–delivered intervention was categorized according to the WHO classification for self-care health interventions [4]: individual agency, health information seeking, social and community support, personal health tracking, self-diagnosis of health conditions, self-management of health, individuals' links to their health systems, and individuals' financial outlays for health. Specific eHealth app quality evaluation dimensions or criteria were extracted and tabulated in a separate table.

Study Assessment

The included studies were assessed using the Appraisal of Guidelines for Research and Evaluation–II (AGREE-II) instrument [42], which is widely used to evaluate guideline development processes. As many of the included studies did not a priori intend to develop an evaluation guideline, this instrument might not have been the best option for all studies. However, as we were mainly interested in the justification and development of the dimensions or criteria used in the studies, this instrument provided us with the best support for evaluating these aspects. The AGREE-II instrument consists of 23 items classified into 6 domains (3 items on scope and purpose; 3 on stakeholder involvement; 8 on rigor of development; 3 on clarity of presentation; 4 on applicability; and 2 on editorial independence, ie, whether funding body and competing interests were reported). We concluded our AGREE-II assessment by rating the degree to which each included study described each domain. For this, we used a 4-descriptor scale: *accurately* (all AGREE-II items fulfilled), *partially* (two-thirds of all AGREE-II items fulfilled), *hardly* (one-third of all AGREE-II items fulfilled), and *not at all* (0 AGREE-II items fulfilled).

Step 2: Development of a Conceptual Framework

The original dimensions of the frameworks reported in the selected articles were listed in a table. Similar descriptions of dimensions were merged. The first draft of the proposed conceptual framework and graphical representations was reviewed and discussed with various stakeholders (researchers, clinicians, designers, and software developers with backgrounds in nursing, medicine, ethics, and informatics). During these discussions, the participants recommended that we distinguish between technical dimensions (eg, design, usability, security, safety, and privacy) and those that focused on content (eg, evidence base and scientific evaluation). It was also recommended that the dimensions be presented as a linear, step-by-step process. During these meetings, the first author (JR) took notes and recorded the proposed changes until consensus was reached on the next version.

The second draft of the framework was discussed with 18 international volunteers (patient representatives, researchers, clinicians, and technology developers) from diverse backgrounds in health care (eg, psychology, nursing, and pharmacy) who were participating in a public webinar on quality evaluation of eHealth technology. As it was a public webinar, only limited data on participant demographics were collected (Table S1 in Multimedia Appendix 2). A survey via AhaSlides (AhaSlides Pte Ltd) to rate the importance and clarity of the dimensions and subgroups (1=not important; 5=very important) and open-ended questions to add missing dimensions or subgroups were used to engage with the participants. In addition, the participants were engaged to add comments orally that the first author (JR) put in writing. The quantitative data were analyzed descriptively (eg, frequency and mean), whereas the qualitative data were analyzed using the mind-mapping technique. The participants found the *technology* dimension too large (ie, covering too many subtopics) and partially unclear. Therefore, it was recommended to split this dimension into technological concerns (eg, technical requirements, security, safety, and privacy) and functional requirements (eg, the user-centeredness and usability of the design). In addition, they understood eHealth evaluation as a cyclical process and recommended presenting the conceptual framework as a continuous circuit as opposed to the initially linear process recommended previously.

Their feedback was used to draft a third version of the framework, which was presented and discussed with 34 researchers, clinicians, and technology developers (Table S2 in Multimedia Appendix 2; only 1 person overlaps with the volunteers from the webinar) mainly with a background in pharmacy who participated in the Next Chapter in Patient Care Conference in April 2022 in Pärnu, Estonia. This time, a survey via Mentimeter (Mentimeter AB) was used to rate the importance and clarity of the dimensions and subgroups (1=not important; 5=very important), and open-ended questions were used to add missing dimensions or subgroups. In addition, the participants were engaged to add comments orally while the

first author (JR) took notes. Quantitative data were analyzed descriptively, whereas qualitative data were analyzed using the mind-mapping technique. The Next Chapter in Patient Care participants recommended highlighting the overarching nature of the ethical, legal, and social aspects, which must be considered in all eHealth smartphone app evaluation dimensions. This resulted in a fourth version reflecting the general character of relevant ethical, legal, and social considerations.

Subsequent rounds of discussion and feedback with the research team focused on the scope and relationships between the dimensions. The participants highlighted the legal, ethical, and social aspects to be treated as part of the context. In addition, they agreed that stakeholder involvement should be seen as another overarching aspect that is important in all dimensions of eHealth app evaluation. This discussion resulted in the final version of our new conceptual framework for evaluating eHealth apps: the eHealth Smartphone App Evaluation (eHAPPI) framework.

Step 3: Development of a Comprehensive Criteria List

The eHAPPI framework was then used to synthesize all the eHealth smartphone app evaluation criteria identified in the selected studies. The redundant criteria were combined. The classification of the criteria according to the eHAPPI dimensions and suggestions for changes and regarding additional or irrelevant criteria were discussed and refined by the research team according to consensus.

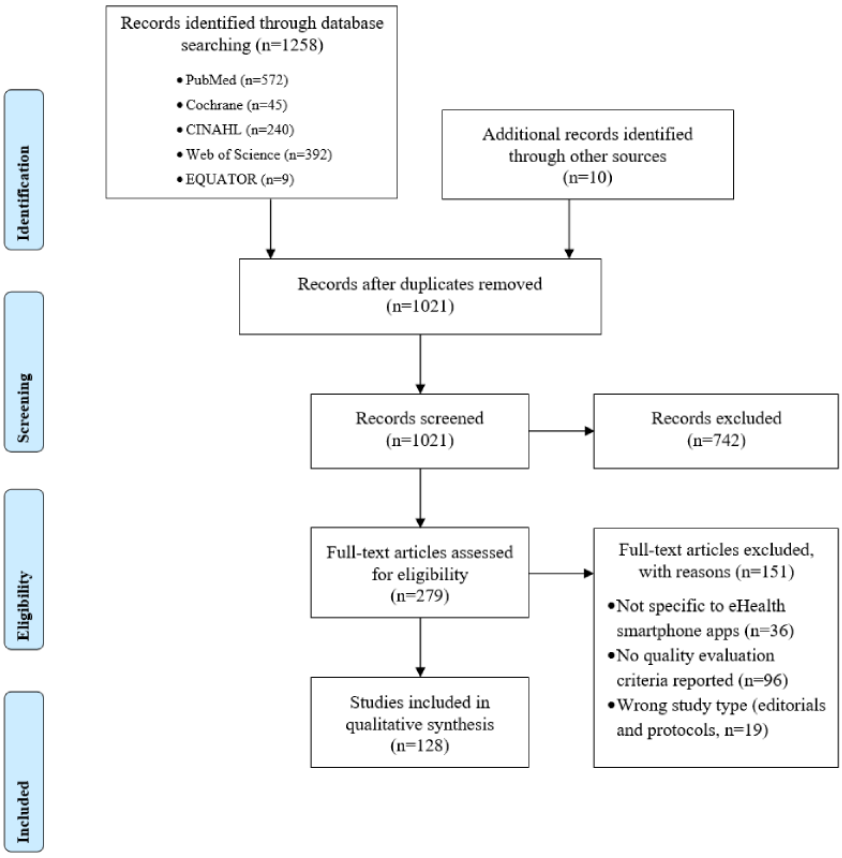
Results

Step 1: Systematic Review

Study Selection

The results of our study selection process are presented in the PRISMA flow diagram (Figure 2). The search strategy described previously yielded 1021 nonduplicate titles. After screening of the titles and abstracts as well as full-text assessment for eligibility, our final analysis included 128 articles that met all the inclusion criteria (Multimedia Appendix 3 [43]).

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



Study Characteristics

General Description

A detailed description of the characteristics of the included studies can be found in Multimedia Appendix 4. The years of publication ranged from 2013 to 2022. All but 1.6% (2/128) of the articles were written in English—1 was in French [44] and the other in Spanish [45]. Four-fifths of the studies (101/128, 78.9%) were conducted in Western Europe, North America, or Australia. The studies used a variety of designs, mainly cross-sectional reviews of existing apps (42/128, 32.8%); reports on various app development and evaluation approaches (29/128, 22.7%); quantitative, qualitative cross-sectional, or longitudinal user testing of a single app (27/128, 21.1%); or different forms of reviews (15/128, 11.7%). The 128 included studies involved apps covering 30 topics, such as mental health, health promotion, or support for specific physical conditions (eg, heart disease and diabetes). However, only 50.8% (65/128) of the studies provided enough detail to categorize them according to

the WHO classification for self-care interventions for health and well-being [4].

Evaluation Tools

In total, the included studies used 142 distinct tools to evaluate eHealth apps. Although 76.6% (98/128) of the included studies used a single evaluation tool, 23.4% (30/128) used multiple tools and scales. Almost one-quarter of the studies (30/128, 23.4%) used an evaluation tool with investigator-developed criteria and then provided only scantily described development processes or theoretical backgrounds for those criteria. The most frequently used tool was the Mobile App Rating Scale (33/128, 25.8%) followed by its adapted versions (8/128, 6.3%) and the System Usability Scale (22/128, 17.2%). Less frequently used tools were the Post-Study System Usability Questionnaire (2/128, 1.6%), the Questionnaire for User Interaction Satisfaction (2/128, 1.6%), and the quality of experience survey (2/128, 1.6%). A total of 43 other tools were used in only 0.8% (1/128) of the studies each. In total, 10.2% (13/128) of the studies used qualitative methods (eg, interviews and focus groups) to generate the app evaluation criteria. In some cases (4/128, 3.1%), the origin of the criteria was unclear, or similar names were used for different tools.

Theoretical Frameworks

Most studies (83/128, 64.8%) did not clearly report a theoretical underpinning. The 32% (41/128) that did used 59 different frameworks, including various non–eHealth-specific behavioral, social, or implementation theories (10/59, 17%), the technology acceptance model (7/59, 12%), heuristic evaluation (5/59, 8%), models of the International Organization for Standardization (3/59, 5%), the (extended) Unified Theory of Acceptance and Use of Technology (3/59, 5%), or user-centered design (2/59, 3%). In total, 29 frameworks were used in only 2% (1/59) of the studies each. Of the 59 frameworks used in the included studies, 16 (27%) guided the development and 43 (73%) guided the evaluation of eHealth smartphone apps. In 1.6% (2/128) of the studies, different frameworks were used for development and evaluation.

Study Assessment

On average, the studies described 8.6 (SD 2.4; range 3-15) of the AGREE-II instrument's 23 items [42]. Few studies described the scientific or theoretical basis and development processes of the app evaluation criteria that they applied. The most completely described or justified domains were related to *scope and purpose* (116/128, 90.6% described it accurately; 11/128, 8.6% described it partially; and 1/128, 0.8% hardly described it), *editorial independence* (91/128, 71.1% described it accurately; 30/128, 23.4% described it partially; and 7/128, 5.5% did not describe it at all), and *stakeholder involvement* (13/128, 10.2% described it accurately; 42/128, 32.8% described it partially; 70/128, 54.7% hardly described it; and 3/128, 2.3% did not describe it at all). The least fulfilled domains were *applicability* (1/128, 0.8% described it accurately; 2/128, 1.6% described it partially; 22/128, 17.2% hardly described it; and 103/128, 80.5% did not describe it at all), *rigor of development* (13/128, 10.2% described it partially; 53/128, 41.4% hardly described it; and 62/128, 48.4% did not describe it at all), and *clarity of presentation* (14/128, 10.9% described it accurately; 32/128, 25% described it partially; 52/128, 40.6% hardly described it; and 30/128, 23.4% did not describe it at all).

Step 2: Development of a Conceptual Framework

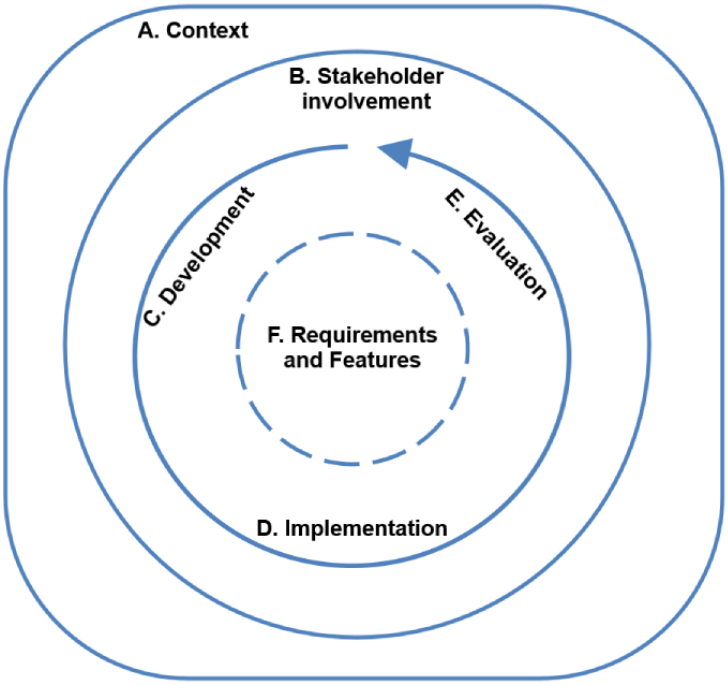
The full list of frameworks and original eHealth evaluation dimensions identified in the selected studies can be found in Multimedia Appendix 5 [43,45-47]. Some dimensions were included in only a few frameworks, and no framework included all possible dimensions.

The condensed dimensions were presented graphically and refined iteratively with the stakeholders until consensus was reached and no further adaptions were needed. The final eHAPPI conceptual framework (Figure 3) consists of six interrelated dimensions: (1) *context*, (2) *stakeholder involvement*, (3) *development processes*, (4) *implementation*, (5) *evaluation*, and (6) *features and requirements*.

A detailed definition of each dimension, including the subgroups, is presented in Textbox 1. *Context* describes a set of unique factors and conditions in which the app will be implemented [48]. This is an overarching dimension that depends on and, in turn, influences the other domains. *Stakeholder involvement* is essential in all aspects of eHealth. It involves the active engagement of relevant partners in all processes of the app life cycle, from conceptualization to sustainable implementation (eg, with end users, HCPs, researchers, and health insurers) [49,50].

Several subgroups were defined to further outline and structure the framework: *basic information* (concerning the app) and *ethical, legal, and social aspects* were seen as relevant subgroups of the *context* dimension. The *featuresandrequirements* dimension was assigned 4 subgroups: *evidence-based content*; *functionality*; *usability*, *privacy*, *and security*; and *performance.* Similarly, the *development process* dimension was divided into *cocreation/user-centered design* and *characteristics of the development team*, and *adoption (integration into daily life)* and *maintenance* were seen as relevant subgroups of the *implementation* dimension. Finally, the *evaluation* dimension included only 1 subgroup: *scientific evaluation*.

Figure 3. The final eHealth Smartphone App Evaluation framework.



Textbox 1. Overview of framework dimensions, subgroups, and definitions.

Context

- *Context* describes the set of unique factors and conditions in which the app is implemented. During the implementation process, the app, implementation, and context interact, influencing, modifying, facilitating, or hindering one another. As an overarching concept, context encompasses a physical location and roles, relationships, and interactions at various levels (ethical, legal, social, geographical, epidemiological, and political) [48,51].
- *Basic information* describes general information concerning the app (eg, name, URL, or available languages).
- *Ethical, legal, and social aspects* describe activities to help evaluators consider ethical, legislative, and social implications of an app's development and implementation. These aspects cannot be completely separated from each other as they overlap somewhat [28].

  Ethics describes activities to understand and explore *the moral life*, wherein *morality* includes beliefs, norms of behaviors, principles, and rules that guide individual and institutional behavior. Morality is a widely shared set of norms that result from a certain social consensus. Ethical aspects include the prevailing moral norms and values during the development and implementation of apps. Although moral values are influenced by cultural, sociopolitical, religious, and economic differences, many ethical reflections are shared by all countries and societies. Regarding apps, important ethical topics include *benefit-harm balance*, *autonomy, respect for persons*, and *justice and equity and legislation* (when using the app, ethical challenges may arise that were not considered in existing legislation and regulations) [28].

  Legal aspects include rules and regulations that must be considered when developing and implementing apps. Rules and regulations are designed to protect the rights and interests of the patients and other members of society (eg, legislation on patients' rights; data protection laws; or the regulations, rights, and obligations of health care professionals [HCPs] in general). Important legal topics are, for example, *autonomy of the patient* (eg, legal requirements), *privacy of the patient* (eg, no use of "unnecessary" patient information), *equality in health care*, *ethical aspects* (eg, impact on achievement of human rights), *authorization and safety* (eg, laws and rules regarding safety), *ownership and liability*, and *regulation of the market* [28].

  Social aspects describe human-centered activities that approach end users and their social networks as reference points in an app's development and implementation. These include groups of patients or individuals who might require special consideration (eg, vulnerable populations, people living in remote communities, people with learning difficulties, older people, ethnic minority individuals, and immigrants). Patients', individuals', and relatives' perspectives should be considered when developing and implementing apps. Some social groups may be important for a particular app (eg, ethnic minority individuals and individuals with disabilities), which should be specified. Important social topics include *patients' perspectives* (eg, expectations and wishes), *social group aspects* (accessibility), and *communication aspects* (eg, explanation of treatment choices) [28]

Stakeholder involvement

- *Stakeholder involvement* describes the active engagement of relevant partners in all processes of the smartphone app life cycle, from conceptualization to sustainable implementation (eg, end users, HCPs, researchers, and health insurers [49,50].

  Note: for end-user involvement in the development process, see the *Cocreation/user-centered design* section under the *Development process* subheading in this textbox).

Development process

- *Development process* describes all activities performed regarding the app's creation, such as activities related to its aim, target population, guiding principles, applied approaches, stakeholder contributions, required changes for subgroups, or continuing uncertainties [52-54].
- Cocreation of technologies or user-centered design: *cocreation* describes a stepwise process in which partners (eg, patients, caregivers, and HCPs) are actively involved in the strategic design and planning of the app's development and implementation at every stage of the process [55].

  User-centered design is a multidisciplinary design approach using cyclic iteration and evaluation. As it actively involves users to improve the developers' understanding of their requirements and wishes, this approach is seen as essential regarding product benefits and usability [56].
- *Characteristics of the development team* describes information about the multidisciplinary group responsible for the creation of the working, validated app. Important characteristics of the development team include the name of the app provider (developer and manufacturer of technology and content), contact details, organization attributes, and funding, as well as any conflicts of interest [57].

Evaluation

- *Evaluation* describes the assessment of the app's efficacy, effectiveness, cost-effectiveness, safety, implementation, and impact [58].
- *Scientific evaluation* is the systematic assessment of the app's efficacy, cost-effectiveness, safety, implementation, and impact through observation, measurement, and experimentation in a scientific study. Such an evaluation is essential to reliably measure an app's effects and outcomes as a basis for decision-making [58-60].

Implementation

- *Implementation* describes the uptake and sustainable integration of evidence-based innovations such as apps into routine use [61].
- *Adoption* (ie, integration into daily life) deals with the app's uptake (ie, activities focused on using the app in everyday life). This includes the user's reaction to the app with respect to desired activities and interactions, such as downloads, clicks, and data entries. User engagement, that is, the user's investment in learning about and participating in the app (eg, time and memory load), is an important prerequisite for the app's long-term integration into daily life [62-64].

- *Maintenance* covers activities that ensure that an app is sustained in an acceptable and safe condition to perform its specified functions. Maintenance also includes repair as a stand-alone action to restore deteriorated or damaged parts. The activities needed to ensure the required performance of the app during its lifetime include updates, performance evaluations, and the planning and execution of necessary repairs [65].

Features and requirements

- *Features and requirements* describe the app and the characteristics or attributes it must provide to meet the users' needs (ie, its *information content*, *functionality*, *usability*, *security*, *privacy*, and *performance*). The indications, performance features, and app options may differ for different generations or versions of an app. Important *features and requirements* topics include app description, claimed benefits, care level of use, reference values, or cutoff points used [28, 66]. A clear description of the app is important as, in rapidly evolving fields, even small changes or improvements to an app can have large impacts on its use and benefits. The description should enable those who are not familiar with the app to understand what it does, how it works, and how it can be used..
- *Evidence-based content* describes information, subject matter, and data content made available by the app based on the best available scientific knowledge and clinical expertise [67].
- *Functionality* describes the properties or functional requirements that affect the app's use (ie, its *features*, *components*, and *usefulness*). In the context of self-monitoring, there are specific requirements related to design issues, particularly those concerning wearable technologies, as well as requirements regarding the behavioral aspects that are part of the design. These requirements need to be considered when evaluating a product or system for self-monitoring. The *functionality* information is intended to summarize the app's overall suitability for use in a particular situation [38,68,69].
- *Usability*, *privacy*, and *security* are nonfunctional requirements (ie, they deal with how the system should operate in terms of usability, security, privacy, design, modularity, modifiability, reliability, availability, portability, and operability).

  Usability describes the extent to which an app can be used by its target users to achieve its intended goals *effectively*, *efficiently*, and *satisfactorily*. Ideally, an app should be *easy to use*, *easy to learn*, and *easy to understand* [70].

  Privacy describes the protection of and control over personal data during the app's processing operations [71].

  Security describes the provision of safeguards that serve the security of the individual or community. For example, this helps prevent or avoid poverty, hardship, theft, or espionage. Security is a major component of a stable, relatively predictable environment in which people can pursue their goals without interference, harm, or the fear of them. This includes protecting the app and associated data from events that could cause loss or severe damage, such as fire, burglary, theft, or vandalism [72,73].
- *Performance* describes whether an app works quickly and without errors and does not cause problems. Important performance topics include reliability and scalability (ie, whether an app still works properly when the number of users increases [74,75]).

Step 3: Development of a Comprehensive Criteria List

In total, 205 criteria for describing and evaluating eHealth smartphone apps were reported in the selected articles. None of the articles included all the possible criteria. The use of terminology differed within the publications or was attributed different meanings. For example, several were named *usability* but referred to different aspects of that criterion, such as *ease of use*, *usefulness*, or *speed*. Others, such as *research-backed*, *scientific references*, *information accuracy*, and *information quality*, were named differently but clearly referred to a single criterion—in this case, *evidence-based information*.

Using the new eHAPPI framework, it is clear that most studies reported criteria that focused on usability, evidence-based content, functionality, or scientific evaluation. However, only 1 reported criterion dealt with context, and only 5 dealt with development processes.

The research team's discussions emphasized that some criteria are objective (eg, *average rating in the app store* and *purpose of the app*). In contrast, others are more subjective (eg, *matching the needs of the target population* and *intention to use*). In addition, several were dependent on the tested app's purpose or content (eg, *whether an app community exists* and *features to support behavior change*).

The research team agreed on how all the criteria were classified into dimensions and how most were formulated. However, to improve understanding or align with known formulations, they suggested revisions to the wording of 12 criteria [28]. On the basis of the research team's recommendations (Multimedia Appendix 6 [25,45-47,76-85]), 11 new criteria were added (Multimedia Appendix 7 [28]). These additions were mainly to the *ethical, legal, and social aspects* section. No criteria were deemed completely irrelevant, so none were removed. Finally, the research team agreed to classify these added criteria into the existing dimensions.

The final list contained 216 criteria for describing and evaluating eHealth smartphone apps (Multimedia Appendix 8). Although this list was comprehensive, our discussions highlighted that it was not practical for use by HCPs in clinical practice. The research team agreed that, in the next phase, a short version with only 1 to 3 items per dimension would be useful to make a quick initial decision (ie, acting as an algorithm to gauge whether an app should be given further consideration). Only if an app passed this pretest would it undergo a more thorough evaluation using detailed criteria and offering nuanced results. In addition, the research team recommended a specific algorithm with thresholds that could be adjusted depending on each tested app's purpose and context.

The research team recommended that the next step for the project's second phase be a Delphi process to condense the list and develop the proposed decision support tool. This process will have two aims: (1) to provide a means to reach a consensus and (2) to develop a useful and feasible (ie, practical for use in

clinical practice) tool to describe and evaluate eHealth smartphone apps.

Moreover, participants expressed concerns that the necessary information to complete the tool may be difficult to find. Therefore, the development of a user guide for HCPs on how to apply the tool and where to typically find the required information was also proposed for the next phase. Finally, the eHealth smartphone app evaluation tool and user guide will need to be pilot-tested with HCPs.

Discussion

Principal Findings

Although the eHealth field is rapidly expanding and evolving, there is no consensus on how the quality of eHealth apps should be defined and evaluated by HCPs [13,19,20,86]. In this paper, we described how we developed a comprehensive list of criteria to evaluate eHealth apps. We used a meticulous methodological approach to derive the list, consisting of a systematic literature review and iterative rounds with stakeholders from various backgrounds to compose a comprehensive framework—the eHAPPI—and use it to synthesize all the criteria identified in the selected studies. We found 6 overarching dimensions (ie, *context*, *stakeholder involvement*, *features and requirements*, *development processes*, *implementation*, and *evaluation*) of eHealth app evaluation and 205 criteria in the literature. A research team discussion resulted in 11 additional criteria, bringing the new total to 216. Using this comprehensive list, HCPs would be able to evaluate eHealth apps designed for diverse health care needs. No original studies included all the dimensions or all the criteria. Most of the selected studies (83/128, 64.8%) did not describe an underlying framework or theoretical guidance or how their criteria were developed. In addition, a general lack of common terminology among the included studies further complicated efforts toward comparison and direct application in clinical practice.

Although the proposed complete list of criteria was comprehensive, applicable to diverse health care needs, and applied consistent terminology, its length and depth limited the feasibility of using it in clinical practice. As a stand-alone reference, this list would be better suited for comprehensive evaluations of a broad range of eHealth apps by a regulatory body. Therefore, we proposed a 2-step approach to developing these criteria for use by HCPs. The first step would be defining a few critical initial decision criteria. A more detailed description and assessment would be useful only for positive evaluations using these first criteria. Using such an algorithm would allow the most obviously inferior apps to be quickly excluded from the process.

However, our progress in this direction was slowed by a lack of consensus regarding which criteria were essential and which would simply be nice to have [23]. As a compromise, we agreed that developing the proposed algorithm would require more expertise than we had and further expert discussion. Therefore, for the second phase of this project, we proposed a Delphi process [87] to guide the further development and fine-tuning of the final evaluation tool.

Although we did not reach a consensus on which criteria qualify as essential, our group discussions provided insights into which qualities to consider in describing and evaluating apps as well as how an optimal tool might be structured. For example, given the dynamic and rapidly evolving use of apps in clinical practice [29,88], flexibility is a significant concern [23,25]. Therefore, the research team recommended an algorithm not only whose cutoff criteria can be modified to fit each evaluated app's purpose and context but also whose overall functionalities can evolve alongside the surrounding technology [89].

Another point of discussion concerned the difficulty we encountered in finding the information to complete this evaluation tool. HCPs who are less familiar with eHealth apps may have difficulty gathering even basic data, such as the name of an app's developer or its latest update [13,21,22]. More technologically savvy individuals may have trouble finding information on that app's scientific basis, how or whether its development processes included stakeholder involvement, or which strategies were used in its implementation. Therefore, we suggest that the proposed tool include a user guide describing why such criteria are important and where to find and how to judge the required information. This echoes a recommendation by the European Network to Advance Best Practices and Technology on Medication Adherence in their Cooperation in Science and Technology Action (CA19132), which facilitates the use of a web-based repository of information on medication adherence technology [90,91]. Although eHealth app developers clearly need to provide relevant details in a clear and easily accessible way, health educators also need to include eHealth evaluation in HCP education and training curricula.

The previous discussion provides the foundation for conducting phase 2 of this study. This phase has three goals: (1) to conduct a Delphi survey to narrow down the number of criteria and develop an algorithm for initial decision-making, (2) to develop a user guide, and (3) to pilot-test the resulting iteration.

Limitations

This study has several limitations. Most notably, at this point in the project, although our list of eHealth app description and evaluation criteria is comprehensive, it remains a preliminary version. That is, despite discussions with various interdisciplinary experts, phase 1 did not produce an in-depth consensus on the essential criteria for a richly detailed but broadly feasible means of evaluating eHealth apps. This drove the decision to design a 2-phase project. In phase 2, which will be a Delphi study [87], we aim to develop a criteria-sorting algorithm. With this in place, the phase will culminate in a version of an eHealth app evaluation tool for pilot-testing.

In addition, all the included studies were assessed using the AGREE-II instrument [42], which was specifically designed to assess clinical practice guideline development reports. Considering the high level of heterogeneity across many of the study characteristics, direct comparability using a single tool was limited. However, although other instruments would have been more suitable in many cases, using various instruments would have yielded equally varied results. As we were primarily interested in the rationales and development processes that

supported the dimensions and criteria, the AGREE-II scales provided a consistent assessment of these aspects.

Finally, the participants in the discussion rounds for the development of the conceptual framework were primarily health care researchers and professionals. There were very few technology developers or industry representatives present, and only 2 patient representatives participated. This may have resulted in a limited consideration of the patient perspective and an increased risk of interventions exacerbating existing health care inequalities [92]. Therefore, we included in the *context* dimension of the eHAPPI framework subgroups focusing on ethical and social aspects. These subgroups aim to underscore the necessity of addressing the risk of intervention-generated inequalities.

Comparison With Prior Work

Although 23 of the existing tools were explicitly intended for HCPs [14,43,46,47,76-86,93-100], none of these were complete; rather, they were too focused on specific conditions, or their theoretical justifications or development processes were not described. Such omissions make it difficult for HCPs in clinical practice to comprehensively but feasibly describe and evaluate eHealth apps in a standardized way to guide the recommendation of relevant, reliable, and high-quality apps to their patients [11,13,14]. Other studies supplemented the dimensions and criteria for describing and evaluating apps. However, it remains unclear which criteria are essential and how detailed they need to be. Recently, there has been much discussion about how to define and evaluate eHealth quality and what criteria are needed for an app to be used in the health care system [23,89,101]. Future findings from the planned phase 2 will likely provide a basis for further discussion on this topic among app developers or providers, HCPs, patients, researchers, and policy makers. Our first comprehensive list of criteria as a result of phase 1 provides an excellent basis for the next steps in phase 2 to develop a new eHealth app evaluation tool.

The need for a tool to describe and evaluate eHealth apps and help HCPs and their patients navigate the digital health ecosystem is pressing [23]. Our path to a proposed resolution has been quite complex as this field is also complex. After listing the criteria identified via a literature review, we developed them through expert discussions, revealing important improvement areas. In particular, compared with recommendations from the Health Technology Assessment Core Model [28], the criteria concerning the ethical, legal, and social aspects of eHealth apps were deemed incomplete. Therefore, in addition to adapting many criteria, we added several.

Contribution of This Study

This study focuses on addressing the rapidly growing and somewhat chaotic field of eHealth, particularly the challenges faced by HCPs when it comes to evaluating and recommending health-related smartphone apps to their patients. This study's contribution lies in its comprehensive methodology for gathering and categorizing existing criteria for evaluating health apps, which is essential as no single framework or evaluation tool effectively serves this purpose. The methodology involved a systematic review of the literature, which resulted in the identification of 216 distinct evaluation criteria organized within a conceptual framework comprising 6 app evaluation dimensions. These dimensions encompass various aspects, including the app's context, stakeholder involvement in its development, features, development processes, implementation, and evaluation. This study highlights the need for a more purpose-built, theory-driven tool to help HCPs assess and recommend apps effectively and outlines plans to create a 2-part app evaluation tool based on the gathered criteria, which will expedite the process of disqualifying unsuitable apps and scrutinizing potential candidates more closely. This study serves as a crucial foundational step toward developing a practical tool that can guide HCPs in evaluating and recommending health-related apps.

Conclusions

Developing a tool comprehensive enough for HCPs to reliably describe and evaluate the full range of eHealth apps yet short enough to be feasible for daily clinical practice is a daunting challenge. After our literature review yielded a list of criteria too bulky for routine use, there was a lack of consensus either on terminology or on relevance to define and evaluate app quality. In this report of phase 1, we provided our initial comprehensive overview of 216 relevant criteria used in the selected studies to describe, evaluate, and recommend eHealth apps. To condense this list to a more manageable size, in phase 2, we will formulate and apply a robust consensus-building process to generate a list of criteria ranked by importance, followed by the creation of an algorithm to produce short- and long-form evaluations to match the characteristics of the apps to be evaluated. In addition, the development of a user guide and pilot-testing of the tool are planned. As a basis for informed guidance and decision-making, such a tool will help HCPs reliably describe and evaluate eHealth apps for their patients.

Acknowledgments

This work was conducted mainly by a PhD student as part of her PhD project funded by the University of Basel. The funders had no role in the study design; collection, analysis, and interpretation of the data; writing of the report; or decision to submit the paper for publication. The authors thank all the participants in the various discussion rounds for their reflections and input for the development of the conceptual framework and criteria. Special thanks also go to the participants of the Research Round Table Meeting seminar for their critical feedback in developing the study methods and manuscript, as well as to Chris Shultis for his language editing.

Authors' Contributions

The study was conceptualized by JR, SDG, and ADD. The literature search strategy was developed by JR, SDG, and FD. Articles were located, identified, and evaluated by JR and TH. Data were extracted and checked by these same authors. The initial dimension and criteria list was composed and discussed by all the authors. The manuscript was drafted and edited by JR with close revision and feedback from SDG as well as review and feedback from all other coauthors.

Conflicts of Interest

None declared.

Multimedia Appendix 1
MEDLINE search strategy.
[PDF File (Adobe PDF File), 153 KB - mhealth_v12i1e48625_app1.pdf]

Multimedia Appendix 2
Demographics of participants.
[PDF File (Adobe PDF File), 184 KB - mhealth_v12i1e48625_app2.pdf]

Multimedia Appendix 3
Included articles.
[PDF File (Adobe PDF File), 470 KB - mhealth_v12i1e48625_app3.pdf]

Multimedia Appendix 4
Characteristics of the included studies.
[PDF File (Adobe PDF File), 289 KB - mhealth_v12i1e48625_app4.pdf]

Multimedia Appendix 5
Original frameworks and dimensions.
[PDF File (Adobe PDF File), 282 KB - mhealth_v12i1e48625_app5.pdf]

Multimedia Appendix 6
Revised criteria.
[DOCX File , 129 KB - mhealth_v12i1e48625_app6.docx]

Multimedia Appendix 7
Additional criteria.
[DOCX File , 45 KB - mhealth_v12i1e48625_app7.docx]

Multimedia Appendix 8
Complete criteria list.
[PDF File (Adobe PDF File), 376 KB - mhealth_v12i1e48625_app8.pdf]

References

1. Using e-health and information technology to improve health 2020. World Health Organization. URL: https://www.who.int/westernpacific/activities/using-e-health-and-information-technology-to-improve-health [accessed 2023-11-29]
2. Bevans M, El-Jawahri A, Tierney D, Wiener L, Wood WA, Hoodin F, et al. National institutes of health hematopoietic cell transplantation late effects initiative: the patient-centered outcomes working group report. Biol Blood Marrow Transplant 2017 Apr;23(4):538-551 [FREE Full text] [doi: 10.1016/j.bbmt.2016.09.011] [Medline: 27660168]
3. Gee PM, Greenwood DA, Paterniti DA, Ward D, Miller LM. The eHealth enhanced chronic care model: a theory derivation approach. J Med Internet Res 2015 Apr 01;17(4):e86 [FREE Full text] [doi: 10.2196/jmir.4067] [Medline: 25842005]
4. Classification of self-care interventions for health: a shared language to describe the uses of self-care interventions. World Health Organization. 2021. URL: https://www.who.int/publications/i/item/9789240039469 [accessed 2023-11-29]
5. Mueller SM, Jungo P, Cajacob L, Schwegler S, Itin P, Brandt O. The absence of evidence is evidence of non-sense: cross-sectional study on the quality of psoriasis-related videos on YouTube and their reception by health seekers. J Med Internet Res 2019 Jan 16;21(1):e11935 [FREE Full text] [doi: 10.2196/11935] [Medline: 30664460]

6. Taylor N, Conner M, Lawton R. The impact of theory on the effectiveness of worksite physical activity interventions: a meta-analysis and meta-regression. Health Psychol Rev 2012 Mar;6(1):33-73 [FREE Full text] [doi: 10.1080/17437199.2010.533441]
7. Dombrowski S, Sniehotta F, Avenell A, Johnston M, MacLennan G, Araújo-Soares V. Identifying active ingredients in complex behavioural interventions for obese adults with obesity-related co-morbidities or additional risk factors for co-morbidities: a systematic review. Health Psychol Rev 2012 Mar;6(1):7-32. [doi: 10.1080/17437199.2010.513298]
8. Schoeppe S, Alley S, Van Lippevelde W, Bray NA, Williams SL, Duncan MJ, et al. Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: a systematic review. Int J Behav Nutr Phys Act 2016 Dec 07;13(1):127 [FREE Full text] [doi: 10.1186/s12966-016-0454-y] [Medline: 27927218]
9. Ahmed I, Ahmad N, Ali S, Ali S, George A, Saleem Danish H, et al. Medication adherence apps: review and content analysis. JMIR Mhealth Uhealth 2018 Mar 16;6(3):e62 [FREE Full text] [doi: 10.2196/mhealth.6432] [Medline: 29549075]
10. Eckert T, Wunsch K, Fiedler J, Woll A. SMARTMOVE-Involving families in the development and implementation of mHealth interventions. Pravention Und Gesundheitsforderung 2022;17(3):313-319 [FREE Full text]
11. Magnol M, Eleonore B, Claire R, Castagne B, Pugibet M, Lukas C, et al. Use of eHealth by patients with rheumatoid arthritis: observational, cross-sectional, multicenter study. J Med Internet Res 2021 Jan 29;23(1):e19998 [FREE Full text] [doi: 10.2196/19998] [Medline: 33512320]
12. Taeger J, Müller-Graff FT, Hagen R, Rak K. [Application areas of medical apps in otolaryngology]. HNO 2021 May 25;69(5):435-444. [doi: 10.1007/s00106-021-01038-9] [Medline: 33768279]
13. Cummings E, Borycki E, Roehrer E. Issues and considerations for healthcare consumers using mobile applications. Stud Health Technol Inform 2013;183:227-231. [Medline: 23388288]
14. Knitza J, Simon D, Lambrecht A, Raab C, Tascilar K, Hagen M, et al. Mobile health usage, preferences, barriers, and eHealth literacy in rheumatology: patient survey study. JMIR Mhealth Uhealth 2020 Aug 12;8(8):e19661 [FREE Full text] [doi: 10.2196/19661] [Medline: 32678796]
15. Kuehnhausen M, Frost V. Trusting smartphone Apps? To install or not to install, that is the question. In: Proceedings of the 2013 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support. 2013 Presented at: CogSIMA '13; February 25-28, 2013; San Diego, CA p. 30-37 URL: https://ieeexplore.ieee.org/document/6523820?arnumber=6523820 [doi: 10.1109/cogsima.2013.6523820]
16. Dinour LM, Pole A. Evaluation of breastfeeding app features: content analysis study. JMIR Pediatr Parent 2022 Oct 26;5(4):e37581 [FREE Full text] [doi: 10.2196/37581] [Medline: 36287596]
17. Chiang M. Networked Life: 20 Questions and Answers. Cambridge, MA: Cambridge University Press; 2012.
18. Girardello A, Michahelles F. AppAware: which mobile applications are hot? In: Proceedings of the 12th international conference on Human computer interaction with mobile devices and services. 2010 Presented at: MobileHCI '10; September 7-10, 2010; Lisbon, Portugal p. 431-434 URL: https://dl.acm.org/doi/abs/10.1145/1851600.1851698 [doi: 10.1145/1851600.1851698]
19. Stach M, Kraft R, Probst T, Messner E, Terhorst Y, Baumeister H. Mobile health app database - a repository for quality ratings of mHealth apps. In: Proceedings of the 2020 IEEE 33rd International Symposium on Computer-Based Medical Systems. 2020 Presented at: CBMS '20; July 28-30, 2020; Rochester, MN p. 427-432 URL: https://ieeexplore.ieee.org/document/9183243 [doi: 10.1109/cbms49503.2020.00087]
20. Krick T. Evaluation frameworks for digital nursing technologies: analysis, assessment, and guidance. An overview of the literature. BMC Nurs 2021 Aug 17;20(1):146 [FREE Full text] [doi: 10.1186/s12912-021-00654-8] [Medline: 34404406]
21. Quinn S, Bond R, Nugent C. Quantifying health literacy and eHealth literacy using existing instruments and browser-based software for tracking online health information seeking behavior. Comput Human Behav 2017 Apr;69:256-267 [FREE Full text] [doi: 10.1016/j.chb.2016.12.032]
22. Lam M, Hines M, Lowe R, Nagarajan S, Keep M, Penman M, et al. Preparedness for eHealth: health sciences students' knowledge, skills, and confidence. J Inf Technol Educ Res 2016;15:305-334 [FREE Full text] [doi: 10.28945/3523]
23. Nebeker C, Bartlett Ellis RJ, Torous J. Development of a decision-making checklist tool to support technology selection in digital health research. Transl Behav Med 2020 Oct 08;10(4):1004-1015 [FREE Full text] [doi: 10.1093/tbm/ibz074] [Medline: 31120511]
24. Dansky KH, Thompson D, Sanner T. A framework for evaluating eHealth research. Eval Program Plann 2006 Nov;29(4):397-404. [doi: 10.1016/j.evalprogplan.2006.08.009] [Medline: 17950868]
25. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
26. Anthony Berauk VL, Murugiah MK, Soh YC, Chuan Sheng Y, Wong TW, Ming LC. Mobile health applications for caring of older people: review and comparison. Ther Innov Regul Sci 2018 May;52(3):374-382. [doi: 10.1177/2168479017725556] [Medline: 29714532]
27. Gomis M, Gil D, Lopez L, Brossa V, Mirabet S, Roig E, et al. Impact of mHealth in heart transplant management (mHeart). Int J Integr Care 2016 Dec 16;16(6):38. [doi: 10.5334/ijic.2981]

28. Work package 8. HTA core model® version 3.0. EUnetHTA Joint Action. 2016. URL: https://www.eunethta.eu/wp-content/uploads/2018/03/HTACoreModel3.0-1.pdf [accessed 2023-11-29]
29. Kuziemsky C, Lau F. Handbook of eHealth Evaluation: An Evidence-based Approach. Victoria, BC: University of Victoria; 2016. URL: https://onlinelibrary.wiley.com/doi/epdf/10.1002/9780470712184
30. Simera I, Moher D, Hirst A, Hoey J, Schulz K, Altman D. Transparent and accurate reporting increases reliability, utility, and impact of your research: reporting guidelines and the EQUATOR Network. BMC Med 2010 Apr 26;8:24 [FREE Full text] [doi: 10.1186/1741-7015-8-24] [Medline: 20420659]
31. Higgins JP, Green SE. Cochrane Handbook for Systematic Reviews of Interventions: Cochrane Book Series. Volume 4. London, UK: The Cochrane Collaboration; 2011. URL: https://onlinelibrary.wiley.com/doi/epdf/10.1002/9780470712184
32. Page MJ, Moher D, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. BMJ 2021 Mar 29;372:n160 [FREE Full text] [doi: 10.1136/bmj.n160] [Medline: 33781993]
33. Ribaut J, DeVito Dabbs A, Teynor A, Dobbels F, De Geest S. Development of an evaluation tool to assess and evaluate the characteristics and quality of eHealth applications: a systematic review and consensus finding. National Institute for Health and Care Research. 2021. URL: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42021227064 [accessed 2023-12-11]
34. McKay FH, Cheng C, Wright A, Shill J, Stephens H, Uccellini M. Evaluating mobile phone applications for health behaviour change: a systematic review. J Telemed Telecare 2018 Jan;24(1):22-30. [doi: 10.1177/1357633X16673538] [Medline: 27760883]
35. Vukovic V, Favaretti C, Ricciardi W, de Waure C. Health technology assessment evidence on e-health/m-health technologies: evaluating the transparency and thoroughness. Int J Technol Assess Health Care 2018 Jan;34(1):87-96. [doi: 10.1017/S0266462317004512] [Medline: 29455685]
36. Chi NC, Demiris G. A systematic review of telehealth tools and interventions to support family caregivers. J Telemed Telecare 2015 Jan;21(1):37-44 [FREE Full text] [doi: 10.1177/1357633X14562734] [Medline: 25475220]
37. Boudreaux ED, Waring ME, Hayes RB, Sadasivam RS, Mullen S, Pagoto S. Evaluating and selecting mobile health apps: strategies for healthcare providers and healthcare organizations. Transl Behav Med 2014 Dec 24;4(4):363-371 [FREE Full text] [doi: 10.1007/s13142-014-0293-9] [Medline: 25584085]
38. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
39. Bessell TL, McDonald S, Silagy CA, Anderson JN, Hiller JE, Sansom LN. Do Internet interventions for consumers cause more harm than good? A systematic review. Health Expect 2002 Mar 23;5(1):28-37 [FREE Full text] [doi: 10.1046/j.1369-6513.2002.00156.x] [Medline: 11906539]
40. Barello S, Triberti S, Graffigna G, Libreri C, Serino S, Hibbard J, et al. eHealth for patient engagement: a systematic review. Front Psychol 2016 Jan 08;6:2013 [FREE Full text] [doi: 10.3389/fpsyg.2015.02013] [Medline: 26779108]
41. Eland-de Kok P, van Os-Medendorp H, Vergouwe-Meijer A, Bruijnzeel-Koomen C, Ros W. A systematic review of the effects of e-health on chronically ill patients. J Clin Nurs 2011 Nov 27;20(21-22):2997-3010. [doi: 10.1111/j.1365-2702.2011.03743.x] [Medline: 21707807]
42. Appraisal of guidelines for research and evaluation (AGREE) II instrument. The AGREE Next Steps Consortium. 2019. URL: https://www.agreetrust.org/wp-content/uploads/2013/10/AGREE-II-Users-Manual-and-23-item-Instrument_2009_UPDATE_2013.pdf [accessed 2023-11-29]
43. Guan Z, Sun L, Xiao Q, Wang Y. Constructing an assessment framework for the quality of asthma smartphone applications. BMC Med Inform Decis Mak 2019 Oct 15;19(1):192 [FREE Full text] [doi: 10.1186/s12911-019-0923-8] [Medline: 31615493]
44. East-Richard C, Laplante L, Vézina J, Cellard C. L'évaluation psychologique par le biais des applications mobiles. Can Psychol 2018 Aug;59(3):262-271 [FREE Full text] [doi: 10.1037/cap0000113]
45. Grau I, Kostov B, Gallego JA, Grajales Iii F, Fernández-Luque L, Sisó-Almirall A. Método de valoración de aplicaciones móviles de salud en español: el índice iSYScore. Semergen 2016 Nov;42(8):575-583. [doi: 10.1016/j.semerg.2015.12.001] [Medline: 26879598]
46. Chan S, Torous J, Hinton L, Yellowlees P. Towards a framework for evaluating mobile mental health apps. Telemed J E Health 2015 Dec;21(12):1038-1041. [doi: 10.1089/tmj.2015.0002] [Medline: 26171663]
47. Wyatt JC, Thimbleby H, Rastall P, Hoogewerf J, Wooldridge D, Williams J. What makes a good clinical app? Introducing the RCP Health Informatics Unit checklist. Clin Med (Lond) 2015 Dec 30;15(6):519-521 [FREE Full text] [doi: 10.7861/clinmedicine.15-6-519] [Medline: 26621937]
48. Pfadenhauer LM, Gerhardus A, Mozygemba K, Lysdahl KB, Booth A, Hofmann B, et al. Making sense of complexity in context and implementation: the Context and Implementation of Complex Interventions (CICI) framework. Implement Sci 2017 Mar 15;12(1):21-17 [FREE Full text] [doi: 10.1186/s13012-017-0552-5] [Medline: 28202031]

49. Pollock A, Campbell P, Struthers C, Synnot A, Nunn J, Hill S, et al. Stakeholder involvement in systematic reviews: a protocol for a systematic review of methods, outcomes and effects. Res Involv Engagem 2017 Apr 21;3:9 [FREE Full text] [doi: 10.1186/s40900-017-0060-4] [Medline: 29062534]
50. Brugha R, Varvasovszky Z. Stakeholder analysis: a review. Health Policy Plan 2000 Sep;15(3):239-246 [FREE Full text]
51. Leppla L, Mielke J, Kunze M, Mauthner O, Teynor A, Valenta S, SMILe study team. Clinicians and patients perspectives on follow-up care and eHealth support after allogeneic hematopoietic stem cell transplantation: a mixed-methods contextual analysis as part of the SMILe study. Eur J Oncol Nurs 2020 Apr;45:101723. [doi: 10.1016/j.ejon.2020.101723] [Medline: 32062362]
52. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, Medical Research Council Guidance. Developing and evaluating complex interventions: the New Medical Research Council Guidance. BMJ 2008 Sep 29;337:a1655 [FREE Full text] [doi: 10.1136/bmj.a1655] [Medline: 18824488]
53. Process development definition. Law Insider. URL: https://www.lawinsider.com/dictionary/process-development [accessed 2023-11-29]
54. Duncan E, O'Cathain A, Rousseau N, Croot L, Sworn K, Turner KM, et al. Guidance for reporting intervention development studies in health research (GUIDED): an evidence-based consensus study. BMJ Open 2020 Apr 08;10(4):e033516 [FREE Full text] [doi: 10.1136/bmjopen-2019-033516] [Medline: 32273313]
55. Brandsen T, Honingh M. Definitions of co-production and co-creation. In: Brandsen T, Verschuere B, Steen T, editors. Co-Production and Co-Creation: Engaging Citizens in Public Services. New York, NY: Routledge; 2018:9-17 URL: https://www.taylorfrancis.com/chapters/oa-edit/10.4324/9781315204956-2/definitions-co-production-co-creation-taco-brandsen-marlies-honingh
56. Mao J, Vredenburg K, Smith PW, Carey T. The state of user-centered design practice. Commun ACM 2005 Mar;48(3):105-109. [doi: 10.1145/1047671.1047677]
57. Development team. Agile Innovative Solutions. URL: https://innolution.com/resources/glossary/development-team [accessed 2023-11-29]
58. PubMed MeSH term: program evaluation. National Center for Biotechnology Information. 1989. URL: https://www.ncbi.nlm.nih.gov/mesh/68015397 [accessed 2023-11-29]
59. Usmani AM, Meo SA. Evaluation of science. Sudan J Paediatr 2011;11(1):6. [doi: 10.24911/sjp]
60. PubMed MeSH term: science. National Center for Biotechnology Information. URL: https://www.ncbi.nlm.nih.gov/mesh/68012586 [accessed 2023-11-29]
61. Eccles MP, Mittman BS. Welcome to implementation science. Implement Sci 2006 Feb 22;1(1):5908 [FREE Full text] [doi: 10.1186/1748-5908-1-1]
62. Proctor E, Silmere H, Raghavan R, Hovmand P, Aarons G, Bunger A, et al. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. Adm Policy Ment Health 2011 Mar;38(2):65-76. [doi: 10.1007/s10488-010-0319-7] [Medline: 20957426]
63. Vande Moere A, Tomitsch M, Wimmer C, Christoph B, Grechenig T. Evaluating the effect of style in information visualization. IEEE Trans Vis Comput Graph 2012 Dec;18(12):2739-2748. [doi: 10.1109/TVCG.2012.221] [Medline: 26357183]
64. Boy J, Detienne F. Storytelling in information visualizations: does it engage users to explore data? In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. 2015 Presented at: CHI '15; April 18-23, 2015; Seoul, Republic of Korea p. 1449-1458 URL: https://dl.acm.org/doi/10.1145/2702123.2702452 [doi: 10.1145/2702123.2702452]
65. Maintenance and repair of concrete structures ? Part 1: general princi-ples (ISO standard no. 16311-1:2014). International Organization for Standardization. 2014. URL: https://www.iso.org/obp/ui/#iso:std:iso:16311:-1:ed-1:v1:en [accessed 2023-11-29]
66. Requirement 2022. Wikipedia. URL: https://en.wikipedia.org/wiki/Requirement [accessed 2023-11-29]
67. PubMed MeSH term: evidence-based practice. National Center for Biotechnology Information. 2009. URL: https://www.ncbi.nlm.nih.gov/mesh/68055317 [accessed 2023-11-29]
68. Cheu LR, Casals A, Cuxart A, Parra A. Towards the definition of a functionality index for the quantitative evaluation of hand-prosthesis. In: Proceedings of the 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems. 2005 Presented at: IROS '05; August 2-6, 2005; Edmonton, AB p. 541-546 URL: https://ieeexplore.ieee.org/document/1545528 [doi: 10.1109/iros.2005.1545528]
69. Martinez WH. The importance of functionality of vegetable protein in foods. In: Wilcke H, editor. Soy Protein and Human Nutrition. New York, NY: Academic Press; 1979:53-77.
70. Ergonomics of human-system interaction: part 11: usability (ISO standard no. 9241-11:2018). International Organization for Standardization. 2018. URL: https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-2:v1:en [accessed 2023-11-29]
71. Privacy protection: Privacy guidelines for smart cities (ISO standard no. 27570:2021). International Organization for Standardization. 2021. URL: https://www.iso.org/obp/ui/#iso:std:iso-iec:ts:27570:ed-1:v1:en [accessed 2023-11-29]
72. Brooks DJ. What is security: definition through knowledge categorization. Secur J 2010;23(3):225-239 [FREE Full text]

73. Cobb M. Physical security. Physical Security TechTarget. URL: https://searchsecurity.techtarget.com/definition/physical-security [accessed 2023-11-29]
74. Lo O, Fan L, Buchanan W, Thuemmler C. Conducting performance evaluation of an e-health platform. In: Information Systems and Technology for Organizations in a Networked Society. New York, NY: IGI Global; 2013:295-315.
75. Martínez-Pérez B, de la Torre-Díez I, López-Coronado M. Experiences and results of applying tools for assessing the quality of a mHealth app named heartkeeper. J Med Syst 2015 Nov;39(11):142. [doi: 10.1007/s10916-015-0303-6] [Medline: 26345452]
76. Jain YS, Garg A, Jhamb D, Jain P, Karar A. Preparing India to leverage power of mobile technology: development of a bilingual mobile health tool for heart patients. Cardiovasc Hematol Agents Med Chem 2019;17(2):125-134. [doi: 10.2174/1871525717666190912152938] [Medline: 31512999]
77. Collins R. Nurses' perceived usefulness of secure texting applications for the purpose of patient car. Online J Nurs Inform 2019;23(1) [FREE Full text]
78. Groen G, Jörns-Presentati A, Dessauvagie A, Seedat S, van den Heuvel LL, Suliman S, et al. Development of a mobile application for detection of adolescent mental health problems and feasibility assessment with primary health care workers. Issues Ment Health Nurs 2022 Nov;43(11):1046-1055. [doi: 10.1080/01612840.2022.2124003] [Medline: 36205922]
79. Sanatkar S, Counson I, Mackinnon A, Bartholomew A, Glozier N, Harvey S. Preliminary investigation of shift, a novel smartphone app to support junior doctors' mental health and well-being: examination of symptom progression, usability, and acceptability after 1 month of use. J Med Internet Res 2022 Sep 21;24(9):e38497 [FREE Full text] [doi: 10.2196/38497] [Medline: 36129745]
80. Kaliyadan F, Ashique K. Use of mobile applications in dermatology. Indian J Dermatol 2020;65(5):371. [doi: 10.4103/ijd.ijd_422_20]
81. Kwan V, Hagen G, Noel M, Dobson K, Yeates K. Healthcare at your fingertips: the professional ethics of smartphone health-monitoring applications. Ethics Behav 2017 Jan 24;27(8):615-631 [FREE Full text] [doi: 10.1080/10508422.2017.1285237]
82. Olfert MD, Barr ML, Hagedorn RL, Long DM, Haggerty TS, Weimer M, et al. Feasibility of a mHealth approach to nutrition counseling in an Appalachian state. J Pers Med 2019 Nov 20;9(4):50 [FREE Full text] [doi: 10.3390/jpm9040050] [Medline: 31757057]
83. Sudol NT, Adams-Piper E, Perry R, Lane F, Chen KT. In search of mobile applications for patients with pelvic floor disorders. Female Pelvic Med Reconstr Surg 2019 May;25(3):252-256. [doi: 10.1097/SPV.0000000000000527] [Medline: 29219859]
84. Vasiloglou MF, Christodoulidis S, Reber E, Stathopoulou T, Lu Y, Stanga Z, et al. What healthcare professionals think of "nutrition and diet" apps: an international survey. Nutrients 2020 Jul 24;12(8):2214 [FREE Full text] [doi: 10.3390/nu12082214] [Medline: 32722339]
85. Yasini M, Marchand G. Mobile health applications, in the absence of an authentic regulation, does the usability score correlate with a better medical reliability? Stud Health Technol Inform 2015;216:127-131. [Medline: 26262024]
86. Johnson E, Emani VK, Ren J. Breadth of coverage, ease of use, and quality of mobile point-of-care tool information summaries: an evaluation. JMIR Mhealth Uhealth 2016 Oct 12;4(4):e117 [FREE Full text] [doi: 10.2196/mhealth.6189] [Medline: 27733328]
87. Fitch K, Bernstein SJ, Aguilar M, Burnand B, LaCalle JR, Lazaro P, et al. The RAND/UCLA Appropriateness Method User's Manual. Cambridge, UK: RAND Corporation; 2001. URL: https://apps.dtic.mil/sti/pdfs/ADA393235.pdf
88. Rooij T, Marsh S. eHealth: past and future perspectives. Per Med 2016 Jan;13(1):57-70. [doi: 10.2217/pme.15.40] [Medline: 29749870]
89. Digital health trends 2021: innovation, evidence, regulation, and adoption. IQVIA Inc. 2021. URL: https://www.iqvia.com/insights/the-iqvia-institute/reports-and-publications/reports/digital-health-trends-2021 [accessed 2023-11-29]
90. van Boven JF, Tsiligianni I, Potočnjak I, Mihajlović J, Dima AL, Nabergoj Makovec U, et al. European network to advance best practices and technology on medication adherence: mission statement. Front Pharmacol 2021;12:748702 [FREE Full text] [doi: 10.3389/fphar.2021.748702] [Medline: 34707502]
91. Nabergoj Makovec U, Goetzinger C, Ribaut J, Barnestein-Fonseca P, Haupenthal F, Herdeiro M, European Network to Advance Best practices and technoLogy on medication adherencE (ENABLE), European Network to Advance Best Practices and TechnoLogy on Medication AdherencE (ENABLE). Developing a medication adherence technologies repository: proposed structure and protocol for an online real-time Delphi study. BMJ Open 2022 Apr 22;12(4):e059674 [FREE Full text] [doi: 10.1136/bmjopen-2021-059674] [Medline: 35459677]
92. Veinot T, Mitchell H, Ancker J. Good intentions are not enough: how informatics interventions can worsen inequality. J Am Med Inform Assoc 2018 Aug 01;25(8):1080-1088 [FREE Full text] [doi: 10.1093/jamia/ocy052] [Medline: 29788380]
93. Ehrler F, Weinhold T, Joe J, Lovis C, Blondon K. A mobile app (BEDSide mobility) to support nurses' tasks at the patient's bedside: usability study. JMIR Mhealth Uhealth 2018 Mar 21;6(3):e57 [FREE Full text] [doi: 10.2196/mhealth.9079] [Medline: 29563074]
94. Shalan A, Abdulrahman A, Habli I, Tew G, Thompson A. YORwalK: desiging a smartphone exercise application for people with intermittent claudication. Stud Health Technol Inform 2018;247:311-315. [Medline: 29677973]

95. Chang O, Patel VL, Iyengar S, May W. Impact of a mobile-based (mHealth) tool to support community health nurses in early identification of depression and suicide risk in Pacific Island Countries. Australas Psychiatry 2021 Apr 22;29(2):200-203. [doi: 10.1177/1039856220956458] [Medline: 32961100]
96. Chirambo GB, Muula AS, Thompson M, Hardy VE, Heavin C, Connor YO, et al. End-user perspectives of two mHealth decision support tools: electronic community case management in Northern Malawi. Int J Med Inform 2021 Jan;145:104323. [doi: 10.1016/j.ijmedinf.2020.104323] [Medline: 33232917]
97. Fijačko N, Masterson Creber R, Gosak L, Kocbek P, Cilar L, Creber P, et al. A review of mortality risk prediction models in smartphone applications. J Med Syst 2021 Nov 04;45(12):107 [FREE Full text] [doi: 10.1007/s10916-021-01776-x] [Medline: 34735603]
98. Jiang S, Lv M, Wu T, Chen W, Zhang J. A smartphone application for remote adjustment of warfarin dose: development and usability study. Appl Nurs Res 2022 Mar;63:151521. [doi: 10.1016/j.apnr.2021.151521] [Medline: 35034699]
99. Shen Z, Zhang Y, Yang C, Liu J, Huang C, Zhang X, et al. A smart-phone app for fluid balance monitoring in patients with heart failure: a usability study. Patient Prefer Adherence 2022 Aug 02;16:1843-1853 [FREE Full text] [doi: 10.2147/PPA.S373393] [Medline: 35942225]
100. Kettlewell J, Phillips J, Radford K, dasNair R. Informing evaluation of a smartphone application for people with acquired brain injury: a stakeholder engagement study. BMC Med Inform Decis Mak 2018 May 30;18(1):33 [FREE Full text] [doi: 10.1186/s12911-018-0611-0] [Medline: 29848312]
101. Kyhlstedt M. The need for action by evaluators and decision makers in Europe to ensure safe use of medical software. Front Med Technol 2022 Nov 29;4:1063622 [FREE Full text] [doi: 10.3389/fmedt.2022.1063622] [Medline: 36523427]

Abbreviations

AGREE-II: Appraisal of Guidelines for Research and Evaluation–II
eHAPPI: eHealth Smartphone App Evaluation
HCP: health care professional
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
WHO: World Health Organization

Edited by M Dorsch; submitted 01.05.23; peer-reviewed by B Kalema, R Arriaga, L Visser; comments to author 06.10.23; revised version received 03.11.23; accepted 09.11.23; published 15.01.24.

Please cite as:
Ribaut J, DeVito Dabbs A, Dobbels F, Teynor A, Mess EV, Hoffmann T, De Geest S
Developing a Comprehensive List of Criteria to Evaluate the Characteristics and Quality of eHealth Smartphone Apps: Systematic Review
JMIR Mhealth Uhealth 2024;12:e48625
URL: https://mhealth.jmir.org/2024/1/e48625
doi:10.2196/48625
PMID:38224477

Original Paper

Evaluation of Chinese HIV Mobile Apps by Researchers and Patients With HIV: Quality Evaluation Study

Peng Liu[1,2], MS; Lingmeng Wang[1], BS; Fuzhi Wang[1,2], PhD

[1]School of Health Management, Bengbu Medical College, Bengbu, China

[2]Innovation Team of Health Information Management and Application Research, Bengbu Medical College, Bengbu, China

Corresponding Author:
Fuzhi Wang, PhD
School of Health Management
Bengbu Medical College
2600 Donghai Road
Bengbu, 233030
China
Phone: 86 0552 3173195
Email: wfz.bbmc@foxmail.com

Abstract

Background: Against the backdrop of globalization, China remains one of the most heavily burdened countries in Asia with regard to AIDS. However, many high-risk groups and patients affected by AIDS may be less likely to actively seek care from medical institutions because of fear of experiencing shame or discrimination. Mobile apps provide a promising avenue for supporting the prevention, diagnosis, and treatment of AIDS. However, a comprehensive systematic evaluation of these mobile apps' functionality and quality has not been conducted yet.

Objective: This study aims to identify the available mobile apps for AIDS in China, assess and discuss the functional features and quality of these Chinese AIDS mobile apps, and offer decision support for patients and clinical practitioners in accessing high-quality AIDS mobile apps. Furthermore, based on the evaluation results, recommendations for improvement will be provided.

Methods: A systematic search was conducted on the Qimai app data platform, the Aladdin WeChat applet data platform, and WeChat to identify mobile apps related to AIDS. A snowball sampling method was used to supplement the potentially overlooked apps. The selected mobile apps underwent a rigorous screening process based on unified criteria. Subsequently, assessments were independently undertaken by 3 separate researchers and 2 patients with HIV, using both the Mobile App Rating Scale (MARS) and the User Mobile App Rating Scale (uMARS). Quantitative interpretations of the data were facilitated by the MedCalc statistical software (version 20.217, MedCalc Software).

Results: A total of 2901 AIDS mobile apps were included in the study, with 2897 identified through information retrieval and an additional 4 added via snowball sampling. After a rigorous selection process, 21 apps were determined to be usable. Among them, the Hong Feng Wan app achieved the highest combined average score, calculated based on the MARS (3.96, SD 0.33) and uMARS (4.47, SD 0.26). Overall, there was no significant correlation between MARS and uMARS ($r_{\text{app quality total score}}$=0.41; P=.07; $r_{\text{subjective quality}}$=0.39; P=.08). A notable issue was the widespread lack of user privacy protection, with only 24% (5/21) of the apps offering this feature.

Conclusions: The number of available Chinese AIDS mobile apps is limited, with WeChat applets dominating the market. Nonetheless, the performance of WeChat mini-apps is generally inferior to that of independent apps, and there may be significant discrepancies between assessments conducted by researchers and those provided by genuine end users, emphasizing the necessity of involving real users in the development and evaluation of HIV mobile apps. In addition, developers of these Chinese HIV mobile apps need to devote attention to improving privacy protection mechanisms, in addition to considering the evaluations of researchers and real users. This will help attract more users and increase user loyalty.

(JMIR Mhealth Uhealth 2024;12:e52573) doi:10.2196/52573

KEYWORDS

HIV; mobile app; evaluation; mobile phone

Introduction

Background

AIDS is a chronic infectious disease caused by HIV infection, resulting in high mortality rates. Since the first reported case of AIDS, it has rapidly spread worldwide, becoming a major public health concern. China, influenced by globalization, is no exception. According to the 2021 National Notifiable Infectious Diseases Summary released by the Chinese Center for Disease Control and Prevention, AIDS has the highest mortality rate and death toll, with 19,623 reported deaths. Available data indicate that as of 2022, the number of confirmed AIDS cases in China has exceeded 1.05 million and is projected to surpass 1.6 million infections by 2023. AIDS remains a significant infectious disease affecting public health and socioeconomic development in China [1].

For individuals who are at a high risk of HIV infection but have not yet been infected, pre-exposure prophylaxis (PrEP) is a highly effective method of preventing HIV through daily medication. However, the current use of and adherence to PrEP remain low. However, mobile apps have shown effectiveness in promoting PrEP use and adherence among high-risk populations [2]. Muessig et al [3] highlighted the advantages of internet and mobile-based interventions, which facilitate wider dissemination of PrEP at a lower cost compared with conventional methods. The MyChoices app developed by Biello KB demonstrated feasibility and potential in improving HIV PrEP use rates among gay and bisexual men in the United States [4]. The prevalence of HIV infections remains disproportionately high among gay, bisexual, and other men who have sex with men (GBMSM) [5]. With the widespread use and convenience of smartphones, sexual networking apps have replaced traditional dating websites as the primary online social platform for GBMSM. As such, apps have become crucial venues for sexual health research [6]. Targeted video and text-based sexually transmitted infection and HIV information provided through mobile apps has proven effective in reducing new infection rates among GBMSM [5].

For people living with HIV, the highly active antiretroviral therapy developed by scientist David Ho represents the most effective treatment method. Highly active antiretroviral therapy can suppress HIV replication and prevent the emergence of drug-resistant viruses. However, strict adherence is required, necessitating patients to adhere to prescribed medication schedules [7]. Despite simplified treatment regimens, adherence remains challenging for some people living with HIV [8]. Evidence suggests that HIV case management can improve treatment adherence and quality of life and reduce risky sexual behavior [9,10]. However, case managers face significant workloads and limited policy and funding support [11,12]. Mobile health (mHealth) apps have demonstrated potential in assisting people living with HIV with effective self-management and delivering personalized interventions. Schnall et al [13,14] identified ideal features for an HIV app, including reminders, health information delivery, medication logs, communication, settings, and search functions. However, comprehensive apps meeting these criteria are scarce in the current market, and there is a lack of rigorously evaluated mHealth apps specifically designed for people living with HIV [13,14]. Yang et al [15] found that most Chinese GBMSM apps, selected based on relaxed inclusion criteria, primarily focused on dating and lacked HIV prevention and health information.

For both high-risk populations and individuals infected with HIV, HIV testing plays a vital role in combating HIV. Early HIV testing allows individuals to learn about their infection status promptly, facilitating timely access to antiretroviral therapy and significantly reducing mortality rates. However, fear of shame or discrimination may discourage many high-risk populations and patients from actively seeking care at health care facilities [16-19]. With the increasing prevalence of mobile phones, various mHealth interventions have been developed to diversify HIV self-testing (HIVST) approaches, including telephone hotlines, SMS text messaging–based interventions, and internet-based interventions. These interventions have shown potential in improving testing rates, particularly among hard-to-reach populations [20-26]. Although these platforms have achieved varying levels of success, the use of mobile apps has emerged as a highly popular trend because of their flexibility and scalability. For instance, the mLab app serves to enhance users' understanding of their HIV test results while facilitating their access to pertinent HIV information and services [27]. Another noteworthy app is Aspect HIVST, which offers an acceptable means of uploading mobile HIVST results and demographic information to a centralized database [28]. In addition, ApiDé serves as a multilingual electronic tool (app) that assists health care providers in offering and explaining HIV screening to immigrants facing language barriers [29].

In conclusion, the use of mobile apps is becoming increasingly prevalent in HIV/AIDS prevention and control. These apps provide a convenient and accessible means for high-risk individuals and patients to access information, consultations, and support related to HIV/AIDS. They serve as effective adjunct measures in improving antiretroviral therapy adherence and reducing AIDS incidence rates, thereby mitigating the currently imbalanced resource allocation between patients and health care providers involved in antiviral treatment. However, research on Chinese AIDS mobile medical apps is relatively scarce and late, and empirical studies evaluating the effectiveness of HIV mobile medical apps remain sparse. Therefore, the effectiveness of Chinese HIV mobile apps in meeting user needs requires further assessment. To evaluate the efficacy of AIDS mobile apps in China, the Mobile App Rating Scale (MARS) and the User Mobile App Rating Scale (uMARS) scales, which are widely accepted and applied for uses such as assessing chronic disease management [30], COVID-19 tracking [31], psychiatric interventions [32], physical exercise among older people [33], and menstrual monitoring [34], were incorporated into this study as standardized assessment tools for the quality of mobile apps.

Objectives

This study aims to achieve the following objectives:

1. To identify a comprehensive list of available Chinese mobile apps for HIV/AIDS.

2. To evaluate the functional features and quality of Chinese HIV/AIDS mobile apps from 2 distinct standpoints: those of researchers and those of individuals diagnosed with HIV, and to provide decision support for AIDS-related groups and health care professionals in accessing high-quality AIDS mobile apps.
3. To conduct a thorough analysis of the evaluation results and to provide improvement recommendations based on the findings, with the ultimate goal of enhancing the quality of Chinese HIV/AIDS mobile apps.

Methods

Search Strategy

Considering the use status of Chinese AIDS apps, a systematic search was conducted on the Qimai app data platform, Aladdin WeChat applet data platform, and WeChat from February 18 to 19, 2023, adhering to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. The specific search strategies are as follows.

Data Collection of Apps

The Qimai platform, a well-known domestic mobile app data analytics platform, was used to retrieve AIDS-related apps. Qimai provides comprehensive data on iOS and Android app markets (including Huawei, Baidu, Xiaomi, Vivo, etc), along with app store optimization and app store search marketing optimization service tools and SearchAds data reports. It also offers professional data analysis and optimization strategies. The platform features an intelligent keyword expansion and association tool, which proved valuable for obtaining the required data for this study. In this research, the Qimai Keyword Expansion Assistant was used to search for AIDS-related terms. We selected keywords with a relevance score exceeding 50% and a search index surpassing 4605 (typically indicating a higher search frequency) to identify relevant apps that met the criteria. The search was conducted on February 18, 2023.

WeChat Applet Data Collection

WeChat applets, launched in China in 2017, are app programs developed on the WeChat platform. These applets offer users the convenience of accessing and using various functionalities and services directly within the WeChat app without requiring any separate downloads or installations. Known for their lightweight, fast, and user-friendly nature, they have gained popularity. To collect WeChat applet data, searches were performed on both the Aladdin Index and WeChat. The Aladdin Index, developed by Beijing Aladdin Future Technology Co, Ltd, serves as a ranking platform for the WeChat applet, providing a comprehensive reference for applet developers' operations nationwide. For this study, keywords such as *HIV*, *AIDS*, *Human Immunodeficiency Virus*, *Acquired Immunodeficiency Syndrome*, as well as their corresponding Chinese keywords with similar meanings, such as *huo-de-xing-mian-yi-que-xian-zong-he-zheng*,*ai-zi*, *ai-zi-bing*, and *ai-si-bing*, were used to search for the WeChat applet related to AIDS in both the Aladdin applet list and WeChat. The search was conducted on February 19, 2023.

Supplementary Method

In collaboration with the HIV prevention and control team at the local Centers for Disease Control and Prevention (CDC), a snowball sampling method was used to distribute electronic questionnaires to individuals living with AIDS during their follow-up visits conducted by the CDC staff. The questionnaires (without personal information and providing adequate privacy protection) aimed to inquire about the HIV/AIDS apps they had used. Participants were encouraged to provide feedback using the Wenjuanxing platform and to share the electronic questionnaire with other individuals living with AIDS they knew. This supplementary method aimed to identify AIDS-related apps that were widely used within the AIDS community but were not captured by the aforementioned search strategies. The feedback collection period was extended from February 20 to May 19.

The exclusion criteria for apps were as follows: (1) nonsimplified Chinese language, (2) games, (3) apps with excessively limited functionality, (4) mobile apps that did not primarily focus on HIV or AIDS-related content, (5) >50% of the content is inaccessible, (6) apps that had terminated their services, (7) apps restricted to internal use by specific personnel (apps had to be accessible and functional for evaluation purposes), and (8) apps predominantly intended for advertising and product sales.

Evaluation Tool

In this study, we selected the MARS and uMARS scales as the evaluation tools for assessing the quality of the apps. These scales were chosen for 2 main reasons. First, these scales have been extensively validated and have demonstrated good reliability and validity across different contexts. Second, in our study, the included apps were a mix of independent apps and WeChat applets, with the majority being WeChat applets rather than independent apps. However, the MARS and uMARS scales offer broader applicability in this regard.

Developed by Zelmer et al [35], the MARS scale provides a standardized set of criteria for evaluating mobile apps in terms of engagement, functionality, esthetics, and information quality. The MARS scale has been widely used in user research, mobile app evaluations, and related fields. Previous studies have indicated that the MARS scale demonstrates good validity and reliability, making it a reliable measurement tool [30-34]. The uMARS scale, developed by Manning et al [36], serves as an extension of the original MARS scale, specifically designed to prioritize the assessments of average users toward mobile apps.

Evaluation Process

The evaluation was conducted by 3 researchers. Rater 1 (PL) holds a master's degree in medical informatics and has extensive experience in medical information analysis. Rater 2 (Bin Li) holds a master's degree in computer science and has abundant software development experience. Rater 3 (FW), an esteemed scholar in medical informatics in China, possesses a master's degree in computer science and a PhD in social medicine. Being an expert in app quality evaluation, rater 3 boasts rich experience and high authority in research on assessing mobile app quality. Both rater 1 and rater 2 were recruited as real patients with HIV.

The evaluations by the researchers and patients with HIV were conducted separately to avoid any interference. Each group followed a 3-step evaluation process. First, basic information for 21 apps, including their names, developers, platforms, and core functionalities, was collected from the SevenMa platform, Aladdin WeChat platform, and WeChat. Second, 2 WeChat applets and 2 independent apps were randomly selected for pilot evaluation. Before conducting the pilot evaluation, all 3 raters watched training videos on the MARS scale to better understand the purpose and significance of each item on the scale. Third, the 21 collected apps underwent a formal evaluation. To ensure consistent evaluation results, rater 1 and rater 2 independently assessed the same samples, completing the evaluation of the nth subdimension for all samples before moving on to evaluate the n+1th dimension. In cases where rater 1 and rater 2 had conflicting evaluations, we applied the Delphi method, which involved multiple rounds of consultation and feedback with experts, to achieve agreement regarding the discrepancies. Rater 3, an expert with relevant qualifications, reassessed the dimension that exhibited disagreement. The evaluation results from rater 3 were then shared with rater 1 and rater 2, along with appropriate explanations. This gave rater 1 and rater 2 the opportunity to revise their ratings and evaluations. If substantial differences persisted even after re-evaluation, the entire process was repeated until their evaluations converged or reached a satisfactory level of agreement.

Statistical Analysis

All analyses were performed using MedCalc software (version 20.217, MedCalc Software). Descriptive scores were derived from the MARS and uMARS scales. To evaluate the reliability of the raters' assessments, Bland-Altman analysis was used to assess both interrater agreement and the range of variability in their scores. Visual representations were also used to visualize the differences in ratings between the 2 evaluators, providing a more concrete quantification of the discrepancies.

Ethical Considerations

This study focused solely on mobile apps and did not involve the collection of any personal information or data. Therefore, there were no ethical concerns or disputes associated with this research.

Results

Overview

A total of 21 apps that met the criteria were obtained through screening, with 18 being WeChat applets, dominating in terms of quantity, whereas only 3 were independent apps (Figure 1). Representative AIDS mobile apps were also displayed, as detailed in Figure 2. The characteristics of these apps are shown in Tables 1 and 2.

The developers of these apps can be categorized into 3 groups: companies (6/21, 29%), government organizations (7/21, 33%), and nonprofit social organizations (8/21, 38%). In addition, of the 21 included apps, 18 (86%) provided HIV/AIDS knowledge dissemination, 12 (57%) offered appointment booking for testing, 11 (52%) had counseling services, 8 (38%) had features related to PrEP and postexposure prophylaxis (PEP), 2 (10%) had community networking functionalities, and only 1 (5%) had live streaming capabilities. Of the 21 apps included in the study, 16 (76%) were able to update their content within a year, whereas 5 (24%) had not been updated for >1 year. Only 6 apps had privacy protection features, all of which provided privacy policies or agreements, with 2 of the apps offering privacy password functionality.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart of the selection process for inclusion of the apps.

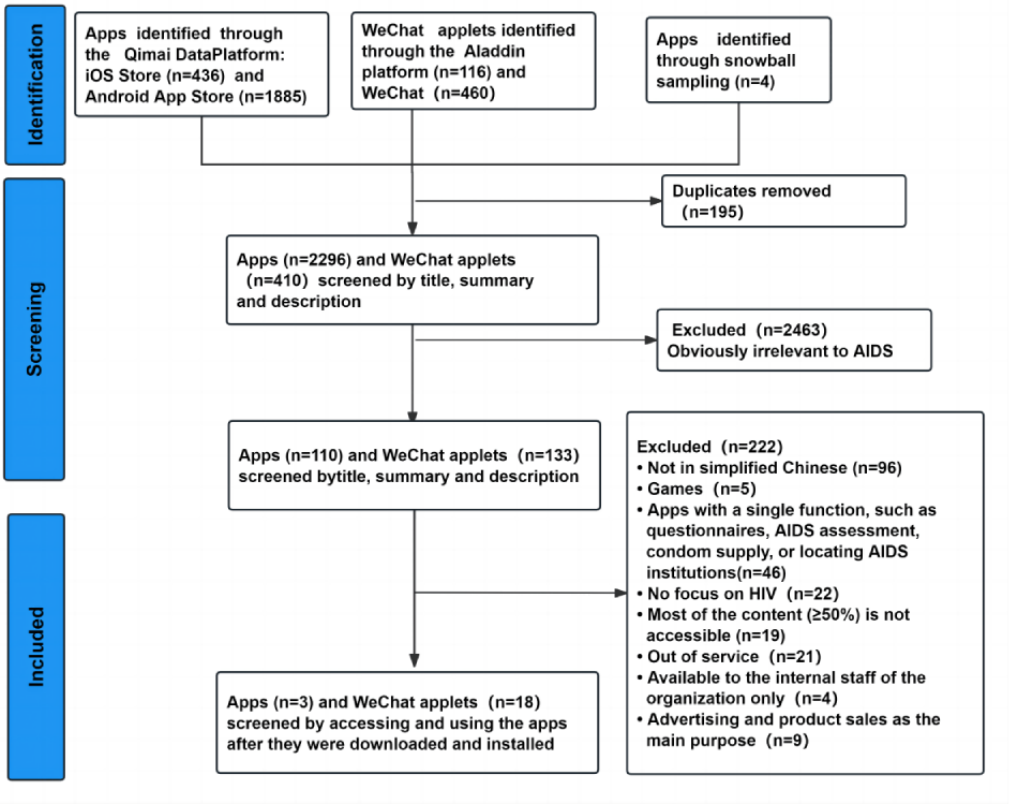


Figure 2. Example of a typical Chinese AIDS mobile app.

Table 1. General characteristics of the AIDS mobile apps (N=21).

Features	Total apps, n (%)	Independent apps (n=3), n (%)	WeChat applets (n=18), n (%)
Affiliation			
Company	6 (29)	3 (14)	3 (14)
Government organization	7 (33)	0 (0)	7 (33)
Nonprofit social organization	8 (38)	0 (0)	8 (38)
Content type			
Live streaming	1 (5)	0 (0)	1 (5)
Appointment for testing	15 (71)	3 (14)	12 (57)
Consultation	10 (48)	2 (10)	8 (38)
AIDS knowledge	18 (86)	2 (10)	16 (76)
PrEP[a] or PEP[b]	8 (38)	3 (14)	5 (24)
Web community	2 (10)	1 (5)	1 (5)
Last updated			
Within 1 month	5 (24)	1 (5)	4 (19)
Within 2 to 6 months	8 (38)	2 (10)	6 (29)
Within 7 to 12 months	3 (14)	0 (0)	3 (14)
>1 year	5 (24)	0 (0)	5 (24)
Privacy			
Privacy policy or agreement	6 (29)	3 (14)	3 (14)
Privacy password	2 (10)	2 (10)	0 (0)

[a]PrEP: pre-exposure prophylaxis.

[b]PEP: postexposure prophylaxis.

Table 2. Basic information about the AIDS mobile apps.

App name	App OR WC[a]	Knowledge sharing	Appointment testing	Live streaming	Consultation	Risk assessments	PrEP[b] and PEP[c]	Web-based community	Privacy protection
Chabei	WC	✓	✓		✓		✓		✓
Danlan Happy Test	WC	✓	✓						
Suzhou Red Ribbon	WC	✓			✓				✓
Baiyin HIV test	WC	✓	✓		✓		✓		
Linqu County CDC[d]	WC	✓	✓			✓			
Wu Ai Fang Hua	WC	✓						✓	
E Ai Jian	WC		✓		✓	✓	✓		
Ai Yi Jian	WC	✓	✓				✓		
Rong Ai Jian	WC	✓	✓		✓	✓			
Liaocheng Dongchangfu District Anti-AIDS Service Platform	WC	✓			✓	✓	✓		
Zhecheng County AIDS consulting and test	WC	✓	✓		✓	✓			
Qingai Health Services	WC	✓	✓		✓				
Ai Zhiku	WC	✓							
Ai Cheng Wang Shi	WC	✓	✓						
Beijing AIDS Association	WC		✓	✓					
Douai Check	WC	✓	✓			✓			
Nanyue Gaozhibao	WC	✓							
Red Ribbon Volunteer House	WC	✓							
Hong Feng Wan	App	✓	✓		✓		✓	✓	✓
Life4me+	App		✓				✓		✓
Xiao Ai	App	✓	✓		✓		✓		✓

[a]WC: WeChat applet.

[b]PrEP: pre-exposure prophylaxis.

[c]PEP: postexposure prophylaxis.

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

Functionality of the App

Most mobile apps (18/21, 86%) provided HIV/AIDS knowledge dissemination, and 71% (15/21) of the apps offered appointment booking for testing. Nearly half of the apps (10/21, 48%) provided counseling services, whereas 38% (8/21) of the apps offered features related to PrEP or PEP. Only 10% (2/21) of the apps had functionalities for live streaming and web-based community.

In addition to these functions related to function evaluation, we also included additional statistics on privacy protection settings. Among the 21 apps included in the study, only 6 (29%) provided privacy protection settings and 2 (10%) designed privacy passwords.

Quality of the App

Overview of App Composite Scores

The composite scores for the app quality total score in MARS and uMARS were obtained by averaging the scores for each app. The overall composite score for the 21 included apps was 3.43 ($mean_{app\ quality\ total\ score\ of\ MARS}$=3.47, SD 0.37; $mean_{app\ quality\ total\ score\ of\ uMARS}$=3.38, SD 0.53). Hong Feng Wan achieved the highest composite score, with a score of 4.22 ($mean_{app\ quality\ total\ score\ of\ MARS}$=3.96; $mean_{app\ quality\ total\ score\ of\ uMARS}$=4.47), followed by Wu Ai Fang Hua with a composite score of 3.9 ($mean_{app\ quality\ total\ score\ of\ MARS}$=4.06; $mean_{app\ quality\ total\ score\ of\ uMARS}$=3.74), and Chabei secured the third rank with a composite score of 3.76 ($mean_{app\ quality\ total\ score\ of\ MARS}$=3.66; $mean_{app\ quality\ total\ score\ of\ uMARS}$=3.86). Suzhou Red Ribbon obtained the lowest composite score, with a score of 2.68 ($mean_{app\ quality\ total\ score\ of\ MARS}$=3.67; $mean_{app\ quality\ total\ score\ of\ uMARS}$=1.68).

Comparative Analysis of MARS Score and uMARS Score

Figure 3 presents a correlation analysis between the scores of MARS and uMARS. No significant relationship was observed between the app total quality scores of MARS and uMARS (0.41; *P*=.07) or between their subjective quality scores (0.39; *P*=.08). The app most reflective of the disparity between MARS and uMARS is Suzhou Red Ribbon, which had the greatest discrepancy between its app total quality scores on MARS (3.67) and uMARS (1.68). Its score on uMARS (1.68) was the lowest in the ranking, with respective dimensional scores of engagement (1.8), functionality (1.5), esthetics (1.67), information (1.75), and subjective quality (1). Conversely, for MARS, it ranked seventh for the app quality total score, with respective dimensional scores of engagement (2.9), functionality (4.38), esthetics (3.5), information (3.92), and subjective quality (2.5). There was considerable variation disclosed in the rankings of the most highly rated apps between MARS and uMARS. The top 5 app total quality scores on MARS were Wu Ai Fang Hua (4.06), Hong Feng Wan (3.96), Ai Yi Jian (3.95), Qing Ai Health Services (3.77), and Ai Zhiku (3.76), whereas the top 5 on uMARS were Hong Feng Wan (4.47), Chabei (3.86), Wu Ai Fang Hua (3.74), Linqu County CDC (3.66), and Zhecheng County AIDS consulting and test (3.62).

Interpreting composite data from Figures 4 and 5 and Table 3 reveal that within the 4 MARS scale dimensions of the evaluated 21 apps, the functionality dimension achieved the highest score (4.28, SD 0.39).

The information (mean 3.82, SD 0.33), aesthetics (mean 3.4, SD 0.49), and engagement (mean 2.4, SD 0.53) dimensions sequentially trailed behind. In contrast, under the uMARS scale's evaluation, the information dimension prevailed with the top score (mean 3.89, SD 0.74), followed by the functionality (mean 3.84, SD 0.59), aesthetics (mean 3.21, SD 0.48), and engagement (mean 2.56, SD 0.63) dimensions.

Figure 3. Results of correlation analysis between the Mobile App Rating Scale (MARS) and the User Mobile App Rating Scale (uMARS) scores.

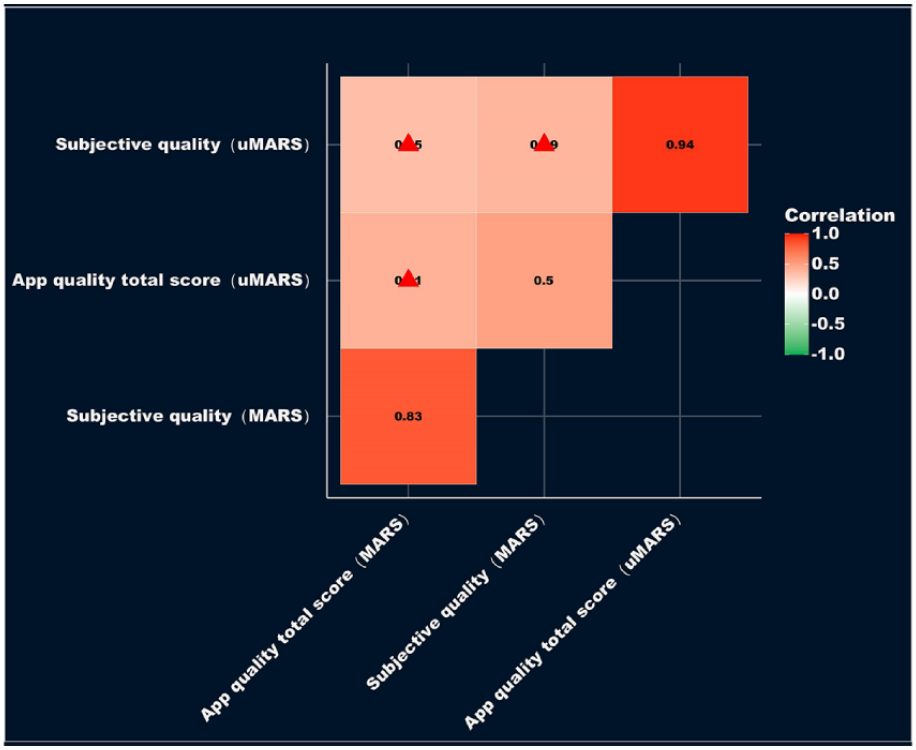


Figure 4. Box plot of the Mobile App Rating Scale (MARS) score.

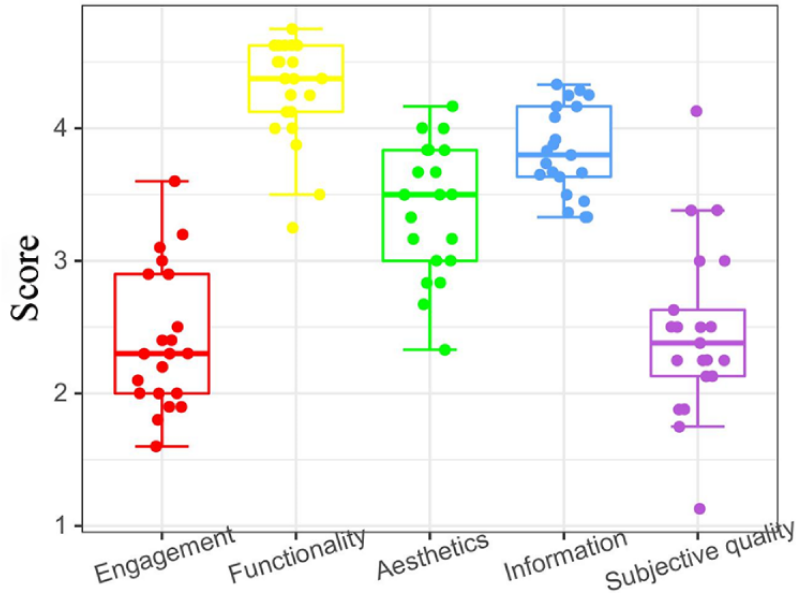


Figure 5. Box plot of the User Mobile App Rating Scale (uMARS) score.

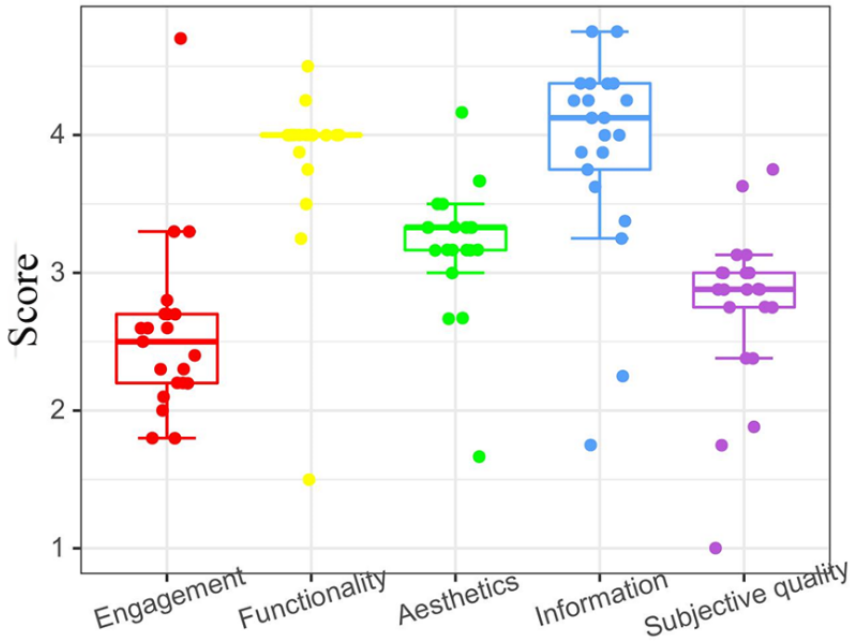


Table 3. The MARS[a] and uMARS[b] scales' scores for apps. The top 5 apps with app quality total score for each dimension are italicized.

App name	App quality ratings											
	Section A: engagement		Section B: functionality		Section C: esthetics		Section D: information		App quality total score		Section E: subjective quality	
	MARS	uMARS	MARS	uMARS	MARS	uMARS	MARS	uMARS	MARS	uMARS	MARS	uMARS
Ai Cheng Wang Shi	2.00	2.10	4.25	4.00	2.33	3.33	3.50	3.88	3.02	3.33	1.88	2.75
Ai Yi Jian[c]	*3.00*	2.80	*4.63*	4.00	*4.00*	3.17	*4.17*	4.13	*3.95*	3.52	3.38	2.75
Ai Zhiku[c]	*2.50*	2.30	*4.38*	4.00	*4.00*	3.17	*4.17*	4.38	*3.76*	3.46	3.00	3.00
Baiyin HIV test	2.30	2.60	4.63	4.00	3.50	3.33	3.67	4.00	3.52	3.48	2.50	2.88
Beijing AIDS Association	2.40	2.20	3.50	4.00	3.33	3.17	3.65	4.25	3.22	3.40	2.50	2.88
Chabei[d]	2.20	*2.70*	4.50	*4.50*	3.67	*3.50*	4.29	*4.75*	3.66	*3.86*	3.00	3.75
Danlan Happy Test	1.60	2.30	3.25	3.75	2.84	3.00	3.33	2.25	2.63	2.83	1.13	2.38
Dou Ai Jian	1.90	1.80	4.13	3.25	3.00	2.67	3.45	3.25	3.12	2.74	2.25	1.75
E Ai Jian	2.30	2.70	4.75	4.00	3.84	3.33	3.74	4.38	3.66	3.60	2.25	3.00
Red Ribbon Volunteer House	2.00	2.50	3.88	4.00	3.00	3.17	3.37	4.13	3.06	3.45	2.13	2.88
Liaocheng Dongchangfu District Anti-AIDS Service Platform	2.00	2.20	4.50	4.00	3.50	3.33	4.25	4.25	3.56	3.45	2.25	2.75
Linqu County CDC[d]	2.10	*2.60*	4.63	*4.00*	3.17	*3.67*	3.80	*4.38*	3.42	*3.66*	2.50	3.00
Nanyue Gaozhibao	1.90	2.20	4.00	4.00	2.84	3.17	3.67	3.75	3.10	3.28	1.75	2.88
Qing Ai Health Services[c]	*2.90*	3.30	*4.50*	4.00	*3.84*	3.17	*3.83*	3.88	*3.77*	3.59	2.63	2.88
Rong Ai Jian	2.30	2.40	4.63	3.88	3.84	3.33	3.33	3.38	3.52	3.25	2.13	2.38
Suzhou Red ribbon	2.90	1.80	4.38	1.50	3.50	1.67	3.92	1.75	3.67	1.68	2.50	1.00
Wu Ai Fang Hua[c,d]	*3.20*	*3.30*	*4.63*	*4.00*	*4.17*	*3.67*	*4.25*	*4.00*	*4.06*	*3.74*	3.38	3.13
Zhecheng County AIDS consulting and test[d]	2.40	*2.60*	4.00	*4.00*	3.17	*3.50*	4.09	*4.38*	3.41	*3.62*	1.88	3.00
Life4me+	1.80	2.00	4.13	3.50	2.67	2.67	3.64	3.63	3.06	2.95	2.25	1.88
Hong Feng Wan[c,d]	*3.60*	*4.70*	*4.25*	*4.25*	*3.67*	*4.17*	*4.33*	*4.75*	*3.96*	*4.47*	4.13	3.63
Xiao Ai	3.10	2.70	4.38	4.00	3.50	3.33	3.88	4.25	3.71	3.57	2.38	3.13

[a]MARS: Mobile App Rating Scale.

[b]uMARS: User Mobile App Rating Scale.

[c]The top 5 apps in terms of app quality total score in MARS.

[d]The top 5 apps in terms of app quality total score in uMARS.

Comparative Analysis of Quality Between WeChat Mini-Apps and Independent Apps

As shown in Table 4, insignificant disparities are discernible between WeChat mini-apps and stand-alone apps regarding 2 metric domains: app quality (P_{MARS}=.70; P_{uMARS}=.54) and subjective app quality (P_{MARS}=.48; P_{uMARS}=.80).

Table 4. Independent samples *t* test for unequal variances.

MARS[a] quality	MARS, mean (SD)	uMARS[b], mean (SD)	MARS, *t* value (*df*)	uMARS, *t* value (*df*)	MARS, *P* value	uMARS, *P* value
App quality	__[c]	—	−0.448 (2.4)	0.73 (2.3)	.70	.54
WeChat applets	3.45 (0.37)	3.33 (0.50)	—	—	—	—
Independent apps	3.58 (0.47)	3.66 (0.76)	—	—	—	—
App subjective quality	—	—	−0.852 (2.2)	0.289 (2.3)	.48	.80
WeChat applets	2.39 (0.57)	2.72 (0.58)	—	—	—	—
Independent apps	2.92 (1.05)	2.88 (0.90)	—	—	—	—

[a]MARS: Mobile App Rating Scale.

[b]uMARS: User Mobile App Rating Scale.

[c]Not available.

Internal Consistency and Reliability Testing for MARS and uMARS Scores

Similarly, we formed 4 Bland-Altman plots using the disparities and mean values between the scorings by reviewers on app quality (both MARS and uMARS) and app subjective quality (MARS and uMARS). The relatively limited range of the 95% limits of agreements suggest that the evaluators' judgment outputs contain minor dissimilarities. A significant proportion of dots in all 4 Bland-Altman plots lie within the concordance interval (Figures 6-9), with their arithmetic means impressionably approaching 0 (Table 5).

This provides evidence of the high degree of internal uniformity and dependability in both MARS and uMARS scores.

Figure 6. Bland-Altman plot of the app quality (Mobile App Rating Scale).

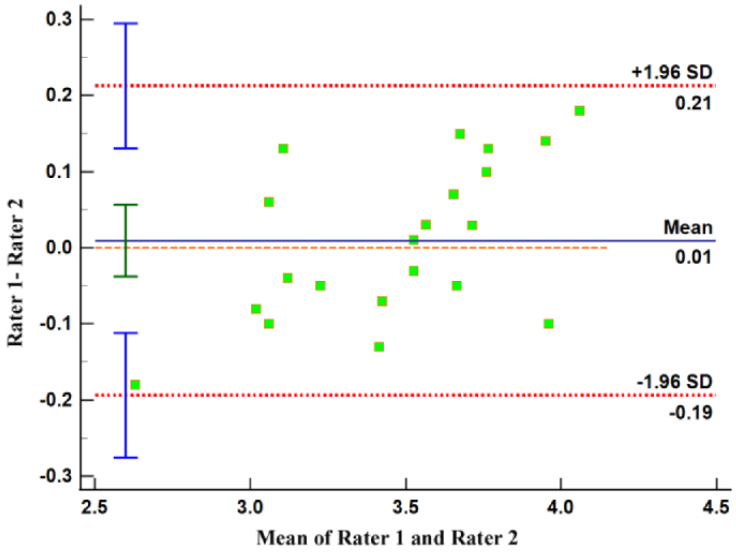


Figure 7. Bland-Altman plot of the app subjective quality (Mobile App Rating Scale).

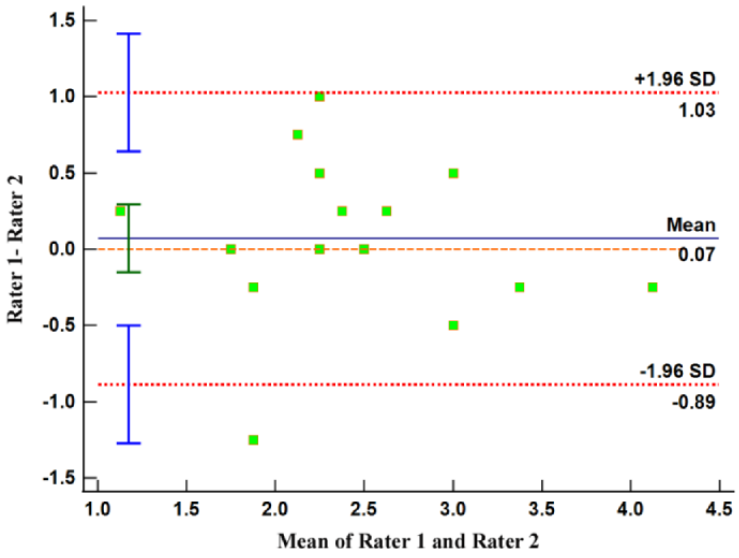


Figure 8. Bland-Altman plot of the app quality (User Mobile App Rating Scale).

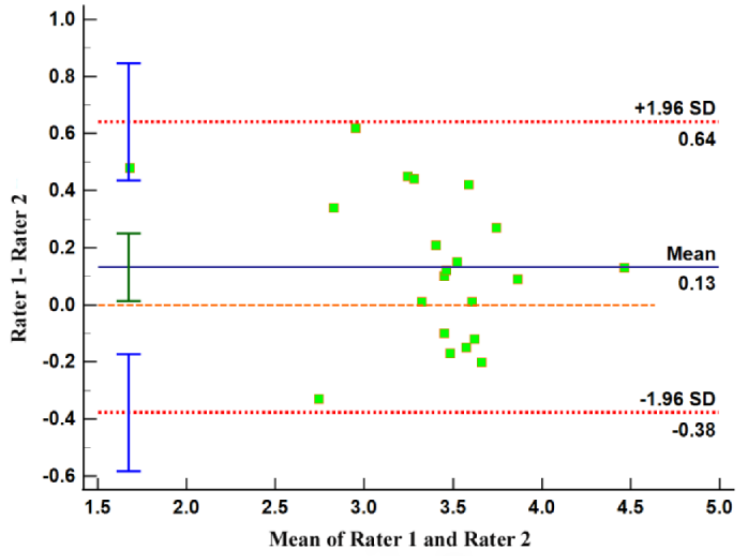


Figure 9. Bland-Altman plot of the app subjective quality (User Mobile App Rating Scale).

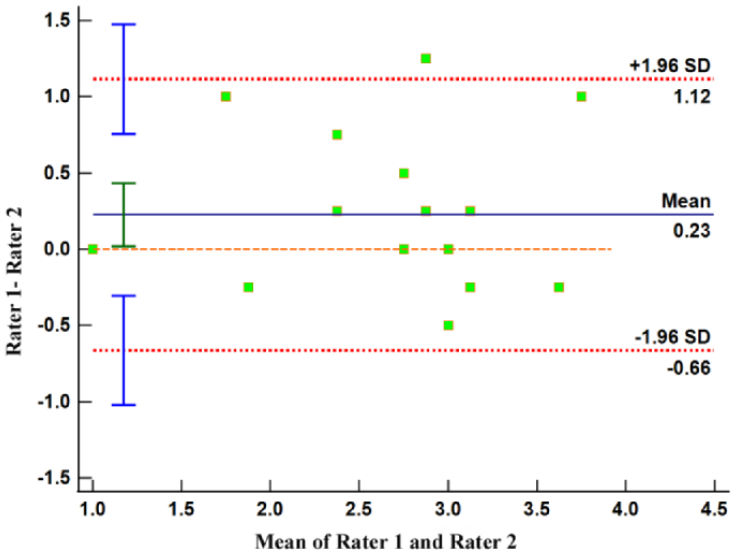


Table 5. Summary of results of Bland-Altman analysis.

Index	95% LoA[a]	95% CI (arithmetic mean)	95% CI (LoA: upper limit)	95% CI (LoA: lower limit)
App quality (MARS[b])	−0.19 to 0.21	−0.04 to 0.06	0.13 to 0.30	−0.28 to −0.11
App subjective quality (MARS)	−0.89 to 1.03	−0.15 to 0.29	0.64 to 1.41	−1.27 to −0.50
App quality (uMARS[c])	−0.38 to 0.64	0.01 to 0.25	0.44 to 0.85	−0.58 to −0.17
App subjective quality (uMARS)	−0.66 to 1.12	0.02 to 0.43	0.76 to 1.47	−1.02 to −0.30

[a]LoA: limits of agreement.

[b]MARS: Mobile App Rating Scale.

[c]uMARS: User Mobile App Rating Scale.

Discussion

Principal Findings

This study provides a comprehensive statistical summary of the functionalities and characteristics of 21 Chinese HIV-related mobile apps, along with their quality assessment using the MARS scale. The primary objectives behind the development of these mobile apps were the dissemination of HIV prevention knowledge (18/21, 86%), appointment booking for testing (15/21, 71%), and counseling services (10/21, 48%). These mobile apps aim to assist users in HIV prevention and treatment while improving the quality of communication and interaction between patients and health care providers.

Overview of App Functions

In recent years, the Chinese government has implemented a series of regulations and policies for HIV/AIDS prevention and has established an effective mechanism for HIV/AIDS control. These efforts have achieved certain progress in HIV/AIDS prevention. However, the burden of AIDS disease continues to increase, and the prevention and control situation remains severe. With the widespread adoption of smartphones and the rapid development of mobile internet, mobile apps have played an increasingly important role in disseminating HIV/AIDS knowledge and raising public awareness of prevention [1,37]. Among the 21 apps included in this study, 18 (86%) had HIV/AIDS knowledge dissemination functionality. This indicates that HIV/AIDS knowledge dissemination is an almost indispensable feature of Chinese HIV/AIDS mobile apps, which have become important tools for promoting HIV/AIDS knowledge in China.

In this study, 17 (81%) of the 21 mobile apps were identified, including 14 (67%) WeChat applets (such as Chabei, Danlan Happy Test, Baiyin HIV test, Suzhou Red Ribbon, Linqu County CDC, E Ai Jian, Ai Yi Jian, Rong Ai Jian, Liaocheng Dongchangfu District Anti-AIDS Service Platform, Zhecheng County AIDS consulting and test, Qing Ai Health Service, Ai Cheng Wang Shi, Beijing AIDS Association, and Dou Ai Jian) and 3 (14%) independent apps (Life4me+, Hong Feng Wan, and Xiao Ai). These apps provided various forms of health services, including appointment booking, counseling, risk assessment, and PrEP or PEP. A total of 2 mobile apps (Wu Ai Fang Hua and Hong Feng Wan) offered web-based community functionalities and scored excellently in terms of app quality and subjective quality, ranking among the top 2 positions. This finding is consistent with that of a previous study by Nour et al [38]. Therefore, we recommend considering the incorporation of web-based community functions when developing and designing mobile apps related to AIDS. This can encourage users to share their experiences, ask questions, provide

suggestions, and offer mutual assistance and support. Organizing regular web-based or offline activities can also promote face-to-face interactions among users, thereby enhancing their sense of belonging and engagement.

Privacy protection is another crucial topic that should be given sufficient attention in AIDS-related mobile apps. As social stigmatization and discrimination are still associated with AIDS, many users are concerned about their personal information being disclosed or misused, posing a threat to their privacy and security [39]. Therefore, unlike general mobile apps, ensuring user privacy protection in AIDS mobile apps is of utmost importance, aligning with previous research findings [40]. However, in this study, of the 21 apps, only 5 (24%) mobile apps, including 2 (10%) WeChat applets (Chabei and Su Zhou Red Ribbon) and 3 (14%) apps (My Life+, Hong Feng Wan, and Xiao Ai), provided privacy policies or agreements. The average subjective quality score (mean 2.85, SD 0.77) was significantly higher than that of the mobile apps without privacy protection features (mean 2.34, SD 0.58). This observed phenomenon may be attributed to the fact that inadequate privacy protection design may lead to user wariness and lower intentions and frequencies of use regarding AIDS mobile apps [41]. Therefore, we strongly suggest that all AIDS-related mobile apps incorporate privacy policies or agreements to alleviate user concerns about privacy protection, enhance user trust, and promote willingness to use these apps. In addition to transparent privacy policies and user agreements, user education is also critical. Developers of AIDS mobile apps should provide users with relevant information and educational resources on privacy protection. This can be achieved through in-app prompts, tutorials, and frequently asked questions. Users should receive clear instructions on configuring their privacy settings, handling their personal data, and the steps to take in reporting privacy concerns or data breaches. These instructions should be easily accessible, comprehensible, and provide specific examples to guide users through each process.

Rating of App Quality

According to the results from MARS scores, among the 4 dimensions of app quality, both researchers and patients with HIV rated 21 mobile apps lowest on engagement ($mean_{MARS}$ 2.4, SD 0.53; $mean_{uMARS}$ 2.56, SD 0.63), especially the WeChat mini-apps, which had the lowest average scores ($mean_{MARS}$ 2.33, SD 0.43; $mean_{uMARS}$ 2.47, SD 0.42). The average engagement scores for the 7 mobile apps developed by the Chinese government were even lower ($mean_{MARS}$ 2.27, SD 0.33; $mean_{uMARS}$ 2.34, SD 0.30). This underlined the concern that the level of engagement would remain low, even for officially developed apps. An analysis of the respective scores for each item within the MARS and uMARS engagement portions revealed a predominant lack of interest and amusement within these mobile apps, thus leading to decreased user engagement. Many AIDS-related mobile apps merely provide basic information and functions without interactive and stimulating designs to attract user participation, lacking a strong appeal to users. In addition, the absence of personalized and customized features limits user engagement. A study has shown a significant correlation between user engagement and an increase in the adoption rate of mobile apps [42]. Therefore, it is recommended that developers not only provide high-quality HIV/AIDS prevention and treatment information but also focus on meeting users' needs in terms of multidimensionality, functionality, and depth. In addition, attention should be paid to design in terms of amusement, entertainment, customization, interactivity, and other participatory aspects. This will help attract more user participation and enhance user stickiness.

This investigation revealed a lack of correlation between the MARS and uMARS scores. Specifically, (1) the app with the lowest overall uMARS scores across all dimensions surprisingly ranked seventh in terms of MARS scores. (2) A notable discrepancy was found in the functionality dimension ratings among patients with HIV and researchers, exposing the highest demographic variance in this attribute. (3) Moreover, the apps predominantly preferred by patients with HIV exhibited robust performance in functionality and information dimensions, with the latter appearing particularly predominant. However, the apps gravitating toward researchers demonstrated high competence in functionality, esthetics, and information, with functionality being the most superior. These data imply possible significant divergences distinct between the evaluations of researchers and genuine users of the apps. To maintain superior app quality and consumer satisfaction, rigorous surveillance of app quality should be sustained from both research and real-user perspectives. This bifocal assessment permits the accurate identification of genuine user requirements and researcher appraisal, providing valuable insights for the pinpoint and scientific enhancement of both app quality and user experience.

In this study, we observed that most Chinese AIDS-related mobile apps are WeChat applets. A reason for this is that the WeChat applet has advantages such as not requiring downloading or installation, having minimal resource consumption, and high user retention rates, which independent apps do not possess [43]. Another reason is that the HIV/AIDS population is relatively niche [44], which means that the market for AIDS mobile apps represents a low-frequency demand niche market. Independent apps often overlook this niche market owing to their high development costs, whereas WeChat applets, with their low development costs and the ability to cater to low-frequency demand niches through segmented scene construction, can effectively meet the needs of this market. Although WeChat mini-apps dominate the landscape of mobile apps for HIV/AIDS in China, their comprehensive performance lags behind that of stand-alone apps. The app quality score ($mean_{MARS}$ 3.45, SD 0.37 vs 3.58, SD 0.47; $mean_{uMARS}$ 3.33, SD 0.50 vs 3.66, SD 0.76) and subjective quality score ($mean_{MARS}$ 2.39, SD 0.57 vs 2.92, SD 1.05; $mean_{uMARS}$ 2.72, SD 0.58 vs 2.88, SD 0.90) for WeChat mini-apps are both lower compared with stand-alone apps. A granular analysis of the scores in different dimensions reveals the greatest discrepancy in the engagement dimension, with WeChat mini-apps scoring markedly lower ($mean_{MARS}$ 2.33, SD 0.43 vs 2.83, SD 0.93; $mean_{uMARS}$ 2.47, SD 0.42 vs 3.13, SD 1.40). This can be attributed to the constraints and limitations imposed by the WeChat platform, which prevent WeChat applets from providing the same user experience as independent apps in terms of

interface design, interaction methods, and customization settings. Improvement in these aspects should be a key focus for the future development of AIDS-related WeChat applets.

Limitations

The mobile apps evaluated in this study represent a snapshot of the current status of Chinese HIV-related mobile apps during the research period. Over time, mobile apps may be removed or updated, so the list of included mobile apps meeting the criteria may change in the future. In addition, some mobile apps intended solely for internal organizational use or no longer available in the market were excluded.

Conclusions

This study presents a systematic introduction to the functionality and quality of the currently available Chinese mobile apps for AIDS, providing valuable decision-making support for AIDS-related groups and health care professionals in accessing high-quality AIDS mobile apps. Through our systematic search and evaluation of existing Chinese mobile apps for AIDS, it has been observed that because of the lower demand frequency of AIDS mobile apps in China, less costly WeChat mini-apps have become the primary mode of app, and the overall quality attains a merely average level. A significant implication of our research is identifying the potentially significant discrepancy between the assessments made by researchers and the authentic users of the apps. Consequently, the inclusion of genuine users during the assessment and refinement stages of HIV apps is crucial. The main purpose of developing these mobile apps is to spread HIV prevention knowledge and facilitate booking appointments for testing and counseling services. However, most of these apps lack privacy protection features. Unlike general mobile apps, privacy protection is especially crucial in AIDS-related mobile apps because it directly affects users' willingness to use them [40]. Therefore, the introduction of legal and ethical frameworks for privacy protection as well as privacy protection technologies is essential. In addition, enhancing user education on privacy protection and ensuring informed consent is of utmost importance. Research related to privacy protection in Chinese AIDS mobile apps may be a vital and urgent topic to address in the future. It is our intent that the findings of our research may function as a road map and reference for the future development of HIV apps in China. Furthermore, we aim to provide crucial decision-making support for individuals living with HIV in their quest for superior HIV apps.

Acknowledgments

The authors sincerely acknowledge the editors and anonymous reviewers for their insights and comments to further improve the quality of the manuscript. The authors thank Wang Xiang, chief of the AIDS Prevention Section of Bengbu Centers for Disease Control and Prevention, for his great help in the investigation. Finally, the first author (PL) would like to personally express his gratitude to his son (Liu Daixin) and his family for their support all through this work. This work was supported by the Anhui Provincial University Philosophy and Social Science Major Project (grant 2023AH040283).

Data Availability

The data sets generated during and/or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

PL and FW initiated this study. PL and FW contributed to the study design. PL drafted the manuscript. LW conducted data sorting and screening. All authors revised the manuscript and approved the final version of the manuscript.

Conflicts of Interest

None declared.

References

1. Wu Z, Chen J, Scott SR, McGoogan JM. History of the HIV epidemic in China. Curr HIV/AIDS Rep 2019 Dec 26;16(6):458-466 [FREE Full text] [doi: 10.1007/s11904-019-00471-4] [Medline: 31773405]
2. Sharpe JD, Kamara MT. A systematic evaluation of mobile apps to improve the uptake of and adherence to HIV pre-exposure prophylaxis. Sex Health 2018 Nov;15(6):587-594. [doi: 10.1071/SH18120] [Medline: 30347177]
3. Muessig KE, Nekkanti M, Bauermeister J, Bull S, Hightow-Weidman LB. A systematic review of recent smartphone, internet and web 2.0 interventions to address the HIV continuum of care. Curr HIV/AIDS Rep 2015 Mar;12(1):173-190 [FREE Full text] [doi: 10.1007/s11904-014-0239-3] [Medline: 25626718]
4. Biello KB, Hill-Rorie J, Valente PK, Futterman D, Sullivan PS, Hightow-Weidman L, et al. Development and evaluation of a mobile app designed to increase HIV testing and pre-exposure prophylaxis use among young men who have sex with men in the united states: open pilot trial. J Med Internet Res 2021 Mar 24;23(3):e25107 [FREE Full text] [doi: 10.2196/25107] [Medline: 33759792]

5. Downing Jr MJ, Wiatrek SE, Zahn RJ, Mansergh G, Olansky E, Gelaude D, et al. Video selection and assessment for an app-based HIV prevention messaging intervention: formative research. Mhealth 2023 Jan;9:2-2 [FREE Full text] [doi: 10.21037/mhealth-21-53] [Medline: 36760783]
6. Grov C, Stief M, Westmoreland DA, MacCrate C, Mirzayi C, Nash D, Together 5000 Study Team. Maximizing response rates to ads for free at-home HIV testing on a men-for-men geosocial sexual networking app: lessons learned and implications for researchers and providers. Health Educ Behav 2020 Feb 03;47(1):5-13 [FREE Full text] [doi: 10.1177/1090198119893692] [Medline: 31896287]
7. Di Mascio M, Markowitz M, Louie M, Hogan C, Hurley A, Chung C, et al. Viral blip dynamics during highly active antiretroviral therapy. J Virol 2003 Nov 15;77(22):12165-12172 [FREE Full text] [doi: 10.1128/jvi.77.22.12165-12172.2003] [Medline: 14581553]
8. Beauchemin M, Gradilla M, Baik D, Cho H, Schnall R. A multi-step usability evaluation of a self-management app to support medication adherence in persons living with HIV. Int J Med Inform 2019 Feb;122:37-44 [FREE Full text] [doi: 10.1016/j.ijmedinf.2018.11.012] [Medline: 30623782]
9. Yang LW, Renfang NI, Xiulan W. The measures and effects evaluation of case management in HIV-infected women during perinatal period. Adv Nurs Pract 2016;16(3):421. [doi: 10.3969/j.issn.1672-1756.2016.03.035]
10. Wen YL, Xian-yu Y, Jing-qiu F. Influence of case intervention on behaviors of homosexual behavior patients with HIV infection. Chin J Nosocomiol 2018;28(22):3413. [doi: 10.11816/cn.ni.2018-173749]
11. Tian Bo LJ, Kun-li W, Rong H. Exploration of case management mode in antiretroviral therapy. J Dermatol Venereol 2018;40(2):173. [doi: 10.3969/j.issn.1002-1310.2018.02.007]
12. Kong F. WH-q: research progress for nursing interventions for HIV/AIDS patients. Chinese Gen Pract Nurs 2018;16(1):24. [doi: 10.3969/j.issn.1674-4748.2018.01.007]
13. Schnall R, Mosley JP, Iribarren SJ, Bakken S, Carballo-Diéguez A, Brown Iii W. Comparison of a user-centered design, self-management app to existing mhealth apps for persons living with HIV. JMIR Mhealth Uhealth 2015 Sep 18;3(3):e91 [FREE Full text] [doi: 10.2196/mhealth.4882] [Medline: 26385783]
14. Schnall R, Rojas M, Bakken S, Brown W, Carballo-Dieguez A, Carry M, et al. A user-centered model for designing consumer mobile health (mHealth) applications (apps). J Biomed Inform 2016 Apr;60:243-251 [FREE Full text] [doi: 10.1016/j.jbi.2016.02.002] [Medline: 26903153]
15. Yang G, Long J, Luo D, Xiao S, Kaminga AC. The characteristics and quality of mobile phone apps targeted at men who have sex with men in China: a window of opportunity for health information dissemination? JMIR Mhealth Uhealth 2019 Mar 27;7(3):e12573 [FREE Full text] [doi: 10.2196/12573] [Medline: 30916658]
16. Beach LB, Greene GJ, Lindeman P, Johnson AK, Adames CN, Thomann M, et al. Barriers and facilitators to seeking HIV services in Chicago among young men who have sex with men: perspectives of HIV service providers. AIDS Patient Care STDS 2018 Nov;32(11):468-476 [FREE Full text] [doi: 10.1089/apc.2018.0094] [Medline: 30398956]
17. Catalani C, Philbrick W, Fraser H, Mechael P, Israelski DM. mHealth for HIV treatment and prevention: a systematic review of the literature. Open AIDS J 2013;7:17-41 [FREE Full text] [doi: 10.2174/1874613620130812003] [Medline: 24133558]
18. Devi BR, Syed-Abdul S, Kumar A, Iqbal U, Nguyen P, Li YJ, et al. mHealth: an updated systematic review with a focus on HIV/AIDS and tuberculosis long term management using mobile phones. Comput Methods Programs Biomed 2015 Nov;122(2):257-265. [doi: 10.1016/j.cmpb.2015.08.003] [Medline: 26304621]
19. Rossman K, Salamanca P, Macapagal K. A qualitative study examining young adults' experiences of disclosure and nondisclosure of LGBTQ identity to health care providers. J Homosex 2017 Jun 28;64(10):1390-1410 [FREE Full text] [doi: 10.1080/00918369.2017.1321379] [Medline: 28459379]
20. Mokgatle MM, Madiba S. High acceptability of HIV self-testing among technical vocational education and training college students in Gauteng and North West province: what are the implications for the scale up in South Africa? PLoS One 2017 Jan 31;12(1):e0169765 [FREE Full text] [doi: 10.1371/journal.pone.0169765] [Medline: 28141858]
21. Majam M, Quaife M, Phatsoane M, Rhagnath N, Quaife M. High self-reporting of HIV self-test results through an interactive voice response telephone line in inner city Johannesburg. Poster published in IAS Conference on HIV Science. 2019. URL: https://programme.ias2019.org/PAGMaterial/eposters/5054.pdf [accessed 2024-01-04]
22. De Boni RB, Lentini N, Santelli AC, Barbosa A, Cruz M, Bingham T, et al. Self-testing, communication and information technology to promote HIV diagnosis among young gay and other men who have sex with men (MSM) in Brazil. J Int AIDS Soc 2018 Jul 22;21 Suppl 5(Suppl Suppl 5):e25116 [FREE Full text] [doi: 10.1002/jia2.25116] [Medline: 30033612]
23. Janssen R, Engel N, Esmail A, Oelofse S, Krumeich A, Dheda K, et al. Alone but supported: a qualitative study of an HIV self-testing app in an observational cohort study in South Africa. AIDS Behav 2020 Feb 2;24(2):467-474 [FREE Full text] [doi: 10.1007/s10461-019-02516-6] [Medline: 31049809]
24. Pant Pai N, Smallwood M, Desjardins L, Goyette A, Birkas KG, Vassal A, et al. An unsupervised smart app-optimized HIV self-testing program in Montreal, Canada: cross-sectional study. J Med Internet Res 2018 Nov 27;20(11):e10258 [FREE Full text] [doi: 10.2196/10258] [Medline: 30465709]
25. Chan PS, Chidgey A, Lau J, Ip M, Lau JT, Wang Z. Effectiveness of a Novel HIV self-testing service with online real-time counseling support (HIVST-online) in increasing HIV testing rate and repeated HIV testing among men who have sex with

men in Hong Kong: results of a pilot implementation project. Int J Environ Res Public Health 2021 Jan 15;18(2):729 [FREE Full text] [doi: 10.3390/ijerph18020729] [Medline: 33467770]
26. Witzel TC, Gabriel MM, McCabe L, Weatherburn P, Gafos M, Speakman A, et al. Pilot phase of an internet-based RCT of HIVST targeting MSM and transgender people in England and Wales: advertising strategies and acceptability of the intervention. BMC Infect Dis 2019 Aug 08;19(1):699 [FREE Full text] [doi: 10.1186/s12879-019-4247-1] [Medline: 31391003]
27. Sanabria G, Scherr T, Garofalo R, Kuhns LM, Bushover B, Nash N, et al. Usability evaluation of the mLab app for improving home HIV testing behaviors in youth at risk of HIV infection. AIDS Educ Prev 2021 Aug;33(4):312-324 [FREE Full text] [doi: 10.1521/aeap.2021.33.4.312] [Medline: 34370566]
28. Gous N, Fischer AE, Rhagnath N, Phatsoane M, Majam M, Lalla-Edward ST. Evaluation of a mobile application to support HIV self-testing in Johannesburg, South Africa. South Afr J HIV Med 2020 Jun 30;21(1):1088 [FREE Full text] [doi: 10.4102/sajhivmed.v21i1.1088] [Medline: 32670629]
29. Thonon F, Fahmi S, Rousset-Torrente O, Bessonneau P, Griffith JW, Brown C, et al. Promoting HIV, Hepatitis B virus, and Hepatitis C virus screening among migrants with a language barrier: protocol for the development and evaluation of an electronic app (Apidé). JMIR Res Protoc 2021 May 05;10(5):e22239 [FREE Full text] [doi: 10.2196/22239] [Medline: 33949963]
30. Itoh N, Mishima H, Yoshida Y, Yoshida M, Oka H, Matsudaira K. Evaluation of the effect of patient education and strengthening exercise therapy using a mobile messaging app on work productivity in Japanese patients with chronic low back pain: open-label, randomized, parallel-group trial. JMIR Mhealth Uhealth 2022 May 16;10(5):e35867 [FREE Full text] [doi: 10.2196/35867] [Medline: 35576560]
31. Ellmann S, Maryschok M, Schöffski O, Emmert M. The German COVID-19 digital contact tracing app: a socioeconomic evaluation. Int J Environ Res Public Health 2022 Nov 02;19(21):14318 [FREE Full text] [doi: 10.3390/ijerph192114318] [Medline: 36361198]
32. Linardon J, King T, Shatte A, Fuller-Tyszkiewicz M. Usability evaluation of a cognitive-behavioral app-based intervention for binge eating and related psychopathology: a qualitative study. Behav Modif 2022 Sep;46(5):1002-1020. [doi: 10.1177/01454455211021764] [Medline: 34075803]
33. Schepens Niemiec SL, Wagas R, Vigen CL, Blanchard J, Barber SJ, Schoenhals A. Preliminary user evaluation of a physical activity smartphone app for older adults. Health Policy Technol 2022 Sep;11(3):100639 [FREE Full text] [doi: 10.1016/j.hlpt.2022.100639] [Medline: 36213682]
34. Luna-Perejon F, Malwade S, Styliadis C, Civit J, Cascado-Caballero D, Konstantinidis E, et al. Evaluation of user satisfaction and usability of a mobile app for smoking cessation. Comput Methods Programs Biomed 2019 Dec;182:105042 [FREE Full text] [doi: 10.1016/j.cmpb.2019.105042] [Medline: 31473444]
35. Zelmer J, van Hoof K, Notarianni M, van Mierlo T, Schellenberg M, Tannenbaum C. An assessment framework for e-mental health apps in Canada: results of a modified Delphi process. JMIR Mhealth Uhealth 2018 Jul 09;6(7):e10016 [FREE Full text] [doi: 10.2196/10016] [Medline: 29986846]
36. Manning JB, Blandford A, Edbrooke-Childs J. Digital companion choice to support teachers' stress self-management: systematic approach through taxonomy creation. JMIR Form Res 2022 Feb 16;6(2):e32312 [FREE Full text] [doi: 10.2196/32312] [Medline: 35171106]
37. Sun X, Lu F, Wu Z, Poundstone K, Zeng G, Xu P, et al. Evolution of information-driven HIV/AIDS policies in China. Int J Epidemiol 2010 Dec 26;39 Suppl 2(Suppl 2):ii4-i13 [FREE Full text] [doi: 10.1093/ije/dyq217] [Medline: 21113036]
38. Nour M, Chen J, Allman-Farinelli M. Young adults' engagement with a self-monitoring app for vegetable intake and the impact of social media and gamification: feasibility study. JMIR Form Res 2019 May 10;3(2):e13324 [FREE Full text] [doi: 10.2196/13324] [Medline: 31094322]
39. Gillette E, Naanyu V, Nyandiko W, Chory A, Scanlon M, Aluoch J, et al. HIV-related stigma shapes research participation for youth living with HIV in Kenya. J Int Assoc Provid AIDS Care 2023 Apr 26;22:23259582231170732 [FREE Full text] [doi: 10.1177/23259582231170732] [Medline: 37101381]
40. Khati A, Wickersham JA, Rosen AO, Luces JR, Copenhaver N, Jeri-Wahrhaftig A, et al. Ethical issues in the use of smartphone apps for HIV prevention in Malaysia: focus group study with men who have sex with men. JMIR Form Res 2022 Dec 23;6(12):e42939 [FREE Full text] [doi: 10.2196/42939] [Medline: 36563046]
41. Bui TM, Hoang MT, Ngo TV, Do CD, Nghiem SH, Byrnes J, et al. Smartphone use and willingness to pay for HIV treatment-assisted smartphone applications among HIV-positive patients in urban clinics of Vietnam. Int J Environ Res Public Health 2021 Feb 04;18(4):1467 [FREE Full text] [doi: 10.3390/ijerph18041467] [Medline: 33557412]
42. Kahnbach L, Lehr D, Brandenburger J, Mallwitz T, Jent S, Hannibal S, et al. Quality and adoption of COVID-19 tracing apps and recommendations for development: systematic interdisciplinary review of European apps. J Med Internet Res 2021 Jun 02;23(6):e27989 [FREE Full text] [doi: 10.2196/27989] [Medline: 33890867]
43. Fan Y, Wang Z, Deng S, Lv H, Wang F. The function and quality of individual epidemic prevention and control apps during the COVID-19 pandemic: a systematic review of Chinese apps. Int J Med Inform 2022 Apr;160:104694 [FREE Full text] [doi: 10.1016/j.ijmedinf.2022.104694] [Medline: 35144100]

44. Mastro TD, Kim AA, Hallett T, Rehle T, Welte A, Laeyendecker O, et al. Estimating HIV incidence in populations using tests for recent infection: issues, challenges and the way forward. J HIV AIDS Surveill Epidemiol 2010 Jan 01;2(1):1-14. [Medline: 21743821]

Abbreviations

CDC: Centers for Disease Control and Prevention
GBMSM: gay, bisexual, and other men who have sex with men
HIVST: HIV self-testing
MARS: Mobile App Rating Scale
mHealth: mobile health
PEP: postexposure prophylaxis
PrEP: pre-exposure prophylaxis
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
uMARS: User Mobile App Rating Scale

Edited by L Buis; submitted 08.09.23; peer-reviewed by Z Niu, R Froutan; comments to author 03.10.23; revised version received 15.11.23; accepted 11.12.23; published 26.01.24.

Please cite as:
Liu P, Wang L, Wang F
Evaluation of Chinese HIV Mobile Apps by Researchers and Patients With HIV: Quality Evaluation Study
JMIR Mhealth Uhealth 2024;12:e52573
URL: https://mhealth.jmir.org/2024/1/e52573
doi:10.2196/52573
PMID:38277215

Original Paper

mHealth Apps for the Self-Management of Low Back Pain: Systematic Search in App Stores and Content Analysis

Tianyu Zhou[1], MSc; David Salman[2], PhD; Alison McGregor[1], PhD

[1]Department of Surgery and Cancer, Imperial College London, London, United Kingdom

[2]Department of Primary Care and Public Health, Imperial College London, London, United Kingdom

Corresponding Author:
Tianyu Zhou, MSc
Department of Surgery and Cancer
Imperial College London
MSk lab, Sir Michael Uren Hub, 86 Wood Lan
London, W12 0BZ
United Kingdom
Phone: 44 7421760832
Email: t.zhou20@imperial.ac.uk

Abstract

Background: With the rapid development of mobile health (mHealth) technology, many health apps have been introduced to the commercial market for people with back pain conditions. However, little is known about their content, quality, approaches to care for low back pain (LBP), and associated risks of use.

Objective: The aims of this research were to (1) identify apps for the self-management of LBP currently on the market and (2) assess their quality, intervention content, theoretical approaches, and risk-related approaches.

Methods: The UK iTunes and Google Play stores were initially searched for apps related to the self-management of LBP in May 2022. A repeat search in June 2023 was conducted to ensure that any relevant new apps developed in the last year were incorporated into the review. A total of 3 keywords recommended by the Cochrane Back and Neck Group were used to search apps “low back pain,” “back pain,” and “lumbago.” The quality of the apps was assessed by using the 5-point Mobile App Rating Scale (MARS).

Results: A total of 69 apps (25 iOS and 44 Android) met the inclusion criteria. These LBP self-management apps mainly provide recommendations on muscle stretching (n=51, 73.9%), muscle strengthening (n=42, 60.9%), core stability exercises (n=32, 46.4%), yoga (n=19, 27.5%), and information about LBP mechanisms (n=17, 24.6%). Most interventions (n=14, 78%) are consistent with the recommendations in the National Institute for Health and Care Excellence (NICE) guidelines. The mean (SD) MARS overall score of included apps was 2.4 (0.44) out of a possible 5 points. The functionality dimension was associated with the highest score (3.0), whereas the engagement and information dimension resulted in the lowest score (2.1). Regarding theoretical and risk-related approaches, 18 (26.1%) of the 69 apps reported the rate of intervention progression, 11 (15.9%) reported safety checks, only 1 (1.4%) reported personalization of care, and none reported the theoretical care model or the age group targeted.

Conclusions: mHealth apps are potentially promising alternatives to help people manage their LBP; however, most of the LBP self-management apps were of poor quality and did not report the theoretical approaches to care and their associated risks. Although nearly all apps reviewed included a component of care listed in the NICE guidelines, the model of care delivery or embracement of care principles such as the application of a biopsychosocial model was unclear.

(JMIR Mhealth Uhealth 2024;12:e53262) doi:10.2196/53262

KEYWORDS

smartphone; mHealth; mobile health; low back pain; self-management; treatment interventions; mobile phone

Introduction

Low back pain (LBP) is a complex multifactorial disorder, often considered a combination of physical, psychological, and social dysfunction [1]. A multidisciplinary self-management intervention based on a biopsychosocial model holds significant potential to manage LBP [2,3] and has been demonstrated to be more effective than unimodal exercise therapy [4]. The

National Institute for Health and Care Excellence (NICE) guidelines recommend using self-management for LBP, which can be described as the patient's proactive adoption of strategies to manage their symptoms and monitor their health and well-being [5]. Despite being a promising approach to managing LBP, it can be challenging for an individual to self-manage any long-term condition [6]. The adherence to self-management strategies is commonly poor, especially without support, feedback, and positive reinforcement [7,8]. A qualitative study also noted poor adherence to advice and exercises as a limiting factor to recovery from LBP [9].

With the increasing popularity of electronic products, digital health solutions such as smartphone apps can be used as an innovative way to support self-management for many conditions, including LBP and may provide a solution to some of the problems outlined above [6,10]. Mobile health (mHealth) apps for pain management may be beneficial to patients [11,12], helping monitor those with acute or chronic pain and providing them with information and support for pain management. However, while many mHealth apps have been introduced into the commercial marketplace to manage pain, most have not been regulated in a uniform or standardized way before being released to the market [12,13]. The involvement of health care professionals in their development and content has been lacking [14]. This has raised concerns about the quality of these mHealth apps and whether their content information is evidence-based [15,16]. It is therefore desirable to assess the quality of current apps in the UK market and whether their content aligns with guideline recommendations. Another concerning issue is the paucity of evidence on specific intervention approaches in developed mHealth apps that have been developed, including underpinning evidence and theory [17] and relative risk management [18]. This potentially impacts on safety and efficacy of health-related smartphone apps [19].

Since there is no unified framework for assessing the theoretical and risk-based approaches associated with LBP self-management applications, we developed a theoretical framework that considers the theoretical care model of the intervention, the personalization of care, and the rate of intervention progression, as well as a risk-related framework that includes the targeted age group and the provision of appropriate safety checks. Interventions based on specific theoretical frameworks are known to be more effective in health care [20]. The Medical Research Council (MRC) guidelines strongly recommend using theoretical approaches in designing complex interventions [21], thus its inclusion in the evaluation. A critical factor in mHealth apps is personalization, where the management or treatment of the disease is tailored to the patient's situation and individual needs, which will make the user feel that it is relevant and meaningful to them [22,23], thus the inclusion of criteria in relation to personalization. Treatment methods such as exercise need to be both adaptable and progressive. This means starting with basic exercises and gradually advancing to more complex levels at a pace that patients with LBP can comfortably handle. Such a progression helps enhance their functional abilities and quality of life, while simultaneously equipping them with strategies to effectively manage their pain. [24]. Since the treatment paths for LBP vary according to different age groups, including children, adults, and the elderly [25], the target user age group must be defined. Finally, patient safety is an essential component of health care provision and is critical to primary care management, which will effectively reduce the clinical risks associated with LBP management [18,26].

Therefore, this app review, aimed to first, identify apps for the self-management of LBP currently on the market and second, to assess their quality (eg, functionality and design), intervention content (compliance with best practice guidelines), underlying theory (eg, theoretical care model), and risk-related approaches (eg, the age group targeted).

Methods

Search Strategy

Apps currently on the market for the self-management of LBP were identified, reviewed, and analyzed using a systematic approach. The UK official app stores for both Apple's iOS and Google's Android OS were used to search for mobile apps. These 2 operating systems currently dominate the marketplace of mobile medical apps [27]. We logged into Apple iTunes and Android Google Play stores in May 2022. A total of 3 keywords recommended by the Cochrane Back and Neck Group [28] were used to search apps "low back pain," "back pain," and "lumbago." A subsequent search in June 2023 was repeated to ensure that any relevant new apps in the last year were incorporated into the review.

Study Selection

Criteria for inclusion in the review were (1) apps were a self-contained product (ie, did not depend on an external device or add-ons), (2) apps offered at least 1 active treatment option for LBP (eg, unsupervised exercise program or patient education), (3) apps only designed for people with LBP, (4) apps created or updated in the last 5 years to ensure software functionality and ongoing technical support, and (5) apps developed in English. Exclusion criteria were (1) apps targeted at managing general chronic pain, (2) apps only focused on risk factors and diagnostic tests for LBP, (3) apps only focused on specific LBP pathologies (eg, lumbar disk herniation), (4) apps designed for clinicians, (5) general back fitness apps with no mention of physiotherapy or physical therapy or musculoskeletal (MSK) conditions, and (6) apps were not downloadable or had restrictions (eg, requiring an activate access password). Apps that incurred a cost were also included; however, when both paid and free versions of an app were available, we reviewed only the paid version to ensure access to the full content. If the same app was available on iOS and Android, the iOS version was kept for inclusion and analysis.

An independent reviewer initially screened the eligible apps based on the apps' names and descriptions and the screenshots provided. After the preliminary screening phase, the same independent reviewer downloaded apps that met the eligibility criteria for a second screening. Concerns regarding inclusion were discussed and resolved within the research group (AM and DS) until a final decision was reached.

Data Extraction

The selected apps were downloaded onto either a Samsung SM-N975F (Android version 7.1.2) or an iPhone 12 (iOS version 16.5) for a complete assessment of eligibility and characteristics. Relevant background information offered in the included apps was extracted, including the name, version, developer, update date, cost, and presence of in-app purchases. We extracted the age or content rating and consumer rating (5-star rating system) when available. In addition, we extracted if the apps contained advertisements and whether these adverts were relevant to their back pain. It was also noted if the apps were asynchronous or synchronous (available with support from a provider), and whether the apps tracked user engagement. The collection of personally identifiable information by apps and whether consent is stated were also checked. The category of management content, specific intervention component, theoretical care model, personalization of care, the rate of intervention progression, the age group targeted, and safety checks were also extracted.

App Content Assessment

Main categories and specific components of LBP app management content were identified and classified. Frequency analysis was performed to determine the number of apps providing these intervention contents. The recently updated 2020 NICE guideline for LBP was used to assess whether the included apps provided evidence-based interventions (categorized as "yes/no") [29]. For this, we mapped app interventions to recommendations listed in the NICE guidelines. This guideline provides the most recent best practice recommendations for assessing and managing LBP and sciatica in people aged 16 years or older [5].

App Quality Assessment

The Mobile App Rating Scale (MARS) was used to assess the quality of included apps in this review. MARS is a brief tool with a 23-item questionnaire to classify and assess the quality of mHealth apps for researchers, professionals, and clinicians [30]. It assesses app quality across 4 domains: engagement, functionality, aesthetics, and information quality. All items are assessed on a 5-point scale (1=inadequate, 2=poor, 3=acceptable, 4=good, and 5=excellent). The MARS has demonstrated excellent internal consistency and interrater reliability for evaluating the quality of mHealth apps [30]. To standardize the quality ratings, the assessor completed a MARS video training recommended by the developers of MARS [30]. A total of 10 apps were randomly selected for training until a consensus on the scores was reached.

App Assessment of the Theoretical and Risk-Related Approaches

A total of 5 aspects considered in our theoretical and risk-related approaches were assessed. These included the underpinning LBP management theory, tailoring of content, the intervention of progression approach, the age group the app targeted, and appropriate safety checks.

Ethical Considerations

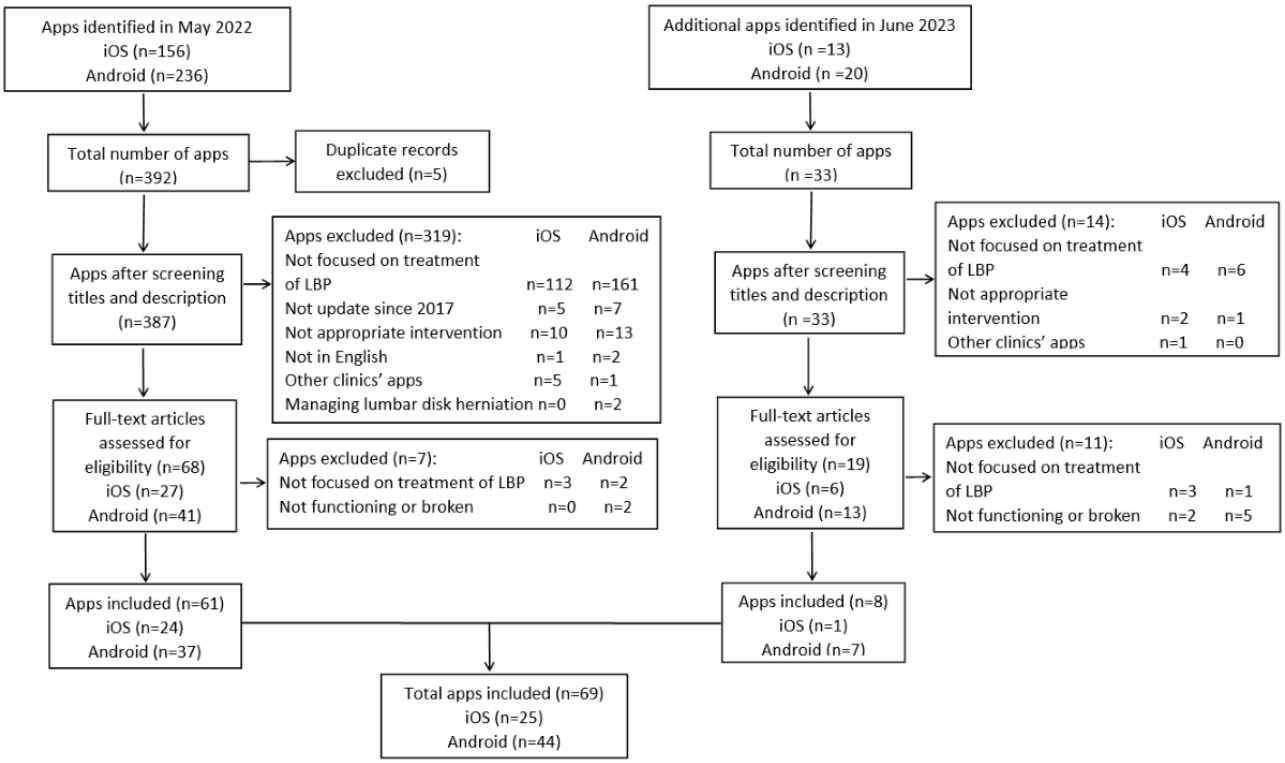
This study does not involve human participants.

Results

App Selection

The searches performed in May 2022 yielded 392 apps from 2 platforms. Of these, 156 apps were identified from iTunes stores, and 236 apps were identified from Google Play stores. A total of 5 duplicates were removed, resulting in 387 apps identified for screening based on the titles and app descriptions. After initial screening, 319 apps were excluded. The eligible 68 apps were downloaded for a full evaluation and further 7 apps were excluded. The subsequent search in June 2023 found a total of 33 newly developed apps based on the initial search, of which 8 apps were newly developed. Finally, 69 apps were included in this review, of which 25 were iOS apps and 44 were Android apps. Figure 1 illustrates the selection procedure of smartphone apps for LBP.

Figure 1. The flowchart of the app selection process. LBP: low back pain.



Characteristics of Included Apps

Of the 69 apps included in this review, 25 (36%) were found on iTunes exclusively, 44 (64%) on Google Play exclusively, and 5 (7%) were found on both app stores. There are 5 (7%) of 69 apps that required payment, ranging in price from US $1.13 to US $22.87 (median US $8.17). The majority (n=64, 93%) of the apps included were free of charge. Of these, 8 offered in-app purchases ranging from US $3.80 to US $12.70. Android apps (n=41, 93%) were more often free to access full functionality than iOS apps (n=15, 60%). Of the 41 apps reviewed by customers on a 5-star rating system, the median customer rating in 19 apps from iTunes (4.3 stars) was higher than in 22 apps from Google Play (4.1 stars). With respect to age or content rating, most of the included iOS apps (n=19, 76%) were downloaded without any age limitation: 4 apps were restricted to those of 12 years or above and 3 apps were restricted to those of at least 17 years old. However, all Android apps were labeled as suitable for all age groups.

In terms of developers, there is a mix of some health care groups and other private companies. A total of 24 (35%) apps contained advertisements, 10 (42%) of which were for products or medical companies targeting MSK disorders, leaving 14 (58%) random advertisements. In addition, all apps were asynchronous which means that they failed to deliver continuously updated application data to users. A total of 9 (13%) apps collect personally identifiable information from users and only 2 (22%) asked for their consent for collection. It appeared that none of the apps tracked user engagement. The characteristics of each app are presented in Multimedia Appendix 1.

MARS Quality Assessment

The mean MARS total score obtained from 69 applications was 2.4 out of 5 (SD 0.44). Table 1 summarizes the MARS total scores for each app. All apps were initially assessed using MARS for engagement, functionality, aesthetics, and information. Mean scores for each subscale (out of 5) were calculated. Of the 5 categories, apps scored highest in the functionality (mean 3.0, SD 0.55) domain, followed by aesthetics (mean 2.6, SD 0.61) and engagement (mean 2.1, SD 0.58). The information domain received the lowest score (mean 2.1, SD 0.46). The MARS total score and domain score for each app are shown in Multimedia Appendix 2.

Table 1. Mobile app rating scale scores (N=69 apps).

MARS[a] subscale	iOS, mean (SD)	Android, mean (SD)	Total, mean (SD)
Engagement	2.4 (0.52)	1.9 (0.55)	2.1 (0.58)
Functionality	3.1 (0.57)	3.0 (0.49)	3.0 (0.55)
Aesthetics	2.9 (0.57)	2.4 (0.47)	2.6 (0.61)
Information	2.3 (0.49)	1.9 (0.40)	2.1 (0.46)
MARS overall score[b]	2.6 (0.43)	2.3 (0.35)	2.4 (0.44)

[a]MARS: Mobile App Rating Scale.

[b]Average of 4 objective subscales.

Intervention Contents for LBP

The LBP interventions embedded with the included apps were mainly classified into 3 categories (Table 2). Of these, most (n=47, 68.1%) of the apps offered only an exercise program, while 14.5% (n=10) apps provided patient education alone, and 13.0% (n=9) apps recommended patient education in combination with an exercise program. The remaining 2.9% (n=2) apps provided some psychological intervention for LBP in combination with an exercise program. Finally, only 1.4% (n=1) app-prescribed combinations of patient education, exercise, and psychological intervention.

Table 2. Number and percentage of category for low back pain interventions used in included apps.

Main category	Value, n (%)
Patient education + exercise program + psychological intervention	1 (1.4)
Exercise program + patient education	9 (13.0)
Exercise program + psychological intervention	2 (2.9)
Patient education only	10 (14.5)
Exercise program only	47 (68.1)

More specifically, of the 69 apps included in this review, most (n=51, 73.9%) apps recommended muscle stretching as a self-management strategy for LBP. A total of 42 (60.9%) apps suggested muscle strengthening, 32 (46.4%) apps offered core stability exercises, and 19 apps (27.5%) recommended using yoga to manage LBP. Also, there were 17 (24.6%) apps providing information about LBP mechanisms, followed by advice to use medication (n=9, 13%), staying active (n=9, 13%), postural therapy (n=8, 11.6%), cold and heat therapy (n=8, 11.6%), and aerobic exercise (n=8, 11.6%). Only some apps mentioned manual therapy (n=4, 5.8%), cognitive behavioral therapy (n=4, 5.8%), meditation (n=4, 5.8%), mindfulness (n=3, 4.3%), McKenzie exercise (n=3, 4.3%), electrotherapy (n=3, 4.3%), acupuncture (n=3, 4.3%), and lifestyle advice (n=2, 2.9%). Concerning app intervention content, 14 (78%) of 18 interventions for LBP were aligned with the NICE guidelines and 4 included interventions that were not endorsed by the NICE guideline: postural therapy, electrotherapy, cold and heat therapy, and acupuncture [29]. Details of the interventions offered for LBP managed in the included apps are summarized in Table 3.

Table 3. Number and percentage of specific component for low back pain interventions used in included apps.

Specific components	Value, n (%)
Understanding LBP[a] mechanisms	17 (24.6)
Staying active	9 (13.0)
Postural therapy	8 (11.6)
Lifestyle advice	2 (2.9)
Electrotherapy	3 (4.3)
Cold and heat therapy	8 (11.6)
Manual therapy	4 (5.8)
Medication use	9 (13.0)
Core stability exercise	32 (46.4)
Muscle strengthening	42 (60.9)
Muscle stretching	51 (73.9)
McKenzie exercise	3 (4.3)
Aerobic exercise	8 (11.6)
Yoga	19 (27.5)
Mindfulness	3 (4.3)
Meditation	4 (5.8)
CBT[b]	4 (5.8)
Acupuncture	3 (4.3)

[a]LBP: low back pain.

[b]CBT: cognitive behavioral therapy.

Theoretical and Risk-Related Approaches

None of the 69 apps included in this review referred to or explained their theoretical care model. Only 1 app considered or incorporated a tailored approach to their intervention. A total of 18 (26%) of these apps provided an intervention program based on principles of gradual intervention progression. No app mentioned the age group for which their intervention content was appropriate and none set an age limit for their use. Even fewer, 11 (16%) apps offered safety checks for app users, including identifying red flags, signs, and symptoms of LBP that required medical attention or providing safety-netting advice if the back pain did not resolve or worsen (Multimedia Appendix 3).

Discussion

Intervention Content in LBP Apps

We attempted to benchmark the content of the included apps against the most recently published best practice guideline from NICE and found that most LBP self-management app components were consistent with those recommended by NICE. The NICE LBP guidelines [29] recommended the use of a group exercise program, including biomechanical, aerobic, mind-body, or a combination of approaches. Accordingly, muscle stretching, strengthening, core stability exercises, and yoga are the most common interventions in self-management apps. The findings of this review support a 2016 systematic review, which found that interventions in LBP selected by app developers were primarily based on clinical practice guidelines [31]. Similarly, the 2021 Cochrane review also reported moderate-certainty evidence suggesting that different types of exercise therapy are effective in treating LBP [32]. This means that most current LBP self-management apps offer evidence-based interventions.

However, while the content of most apps appears evidence-based, they are not often delivered in the context of a complex intervention and as such do not reflect the current LBP care models. We found that most self-management apps rely on exercise interventions, and very few apps incorporate social and psychological interventions for managing LBP. This suggests that current self-management apps emphasize a more biological care model to manage LBP rather than considering the influence of psychological and social factors in the development and maintenance of pain [33]. Research has emphasized interrelationships among biological changes, psychological status, and the sociocultural context and as such, they all need to be considered to understand the impact of chronic pain and its subsequent management [34]. The biopsychosocial model has been widely accepted as a holistic approach to increase efficacy and outcomes in managing chronic LBP [2,35]. Also, a complex multidisciplinary approach with a biopsychosocial model has been recommended in the early stages of LBP to reduce the likelihood of chronicity following acute LBP [36]. Thus, digital self-management interventions for LBP should consider adopting this model.

Quality Assessment in Self-Management Apps

Generally, apps for the self-management of LBP are of poor quality as assessed by MARS. Functionality (mean 3.0, SD 0.55) was the domain that scored the highest on the MARS test, as described by other authors [31,37]. This implies that the apps are functioning well, easy to learn and navigate, and efficient. However, the quality assessment revealed that these apps had low scores on the domains of "engagement" (mean 2.1, SD 0.58) and "aesthetics" (mean 2.6, SD 0.61). This indicates that the features that make the app equally engaging and important to a wide user base may have been overlooked. This was partly because most apps did not consider using specific strategies to increase involvement and aesthetics from the user's point of view (eg, entertainment, interactivity, customization, layout, and graphics). Involving patient users during the development of these apps might better identify their needs and characteristics, and increase adherence to improve self-management and health outcomes [38].

The lowest score on information (mean 2.1, SD 0.46) indicated weakness in the quality and trustworthiness of information presented in the included apps. This was evidenced through assessing credibility (MARS item 18) and evidence base (MARS item 19). In terms of MARS item 18, most apps were developed by either commercial businesses with a vested interest or a legitimate source without verification (eg, has no web page), yet few are developed by credible health organizations (eg, government or universities). The lack of health professional involvement is a consistent issue highlighted by Rizwana, who has expressed concern about the accuracy and trustworthiness of in-app information [39]. In addition, none of the apps available for LBP management were evaluated using a randomized controlled trial. This was evident in MARS item 19, which assesses whether the app has been trialed or tested in the published scientific literature, and therefore the effectiveness and safety of these apps remain unknown. These results align with previous reviews on mHealth apps directed at pain management, in which a lack of scientific basis of the outcomes was found to support the use of such apps [40,41]. A possible explanation could be that most of the apps are of commercial rather than scientific origin, which suggests that the need to promote cross-disciplinary collaboration between academic and commercial institutions might help develop the evidence base for using such apps [37,40].

Theoretical and Risk-Related Approaches

Current self-management apps do not appear to have used a theoretical rationale in their development. This aligns with the findings from a 2018 review, which found the development of current self-management mHealth apps for patient education programs lacks the support of underpinning theory or framework [42]. An underpinning theoretical model is widely recognized as a crucial component of health interventions and is important when trying to understand key components of the intervention, how they interact, and the mechanisms of the intervention [43-45]. Systematic reviews of existing evidence demonstrate that interventions underpinned by theory are more effective than those that are not [46,47]. Additionally, as a complex multifactorial condition, the management of LBP should consider theory development and identifying the evidence base in accordance with the MRC framework for complex interventions [48].

Regarding personalized care, almost all included apps provided limited customized service. Tailored communications provide individuals with information that is relevant to them and that fits their situation. This can lead to increased perceived personal relevance, user engagement, more in-depth processing of information, and consequently, more desire and motivation to engage in the health behavior change [49]. The importance of personalization of mHealth apps was emphasized in a qualitative review in 2019, which expressed that mHealth apps should meet patients' needs since they were created for use [22]. Also, our review suggests that the principle of gradual progression from simple to more advanced levels of intervention is not universal. Providing tools that help the user implement exercise progression ensures that the intervention progressively becomes more challenging to continually stimulate adaptations and maintain interest in the program [24]. The Coventry, Aberdeen, and London—Refined (CALO-RE) taxonomy of behavior change techniques highlighted the importance of setting graded tasks, breaking targeted behaviors into smaller, more manageable tasks, and facilitating progress in small increments [50].

Current apps do not evaluate the age limitation of intervention content for which LBP advice and treatments are appropriate, which may pose some risks for the users. Although age ratings in the App Store are often reported, it is based primarily on the degree to which an app contains sensitive information rather than on the applicability of an intervention to different age groups. It is also worth noting that the current NICE guideline for LBP published in 2016 is based on an adult population. A review from the Lancet showed that the evidence underpinning LBP guidelines is drawn almost exclusively from clinical trials of adults, and there are few trials on the treatment of back pain in children [51]. Thus, adult LBP care pathways may not be suitable for adolescents and children [52,53]. When applying these LBP management apps to nonadults, their scope of application needs to be considered. In addition, most included apps performed poorly in providing safety considerations. Safety check advice, including information on the natural history of the illness, advice on worrying symptoms to watch out for, and specific information on how and when to seek help, as well as advice about the follow-up of investigations and hospital referrals, can effectively address uncertainty in the process of LBP management [54]. A scoping review in 2020 discussed that safety concerns within apps were a primary concern [18]. Thus, such approaches will improve the likelihood of providing users with appropriate care and reduce clinical risks associated with self-management [55].

Evaluation Tool for LBP Apps

Although some self-management apps show relatively high scores on the MARS score, such as "Back Doctor/Pain Relief-1.03.24," "Perfect Posture&Healthy back-1.5.2," and "Back pain exercise at home-1.0.99," they perform poorly on the theoretical and risk-related framework, such as the personalization of care and the age group targeted. Conversely,

some self-management apps (eg, BackTrainer-2.0) adopted a tailored approach and addressed risk-related issues, but MARS' quality assessment scores were not high. Additionally, MARS has not effectively evaluated or included the biopsychosocial care model commonly used in chronic pain [34]. This indicates that current evaluation tools may be limited in their ability to assess LBP self-management apps comprehensively, and consequently, further study is needed to explore whether a holistic tool to evaluate LBP self-management apps is required.

Barriers to Holistic Digital Apps

Despite the overwhelming evidence recommending the use of holistic multidisciplinary interventions based on a biopsychosocial model for the self-management of LBP, the proliferation of back pain apps on the market that use simple interventions is concerning. It is hard to speculate why this is the case given the strong evidence for the use of a biopsychosocial model in the management of LBP. This may relate to current regulatory approvals. For an app to deliver a complex biopsychosocial intervention it would be classified as a medical device requiring such approvals. Further work is needed to explain why there are so many apps on the market that fail to address the model of care we seek in medical practice, and how current regulatory processes affect this.

Moreover, digital health interventions have been strongly advocated for and implemented in other domains and countries. Notably, draft guidance from NICE has informed digital programs using multidisciplinary models to assist the NHS in delivering specialized services for weight management [56]. Furthermore, the Federal Institute for Drugs and Medical Devices in Germany has authorized the entry of multidisciplinary health care apps with robust trial data into the market, with the overarching objective of empowering clinicians to prescribe health care apps for their patients [57,58]. This illustrates the feasibility of developing multidisciplinary mHealth apps for health care practitioners to recommend. However, it remains imperative to conduct a comprehensive evaluation of the apps using clinical studies to ascertain their efficacy and suitability for widespread prescription.

Limitations

Apps not specifically targeted for the self-management of LBP (eg, chronic pain, pain management, or MSK apps) were excluded from this review, which may result in potentially eligible apps being missed. While many LBP apps in different countries are available in app stores, our search was limited to the UK iTunes and Google Play stores for practical reasons relating to data-capturing capacity. Therefore, this review did not include apps exclusively available in other countries, and therefore may not represent the broader landscape relating to digital approaches for management of LBP. Only 1 independent reviewer assessed app quality using MARS and discussed uncertainties with 2 other research members until a consensus was reached, which may impact the reliability of the assessments.

Conclusions

In this review, we identified 69 apps related to the self-management of LBP and rated them using MARS. Most apps scored poor quality due to their approaches to engagement and information, and many emerged as tools for delivering passive information rather than active management apps. Most apps were aligned with guideline-based care. However, no app offered a holistic self-management intervention approach incorporating a biopsychological model. Most apps underestimated the importance of theoretical and risk-related aspects. Thus, current self-management apps for LBP are limited in what they offer.

Acknowledgments

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. Imperial College London is the sponsor for the study and has no influence on the direction or content of the work. We are grateful to Rebecca Jones (Library Manager in Imperial College London) who offered advice and support for the building search strategy and screening process. We also wish to thank our colleagues in the MSK laboratory for their valuable and helpful comments and suggestions. Without these people, studying would have been impossible.

Data Availability

The data sets generated and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

TZ, DS, and AM conceived and designed the study. TZ takes responsibility for the acquisition of the data and the analysis of the data. TZ drafted the paper. DS and AM revised the paper critically for important intellectual content. All authors approved the final version to be published.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Characteristics of apps for the self-management of low back pain.

[DOC File , 107 KB - mhealth_v12i1e53262_app1.doc]

Multimedia Appendix 2
The Mobile App Rating Scale overall and subscale scores.
[DOC File , 117 KB - mhealth_v12i1e53262_app2.doc]

Multimedia Appendix 3
The content and approach for low back pain interventions used in the included apps.
[DOC File , 133 KB - mhealth_v12i1e53262_app3.doc]

References

1. Costa LDCM, Maher CG, McAuley JH, Hancock MJ, Herbert RD, Refshauge KM, et al. Prognosis for patients with chronic low back pain: inception cohort study. BMJ 2009;339:b3829 [FREE Full text] [doi: 10.1136/bmj.b3829] [Medline: 19808766]
2. Kamper SJ, Apeldoorn AT, Chiarotto A, Smeets RJEM, Ostelo RWJG, Guzman J, et al. Multidisciplinary biopsychosocial rehabilitation for chronic low back pain: cochrane systematic review and meta-analysis. BMJ 2015;350:h444 [FREE Full text] [doi: 10.1136/bmj.h444] [Medline: 25694111]
3. Igwesi-Chidobe CN, Kitchen S, Sorinola IO, Godfrey EL. Evidence, theory and context: using intervention mapping in the development of a community-based self-management program for chronic low back pain in a rural African primary care setting - the good back program. BMC Public Health 2020 Mar 17;20(1):343 [FREE Full text] [doi: 10.1186/s12889-020-8392-7] [Medline: 32183758]
4. Igwesi-Chidobe CN, Kitchen S, Godfrey E, Sorinola I. Feasibility of a physiotherapist-led community-based self-management programme for reducing non-specific chronic low back pain disability in rural Nigeria. Physiotherapy 2016 Nov;102:e61. [doi: 10.1016/j.physio.2016.10.384]
5. Bernstein IA, Malik Q, Carville S, Ward S. Low back pain and sciatica: summary of NICE guidance. BMJ 2017;356:i6748. [doi: 10.1136/bmj.i6748] [Medline: 28062522]
6. Nordstoga AL, Bach K, Sani S, Wiratunga N, Mork PJ, Villumsen M, et al. Usability and acceptability of an app (SELFBACK) to support self-management of low back pain: mixed methods study. JMIR Rehabil Assist Technol 2020;7(2):e18729 [FREE Full text] [doi: 10.2196/18729] [Medline: 32902393]
7. Mork PJ, Bach K, selfBACK Consortium. A decision support system to enhance self-management of low back pain: protocol for the selfBACK project. JMIR Res Protoc 2018;7(7):e167 [FREE Full text] [doi: 10.2196/resprot.9379] [Medline: 30030208]
8. Schaller A, Froboese I. Movement coaching: study protocol of a randomized controlled trial evaluating effects on physical activity and participation in low back pain patients. BMC Musculoskelet Disord 2014;15:391 [FREE Full text] [doi: 10.1186/1471-2474-15-391] [Medline: 25416514]
9. Liddle SD, Baxter GD, Gracey JH. Chronic low back pain: patients' experiences, opinions and expectations for clinical management. Disabil Rehabil 2007;29(24):1899-1909. [doi: 10.1080/09638280701189895] [Medline: 17852259]
10. mHealth New Horizons for Health Through Mobile Technologies. Geneva: World Health Organization; 2011.
11. Thurnheer SE, Gravestock I, Pichierri G, Steurer J, Burgstaller JM. Benefits of mobile apps in pain management: systematic review. JMIR Mhealth Uhealth 2018;6(10):e11231 [FREE Full text] [doi: 10.2196/11231] [Medline: 30348633]
12. Rosser BA, Eccleston C. Smartphone applications for pain management. J Telemed Telecare 2011;17(6):308-312 [FREE Full text] [doi: 10.1258/jtt.2011.101102] [Medline: 21844177]
13. Day S, Shah V, Kaganoff S, Powelson S, Mathews SC. Assessing the clinical robustness of digital health startups: cross-sectional observational analysis. J Med Internet Res 2022;24(6):e37677 [FREE Full text] [doi: 10.2196/37677] [Medline: 35723914]
14. Lalloo C, Jibb LA, Rivera J, Agarwal A, Stinson JN. "There's a pain app for that": review of patient-targeted smartphone applications for pain management. Clin J Pain 2015;31(6):557-563. [doi: 10.1097/AJP.0000000000000171] [Medline: 25370138]
15. Nouri R, Kalhori SRN, Ghazisaeedi M, Marchand G, Yasini M. Criteria for assessing the quality of mHealth apps: a systematic review. J Am Med Inform Assoc 2018;25(8):1089-1098 [FREE Full text] [doi: 10.1093/jamia/ocy050] [Medline: 29788283]
16. BinDhim NF, Hawkey A, Trevena L. A systematic review of quality assessment methods for smartphone health apps. Telemed J E Health 2015;21(2):97-104. [doi: 10.1089/tmj.2014.0088] [Medline: 25469795]
17. Chib A, Lin SH. Theoretical advancements in mHealth: a systematic review of mobile apps. J Health Commun 2018;23(10-11):909-955. [doi: 10.1080/10810730.2018.1544676] [Medline: 30449261]
18. Akbar S, Coiera E, Magrabi F. Safety concerns with consumer-facing mobile health applications and their consequences: a scoping review. J Am Med Inform Assoc 2020;27(2):330-340 [FREE Full text] [doi: 10.1093/jamia/ocz175] [Medline: 31599936]

19. Kao CK, Liebovitz DM. Consumer mobile health apps: current state, barriers, and future directions. PM R 2017;9(5S):S106-S115. [doi: 10.1016/j.pmrj.2017.02.018] [Medline: 28527495]
20. Dyson J, Cowdell F. How is the theoretical domains framework applied in designing interventions to support healthcare practitioner behaviour change? A systematic review. Int J Qual Health Care 2021;33(3):mzab106 [FREE Full text] [doi: 10.1093/intqhc/mzab106] [Medline: 34279637]
21. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, Medical Research Council Guidance. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ 2008;337:a1655 [FREE Full text] [doi: 10.1136/bmj.a1655] [Medline: 18824488]
22. Vo V, Auroy L, Sarradon-Eck A. Patients' perceptions of mHealth apps: meta-ethnographic review of qualitative studies. JMIR Mhealth Uhealth 2019;7(7):e13817 [FREE Full text] [doi: 10.2196/13817] [Medline: 31293246]
23. Floch J, Zettl A, Fricke L, Weisser T, Grut L, Vilarinho T, et al. User needs in the development of a health app ecosystem for self-management of cystic fibrosis: user-centered development approach. JMIR Mhealth Uhealth 2018;6(5):e113 [FREE Full text] [doi: 10.2196/mhealth.8236] [Medline: 29739742]
24. Brody LT. Effective therapeutic exercise prescription: the right exercise at the right dose. J Hand Ther 2012;25(2):220-231; quiz 232. [doi: 10.1016/j.jht.2011.09.009] [Medline: 22212491]
25. Casiano VE, Sarwan G, Dydyk AM, Varacallo M. Back Pain. Treasure Island (FL): StatPearls Publishing; 2019.
26. Wetzels R, Wolters R, van Weel C, Wensing M. Mix of methods is needed to identify adverse events in general practice: a prospective observational study. BMC Fam Pract 2008;9:35 [FREE Full text] [doi: 10.1186/1471-2296-9-35] [Medline: 18554418]
27. Hanrahan C, Dy Aungst T, Cole S. Evaluating Mobile Medical Applications. Bethesda, MD: American Society of Health-System Pharmacists; 2014.
28. Furlan AD, Malmivaara A, Chou R, Maher CG, Deyo RA, Schoene M, et al. 2015 updated method guideline for systematic reviews in the cochrane back and neck group. Spine (Phila Pa 1976) 2015;40(21):1660-1673. [doi: 10.1097/BRS.0000000000001061] [Medline: 26208232]
29. de Campos TF. Low back pain and sciatica in over 16s: assessment and management NICE guideline [NG59]. J Physiother 2017;63(2):120 [FREE Full text] [doi: 10.1016/j.jphys.2017.02.012] [Medline: 28325480]
30. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
31. Machado GC, Pinheiro MB, Lee H, Ahmed OH, Hendrick P, Williams C, et al. Smartphone apps for the self-management of low back pain: a systematic review. Best Pract Res Clin Rheumatol 2016;30(6):1098-1109. [doi: 10.1016/j.berh.2017.04.002] [Medline: 29103552]
32. Hayden JA, Ellis J, Ogilvie R, Malmivaara A, van Tulder MW. Exercise therapy for chronic low back pain. Cochrane Database Syst Rev 2021;9(9):CD009790 [FREE Full text] [doi: 10.1002/14651858.CD009790.pub2] [Medline: 34580864]
33. Foster NE. Barriers and progress in the treatment of low back pain. BMC Med 2011;9:108 [FREE Full text] [doi: 10.1186/1741-7015-9-108] [Medline: 21943396]
34. Gatchel RJ, Peng YB, Peters ML, Fuchs PN, Turk DC. The biopsychosocial approach to chronic pain: scientific advances and future directions. Psychol Bull 2007;133(4):581-624. [doi: 10.1037/0033-2909.133.4.581] [Medline: 17592957]
35. van Erp RMA, Huijnen IPJ, Jakobs MLG, Kleijnen J, Smeets RJEM. Effectiveness of primary care interventions using a biopsychosocial approach in chronic low back pain: a systematic review. Pain Pract 2019;19(2):224-241 [FREE Full text] [doi: 10.1111/papr.12735] [Medline: 30290052]
36. Chou R, McCarberg B. Managing acute back pain patients to avoid the transition to chronic pain. Pain Manag 2011;1(1):69-79. [doi: 10.2217/pmt.10.7] [Medline: 24654586]
37. Escriche-Escuder A, De-Torres I, Roldán-Jiménez C, Martín-Martín J, Muro-Culebras A, González-Sánchez M, et al. Assessment of the quality of mobile applications (apps) for management of low back pain using the Mobile App Rating Scale (MARS). Int J Environ Res Public Health 2020 Dec 09;17(24):9209 [FREE Full text] [doi: 10.3390/ijerph17249209] [Medline: 33317134]
38. Molina-Recio G, Molina-Luque R, Jiménez-García AM, Ventura-Puertos PE, Hernández-Reyes A, Romero-Saldaña M. Proposal for the user-centered design approach for health apps based on successful experiences: integrative review. JMIR Mhealth Uhealth 2020;8(4):e14376 [FREE Full text] [doi: 10.2196/14376] [Medline: 32319965]
39. Biviji R, Vest JR, Dixon BE, Cullen T, Harle CA. Factors related to user ratings and user downloads of mobile apps for maternal and infant health: cross-sectional study. JMIR Mhealth Uhealth 2020;8(1):e15663 [FREE Full text] [doi: 10.2196/15663] [Medline: 32012107]
40. Salazar A, de Sola H, Failde I, Moral-Munoz JA. Measuring the quality of mobile apps for the management of pain: systematic search and evaluation using the mobile app rating scale. JMIR Mhealth Uhealth 2018;6(10):e10718 [FREE Full text] [doi: 10.2196/10718] [Medline: 30361196]
41. Lalloo C, Shah U, Birnie KA, Davies-Chalmers C, Rivera J, Stinson J, et al. Commercially available smartphone apps to support postoperative pain self-management: scoping review. JMIR Mhealth Uhealth 2017;5(10):e162 [FREE Full text] [doi: 10.2196/mhealth.8230] [Medline: 29061558]

42. Bashi N, Fatehi F, Fallah M, Walters D, Karunanithi M. Self-management education through mHealth: review of strategies and structures. JMIR Mhealth Uhealth 2018;6(10):e10771 [FREE Full text] [doi: 10.2196/10771] [Medline: 30341042]
43. Lippke S, Ziegelmann JP. Theory-based health behavior change: developing, testing, and applying theories for evidence-based interventions. Appl Psychol 2008;57(4):698-716. [doi: 10.1111/j.1464-0597.2008.00339.x]
44. Skivington K, Matthews L, Simpson SA, Craig P, Baird J, Blazeby JM, et al. A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. BMJ 2021;374:n2061 [FREE Full text] [doi: 10.1136/bmj.n2061] [Medline: 34593508]
45. Funnell SC, Rogers PJ. Purposeful Program Theory: Effective Use of Theories of Change and Logic Models. Hoboken, NJ: John Wiley & Sons; 2011.
46. Taylor N, Conner M, Lawton R. The impact of theory on the effectiveness of worksite physical activity interventions: a meta-analysis and meta-regression. Health Psychol Rev 2012;6(1):33-73. [doi: 10.1080/17437199.2010.533441]
47. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
48. Hush JM. Low back pain: it is time to embrace complexity. Pain 2020;161(10):2248-2251. [doi: 10.1097/j.pain.0000000000001933] [Medline: 32453129]
49. Bol N, Smit ES, Lustria MLA. Tailored health communication: opportunities and challenges in the digital era. Digit Health 2020;6:2055207620958913 [FREE Full text] [doi: 10.1177/2055207620958913] [Medline: 33029355]
50. Hagger MS, Keatley DA, Chan D. CALO-RE taxonomy of behavior change techniques. In: Tenenbaum G, Eklund RC, editors. Encyclopedia of Sport and Exercise Psychology. Thousand Oaks, CA: Sage Publications, Inc; 2014:100-104.
51. Foster NE, Anema JR, Cherkin D, Chou R, Cohen SP, Gross DP, et al. Prevention and treatment of low back pain: evidence, challenges, and promising directions. Lancet 2018;391(10137):2368-2383. [doi: 10.1016/S0140-6736(18)30489-6] [Medline: 29573872]
52. O'Sullivan P, Smith A, Beales D, Straker L. Understanding adolescent low back pain from a multidimensional perspective: implications for management. J Orthop Sports Phys Ther 2017;47(10):741-751 [FREE Full text] [doi: 10.2519/jospt.2017.7376] [Medline: 28898135]
53. O'Sullivan K, O'Keeffe M, Forster BB, Qamar SR, van der Westhuizen A, O'Sullivan PB. Managing low back pain in active adolescents. Best Pract Res Clin Rheumatol 2019;33(1):102-121. [doi: 10.1016/j.berh.2019.02.005] [Medline: 31431266]
54. Jones D, Dunn L, Watt I, Macleod U. Safety netting for primary care: evidence from a literature review. Br J Gen Pract 2019;69(678):e70-e79 [FREE Full text] [doi: 10.3399/bjgp18X700193] [Medline: 30510099]
55. Cook CE, George SZ, Reiman MP. Red flag screening for low back pain: nothing to see here, move along: a narrative review. Br J Sports Med 2018;52(8):493-496 [FREE Full text] [doi: 10.1136/bjsports-2017-098352] [Medline: 28923844]
56. Digital services to enable easier access to weight management support. National Institute for Health and Care Excellence. 2023. URL: https://www.nice.org.uk/news/article/digital-services-to-enable-easier-access-to-weight-management-support [accessed 2023-11-27]
57. Toelle TR, Utpadel-Fischler DA, Haas KK, Priebe JA. App-based multidisciplinary back pain treatment versus combined physiotherapy plus online education: a randomized controlled trial. NPJ Digit Med 2019;2:34 [FREE Full text] [doi: 10.1038/s41746-019-0109-x] [Medline: 31304380]
58. German doctors to prescribe health apps in 'world first'. HealthManagement.org. 2020. URL: https://healthmanagement.org/c/it/news/german-doctors-to-prescribe-health-apps-in-world-first [accessed 2023-11-28]

Abbreviations

CALO-RE: Coventry, Aberdeen, and London—Refined
LBP: low back pain
MARS: Mobile App Rating Scale
mHealth: mobile health
MRC: Medical Research Council
MSK: musculoskeletal
NICE: National Institute for Health and Care Excellence

Edited by L Buis; submitted 04.10.23; peer-reviewed by V Sparkes, R Jamison, C Eccleston; comments to author 02.11.23; revised version received 06.12.23; accepted 21.12.23; published 01.02.24.

Please cite as:
Zhou T, Salman D, McGregor A
mHealth Apps for the Self-Management of Low Back Pain: Systematic Search in App Stores and Content Analysis
JMIR Mhealth Uhealth 2024;12:e53262
URL: https://mhealth.jmir.org/2024/1/e53262
doi:10.2196/53262
PMID:38300700

Original Paper

Evaluation of Patient-Facing Mobile Apps to Support Physiotherapy Care: Systematic Review

Mark Merolli[1,2], BPhysio (Hons), PhD; Jill J Francis[3], BA, PhD; Patrick Vallance[3,4], BPTh; Kim L Bennell[1], BPTh, PhD; Peter Malliaras[4], BPTh (Hons), PhD; Rana S Hinman[1], BPhysio (Hons), PhD

[1]Centre for Health, Exercise & Sports Medicine, Department of Physiotherapy, School of Health Sciences, The University of Melbourne, Melbourne, Australia

[2]Centre for Digital Transformation of Health, The University of Melbourne, Melbourne, Australia

[3]School of Health Sciences, The University of Melbourne, Melbourne, Australia

[4]Department of Physiotherapy, Podiatry, Prosthetics and Orthotics, School of Allied Health, Human Service and Sport, La Trobe University, Melbourne, Australia

Corresponding Author:
Mark Merolli, BPhysio (Hons), PhD
Centre for Health, Exercise & Sports Medicine
Department of Physiotherapy, School of Health Sciences
The University of Melbourne
161 Barry Street, Parkville
Melbourne, 3052
Australia
Phone: 61 383443689
Email: merollim@unimelb.edu.au

Abstract

Background: Mobile health interventions delivered through mobile apps are increasingly used in physiotherapy care. This may be because of the potential of apps to facilitate changes in behavior, which is central to the aims of care delivered by physiotherapists. A benefit of using apps is their ability to incorporate behavior change techniques (BCTs) that can optimize the effectiveness of physiotherapeutic interventions. Research continues to suggest that despite their importance, behavior change strategies are often missing in patient management. Evaluating mobile apps that physiotherapists can use to drive behavior change may inform clinical practice and potentially improve patient outcomes. Examining the quality of apps and exploring their key features that can support behavior change and physiotherapy care are important aspects of such an evaluation.

Objective: The primary aim of this study was to describe the range of mobile apps in app stores that are intended for use by patients to support physiotherapy care. The secondary aims were to assess app quality, BCTs, and their behavior change potential.

Methods: A systematic review of mobile apps in app stores was undertaken. The Apple App Store and Google Play were searched using a 2-step search strategy, using terms relevant to the physiotherapy discipline. Strict inclusion and exclusion criteria were applied: apps had to be intended for use by patients and be self-contained (or stand-alone) without the requirement to be used in conjunction with a partner wearable device or another plugin. Included apps were coded for BCTs using the Behavior Change Technique Taxonomy version 1. App quality was assessed using the Mobile App Rating Scale, and the App Behavior Change Scale was used to assess the app's potential to change behavior.

Results: In total, 1240 apps were screened, and 35 were included. Of these 35 apps, 22 (63%) were available on both the Apple App Store and Google Play platforms. In total, 24 (69%) were general in their focus (eg, not condition-specific), with the remaining 11 (31%) being more specific (eg, knee rehabilitation and pelvic floor training). The mean app quality score (Mobile App Rating Scale) was 3.7 (SD 0.4) of 5 (range 2.8-4.5). The mean number of BCTs identified per app was 8.5 (SD 3.6). BCTs most frequently included in the apps were instruction on how to perform a behavior (n=32), action planning (n=30), and self-monitoring of behavior (n=28). The mean behavior change potential score (App Behavior Change Scale) was 8.5 (SD 3.1) of 21 (range 3-15).

Conclusions: Mobile apps available to support patient care received from a physiotherapist are of variable quality. Although they contain some BCTs, the potential for behavior change varied widely across apps.

International Registered Report Identifier (IRRID): RR2-10.2196/29047

(JMIR Mhealth Uhealth 2024;12:e55003) doi:10.2196/55003

KEYWORDS

physiotherapy; physical therapy; digital health intervention; mobile app; behavior change technique; behavior change; exercise; systematic review; quality; rehabilitation; BCT; mHealth; mobile health; app; apps; physical activity; fitness; synthesis; syntheses; review methods; review methodology; search; searches; searching; systematic; mobile phone

Introduction

Background

Digital health in physiotherapy is an emergent area. It continues to gather speed with its use in both clinical practice and research growing exponentially [1-3]. Many digital modalities, with varied functions, are relevant to physiotherapy practice; these might include the delivery of timely digital patient information and resources through websites and patient portals [4], digital patient assessment using connected Bluetooth and wireless-enabled devices [5], digital models of care (telehealth) supported by video-based consultation [6-8], remote monitoring of patient status through wearables [9,10], and mobile health (mHealth) apps used to prescribe, monitor, and support home exercise programs [1,11].

Mobile apps are one type of digital health modality of particular interest because of the ubiquity of smartphone use and their ability to deliver digital therapeutics [12-14]. They are relatively inexpensive and thus scalable [15]. They are also worthy of scientific attention within the physiotherapy community because of their ability to support interventions targeting several aspects of behavior change, otherwise known as behavior change techniques (BCTs) [16-19]. Facilitating positive behavior change in people with health conditions is often central to the management approach of physiotherapists, aimed at deriving the best possible therapeutic benefit from an intervention. For example, behavior change may be required for a person to follow physiotherapist advice regarding activity pacing or to adhere to exercise and physical activity recommendations [1,16,19,20]. BCTs can be thought of as the components of an intervention that regulate the behavior by altering cause and effect [21]. As defined by Michie and Johnston [22], they may be thought of as the "active ingredients" that facilitate intended behaviors. Apps have the ability to incorporate and deliver numerous BCTs, which can mediate a behavioral target of physiotherapy care [23-25].

The app space continues to grow exponentially, with well over 300,000 health apps now available and over 200 health apps added to major app stores daily [18]. Concurrently, despite a growing evidence base regarding the role of mobile apps in physiotherapy contexts, there remains a dearth of high-quality research using validated and rigorous scientific methods assessing the quality of physiotherapy-specific mobile apps and their behavior change aspects (eg, BCTs used and behavior change potential) [11,25].

Objectives

The primary study objective was to describe mobile apps, intended for use by patients, that can support physiotherapy care. The secondary objective was to evaluate app quality [26] and any BCTs contained within them [21], and to evaluate the behavior change potential of the apps [24].

Methods

Study Design

The detailed protocol for this systematic review was previously published [25]. As this study is a systematic review of apps (in app stores) and does not involve human participants, ethics approval was not required. The review followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement's systematic review reporting principles, with minor adaptations as relevant for our search of app stores rather than research literature [27]. This method is common and adopted in other similar studies [18,28,29] (Multimedia Appendix 1) [30].

Search Strategy

Overview

A 2-phase search strategy was used to search the popular app stores (Apple App Store and Google Play), which was in line with other rigorous systematic reviews of health apps in app stores [29]. App store search strategy, keywords used, and steps can be seen in Textbox 1 [25]. Initial search was conducted in March 2021 and rerun in February 2023 to ensure further up-to-date coverage. To maintain the feasibility of the search, each search term combination was limited to the first 100 apps retrieved. This was also done to maintain the relevance of apps retrieved, which diminishes as the end of the search list is refreshed [28].

Textbox 1. Search strategy.

Step 1

"physiotherapy," "physio," "physical therapy," "physiotherapist," "physical therapist."

Step 2

"physiotherapy," "physio," "physical therapy," "physiotherapist," "physical therapist."

and

"assessment," "diagnosis," "digital," "eHealth," "evaluation," "examination," "exercise," "health promotion," "intervention," "physical activity," "plan," "care," "prevention," "rehabilitation," "screening," "pain," "self-management," "treatment," "support," "adherence."

Additionally, the websites of the top 10 member organizations of World Physiotherapy (the peak international physiotherapy body) based on number of members were searched for recommendation or endorsement of any specific apps that met our criteria. Only websites in English and with the relevant section not behind a paywall were searched.

Selection Process

The search was performed independently by 2 reviewers (MM and PV). The same reviewers independently screened the apps using 1 Apple and 1 Android device each. Screening was based on information in the respective app stores, including app title, description, and screenshots [31]. Apps eligible for full analysis were downloaded onto the devices for review by each reviewer independently, and any disagreements were resolved by discussion. A third reviewer (PM) was flagged as a mediator to resolve any nonconsensus, but this was not required [28].

Data Extraction

Overview

The 2 reviewers (MM and PV) independently extracted descriptive data about each app from the app stores, within the apps themselves, or from official websites of the apps (if readily apparent from information in the app store). All extracted data were computed in Microsoft Excel (Microsoft Corp). A full list of extracted descriptive characteristics is shown in Multimedia Appendix 2 [25].

Both reviewers (MM and PV) independently engaged with all of the app functions for a minimum of 10 minutes on each device, for familiarization. This allowed each reviewer to independently code and score the app quality, BCTs, and their behavior change potential.

Mobile App Quality

To appraise the quality of included apps, the 23-item Mobile App Rating Scale (MARS) was used [26]. MARS was used because of its reliability, simplicity, and wide applicability in mHealth research [26,32]. Its primary domains are engagement, functionality, aesthetics, and information, as well as a subjective rating of quality. Each domain has questions coded on a 5-point Likert scale (ranging from 1=inadequate to 5=excellent). Mean scores are calculated for each of the 4 domains, which are tallied and divided by the number of domains to produce an overall mean quality score out of 5. Mean scoring is used instead of total scores because individual items can be rated as "not applicable" [26].

Calculating the mean scores of the engagement, functionality, aesthetics, and information quality objective subscales and an overall mean app quality total score is how the MARS is scored. Mean scores instead of total scores are used because an item can be rated as "not applicable."

Both reviewers (MM and PV) completed MARS training before scoring each app independently. Any disagreements were rectified via discussion. A third reviewer (PM) was flagged as a mediator to resolve any nonconsensus, but this was not required.

Coding BCTs

Both reviewers (MM and PV) underwent and received certification in coding BCTs using the Behavior Change Technique Taxonomy version 1 (BCTTv1). This was to increase their competence in identifying and assessing BCTs and to improve coding agreement [33]. Based on the minimum of 10 minutes they engaged with app functions, each reviewer independently coded the BCTs incorporated in each app using the BCTTv1, a framework of 93 BCTs created for investigating behavior change interventions using valid and reliable methods [21]. Furthermore, the BCTs in the BCTTv1 are further arranged into 16 clusters, each including similar BCTs [21]. This clustering was helpful for coders when developing a novel codebook (Multimedia Appendix 3), as it helped clarify coding decisions when examining the behavior change potential of included apps [25]. Any disagreements were discussed to achieve consensus. Where consensus was not able to be achieved, a third reviewer (JJF; who is a behavior change expert and one of the original developers of the BCTTv1) helped resolve any disagreements.

Behavior Change Potential

The App Behavior Change Scale (ABACUS) was used to evaluate the behavior change "potential" of each app, which is a scale of 21 items [24]. Apps are scored by identifying any of the 21 items and summing these item scores to give a score out of 21. The ABACUS focuses on 4 BCT clusters: knowledge and information, goals and planning, feedback and monitoring, and actions [24]. Using clustered BCTs identified, we developed the aforementioned novel coding rulebook to support reporting, and the 2 reviewers (MM and PV) scored behavior change potential together using the ABACUS. Any disagreements were discussed at the time, and a third reviewer (JJF) decided in the case of nonconsensus.

Data Synthesis

App characteristics are reported as descriptive and categorical data and a proportion (percentage), including app name, data privacy transparency (yes or no), companion app (yes or no), platform (Apple App Store, Google Play, or both), focus of the app (specific condition or region or whether more general), physiotherapy specialty (eg, musculoskeletal, pelvic health, and pediatric), target behaviors, simplified target behaviors, country of origin, developer qualifications (not clear or health care professional [HCP] or non-HCP), app version number, payment method (free or in-app purchases or one-off payment), and cost. Overall mean (SD) and individual domain scores are presented for MARS and ABACUS scores, as well as total BCT identification frequency and BCTs coded per app.

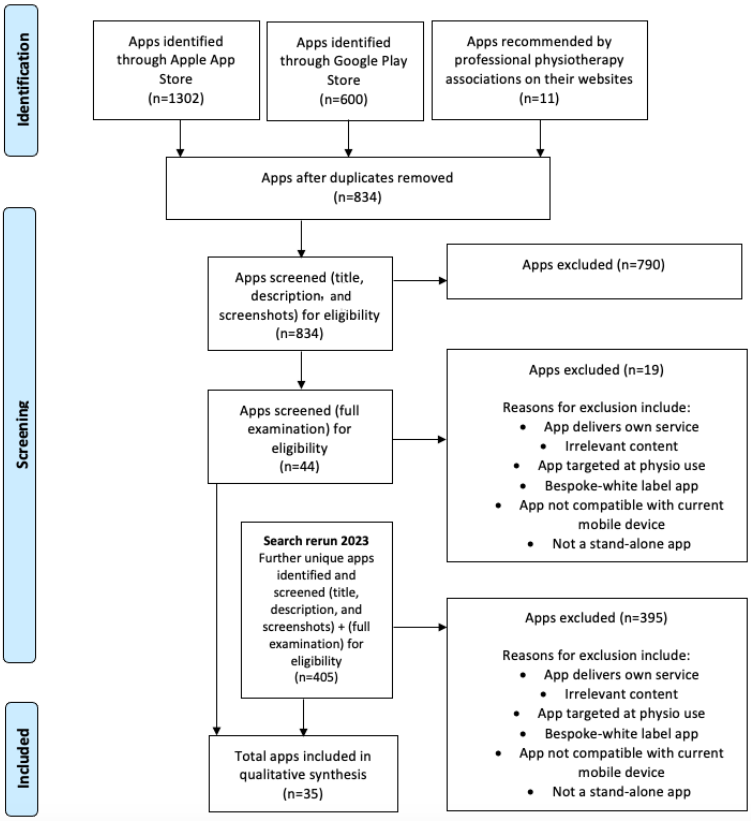
Results

Search

Our initial search identified a total of 1913 apps (Figure 1). This included 1302 unique apps from the Apple App Store, 600 apps from the Google Play Store, and a further 11 apps identified on the websites of professional physiotherapy associations. Overall, 834 were left after duplicates were removed. Following screening of app name, description, and screenshots, a further

790 were excluded. In total, 44 apps were screened fully, of which 19 were excluded. Reasons for exclusion included non-English language, costing more than Aus $10 (a currency exchange rate of Aus $1=US $0.72 is applicable), delivering its own service, irrelevant content, targeted at physiotherapist (clinician use) rather than the patient, being a bespoke or white-labeled app (ie, standard architectures licensed to a business or brand for private use), incompatibility with a current mobile device, and not a stand-alone app. This left 25 included apps for data extraction. Search update was rerun in February 2023, yielding an additional 405 unique apps. Following screening, this yielded a further 10 included apps for data extraction. Thus, a total of 35 apps were examined.

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



Descriptive Characteristics of Included Apps

App characteristics are described in Multimedia Appendix 2. Overall, 22 (63%) apps were available on both Apple App Store and Google Play platforms, and 12 (34%) only on Apple App Store. Only 1 (3%) app was solely available on Google Play. Country of origin was clear for 24 (69%) apps, with 6 apps originating from the United Kingdom, 6 from the United States, and 5 from Australia. Regarding price, 26 (74%) were free to download, and 9 had a one-off payment price ranging from Aus $1.49 to $9.99. Additionally, 10 apps offered in-app purchases (including subscriptions). Just under half of the apps (n=16, 46%) were not clear about the developer's qualifications, a further 17 (49%) were developed by HCPs (ie, physiotherapists, orthopedic surgeons, exercise physiologists, and chiropractors), and 2 (6%) were developed by non-HCPs. Overall, 15 (43%) apps had an obvious physiotherapist companion feature (ie, dashboard, desktop software, or precise login for therapists to access and engage with patients), and 31 (89%) apps had information about privacy and security in the app store. The focus of most apps was general in nature (eg, not condition-specific; n=24, 69%), while 11 (31%) had a specific target (eg, knee rehabilitation and pelvic floor training). This included 3 apps focusing on pelvic health and 1 app each focusing on wrist or hand, knee, hemophilia, core stability, wheelchair-bound individuals, cerebral palsy, osteoarthritis, and back pain.

Assessment of Quality (MARS)

Table 1 shows app quality scores using the MARS. Individual app quality scores ranged from 2.8 to 4.5, with a mean score of 3.7 (SD 0.4) of a maximum of 5. The apps with the highest MARS scores were Squeezy: CF (4.5), Squeezy (4.4), AllyCare (4.3), Squeezy for Men (4.2), and TeleHab (4.2). Of the appraised MARS domains, "aesthetics" (mean 3.9, SD 0.6) and "functionality" (mean 3.8, SD 0.5) were the highest scoring sections, while "subjective app quality" (mean 3.4, SD 0.8) and "engagement" (mean 3.4, SD 0.5) were the lowest. To the best of our knowledge, only 4 (11%) apps met criterion 19 (has the app been trialed or tested and published in scientific literature), including ReHand, Hand Rehabilitation, Squeezy, Embodia, and PhysiApp. Only 1 (2.9%) app (ReHand, Hand Rehabilitation) has been evaluated in a randomized controlled trial (RCT), showing positive outcomes for physiotherapy patients [34,35].

Table 1. Mobile App Rating Scale scoring.

App	Section A: engagement	Section B: functionality	Section C: aesthetics	Section D: information	Mean app quality score (section A to D)	Section E: subjective app quality
A Rehab Diary	3.6	3.0	2.7	3.2	3.1	2.5
AllyCare	4.4	4.0	4.7	4.2	4.3	4.5
Back Pain Diary	3.4	4.0	3.7	2.8	3.5	2.5
BlueJay Engage - Patient	3.8	4.0	4.3	4.0	4.0	4.0
ComplexCore	2.4	4.0	4.0	3.3	3.4	3.0
CP-Fit	3.2	3.5	4.0	3.3	3.5	3.0
Embodia	3.8	4.0	4.3	3.6	3.9	4.0
ExorLive	3.6	3.3	3.0	3.1	3.3	3.0
Extensor	3.4	3.5	3.3	3.3	3.4	3.0
Guided Physio	3.4	4.3	4.3	3.8	4.0	4.0
HaemActive	3.6	3.5	4.7	3.6	3.8	4.0
Home Physio	3.0	4.0	4.3	3.3	3.7	2.5
My Exercise Messages	3.4	4.3	4.0	4.8	4.1	3.5
My Exercise Program	3.6	3.8	4.0	3.8	3.8	3.3
My Injury Exercises	2.8	4.0	4.0	3.5	3.6	3.3
OT App Lite	2.6	3.5	4.0	3.0	3.3	2.0
PhysiApp	3.4	4.3	4.3	3.7	3.9	4.0
Physiotools Trainer	3.0	3.5	3.3	3.3	3.3	3.3
Pocket Physio	2.4	2.5	3.0	3.3	2.8	2.3
PT Timer: Stretch & Exercise	3.0	3.5	3.0	3.2	3.2	2.0
PT-Helper Pro	3.0	3.3	3.3	3.0	3.2	2.5
RecovAware Knee Health Fitness	4.2	3.8	4.3	4.2	4.1	4.0
Rehab Guru Client	3.8	3.8	4.3	3.8	3.9	4.0
ReHand, Hand Rehabilitation	3.6	4.3	4.3	4.1	4.1	4.0
Smart Therapist	3.4	3.5	3.3	3.5	3.4	3.3
Squeezy: CF	4.0	5.0	4.3	4.5	4.5	4.5
Squeezy for Men	3.6	4.8	4.3	4.3	4.2	4.5
Squeezy	4.0	5.0	4.3	4.3	4.4	5.0
Switchback Health	3.2	3.8	3.7	3.5	3.5	3.0
TeleHab	4.0	4.5	4.7	3.7	4.2	4.0
Track Rehab	3.4	3.3	3.3	3.8	3.5	3.0
TrackActive Pro - Patient App	3.4	3.5	3.7	3.8	3.6	3.5
VR steps Home rehabilitation	4.0	3.8	3.7	3.3	3.7	3.3
Wheelchair Exercises	3.0	3.8	4.0	3.5	3.6	3.3
YRMOVE	3.0	3.5	3.0	2.6	3.0	2.3
Mean (SD)	3.4 (0.5)	3.8 (0.5)	3.9 (0.6)	3.6 (0.5)	3.7 (0.4)	3.4 (0.8)

Behavior Change: Target Behaviors and BCTs

The most frequently observed target behaviors in the apps included recording information about exercise (n=28, 80%), performing therapeutic exercise (general; n=25, 71%), and communicating with a health professional (n=12, 34%; Multimedia Appendix 2). Other observed target behaviors included performing therapeutic exercise (men's health, women's health, hand therapy, knee, and perioperative), connecting with a health professional (make an appointment), connecting with friends, general self-care, and postsurgical self-care.

Coded BCTs by frequency can be observed in Multimedia Appendix 4, with a glossary of BCTs presented in Multimedia Appendix 5. The mean number of BCTs identified per app was 8.5 (SD 3.6). The apps with the highest number of unique BCTs were AllyCare (16), My Exercise Messages (14), ReHand, Hand Rehabilitation (14), and PhysiApp, BlueJay Engage, A Rehab Diary, and TeleHab (all with 13 BCTs identified). The most frequently coded BCTs were BCT 4.1 (instruction on how to perform a behavior; coded in n=32, 91% apps), BCT 1.4 (action planning; coded in n=30, 86% apps), BCT 2.3 (self-monitoring of behavior; coded in n=28, 80% apps), BCT 2.2 (feedback on behavior; coded in n=27, 77% apps), BCT 6.1 (demonstration of the behavior; coded in n=27, 77% apps), BCT 7.1 (prompts or cues; coded in n=25, 71% apps), and BCT 9.1 (credible source; coded in n=25, 71% apps).

Assessment of Behavior Change Potential (ABACUS)

Table 2 shows ABACUS scores for each app. The behavior change potential of included apps was a mean of 8.5 (SD 3.1) of a maximum of 21 (range 3-15). Of the 4 domains assessed by the ABACUS, section 1 (knowledge and information: mean score 3.0, SD 0.9) and section 3 (feedback and monitoring: mean score 2.9, SD 1.6) scored the highest, while section 2 (goals and planning: mean score 0.3, SD 0.8) and section 4 (actions: mean score 2.4, SD 1.0) scored the lowest. The apps (n=21) with the highest ABACUS scores included AllyCare (n=15, 71%), My Exercise Messages (n=14, 67%), ExorLive Go (n=14, 67%), A Rehab Diary (n=13, 62%), and both BlueJay Engage (n=12, 57%) and ReHand, Hand Rehabilitation (n=12, 57%).

Table 2. App Behavior Change Scale scoring.

	Section 1: Knowledge and information	Section 2: Goals and planning	Section 3: Feedback and monitoring	Section 4: Actions	Total
A Rehab Diary	3	2	5	3	13
AllyCare	4	2	6	3	15
Back Pain Diary	1	0	3	0	4
BlueJay Engage - Patient	4	1	4	3	12
ComplexCore	3	0	1	2	6
CP-Fit	2	0	3	2	7
Embodia	4	0	4	3	11
ExorLive	4	2	5	3	14
Extensor	3	0	4	2	9
Guided Physio	4	0	2	0	6
HaemActive	4	0	2	3	9
Home Physio	2	0	0	3	5
My Exercise Messages	3	3	4	4	14
My Exercise Program	3	0	3	2	8
My Injury Exercises	3	0	0	1	4
OT App Lite	2	0	0	1	3
PhysiApp	4	0	4	3	11
Physiotools Trainer	2	0	3	3	8
Pocket Physio	4	0	0	3	7
PT Timer: Stretch & Exercise	2	0	3	3	8
PT-Helper Pro	3	0	3	3	9
RecovAware Knee Health Fitness	3	0	3	3	9
Rehab Guru Client	2	0	4	3	9
ReHand, Hand Rehabilitation	3	0	5	4	12
Smart Therapist	2	0	3	2	7
Squeezy: CF	4	0	2	3	9
Squeezy for Men	4	0	2	3	9
Squeezy	4	0	2	3	9
Switchback Health	2	0	4	3	9
TeleHab	4	0	5	2	11
Track Rehab	4	0	4	3	11
TrackActive Pro - Patient App	2	0	3	2	7
VR steps Home rehabilitation	3	0	3	1	7
Wheelchair Exercises	2	0	0	1	3
YRMOVE	2	0	1	1	4
Mean (SD)	3.0 (0.9)	0.3 (0.8)	2.9 (1.6)	2.4 (1.0)	8.5 (3.1)

Discussion

Principal Findings

Implications

The primary aim of this study was to describe apps, intended for use by patients, that can support physiotherapy care. The secondary aims were to evaluate app quality, examine BCTs they contained, and evaluate behavior change potential. The findings of this study offer valuable insights into the current landscape of apps that may be used to support physiotherapy care and shed light on their ability to facilitate behavior change and potentially improve patient outcomes.

We identified 35 eligible apps, highlighting the popularity and increasing recognition of digital strategies for supplementing physiotherapy care [2,36]. It must also be noted that there is a growing number of apps within the broader exercise medicine and rehabilitation science space that may not have met the specific eligibility criteria for this systematic review but may still be of value in supporting physiotherapy care [1,18]. The rising number of available mobile apps necessitates careful and considered evaluation and selection of apps by clinicians and researchers to ensure that patients are accessing high-quality, effective, and safe apps that are evidence-based [1,11,37,38].

App Characteristics

Analysis of the app descriptive characteristics revealed interesting findings. Just under two-thirds of apps were cross-platform compatible (available on both the Google Play Store and Apple App Store). As smartphone ubiquity continues to grow, broader availability may contribute to wider accessibility and increase the uptake of apps to supplement physiotherapy care by both patients and physiotherapists [39]. The physiotherapy profession has recognized the value of mHealth tools to serve under-resourced communities and regions [40]. It is encouraging to note that in terms of price, the majority of the apps included in our review were free to download for patients. It must also be acknowledged that our review only examined apps that cost Aus $10 or less, which is in line with similar studies of apps for the management of arthritis, back pain, and persistent pain [29,31,32,41]. This follows similar research, suggesting that health consumers are less likely to buy health apps costing more than this [42]. However, there may well be more expensive apps available for purchase that were not included in our review. Conversely, while there are many seemingly free-to-download apps, several of the included apps require the physiotherapist to prescribe the patient a program from a companion app or dashboard before they can use it (eg, PhysiApp and TeleHab). In situations like this, the use of apps to support care is contingent on factors such as willingness to engage with digital health, acceptance, and digital health literacy of treating physiotherapists [2,37,38].

Relevant to an evidence-based profession like physiotherapy, the credibility of apps to support care remains an important consideration, and research suggests that this is a central factor in user engagement and satisfaction [43,44]. Less than half of the apps 16 (46%) provided clear information about the qualifications or background of the developers. Similarly, only 17 (49%) apps were clearly developed by health professionals (including physiotherapists). Research suggests that apps developed by or with health professionals may contribute to improved efficacy and safety, outcomes, and evidence-based care delivered by the apps [45]. However, while this may be true, a counterargument stands that while said to be developed by health professionals, it remains unclear what evidence was used in developing these apps.

Quality

A key aim of this study was to assess the quality of mobile apps relevant to physiotherapy care. While app quality is somewhat subjective, and measurement scales continue to emerge (eg, the recently released Deakin Health E-technologies Assessment Lab framework) [46], the MARS remains the most used, published, and validated [1,26,29]. Overall mean app quality in this study was 3.7 (SD 0.4), which is similar to previous research validating the MARS [47]. The authors reported a mean health app quality score of 3.74 (SD 0.6) of 5 and considered this as “moderate” quality. Notably, there is no universally accepted threshold for interpreting MARS scores. This suggests that the current landscape of physiotherapy apps has a little way to go in terms of quality. Our MARS data for the domains of “aesthetics” and “functionality” suggested that the included apps generally offer a good user experience and technical performance. These findings align with a quality appraisal of mobile apps for self-management of persistent pain conditions [29] but contrast to the relatively low scores obtained for apps specific to the management of low back pain [32].

Given that the physiotherapy profession is an evidence-based discipline, it is alarming that we found that only 4 (11%) of included apps met MARS criterion 19 (has the app been trialed or tested and published in scientific literature). Further, only 1 (3%; ReHand, Hand Rehabilitation) has been examined in an RCT [34,35,48]. However, the SMS-based precursor, which informed the development of My Exercise Messages, has undergone rigorous RCTs, showing positive patient outcomes [48,49]. An RCT studying the My Exercise Messages app is currently underway, and the protocol has been published [50]. As digital physiotherapy practice evolves, a combination of (1) greater assessment and scrutiny of digital health tools (eg, apps) through standardized validated measures and (2) high-quality well-designed RCTs and systematic reviews will assist in building evidence for and ultimately trust in digital tools that can support physiotherapy care [51-53].

Behavior Change

Another aim was to determine the BCTs included in apps and the apps’ potential for behavior change. Facilitating patient behavior change is an important aspect of physiotherapy care that may determine the effectiveness of many physiotherapy interventions [54-56]. Analysis of target behaviors suggested that the apps included in this review most commonly targeted behaviors linked to recording information about exercise, performing therapeutic exercise, and communicating with a health professional. This is encouraging, as these behaviors are often central to the core aims of many physiotherapeutic interventions [55,57]. The most frequently occurring BCTs coded within the included apps were instruction on how to

perform a behavior, action planning, self-monitoring of behavior, feedback on behavior, demonstration of the behavior, prompts or cues, and credible source, and these align very well to digital behavior change interventions using apps reported in literature on physical activity in osteoarthritis, various musculoskeletal conditions, and falls prevention exercise [58-60].

The potential for behavior change (based on ABACUS findings) could only be considered modest, given the mean ABACUS score of 8.5/21 (SD 3.1; 41%) across the 35 apps evaluated. This is similar to a recent comprehensive analysis of mobile apps to support behavior change in patients with a chronic disease or multimorbidity. In that study, mean ABACUS score was 8.07/21 (38%) [19]. Given that the developers of the ABACUS [24] have highlighted that the clinical implications of ABACUS score are still to be determined, it is not possible to determine whether the apps included in this study are effective or not at changing behavior or if they are likely to be effective at improving clinical outcomes from physiotherapy care. Future prospective research is required to evaluate if the apps included in our review can change patient behavior over time.

Limitations

The study has limitations. The review only analyzed apps in the English language, which limits the generalizability of the findings. Importantly, the search strategy had specific and stringent inclusion or exclusion criteria, which may lead to interpretative bias. For instance, apps had to be directly identifiable as being supportive of care between a patient and their physiotherapist. While several apps in app stores may indirectly support physiotherapy (eg, general exercise or informational apps), these were out of scope. Similarly, our search was based on a specified set of keywords, and thus, retrieval may not have captured all relevant apps. Furthermore, some apps originally identified in the search strategy were not able to be trialed for several reasons such as no free trial readily available, not available in the select region, or no response by developers for limited-time access by the study authors. A similar scenario occurred for apps that were subscription-based. In these cases, free trials were examined where possible, which may have not included full app functionality, thus potentially leading to more conservative quality, BCT, and behavioral potential scores. Regarding app quality, this was not appraised by patients themselves in this study. Should this have occurred, a different version of the MARS, the user-MARS, would have been required [61]. It is possible that patients may achieve a different quality rating for the apps. This is an area worthy of future research.

In addition, our protocol involved apps being trialed for a minimum of 10 minutes [25,29]. Vaghefi and Tulu [62] suggest "... most users tend to withdraw from mHealth apps before the end of the first week." This raises the possibility that engaging with each app for a limited amount of time can make it difficult to get a complete picture of an app, with certain features and BCTs not being immediately apparent. This was observed firsthand in the case of "My Exercise Messages," which scored 14/21 on the ABACUS in this study but scored 17/21 when rated by its developers who are intimately familiar with all the features of the entire 24-week app program [50]. The discrepancy is likely due to the fact that the more BCTs become apparent, the longer a user engages with the app over 6 months. A similar issue lies in the possibility that important descriptive data about an app may be missed or inaccurate because we relied on limited data sources. For instance, not all the necessary detailed information about an app can be found in the app itself or within its app store description. In select cases, this information may be identifiable if publications about the app exist. However, our protocol was deliberately designed to more closely mimic how a lay user would likely find and interact with an app through an app store. For these reasons, the findings reported in this study may be conservative for some apps, and it is possible that quality and behavior change scores would be higher with proper in-depth use. We thus urge readers to use caution in interpreting our findings.

Finally, this study focused primarily on app descriptive characteristics, quality, BCTs, and behavior change potential. It was not an aim of this review to evaluate the clinical effectiveness of these apps in changing behavior or their efficacy toward improving patient health outcomes.

Conclusions

Mobile apps available to support patient care received from a physiotherapist are of variable quality and contain relatively few BCTs. The potential for behavior change varied widely across apps. This study has provided the first comprehensive examination of mobile apps specifically supporting the care of patients receiving physiotherapy.

Acknowledgments

This study was supported by a University of Melbourne Early Career Researcher Grant Scheme (MM). KLB is supported by a National Health & Medical Research Council Leadership Investigator Grant (1174431). RSH is supported by a National Health & Medical Research Council Senior Research Fellowship (1154217).

Conflicts of Interest

RSH and KLB designed the app My Exercise Messages, which was included in this systematic review. They were not involved in the screening or data extraction or analysis processes for any app in this review. KLB also reports receiving royalties from Wolter Kluwers for production of the UpToDate knee osteoarthritis clinical guidelines.

Multimedia Appendix 1

PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[PDF File (Adobe PDF File), 668 KB - mhealth_v12i1e55003_app1.pdf]

Multimedia Appendix 2
App descriptive data.
[PDF File (Adobe PDF File), 151 KB - mhealth_v12i1e55003_app2.pdf]

Multimedia Appendix 3
Rulebook to support App Behavior Change Scale coding.
[PDF File (Adobe PDF File), 55 KB - mhealth_v12i1e55003_app3.pdf]

Multimedia Appendix 4
Coded behavior change techniques by frequency.
[DOCX File , 35 KB - mhealth_v12i1e55003_app4.docx]

Multimedia Appendix 5
Coded behavior change techniques and glossary.
[PDF File (Adobe PDF File), 239 KB - mhealth_v12i1e55003_app5.pdf]

References

1. Agnew JMR, Hanratty CE, McVeigh JG, Nugent C, Kerr DP. An investigation into the use of mHealth in musculoskeletal physiotherapy: scoping review. JMIR Rehabil Assist Technol 2022;9(1):e33609 [FREE Full text] [doi: 10.2196/33609] [Medline: 35275089]
2. Merolli M, Gray K, Choo D, Lawford BJ, Hinman RS. Use, and acceptability, of digital health technologies in musculoskeletal physical therapy: a survey of physical therapists and patients. Musculoskelet Care 2022;20(3):641-659 [FREE Full text] [doi: 10.1002/msc.1627] [Medline: 35278266]
3. Estel K, Scherer J, Dahl H, Wolber E, Forsat ND, Back DA. Potential of digitalization within physiotherapy: a comparative survey. BMC Health Serv Res 2022;22(1):496 [FREE Full text] [doi: 10.1186/s12913-022-07931-5] [Medline: 35418069]
4. Slater H, Stinson JN, Jordan JE, Chua J, Low B, Lalloo C, et al. Evaluation of digital technologies tailored to support young people's self-management of musculoskeletal pain: mixed methods study. J Med Internet Res 2020;22(6):e18315 [FREE Full text] [doi: 10.2196/18315] [Medline: 32442143]
5. Rose MJ, Costello KE, Eigenbrot S, Torabian K, Kumar D. Inertial measurement units and application for remote health care in hip and knee osteoarthritis: narrative review. JMIR Rehabil Assist Technol 2022;9(2):e33521 [FREE Full text] [doi: 10.2196/33521] [Medline: 35653180]
6. Hörder H, Nero H, Ignjatovic MM, Kiadaliri A, Lohmander LS, Dahlberg LE, et al. Digitally delivered exercise and education treatment program for low back pain: longitudinal observational cohort study. JMIR Rehabil Assist Technol 2022;9(2):e38084 [FREE Full text] [doi: 10.2196/38084] [Medline: 35727622]
7. Malliaras P, Merolli M, Williams CM, Caneiro JP, Haines T, Barton C. 'It's not hands-on therapy, so it's very limited': telehealth use and views among allied health clinicians during the coronavirus pandemic. Musculoskelet Sci Pract 2021;52:102340 [FREE Full text] [doi: 10.1016/j.msksp.2021.102340] [Medline: 33571900]
8. Bennell KL, Lawford BJ, Metcalf B, Mackenzie D, Russell T, van den Berg M, et al. Physiotherapists and patients report positive experiences overall with telehealth during the COVID-19 pandemic: a mixed-methods study. J Physiother 2021;67(3):201-209 [FREE Full text] [doi: 10.1016/j.jphys.2021.06.009] [Medline: 34147399]
9. Ummels D, Beekman E, Moser A, Braun SM, Beurskens AJ. Patients' experiences with commercially available activity trackers embedded in physiotherapy treatment: a qualitative study. Disabil Rehabil 2020;42(23):3284-3292 [FREE Full text] [doi: 10.1080/09638288.2019.1590470] [Medline: 30973026]
10. Vallati C, Virdis A, Gesi M, Carbonaro N, Tognetti A. ePhysio: a wearables-enabled platform for the remote management of musculoskeletal diseases. Sensors (Basel) 2018;19(1):2 [FREE Full text] [doi: 10.3390/s19010002] [Medline: 30577467]
11. Thompson D, Rattu S, Tower J, Egerton T, Francis J, Merolli M. Mobile app use to support therapeutic exercise for musculoskeletal pain conditions may help improve pain intensity and self-reported physical function: a systematic review. J Physiother 2023;69(1):23-34 [FREE Full text] [doi: 10.1016/j.jphys.2022.11.012] [Medline: 36528508]
12. Arntz A, Weber F, Handgraaf M, Lällä K, Korniloff K, Murtonen KP, et al. Technologies in home-based digital rehabilitation: scoping review. JMIR Rehabil Assist Technol 2023;10:e43615 [FREE Full text] [doi: 10.2196/43615] [Medline: 37253381]
13. Elgert L, Steiner B, Saalfeld B, Marschollek M, Wolf KH. Health-enabling technologies to assist patients with musculoskeletal shoulder disorders when exercising at home: scoping review. JMIR Rehabil Assist Technol 2021;8(1):e21107 [FREE Full text] [doi: 10.2196/21107] [Medline: 33538701]

14. Bailey JF, Agarwal V, Zheng P, Smuck M, Fredericson M, Kennedy DJ, et al. Digital care for chronic musculoskeletal pain: 10,000 participant longitudinal cohort study. J Med Internet Res 2020;22(5):e18250 [FREE Full text] [doi: 10.2196/18250] [Medline: 32208358]
15. Silva BMC, Rodrigues JJPC, de la Torre Díez I, López-Coronado M, Saleem K. Mobile-health: a review of current state in 2015. J Biomed Inform 2015;56:265-272 [FREE Full text] [doi: 10.1016/j.jbi.2015.06.003] [Medline: 26071682]
16. Greenstein J, Topp R, Etnoyer-Slaski J, Staelgraeve M, McNulty J. Effect of a mobile health app on adherence to physical health treatment: retrospective analysis. JMIR Rehabil Assist Technol 2021;8(4):e31213 [FREE Full text] [doi: 10.2196/31213] [Medline: 34655468]
17. Rintala A, Rantalainen R, Kaksonen A, Luomajoki H, Kauranen K. mHealth apps for low back pain self-management: scoping review. JMIR Mhealth Uhealth 2022;10(8):e39682 [FREE Full text] [doi: 10.2196/39682] [Medline: 36018713]
18. Ryan S, Chasaide NN, O' Hanrahan S, Corcoran D, Caulfield B, Argent R. mHealth apps for musculoskeletal rehabilitation: systematic search in app stores and content analysis. JMIR Rehabil Assist Technol 2022;9(3):e34355 [FREE Full text] [doi: 10.2196/34355] [Medline: 35916688]
19. Bricca A, Pellegrini A, Zangger G, Ahler J, Jäger M, Skou ST. The quality of health apps and their potential to promote behavior change in patients with a chronic condition or multimorbidity: systematic search in App Store and Google Play. JMIR Mhealth Uhealth 2022;10(2):e33168 [FREE Full text] [doi: 10.2196/33168] [Medline: 35119367]
20. Chen M, Wu T, Lv M, Chen C, Fang Z, Zeng Z, et al. Efficacy of mobile health in patients with low back pain: systematic review and meta-analysis of randomized controlled trials. JMIR Mhealth Uhealth 2021;9(6):e26095 [FREE Full text] [doi: 10.2196/26095] [Medline: 34114965]
21. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013;46(1):81-95 [FREE Full text] [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
22. Michie S, Johnston M. Behavior change techniques. In: Gellman MD, Turner JR, editors. Encyclopedia of Behavioral Medicine. New York, NY: Springer; 2013:182-187.
23. Willett M, Duda J, Gautrey C, Fenton S, Greig C, Rushton A. Effectiveness of behavioural change techniques in physiotherapy interventions to promote physical activity adherence in patients with hip and knee osteoarthritis: a systematic review protocol. BMJ Open 2017;7(6):e015833 [FREE Full text] [doi: 10.1136/bmjopen-2017-015833] [Medline: 28667221]
24. McKay FH, Slykerman S, Dunn M. The app behavior change scale: creation of a scale to assess the potential of apps to promote behavior change. JMIR Mhealth Uhealth 2019;7(1):e11130 [FREE Full text] [doi: 10.2196/11130] [Medline: 30681967]
25. Merolli M, Francis JJ, Vallance P, Bennell KL, Malliaras P, Hinman RS. Patient-facing mobile apps to support physiotherapy care: protocol for a systematic review of apps within app stores. JMIR Res Protoc 2021;10(12):e29047 [FREE Full text] [doi: 10.2196/29047] [Medline: 34889767]
26. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
27. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
28. Richardson B, Dol J, Rutledge K, Monaghan J, Orovec A, Howie K, et al. Evaluation of mobile apps targeted to parents of infants in the neonatal intensive care unit: systematic app review. JMIR Mhealth Uhealth 2019;7(4):e11620 [FREE Full text] [doi: 10.2196/11620] [Medline: 30985282]
29. Devan H, Farmery D, Peebles L, Grainger R. Evaluation of self-management support functions in apps for people with persistent pain: systematic review. JMIR Mhealth Uhealth 2019;7(2):e13080 [FREE Full text] [doi: 10.2196/13080] [Medline: 30747715]
30. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
31. Geuens J, Swinnen TW, Westhovens R, de Vlam K, Geurts L, Abeele VV. A review of persuasive principles in mobile apps for chronic arthritis patients: opportunities for improvement. JMIR Mhealth Uhealth 2016;4(4):e118 [FREE Full text] [doi: 10.2196/mhealth.6286] [Medline: 27742604]
32. Machado GC, Pinheiro MB, Lee H, Ahmed OH, Hendrick P, Williams C, et al. Smartphone apps for the self-management of low back pain: a systematic review. Best Pract Res Clin Rheumatol 2016;30(6):1098-1109. [doi: 10.1016/j.berh.2017.04.002] [Medline: 29103552]
33. Wood CE, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. Applying the behaviour change technique (BCT) taxonomy v1: a study of coder training. Transl Behav Med 2015;5(2):134-148 [FREE Full text] [doi: 10.1007/s13142-014-0290-z] [Medline: 26029276]
34. Blanquero J, Cortés-Vega MD, Rodríguez-Sánchez-Laulhé P, Corrales-Serra B, Gómez-Patricio E, Díaz-Matas N, et al. Feedback-guided exercises performed on a tablet touchscreen improve return to work, function, strength and healthcare

usage more than an exercise program prescribed on paper for people with wrist, hand or finger injuries: a randomised trial. J Physiother 2020;66(4):236-242 [FREE Full text] [doi: 10.1016/j.jphys.2020.09.012] [Medline: 33069608]
35. Blanquero J, Cortés-Vega MD, García-Frasquet MA, Sánchez-Laulhé PR, de Los Bernardos MIND, Suero-Pineda A. Exercises using a touchscreen tablet application improved functional ability more than an exercise program prescribed on paper in people after surgical carpal tunnel release: a randomised trial. J Physiother 2019;65(2):81-87 [FREE Full text] [doi: 10.1016/j.jphys.2019.02.008] [Medline: 30926400]
36. Keel S, Schmid A, Keller F, Schoeb V. Investigating the use of digital health tools in physiotherapy: facilitators and barriers. Physiother Theory Pract 2023;39(7):1449-1468 [FREE Full text] [doi: 10.1080/09593985.2022.2042439] [Medline: 35293846]
37. Alkhaldi O, McMillan B, Maddah N, Ainsworth J. Interventions aimed at enhancing health care providers' behavior toward the prescription of mobile health apps: systematic review. JMIR Mhealth Uhealth 2023;11:e43561 [FREE Full text] [doi: 10.2196/43561] [Medline: 36848202]
38. Giebel GD, Speckemeier C, Abels C, Plescher F, Börchers K, Wasem J, et al. Problems and barriers related to the use of digital health applications: scoping review. J Med Internet Res 2023;25:e43808 [FREE Full text] [doi: 10.2196/43808] [Medline: 37171838]
39. Biørn-Hansen A, Grønli TM, Ghinea G, Alouneh S. An empirical study of cross-platform mobile development in industry. Wirel Commun Mob Comput 2019;2019:1-12 [FREE Full text] [doi: 10.1155/2019/5743892]
40. Balikuddembe JK, Reinhardt JD. Can digitization of health care help low-resourced countries provide better community-based rehabilitation services? Phys Ther 2020;100(2):217-224 [FREE Full text] [doi: 10.1093/ptj/pzz162] [Medline: 31680158]
41. Salazar A, de Sola H, Failde I, Moral-Munoz JA. Measuring the quality of mobile apps for the management of pain: systematic search and evaluation using the mobile app rating scale. JMIR Mhealth Uhealth 2018;6(10):e10718 [FREE Full text] [doi: 10.2196/10718] [Medline: 30361196]
42. Mani M, Kavanagh DJ, Hides L, Stoyanov SR. Review and evaluation of mindfulness-based iPhone apps. JMIR Mhealth Uhealth 2015;3(3):e82 [FREE Full text] [doi: 10.2196/mhealth.4328] [Medline: 26290327]
43. Wang C, Qi H. Influencing factors of acceptance and use behavior of mobile health application users: systematic review. Healthcare (Basel) 2021;9(3):357 [FREE Full text] [doi: 10.3390/healthcare9030357] [Medline: 33809828]
44. Handayani PW, Gelshirani NB, Azzahro F, Pinem AA, Hidayanto AN. The influence of argument quality, source credibility, and health consciousness on satisfaction, use intention, and loyalty on mobile health application use. Inform Med Unlocked 2020;20:100429 [FREE Full text] [doi: 10.1016/j.imu.2020.100429]
45. Akbar S, Coiera E, Magrabi F. Safety concerns with consumer-facing mobile health applications and their consequences: a scoping review. J Am Med Inform Assoc 2020;27(2):330-340 [FREE Full text] [doi: 10.1093/jamia/ocz175] [Medline: 31599936]
46. Dona SWA, Nguyen D, Angeles MR, Cooper P, Winter N, Chatterton ML, et al. Developing and testing a health app evaluation framework for organisations to recommend the best health apps for consumers in the Australian setting: executive summary. Deakin University—Deakin Health Economics, School of Health and Social Development, Institute for Health Transformation. 2022. URL: https://iht.deakin.edu.au/wp-content/uploads/sites/153/2022/07/D-HEAL-Final-Report.pdf [accessed 2024-01-27]
47. Terhorst Y, Philippi P, Sander LB, Schultchen D, Paganini S, Bardus M, et al. Validation of the Mobile Application Rating Scale (MARS). PLoS One 2020;15(11):e0241480 [FREE Full text] [doi: 10.1371/journal.pone.0241480] [Medline: 33137123]
48. Nelligan RK, Hinman RS, Kasza J, Crofts SJC, Bennell KL. Effects of a self-directed web-based strengthening exercise and physical activity program supported by automated text messages for people with knee osteoarthritis: a randomized clinical trial. JAMA Intern Med 2021;181(6):776-785 [FREE Full text] [doi: 10.1001/jamainternmed.2021.0991] [Medline: 33843948]
49. Bennell K, Nelligan RK, Schwartz S, Kasza J, Kimp A, Crofts SJ, et al. Behavior change text messages for home exercise adherence in knee osteoarthritis: randomized trial. J Med Internet Res 2020;22(9):e21749 [FREE Full text] [doi: 10.2196/21749] [Medline: 32985994]
50. Hinman RS, Nelligan RK, Campbell PK, Kimp AJ, Graham B, Merolli M, et al. Exercise adherence mobile app for knee osteoarthritis: protocol for the MappKO randomised controlled trial. BMC Musculoskelet Disord 2022;23(1):874 [FREE Full text] [doi: 10.1186/s12891-022-05816-6] [Medline: 36127677]
51. Merolli M, Hinman RS, Lawford BJ, Choo D, Gray K. Digital health interventions in physiotherapy: development of client and health care provider survey instruments. JMIR Res Protoc 2021;10(7):e25177 [FREE Full text] [doi: 10.2196/25177] [Medline: 34319242]
52. Zanaboni P, Ngangue P, Mbemba GIC, Schopf TR, Bergmo TS, Gagnon MP. Methods to evaluate the effects of internet-based digital health interventions for citizens: systematic review of reviews. J Med Internet Res 2018;20(6):e10202 [FREE Full text] [doi: 10.2196/10202] [Medline: 29880470]
53. Blandford A, Gibbs J, Newhouse N, Perski O, Singh A, Murray E. Seven lessons for interdisciplinary research on interactive digital health interventions. Digit Health 2018;4:2055207618770325 [FREE Full text] [doi: 10.1177/2055207618770325] [Medline: 29942629]

54. Kunstler BE, Cook JL, Freene N, Finch CF, Kemp JL, O'Halloran PD, et al. Physiotherapists use a small number of behaviour change techniques when promoting physical activity: a systematic review comparing experimental and observational studies. J Sci Med Sport 2018;21(6):609-615. [doi: 10.1016/j.jsams.2017.10.027] [Medline: 29233466]
55. Söderlund A, Elvén M, Sandborgh M, Fritz J. Implementing a behavioral medicine approach in physiotherapy for patients with musculoskeletal pain: a scoping review. Pain Rep 2020;5(5):e844 [FREE Full text] [doi: 10.1097/PR9.0000000000000844] [Medline: 33490840]
56. Hay-Smith EJC, McClurg D, Frawley H, Dean SG. Exercise adherence: integrating theory, evidence and behaviour change techniques. Physiotherapy 2016;102(1):7-9. [doi: 10.1016/j.physio.2015.08.006] [Medline: 26549601]
57. Bassett SF. Bridging the intention-behaviour gap with behaviour change strategies for physiotherapy rehabilitation non-adherence. N Z J Physiother 2015;43(3):105-111 [FREE Full text] [doi: 10.15619/nzjp/43.3.05]
58. Arkkukangas M, Cederbom S, Tonkonogi M, Carlsson Õ. Older adults' experiences with mHealth for fall prevention exercise: usability and promotion of behavior change strategies. Physiother Theory Pract 2021;37(12):1346-1352. [doi: 10.1080/09593985.2020.1712753] [Medline: 31910707]
59. Berry A, McCabe CS, Muir S, Walsh N. Digital behaviour change interventions to facilitate physical activity in osteoarthritis: a systematic review. Phys Ther Rev 2018;23(3):197-206. [doi: 10.1080/10833196.2018.1470747]
60. Beresford L, Norwood T. The effect of mobile care delivery on clinically meaningful outcomes, satisfaction, and engagement among physical therapy patients: observational retrospective study. JMIR Rehabil Assist Technol 2022;9(1):e31349 [FREE Full text] [doi: 10.2196/31349] [Medline: 35107436]
61. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and validation of the user version of the Mobile Application Rating Scale (uMARS). JMIR Mhealth Uhealth 2016;4(2):e72 [FREE Full text] [doi: 10.2196/mhealth.5849] [Medline: 27287964]
62. Vaghefi I, Tulu B. The continued use of mobile health apps: insights from a longitudinal study. JMIR Mhealth Uhealth 2019;7(8):e12983 [FREE Full text] [doi: 10.2196/12983] [Medline: 31469081]

Abbreviations

ABACUS: App Behavior Change Scale
BCT: behavior change technique
BCTTv1: Behavior Change Technique Taxonomy version 1
MARS: Mobile App Rating Scale
mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial

Edited by L Buis, G Eysenbach; submitted 29.11.23; peer-reviewed by TAR Sure, D Singh; comments to author 29.12.23; revised version received 08.01.24; accepted 23.01.24; published 04.03.24.

Please cite as:
Merolli M, Francis JJ, Vallance P, Bennell KL, Malliaras P, Hinman RS
Evaluation of Patient-Facing Mobile Apps to Support Physiotherapy Care: Systematic Review
JMIR Mhealth Uhealth 2024;12:e55003
URL: https://mhealth.jmir.org/2024/1/e55003
doi:10.2196/55003
PMID:38437018

Review

Quality and Accessibility of Home Assessment mHealth Apps for Community Living: Systematic Review

Jung-hye Shin[1], BFA, MS, PhD; Rachael Shields[1], BS, MLA; Jenny Lee[1]; Zachary Skrove[2], BS; Ross Tredinnick[3], BS, MS; Kevin Ponto[1,3], BS, MS, PhD; Beth Fields[2], BS, MS, PhD

[1]Department of Design Studies, University of Wisconsin-Madison, Madison, WI, United States

[2]Department of Kinesiology, University of Wisconsin-Madison, Madison, WI, United States

[3]Wisconsin Institute for Discovery, University of Wisconsin-Madison, Madison, WI, United States

Corresponding Author:
Jung-hye Shin, BFA, MS, PhD
Department of Design Studies
University of Wisconsin-Madison
1300 Linden Dr
Madison, WI, 53706-1524
United States
Phone: 1 608 890 3704
Email: jshin9@wisc.edu

Abstract

Background: Home assessment is a critical component of successful home modifications, enabling individuals with functional limitations to age in place comfortably. A high-quality home assessment tool should facilitate a valid and reliable assessment involving health care and housing professionals, while also engaging and empowering consumers and their caregivers who may be dealing with multiple functional limitations. Unlike traditional paper-and-pencil assessments, which require extensive training and expert knowledge and can be alienating to consumers, mobile health (mHealth) apps have the potential to engage all parties involved, empowering and activating consumers to take action. However, little is known about which apps contain all the necessary functionality, quality appraisal, and accessibility.

Objective: This study aimed to assess the functionality, overall quality, and accessibility of mHealth home assessment apps.

Methods: mHealth apps enabling home assessment for aging in place were identified through a comprehensive search of scholarly articles, the Apple (iOS) and Google Play (Android) stores in the United States, and fnd.io. The search was conducted between November 2022 and January 2023 following a method adapted from PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses). Reviewers performed a content analysis of the mobile app features to evaluate their functionality, overall quality, and accessibility. The functionality assessment used a home assessment component matrix specifically developed for this study. For overall quality, the Mobile Application Rating Scale (MARS) was used to determine the apps' effectiveness in engaging and activating consumers and their caregivers. Accessibility was assessed using the Web Content Accessibility Guidelines (WCAG) 2.1 (A and AA levels). These 3 assessments were synthesized and visualized to provide a comprehensive evaluation.

Results: A total of 698 apps were initially identified. After further screening, only 6 apps remained. Our review revealed that none of the apps used thoroughly tested assessment tools, offered all the functionality required for reliable home assessment, achieved the "good" quality threshold as measured by the MARS, or met the accessibility criteria when evaluated against WCAG 2.1. However, DIYModify received the highest scores in both the overall quality and accessibility assessments. The MapIt apps also showed significant potential due to their ability to measure the 3D environment and the inclusion of a desktop version that extends the app's functionality.

Conclusions: Our review revealed that there are very few apps available within the United States that possess the necessary functionality, engaging qualities, and accessibility to effectively activate consumers and their caregivers for successful home modification. Future app development should prioritize the integration of reliable and thoroughly tested assessment tools as the foundation of the development process. Furthermore, efforts should be made to enhance the overall quality and accessibility of these apps to better engage and empower consumers to take necessary actions to age in place.

(JMIR Mhealth Uhealth 2024;12:e52996) doi:10.2196/52996

KEYWORDS

mobile app; mobile applications; mHealth; mobile health; app; apps; application; applications; mobile phone; mobile app rating system; occupational therapy; home assessment; web accessibility; aging in place; accessible; accessibility; quality; rating; gerontology; geriatric; geriatrics; older adult; older adults; elder; elderly; older person; older people; ageing; aging; systematic; synthesis; syntheses; PRISMA; Google Play; content analysis; functionality; WCAG

Introduction

Home modifications are essential procedures for individuals with various functional limitations, enabling them to live independently within their own community. Traditionally seen as targeted biopsychosocial interventions, these modifications aim to address the functional limitations experienced by aging adults and individuals living with disabilities in their home environments. Additionally, home modifications are frequently used as part of hospital discharge planning after medical treatments such as geriatric and stroke rehabilitation [1,2]. Conducting a timely home assessment using a valid instrument and promptly implementing home modifications is crucial in assisting individuals recovering from medical procedures. These steps help maintain their functional abilities and ensure a reasonable quality of life in their homes [1,3,4].

The current gold standard for home modifications necessitates a systematic home assessment conducted by trained professionals, often performed by occupational therapists, as a prerequisite [3]. However, accessing such services remains challenging for many consumers [2,5,6]. Several contributing factors include (1) the lack of standardized and validated assessments and limited knowledge of best practices [7]; (2) the shortage of professionals trained to conduct these assessments, particularly in rural areas [6,8-10]; (3) the consumer's perceived burden from participating in comprehensive assessments [11,12]; and (4) the complexity and cost involved in conducting home assessments [13-16].

Reliable and validated home assessment tools do exist, albeit in a paper-and-pencil format. In a systematic review of the psychometric properties of available home accessibility assessment tools, Patry et al [7] identified several tools that meet critical psychometric properties, including In-Home Occupational Performance Evaluation (I-HOPE) [17], I-HOPE Assist [18], Housing Enabler (HE) [19,20], Comprehensive Assessment and Solution Process for Aging Residents [21], and Home and Community Environment [22]. However, challenges still exist in using these tools, such as the laborious and time-consuming measurement process and the difficulty in getting reimbursement for the cost of an occupational therapist's time [23]. The lack of objective environmental measurement has also been identified as a weakness of these tools, except for the HE [24].

The cost barrier and limitations in objectively measuring the physical environment have prompted researchers to explore the use of teleconferencing for remote home assessment tools [8,9,25]. More recently, several entities have started developing mobile health (mHealth) apps that integrate 3D modeling [26], virtual reality [27], and augmented reality (AR) tools [28,29] to measure, store, and share spatial data required for home modification solutions. However, what remains less understood and documented are the functionality, quality, and accessibility of mHealth apps. This is problematic as the number of mHealth apps for home assessment continues to increase, and there is no available evidence-informed guidance on which ones to use and why.

Therefore, the objective of this study was to systematically identify and evaluate publicly available mHealth apps available in the United States that focus on home modification in the context of aging in place, using three important criteria: (1) comprehensiveness of functionality, (2) overall quality leading to consumer engagement and follow-up actions, and (3) accessibility. A well-developed tool with all necessary functions can help professionals and consumers perceive that the app possesses the features and qualities they need to support collaboration with home modification providers in achieving desired goals [10,30].

Methods

App Identification

The research team conducted a systematic search across multiple information sources, including a database search of peer-reviewed journal papers on home assessment, the Apple (iOS) and Google Play (Android) stores in the United States, and fnd.io. The team followed PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses [31]) guidelines whenever applicable (Multimedia Appendices 1 and 2) and referred to reviews focused on mHealth apps [32-34]. The initial search was conducted between November 2022 and January 2023.

The first author (JS) conducted a search of 4 databases, including Academic Search Premier, APA PsycInfo, Consumer Health Complete—EBSCOhost, and MEDLINE. The search terms used were ("home assessment" or "home modifications" or "home mods" or "home adaptations") AND "occupational therapy" AND (technology or application or computer or tablet or mobile phone or smartphone or internet). Titles and abstracts were then reviewed using the following criteria:

- Years considered: 1990-2022
- Language: English
- Publication status: Published
- Publication type: Includes articles in peer-reviewed journals, encompassing all types of publications
- Home assessment focus: Accessibility, covering parts of or the entire house
- Types of functional limitations: All forms of functional limitations, including both physical and mental

- Exclusion criteria: Gaming apps for occupational therapy or medical training and exclusive use of technology for communication (telehealth)

In addition to articles on individual apps, the database search yielded 3 recent research publications conducting a meta-analysis of home assessment tools [10,23,35], prompting further manual searches.

The app store search was carried out by JL. JL systematically conducted individual searches on the Google Play store using a Samsung Galaxy S21 phone and on the Apple App Store using an iPhone 11. The search terms used were consistent with those used in the database search. To ensure comprehensive coverage of potentially relevant apps, the same method was applied to searches on the fnd.io website by the author RS.

All 3 reviewers (JS, RS, and JL) convened to establish common exclusion criteria and reach consensus on the final list of apps for analysis. The exclusion criteria encompassed (1) apps not intended for home assessment; (2) unstable operations that impeded effective use, such as frequent crashes and errors; (3) regional restrictions that limited access to certain apps for US users; and (4) apps with dubious objectives, such as prioritizing product promotion over facilitating home modifications for enhanced accessibility.

Data Extraction

Between January 2023 and March 2023, three distinct tools were used to evaluate the quality of the chosen apps: an app component matrix developed for this study, the Mobile App Rating Scale (MARS) [36], and the Web Content Accessibility Guidelines (WCAG) 2.1 (created by the World Wide Web Consortium) [37].

The analysis of app components focused on evaluating the features, capabilities, and operations of each home assessment tool to understand its usefulness compared with traditional paper-and-pencil evaluations [23]. The primary objective of this evaluation was to assess the potential of each app to be effectively used by professionals or consumers in the field. Traditional and validated home assessment tools, such as I-HOPE [17] and HE [38,39], typically enable evaluators to assess the functional limitations of residents and evaluate the physical aspects of the home environment to identify necessary adjustments for the identified functional limitations. These assessments typically do not encompass suggestions for subsequent home modifications, as such considerations lie beyond the purview of home assessments. Nonetheless, given that a significant number of the reviewed apps featured recommendations, the incorporation of recommendations was introduced into the component matrix for this study. Overall, the examination centered on assessing whether each app empowers users to appraise functional limitations, conduct home environment assessments (through checklists or measurements), generate assessment outcome reports, and offer recommendations.

The MARS is a reliable and objective tool used for classifying and assessing the overall quality of mHealth apps [36]. Unlike star ratings in app stores or subjective app reviews, the MARS provides a systematic approach to evaluate mHealth apps, offering a more comprehensive and reliable measure of their quality. Studies conducted by Stoyanov et al [36] and Terhorst et al [40] have reported a high level of construct and concurrent validity, as well as reliability and objectivity, with an intraclass correlation coefficient ranging between 0.82 and 0.85. This indicates a strong level of consistency among different MARS raters, further highlighting the reliability of the scale. Moreover, the MARS has been used by researchers to assess the ability of mHealth apps to engage and activate patients [34].

The MARS consists of 3 main components: App Quality Questions, App Subjective Quality Questions, and App Specific Questions. The App Quality Questions cover various categories to provide a comprehensive evaluation of the app. These categories include engagement (A), functionality (B), aesthetics (C), and information quality (D). The engagement category assesses factors such as fun, interest, individual adaptability, interactivity, and target group. Functionality focuses on the app's performance, usability, navigation, and gestural design. Aesthetics evaluates the layout, graphics, and visual appeal of the app. Information quality examines the accuracy, quantity, and quality of information provided, including the credibility and evidence base of the app.

In addition to the App Quality Questions, the MARS includes App Subjective Quality Questions to capture the reviewer's personal opinion and the perceived impact on the user. The reviewer's personal opinion (E) covers aspects such as app recommendation, willingness to pay for it, anticipated frequency of usage, and an assigned star rating. The perceived impact on the user (F) assesses how the app affects the user's knowledge, attitudes, intentions to change, and the likelihood of actual change in the target health behavior.

Each question in the MARS is aligned with a 5-point scale (1=inadequate, 2=poor, 3=acceptable, 4=good, 5=excellent). This scale provides a standardized framework for rating the app's performance across different categories. The unique structure and scale of the MARS allow reviewers to holistically evaluate mobile apps, considering both objective quality indicators and subjective assessments.

The research team also used the WCAG 2.1 [37]. These guidelines are designed to make web content more accessible to all users but is highly relevant to web and nonweb mobile phone content [41]. Considering that the target population of home assessment apps includes people with functional limitations, ensuring accessibility is deemed critical for the successful deployment of these apps. In comparison with the MARS rating, which primarily provides a general evaluation of mHealth apps, the WCAG assessment offers a comprehensive framework for evaluating accessibility criteria.

The WCAG 2.1 has four criteria categories: (1) perceivable, (2) operable, (3) understandable, and (4) robust. The "perceivable" category highlights the importance of users being able to perceive all presented information using their available senses. "Operable" refers to an interface that is easily navigable and usable by a wide range of users. The "understandable" category includes criteria ensuring that users should have no difficulty comprehending both the content and the user interface. Lastly, "robust" ensures that the content is consistently and accurately

interpreted by a diverse range of user agents, including assistive technologies.

WCAG 2.0 was initially released in December 2008 and was adopted into Section 508 of the Rehabilitation Act (29 USC 794d) in 2018. Any project receiving federal funds must adhere to Section 508 of the Rehabilitation Act. WCAG has 3 conformance levels: A, AA, and AAA, with level AAA being the highest level. To understand how WCAG operates, it is crucial to recognize that content meeting a higher level of compliance also satisfies all the criteria of a lower level. Additionally, it is worth highlighting that achieving full compliance with high-level standards is uncommon among apps, mainly due to the inherent difficulties and resource-intensive nature involved in meeting these criteria. This study used both the A and AA levels of WCAG 2.1 to evaluate the accessibility of all identified apps.

Data Analysis

The identified mHealth apps underwent testing and assessment by 3 reviewers (RS, JL, and ZS), with each being evaluated one at a time. All apps were installed and tested using an iPhone 13. The diverse training backgrounds of the 3 reviewers represented a mix of professionals likely to use or be involved in app development. ZS brings 2 years of experience in occupational therapy. RS has training and professional experience in architecture and landscape architecture, with substantial knowledge of building accessibility. JL's training encompasses interior design, user interface design, and graphic design expertise; she also has substantial knowledge of building accessibility.

The reviewers individually examined each app and determined if it included assessment components commonly found in traditional paper-and-pencil home assessment tools: functional limitations, physical environment assessment area, mode of measurement (checklist vs measurement), final report, and recommendations. As they evaluated these components, they considered observations during app usage and information from the app store. The components were agreed on during a postassessment meeting before being included in a matrix.

The MARS rating procedures followed the recommendations specified in the original study [36]. All 3 reviewers familiarized themselves with the MARS and watched the training video to gain a thorough understanding of its components and dimensions. Before assessing apps, they engaged in extensive discussion and consensus building to clarify the meaning and relevance of each MARS assessment item.

Subsequently, the reviewers extensively used each selected app for review to gain a comprehensive understanding of its features, functionality, content, and user experience. After individually assessing the apps, the reviewers convened to reach a consensus on the final scores and tabulated the mean scores. The mean scores were compiled from each section of the MARS, namely, (A) engagement, (B) functionality, (C) aesthetics, and (D) information. The app quality mean score was calculated by averaging the scores from these 4 sections. App subjective quality (E) and app-specific (F) scores were separately averaged and treated as distinct measures. The questions in the (F) section, which evaluate the perceived impact of the app on the user's knowledge, attitudes, intentions to change, and the likelihood of actual change in the target health behavior, were specifically responded to considering the target behavior of home assessment and modification, aligning them with the scope of this study. The agreed-on ratings were then compiled into a matrix to facilitate comparisons and analysis of ratings and data across the apps.

The accessibility assessment involved checking apps against all WCAG 2.1 criteria. During a preassessment meeting, the reviewers engaged in a detailed discussion of each of the 50 WCAG criteria. They collaboratively developed concise descriptions in the newly created WCAG evaluation form (Multimedia Appendix 3) and established a consensus on how to assess the apps based on these criteria. Individually, the reviewers assigned a pass, fail, or NA (not applicable) designation to each of the 50 WCAG criteria. The NA designation was used when a particular criterion covered a function that was not present in the app being evaluated.

In the postassessment meeting for each app, the reviewers convened to deliberate on the designations for each WCAG criterion and worked together to reach a consensus. Subsequently, the passing percentage for each WCAG category (perceivable, operable, understandable, and robust) was calculated, along with an overall passing percentage. The passing percentages were determined by excluding any criteria assigned an NA designation from the calculation.

The results of the quality and accessibility appraisal were visualized in a map. The map represents the quality appraisal using the MARS on the horizontal axis and WCAG 2.1 on the vertical axis. To enhance the objectivity of the MARS as a measure of app quality [36], we used the average score from the objective MARS items on the horizontal axis. On the vertical axis, we considered 2 levels of WCAG 2.1, namely, A and AA. Apps that scored higher on both parameters indicate high quality and accessibility, suggesting a greater potential for user engagement and activation [34,36].

Results

Identification

The initial database search yielded 104 articles, which were then reduced to 36 after eliminating duplicates. Among these, 3 meta-analyses on home assessment tools were included. After a comprehensive review of all identified articles from the database search, 3 mHealth apps were identified: HESTIA, MapIt Mobile, and Magicplan. After an initial assessment of downloadable apps, HESTIA was excluded due to its ongoing development and unavailability for download.

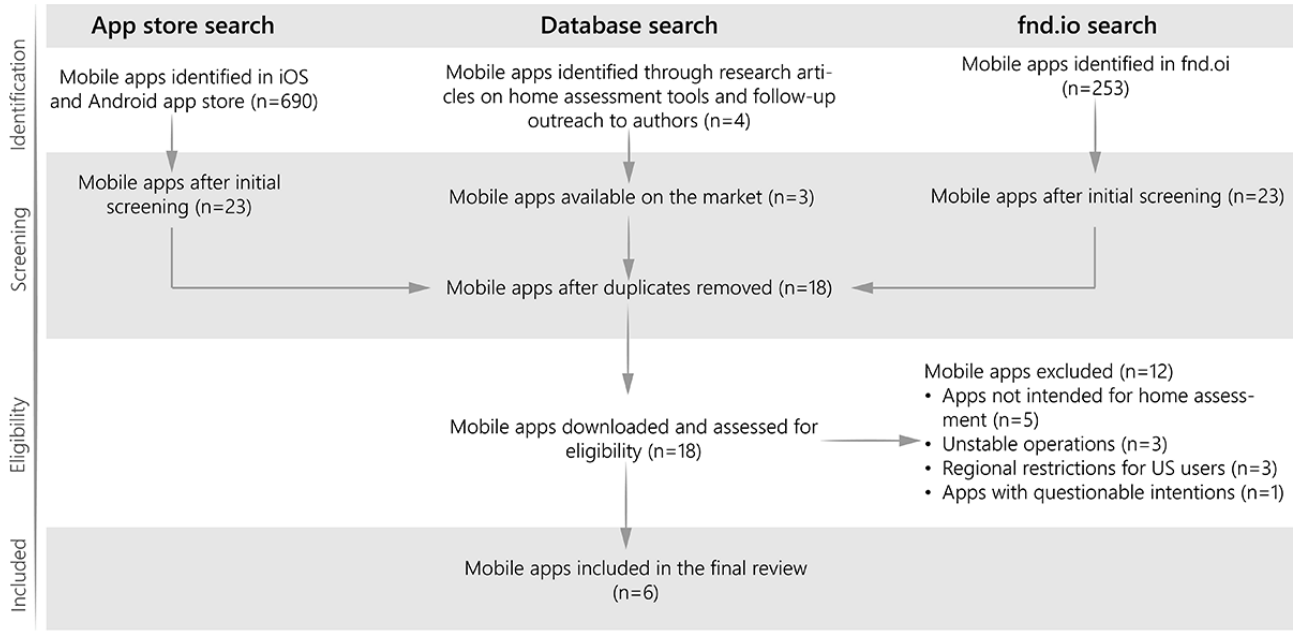
On engaging with the development team of MapIt, it was discovered that the desktop version of the app offered significantly enhanced functionality compared with its mobile counterpart. Because of the distinct interfaces between the 2 versions, separate evaluations were conducted for each. Although MapIt Desktop is not classified as an mHealth app, it was included in the evaluation due to its close association with the mobile app and its robustness in offering 3D

measurement capabilities in the context of home assessment, which were not found in any other apps.

The initial app store search yielded a substantial 690 apps. After reviewing titles, descriptions, preview photos, and keywords, 23 apps remained. The fnd.io search yielded a total of 253 apps, which were then narrowed down to 16 apps using the same screening method used for the app store search.

After removing duplicates from all 3 sources, 18 apps remained, all of which were downloaded for eligibility evaluation. Eventually, 12 apps were eliminated due to exclusion criteria, leaving 6 apps for detailed analysis: BEAT-D, DC Carehomes, DIYModify, HomeFit AR, MapIt Mobile, and MapIt Desktop (Figure 1).

Figure 1. Identification process of mobile health apps for home assessment in the United States.



Characteristics of the Included Apps

The characteristics of the included apps are presented in Table 1. Of the 6 apps reviewed, 4 (67%) apps were developed in university settings, 1 (17%) app was developed by a private company, and 1 app was developed by AARP, a nonprofit organization in the United States. Geographically, 3 (50%) apps (BEAT-D, DC Carehomes, and DIYModify) were developed in Australia, whereas 2 (33%) apps (MapIt Mobile and MapIt Desktop) were developed in Canada. The latter 2 apps were developed by the same entity, and they are functionally complementary to each other. Of the 6 apps, 1 (17%) app (HomeFit AR) was developed in the United States.

All the apps were designed for the iOS environment. Moreover, DC Carehomes and DIYModify were also developed for the Android platform. All apps were free to download. None of the reviewed apps were categorized as medical products nor had they published trials evaluating the effectiveness of the apps.

Among the 6 apps, only 2 (33%) apps (MapIt Mobile and MapIt Desktop) allowed the users to take actual measurements of the environment. These apps provided a feature for users to gather specific measurements. In contrast, all other apps were questionnaire-based interactive decision-making tools that asked a series of questions to the users as part of the home assessment process (Tables 1 and 2).

Table 1. Operating characteristics and summary of identified mobile health apps.

App name	Logo	Platform and operating system	Developer	Assessment characteristics: app summaries
BEAT-D		Mobile, tablet, and iOS	University of Wollongong, Australia	Questionnaire: a guided questionnaire assessment for buildings designed to accommodate individuals with dementia. The user's responses are compiled into a comprehensive report, which identifies areas that need improvement in order to reduce confusion, agitation, and depression.
DC Carehomes		Mobile, web, tablet, iOS, and Android	Private Company: Hammond Care, Australia	Questionnaire: a guided questionnaire assessment for care homes, units, or households catering specifically to individuals with dementia. The app generates a comprehensive report that offers recommendations based on the assessment findings.
DIYModify		Mobile, iOS	University of New South Wales, Australia	Questionnaire: an interactive decision-making tool that concentrates on 5 particular home modifications and provides guidance. It helps users select appropriate product types that match their needs and offers instructions on taking necessary measurements before shopping for home modifications. The app includes real-life stories of individuals who have undergone these specific adaptations, allowing users to learn from their experiences.
HomeFit		Mobile, tablet, and iOS	Nonprofit: AARP, United States	Questionnaire: an interactive decision-making tool assists users in identifying potential home improvements for aging in place. The app generates a comprehensive report, including tips, suggestions, and a checklist, based on the user's responses. The checklist distinguishes between tasks suitable for do-it-yourself and those requiring professional assistance. While the app uses AR[a] to recognize specific features such as a kitchen sink, it does not use AR for actual measurement purposes.
MapIt		Mobile, tablet, iOS, and Android	University of Sherbrooke, Canada	Measurement: the app uses AR and the LiDAR[b] sensor on the phone to create a 3D scan of a room. Users can then add measurements to specific areas of interest within the scan, catering to accessibility needs. The scan can be exported and viewed in the MapIt Desktop version, enhancing the overall measurement experience.
MapIt Desktop		Windows desktop, Mac OS X, and Windows	University of Sherbrooke, Canada	Measurement: this desktop app leverages MapIt on an iPhone to capture 3D scans of rooms, which can subsequently be imported to facilitate space measurements.

[a]AR: augmented reality.

[b]LiDAR: Light Detection and Ranging.

Table 2. Functional components and mode of assessment of the reviewed apps.

App name	Functional assessment	Environmental assessment				Summary report	Recommendation	Mode of assessment
		Entrance	Bathroom	Kitchen	Bedroom			
BEAT-D	—[a]	N[b]	N	N	N	Yes	Yes	Checklist
DC Carehomes	—	N	N	N	N	Yes	Yes	Checklist
DIYModify	—	T[c]	T	—	—	—	Yes	Checklist
HomeFit	—	—	T	T	T	Yes	Yes	Checklist
MapIt Mobile	—	Generic	Generic	Generic	Generic	—	—	Measurement
MapIt Desktop	—	Generic	Generic	Generic	Generic	—	—	Measurement

[a]Not available.

[b]N: nontargeted.

[c]T: targeted.

Component Analysis of the Included Apps

Table 2 includes a summary of the component analysis. Among the 6 apps assessed, none of them considered the functional limitations of the resident. Consequently, none of the apps enabled the evaluation of the physical environment tailored to the resident's individual functional capacities.

DIYModify and HomeFit AR stood out by delivering concentrated assessments for certain critical spaces, notably the entrance, bathroom, kitchen, and bedroom, designated as "targeted" (T) in Table 2. DIYModify, in particular, enabled users to assess essential elements within entrance and bathroom areas, while HomeFit emphasized assessment of the bathroom, kitchen, and bedroom areas. However, it is worth noting that no single app assessed all of the key areas comprehensively. Conversely, the remaining apps either lacked specific evaluations for the designated target spaces—although they did include questions related to those areas, classified as "nontargeted" (N)—or presented generalized assessment tools adaptable to any area, categorized as "Generic" in Table 2.

The reporting modules within these apps should ideally furnish a concise overview of assessment findings on each evaluation's conclusion, aiding users in comprehending the assessment outcomes and strategizing potential home modifications. Among the 6 apps under scrutiny, only HomeFit, DC Carehomes, and BEAT-D yielded a comprehensive report after completion of the assessment. Notably, HomeFit AR, DIYModify, DC Carehomes, and BEAT-D offered recommendations. Conversely, DIYModify omitted the provision of a report, whereas both MapIt Mobile and Desktop were deficient in both report and recommendation functionalities.

Quality Appraisal of the Included Apps: MARS

The MARS scores for the 6 parameters of (A) engagement, (B) functionality, (C) aesthetics, (D) information, (E) subjective rating of the app overall, and (F) subjective ratings of app-specific features are presented in Table 3. Very few apps received a rating of "good" (4 or above [36]) across the measured parameters, although many of them achieved an "acceptable" (3 or above and below 4) range.

Table 3. Mobile Application Rating Scale (MARS) objective and subjective quality criteria and the assessment result.

App name	Objective quality					Subjective quality		
	(A) engagement	(B) functionality	(C) aesthetics	(D) information	Overall objective quality	(E) app overall	(F) app specific	Overall subjective quality
BEAT-D	1.6	3.75	3.3	3.4	3.0	2	3.0	2.5
DC Carehomes	2	3.5	3.3	4[a]	3.2	2	3.2	2.6
DIYModify	3	3.75	4[a]	4[a]	3.7	3.5	4.7[a]	4.1[a]
HomeFit AR	2.8	3	3.7	2.83	3.1	1.75	2.7	2.2
MapIt Mobile	3	2.75	3	3.25	3.0	3	1.5	2.3
MapIt Desktop	3.2	3.5	3.7	3.25	3.4	3.5	2.5	3

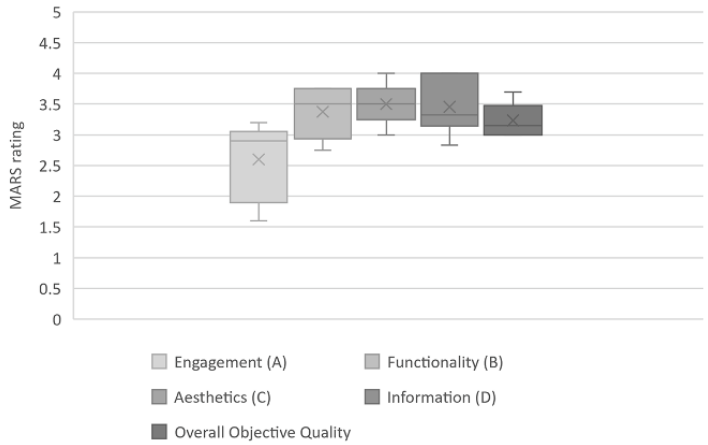
[a]Apps considered "good" (4 or above) under subcategories of the MARS assessment criteria.

When using the MARS objective quality criteria, none of the apps were rated as good in the categories of (A) engagement and (B) functionality. In the (C) aesthetic category, only DIYModify achieved a good rating with a score of 4. In the (D) information category, 2 apps, namely DC Carehomes and DIYModify, scored 4 and were thus classified as good. However, when considering the overall objective quality of the apps (mean of A, B, C, and D), none of the apps reached the threshold of 4.

Ratings for the MARS subjective quality were even lower. None of the reviewed apps reached the threshold of "good" in the (E) subjective rating of the app overall, and only half of them reached an "acceptable" level. Apps performed similarly in (F) subjective ratings on the app-specific features. However, DIYModify achieved an exceptionally high score of 4.7 out of 5. The incorporation of real-life video stories showcasing practical and effective home modifications within the app contributed to the high score, as it significantly enhanced the app's potential to positively influence the user's knowledge, attitudes, and actual behavior change in relation to home assessment.

When considering all reviewed apps together and focusing solely on the MARS objective quality criteria, the apps showed a tendency to perform better in the (C) aesthetics category (range 3-4; mean 3.50, SD 0.35) and (D) information category (range 2.84-4; mean 3.46, SD 0.46). However, their performances in (A) engagement were much lower (range 1.6-3.2; mean 2.6, SD 0.64; Figure 2). This discrepancy may be attributed to the "customization" score under the (A) engagement category, which scored the lowest (mean 1.33, SD 0.82 out of 5). On the other hand, the "gestural design" under the (B) functionality category scored the highest (mean 4, SD 0.63 out of 5), contributing to a slightly higher overall "functionality" score.

Figure 2. Box and whisker plot of the MARS objective quality assessments of all apps. MARS: Mobile App Rating Scale.



Accessibility Appraisal of the Included Apps: WCAG 2.1

The results from the accessibility evaluation using WCAG 2.1 are presented in Table 4. None of the reviewed apps conformed to either the A or AA version of WCAG 2.1. Conformance to these standards means that there is no content that violates the success criteria [42]. Considering that the multitude of criteria and any single issue such as a broken link or lack of voice-over recognition of a piece of text can cause an app to fail, this outcome is not surprising.

When evaluated against the overall success criteria for WCAG 2.1A, which is the most used standard in the field to meet basic accessibility requirements, apps achieved a conformance rate ranging from 65% to 86%. However, this range dropped to 53% to 71% when evaluated against WCAG 2.1AA. BEAT-D and DIYModify received the highest ratings in both evaluations (Table 4).

It is worth noting that all apps fulfilled at least 1 or 2 subcriteria of the WCAG. For example, DIYModify met all criteria under the "understandable" and "robust" categories of both the A and AA versions of WCAG 2.1. BEAT-D also met both criteria but only for WCAG 2.1A. Furthermore, all apps met the "robust" criteria of WCAG 2.1A (Table 4).

On examining the individual assessment items, we found that all apps passed at least a couple of items in each success criterion. In the "perceivable" category, all apps successfully met the assessment items of info and relationships (1.3.1) and meaningful sequence (1.3.2). Similarly, in the "operable" category, all apps passed the assessment items of 3 flashes (2.3.1) and pointer cancellation (2.5.2). Moving to the "understandable" category, the apps fulfilled the assessment items of language of page (3.1.1), language of parts (3.1.2), on focus (3.2.1), consistent navigation (3.2.3), and consistent identification (3.2.4). Lastly, in the "robust" category, the apps satisfied the assessment items of parsing (4.1.1) and name, role, value (4.1.2).

However, none of the reviewed apps managed to pass 3 assessment items. These items include resize text (1.4.4), which evaluates the ability to zoom in and enlarge text; reflow (1.4.10), which assesses the ability to reflow and adjust the content to fit the screen when zoomed in; and text spacing (1.4.12), which examines the ability to customize text characteristics. All these criteria are measured against the WCAG 2.1 AA level standards.

Table 4. Web Content Accessibility Guidelines (WCAG) 2.1 assessment success criteria and the assessment.

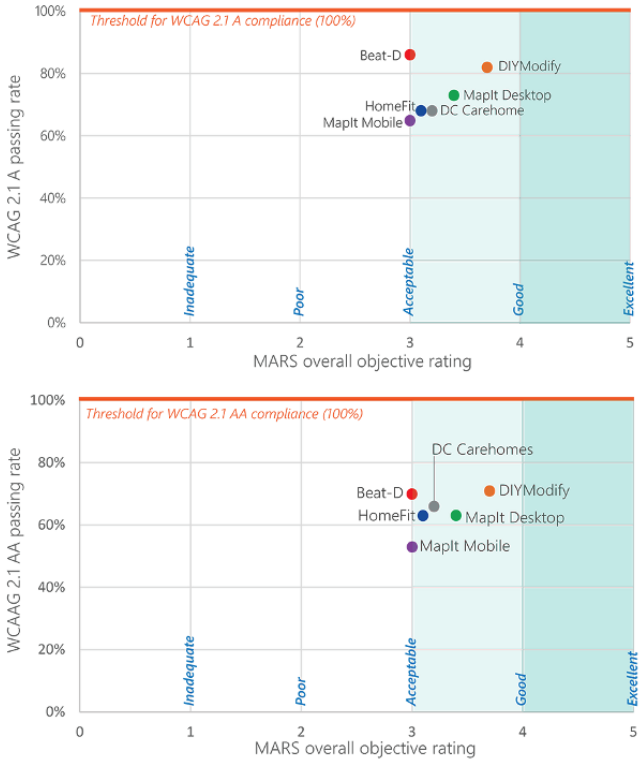
App name	Success criteria for WCAG 2.1 A					Success criteria for WCAG 2.1 AA				
	Perceivable	Operable	Understand-able	Robust	Overall passing	Perceivable	Operable	Understand-able	Robust	Overall passing
BEAT-D	4/5	7/9	5/5[a]	2/2[a]	0.86	6/13	9/12	9/9[a]	2/3	0.70
DC Carehomes	4/5	6/10	73/5	2/2[a]	0.68	8/13	8/13	7/9	2/3	0.66
DIYModify	6/9	6/7	4/4[a]	2/2[a]	0.82	8/17	8/9	7/7	2/2[a]	0.71
HomeFit AR	5/5[a]	4/8	2/4	2/2[a]	0.68	8/12	4/10	6/8	2/2[a]	0.63
MapIt Desktop	1/4	7/10	3/4	2/2[a]	0.65	1/11	8/13	6/7	3/3[a]	0.53
MapIt Mobile	1/4	8/8[a]	3/5	2/2[a]	0.73	4/11	8/10	6/9	2/2[a]	0.63

[a]Apps fulfilled a subcriterion of the WCAG 2.1. These items show that all apps met at least 1 or 2 subcriteria of the WCAG 2.1 despite their failure to meet WCAG 2.1 in its entirety.

The Objective Quality and Accessibility: Visual Synthesis of MARS×WCAG

The synthesis of the results from the MARS objective quality and WCAG assessments is visually represented in Figure 3. This visualization illustrates the performance of different apps based on the MARS overall objective quality appraisal and the WCAG 2.1 accessibility criteria. Although none of the apps met the thresholds to be considered both accessible and "good" according to WCAG 2.1 A and MARS, respectively, all of them were fairly close to these thresholds. Notably, DIYModify and MapIt Desktop achieved high scores for both accessibility and the MARS objective quality. While BEAT-D performed well in terms of accessibility, there is room for improvement in its overall MARS quality. On the other hand, HomeFit AR, MapIt Mobile, and DC Carehomes scored lower in both accessibility and MARS objective quality (Figure 3).

Figure 3. The visualization of apps' MARS quality and accessibility using the WCAG 2.1 A and AA criteria. This visualization indicates that all reviewed apps fall within the "acceptable" range, yet they fall short of achieving the "good" range as measured by the MARS tool. None of the apps managed to meet accessibility compliance when evaluated against both WCAG 2.1 A and AA standards. The lower visualization illustrates a greater level of challenge in meeting the WCAG 2.1 AA standard, which is a more stringent criterion than WCAG 2.1 A. MARS: Mobile App Rating Scale; WCAG: Web Content Accessibility Guidelines.



Discussion

Principal Findings

mHealth apps are expected to empower users with the capability to assess an individual's functional capacities and the environmental conditions crucial for comprehensive home evaluations [43]. Key areas such as the entrance, bathroom, kitchen, and bedroom hold significant importance in home assessments, which is reflected in conventional paper-and-pencil assessment tools and should thus be integral components of the app [21,38]. Furthermore, these apps should not only offer a succinct summary of the assessment but also motivate users to take up the subsequent steps after assessment, fostering engagement among all parties involved, including health care and housing professionals, as well as consumers and their caregivers. The manner in which these apps facilitate these processes should embody comprehensiveness, engagement, and accessibility.

Our findings demonstrate that, currently, there are no apps available in the United States that meet all of these criteria. Specifically, none of the apps allowed for the assessment of functional limitations of consumers, a crucial element in identifying areas requiring assessment. The MARS ratings revealed that all the apps were near the lower threshold of the acceptable range, with the exception of DIYModify; however, none of them reached the "good" range. The low scores in the "engagement" category, in particular, need further exploration as significant factors contributing to these apps' lower ratings. Furthermore, none of the reviewed apps met the accessibility criteria.

Despite these findings, our team notes that DIYModify scored the highest in our multidimensional assessment. We also observed that when used collectively, MapIt Mobile, MapIt Desktop, and the creator's instruction website demonstrate strong potential. Although we assessed them separately based on the parameters of our review, we found that using the entire suite together was highly effective for visualizing multiple measurements on a 3D scan. Offering an alternative viewing option on a larger screen device could prove beneficial for apps with complex user interfaces or content. On the other hand, the other apps (HomeFit AR, BEAT-D, and DC Carehomes) were

primarily questionnaire based and could have been easily accomplished without the need for an app.

Comparison With Prior Work

Our findings align with previous meta-analyses of home assessment tools, encompassing both paper-and-pencil formats as well as technology-assisted formats [10,23]. These studies have consistently revealed a lack of comprehensive and user-friendly technology tools that can be used to assess the home environment in relation to the functional abilities of its residents. Using technology to assess home environments in order to enhance accessibility and prevent falls and other injuries has remained challenging despite the rapid advancements in 3D modeling, virtual reality, and AR over the past few decades [10,23].

Within the limited pool of tools identified in previous studies, the majority were either pilot studies or exploratory qualitative studies [10]. Only a small fraction of these tools successfully transitioned into commercially available products, as confirmed by our search process across multiple app stores. Furthermore, review studies evaluating the efficacy of home assessment tools in all formats consistently demonstrated that the traditional paper-and-pencil assessment method was more effective in identifying issues [10]. This indicates a continued preference for the traditional method over digital alternatives among occupational therapy professionals. The findings of our study, which revealed that none of the reviewed apps allowed users to assess functional limitations, are in line with these earlier observations. This limitation hampers the effectiveness of the apps in detecting problematic areas and assessing accessibility.

Our study expands on previous research by evaluating the overall quality and accessibility of mHealth home assessment tools, emphasizing their significance in promoting consumer engagement and their follow-up actions. Specifically, we found that all of the reviewed apps met the minimum acceptable quality, but none reached the threshold of “good” quality. Additionally, none of the apps met the accessibility criteria as measured by the WCAG. This is a glaring omission as the home assessment and the subsequent modification are to help aid individuals with functional limitations in their home. To facilitate the active engagement and informed decision-making of older adults and individuals with functional limitations in their health care, as well as to empower them to undertake necessary home modifications, the development of apps that prioritize engagement, activation, and accessibility becomes imperative.

We anticipate that meeting this goal will remain challenging in the foreseeable future. Developing a mobile app may seem straightforward on the surface but can quickly become a multimillion-dollar project for several reasons. First, the development of mHealth apps for home assessment, with a focus on reliability, precision, and user-friendliness, necessitates robust interdisciplinary collaboration involving occupational therapists, building professionals, and user interface and user experience design experts.

Second, the inherent limitations in precision with current 3D scanning and AR technology, along with the need to meet the requirements of various devices, quickly add another layer of complexity to the endeavor [44-47]. Third, creating user-friendly apps requires extensive usability testing across diverse populations. In the case of home assessment, this involves testing with individuals exhibiting various functional limitations and their caregivers, contributing to the overall complexities and high cost of conducting such studies.

In contrast, the current funding landscape shows a tendency to prioritize research emphasizing cutting-edge scientific discovery or direct health outcomes with large-scale clinical trials. While the benefits of home modification have been studied either through indirect measures such as falls and emergency department visits via secondary data analysis [48,49], or a clinical trial [50], measuring the direct health outcome of the home assessment itself remains rather obscure.

Additionally, the transition from discovery to commercialization, as discussed earlier, introduces additional intricacies. Deploying the app in the market sustainably necessitates ongoing support, updates, and maintenance, contributing to the long-term cost of app development, which academic endeavors are not well suited for, often impeding the provision of free or affordable consumer apps. Above all, the lack of awareness and appreciation of the benefits that come from home assessments and home modification in the general public appears to be a key hindrance [11,12], discouraging adequate investment in this critical domain. With the rapid aging of the population and increasing interest and awareness from both the consumer market and the government alike, we hope that adequate resources are invested, fostering innovations in academia and the commercial realm alike.

Strengths and Limitations

Our review boasts several strengths. First, our comprehensive search encompassed scholarly databases, the US Google Play Store, Apple App Store, and fnd.io, ensuring a thorough exploration of available resources. The convergence of these searches instilled confidence in the thoroughness of our efforts. Furthermore, the significant disparity between the results of the scholarly database search and the app store search shed light on the challenges associated with translating scholarly endeavors into practical applications through the commercialization process.

Another strength lies in the complementary use of multiple rigorous assessments, focusing on both quality and accessibility. For instance, the BEAT-D app achieved a high passing percentage for WCAG criteria, yet it ranked lower in the MARS evaluation. On the other hand, MapIt Desktop had a lower score in WCAG, but ranked higher in the MARS. This discrepancy emphasizes the importance of conducting both assessments, particularly in the context of home assessment tools where individuals with functional limitations play a crucial role.

Finally, while evaluating apps with the MARS assessment is standardized and relatively straightforward, assessing apps based on WCAG criteria requires a more substantial time investment to grasp each of the 50 criteria, which is extensively explained on the WCAG’s website. To facilitate the use of WCAG, our research team has developed an evaluation form that includes

concise summaries of each criterion, which can be used in future studies, streamlining the evaluation process (Multimedia Appendix 1).

The review also presents certain weaknesses that deserve attention. First, despite our efforts to conduct a comprehensive search for all available home assessment tools, the inclusion of apps was limited by their availability in the US market. It is worth highlighting that several apps, discovered via our database search, fnd.io search, and personal connections, demonstrate promise but remain unavailable on US app stores.

Second, our testing of multiplatform apps was focused solely on the iPhone versions of all reviewed apps. This limitation was observed during the assessment of WCAG criterion 1.3.4 (orientation), which assesses the device's capacity to transition between portrait and landscape modes. Notably, the iPhone variant of BEAT-D did not meet this criterion, while the iPad version might have passed had it been evaluated. To ensure a more comprehensive assessment, it would be advantageous to test these apps on all compatible devices, including those running on Android operating systems.

Finally, while assessing apps with the MARS and WCAG provided valid insights into the overall quality and accessibility of apps based on established criteria, future studies will have to take into consideration consumer-level feedback, particularly focusing on the firsthand experiences of those with various functional capacities and their caregivers.

Conclusions

A proficient home assessment tool, designed to engage consumers, health care providers, and housing professionals, should offer reasonable functionality and possess objective quality and accessibility. However, our findings bring to light that none of the currently available home assessment mHealth apps in the United States align with these benchmarks. None of the apps offered sufficient methods to assess individuals' functional capacity and conduct comprehensive environmental assessments, and they fell short of meeting the WCAG accessibility criteria. Furthermore, although every app reached an "acceptable" level, none of them attained a "good" level in the MARS quality evaluations.

Acknowledgments

The review received funding from the RRF Foundation of Aging, the Fall Research Competition, and the School of Human Ecology at the University of Wisconsin-Madison.

Authors' Contributions

JS and RS played a leading role in conceptualizing, designing, and planning the review. App identification involved the contributions of JS, RS, and JL. Content analysis and quality or accessibility appraisal were conducted by RS, JL, and ZS under the supervision of JS. The initial draft of the manuscript was prepared by JS, incorporating the key technical findings provided by RS. All authors contributed to the manuscript by providing comments and suggestions for improvement, ultimately approving the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 checklist.
[PDF File (Adobe PDF File), 106 KB - mhealth_v12i1e52996_app1.pdf]

Multimedia Appendix 2
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 for abstracts checklist.
[PDF File (Adobe PDF File), 65 KB - mhealth_v12i1e52996_app2.pdf]

Multimedia Appendix 3
Accessibility evaluation form adapted from Web Content Accessibility Guidelines (WCAG) 2.1.
[DOC File , 101 KB - mhealth_v12i1e52996_app3.doc]

References

1. Schulz CH, Hersch GI, Foust JL, Wyatt AL, Godwin KM, Virani S, et al. Identifying occupational performance barriers of stroke survivors: utilization of a home assessment. Phys Occup Ther Geriatr 2012;30(2):109-123 [FREE Full text] [doi: 10.3109/02703181.2012.687441] [Medline: 24285912]
2. Lannin NA, Clemson L, McCluskey A. Survey of current pre-discharge home visiting practices of occupational therapists. Aust Occup Ther J 2011;58(3):172-177. [doi: 10.1111/j.1440-1630.2010.00911.x] [Medline: 21599682]
3. Stark S, Keglovits M, Arbesman M, Lieberman D. Effect of home modification interventions on the participation of community-dwelling adults with health conditions: a systematic review. Am J Occup Ther 2017;71(2):7102290010p1-7102290010p11. [doi: 10.5014/ajot.2017.018887] [Medline: 28218595]

4. Gitlin LN, Corcoran M, Winter L, Boyce A, Hauck WW. A randomized, controlled trial of a home environmental intervention: effect on efficacy and upset in caregivers and on daily function of persons with dementia. Gerontologist 2001;41(1):4-14 [FREE Full text] [doi: 10.1093/geront/41.1.4] [Medline: 11220813]
5. Greiman L, Ravesloot C, Goddard KS, Ward B. Effects of a consumer driven home modification intervention on community participation for people with mobility disabilities. Disabil Health J 2022;15(1S):101210 [FREE Full text] [doi: 10.1016/j.dhjo.2021.101210] [Medline: 34521604]
6. Carrington M, Islam MS. The use of telehealth to perform occupational therapy home assessments: an integrative literature review. Occup Ther Health Care 2023;37(4):648-663. [doi: 10.1080/07380577.2022.2056779] [Medline: 35357275]
7. Patry A, Vincent C, Duval C, Careau E. Psychometric properties of home accessibility assessment tools: a systematic review. Can J Occup Ther 2019;86(3):172-184. [doi: 10.1177/0008417418824731] [Medline: 31023068]
8. Latulippe K, Provencher V, Boivin K, Vincent C, Guay M, Kairy D, et al. Using an electronic tablet to assess patients' home environment by videoconferencing prior to hospital discharge: protocol for a mixed-methods feasibility and comparative study. JMIR Res Protoc 2019;8(1):e11674 [FREE Full text] [doi: 10.2196/11674] [Medline: 31344677]
9. Sanford JA, Jones M, Daviou P, Grogg K, Butterfield T. Using telerehabilitation to identify home modification needs. Assist Technol 2004;16(1):43-53. [doi: 10.1080/10400435.2004.10132073] [Medline: 15357147]
10. Ninnis K, Van Den Berg M, Lannin NA, George S, Laver K. Information and communication technology use within occupational therapy home assessments: a scoping review. Br J Occup Ther 2018;82(3):141-152. [doi: 10.1177/0308022618786928]
11. Golant SM. Resistance to assistance. CSA J 2018;1:27-31.
12. Kruse RL, Moore CM, Tofle RB, LeMaster JW, Aud M, Hicks LL, et al. Older adults' attitudes toward home modifications for fall prevention. J Hous Elderly 2010;24(2):110-129. [doi: 10.1080/02763891003757031]
13. Aplin T, de Jonge D, Gustafsson L. Understanding the dimensions of home that impact on home modification decision making. Aust Occup Ther J 2013;60(2):101-109. [doi: 10.1111/1440-1630.12022] [Medline: 23551003]
14. Scheckler S, Molinsky J, Airgood-Obrycki W. How well does the housing stock meet accessibility needs? an analysis of the 2019 American housing survey. Joint Center for Housing Studies of Harvard University, Cambridge, MA. 2022. URL: https://www.jchs.harvard.edu/sites/default/files/research/files/harvard_jchs_housing_stock_accessibility_scheckler_2022_0.pdf [accessed 2024-02-01]
15. Wiseman JM, Stamper DS, Sheridan E, Caterino JM, Quatman-Yates CC, Quatman CE. Barriers to the initiation of home modifications for older adults for fall prevention. Geriatr Orthop Surg Rehabil 2021;12:21514593211002161 [FREE Full text] [doi: 10.1177/21514593211002161] [Medline: 33868764]
16. Lanfranchi V, Jones N, Read J, Fegan C, Field B, Simpson E, et al. User attitudes towards virtual home assessment technologies. J Med Eng Technol 2022;46(6):536-546 [FREE Full text] [doi: 10.1080/03091902.2022.2089250] [Medline: 35730495]
17. Stark SL, Somerville EK, Morris JC. In-Home Occupational Performance Evaluation (I-HOPE). Am J Occup Ther 2010;64(4):580-589 [FREE Full text] [doi: 10.5014/ajot.2010.08065] [Medline: 20825129]
18. Keglovits M, Somerville E, Stark S. In-Home Occupational Performance Evaluation for providing assistance (I-HOPE assist): an assessment for informal caregivers. Am J Occup Ther 2015;69(5):6905290010 [FREE Full text] [doi: 10.5014/ajot.2015.015248] [Medline: 26356664]
19. Helle T, Nygren C, Slaug B, Brandt A, Pikkarainen A, Hansen AG, et al. The Nordic Housing Enabler: inter-rater reliability in cross-Nordic occupational therapy practice. Scand J Occup Ther 2010;17(4):258-266. [doi: 10.3109/11038120903265014] [Medline: 20078255]
20. Iwarsson S, Nygren C, Slaug B. Cross-national and multi-professional inter-rater reliability of the Housing Enabler. Scand J Occup Ther 2005;12(1):29-39. [doi: 10.1080/11038120510027144] [Medline: 16389996]
21. Sanford JA, Pynoos J, Tejral A, Browne A. Development of a comprehensive assessment for delivery of home modifications. Phys Occup Ther Geriatr 2009;20(2):43-55. [doi: 10.1080/j148v20n02_03]
22. Keysor J, Jette A, Haley S. Development of the Home and Community Environment (HACE) instrument. J Rehabil Med 2005;37(1):37-44 [FREE Full text] [doi: 10.1080/16501970410014830] [Medline: 15788331]
23. Mihandoust S, Joshi R, Joseph A, Madathil KC, Dye CJ, Machry H, et al. Identifying key components of paper-based and technology-based home assessment tools using a narrative literature review. J Aging Environ 2020;35(4):1-28. [doi: 10.1080/26892618.2020.1856754]
24. Swenor BK, Yonge AV, Goldhammer V, Miller R, Gitlin LN, Ramulu P. Evaluation of the Home Environment Assessment for the Visually Impaired (HEAVI): an instrument designed to quantify fall-related hazards in the visually impaired. BMC Geriatr 2016;16(1):214 [FREE Full text] [doi: 10.1186/s12877-016-0391-2] [Medline: 27938346]
25. Nix J, Comans T. Home quick—occupational therapy home visits using mHealth, to facilitate discharge from acute admission back to the community. Int J Telerehabil 2017;9(1):47-54 [FREE Full text] [doi: 10.5195/ijt.2017.6218] [Medline: 28814994]
26. Hamm J, Money AG, Atwal A, Ghinea G. Mobile three-dimensional visualisation technologies for clinician-led fall prevention assessments. Health Informatics J 2019;25(3):788-810 [FREE Full text] [doi: 10.1177/1460458217723170] [Medline: 28816091]

27. Threapleton K, Newberry K, Sutton G, Worthington E, Drummond A. Virtually home: exploring the potential of virtual reality to support patient discharge after stroke. Br J Occup Ther 2016;80(2):99-107. [doi: 10.1177/0308022616657111]
28. Chandrasekera T, Kang M, Hebert P, Choo P. Augmenting space: enhancing health, safety, and well-being of older adults through hybrid spaces. Technol Disabil 2017;29(3):141-151. [doi: 10.3233/tad-170159]
29. Guay M, Labbé M, Séguin-Tremblay N, Auger C, Goyer G, Veloza E, et al. Adapting a person's home in 3D using a mobile app (MapIt): participatory design framework investigating the app's acceptability. JMIR Rehabil Assist Technol 2021;8(2):e24669 [FREE Full text] [doi: 10.2196/24669] [Medline: 33973867]
30. Vo V, Auroy L, Sarradon-Eck A. Patients' perceptions of mHealth apps: meta-ethnographic review of qualitative studies. JMIR Mhealth Uhealth 2019;7(7):e13817 [FREE Full text] [doi: 10.2196/13817] [Medline: 31293246]
31. Page MJ, Moher D, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. BMJ 2021;372:n160 [FREE Full text] [doi: 10.1136/bmj.n160] [Medline: 33781993]
32. Jacob C, Sezgin E, Sanchez-Vazquez A, Ivory C. Sociotechnical factors affecting patients' adoption of mobile health tools: systematic literature review and narrative synthesis. JMIR Mhealth Uhealth 2022;10(5):e36284 [FREE Full text] [doi: 10.2196/36284] [Medline: 35318189]
33. Grainger R, Devan H, Sangelaji B, Hay-Smith J. Issues in reporting of systematic review methods in health app-focused reviews: a scoping review. Health Informatics J 2020;26(4):2930-2945 [FREE Full text] [doi: 10.1177/1460458220952917] [Medline: 32914696]
34. Cozad MJ, Crum M, Tyson H, Fleming PR, Stratton J, Kennedy AB, et al. Mobile health apps for patient-centered care: review of United States rheumatoid arthritis apps for engagement and activation. JMIR Mhealth Uhealth 2022;10(12):e39881 [FREE Full text] [doi: 10.2196/39881] [Medline: 36469397]
35. Hwang NK, Shim SH. Use of virtual reality technology to support the home modification process: a scoping review. Int J Environ Res Public Health 2021;18(21):11096 [FREE Full text] [doi: 10.3390/ijerph182111096] [Medline: 34769616]
36. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
37. W3C. Web Content Accessibility Guidelines 2.1, W3C World Wide Web Consortium Recommendation, Level A & Level AA Success Criteria. URL: https://www.w3.org/TR/WCAG21/ [accessed 2024-02-01]
38. Iwarsson S. The housing enabler: an objective tool for assessing accessibility. Br J Occup Ther 2016;62(11):491-497. [doi: 10.1177/030802269906201104]
39. Iwarsson S, Slaug B. Housing Enabler: A Method for Rating/Screening and Analysing Accessibility Problems in Housing, Manual for the Complete Instrument and Screening Tool, 2nd Edition. Lund and Staffanstorp, Sweden: Veten & Skapen HD and Slaug Enabling Development; 2010.
40. Terhorst Y, Philippi P, Sander LB, Schultchen D, Paganini S, Bardus M, et al. Validation of the Mobile Application Rating Scale (MARS). PLoS One 2020;15(11):e0241480 [FREE Full text] [doi: 10.1371/journal.pone.0241480] [Medline: 33137123]
41. Patch K, Spellman J, Wahlbin K. Mobile accessibility: how WCAG 2.0 and other W3C/WAI guidelines apply to mobile. World Wide Web Consortium. 2015. URL: https://www.w3.org/TR/mobile-accessibility-mapping/ [accessed 2024-02-01]
42. Cooper M, Kirkpatrick A, O'Connor J. Understanding conformance, understanding WCAG 2.0: a guide to understanding and implementing web content accessibility guidelines 2.0. World Wide Web Consortium. 2015. URL: https://www.w3.org/TR/UNDERSTANDING-WCAG20/conformance.html [accessed 2024-02-01]
43. Lawton MP, Nahemow L. Ecology and the aging process. In: Eisdorfer C, Lawton MP, editors. The Psychology of Adult Development and Aging. Washington, DC: American Psychological Association; 1973:619-674.
44. Boonbrahm S, Boonbrahm P, Kaewrat C. The use of marker-based augmented reality in space measurement. Procedia Manufacturing 2020;42:337-343 [FREE Full text] [doi: 10.1016/j.promfg.2020.02.081]
45. Marchand E, Uchiyama H, Spindler F. Pose estimation for augmented reality: a hands-on survey. IEEE Trans Vis Comput Graph 2016;22(12):2633-2651. [doi: 10.1109/TVCG.2015.2513408] [Medline: 26731768]
46. Ko SM, Chang WS, Ji YG. Usability principles for augmented reality applications in a smartphone environment. Int J Hum-Comput Int 2013;29(8):501-515. [doi: 10.1080/10447318.2012.722466]
47. Börsting I, Karabulut C, Fischer B, Gruhn V. Design patterns for mobile augmented reality user interfaces—an incremental review. Information 2022;13(4):159 [FREE Full text] [doi: 10.3390/info13040159]
48. Keall MD, Pierse N, Howden-Chapman P, Cunningham C, Cunningham M, Guria J, et al. Home modifications to reduce injuries from falls in the Home Injury Prevention Intervention (HIPI) study: a cluster-randomised controlled trial. Lancet 2015;385(9964):231-238. [doi: 10.1016/S0140-6736(14)61006-0] [Medline: 25255696]
49. Eriksen MD, Greenhalgh-Stanley N, Engelhardt GV. Home safety, accessibility, and elderly health: evidence from falls. J Urban Econ 2015;87:14-24. [doi: 10.1016/j.jue.2015.02.003]
50. Petersson I, Kottorp A, Bergström J, Lilja M. Longitudinal changes in everyday life after home modifications for people aging with disabilities. Scand J Occup Ther 2009;16(2):78-87. [doi: 10.1080/11038120802409747] [Medline: 18821447]

Abbreviations

AR: augmented reality
HE: Housing Enabler
I-HOPE: In-Home Occupational Performance Evaluation
MARS: Mobile Application Rating System
mHealth: mobile health
NA: not applicable
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
WCAG: Web Content Accessibility Guidelines

Edited by L Buis; submitted 21.09.23; peer-reviewed by G Lillie, A Atwal, V Vo; comments to author 07.11.23; revised version received 18.11.23; accepted 19.01.24; published 11.03.24.

Please cite as:
Shin JH, Shields R, Lee J, Skrove Z, Tredinnick R, Ponto K, Fields B
Quality and Accessibility of Home Assessment mHealth Apps for Community Living: Systematic Review
JMIR Mhealth Uhealth 2024;12:e52996
URL: https://mhealth.jmir.org/2024/1/e52996
doi:10.2196/52996
PMID:38466987

Original Paper

Assessment of Eye Care Apps for Children and Adolescents Based on the Mobile App Rating Scale: Content Analysis and Quality Assessment

Ming Liu[1], MS; Xiaoqian Wu[2], MS; Ziyu Li[1], MS; Dongmei Tan[1], MS; Cheng Huang[1], MD

[1]College of Medical Informatics, Chongqing Medical University, Chongqing, China

[2]Department of Quality Management, Daping Hospital, Army Medical University (The Third Military Medical University), Chongqing, China

Corresponding Author:
Cheng Huang, MD
College of Medical Informatics
Chongqing Medical University
1 Yixueyuan Road Yuzhong District
Chongqing, 400016
China
Phone: 86 023 6848 0060
Email: huangcheng@cqmu.edu.cn

Abstract

Background: In China, the current situation of myopia among children and adolescents is very serious. Prevention and control of myopia are inhibited by the lack of medical resources and the low awareness about eye care. Nevertheless, mobile apps provide an effective means to solve these problems. Since the health app market in China is still immature, it has become particularly important to conduct a study to assess the quality of eye-care apps to facilitate the development of better eye-care service strategies.

Objective: This study aimed to evaluate the quality, functionality, medical evidence, and professional background of eye-care apps targeting children and adolescents in the Chinese app stores.

Methods: A systematic search on iOS and Android app stores was performed to identify eye-care apps for children and adolescents. The general characteristics, development context, and functional features of the apps were described. Quality assessment of the apps was completed by 2 independent researchers using the Mobile App Rating Scale.

Results: This study included 29 apps, of which 17 (59%) were developed by commercial organizations and 12 (41%) had a design with relevant scientific basis. The main built-in functions of these apps include self-testing (18/29, 62%), eye exercises (16/29, 55%), and eye-care education (16/29, 55%). The mean overall quality of eye-care apps was 3.49 (SD 0.33), with a score ranging from 2.89 to 4.39. The overall Mobile App Rating Scale score exhibited a significant positive correlation with the subscale scores (r=0.81-0.91; P<.001). In addition, although most apps provided basic eye-care features, there are some deficiencies. For example, only a few apps were developed with the participation of medical organizations or professional ophthalmologists, and most of the apps were updated infrequently, failing to provide the latest eye-care information and technology in a timely manner.

Conclusions: In general, the quality of eye-care apps for children and teenagers in Chinese app stores is good. These apps fulfill users' needs for eye-care services to a certain extent, but they still suffer from insufficient medical background, low user engagement, and untimely updates. In order to further improve the effectiveness of eye-care apps, cooperation with medical institutions and professional ophthalmologists should be strengthened to enhance the scientific and authoritative nature of the apps. At the same time, interactive features and regular updates should be added to enhance user participation and the continuity of the apps. This study provides a reference for future development or improvement of eye-care apps, which can help promote myopia prevention and control.

(JMIR Mhealth Uhealth 2024;12:e53805) doi:10.2196/53805

KEYWORDS

Mobile apps; Eye care; Myopia; MARS; Children; Adolescent; Quality Assessment; Content Analysis; mhealth; China

Introduction

Children and adolescents are currently facing complex and serious health challenges, such as obesity, myopia, and anxiety and depression [1,2]. Children and adolescents represent the future of society; thus, more attention should be paid to their growth and development. Nearsightedness has been recognized as the most common eye disorder among children and adolescents [3]. Patients with shortsightedness experience blurred vision at a distance. High myopia is also associated with pathologic myopia, which can lead to complications such as retinal tear and detachment, macular degeneration, and irreversible vision damage [4]. Meanwhile, there is a link between shortsightedness, astigmatism, anisometropia, and amblyopia, which may interact with each other and work together to affect the visual health of children and adolescents [5,6].

A government statistical report in 2021 stated that the overall incidence of nearsightedness among Chinese children and adolescents was 52.7% and that the number of patients was about 100 million; shortsightedness is highly prevalent among children and adolescents and shows a trend of decreasing age of incidence [7]. However, prevention and control of myopia are hindered by insufficient medical resources, regional limitations, insufficient existing knowledge of healthy eye use, and low awareness about myopia [8-10]. Nevertheless, mobile health (mHealth) apps bring a new possibility to solve the aforementioned problems [10,11].

With their robust functionality, mobile apps have become a valuable tool for improving health care delivery and providing low-cost interventions [12,13]. In the domains of hypertension [14], nutrition and diet [15], exercise and medication adherence [16,17], chronic disease management [18,19], and so on, the significant effects of the use of apps on clinical, knowledge, behavioral, and psychosocial outcome improvement have been demonstrated [20,21]. Mobile apps have not only been recognized as an effective means to raise awareness of eye health [22,23], but have also efficiently provided myopia screening services and vision monitoring services [24,25].

Further integration of mobile apps with the internet has also significantly improved mHealth app accessibility [26]. Meanwhile, the number and types of apps have rapidly increased. The use of mobile apps is not limited to a specific time and place [27], which can help effectively overcome barriers such as insufficient health care resources and regional restrictions. In particular, eye-care app enables children and adolescents with no access to specialized eye education due to geographic location and economic status to receive high-quality eye-care services. Eye-care apps have attracted increasing attention with the discovery of the many advantages of mobile apps and the growing demand for eye-care services [28].

Nowadays, studies on the content and quality of eye-care apps are scarce [29]. In a study of eye-care–related apps in Chinese app stores, the researchers assessed the quality of the app through its general characteristics [30]. However, the professional background and built-in functionality of the app are also significant. The built-in functions of an app are an important reference for people to choose them, and the presence or absence of a professional background affects the user's trust on the app [31-33]. Due to the immature health app market in China [34], it becomes particularly important to conduct assessment studies on the quality of eye-care apps to facilitate the development of better strategies for eye-care services [35,36].

Therefore, this study aimed to summarize the general characteristics of eye-care apps for children and adolescents in Chinese app stores and to assess the quality of these apps. This will help users identify high-quality eye-care apps and provide a realistic basis for app development and improvement.

Methods

Systematic Search Strategy

In this study, a systematic search was performed on Chinese app stores to identify eye-care apps applicable to children and adolescents on February 15, 2023. At present, the sources of app stores are mainly categorized into third party companies and mobile phone manufacturers. According to the comprehensive ranking of Chinese app stores [37], Huawei AppGallery is the largest Android app store, and My app is the largest third-party app store in China. Therefore, Apple App Store (iOS) as well as Tencent My App and Huawei AppGallery (Android) were searched.

Through literature search and group discussion, the following keywords were used: eye care, myopia, eye health, vision, adolescent eye care, eye protection, and childhood myopia. Combining the age definition of child and adolescent as well as the age-applicable interpretation of app stores, we have added review criteria for targeted apps. The aforementioned keywords were searched in the app stores using an account that was not logged into any user, and all search results were recorded. If an app existed on both iOS and Android with identical design and content, considering the dominance of Apple App Store in app development [38], we evaluated the iOS version of the app.

Eligibility Criteria

After eliminating duplicate apps from the recorded search results, 2 researchers performed 2 rounds of app screening based on software description and usage. In the first round of screening, apps that satisfied all of the following conditions were included (Textbox 1):

Textbox 1.

Inclusion criteria

- The topic and content of the apps are related to eye care and the apps provide eye-care services
- The target audience of the apps is people needing eye-care services
- The target users of the apps are people aged 18 years or younger, and
- The apps are available in Chinese version

Exclusion criteria

- The apps focus on content unrelated to eye-care services, such as games, timers, and brightness adjustment on cell phones that do not provide eye care
- The apps are targeted only at eye-care service providers, such as doctors, nurses, research scholars, and other health professionals
- The apps do not have a Chinese version

In the second round of screening, apps were excluded if they were not successfully downloaded (after 2 attempts) or could not be logged in for use due to technical errors or special authentication.

Information Extraction

Information of the apps was extracted by 2 researchers, which included platform, developer, applicable age, last update time, star rating, number of downloads, and professional development background (Table 1). The general characteristics of the apps were described based on the information of the apps was extracted. Furthermore, through literature review and group discussion, the main functions of the eye-care apps were categorized into self-testing, eye exercises, monitoring of eye use data, eye-care education, patient record management, physical eye protection, medical consultation, online community, and eye-care mall. All disagreements were resolved through discussion between the researchers.

Table 1. General information collected for each app.

Assessment measure	Definition and values
Platform	Huawei (AppGallery), Apple (Apple App Store), Tencent (My App).
Developer	Commercial organizations, professional ophthalmology organization, individual developer.
Applicable age	Children (12 years or younger), adolescents (13-17 years).
Last update time	The latest update time when retrieving.
Star rating	Star rating (out of 5) left by users in the app store.
Downloads	Number of app downloads in the app store.
Evidence-based and professional background	Does the app claim that the design is based on relevant eye care theory, academic research, or the opinions of health professionals?

App Quality Assessment

To assess the quality of the apps more objectively, the Mobile App Rating Scale (MARS), a reliable tool for classifying and assessing the quality of mHealth apps [39], was used. MARS has been used to assess the quality of different apps, such as mental health [27], nutritional and diet-related [40,41], and chronic disease management [42,43] apps. It consists of 4 objective quality subscales of engagement, functionality, aesthetics, and information quality and 1 subjective quality subscale. As this study aimed to objectively assess the quality of apps, the subjective quality subscales were discarded. Before the formal scoring, 2 researchers randomly selected 5 apps to determine the MARS scoring rules so as to ensure that their understanding of the MARS questions and scoring criteria was consistent. To fully experience the services offered by the apps, the 2 researchers downloaded and used each app for at least 20 minutes during the formal scoring.

Data Analysis

To ensure the quality assessment reliability of the 2 researchers, within-group correlation coefficients (2-way random, average measure and absolute agreement) were used to assess researchers' agreement at the subscale and overall score level [44]. Pearson's correlation coefficients were used to explore the correlation among the overall MARS score, subscale scores, and user ratings. All statistical analyses were conducted in SPSS (version 26, IBM Corp).

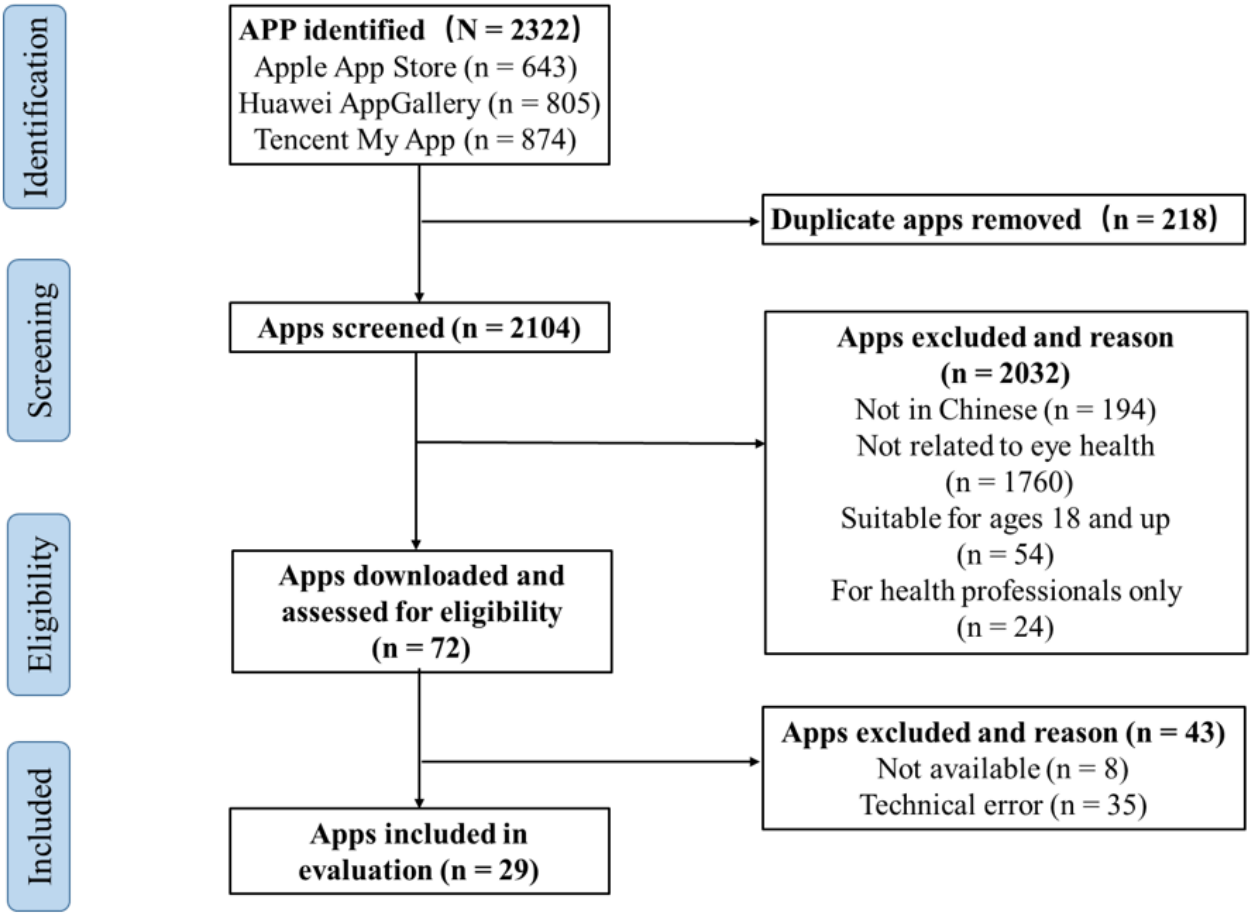
Results

App Selection

Through keyword search, 2322 apps were identified. Tencent My Apps had the highest number of apps (n=874, 37.6%), followed by Huawei AppGallery (n=805, 34.7%) and then Apple App Store (n=643, 27.7%). First, the search results of the apps were consolidated, and 218 apps were excluded due to repeated occurrence. Subsequently, 2032 (87.5%) apps were also

excluded based on the first round of screening criteria. The second round of operability screening was performed on the remaining 72 apps. Of the apps excluded in this round, 35 (48.7%) could not be downloaded or properly used due to technical reasons and 8 (11.1%) required special authentication to log in for use. Finally, after 2 rounds of screening, 29 (40.2%) apps were included in the study (Figure 1).

Figure 1. Flowchart for the systematic search and selection of apps.



General Characteristics of the App

The sources of the included apps were counted, and it was found that 18 (62%) apps were from Apple App Store, 8 (28%) from Tencent My Apps, and 3 (10%) from the Huawei AppGallery. Furthermore, more than half of the apps (17/29, 58%) were developed by commercial organizations, 8 (28%) by individual developers, and 4 (13%) by professional ophthalmology organizations. Simultaneously, based on the update time of the apps, 3 (10%) apps were updated within 1 month; 16 (55%), more than 1 month but less than 1 year; and 10 (34%), before 1 year (Table 2).

Table 2. Descriptive information on the apps reviewed.

Assessment measure	Apps, n (%)
Platform	
Apple App Store	18 (62)
Huawei AppGallery	3 (10)
Tencent My App	8 (28)
Developer	
Commercial organizations	17 (59)
Professional ophthalmology organizations	4 (13)
Individual developer	8 (28)
Applicable age	
Children	23 (80)
Adolescents	29 (100)
Last updated	
<1 month	3 (10)
>1 month and <1 year	16 (55)
>1 year	10 (35)
Evidence-based and professional background	
Participation of professional ophthalmologists	8 (28)
Verified methods and theories	3 (10)
Peer-reviewed academic research	1 (3)
Not mentioned	17 (59)

Evidence-Based and Professional Background

The details of the apps were checked, which indicated that 17 (59%) of them did not have an app design with a relevant scientific basis. As for the remaining 12 (41%) apps, their details indicated that their design was based on relevant scientific foundation: 8 (28%), designed with the involvement of medical professionals (eg, clinicians, ophthalmologists); 3 (10%), designed using a validated methodology such as ocular muscle training; and 1, designed based on scientific basis coming from peer-reviewed academic research.

Functional Review

The results of the statistical analysis indicated that eye-care apps provide a wide range of eye-care services for children and adolescents. More than half of the apps offered functions including self-testing (18/29, 62%), eye exercises (16/29, 55%), and eye-care education (16/29, 55%). About a quarter of the apps offered functions including patient record management (10/29, 34%), online community (7/29, 24%), and monitoring of eye use data (7/29, 24%). Furthermore, less than a quarter of the apps offered functions such as eye-care mall (6/29, 21%), medical consultation (6/29, 21%), and physical eye protection (5/29, 17%; Multimedia Appendix 1). The most common combination among multifunctional apps was that of self-testing and eye-care education (12/29, 41%), followed by a combination of self-testing and eye exercises (7/29, 24%).

Self-Testing

The self-testing function is mainly used to detect the user's vision level and other eye disorders. The vision testing function (18/18, 100%) was developed based on the E Vision Scale or C Vision Scale. The users participated in the test through manual selection as well as gesture and voice recognition. Meanwhile, the eye disease screening function (10/18, 56%) includes screening for color blindness, color deficiency, astigmatism, macular degeneration, and other diseases. Regular visual acuity screening of children and adolescents provides timely information on the current status of their vision and enables detection of eye disorders such as astigmatism, macular degeneration, and refractive errors.

Eye Exercises

A total of 16 apps offered the eye exercise function. The services included in this function were eye relaxation exercises, eye focus training, and amblyopia correction training. Eye exercise function is presented in the form of animations (12/16, 75%), games (6/16, 37.5%), and eye exercises (5/16, 31%). As one of the most important functions in eye-care apps, whether or not the eye exercise function has a scientific basis affects users' trust and willingness to use it. After reading the introduction of the apps offering eye exercises, only 6 apps were found to have functions based on scientific evidence, such as eye muscle training and the theory of eye exercises. Furthermore, details of the "FireEye" app clearly state that its function is developed

based on the results of big data analysis and professional optometry academic research.

Eye-Care Education

More than half of the apps (16/29, 55%) offer eye-care education function. The types of eye-care information provided include eye-care science articles (16/16, 100%), eye-care news (4/16, 25%), and eye-care instructional videos (7/16, 44%). Articles on eye-care science had the highest readership and attention. These articles focused on how to use eyes healthily and were presented in a combination of text and animation. In addition, based on the quality assessment results of the apps, eye-care education in the form of videos was found to be more attractive owing to its vivid sound and picture effects than eye-care education in the form of text.

Patient Record Management

The patient record management function is mainly used to record the vision conditions of children and adolescents at different periods. The eye diseases that can be recorded by this function include nearsightedness, hyperopia, strabismus, and amblyopia. A total of 10 apps offered the vision record management function. Further analysis of this function revealed that 6 apps not only recorded basic information, such as name, age, and recording time, but also provided patients with treatment reports or training plans based on the results of the vision records.

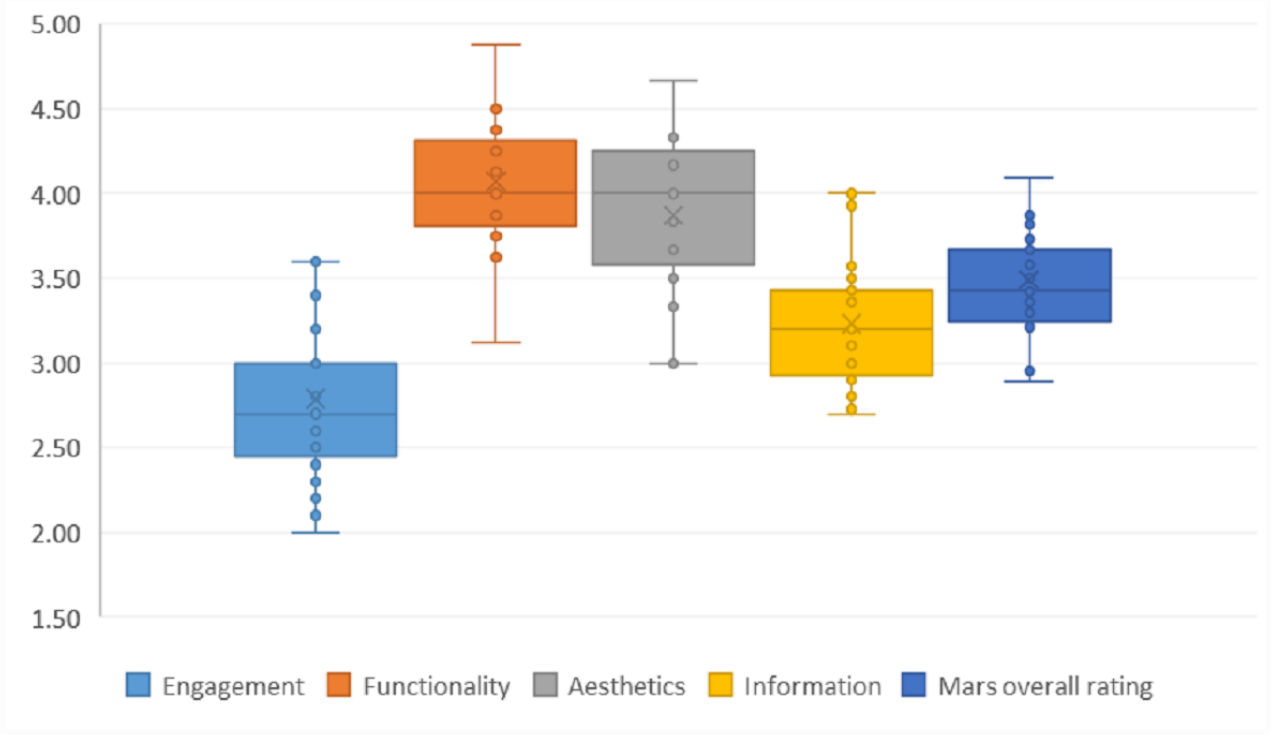
Online Community

The online community function embodies the great advantages of eye-care apps: proximity, interactivity, and convenience. The online community function exists in 7 (24%) apps, providing a social platform for eye care. In this community, doctors, patients, and others can communicate with each other through text and images. This function can effectively solve problems on consultation fees, long waiting time, and communication barriers that may exist in the actual medical service area. When users and doctors interact with each other through the platform, the distance between them is significantly shortened. Furthermore, patients suffering from the same disorder can come together to discuss the symptoms and treatment of the disease through this platform [45].

Mobile App Rating Scale Evaluation

The overall MARS scores exhibited high inter-reviewer reliability (intraclass correlation coefficient [ICC] 0.95, 95% CI 0.90-0.98). Meanwhile, the 4 subscales demonstrated good consistency: engagement ICC of 0.91 (95% CI 0.82-0.96), functionality ICC of 0.81 (95% CI 0.60-0.91), aesthetics ICC of 0.94 (95% CI 0.87-0.97), and information ICC of 0.90 (95% CI 0.79-0.95). The overall average MARS score of the eye-care apps was 3.49 (SD 0.33). The span of score between the highest scoring app (“Plano Eyes,” 4.39) and the lowest scoring app (“Eye 3D”, 2.89) was 1.5 (on 5-point Likert scale). The average scores for each subscale were as follows: engagement score, 2.79 (SD 0.47); functional score, 4.07 (SD 0.36); aesthetic score 3.87, (SD 0.43); and information score, 3.23 (SD 0.36). The engagement subscale had the largest span of scores (2.0-4.0). In terms of the app score distribution, 2 apps had MARS scores above 4 and 2 apps with scores below 3. Figure 2 presents the distribution of the overall quality of the apps and the scores on the 4D subscales.

Figure 2. Graphical representation of the distribution of the Mobile App Rating Scale overall and subscale score (N=29).



Correlation of Mobile App Rating Scale Scores and User Ratings

Further analysis of the correlation among the MARS overall score, subscale scores, and user ratings on the app store revealed that the MARS overall and subscale scores were positively correlated (r=0.81-0.91, P<.001). Table 3 summarizes the results of the correlation analysis.

By combining the professional background of the apps and update time, apps with an update time of less than 1 month (3.66) or with a medical professional development background (3.95) were found to have an overall average MARS score higher than the overall average MARS score (3.49). Meanwhile, the overall average MARS score for apps that were updated for more than 1 year (3.33) or did not have a background in medical specialty development (3.39) was lower than the overall average MARS score.

Table 3. Correlation between the Mobile App Rating Scale subscale and the overall score and the user star score.

Characteristic	Engagement, correlation	Functionality, correlation	Aesthetics, correlation	Information, correlation	Overall rating. correlation
Engagement	—[a]	—[a]	—[a]	—[a]	—[a]
Functionality	0.40	—[a]	—[a]	—[a]	—[a]
Aesthetics	0.46[b]	0.54[b]	—[a]	—[a]	—[a]
Information	0.73[c]	0.60[c]	0.66[c]	—[a]	—[a]
Overall rating	0.81[c]	0.75[c]	0.82[c]	0.91[c]	—[a]
User star rating[d]	0.33	0.34	0.02	0.33	0.31

[a]Not applicable.

[b]P≤.01.

[c]P≤.001.

[d]Apps with zero user star ratings were excluded.

Discussion

Principal Findings

This study discussed the general characteristics, main functions, and professional context of eye-care apps and also assessed the quality of these apps.

The main functions of the eye-care app included 9 categories: self-testing, eye exercises, monitoring of eye use data, eye-care education, patient record management, physical eye protection, medical consultation, online community, and eye-care mall. In terms of the number of built-in functions in apps, the average number of built-in functions was 3.1. In addition, the app with the most built-in functions contains 6 of the 9 categories of functions. The most common combination of functions was that of vision tests and screenings and eye-care information. The distribution of features in eye-care apps was broadly similar compared with previous studies, with an increase of 0.8 in the average number of built-in features in apps and an increase in the frequency of online community features [30,46,47]. There is a trend toward multifunctionality combination in the development of eye-care apps. Most eye-care apps in Chinese app stores include multiple eye-care–need services in one app. However, in the global app stores, these apps are usually considered to be designed for specific diseases or ophthalmic treatments, such as myopia surgery and cataract surgery [8,48].

In our count of app development backgrounds, only 1 app (FireEye) explicitly stated that its usability had been confirmed by peer-reviewed academic research. However, apps' lack of medical context can affect users' trust in them. Summary studies of apps for mental health, chronic disease management, exercise, and medication adherence found that the lack of scientific validation of apps was a recurring issue [14,18,19,22,23]. Of the 29 app developer statistics, 17 (59%) apps were developed by commercial organizations; 4 (13%), by professional ophthalmology organizations; and 8 (28%), by individual developers. These statistics indicated that the lack of involvement of medical professionals or scientific theories in the development of apps is a significant issue. Thus, the involvement of medical professionals in app design and function development is an important means to improve the professionalism of apps [15,40,49].

By counting the last update time of the apps, it was found that only 3 (10%) had been updated within 1 month and 10 (34%) for more than 1 year. Apps that are not updated for a long time may exhibit system vulnerabilities that are more easily exploited by lawbreakers, which in turn threatens the security of users' personal information [28,50]. By updating apps in a timely manner, developers can not only ensure the security of users' information but also increase the attractiveness of the apps to users by adding function modules to the apps [27,51].

Through an in-depth exploration of the functions of eye-care apps, one of the apps with a rating of more than 4 (Eye Nurse) has developed intelligence-guided diagnosis function based on artificial intelligence (AI). In this function, users can initially diagnose their disease by selecting age, gender, relevant symptoms, and disease duration. Simultaneously, the AI customer service generates a self-diagnostic report that includes a review of the condition, possible diseases, and doctor's recommendations. Initial diagnosis of patients using the intelligence-guided diagnosis function can effectively alleviate

the problem of insufficient health care resources. Existing research also confirms the feasibility of AI application to the diagnosis and treatment of relevant diseases [52-54]. The overall MARS scores for these apps ranged from 2.89 to 4.39, suggesting that the quality gap for apps was within the normal range [55]. However, these apps had lower mean scores on the engagement and information subscales (2.79, 3.23). This indicates that developers should focus on the improvement of user engagement and information quality during app development. In terms of how to improve user engagement, one of the apps with an engagement score of 4 (Plano Eyes) provided the following strategies: on the one hand, the app designed a punch-card module with daily life tasks, such as outdoor exercise, eye-care habits, and electronic use, which enhances the interaction between the app and the user. On the other hand, the app used audiovisual animations suitable for children and adolescents in eye exercises and eye-care education functions and constructed the eye exercise function in the form of a minigame. Related studies have also demonstrated that the form in which the functions of the apps are presented is an important factor influencing user engagement [56].

The mean MARS score suggests that the overall quality of eye-care apps for children and adolescents in Chinese app stores is acceptable. Meanwhile, by grouping and comparing the 2 factors of update time and presence or absence of a professional background, these 2 factors were found to have a potential association with the quality of the app. Furthermore, there may not be a correlation between user star ratings and MARS, which is consistent with the findings of quality correlation studies with mental health, family assessment, and breast cancer management apps [27,57,58], but food allergy app studies have found a strong correlation [8]. Perhaps the star ratings in the app store came from earlier versions of the app or may have been influenced by other factors and therefore do not fully represent the quality of the app [59]. Previous studies have also found that app quality depends not only on functionality but also on the content and design approach [60].

Contribution

In this study, a content analysis of eye-care apps for children and adolescents in Chinese app stores was conducted and the quality of these apps was assessed using a scale. Through the results of the study, deficiencies in eye-care apps at this stage, such as insufficient professional medical background, untimely updates, and low user engagement, were identified. At the same time, reference suggestions for the upgrading of eye-care apps were provided. In addition, we found that incorporation of AI technology may lead to technological breakthroughs in preliminary diagnosis.

Limitations and Future Work

This study has some limitations. We cannot guarantee that all apps are included or that all services (eg, connecting with a certified professional) are provided as requested. Furthermore, the app stores are constantly changing, and apps are on and off the shelves of the app store at all times. In addition, our research was conducted by only 2 researchers; thus, we were unable to advise on app development from the perspective of the target users. Therefore, as a next step, we intend to approach the subjective perspective of users requiring eye-care services to obtain the most realistic views of actual users of eye-care apps. Certainly, evaluation of the efficacy of a certain function in the app, such as relieving eyestrain and self-testing, would also be a good research direction.

Conclusions

This study analyzed the content and quality of 29 apps in Chinese app stores that are suitable for eye care for children and adolescents. The results of the study indicate that the quality of these apps is acceptable. However, the lack of medical background in apps affects user trust, leads to low app engagement, and reduces user dependence. Furthermore, apps that are not updated for a long time can threaten the security of users' information. Meanwhile, combining eye-care apps with AI technology may bring about dramatic changes in the popularity and quality of eye-care services in the foreseeable future. These findings provide a realistic basis for the improvement and development of eye-care apps in the future.

Acknowledgments

This research was supported by the Intelligent Medicine Research Project of Chongqing Medical University (ZHYX202105).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Functional review results of the eye care apps.
[PNG File , 86 KB - mhealth_v12i1e53805_app1.png]

References

1. Shetty AK. Global maternal, newborn, and child health: successes, challenges, and opportunities. Pediatr Clin North Am 2016;63(1):1-18. [doi: 10.1016/j.pcl.2015.08.001] [Medline: 26613686]
2. Costello A, Naimy Z. Maternal, newborn, child and adolescent health: challenges for the next decade. Int Health 2019;11(5):349-352. [doi: 10.1093/inthealth/ihz051] [Medline: 31529111]

3. Bremond-Gignac D. Myopia in children. Med Sci (Paris) 2020;36(8-9):763-768 [FREE Full text] [doi: 10.1051/medsci/2020131] [Medline: 32821053]
4. Baird PN, Saw S, Lanca C, Guggenheim JA, Smith Iii EL, Zhou X, et al. Myopia. Nat Rev Dis Primers 2020;6(1):99. [doi: 10.1038/s41572-020-00231-4] [Medline: 33328468]
5. Bastias GM, Villena MR, Dunstan EJ, Zanolli SM. Myopia and myopic astigmatism in school-children. Andes Pediatr 2021;92(6):896-903. [doi: 10.32641/andespediatr.v92i6.3527] [Medline: 35506801]
6. Gong W, Zhu Z, Bulloch G, Wang J, Chen J, Du L, et al. Anisometropia and its association with refraction development in highly myopic children. Clin Exp Optom 2024;107(1):58-65. [doi: 10.1080/08164622.2023.2198635] [Medline: 37078165]
7. The overall myopia rate of children and adolescents in China will be 52.7% in 2020, and the problem of myopia at a younger age remains prominent.: The Central People's Government of the People's Republic of China; 2020. URL: https://www.gov.cn/xinwen/2021-07/13/content_5624709.htm [accessed 2022-08-16]
8. Mandracchia F, Llauradó E, Tarro L, Valls RM, Solà R. Mobile phone apps for food allergies or intolerances in app stores: systematic search and quality assessment using the mobile app rating scale (MARS). JMIR Mhealth Uhealth 2020;8(9):e18339 [FREE Full text] [doi: 10.2196/18339] [Medline: 32936078]
9. Capó H, Briceño CA. Impact of health literacy on eye health disparities. Ophthalmology 2022;129(10):1083-1084. [doi: 10.1016/j.ophtha.2022.06.031] [Medline: 36058734]
10. Capó H, Edmond JC, Alabiad CR, Ross AG, Williams BK, Briceño CA. The importance of health literacy in addressing eye health and eye care disparities. Ophthalmology 2022;129(10):e137-e145. [doi: 10.1016/j.ophtha.2022.06.034] [Medline: 36058736]
11. Walline JJ. Myopia control: a review. Eye Contact Lens 2016;42(1):3-8. [doi: 10.1097/ICL.0000000000000207] [Medline: 26513719]
12. Free C, Phillips G, Watson L, Galli L, Felix L, Edwards P, et al. The effectiveness of mobile-health technologies to improve health care service delivery processes: a systematic review and meta-analysis. PLoS Med 2013;10(1):e1001363 [FREE Full text] [doi: 10.1371/journal.pmed.1001363] [Medline: 23458994]
13. Zilch A, Tschirsich M. Data protection and information security of digital health applications (DiGA). Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2021;64(10):1254-1261. [doi: 10.1007/s00103-021-03412-y] [Medline: 34524475]
14. Xu H, Long H. The effect of smartphone app-based interventions for patients with hypertension: systematic review and meta-analysis. JMIR Mhealth Uhealth 2020;8(10):e21759 [FREE Full text] [doi: 10.2196/21759] [Medline: 33074161]
15. Sandborg J, Söderström E, Henriksson P, Bendtsen M, Henström M, Leppänen MH, et al. Effectiveness of a smartphone app to promote healthy weight gain, diet, and physical activity during pregnancy (HealthyMoms): randomized controlled trial. JMIR Mhealth Uhealth 2021;9(3):e26091 [FREE Full text] [doi: 10.2196/26091] [Medline: 33704075]
16. Yerrakalva D, Yerrakalva D, Hajna S, Griffin S. Effects of mobile health app interventions on sedentary time, physical activity, and fitness in older adults: systematic review and meta-analysis. J Med Internet Res 2019;21(11):e14343 [FREE Full text] [doi: 10.2196/14343] [Medline: 31778121]
17. Al-Arkee S, Mason J, Lane DA, Fabritz L, Chua W, Haque MS, et al. Mobile apps to improve medication adherence in cardiovascular disease: systematic review and meta-analysis. J Med Internet Res 2021;23(5):e24190 [FREE Full text] [doi: 10.2196/24190] [Medline: 34032583]
18. Agarwal P, Mukerji G, Desveaux L, Ivers NM, Bhattacharyya O, Hensel JM, et al. Mobile app for improved self-management of type 2 diabetes: multicenter pragmatic randomized controlled trial. JMIR Mhealth Uhealth 2019;7(1):e10321 [FREE Full text] [doi: 10.2196/10321] [Medline: 30632972]
19. Wang Y, Min J, Khuri J, Xue H, Xie B, A Kaminsky L, et al. Effectiveness of mobile health interventions on diabetes and obesity treatment and management: systematic review of systematic reviews. JMIR Mhealth Uhealth 2020;8(4):e15400 [FREE Full text] [doi: 10.2196/15400] [Medline: 32343253]
20. Cotter AP, Durant N, Agne AA, Cherrington AL. Internet interventions to support lifestyle modification for diabetes management: a systematic review of the evidence. J Diabetes Complications 2014;28(2):243-251 [FREE Full text] [doi: 10.1016/j.jdiacomp.2013.07.003] [Medline: 24332469]
21. Liu K, Xie Z, Or CK. Effectiveness of mobile app-assisted self-care interventions for improving patient outcomes in type 2 diabetes and/or hypertension: systematic review and meta-analysis of randomized controlled trials. JMIR Mhealth Uhealth 2020;8(8):e15779 [FREE Full text] [doi: 10.2196/15779] [Medline: 32459654]
22. Sharma IP, Chaudhry M, Sharma D, Kaiti R. Mobile health intervention for promotion of eye health literacy. PLOS Glob Public Health 2021;1(10):e0000025 [FREE Full text] [doi: 10.1371/journal.pgph.0000025] [Medline: 36962088]
23. Narrillos-Moraza Á, Gómez-Martínez-Sagrera P, Amor-García M, Escudero-Vilaplana V, Collado-Borrell R, Villanueva-Bueno C, et al. Mobile apps for hematological conditions: review and content analysis using the mobile app rating scale. JMIR Mhealth Uhealth 2022;10(2):e32826 [FREE Full text] [doi: 10.2196/32826] [Medline: 35171109]
24. Ma S, Guan Y, Yuan Y, Tai Y, Wang T. A one-step, streamlined children's vision screening solution based on smartphone imaging for resource-limited areas: design and preliminary field evaluation. JMIR Mhealth Uhealth 2020;8(7):e18226 [FREE Full text] [doi: 10.2196/18226] [Medline: 32673243]

25. Luo G, Lee CY, Shivshanker P, Cheng W, Wang J, Marusic S, et al. Preliminary evaluation of a smartphone app for refractive error measurement. Transl Vis Sci Technol 2022;11(2):40 [FREE Full text] [doi: 10.1167/tvst.11.2.40] [Medline: 35703567]
26. Xiao P, Chen L, Dong X, Zhao Z, Yu J, Wang D, et al. Anxiety, depression, and satisfaction with life among college students in China: nine months after initiation of the outbreak of COVID-19. Front Psychiatry 2022;12:777190 [FREE Full text] [doi: 10.3389/fpsyt.2021.777190] [Medline: 35126198]
27. Wu X, Xu L, Li P, Tang T, Huang C. Multipurpose mobile apps for mental health in Chinese app stores: content analysis and quality evaluation. JMIR Mhealth Uhealth 2022;10(1):e34054 [FREE Full text] [doi: 10.2196/34054] [Medline: 34982717]
28. Grundy Q. A review of the quality and impact of mobile health apps. Annu Rev Public Health 2022;43:117-134 [FREE Full text] [doi: 10.1146/annurev-publhealth-052020-103738] [Medline: 34910582]
29. Suo L, Ke X, Zhang D, Qin X, Chen X, Hong Y, et al. Use of mobile apps for visual acuity assessment: systematic review and meta-analysis. JMIR Mhealth Uhealth 2022;10(2):e26275 [FREE Full text] [doi: 10.2196/26275] [Medline: 35156935]
30. Karthikeyan SK, Thangarajan R, Theruvedhi N, Srinivasan K. Android mobile applications in eye care. Oman J Ophthalmol 2019;12(2):73-77 [FREE Full text] [doi: 10.4103/ojo.OJO_226_2018] [Medline: 31198290]
31. Lagan S, Sandler L, Torous J. Evaluating evaluation frameworks: a scoping review of frameworks for assessing health apps. BMJ Open 2021;11(3):e047001 [FREE Full text] [doi: 10.1136/bmjopen-2020-047001] [Medline: 33741674]
32. Sekara V, Alessandretti L, Mones E, Jonsson H. Temporal and cultural limits of privacy in smartphone app usage. Sci Rep 2021;11(1):3861 [FREE Full text] [doi: 10.1038/s41598-021-82294-1] [Medline: 33594096]
33. Pepper C, Reyes-Cruz G, Pena AR, Dowthwaite L, Babbage CM, Wagner H, et al. Understanding trust and changes in use after a year with the NHS COVID-19 contact tracing app in the United Kingdom: longitudinal mixed methods study. J Med Internet Res 2022;24(10):e40558 [FREE Full text] [doi: 10.2196/40558] [Medline: 36112732]
34. Xian Y, Ye Y, Liu F, Zhou X, Zhao J. Agreement between a mobile applet-based visual acuity self-test program and the conventional method for distance and near visual acuity tests. Clin Exp Ophthalmol 2023;51(2):112-121. [doi: 10.1111/ceo.14194] [Medline: 36478363]
35. Ventola CL. Mobile devices and apps for health care professionals: uses and benefits. PT 2014;39(5):356-364 [FREE Full text] [Medline: 24883008]
36. Giunti G, Giunta DH, Guisado-Fernandez E, Bender JL, Fernandez-Luque L. A biopsy of breast cancer mobile applications: state of the practice review. Int J Med Inform 2018;110:1-9 [FREE Full text] [doi: 10.1016/j.ijmedinf.2017.10.022] [Medline: 29331247]
37. App store index.: AppInChina; 2022. URL: https://www.appinchina.co/market/app-stores/ [accessed 2022-08-16]
38. Sadare O, Melvin T, Harvey H, Vollebregt E, Gilbert S. Can Apple and Google continue as health app gatekeepers as well as distributors and developers? NPJ Digit Med 2023;6(1):8 [FREE Full text] [doi: 10.1038/s41746-023-00754-6] [Medline: 36721023]
39. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
40. Martinon P, Saliasi I, Bourgeois D, Smentek C, Dussart C, Fraticelli L, et al. Nutrition-related mobile apps in the French app stores: assessment of functionality and quality. JMIR Mhealth Uhealth 2022;10(3):e35879 [FREE Full text] [doi: 10.2196/35879] [Medline: 35285817]
41. McAleese D, Linardakis M, Papadaki A. Quality and presence of behaviour change techniques in mobile apps for the mediterranean diet: a content analysis of android google play and apple app store apps. Nutrients 2022;14(6):1290 [FREE Full text] [doi: 10.3390/nu14061290] [Medline: 35334947]
42. Jamaladin H, van de Belt TH, Luijpers LC, de Graaff FR, Bredie SJ, Roeleveld N, et al. Mobile apps for blood pressure monitoring: systematic search in app stores and content analysis. JMIR Mhealth Uhealth 2018;6(11):e187 [FREE Full text] [doi: 10.2196/mhealth.9888] [Medline: 30429116]
43. Selvaraj SN, Sriram A. The quality of Indian obesity-related mHealth apps: precede-proceed model-based content analysis. JMIR Mhealth Uhealth 2022;10(5):e15719 [FREE Full text] [doi: 10.2196/15719] [Medline: 35544318]
44. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016;15(2):155-163 [FREE Full text] [doi: 10.1016/j.jcm.2016.02.012] [Medline: 27330520]
45. Wu H, Lu N. Service provision, pricing, and patient satisfaction in online health communities. Int J Med Inform 2018;110:77-89. [doi: 10.1016/j.ijmedinf.2017.11.009] [Medline: 29331257]
46. Zhao L, Stinnett SS, Prakalapakorn SG. Visual acuity assessment and vision screening using a novel smartphone application. J Pediatr 2019;213:203-210.e1. [doi: 10.1016/j.jpeds.2019.06.021] [Medline: 31326117]
47. Aruljyothi L, Janakiraman A, Malligarjun B, Babu BM. Smartphone applications in ophthalmology: a quantitative analysis. Indian J Ophthalmol 2021;69(3):548-553 [FREE Full text] [doi: 10.4103/ijo.IJO_1480_20] [Medline: 33595469]
48. Lunde P, Nilsson BB, Bergland A, Kværner KJ, Bye A. The effectiveness of smartphone apps for lifestyle improvement in noncommunicable diseases: systematic review and meta-analyses. J Med Internet Res 2018;20(5):e162 [FREE Full text] [doi: 10.2196/jmir.9751] [Medline: 29728346]

49. Paganini S, Terhorst Y, Sander LB, Catic S, Balci S, Küchler AM, et al. Quality of physical activity apps: systematic search in app stores and content analysis. JMIR Mhealth Uhealth 2021;9(6):e22587 [FREE Full text] [doi: 10.2196/22587] [Medline: 34106073]
50. Sweeney M, Barton W, Nebeker C. Evaluating mobile apps targeting older adults: descriptive study. JMIR Form Res 2023;7:e37329 [FREE Full text] [doi: 10.2196/37329] [Medline: 37103995]
51. Varma M, Washington P, Chrisman B, Kline A, Leblanc E, Paskov K, et al. Identification of social engagement indicators associated with autism spectrum disorder using a game-based mobile app: comparative study of gaze fixation and visual scanning methods. J Med Internet Res 2022;24(2):e31830 [FREE Full text] [doi: 10.2196/31830] [Medline: 35166683]
52. Bakker D, Kazantzis N, Rickwood D, Rickard N. Mental health smartphone apps: review and evidence-based recommendations for future developments. JMIR Ment Health 2016;3(1):e7 [FREE Full text] [doi: 10.2196/mental.4984] [Medline: 26932350]
53. Tan TF, Thirunavukarasu AJ, Jin L, Lim J, Poh S, Teo ZL, et al. Artificial intelligence and digital health in global eye health: opportunities and challenges. Lancet Glob Health 2023;11(9):e1432-e1443 [FREE Full text] [doi: 10.1016/S2214-109X(23)00323-6] [Medline: 37591589]
54. Vilela MAP, Arrigo A, Parodi MB, da Silva Mengue C. Smartphone eye examination: artificial intelligence and telemedicine. Telemed J E Health 2024;30(2):341-353. [doi: 10.1089/tmj.2023.0041] [Medline: 37585566]
55. Song Y, Chen H. Evaluating Chinese mobile health apps for ankylosing spondylitis management: systematic app search. JMIR Mhealth Uhealth 2021;9(7):e27234 [FREE Full text] [doi: 10.2196/27234] [Medline: 34259644]
56. Zhai Y, Song X, Chen Y, Lu W. A study of mobile medical app user satisfaction incorporating theme analysis and review sentiment tendencies. Int J Environ Res Public Health 2022;19(12):7466 [FREE Full text] [doi: 10.3390/ijerph19127466] [Medline: 35742713]
57. Yang S, Bui CN, Park K. Mobile health apps for breast cancer: content analysis and quality assessment. JMIR Mhealth Uhealth 2023;11:e43522 [FREE Full text] [doi: 10.2196/43522] [Medline: 36821352]
58. Shin JH, Shields R, Lee J, Skrove Z, Tredinnick R, Ponto K, et al. Quality and accessibility of home assessment mhealth apps for community living: systematic review. JMIR Mhealth Uhealth 2024;12:e52996 [FREE Full text] [doi: 10.2196/52996] [Medline: 38466987]
59. Jahnsen R, Aamodt G, Rosenbaum P. Gross motor function classification system used in adults with cerebral palsy: agreement of self-reported versus professional rating. Dev Med Child Neurol 2006;48(9):734-738 [FREE Full text] [doi: 10.1017/S0012162206001575] [Medline: 16904019]
60. Wilks CR, Chu C, Sim D, Lovell J, Gutierrez P, Joiner T, et al. User engagement and usability of suicide prevention apps: systematic search in app stores and content analysis. JMIR Form Res 2021;5(7):e27018 [FREE Full text] [doi: 10.2196/27018] [Medline: 34259163]

Abbreviations

AI: artificial intelligence
ICC: intraclass correlation coefficient
MARS: Mobile Application Rating Scale
mHealth: mobile health

Edited by L Buis; submitted 19.10.23; peer-reviewed by TAR Sure, MR Katigari; comments to author 19.03.24; revised version received 11.04.24; accepted 08.08.24; published 13.09.24.

Please cite as:
Liu M, Wu X, Li Z, Tan D, Huang C
Assessment of Eye Care Apps for Children and Adolescents Based on the Mobile App Rating Scale: Content Analysis and Quality Assessment
JMIR Mhealth Uhealth 2024;12:e53805
URL: https://mhealth.jmir.org/2024/1/e53805
doi:10.2196/53805
PMID:

Original Paper

Self-Management Support Apps for Spinal Cord Injury: Results of a Systematic Search in App Stores and Mobile App Rating Scale Evaluation

Renaldo M Bernard[1†], PhD; Vanessa Seijas[1,2,3], MPH, MD; Micheal Davis[1,2], MSc; Anel Volkova[1,2], BSc; Nicola Diviani[1,2], PhD; Janina Lüscher[1,2], PhD; Carla Sabariego[1,2,3], PhD

[1]Swiss Paraplegic Research, Nottwil, Switzerland

[2]Faculty of Health Sciences and Medicine, University of Lucerne, Lucerne, Switzerland

[3]Center for Rehabilitation in Global Health Systems, University of Lucerne, Lucerne, Switzerland

[†]deceased

Corresponding Author:
Carla Sabariego, PhD
Swiss Paraplegic Research
Guido A Zäch-Strasse 4
Nottwil, 6207
Switzerland
Phone: 41 419396654
Email: carla.sabariego@paraplegie.ch

Abstract

Background: The use of mobile technology to meet health needs, widely referred to as mobile health (mHealth), has played a critical role in providing self-management support for chronic health conditions. However, despite its potential benefits, mHealth technologies such as self-management support apps for spinal cord injury (SCI) have received little research attention, and an understanding of their public availability is lacking. Therefore, an overview of these apps is needed to complement findings from the literature for a complete understanding of mHealth self-management support tools for SCI to support the selection and improvement of existing apps and the development of new ones.

Objective: This study aimed to identify and describe quantity, quality, focus, strengths, and weaknesses of self-management support apps for SCI available on major mobile app digital distribution platforms.

Methods: A systematic search of the Google Play Store and Apple App Store was conducted to identify and summarize apps for SCI that have been updated since 2017. A supplementary systematic literature review was conducted across 11 bibliographic databases to identify publications that provided more detailed descriptions of the identified apps than what is typically available in app stores. The data synthesis was guided by self-management tasks and skills taxonomies. The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines informed the reporting.

Results: The 13 apps included in the final synthesis were launched between 2013 and 2021, mostly originating in the United States, with availability in 72 countries and support for 14 languages. Most apps used the Android operating system (10/13, 77%), while 31% (4/13) used iOS. The identified apps mainly focused on activities of daily living, physical activity promotion, health literacy, and therapeutic exercise. All 3 self-management tasks (medical, role, and emotional management) and most self-management skills and support activities were supported by the apps. The mean Mobile App Rating Scale score was 3.86 (SD 0.54), indicating good overall quality. No publications were found describing these apps.

Conclusions: Despite their good overall quality, as measured by the Mobile App Rating Scale assessment, the 13 identified apps, alone or combined, do not appear to offer a comprehensive self-management approach that incorporates theory-based strategies. Besides working to improve comprehensiveness, future research and practice should consider adopting new technologies, such as artificial intelligence, to enhance future self-management support apps for SCI. Furthermore, adopting new app development methods, such as low-code development platforms, could help reduce barriers to development, such as time, cost, and securing scarce expertise.

(JMIR Mhealth Uhealth 2024;12:e53677) doi:10.2196/53677

KEYWORDS

mobile phone; mobile health; mHealth; eHealth; telemedicine; telehealth; spinal cord injury; self-management; internet-based intervention; world wide web; systematic review; review

Introduction

Background

Spinal cord injury (SCI) is a life-altering condition caused by traumatic or nontraumatic events, leading to temporary or permanent disruption of the normal motor, sensory, or autonomic functions of the spinal cord. It affects the physical, social, and psychological well-being of the patient and places an important burden on health care systems, families, and communities [1]. SCI is a chronic health condition that causes multiple limitations in daily functioning, such as limitations in walking, eating, grooming, working, or caring for oneself, and secondary health conditions, such as spasticity, constipation, urinary tract infections, chronic pain, sexual dysfunction, fatigue, and mental health disorders [2,3]. Therefore, people with SCI frequently rely on specialized and costly health care services, often facing significant unmet health care needs and challenges, particularly in low- and middle-income countries. These challenges are primarily due to the costs of health care services, transportation, and limited service availability [4]. Self-management support, such as the provision of social support and equipment [5], is essential for people with SCI to independently manage the challenging symptoms, treatment, and lifestyle changes associated with having a SCI [6-8].

Besides traditional self-management support, such as in-person counseling sessions, the adoption of mobile health (mHealth) self-management support is growing in the self-management of chronic health conditions such as SCI [9]. mHealth involves using mobile and wireless information and communication technologies, including smartphones, tablets, and wearables, to support meeting health needs [10]. Therefore, mHealth provides more person-centered, available, accessible, and scalable self-management support options than many traditional alternatives, such as institutional- and paper-based options [11]. Smartphones have one of the highest adoption rates among mHealth technology [12,13] and are essential tools for improving diagnostics, personal health monitoring and tracking, access to health care, and the development of health literacy [14].

However, mHealth self-management support tools for SCI have seemingly received little research attention despite the unique self-management support necessitated by the distinctive pathophysiology of SCI. To the best of our knowledge, the most relevant overviews of these tools are provided by systematic literature reviews by Wellbeloved-Stone et al [15] and Bernard et al [16]. Although insightful, the first study does not account for the expected rapid increase in the development of mHealth tools over the last 6 years and identifies 1 tool [15], and both studies are biased toward published literature as they potentially neglect apps that are available through digital distribution platforms and are not mentioned in published literature [15,16]. While mHealth self-management support apps for chronic health conditions aim to assist people in honing self-management skills, there exists a need for apps targeting SCI due to its distinct pathophysiology [17]. For example, self-management for bladder impairments mainly involves managing incontinence in prostate cancer, Parkinson disease, and stroke, compared to bladder voiding dysfunction, urinary tract infections, urinary stones, and renal impairment in SCI [18].

Therefore, identifying available self-management support apps for SCI is important for achieving a more complete overview of mHealth self-management support tools for SCI that complements related findings from the literature. This overview can also be used as a basis for the development of a convenient reference collection of self-management support apps for SCI to support the selection and improvement of existing apps and the development of new ones.

Objectives

The objective of this study was to identify and describe self-management support apps for SCI available on mobile app digital distribution platforms. This study aimed to (1) describe these apps in relation to their quantity and quality, coverage of self-management support activities and skills, and self-management areas targeted by self-management approaches; and (2) highlight associated strengths and weaknesses.

Methods

Overview

The study was conducted in 2 phases. In phase 1, a systematic search of app stores was conducted to identify and summarize self-management support apps for SCI. In phase 2, a systematic literature review was conducted to complement the work done in phase 1. This review aimed to identify publications that focused on designing, describing, evaluating, piloting, implementing, or improving apps previously identified in phase 1. The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [19] and their extension for literature searches [20] were used to guide reporting in both phases. There were deviations from the study protocol [21] due to unforeseen resource constraints. The study's aims were narrowed in scope. When available in both stores, apps were reviewed based on device availability rather than prioritizing apps from a specific store. The IMS Institute for Healthcare Informatics app functionality scale [22] and the behavior change technique taxonomy (version 1) [23] were not used for app evaluation and synthesis, respectively.

Search Strategy

Phase 1

Similar to recent systematic searches of app stores [24-26], the Google Play Store and Apple App Store were systematically searched via an application programming interface provided by 42Matters [27] using keywords for SCI to find eligible apps across regional marketplaces (Multimedia Appendix 1). These digital distribution platforms host the largest number of publicly

available mobile apps globally for smartphones, tablets, and wearables [28]. Given the limited documentation for the application programming interface of 42Matters, researchers conducted preliminary searches and communicated with company representatives to enhance their understanding. To avoid missing relevant apps, the search was comprehensive, covering app titles, descriptions, and developer names. Additional apps were identified by looking at other apps by the same developers and through the "similar apps" and "you might also like" features on both app stores.

Phase 2

MEDLINE, CINAHL Complete, PsycINFO, IEEE Xplore Digital Library, ACM Digital Library, Embase, Scopus, Web of Science Core Collection, Academic Search Premier, LISTA, and Business Source Premier were searched to find eligible literature (Multimedia Appendix 1). The search terms used identified app titles and developer names from phase 1. The reference lists of the included articles were planned to be hand searched.

Eligibility Criteria

Phase 1

To be eligible for inclusion, apps should target individual users with SCI for use primarily outside a clinical setting; be supportive of self-management in SCI; be available to all users within the corresponding regional marketplace; be accessible via a smartphone, tablet, or wearable (eg, smartwatch); be available in English; and be last updated after July 2017 and available up to July 2022. The fully featured version of the app was preferred over the limited version when priced at ≤US $10. In addition, the smartphone version was chosen over the tablet version, and the tablet was favored over the wearable version. Apps were excluded if a technical malfunction prevented access or use after 2 attempts. These criteria were selected as practical choices based on available resources to ensure the inclusion of apps that are relevant to individuals with SCI, accessible to a wide user base, up-to-date, and user-friendly across various device types, all while considering cost and technical functionality.

Phase 2

Publications were eligible for inclusion if they described an identified app from phase 1. Publications available in English were considered. Primary research studies, books, and gray literature (eg, conference proceedings, company websites, and professional publications) were also considered. Broad eligibility criteria were used to maximize the possibility of finding additional information to describe apps identified in phase 1.

Eligibility Assessment

Three researchers (AV, MD, and RMB), having expertise in health science, psychology, and health technology, were involved in the screening process. They attended training sessions to ensure consistency in screening during phases 1 and 2, using a web-based spreadsheet and the web-based service Rayyan [29], without using its artificial intelligence features. Screeners completed a training set of 20% of assigned apps and publications. For phase 1, screeners (AV and RMB) were randomly assigned a screening set of app titles, descriptions, and screenshots, as well as of apps to download and screen their content. For phase 2, screeners (MD and RMB) were randomly assigned a screening set of publication titles, abstracts or summaries, and full texts. A second screening of at least 20% was conducted as part of the training set for apps and publications. Screening was independently conducted to reduce the risk of reviewer bias [30]. Conflicting screening decisions (ie, *include*, *maybe*, or *exclude*) were resolved collaboratively.

Quality Assessment

Two reviewers (AV and MD), both health scientists, independently assessed the identified apps using the 23-item Mobile App Rating Scale (MARS) [31]. The MARS is a widely adopted and reliable tool for classifying and assessing mHealth app quality using 5 subscales: engagement (5 items), functionality (4 items), aesthetics (3 items), information quality (7 items), and a subjective app quality score (4 items) [32]. The review excluded the subjective quality dimension to ensure objectivity in the quality assessment process. Inconsistent ratings were resolved collaboratively. Interrater reliability was calculated for total and dimensional scores using the intraclass correlation coefficient [33]. A quality assessment was not planned for publications, as phase 2 aimed to provide supplementary descriptive information and not quality judgments.

Data Extraction and Synthesis

The same researchers who were involved in the screening process completed data extraction. These researchers attended a training session to help ensure consistency and reliability in data extraction using a web-based data extraction form. The form was discussed and modified for increased clarity. One researcher extracted data from the identified apps, corresponding app store pages, and developer websites, and then, another researcher reviewed and verified the extracted data. The extracted data were collated and summarized by AV and RMB. Any discrepancies in data extraction among researchers were resolved collaboratively.

A widely accepted framework or clinical guideline regarding the key components of a SCI self-management intervention could not be found. Therefore, data extraction and synthesis were guided by the self-management approaches for people with chronic conditions by Barlow et al [6], the self-management skills taxonomy by Lorig and Holman [34] (detailed in Textbox 1), and the Practical Reviews in Self-Management Support taxonomy of self-management support activities [5] to help organize its findings.

Textbox 1. Self-management task and skill frameworks.

Self-management tasks [6]

- Medical management: making health-related appointments, following treatment plans, tracking symptoms, and taking medication as directed
- Role management: organizing and coordinating the various everyday roles and responsibilities related to work, family, community, and self-care and adapting these roles as needed
- Emotional management: regulating and coping with emotions resulting from living with a condition in a healthy and effective manner

Self-management skills [34]

- Problem-solving: identifying problems and finding, implementing, and evaluating solutions
- Decision-making: weighing options and choosing the best course of action in response to changes in their condition
- Resource use: finding and effectively using resources
- Forming patient-provider partnerships: learning from and partnering with health care professionals to understand the patterns experienced with a condition, make informed decisions, and discuss related issues
- Action planning: developing a realistic action plan that can be confidently used to achieve a set goal
- Self-tailoring: developing and implementing personalized self-management strategies as needed

Results

Overview

A total of 13 apps were included in the final synthesis, and no publications describing them were identified. Figures 1 and 2 detail the corresponding methodological processes (see Multimedia Appendix 2 for the PRISMA checklist).

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart of the app store search, selection, and inclusion process.

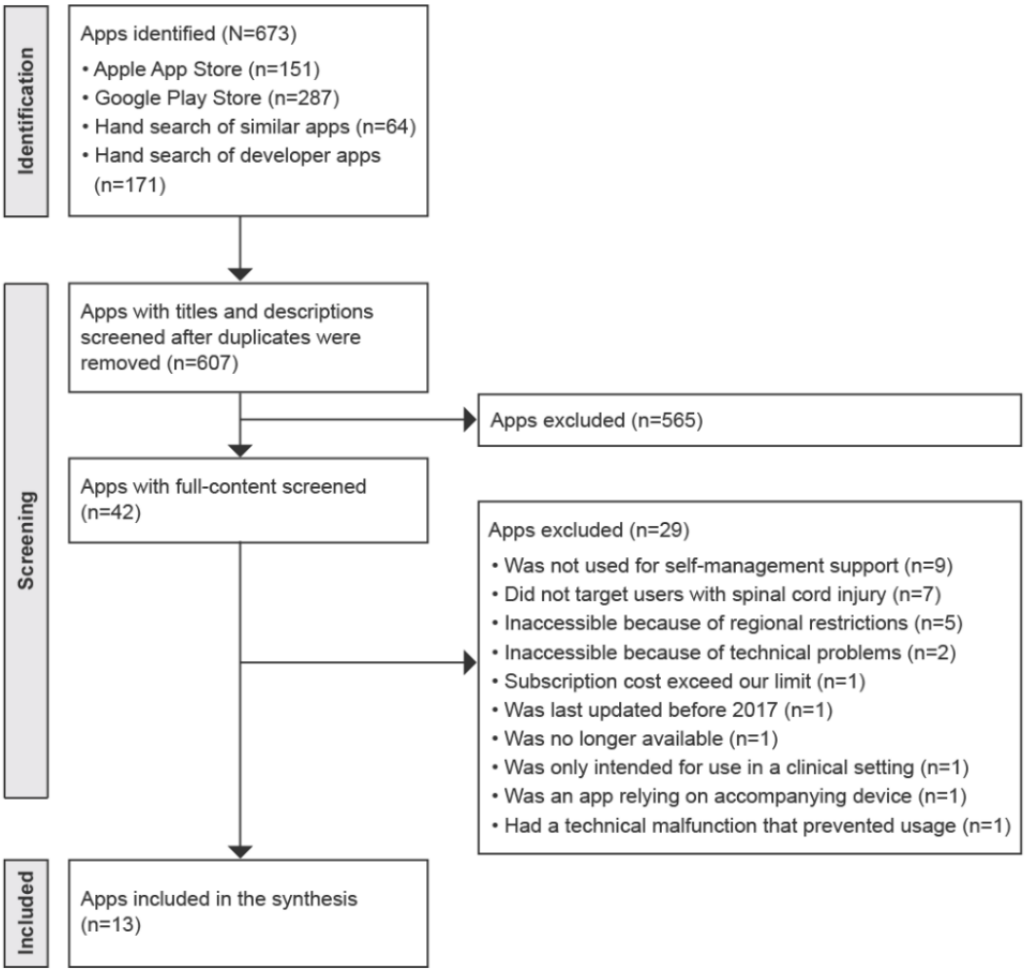
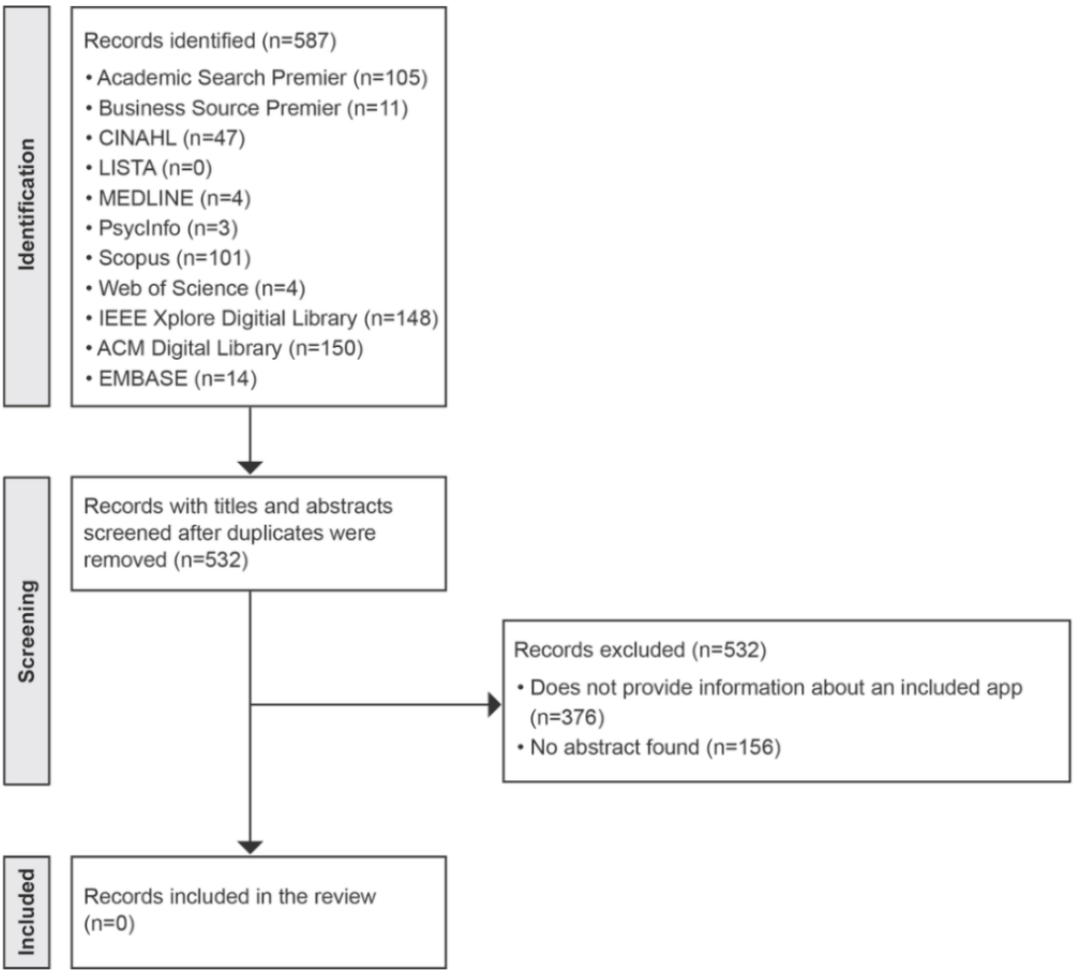


Figure 2. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart of the literature search, selection, and inclusion process.



App Characteristics

Overview

The 13 identified apps are summarized in Multimedia Appendix 3. They were launched between 2013 and 2021, with the majority (9/13, 69%) released from 2019 onward [35-43] (Table 1). These apps mainly originated from the United States (8/13, 62%) [35,38-41,44-46], followed by Australia (2/13, 15%) [37,42] and 1 (1/13, 8%) each from Belarus [47], France [36], and Italy [43]. They were mainly published under the Health category (8/13, 62%) [37-40,42,43,46,47], followed by 1 (1/13, 8%) app each under the Tools [35], News and magazines [44], Sports [45], Communication [36], and Lifestyle [41] categories across both app stores. The identified apps supported 14 languages, with English having full support due to the language eligibility requirement, followed by French (4/13, 31%); German and Italian (3/13, 23%); Spanish and Portuguese (2/13, 15%); and several other European languages, Turkish, and Japanese (1/13, 8%). The identified apps were available in 72 countries, covering all 6 major world regions and 16 (73%) of the 22 subregions globally [48]. They were unavailable in countries in central Asia, middle Africa, eastern Africa, Melanesia, Polynesia, and Micronesia. Where reported (8/13, 62%) [35-39,44,45,47], 15% (2/13) of the identified apps had between 500,001 and 1,000,000 [35,47], 1001 and 5000 [36,44], 101 and 1000 [37,45], and 6 and 50 [38,39] downloads each. The number of downloads for the 5 Apple App Store apps was not reported. Target user ages were not reported, but most apps had content deemed appropriate for all ages (9/13, 69%) [35-40,44,45,47] and a few for adults only (4/13, 31%) [41-43,46].

Table 1. Characteristics of the 13 identified self-management apps for spinal cord injury by frequency.

Characteristic	Apps, n (%)
Year of launch	
2020 [35-37,39]	4 (31)
2019 [38,40-42]	4 (31)
2013 [46,47]	2 (15)
2017 [44,45]	2 (15)
2021 [43]	1 (8)
Country of origin	
United States [35,38-41,44-46]	8 (62)
Australia [37,42]	2 (15)
Belarus [47]	1 (8)
France [36]	1 (8)
Italy [43]	1 (8)
Regional availability	
Australia and New Zealand [35-47]	13 (100)
Northern America [35-47]	13 (100)
Western Europe [35-47]	13 (100)
Northern Africa [35,36,38-47]	12 (92)
Northern Europe [35,36,38-47]	12 (92)
South America [35,36,38-47]	12 (92)
Central America [35,38-47]	11 (85)
Southern Africa [35,38-47]	11 (85)
Western Africa [35,38-47]	11 (85)
Eastern Asia [35,38-47]	11 (85)
Southeastern Asia [35,38-47]	11 (85)
Southern Asia [35,38-47]	11 (85)
Western Asia [35,38-47]	11 (85)
Eastern Europe [35,38-47]	11 (85)
Southern Europe [35,38-47]	11 (85)
Caribbean [40-42,46]	4 (31)
Supported languages	
English [35-47]	13 (100)
French [35,36,38,47]	4 (31)
Italian [35,43,47]	3 (23)
German [35,36,47]	3 (23)
Spanish [35,47]	2 (15)
Portuguese [35,47]	2 (15)
Czech [47]	1 (8)
Danish [47]	1 (8)
Greek [47]	1 (8)
Japanese [35]	1 (8)
Russian [47]	1 (8)
Slovak [47]	1 (8)

Characteristic	Apps, n (%)
Swedish [47]	1 (8)
Turkish [47]	1 (8)
Operating system	
Android [35-39,44,45,47]	8 (62)
iOS [40-43,46]	5 (38)
App store category	
Health and fitness [37-40,42,47]	6 (46)
Medical [43,46]	2 (15)
Tools [35]	1 (8)
News and magazines [44]	1 (8)
Sports [45]	1 (8)
Communication [36]	1 (8)
Lifestyle [41]	1 (8)
Year of last update	
2020 [39-41,44,45]	5 (38)
2022 [35-37,43,46]	5 (38)
2019 [38,42,47]	3 (23)
Cost to download (US $)	
0.00 [35,37,39,41-44,46,47]	6 (46)
≤5.00 [36,38,40]	3 (23)
≤10.00 [39,47]	2 (15)
≤15.00 [45]	1 (8)
≤20.00 [44]	1 (8)
In-app purchases	
No [35,37-43,47]	9 (69)
Yes [36,44-46]	4 (31)
Downloads	
NR[a] [40-43,46]	5 (38)
≥500,000 [35,47]	2 (15)
≥1000 [36,44]	2 (15)
≥100 [45]	1 (8)
≥500 [37]	1 (8)
≥10 [38]	1 (8)
≥5 [39]	1 (8)

[a]NR: not reported.

Platforms and Functionality

Of the 13 apps included in this study, 10 (77%) used the Android operating system [49-56] and 4 (31%) used iOS [57-59]. Of the 13 apps, 2 (15%) used the Android operating system only [49,54], 3 (23%) used iOS only [57-59], and 6 (46%) used both Android and iOS [50-53,55,56,60-65]. The earliest operating system versions supported by the identified apps were Android version 4 (5/13, 38%) [38,39,44,45,47] and iOS version 9 (1/13, 8%) [40]. The identified apps had a file size ranging from 3 MB to 117 MB, with the average size being 31 (SD 35.18) MB for Android and 47 (SD 28.15) MB for iOS. Of the identified apps, half (7/13, 54%) had a tablet version [35,37,39,40,44-46]. Most of the identified apps (9/13, 69%) were functional without an internet connection [35-40,42,43,47]. The identified apps had no discernible way for users to access their data and freely export their data. Of the 13 apps, 1 (8%) incorporated gamification [42], 2 (15%) included assistive videos [42,47], and 3 (23%) provided music or audio [39,42,47].

Very few of the included apps (3/13, 23%) had a wide range of features. Of the 13 apps, only 2 (15%) offered tracking functionality for physical activity [37,42], and 1 (8%) app had features to support mindfulness and deep breathing exercises as a form of sleep therapy [39]. None of the apps provided a platform for peer support, though 1 (8%) of the 13 apps did provide a connection to a support professional [37]. No apps provided features for psychoeducation, journaling, goal setting, habits, or chatbot interactions.

Privacy and Support

Most of the identified apps (8/13, 62%) provided accessible privacy policies [35-41,43]. The remaining apps either had broken web page links to the privacy policies [42,44-46] or provided none [47]. When provided, privacy policies were accessible from the app store and app (4/13, 31%) [37,39-41] and app store alone (8/13, 62%) [35-41,43]. No clarification on personal data processing and data security measures was provided. Of the 13 apps, 1 (8%) [37] was identified as sharing personal health information as it connected users to health care providers. No deidentified, anonymized, or aggregated data were seemingly shared by the identified apps, and they did not claim to meet any regulation targeting data privacy or security such as the Health Insurance Portability and Accountability Act of 1996 in the United States. Data collection was mandatory for the identified apps. However, it was difficult to distinguish between optional and mandatory data collection. As a result, this distinction was not included in the synthesis. See Table 2 for more information.

Table 2. User data collected by the 13 identified self-management apps for spinal cord injury by purpose.

Data collected	Purpose	Apps, n (%)
Network connections	Analytics [35] and unknown [36-39,44,45,47]	8 (62)
Email address	Analytics [35-37,39], app functionality [36-39,41], developer communications [36,37,39], security and compliance [36,37,39], personalization [36,39], and account management [36,39]	6 (46)
App interactions	Analytics [35,36], app functionality [36], developer communications [36], and unknown [38,41]	4 (31)
Contact list	App functionality [37,41] and personalization [37]	4 (31)
Device ID	Analytics [35,37,38] and security and compliance [36-38]	4 (31)
IP address	Analytics [37-40] and security and compliance [38]	4 (31)
Credit card information	App functionality [37] and unknown [38,41]	3 (23)
Location	App functionality [37], personalization [37], and unknown [39]	2 (15)
Log data	Analytics [36,37,40]	3 (23)
Name	App functionality [37,38,41] and personalization [37,38]	3 (23)
Phone number	Analytics [36], app functionality [36], developer communications [36,37], security and compliance [36,37,41], personalization [36,41], account management [36], and unknown [36]	3 (23)
Age	Analytics [37], app functionality [37,38], and personalization [37,38]	2 (15)
Crash logs	Analytics [35,36], app functionality [36], and developer communications [36]	2 (15)
Date of birth	Analytics [37,41], app functionality [37], and personalization [37,41]	2 (15)
Device type	Analytics [37,41] and security and compliance [37,41]	2 (15)
Diagnostics	Analytics [35,36], app functionality [36], and developer communications [36]	2 (15)
Postal address	Analytics [37], app functionality [37], developer communications [37], and personalization [41]	2 (15)
User ID	Analytics [35,36], app functionality [36], developer communications [36], security and compliance [36], personalization [36], and account management [36]	2 (15)
Gender	Analytics [41] and personalization [41]	1 (8)
In-app search history	Analytics [35]	1 (8)
Other user-generated content	Analytics [35] and app functionality [35]	1 (8)
Statistics on page views	Analytics [37]	1 (8)

Most apps provided support options (12/13, 92%) and used email (6/13, 46%) [35,36,39,41,42,47]; website-based information (6/13, 46%) [36,37,40,42,44,45]; tutorials, guides, or manuals (6/13, 46%) [36,37,40,44-46]; frequently asked

questions (3/13, 23%) [36,37,41]; phone (2/13, 15%) [35,42]; web chat (1/13, 8%) [35]; contact forms (1/13, 8%) [42]; and social media platforms (1/13, 8%) [37] for this purpose. Of the 13 apps, 4 (31%) provided 2 options [40,41,44,45], 3 (23%) provided 4 options [36,37,42] and 1 option [39,46,47], and 1 (8%) provided 3 options [35].

Cost

Of the 13 apps, 5 (38%) did not request payment [35,37,41-43] and the remaining 8 (62%) accepted payment [36,38-40,44-47] to download [36,38-40,44,45,47], while 4 (31%) accepted payment for one-time or recurring in-app purchases to access full features [36,44-46]. The cost to download ranged from US $1.99 to US $19.99.

Developers

Developers were mainly for-profit companies (10/11, 91%) [49,50,52-60,62-65] except for 1 (9%) nonprofit company [51,61]. Developers published 246 apps across both app stores [49-65], with most of them being published in Google Play Store (137/246, 55.7%) [49-56] and the remainder in the Apple App Store (109/246, 44.3%) [57-65] (Table 3). Half of the developers focused on publishing health-related apps (6/11, 55%) [49,53-55,57,59,63,64], and the others published apps across a wide range of categories in the app store [50-52,56,58,60-62,65]. The Paralyzed Veterans of America [51,61] published 3 (23%) of the 13 identified apps [44-46]; other developers (10/11, 91%) published 1 each [35-43,47].

Table 3. Characteristics of the developers of the identified self-management apps for spinal cord injury (N=11).

Developer name and citation	Country	Company type	App store categories with published apps
DSN Inc [49]	Belarus	For-profit company	Health and fitness and puzzle
Google LLC [50,60]	United States	For-profit company	Auto and vehicles, books and reference, business, communication, education, entertainment, finance, health and fitness, libraries and demo, lifestyle, medical, music and audio, news and magazines, personalization, photo and video, photography, productivity, simulation, social, tools, travel and local, utilities, utilities, video players and editors, and weather
Paralyzed Veterans of America [51,61]	United States	Nonprofit company	Business, medical, news and magazines, and sports
JIB Smart Home [52,62]	France	For-profit company	Communication, house and home, and style
Maslow For People and Ilya Thai [53,63]	Australia	For-profit company	Health and fitness
Kinnereth LLC App Dev [54]	United States	For-profit company	Health and fitness
Injectful LLC [55,64]	United States	For-profit company	Health and fitness
Cordilac LC [57]	United States	For-profit company	Health and fitness
iAccess Innovations [56,65]	United States	For-profit company	Lifestyle
Monster Hub [58]	Australia	For-profit company	Business, education, entertainment, games, health and fitness, lifestyle, productivity, shopping, social networking, sports, and travel
Giorgio Lofrese [59]	Italy	For-profit company	Medical

Characteristics of Approaches Providing mHealth Self-Management Support for SCI

All self-management tasks were supported by the identified apps (Table 4). Role management received the most support (11/13, 85%) [35-38,40-42,44-47], followed by medical (7/13, 54%) [37,38,40,42,43,45,47] and emotional (4/13, 31%) [39,42,44,45] management (Table 5). Support for a single task was most common (6/13, 46%) [35,36,39,41,43,46], followed by 2 (5/13, 38%) [37,38,40,44,47] and 3 (2/13, 15%) [42,45] tasks. When combined, role and medical management were most often supported together (4/13, 31%) [37,38,40,47], followed by role and emotional management (1/13, 8%) [44].

Table 4. Identified self-management apps for spinal cord injury by self-management approach (N=13).

App name and citation	Self-management focus area	Relevant self-management tasks	Relevant self-management skills	Relevant self-management support components
Pilates [47]	Physical activity promotion	Role management and medical management	Action planning	Training and rehearsal for practical self-management activities
Action Blocks [35]	Activities of daily living	Role management	Resource use	Training and rehearsal for practical self-management activities, training and rehearsal for everyday activities, and lifestyle advice and support
PN–Paraplegia News [44]	Health literacy and activities of daily living	Role management and emotional management	Problem-solving, action planning, and resource use	Information about the condition and or its management, information about available resources, and lifestyle advice and support
SNS Digital [45]	Physical activity promotion, health literacy, and activities of daily living	Role management, emotional management, and medical management	Problem-solving, action planning, and resource use	Information about available resources and lifestyle advice and support
JIB CALLS [36]	Activities of daily living	Role management	Resource use	Training and rehearsal for everyday activities
Disability Care App [37]	Therapeutic exercise and activities of daily living	Role management and medical management	Action planning, maintaining patient-provider partnership, and problem-solving	Practical support with adherence (medication or behavioral), training and rehearsal for practical self-management activities, and training and rehearsal to communicate with health care professionals
Dietitian's Tools [38]	Medicating and dieting	Role management and medical management	Action planning	Training and rehearsal for practical self-management activities and training and rehearsal to communicate with health care professionals
Injectful [39]	Pain management	Emotional management	Problem-solving	Training and rehearsal for psychological strategies
PVA ePubs [46]	Health literacy and activities of daily living	Role management	Resource use and action planning	Information about available resources and lifestyle advice and support
AccessiRep [40]	Physical activity promotion	Role management and medical management	Problem-solving and action planning	Training and rehearsal for practical self-management activities
iAccess Life–Accessibility [41]	Mobility	Role management	Resource use and action planning	Lifestyle advice and support, social support, and information about available resources
Neuro Therapy [42]	Therapeutic exercise and physical activity promotion	Role management, emotional management, and medical management	Action planning and maintaining patient-provider partnership	Training and rehearsal for practical self-management activities, practical support with adherence (medication or behavioral), provision of easy access to advice or support when needed, and social support

App name and citation	Self-management focus area	Relevant self-management tasks	Relevant self-management skills	Relevant self-management support components
Spine Fine [43]	Therapeutic exercise and physical activity promotion	Medical management	Action planning	Training and rehearsal for practical self-management activities, monitoring of the condition with feedback, and provision of easy access to advice or support when needed

Table 5. Characteristics of approaches to self-management support for spinal cord injury by frequency (N=13).

Characteristic	Apps, n (%)
Supported self-management tasks	
Role management [35-38,40-42,44-47]	11 (85)
Emotional management [39,42,44,45]	4 (31)
Medical management [37,38,40,42,43,45,47]	7 (54)
Supported self-management skills	
Action planning [37,38,40-47]	10 (77)
Resource use [35,36,41,44-46]	6 (46)
Problem-solving [37,39,40,44,45]	5 (38)
Maintaining patient-provider partnership [37,42]	2 (15)
Incorporated self-management support components	
Training and rehearsal for practical self-management activities [35,37,38,40,42,43,47]	7 (54)
Lifestyle advice and support [35,41,44-46]	5 (38)
Information about available resources [41,44-46]	4 (31)
Social support [41,42]	2 (15)
Practical support with adherence (medication or behavioral) [37,42]	2 (15)
Provision of easy access to advice or support when needed [42,43]	2 (15)
Training and rehearsal for everyday activities [35,36]	2 (15)
Training and rehearsal to communicate with health care professionals [37,38]	2 (15)
Monitoring of the condition with feedback [43]	1 (8)
Training and rehearsal for psychological strategies [39]	1 (8)
Targeted self-management focus areas	
Activities of daily living [35-37,44-46]	6 (46)
Physical activity promotion [40,42,43,45,47]	5 (38)
Health literacy [44-46]	3 (23)
Therapeutic exercise [37,42,43]	3 (23)
Mobility [41]	1 (8)
Medicating and dieting [38]	1 (8)
Pain management [39]	1 (8)

Most self-management skills were supported by the identified apps (4/6, 67%; Table 5). *Action planning* received support from most of the identified apps (10/13, 77%) [37,38,40-47], whereas *maintaining patient-provider partnership* received the least support (2/13, 15%) [37,42]. Support for a single skill was most common (6/13, 46%) [35,36,38,39,43,47], followed by 2 (4/13, 31%) [40-42,46] and 3 (3/13, 23%) [37,44,45] skills.

Incorporated self-management support components included 71% (10/14) of the Practical Reviews in Self-Management Support components (Table 5). The top 3 components were

incorporated more than the average number of times and accounted for 85% (11/13) of the identified apps. Information regarding the condition and its management, including equipment details, specific clinical action plans, rescue medications, and regular clinical reviews, was not included. Of the 13 apps, 6 (46%) incorporated 1 component [35,36,38,39,43,47], 4 (31%) incorporated 2 components [40-42,46], and 3 (23%) incorporated 3 components [37,44,45].

The adopted approaches to providing self-management support mainly focused on activities of daily living, physical activity promotion, health literacy, and therapeutic exercise (Table 5). Mobility, medicating and dieting, and pain management were targeted to a lesser extent.

MARS Evaluation

The mean MARS score for the 13 identified apps was 3.86 (SD 0.54), with a maximum of 4.52 for Action Blocks [35] and a minimum of 2.72 for AccessiRep [40], indicating good overall quality (Table 6). On average, the best-rated section was functionality (mean 4.56, SD 0.33), followed by information quality (mean 3.71, SD 0.78), aesthetics (mean 3.62, SD 0.91), engagement (mean 3.54, SD 0.61), app subjective quality (mean 3.39, SD 0.98), and app specific quality (mean 2.73, SD 0.97). The mean MARS score for the apps from Google Play and Apple App Store were 3.84 (SD 0.47; 8/13, 61%) and 3.88 (SD 0.71; 5/13, 39%), respectively. The interrater reliability as assessed by the intraclass correlation coefficient was 0.80 (95% CI 0.49-0.94), indicating a good level of agreement in the scoring between the 2 raters (AV and MD) [66].

Table 6. Mobile App Rating Scale scores (overall score and 4 subscales) of the 13 identified self-management apps for spinal cord injury.

App name and citation; developer and citation	Overall score	Engagement score	Functionality score	Aesthetics score	Information quality score
Action Blocks [35]; Google LLC [50,60]	4.52	4.40	4.75	4.50	4.42
PVA ePubs [46]; Paralyzed Veterans of America [51,61]	4.48	3.50	4.88	4.83	4.70
Neuro Therapy [42]; Monster Hub [58]	4.44	4.30	4.38	4.50	4.60
Disability Care App [37]; Maslow For People [53,63]	4.19	4.20	4.63	4.67	3.25
JIB CALLS [36]; JIB Smart Home [52,62]	3.98	3.80	4.88	4.00	3.25
Injectful [39]; Injectful LLC [55,64]	3.93	3.50	4.63	4.00	3.60
iAccess Life–Accessibility [41]; iAccess Innovations [56,65]	3.92	3.80	4.88	3.50	3.50
PN–Paraplegia News [44]; Paralyzed Veterans of America [51,61]	3.85	3.40	4.50	3.00	4.50
SNS Digital [45]; Paralyzed Veterans of America [51,61]	3.85	3.40	4.50	3.00	4.50
Spine Fine [43]; Giorgio Lofrese [59]	3.83	3.60	4.38	3.83	3.50
Dietitian's Tools [38]; Kinnereth LLC App Dev [54]	3.46	2.80	4.88	2.67	3.50
Pilates [47]; DSN Inc [49]	2.96	3.10	4.25	2.17	2.33
AccessiRep [40]; Cordilac LC [57]	2.72	2.20	3.75	2.33	2.58

Regarding *engagement*, most of the identified apps were deemed sufficiently customizable, interactive, and focused on the targeted audience [35-38,41,42,47], as indicated by the minimum acceptability score of 3.0 established by the MARS. Nonetheless, several apps were rated lower due to lacking sufficient settings for customization [39,40,43-47]. In terms of *functionality*, the identified apps were generally easy to use and had good navigation and gestural design. Similarly, the identified apps received high scores for aesthetics, as they featured appealing graphic designs, color schemes, and layouts that were consistent throughout. Concerning *information quality*, most apps had accurate descriptions and presented enough information of good quality. However, low or no scores were noted regarding the presence of specific, measurable, and achievable goals as well as an evidence base supported by scientific research or expert-reviewed guidelines for the app's content and functionality.

Discussion

Principal Findings and Comparison With Prior Work

This study identified 13 self-management support apps for SCI available on the Google Play Store and Apple App Store that meet this study's inclusion criteria. No scientific literature was found that focused on designing, describing, evaluating, piloting, implementing, or improving them. This emphasizes the critical

shortage of evidential support in literature for these apps and underscores the uncertainty around their quality and effectiveness. Consequently, it becomes challenging to confidently recommend any of them or identify which aspects require improvement. Unsurprisingly, none of the apps were identified by an earlier systematic literature review with similar inclusion criteria [67]. This highlights the need for greater collaboration between researchers and developers to improve the dissemination and availability of theory- and evidence-based self-management support apps for SCI.

Self-Management Support Apps for SCI

This study identified 11 self-management support apps launched on the Google Play Store and Apple App Store between 2017 and 2021, similar in number to the 11 apps identified by the earlier systematic literature review [67]. As all apps were unique, a total of 22 self-management support apps for SCI were available during this period. The United States continues to contribute most of the apps identified, which may result in a potential bias toward the cultural, social, and demographic context of the country and may not fully capture the diverse needs and preferences of users from other regions. Although apps from Canada, the Netherlands, Switzerland, and Thailand were discussed in the literature, no similar apps from these countries were found on app distribution platforms. According to this review, Australia, Belarus, and France were also represented for the first time. The number of downloads for Apple App Store apps was not reported, limiting our understanding of the popularity and reach of these apps. It was noted that most apps did not require an internet connection, which is beneficial for areas with poor or no internet connectivity.

mHealth Self-Management Support Approaches in SCI

This study identified self-management focus areas that align with those found in the earlier systematic literature review [67]. These included physical activity promotion, pain management, therapeutic exercise, medication, and dieting. The identified apps in this review also supported all self-management tasks; emotional management received the least amount of support, albeit more than in the earlier review. While the support for role management was slightly less but almost equal to medical management in the earlier review, the identified apps in this review provided more resources in support of role management. The general trend in the coverage of support for self-management tasks remains consistent with the earlier review. However, the apps identified in this study provide more comprehensive support for these tasks compared to the previous review. Unlike in the earlier review, *decision-making* and *self-tailoring* self-management skills were unsupported by apps identified in this study. They were also 2 (50%) of the top 4 skills supported by mHealth covered in literature. The trend in the coverage of support for self-management skills and support activities remains consistent with the earlier review, but the identified apps provide less coverage for these skills and activities.

Quality of Self-Management Support Apps for SCI

Similar to findings from this study, recent reviews on self-management support apps for food allergies or intolerances [68], depression [69], rare diseases [70], and diabetes (3.42) [71] also report that the overall quality of self-management support apps assessed are generally acceptable. App *functionality* also received the highest ratings across these studies, which suggests that self-management support apps tend to be user-friendly. A higher *engagement* score was reported by this study compared to the other aforementioned studies, suggesting that self-management support apps for SCI are more engaging on average. *Aesthetics* received high mean scores across the studies as well, suggesting that self-management support apps across varying health conditions had appealing graphic designs, color schemes, and layouts. *Information quality* scores were also similar among these studies.

Implications for Future Practice and Research

More apps that provide comprehensive self-management support for SCI are needed. Adopting new development methods could help reduce barriers to development, such as time, cost, and securing scarce expertise. For example, low-code development platforms provide prebuilt templates and drag-and-drop interfaces, allowing apps to be created with minimal hand-coding. In addition, customizing white-label solutions can further streamline the development process. Adopting an iterative co-design approach involving key stakeholders, such as persons with SCI, health professionals, app developers, carers, and researchers, would also be key for app success. This approach could also incorporate validation studies with health professionals, efficacy and effectiveness studies through collaborations with researchers, and quality evaluations with a wide range of key stakeholders using tools such as MARS [72].

Complying with best practices should also be considered. App store categorization should be more consistent to improve the discoverability and accessibility for potential users. Privacy notices must be accessible via app stores so users can make an informed decision at the point of download and in the app. Providing information about the required device permissions is also useful.

It is critical for future research to provide a supportive evidence base for the development of high-quality self-management support apps for SCI, as none appears to exist. Investigations using app use data are needed to adequately determine their feasibility. App download counts are insufficient as they only indicate awareness of the app and potentially the initial willingness to use it. Further investigation into the data collection practices of these apps is warranted due to the challenges in distinguishing between optional and mandatory data collection and the associated concerns about the potential data privacy risks and implications. Given the high turnover of available apps on these platforms, future systematic searches in app stores need to continuously monitor and capture this information to increase the awareness of available high-quality self-management support apps that could benefit people with SCI. Persons with SCI, their carers, and related health professionals should be asked to recommend self-management support apps for consideration. Searching a wider range of

mobile app digital distribution platforms could also be beneficial for identifying more self-management support apps. Furthermore, it could be beneficial to validate the reported findings with datasets from these other platforms. Research should also consider supplementing the MARS with other instruments for a more comprehensive assessment and description of apps, such as the app evaluation model by the American Psychiatric Association [73], behavior change technique taxonomy [23], and IMS Institute for Healthcare Informatics app functionality scale [22].

Limitations

Although the most popular by a wide margin, only 2 of the many mobile app digital distribution platforms (eg, Microsoft Store, BlackBerry World, Huawei, and AppGallery) were searched, and some eligible apps could have been overlooked. The app store search keywords were broad but also limited to ensure that screening was feasible for the small team, and it cannot be guaranteed that all eligible apps were retrieved. Using a proprietary application programming interface potentially imposes limitations in terms of transparency relating to search accuracy, comprehensiveness, and documentation, which could affect the reproducibility of this study. Although unlikely, it is acknowledged that some eligible apps might have been missed due to the English language restriction. Not considering all app versions across both platforms could have biased the results, but this was unlikely in this synthesis as the identified apps were identical versions and were almost indistinguishable beyond differences imposed by the platform. Assessing identified apps in the English language could have also biased the results as other translations might not provide a similar user experience. While the inclusion of privacy policies was noted, analyzing their content and drawing conclusions about their conciseness, transparency, intelligibility, and accessibility was beyond the scope of this synthesis. The MARS neglects important aspects of app quality, such as data privacy, security, and accessibility, but provides sufficient scope with its 5 dimensions for the purposes of this initial overview of self-management support apps for SCI. Due to the exclusion of several inapplicable items from the mean score of the *information* dimension, comparisons between apps with and without complete data, especially for this dimension, should be cautiously interpreted. As mobile app digital distribution platforms constantly undergo swift changes, it is acknowledged that these findings provide a snapshot. The identified apps may have been updated or no longer available, their content may have been modified, or new mobile apps may have been developed since this systematic search was conducted, and this can limit the generalizability of this study's findings. Furthermore, these findings may not be generalizable to all self-management support apps for SCI, as the inclusion criteria were specific to certain types of apps, including those available in English and in certain app stores, such as Google Play Store and Apple App Store.

Conclusions

This study contributes to a broader understanding of mHealth self-management support tools for SCI by identifying and evaluating 13 self-management support apps for SCI. While these apps had good overall quality as measured by the MARS assessment, greater collaboration between developers and researchers is needed to improve the dissemination and availability of theory- and evidence-based self-management support apps for SCI. Failure to adopt this approach could limit the potential impact of these apps and result in missed opportunities to improve self-management support for people living with SCI. Furthermore, exploring new app development methods, such as using low-code platforms, could help promote stakeholder inclusivity in the development process and reduce barriers to app development, including time, cost, and expertise.

Acknowledgments

We dedicate this work to the memory of our dear Renaldo M Bernard, who passed away before the final publication of this study, which he spearheaded. His death is a great loss to the scientific community and to all those who knew him as a kind, joyful, and bright researcher.

The authors thank Dr Ana Ona Macías for her generous contribution to the Mobile App Rating Scale analyses.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search concepts and terms and search strategies for queried databases.
[DOCX File , 49 KB - mhealth_v12i1e53677_app1.docx]

Multimedia Appendix 2
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOCX File , 26 KB - mhealth_v12i1e53677_app2.docx]

Multimedia Appendix 3
Characteristics of the 13 identified self-management apps for spinal cord injury.
[DOCX File , 21 KB - mhealth_v12i1e53677_app3.docx]

References

1. Singh A, Tetreault L, Kalsi-Ryan S, Nouri A, Fehlings MG. Global prevalence and incidence of traumatic spinal cord injury. Clin Epidemiol 2014 Sep;6:309-331 [FREE Full text] [doi: 10.2147/CLEP.S68889] [Medline: 25278785]
2. International perspectives on spinal cord injury. World Health Organization. 2013 Nov 11. URL: https://www.who.int/publications/i/item/international-perspectives-on-spinal-cord-injury [accessed 2024-11-25]
3. Tackett MJ, Ullrich PM. Spinal cord injuries. In: Brenner LA, Reid-Arndt SA, Elliott TR, Frank RG, Caplan B, editors. Handbook of Rehabilitation Psychology. Washington, DC: American Psychological Association; 2019.
4. Pacheco Barzallo D, Oña A, Gemperli A. Unmet health care needs and inequality: a cross-country comparison of the situation of people with spinal cord injury. Health Serv Res 2021 Dec 12;56 Suppl 3(Suppl 3):1429-1440 [FREE Full text] [doi: 10.1111/1475-6773.13738] [Medline: 34386981]
5. Pearce G, Parke HL, Pinnock H, Epiphaniou E, Bourne CL, Sheikh A, et al. The PRISMS taxonomy of self-management support: derivation of a novel taxonomy and initial testing of its utility. J Health Serv Res Policy 2016 Apr 15;21(2):73-82. [doi: 10.1177/1355819615602725] [Medline: 26377727]
6. Barlow J, Wright C, Sheasby J, Turner A, Hainsworth J. Self-management approaches for people with chronic conditions: a review. Patient Educ Couns 2002 Oct;48(2):177-187. [doi: 10.1016/s0738-3991(02)00032-0] [Medline: 12401421]
7. Yi JY, Kim Y, Cho YM, Kim H. Self-management of chronic conditions using mHealth interventions in Korea: a systematic review. Healthc Inform Res 2018 Jul;24(3):187-197 [FREE Full text] [doi: 10.4258/hir.2018.24.3.187] [Medline: 30109152]
8. Conti A, Clari M, Kangasniemi M, Martin B, Borraccino A, Campagna S. What self-care behaviours are essential for people with spinal cord injury? A systematic review and meta-synthesis. Disabil Rehabil 2022 Apr 30;44(7):991-1006. [doi: 10.1080/09638288.2020.1783703] [Medline: 32602384]
9. Classification of digital health interventions v1.0: a shared language to describe the uses of digital technology for health. World Health Organization. 2018 Jan 1. URL: https://www.who.int/publications/i/item/WHO-RHR-18.06 [accessed 2024-11-25]
10. mHealth: new horizons for health through mobile technologie. World Health Organization. URL: https://www.afro.who.int/publications/mhealth-new-horizons-health-through-mobile-technologie [accessed 2024-11-25]
11. mHealth: use of appropriate digital technologies for public health: report by the Director-General. World Health Organization. 2017. URL: https://iris.who.int/handle/10665/274134 [accessed 2024-11-25]
12. Smartphones in Europe - statistics and facts. Statista. URL: https://www.statista.com/topics/3341/smartphone-market-in-europe/#topicOverview [accessed 2024-11-25]
13. Penetration rate of smartphones in selected countries. Statista. URL: https://www.statista.com/search/?q=penetration+rate+smartphones&Search=&p=1 [accessed 2024-11-25]
14. Agnihothri S, Cui L, Delasay M, Rajan B. The value of mHealth for managing chronic conditions. Health Care Manag Sci 2020 Jun 31;23(2):185-202. [doi: 10.1007/s10729-018-9458-2] [Medline: 30382448]
15. Wellbeloved-Stone CA, Weppner JL, Valdez RS. A systematic review of telerehabilitation and mHealth interventions for spinal cord injury. Curr Phys Med Rehabil Rep 2016 Nov 30;4(4):295-311. [doi: 10.1007/s40141-016-0138-1]
16. Bernard RM, Davis M, Volkova A, Vanessa S, Sabariego C. Mobile health technologies for self-management in spinal cord injuries: a systematic literature review. OSF Registries. 2022 Apr 14. URL: https://osf.io/nwh63 [accessed 2022-05-14]
17. Ahuja CS, Nori S, Tetreault L, Wilson J, Kwon B, Harrop J, et al. Traumatic spinal cord injury-repair and regeneration. Neurosurgery 2017 Mar 01;80(3S):S9-22 [FREE Full text] [doi: 10.1093/neuros/nyw080] [Medline: 28350947]
18. Stoffel JT. Detrusor sphincter dyssynergia: a review of physiology, diagnosis, and treatment strategies. Transl Androl Urol 2016 Feb;5(1):127-135 [FREE Full text] [doi: 10.3978/j.issn.2223-4683.2016.01.08] [Medline: 26904418]
19. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Int J Surg 2021 Apr;88:105906 [FREE Full text] [doi: 10.1016/j.ijsu.2021.105906] [Medline: 33789826]
20. Rethlefsen M, Kirtley S, Waffenschmidt S, Ayala AP, Moher D, Page MJ, et al. PRISMA-S: an extension to the PRISMA statement for reporting literature searches in systematic reviews. OSF Preprints Preprint posted online on March 20, 2019 [FREE Full text] [doi: 10.31219/osf.io/sfc38]
21. Apps for self-management in spinal cord injury: protocol for a systematic search of app stores. OSF Registries. 2022 Jul 14. URL: https://osf.io/wcvdm [accessed 2022-10-31]
22. Aitken M, Gauntlett C. Patient apps for improved healthcare: from novelty to mainstream. IMS Institute for Healthcare Informatics. 2013 Oct. URL: https://ignacioriesgo.es/wp-content/uploads/2014/03/iihi_patient_apps_report_editora_39_2_1.pdf [accessed 2024-11-25]
23. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug 20;46(1):81-95 [FREE Full text] [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
24. Alfawzan N, Christen M, Spitale G, Biller-Andorno N. Privacy, data sharing, and data security policies of women's mHealth apps: scoping review and content analysis. JMIR Mhealth Uhealth 2022 May 06;10(5):e33735 [FREE Full text] [doi: 10.2196/33735] [Medline: 35522465]

25. Martinengo L, Stona A, Griva K, Dazzan P, Pariante CM, von Wangenheim F, et al. Self-guided cognitive behavioral therapy apps for depression: systematic assessment of features, functionality, and congruence with evidence. J Med Internet Res 2021 Jul 30;23(7):e27619 [FREE Full text] [doi: 10.2196/27619] [Medline: 34328431]
26. Flors-Sidro JJ, Househ M, Abd-Alrazaq A, Vidal-Alaball J, Fernandez-Luque L, Sanchez-Bocanegra CL. Analysis of diabetes apps to assess privacy-related permissions: systematic search of apps. JMIR Diabetes 2021 Jan 13;6(1):e16146 [FREE Full text] [doi: 10.2196/16146] [Medline: 33439129]
27. All the mobile and CTV app intelligence you need. 42matters. URL: https://42matters.com/ [accessed 2022-06-15]
28. Number of apps available in leading app stores as of August 2024. Statista. URL: https://www.statista.com/statistics/276623/number-of-apps-available-in-leading-app-stores/#:~:text=In%20the%20third%20quarter%20of,of%20the%20Google%20Play%20platform [accessed 2024-11-25]
29. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016 Dec 05;5(1):210 [FREE Full text] [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
30. Denison HJ, Dodds RM, Ntani G, Cooper R, Cooper C, Sayer AA, et al. How to get started with a systematic review in epidemiology: an introductory guide for early career researchers. Arch Public Health 2013 Aug 07;71(1):21 [FREE Full text] [doi: 10.1186/0778-7367-71-21] [Medline: 23919540]
31. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
32. Terhorst Y, Philippi P, Sander LB, Schultchen D, Paganini S, Bardus M, et al. Validation of the mobile application rating scale (MARS). PLoS One 2020 Nov 2;15(11):e0241480 [FREE Full text] [doi: 10.1371/journal.pone.0241480] [Medline: 33137123]
33. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979 Mar;86(2):420-428. [doi: 10.1037//0033-2909.86.2.420] [Medline: 18839484]
34. Lorig KR, Holman HR. Self-management education: history, definition, outcomes, and mechanisms. Ann Behav Med 2003 Aug;26(1):1-7. [doi: 10.1207/s15324796abm2601_01]
35. Action Blocks. Google Play. URL: https://play.google.com/store/apps/details/Action_Blocks?id=com.google.android.apps.accessibility.maui.actionblocks&hl=en_IN [accessed 2024-11-25]
36. JIB CALLS. Google Play. URL: https://play.google.com/store/apps/details?id=com.jib.calls&hl=en_IN [accessed 2024-11-25]
37. Maslow - collaborative care. Google Play. URL: https://play.google.com/store/apps/details?id=io.maslow.maslow&hl=en_IN [accessed 2024-11-25]
38. Dietitian's Tools. Kinnereth LLC App Dev. 2019.
39. Injectful. Injectful LLC. 2020.
40. AccessiRep. Cordilac LC. 2020. URL: https://apps.apple.com/ch/app/accessirep/id1451239112?l=en-GB [accessed 2024-12-11]
41. iAccess life - accessibility. Google Play. URL: https://play.google.com/store/apps/details?id=com.iaccessapp&hl=en_IN [accessed 2024-11-25]
42. Monster Hub. Neuro Therapy. App Store. 2019. URL: https://apps.apple.com/ch/app/neuro-therapy/id1479833318?l=en-GB [accessed 2024-12-11]
43. Spine Fine by Giorgio Lofrese. AppAdvice. URL: https://appadvice.com/app/spine-fine/1601293313 [accessed 2024-11-25]
44. PVA events. Google Play. URL: https://play.google.com/store/apps/details?id=com.aventri.pva300009319&hl=en_IN [accessed 2024-11-25]
45. SNS Digital. Paralyzed Veterans of America. 2020.
46. Paralyzed Veterans of America. PVA ePubs. App Store. 2022. URL: https://apps.apple.com/ch/app/pva-epubs/id596481957?l=en-GB [accessed 2024-12-11]
47. Pilates. DSN Inc. 2019.
48. Methodology: standard country or area codes for statistical use (M49). United Nations Statistics Division. URL: https://unstats.un.org/unsd/methodology/m49/ [accessed 2024-10-30]
49. Android Apps by DSN Inc. DSN Inc.
50. Android Apps by Google LLC on Google Play. Google Play. URL: https://play.google.com/store/apps/dev?id=5700313618786177705&hl=en_IN [accessed 2024-10-22]
51. Android apps by paralyzed veterans of America on Google Play. App Store. URL: https://apps.apple.com/nr/developer/paralyzed-veterans-of-america/id596481960 [accessed 2024-12-11]
52. Google P. Android Apps by JIB Smart Home on Google Play. 2022 24. URL: https://play.google.com/store/apps/developer?id=JIB+Smart+Home&hl=en [accessed 2024-10-30]
53. Maslow for people. Google Play. URL: https://play.google.com/store/apps/developer?id=Maslow+For+People [accessed 2024-11-25]
54. Android Apps by Kinnereth LLC App Dev on Google Play. Kinnereth LLC.
55. Android Apps by Injectful LLC on Google Play. Injectful LLC.

56. iAccess innovations. Google Play. URL: https://play.google.com/store/apps/developer?id=iAccess+Innovations [accessed 2024-11-25]
57. Cordilac LC Apps on the App Store. App Store. URL: https://apps.apple.com/us/developer/cordical-lc/id1451239111 [accessed 2024-11-25]
58. Monster Hub Apps on the App Store. App Store. URL: https://apps.apple.com/us/developer/monster-hub/id1455953573 [accessed 2024-11-25]
59. Giorgio Lofrese Apps on the App Store. App Store. URL: https://apps.apple.com/us/developer/giorgio-lofrese/id1601293315 [accessed 2024-11-25]
60. iPad and iPhone. App Store. URL: https://apps.apple.com/us/developer/google-llc/id281956209?see-all=i-phonei-pad-apps [accessed 2024-11-25]
61. Paralyzed Veterans of America Apps on the App Store. App Store. URL: https://apps.apple.com/us/developer/paralyzed-veterans-of-america/id596481960 [accessed 2024-11-25]
62. JIB Smart Home SAS. App Store. URL: https://apps.apple.com/ch/developer/jib-smart-home-sas/id1435310186 [accessed 2024-11-25]
63. iThai - Thailand Official Guide. App Store. URL: https://apps.apple.com/au/developer/ilya-thai/id1495108430 [accessed 2024-11-25]
64. App S. Android Apps by Injectful LLC on Google Play. 2022 24. URL: https://apps.apple.com/us/developer/injectful-llc/id1536724322 [accessed 2024-10-30]
65. iAccess Innovations Apps on the App Store. App Store. URL: https://apps.apple.com/us/developer/iaccess-innovations/id1453988842 [accessed 2024-11-25]
66. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016 Jun;15(2):155-163 [FREE Full text] [doi: 10.1016/j.jcm.2016.02.012] [Medline: 27330520]
67. Bernard RM, Seijas V, Davis M, Volkova A, Diviani N, Lüscher J, et al. Mobile health self-management support for spinal cord injury: systematic literature review. JMIR Mhealth Uhealth 2023 Apr 26;11:e42679 [FREE Full text] [doi: 10.2196/42679] [Medline: 37099372]
68. Mandracchia F, Llauradó E, Tarro L, Valls RM, Solà R. Mobile phone apps for food allergies or intolerances in app stores: systematic search and quality assessment using the mobile app rating scale (MARS). JMIR Mhealth Uhealth 2020 Sep 16;8(9):e18339 [FREE Full text] [doi: 10.2196/18339] [Medline: 32936078]
69. Myers A, Chesebrough L, Hu R, Turchioe MR, Pathak J, Creber RM. Evaluating commercially available mobile apps for depression self-management. AMIA Annu Symp Proc 2020;2020:906-914 [FREE Full text] [Medline: 33936466]
70. Hatem S, Long JC, Best S, Fehlberg Z, Nic Giolla Easpaig B, Braithwaite J. Mobile apps for people with rare diseases: review and quality assessment using mobile app rating scale. J Med Internet Res 2022 Jul 26;24(7):e36691 [FREE Full text] [doi: 10.2196/36691] [Medline: 35881435]
71. Gong E, Zhang Z, Jin X, Liu Y, Zhong L, Wu Y, et al. Quality, functionality, and features of Chinese mobile apps for diabetes self-management: systematic search and evaluation of mobile apps. JMIR Mhealth Uhealth 2020 Apr 07;8(4):e14836 [FREE Full text] [doi: 10.2196/14836] [Medline: 32255432]
72. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and validation of the user version of the mobile application rating scale (uMARS). JMIR Mhealth Uhealth 2016 Jun 10;4(2):e72 [FREE Full text] [doi: 10.2196/mhealth.5849] [Medline: 27287964]
73. Lagan S, Aquino P, Emerson MR, Fortuna K, Walker R, Torous J. Actionable health app evaluation: translating expert frameworks into objective metrics. NPJ Digit Med 2020 Jul 30;3(1):100 [FREE Full text] [doi: 10.1038/s41746-020-00312-4] [Medline: 32821855]

Abbreviations

MARS: Mobile App Rating Scale
mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
SCI: spinal cord injury

Edited by L Buis; submitted 15.10.23; peer-reviewed by S Newman, A Van de Winckel; comments to author 26.03.24; revised version received 22.08.24; accepted 11.10.24; published 19.12.24.

Please cite as:
Bernard RM, Seijas V, Davis M, Volkova A, Diviani N, Lüscher J, Sabariego C
Self-Management Support Apps for Spinal Cord Injury: Results of a Systematic Search in App Stores and Mobile App Rating Scale Evaluation
JMIR Mhealth Uhealth 2024;12:e53677
URL: https://mhealth.jmir.org/2024/1/e53677
doi:10.2196/53677
PMID:

Exploring and Predicting HIV Preexposure Prophylaxis Adherence Patterns Among Men Who Have Sex With Men: Randomized Controlled Longitudinal Study of an mHealth Intervention in Western China

Bing Lin[1,2], PhD; Jiayan Li[1,2], MPH; Jiaxiu Liu[3], PhD; Wei He[4], MPH; Haiying Pan[4], MPH; Xiaoni Zhong[1,2], MPH

[1]School of Public Health, Chongqing Medical University, Chongqing, China

[2]Research Center for Medicine and Social Development, Chongqing, China

[3]School of Medical Informatics, Chongqing Medical University, Chongqing, China

[4]Jiulongpo District Center for Disease Control and Prevention, Chongqing, China

Corresponding Author:
Xiaoni Zhong, MPH
School of Public Health, Chongqing Medical University, Chongqing, China

Abstract

Background: Preexposure prophylaxis (PrEP) is an effective strategy to reduce the risk of HIV infection. However, the efficacy of PrEP is highly dependent on adherence. Meanwhile, adherence changes over time, making it difficult to manage effectively.

Objective: Our study aimed to explore and predict the patterns of change in PrEP adherence among men who have sex with men (MSM) and evaluate the impact of the WeChat-based reminder intervention on adherence, thus providing more information for PrEP implementation strategies.

Methods: From November 2019 to June 2023, in a randomized controlled longitudinal study of the PrEP demonstration project in Western China (Chongqing, Sichuan, and Xinjiang) based on a mobile health (mHealth) reminder intervention, participants were randomly divided into reminder and no-reminder groups, with those in the reminder group receiving daily reminders based on the WeChat app. Participants were followed up and self-reported their medication adherence every 12 weeks for a total of 5 follow-up visits. We used the growth mixture model (GMM) to explore potential categories and longitudinal trajectories of adherence among MSM, and patterns of change in PrEP adherence were predicted and evaluated based on the decision tree.

Results: A total of 446 MSM were included in the analysis. The GMM identified 3 trajectories of adherence: intermediate adherence group (n=34, 7.62%), low adherence ascending group (n=126, 28.25%), and high adherence decline group (n=286, 64.13%). We included 8 variables that were significant in the univariate analysis in the decision tree prediction model. We found 4 factors and 8 prediction rules, and the results showed that HIV knowledge score, education attainment, mHealth intervention, and HIV testing were key nodes in the patterns of change in adherence. After 10-fold cross-validation, the final prediction model had an accuracy of 75%, and the classification accuracy of low and intermediate adherence was 78.12%.

Conclusions: The WeChat-based reminder intervention was beneficial for adherence. A short set of questions and prediction rules, which can be applied in future large-scale validation studies, aimed at developing and validating a short adherence assessment tool and implementing it in PrEP practices among MSM.

Trial Registration: Chinese Clinical Trial Registry ChiCTR1900026414; https://www.chictr.org.cn/showproj.html?proj=35077

(JMIR Mhealth Uhealth 2024;12:e58920) doi:10.2196/58920

KEYWORDS

preexposure prophylaxis; adherence; trajectory analysis; men who have sex with men; mHealth; mobile health; mHealth intervention; decision tree

Introduction

In China, men who have sex with men (MSM) are the high-risk population for HIV infection. Previous studies have estimated the overall national HIV prevalence among MSM in China to be 5.7% from 2001 to 2018 [1]. The annual number of newly diagnosed HIV infections through homosexual transmission in China increased from 2.5% in 2006 to 25.6% in 2022 [2]. To reduce the persistently high incidence of new infections among high-risk populations, we need to focus on HIV prevention. Preexposure prophylaxis (PrEP) is a biomedical HIV prevention intervention that consists primarily of the prophylactic daily use of antiretroviral medications to reduce the risk of infection

in the event of HIV exposure [3]. Effective PrEP services for at-risk populations, including MSM, are key to reducing new HIV infections [4].

However, a growing number of studies have shown that the efficacy of PrEP is highly dependent on medication adherence [5,6]. PrEP adherence is defined as the users taking the medication as prescribed by the clinician, occurring behaviors consistent with the clinician's orders, and vice versa, known as nonadherence. A previous systematic review and meta-analysis study showed that the PrEP strategy was highly effective among MSM. Meanwhile, PrEP efficacy was strongly correlated with adherence, with the RR decreasing by 0.13 as adherence increased from 50% to 60%. It was shown that, on average, a 10% decrease in adherence would correspond to a 13% decrease in PrEP efficacy [7]. In summary, achieving a high level of adherence is necessary for PrEP to be completely effective. This highlights the need for good adherence in ensuring the efficacy of PrEP among the MSM population.

Moreover, earlier studies have acknowledged the time-varying nature of adherence, both within and between individuals over time [8]. Exploring and elucidating patterns of change in adherence can contribute to understanding the reasons for users' behavioral changes and enhance the capacity to recognize individuals who are more susceptible to nonadherence. Studies assessing long-term behavioral patterns of PrEP adherence have been conducted in many countries, including Australia [9], America [10], and South Africa [11]. However, at present, few studies in China have assessed patterns of change in adherence. Taken together, several previous empirical studies have shown that adherence is critical for PrEP efficacy in high-risk populations and deserves further attention. Future PrEP strategies will be more effective if we are able to accurately recognize patterns of change in adherence and develop predictive models that reliably identify individuals who engage in nonadherent behavior. Therefore, our study aimed to explore and predict the patterns of change in PrEP adherence based on the decision tree model, thus providing more information for PrEP implementation strategies. Our study was a preliminary application of decision tree prediction modeling to PrEP adherence among the MSM population in Western China. More importantly, recent simulation studies have shown that machine learning techniques, for example, decision tree modeling, can be reliably applied to relatively small samples and are more applicable to our particular study population [12,13].

In addition, mobile health (mHealth) has emerged as a promising tool for improving health care access and delivery globally [14]. Previous studies have shown that mHealth could reach and engage MSM in HIV prevention and care, which suggested that mHealth interventions have the potential to improve HIV prevention efforts [15]. At the same time, willingness to use mHealth intervention for HIV prevention was high among the MSM population [16]. Although research on adherence interventions for PrEP was in its early stages, some mHealth intervention methods showed promise. The WeChat app (Tencent Holdings Ltd) is the most popular social media platform in China, with 1.27 billion monthly active users [17]. Thus, a WeChat-based medication reminder service is a low-cost and scalable intervention.

In summary, based on the randomized controlled longitudinal study of the PrEP demonstration project through the mHealth intervention, the aims of our study were to (1) identify patterns of change in PrEP adherence among the MSM population; (2) provide an initial basis for identifying a set of questions and decision rules that can accurately predict individuals at high risk for low PrEP adherence behaviors among the MSM population based on the decision tree prediction model; and (3) evaluate the impact of the mHealth intervention on adherence. Validation in larger samples in the future could contribute to the creation of a comprehensive, brief adherence screening tool based on the results of our study. In the long term, our study will help administrators target individuals with low adherence to preventive interventions and, at the same time, provide a direct basis for policy development and service planning for the implementation of PrEP among MSM in China, which is of great practical significance.

Methods

Participants and Procedure

From November 2019 to June 2023, our study was conducted among the MSM population in Western China (Chongqing, Sichuan, and Xinjiang). The randomized controlled longitudinal study of the PrEP demonstration project was based on a mobile phone intervention, accompanied by the Chinese 13th Five-Year Plan for AIDS Prevention and Control carried out by the Ministry of Science and Technology (Chinese Clinical Trial Registration Number: ChiCTR1900026414).

The participants were HIV-negative MSM individuals at higher risk of HIV infection. We recruited eligible MSM individuals through collaboration with local nongovernmental organizations and peer referrals. The inclusion criteria included (1) assigned male sex at birth; (2) age 18 - 65 years; (3) negative HIV antigen antibody test; (4) self-reported high-risk sexual activity with a male partner (both temporary and regular) in the last 6 months; (5) no serious heart, liver, or kidney disease or hematopoietic dysfunction, bleeding tendency, and bleeding disease; (6) willing to use the experimental drug under guidance and comply with follow-up arrangements; and (7) signed informed consent.

After completing the baseline questionnaire, participants who met the inclusion criteria were randomly divided into a reminder group and a no-reminder group. Participants in the reminder group received daily mobile phone–based reminder messages and took lamivudine-tenofovir (PrEP drug) orally daily. The reminder information relies on the internet and an intelligent medication information tracking management system using cloud computing, big data, intelligent hardware, and other new-generation IT products to collect medication plans, medication reminders, medication logs, electronic instructions for medicines, adverse reactions, health knowledge, and other process information, and through the collection of information, it provides information for the administrator to provide the user with effective medication management to achieve the expected preventive effect. The reminder messages were sent at regular times daily through the WeChat app to remind users to take PrEP. At the same time, in order to protect the user's privacy, the content of the messages we sent were "processed." For

example, when users received "You need to learn," it meant that they were reminded to take medication; "Time for examination" meant that it was time for the next follow-up visit. The presentation of the reminder messages on WeChat is shown in Figures S1 and S2 in Multimedia Appendix 1. Participants in the no-reminder group who took PrEP daily did not receive reminder messages. Participants were followed up and self-reported their medication adherence every 12 weeks for a total of 5 follow-up visits.

Measurement

Demographic variables of MSM included age, place of residence, ethnicity, education attainment, employment status, marital status, and monthly disposable income.

The HIV Knowledge Scale (Cronbach α=.672) was used to measure knowledge mastery in the MSM population. The scale consisted of 13 questions and was based on the revised Universal Scale of the International HIV Knowledge Survey [18]. The answers included "correct," "wrong," and "I don't know." One point was awarded for a correct answer and zero points for an incorrect or unknown answer. A higher score indicated a higher level of knowledge about HIV. We believed that a score ≥11 indicates a high level of knowledge [19,20].

Meanwhile, we also measured the level of risk perception of the participants. They were asked, "How likely do you think you are to contract HIV? How serious do you think AIDS is? How much of a threat do you think AIDS poses to you?" These 3 items were used to measure their perceived risk. Participants rated their perceived risk on a scale of 1 - 5, with 1 - 2 representing a low level of perception, 3 representing a moderate level of perception, and 4 - 5 representing a high level of perception.

Previous literature mentioned that MSM may show different preferences and willingness to use PrEP and other biomedical prevention strategies based on the role of anal sex [21]. Therefore, our study included the variable "sexual role" to explore whether differences in sexual roles might lead to differences in adherence to PrEP in the MSM population. Sexual role was the way of performing sexual behavior with a male sexual partner, such as inserter (like "top") and receiver (like "bottom").

We also collected information about HIV testing, HIV counseling, the number of sexual partners, condom use, internet searches for sexual partners, sexually transmitted diseases history, commercial sex, recreational drug use, attitudes, and potential effects of PrEP use among male sexual partners among the participants.

Adherence

During the follow-up, participants self-reported the number of days they missed taking their medication in the previous 2 weeks, and adherence was equal to the proportion of days medication was taken.

Growth Mixture Model

The growth mixture model (GMM) was used in our study to determine potential class trajectories for adherence. The model used Akaike information criteria (AIC), Bayesian information criteria (BIC), sample size–adjusted BIC (aBIC), and entropy to explore the optimal trajectory of internalization and externalization problems. The AIC, BIC, and aBIC statistics were all used to determine how well the model fits by comparing the difference between the expected and actual values, with lower values indicating a good model fit. The entropy took values in the range 0 - 1, with closer to 1 indicating a clearer group classification. The Vuong-Lo-Mendell-Rubin likelihood ratio test and bootstrapped likelihood ratio test were used to compare the differences in fit between the k-1 and k-category models. When the *P* value was significant, the model was considered well fitted for the k-category.

Decision Tree

The decision tree is based on the known probability of the occurrence of various situations, through the composition of the decision rule to determine the probability and evaluate and judge the risk decision analysis method, which is the intuitive use of probability analysis using a graphical method. This decision branching is drawn graphically, much like the branches of a tree. The top decision node in the tree is called the root node; it corresponds to the best predictor. The final nodes are called terminal nodes or leaves. The numbers on the terminal nodes indicate the probability of each class and are what determine the final classification. In our study, the Gini index was used as a criterion for selecting features. The magnitude of the Gini index indicated the complexity of the data, and when the Gini index was less than a threshold, it indicated that the purity of the data was greater and the nodes of the tree stopped splitting. The decision tree model was pruned according to the complexity parameter (Cp) value, and the Cp value corresponding to the smallest error was selected to obtain the optimal decision tree model.

Statistical Analysis

Longitudinal data collected on repeated measures of PrEP adherence among the MSM population were analyzed using the GMM to explore potential categories and longitudinal trajectories of adherence, and statistical metrics were used to measure the degree of model fit. Based on the different potential categories of adherence, relevant information was compared using univariate analysis, and predictors with a *P* value ≤.05 were screened and included in the decision tree prediction model. The SMOTE algorithm in the "smotefamily" package of the R program was used to balance the adherence classification data; the "rpart" package was used to construct the decision tree prediction model. The predictive power of the final model was assessed by three criteria: (1) "accuracy" indicates the overall percentage of correct classifications; (2) "positive classification accuracy" refers to the percentage of predictions that the model accurately classifies as positive; and (3) "negative classification accuracy" refers to the percentage of predictions that the model accurately classifies as negative. For each of the 3 evaluation criteria, 10-fold cross-validation was used, and their estimates were computed as the average of the 10 test samples. At the same time, missing values were considered to be informative indicators, and therefore were considered as potential predictors in modeling [22]. *P* value

≤.05 represented a statistical difference. All analyses were completed using Mplus (Muthén & Muthén) and R software (R Foundation for Statistical Computing).

Ethical Considerations

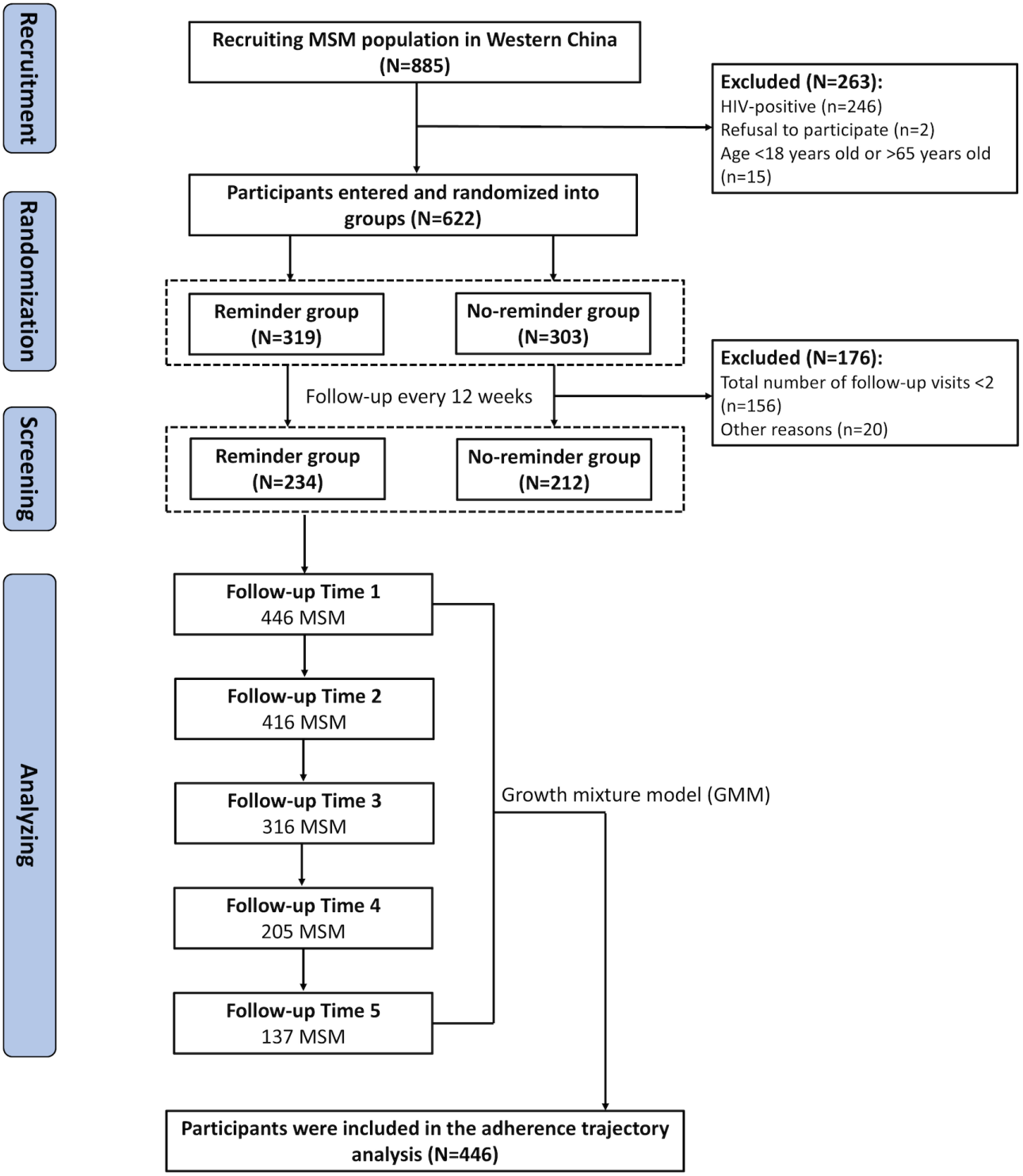
The study was approved by the Ethics Committee of Chongqing Medical University (reference number: 2019001). Before participating in this research, the participants were fully informed about the purpose, significance, voluntary participation, and confidentiality of the research. Each participant signed an informed consent form.

Results

Basic Demographic Information

A total of 885 MSM were recruited in Western China. Of these, 263 MSM were excluded because they did not meet the requirements for inclusion in the cohort. A total of 622 MSM were enrolled in the longitudinal study and were randomized into a reminder group (n=319) and a no-reminder group (n=303). Due to the need to model longitudinal trajectories of adherence, we excluded a subset of participants with less than 2 follow-up visits. A total of 446 MSM were eventually included in the statistical analysis, with 234 and 212 MSM in the reminder and no-reminder groups, respectively. The detailed screening process is shown in Figure 1. We compared basic demographic information between the reminder and no-reminder groups (Table S1 in Multimedia Appendix 1). According to the results of the univariate analysis, none of the variables differed statistically in both groups. We performed descriptive analyses of the number of MSM and adherence in each follow-up period (Table S2 in Multimedia Appendix 1).

Figure 1. Flowchart for the screening of study participants. MSM: men who have sex with men; GMM: growth mixture model.



Exploring Patterns of Change in Adherence

Based on the GMM, we explored 1, 2, 3, and 4 potential categories of adherence (Table 1). Considering the parameters, the scheme for category 4 is reasonable. However, Class 2 accounted for 3.81% (n=17) of the scheme for category 4, which is too small a number to carry out subsequent analyses, and the trajectories of Class 1 and Class 2 were partially close to overlapping. Therefore, we chose the scheme with class probability and trajectory distribution, which was more reasonable for category 3. The parameters of the model also showed good model fit. The adherence trajectories for each categorical scheme were plotted in Figure 2.

Table . Exploration of potential categories for PrEP[a] adherence among MSM[b] population based on GMM[c].

Class	AIC[d]	BIC[e]	aBIC[f]	Entropy	VLRT[g]	BLRT[h]	Class probability
1C[i]	1119.749	1168.953	1130.870	—[j]	—	—	1
2C	550.175	611.679	564.076	0.989	<.001	<.001	0.7108/0.2892
3C	356.855	430.661	373.537	0.972	<.001	<.001	0.0762/0.2825/0.6413
4C	145.987	232.094	165.449	0.969	<.001	<.001	0.2490/0.0381/0.6121/0.1009

[a]PrEP: preexposure prophylaxis.

[b]MSM: men who have sex with men.

[c]GMM: growth mixture model.

[d]AIC: Akaike information criteria.

[e]BIC: Bayesian information criteria.

[f]aBIC: sample size–adjusted BIC.

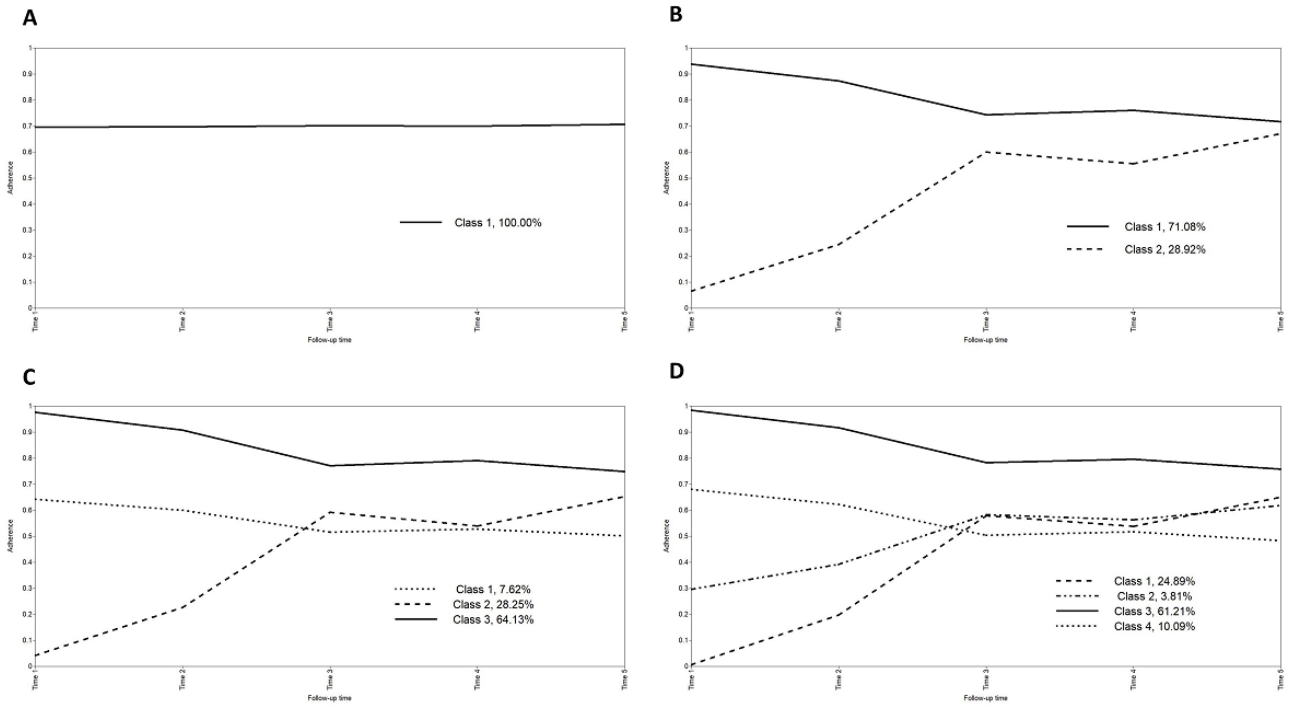
[g]VLRT: Vuong-Lo-Mendell-Rubin likelihood ratio test.

[h]BLRT: bootstrapped likelihood ratio test.

[i]C: category.

[j]Not applicable.

Figure 2. Trajectory of PrEP adherence potential categories in the MSM population. A, B, C, and D correspond to schemes with 1, 2, 3, and 4 adherence potential categories, respectively. PrEP: preexposure prophylaxis; MSM: men who have sex with men.



Based on the descriptions of the 3 latent categories of adherence (Table 2), Class 1 (n=34, 7.62%) had mean values of intercept (I) and slope (S) of 0.642 (P<.001) and –0.042 (P=.35), respectively, and was named "Intermediate adherence group." Class 2 (n=126, 28.25%) had mean values of intercept (I) and slope (S) of 0.042 (P<.001) and 0.183 (P<.001), respectively, and was named "Low adherence ascending group." Class 3 (n=286, 64.13%) had mean values of intercept (I) and slope (S) of 0.976 (P<.001) and –0.068 (P<.001), respectively, and was named "High adherence decline group."

Table . Description of GMM[a] parameters for 3 potential categories of PrEP[b] adherence among MSM[c] population.

Variables	Estimate	SE	*P* value
Class 1: Intermediate adherence group			
Mean I[d]	0.642	0.039	<.001
Mean S[e]	–0.042	0.045	.35
Class 2: Low adherence ascending group			
Mean I	0.042	0.010	<.001
Mean S	0.183	0.020	<.001
Class 3: High adherence decline group			
Mean I	0.976	0.007	<.001
Mean S	–0.068	0.011	<.001

[a]GMM: growth mixture model.

[b]PrEP: preexposure prophylaxis.

[c]MSM: men who have sex with men.

[d]I: intercept.

[e]S: slope.

Differences Between Adherence Potential Categories

We performed a univariate analysis of the relevant variables (Table 3). According to the results of the χ^2 test, educational attainment (*P*=.007), HIV knowledge score (*P*<.001), HIV testing (*P*=.05), HIV counseling (*P*=.02), number of male sexual partners (*P*=.003), attitudes of male sexual partners (*P*=.04), PrEP use (*P*=.05), and groups (reminder and no-reminder groups, *P*=.01) showed significant variability between patterns of change in adherence.

Table . Univariate analysis of different latent categories of PrEP[a] adherence among MSM[b] population.

Variables	Total (N=446)	Intermediate adherence group (n=34)	Low adherence ascending group (n=126)	High adherence decline group (n=286)	*P* value
Age, n (%)					.18
18 - 25	70 (15.70)	8 (23.53)	26 (20.63)	36 (12.59)	
25 - 35	179 (40.13)	11 (32.35)	48 (38.10)	120 (41.96)	
≥35	197 (44.17)	15 (44.12)	52 (41.27)	130 (45.45)	
Place of residence, n (%)[c]					.39
Urban	329 (74.43)	27 (79.41)	87 (70.16)	215 (75.70)	
Rural	113 (25.57)	7 (20.59)	37 (29.84)	69 (24.30)	
Ethnicity, n (%)					.39
Ethnic Han	413 (92.60)	33 (97.06)	114 (90.48)	266 (93.01)	
Ethnic minorities	33 (7.40)	1 (2.94)	12 (9.52)	20 (6.99)	
Education attainment, n (%)					.007[d]
Primary school or below	4 (0.90)	1 (2.94)	3 (2.38)	0 (0.00)	
Junior high	24 (5.38)	0 (0.00)	11 (8.73)	13 (4.55)	
High school or vocational high school	93 (20.85)	8 (23.53)	32 (25.40)	53 (18.53)	
College or university degree or above	325 (72.87)	25 (73.53)	80 (63.49)	220 (76.92)	
Employment status, n (%)					.82[d]
Employed	384 (86.10)	28 (82.36)	108 (85.71)	248 (86.72)	
Unemployed or retired	33 (7.40)	4 (11.76)	10 (7.94)	19 (6.64)	
Student	29 (6.50)	2 (5.88)	8 (6.35)	19 (6.64)	
Marital status, n (%)					.53
Married	60 (13.45)	3 (8.82)	15 (11.90)	42 (14.69)	
Unmarried or divorced	386 (86.55)	31 (91.18)	111 (88.10)	244 (85.31)	
Monthly disposable income, n (%)[c]					.69
1000 - 3000 CNY[e]	104 (23.37)	7 (20.59)	35 (28.00)	62 (21.68)	
3000 - 10,000 CNY	310 (69.66)	25 (73.53)	82 (65.60)	203 (70.98)	
≥10,000 CNY	31 (6.97)	2 (5.88)	8 (6.40)	21 (7.34)	
HIV knowledge score, n (%)					<.001
<11	204 (45.74)	12 (35.29)	80 (63.49)	112 (39.16)	
≥11	242 (54.26)	22 (64.71)	46 (36.51)	174 (60.84)	
HIV testing, n (%)[c]					.05
Yes	360 (82.38)	30 (88.24)	88 (75.21)	242 (84.62)	
No	77 (17.62)	4 (11.76)	29 (24.79)	44 (15.38)	
HIV counseling, n (%)					.02
Yes	327 (73.32)	24 (70.59)	81 (64.29)	222 (77.62)	
No	119 (26.68)	10 (29.41)	45 (35.71)	64 (22.38)	
How likely do you think you are to contract HIV?, n (%)					.78
Low level	263 (58.97)	21 (61.76)	79 (62.70)	163 (56.99)	
Moderate level	140 (31.39)	9 (26.48)	37 (29.37)	94 (32.87)	

Variables	Total (N=446)	Intermediate adherence group (n=34)	Low adherence ascending group (n=126)	High adherence decline group (n=286)	*P* value
High level	43 (9.64)	4 (11.76)	10 (7.93)	29 (10.14)	
How serious do you think AIDS is?, n (%)[c]					.54[d]
Low level	1 (0.23)	0 (0.00)	1 (0.83)	0 (0.00)	
Moderate level	42 (9.68)	3 (9.68)	10 (8.26)	29 (10.28)	
High level	391 (90.09)	28 (90.32)	110 (90.91)	253 (89.72)	
How much of a threat do you think AIDS poses to you?, n (%)[c]					.61[d]
Low level	40 (9.01)	4 (11.76)	13 (10.40)	23 (8.07)	
Moderate level	61 (13.74)	3 (8.82)	14 (11.20)	44 (15.44)	
High level	343 (77.25)	27 (79.42)	98 (78.40)	218 (76.49)	
Sexual role, n (%)[c]					.97
Mainly “bottom”	124 (28.51)	8 (25.00)	32 (27.12)	84 (29.47)	
Both	88 (20.23)	7 (21.88)	25 (21.19)	56 (19.65)	
Mainly “top”	223 (51.26)	17 (53.12)	61 (51.69)	145 (50.88)	
Number of male sexual partners last month (both temporary and regular), n (%)[c]					.003
0	41 (9.31)	3 (8.83)	22 (17.89)	16 (5.65)	
1	251 (57.05)	21 (61.76)	63 (51.22)	167 (59.01)	
2 or more	148 (33.64)	10 (29.41)	38 (30.89)	100 (35.34)	
Condom use during sex with a male partner, n (%)[c]					.58
Use every time	298 (70.78)	24 (77.42)	74 (65.49)	200 (72.20)	
Sometime use	92 (21.86)	5 (16.13)	28 (24.78)	59 (21.30)	
Never use	31 (7.36)	2 (6.45)	11 (9.73)	18 (6.50)	
Number of female sexual partners last month (both temporary and regular), n (%)[c]					.51[d]
0	368 (85.98)	26 (83.87)	96 (82.76)	246 (87.54)	
1	41 (9.58)	4 (12.90)	12 (10.34)	25 (8.90)	
2 or more	19 (4.44)	1 (3.23)	8 (6.90)	10 (3.56)	
Internet searches for sexual partners, n (%)[c]					.67
No	152 (36.19)	10 (30.30)	47 (38.52)	95 (35.85)	
Yes	268 (63.81)	23 (69.70)	75 (61.48)	170 (64.15)	
Have been diagnosed by a doctor with an STD[f] (eg, syphilis, genital herpes, gonorrhea, etc), n (%)[c]					.18
Yes	31 (7.03)	5 (14.71)	7 (5.74)	19 (6.67)	
No	410 (92.97)	29 (85.29)	115 (94.26)	266 (93.33)	
Commercial sex, n (%)[c]					.91[d]
Yes	15 (3.39)	1 (2.94)	5 (4.00)	9 (3.17)	
No	428 (96.61)	33 (97.06)	120 (96.00)	275 (96.83)	
Recreational drug use, n (%)[c]					.22[d]
No	435 (97.97)	34 (100.00)	119 (95.97)	282 (98.60)	
Yes	9 (2.03)	0 (0.00)	5 (4.03)	4 (1.40)	
What would be the attitude of the male sexual partner if they knew you were using PrEP?, n (%)					.04
Negative	43 (9.64)	3 (8.82)	18 (14.29)	22 (7.69)	
Neutral	181 (40.58)	16 (47.06)	58 (46.03)	107 (37.41)	

Variables	Total (N=446)	Intermediate adherence group (n=34)	Low adherence ascending group (n=126)	High adherence decline group (n=286)	*P* value
Positive	222 (49.78)	15 (44.12)	50 (39.68)	157 (54.90)	
Do male sexual partner attitudes potentially affect your use of PrEP?, n (%)[c]					.05
No	216 (49.32)	15 (44.12)	50 (42.37)	151 (52.80)	
Neutral	89 (20.32)	5 (14.71)	34 (28.81)	50 (17.48)	
Yes	133 (30.37)	14 (41.18)	34 (28.81)	85 (29.72)	
Group, n (%)					.01
Reminder group	234 (52.47)	19 (55.88)	52 (41.27)	163 (59.99)	
No-reminder group	212 (47.53)	15 (44.12)	74 (58.73)	123 (43.01)	

[a]PrEP: preexposure prophylaxis

[b]MSM: men who have sex with men.

[c]Indicates missing data.

[d]Indicates the Fisher exact test.

[e]CNY: Chinese yuan (¥1=US $0.14).

[f]STD: sexually transmitted disease.

Predicting the Patterns of Change in Adherence

In general, only levels of adherence ≥80% are recognized as having a preventive effect [23,24]. For the "intermediate adherence group" and the "low adherence ascending group," both groups had adherence rates less than 80% and were considered to be at high risk for inadequate prevention. In our modeling, we prioritized these 2 groups together for prediction. We included all 8 variables that were significant in the univariate analysis in the data-balanced decision tree prediction model. The results of the decision tree analysis are shown in Figure 3.

In the final model, HIV knowledge score, education attainment, reminder interventions, and HIV testing were key nodes in the patterns of change in adherence. We extracted a total of 4 questions and 8 prediction rules. For example, we selected 2 of these rules for elaboration (Figure 4). If the MSM's HIV knowledge score (NODE 0) was ≥11, education level (NODE 1) was "Junior high; College or University degree or above," the MSM belonged to the reminder group (NODE 3), and the MSM had not been tested for HIV (NODE 4), the MSM had a 0.636 probability of belonging to the "low and intermediate adherence" groups. For the second prediction rule, if the MSM's HIV knowledge score (NODE 0) was <11, education level (NODE 10) was "Junior high; College or University degree or above," and the MSM did not receive mobile phone–based daily reminders (NODE 12), there was a 0.684 probability that the MSM would belong to the "low and intermediate adherence" groups.

After 10-fold cross-validation, the final prediction model had an accuracy of 75%; the classification accuracy of "low and intermediate adherence" was 78.12%, and that of "high adherence decline group" was 71.43%.

Figure 3. Analytical results of the decision tree prediction model.

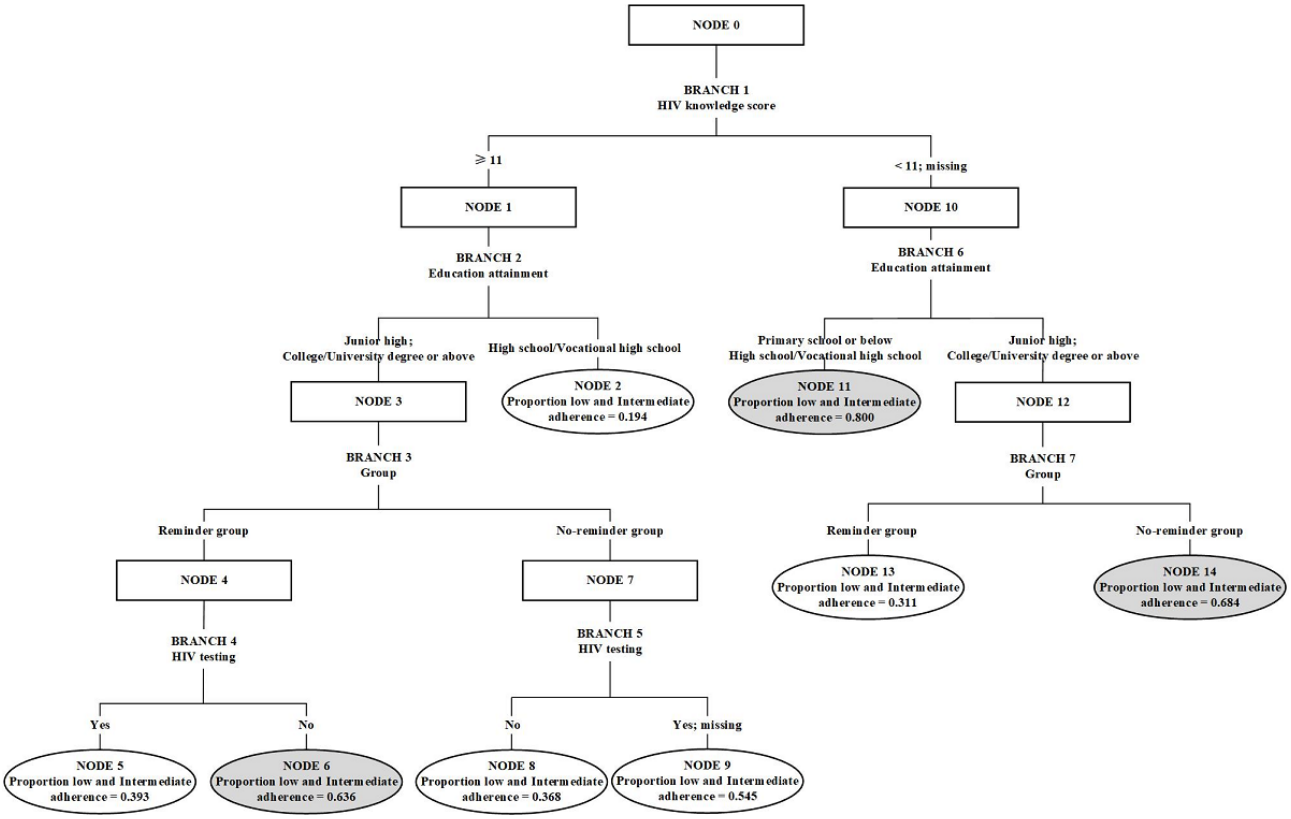
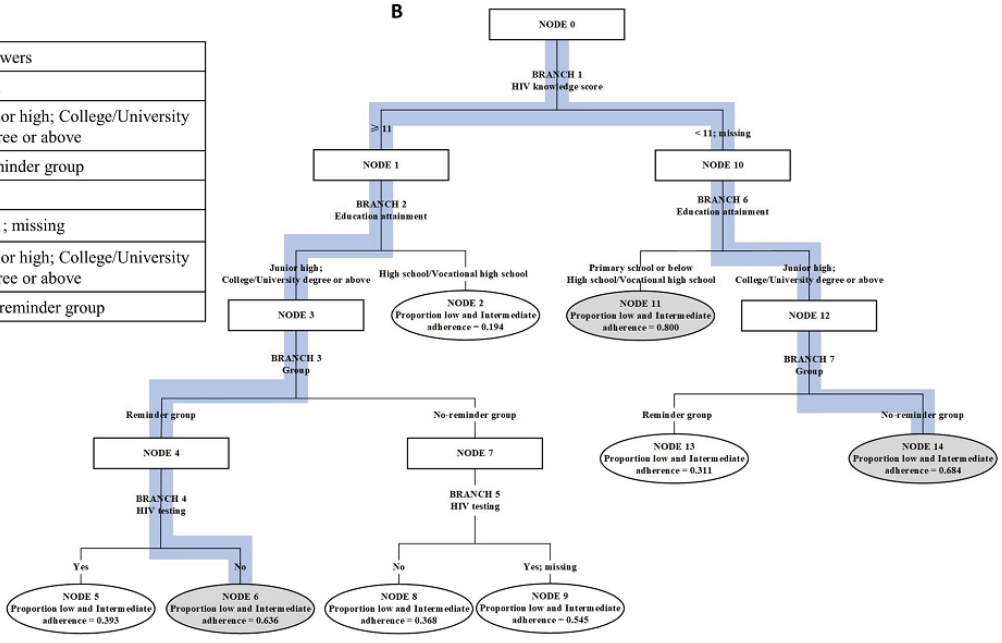


Figure 4. Two prediction rules of the decision tree prediction model. "A" indicates the description of the prediction path, and "B" indicates the prediction path in the decision tree.

A

	Questions	Answers
Rule 1	HIV knowledge score	≥ 11
	Educational attainment	Junior high; College/University degree or above
	Group	Reminder group
	HIV testing	No
Rule 2	HIV knowledge score	< 11; missing
	Educational attainment	Junior high; College/University degree or above
	Group	No-reminder group

B



Discussion

Principal Findings

Our study was one of the few in China to assess and predict patterns of change in PrEP adherence among the MSM population. The results of the decision tree prediction model showed good and stable prediction results, identifying individuals at risk for low levels of adherence. At the same time, our research results also indirectly demonstrated the role of mHealth in PrEP adherence interventions. It revealed the great potential of digital health tools (for example, mobile phone apps or web-based) and interventions for HIV prevention and overall well-being in the MSM community.

Three Adherence Potential Categories Among MSM

We selected a 3-category scheme, in which the class probability and trajectory distribution were the most reasonable. The results

of the study showed that MSM in the "high adherence decline group" had a higher overall adherence level (mean adherence >80%) and accounted for the largest proportion (64.13%), which indicated that this PrEP strategy had a high acceptability among MSM in Western China and that most MSM could successfully adhere to PrEP. Several previous studies have also explored patterns of change in adherence. For example, in the CROPrEP project conducted in China (Beijing, Shenyang, Shenzhen, and Chongqing), the researchers identified 2 types of adherence trajectories based on a group-based trajectory model, which were named "High adherence" and "Low adherence" [25]. In a randomized controlled study based on an MSM population in the United States, the authors identified subgroups of individuals with similar adherence trajectories, also based on growth mixture modeling [8]. At the same time, the trajectory for this initial state at a high level and subsequent decline in adherence have been seen in previous studies [10,11]. Previous studies have conducted qualitative interviews with this subset of the population and found that the most consistent topic of declining adherence was about the frequency of visits [11]. Users felt unsupported and overwhelmed during the long visit intervals, thus failing to adhere to their daily medication. These findings were consistent with evidence from previous studies, which suggested that visit frequency may be associated with medication adherence [26,27]. It is suggested that appropriately increasing the frequency of visits (follow-up) and giving encouragement and support to users will help to improve adherence and avoid a decline. In summary, our study used a person-centered approach based on the GMM that took into account individual differences and presented trajectories based on individual characteristics. Identifying these subgroups of discrete individuals who follow similar adherence trajectories can help identify those at risk for nonadherence who may require increased intervention support. We extended the literature on PrEP adherence by assessing patterns of change in PrEP adherence.

Decision Tree to Predict Patterns of Change in Adherence

Over the past few years, machine learning methods have been increasingly used in a variety of fields and are equally feasible in small samples [28,29]. The MSM population is a sexual minority group and attaches greater importance to privacy issues, and we were unable to conduct a large-scale survey to get a large sample of data. Therefore, decision tree-based machine learning prediction method was more suitable for our study population. In addition, we combined the "intermediate adherence group" and "low adherence ascending group" identified by the GMM for prediction and prioritized the classification accuracy of this group. This decision was made to avoid incorrect negative categorization, which could result in MSM individuals with low adherence not receiving targeted interventions and guidance, increasing the risk of HIV infection. Indeed, many researchers advocated prioritizing sensitivity (classification accuracy for low and intermediate adherence groups) over specificity to ensure that the largest possible proportion of at-risk individuals were classified [30]. However, it should be noted that in our study, the classification accuracy of the "high adherence decline group" was 71.43%, which is acceptable in terms of specificity in addition to good sensitivity. Our study constructed a predictive model for PrEP adherence among the MSM population based on the decision tree and achieved good predictive results. We identified short questions and prediction rules that will facilitate the identification of individuals at risk for low adherence among PrEP users and targeted interventions that will be beneficial in improving adherence levels in high-risk populations.

Factors Influencing Patterns of Change in Adherence Among MSM

At the same time, decision tree also provided insight and understanding of the predictor variables of the data. The variable in the root node was the most influential variable in the categorization of observations, while the other nodes contained variables that have influence on a subset of the data [31]. In the final model of the decision tree, we found that the HIV knowledge score (root node) was the most influential predictor variable, followed by educational attainment, group (reminder and nonreminder groups), and HIV testing. The results showed a significant correlation between HIV knowledge and adherence among the MSM population. According to the Information-Motivation-Behavioral Skills Model, knowledge had a significant impact on individual behavior change and guided HIV risk reduction interventions [32], which provided a theoretical explanation for our findings.

Our findings suggested that the mHealth intervention would be beneficial in improving PrEP medication adherence in the MSM population, which was also consistent with previous studies [33]. A daily mobile phone–based reminder is a promising intervention based on the actual needs and wishes of MSM. The importance of serving the unique needs and preferences of this marginalized population based on mHealth technology was also highlighted in a study in Nepal MSM [16]. In future studies, more multifaceted features and more reliable data collection approaches should be considered for mobile phone–based interventions. mHealth interventions should prioritize privacy and confidentiality, be intuitive, be user-friendly, and be tailored to the specific needs and preferences of MSM, thereby optimizing the functionality and acceptability of these mHealth care interventions.

Furthermore, Mujugira et al [34] mentioned in their study that HIV self-testing and PrEP are supplementary tools that can empower individuals to take control of their HIV protection. Regular HIV testing was recommended as part of the PrEP implementation guidelines. Our findings also suggested that HIV testing could provide an opportunity to promote PrEP adherence in the Chinese MSM population. Based on the results of the decision tree model, interventions targeting knowledge, medication reminders, and HIV testing would be beneficial in improving adherence in the MSM population.

Limitations

Our study was a preliminary application of decision tree prediction modeling to PrEP adherence among MSM population in Western China. However, there are some limitations in our study. First, we measured adherence through participants' self-reports, which may have been influenced by their recall

bias, social desirability bias, and thus overestimation of adherence. However, a significant correlation between self-reported adherence and medication levels was found in a study of the PrEP program in a clinical setting, suggesting that self-reporting can be used to predict PrEP adherence [35]. Second, participants in our mHealth trial were rarely blinded, as they may have interacted with other participants in the survey field, thereby learning about the intervention, which is a typical shortcoming in mHealth trials. Although the results of our additional analyses indicated that there were no statistically significant differences between the intervention and control groups (Table S1 in Multimedia Appendix 1), the unblinded categorization of the 2 groups could have led to a potential impact on the results of the study. Future mHealth-based studies should fully consider the issue of blinding of interventions so as to ensure the accuracy of the findings. Meanwhile, there were the significant differences in baseline characteristics between included and excluded participants (Table S3 in Multimedia Appendix 1). According to the results of our analysis, age, place of residence, and employment status were statistically different in the inclusion and exclusion groups. The same shortcomings have been mentioned in previous studies [36]. The sample selected in their study was predominantly white and highly educated, which may have led to high estimates of self-reported adherence. However, we and previous researchers likewise agreed that such a bias was acceptable because these significantly different baseline characteristics were not the most significant factor influencing adherence. Meanwhile, for the evaluation metrics of the prediction model, we only selected 3 standard metrics and did not further adopt more metrics for assessment. However, the 10-fold cross-validation results had demonstrated the stability of the prediction results, and for small-sample machine learning predictions, we believed that the current metrics were sufficiently persuasive, and previous studies had used such an evaluation [22]. Since the MSM population is sensitive and concealed, this made our investigation process very difficult. Therefore, we did not have enough time to conduct external validation in this study. In future studies, we will validate the model more fully if the opportunity is available. Finally, further future research should be applied to larger samples while inputting more different variables so as to explore the accuracy of the predictive model.

Conclusions

Awareness and utilization of PrEP are increasing in China, and information to support improved PrEP adherence is becoming more important. Our study modeled the longitudinal trajectory of adherence among the MSM population in Western China and predicted patterns of change in adherence. Accurately identifying patterns of change in adherence based on the decision tree and its predictive pathways can help administrators increase their focus on individuals who need preventive interventions that are conducive to improving PrEP adherence. This study was an important first step in implementing predictors of adherence among the MSM population using PrEP. More specifically, we proposed a short set of questions and prediction rules that could hopefully be assessed in future large-scale validation studies to prospectively predict patterns of change in adherence in high-risk populations. Meanwhile, interventions targeting knowledge, mHealth reminders, and HIV testing will also be beneficial in improving adherence among the MSM population. Future studies should continue to investigate the dynamic trajectories of adherence, assess patterns of change more frequently, and develop more accurate predictive models to better understand PrEP adherence in high-risk populations in the current environmental context.

Acknowledgments

This work was supported by the National Key Project for Infectious Diseases of the Ministry of Science and Technology of China (2018ZX10721102-005).

Authors' Contributions

BL was responsible for the analysis of the data, interpretation of the results, and the draft and revision of the manuscript. XZ, J Li, and J Liu advised on the analysis approach, interpretation of the results, and the revision of the manuscript. BL, WH, and HP were responsible for data collection and data cleaning. XZ was responsible for funding acquisition and project administration and participated in the study conceptualization and supervision.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary materials regarding presentation of the reminder messages on WeChat, comparison of basic information between the reminder and no-reminder groups, descriptive analysis of adherence, and univariate analysis of inclusion and exclusion groups.
[DOCX File, 249 KB - mhealth_v12i1e58920_app1.docx]

Checklist 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File, 1210 KB - mhealth_v12i1e58920_app2.pdf]

References

1. Dong MJ, Peng B, Liu ZF, et al. The prevalence of HIV among MSM in China: a large-scale systematic analysis. BMC Infect Dis 2019 Dec;19(1):1000. [doi: 10.1186/s12879-019-4559-1]
2. He L, Jiang T, Chen W, et al. Examining HIV testing coverage and factors influencing first-time testing among men who have sex with men in Zhejiang Province, China: cross-sectional study based on a large internet survey. JMIR Public Health Surveill 2024 Jun 14;10:e56906. [doi: 10.2196/56906] [Medline: 38875001]
3. Davies O, Ustianowski A, Fox J. Pre-exposure prophylaxis for hiv prevention: why, what, who and how. Infect Dis Ther 2016 Dec;5(4):407-416. [doi: 10.1007/s40121-016-0128-8] [Medline: 27677264]
4. Sun Z, Gu Q, Dai Y, et al. Increasing awareness of HIV pre-exposure prophylaxis (PrEP) and willingness to use HIV PrEP among men who have sex with men: a systematic review and meta-analysis of global data. J Int AIDS Soc 2022 Mar;25(3):e25883. [doi: 10.1002/jia2.25883] [Medline: 35255193]
5. Grant RM, Lama JR, Anderson PL, et al. Preexposure chemoprophylaxis for HIV prevention in men who have sex with men. N Engl J Med 2010 Dec 30;363(27):2587-2599. [doi: 10.1056/NEJMoa1011205]
6. Musinguzi N, Muganzi CD, Boum Y, et al. Comparison of subjective and objective adherence measures for preexposure prophylaxis against HIV infection among serodiscordant couples in East Africa. AIDS 2016;30(7):1121-1129. [doi: 10.1097/QAD.0000000000001024]
7. O Murchu E, Marshall L, Teljeur C, et al. Oral pre-exposure prophylaxis (PrEP) to prevent HIV: a systematic review and meta-analysis of clinical effectiveness, safety, adherence and risk compensation in all populations. BMJ Open 2022 May 11;12(5):e048478. [doi: 10.1136/bmjopen-2020-048478] [Medline: 35545381]
8. Pasipanodya EC, Jain S, Sun X, et al. Trajectories and predictors of longitudinal preexposure prophylaxis adherence among men who have sex with men. J Infect Dis 2018 Oct 5;218(10):1551-1559. [doi: 10.1093/infdis/jiy368] [Medline: 30295803]
9. Jin F, Amin J, Guy R, et al. Adherence to daily HIV pre-exposure prophylaxis in a large-scale implementation study in New South Wales, Australia. AIDS 2021;35(12):1987-1996. [doi: 10.1097/QAD.0000000000002970]
10. Pyra M, Brewer R, Rusie L, Kline J, Willis I, Schneider J. Long-term HIV pre-exposure prophylaxis trajectories among racial & ethnic minority patients: short, declining, & sustained adherence. J Acquir Immune Defic Syndr 2022;89(2):166-171. [doi: 10.1097/QAI.0000000000002833]
11. Stoner MCD, Rucinski KB, Giovenco D, et al. Trajectories of PrEP adherence among young women aged 16 to 25 in Cape Town, South Africa. AIDS Behav 2021 Jul;25(7):2046-2053. [doi: 10.1007/s10461-020-03134-3] [Medline: 33389323]
12. Ryo M, Rillig MC. Statistically reinforced machine learning for nonlinear patterns and variable interactions. Ecosphere 2017 Nov;8(11):11 [FREE Full text] [doi: 10.1002/ecs2.1976]
13. Mazumdar M, Lin JYJ, Zhang W, et al. Comparison of statistical and machine learning models for healthcare cost data: a simulation study motivated by Oncology Care Model (OCM) data. BMC Health Serv Res 2020 Dec;20(1):350. [doi: 10.1186/s12913-020-05148-y]
14. Gagnon MP, Ngangue P, Payne-Gagnon J, Desmartis M. m-Health adoption by healthcare professionals: a systematic review. J Am Med Inform Assoc 2016 Jan 1;23(1):212-220. [doi: 10.1093/jamia/ocv052]
15. Shrestha R, Altice FL, Khati A, et al. Clinic-integrated smartphone app (JomPrEP) to improve uptake of HIV testing and pre-exposure prophylaxis among men who have sex with men in Malaysia: mixed methods evaluation of usability and acceptability. JMIR Mhealth Uhealth 2023 Feb 16;11:e44468. [doi: 10.2196/44468] [Medline: 36795465]
16. Gautam K, Paudel K, Ahmed A, et al. High interest in the use of mHealth platform for HIV prevention among men who have sex with men in Nepal. J Community Health 2024 Aug;49(4):575-587. [doi: 10.1007/s10900-024-01324-x] [Medline: 38281283]
17. Ge Y, Yao D, Ung COL, et al. Digital medical information services delivered by pharmaceutical companies via WeChat: qualitative analytical study. J Med Internet Res 2023;25:e43812. [doi: 10.2196/43812]
18. DiClemente RJ, Zorn J, Temoshok L. Adolescents and AIDS: a survey of knowledge, attitudes and beliefs about AIDS in San Francisco. Am J Public Health 1986 Dec;76(12):1443-1445. [doi: 10.2105/AJPH.76.12.1443]
19. Lin B, Liu J, Ma Y, Zhong X. Factors influencing HIV testing and counselling services among men who have sex with men in Western China: a cross-sectional study based on Andersen's Behavioral Model. Environ Health Prev Med 2022;27(0):21-21. [doi: 10.1265/ehpm.22-00066]
20. Lin B, Liu J, Zhong X. PrEP risk perception and adherence among men who have sex with men: a prospective cohort study based on growth mixture model. BMC Infect Dis 2022 Dec 30;22(1):969. [doi: 10.1186/s12879-022-07966-3] [Medline: 36585642]
21. Goedel WC, Schneider JA, Hambrick HR, et al. Are anal sex roles associated with preferences for pre-exposure prophylaxis administration modalities among men who have sex with men? Arch Sex Behav 2018 Oct;47(7):2123-2133. [doi: 10.1007/s10508-017-1083-5]
22. Rousseau S, Polachek IS, Frenkel TI. A machine learning approach to identifying pregnant women's risk for persistent post-traumatic stress following childbirth. J Affect Disord 2022 Jan 1;296:136-149. [doi: 10.1016/j.jad.2021.09.014] [Medline: 34601301]

23. Haberer JE, Baeten JM, Campbell J, et al. Adherence to antiretroviral prophylaxis for HIV prevention: a substudy cohort within a clinical trial of serodiscordant couples in East Africa. PLoS Med 2013;10(9):e1001511. [doi: 10.1371/journal.pmed.1001511] [Medline: 24058300]
24. Wu D, Tao H, Dai J, Liang H, Huang A, Zhong X. Study on pre-exposure prophylaxis regimens among men who have sex with men: a prospective cohort study. Int J Environ Res Public Health 2019 Dec 9;16(24):4996. [doi: 10.3390/ijerph16244996] [Medline: 31818006]
25. Chen S, Zhu YY, Chu ZX, et al. Trajectories of anxiety and depression among Chinese men who have sex with men on pre-exposure prophylaxis: a group-based trajectory model approach. BMC Public Health 2024;24(1):365-365. [doi: 10.1186/s12889-024-17854-x]
26. Hosek SG, Rudy B, Landovitz R, et al. An HIV preexposure prophylaxis demonstration project and safety study for young MSM. J Acquir Immune Defic Syndr 2017;74(1):21-29. [doi: 10.1097/QAI.0000000000001179]
27. Hosek SG, Landovitz RJ, Kapogiannis B, et al. Safety and feasibility of antiretroviral preexposure prophylaxis for adolescent men who have sex with men aged 15 to 17 years in the United States. JAMA Pediatr 2017 Nov 1;171(11):1063-1071. [doi: 10.1001/jamapediatrics.2017.2007] [Medline: 28873128]
28. Aafjes-van Doorn K, Kamsteeg C, Bate J, Aafjes M. A scoping review of machine learning in psychotherapy research. Psychother Res 2021 Jan;31(1):92-116. [doi: 10.1080/10503307.2020.1808729] [Medline: 32862761]
29. Mak KK, Lee K, Park C. Applications of machine learning in addiction studies: a systematic review. Psychiatry Res 2019 May;275:53-60. [doi: 10.1016/j.psychres.2019.03.001] [Medline: 30878857]
30. Lewis KC, Good EW, Tillman JG, Hopwood CJ. Assessment of psychological pain in clinical and non-clinical samples: a preliminary investigation using the Psychic Pain Scale. Arch Suicide Res 2021;25(3):552-569. [doi: 10.1080/13811118.2020.1729914] [Medline: 32089105]
31. Grabauskytė I, Tamošiūnas A, Kavaliauskas M, Radišauskas R, Bernotienė G, Janilionis V. A comparison of decision tree induction with binary logistic regression for the prediction of the risk of cardiovascular diseases in adult men. Inf 2019;29(4):675-692. [doi: 10.15388/Informatica.2018.187]
32. Qu D, Zhong X, Xiao G, Dai J, Liang H, Huang A. Adherence to pre-exposure prophylaxis among men who have sex with men: a prospective cohort study. Int J Infect Dis 2018 Oct;75:52-59. [doi: 10.1016/j.ijid.2018.08.006] [Medline: 30125688]
33. Baxi SM, Liu A, Bacchetti P, et al. Comparing the novel method of assessing PrEP adherence/exposure using hair samples to other pharmacologic and traditional measures. J Acquir Immune Defic Syndr 2015;68(1):13-20. [doi: 10.1097/QAI.0000000000000386]
34. Mujugira A, Nakyanzi A, Nabaggala MS, et al. Effect of HIV self-testing on PrEP adherence among gender-diverse sex workers in Uganda: a randomized trial. J Acquir Immune Defic Syndr 2022;89(4):381-389. [doi: 10.1097/QAI.0000000000002895]
35. Montgomery MC, Oldenburg CE, Nunn AS, et al. Adherence to pre-exposure prophylaxis for HIV prevention in a clinical setting. PLoS One 2017;11(6):e0157742. [doi: 10.1371/journal.pone.0157742]
36. Monteiro Spindola Marins L, Silva Torres T, Luz PM, et al. Factors associated with self-reported adherence to daily oral pre-exposure prophylaxis among men who have sex with man and transgender women: PrEP Brasil study. Int J STD AIDS 2021 Nov;32(13):1231-1241. [doi: 10.1177/09564624211031787] [Medline: 34311605]

Abbreviations

aBIC: sample size–adjusted Bayesian information criteria
AIC: Akaike information criteria
BIC: Bayesian information criteria
Cp: complexity parameter
GMM: growth mixture model
mHealth: mobile health
MSM: men who have sex with men
PrEP: preexposure prophylaxis

Edited by L Buis; submitted 27.03.24; peer-reviewed by C Fisher, C Lei; revised version received 26.08.24; accepted 14.10.24; published 12.12.24.

Please cite as:
Lin B, Li J, Liu J, He W, Pan H, Zhong X
Exploring and Predicting HIV Preexposure Prophylaxis Adherence Patterns Among Men Who Have Sex With Men: Randomized Controlled Longitudinal Study of an mHealth Intervention in Western China
JMIR Mhealth Uhealth 2024;12:e58920
URL: https://mhealth.jmir.org/2024/1/e58920
doi:10.2196/58920

User Experience of Persons Using Ingestible Sensor–Enabled Pre-Exposure Prophylaxis to Prevent HIV Infection: Cross-Sectional Survey Study

Sara Browne[1,2*], MD, MPH; Anya Umlauf[3*], MS; David J Moore[3], PhD; Constance A Benson[1], MD; Florin Vaida[4], PhD

[1]Division of Infectious Diseases and Global Public Health, University of California San Diego, San Diego, CA, United States

[2]Specialists in Global Health, Encinitas, CA, United States

[3]Department of Psychiatry, University of California San Diego, San Diego, CA, United States

[4]Department of Family Medicine and Public Health, University of California San Diego, San Diego, CA, United States

[*]these authors contributed equally

Corresponding Author:
Sara Browne, MD, MPH
Division of Infectious Diseases and Global Public Health, University of California San Diego, , San Diego, CA, , United States

Abstract

Background: A digital health technology's success or failure depends on how it is received by users.

Objectives: We conducted a user experience (UX) evaluation among persons who used the Food and Drug Administration–approved Digital Health Feedback System incorporating ingestible sensors (ISs) to capture medication adherence, after they were prescribed oral pre-exposure prophylaxis (PrEP) to prevent HIV infection. We performed an association analysis with baseline participant characteristics, to see if "personas" associated with positive or negative UX emerged.

Methods: UX data were collected upon exit from a prospective intervention study of adults who were HIV negative, prescribed oral PrEP, and used the Digital Health Feedback System with IS-enabled tenofovir disoproxil fumarate plus emtricitabine (IS-Truvada). Baseline demographics; urine toxicology; and self-report questionnaires evaluating sleep (Pittsburgh Sleep Quality Index), self-efficacy, habitual self-control, HIV risk perception (Perceived Risk of HIV Scale 8-item), and depressive symptoms (Patient Health Questionnaire–8) were collected. Participants with ≥28 days in the study completed a Likert-scale UX questionnaire of 27 questions grouped into 4 domain categories: overall experience, ease of use, intention of future use, and perceived utility. Means and IQRs were computed for participant total and domain subscores, and linear regressions modeled baseline participant characteristics associated with UX responses. Demographic characteristics of responders versus nonresponders were compared using the Fisher exact and Wilcoxon rank-sum tests.

Results: Overall, 71 participants were enrolled (age: mean 37.6, range 18-69 years; n=64, 90% male; n=55, 77% White; n=24, 34% Hispanic; n=68, 96% housed; and n=53, 75% employed). No demographic differences were observed in the 63 participants who used the intervention for ≥28 days. Participants who completed the questionnaire were more likely to be housed (52/53, 98% vs 8/10, 80%; *P*=.06) and less likely to have a positive urine toxicology (18/51, 35% vs 7/10, 70%; *P*=.08), particularly methamphetamine (4/51, 8% vs 4/10, 40%; *P*=.02), than noncompleters. Based on IQR values, ≥75% of participants had a favorable UX based on the total score (median 3.78, IQR 3.17-4.20), overall experience (median 4.00, IQR 3.50-4.50), ease of use (median 3.72, IQR 3.33-4.22), and perceived utility (median 3.72, IQR 3.22-4.25), and ≥50% had favorable intention of future use (median 3.80, IQR 2.80-4.40). Following multipredictor modeling, self-efficacy was significantly associated with the total score (0.822, 95% CI 0.405-1.240; *P*<.001) and all subscores (all *P*<.05). Persons with more depressive symptoms reported better perceived utility (*P*=.01). Poor sleep was associated with a worse overall experience (−0.07, 95% CI −0.133 to −0.006; *P*=.03).

Conclusions: The UX among persons using IS-enabled PrEP (IS-Truvada) to prevent HIV infection was positive. Association analysis of baseline participant characteristics linked higher self-efficacy with positive UX, more depressive symptoms with higher perceived utility, and poor sleep with negative UX.

Trial Registration: ClinicalTrials.gov NCT03693040; https://clinicaltrials.gov/study/NCT03693040

(JMIR Mhealth Uhealth 2024;12:e53596) doi:10.2196/53596

KEYWORDS

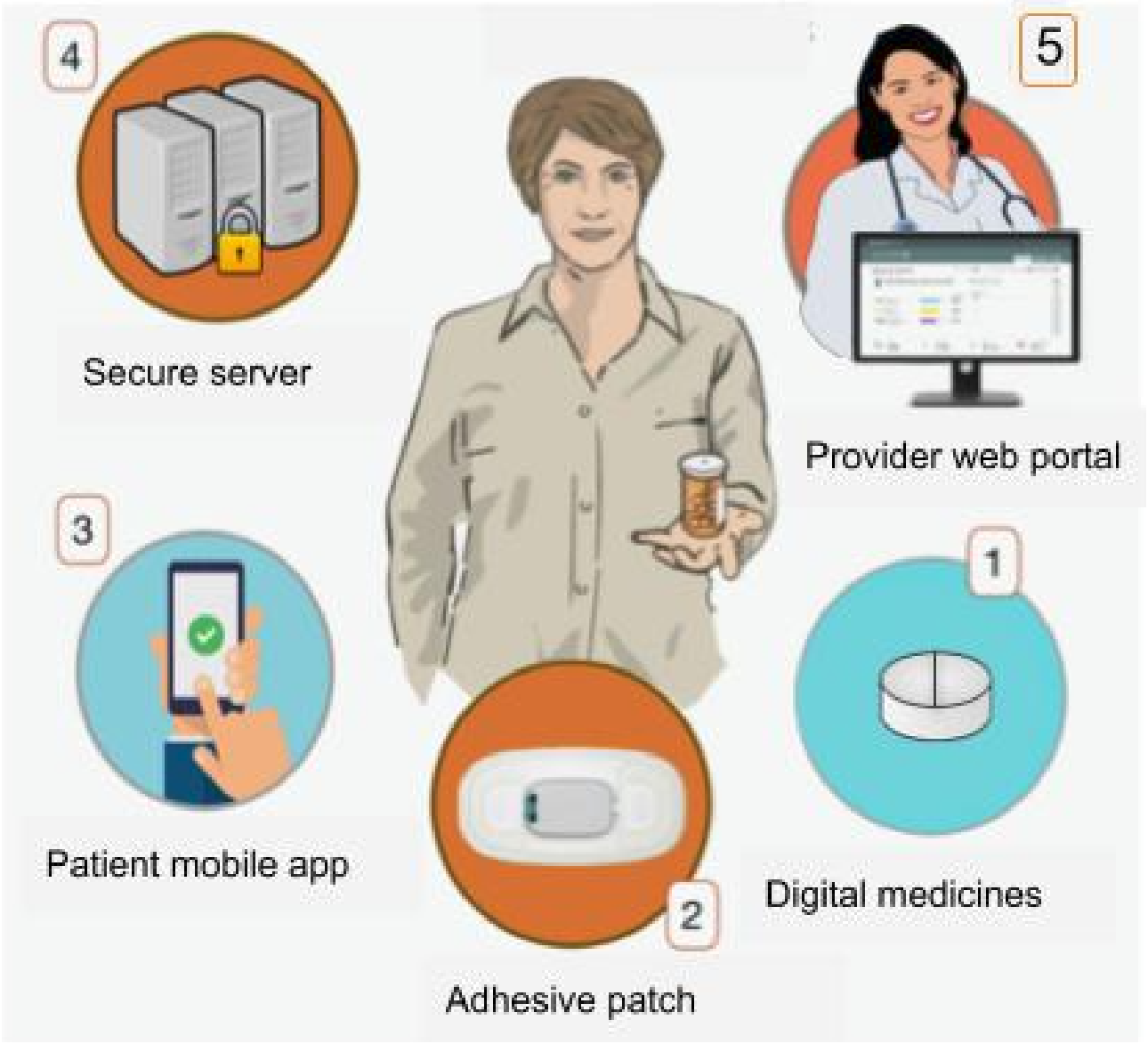
ingestible sensor; sensor; sensors; oral; UX; user experience; HIV prevention; medication adherence; HIV; prevention; prophylaxis; STI; STD; sexually transmitted; sexual transmission; drug; drugs; pharmacy; pharmacies; pharmacology; pharmacotherapy;

pharmaceutic; pharmaceutics; pharmaceuticals; pharmaceutical; medication; medications; adherence; compliance; sexually transmitted infection; sexually transmitted disease

Introduction

The first ingestible sensor (IS) technology to capture oral medication adherence was approved by the Food and Drug Administration (FDA) in 2015 [1], followed by the approval of a second sensor variety in 2019 [2]. The major advancement associated with these medical devices is their capture of remote real-time data on actual drug ingestion, in some cases with simultaneous physiological data [3]. These novel digital technologies, in addition to accurate oral dose ingestion confirmation, may also allow bidirectional treatment adherence support [4-6]. The FDA-approved Digital Health Feedback System (DHFS) consists of an IS, external wearable patch, and paired mobile device [1]. It detects and records the timing of ingestion events and physiologic measures [3], which are then automatically uploaded to a secure internet server, allowing patients and health care providers to follow medication taking in real time and facilitate patient-provider communication (see Figure 1) [1,3-6].

Figure 1. Depiction of the components of the Digital Health Feedback System (DHFS). (1) At home, the patient takes the digitized medicine. The ingestible sensor activates in the stomach, and its serial number is captured and stored by the patch. (2) Patch data are transferred by Bluetooth to an app on the patient's paired mobile device. (3) Patients can follow their own medication taking (the DHFS has the capacity to send patients tailored automated reminder messages). (4) Data are transferred to secure servers. (5) Patient-approved health care workers can remotely receive real-time treatment adherence data and follow large cohorts of patients using the secure web-based dashboard.



IS-enabled tenofovir disoproxil fumarate plus emtricitabine (IS-Truvada) with the DHFS has recently been deployed to capture medication adherence behavior in persons starting oral pre-exposure prophylaxis (PrEP) [7]. This study confirmed that the DHFS is highly accurate, providing valid measures of ingestion with 99.3% (95% CI 97.2%-100%) reliability versus direct observation, similar to findings in a population of patients with tuberculosis (TB) [6,7]. Clinical trials evaluating DHFS

use in TB and hepatitis C treatments have demonstrated persistence use, efficacy, and even superiority versus direct observation [5,6]. Adverse events were few and mild, involving skin reactions to the patch, similar to reports from studies in chronic cardiometabolic disease management [3,8].

In the arena of HIV prevention, both providers and patients have described concerns with oral PrEP medication adherence as a barrier to successful implementation [9,10]. However, there are still no highly accurate real-time adherence measurement tools used in clinical practice to guide patient and physician discussions around these concerns. Providers continue to depend principally on self-reported adherence, which is subject to recall bias [11-13] and patient's efforts to avoid potential negative interactions with their physicians when disclosing nonadherence [14,15], or pharmacy refills, which indicate what a patient has on hand but provide no information on if and when medications are taken [16]. Directly observed therapy, which is primarily used in TB treatment, where a person is observed taking their medication, is highly reliable but personnel and resource intensive, time-consuming, costly [6,17,18], and impractical in chronic management for HIV prevention.

Multiple earlier technologies that supply data on oral medication adherence are available. Medication event monitoring system devices use openings of electronic containers and lids but have well-documented inaccuracies based on mismatches between openings and actual pill taking [19-24]. Smartphone apps incorporating SMS text messaging are based on self-report [25] or send videos [26] for later viewing and assessment. In contrast to these surrogate technologies, ISs signal when medication reaches the gastric track and are able to capture individual daily behavior patterns in real time, providing insight into variations in daily medication adherence. This capacity even has advantages over cumulative metabolite-based adherence measures developed and evaluated in the HIV prevention arena, such as dried blood spots and hair analysis [27,28], which do not allow real-time intervention or capture pattern variations in medication ingestion over time [7]. Such patterns are of importance based on the postdose durations of the therapeutic drug [29,30], which alters the risk of acquiring HIV infection, for example, a week where PrEP is taken once, followed by a week where it is taken daily [31-33].

Regardless of the superior capacities of the DHFS, the success or failure of any digital health innovation often depends on how it is received by the user [34]. Limited medical research exists on the user experience (UX) of persons using IS-based digital technology. UX is considered crucial to product design [34]. Classical consumer research ranks products according to levels of utility to consumers, which are subjective individual tastes; however, individuals change over time, and effort is devoted to developing an understanding of current and future user "personas" [35]. In contrast, traditional medical research, particularly in the infectious disease arena, historically looks for programmatic implementation of adherence technology, with the implicit assumption that "one size" should or could "fit all." The development of technology acceptance models has underscored the importance of understanding how personal attitudes contribute to behavioral intention on technology use [36-41], particularly how perceived ease of use and utility influences individuals' willingness to adopt and continue to use a given technology [39]. Research on the contribution of personal attitudes and characteristics is now expanding into health care technology use [42] but is entirely novel in the arena of digital adherence technology.

We conducted a detailed UX evaluation among persons prescribed PrEP to prevent HIV infection who used the DHFS with IS-Truvada. Our study evaluated the overall experience, ease of use, perceived utility, and intention of future use. We captured baseline demographics and urine toxicology screening (UTOX). In addition, we collected detailed self-report questionnaires to evaluate depressive symptoms, HIV risk perception, sleep, and individual self-efficacy in the context of medication taking [43]. Self-efficacy is defined as a person's belief in their capability to succeed and achieve a given level of performance [44] and is considered to be connected to motivation, achievement, and self-regulation [45,46]. We then conducted an association analysis of our UX findings with individual participant characteristics captured at baseline, to see if current and future user "personas" emerged.

Methods

Overview

UX data were collected upon exit from a prospective, single-arm, open-label intervention study of participants using the DHFS (manufacturers: Proteus Digital Health and Otsuka Pharma) with IS-enabled tenofovir disoproxil fumarate plus emtricitabine for up to 12 weeks. The parent intervention study evaluated DHFS adherence measurements, ability to capture patterns of adherence behavior, and the association of predictors with adherence behavior among persons starting PrEP [7].

Ethical Considerations

The study protocol was approved by the University of California San Diego (UCSD) Institutional Review Board (#161618), was conducted in accordance with Good Clinical Practice principles, and was registered on ClinicalTrials.gov (NCT03693040). Participants signed an informed consent. All data derived from this study were deidentified. Participants were compensated with a gift card equivalent in value to US $50 on the completion of all activities associated with the baseline and study exit visits, which included survey completion.

Participants

Eligible participants were HIV and hepatitis B seronegative, aged ≥18 years old, were at risk for HIV, and desired oral PrEP. Participants were recruited from the UCSD AntiViral Research Center, UCSD Owen Clinic, or other primary care clinics in San Diego. Participant procedures were as follows. Baseline laboratory evaluations were required within the defined parameters; participants needed to be able to use mobile devices (these were provided by the study if they did not have them), be willing to use the DHFS, and have no known skin adhesive hypersensitivity. Baseline demographics, UTOX, and self-reported questionnaires were collected. Participants were instructed on DHFS use at baseline; this instruction comprised how to place and change the patch, how to pair the patch with the mobile device, and how to connect the mobile device to

Wi-Fi. During the trial, participants changed the monitor patch themselves as needed and could view the medication ingestion log on their mobile device. Study staff counseled participants on wearing the patch and keeping their paired mobile device consistently charged. After the intervention, participants underwent repeat HIV testing and continued on PrEP as prescribed by their practitioner. Participants with ≥28 days in study (DHFS with IS-Truvada) completed the detailed exit questionnaire and formed the cohort analyzed.

Measures

Baseline self-report questionnaires evaluated habitual self-control [47], self-efficacy beliefs [43], depression (Patient Health Questionnaire–8 [PHQ-8]) [48], sleep (Pittsburgh Sleep Quality Index [PSQI]) [49], and HIV risk perception (Perceived Risk of HIV Scale [PRHS] 8-item) [50]. On study exit, the detailed UX questionnaire was completed. The UX questionnaire consisted of 27 questions with responses coded from 1 to 5 and included reverse scored and related questions to ensure validity [51]. Of the items scored on the Likert scale, 2 questions assessed satisfaction, from 1=extremely unsatisfied to 5=extremely satisfied; 5 questions asked participants to rate various aspects of the system, from 1=extremely unhelpful to 5=extremely helpful; and the responses to the remaining questions ranged from 1=strongly disagree to 5=strongly agree. The UX questions were grouped into 4 domain categories: overall experience, ease of use, intention of future use, and perceived utility. Textbox 1 shows the questions, domains, and the number of questions per domain.

Textbox 1. User experience questionnaire items for the Digital Health Feedback System (DHFS). Domain categories are shown, with questions grouped by category, not in the originally administered order.

Overall experience

1. How would you rate your overall experience with participating in this mediation adherence study?
2. How would you rate your overall satisfaction with the DHFS (the iPad and patch system)?
3. Overall, this experience using the DHFS was positive.
4. Overall, this experience using the DHFS was challenging. (reverse scored)

Ease of use

5. When you started the study, how helpful was the Patient Information Booklet?
6. How helpful were the Proteus app instructions?
7. I was very comfortable changing the patch on my own.
8. The instructions for changing the patch were easy to follow.
9. The patch is comfortable to wear.
10. Wearing the patch interfered with my daily activities. (reverse scored)
11. The Proteus app was difficult to navigate. (reverse scored)
12. Accessing my medication ingestion report was difficult. (reverse scored)
13. Technical difficulties were easily resolved.

Intention of future use

14. I would use the DHFS in the future.
15. I would use the DHFS in the future to keep track of my treatment.
16. I would use the DHFS in the future if I had problems following my treatment.
17. I would recommend that others use the DHFS.
18. I would recommend use of the DHFS to others if they are having problems following their treatment.

Perceived utility

19. How helpful was participating in the study for your medication adherence?
20. How helpful was the DHFS in helping you follow your medication adherence?
21. I used the DHFS app frequently to follow my medication taking.
22. I used the DHFS app frequently to follow my activity and rest.
23. The DHFS was useful.
24. The DHFS made taking my medication easier for me.
25. Using the DHFS interfered with how I typically manage my medications. (reverse scored)
26. Referring to the Proteus app during the study helped me track my medication adherence.
27. The DHFS improved my medication adherence.

Statistical Analysis

A total of 27 UX questionnaire item responses were scored (using reverse scoring where necessary), so that higher scores mean higher levels of satisfaction. The questions' average values were used for the total summary score and each of the 4 domain summary subscores. Demographic characteristics of participants who completed the UX questionnaire were compared to those of nonresponders using the Wilcoxon rank-sum test and Fisher exact test as appropriate.

Single- and multipredictor linear models were used to analyze the 5 summary scores for their association with demographic and other baseline characteristics, including age, gender, race and ethnicity (non-Hispanic White, Asian, Black, or Hispanic), UTOX results (positive or negative), number of substances detected in UTOX, sleep (PSQI), self-efficacy, habitual self-control, HIV risk perception (PRHS 8-item), and depression (PHQ-8). Prior to regression analyses, the multiple imputation by chained equations method with 10 imputations was used to impute missing values in the predictors. All model estimates were based on pooled results, using the rules from Rubin [52]. Predictors with *P* values <.20 in the univariable analyses were included for consideration into multivariable analyses. Backward model selection with a .20 threshold for predictor significance was used to select the final multivariable models. The *Result*s section reports model coefficients, their 95% CIs, and the relevant *P* values. The CIs for effects of ethnicity were Bonferroni-adjusted for multiple comparisons. PHQ-8 exhibited nonlinear association with the summary total score and subscores. Natural cubic splines, with the number of knots determined by minimizing the Akaike information criterion, were used to model these associations. Analyses and figures were done using R (version 4.0.3; R Foundation for Statistical Computing) [53]. The R package *mice* was used for multiple imputation analyses [54].

Results

Study Cohort Description

Study Enrollment, Demographics, and Cohort Description

Overall, 71 persons were enrolled in the intervention using the DHFS with IS-Truvada. Participants had a mean age of 37.6 (range 18-69) years and were mostly male (n=64, 90%), White (n=55, 77%; n=24, 34% were Hispanic), housed (n=68, 96%), and employed (n=53, 75%). Baseline toxicology was positive in 41% (n=28) of participants, with marijuana (n=17, 25%), amphetamines (n=10, 14%), and methamphetamines (n=8, 12%). A total of 63 participants used the DHFS with IS-Truvada for ≥28 days, and there were no significant differences in baseline demographics compared to enrolled participants who dropped out early (n=8) [7]. Of the 63 participants, 53 (84%) fully or partially completed the comprehensive UX exit questionnaire. Table 1 shows the demographic characteristics of participants at baseline and includes the comparison of participants who completed the UX questionnaire and those that did not. Questionnaire respondents did not differ statistically from nonrespondents on age (37.5 vs 33.9 y; *P*=.34), sex (48/53, 91% vs 9/10, 90% male; *P*>.99), employment status (39/53, 74% vs 7/10, 70% employed; *P*>.99), or race and ethnicity (30/53, 57% vs 5/10, 50% non-Hispanic White; *P*=.43). However, questionnaire respondents were more likely to have stable housing (52/53, 98% vs 8/10, 80%; *P*=.06) and less likely to test positive on UTOX (18/51, 35% vs 7/10, 70%; *P*=.08), particularly for methamphetamine (4/51, 8% vs 4/10, 40%; *P*=.02). No significant difference was observed in self-report questionnaire scores between respondents and nonrespondents (all *P*>.05).

Table . Baseline cohort characteristics and comparison between completers and noncompleters of the user experience questionnaire. *P* values are based on Wilcoxon rank-sum test (numeric variables) and Fisher exact test (categorical variables).

Variable		Completed exit survey (n=53)	Did not complete exit survey (n=10)	*P* value
Age (years), mean (SD)		37.5 (10.8)	33.9 (11.2)	.34
Gender, n (%)				>.99
	Male	48 (91)	9 (90)	
	Female or transgender	5 (9)	1 (10)	
Race and ethnicity, n (%)				.43
	Asian, non-Hispanic	4 (8)	0 (0)	
	Black, non-Hispanic	3 (6)	2 (20)	
	Hispanic	18 (30)	3 (30)	
	White, non-Hispanic	30 (57)	5 (50)	
Positive drug screen (any drugs), n (%)		18 (35)[a]	7 (70)	.08
Number of drugs identified on toxicology screen, median (IQR)		0.00 (0.00-1.00)[a]	1.00 (0.25-2.75)	.02[b]
Positive methamphetamine toxicology screen, n (%)		4 (8)[a]	4 (40)	.02[b]
Employment, n (%)				>.99
	Full or part time	39 (74)	7 (70)	
	Unemployed, retired, or disabled	14 (26)	3 (30)	
Housing, n (%)				.06
	Stable	52 (98)	8 (80)	
	Transient or homeless	1 (2)	2 (20)	
Global PSQI[c] score, mean (SD)		6.35 (3.17)[d]	5.29 (3.55)[e]	.44
Self-efficacy, mean (SD)		4.47 (0.41)[f]	4.42 (0.43)[g]	.67
Habitual self-control, mean (SD)		3.68 (0.60)[h]	3.83 (0.29)[g]	.61
HIV risk perception (PRHS[i] 8-item), mean (SD)		23.1 (5.49)[d]	24 (8.67)[g]	.50
PHQ-8[j] total (8-item), median (IQR)		3.0 (1.0-5.0)[k]	1.0 (0.0-2.0)[l]	.07

[a]n=51.

[b]*P*<.05.

[c]PSQI: Pittsburgh Sleep Quality Index (higher score=worse).

[d]n=49.

[e]n=7.

[f]n=45.

[g]n=8.

[h]n=43.

[i]PRHS: Perceived Risk of HIV Scale (higher score=worse).

[j]PHQ-8: Patient Health Questionnaire–8 (higher score=worse).

[k]n=50.

[l]n=9.

UX Questionnaire Summary Scores

Table 2 lists the summary statistics, including mean and range, for the total summary score and the 4 themed subscores. On average, participants expressed an overall satisfaction with the DHFS system (total summary score: mean 3.74, SD 0.70). On average, the participants had the highest summary score for overall experience (mean 3.89, SD 0.87), followed by ease of use (mean 3.74, SD 0.65), perceived utility (mean 3.73, SD 0.76), and intention of future use (mean 3.58, SD 1.08). Based on IQR values, at least 75% of participants provided favorable feedback for the total score (IQR 3.17-4.20), overall experience (IQR 3.50-4.50), ease of use (IQR 3.33-4.22), and perceived

utility (IQR 3.22-4.25), and at least 50% of participants expressed favorable feedback on intention of future use (IQR 2.80-4.40).

Table . Summary scores from the user experience questionnaire.

Summary score	Participants, n	Value, mean (SD)	Value, median (IQR)	Value, range
Overall experience	53	3.89 (0.87)	4.00 (3.50-4.50)	2.25-5.00
Ease of use	52	3.74 (0.65)	3.72 (3.33-4.22)	2.33-5.00
Intention of future use	53	3.58 (1.08)	3.80 (2.80-4.40)	1.00-5.00
Perceived utility	48	3.73 (0.76)	3.72 (3.22-4.25)	1.44-5.00
Total	47	3.74 (0.70)	3.78 (3.17-4.20)	2.00-4.96

Participant Characteristics as Predictors of UX Score

Total Summary Score

The single-predictor analyses showed that higher total summary scores were associated with better self-efficacy rating (0.822 per point, 95% CI 0.405-1.240; *P*<.001) and PHQ-8 score with a nonlinear trend (natural cubic spline with 1 knot). Only self-efficacy remained in the model after multipredictor model selection. Regression analyses for predictors of total summary score are shown in Table 3.

Table . Single- and multipredictor linear regression associations of demographics and baseline characteristics with the total user experience score.

Variable		Single-predictor model		Multipredictor model	
		Coefficient (95% CI)	*P* value	Coefficient (95% CI)	*P* value
Age (per year)		0.008 (−0.011 to 0.027)	.38	—[a]	—
Gender			.60		—
	Male	Reference		—	
	Female or transgender	−0.197 (−0.944 to 0.549)		—	
Race and ethnicity			.71		—
	Asian, non-Hispanic	−0.122 (−0.944 to 0.701)		—	
	Black, non-Hispanic	0.320 (−0.705 to 1.344)		—	
	Hispanic	0.209 (−0.377 to 0.796)		—	
	White, non-Hispanic	Reference		—	
Positive drug screen (any drugs)		0.214 (−0.239 to 0.666)	.35	—	—
Number of drugs (per drug)		0.103 (−0.139 to 0.346)	.39	—	—
Global PSQI[b] score (per point)		−0.018 (−0.082 to 0.047)	.58	—	—
Self-efficacy (per point)		0.822 (0.405 to 1.240)	<.001	0.822 (0.405 to 1.240)	<.001
Habitual self-control (per point)		0.224 (−0.134 to 0.581)	.21	—	—
HIV risk perception (PRHS[c] 8-item; per point)		−0.015 (−0.055 to 0.025)	.45	—	—
PHQ-8[d] total (per point)		—[e]	.16	—	—

[a]Not applicable.

[b]PSQI: Pittsburgh Sleep Quality Index.

[c]PRHS: Perceived Risk of HIV Scale.

[d]PHQ-8: Patient Health Questionnaire–8.

[e]Nonlinear terms using cubic spline curves.

Overall Experience Score

In the univariable analyses, associations were found at the .20 significance level between a higher overall experience score and male gender (mean difference 0.543, 95% CI −0.274 to 1.360; *P*=.19 vs female or transgender), Hispanic ethnicity (mean difference 0.654, 95% CI 0.026-1.282 vs non-Hispanic White), better PSQI score (−0.080 per point, 95% CI −0.152 to −0.007; *P*=.03), better self-efficacy rating (0.688 per point, 95% CI 0.146-1.230; *P*=.01), stronger habitual self-control (0.384

per point, 95% CI −0.008 to 0.776; *P*=.06), and PHQ-8 score with a nonlinear trend (natural cubic spline with 1 knot; *P*=.09). In the multivariable analyses, incorporating the above univariable associations of .20 significance and higher, the overall experience score was associated with race and ethnicity (*P*=.02), lower PSQI score indicating better sleep (−0.070 per point, 95% CI −0.133 to −0.006; *P*=.03), and higher self-efficacy rating (0.771 per point, 95% CI 0.292-1.250; *P*=.002; see Table 4).

Table . Multipredictor linear regression associations of demographic, baseline characteristics, and percentage of confirmed doses with subscores. See Figure 2 for the perceived utility model.

Variable		Overall experience		Ease of use		Intent of future use		Perceived utility	
		Coefficient (95% CI)	*P* value	Coefficient (95% CI)	*P* value	Coefficient (95% CI)	*P* value	Coefficient (95% CI)	*P* value
Race and ethnicity			.02		—[a]		—		—
	Asian, non-Hispanic	−0.195 (−1.094 to 0.704)		—		—		—	
	Black, non-Hispanic	0.937 (−0.209 to 2.083)		—		—		—	
	Hispanic	0.606 (0.002 to 1.210)		—		—		—	
	White, non-Hispanic	Reference		—		—		—	
Global PSQI[b] score (per point)		−0.070 (−0.133 to −0.006)	.03	—	—	—	—	—	—
Self-efficacy (per point)		0.771 (0.292 to 1.250)	.002	0.750 (0.375 to 1.126)	<.001	0.885 (0.213 to 1.557)	.01	0.901 (0.411 to 1.391)	<.001
PHQ-8[c] total (per point)		—	—	—	—	—	—	—[d]	.01

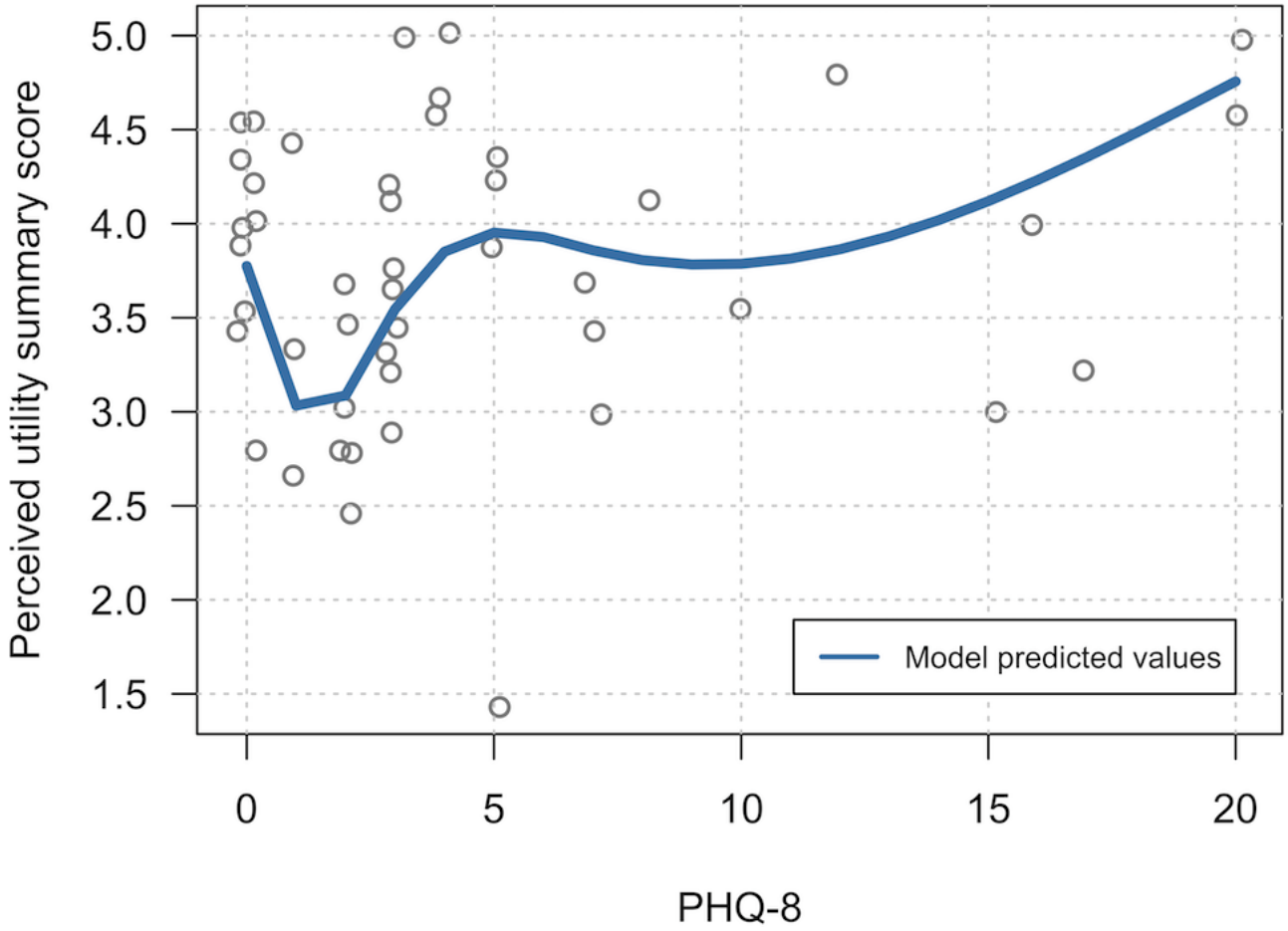
[a]Not applicable.

[b]PSQI: Pittsburgh Sleep Quality Index.

[c]PHQ-8: Patient Health Questionnaire–8.

[d]Nonlinear terms using cubic spline curves.

Figure 2. Observed values (points) and the predicted spline (line) showing association between PHQ-8 and perceived utility summary score, adjusted for self-efficacy. PHQ-8: Patient Health Questionnaire–8.



Ease of Use Score

The ease of use score showed correlation with better self-efficacy rating (0.750 per point, 95% CI 0.375-1.126; *P*<.001) and PHQ-8 score with a nonlinear trend (natural cubic spline with 1 knot; *P*=.16). The multipredictor model on the ease of use score retained only self-efficacy as a predictor (see Table 4).

Intention of Future Use Score

Only the self-efficacy rating was associated with the intention of future use score (0.885 per point, 95% CI 0.213-1.557; *P*=.01; see Table 4)

Perceived Utility Score

The single-predictor analyses showed association between the perceived utility score and self-efficacy rating (0.672 per point, 95% CI 0.187-1.157; *P*=.008), as well as PHQ-8 score (natural cubic spline with 3 knots; *P*=.10). The multipredictor model retained both predictors and showed that better perceived utility score was associated with better self-efficacy rating (0.901 per point, 95% CI 0.411-1.391; *P*<.001) and worse PHQ-8 score (natural cubic spline with 3 knots; overall *P*=.01; see Table 4).

Discussion

Principal Findings

The detailed UX analysis indicated that greater than 75% of participants who used the DHFS with IS-Truvada reported positive experiences, based on the total score and the subscore analyses (overall experience, ease of use, and perceived utility analyses); 70% of participants responded positively toward the intention of future use. Multivariable linear regression analyses of participant characteristics found that having a higher baseline self-efficacy concept score was associated with more positive UX in the total score and across all subscores. In addition, Hispanic ethnicity and more depressive symptoms were associated with reporting positive overall experience and higher DHFS utility, respectively. Poor sleep (captured by the global PSQI score) was associated with a worse overall experience.

UX research is central to the process of developing user-centric technology integration into clinical arenas serving different patient populations [34]. Our research indicates that among participants prescribed PrEP, the DHFS with IS-Truvada was well received. Meta-analysis of UX with mobile health technology repeatedly finds the following themes as being critical to the end user: functionality (related to experiences supporting self-management); acceptance (related to usability and feasibility); perceptions of benefit (related to self-efficacy and empowerment); and importance of co-design [55]. From

this perspective, our findings indicate that the DHFS met critical end-user needs of functionality, acceptance, and perception of benefit in the population using PrEP in this study.

Our data currently represent one of the largest and most detailed study available on UX in persons with any medical diagnosis who have experience using an IS medication adherence system. Our findings are in line with those of Chai et al [56], who reported from qualitative interviews that 15 out of 90 persons using a digital pill system (DPS) [2] for PrEP adherence measurement perceived the device as acceptable, novel, and valuable. Interestingly, men who have sex with men (MSM) taking PrEP who reported substance use were also found to have positive perceptions toward using the DPS in the future [57]. Notably, the PrEP cohort we studied had significantly higher levels of UX satisfaction with the DHFS than that reported in participants in a psychiatric study population (greater than 75% vs 53%, respectively) [58]. The authors noted that their study population included participants with acute psychotic illness; in contrast, participants on stable antipsychotic doses without psychotic symptoms in a prior study [59] reported 70% satisfaction and 78% utility in response to single-question item. The findings from a stable psychiatric study population are close to those found by our detailed UX analysis in persons starting PrEP.

At the outset, we investigated whether baseline demographics and self-report questionnaires could be used to inform current and future use "personas" among our study population, following the expectation that with technology use, "one size may not fit all." A significant association with positive UX for the total score and all subscores was the participants' sense of self-efficacy. Self-efficacy is defined as a person's belief in their capability to succeed and achieve a given level of performance [44]. Self-efficacy is considered to be connected to motivation, achievement, and self-regulation [45,46]. We used an established scale for capturing self-efficacy in the context of medication taking [43] and found an association between the self-efficacy concept and experience of DHFS technology functionality, acceptance, and perception of benefit. Our findings indicate that the self-efficacy concept is directly related to the use of the health care technology tested, and our findings are in line with prior reports showing that self-efficacy beliefs can affect perceived usefulness and perceived ease of use of technology in general, and health informatics and digital health social media applications in particular [60,61].

Persons having lower sleep quality at baseline reported a worse overall experience with the DHFS, which may be related to the requirement of a patch worn on the torso in the system tested and suggests that sleep quality should be evaluated before using the DHFS in a clinical study or practice. The role of depressive symptoms on the UX with the DHFS needs more evaluation. Current evidence indicates that persons with depressive symptoms have significantly higher use of health information technology than persons with other chronic diseases [62], and a plethora of research exists on both digital data-gathering and web-based intervention tools for depression [63]. No published literature appears to be available on what persons with depressive symptoms or a diagnosis of depression *want* from digital health technology. It is likely such data are collected during technology "co-design" efforts, but these data may be analyzed as chronic disease comorbidities or general mental health associations. Our findings suggest that specific criteria for digital health technology may be important to persons with depressive symptoms and support separate analysis of user requirements and experience for these persons.

Clinical Implications

As with any long-term therapy, successful provision of PrEP to prevent HIV infection requires a compassionate mindset, involving a highly individualized series of investigations into how each patient and their disease risk intertwine over time, with trust and honesty between both patients and physicians [64]. Patients need and have choices on PrEP delivery. Based on our UX evaluation of the DHFS with IS-Truvada, this technology is acceptable to patients prescribed PrEP and may be useful to provide insights for both patients and providers on optimal PrEP treatment modalities for individuals over time [7].

Limitations

The study sample was almost exclusively comprised of MSM, and the findings are not generalizable to other populations of patients using PrEP. The study duration was limited to 3 months, and our analysis included only those who persisted with the technology for at least 1 month, 84% (53/63) of whom completed the detailed UX questionnaire. As is expected with a detailed questionnaire, some participants omitted answers to some questions. However, the use of reverse scored and related questions, while primarily designed to ensure questionnaire validity, also served to expand the capture of concepts where questions were left unanswered. Significant differences were observed between questionnaire nonrespondents and respondents, limiting the generalizability of our findings to persons taking PrEP who are homeless or use methamphetamine. In the latter regard, it is notable that Chai et al [56,57] report positive UX and attitude findings regarding DPS technology in MSM taking PrEP who use substances. Despite being one of the largest detailed study available on UX in persons who used an IS medication adherence system, our sample size was limited, and as the probability of reporting a positive experience was observed in >75% of the study population, the likelihood of identifying "current and future use personas" across our study was restricted by our sample size.

Conclusions

The evaluated IS technology met critical end-user needs of functionality, acceptance, and perception of benefit in the population using PrEP in this study. These findings support the continued evaluation of IS adherence technologies in this patient population. Oral medication adherence is a complex behavior. Increased focus on understanding and matching the needs of individual patients to available digital adherence technology options may improve the impact of these technologies on adherence monitoring and support, as well as inform optimal PrEP treatment modalities for individual patients over time.

Acknowledgments

This work was supported in by the US National Institutes of Health through an R01 grant from the National Institute of Mental Health (MH110057), awarded to SB and CAB. Additional support was provided by the nonprofit Specialists in Global Health. The study medication was kindly provided by Gilead Sciences Inc.

Conflicts of Interest

CAB serves as an expert consultant for NDA Partners, Inc. CAB is the Director of the University of California San Diego (UCSD) AntiViral Research Center (AVRC) and receives grant and contract support from Gilead Sciences for the conduct of a clinical trial. SB receives grant funding from Samsung Electronics, South Korea, for the conduct of a clinical trial. All the above authors state that none of these activities represent any conflicts or competing interests relevant to this research effort. All other authors declare no competing interests.

References

1. 510(K) Premarket notification: ingestible event marker K150494, Proteus Digital Health Feedback Device. U.S. Food and Drug Administration. URL: https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K150494 [accessed 2023-08-01]
2. 510(K) Premarket notification: ingestible event marker K183052, ID-Cap System. U.S. Food and Drug Administration. URL: https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K183052 [accessed 2023-08-01]
3. Browne SH, Behzadi Y, Littlewort G. Let visuals tell the story: medication adherence in patients with type II diabetes captured by a novel ingestion sensor platform. JMIR Mhealth Uhealth 2015 Dec 31;3(4):e108. [doi: 10.2196/mhealth.4292] [Medline: 26721413]
4. Hafezi H, Robertson TL, Moon GD, Au-Yeung KY, Zdeblick MJ, Savage GM. An ingestible sensor for measuring medication adherence. IEEE Trans Biomed Eng 2015 Jan;62(1):99-109. [doi: 10.1109/TBME.2014.2341272] [Medline: 25069107]
5. Sulkowski M, Luetkemeyer AF, Wyles DL, et al. Impact of a digital medicine programme on hepatitis C treatment adherence and efficacy in adults at high risk for non-adherence. Aliment Pharmacol Ther 2020 Jun;51(12):1384-1396. [doi: 10.1111/apt.15707] [Medline: 32352586]
6. Browne SH, Umlauf A, Tucker AJ, et al. Wirelessly observed therapy compared to directly observed therapy to confirm and support tuberculosis treatment adherence: a randomized controlled trial. PLoS Med 2019 Oct 4;16(10):e1002891. [doi: 10.1371/journal.pmed.1002891] [Medline: 31584944]
7. Browne SH, Vaida F, Umlauf A, Tucker AJ, Blaschke TF, Benson CA. Supporting the art: medication adherence patterns in persons prescribed ingestible sensor-enabled oral pre-exposure prophylaxis to prevent human immunodeficiency virus infection. Clin Infect Dis 2023 Jan 6;76(1):134-143. [doi: 10.1093/cid/ciac280] [Medline: 36484300]
8. Frias J, Virdi N, Raja P, Kim Y, Savage G, Osterberg L. Effectiveness of digital medicines to improve clinical outcomes in patients with uncontrolled hypertension and type 2 diabetes: prospective, open-label, cluster-randomized pilot clinical trial. J Med Internet Res 2017 Jul 11;19(7):e246. [doi: 10.2196/jmir.7833] [Medline: 28698169]
9. Karris MY, Beekmann SE, Mehta SR, Anderson CM, Polgreen PM. Are we prepped for preexposure prophylaxis (PrEP)? provider opinions on the real-world use of PrEP in the United States and Canada. Clin Infect Dis 2014 Mar;58(5):704-712. [doi: 10.1093/cid/cit796] [Medline: 24319083]
10. Wade Taylor S, Mayer KH, Elsesser SM, Mimiaga MJ, O'Cleirigh C, Safren SA. Optimizing content for pre-exposure prophylaxis (PrEP) counseling for men who have sex with men: perspectives of PrEP users and high-risk PrEP naive men. AIDS Behav 2014 May;18(5):871-879. [doi: 10.1007/s10461-013-0617-7] [Medline: 24077928]
11. Agot K, Taylor D, Corneli AL, et al. Accuracy of self-report and pill-count measures of adherence in the FEM-PrEP clinical trial: implications for future HIV-prevention trials. AIDS Behav 2015 May;19(5):743-751. [doi: 10.1007/s10461-014-0859-z] [Medline: 25100053]
12. van Damme L, Corneli A, Ahmed K, et al. Preexposure prophylaxis for HIV infection among African women. N Engl J Med 2012 Aug 2;367(5):411-422. [doi: 10.1056/NEJMoa1202614] [Medline: 22784040]
13. Marrazzo JM, Ramjee G, Richardson BA, et al. Tenofovir-based preexposure prophylaxis for HIV infection among African women. N Engl J Med 2015 Feb 5;372(6):509-518. [doi: 10.1056/NEJMoa1402269] [Medline: 25651245]
14. Ingersoll KS, Heckman CJ. Patient-clinician relationships and treatment system effects on HIV medication adherence. AIDS Behav 2005 Mar;9(1):89-101. [doi: 10.1007/s10461-005-1684-1] [Medline: 15812616]
15. Quirk A, Chaplin R, Hamilton S, Lelliott P, Seale C. Communication about adherence to long-term antipsychotic prescribing: an observational study of psychiatric practice. Soc Psychiatry Psychiatr Epidemiol 2013 Apr;48(4):639-647. [doi: 10.1007/s00127-012-0581-y] [Medline: 23007293]
16. Laine C, Newschaffer CJ, Zhang D, Cosler L, Hauck WW, Turner BJ. Adherence to antiretroviral therapy by pregnant women infected with human immunodeficiency virus: a pharmacy claims-based analysis. Obstet Gynecol 2000 Feb;95(2):167-173. [doi: 10.1016/s0029-7844(99)00523-2] [Medline: 10674574]
17. Weis SE, Foresman B, Matty KJ, et al. Treatment costs of directly observed therapy and traditional therapy for Mycobacterium tuberculosis: a comparative analysis. Int J Tuberc Lung Dis 1999 Nov;3(11):976-984. [Medline: 10587319]

18. Vernon AA, Iademarco MF. In the treatment of tuberculosis, you get what you pay for. Am J Respir Crit Care Med 2004 Nov 15;170(10):1040-1042. [doi: 10.1164/rccm.2409005] [Medline: 15533952]
19. Ingerski LM, Loew M, Porter JS, et al. Use of wise device technology to measure adherence to hydroxyurea therapy in youth with sickle cell disease. J Pediatr Hematol Oncol 2021 Jan;43(1):e19-e25. [doi: 10.1097/MPH.0000000000001997] [Medline: 33235145]
20. Vrijens B, Urquhart J, White D. Electronically monitored dosing histories can be used to develop a medication-taking habit and manage patient adherence. Expert Rev Clin Pharmacol 2014 Sep;7(5):633-644. [doi: 10.1586/17512433.2014.940896] [Medline: 25088004]
21. Cattamanchi A, Crowder R, Kityamuwesi A, et al. Digital adherence technology for tuberculosis treatment supervision: a stepped-wedge cluster-randomized trial in Uganda. PLoS Med 2021 May;18(5):e1003628. [doi: 10.1371/journal.pmed.1003628] [Medline: 33956802]
22. Bova CA, Fennie KP, Knafl GJ, Dieckhaus KD, Watrous E, Williams AB. Use of electronic monitoring devices to measure antiretroviral adherence: practical considerations. AIDS Behav 2005 Mar;9(1):103-110. [doi: 10.1007/s10461-005-1685-0] [Medline: 15812617]
23. Denhaerynck K, Schäfer-Keller P, Young J, Steiger J, Bock A, de Geest S. Examining assumptions regarding valid electronic monitoring of medication therapy: development of a validation framework and its application on a European sample of kidney transplant patients. BMC Med Res Methodol 2008 Feb 19;8:5. [doi: 10.1186/1471-2288-8-5] [Medline: 18284675]
24. Olivieri NF, Matsui D, Hermann C, Koren G. Compliance assessed by the Medication Event Monitoring System. Arch Dis Child 1991 Dec;66(12):1399-1402. [doi: 10.1136/adc.66.12.1399] [Medline: 1776885]
25. Moore DJ, Jain S, Dubé MP, et al. Randomized controlled trial of daily text messages to support adherence to preexposure prophylaxis in individuals at risk for human immunodeficiency virus: the TAPIR study. Clin Infect Dis 2018 May 2;66(10):1566-1572. [doi: 10.1093/cid/cix1055] [Medline: 29228144]
26. Bendiksen R, Ovesen T, Asfeldt AM, Halvorsen DS, Gravningen K. Use of video directly observed treatment for tuberculosis in northern Norway. Tidsskr Nor Laegeforen 2020 Jan 13;140(1). [doi: 10.4045/tidsskr.19.0322] [Medline: 31948218]
27. Castillo-Mancilla JR, Zheng JH, Rower JE, et al. Tenofovir, emtricitabine, and tenofovir diphosphate in dried blood spots for determining recent and cumulative drug exposure. AIDS Res Hum Retroviruses 2013 Feb;29(2):384-390. [doi: 10.1089/AID.2012.0089] [Medline: 22935078]
28. Okochi H, Louie A, Phung N, et al. Tenofovir and emtricitabine concentrations in hair are comparable between individuals on tenofovir disoproxil fumarate versus tenofovir alafenamide-based ART. Drug Test Anal 2021 Jul;13(7):1354-1370. [doi: 10.1002/dta.3033] [Medline: 33742745]
29. Blaschke TF, Osterberg L, Vrijens B, Urquhart J. Adherence to medications: insights arising from studies on the unreliable link between prescribed and actual drug dosing histories. Annu Rev Pharmacol Toxicol 2012;52:275-301. [doi: 10.1146/annurev-pharmtox-011711-113247] [Medline: 21942628]
30. Osterberg LG, Urquhart J, Blaschke TF. Understanding forgiveness: minding and mining the gaps between pharmacokinetics and therapeutics. Clin Pharmacol Ther 2010 Oct;88(4):457-459. [doi: 10.1038/clpt.2010.171]
31. Grant RM, Anderson PL, McMahan V, et al. Uptake of pre-exposure prophylaxis, sexual practices, and HIV incidence in men and transgender women who have sex with men: a cohort study. Lancet Infect Dis 2014 Sep;14(9):820-829. [doi: 10.1016/S1473-3099(14)70847-3] [Medline: 25065857]
32. Liu AY, Cohen SE, Vittinghoff E, et al. Preexposure prophylaxis for HIV infection integrated with municipal- and community-based sexual health services. JAMA Intern Med 2016 Jan;176(1):75-84. [doi: 10.1001/jamainternmed.2015.4683] [Medline: 26571482]
33. Martin M, Vanichseni S, Suntharasamai P, et al. The impact of adherence to preexposure prophylaxis on the risk of HIV infection among people who inject drugs. AIDS 2015 Apr 24;29(7):819-824. [doi: 10.1097/QAD.0000000000000613] [Medline: 25985403]
34. Gilbert RM. Reimagining digital healthcare with a patient-centric approach: the role of user experience (UX) research. Front Digit Health 2022 Aug;4:899976. [doi: 10.3389/fdgth.2022.899976] [Medline: 36016600]
35. Kotler PT, Armstrong G. Principles of Marketing, 17th Global Edition: Pearson; 2010:160.
36. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly 1989 Sep;13(3):319-340. [doi: 10.2307/249008]
37. Davis FD, Bagozzi RP, Warshaw PR. User acceptance of computer technology: a comparison of two theoretical models. Manage Sci 1989 Aug 1;35(8):902-1028. [doi: 10.1287/mnsc.35.8.982]
38. Venkatesh V, Davis FD. A theoretical extension of the technology acceptance model: four longitudinal field studies. Manage Sci 2000 Feb 1;46(2):169-332. [doi: 10.1287/mnsc.46.2.186.11926]
39. Marangunić N, Granić A. Technology acceptance model: a literature review from 1986 to 2013. Univ Access Inf Soc 2015 Mar;14:81-95. [doi: 10.1007/s10209-014-0348-1]
40. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Quarterly 2003 Sep;27(3):425-478. [doi: 10.2307/30036540]

41. Pan X. Technology acceptance, technological self-efficacy, and attitude toward technology-based self-directed learning: learning motivation as a mediator. Front Psychol 2020 Oct 27;11:564294. [doi: 10.3389/fpsyg.2020.564294] [Medline: 33192838]
42. AlQudah AA, Al-Emran M, Shaalan K. Technology acceptance in healthcare: a systematic review. Appl Sci 2021 Nov 9;11(22):10537. [doi: 10.3390/app112210537]
43. Johnson MO, Neilands TB, Dilworth SE, Morin SF, Remien RH, Chesney MA. The role of self-efficacy in HIV treatment adherence: validation of the HIV Treatment Adherence Self-Efficacy Scale (HIV-ASES). J Behav Med 2007 Oct;30(5):359-370. [doi: 10.1007/s10865-007-9118-3] [Medline: 17588200]
44. Bandura AJ. Self-efficacy: toward a unifying theory of behavioral change. Psychol Rev 1977 Mar;84(2):191-215. [doi: 10.1037//0033-295x.84.2.191] [Medline: 847061]
45. Bandura A. Guide for constructing self-efficacy scales. In: Pajares F, Urdan T, editors. Self-Efficacy Beliefs of Adolescents: Information Age Publishing; 2006, Vol. 5:307-337.
46. Bandura A. Self-efficacy. In: The Corsini Encyclopedia of Psychology: John Wiley & Sons, Inc; 2010:1-3. [doi: 10.1002/9780470479216]
47. Schroder KEE, Schwarzer R. Habitual self-control and the management of health behavior among heart patients. Soc Sci Med 2005 Feb;60(4):859-875. [doi: 10.1016/j.socscimed.2004.06.028] [Medline: 15571902]
48. Kroenke K, Strine TW, Spitzer RL, Williams JBW, Berry JT, Mokdad AH. The PHQ-8 as a measure of current depression in the general population. J Affect Disord 2009 Apr;114(1-3):163-173. [doi: 10.1016/j.jad.2008.06.026] [Medline: 18752852]
49. Buysse DJ, Reynolds CFIII, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index (PSQI): a new instrument for psychiatric research and practice. Psychiatry Res 1989 May;28(2):193-213. [doi: 10.1016/0165-1781(89)90047-4] [Medline: 2748771]
50. Napper LE, Fisher DG, Reynolds GL. Development of the Perceived Risk of HIV Scale. AIDS Behav 2012 May;16(4):1075-1083. [doi: 10.1007/s10461-011-0003-2] [Medline: 21785873]
51. Levin JB, Sajatovic M, Rahman M, et al. Outcomes of psychoeducation and a text messaging adherence intervention among individuals with hypertension and bipolar disorder. Psychiatr Serv 2019 Jul 1;70(7):608-612. [doi: 10.1176/appi.ps.201800482] [Medline: 30991908]
52. Rubin DB. Multiple imputation after 18+ years. J Am Stat Assoc 1996;91(434):473-489. [doi: 10.1080/01621459.1996.10476908]
53. The R project for statistical computing. R project. 2020. URL: https://www.R-project.org/ [accessed 2024-04-19]
54. van Buuren S, Groothuis-Oudshoorn K. mice: multivariate imputation by chained equations in R. J Stat Softw 2011 Dec 12;45(3):1-67. [doi: 10.18637/jss.v045.i03]
55. Slater H, Campbell JM, Stinson JN, Burley MM, Briggs AM. End user and implementer experiences of mHealth technologies for noncommunicable chronic disease management in young adults. J Med Internet Res 2017 Dec 12;19(12):e406. [doi: 10.2196/jmir.8888] [Medline: 29233804]
56. Chai PR, Goodman GR, Bronzi O, et al. Real-world user experiences with a digital pill system to measure prep adherence: perspectives from MSM with substance use. AIDS Behav 2022 Jul;26(7):2459-2468. [doi: 10.1007/s10461-022-03594-9] [Medline: 35089449]
57. Chai P, De D, Albrechta H, et al. Attitudes towards participating in research involving digital pill systems to measure oral HIV pre-exposure chemoprophylaxis: a cross-sectional study among men who have sex with men with substance use in the USA. BMJ Open 2023 Jan 30;13(1):e067549. [doi: 10.1136/bmjopen-2022-067549] [Medline: 36717151]
58. Fowler JC, Cope N, Knights J, et al. Hummingbird study: results from an exploratory trial assessing the performance and acceptance of a digital medicine system in adults with schizophrenia, schizoaffective disorder, or first-episode psychosis. Neuropsychiatr Dis Treat 2021 Feb 12;17:483-492. [doi: 10.2147/NDT.S290793] [Medline: 33603385]
59. Peters-Strickland T, Pestreich L, Hatch A, et al. Usability of a novel digital medicine system in adults with schizophrenia treated with sensor-embedded tablets of aripiprazole. Neuropsychiatr Dis Treat 2016 Oct 11;12:2587-2594. [doi: 10.2147/NDT.S116029] [Medline: 27785036]
60. Tetri B, Juujärvi S. Self-efficacy, internet self-efficacy, and proxy efficacy as predictors of the use of digital social and health care services among mental health service users in Finland: a cross-sectional study. Psychol Res Behav Manag 2022 Feb 15;15:291-303. [doi: 10.2147/PRBM.S340867] [Medline: 35210878]
61. Kim NE, Han SS, Yoo KH, Yun EK. The impact of user's perceived ability on online health information acceptance. Telemed J E Health 2012 Nov;18(9):703-708. [doi: 10.1089/tmj.2011.0277] [Medline: 23072632]
62. Rajamani G, Lindemann E, Evans MD, Pillai R, Badlani S, Melton GB. Health information technology use among chronic disease patients: an analysis of the United States Health Information National Trends Survey. Appl Clin Inform 2022 May;13(3):752-766. [doi: 10.1055/s-0042-1751305] [Medline: 35952679]
63. Lattie EG, Adkins EC, Winquist N, Stiles-Shields C, Wafford QE, Graham AK. Digital mental health interventions for depression, anxiety, and enhancement of psychological well-being among college students: systematic review. J Med Internet Res 2019 Jul 22;21(7):e12869. [doi: 10.2196/12869] [Medline: 31333198]

64. Richey AG, Kovacs I, Browne S. Use of an ingestible, sensor-based digital adherence system to strengthen the therapeutic relationship in serious mental illness. JMIR Ment Health 2022 Dec 2;9(12):e39047. [doi: 10.2196/39047] [Medline: 36459392]

Abbreviations

DHFS: Digital Health Feedback System
DPS: digital pill system
FDA: Food and Drug Administration
IS: ingestible sensor
MSM: men who have sex with men
PHQ-8: Patient Health Questionnaire–8
PrEP: pre-exposure prophylaxis
PRHS: Perceived Risk of HIV Scale
PSQI: Pittsburgh Sleep Quality Index
TB: tuberculosis
UCSD: University of California San Diego
UTOX: urine toxicology screening
UX: user experience

Edited by R Shrestha; submitted 16.10.23; peer-reviewed by G Goodman, P Cook; revised version received 26.02.24; accepted 01.04.24; published 03.05.24.

Please cite as:
Browne S, Umlauf A, Moore DJ, Benson CA, Vaida F
User Experience of Persons Using Ingestible Sensor–Enabled Pre-Exposure Prophylaxis to Prevent HIV Infection: Cross-Sectional Survey Study
JMIR Mhealth Uhealth 2024;12:e53596
URL: https://mhealth.jmir.org/2024/1/e53596
doi:10.2196/53596

Original Paper

Short-Term Effect of a Health Promotion Intervention Based on the Electronic 12-Hour Dietary Recall (e-12HR) Smartphone App on Adherence to the Mediterranean Diet Among Spanish Primary Care Professionals: Randomized Controlled Clinical Trial

Luis María Béjar[1], PhD; Pedro Mesa-Rodríguez[2], BSc; María Dolores García-Perea[3], PhD

[1]Department of Preventive Medicine and Public Health, Institute of Anatomy, School of Medicine, University of Seville, Seville, Spain

[2]Camas Health Center, Seville, Spain

[3]Virgen Macarena University Hospital, Seville, Spain

Corresponding Author:
Luis María Béjar, PhD
Department of Preventive Medicine and Public Health
Institute of Anatomy
School of Medicine, University of Seville
3rd Floor, Sánchez-Pizjuán Avenue
Seville, 41009
Spain
Phone: 34 954551771
Fax: 34 954556481
Email: lmbprado@us.es

Abstract

Background: The World Health Organization has called for addressing the growing burden of noncommunicable diseases (NCDs) by promoting healthy lifestyles among the population. Regarding patient health, primary care professionals (PCPs) are the first line of care who can positively influence patients' behavior and lifestyle habits. However, a significant percentage of PCPs do not lead a healthy lifestyle. Therefore, addressing their health behaviors may be the key to substantially increasing health promotion advice in general practice. The Mediterranean diet has been extensively studied, and there is strong evidence of it being a dietary pattern for the prevention of NCDs, in addition to its significant environmental, sociocultural, and local economics benefits.

Objective: This study focused only on the dietary aspect of the PCPs' lifestyle. The primary objective was to evaluate the effect of using the Electronic 12-Hour Dietary Recall (e-12HR) smartphone app to improve diet, specifically to promote adherence to the Mediterranean diet (AMD), among PCPs. The secondary objectives were to establish the usability of the e-12HR app and to determine AMD among PCPs.

Methods: An individual-level randomized, controlled, and single-blind clinical trial was conducted with 2 parallel groups: a control group (CG), using the nonfeedback version of the e-12HR app, and an intervention group (IG), using the feedback version of the e-12HR app. The level of human involvement was fully automated through the use of the app. There was a 28-day follow-up period. Participants were PCPs (medicine or nursing) recruited offline at one of the selected primary care centers (Andalusia, Spain, Southern Europe), of both sexes, over 18 years old, possessing a smartphone, and having smartphone literacy.

Results: The study response rate was 73% (71 of 97 PCPs), with 27 (38%) women and 44 (62%) men: 40 (56%) PCPs in the CG and 31 (44%) in the IG. At baseline, AMD was medium (mean Mediterranean Diet Serving Score [MDSS] index 9.45, range 0-24), with 47 (66%) PCPs with a medium/high MDSS index. There were significant statistical improvements (CG vs IG, in favor of the IG) at week 4 (no significant statistical differences at baseline): +25.6% for the MDSS index (P=.002) and +213.1% for the percentage with a medium/high MDSS index (P=.001). In relation to specific food groups, there were significant statistical improvements for fruits (+33.8%, P=.02), vegetables (+352%, P=.001), nuts (+184%, P=.02), and legumes (+75.1%, P=.03). The responses to the usability rating questionnaire were satisfactory.

Conclusions: The results support recommending the use of the e-12HR app as a tool to contribute to improving diet and preventing NCDs among PCPs, while positively influencing patient dietary behavior and preventing diet-related NCDs among patients.

Trial Registration: ClinicalTrials.gov NCT05532137; https://clinicaltrials.gov/study/NCT05532137

(JMIR Mhealth Uhealth 2024;12:e49302) doi:10.2196/49302

KEYWORDS

primary care professionals; Mediterranean diet; smartphone applications; smartphone apps; health promotion; Mediterranean diet adherence; food group

Introduction

Major chronic noncommunicable diseases (NCDs; eg, cardiovascular diseases, cancer, chronic respiratory diseases, and diabetes) are responsible for 74% of deaths worldwide, making them the leading cause of preventable mortality [1]; they also reduce patients' well-being and activity, which contributes to a poor quality of life, disability, and reduced productivity [1]. Notably, these NCDs share common risk factors, such as an unhealthy diet, smoking, harmful alcohol consumption, and physical inactivity [1], all of which are behavioral or lifestyle-related factors that are potentially modifiable [2]. For this reason, the World Health Organization has called for addressing the growing burden of NCDs by promoting healthy lifestyles [1,3].

Health care workers are a subgroup of the population where lifestyle promotion is essential for 3 main reasons. The first reason is their own health: although health care professionals do their best to provide exceptional patient care, they often fail to prioritize their own self-care [4], resulting in unhealthy behaviors associated with a high risk of NCDs. Studies of health care workers' lifestyles in hospitals, for example, showed high rates of overweight/obesity [5-7], low fruit and vegetable consumption [6,7], low physical activity [5-8], tobacco [5,7] and alcohol consumption [6,7], and high levels of stress and insufficient rest [4-6]. In addition, almost half had more than 2 of these risk factors [5,6].

The second reason is the success of health care organizations: health care workers with better and healthier habits have been shown to have higher personal and job satisfaction and fewer sickness-related absences from work [9]. The third reason is patient health: several studies have described that health care professionals who practice healthy behaviors offer more advice on healthy habits to patients who come to their practice, that they have more assertive attitudes when counselling patients, and they provide more comprehensive and aggressive counselling, which can positively influence patients' health [10-14].

Regarding patient health, primary care is the cornerstone for preventing NCDs through health education and plays an essential role in the success of therapeutic medicine [11]. Primary care professionals (PCPs) are the first line of care who can positively influence patient behavior and lifestyle habits [15], and indeed, patients perceive PCPs (doctors and nurses) as the most trusted source of health information and advice on healthy lifestyles [13]. However, a significant percentage of PCPs, as with hospital health care professionals, do not lead a healthy lifestyle: there is a high prevalence of NCD risk factors, such as smoking [10,16], alcohol abuse [10,16], overweight/obesity [17], physical inactivity [5,10,17], inadequate fruit and vegetable intake [10,11,17], added salt intake [17], and high consumption of ultraprocessed foods [18], sugars, and fats [10]. Therefore, addressing PCPs' health behaviors may be the key to substantially increasing health promotion advice in general practice [12].

As the evidence shows, PCPs are exposed to several risk factors of NCDs; this study focused only on the dietary aspect of the PCPs' lifestyle. As a starting point, the research team posed the following question: How can the dietary habits of PCPs be improved?

The research team has previously evaluated the effectiveness of a smartphone app called the Electronic 12-Hour Dietary Recall (e-12HR) in improving diet in Spanish university students (health science [19,20] and non–health science [20]). The main hypothesis of this study was that the use of this app among PCPs can have an influence on improving their diet, as has already been evidenced in previous research among university students [19,20], the null hypothesis being that the use of the app has no influence on improving the diet of PCPs.

To the best of our knowledge, this study is the first to evaluate the use of a smartphone app to improve diet among PCPs. The primary study objective was to evaluate the effect of using the e-12HR app on improving diet among PCPs, specifically to promote adherence to the Mediterranean diet (AMD). The Mediterranean diet has been extensively studied, and there is strong evidence of it being a dietary pattern for the prevention of NCDs [21-25], in addition to its significant environmental, sociocultural, and local economics benefits [26-28]. In addition, secondary objectives were to establish the usability of the e-12HR app and determine AMD among PCPs.

Methods

Overview of the Study

The study was an individual-level randomized, controlled, and single-blind clinical trial with 2 parallel groups: a control group (CG) and an intervention group (IG). All participants used an app called the e-12HR app, with different versions for the CG and the IG. In the CG, participants used the nonfeedback version of the e-12HR app, and in the IG, participants used the feedback version of the e-12HR app (see the *Intervention* section). The

level of human involvement was fully automated through the use of the app.

All research was carried out in 3 basic health zones of the Andalusian Health Service (Andalusia, Spain, South of Europe): Camas, Coria del Río, and San Juan de Aznalfarache. Several primary care centers were selected in the 3 zones: the Camas, Santiponce, Valencina de la Concepción, Castilleja de Guzmán, Coca de la Piñera, Carambolo, and Pañoleta health centers in Camas; the Coria del Río health center in Coria del Río; and the San Juan de Aznalfarache and Gelves health centers in San Juan de Aznalfarache. The study ran for 28 days, and participant recruitment took place offline from September to October 2022.

Inclusion criteria for the study were both sexes, age over 18 years, possession of a smartphone (iOS or Android operating system), smartphone literacy, and a PCP (medicine or nursing) at one of the selected primary care centers. Exclusion criteria included food intolerance, chronic disease, or pregnancy (due to the possibility of requiring specific dietary recommendations).

Ethical Considerations

The study was conducted according to the guidelines laid down in the Declaration of Helsinki, and all procedures involving study participants were approved by the Andalusian Biomedical Research Ethics Portal (PEIBA) on March 30, 2022 (identifier: 2813-N-21). The trial was registered at ClinicalTrials (identifier: NCT05532137). Written informed consent was obtained from all participants.

Participant Enrollment

The project was publicized in the selected primary care centers by a member of the research team, and individual talks were scheduled with interested health care professionals. In each of the individual talks, the study protocol was explained, including the objectives, risks, and benefits of the research, and an email address for the study was provided.

To participate in the study, it was necessary for interested health care professionals to send an email to the designated address, indicating their “interest in participating in the study” and the primary care center where they worked. After receiving the email, a member of the research group sent candidates a series of documents necessary to be able to participate in the study: (1) an informed consent form; (2) a form with personal information (sex, date of birth, primary care center, weight, height, and smoking status), with documents 1 and 2 to be completed, signed, and returned to the same email address; (3) instructions for downloading the e-12HR app (free to download from the App Store or the Play Store); (4) an image of the Mediterranean diet pyramid (with recommendations for consumption by food group); (5) a personal alphanumeric code; and (6) a user’s guide with detailed information for using the app. Document 6 was the only one that differed depending on whether it was intended for participants in the CG (nonfeedback e-12HR version) or the IG (feedback e-12HR version), and obviously, document 5, which contained the personal code, was unique for each participant.

Throughout the study, participants could contact the research team by email with any questions, including questions to reduce the likelihood of harm.

Participant Allocation

In each basic health zone, the recruited participants were randomized into 1 of 2 groups (CG or IG) in a ratio of 1:1 as follows: the participant who sent the first email was assigned to the CG, the participant who sent the second email was assigned to the IG, and so on.

This study was single-blind because, due to the nature of the intervention, the PCPs could not be blinded. However, the investigator who performed the statistical analysis of the data was blinded throughout the study. In addition, each participant only had access to 1 version of the app: the personal codes of the participants assigned to the CG activated only the nonfeedback e-12HR version, while the personal codes of the participants assigned to the IG activated only the feedback e-12HR version.

Intervention

The structure and functions of the e-12HR app (nonfeedback and feedback versions) have been described in detail by Béjar et al [19]. In this study, we used e-12HR version 3.0. The e-12HR app did not undergo changes throughout the study. In brief, the nonfeedback e-12HR version allows the user to collect food consumption data; however, this version does not provide any feedback to users to promote the Mediterranean diet (ie, this version of the app presents a single function: diet determination). The feedback e-12HR version allows for the collection of food consumption data, and as an additional automatic function, every 7 days the app issues personalized feedback on how to improve AMD (ie, this version of the app has 2 functions: determining the diet and providing feedback to improve AMD). The feedback provided came in 3 parts: (1) the AMD index score: specifically, the Mediterranean Diet Serving Score (MDSS) index [29] (range 0-24); (2) the image of a traffic light: the MDSS index score was divided into 1 of 3 levels (low: score 0-8, red light; medium: score 9-15, orange light; high: score 16-24, green light) [30]; and (3) recommendations for consumption by food group [19]. The nonfeedback e-12HR version did not provide any of the 3 parts of the feedback, as they were exclusive to the e-12HR feedback version. See Multimedia Appendix 1 for real images of the e-12HR app (nonfeedback and feedback versions).

Follow-up and Outcome Measures

To analyze the effect of the intervention (CG nonfeedback e-12HR vs IG feedback e-12HR), 4 follow-up points were established: week 1 (baseline), week 2, week 3, and week 4. At each follow-up point, the research team manually calculated the MDSS index for each of the 2 groups from the data provided by the e-12HR app. The method for calculating the MDSS index has been described in detail elsewhere [19].

The main result variable was the change in the total MDSS index at weeks 2, 3, and 4 of monitoring, while the secondary result variables were the personal information variables, the MDSS index at week 1 (baseline), and the answers to the

usability rating questionnaire for the e-12HR app (see the *Usability Rating Questionnaire for the e-12HR* section).

The MDSS index at week 1 (baseline) was used to determine AMD among the PCPs (a secondary objective of the study). To relativize the data, the MDSS index of the PCPs was compared to the MDSS index of health sciences students. For a proper comparison, the MDSS index was obtained using the same app (e-12HR) and during the same follow-up period (recruitment period: September-October 2022) for both students and health care professionals.

Usability Rating Questionnaire for the e-12HR

After the 4-week study period, a member of the research team sent a new email to each PCP who had completed the follow-up; this new email contained a usability rating questionnaire for the e-12HR app [19] (see Multimedia Appendix 2).

Statistical Analysis

The sample size was estimated for the main result variable. Assuming SD=2.7 points, dropout rate=20.6% (from a previous study on use of the e-12HR app among health science university students [19]), α=.05, and β=.20 (bilateral test), 82 participants (n=41, 50%, per group) were needed to detect an increase of 2 points in the MDSS index (CG versus IG). The sample size was calculated using nQuery Advisor Release 7.0 (Statsols).

Quantitative variables were expressed as means (SD), and qualitative variables were displayed as numbers (percentages). The nonparametric Kolmogorov-Smirnov test was used for the test for normality.

For unpaired samples and quantitative variables, the Student *t* test or the nonparametric Mann-Whitney *U* test was used, and the chi-square test (or Fisher exact test) was used for the comparison of proportions.

For paired samples, quantitative variables, and two groups, the Student *t* test or the nonparametric Wilcoxon test was used, penalizing *P* values with Bonferroni adjustment for multiple comparisons. For 3 or more groups, the ANOVA test or the nonparametric Friedman test was used.

P<.05 was considered significant, except for multiple comparisons using Bonferroni penalization: *P*<.02 (.05/3).

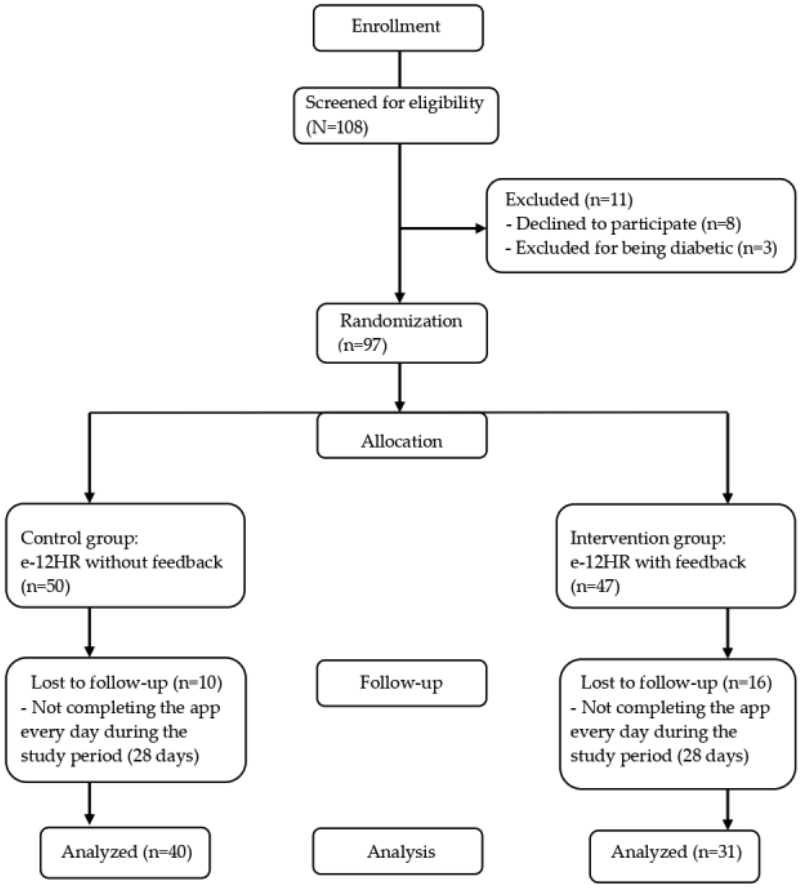
All statistical analyses were performed using the SPSS statistical software package version 26.0 (SPSS Inc).

Results

Sample and Adherence to the Study

The sequence of allocating participants to the 2 study groups is detailed in Figure 1. In total, 97 PCPs signed the informed consent form (n=50, 52%, in the CG and n=47, 48%, in the IG). Of them, 26 (27%; n=10, 38%, in the CG and n=16, 62%, in the IG) were considered nonresponsive because they did not complete the study's 4-week follow-up period (Figure 1). The nonresponsive individuals were not included in the later statistical analysis (ie, in this study, per protocol analysis was applied).

Figure 1. Flowchart of the study. e-12HR: Electronic 12-Hour Dietary Recall.



The study response rate was 73% (71 of 97 participants), with 40 of 50 (80%) participants in the CG and 31 of 47 (66%) participants in the IG (Figure 1). Participants did not report any harm or unintended effects throughout the study.

Personal Information of the Participants

Table 1 shows the personal information of the PCPs who completed the study (CG and IG).

Table 1. Characteristics of participants who completed the study on the short-term effects of a health promotion intervention based on the e-12HR[a] smartphone app on AMD[b] among Spanish PCPs[c].

Characteristics	All participants (N=71)	CG[d] (n=40)	IG[e] (n=31)	*P* value[f]
Age (years)				
Mean (SD)	43.2 (11.6)	45.1 (11.1)	40.7 (12.1)	.07[g]
<40, n (%)	34 (47.9)	16 (40.0)	18 (58.1)	.13[h]
≥40, n (%)	37 (52.1)	24 (60.0)	13 (41.9)	—[i]
Sex, n (%)				
Female	27 (38.0)	16 (40.0)	11 (35.5)	.70[h]
Male	44 (62.0)	24 (60.0)	20 (64.5)	—
BMI (kg/m^2)				
Mean (SD)	25.0 (4.0)	25.3 (4.4)	24.7 (3.4)	.76[g]
<25, n (%)	41 (57.7)	23 (57.5)	18 (58.1)	.96[h]
≥25, n (%)	30 (42.3)	17 (42.5)	13 (41.9)	—
Smoking status, n (%)				
No	64 (90.1)	37 (92.5)	27 (87.1)	.45[h]
Yes	7 (9.9)	3 (7.5)	4 (12.9)	—
Physical activity status (minutes/week), n (%)				
≥150	45 (63.4)	24 (60.0)	21 (67.7)	.50[h]
<150	26 (36.6)	16 (40.0)	10 (32.3)	—

[a]e-12HR: Electronic 12-Hour Dietary Recall.

[b]AMD: adherence to the Mediterranean diet.

[c]PCP: primary care professional.

[d]CG: control group.

[e]IG: intervention group.

[f]*P*<.05 considered significant.

[g]Evaluated with the Mann-Whitney *U* test.

[h]Evaluated with the chi-square test.

[i]Not applicable.

No significant statistical differences were observed in the personal variables studied between the CG and the IG (Table 1). PCPs (n=47, 66%, doctors and n=24, 34%, nurses) recorded their daily consumption using the e-12HR app for a total of 1988 days (N=71 participants × 28 recording days). The app differentiated 19 food groups. Thus, a total of 37,772 daily food group consumption data points were collected during the study.

There were no significant statistical differences in the personal variables between responsive (those who completed the study) and nonresponsive (those who did not complete the study) participants.

MDSS Index

As previously mentioned, the MDSS index was calculated manually by the research team (CG and IG, weeks 1, 2, 3, and 4) [19]. During the process, the research team corrected obvious errors made by the PCPs: for example, when the app asked for the number of standard servings of a certain food group consumed on the current day, if the participant entered a value of 150, it was considered that the data referred to milliliters or grams (instead of standard servings). In any case, only 0.2% (91 of 37,772) of the recorded data were corrected.

At week 1 of the monitoring period (baseline), that is, before the IG received the first feedback from the e-12HR app (only week 1 was considered because the feedback for IG in weeks 2, 3, and 4 could affect the alteration of the usual dietary intake),

the PCPs had a mean MDSS index of 9.45 (SD 2.32), which corresponds to a medium level of adherence [30]; moreover, two-thirds of them (n=47, 66%) had a medium/high MDSS value (≥9) at baseline (week 1) [30].

Effect of the Intervention With the e-12HR App in Terms of Variation in the MDSS Index and Number of Participants With Medium/High (≥12) MDSS Index

Tables 2 and 3 show the MDSS index, and Table 2 also shows the number of participants with a medium/high MDSS index (≥12) in the CG and the IG throughout the 4 weeks of follow-up. We decided to use the value of 12 (instead of 9) [30] due to the high percentage of PCPs (n=47, 66%) who at baseline (week 1) already had an MDSS index≥9; therefore, using the value of 9 would have made it difficult to observe statistically significant differences between the CG and the IG.

Table 2. MDSS[a] index for the CG[b] and the IG[c] and number of participants with a medium/high (≥12) MDSS index throughout the 4 weeks of follow-up.

Variables and week number	CG (n=40)	IG (n=31)	*P* value[d]
MDSS index, mean (SD)			
Week 1	9.30 (2.40)	9.65 (2.23)	.54[e]
Week 2	8.98 (2.84)	10.81 (2.82)	.009[e]
Week 3	9.08 (2.45)	10.94 (3.05)	.008[e]
Week 4	9.30 (2.59)	11.68 (3.61)	.002[f]
P value[g]	.99	.01	—[h]
Participants with a medium/high (≥12) MDSS index, n (%)			
Week 1	10 (25.0)	10 (32.3)	.50
Week 2	7 (17.5)	13 (41.9)	.02
Week 3	5 (12.5)	13 (41.9)	.005
Week 4	7 (17.5)	17 (54.8)	.001
P value	.41	.07	—

[a]MDSS: Mediterranean Diet Serving Score.

[b]CG: control group.

[c]IG: intervention group.

[d]*P* value in columns: MDSS index—intragroup differences (CG and IG) throughout the 4 weeks of follow-up in the study, evaluated with the Friedman test; number of participants with a medium/high (≥12) MDSS index—intragroup differences (CG and IG) in week 1 versus week 4, evaluated with the chi-square test. $P<.05$ was considered significant.

[e]Evaluated with the Student *t* test.

[f]Evaluated with the Mann-Whitney *U* test.

[g]*P* value in rows: MDSS index—intergroup differences (CG versus IG) in each of the 4 study weeks; number of participants with a medium/high (≥12) MDSS index—intergroup differences (CG versus IG) in each of the 4 study weeks, evaluated with the chi-square test. $P<.05$ was considered significant.

[h]Not applicable.

Table 3. Comparison of the MDSS[a] index in weeks 2, 3, and 4 of follow-up with that in week 1 (baseline) for the CG[b] and the IG[c].

Group and week	MDSS index, mean (SD)	*P* value[d]
CG		
Week 1	9.30 (2.40)	Reference
Week 2	8.98 (2.84)	.34[e]
Week 3	9.08 (2.45)	.58[e]
Week 4	9.30 (2.59)	.99[f]
IG		
Week 1	9.65 (2.23)	Reference
Week 2	10.81 (2.82)	.005[f]
Week 3	10.94 (3.05)	.004[e]
Week 4	11.68 (3.61)	.001[f]

[a]MDSS: Mediterranean Diet Serving Score.

[b]CG: control group.

[c]IG: intervention group.

[d]Intragroup differences (CG and IG) in week 1 versus weeks 2, 3, and 4. $P<.02$ (.05/3) was considered significant (penalizing *P* values with Bonferroni adjustment for multiple comparisons).

[e]Evaluated with the Wilcoxon test.

[f]Evaluated with the Student *t* test.

Regarding intragroup modifications, there were no significant statistical differences in the MDSS index in the CG, either throughout the 4 weeks of study, ranging from mean 9.30 (SD 2.59) in weeks 1 and 4 to mean 8.98 (SD 2.84) in week 2 (Table 2) or in weeks 2, 3, and 4 when compared to week 1 (baseline; Table 3). In the IG, there were significant statistical differences in the MDSS index throughout the 4 weeks of study, ranging from mean 9.65 (2.23) in week 1 to mean 11.68 (SD 3.61) in week 4 (Table 2). Compared to week 1 (baseline), the differences were statistically significant from week 2 onward: with 1.16, 1.29, and 2.03 points of improvement at weeks 2, 3, and 4, respectively (Table 3). There were no significant statistical differences in the number of participants with a medium/high (≥12) MDSS index (week 1 vs week 4) in either the CG or the IG.

Regarding intergroup modifications, there were significant statistical differences for both the MDSS index and the number of participants with a medium/high (≥12) MDSS index in the CG versus the IG (in favor of the IG) from week 2 onward (no significant differences in week 1). For the MDSS index, we found 1.83, 1.86, and 2.38 points of improvement at weeks 2, 3, and 4, respectively; for the number of participants with a medium/high (≥12) MDSS index, we found 24.4, 29.4, and 37.3 percentage points of improvement at weeks 2, 3, and 4, respectively (Table 2).

Effect of the Intervention With the e-12HR App in Terms of Variation in Food Groups

Table 4 shows the number of participants who met the consumption criteria for each food group [19] in the CG and the IG throughout the 4 weeks of follow-up.

Table 4. Number of participants who met the consumption criteria of the MDSS[a] index for each food group throughout the 4 weeks of follow-up (CG[b] n=40, IG[c] n=31).

Food group MDSS index consumption criteria and study group	Week				*P* value[d]
	Week 1	Week 2	Week 3	Week 4	
Fruits (1-6 servings/day)					
CG, n (%)	26 (65.0)	27 (67.5)	27 (67.5)	27 (67.5)	.81[e]
IG, n (%)	24 (77.4)	27 (87.1)	28 (90.3)	28 (90.3)	.17[e]
P value[f]	.26[g]	.06[g]	.02[g]	.02[g]	—[h]
Vegetables (≥2 servings/day)					
CG, n (%)	1 (2.5)	3 (7.5)	3 (7.5)	4 (10.0)	.36[i]
IG, n (%)	3 (9.7)	8 (25.8)	10 (32.3)	14 (45.2)	.002[e]
P value	.31[j]	.05[j]	.01[g]	.001[g]	—
Cereals (1-6 servings/day of breakfast cereals, pasta, rice, and bread)					
CG, n (%)	40 (100.0)	37 (92.5)	39 (97.5)	37 (92.5)	.241[i]
IG, n (%)	31 (100.0)	30 (96.8)	30 (96.8)	30 (96.8)	.99[i]
P value	—	.63[j]	.99[j]	.63[j]	—
Olive oil (1-4 servings/day)					
CG, n (%)	31 (77.5)	33 (82.5)	33 (82.5)	34 (85.0)	.390[e]
IG, n (%)	26 (83.9)	26 (83.9)	28 (90.3)	27 (87.1)	.99[h]
P value	.50[f]	.88[f]	.50[i]	.99[i]	—
Milk and dairy products (1-3 servings/day)					
CG, n (%)	33 (82.5)	32 (80.0)	32 (80.0)	33 (82.5)	.99[e]
IG, n (%)	26 (83.9)	27 (87.1)	27 (87.1)	28 (90.3)	.71[i]
P value	.88[g]	.43[g]	.43[g]	.50[j]	—
Nuts (1-2 servings/day)					
CG, n (%)	3 (7.5)	2 (5.0)	2 (5.0)	5 (12.5)	.71[i]
IG, n (%)	3 (9.7)	5 (16.1)	3 (9.7)	11 (35.5)	.02[e]
P value	.99[j]	.23[j]	.65[j]	.02[g]	—
Fermented beverages (0-2 serving/day of wine and beer)					
CG, n (%)	36 (90.0)	36 (90.0)	37 (92.5)	37 (92.5)	.99[i]
IG, n (%)	27 (87.1)	27 (87.1)	29 (93.5)	29 (93.5)	.67[i]
P value	.72[j]	.72[j]	.99[j]	.99[j]	—
Potatoes (≤3 servings/week)					
CG, n (%)	32 (80.0)	27 (67.5)	26 (65.0)	25 (62.5)	.08[e]
IG, n (%)	23 (74.2)	19 (61.3)	17 (54.8)	19 (61.3)	.28[e]
P value	.56[g]	.59[g]	.39[g]	.92[g]	—
Legumes (≥2 servings/week)					
CG, n (%)	15 (37.5)	13 (32.5)	17 (42.5)	14 (35.0)	.82[e]
IG, n (%)	13 (41.9)	18 (58.1)	22 (71.0)	19 (61.3)	.13[e]

Food group MDSS index consumption criteria and study group	Week				*P* value[d]
	Week 1	Week 2	Week 3	Week 4	
P value	.70[g]	.03[g]	.02[g]	.02[g]	—
Eggs (2-4 servings/week)					
CG, n (%)	21 (52.5)	24 (60.0)	22 (55.0)	26 (65.0)	.26[e]
IG, n (%)	19 (61.3)	20 (64.5)	17 (54.8)	17 (54.8)	.61[e]
P value	.46[g]	.70[g]	.99[g]	.39[g]	—
Fish (≥2 servings/week)					
CG, n (%)	31 (77.5)	28 (70.0)	32 (80.0)	35 (87.5)	.24[e]
IG, n (%)	28 (90.3)	27 (87.1)	26 (83.9)	27 (87.1)	.99[h]
P value	.15[g]	.09[g]	.68[g]	.99[j]	—
White meat (2-3 servings/week)					
CG, n (%)	20 (50.0)	11 (27.5)	10 (25.0)	11 (27.5)	.04[e]
IG, n (%)	10 (32.3)	9 (29.0)	12 (38.7)	5 (16.1)	.14[e]
P value	.13[g]	.89[g]	.22[g]	.26[g]	—
Red meat (<2 servings/week of pork, beef, lamb, and processed meat)					
CG, n (%)	9 (22.5)	9 (22.5)	7 (17.5)	6 (15.0)	.39[e]
IG, n (%)	6 (19.4)	10 (32.3)	7 (22.6)	8 (25.8)	.54[e]
P value	.75[g]	.36[g]	.59[g]	.26[g]	—
Sweets (≤2 servings/week)					
CG, n (%)	19 (47.5)	15 (37.5)	20 (50.0)	20 (50.0)	.82[e]
IG, n (%)	15 (48.4)	15 (48.4)	14 (45.2)	17 (54.8)	.61[e]
P value	.94[g]	.36[g]	.69[g]	.69[g]	—

[a]MDSS: Mediterranean Diet Serving Score.

[b]CG: control group.

[c]IG: intervention group.

[d]*P* value in columns: intergroup differences (CG vs IG) in each of the 4 study weeks. *P*<.05 was considered significant.

[e]Evaluated with the chi-square test.

[f]*P* value in rows: intragroup differences (CG versus IG) in week 1 versus week 4. *P*<.05 was considered significant.

[g]Evaluated with the chi-square test.

[h]Not applicable.

[i,j]Evaluated with the Fisher exact test.

Regarding intergroup modifications, statistically significant differences (CG vs GI) were observed throughout the study period in 4 food groups: fruits (weeks 3 and 4), vegetables (weeks 2, 3, and 4), nuts (week 4), and legumes (weeks 2, 3, and 4). In these 4 food groups, at week 4, the number of participants meeting the recommendations was higher in the IG compared to the CG, with 22.8% for fruits, 35.2% for vegetables, 23.0% for nuts, and, finally, 26.3% for legumes (Table 4).

Regarding intragroup modifications, statistically significant differences (week 1 vs week 4) were observed in the CG for white meat and in the IG for vegetables and nuts.

Usability Rating Questionnaire for the e-12HR App

Of 71 participants, 45 (63%) returned the completed questionnaire: 25 (63%) of 40 from the CG and 20 (65%) of 31 from the IG. The responses of the 45 PCPs are shown in Table 5.

Table 5. Responses to the usability rating questionnaire for the e-12HR[a] app (CG[b] n=25, IG[c] n=20).

Questions and groups	Answers
1. Easy to complete (strongly agree + agree)	
CG, n (%)	25 (100)
IG, n (%)	20 (100)
P value[d]	—[e]
2. Understandable questions (strongly agree + agree)	
CG, n (%)	24 (96)
IG, n (%)	20 (100)
P value	.99[f]
3. Understandable feedback only for the IG (strongly agree + agree)	
CG, n (%)	—
IG, n (%)	18 (90)
P value	—
4. I would be willing to complete again (strongly agree + agree)	
CG, n (%)	20 (80)
IG, n (%)	12 (60)
P value	0.141[g]
5. Time to complete (≤3 minutes/day)	
CG, n (%)	19 (76)
IG, n (%)	15 (75)
P value	.99[f]

[a]e-12HR: Electronic 12-Hour Dietary Recall.

[b]CG: control group.

[c]IG: intervention group.

[d]Differences between subgroups. $P<.05$ was considered significant.

[e]Not applicable.

[f]Evaluated with the Fisher exact test.

[g]Evaluated with the chi-square test.

No statistically significant differences were observed for any of the questions on the questionnaire (CG vs IG). All PCPs indicated that the e-12HR app was easy to complete, and most of them responded that the app contained questions that were understandable (CG n=24, 96%; IG n=20, 100%) and that feedback was understandable (only for the IG, n=18, 90%). Furthermore, a large percentage were willing to use the app again (CG n=20, 80; IG n=12, 60%). Completion of the app could be considered to have taken 3 minutes or less (CG n=19, 76%; IG n=15, 75%).

Discussion

Principal Findings

In relation to the main objective, there were significant statistical differences between the 2 groups in this study. At week 4 (no significant differences in week 1, baseline), the values were higher in the IG compared to the CG by 25.6% for the MDSS index (Table 2); by 213.1 for the number of participants with a medium/high (≥12) MDSS index (Table 2); and by 33.8% for fruits, 352.0% for vegetables, 184.0% for nuts, and 75.1% for legumes for the number of participants meeting the recommendations for specific food groups (Table 4).

Regarding the secondary objectives, first, the answers to the questions of the usability rating questionnaire for the e-12HR app were satisfactory. According to the questionnaire, the daily use time of the app was about 3 minutes or less per day for most respondents (Table 5). When working with smartphone apps, usability is an important aspect to consider. According to health care professionals, there are 3 principal criteria for selecting a "nutrition and diet" app for clients/patients, which are [31] ease of use (satisfactory data were obtained in this study), free of charge, and validation (the e-12HR app is free to download and has been previously validated [32-36]). Second, at baseline (week 1), AMD for PCPs was medium (mean MDSS index 9.45, SD 2.32) and 66% of participants had a medium/high MDSS index (≥9).

Overview

To begin with, workplace interventions are an excellent strategy to promote a healthy diet, considering that health care professionals spend long hours in their professional activity and often have 1 or more meals during their working day. At the hospital level, interventions have been implemented to facilitate access to and choice of healthy foods during the working day, such as modifying the availability of foods served in the canteen, subsidizing the cost of fresh fruits and vegetables (which are often more expensive than less healthy alternatives) [37], or implementing traffic light labeling (green: healthy; yellow: less healthy; red: unhealthy) [37,38]. In a study by Thorndike et al. [38] (in a hospital in Boston, Massachusetts, USA), their intervention also included personalized automated messages using a platform that automatically generated 2 weekly emails with feedback on previous purchases in the hospital cafeteria and lifestyle advice. Significant statistical increases were observed in green-labeled food purchases and decreases in red-labeled food purchases among the IG compared to the CG in the hospital cafeteria throughout the study period.

Thorndike et al's [38] intervention was based on information about food eaten only in the hospital cafeteria (without considering other food consumed outside the hospital), so its scope was limited. However, to date, no interventions to promote a healthy diet among PCPs have been implemented; for example, the workplace intervention strategies discussed before would be difficult to implement in Spain because health centers are widely distributed throughout the territory and do not usually have a cafeteria or restaurant. Considering these difficulties as a possible alternative strategy, this study was the first to assess the ability of a smartphone app to improve the dietary habits among health care professionals (specifically, PCPs). Several randomized controlled clinical trials have used an app to improve AMD in Spanish adults, such as patients of health care centers (the EVIDENT II app [39,40] and the SalBi Educa Nutrition app [41]) and patients with type 2 diabetes mellitus (the EVIDENT II app [42]), but not in health care professionals.

Comparison With Prior Work

As previously mentioned, the e-12HR app has also been evaluated among Spanish university students (health sciences and non–health sciences) [19,20]. In relation to the main objective, the results obtained by the e-12HR app among PCPs compared to university students were (1) similar for the MDSS index (the increase among PCPs was 25.6% for the MDSS index, as shown in Table 2, and among university students was 17.4% [19] or 25.7% [20]), (2) more positive for the number of participants with a medium/high MDSS index (the increase among PCPs was 213.1%, as shown in Table 2, and among university student was 61.9% [19] or 74.5% [20]), and (3) less positive regarding the number of participants meeting the recommendations for specific food groups (improvements in 4 food group among PCPs, as shown in Table 4, and in 7 food groups among Spanish university students).

Regarding the secondary objectives, first, similar results were obtained among Spanish university students for the answers to the questions of the usability rating questionnaire for the e-12HR app. Second, at baseline (week 1), the mean MDSS index of 9.45 (SD 2.32) and the number of participants with a medium/high MDSS index (66%) among PCPs (Table 1) were higher compared with the data from health sciences students [20] during the same follow-up period (mean MDSS index 7.59, SD 2.72; percentage of participants with a medium/high MDSS index=33.4%). Significant statistical differences were found (P<.05) for both the MDSS index, which was evaluated with the Mann-Whitney U test, and the number of participants with a medium/high MDSS index, which was evaluated with the chi-square test: PCPs showed an improvement of 24.5% for the MDSS index and 98.2% for the number of participants with a moderate/high MDSS index (PCPs vs health science university students). This comparison must be made with caution, since the sample of PCPs was made up of doctors and nurses and the sample of health sciences university students was made up of students from the faculties of medicine and pharmacy. In addition, in a previous study by Sentenach-Carbo et al [43] among Spanish PCPs, the number of participants with medium/high AMD was lower: 55% versus 66.2%. It should, however, be considered that the adherence index used by both studies was different: the MDSS index was used in this study, and the validated 14-point Mediterranean diet adhesion screener was used in the Prevention with Mediterranean Diet (PREDIMED) study.

Limitations

This study presents several limitations, and the first is internal validation. This included the fact that the e-12HR app is a self-reporting method and presents the limitations inherent in this type of tool, which have been amply described [44-50]. Due to the nature of the intervention, on the one hand, only the investigator who performed the statistical analysis of the data was blinded (but not the PCPs) and, on the other hand, it was not possible to guarantee that the participants were not using another nutrition app during the study period.

Regarding external validation, the dietary program was short (4 weeks), and the long-term evolution of the study variables is unknown. In addition, the evaluation of the usability of the app was based on the responses of those participants who completed the study; however, there could be differences in the perception of usability between responsive (those who completed the study) and nonresponsive (those who did not) participants.

Future Research

According to Recio-Rodríguez et al [40], future research related to the effectiveness of apps to improve diet should clarify the possible effects of certain factors (eg, age, gender, or educational level). Therefore, in future studies, the research team intends to evaluate the effectiveness of the e-12HR app in increasing the MDSS index in different strata of PCPs—for example, examining results according to gender, age, occupational category, and the BMI as possible moderating variables and according to technological perception and technological familiarity as possible mediating variables.

Conclusion

At baseline, Spanish PCPs presented medium AMD (measured as the MDSS index and the number of participants with a

medium/high MDSS index). Throughout the study period, in the short term, the use of the e-12HR app (an easy-to-implement and low-cost intervention) showed moderate improvements in the MDSS index and remarkable improvements in the number of participants with a medium/high MDSS index; in addition, PCPs responded positively to questions about the usability of the app. These results support recommending the use of the e-12HR app as a tool to contribute to improving diet and preventing NCDs among PCPS, which, at the same time, could positively influence patient dietary behavior and prevent diet-related NCDs among the patients. From the point of view of health care organizations, the prevention of NCDs among PCPs could, in addition, lead to higher personal and job satisfaction and fewer sickness-related absences from work; for this reason, health organizations themselves should be more involved in the recommendations to use tools such as the one analyzed in this study among their own workers.

Acknowledgments

This work was supported in part by the SAMFyC Foundation (grant number TD180/21). The SAMFyC Foundation had no role in the design, analysis or writing of this article. The research team would like to thank the PCPs who participated in this study.

Authors' Contributions

LMB performed the conception and design of the study, developed the app, analyzed and interpreted data, and wrote the paper. PMR and MDGP were involved in data collection and interpretation of the data and contributed to drafting the manuscript. The authors/evaluators are the owners and developers of the e-12HR app.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Real images of the e-12HR app (nonfeedback and feedback versions). e-12HR: Electronic 12-Hour Dietary Recall.
[DOCX File , 605 KB - mhealth_v12i1e49302_app1.docx]

Multimedia Appendix 2
Usability rating questionnaire for the e-12HR app. e-12HR: Electronic 12-Hour Dietary Recall.
[DOCX File , 28 KB - mhealth_v12i1e49302_app2.docx]

Multimedia Appendix 3
Consort checklist.
[PDF File (Adobe PDF File), 3262 KB - mhealth_v12i1e49302_app3.pdf]

References

1. Noncommunicable diseases progress monitor 2022. Global report. World Health Organization. 2022. URL: https://www.who.int/publications/i/item/9789240047761 [accessed 2023-05-01]
2. Noble N, Paul C, Turon H, Oldmeadow C. Which modifiable health risk behaviours are related? A systematic review of the clustering of smoking, nutrition, alcohol and physical activity ('SNAP') health risk factors. Prev Med 2015 Dec;81:16-41. [doi: 10.1016/j.ypmed.2015.07.003] [Medline: 26190368]
3. Malekzadeh A, Michels K, Wolfman C, Anand N, Sturke R. Strengthening research capacity in LMICs to address the global NCD burden. Glob Health Action 2020 Dec 31;13(1):1846904 [FREE Full text] [doi: 10.1080/16549716.2020.1846904] [Medline: 33373280]
4. Melnyk BM, Kelly SA, Stephens J, Dhakal K, McGovern C, Tucker S, et al. Interventions to improve mental health, well-being, physical health, and lifestyle behaviors in physicians and nurses: a systematic review. Am J Health Promot 2020 Nov;34(8):929-941 [FREE Full text] [doi: 10.1177/0890117120920451] [Medline: 32338522]
5. Pardo A, McKenna J, Mitjans A, Camps B, Aranda-García S, Garcia-Gil J, et al. Physical activity level and lifestyle-related risk factors from Catalan physicians. J Phys Act Health 2014 Jul;11(5):922-929. [doi: 10.1123/jpah.2012-0249] [Medline: 23799257]
6. Mittal TK, Cleghorn CL, Cade JE, Barr S, Grove T, Bassett P, et al. A cross-sectional survey of cardiovascular health and lifestyle habits of hospital staff in the UK: do we look after ourselves? Eur J Prev Cardiol 2018 Mar;25(5):543-550 [FREE Full text] [doi: 10.1177/2047487317746320] [Medline: 29198137]
7. González-Sosa S, Ruiz-Hernández JJ, Domínguez-Rivero S, Águila-Fernández E, Godoy-Díaz D, Santana-Báez S, et al. Adherence to the Mediterranean diet in health personnel from the province of Las Palmas. Rev Clin Esp (Barc) 2021 Dec;221(10):569-575. [doi: 10.1016/j.rceng.2021.01.008] [Medline: 34305037]

8. Worley V, Fraser P, Allender S, Bolton KA. Describing workplace interventions aimed to improve health of staff in hospital settings - a systematic review. BMC Health Serv Res 2022 Apr 07;22(1):459 [FREE Full text] [doi: 10.1186/s12913-021-07418-9] [Medline: 35392894]
9. Blake H, Zhou D, Batt ME. Five-year workplace wellness intervention in the NHS. Perspect Public Health 2013 Sep;133(5):262-271. [doi: 10.1177/1757913913489611] [Medline: 23771680]
10. Mahler L, Sebo P, Favrod-Coune T, Moussa A, Cohidon C, Broers B. The prevalence of five lifestyle risk factors in primary care physicians: a cross-sectional study in Switzerland. Prev Med Rep 2022 Apr;26:101740 [FREE Full text] [doi: 10.1016/j.pmedr.2022.101740] [Medline: 35251911]
11. Znyk M, Polańska K, Wojtysiak P, Szulc M, Bąk-Romaniszyn L, Makowiec-Dąbrowska T, et al. Predictors of counselling related to a healthy lifestyle carried out by a general practitioner. Int J Environ Res Public Health 2019 Nov 14;16(22):4475 [FREE Full text] [doi: 10.3390/ijerph16224475] [Medline: 31739411]
12. Rippe JM. Lifestyle medicine: the health promoting power of daily habits and practices. Am J Lifestyle Med 2018;12(6):499-512 [FREE Full text] [doi: 10.1177/1559827618785554] [Medline: 30783405]
13. Oberg EB, Frank E. Physicians' health practices strongly influence patient health practices. J R Coll Physicians Edinb 2009 Dec;39(4):290-291 [FREE Full text] [doi: 10.4997/JRCPE.2009.422] [Medline: 21152462]
14. Frank E, Segura C, Shen H, Oberg E. Predictors of Canadian physicians' prevention counseling practices. Can J Public Health 2010;101(5):390-395 [FREE Full text] [doi: 10.1007/BF03404859] [Medline: 21214054]
15. Vrkatić A, Grujičić M, Jovičić-Bata J, Novaković B. Nutritional knowledge, confidence, attitudes towards nutritional care and nutrition counselling practice among general practitioners. Healthcare (Basel) 2022 Nov 07;10(11):2222 [FREE Full text] [doi: 10.3390/healthcare10112222] [Medline: 36360563]
16. Sebo P, Bouvier Gallacchi M, Goehring C, Künzi B, Bovier PA. Use of tobacco and alcohol by Swiss primary care physicians: a cross-sectional survey. BMC Public Health 2007 Jan 12;7:5 [FREE Full text] [doi: 10.1186/1471-2458-7-5] [Medline: 17222332]
17. Faruque M, Barua L, Banik PC, Sultana S, Biswas A, Alim A, et al. Prevalence of non-communicable disease risk factors among nurses and para-health professionals working at primary healthcare level of Bangladesh: a cross-sectional study. BMJ Open 2021 Mar 19;11(3):e043298 [FREE Full text] [doi: 10.1136/bmjopen-2020-043298] [Medline: 33741665]
18. Campos SF, Dos Santos LC, Lopes MS, de Freitas PP, Lopes AC. Consumption of ultra-processed foods and nutritional profile in a health promotion service of primary care. Public Health Nutr 2021 Oct;24(15):5113-5126. [doi: 10.1017/S1368980021002202] [Medline: 34034849]
19. Béjar LM, García-Perea MD, Mesa-Rodríguez P. Evaluation of an application for mobile telephones (e-12HR) to increase adherence to the Mediterranean diet in university students: a controlled, randomized and multicentric study. Nutrients 2022 Oct 08;14(19):4196 [FREE Full text] [doi: 10.3390/nu14194196] [Medline: 36235848]
20. Béjar LM, Mesa-Rodríguez P, Quintero-Flórez A, Ramírez-Alvarado MDM, García-Perea MD. Effectiveness of a smartphone app (e-12HR) in improving adherence to the Mediterranean diet in Spanish university students by age, gender, field of study, and body mass index: a randomized controlled trial. Nutrients 2023 Mar 30;15(7):1688 [FREE Full text] [doi: 10.3390/nu15071688] [Medline: 37049528]
21. WHO and FAO. Diet, nutrition, and the prevention of chronic diseases (report of a joint WHO and FAO expert consultation). World Health Organization. URL: http://apps.who.int/iris/bitstream/handle/10665/42665/WHO_TRS_916.pdf;jsessionid=C124FDE9271A89C8E749C555CB173BCB?sequence=1 [accessed 2023-05-01]
22. Gallus S, Bosetti C, La Vecchia C. Mediterranean diet and cancer risk. Eur J Cancer Prev 2004 Oct;13(5):447-452. [doi: 10.1097/00008469-200410000-00013] [Medline: 15452458]
23. Salas-Salvadó J, Fernández-Ballart J, Ros E, Martínez-González MA, Fitó M, Estruch R, PREDIMED Study Investigators. Effect of a Mediterranean diet supplemented with nuts on metabolic syndrome status: one-year results of the PREDIMED randomized trial. Arch Intern Med 2008 Dec 08;168(22):2449-2458. [doi: 10.1001/archinte.168.22.2449] [Medline: 19064829]
24. Estruch R, Ros E, Salas-Salvadó J, Covas M, Corella D, Arós F, PREDIMED Study Investigators. Primary prevention of cardiovascular disease with a Mediterranean diet supplemented with extra-virgin olive oil or nuts. N Engl J Med 2018 Jun 21;378(25):e34 [FREE Full text] [doi: 10.1056/NEJMoa1800389] [Medline: 29897866]
25. Piepoli MF, Hoes AW, Agewall S, Albus C, Brotons C, Catapano AL, ESC Scientific Document Group. Eur Heart J 2016 Aug 01;37(29):2315-2381 [FREE Full text] [doi: 10.1093/eurheartj/ehw106] [Medline: 27222591]
26. Dernini S, Berry EM, Serra-Majem L, La Vecchia C, Capone R, Medina FX, et al. Med Diet 4.0: the Mediterranean diet with four sustainable benefits. Public Health Nutr 2017 May;20(7):1322-1330 [FREE Full text] [doi: 10.1017/S1368980016003177] [Medline: 28003037]
27. Berry EM, Dernini S, Burlingame B, Meybeck A, Conforti P. Food security and sustainability: can one exist without the other? Public Health Nutr 2015 Sep;18(13):2293-2302 [FREE Full text] [doi: 10.1017/S136898001500021X] [Medline: 25684016]
28. Burlingame B, Gitz V, Meybeck A, Office of Assistant Director-General (Agriculture Department). Mediterranean Food Consumption Patterns. Diet, Environment, Society, Economy and Health. Rome, Italy: Food and Agriculture Organization; 2015.

29. Monteagudo C, Mariscal-Arcas M, Rivas A, Lorenzo-Tovar ML, Tur JA, Olea-Serrano F. Proposal of a Mediterranean diet serving score. PLoS One 2015;10(6):e0128594 [FREE Full text] [doi: 10.1371/journal.pone.0128594] [Medline: 26035442]
30. Vasiloglou MF, Lu Y, Stathopoulou T, Papathanail I, Fäh D, Ghosh A, et al. Assessing Mediterranean diet adherence with the smartphone: the Medipiatto Project. Nutrients 2020 Dec 07;12(12):3763 [FREE Full text] [doi: 10.3390/nu12123763] [Medline: 33297550]
31. Vasiloglou MF, Christodoulidis S, Reber E, Stathopoulou T, Lu Y, Stanga Z, et al. Nutrients 2020 Jul 24;12(8):2214 [FREE Full text] [doi: 10.3390/nu12082214] [Medline: 32722339]
32. Bejar LM, Sharp BN, García-Perea MD. The e-EPIDEMIOLOGY mobile phone app for dietary intake assessment: comparison with a food frequency questionnaire. JMIR Res Protoc 2016 Nov 02;5(4):e208 [FREE Full text] [doi: 10.2196/resprot.5782] [Medline: 27806922]
33. Béjar LM, Vázquez-Limón E. Is there any alternative to traditional food frequency questionnaire for evaluating habitual dietary intake? Nutr Hosp 2017 Jul 28;34(4):990-888. [doi: 10.20960/nh.650] [Medline: 29095013]
34. Béjar LM. First evaluation steps of a new method for dietary intake estimation regarding a list of key food groups in adults and in different sociodemographic and health-related behaviour strata. Public Health Nutr 2017 Oct;20(15):2660-2669 [FREE Full text] [doi: 10.1017/S1368980017001641] [Medline: 28789723]
35. Béjar LM, Reyes Ó, García-Perea MD. Electronic 12-hour dietary recall (e-12HR): comparison of a mobile phone app for dietary intake assessment with a food frequency questionnaire and four dietary records. JMIR Mhealth Uhealth 2018 Jun 15;6(6):e10409 [FREE Full text] [doi: 10.2196/10409] [Medline: 29907555]
36. Béjar LM, García-Perea MD, Reyes Ó, Vázquez-Limón E. Relative validity of a method based on a smartphone app (electronic 12-hour dietary recall) to estimate habitual dietary intake in adults. JMIR Mhealth Uhealth 2019 Apr 11;7(4):e11531 [FREE Full text] [doi: 10.2196/11531] [Medline: 30973343]
37. Thorndike AN, Riis J, Levy DE. Social norms and financial incentives to promote employees' healthy food choices: a randomized controlled trial. Prev Med 2016 May;86:12-18 [FREE Full text] [doi: 10.1016/j.ypmed.2016.01.017] [Medline: 26827617]
38. Thorndike AN, McCurley JL, Gelsomin ED, Anderson E, Chang Y, Porneala B, et al. Automated behavioral workplace intervention to prevent weight gain and improve diet: the ChooseWell 365 randomized clinical trial. JAMA Netw Open 2021 Jun 01;4(6):e2112528 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.12528] [Medline: 34097048]
39. Recio-Rodriguez JI, Agudo-Conde C, Martin-Cantera C, González-Viejo MN, Fernandez-Alonso MDC, Arietaleanizbeaskoa MS, EVIDENT Investigators. Short-term effectiveness of a mobile phone app for increasing physical activity and adherence to the Mediterranean diet in primary care: a randomized controlled trial (EVIDENT II Study). J Med Internet Res 2016 Dec 19;18(12):e331 [FREE Full text] [doi: 10.2196/jmir.6814] [Medline: 27993759]
40. Recio-Rodriguez JI, Agudo Conde C, Calvo-Aponte MJ, Gonzalez-Viejo N, Fernandez-Alonso C, Mendizabal-Gallastegui N, EVIDENT Investigators. The effectiveness of a smartphone application on modifying the intakes of macro and micronutrients in primary care: a randomized controlled trial. The EVIDENT II Study. Nutrients 2018 Oct 10;10(10):1473 [FREE Full text] [doi: 10.3390/nu10101473] [Medline: 30309008]
41. Gonzalez-Ramirez M, Sanchez-Carrera R, Cejudo-Lopez A, Lozano-Navarrete M, Salamero Sánchez-Gabriel E, Torres-Bengoa MA, et al. Short-term pilot study to evaluate the impact of SalBi Educa nutrition app in macronutrients intake and adherence to the Mediterranean diet: randomized controlled trial. Nutrients 2022 May 14;14(10):2061 [FREE Full text] [doi: 10.3390/nu14102061] [Medline: 35631202]
42. Alonso-Domínguez R, García-Ortiz L, Patino-Alonso MC, Sánchez-Aguadero N, Gómez-Marcos MA, Recio-Rodríguez JI. Effectiveness of A a multifactorial intervention in increasing adherence to the Mediterranean diet among patients with diabetes mellitus type 2: a controlled and randomized study (EMID Study). Nutrients 2019 Jan 14;11(1):162 [FREE Full text] [doi: 10.3390/nu11010162] [Medline: 30646500]
43. Sentenach-Carbo A, Batlle C, Franquesa M, García-Fernandez E, Rico L, Shamirian-Pulido L, et al. Adherence of Spanish primary physicians and clinical practise to the Mediterranean diet. Eur J Clin Nutr 2019 Jul;72(Suppl 1):92-98. [doi: 10.1038/s41430-018-0314-8] [Medline: 30487567]
44. Willett W. Nutritional Epidemiology. 3rd Ed. New York, NY: Oxford University Press; 2013.
45. Forster H, Fallaize R, Gallagher C, O'Donovan CB, Woolhead C, Walsh MC, et al. Online dietary intake estimation: the Food4Me food frequency questionnaire. J Med Internet Res 2014 Jun 09;16(6):e150 [FREE Full text] [doi: 10.2196/jmir.3105] [Medline: 24911957]
46. Rutishauser IHE. Dietary intake measurements. Public Health Nutr 2005 Oct;8(7A):1100-1107. [doi: 10.1079/phn2005798] [Medline: 16277822]
47. Tucker KL, Smith CE, Lai C, Ordovas JM. Quantifying diet for nutrigenomic studies. Annu Rev Nutr 2013;33:349-371 [FREE Full text] [doi: 10.1146/annurev-nutr-072610-145203] [Medline: 23642200]
48. Gibson R. Principles of Nutritional Assessment, 2nd Ed. New York, NY: Oxford University Press; 2005.
49. Martín-Moreno JM, Gorgojo L. [Assessment of dietary intake at the population level through individual questionnaires: methodological shadows and lights]. Rev Esp Salud Publica 2007;81(5):507-518 [FREE Full text] [doi: 10.1590/s1135-57272007000500007] [Medline: 18274354]

50. Dhurandhar NV, Schoeller D, Brown AW, Heymsfield SB, Thomas D, Sørensen TIA, Energy Balance Measurement Working Group. Energy balance measurement: when something is not better than nothing. Int J Obes (Lond) 2015 Jul 13;39(7):1109-1113 [FREE Full text] [doi: 10.1038/ijo.2014.199] [Medline: 25394308]

Abbreviations

AMD: adherence to the Mediterranean diet
CG: control group
e-12HR: Electronic 12-Hour Dietary Recall
IG: intervention group
MDSS: Mediterranean Diet Serving Score
NCD: noncommunicable disease
PCP: primary care professional

Edited by L Buis; submitted 24.05.23; peer-reviewed by K DiFilippo, E Hernández, M Andrade; comments to author 01.09.23; revised version received 14.09.23; accepted 20.12.23; published 08.01.24.

Please cite as:
Béjar LM, Mesa-Rodríguez P, García-Perea MD
Short-Term Effect of a Health Promotion Intervention Based on the Electronic 12-Hour Dietary Recall (e-12HR) Smartphone App on Adherence to the Mediterranean Diet Among Spanish Primary Care Professionals: Randomized Controlled Clinical Trial
JMIR Mhealth Uhealth 2024;12:e49302
URL: https://mhealth.jmir.org/2024/1/e49302
doi:10.2196/49302
PMID:38190226

Original Paper

Quality Assessment of Smartphone Medication Management Apps in France: Systematic Search

Mickael Toïgo[1], MD; Julie Marc[1], MSc, MD; Maurice Hayot[2], MD, PhD; Lionel Moulis[3,4], MD; Francois Carbonnel[1,5], MD, PhD

[1]Department of General Practice, Univ Montpellier, Montpellier, France

[2]PhyMedExp, Univ Montpellier, CNRS, INSERM, CHU Montpellier, Montpellier, France

[3]Clinical Research and Epidemiology Unit, Department of Public Health, Univ Montpellier, CHU Montpellier, Montpellier, France

[4]Pathogenesis and Control of Chronic and Emerging Infections, Univ Montpellier, INSERM, EFS, University of Antilles, Montpellier, France

[5]Desbrest Institute of Epidemiology and Public Health, Univ Montpellier, INSERM, Montpellier, France

Corresponding Author:
Francois Carbonnel, MD, PhD
Desbrest Institute of Epidemiology and Public Health
Univ Montpellier, INSERM
Campus Santé, IURC, 641 avenue du Doyen Gaston Giraud
Montpellier, 34090
France
Phone: 33 684014834
Email: Francois.carbonnel@umontpellier.fr

Abstract

Background: Adherence to medication is estimated to be around 50% for chronically ill patients in high-income countries. Improving the effectiveness of adherence interventions could have a far greater impact on population health than any improvement in specific medical treatments. Mobile health (mHealth) is one of the most effective solutions for helping patients improve their medication intake, notably through the use of mobile apps with reminder systems. With more than 327,000 apps available in the mHealth field, it is difficult for health care professionals and patients alike to choose which apps to recommend and use.

Objective: We aim to carry out a systematic search of medication management smartphone apps available in France that send reminders to patients and assess their quality using a validated scale.

Methods: Mobile apps were identified in October and November 2022 after a systematic keyword search on the 2 main app download platforms: App Store (Apple Inc) and Google Play Store. Inclusion criteria were free availability, date of last update, and availability in French. Next, 2 health care professionals independently evaluated the included apps using the French version of the Mobile App Rating Scale (MARS-F), an objective scoring system validated for assessing the overall quality of apps in the mHealth field. An intraclass correlation coefficient was calculated to determine interrater reliability.

Results: In total, 960 apps were identified and 49 were selected (25 from the App Store and 24 from the Google Play Store). Interrater reliability was excellent (intraclass correlation coefficient 0.92; 95% CI 0.87-0.95; P<.001). The average MARS-F score was 3.56 (SD 0.49) for apps on the App Store and 3.51 (SD 0.46) for those on the Google Play Store, with 10 apps scoring above 4 out of 5. Further, 2 apps were tested in at least one randomized controlled trial and showed positive results. The 2 apps with the highest ratings were *Mediteo rappel de médicaments* (Mediteo GmbH) and *TOM rappel medicaments, pilule* (Innovation6 GmbH), available on both platforms. Each app's MARS-F score was weakly correlated with user ratings on the App Store and moderately correlated on the Google Play Store.

Conclusions: To our knowledge, this is the first study that used a validated scoring system to evaluate medication management apps that send medication reminders. The quality of the apps was heterogeneous, with only 2 having been studied in a randomized controlled trial with positive results. The evaluation of apps in real-life conditions by patients is necessary to determine their acceptability and effectiveness. Certification of apps is also essential to help health care professionals and patients identify validated apps.

(JMIR Mhealth Uhealth 2024;12:e54866) doi:10.2196/54866

KEYWORDS

medication adherence; mobile apps; telemedicine; reminder system; behavioral therapy; mHealth; mobile health; app; apps; applications; smartphone apps; medication; medications; adherence; search; searches; searching; systematic; App Store; Google Play; French; reminder; reminders; MARS; quality; Mobile App Rating Scale; mobile phone

Introduction

Therapeutic adherence is defined by the World Health Organization as "the extent to which the behaviors of a person required to take medication, follow a diet and/or change lifestyle correspond to the recommendations agreed with a healthcare professional" [1]. It is estimated to be around 50% for people with chronic diseases in high-income countries [1]. The result is an increased risk of complications, hospitalization, and mortality for these patients, as well as consequently higher health care costs [2-5]. Improving the effectiveness of adherence interventions could potentially have a far greater impact on population health than any improvement in specific medical treatments [1].

Mobile health (mHealth) is a promising strategy to optimize therapeutic adherence [6]. mHealth covers medical and public health practices based on mobile devices such as cell phones, patient monitoring systems, personal digital assistants, and other wireless devices [7]. It is dominated by the use of health and wellness smartphone apps, the number of which continues to grow yearly [8]. In 2021, over 327,000 mHealth apps were listed in digital stores [8].

In France, almost 90% of French people aged older than 12 years owned a smartphone in 2022 and 72% of them had downloaded apps [9]. For their part, French general practitioners seem to be in favor of mHealth apps or devices and ready to prescribe these in their practice, but at the same time stress the importance of validating the use of these apps through randomized clinical trials and certification by health care professionals [10].

Several studies have confirmed that the use of a medication management app improves adherence to therapy, notably by sending reminders, even in older patients with no experience of using new technologies [11-14]. However, no study evaluating the quality of medication management apps via a validated score has been published to date.

The main objective of this study was to identify medication management apps to improve patient adherence and assess their quality, using a validated scale. The secondary objective of this study was to identify high-quality medication reminder apps and to provide recommendations to any patient needing to take one or more medications, regardless of pathology.

Methods

Overview

This involved a systematic search of smartphone apps with content evaluation, carried out between October 1, 2022, and June 20, 2023. It was reported in line with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist for reporting a systematic review (items not pertinent to a systematic search of apps were considered not applicable) [15].

App Selection

Apps were searched from October 1 to November 30, 2022, on the App Store and Google Play Store. These 2 platforms were used by over 99% of mobile users to download apps in 2022 [16].

The keywords searched on the download platforms were "medication reminder" ("rappel de médicament"), "medication monitoring" ("suivi de médicament"), "treatment reminder" ("rappel de traitement"), "treatment monitoring" ("suivi de traitement"), and "pillbox" ("pilulier"). The selection of search terms was based on existing studies, which were adapted after several prior search attempts on digital download platforms [17]. We also included native health apps, automatically present on iOS and Android phones without the need to download them. Being preinstalled on smartphones, they cannot be identified by our search equations on app stores. However, they are available and easily accessible to users. Android currently has no native health app.

Inclusion criteria were as follows: (1) medication reminder app; (2) availability in the French language; (3) free (apps free for only a trial period were not selected); (4) updated within the last 2 years (since 2022, Google Play Store removes apps without updates for more than 2 years [18], and for the App Store, this period is 3 years for apps with few downloads in the last 12 months [19]); and (5) not specific to a single treatment (eg, the contraceptive pill) or to the treatment of a particular pathology. Inclusion criteria were based on the data available on the apps' presentation page, that is, the title, description, images, and general information about the app. Apps were excluded after download in the event of (1) or (2) unavailability on download platforms at the time of evaluation by one of the evaluators.

The decision to assess exclusively free apps was grounded in the primary target audience of patients encountering diverse and, at times, multiple impediments to consistent medication adherence. The high cost of an app is one of the main reasons why it is not downloaded [20]. In addition, users frequently report hidden costs as the main reason for discontinuation [20]. So, considering the price as a dissuading factor for the adoption of health apps, proposing a paid app to this population appeared inadequate [21].

Evaluation via the French Version of the Mobile App Rating Scale App

The Mobile App Rating Scale (MARS) was used to evaluate the apps. This score was previously validated for the overall quality assessment of an app in the field of mHealth [22]. The French version of this score, MARS-F, has also been validated for use with French-language apps [23].

MARS-F is a 19-item questionnaire divided into 4 objective evaluation sections (A to D): engagement (5 items: entertainment, interest, personalization, interactivity, and adaptation to target group), functionality (4 items: app performance, ease of use, navigation, and app gesture design), aesthetics (3 items: layout, graphic design, and overall visual appeal), and information (7 items: accuracy of app description, precise app objectives, quality of information, quantity of information, visual information, credibility of information, and scientific evidence; Multimedia Appendix 1). Each item is rated with a Likert scale from 1 (“inadequate”) to 5 (“excellent”). When specific objectives (item 14) were not delineated, and pertinent information (items 15, 16, and 17) or scientific evidence (item 19) was lacking, the respective items were rated as “not applicable.” Consequently, these items were excluded from consideration in the overall scoring process. In total, MARS-F is a score out of 5, corresponding to the average of these 4 sections. Section E, not included in the overall MARS-F calculation, evaluates the apps’ subjective quality via 4 items and is described separately. The MARS-F also includes a preliminary descriptive section on the apps’ characteristics, including users’ ratings on the platforms, the apps’ target objectives, the strategies used, its affiliation (commercial, governmental, and academic), as well as the technical aspect (password protection, possible sharing, and internet access required).

In total, 2 family medicine residents evaluated each app independently. Before starting, they assisted with a training video on the use of MARS-F and trained in real-life conditions by evaluating 3 apps not selected for this study. The scores for each item were then discussed one by one to ensure a similar understanding on the part of the 2 evaluators.

Apps available on the 2 platforms were evaluated independently on iOS and Android. Each app was tested and used for at least 15 minutes. Evaluations took place from February to June 2023.

Statistical Analysis

App characteristics were described using frequencies and proportions. To determine interrater reliability, a 2-way mixed-effects model intraclass correlation coefficient (ICC) was calculated for the mean of the raters. An ICC below 0.5 was considered poor, moderate between 0.5 and 0.75, good between 0.75 and 0.90, and excellent if above 0.90 [24].

The average of the scores given by the 2 raters or evaluators was used for the final rating of each app. Apps were compared according to their respective MARS-F quality score. The results were presented as mean (SD) and median and quartiles.

The correlation between the average rating of app users on download platforms and the MARS-F obtained was measured by Spearman correlation after a normality test. Regarding the correlation between MARS-F and user ratings, we excluded apps with a limited number of ratings on platforms, as they may not be highly representative of users. The choice of the threshold was determined following a sensitivity analysis, wherein the correlation was computed for various thresholds: five ratings on platforms, 10, 20, 30, 50, and so on. The selected threshold was the lowest one for which both the lower and upper thresholds yielded similar results. The threshold of 20 user ratings was finally selected. The correlation between reviewers’ subjective assessment of the apps via item 23 (“What is your overall star rating for the application?”) and the MARS-F obtained for each app was also calculated by Pearson correlation after a normality test. The correlation was judged as very strong from 1 to 0.9, strong from 0.9 to 0.7, moderate from 0.7 to 0.5, weak from 0.5 to 0.3, and very weak from 0.3 to 0.

All statistical analyses were performed with EasyMedStat (version 3.29; EasyMedStat).

Ethical Considerations

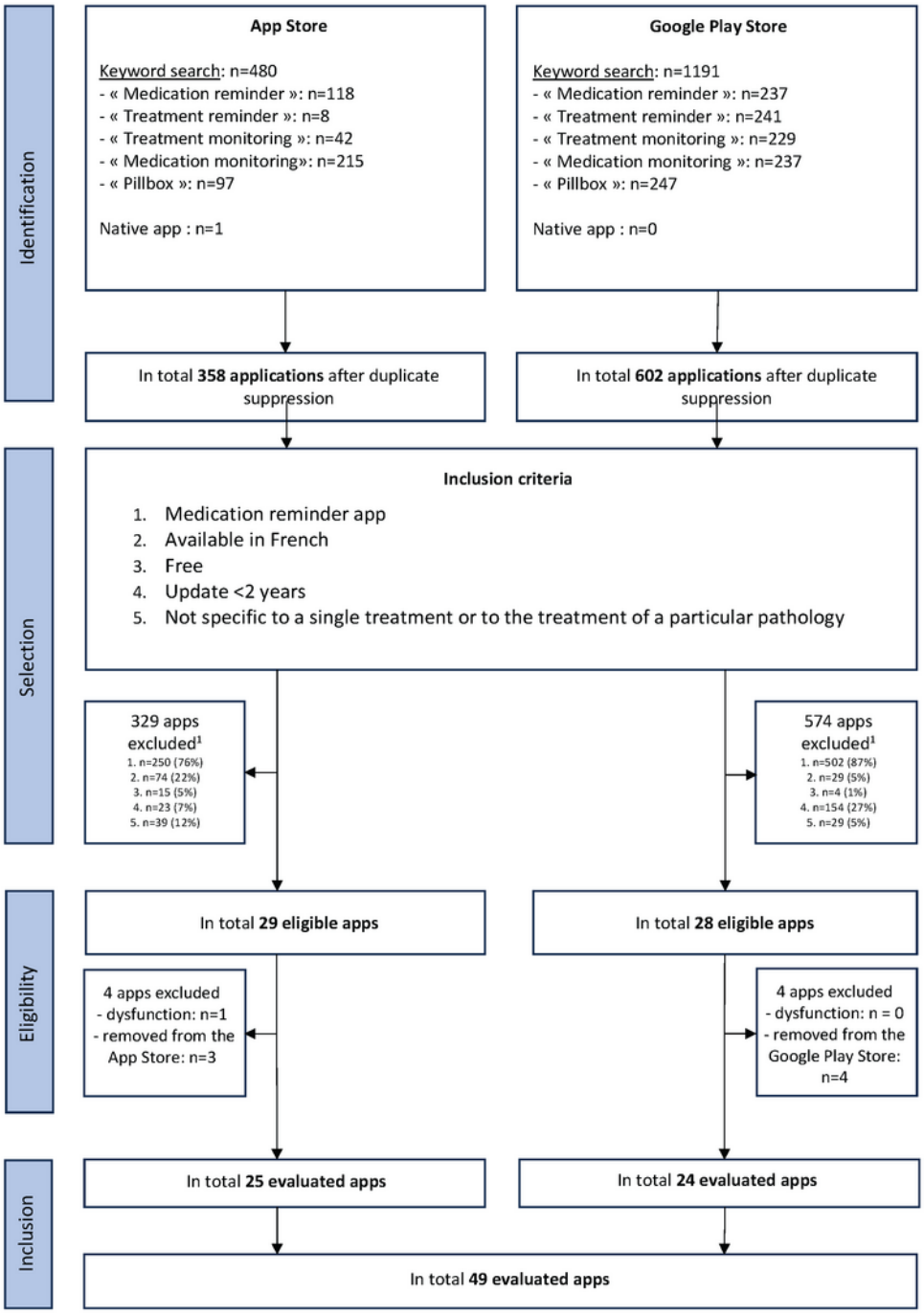
The Research Ethics Committee of the University of Montpellier approved this research project (UM 2022-006bis; Multimedia Appendix 2). This study was not funded.

Results

App Selection

After a keyword search and the addition of native apps, 480 apps were identified on the App Store and 1191 on the Google Play Store. A total of 25 apps meeting the inclusion and exclusion criteria were selected from the App Store and 24 from the Google Play Store (Figure 1).

Figure 1. Flowchart showing app selection. The numbers under "selection" refer to the inclusion criteria. The sum of the percentages exceeds 100% because a single app may not meet different inclusion criteria.



Description of the Apps

In total, 51% (n=25) of the apps evaluated were available on the App Store and 49% (n=24) on the Google Play Store, 10 of which were common and available on both platforms. Further, 40 had been rated by users on both platforms (23 on the App Store and 17 on the Google Play Store).

Although all the apps were available free of charge, some had a paid version, as in the case of 15 (60%) App Store apps and 12 (50%) Google Play Store apps. The affiliation of the developers was not always known, but the majority were commercial (24/25, 96% on the App Store; 21/24, 87.5% on the Google Play Store).

Technical aspects according to the download platform are summarized in Table 1.

Table 1. Technical aspects of apps according to platform.

	App Store (n=25), n (%)	Google Play (n=24), n (%)
Allows sharing	12 (48)	10 (42)
Has an app-linked community	3 (12)	1 (4)
Has the ability to password-protect	10 (40)	3 (12.5)
Requires a login ID	4 (16)	2 (8)
Sends reminders	25 (100)	24 (100)
Needs internet access to work	10 (40)	8 (33)

Evaluation Reliability

MARS-F interrater reliability for all apps was excellent (ICC 0.92; 95% CI 0.87-0.95; *P*<.001).

For each section, interrater reliability was excellent for engagement (ICC 0.92; 95% CI 0.88-0.95; *P*<.001) and for subjective app quality (ICC 0.95; 95% CI 0.92-0.97; *P*<.001). It was good for aesthetics (ICC 0.81; 95% CI 0.70-0.89; *P*<.001) and the information section (ICC 0.86; 95% CI 0.78-0.92; *P*<.001). Finally, interrater reliability was only moderate for the functionality part (ICC 0.70; 95% CI 0.52-0.82; *P*<.001).

Quality of Apps Based on MARS-F

The mean MARS-F score was 3.56 (SD 0.49) for iOS apps and 3.51 (SD 0.46) for Android apps (Table 2). The full set of scores by the app is presented in Multimedia Appendix 3.

Table 2. Average score, range, and median per section.

Variable	iOS (n=25)		Android (n=24)	
	Score, mean (SD)	Score, median (Q1[a]-Q3)	Score, mean (SD)	Score, median (Q1-Q3)
Section A: engagement	3.19 (0.616)	3.3 (2.8-3.6)	3.15 (0.628)	3.15 (2.8-3.52)
Section B: functionality	4.26 (0.448)	4.38 (4-4.62)	4.17 (0.431)	4.25 (3.84-4.53)
Section C: aesthetics	3.5(0.555)	3.67 (3.17-4)	3.38 (0.599)	3.5 (3-3.83)
Section D: information	3.29 (0.708)	3.3 (2.83-3.8)	3.33 (0.561)	3.17 (2.83-3.81)
MARS-F[b]: global quality	3.56 (0.485)	3.6 (3.21-3.98)	3.51 (0.455)	3.48 (3.29-3.88)
Section E: subjective quality	2.52 (1.12)	2.38 (1.5-3.62)	2.55 (1.19)	2.25 (1.72-3.22)

[a]Q: quartile.

[b]MARS-F: Mobile App Rating Scale, French version.

General App Features

The iOS apps with the highest MARS-F scores were *TOM rappel medicament, Pilule* (MARS-F score: 4.37; Innovation6 GmbH), then *Mediteo rappels de médicaments* from (MARS-F score: 4.34; Mediteo GmbH), followed by *Rappels de médicaments* from (MARS-F at 4.13; smartpatient GmbH). The Android apps with the highest MARS-F scores were *Mediteo rappels de médicaments* (MARS-F score: 4.32), followed by *TOM Rappel medicaments, pilule* (MARS-F score: 4.24) and *Rappel de pilule et medicament* (MARS-F score: 4.09; Medisafe). The MARS-F ranking of apps by platform is available in Multimedia Appendix 4.

In total, 84% (n=21) of the apps available on the App Store had a MARS-F score above 3, with 5 (20%) scoring above 4. On Google Play Store, 79% (n=19) of apps had a MARS-F score higher than 3, including 5 (21%) with a score higher than 4.

The "scientific evidence" item was completed for 4 (16%) apps on the App Store and 2 (8%) apps on the Google Play Store, 2 of which were common to both platforms.

Correlation Between MARS-F and User Ratings

In total, 23 (92%) apps were rated by users on the App Store, with an average rating of 4.38/5 (range 2.9-5.0) and an average number of ratings of 475 (range 1-3900; Multimedia Appendix 5). On Google Play Store, 17 (71%) apps were rated by users, with an average score of 4.17/5 (range 3.3-4.8) and an average number of ratings of 25,062 (range 24-223,000).

After performing sensitivity analyses, tests looking for a correlation between a given app's MARS-F and average user rating were carried out on apps with at least 20 reviews on the platforms. The correlation was weak on the App Store (ρ=0.46; *P*=.12) and moderate on the Google Play Store (ρ=0.55; *P*=.02).

Correlation Between MARS-F and Reviewers' Subjective Evaluation

The correlation between MARS-F and item 23 ("What is your overall star rating of the app?") was considered very strong for iOS apps (ρ=0.93; *P*<.001) and strong for Android apps (ρ=0.88; *P*<.001).

The mean score for this item was 2.90 and 2.98 on iOS and Android, respectively, which is below the respective mean MARS-F.

Discussion

Principal Results

The main objective of our study was to identify medication management apps to improve patient adherence. In total, 960 apps were identified: 358 (37.3%) on App Store and 602 (62.7%) on Google Play Store. This imbalance in favor of the Google Play Store has already been observed in several studies [25,26]. However, after selection, the number of apps was balanced 51% (n=25) on the App Store and 49% (n=24) on the Google Play Store, so it seems that the Google Play Store search engine offers more apps for the same keyword. This balanced proportion is consistent with a 2019 medication management app evaluation study in which 328 apps available in English were studied [17]. In this study, 53.4% and 46.6% of apps were retrieved from the Google Play Store and the App Store, respectively [17]. Our study evaluated the content of these apps according to the information available on app download stores without downloading them.

Our study highlighted a cybersecurity issue: only 40% (n=10) of apps identified on iOS had password-protected access, and 4% (n=1) required a login ID. These figures were even lower for Android apps, at 12.5% (n=3) and 8.3% (n=2), respectively. This was also the case for pain management apps, with 39% of apps evaluated allowing password protection and 44% requiring login [27].

The overall quality of the medication reminder apps evaluated is considered acceptable to moderate, with 84% (n=21) of apps on the App Store achieving a MARS-F higher than 3/5 with an average of 3.56 (SD 0.49) and 79% (n=19) on Google Play Store with an average MARS-F of 3.51 (SD 0.46). We have no point of comparison in the absence of any previous French-speaking or English-speaking study evaluating medication reminder apps by a validated score. Nevertheless, these results were expected, as they are consistent with other French studies on health apps for smoking cessation (mean MARS-F of 3.49, SD 0.57) for 14 apps [28], nutrition (mean MARS-F of 3.26, SD 0.43) for 15 apps [29], mental health (mean MARS-F of 3.16, SD 0.43) for 12 apps [30], or oral hygiene (mean MARS-F of 3.20, SD 0.38) for 9 apps [31].

All the other studies reported that the functionality section of the MARS scale had the highest ratings, which was also the case in our study, with an average of 4.26 (SD 0.45) for the App Store and 4.17 (SD 0.43) for the Google Play Store for this section. This shows that most apps are functional, which also coincides with the fact that only 1 app was excluded for malfunction in our study. In contrast, the section with the lowest average score was the engagement section (average of 3.19, SD 0.62, on the App Store and 3.15, SD 0.63, on the Google Play Store), followed by the information section (average of 3.29, SD 0.71, and 3.33, SD 0.56, on the App Store and Google Play Store, respectively), in line with the literature [32-37]. Several other studies have noted lower scores for the information section [25,27,30,31,38-41]. Nevertheless, these 2 sections are regularly cited as being those for which we find lower averages. Further, 1 exception is the study by Salehinejad et al [42], for which the highest-rated section was the information section. This is a special case, as the evaluation was concerned with COVID-19 management apps, which were probably created for information purposes in the first place.

About the information section, the lack of involvement from health care professionals, universities, or governmental organizations, may explain the observed low scores, thereby lowering the credibility of the apps. This point is assessed in the MARS-F by item 18 (“Does the application come from a legitimate source (specified in the application store description or in the application itself)?”). Often difficult to ascertain [16], most apps were affiliated with a commercial company, with only 14.6% of medication management apps and 15.2% of apps for patients with genitourinary tumors showing involvement by health care professionals [43]. Involvement by health care professionals was sometimes mentioned on the apps’ websites, but without explaining their precise role or degree of involvement. The “scientific evidence” item (item 19) was completed for 4 different apps out of the 49 evaluated in our study, 2 of which were common to both the App Store and Google Play Store. These 2 apps are the only ones to have been tested in at least one randomized controlled trial with positive results: smartpatient GmbH or MyTherapy’s *Rappels de médicaments* [44] and Medisafe’s *Rappel de pilule et medicament* [45]. They are ranked among the best apps on the 2 platforms according to their MARS-F obtained in this study. This lack of scientific validation of health care apps is a recurring theme [25,29,32,33,37].

In our study, user ratings of digital app stores were higher than MARS-F, as expected [27,29]. The absence or low correlation between user rating and MARS score has also been described previously [33,46]. To our knowledge, only the study by Chen et al [47] found a link between the quality of apps assessed by the MARS score and user rating for drug information apps. The average MARS score of apps with higher user star ratings was significantly higher than for apps with lower user star ratings (3.38, SD 0.64, vs 3.05, SD 0.64, $P<.001$) [47]. However, this may be explained by their study design: 3 out of 7 assessors were not health care professionals, which may be a bias.

The correlation between item 23 (“What is your overall star rating for the app?”) and the MARS-F was very strong for apps on iOS and strong for apps on Android. These results are consistent with other studies dealing with nutrition and mental health apps, respectively [29,30]. It is interesting to note that the subjective rating of the reviewers correlated with the overall quality of the app assessed by an objective scale, unlike the user rating. It is challenging to predict whether this is due to the evaluators’ experience or the fact that they have delved deeper into the evaluation of apps using the MARS score.

Strengths and Limitations

The main limitation of this study is the mobile app sector itself since this study is a reflection of the supply and quality of apps for a specific period. We note, for example, that 6 apps (3 on the App Store and 3 on the Google Play Store) were excluded because they were no longer available a few months after they were identified and downloaded for evaluation. This difficulty had already been highlighted in a study that aimed to evaluate apps for pregnant women at 2-year intervals, in 2017 and then

in 2019. One of the things that stood out was that the best app in 2017 was no longer available in 2019 at the time of the second evaluation [40]. The presence of new versions following updates is another element that can influence the quality of an app over time. App selection by only one of the evaluators is also a limitation, as it is possible that some eligible apps have not been identified. In terms of evaluation, the apps were assessed after they had been in use for a minimum of 15 minutes, so additional strengths or weaknesses of each app may not have been detected due to this limited duration.

Our choice to evaluate only free apps is also open to criticism. It could be argued that this choice is induced by a bias of the health care system in which the authors of this paper, who are French, operate. France has indeed the lowest share of out-of-pocket health expenditure among all EU countries [48]. Yet, medication nonadherence is a global problem. Costs attributed to "all causes" nonadherence range from US $5271 to US $52,341 [49]. The high cost of a paid app is dissuasive and hidden costs have been demonstrated as one of the main reasons for discontinuation of the use of an app [20,21]. To our knowledge, the superiority of a paid medication management app over a free one has yet to be demonstrated.

Finally, the MARS score was used because its use is simple, validated, and widespread in the evaluation of health care apps. However, it has several limitations. The first is the absence of data security and privacy evaluation criteria, although the presence of a password and login is indicated in the descriptive section. These points are nevertheless a concern for users of health care apps [50]. The second limitation is the absence of a threshold for judging app quality. We have described it here as moderate, as in several other studies, which found similar average scores, but this was not described when this tool was created [51].

To our knowledge, no other study has evaluated medication management apps using a validated scale. The identification method combined with the inclusion and exclusion criteria enabled an exhaustive analysis of free medication reminder apps available to French patients. These were not selected based on user ratings or the number of downloads from digital stores, which made it possible to evaluate apps that are not promoted on platforms but may nevertheless be relevant to patients. The weak correlation between app quality and user rating found in this study supports this approach. Apps available for iOS and Android devices were evaluated independently on each operating system, as there may be differences in terms of updating or functionality depending on the device used. Finally, the independent testing by 2 evaluators is a strong point, particularly with the observed excellent interevaluator reliability, enabling result confirmation.

Perspectives

This work is the first step toward facilitating more in-depth studies on top-rated and best-quality apps. App evaluation by patients with the user version of the MARS [52] would be relevant, even though there is currently no validated French version of this scale. The long-term use and benefits of these apps need to be studied in randomized clinical trials, to verify their acceptability and whether or not they improve therapeutic adherence and clinical outcomes in patients undergoing long-term treatment.

The results of this study, added to other studies on app evaluation in the health care field, show the possible ways to improve existing apps and give leads for the creation of new ones. Functionality is paramount, and this point already seems to have been achieved for the majority of apps currently available. The areas that need improvement relate above all to engagement and the information available, of which gamification is 1 avenue to be explored [53]. Data security and privacy protection are also important for patients and should not be neglected.

Currently, France is investing in the use of digital health, particularly for patients with chronic diseases. There is an ongoing project to list more than 50 apps offering exchanges with *Mon Espace Santé* by 2026. Therefore, it is vital to set up certification for existing apps to help doctors in their recommendations [10]. The creation of an app with the help of health care professionals and validated by the Agence du Numérique en Santé would also be a solution that would enable doctors to know which apps to recommend, thus facilitating its use by patients.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Mobile App Rating Scale, French version (MARS-F).
[PDF File (Adobe PDF File), 190 KB - mhealth_v12i1e54866_app1.pdf]

Multimedia Appendix 2
Ethics committee approval.
[PDF File (Adobe PDF File), 280 KB - mhealth_v12i1e54866_app2.pdf]

Multimedia Appendix 3
The full set of scores by app.
[DOCX File , 40 KB - mhealth_v12i1e54866_app3.docx]

Multimedia Appendix 4
Ranking of apps by platform according to the Mobile App Rating Scale, French version (MARS-F).
[DOCX File , 24 KB - mhealth_v12i1e54866_app4.docx]

Multimedia Appendix 5
User ratings and number of evaluation of apps by platform.
[DOCX File , 27 KB - mhealth_v12i1e54866_app5.docx]

Multimedia Appendix 6
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOCX File , 32 KB - mhealth_v12i1e54866_app6.docx]

References

1. Sabaté E. Adherence to Long-Term Therapies: Evidence for Action. World Health Organization. 2003. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6147925/ [accessed 2024-02-28]
2. Xu T, Yu X, Ou S, Liu X, Yuan J, Tan X, et al. Adherence to antihypertensive medications and stroke risk: a dose-response meta-analysis. J Am Heart Assoc 2017;6(7):e006371 [FREE Full text] [doi: 10.1161/JAHA.117.006371] [Medline: 28743788]
3. Egede LE, Gebregziabher M, Dismuke CE, Lynch CP, Axon RN, Zhao Y, et al. Medication nonadherence in diabetes: longitudinal effects on costs and potential cost savings from improvement. Diabetes Care 2012;35(12):2533-2539 [FREE Full text] [doi: 10.2337/dc12-0572] [Medline: 22912429]
4. Currie CJ, Peyrot M, Morgan CL, Poole CD, Jenkins-Jones S, Rubin RR, et al. The impact of treatment noncompliance on mortality in people with type 2 diabetes. Diabetes Care 2012;35(6):1279-1284 [FREE Full text] [doi: 10.2337/dc11-1277] [Medline: 22511257]
5. Hiligsmann M, Cornelissen D, Vrijens B, Abrahamsen B, Al-Daghri N, Biver E, et al. Determinants, consequences and potential solutions to poor adherence to anti-osteoporosis treatment: results of an expert group meeting organized by the European Society for Clinical and Economic Aspects of Osteoporosis, Osteoarthritis and Musculoskeletal Diseases (ESCEO) and the International Osteoporosis Foundation (IOF). Osteoporos Int 2019;30(11):2155-2165 [FREE Full text] [doi: 10.1007/s00198-019-05104-5] [Medline: 31388696]
6. Pouls BPH, Vriezekolk JE, Bekker CL, Linn AJ, van Onzenoort HAW, Vervloet M, et al. Effect of interactive eHealth interventions on improving medication adherence in adults with long-term medication: systematic review. J Med Internet Res 2021;23(1):e18901 [FREE Full text] [doi: 10.2196/18901] [Medline: 33416501]
7. Global Observatory for eHealth Series—Volume 3. MHealth: New Horizons for Health Through Mobile Technologies. World Health Organization. 2011. URL: https://iris.who.int/bitstream/handle/10665/44607/9789241564250_eng.pdf?sequence=1 [accessed 2024-02-28]
8. Évaluation des Applications dans le champ de la santé mobile (mHealth)—État des lieux et critères de qualité du contenu médical pour le référencement des services numériques dans l'espace numérique de santé et le bouquet de services des professionnels. La Plaine, Saint-Denis: Haute Autorité de Santé; 2021.
9. Baromètre du numérique. Centre de recherche pour l'étude et l'observation des conditions de vie. 2022. URL: https://www.economie.gouv.fr/files/files/directions_services/cge/barometre-numerique-2022.pdf?v=1684931616 [accessed 2024-01-18]
10. Vecchia CD, Leroy T, Bauquier C, Pannard M, Sarradon-Eck A, Darmon D, et al. Willingness of French general practitioners to prescribe mHealth apps and devices: quantitative study. JMIR mHealth uHealth 2022;10(2):e28372 [FREE Full text] [doi: 10.2196/28372] [Medline: 35147508]
11. Al-Arkee S, Mason J, Lane DA, Fabritz L, Chua W, Haque MS, et al. Mobile apps to improve medication adherence in cardiovascular disease: systematic review and meta-analysis. J Med Internet Res 2021;23(5):e24190 [FREE Full text] [doi: 10.2196/24190] [Medline: 34032583]
12. Zhou TT, Wang R, Gu SJ, Xie LL, Zhao QH, Xiao MZ, et al. Effectiveness of mobile medical apps in ensuring medication safety among patients with chronic diseases: systematic review and meta-analysis. JMIR mHealth uHealth 2022;10(11):e39819 [FREE Full text] [doi: 10.2196/39819] [Medline: 36413386]
13. Fenerty SD, West C, Davis SA, Kaplan SG, Feldman SR. The effect of reminder systems on patients' adherence to treatment. Patient Prefer Adher 2012;6:127-135 [FREE Full text] [doi: 10.2147/PPA.S26314] [Medline: 22379363]
14. Mira JJ, Navarro I, Botella F, Borrás F, Nuño-Solinís R, Orozco D, et al. A Spanish pillbox app for elderly patients taking multiple medications: randomized controlled trial. J Med Internet Res 2014;16(4):e99 [FREE Full text] [doi: 10.2196/jmir.3269] [Medline: 24705022]
15. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. PLoS Med 2021;18(3):e1003583 [FREE Full text] [doi: 10.1371/journal.pmed.1003583] [Medline: 33780438]
16. Mobile operating system market share France. StatCounter Global Stats. URL: https://gs.statcounter.com/os-market-share/mobile/france [accessed 2024-01-18]

17. Tabi K, Randhawa AS, Choi F, Mithani Z, Albers F, Schnieder M, et al. Mobile apps for medication management: review and analysis. JMIR mHealth uHealth 2019;7(9):e13608 [FREE Full text] [doi: 10.2196/13608] [Medline: 31512580]
18. Developer program policy. Google Play Console. URL: https://support.google.com/googleplay/android-developer/answer/13837496?hl=en&ref_topic=9877065&sjid=13400681840324225447-EU [accessed 2024-01-18]
19. Clarifying criteria and new timing extension for App Store improvements process. Apple Developer. URL: https://developer.apple.com/news/?id=gi6npkmf [accessed 2024-01-18]
20. Krebs P, Duncan DT. Health app use among US mobile phone owners: a national survey. JMIR mHealth uHealth 2015;3(4):e101 [FREE Full text] [doi: 10.2196/mhealth.4924] [Medline: 26537656]
21. Folkvord F, Bol N, Stazi G, Peschke L, Lupiáñez-Villanueva F. Preferences in the willingness to download an mHealth app: discrete choice experimental study in Spain, Germany, and the Netherlands. JMIR Form Res 2023;7:e48335 [FREE Full text] [doi: 10.2196/48335] [Medline: 38145470]
22. Muro-Culebras A, Escriche-Escuder A, Martin-Martin J, Roldán-Jiménez C, De-Torres I, Ruiz-Muñoz M, et al. Tools for evaluating the content, efficacy, and usability of mobile health apps according to the consensus-based standards for the selection of health measurement instruments: systematic review. JMIR mHealth uHealth 2021;9(12):e15433 [FREE Full text] [doi: 10.2196/15433] [Medline: 34855618]
23. Saliasi I, Martinon P, Darlington E, Smentek C, Tardivo D, Bourgeois D, et al. Promoting health via mHealth applications using a French version of the mobile app rating scale: adaptation and validation study. JMIR mHealth uHealth 2021;9(8):e30480 [FREE Full text] [doi: 10.2196/30480] [Medline: 34463623]
24. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016;15(2):155-163 [FREE Full text] [doi: 10.1016/j.jcm.2016.02.012] [Medline: 27330520]
25. Narrillos-Moraza Á, Gómez-Martínez-Sagrera P, Amor-García M, Escudero-Vilaplana V, Collado-Borrell R, Villanueva-Bueno C, et al. Mobile apps for hematological conditions: review and content analysis using the mobile app rating scale. JMIR mHealth uHealth 2022;10(2):e32826 [FREE Full text] [doi: 10.2196/32826] [Medline: 35171109]
26. Kanjak J, Likitdee N, Kietpeerakool C, Temtanakitpaisan A. Evaluation of mobile health applications for cervical cancer in the digital marketplace. Obstet Gynecol Sci 2022;65(3):244-255 [FREE Full text] [doi: 10.5468/ogs.22037] [Medline: 35381627]
27. Salazar A, de Sola H, Failde I, Moral-Munoz JA. Measuring the quality of mobile apps for the management of pain: systematic search and evaluation using the mobile app rating scale. JMIR mHealth uHealth 2018;6(10):e10718 [FREE Full text] [doi: 10.2196/10718] [Medline: 30361196]
28. Bustamante LA, Ménard CG, Julien S, Romo L. Behavior change techniques in popular mobile apps for smoking cessation in France: content analysis. JMIR mHealth uHealth 2021;9(5):e26082 [FREE Full text] [doi: 10.2196/26082] [Medline: 33983130]
29. Martinon P, Saliasi I, Bourgeois D, Smentek C, Dussart C, Fraticelli L, et al. Nutrition-related mobile apps in the French App Stores: assessment of functionality and quality. JMIR mHealth uHealth 2022;10(3):e35879 [FREE Full text] [doi: 10.2196/35879] [Medline: 35285817]
30. Carrouel F, du Sartz de Vigneulles B, Bourgeois D, Kabuth B, Baltenneck N, Nusbaum F, et al. Mental health mobile apps in the French App Store: assessment study of functionality and quality. JMIR mHealth uHealth 2022;10(10):e41282 [FREE Full text] [doi: 10.2196/41282] [Medline: 36223178]
31. Carrouel F, Bourgeois D, Clément C, Tardivo D, Martinon P, Guiral S, et al. Oral-hygiene-related mobile apps in the French App Stores: assessment of functionality and quality. Int J Environ Res Public Health 2022;19(12):7293 [FREE Full text] [doi: 10.3390/ijerph19127293] [Medline: 35742541]
32. Gutiérrez NB, Ramallo HR, González MF, Martín LAK. Smartphone apps for patients with hematologic malignancies: systematic review and evaluation of content. JMIR mHealth uHealth 2022;10(9):e35851 [FREE Full text] [doi: 10.2196/35851] [Medline: 36125860]
33. Paganini S, Terhorst Y, Sander LB, Catic S, Balci S, Küchler AM, et al. Quality of physical activity apps: systematic search in App Stores and content analysis. JMIR mHealth uHealth 2021;9(6):e22587 [FREE Full text] [doi: 10.2196/22587] [Medline: 34106073]
34. Sharpe JD, Kamara MT. A systematic evaluation of mobile apps to improve the uptake of and adherence to HIV pre-exposure prophylaxis. Sex Health 2018;15(6):587-594. [doi: 10.1071/SH18120] [Medline: 30347177]
35. Seo S, Cho SI, Yoon W, Lee CM. Classification of smoking cessation apps: quality review and content analysis. JMIR mHealth uHealth 2022;10(2):e17268 [FREE Full text] [doi: 10.2196/17268] [Medline: 35175213]
36. Kim BY, Sharafoddini A, Tran N, Wen EY, Lee J. Consumer mobile apps for potential drug-drug interaction check: systematic review and content analysis using the Mobile App Rating Scale (MARS). JMIR mHealth uHealth 2018;6(3):e74 [FREE Full text] [doi: 10.2196/mhealth.8613] [Medline: 29592848]
37. McAleese D, Linardakis M, Papadaki A. Quality and presence of behaviour change techniques in mobile apps for the Mediterranean diet: a content analysis of Android Google Play and Apple App store apps. Nutrients 2022;14(6):1290 [FREE Full text] [doi: 10.3390/nu14061290] [Medline: 35334947]

38. Mohammadzadeh N, Khenarinezhad S, Ghazanfarisavadkoohi E, Safari MS, Pahlevanynejad S. Evaluation of M-Health applications use in epilepsy: a systematic review. Iran J Public Health 2021;50(3):459-469 [FREE Full text] [doi: 10.18502/ijph.v50i3.5586] [Medline: 34178793]
39. Gerner M, Vuillerme N, Aubourg T, Messner EM, Terhorst Y, Hörmann V, et al. Review and analysis of German mobile apps for inflammatory bowel disease management using the mobile application rating scale: systematic search in App Stores and content analysis. JMIR mHealth uHealth 2022;10(5):e31102 [FREE Full text] [doi: 10.2196/31102] [Medline: 35503246]
40. Musgrave LM, Kizirian NV, Homer CSE, Gordon A. Mobile phone apps in Australia for improving pregnancy outcomes: systematic search on App Stores. JMIR mHealth uHealth 2020;8(11):e22340 [FREE Full text] [doi: 10.2196/22340] [Medline: 33196454]
41. Gupta K, Roy S, Altameem A, Kumar R, Saudagar AKJ, Poonia RC. Usability evaluation and classification of mHealth applications for type 2 diabetes mellitus using MARS and ID3 algorithm. Int J Environ Res Public Health 2022;19(12):6999 [FREE Full text] [doi: 10.3390/ijerph19126999] [Medline: 35742248]
42. Salehinejad S, Niakan Kalhori SR, Gohari SH, Bahaadinbeigy K, Fatehi F. A review and content analysis of national apps for COVID-19 management using Mobile Application Rating Scale (MARS). Inform Health Soc Care 2021;46(1):42-55. [doi: 10.1080/17538157.2020.1837838] [Medline: 33164594]
43. Amor-García M, Collado-Borrell R, Escudero-Vilaplana V, Melgarejo-Ortuño A, Herranz-Alonso A, Arija JAA, et al. Assessing apps for patients with genitourinary tumors using the Mobile Application Rating Scale (MARS): systematic search in App Stores and content analysis. JMIR mHealth uHealth 2020;8(7):e17609 [FREE Full text] [doi: 10.2196/17609] [Medline: 32706737]
44. Nurakysh S, Kurakbayev K, Kosherbaeva L, Tazhiyeva A, Aimakhanova A, Kulkaeva G, et al. Evaluation of the effectiveness of the mobile application on adherence of patients with arterial hypertension. Acta Inform Med 2022;30(1):18-24 [FREE Full text] [doi: 10.5455/aim.2022.30.18-24] [Medline: 35800910]
45. Morawski K, Ghazinouri R, Krumme A, Lauffenburger JC, Lu Z, Durfee E, et al. Association of a smartphone application with medication adherence and blood pressure control: the MedISAFE-BP randomized clinical trial. JAMA Intern Med 2018;178(6):802-809 [FREE Full text] [doi: 10.1001/jamainternmed.2018.0447] [Medline: 29710289]
46. Wu X, Xu L, Li PF, Tang TT, Huang C. Multipurpose mobile apps for mental health in Chinese App Stores: content analysis and quality evaluation. JMIR mHealth uHealth 2022;10(1):e34054 [FREE Full text] [doi: 10.2196/34054] [Medline: 34982717]
47. Chen YC, Liao WW, Su MC, Lin YH. Personalized and self-management: systematic search and evaluation quality factors and user preference of drug reference apps in Taiwan. J Pers Med 2021;11(8):790 [FREE Full text] [doi: 10.3390/jpm11080790] [Medline: 34442435]
48. State of Health in the EU. France: Country Health Profile 2023. OECD, Paris/European Observatory on Health Systems and Policies. 2023. URL: https://eurohealthobservatory.who.int/docs/librariesprovider3/country-health-profiles/chp2023pdf/chp-france2023.pdf?sfvrsn=a15a53aa_3&download=true [accessed 2024-01-18]
49. Cutler RL, Fernandez-Llimos F, Frommer M, Benrimoj C, Garcia-Cardenas V. Economic impact of medication non-adherence by disease groups: a systematic review. BMJ Open 2018;8(1):e016982 [FREE Full text] [doi: 10.1136/bmjopen-2017-016982] [Medline: 29358417]
50. Zhou L, Bao J, Watzlaf V, Parmanto B. Barriers to and facilitators of the use of mobile health apps from a security perspective: mixed-methods study. JMIR mHealth uHealth 2019;7(4):e11223 [FREE Full text] [doi: 10.2196/11223] [Medline: 30990458]
51. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR mHealth uHealth 2015;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
52. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and validation of the User Version of the Mobile Application Rating Scale (uMARS). JMIR mHealth uHealth 2016;4(2):e72 [FREE Full text] [doi: 10.2196/mhealth.5849] [Medline: 27287964]
53. Abraham O, LeMay S, Bittner S, Thakur T, Stafford H, Brown R. Investigating serious games that incorporate medication use for patients: systematic literature review. JMIR Serious Games 2020;8(2):e16096 [FREE Full text] [doi: 10.2196/16096] [Medline: 32347811]

Abbreviations

ICC: intraclass correlation coefficient
MARS: Mobile App Rating Scale
MARS-F: Mobile App Rating Scale, French version
mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by L Buis; submitted 09.12.23; peer-reviewed by L Bustamante, T Orlemann; comments to author 10.01.24; revised version received 28.01.24; accepted 06.02.24; published 18.03.24.

Please cite as:
Toïgo M, Marc J, Hayot M, Moulis L, Carbonnel F
Quality Assessment of Smartphone Medication Management Apps in France: Systematic Search
JMIR Mhealth Uhealth 2024;12:e54866
URL: https://mhealth.jmir.org/2024/1/e54866
doi:10.2196/54866
PMID:38498042

Original Paper

Understanding Patient Perspectives on the Use of Gamification and Incentives in mHealth Apps to Improve Medication Adherence: Qualitative Study

Steven Tran[1], BPharm; Lorraine Smith[1], BA, PhD; Stephen Carter[1], BPharm, MSc, PhD

School of Pharmacy, Faculty of Medicine and Health, University of Sydney, Camperdown, Australia

Corresponding Author:
Steven Tran, BPharm
School of Pharmacy
Faculty of Medicine and Health
University of Sydney
A15, Science Rd
Camperdown, 2050
Australia
Phone: 61 93512222
Email: steventran@hotmail.com.au

Abstract

Background: Medication nonadherence remains a significant health and economic burden in many high-income countries. Emerging smartphone interventions have started to use features such as gamification and financial incentives with varying degrees of effectiveness on medication adherence and health outcomes. A more consistent approach to applying these features, informed by patient perspectives, may result in more predictable and beneficial results from this type of intervention.

Objective: This qualitative study aims to identify patient perspectives on the use of gamification and financial incentives in mobile health (mHealth) apps for medication adherence in Australian patients taking medication for chronic conditions.

Methods: A total of 19 participants were included in iterative semistructured web-based focus groups conducted between May and December 2022. The facilitator used exploratory prompts relating to mHealth apps, gamification, and financial incentives, along with concepts raised from previous focus groups. Transcriptions were independently coded to develop a set of themes.

Results: Three themes were identified: purpose-driven design, trust-based standards, and personal choice. All participants acknowledged gamification and financial incentives as potentially effective features in mHealth apps for medication adherence. However, they also indicated that the effectiveness heavily depended on implementation and execution. Major concerns relating to gamification and financial incentives were perceived trivialization and potential for medication abuse, respectively.

Conclusions: The study's findings provide a foundation for developers seeking to apply these novel features in an app intervention for a general cohort of patients. However, the study highlights the need for standards for mHealth apps for medication adherence, with particular attention to the use of gamification and financial incentives. Future research with patients and stakeholders across the mHealth app ecosystem should be explored to formalize and validate a set of standards or framework.

(JMIR Mhealth Uhealth 2024;12:e50851) doi:10.2196/50851

KEYWORDS

qualitative; patient; perspectives; gamification; incentives; mobile app; mobile health; mHealth; medication adherence; mobile phone

Introduction

Background

Medication adherence is defined by the World Health Organization as the extent to which a person's behavior corresponds to their agreed health recommendations from a health care provider [1]. In high-income countries, such as Australia and the United States, the estimated average adherence rate has remained at approximately 50% over the last two decades [2]. Adherence rates have also been measured to be much lower in low-income countries [1]. The direct consequences of medication nonadherence are suboptimal clinical benefits leading to disease progression and

complications, which in turn has an impact on patients' quality of life. The impact of medication nonadherence on health is even more apparent in older or low-income people, with a substantial association between higher all-cause hospitalization and mortality [3,4].

Medication nonadherence also contributes to medication wastage and a substantial economic burden arising from the medical resources consumed to treat preventable health events, productivity loss, and loss of life. In Australia, medication nonadherence was estimated to have an economic burden of approximately Aus $10 billion (US $6.5 billion) annually in 2018 [5]. This economic burden is expected to be much higher since then, exacerbated by additional barriers to medication adherence, such as travel restrictions and medication shortages, from the recent COVID-19 pandemic [6] and financial stressors of a potential economic recession [7,8]. In light of the significant economic burden and health impact on patient lives, there is a pressing need to address medication nonadherence through health care–provided interventions [9].

More recently, mobile health (mHealth) apps have been used to support medication management and promote medication adherence [10]. Some mHealth apps use gamification to enhance user engagement and some also provide direct-to-patient financial incentives, which are funded by the government or third-party interest groups such as health insurers [11]. Gamification (ie, the use of game elements in nongame contexts) and financial incentives (ie, the provision of an item with real-world economic value in exchange for a completed task) have been applied across many health and nonhealth domains. Two notable examples of gamification and financial incentives are Duolingo (Duolingo, Inc) and the "Incentive to Quit" trial, respectively.

Duolingo is an educational technology company that has an app under the same name offering courses in multiple languages, music, and math. Currently, the app has over 21 million daily active users and has retained a high level of engagement over the last decade of operations attributed to its use of gamification such as point-based systems, social leaderboards, and interactive storytelling [12]. A meta-analysis on the gamification of learning [13] suggests that in addition to gamified storytelling, social interactions including competition and collaboration were particularly effective in facilitating behavioral and motivational learning outcomes. While it cannot be assumed that the competitive and collaborative aspects of gamification would foster behavioral or motivational outcomes for medication adherence, the meta-analysis [13] also notes that the self-determination theory by Ryan and Deci [14] can be used to explain the mechanistic effects of gamification in the context of learning. The self-determination theory has also been applied to health behavior change including medication adherence [15]. The theoretical framework outlines 3 psychological needs (ie, competence, autonomy, and relatedness) required for intrinsic motivation. This intrinsic motivation can subsequently result in volitional behavior such as improved medication adherence.

A Cochrane review [16] into incentives for smoking cessation found that not only the use of financial incentives, either monetary or vouchers, were accepted in multiple mixed-population settings but also there was sufficient evidence to indicate that incentives improved long-term smoking cessation rates. This outcome was also sustained after the withdrawal of the incentives. The result from the Cochrane review may have contributed to the inception of the "Incentive to Quit" trial, a government-funded program in Australia using supermarket vouchers for reaching smoking cessation milestones. The program will cost Aus $500,000 (US $324,255) with the aim to recoup the amount by reducing smoking-related public health costs that are estimated to be Aus $1.5 billion (US $972 million) nationally each year [17]. This signals a potential sustainable solution to funding concerns for financial incentive programs in countries providing universal health care.

In the domain of medication adherence, our scoping review [18] identified limited evidence for the benefit of gamification with incentives. We also highlighted a wide variation in app content, design, and development processes; the use of behavioral theories or frameworks; evaluation methods; and outcomes. The review also found that when there was no patient involvement, it was likely that design and implementation decisions were largely made by developers or researchers. This lack of patient involvement could lead to the misalignment of patients' goals and bias in the selection and use of gamified and incentivized app features.

Several qualitative studies [19-21] have explored patient perspectives on the use of technologies such as mHealth apps for medication adherence. The studies include a range of findings including observed benefits of a single location to manage their whole regimen, the value of personalization and utility, and an ambivalent feeling toward apps in health. A limitation of these study findings is that they were specifically designed to investigate a particular chronic condition and may not apply to an app for a generalized cohort. Gamification and incentive-containing apps were also not the focus of those studies, limiting the ability to draw conclusions about how patients feel about apps using these features [19-21].

As there is insufficient knowledge in the existing literature pertaining to patient perspectives on gamification or financial incentives in a generalized cohort for medication adherence, a study was conducted to address this gap and disseminate the findings through publication.

Objectives

This qualitative study aims to identify and understand patient perspectives of Australian patients taking medication for chronic conditions on the use of gamification and financial incentives in mHealth apps for medication adherence. The perspectives gained from this study will help identify barriers, identify potential opportunities, and provide a foundation for developers seeking to apply these novel features in an app intervention for a general cohort of patients.

Methods

The COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist was used to guide reporting [22]. A completed COREQ checklist applied to this study is available in Multimedia Appendix 1 [22].

Ethical Considerations

Research involving humans in Australia is reviewed by an independent group of people called a Human Research Ethics Committee. The ethical aspects of this study have been approved by the Human Research Ethics Committee of the University of Sydney (project number 2022/061).

Recruitment

Participants were passively recruited through study posters and flyers displayed by consenting pharmacies and medical practices and email newsletters distributed by patient advocacy groups (eg, organizations comprised of mostly patients or caregivers to represent and promote the needs and priorities of patients) to their members advertising the study. Participants who were interested self-enrolled via a QR code displayed on the recruitment materials, which directed them to a screening questionnaire. The screening questionnaire was based on the following inclusion criteria: aged ≥18 years, competent in English, taking ≥1 medication for an ongoing medical condition for ≥3 months, not being given end-of-life care or in palliative care, and not being a health care professional. Eligible participants were then contacted via phone by the first author (ST). The first author introduced himself by explaining his own professional background and personal motivations for conducting the study including that the research would contribute to a higher degree. The first author explained the purpose of the study and potential impact before confirming enrollment and availabilities for focus group sessions. Participants were assigned to groups based on their availability when at least 3 participants were available to attend a common session. An email was subsequently sent to the participant containing the link to the web-based focus group (Zoom; Zoom Video Communications Inc) and a link to a web-based survey (REDCap [Research Electronic Data Capture]; Vanderbilt University) to capture consent to record and the baseline characteristic questions.

Focus Groups

In line with the COREQ checklist, the first author (ST), having worked in community pharmacies for 6 years, facilitated all the focus groups after undertaking formal training and orientation by the research team. Apart from the first author and participants, there were no other parties in any of the focus groups. All focus groups commenced with an introduction of the study topic and rules and guidelines for the focus group before the recordings were initiated. An ice breaker “1-fun fact” question was asked for all participants starting with the facilitator followed by a list of open-ended questions covering topics across medication adherence, gamification, and financial incentives. All focus groups were limited to a duration of 1 hour. Participants who attended the web-based focus group was compensated with an Aus $50 (US $32) e-gift card.

The semistructured focus group guide was created and reviewed by all authors to reduce assumptions and potential bias of the first author. The semistructured focus group guide is available in Multimedia Appendix 2. Using the constant comparative method, concepts raised in a focus group were used as additional prompts in the subsequent focus groups if it was not identified before moving on to the next topic. The additional prompts were neutrally phrased and open ended to limit bias. An example of an additional prompt is as follows: “In a previous focus group data management and privacy was mentioned, what are your thoughts on data management and privacy in an app like this?” The facilitator also actively aimed to explore positive and negative perspectives equally.

Data Collection and Analysis

The web-based focus group sessions were audio and video recorded using Zoom and stored on a secured university-licensed cloud service (OneDrive; Microsoft Corp). Each recording was auto-transcribed (Adobe Premiere Pro; Adobe, Inc) before undergoing manual transcription by the first author for familiarization. Notes made by the first author during the focus groups were also annotated in the transcripts. A senior author (SC) reviewed the audio record and transcripts of the first 2 focus groups. Before conducting the remainder of the focus groups, the team discussed the conduct of the focus groups, discussed the preliminary findings, and provided advice to the first author regarding the use of prompts and pauses. The transcripts from 2 focus groups were independently reviewed and iteratively coded using NVivo (Release 1.7.1; Lumivero) into concepts by 2 of the authors (ST and SC). The research team then compared and discussed the concepts to generate a list of themes and subthemes. The subthemes and themes were evaluated and revised 3 times before being applied to the transcripts again for validation. Having decided on an agreed coding framework, this framework was applied to the remaining transcripts. Additional subthemes were added as required; however, the recruitment of the focus groups was discontinued when the last transcript did not generate any unique concepts indicating that the study was approaching data saturation. The last focus group, after recruitment ended, further indicated this as it also did not generate any unique concepts. The coding mapping tree is illustrated in Multimedia Appendix 3. Participants were informed of the preliminary findings, themes, and subthemes as a study summary and were invited to review the transcripts for commentary and correction. One participant concurred with our findings, and no other participant provided any feedback.

Results

Overview

Of the 20 participants who showed interest in participating in the study, 1 participant was excluded due to the contact being unreachable or unresponsiveness. A total of 19 participants were included in the 5 web-based focus groups (via Zoom) conducted between May 2022 and December 2022. The mean age of the participants was 40 (SD 17; range 19-71) years. All participants reported that they used their smartphone daily, while more than half reported playing games (13/19, 69%) and using loyalty rewards (12/19, 63%) on a weekly or daily basis. More than a third of the participants (7/19, 37%) were taking ≥3 medications. Further details on participant characteristics (Table 1) where collected and tabulated, such as self-reported clinical characteristics.

In total, 3 main themes were identified, along with 8 subthemes. Many of the concepts derived in the subthemes were interconnected and overlapped across the 3 main themes. Additional quotes from the participants corresponding to the subthemes are available in Multimedia Appendix 4.

Table 1. Demographic characteristics of the participants (N=19).

Characteristics	Values
Age (years), mean (SD; range)	40 (17; 19-71)
Sex (female), n (%)	12 (63)
Frequency of phone use, n (%)	
Rare	0 (0)
Weekly	0 (0)
Daily	19 (100)
Frequency of games played, n (%)	
Never	0 (0)
Rarely	6 (32)
Weekly	7 (37)
Daily	6 (32)
Frequency of loyalty rewards use, n (%)	
Never	0 (0)
Rarely	7 (37)
Weekly	9 (47)
Daily	2 (16)
Self-reported clinical characteristics, n (%)	
Arthritis and joint pain	4 (21)
Cancer	1 (5)
Cardiovascular disease	1 (5)
Diabetes	3 (16)
Kidney disease	2 (11)
Mental and behavioral conditions	7 (37)
Respiratory conditions	2 (11)
Prefer not to disclose	2 (11)
Other	7 (37)
Number of medications, n (%)	
1	8 (42)
2	4 (21)
3	3 (16)
4	1 (5)
≥5	3 (16)
Prefer not to disclose	0 (0)

Theme 1: Novelty of Gamification and Incentives Require a Purpose-Driven Design

Overview

Game features designed to improve medication and health knowledge, the process of goal setting, and a sense of empowerment were recognized as potentially beneficial. The use of incentives arising from game features to drive medication adherence was a novel concept that was acceptable to some but not universally embraced. Given the novelty, participants expected apps with these features to have excellent functionality and reliability. They noted that potentially complex language, inaccessible terminology, and complicated medication use processes require that the digital usability of apps be optimized.

Gamification for Knowledge, Empowerment, and Goal Setting

Participants' views about the use of gamification to drive medication adherence in apps were formulated from their knowledge and experience of using health apps in general. Participants' expectations about apps for medication management centered on having reminders and features to help with the organization of supply. Some also had prior experience with medication apps that included resources to help them improve their medication knowledge. Those with no prior experience rationalized that this was a benefit for some but not necessarily for themselves. Participants believed that gamification could allow users to "test your knowledge before and after" use. One participant predicted that game features designed to educate about drug interactions could empower consumers to check, understand, and respond to drug interactions:

> *...to check the medications and the side effects and everything and whether if there's any food or medicine interaction.* [Participant 12]

However, another participant doubted their capacity to benefit from this:

> *My chemist is frequently telling me don't take this tablet with that tablet because they interact...If I had a more educational interactive session, I'm not sure the message might get through to an old head like this.* [Participant 11]

Participants expected game elements to ideally elicit and respond to the individuals' specific health-related goals and personalized needs. In the following example, a participant shares how gamification had helped them to engage with yoga:

> *I think that gamification really was a selling point for me because to be honest. I...I wanted to be in the top batch (of users) and to do that I can't be missing my exercise.* [Participant 12]

Failing to support or empower the user in response to individualized and time-sensitive goals would result in a lack of motivation to engage with gamification as expressed by Participant 2 in one focus group, which was endorsed by Participant 1 and 4:

> *I think when the goal goes away, I don't feel the need to use the app anymore, I just stop using it completely.* [Participant 2]

The use of gamification to promote adherence seemed foreign to some; the following participant explained their concerns:

> *I have to say that it wouldn't put me off, but I could see that some people would see it [gamification] as a trivialization of the process. And you know, on that, I don't think it would encourage me either.* [Participant 9]

Incentives or Rewards for Driving Medication Adherence Is Novel

Most participants had experience with accruing and redeeming incentives in apps and understood that incentives create motivation to drive goals:

> *We are incentive driven as a race or as a people.* [Participant 14]

Participants expected incentivization to be facilitated by the accrual of points to culminate in a "tangible kind of incentive," such as non–cash-purchasing power or other reward. There were mixed views about the notion of receiving monetary rewards for adherent behavior. Some embraced the notion, as can be seen in the following statement:

> *Something like that, that's not cash or monetary appeals to me because it's something I do to get a reward.* [Participant 14]

However, some explained that incentives were simply not needed:

> *And the only incentive for me is, it's my health, it's the motivating factor.* [Participant 13]

There was some innate hesitancy toward the concept of people receiving personal benefit for good medication adherence behavior:

> *I have a suspicion that I'm probably more adverse to having anything monetary related on apps, I think I suspect that I probably have a stronger aversion. So if you're asking me if there was one, I probably would not use it. However, I don't think the average population would have an issue if you combined financial incentive with nonfinancial incentive [features] or you could switch the feature on and off.* [Participant 17]

Some felt that simply rewarding the quantity of drugs dispensed could lead to perverse incentives:

> *...abuse of medications by people taking medications, you know, two or three times today when they're only prescribed to take it once a day to reach a financial incentive.* [Participant 14]

Furthermore, participants expressed the notion that receiving rewards for medication adherence could be problematic for people with gambling problems:

> *I guess one of my worries is that it can become a bit like gambling at some point, especially for people that are addiction prone, that's the main thing that's a worry for me when it comes to this sort of concept for medication adherence.* [Participant 6]

Regarding how the points were to be redeemed, some participants expressed that the accrual of points could be turned into financial rewards, redeemable in the pharmacy:

> *Maybe you get to 10% discount on your next to medication script.* [Participant 12]

Other participants wanted accrued points to turn into a notification that provides rewards at other vendors:

> *I'd much prefer to see a nonfinancial incentive something like your app flushes up a barcode for free coffee or a donut or whatever it is, you know, rather than a dollar value.* [Participant 14]

Some others wanted points to be converted to financial benefits for charitable organizations:

> *I would rather that I contributed to something that was part of a bigger pool of money to help people in need.* [Participant 13]

Functionality and Reliability

The participants also highlighted the importance of having a reliable and functional mHealth app to support medication adherence. This meant the need for an app to provide accurate information about their medication schedule and dosage, accurately keep track of their repeats, and keep reminding them when it was time to take their medication or when to fill their repeats. Apps that were slow and clunky also impacted their perception of its effectiveness to keep track of their medication. Overall, the participants were clear that a reliable and functional app was critical to their ability to successfully manage their medication adherence:

> *My scripts are different quantities, some of them fall due in like three weeks and some of them fall due in four weeks...the app wasn't able to manage that [different frequencies of script repeat reminders], and it was easier for me to have the scripts (and manage it) myself.* [Participant 14]

The importance of reliability and consistency seemed particularly poignant when applied to the collation of rewards and incentives:

> *If there was like a problem with the app but their points whatever weren't going through or they also had lots of the rewards got sold out and they couldn't get what they wanted. I think those sorts of things could make people pretty unhappy and then they might just not use the app at all.* [Participant 18]

Digital Usability

It was considered imperative for an app to use simple, lay language to help break down complex medical topics or terminology. In addition, the app needs a simple user interface and intuitive features. Participants who identified as technology avoiders, due to a desire to be a digital minimalist or having concerns about their ability to use an app effectively, suggested that they would consider an app if the content was easy to understand and if it was not difficult to use. This can be further supported externally by a health care professional or care provider to guide the user or provide a tutorial or introduction to the app:

> *I've never used an app. I'm not very tech savvy. I'm probably tech phobic. I'm 60 and I've only had a mobile phone for the last two years... I would probably just need to someone to show me the benefits. How to make it [medication management] easy for me, how to use it [the app]. Probably they're the main things for me.* [Participant 13]

Participants noted that digital usability can be enhanced using gamified graphical or visual representations to summarize complex topics and medication adherence numbers into easy-to-understand tallies and metrics:

> *...a good way to visualize a lot of like statistics and stuff rather than having it in words. And it's all just this simple and easy to read, you know, you don't have to have like a stats background.* [Participant 19]

Theme 2: Trust-Based Standards

Participants expressed that their attitudes toward using gamification or financial incentives for medication adherence was highly influenced by their thoughts about whether the app was created and curated with trust-based standards. These standards related to the perceived credibility of the app and its ecosystem and the policies and governance relating to the user's data.

Credibility of the App Ecosystem

An app ecosystem refers to the intricate network of connections among the app, devices, databases, and various stakeholders such as end users, developers, and app owners. Participants expressed that the credibility of an app's ecosystem was determined by multiple factors such as brand image, mission statement, and history.

For example, some participants seemed reluctant or skeptical about the transfer of rewards being managed between commercial entities:

> *...the chemist to interact with the coffee vendor. And I can't see that...business managing to cooperate appropriately.* [Participant 11]

Some expressed concerns that a transfer to charitable organizations is potentially problematic:

> *Really concerns me that the money wouldn't get to where it's supposed to go, that you're adding more and more middlemen to it and everybody takes their cut.* [Participant 14]

Credibility was noticeably different when financial incentives were provided by a for-profit versus a not-for-profit organization:

> *If it had some sort of backing from a site like GP's or was from the government or from medical institutions or from medical groups, I would be more comfortable with it. And if it was just a private initiative, I wouldn't feel as comfortable around it, to be honest.* [Participant 17]

Generally, financial incentives provided by not-for-profit organizations were found to be more credible with most participants likely to engage with an app if it was managed by a not-for-profit organization. In contrast, participants were conflicted by for-profit organizations offering financial incentives as the app health benefits, enhanced by rewards, can be compromised and exploited. This was attributed to the lack of transparency over the economic sustainability of financial incentives. Credibility was also valuable in apps that did not offer financial incentives as it informed participants on the

trustworthiness of educational medical content provided via gamified features.

Governance Over One's Data

The notion of having to share one's health data to engage with gamification and incentives was seen as highly sensitive. Most participants reported being somewhat comfortable with sharing some personal health data on the mHealth apps they had used. Moreover, when the practicalities of sharing private data to redeem rewards were discussed, participants expressed some skepticism:

> *When you try to fill in the forms [for financial incentives] and they ask you a lot of questions like your demographic, your age group and all. Even if they ask for my email and my name, that should be fine. But sometimes they've asked too many questions, and that won't appeal to me anymore because of data privacy concerns.* [Participant 12]

> *I would be averse to the model you proposed there. I don't like the concept. It's morally repugnant and that it comes to the measure of how much information do they wish me to provide at the point of redemption.* [Participant 11]

It was also noted that access to social game elements, including financial incentives or rewards, should not be dependent on the user's decision to provide specific personal data. Having the ability for the app user to selectively choose what data are shared, including how far they are shared and for what purpose, was also suggested as a best practice by some participants. Participants wanted to be informed of what data would be shared before "consent to share" was requested.

Theme 3: Personal Choice

The last theme was related to participants' views about personal choice, that is, having the option to decide whether they should use an app and then how and when they would like to use the app with games and incentives.

Choice to Use the App

Many participants felt there was the potential for mHealth apps with gamification to effectively support medication adherence. Some would choose to engage with incentives. Participants also expected that medication management apps with gamification and incentives would come at no financial cost to them. Participants wanted the researchers to know that consumers did not want to feel forced or coerced into using these apps, especially by their treating health care provider or insurance provider:

> *I think it should be an option for everyone to use an app. So, whether they like it, I don't think it should be forced upon anyone.* [Participant 19]

Ability to Customize and Choose What Features to Use in the App

There was significant variability among the participants on the extent to which they were required to engage with gamification and the accrual of incentives. It was evident that participants wanted the flexibility to decide how and when they wanted to use the app.

The participants noted that they wanted the option to toggle on and off features that they want to use or hide, respectively. An example of a specific feature customization is being able to change the esthetics and cosmetic features of the app to the user's personal preference:

> *I mean you don't want bright colors and childish kind of images for like a seventy-year-old. I feel like they want something more mellow and relaxed and I don't know, I guess also like a customization of the actual theme. You know how sometimes like you can customize how you want something to look and make you feel that if you put it in the hands of the consumer, it can customize it. They'll be more like your feel, more personal and they'll enjoy using the app more.* [Participant 19]

While a high level of customizability was desired, it was acknowledged that this could increase complexity and potentially reduce digital usability. A balanced approach to customizability was suggested where the user would be provided with default "recommended" or standard features based on some characteristics of the user. For example, it was mentioned that it could be possible to provide a relatively younger person with a high degree of flexibility on first use. Whereas for an older person, the initial level of customization could be limited, while allowing options for further customization once the user was familiar with the basic functionality:

> *What puts me off is probably if it takes a long time to set up any customization or, you know, Avatar, that sort of stuff. And it took a long time for me, again, being very time poor, I think that would put me off. So maybe something very easy to navigate and stuff like that.* [Participant 7]

Discussion

Principal Findings

The themes synthesized in this study provide a preliminary understanding of patient perspectives on the use of gamification and financial incentives in mHealth apps to promote medication adherence. To summarize, participants expressed that apps that allow users access to game features, which are designed to promote knowledge about medicines, and allow the self-monitoring of medication taking were broadly accepted and even desirable. However, the notion that a user's behavior would be monitored by a third party and that users may accrue financial or nonfinancial benefits as rewards for adherent behavior created some skepticism. An app would need to be underpinned by good governance that is built on trust in the sponsor for it to be adopted by the public. Participants believed that the adoption of apps with game features and incentives will require that users maintain a high level of personal choice in the selection of app features, desired engagement levels, and ways to redeem rewards.

In the focus groups, participants predicted that medication apps could be helpful for individuals taking multiple medications,

for example, to help them overcome problems with forgetfulness or disarray, using reminders and scheduling. Participants in the focus groups tended to support the option of patients self-recording their medication consumption using a medication app. Previous research [15] suggests behavior change can be achieved with interventions based on the Self-Regulation Theory [23], for example, by optimizing attributes such as competence and autonomy. This could be explored through the self-monitoring of medication consumption using an app. With regard to gamification, participants saw potential to improve medication knowledge through engagement with gamified learning modules, which in turn could help patients to remain adherent to medicines for chronic diseases.

When the concept of potentially rewarding good medication adherence with financial or nonfinancial benefits was considered, participants discussed that to trigger and realize the incentives, their behavior would be observed by a third party that may or may not be their help providers. In research settings, the monitoring of adherence to provide incentives is achieved by directly observing consumption (ie, watching actual consumption in a community pharmacy), conducting pill counts of returned containers, and using electronic monitoring devices such as MEMS Caps. While those methods are accepted as accurate methods of measuring adherence [24], they are resource intensive, relatively intrusive, and unsuited for widespread adoption for the management of most chronic diseases. Therefore, while medications apps could be designed to monitor and incentivize adherence by communicating with an app, to the participants in this study, the notion of having their medication taking being monitored and receiving incentives for interacting with such an app was novel and generated several controversial discussions about appropriateness and ethicality of providing rewards for a health behavior. Much of the skepticism expressed was related to concerns about trust in third-party app providers.

The theme of trust-based standards aligns with existing and burgeoning concerns about data privacy and security across all mHealth apps; specifically with regard to inconsistencies in the way data privacy is applied [25]. In addition, Schroeder et al [26] found that the major concerns patients had on data privacy within mHealth apps were related to the potential risk of misuse of their personal health data and the fear of receiving personalized advertisements. In this study, participants actively vocalized a fear regarding which actors would have access to their health data and how private companies could monetize and gain from access to personal data. Commercial advertising and exploitation in an mHealth app setting raises substantial concerns on an ethical level, and revisions to consumer law are urgently needed to protect the user [27]. It is essential that apps adhere to established and emerging data privacy standards to establish a level of trust. While most health care apps do comply with the existing compliance standards [25,28], such as the General Data Protection Regulation and the Health Insurance Portability and Accountability Act, this is generally unbeknown to the public and considered inadequate by individuals when personally exposed to or made aware of events relating to security data breaches. This was particularly emphasized in our findings due to the recency of several major data breaches in Australia from a telecom company, Optus, and a health insurance company, Medibank, during the study period [29]. While this concern may have been inflated by recency bias, studies have reported increases in data breaches globally, indicating that these concerns may be warranted [30,31].

Further improvements in transparency and stronger preventive and remedial processes relating to data breach events or unauthorized access of data may alleviate this concern for the individual. Being data-privacy compliant and transparent adds to the security for an individual to carry out health activities within the app by mitigating the risk of damage to their identity and dignity [32]. When data privacy and security is applied and communicated effectively, the collection, analysis, use, and evaluation of personal health information have potential to generate new preventive and curative therapies, diagnostics, and the optimized delivery of health care. The well-managed use of app data when well managed can be fed back to the individuals as a form of digitally assisted precision medicine [33]. The concept of securely sharing data for the personal benefit of one's medical treatment and no other uninformed purpose has been identified by our participants as an acceptable approach if it is provided as an opt-in option instead of a mandate of the app.

The likely success of mHealth apps with gamification and incentives largely depends on their ability to engage with their intended users and align with their needs and preferences. A SWOT (strengths, weaknesses, opportunities, and threats) analysis in the study by Hein et al [34] reported that while digital solutions add complexity to health regimens, they also have the potential to provide high-quality medication management, enable additional services, and drive innovation in health care. The added complexity of digital solutions can be contributed by potential issues with regard to functionality and reliability or poor digital usability. Ensuring the functionality, reliability, and digital usability of apps is paramount to all mHealth and IT solutions and is not exclusive to integrating gamification and financial incentives into mHealth apps for medication adherence. The findings of this study provide some anecdotal evidence of poor experiences with apps used in health and other areas. While general IT standards such as SQuaRE (System and Software Quality Requirements and Evaluation) exist, there are no gold standards to qualitatively evaluate software apps with regard to digital usability specifically for health [35-37]. The development of a framework to include and contextualize existing IT standards for health may provide further guidance when developing such interventions. This may include recommendations such as more rigorous testing, user feedback loops, and mechanisms to allow developers to identify and address any technical issues or usability challenges promptly. Furthermore, involving health care professionals and patients in the app's design process can provide valuable insights into specific needs and preferences, ultimately enhancing engagement and adherence [38].

This study shows that potential users will want to maintain a high level of personal choice over how they opt in, configure, use, share their personal data, and opt out of mHealth apps with gamification and incentives. This finding highlights the critical role of ensuring that patients have a sense of autonomy over

their health behavior [39]. As such, mHealth apps should be designed to empower patients to become active partners in their health care while ethically informing health care providers of patient-specific data. People who use mHealth apps have been shown to have increased satisfaction with their overall care and have reported improved interactions with their health care provider [40]. In addition, health care providers feel that the use of an mHealth app ensures better clinical decision-making and patient outcomes [41-43]. During clinical practice, the benefits of using an mHealth app should be outlined and promoted alongside their medical treatment. A holistic approach removes coercion and respects the patient's autonomy over the app and their medical treatment. Sax et al [27] similarly argue that patients should have autonomy over their health care decisions and that more attention should be provided to mHealth apps due to the increasingly persuasive methods the apps use to influence the behavior of users potentially for economic gains. To analyze the ethical impact of an app on autonomy, Sax et al [27] further suggested a framework that considers 3 requirements: independence, authenticity, and options. Considering the role of the health care provider in mHealth, further research into best clinical practices on how to support mHealth adoption through these 3 requirements is needed.

In addition, personal choice is closely related to the self-determination theory by Deci and Ryan [23], which emphasizes the importance of autonomy, competence, and relatedness in motivation and behavior change. In the context of mHealth apps for medication adherence, this theory suggests that patients should be given a sense of control over their health care decisions and that the use of financial incentives and gamification should be aligned with their personal values and goals. From our scoping review [18], we analyzed various underpinning theories or frameworks used for app development and found that the self-determination theory was used among the included studies. This indicates that the self-determination theory could be a useful framework for bolstering an mHealth app designed to address medication adherence. Such an app might incorporate gamification and financial incentives, where participants noted that social features foster a sense of community, encouraging connection and relatedness. A further analysis of patient perspectives on the various gamified and incentivized features in the context of this framework may help generate the default "recommended" or standard features of an app. Research into best practices specifically for financial incentives in medication adherence would also help address the concerns of exploitation and abuse, which have been identified in our findings and the literature, to undermine intrinsic motivation [44].

Government agencies are making efforts to manage the challenges and facilitate the opportunities associated with mHealth apps in general, which will be important to the application of gamification and financial incentives to promote medication adherence. For example, during the study and analysis, the Australian Digital Health Agency published an assessment framework for mHealth apps [45] in December 2022. The Australian Digital Health Agency is a statutory agency of the Australian government aimed at accelerating the adoption and innovation of digital technologies for health. The framework is intended to be used as a reference tool for app developers working on health apps in Australia. While not specific to medication adherence, gamification, or financial incentives, we found that there were similarities between the derived themes and subthemes in this study and the assessment domains (namely acceptability, safety and trust, ease of use, privacy and security, and technical quality assurance).

Contribution to Research

Overall, this study builds upon existing literature across medical modalities and conditions in mHealth apps to explore consumer perceptions of gamification or financial incentives for medication adherence. While some consumers appear ready to embrace the concept, it was surprising that participants ascribed such high importance to the notion that the widespread uptake of these features will require excellent governance and oversight. The themes have practical implications as a foundation for providing guidance to aspiring app developers. Further consultation with consumers, the industry, the government, and health providers will be required to ensure that apps using gamification and financial incentives are created and curated with clear guidelines and standards. It is also recommended that as the industry adopts these standards to create a specific app or app features, consumers are consulted throughout the design process, from conceptualization through early adoption and delivery to quality assurance. Future research is still warranted to discuss and evaluate the implementation of standards or frameworks in app development such as that provided by the Australian Digital Health Agency by consulting with more patients and industry stakeholders. This will ensure that the apps are acceptable and remain relevant and motivational to improve uptake.

Limitations

Two main limitations were identified in the design and analysis of this study that may have affected the interpretation of reported results. The first limitation source comes from the recruitment strategy. The need to scan a QR code to self-register interest in the study may have excluded certain participant groups such as those who identify as digital minimalists or technology adverse despite meeting the inclusion criteria. While the QR code ensures we recruited appropriately skilled participants to the study, there is a risk of unintentionally excluding certain participants resulting in the loss of some opinions on this topic. In addition, our recruitment strategy was not specifically designed to support a demographically diverse set of participants, with regard to language or the cultural background. The small sample size may have further impacted the lack of diversity. The use of focus groups, while benefiting from fruitful discussions among participants, also inherently includes limitations in the form of conformity bias as there may be disproportionate contributions by the participants and heterogeneity in character personalities.

The second limitation arose because the facilitator, analysts, and authors were from health care backgrounds, which potentially introduces bias in study findings through moderating the focus groups and analysis of the data. While all authors were also health care consumers, future studies could include a

stronger focus on personal reflexivity and involve a non–health professional and consumer in the research team.

Ultimately, while recruitment stopped because of signals indicating data saturation, the findings of this study are not intended to be taken as a final, comprehensive, and conclusive report on patient perspectives. As technology advances and changes, so do the public and individual perceptions of it. Since this study did not present participants with a working or beta model of a particular app, the findings are mainly limited to perspectives of the concepts of gamification and incentives in general, rather than any specific app.

Conclusions

Our findings provide an introductory understanding of patient perspectives on mHealth apps including gamification or financial incentives for medication adherence. Developers seeking to apply gamification and incentives in a general cohort of patients should strive to involve patients and their perspectives in all stages to inform design and development. This is critical, given the variation in consumer attitudes to the way incentives could be operationalized and concerns about sharing their personal data.

In addition, trust-based standards and personal choice and autonomy should be respected to support optimal app development in this modality. These considerations can effectively be summarized by using a comprehensive and validated framework. This field would gain from further research in the discussion and evaluation of such frameworks and share best practices with further patients and industry stakeholders.

Acknowledgments

The participants did not have a professional or personal relationship with the authors before the study.

Authors' Contributions

ST, a male industry professional, designed the study, collected data, and performed coding and analysis and drafting of the manuscript. SC, a male senior lecturer, designed the study, performed coding and analysis, and performed a review of the manuscript. LS, a female professor, designed the study, adjudicated the coding and analysis, and performed a review of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist applied to “understanding patient perspectives on the use of gamification and incentives in mobile health applications to improve medication adherence.”.
[DOCX File , 30 KB - mhealth_v12i1e50851_app1.docx]

Multimedia Appendix 2
Semistructured focus group guide.
[DOCX File , 21 KB - mhealth_v12i1e50851_app2.docx]

Multimedia Appendix 3
Coding mapping tree.
[PDF File (Adobe PDF File), 45 KB - mhealth_v12i1e50851_app3.pdf]

Multimedia Appendix 4
Additional quotes from participants.
[DOCX File , 26 KB - mhealth_v12i1e50851_app4.docx]

References

1. Adherence to long-term therapies: evidence for action. World Health Organization. 2003. URL: https://iris.who.int/bitstream/handle/10665/42682/9241545992.pdf [accessed 2024-05-01]
2. Saeed A, Amin QK, Saeed R, Yousafzai ZA. Comparing medication non-adherence in cardiovascular disease patients at public and private hospitals in Peshawar: a cross-sectional study of prevalence and contributing factors. Cureus 2023 Mar;15(3):e36345 [FREE Full text] [doi: 10.7759/cureus.36345] [Medline: 37082502]
3. Jimmy B, Jose J. Patient medication adherence: measures in daily practice. Oman Med J 2011 May;26(3):155-159 [FREE Full text] [doi: 10.5001/omj.2011.38] [Medline: 22043406]

4. Walsh CA, Cahir C, Tecklenborg S, Byrne C, Culbertson MA, Bennett KE. The association between medication non-adherence and adverse health outcomes in ageing populations: a systematic review and meta-analysis. Br J Clin Pharmacol 2019 Nov;85(11):2464-2478 [FREE Full text] [doi: 10.1111/bcp.14075] [Medline: 31486099]
5. Cutler RL, Torres-Robles A, Wiecek E, Drake B, Van der Linden N, Benrimoj SI, et al. Pharmacist-led medication non-adherence intervention: reducing the economic burden placed on the Australian health care system. Patient Prefer Adherence 2019;13:853-862 [FREE Full text] [doi: 10.2147/PPA.S191482] [Medline: 31213779]
6. Ruksakulpiwat S, Zhou W, Niyomyart A, Wang T, Kudlowitz A. How does the COVID-19 pandemic impact medication adherence of patients with chronic disease?: a systematic review. Chronic Illn 2023 Sep;19(3):495-513 [FREE Full text] [doi: 10.1177/17423953221110151] [Medline: 35971949]
7. Monokroussou M, Siati G, Theodorou M, Siskou O. Patient adherence to pharmaceutical care amid the economic crisis in Greece: underlying priority areas for policy makers. Health Policy 2020 Sep;124(9):971-976. [doi: 10.1016/j.healthpol.2020.05.025] [Medline: 32620402]
8. Nguyen A, Guttentag A, Li D, Meijgaard JV. The impact of job and insurance loss on prescription drug use: a panel data approach to quantifying the health consequences of unemployment during the Covid-19 pandemic. Int J Health Serv 2022 Jul;52(3):312-322 [FREE Full text] [doi: 10.1177/00207314221078749] [Medline: 35167394]
9. Laba TL, Lung T, Jan S, Scaria A, Usherwood T, Brien JA, et al. Medication non-adherence in a cohort of chronically ill Australians: a case of missed opportunities. Aust J Gen Pract 2018 Aug 01;47(8):556-562. [doi: 10.31128/ajgp-10-17-4351]
10. Backes C, Moyano C, Rimaud C, Bienvenu C, Schneider MP. Digital medication adherence support: could healthcare providers recommend mobile health apps? Front Med Technol 2020 Feb 17;2:616242 [FREE Full text] [doi: 10.3389/fmedt.2020.616242] [Medline: 35047896]
11. Li A, Del Olmo MG, Fong M, Sim K, Lymer SJ, Cunich M, et al. Effect of a smartphone application (Perx) on medication adherence and clinical outcomes: a 12-month randomised controlled trial. BMJ Open 2021 Aug 09;11(8):e047041 [FREE Full text] [doi: 10.1136/bmjopen-2020-047041] [Medline: 34373299]
12. Shortt M, Tilak S, Kuznetcova I, Martens B, Akinkuolie B. Gamification in mobile-assisted language learning: a systematic review of Duolingo literature from public release of 2012 to early 2020. Comput Assist Lang Learn 2021 Jul 05;36(3):517-554. [doi: 10.1080/09588221.2021.1933540]
13. Sailer M, Homner L. The gamification of learning: a meta-analysis. Educ Psychol Rev 2019 Aug 15;32(1):77-112. [doi: 10.1007/s10648-019-09498-w]
14. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000;55(1):68-78. [doi: 10.1037/0003-066X.55.1.68]
15. Ryan RM, Patrick H, Deci EL, Williams GC. Facilitating health behaviour change and its maintenance: interventions based on Self-Determination Theory. Eur Psychol 2008;10(1):2-5.
16. Notley C, Gentry S, Livingstone-Banks J, Bauld L, Perera R, Hartmann-Boyce J. Incentives for smoking cessation. Cochrane Database Syst Rev 2019 Jul 17;7(7):CD004307 [FREE Full text] [doi: 10.1002/14651858.CD004307.pub6] [Medline: 31313293]
17. Financial incentives on offer in SA to quit smoking. Government of South Australia. 2023 Apr 13. URL: https://tinyurl.com/kkdmbhfc [accessed 2024-05-01]
18. Tran S, Smith L, El-Den S, Carter S. The use of gamification and incentives in mobile health apps to improve medication adherence: scoping review. JMIR Mhealth Uhealth 2022 Feb 21;10(2):e30671 [FREE Full text] [doi: 10.2196/30671] [Medline: 35188475]
19. Bezerra Giordan L, Ronto R, Chau J, Chow C, Laranjo L. Use of mobile apps in heart failure self-management: qualitative study exploring the patient and primary care clinician perspective. JMIR Cardio 2022 Apr 20;6(1):e33992 [FREE Full text] [doi: 10.2196/33992] [Medline: 35442205]
20. Watanabe AH, Willis C, Ragsdale R, Biskupiak J, Moore K, Brixner D, et al. Patient perspectives on the use of digital technology to help manage cystic fibrosis. Pulm Med 2023 Jan 23;2023:5082499-5082497 [FREE Full text] [doi: 10.1155/2023/5082499] [Medline: 36727045]
21. Morrissey EC, Casey M, Glynn LG, Walsh JC, Molloy GJ. Smartphone apps for improving medication adherence in hypertension: patients' perspectives. Patient Preference Adher 2018 May;Volume 12:813-822. [doi: 10.2147/ppa.s145647]
22. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
23. Deci EL, Ryan RM. The "what" and "why" of goal pursuits: human needs and the self-determination of behavior. Psychol Inquiry 2000 Oct;11(4):227-268. [doi: 10.1207/s15327965pli1104_01]
24. Vrijens B, Antoniou S, Burnier M, de la Sierra A, Volpe M. Current situation of medication adherence in hypertension. Front Pharmacol 2017;8:100 [FREE Full text] [doi: 10.3389/fphar.2017.00100] [Medline: 28298894]
25. Tangari G, Ikram M, Ijaz K, Kaafar MA, Berkovsky S. Mobile health and privacy: cross sectional study. BMJ 2021 Jun 16;373:n1248 [FREE Full text] [doi: 10.1136/bmj.n1248] [Medline: 34135009]

26. Schroeder T, Haug M, Gewald H. Data privacy concerns using mHealth apps and smart speakers: comparative interview study among mature adults. JMIR Form Res 2022 Jun 13;6(6):e28025 [FREE Full text] [doi: 10.2196/28025] [Medline: 35699993]
27. Sax M, Helberger N, Bol N. Health as a means towards profitable ends: mHealth apps, user autonomy, and unfair commercial practices. J Consum Policy 2018 May 22;41(2):103-134. [doi: 10.1007/s10603-018-9374-3]
28. Galvin HK, DeMuro PR. Developments in privacy and data ownership in mobile health technologies, 2016-2019. Yearb Med Inform 2020 Aug;29(1):32-43 [FREE Full text] [doi: 10.1055/s-0040-1701987] [Medline: 32823298]
29. Mason M. Medibank, Optus hacks hurt, but Aussies still at risk. Financial Review. 2023 Aug 8. URL: https://tinyurl.com/3c9bue9e [accessed 2024-05-01]
30. Seh AH, Zarour M, Alenezi M, Sarkar AK, Agrawal A, Kumar R, et al. Healthcare data breaches: insights and implications. Healthcare (Basel) 2020 May 13;8(2):133 [FREE Full text] [doi: 10.3390/healthcare8020133] [Medline: 32414183]
31. Coventry L, Branley D. Cybersecurity in healthcare: a narrative review of trends, threats and ways forward. Maturitas 2018 Jul;113:48-52 [FREE Full text] [doi: 10.1016/j.maturitas.2018.04.008] [Medline: 29903648]
32. Wilkowska W, Ziefle M. Privacy and data security in E-health: requirements from the user's perspective. Health Informatics J 2012 Sep;18(3):191-201 [FREE Full text] [doi: 10.1177/1460458212442933] [Medline: 23011814]
33. Institute of Medicine. Beyond the HIPAA Privacy Rule: Enhancing Privacy, Improving Health Through Research. Washington, DC: The National Academies Press; 2009.
34. Hein AE, Vrijens B, Hiligsmann M. A digital innovation for the personalized management of adherence: analysis of strengths, weaknesses, opportunities, and threats. Front Med Technol 2020 Dec 14;2:604183 [FREE Full text] [doi: 10.3389/fmedt.2020.604183] [Medline: 35047888]
35. Koepp J, Baron MV, Hernandes Martins PR, Brandenburg C, Kira AT, Trindade VD, et al. The quality of mobile apps used for the identification of pressure ulcers in adults: systematic survey and review of apps in app stores. JMIR Mhealth Uhealth 2020 Jun 16;8(6):e14266 [FREE Full text] [doi: 10.2196/14266] [Medline: 32470916]
36. Young SR, Dworak EM, Byrne GJ, Jones CM, Yao L, Yoshino Benavente JN, et al. Remote self-administration of cognitive screeners for older adults prior to a primary care visit: pilot cross-sectional study of the reliability and usability of the MyCog mobile screening app. JMIR Form Res 2024 Feb 07;8:e54299 [FREE Full text] [doi: 10.2196/54299] [Medline: 38324368]
37. ISO / IEC 25010 : 2011 systems and software engineering — systems and software quality requirements and Evaluation (SQuaRE) — system and software quality models. International Organization for Standardization. URL: https://www.iso.org/standard/35733.html [accessed 2024-05-01]
38. Sanz M, Acha B, García MF. Co-design for people-centred care digital solutions: a literature review. Int J Integr Care 2021 Apr 30;21(2):16 [FREE Full text] [doi: 10.5334/ijic.5573] [Medline: 33981193]
39. Entwistle VA, Carter SM, Cribb A, McCaffery K. Supporting patient autonomy: the importance of clinician-patient relationships. J Gen Intern Med 2010 Jul;25(7):741-745 [FREE Full text] [doi: 10.1007/s11606-010-1292-2] [Medline: 20213206]
40. Cao J, Zhang G, Liu D. The impact of using mHealth apps on improving public health satisfaction during the COVID-19 pandemic: a digital content value chain perspective. Healthcare (Basel) 2022 Mar 04;10(3):479 [FREE Full text] [doi: 10.3390/healthcare10030479] [Medline: 35326957]
41. Ventola CL. Mobile devices and apps for health care professionals: uses and benefits. P T 2014 May;39(5):356-364 [FREE Full text] [Medline: 24883008]
42. Alwashmi MF, Fitzpatrick B, Davis E, Gamble JM, Farrell J, Hawboldt J. Perceptions of health care providers regarding a mobile health intervention to manage chronic obstructive pulmonary disease: qualitative study. JMIR Mhealth Uhealth 2019 Jun 10;7(6):e13950 [FREE Full text] [doi: 10.2196/13950] [Medline: 31199330]
43. Sivakumar B, Lemonde M, Stein M, Goldstein S, Mak S, Arcand J. Evaluating health care provider perspectives on the use of mobile apps to support patients with heart failure management: qualitative descriptive study. JMIR Cardio 2022 Oct 26;6(2):e40546 [FREE Full text] [doi: 10.2196/40546] [Medline: 36287588]
44. Promberger M, Marteau TM. When do financial incentives reduce intrinsic motivation? Comparing behaviors studied in psychological and economic literatures. Health Psychol 2013 Sep;32(9):950-957 [FREE Full text] [doi: 10.1037/a0032727] [Medline: 24001245]
45. Assessment framework for mHealth apps. Australian Digital Health Agency. URL: https://www.digitalhealth.gov.au/about-us/strategies-and-plans/assessment-framework-for-mhealth-apps [accessed 2024-05-01]

Abbreviations

COREQ: Consolidated Criteria for Reporting Qualitative Research
mHealth: mobile health
REDCap: Research Electronic Data Capture
SQuaRE: System and Software Quality Requirements and Evaluation
SWOT: strengths, weaknesses, opportunities, and threats

Edited by L Buis; submitted 14.07.23; peer-reviewed by E Wiecek, V Perez Jover, M Haug, C Zou; comments to author 03.11.23; revised version received 20.12.23; accepted 27.03.24; published 14.05.24.

Please cite as:
Tran S, Smith L, Carter S
Understanding Patient Perspectives on the Use of Gamification and Incentives in mHealth Apps to Improve Medication Adherence: Qualitative Study
JMIR Mhealth Uhealth 2024;12:e50851
URL: https://mhealth.jmir.org/2024/1/e50851
doi:10.2196/50851
PMID:38743461

Original Paper

Tuberculosis Treatment Compliance Under Smartphone-Based Video-Observed Therapy Versus Community-Based Directly Observed Therapy: Cluster Randomized Controlled Trial

Ponlagrit Kumwichar[1], MD, PhD; Tagoon Prappre[1], BSc; Virasakdi Chongsuvivatwong[1], MD, PhD

Department of Epidemiology, Faculty of Medicine, Prince of Songkla University, Hat Yai, Thailand

Corresponding Author:
Virasakdi Chongsuvivatwong, MD, PhD
Department of Epidemiology, Faculty of Medicine, Prince of Songkla University
15 Kanjanavanich Rd, Kho Hong
Hat Yai, 90110
Thailand
Phone: 66 74451165
Fax: 66 74451165
Email: cvirasak@medicine.psu.ac.th

Abstract

Background: There are no recent studies comparing the compliance rates of both patients and observers in tuberculosis treatment between the video-observed therapy (VOT) and directly observed therapy (DOT) programs.

Objective: This study aims to compare the average number of days that patients with pulmonary tuberculosis and their observers were compliant under VOT and DOT. In addition, this study aims to compare the sputum conversion rate of patients under VOT with that of patients under DOT.

Methods: Patient and observer compliance with tuberculosis treatment between the VOT and DOT programs were compared based on the average number of VOT and DOT compliance days and sputum conversion rates in a 60-day cluster randomized controlled trial with patients with pulmonary tuberculosis (VOT: n=63 and DOT: n=65) with positive sputum acid-fast bacilli smears and 38 observers equally randomized into the VOT and DOT groups (19 observers per group and n=1-5 patients per observer). The VOT group submitted videos to observers via smartphones; the DOT group followed standard procedures. An intention-to-treat analysis assessed the compliance of both the patients and the observers.

Results: The VOT group had higher average compliance than the DOT group (patients: mean difference 15.2 days, 95% CI 4.8-25.6; *P*=.005 and observers: mean difference 21.2 days, 95% CI 13.5-28.9; *P*<.001). The sputum conversion rates in the VOT and DOT groups were 73% and 61.5%, respectively (*P*=.17).

Conclusions: Smartphone-based VOT significantly outperformed community-based DOT in ensuring compliance with tuberculosis treatment among observers. However, the study was underpowered to confirm improved compliance among patients with pulmonary tuberculosis and to detect differences in sputum conversion rates.

Trial Registration: Thai Clinical Trials Registry (TCTR) TCTR20210624002; https://tinyurl.com/3bc2ycrh

International Registered Report Identifier (IRRID): RR2-10.2196/38796

(JMIR Mhealth Uhealth 2024;12:e53411) doi:10.2196/53411

KEYWORDS

video-enhanced therapy; tuberculosis; health care system; observed therapy; treatment compliance; lung disease; randomized trial; digital health; telehealth; telemedicine; mobile phone

Introduction

Video-observed therapy (VOT) facilitates remote monitoring of patients with tuberculosis [1] and constitutes an alternative program to directly observed therapy (DOT) [2]. VOT has 2 forms: synchronous VOT (S-VOT) and asynchronous VOT (A-VOT) [1]. In S-VOT, observers video call their patients for real-time observation of drug administration, whereas with A-VOT, observers can review the video sent by the patients at any time. Globally, A-VOT is preferred over S-VOT because

it allows patients the flexibility to record drug administration sessions, and the video can be reviewed multiple times [3].

In Thailand, approximately 80,000 tuberculosis cases are reported annually [4]. Since 1996, the country has implemented DOT to ensure treatment adherence [5]. Despite evidence from 2 previous studies indicating the poor sustainability of DOT, no changes have been made due to the lack of alternative strategies and resources [6,7]. The National Tuberculosis Control Program Guideline recommends community-based DOT, observed by health personnel, as the preferred approach [8]. However, in 60% to 75% of tuberculosis cases, family-based DOT is used instead of health personnel observation, reflecting the complacency of the health care system [9,10]. Since 2015, VOT has been used in some areas, without, however, using an accountability system [11]. This system included irregular S-VOT using the LINE (Line Corporation) app or an offline A-VOT that could not be audited daily [11]. The Thai VOT (TH VOT) system, an A-VOT system, has been devised and implemented in Songkhla province, serving as a testing area for the A-VOT system [12,13]. Rather than visiting patient homes in the community as in traditional DOT or performing irregular VOT as previously done, observers can feasibly use the TH VOT system to reduce their travel expenses, and each VOT session performed can be audited daily [12,13]. The TH VOT system is usable and convenient for patients, especially for those who usually take medication late at night [13]. However, the system's effectiveness in improving medication adherence compared to the traditional community-based DOT remains unknown.

Prior research in Western countries has shown that A-VOT surpasses DOT in ensuring patient adherence, cost-effectiveness, and overall acceptance [14-19]. These investigations evaluated A-VOT based on observation counts. For comparison, counts under DOT, which follow strict regulations, such as those in the United Kingdom and the United States, were also examined [15,18]. In contrast, the observation counts reported from DOT in Thailand are irregular, owing to a low level of accountability among observers [6,7,12]. Therefore, to evaluate the effectiveness of VOT compared to DOT, it is important to consider compliance from both the patients' and the observers' perspectives, unlike what has been evaluated in the prior studies [15-18].

The primary objective of this trial was to compare the average number of days that patients with pulmonary tuberculosis and their observers were compliant under VOT and DOT. This was conducted during the intensive phase of treatment, which lasted 60 days, and followed the published protocol [20]. We assumed that medication adherence depended on the compliance of both patients and observers in both the VOT and DOT programs. The results of this study will help determine whether A-VOT can completely replace conventional DOT in Thailand. This is in line with the "Thailand Operation Plan To End TB (2023 to 2027)," which aims at using innovative technology to control tuberculosis [21].

The secondary objective of this trial was to compare the clinical outcomes between the VOT and DOT groups. The clinical outcomes were sputum conversion and reporting of adverse events. This is useful for the future planning of the A-VOT system and for conducting further studies on a larger scale. The eligibility criteria and outcomes were registered before the commencement of the study (TCTR20210624002) [22].

Methods

Study Design

We conducted a cluster randomized controlled trial (RCT) in which an observer was assigned to a cluster of patients with pulmonary tuberculosis living in the same jurisdiction, using either DOT or VOT. The trial was registered in the Thai Clinical Trials Registry (TCTR20210624002). The trial protocol followed the CONSORT (Consolidated Standards of Reporting Trials) 2010 statement [23] and is available online [20].

Study Setting

In Thailand, individuals diagnosed with tuberculosis receive definitive diagnostic evaluations at a hospital located within the jurisdiction of their place of residence. Subsequently, specialized nurses trained in tuberculosis care (tuberculosis nurse) at each hospital delegate a tuberculosis staff member (DOT observer) to administer DOT services to patients located within the corresponding primary care unit (PCU) that aligns with the pertinent jurisdiction.

This study was conducted in the Hat Yai and Meuang Songkhla districts of Songkhla province, Southern Thailand, where a robust internet network is available. Regarding telecommunications services in Thailand, both AIS and TrueMove H corporations offer 4G and 5G networks with speeds that are well above the minimum requirement of 10 Mbps bandwidth for uploading videos from mobile phones [24]. All tuberculosis staff who served as the observers of this study had worked as DOT observers for at least 2 years. All participants were regular smartphone users.

Background of the Existing A-VOT in Thailand

In Thailand, the TH VOT mobile web system was developed for remote monitoring of antituberculosis drug adherence [12]. This system is accessible through any mobile web browser and is available as an app on Google Play Store [25]. It uses user authentication via the widely used LINE app, with daily LINE notifications for setting the times agreed upon [26-28]. Patients upload videos of their medication intake to the server, which then alerts observers to review these videos. Our previous study, conducted in November 2021, found high patient compliance (approximately 70%) and moderate observer compliance (approximately 50%-65%), despite the challenges posed by the Delta variant of the novel SARS-CoV-2 [13,29]. The system, requiring approximately 1 minute for patient video recording and 1.5 minutes for observer review, proved effective and faced no technical issues in areas with robust internet. It was particularly beneficial for patients taking medication late at night, allowing observers to review videos the following morning [13]. More details on system functions and usability are available on the internet in our previous studies [12,13,20].

Sampling Method and Recruitment Procedures

In our study area, we randomly selected 53 PCUs using a computer-generated list of random invitations. From these, we invited 38 observers from 38 PCUs based on the list of invitations, leaving 15 PCUs uninvolved. The randomized allocation lists, used for assigning PCUs at a 1:1 ratio to either the VOT or DOT group, were also generated by a computer. Patients with tuberculosis under the jurisdiction of the selected 38 PCUs were assessed for eligibility and subsequently invited to provide informed consent to participate in the study.

Participants

Observers (Cluster Level)

All 38 observers from randomly selected 38 PCUs consented to participate in the study; 19 observers were allocated to the VOT group, and the remaining 19 observers were allocated to the DOT group.

Patients (Individual Level)

Patients were considered eligible if they had newly active pulmonary tuberculosis with a positive acid-fast bacilli (AFB) sputum smear, were aged >18 years, owned a smartphone, could use the LINE app, and resided in the same jurisdiction as the observer. Participants were excluded if they had a condition that required specialist intervention, which precluded the 60-day follow-up in intensive phase, rifampicin-resistant tuberculosis evaluated by a cartridge-based nucleic acid amplification test (Xpert MTB and RIF, Cepheid), were unable to continue the treatment for 60 days, or had alcohol dependence.

Cointerventions

The patients were provided zipped bags daily for 60 days, each with a daily dose of their HRZE (isoniazid, rifampicin, pyrazinamide, and ethambutol) drug regimen [12]. Patients whose consent was registered in the database by a tuberculosis nurse were scheduled to take their medication (HRZE regimen) once daily. After each patient registration, the observer in the jurisdiction where the patient resided was notified through an autonotification of the official LINE (either DOT or VOT).

For monetary compensation, the patients received 300 baht (US $8.68) immediately after registration to cover the cellular internet cost for the first month. Further compensation was paid once the patients completed their 60-day intensive treatment without discontinuing the assigned intervention. They received 300 baht (US $8.68) as a reimbursement for cellular internet cost in the second month and 400 baht (US $11.57) for transportation of the sputum specimen on 3 consecutive days.

The observers who observed medication administration among patients for at least 15 daily sessions out of 60 sessions were compensated with 600 baht (US $17.36). They were also compensated for the cost of travel to visit their patients (4 baht [US $0.12] per km).

Assigned Interventions

Cluster Level

VOT for Observers

To avoid a learning curve on the VOT side, the observers performed real or simulated activities for 1 month before the trial [13].

After being notified of patient recruitment, the observers visited the patients at home on the first day. The observer would instruct the patient to redemonstrate the learned procedures [12] as a means of verifying their correct understanding of how to record and upload the video. This training and validation process for independent execution typically required approximately 30 minutes. The observer and patient set a time range for taking the medication, after which the system would send reminder notifications to both the patient and the observer via LINE. Next, the patient maintained a daily record of the drug-taking session, noted any adverse events, and sent a video to the observer through the TH VOT system. The observer reviewed the video, approved the session, and provided necessary advice through the LINE chat box. The observer followed up with a phone call if the patient failed to send the video within 30 minutes of the appointment. If the observers detected any mistakes performed by the patients, they would conduct a video call via LINE to correct the process; these video calls would take approximately 15 minutes.

DOT for Observers

Each patient and observer received a session booklet (more details in the protocol by Kumwichar et al [20]). After being notified by the automatic system, the observer conducted a home-visit DOT as a routine service. To validate the observers' recorded information, the patient and observer were requested to take a photo of the most recent page of the booklet and send it to the auditor through the official TH VOT LINE system every weekend. The auditor reviewed and recorded the number of daily compliance sessions in the database. Note that the observer was independently responsible for managing appointment times with their patients. There was no system support for scheduling appointments to mimic a conventional DOT.

Individual Level

VOT for Patients

After registration, patients in the VOT group were trained by their observer to record and upload a drug-taking video session according to the standard operating procedure [12]. Briefly, the patients had to set their video frame so that their face was clearly visible. All tablets and capsules should also be clearly visible. They then had to click the "record video" button to start video recording, noting that there is a warning below the button to complain of any nonserious adverse events that may have occurred during the video recording before taking the medication. Patients had to then pick up the pills and place them on their tongue. Next, they swallowed the pills using clear water from a (clear) glass, raised their tongue to show the sublingual area, and stuck out their tongue to show the palatal area. After the drug-taking process was completed, they had to click the

"end recording" button to upload the video. After uploading the video to the TH VOT system, the patients could watch an instructional video to remind themselves of the serious adverse effects, which, should they experience, they must stop the medication and call the observer immediately.

DOT for Patients

For patients in the DOT group, the tuberculosis nurse provided a booklet to record their daily drug intake and whether the intake was observed by the assigned observer. The tuberculosis nurse requested the patients to return the booklet and all zipped bags on the follow-up day to claim compensation. Each weekend, the auditor notified the patients to capture and send a recent booklet page to the official LINE chat, to which the observers did not have access. All daily reports from patients were recorded without verification, treating them as self-administered treatment.

Procedures for the Auditor to Review Each Video or Picture Session

Sessions in the VOT Group

"Day" was used as the time unit for judging compliance, and local times (GMT +7 hours) were recorded. The *morning* began at midnight, and the evening ended at 11:59 PM. However, daily compliance was judged as "achieved within the cut-off time" if the patients took their medication and submitted their videos before 6 AM on the following day. The auditor assessed the daily video sessions for both the patients and the observers based on the protocol [20].

Sessions in the DOT Group

The auditor scored daily compliance weekly based on booklet photos sent by patients with tuberculosis and their observers. The patients were considered to have daily compliance as reported (no audit). The auditor would make a phone call to patients with tuberculosis to confirm whether they were observed as reported by their observer and to remind them to safely store the booklet and all zipped bags received from the tuberculosis clinic, as per the protocol [20].

Follow-Ups

Each patient was scheduled to return to the tuberculosis clinic for follow-up on day 61. One day before the scheduled visit, the tuberculosis nurse reminded the patients in the DOT group to return the booklet and zipped bags. A deep cough specimen was collected early in the morning for 3 consecutive days (from days 61 to 63). The sputum specimens were subjected to the AFB test. The patients were requested to notify their physicians about all adverse events that occurred at the start of treatment. Physicians recorded the reported adverse events from history using the electronic health record (EHR) system and suggested appropriate treatment. If a patient missed their follow-up appointment, the responsible tuberculosis nurse contacted them and recorded their reasons in the EHR.

Data Collection

Data regarding observational activities were recorded in the database, and data regarding clinical outcomes were documented in the EHR system of the participating hospitals. The records were retrieved for analysis at the end of the follow-up period.

Outcomes

Primary Outcomes

The data recorded by the auditor were compiled to understand patient and observer compliance in each arm. For the compliance of individual patients, the daily compliance scores rated by the auditor were summed. The mean number of compliance days was calculated for all patients.

Similarly, for compliance of individual observers, daily compliance scores rated by the auditor were calculated. A higher number of patient doses observed increased the mean number of compliance days for the entire group of observers (VOT or DOT).

Secondary Outcomes

The clinical outcomes retrieved from the EHR, the conversion of the AFB smear (3 negative sputum smears), the reporting of adverse events, missing follow-up visits, and death during the 60-day follow-up period were compared between the 2 groups.

The information retrieved from the EHR system was used to compare the reporting of adverse events by observers in the VOT and DOT groups.

Sample Size

Each jurisdictional area comprised 1000 to 5000 individuals. With an approximate annual tuberculosis incidence of 130 per 100,000 individuals in the Songkhla province [30], the sample size estimate was based on the assumption that each cluster could recruit approximately 1 to 5 (mean 3) patients with tuberculosis within 9 months.

The sample size was calculated using the group RCT calculator [31]. The parameters are shown in the protocol [20]. The required number of clusters for each arm was 19. Thus, the number of patients with tuberculosis in each group was 57 (19×3). Using a sample size inflation factor of 20% to compensate for the uncertainty of the tuberculosis incidence in each jurisdictional area, a sample size of 70 patients with tuberculosis was estimated for each arm.

Cluster Randomized Allocation

Observers who consented to participate were randomly allocated to either the VOT or DOT groups using a file generated using the R software (R Foundation for Statistical Computing). The sequences were stored on a study server. Following the trial protocol, the participating observers registered themselves in the LINE system. After they pressed the "accept" button, the observers were informed about their allocated intervention group through the study LINE system.

Implementation of the Trial and Patient Information

The new patients with pulmonary tuberculosis were recruited to the VOT or DOT group by a tuberculosis nurse, depending on the jurisdiction of the observer's residence. Relevant information regarding the study was provided to the potential patients before the start of the trial, including highlighting who could observe them taking medication (their observer and

auditor) along with possible assigned interventions (VOT or DOT). The observers' intervention group was excluded before they consented to participate. If the patients consented to participate, they were assigned to the same intervention group as the observer in their jurisdiction. The participants were free to refuse the intervention at any point after receiving instructions from the tuberculosis nurse. Those who refused to participate or withdrew from the study continued the traditional DOT without data collection compliance. However, clinical data were collected as permitted in accordance with the Thai Personal Data Protection Act, 2019.

Blinding

The observers disclosed their assigned interventions to auditors, tuberculosis nurses, and researchers. Next, the researchers trained the VOT observers to familiarize themselves with the TH VOT system [13]; the DOT observers were requested to perform traditional DOT as routine care. Therefore, none of the researchers or staff involved in the study were blinded to the assigned interventions.

Statistical Analysis

Patient and observer background information was summarized using descriptive statistics. An intention-to-treat analysis was conducted according to a randomized allocation. Thus, the participants were classified according to the intervention group to which they were assigned, regardless of whether they changed observation modality. We compared the mean number of compliance days between the 2 groups 60 days after treatment initiation. For a straightforward discussion, we also calculated the compliance rate (%) of each group using the following formula: sum of compliant days for each group×100/(number of patients in each group×60)

Our study was a cluster RCT; thus, the number of compliance days of patients and observers was nested in clusters. We analyzed the 60-day compliance, considering that the same observer may monitor >1 patient. The intervention effect was based on a linear mixed effects model [32]. According to our study design, the intervention was a fixed effect, whereas the cluster level was a random effect. The estimated mean numbers of compliance days along with their SEs for 4 groups of patients under VOT and those under DOT were derived from the model that accounted for the cluster effect.

The number of compliance days for each individual, adjusted for clustering effects, would be calculated based on the specified model. Subsequently, a quantile-quantile (Q-Q) plot would assess the normality of the estimated numbers for each group. If data points predominantly align with a reference line, suggesting normality, a 2-tailed *t* test would be justified for comparing the estimated numbers under VOT versus DOT. Before conducting the *t* test, the Breusch-Pagan test would evaluate homoscedasticity [33]; a *P* value <.05 indicating heteroscedasticity necessitates the use of the Welch *t* test, whereas homoscedastic conditions would permit the application of the Student *t* test. In cases where data markedly deviate from the reference line, indicating non-normality, nonparametric methods would be applied. Data visualization for this study was conducted as outlined in the protocol [20].

Only descriptive statistical methods were used for the secondary outcomes because we did not have sufficient statistical power to detect small differences. The chi-square test was used for comparison; nevertheless, when the expected counts were <5, the Fisher exact test was used.

We also conducted a power analysis using the methodology proposed by Rutterford et al [34] to assess the robustness of our findings. For this analysis, the mean number of patients per cluster was used, and an estimated intraclass correlation coefficient of 0.2 was used to calculate the statistical power, expressed as a percentage. Power calculations were not performed for rare outcomes in which no events occurred in either the DOT or VOT groups.

All analyses were performed using the *epiDisplay* (version 3.5.0.2) [35], *tidyverse* (version 1.3.1) [36], *lmerTest* (version 3.1.3) [37], and *car* (version 3.1-2) [38] packages in R language and environment (version 4.1.1; R Core Team). Statistical significance was established at a 2-sided *P* value of <.05.

Deviation From the Registered Protocol

In this study, we added calculations for the compliance rate to provide more detail and readability than provided in the statistical analysis outlined in the registered protocol.

Ethical Considerations

The Human Research Ethics Committee of the Faculty of Medicine, Prince of the Songkla University, approved the trial on February 19, 2021 (REC 64-036-18-9). All participants consented to participate in the trial, allowed access to their data in the EHRs for this research, and consented to the reporting of results in a format in which individuals cannot be identified.

Results

Participants

Between January 2022 and May 2023, a total of 38 observers from 38 PCUs participated in cluster randomization, with 19 (50%) assigned to the VOT group and 19 (50%) to the control group. The trial ended in July 2023 because of a limited budget. A flow diagram of the observers and patients is shown in Figure 1. Eventually, 62.6% (92/147) of the eligible patients in the VOT group and 82.8% (106/128) of the patients in the DOT group consented. Exclusion of patients who consented occurred mainly because they required hospitalization and were transferred to the internal medicine department, which precluded participation in the 60-day intensive phase of VOT or DOT. Of the rest, 63 and 65 patients were recruited in the VOT and DOT groups, respectively. None of the patients changed the modality of observation. However, 21% (13/63) of the patients in the VOT group refused to record videos because of miscommunication with the observers and lack of training with respect to video recording. In the DOT group, 2% (1/65) of the patients refused the intervention after receiving a tutorial on the procedure because of personal concerns. One patient died of severe superimposed pneumonia.

Finally, 87% (55/63) of the individuals in the VOT group returned for a follow-up visit, compared with 74% (48/65) of the individuals in the DOT group. The number of missing

follow-up cases was not significantly different between the 2 groups.

All patients assigned to the allocations (63 in the VOT group and 65 in the DOT group) were followed up until the end of the trial. Data from all patients, including those who refused the intervention or died, were analyzed using an intention-to-treat approach.

Table 1 compares the baseline characteristics of the observers and patients in the VOT and DOT groups. The observers in both groups supervised a median of 4 patients each. None of the characteristics showed significant differences in distribution.

Figure 1. Study flow. Condition that requires specialist: all patients required hospitalization and were transferred to the internal medicine department; follow-up visit: sputum collection and clinical evaluation by a physician. DOT: directly observed therapy; VOT: video-observed therapy.

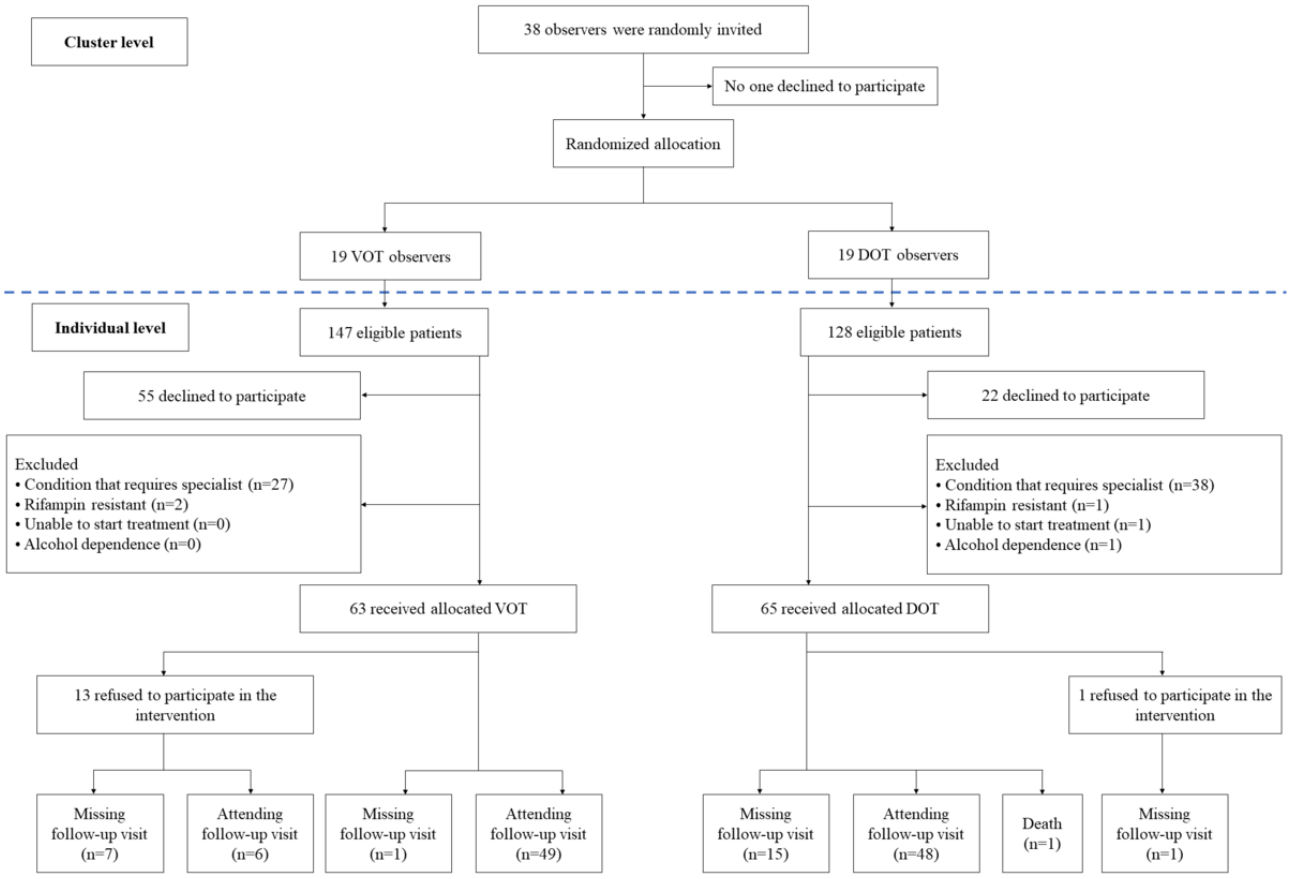


Table 1. Baseline characteristics of the participants.

Characteristics	VOT[a]	DOT[b]
Observers		
Total, n	19	19
Age (y), mean (SD)	37.6 (4.9)	35.6 (7.1)
Sex, n (%)		
Female	11 (58)	13 (68)
Male	8 (42)	6 (32)
Number of patients under supervision, median (IQR)	4 (2-4)	4 (3-4)
Patients with pulmonary tuberculosis		
Total, n	63	65
Age (y), mean (SD)	46.3 (14.2)	51.2 (16)
Sex, n (%)		
Female	21 (33.3)	17 (26.2)
Male	42 (66.7)	48 (73.8)
Weight (kg), mean (SD)	54.3 (11.1)	52.6 (8.5)
Height (cm), mean (SD)	163.2 (10.9)	159.9 (9.3)
BMI (kg/m^2), mean (SD)	20.4 (3.6)	20.5 (2.6)
Pulmonary lesion, n (%)		
Left	35 (55.6)	27 (41.5)
Right	17 (27)	20 (30.8)
Both	11 (17.5)	18 (27.7)
Cavitary lesion, n (%)	26 (41.3)	22 (33.8)
Underlying disease, n (%)		
Diabetes mellitus	11 (17.5)	15 (23.1)
HIV infection	3 (4.8)	4 (6.2)
COPD[c]	2 (3.2)	1 (1.5)
Any cancer	1 (1.6)	1 (1.5)
Number of tablets and capsules prescribed for daily administration, median (IQR)		
Isoniazid	3 (2-3)	3 (2-3)
Rifampicin	2 (1-2)	2 (1-2)
Pyrazinamide	3 (2-3)	3 (2-3)
Ethambutol	2 (2-3)	2 (2-2)

[a]VOT: video-observed therapy.

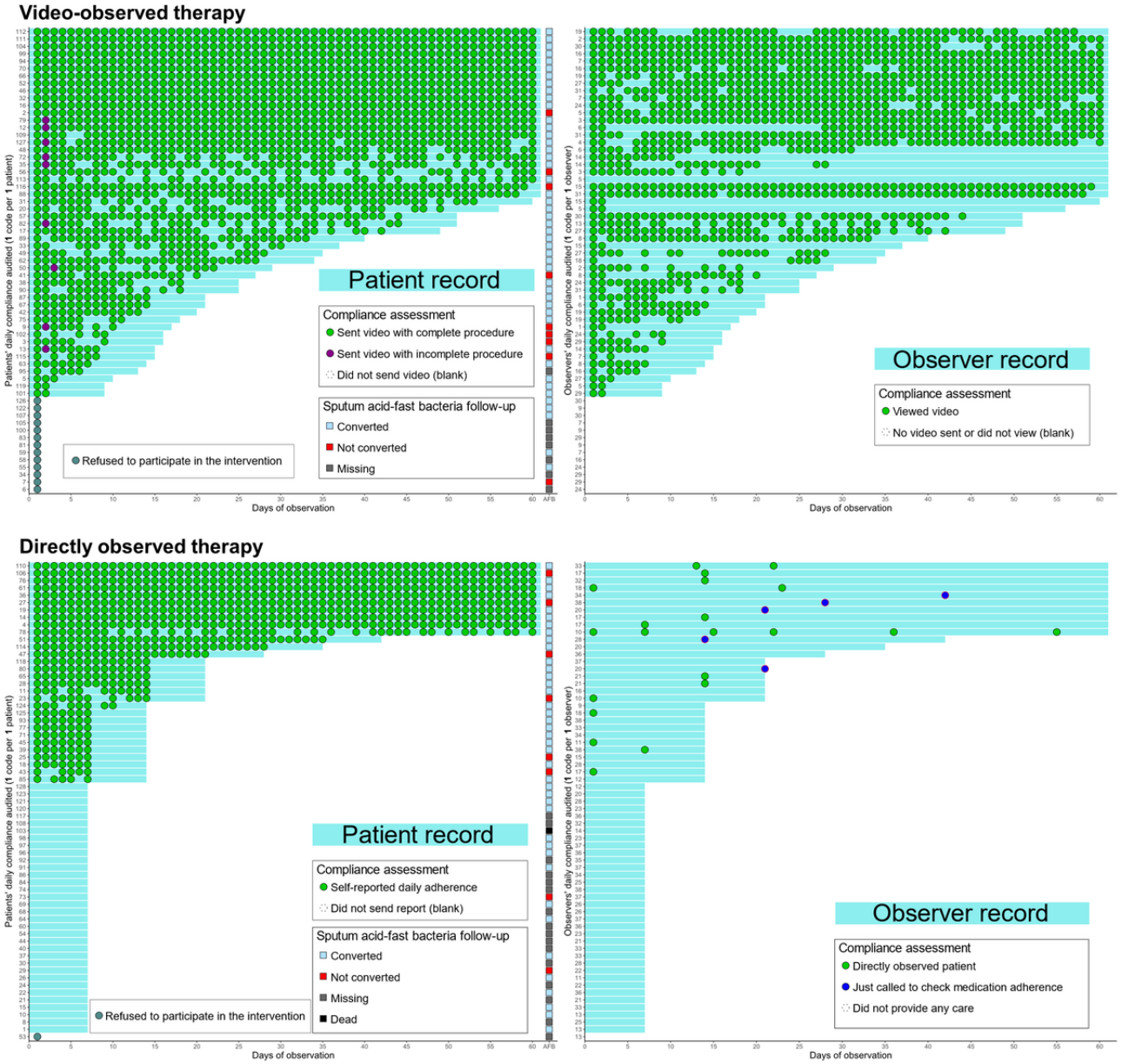
[b]DOT: directly observed therapy.

[c]COPD: chronic obstructive pulmonary disease.

Overall Patient and Observer Compliance

Figure 2 shows the 60-day treatment compliance of the patients (left column) and their observers (right column) in the time series of green dots along the x-axis. The y-axis indicates individual records. On the patient side, the last point of the individual time series represents the follow-up AFB smear. Overall, patient compliance correlated with observer compliance. More than half of the patients were unavailable to answer phone calls twice, at which point, they and their observers were discontinued from the compliance assessment.

Figure 2. Schematic depicting the 60-day treatment compliance of the patients and their observers in the video-observed therapy and directly observed therapy groups as well as the sputum follow-up results. The blue dots (representing only making a call and not conducting a home visit) were not counted as a compliance day.



Within the VOT group, no patient deviated from the recording protocol on day 1 of the compliance assessment because of the oversight provided by their at-home observer. Subsequently, 9 patients failed to comply with the recording protocol but were promptly corrected by their observers. Thereafter, the patients made no further mistakes. Remarkably, without considering the cluster effect, patient compliance rate in the VOT group approached 45.13% (1706/3780) over the 60-day period. In contrast, observer compliance rate was 35.19% (1330/3780).

In contrast, patient compliance rate in the DOT group was 20.9% (815/3900) over the 60-day period. However, the compliance rate of their observers (not including the blue dots in Figure 2) was only 0.54% (21/3900). More than half of the patients (59/65, 91%) initially received no attention. Only 21% (14/65) of the patients were observed at least once. Within 60 days, the number of compliance days for the observers ranged from 0 to 6. Furthermore, some observers simply made a call to check whether their patients had taken their medication without conducting a home visit but reported that the observation was complete (blue dots).

Outcomes

The cluster-adjusted number of compliance days for each group, as evidenced by the Q-Q plots in Multimedia Appendix 1, is assumed to follow a normal distribution. For VOT versus DOT group comparison, there was evidence of heteroscedasticity by the Breusch-Pagan test as $P<.001$ for patient and observer comparisons. The Welch t test was used for comparison of the primary outcomes. Table 2 presents a comparison of the primary and secondary outcomes of the VOT and DOT groups. When comparing the primary outcomes, the average compliance days adjusted for clustering for patients in the VOT group were significantly higher than for those in the DOT group, with a

mean difference of 15.2 (95% CI 4.8-25.6). Similarly, VOT observers reported significantly higher average compliance days compared with almost none for DOT observers, with a mean difference of 21.2 (95% CI 13.5-28.9). With the mean number of patients per cluster equal to 3, our study demonstrated sufficient statistical power (>80%) for detecting differences in compliance among observers.

Assessment of the follow-up secondary outcomes showed that 73% (46/63) of the patients in the VOT group achieved sputum AFB smear conversion, compared with 62% (40/65) of the patients in the DOT group (Table 2). The percentage of missed follow-up appointments was notably higher in the DOT cohort, at 25% (16/65), compared with the VOT cohort, at 13% (8/63). The number of adverse events reported by attending physicians was higher than that reported by observers. Overall, in the VOT group, 17% (5/29) of the adverse events recorded by physicians were detected by observers. Conversely, all the adverse events in the DOT group were overlooked by the observers. No statistical significance was detected for comparisons of all secondary outcomes.

Table 2. Primary and secondary outcomes evaluated in this study.

Outcomes	VOT[a] (n=63)	DOT[b] (n=65)	*P* value	Power (%)
Primary outcomes				
Estimated mean compliance days per person, mean (SE)[c]				
Patients	27.6 (4.4)	12.4 (3.0)	.005	67.5
Observers	21.5 (3.5)	0.3 (1.8)	<.001	90.4
Secondary outcomes				
Status at follow-up, n (%)				
AFB[d] smear converted[e]	46 (73)	40 (62)	.17	21.3
Missed the follow-up visit[e]	8 (13)	16 (25)	.08	30.6
Death[f]	0 (0)	1 (1.5)	—[g]	—
Reported adverse events during history taking by a physician, n (%)				
Nausea[e]	10 (16)	6 (9)	.26	16.0
Rash[e]	9 (14)	3 (5)	.06	35.2
Pruritus[f]	5 (8)	4 (6)	.74	5.0
Fatigue[f]	3 (5)	1 (2)	.36	14.5
Blurred vision[f]	1 (2)	0 (0)	.49	—
Numbness	1 (2)	0 (0)	.49	—
Adverse event reported by observers, n (%)				
Nausea[f]	3 (5)	0 (0)	.12	—
Rash[f]	1 (2)	0 (0)	.49	—
Pruritus	0 (0)	0 (0)	—	—
Fatigue[f]	1 (2)	0 (0)	.49	—
Blurred vision	0 (0)	0 (0)	—	—
Numbness	0 (0)	0 (0)	—	—

[a]VOT: video-observed therapy.

[b]DOT: directly observed therapy.

[c]The mean (SE) was calculated using mixed model linear regression while considering clustering. The Welch *t* test was performed.

[d]AFB: acid-fast bacilli.

[e]Chi-square test was performed.

[f]The Fisher exact test was performed.

[g]Insufficient data for statistical testing.

Discussion

Principal Findings

This study assessed medication adherence by evaluating compliance with the experimental intervention in both patients and their observers, using this as a surrogate outcome measure. We assumed that the higher compliance rates among both patients and observers would indicate greater medication adherence in patients. Overall, patients in the VOT group had a notably higher average number of compliance days than those in the DOT group, consistent with the compliance of the observers. However, this study was underpowered to detect improved compliance among the patients. The consent rate in the VOT group (92/147, 62.6%) was lower than that in the DOT group (106/128, 82.8%). This, combined with the fact that 13 patients in the VOT group refused the intervention after consenting due to a lack of support from their observers, implies a need for increased effort from the tuberculosis nurse at the tuberculosis clinic and the observers to enhance patient acceptability of the A-VOT system.

The TH VOT system includes an onscreen reminder before the "record video" button is clicked, prompting patients to report any adverse events in the video before taking their medication. This reminder process was absent in the DOT group. As shown in Table 2, none of the reported adverse events were serious; most patients could endure them. However, without active inquiry about these events by their observers, the patients did not report them. The failure of patients to communicate their adverse events to the observers could be an important factor related to low treatment compliance. Without this care from the observers, patients were less likely to engage in observation therapy, as it would not differ from self-administered therapy (SAT). This issue could be partially mitigated by the reminder interface in the TH VOT system, which prompts users to record a daily video. However, the observers were able to detect only 17% (5/29) of the adverse events retrospectively identified by physicians.

In this RCT, the patient characteristics were well balanced between the VOT and DOT groups. In addition, compliance in both patients and their observers was assessed within clusters, which is more practical than assessing individual effects in a community-based DOT setting [39]. In the VOT group, better compliance was observed and a higher percentage of patients achieved positive AFB smear conversion compared with the DOT group. However, this difference was not significant, potentially due to the limited sample size. The results also suggest a correlation between the compliance of patients and observers. This may indicate the influence of observers on patients in persuading them to adhere to the observation process. Observers in the VOT group reported approximately one-fifth of the adverse events recorded by physicians. In contrast, all these events were completely ignored by the observers in the DOT group. Therefore, training for both VOT and DOT as well as quality control of the observers are of utmost importance.

Previous studies in the United Kingdom and the United States have shown that more than half of the patients received successful observations in the intensive phase at a rate of ≥80% [15,18]. However, our results are less than half of those reported. The possible reasons for this discrepancy include that in Thailand, compulsory observations are required only for patients with extensively drug-resistant tuberculosis [40]. The treatment efficacy of DOT in Thailand has shown no significant difference compared with that of SAT, particularly in nonclinic-based DOT, which often transitions into SAT [5-7]. Given this context, more suitable comparative studies for the findings with the intended-but-failed DOT would have been those exploring the differences in treatment outcomes between VOT and SAT. However, our literature search did not yield any published studies presenting this comparison. Consequently, we referred to data from the United States, indicating that DOT, compared with SAT, resulted in a 40% increase in complete treatment (estimated as odds ratio−1) for individuals with latent tuberculosis infection [41]. Assuming that VOT would be as effective as DOT (as VOT in the United States has been shown to be equivalent to DOT [18,19]), we anticipated that VOT would similarly result in a 40% improvement in complete treatment over SAT in the US context. For our 2-month follow-up, which evaluated sputum conversion rates as a surrogate outcome for complete treatment, we hypothesized that these outcomes would parallel those in the complete treatment observed in the US study. However, the TH VOT system demonstrated only an 18.7% improvement in successful treatment (calculated as [73−61.5]/61.5×100). Compared with those in the United States, this finding underscores the need for more concerted efforts and regulation to enhance treatment success rates in Thailand.

In addition, this study was conducted during the COVID-19 pandemic. Observers might have used the pandemic as a pretext for poor compliance on observation [29]. However, the COVID-19 pandemic was not the main cause of poor compliance among observers, as we noted the poor compliance in our pilot study even before the pandemic began [12]. In addition, there has been evidence of poor compliance with DOT services among observers for >20 years [6,7].

Limitations

The main limitation of this study was that the trial period was restricted to the first 2 months of tuberculosis treatment (intensive phase). However, although sputum conversion is an uncertain surrogate for successful treatment, many studies have shown that this rate correlates well with treatment success [42-46]. Moreover, blinding was not possible. Nevertheless, both groups were monitored by the same auditor; consequently, the Hawthorne effect should be balanced [47]. In addition, we could not differentiate between the daily doses that were not observed and those that were not taken. Finally, we inferred that the higher compliance of both patients and observers with the assigned intervention indicated better medication adherence. The difference in sputum conversion rate and reporting of adverse events should be interpreted with caution due to inadequate sample size.

The compliance of patients in the VOT group was directly recorded on video to ensure accuracy. In contrast, compliance in the DOT group was based solely on patient reports, which could not be verified. The statistics based on potential

overreporting in the DOT group may have biased our results toward the underestimation of the superiority of the A-VOT system over DOT. This disparity in compliance quality limited the comparison of the 2 groups. Even so, the reported compliance in the DOT group may have been overestimated; however, it was still lower than that in the VOT group. This might be because the VOT system is more feasible than the traditional DOT and its notification system can enhance observer compliance [13]. Consequently, it may indirectly improve patient compliance through increased encouragement and response from observers [48].

Staff time and effort to train and supervise drug intake should be considered during the implementation of the A-VOT system. The A-VOT could reduce the travel time of the observers substantially. On the contrary, effort to train patients to use the A-VOT system, especially patients who are from a low socioeconomic background, and session recording supervision by S-VOT may be necessary for the first few days, when patients are still unfamiliar with the system.

Conclusions

In Thailand, although A-VOT requires more initial effort and has lower acceptability, it was superior to traditional community-based DOT in ensuring treatment compliance among observers. Nonetheless, the study lacked the statistical power to validate enhanced adherence to treatment among patients with pulmonary tuberculosis and to detect differences in sputum conversion rates. In community-based DOT settings with robust internet availability, replacing the DOT program with the A-VOT system may improve medication adherence among patients with tuberculosis, although a more accountable system for the observers is needed.

Acknowledgments

The authors would like to sincerely thank all the tuberculosis staff in the Hat Yai and Mueang Songkhla districts for their engagement in the implementation of the Thai video-observed therapy system. Development of the mobile app was supported by the Health Systems Research Institute, Thailand. This study was funded by the Fogarty International Center and the US National Institute of Allergy and Infectious Diseases of the US National Institutes of Health under award number D43 TW009522. The content of the manuscript is solely the responsibility of the authors, and it does not necessarily represent the official views of the US National Institutes of Health. In addition, the authors would like to express gratitude to Kannarin Bunmee and Asawin Sarabae from Hat Yai Hospital for their excellent coordination of the research project.

Data Availability

The data used in this study were deidentified to ensure patient confidentiality in accordance with the Personal Data Protection Act BE 2562 of Thailand. Therefore, the data could not be used to indirectly identify a person; however, they were sufficiently detailed to enable analysis and yield results consistent with the original data. The research data have been shared on GitHub [49].

Authors' Contributions

PK defined the conceptual framework of the study, led the development of video-observed therapy, and handled data analysis. PK and VC, who were responsible for obtaining the necessary data permissions, also played key roles in interpreting the results of the analysis. The process of data specification and curation involved collaboration between PK and TP. The first draft of the manuscript was created by PK, whereas VC played an essential role by contributing to the academic discourse and offering critical evaluations of subsequent drafts. Final approval of the manuscript was obtained from all authors.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary figures.
[DOCX File , 138 KB - mhealth_v12i1e53411_app1.docx]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 3775 KB - mhealth_v12i1e53411_app2.pdf]

References

1. Garfein RS, Doshi RP. Synchronous and asynchronous video observed therapy (VOT) for tuberculosis treatment adherence monitoring and support. J Clin Tuberc Other Mycobact Dis 2019 Apr 01;17:100098 [FREE Full text] [doi: 10.1016/j.jctube.2019.100098] [Medline: 31867442]
2. Bailey WC, Sbarbaro JA. Controversies in pulmonary medicine. All patients should receive directly observed therapy in tuberculosis. Am Rev Respir Dis 1988 Oct;138(4):1075-1076. [Medline: 3202443]

3. Truong CB, Tanni KA, Qian J. Video-observed therapy versus directly observed therapy in patients with tuberculosis. Am J Prev Med 2022 Mar;62(3):450-458. [doi: 10.1016/j.amepre.2021.10.013] [Medline: 34916094]
4. Global tuberculosis report 2023. World Health Organization. URL: https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2023 [accessed 2024-05-20]
5. Kamolratanakul P, Sawert H, Lertmaharit S, Kasetjaroen Y, Akksilp S, Tulaporn C, et al. Randomized controlled trial of directly observed treatment (DOT) for patients with pulmonary tuberculosis in Thailand. Trans R Soc Trop Med Hyg 1999;93(5):552-557. [doi: 10.1016/s0035-9203(99)90379-6] [Medline: 10696421]
6. Pungrassami P, Johnsen SP, Chongsuvivatwong V, Olsen J. Has directly observed treatment improved outcomes for patients with tuberculosis in southern Thailand? Trop Med Int Health 2002 Mar;7(3):271-279 [FREE Full text] [doi: 10.1046/j.1365-3156.2002.00849.x] [Medline: 11903990]
7. Pungrassami P, Chongsuvivatwong V. Are health personnel the best choice for directly observed treatment in southern Thailand? A comparison of treatment outcomes among different types of observers. Trans R Soc Trop Med Hyg 2002;96(6):695-699. [doi: 10.1016/s0035-9203(02)90354-8] [Medline: 12625153]
8. National tuberculosis control programme guideline, Thailand 2018. Division of Tuberculosis, Department of Disease Control Thailand. 2018. URL: https://www.tbthailand.org/download/Manual/NTP2018.pdf [accessed 2024-05-20]
9. Pungrassami P, Johnsen SP, Chongsuvivatwong V, Olsen J, Sørensen HT. Practice of directly observed treatment (DOT) for tuberculosis in southern Thailand: comparison between different types of DOT observers. Int J Tuberc Lung Dis 2002 May;6(5):389-395. [Medline: 12019914]
10. Anuwatnonthakate A, Limsomboon P, Nateniyom S, Wattanaamornkiat W, Komsakorn S, Moolphate S, et al. Directly observed therapy and improved tuberculosis treatment outcomes in Thailand. PLoS One 2008 Aug 28;3(8):e3089 [FREE Full text] [doi: 10.1371/journal.pone.0003089] [Medline: 18769479]
11. Rattanasuwan P. Tuberculosis treatment: DOT vs. VOT. J Bamrasnaradura Infect Dis Inst 2018;12(1):43-47.
12. Kumwichar P, Chongsuvivatwong V, Prappre T. Development of a video-observed therapy system to improve monitoring of tuberculosis treatment in Thailand: mixed-methods study. JMIR Form Res 2021 Jul 27;5(7):e29463 [FREE Full text] [doi: 10.2196/29463] [Medline: 34313602]
13. Kumwichar P, Chongsuvivatwong V, Prappre T. Video-observed therapy with a notification system for improving the monitoring of tuberculosis treatment in Thailand: usability study. JMIR Form Res 2022 May 06;6(5):e35994 [FREE Full text] [doi: 10.2196/35994] [Medline: 35522469]
14. Beeler Asay GR, Lam CK, Stewart B, Mangan JM, Romo L, Marks SM, et al. Cost of tuberculosis therapy directly observed on video for health departments and patients in New York City; San Francisco, California; and Rhode Island (2017-2018). Am J Public Health 2020 Nov;110(11):1696-1703. [doi: 10.2105/AJPH.2020.305877] [Medline: 32941064]
15. Story A, Aldridge RW, Smith CM, Garber E, Hall J, Ferenando G, et al. Smartphone-enabled video-observed versus directly observed treatment for tuberculosis: a multicentre, analyst-blinded, randomised, controlled superiority trial. Lancet 2019 Mar 23;393(10177):1216-1224 [FREE Full text] [doi: 10.1016/S0140-6736(18)32993-3] [Medline: 30799062]
16. Ravenscroft L, Kettle S, Persian R, Ruda S, Severin L, Doltu S, et al. Video-observed therapy and medication adherence for tuberculosis patients: randomised controlled trial in Moldova. Eur Respir J 2020 Aug 06;56(2):2000493 [FREE Full text] [doi: 10.1183/13993003.00493-2020] [Medline: 32381495]
17. Doltu S, Ciobanu A, Sereda Y, Persian R, Ravenscroft L, Kasyan L, et al. Short and long-term outcomes of video observed treatment in tuberculosis patients, the Republic of Moldova. J Infect Dev Ctries 2021 Sep 29;15(9.1):17S-24S [FREE Full text] [doi: 10.3855/jidc.14601] [Medline: 34609956]
18. Burzynski J, Mangan JM, Lam CK, Macaraig M, Salerno MM, deCastro BR, et al. In-person vs electronic directly observed therapy for tuberculosis treatment adherence: a randomized noninferiority trial. JAMA Netw Open 2022 Jan 04;5(1):e2144210 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.44210] [Medline: 35050357]
19. Perry A, Chitnis A, Chin A, Hoffmann C, Chang L, Robinson M, et al. Real-world implementation of video-observed therapy in an urban TB program in the United States. Int J Tuberc Lung Dis 2021 Aug 01;25(8):655-661 [FREE Full text] [doi: 10.5588/ijtld.21.0170] [Medline: 34330351]
20. Kumwichar P, Chongsuvivatwong V, Prappre T. Tuberculosis treatment compliance under smartphone-based video-observed therapy versus community-based directly observed therapy: protocol for a cluster randomized controlled trial. JMIR Res Protoc 2022 Jul 08;11(7):e38796 [FREE Full text] [doi: 10.2196/38796] [Medline: 35802403]
21. Thailand operation plan to end tuberculosis, phase 2 (2023-2027). Division of Tuberculosis, Department of Disease Control, Ministry of Public Health, Thailand. URL: https://tinyurl.com/23xvaake [accessed 2024-05-20]
22. Tuberculosis treatment compliance under smartphone-based video-observed therapy versus community-based directly observed therapy. Thai Clinical Trials Registry. 2021. URL: https://www.thaiclinicaltrials.org/show/TCTR20210624002 [accessed 2023-08-28]
23. Campbell MK, Piaggio G, Elbourne DR, Altman DG, CONSORT Group. Consort 2010 statement: extension to cluster randomised trials. BMJ 2012 Sep 04;345:e5661. [doi: 10.1136/bmj.e5661] [Medline: 22951546]
24. Wyrzykowski R. Thailand, 5G experience report. Opensignal. 2021 Nov. URL: https://www.opensignal.com/reports/2021/11/thailand/mobile-network-experience-5g [accessed 2023-12-14]
25. Android apps on Google Play. Google Play. URL: https://play.google.com/store?hl=th&gl=TH [accessed 2022-03-20]

26. Choemprayong S, Charoenlap C, Piromsopa K. Exploring usability issues of a smartphone-based physician-to-physician teleconsultation app in an orthopedic clinic: mixed methods study. JMIR Hum Factors 2021 Dec 20;8(4):e31130 [FREE Full text] [doi: 10.2196/31130] [Medline: 34931991]
27. Songsermpong S, Bunluesin S, Khomgongsuwan P, Junthon S, Cazabon D, Moran AE, et al. Innovations to sustain non-communicable disease services in the context of COVID-19: report from Pakkred district, Nonthaburi province, Thailand. Glob Heart 2021 Jun 10;16(1):44 [FREE Full text] [doi: 10.5334/gh.1003] [Medline: 34211830]
28. Amatavete S, Lujintanon S, Teeratakulpisarn N, Thitipatarakorn S, Seekaew P, Hanaree C, et al. Evaluation of the integration of telehealth into the same-day antiretroviral therapy initiation service in Bangkok, Thailand in response to COVID-19: a mixed-method analysis of real-world data. J Int AIDS Soc 2021 Oct;24 Suppl 6(Suppl 6):e25816 [FREE Full text] [doi: 10.1002/jia2.25816] [Medline: 34713623]
29. Pai M, Kasaeva T, Swaminathan S. Covid-19's devastating effect on tuberculosis care - a path to recovery. N Engl J Med 2022 Apr 21;386(16):1490-1493. [doi: 10.1056/NEJMp2118145] [Medline: 34986295]
30. National Tuberculosis Information Program. Division of Tuberculosis, Department of Disease Control Thailand. 2020. URL: https://lookerstudio.google.com/reporting/06209f22-6c12-4638-80c4-ee17400dddf4/page/qYUAD [accessed 2020-12-15]
31. Parallel group- or cluster-randomized trials (GRTs). National Institutes of Health. URL: https://researchmethodsresources.nih.gov/methods/grt [accessed 2022-03-02]
32. Schunck R, Perales F. Within- and between-cluster effects in generalized linear mixed models: a discussion of approaches and the Xthybrid command. Stata J 2017;17(1):89-115. [doi: 10.1177/1536867x1701700106]
33. Breusch TS, Pagan AR. A simple test for heteroscedasticity and random coefficient variation. Econometrica 1979 Sep;47(5):1287-1294. [doi: 10.2307/1911963]
34. Rutterford C, Copas A, Eldridge S. Methods for sample size determination in cluster randomized trials. Int J Epidemiol 2015 Jun;44(3):1051-1067 [FREE Full text] [doi: 10.1093/ije/dyv113] [Medline: 26174515]
35. Chongsuvivatwong V. epiDisplay: epidemiological data display package. The Comprehensive R Archive Network. 2022 May 18. URL: https://CRAN.R-project.org/package=epiDisplay [accessed 2022-11-29]
36. Wickham H. tidyverse: easily install and load the 'tidyverse'. The Comprehensive R Archive Network. 2023 Feb 22. URL: https://CRAN.R-project.org/package=tidyverse [accessed 2024-05-20]
37. Kuznetsova A, Brockhoff PB, Christensen RH, Jensen SP. lmerTest: tests in linear mixed effects models. The Comprehensive R Archive Network. 2020 Oct 23. URL: https://cran.r-project.org/web/packages/lmerTest/index.html [accessed 2023-08-06]
38. Fox J, Weisberg S, Price B, Adler D, Bates D, Baud-Bovy G, et al. car: companion to applied regression. The Comprehensive R Archive Network. 2023 Mar 30. URL: https://CRAN.R-project.org/package=car [accessed 2023-03-31]
39. Dron L, Taljaard M, Cheung YB, Grais R, Ford N, Thorlund K, et al. The role and challenges of cluster randomised trials for global health. Lancet Glob Health 2021 May;9(5):e701-e710 [FREE Full text] [doi: 10.1016/S2214-109X(20)30541-6] [Medline: 33865475]
40. Communicable Disease Act B.E. 2558 (A.D. 2015) (Tentative Translation*). Department of Disease Control. URL: https://ddc.moph.go.th/law_gcd_eng.php [accessed 2023-08-06]
41. Onwubiko U, Wall K, Sales RM, Holland DP. Using Directly Observed Therapy (DOT) for latent tuberculosis treatment - a hit or a miss? A propensity score analysis of treatment completion among 274 homeless adults in Fulton County, GA. PLoS One 2019 Jun 21;14(6):e0218373 [FREE Full text] [doi: 10.1371/journal.pone.0218373] [Medline: 31226132]
42. Kayigamba FR, Bakker MI, Mugisha V, Gasana M, Schim van der Loeff MF. Sputum completion and conversion rates after intensive phase of tuberculosis treatment: an assessment of the Rwandan control program. BMC Res Notes 2012 Jul 16;5:357 [FREE Full text] [doi: 10.1186/1756-0500-5-357] [Medline: 22800438]
43. Bisognin F, Amodio F, Lombardi G, Bacchi Reggiani ML, Vanino E, Attard L, et al. Predictors of time to sputum smear conversion in patients with pulmonary tuberculosis under treatment. New Microbiol 2019 Jul;42(3):171-175 [FREE Full text] [Medline: 31157399]
44. Azarkar Z, Sharifzadeh G, Ebrahimzadeh A, Olumi S. Time to sputum smear conversion in smear-positive pulmonary tuberculosis patients and factors for delayed conversion. Iran J Med Sci 2016 Jan;41(1):44-47 [FREE Full text] [Medline: 26722144]
45. Lv L, Li T, Xu K, Shi P, He B, Kong W, et al. Sputum bacteriology conversion and treatment outcome of patients with multidrug-resistant tuberculosis. Infect Drug Resist 2018 Jan 23;11:147-154 [FREE Full text] [doi: 10.2147/IDR.S153499] [Medline: 29416359]
46. Pefura-Yone EW, Kengne AP, Kuaban C. Non-conversion of sputum culture among patients with smear positive pulmonary tuberculosis in Cameroon: a prospective cohort study. BMC Infect Dis 2014 Mar 11;14:138 [FREE Full text] [doi: 10.1186/1471-2334-14-138] [Medline: 24618155]
47. McCambridge J, Witton J, Elbourne DR. Systematic review of the Hawthorne effect: new concepts are needed to study research participation effects. J Clin Epidemiol 2014 Mar;67(3):267-277 [FREE Full text] [doi: 10.1016/j.jclinepi.2013.08.015] [Medline: 24275499]

48. Clarke M, Dick J, Zwarenstein M, Lombard CJ, Diwan VK. Lay health worker intervention with choice of DOT superior to standard TB care for farm dwellers in South Africa: a cluster randomised control trial. Int J Tuberc Lung Dis 2005 Jun;9(6):673-679. [Medline: 15971396]
49. Kumwichar P. Analysis of tuberculosis treatment compliance under smartphone-based video-observed therapy versus community-based directly observed therapy: A cluster randomized controlled trial. GitHub. 2022. URL: https://github.com/ponlagrit/Cluster_RCT_VOT_vs_DOT [accessed 2024-05-20]

Abbreviations

AFB: acid-fast bacilli
A-VOT: asynchronous video-observed therapy
CONSORT: Consolidated Standards of Reporting Trials
DOT: directly observed therapy
EHR: electronic health record
HRZE: isoniazid, rifampicin, pyrazinamide, and ethambutol
PCU: primary care unit
RCT: randomized controlled trial
SAT: self-administered therapy
S-VOT: synchronous video-observed therapy
TH VOT: Thai video-observed therapy
VOT: video-observed therapy

Edited by L Buis; submitted 05.10.23; peer-reviewed by S Escobar-Agreda, M Benton, C Lam; comments to author 05.12.23; revised version received 26.12.23; accepted 08.05.24; published 03.06.24.

Please cite as:
Kumwichar P, Prappre T, Chongsuvivatwong V
Tuberculosis Treatment Compliance Under Smartphone-Based Video-Observed Therapy Versus Community-Based Directly Observed Therapy: Cluster Randomized Controlled Trial
JMIR Mhealth Uhealth 2024;12:e53411
URL: https://mhealth.jmir.org/2024/1/e53411
doi:10.2196/53411
PMID:38830205

Use of Electronic Patient Messaging by Pregnant Patients Receiving Prenatal Care at an Academic Health System: Retrospective Cohort Study

Kai Holder[1], BA; Joe Feinglass[2], PhD; Charlotte Niznik[3], APRN, CDCES; Lynn M Yee[3], MD, MPH

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

[2]Division of General Internal Medicine and Geriatrics, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

[3]Division of Maternal–Fetal Medicine, Department of Obstetrics and Gynecology, Northwestern Medicine, Chicago, IL, United States

Corresponding Author:
Kai Holder, BA
Feinberg School of Medicine, Northwestern University, , Chicago, IL, , United States

Abstract

Background: The COVID-19 pandemic accelerated telemedicine and mobile app use, potentially changing our historic model of maternity care. MyChart is a widely adopted mobile app used in health care settings specifically for its role in facilitating communication between health care providers and patients with its messaging function in a secure patient portal. However, previous studies analyzing portal use in obstetric populations have demonstrated significant sociodemographic disparities in portal enrollment and messaging, specifically showing that patients who have a low income and are non-Hispanic Black, Hispanic, and uninsured are less likely to use patient portals.

Objective: The study aimed to estimate changes in patient portal use and intensity in prenatal care before and during the pandemic period and to identify sociodemographic and clinical disparities that continued during the pandemic.

Methods: This retrospective cohort study used electronic medical record (EMR) and administrative data from our health system's Enterprise Data Warehouse. Records were obtained for the first pregnancy episode of all patients who received antenatal care at 8 academically affiliated practices and delivered at a large urban academic medical center from January 1, 2018, to July 22, 2021, in Chicago, Illinois. All patients were aged 18 years or older and attended ≥3 clinical encounters during pregnancy at the practices that used the EMR portal. Patients were categorized by the number of secure messages sent during pregnancy as nonusers or as infrequent (≤5 messages), moderate (6-14 messages), or frequent (≥15 messages) users. Monthly portal use and intensity rates were computed over 43 months from 2018 to 2021 before, during, and after the COVID-19 pandemic shutdown. A logistic regression model was estimated to identify patient sociodemographic and clinical subgroups with the highest portal nonuse.

Results: Among 12,380 patients, 2681 (21.7%) never used the portal, and 2680 (21.6%), 3754 (30.3%), and 3265 (26.4%) were infrequent, moderate, and frequent users, respectively. Portal use and intensity increased significantly over the study period, particularly after the pandemic. The number of nonusing patients decreased between 2018 and 2021, from 996 of 3522 (28.3%) in 2018 to only 227 of 1743 (13%) in the first 7 months of 2021. Conversely, the number of patients with 15 or more messages doubled, from 642 of 3522 (18.2%) in 2018 to 654 of 1743 (37.5%) in 2021. The youngest patients, non-Hispanic Black and Hispanic patients, and, particularly, non–English-speaking patients had significantly higher odds of continued nonuse. Patients with preexisting comorbidities, hypertensive disorders of pregnancy, diabetes, and a history of mental health conditions were all significantly associated with higher portal use and intensity.

Conclusions: Reducing disparities in messaging use will require outreach and assistance to low-use patient groups, including education addressing health literacy and encouraging appropriate and effective use of messaging.

(JMIR Mhealth Uhealth 2024;12:e51637) doi:10.2196/51637

KEYWORDS

patient portal; secure messaging; telehealth; health literacy; health disparities; disparity; disparities; telemedicine; information and communication technology; ICT; portals; messaging; use; technology use; pregnant; pregnancy; maternal; obstetric; obstetrics; prenatal; antenatal; demographic; demographics

Introduction

As health systems have adopted electronic health records, patient portal platforms, including widely used mobile apps such as MyChart, have proliferated in both primary care and medical specialties, including obstetrics [1,2]. A patient portal is a secure web-based interface connected to the personal electronic health record. Through the patient portal, individuals are able to review their health record, schedule appointments, refill prescriptions, and conduct secure direct messaging with health care providers [3]. The advent of mobile apps such as MyChart, recognized as the number-one medical app in terms of downloads on the Apple App Store and amassing more than 10 million downloads on Google Play, underscores the portability of these features, facilitating secure communication between patients and health care providers. Patient portals, augmented by mobile apps such as MyChart, provide patients with a resource where they can readily communicate with providers and actively participate in their health care [4]. The COVID-19 pandemic has accelerated the use of telemedicine and eHealth, potentially changing our historic model of maternity care [5].

Patient portal use has been steadily increasing since the 2009 Health Information Technology for Economic and Clinical Health (HITECH) Act and government financial incentives motivating adoption of required technology such as electronic patient portals [6]. Since patient portals have become more common in medical care, several studies have analyzed the effects of patient portals on clinical outcomes. In outpatient, primary care practices, the nonobstetrics literature has provided evidence that electronic portal use is associated with improved patient satisfaction and patient-provider communication [6-9]. Secure messaging within electronic patient portals has also been associated with positive clinical outcomes; for instance, diabetes management patients who used a secure messaging feature were found to have lower hemoglobin A_{1c} values [6,8-15]. Additionally, pilot studies analyzing portal use in obstetric populations have suggested that portals can be a useful tool for management of complex medical comorbidities. A study evaluating portal use and glucose control in an obstetric population found that patients who were active portal users were less likely to have within-goal glycemic control, suggesting that pregnant patients with suboptimal glycemic control may have been more readily engaged in secure messaging with providers [16]. However, previous studies analyzing portal use in obstetric populations have demonstrated significant sociodemographic disparities in portal enrollment and messaging, specifically showing that patients who are non-Hispanic Black, Hispanic, uninsured, and have low income were less likely to enroll in and use patient portals [15]. The rapid expansion of patient portals raises concern that disparities in health outcomes and health care access may be further exacerbated by disparities in technology access and electronic health literacy, which has been demonstrated in prior research [5,6]. Because of findings from previous studies that have suggested associations between patient portal use and improved clinical outcomes, there is a need to gain a deeper understanding of use patterns and the factors influencing the use of patient portals during pregnancy [6-16].

It was therefore of interest to evaluate our health system's use of an obstetric patient portal in the context of the rapid expansion of telemedicine since the start of the COVID-19 pandemic in 2020. This study aimed to estimate changes in patient portal use and intensity in prenatal care before and during the pandemic period and to identify sociodemographic and clinical disparities that continued during the pandemic.

Methods

Overview

This was a retrospective cohort study using electronic medical record (EMR) and administrative data from our health system's Enterprise Data Warehouse. Records were obtained for the first pregnancy episode of all patients who received antenatal care at 8 academically affiliated practices and delivered at a large urban academic medical center from January 1, 2018, to July 22, 2021, in Chicago, Illinois. The study included all patients aged 18 years or older who attended at least 3 clinical encounters during pregnancy at the practices that used the EMR portal.

Ethical Considerations

The study was approved by the Institutional Review Board of Northwestern University with waiver of informed consent (STU00202847). This study follows the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines for reporting observational research.

Trends in Patient Portal Use and Message Frequency

Our health system faculty practices use the EpicCare EMR and MyChart, the associated commercial patient portal. Patients can view records, review laboratory and imaging results, send messages to providers, schedule appointments, and request medication refills via MyChart. During a clinical encounter, a provider can generate an individualized access code designated for portal enrollment, or patients can self-enroll in the portal via email without a prior access code. To gain access to portal use and functions, patients must activate their MyChart account through the portal website. The patient can then access MyChart through web-based interfaces or mobile apps. A formal mechanism for declining the MyChart invitation is unavailable, and all patients are considered to be enrolled in the portal; thus, we were unable to analyze patients who were provided with a MyChart invitation but chose not to use the portal.

Only patient portal use for communication with obstetric providers (physicians, nurse practitioners, certified nurse midwives, or nurses within the Department of Obstetrics and Gynecology) was considered active use for this study. Communication with nonprenatal care providers was excluded from this analysis, as we solely analyzed messaging use within the patients' prenatal practice. Patients were considered enrolled in the portal if they had an account at the time of delivery. Portal enrollment was not specifically analyzed in this study. Patients were considered portal users if they sent at least 1 secure message during pregnancy. Portal users were further categorized by intensity, which was classified by the number of secure messages sent during pregnancy as infrequent (≤5 messages), moderate (6-14 messages), or frequent (≥15 messages). The categorization of portal use into infrequent, moderate, or

frequent categories was determined by assessing patient portal use patterns within our clinic's prenatal practices. Patients sending fewer than 5 secure messages were classified as infrequent users, signifying limited engagement with the portal. Those sending between 6 and 14 messages were categorized as moderate users, indicating a moderate level of portal interaction. Patients sending more than 15 secure messages were considered frequent users, reflecting active and regular portal engagement. These specific cutoff points were chosen to distinguish between different levels of patient portal engagement. Our data did not permit identification of "threads" across messages involving multiple messages on the same issue, as this would require natural language processing of message texts that could identify specific issue content and the duration of threads across a given prenatal time interval.

Patient Portal Use by Patient Sociodemographic and Clinical Characteristics

Patient sociodemographic and clinical data were categorized using hospital administrative data and *International Classification of Diseases, Tenth Revision* (*ICD-10*) diagnosis and procedure codes, as well as prenatal visits, characterized as <9, 9-12 or >12 visits. Maternal age was categorized as <20, 20-24, 25-29, 30-34, 35-39, and ≥40 years at the time of prenatal care enrollment. Race and ethnicity were categorized as Hispanic, non-Hispanic Black, non-Hispanic White, Asian, and other/unknown. Additional sociodemographic characteristics that were collected included preference for a non-English language, antenatal care insurance status (Medicaid versus private or other), and residential zip code. Maternal residential zip codes in Illinois were matched to census zip code tabulation areas (ZCTAs) using the 2020 American Community Survey for the percentage of households living at or below the poverty level. Individuals were categorized as living in ZCTAs with <5%, 5%-9.99%, 10%-19.99%, or >20% of households living at or below the poverty level, or being non–Illinois residents [17].

In addition to multifetal gestation and prior cesarean delivery, parity was categorized as either nulliparous, 1 pregnancy, or 2 or more pregnancies. Maternal BMI at birth was categorized as normal weight (≤24 kg/m^2), overweight (25-29 kg/m^2), or obese (≥30 kg/m^2). We used *ICD-10* data from the delivery admission to characterize the prevalence of any of a number of chronic conditions, including cardiac disease, bleeding disorder, pulmonary hypertension, chronic renal disease, gastrointestinal disease, HIV/AIDS, bariatric surgery, asthma, connective tissue or autoimmune disease, neuromuscular disease, and thyrotoxicosis. We also categorized patients as having a history of depression or severe mental illness (schizophrenia, bipolar disorder, or psychosis) or substance use. We found that hospital *ICD-10* coders had incorrectly coded *both* gestational and preexisting diabetes and hypertension for a substantial number of delivery admissions. We therefore characterized diabetes and hypertension codes as either uniquely preexisting only, gestational only, or coded for both.

Statistical Analysis

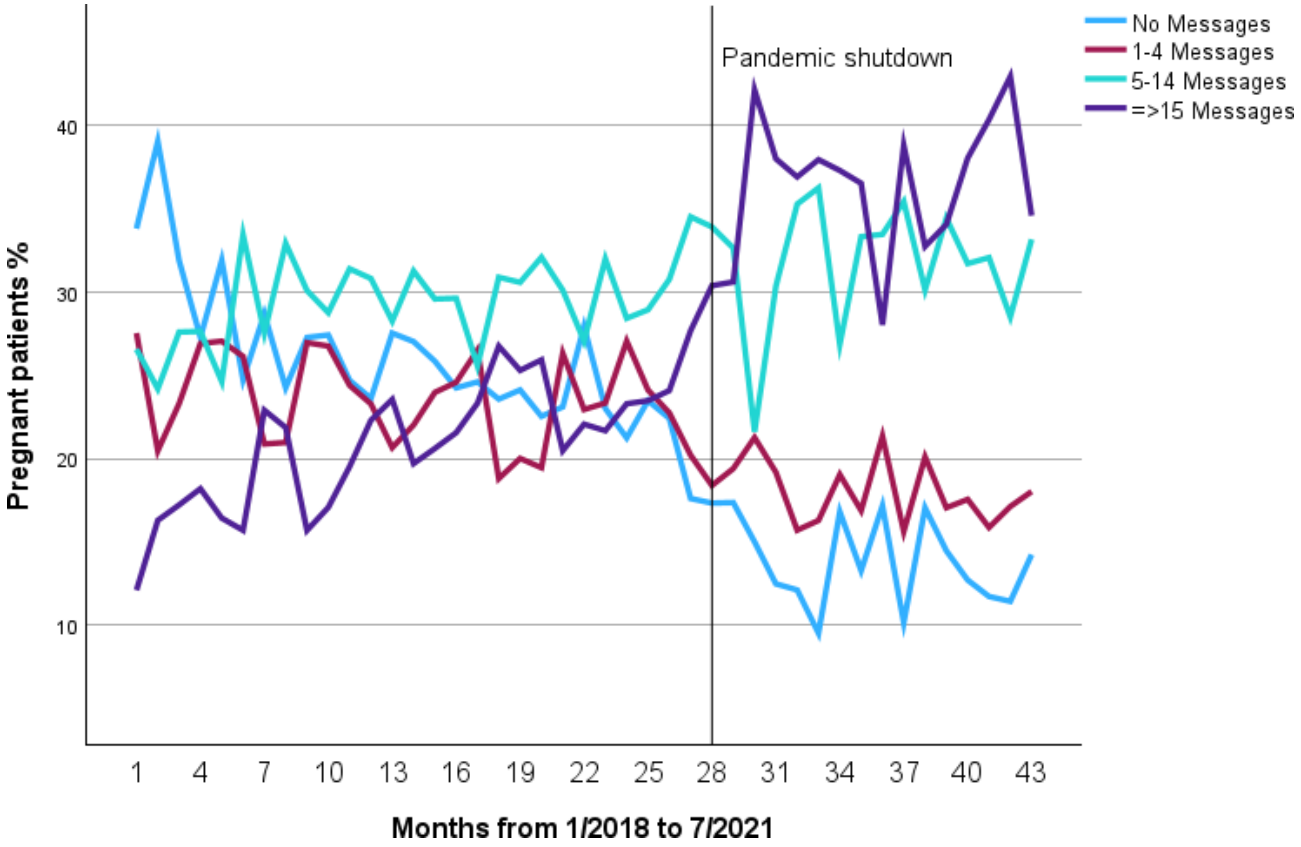
Monthly messaging intensity rates were computed to assess changes before and after the March 2020 pandemic. We determined the significance of bivariate differences in messaging frequency using χ^2 tests. We estimated a logistic regression model of the likelihood of zero portal use for patients who delivered from April 2020 during the 15 months after the end of the pandemic in March 2022. Analyses were performed using SPSS Statistics (version 28; IBM Corp).

Results

Trends in Portal Use and Message Frequency

A total of 12,380 patients were eligible for inclusion. In the total study period, 2681 (21.7%) of patients were nonusers of the portal, 2680 (21.6%) were infrequent users, 3754 (30.3%) were moderate users, and 3265 (26.4%) were frequent users. Figure 1 displays portal use rates by delivery date over the 43-month study period, denoting April 2020 (month 28) as the full onset of the COVID-19 pandemic. There was a modest trend toward increased use in the first study years, followed by a rapid increase in use and a correspondingly dramatic reduction in nonuse after the onset of the COVID-19 pandemic. The number of patients who were nonusers decreased between 2018 and 2021, from 996 of 3522 (28.3%) in 2018 to only 227 of 1743 (13%) in the first 7 months of 2021. Conversely, the number of patients with 15 or more messages doubled, from 642 of 3522 (18.2%) in 2018 to 654 of 1743 (37.5%) in 2021.

Figure 1. Pregnant patient portal messaging intensity. There were 12,380 patients with at least 3 prenatal visits with deliveries between January 1, 2018, and July 22, 2021.



Patient Portal Use by Sociodemographic and Clinical Characteristics

Table 1 displays patient characteristics by messaging intensity. The proportion of those with no portal use was correlated with the number of prenatal visits. Of the 2430 of 12,380 (19.6%) study patients with 8 or fewer visits, 1123 (46.2%) were nonusers overall, as compared to only 488 (10.8%) nonusers among 4520 patients (36.5%) in the sample with 13 or more visits. Those with the most prenatal visits had over 3 times the proportion of frequent messaging than those with the fewest visits. Higher parity was similarly correlated with lower message intensity.

Table . Obstetric patient electronic medical record portal use by number of messages sent during pregnancy among patients with at least 3 prenatal visits. There were 12,380 patients with deliveries from January 1, 2018, to July 22, 2021. All comparison were at *P*<.001 except preeclampsia (*P*=.03).

			Total, n (%)	No messages, n (%)[a]	≤5 messages, n (%)[a]	6-14 messages, n (%)[a]	≥15 messages, n (%)[a]
Overall			12,380 (100)	2681 (21.7)	2680 (21.6)	3754 (30.3)	3265 (26.4)
Delivery year							
	2019-2020		3522 (28.4)	996 (28.3)	863 (24.5)	1021 (29)	642 (18.2)
	2020-2021		4274 (34.5)	1038 (24.3)	984 (23)	1267 (29.6)	985 (23)
	2021-7/2022		4584 (37)	647 (14.1)	833 (18.2)	1466 (32)	1638 (35.7)
Prenatal visits (n)							
	<9		2430 (19.6)	1123 (46.2)	523 (21.5)	514 (21.2)	270 (11.1)
	9-12		5430 (43.9)	1070 (19.7)	1289 (23.7)	1709 (31.5)	1362 (25.1)
	>12		4520 (36.5)	488 (10.8)	868 (19.2)	1531 (33.9)	1633 (36.1)
Route of delivery							
	Vaginal delivery		9235 (74.6)	1974 (21.4)	2071 (22.4)	2865 (31)	2325 (25.2)
	Cesarean section		3145 (25.4)	707 (22.5)	609 (19.4)	889 (28.3)	940 (29.9)
Prior cesarean section			1781 (14.4)	509 (28.6)	379 (21.3)	494 (27.7)	399 (22.4)
Multiple gestation			448 (3.6)	118 (26.3)	92 (20.5)	116 (25.9)	122 (27.2)
Parity (n)							
	0		7266 (58.7)	1088 (15)	1451 (20)	2366 (32.6)	2361 (32.5)
	1		3458 (27.9)	891 (25.8)	849 (24.6)	1026 (29.7)	692 (20)
	≥2		1656 (13.4)	702 (42.4)	380 (22.9)	362 (21.9)	212 (12.8)
Sociodemographic characteristics							
	Age (years)						
		<20	74 (0.6)	52 (70.3)	18 (24.3)	2 (2.7)	2 (2.7)
		20-24	601 (4.9)	325 (54.1)	132 (22)	80 (13.3)	64 (10.6)
		25-29	1876 (15.2)	584 (31.1)	454 (24.2)	499 (26.6)	339 (18.1)
		30-34	5263 (42.5)	934 (17.7)	1155 (21.9)	1753 (33.3)	1421 (27)
		35-39	3693 (29.8)	624 (16.9)	758 (20.5)	1185 (32.1)	1126 (30.5)
		≥40	873 (7.1)	162 (18.6)	163 (18.7)	235 (26.9)	313 (35.9)
	Race/ethnicity						
		Asian/Pacific Islander	1264 (10.2)	212 (16.8)	243 (19.2)	429 (33.9)	380 (30.1)
		Hispanic	1616 (13.1)	572 (35.4)	365 (22.6)	388 (24)	291 (18)
		Non-Hispanic Black	1770 (14.3)	778 (44)	365 (20.6)	355 (20.1)	272 (15.4)
		Non-Hispanic White	6238 (50.4)	832 (13.3)	1355 (21.7)	2111 (33.8)	1940 (31.1)
		Other/unknown	1492 (12.1)	287 (19.2)	352 (23.6)	471 (31.6)	382 (25.6)
	Medicaid		1730 (14)	1095 (63.3)	349 (20.2)	179 (10.3)	107 (6.2)
	Non–English speaking		400 (3.2)	(44.8)	(18)	(23)	(14.2)
	Zip-code–level household poverty						
		<5%	4853 (39.2)	721 (14.9)	1000 (20.6)	1679 (34.6)	1453 (29.9)
		5%-9.99%	2973 (24)	568 (19.1)	667 (22.4)	927 (31.2)	811 (27.3)
		10%-19.99%	3105 (25.1)	863 (27.8)	710 (22.9)	809 (26.1)	723 (23.3)

		Total, n (%)	No messages, n (%)[a]	≤5 messages, n (%)[a]	6-14 messages, n (%)[a]	≥15 messages, n (%)[a]
	≥20%	1077 (8.7)	460 (42.7)	229 (21.3)	214 (19.9)	174 (16.2)
	Non-Illinois resident	372 (3)	69 (18.5)	74 (19.9)	125 (33.6)	104 (28)
Clinical characteristics						
BMI at delivery (kg/m^2)						
	<18.5	1377 (11.1)	260 (18.9)	292 (21.2)	430 (31.2)	395 (28.7)
	18.5-29.9	4793 (38.7)	848 (17.7)	1105 (23.1)	1542 (32.2)	1298 (27.1)
	>30	5498 (44.4)	1383 (25.2)	1136 (20.7)	1584 (28.8)	1395 (25.4)
Anemia		1559 (12.6)	451 (28.9)	302 (19.4)	392 (25.1)	414 (26.6)
Diabetes						
	Preexisting diabetes	291 (2.4)	70 (24.1)	41 (14.1)	71 (24.4)	109 (37.5)
	Gestational diabetes	1087 (8.8)	128 (11.8)	155 (14.3)	354 (32.6)	450 (41.4)
	Preexisting diabetes and gestational diabetes	332 (2.7)	63 (19)	36 (10.8)	96 (28.9)	137 (41.3)
Hypertension						
	Preexisting hypertension	672 (5.4)	185 (27.5)	126 (18.8)	155 (23.1)	206 (30.7)
	Gestational hypertension	677 (5.5)	168 (24.8)	128 (18.9)	194 (28.7)	187 (27.6)
	Preexisting hypertension and gestational hypertension	316 (2.6)	83 (26.3)	46 (14.6)	89 (28.2)	98 (31)
Preeclampsia		141 (1.1)	39 (27.7)	25 (17.7)	31 (22)	46 (32.6)
Preexisting comorbidity[b]		1973 (15.9)	400 (20.3)	409 (20.7)	544 (27.6)	620 (31.4)
History of depression or severe mental illness		3950 (31.9)	666 (16.9)	729 (18.5)	1193 (30.2)	1362 (34.5)
Substance use		312 (2.5)	113 (36.2)	55 (17.6)	69 (22.1)	75 (24)

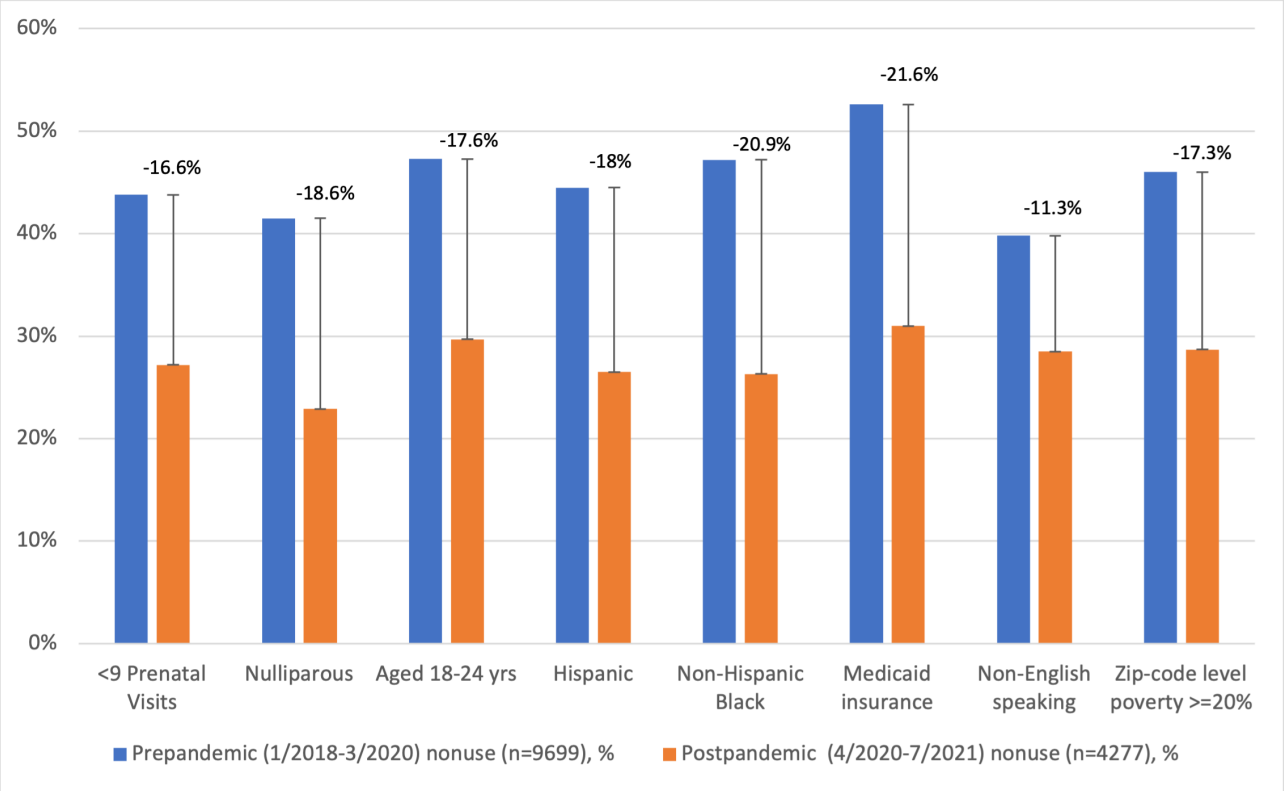
[a]Percentage calculated against the total for each characteristic.

[b]Preexisting comorbidities included cardiac disease, bleeding disorder, pulmonary hypertension, chronic renal disease, gastrointestinal disease, HIV/AIDS, bariatric surgery, asthma, connective tissue or autoimmune disease, neuromuscular disease, thyrotoxicosis, history of depression or severe mental illness (schizophrenia, bipolar disorder, or psychosis), or substance use.

Portal use was more frequent among those who were older, with the lowest messaging intensity among the youngest patients (younger than 25 years), who represented only 675 (5.5%) of all births. The highest messaging intensity was among patients older than 30 years. Non-Hispanic Black patients had the highest proportion of nonuse at 783 of 1784 (44%), followed by Hispanic patients at 581 of 1629 (35.4%) and non–English-speaking patients at 180 of 401 (44.8%). While patients with Medicaid comprised only 14% of the patient population, they had the highest proportion of nonuse among all patient subgroups at 1095 of 1730 (63.3%), and there was a clear poverty-level gradient in messaging intensity across ZCTAs.

Figure 2 presents change in portal use over time for patients with sociodemographic and clinical characteristics associated with high rates of portal nonuse who delivered after April 2020. Patients with Medicaid births had a 21.6% reduction in their nonuse rate after the pandemic. Similar large increases in use occurred among all the groups profiled, including those with fewer prenatal visits, as well as younger, minority, and lower-income patients.

Figure 2. Change in the proportion of patients with highest portal nonuse after the April 2020 pandemic shutdown.



Odds of Pandemic Portal Nonuse

Our logistic model of nonuse, based on the 4277 patients with postpandemic deliveries, showed that the number of visits and parity remained strongly correlated with the likelihood of portal nonuse (Table 2). Those with <8 visits were 97% more likely to have been nonusers than those with >12 prenatal visits, and patients with ≥2 parity were over 3 times more likely to be nonusers than nulliparous patients. Confirming bivariate results for the whole period, the youngest patients, non-Hispanic Black patients, and Hispanic patients, particularly non–English speaking patients, had the highest odds of continued nonuse. Patients with gestational diabetes were significantly more likely to be portal users, while patients with other comorbid conditions and those with preexisting mental health conditions were also more likely to use the portal.

Table . Logistic regression results for postpandemic obstetric patient nonuse of electronic medical record portal messaging among patients (n=4277) with deliveries from April, 2020, to July 22, 2021.

Characteristics			Odds ratio (95% CI)
Sociodemographic characteristics			
	Age (years)		
		<20	3.58 (2.02-6.33)
		20-24	1.90 (1.50-2.40)
		25-29	1.24 (1.06-1.45)
		30-34	Reference
		35-39	0.90 (0.79-1.03)
		≥40	0.84 (0.67-1.06)
	Race/ethnicity		
		Asian/Pacific Islander	1.00 (0.82-1.22)
		Hispanic	1.58 (1.33-1.87)
		Non-Hispanic Black	1.83 (1.52-2.20)
		Non-Hispanic White	Reference
		Other/unknown	1.18 (0.99-1.40)
	Medicaid insurance		4.27 (3.64-5.01)
	Non–English speaking		2.52 (1.93-3.31)
	Zip-code–level household poverty		
		<5%	Reference
		5%-9.99%	1.03 (0.89-1.19)
		10%-19.99%	1.14 (0.98-1.31)
		≥20%	1.15 (0.93-1.43)
		Non–Illinois resident	1.00 (0.73-1.37)
Clinical characteristics			
	Anemia		0.95 (0.81-1.12)
	BMI at delivery (kg/m^2)		
		<18.5	Reference
		18.5-29.9	0.87 (0.74-1.0)
		≥30	0.96 (0.82-1.12)
	Diabetes coding		
		Preexisting diabetes	0.77 (0.53-1.12)
		Gestational diabetes	0.34 (0.27-0.43)
		Preexisting diabetes and gestational diabetes	0.63 (0.44-0.90)
	Hypertension coding		
		Preexisting hypertension	1.16 (0.91-1.79)
		Gestational hypertension	1.11 (0.88-1.41)
		Preexisting hypertension and gestational hypertension	1.09 (0.86-1.38)
		Preeclampsia	0.94 (0.58-1.54)
		Preexisting comorbidity[a]	0.81 (0.69-0.94)
		History of depression or severe mental illness	0.69 (0.61-0.78)

[a]Preexisting comorbidities included cardiac disease, bleeding disorder, pulmonary hypertension, chronic renal disease, gastrointestinal disease, HIV/AIDS, bariatric surgery, asthma, connective tissue or autoimmune disease, neuromuscular disease, thyrotoxicosis, history of depression or severe mental illness (schizophrenia, bipolar disorder, or psychosis), or substance use.

Discussion

Principal Findings

This study was undertaken to guide efforts to improve prenatal portal use by addressing what were already known to be significant socioeconomic disparities in use. We found that messaging frequency significantly increased over the 43-month study period. Most notably, there was a decline in nonuse in the immediate postpandemic period, approximately cutting in half the previous nonuser population from 25.8% to 13.9%. However, pandemic nonuse remained concentrated among patients with fewer prenatal visits, higher parity, and public insurance, as well as those who were younger, were not English speakers, had lower income, and identified as non-Hispanic Black or Hispanic. As hypothesized, patients with high-risk clinical conditions, such as history of depression or severe mental illness, preexisting comorbidities, or gestational or preexisting diabetes, were more likely to use the portal and to be frequent messengers.

The socioeconomic disparities found in this cohort are consistent with prior literature analyzing patient portal use and messaging in obstetric and nonobstetric populations. It has been suggested in several studies that those who identify as non-White, have low income, and are publicly insured are less likely to use portals than their counterparts [5,15,18-22]. Numerous studies have demonstrated that non-Hispanic Black, Hispanic, low-income, and publicly insured or uninsured patients were less likely to activate and use portal messaging. A study of more than 1700 primary care patients receiving care at Kaiser Permanente Georgia showed that compared to non-Hispanic White participants, non-Hispanic Black patients were less likely to register for electronic patient portals, even after controlling for differences in education, income, or internet access, which were also associated with portal registration [4,15]. Previous studies have shown conflicting conclusions regarding the associations of high-risk clinical characteristics and portal use intensity [10,23-25]. For example, 2 studies in a primary care setting found that patients with chronic conditions and conditions associated with increased morbidity used the portal more often than healthier patients [4,24]. However, another study in a similar primary care setting conversely found that patients with fewer medical problems used the portal more often than patients with chronic conditions [25].

Potential Value of Increasing Patient Portal Use in Obstetrics

During the COVID-19 pandemic, health care providers increasingly relied on telehealth and electronic communication methods for health care delivery and patient care [5,19,20]. Telehealth and electronic patient communication were especially important for obstetrics, due to the necessity to continue scheduled prenatal care visits and screenings. The findings in this study demonstrate that messaging may have become a more important care function in the postpandemic period, despite the return to in-person visits.

The antenatal period represents a critical time, when health care engagement may influence both maternal and infant outcomes; however, disparities in antenatal portal use have been identified. Several studies that analyzed portal use in the setting of obstetrics demonstrate greater portal use in antenatal patients who have chronic conditions and are at higher risk for pregnancy complications; however, other studies suggest that medically high-risk obstetric patients were less likely to enroll and use portals [15,16,26]. This cohort's findings align with the former, demonstrating that participants with high-risk clinical conditions, including preexisting comorbidities, preexisting mental illness, and gestational diabetes, were more likely to use the portal and use the portal frequently.

Implications for Improving Patient Communication With Providers in Obstetrics

With secure messaging, patients can discuss a variety of topics, such as a change in condition, a new condition, laboratory results, and prescription concerns. A clinical example of the impact of secure messaging is the management of gestational diabetes mellitus. In this patient population, secure messaging serves as a platform for health care providers to stimulate patient engagement and self-management during a very disruptive obstetric complication.

Upon diagnosis of gestational diabetes mellitus, patients are bewildered as to how to alter lifestyle habits, shop, and prepare for meals, let alone monitor blood glucose to keep their child safe. Unlike an office visit, secure messaging allows health care providers to instantly explain the diagnosis and send patient-education tools, guides, and even YouTube instructions so the patient can immediately launch into self-care management action. This patient action empowers patients to use secure messaging to maintain close communication with the health care providers. Secure messaging solidifies diabetes self-management for weekly review of blood glucose in order to determine the need for insulin therapy. Another challenge in pregnancy if insulin therapy is necessary is for patients to continue to engage in health restoration for themselves and their child. Patient education on insulin administration can be provided via telemedicine or office visits, with weekly glucose review for any insulin adjustment—all through secure messaging for the remainder of the pregnancy. Though the literature has demonstrated telemedicine and mobile health as modalities to improve diabetes control, evidence for the significance of clinical metrics for patient satisfaction with secure messaging is scant, although it has been shown to cause no harm. There is need for more research in this area.

It is clear that patient portals provide an additional avenue for patient-provider communication; however, the disparities found in this paper show that the portal is not equally accessible to several sociodemographic groups.

An early obstetrics encounter is the best time to introduce patients to the value and usefulness of the patient portal. Often, patients need assistance with setting up the patient portal. A

mixed methods study analyzing factors affecting patient portal use among pregnant women with low income demonstrated that 33% of participants did not use the portal because they were unsure how to use it [26]. A practice message is always good to send while in the obstetrics office so the patient is comfortable with the EMR messaging application. Discussion of what will be available to the patient, such as lab results and appointment reminders, can lead to discussing the benefits of sharing blood glucose readings on a weekly basis. This can be crucial for medication and nutrition adjustment to improve diabetes self-care during pregnancy.

It is apparent that there is a need for innovative strategies and interventions aimed at enrolling Medicaid beneficiaries, those in poverty, young patients, and racial/ethnic minorities. With increased mobile phone use rates among women with low income, there is potential to increase portal use through interventions aimed at portal use via mobile apps.

Limitations

Given that this is a retrospective study, the associations made in this study can not be assumed to be causal, and there is potential for unmeasured confounding. For example, this study did not have direct measures for health literacy, internet access, education level, or self-care behaviors, which may play a role in portal use. Additionally, although this sample was large and diverse, patients in this study received care at a large academic tertiary care center and findings may not be fully generalizable to other health settings. Furthermore, portal enrollment was not specifically studied with this data set, which is a parameter that has potential to highlight further sociodemographic and clinical disparities in portal use. As stated in the Methods section, our data did not permit identification of “threads” across multiple messages on the same issue, which would require natural language processing. This could be an important next step in messaging research.

Conclusion

Portal use has major implications for redesigning obstetric delivery systems. This study documents that portal use and messaging frequency at our center have significantly increased in recent years, especially after the COVID-19 pandemic. Obstetric patients with high-risk chronic conditions are now more likely to use patient portals and to message frequently after the pandemic. However, sociodemographic disparities continue to exist fo portal use and intensity. Patient portals have the ability to encourage patient engagement and improve patient-provider communication and shared decision-making. Inclusive health literacy strategies developed for obstetric patients, such as training for and encouragement of portal use, may be a strategy to reduce disparities and improve outcomes. Future studies should focus on evaluating health and digital literacy interventions to address disparities in portal use.

Data Availability

The data sets generated during and/or analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

References

1. Huennekens K, Oot A, Lantos E, Yee LM, Feinglass J. Using electronic health record and administrative data to analyze maternal and neonatal delivery complications. Jt Comm J Qual Patient Saf 2020 Nov;46(11):623-630. [doi: 10.1016/j.jcjq.2020.08.007] [Medline: 32921579]
2. Wright A, Feblowitz J, Samal L, McCoy AB, Sittig DF. The Medicare electronic health record incentive program: provider performance on core and menu measures. Health Serv Res 2014 Feb;49(1 Pt 2):325-346. [doi: 10.1111/1475-6773.12134] [Medline: 24359554]
3. Irizarry T, DeVito Dabbs A, Curran CR. Patient portals and patient engagement: a state of the science review. J Med Internet Res 2015 Jun 23;17(6):e148. [doi: 10.2196/jmir.4255] [Medline: 26104044]
4. Roblin DW, Houston TK, Allison JJ, Joski PJ, Becker ER. Disparities in use of a personal health record in a managed care organization. J Am Med Inform Assoc 2009;16(5):683-689. [doi: 10.1197/jamia.M3169] [Medline: 19567790]
5. Ukoha EP, Davis K, Yinger M, et al. Ensuring equitable implementation of telemedicine in perinatal care. Obstet Gynecol 2021 Mar 1;137(3):487-492. [doi: 10.1097/AOG.0000000000004276] [Medline: 33543895]
6. Kuo A, Dang S. Secure messaging in electronic health records and its impact on diabetes clinical outcomes: a systematic review. Telemed J E Health 2016 Sep;22(9):769-777. [doi: 10.1089/tmj.2015.0207] [Medline: 27027337]
7. Kruse CS, Bolton K, Freriks G. The effect of patient portals on quality outcomes and its implications to meaningful use: a systematic review. J Med Internet Res 2015 Feb 10;17(2):e44. [doi: 10.2196/jmir.3171] [Medline: 25669240]
8. Goldzweig CL, Towfigh AA, Paige NM, et al. Systematic Review: Secure Messaging Between Providers and Patients, and Patients' Access to Their Own Medical Record: Evidence on Health Outcomes, Satisfaction, Efficiency and Attitudes: Department of Veterans Affairs (US); 2012.
9. Osborn CY, Mayberry LS, Mulvaney SA, Hess R. Patient web portals to improve diabetes outcomes: a systematic review. Curr Diab Rep 2010 Dec;10(6):422-435. [doi: 10.1007/s11892-010-0151-1] [Medline: 20890688]

10. Schickedanz A, Huang D, Lopez A, et al. Access, interest, and attitudes toward electronic communication for health care among patients in the medical safety net. J Gen Intern Med 2013 Jul;28(7):914-920. [doi: 10.1007/s11606-012-2329-5] [Medline: 23423453]
11. Harris LT, Haneuse SJ, Martin DP, Ralston JD. Diabetes quality of care and outpatient utilization associated with electronic patient-provider messaging: a cross-sectional analysis. Diabetes Care 2009 Jul;32(7):1182-1187. [doi: 10.2337/dc08-1771] [Medline: 19366959]
12. Harris LT, Koepsell TD, Haneuse SJ, Martin DP, Ralston JD. Glycemic control associated with secure patient-provider messaging within a shared electronic medical record: a longitudinal analysis. Diabetes Care 2013 Sep;36(9):2726-2733. [doi: 10.2337/dc12-2003] [Medline: 23628618]
13. Petullo B, Noble B, Dungan KM. Effect of electronic messaging on glucose control and hospital admissions among patients with diabetes. Diabetes Technol Ther 2016 Sep;18(9):555-560. [doi: 10.1089/dia.2016.0105] [Medline: 27398824]
14. Shimada SL, Allison JJ, Rosen AK, Feng H, Houston TK. Sustained use of patient portal features and improvements in diabetes physiological measures. J Med Internet Res 2016 Jul 1;18(7):e179. [doi: 10.2196/jmir.5663] [Medline: 27369696]
15. Ukoha EP, Feinglass J, Yee LM. Disparities in electronic patient portal use in prenatal care: retrospective cohort study. J Med Internet Res 2019 Sep 23;21(9):e14445. [doi: 10.2196/14445] [Medline: 31586367]
16. Holder K, Ukoha E, Feinglass J, Yee LM. Relationship between patient portal utilization and glycemic control outcomes during pregnancy. J Diabetes Sci Technol 2021 Jul;15(4):970-971. [doi: 10.1177/1932296821998742] [Medline: 33736473]
17. American community survey data. United States Census Bureau. 2020. URL: https://www.census.gov/programs-surveys/acs/data.html [accessed 2021-08-21]
18. Francke JA, Groden P, Ferrer C, et al. Remote enrollment into a telehealth-delivering patient portal: barriers faced in an urban population during the COVID-19 pandemic. Health Technol (Berl) 2022;12(1):227-238. [doi: 10.1007/s12553-021-00614-x] [Medline: 34777935]
19. Huang M, Khurana A, Mastorakos G, et al. Patient portal messaging for asynchronous virtual care during the COVID-19 pandemic: retrospective analysis. JMIR Hum Factors 2022 May 5;9(2):e35187. [doi: 10.2196/35187] [Medline: 35171108]
20. Nishii A, Campos-Castillo C, Anthony D. Disparities in patient portal access by US adults before and during the COVID-19 pandemic. JAMIA Open 2022 Dec;5(4):ooac104. [doi: 10.1093/jamiaopen/ooac104] [Medline: 36540762]
21. Singh S, Polavarapu M, Arsene C. Changes in patient portal adoption due to the emergence of COVID-19 pandemic. Inform Health Soc Care 2023 Apr 3;48(2):125-138. [doi: 10.1080/17538157.2022.2070069] [Medline: 35473512]
22. Wade-Vuturo AE, Mayberry LS, Osborn CY. Secure messaging and diabetes management: experiences and perspectives of patient portal users. J Am Med Inform Assoc 2013 May 1;20(3):519-525. [doi: 10.1136/amiajnl-2012-001253] [Medline: 23242764]
23. Ancker JS, Barrón Y, Rockoff ML, et al. Use of an electronic patient portal among disadvantaged populations. J Gen Intern Med 2011 Oct;26(10):1117-1123. [doi: 10.1007/s11606-011-1749-y] [Medline: 21647748]
24. Carrell D, Ralston JD. Variation in adoption rates of a patient web portal with a shared medical record by age, gender, and morbidity level. AMIA Annu Symp Proc 2006;2006:871. [Medline: 17238491]
25. Weingart SN, Rind D, Tofias Z, Sands DZ. Who uses the patient internet portal? The PatientSite experience. J Am Med Inform Assoc 2006 Jan;13(1):91-95. [doi: 10.1197/jamia.M1833] [Medline: 16221943]
26. Kim J, Mathews H, Cortright LM, Zeng X, Newton E. Factors affecting patient portal use among low-income pregnant women: mixed-methods pilot study. JMIR Form Res 2018 Mar 22;2(1):e6. [doi: 10.2196/formative.5322] [Medline: 30684413]
27. Holmgren AJ, Byron ME, Grouse CK, Adler-Milstein J. Association between billing patient portal messages as e-visits and patient messaging volume. JAMA 2023 Jan 24;329(4):339-342. [doi: 10.1001/jama.2022.24710] [Medline: 36607621]

Abbreviations

EMR: electronic medical record
HITECH: Health Information Technology for Economic and Clinical Health
***ICD-10*:** *International Classification of Diseases, Tenth Revision*
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology
ZCTA: zip code tabulation area

Edited by L Buis; submitted 06.08.23; peer-reviewed by A Lewkowitz, R Blair; revised version received 30.11.23; accepted 01.12.23; published 26.04.24.

Please cite as:
Holder K, Feinglass J, Niznik C, Yee LM
Use of Electronic Patient Messaging by Pregnant Patients Receiving Prenatal Care at an Academic Health System: Retrospective Cohort Study
JMIR Mhealth Uhealth 2024;12:e51637
URL: https://mhealth.jmir.org/2024/1/e51637
doi:10.2196/51637

Original Paper

The Association of eHealth Literacy Skills and mHealth Application Use Among US Adults With Obesity: Analysis of Health Information National Trends Survey Data

George Shaw Jr[1], PhD; Bianca A Castro[2], MSc; Laura H Gunn[1], PhD; Keith Norris[3], MD, PhD; Roland J Thorpe Jr[4], PhD

[1]Department of Public Health Sciences, School of Data Science, University of North Carolina at Charlotte, Charlotte, NC, United States

[2]Department of Public Health Sciences, University of North Carolina at Charlotte, Charlotte, NC, United States

[3]The University of California Los Angeles Division of General Internal Medicine and Health Services Research, University of California, Los Angeles, Los Angeles, CA, United States

[4]Department of Health, Behavior and Society, Johns Hopkins University, Baltimore, MD, United States

Corresponding Author:
George Shaw Jr, PhD
Department of Public Health Sciences
School of Data Science
University of North Carolina at Charlotte
9201 University City Blvd.
Charlotte, NC, 28223-0001
United States
Phone: 1 7045620021
Email: gshaw11@charlotte.edu

Abstract

Background: Physical inactivity and a poor diet are modifiable behaviors that contribute to obesity. Obesity is a well-recognized risk factor for chronic diseases, including diabetes. Mobile health (mHealth) apps can play an important adjuvant role in preventing and treating chronic diseases and promoting positive health behavior change among people with obesity, and eHealth literacy skills have the potential to impact mHealth app use.

Objective: The purpose of this study was to explore the associations between the 2 dimensions, access and application, of eHealth literacy skills and mHealth app use among US adults (≥18 years of age) with obesity (BMI ≥30 kg/m^2).

Methods: Data were obtained from February to June 2020 using the Health Information National Trends Survey 5. A total of 1079 respondents met the inclusion criteria of adults with obesity and owners of smartphones. Individual associations between mHealth app use and sociodemographic variables were explored using weighted chi-square and 2-tailed *t* tests. A multivariable weighted logistic regression model was fitted, and adjusted odds ratios (ORs) of using mHealth apps with corresponding 95% CIs were reported across multiple sociodemographic variables. An Ising model-weighted network visualization was produced. A receiver operating characteristic curve was calculated, and the area under the curve was reported with the corresponding Delong 95% CI.

Results: A majority of respondents were female (550/923, 59.6%) or non-Hispanic White (543/923, 58.8%). Individuals in households earning less than US $50,000 comprised 41.4% (382/923) of the sample. All sociodemographic variables were found to be univariately significant at the 5% level, except employment and region. Results from the multivariable weighted logistic regression model showed that the adjusted odds of using an mHealth app are 3.13 (95% CI 1.69-5.80) and 2.99 (95% CI 1.67-5.37) times higher among those with an access eHealth literacy skill of using an electronic device to look for health or medical information for themselves and an application eHealth literacy skill of using electronic communications with a doctor or doctor's office, respectively. Several sociodemographic variables were found to be significant, such as education, where adjusted ORs comparing subgroups to the lowest educational attainment were substantial (ORs ≥7.77). The network visualization demonstrated that all eHealth literacy skills and the mHealth app use variable were positively associated to varying degrees.

Conclusions: This work provides an initial understanding of mHealth app use and eHealth literacy skills among people with obesity, identifying people with obesity subpopulations who are at risk of a digital health divide. Future studies should identify equitable solutions for people with obesity (as well as other groups) and their use of mHealth apps.

(JMIR Mhealth Uhealth 2024;12:e46656) doi:10.2196/46656

KEYWORDS

accessibility; eHealth literacy; mHealth; multivariable logistic regression; obesity; smartphones

Introduction

Overview

Physical inactivity and poor dietary behaviors are modifiable behaviors that contribute to obesity [1,2]. Recent studies show that obesity affects nearly 42% of the US population aged 20 years or older [3], with an associated excess annual estimated medical cost of upwards of US $170 billion in 2019 [4]. Obesity is a well-recognized risk factor for chronic diseases, including diabetes, cardiovascular diseases, and cancer, and a significant cause of premature morbidity and mortality [2,5]. Wang et al [6] demonstrated the importance of reducing the weight of patients classified as obese, which is a major contributor to the increased incidence of type 2 diabetes mellitus. Obesity is a complex and multifaceted disease, extending far beyond the realm of individual behavior or mere lifestyle choices. Genetic, environmental, and socioeconomic factors in the development of obesity [7]. However, mobile health (mHealth) apps can play an important adjuvant role in preventing and treating chronic diseases and promoting positive health behavior change among individuals with obesity [8,9], with several factors influencing users' acceptance of and engagement with mHealth apps [10-12]. Previous intervention studies have used mHealth apps to promote health behavior change among individuals with obesity [13,14]. Participants from the reference studies found that the use of mHealth made the intervention helpful and benefited weight loss when used in conjunction with other weight loss intervention methods. mHealth apps provide access to health information and can extend this access to underserved groups, particularly those at higher risk of chronic diseases [15]. However, many individuals with chronic diseases like obesity fail to engage in mHealth app use [16].

mHealth and eHealth Literacy

Istepanian et al [17] defined mHealth as mobile computing, medical sensors, or communication technologies designed for health care. Recent interventions involving mHealth apps provide evidence of improvement in participants' self-care and disease self-management [18,19]. It is estimated that more than 350,000 mHealth apps are available on the market [20]. Despite the number of digital health interventions targeting weight management, the level to which users actively and regularly engage with those apps entails user engagement early in the intervention design process [21]. While health literacy is a broader concept associated with the ability of individuals to obtain and understand health information to make rational health decisions, eHealth literacy comprises the complex navigation of health care information from internet sources [22,23]. Kontos et al [24] showed that people with lower levels of education were less likely to use the internet to communicate with a doctor or use health information on their mobile devices. Moreover, national and international studies have shown that weight-management apps are beneficial for improving weight loss [25,26]. Previous studies have recommended focusing on understanding the level of health literacy of recipients who may use these apps, particularly those interested in weight loss interventions. Understanding eHealth literacy skills is critical when evaluating health information from mHealth apps and the subsequent application of the knowledge gained [27,28].

eHealth literacy is defined as seeking, finding, understanding, and appraising health information from electronic sources and applying the knowledge acquired to address or solve a health-related problem [29,30]. More recently, eHealth literacy has been conceptualized in dimensions including access to digital services and the application of services and information that satisfy users [31]. People lacking or with a low level of health literacy or eHealth literacy skills benefit less from digital health information and health informatics interventions [32,33], as low eHealth literacy skills were found to impact mHealth app use when mediated through mHealth app efficacy [23]. By contrast, those with higher levels of health literacy and eHealth literacy skills report a positive connection between mHealth app use and health outcomes [34].

mHealth apps are operated across a diverse group of users. This includes individuals with varying incomes, ages, races, ethnicities, and educations [35]. Curating data that include this and other defining personal characteristics require a significant number of resources. Few studies of eHealth literacy and mHealth app use have included a diverse group of participants. Through a national data set of noninstitutionalized adults, this study uses a diverse national data set. The purpose of this study is to explore the potential associations between the 2 dimensions, access and application, of eHealth literacy skills and mHealth app use among a diverse group of US adults (aged ≥18 years) with obesity (BMI ≥30 kg/m^2).

Theoretical Underpinning

Seeking health information has become a typical behavior among people of all ages and health conditions [36]. The information obtained when factoring in a person's eHealth literacy skills, has the potential to influence health outcomes. This is particularly important for people with obesity. People with obesity have lower self-confidence in managing their health. However, people with obesity view communication with their physician as helpful with self-care weight management [37]. Therefore, people with obesity may improve their confidence over time in managing their weight if they have the eHealth literacy skills to seek information from their physician electronically. There have been numerous information behavior theories and models developed to understand how individuals seek and use information [38]. For example, Zare-Farashbandi and Lalazaryan [39] designed their health information acquisition model based on 6 stages of information seeking. The model acknowledges that the information-seeking process can be iterative and that there is a need for a feedback loop in the search process. However, the model does not consider personal or contextual factors affecting information seeking.

Longo's model of health information considers the effects of personal and contextual factors on the information-seeking behavior of patients [40]. Focusing on patients with chronic diseases, the model was significant in depicting the output process of information-seeking for patients [39]. These theories and models have also considered various social contexts and population groups, such as older individuals, patients with cancer, prisoners, and diabetics [38,41-43]. However, this study is informed by Lenz's [41] Information Seeking Model, which is the foundation of many fundamental tenets of recent models and instruments aimed at measuring eHealth literacy skills.

According to this framework, in which information gathering is part of the decision-making process, individuals follow 6 stages to seek health information. First, they receive a stimulus from their previous disease experience or the environment. Second, they establish their informational goals, including sources, time available, and the type of information needed. Third, the person decides whether or not to actively access the information they want. The decision is based on the individual's previous knowledge, background, and the expected cost-benefit of the action. The fourth stage is of particular interest, characterized by the information-seeking action itself. This step is correlated to the eHealth literacy dimension of access, and it could be an in-depth search or superficial information gathering, depending on the person's need and previous attempts. The access dimension corresponds to having the availability of digital services that suit people's needs and work correctly [44]. The fifth stage corresponds to information achievement and interpretation. This step is related to the eHealth literacy dimension of application, in which the individual understands and appraises or applies the information obtained. In this final stage, people may have to decide on the adequacy of the acquired information [39]. Understanding health information seeking through the potential associations of eHealth literacy skills and mHealth app use may provide insights into how population groups with health disparities with chronic conditions such as obesity can access and apply the information they seek [45]. Using this framework, the objective of this study is to explore the potential associations between the 2 dimensions, access and application, of eHealth literacy skills and mHealth app use among a diverse group of US adults with obesity.

Methods

Data

The Health Information National Trends Survey (HINTS) was used to explore the potential association between mHealth app use and eHealth literacy skills. HINTS has been administered every few years by the National Cancer Institute since 2003, and the data sets that have been made publicly available are used for evaluating health information access and use among US adults [24,46,47]. HINTS collects representative data about noninstitutionalized US adults' knowledge, access, attitudes, and use of cancer- and health-related information. The survey uses a 2-stage stratified random sampling that selects households from residential addresses in the United States and then selects 1 adult within each household [24,46].

This study used the HINTS 5, Cycle 4 data set. The data were collected from February to June 2020 and comprised responses from 3865 participants. Despite the COVID-19 impact on society, the response rate for the survey remained high. The response rate (37%) for the survey remained relatively high and was even higher than prepandemic HINTS 5 surveys, which experienced response rates of at most 33% [48]. However, COVID-19 impacted the time frame in which the data are typically collected. Individuals included in this analysis were those who indicated ownership of a tablet, smartphone, or both and self-declared a BMI ≥30 kg/m^2 (obese). The dependent variable was based on respondents' answers to the following item: "In the past 12 months, have you used any of these health or wellness apps?" The binary variable derived was used to indicate those who reported using any health or wellness apps within the past 12 months and those who did not.

The main independent variables representing eHealth literacy skill's access and application dimensions were the following four items pertaining to eHealth information and services, connecting to the common stem of "In the past 12 months, have you used a computer, smartphone, or other electronic means to do any of the following: (1) looked for health or medical information for yourself; (2) used email or the internet to communicate with a doctor or doctor's office; (3) looked medical test results; and (4) made appointments with a health care provider?" Access within this context is the information-searching behavior involved in accessing information. Application within this context is defined as the interpretation and appraisal of information aimed at completing an action. Additional covariates extracted include age in years, health insurance status, sex at birth, employment status, marital status, education, annual household income (in ranges), race and ethnicity, and US Census region. These variables have been used in previous studies to evaluate mHealth app use or can be relevant confounders regarding the associations between eHealth literacy skills and mHealth app use [49,50].

Due to low counts, the following categories were combined: employment status of unemployed across lengths of unemployment; employment status of students and others; marital status of separated and divorced; marital status of married and those living as married or with a romantic partner; education categories below 11 years of education; and race and ethnicity categories of non-Hispanic Native Hawaiian or Other Pacific Islander and American Indian or Alaska Native.

Statistical Analysis

In order to investigate the research objective and hypothesis, a comprehensive statistical analysis was performed on the collected data using univariate and multiple logistic regression modeling. The weights provided by HINTS were used to perform all analyses and adjust for sampling biases [51]. A weighted complete case analysis was performed on the data. Characteristics were summarized using means, SDs, counts, and percentages as appropriate. Weighted chi-square and 2-tailed *t* tests were used to explore univariate associations between each of the covariates and mHealth app use, with test statistics and corresponding *P* values tabulated. Visualizations were created to explore associations, including (1) Ising model

network weighted analysis of the associations between the main independent variables (eHealth literacy skill's access and application covariates) and the outcome; (2) weighted box plot for the continuous covariate (age) and the dependent variable; and (3) multiple weighted 100% stacked bar charts across the main independent variables and the dependent variable. Additional weighted 100% stacked bar charts were constructed (Multimedia Appendix 1) to visualize the sociodemographic variables and the outcome.

The primary study aim is to assess associations between mHealth app use (binary outcome) and each of the eHealth literacy skills dimensions of access and application (main covariates). Univariate analysis is included to provide a comprehensive description of the individual variables in the study and establish a foundation for more complex multivariable analyses. These were further examined using a multivariable weighted logistic regression adjusted for the aforementioned sociodemographic factors. Adjusted odds ratios (ORs), corresponding 95% CIs, and *P* values were reported across eHealth literacy skills dimensions and sociodemographic variables. Results were tabulated and highlighted using a significance level of 5%. A pseudo-R^2 was calculated. The receiver operating characteristic (ROC) is a common approach used to measure the sensitivity versus specificity of logistic models. Additionally, the area under the curve (AUC) is a single metric for that trade-off, with AUC=1 meaning that the model perfectly fits the data and AUC=0.5 indicating there is a split chance that the model fits the data. Both of these approaches are used to evaluate the performance of logistic models. The ROC curve was estimated, and the corresponding AUC value and Delong 95% CI were reported. R software (version 4.0.3; R Foundation for Statistical Computing) was used for statistical analyses.

Ethical Considerations

This research was approved by the institutional review board (IRB) of the University of North Carolina at Charlotte (study #IRB-22-0585). This data set consisted of deidentified, aggregated data. The IRB approval process did not require additional consent from the respondents representing the data.

Results

A total of 1079 participants were identified as obese and owners of a smartphone, tablet, or both. Fewer than 15% (156/1079) of the responses were removed due to incomplete or incoherent data, resulting in 923 complete observations, with mHealth app use (dependent variable), eHealth literacy skills dimensions (main independent variable), and additional covariates summarized in Table 1. The average age was 53.51 (SD 14.91) years, and most participants were female (550/923, 59.6%) and non-Hispanic White (543/923, 58.8%). A college degree or above was the highest level of education for 43.3% (400/923) of study participants, and they were mainly employed (with a single or multiple employer; 594/923, 64.4%) and covered by health insurance (872/923, 94.5%). The South contained the largest percentage of participants (436/923, 46.7%), which also corresponds to the nation's most populous region [52]. Individuals in households earning less than US $50,000 comprised 41.4% (382/923) of the sample, and 18.9% (174/923) of participants had an annual household income that fell within the range, containing the median annual household income in the United States of US $67,521 in 2020 [53].

The majority of participants (482/923, 52.2%) did not use mHealth apps, resulting in a balanced outcome variable. Within the eHealth literacy skills access dimension, 77.5% (715/923) of respondents used an electronic device to look for health or medical information for themselves within the past 12 months, and approximately half (468/923, 50.7%) used electronic means to look up medical test results, also within the past 12 months. Within the eHealth literacy skills application dimension, 55.6% (513/923) of respondents used email or the internet to communicate with a doctor or doctor's office within the past 12 months, and 53.3% (429/923) made an appointment with a health care provider through electronic means in that same time period. We also examined the univariate association between mHealth app use, covariates, and main independent covariates.

Table 2 summarizes results from weighted chi-square and *t* tests for univariate associations between mHealth app use and each of the covariates. Most covariates and all the main covariates were found to be significant at the 5% level. All eHealth literacy skills dimensions were found to be significantly associated with mHealth app use based on univariate weighted chi-square tests ($P<.001$). Similarly, all demographic factors were found to be significantly associated with mHealth app use except for employment status ($P=.20$) and Census region ($P=.16$). Figure 1 displays pairwise weighted 100% stacked bar charts for each of the eHealth literacy skills dimensions versus mHealth app use. Figure 2 portrays a joint network representation of the weighted associations between the eHealth literacy skills dimensions and mHealth app use, which demonstrate strong positive associations both between the skills dimensions as well as between those and the outcome (mHealth app use). Figures S1-S9 in Multimedia Appendix 1 include a weighted box plot (age) and weighted 100% stacked bar charts visualizing the univariate associations with mHealth app use.

Table 3 presents the results of the multivariable weighted logistic regression model. The adjusted odds of using an mHealth app are 3.13 (95% CI 1.69-5.80) times higher among those who responded with an access eHealth literacy skill of using an electronic device to look for health or medical information for themselves within the past 12 months. Similarly, those with an application eHealth literacy skill of using email or the internet to communicate with a doctor or doctor's office within the past 12 months experience 2.99 (95% CI 1.67-5.37) times higher odds of using an mHealth app compared to those without this skill. Sociodemographic factors found to be significantly associated with mHealth app use include age, disabled or retired status, single or never married or widowed, education, and Hispanic ethnicity. Each additional year of age is associated with 4% lower odds of using mHealth apps (OR 0.96, 95% CI 0.94-0.98). Disabled and retired participants experienced 4.21 (95% CI 1.28-13.82) and 2.53 (95% CI 1.14-5.60) higher odds, respectively, of using mHealth apps compared to those who were employed. Single or never married and widowed participants experienced 49% and 81% lower odds of mHealth app use, respectively, than those who are married, living as married, or living with romantic partners. Previous work has

indicated that surrogate seekers, those who may seek health information on behalf of others, were more likely to be married or have someone close to them with a chronic illness [54]. Those who received more than 11 years of formal education experienced higher odds of mHealth app use than those with 11 years or less, with OR estimates ranging from 7.77 to 17.24, though with substantially wide CIs. Hispanic participants experienced higher odds of using mHealth apps than non-Hispanic White participants (OR 2.61, 95% CI 1.28-5.33). Insurance status, sex at birth, annual household income, and Census region were not found statistically significant upon adjusting for the other covariates, though there is some level of collinearity present among sociodemographic covariates, as demonstrated in the univariate significance of some of these variables. The multivariable weighted logistic regression adjusted for sociodemographic characteristics showed relatively strong explanatory power with a pseudo-R^2 of 0.32 and AUC of 0.7957 (95% CI 0.7671-0.8243). The corresponding ROC is included in Multimedia Appendix 1.

Table 1. Unweighted characteristics of study participants (n=923) using the 2020 Health Information National Trends Survey data set.

Sociodemographic variables	Values
Age (years), mean (SD)	53.51 (14.91)
Health insurance, n (%)	
Insured	872 (94.5)
Uninsured	51 (5.5)
Sex at birth, n (%)	
Female	550 (59.6)
Male	373 (40.4)
Employment status, n (%)	
Disabled	67 (7.3)
Employed	494 (53.5)
Homemaker	25 (2.7)
Multiple	100 (10.8)
Retired	183 (19.8)
Unemployed	42 (4.6)
Other	12 (1.3)
Marital status, n (%)	
Married or living as married or with a romantic partner	518 (56.1)
Separated or divorced	180 (19.5)
Single or never married	159 (17.2)
Widowed	66 (7.2)
Education, n (%)	
≤11 years	49 (5.3)
12 years or completed high school	174 (18.9)
Post–high school training other than college (vocational or technical)	68 (7.4)
Some college	232 (25.1)
College graduate	241 (26.1)
Postgraduate	159 (17.2)
Annual household income (US $), n (%)	
0-9999	51 (5.5)
10,000-14,999	49 (5.3)
15,000-19,999	37 (4)
20,000-34,999	113 (12.2)
35,000-49,999	132 (14.3)
50,000-74,999	174 (18.9)
75,000-99,999	128 (13.9)
100,000-199,999	198 (21.5)
≥200,000	41 (4.4)
Race and ethnicity, n (%)	
Black or African American	148 (16)
Hispanic	171 (18.5)
Non-Hispanic Asian	18 (2)
Non-Hispanic Native Hawaiian or Other Pacific Islander or American Indian or Alaska Native	9 (1)

Sociodemographic variables	Values
Non-Hispanic White	543 (58.8)
Non-Hispanic multiple races	34 (3.7)
Census region, n (%)	
Midwest	155 (16.8)
Northeast	141 (15.3)
South	431 (46.7)
West	196 (21.2)
Outcome variable, n (%)	
mHealth app use	
No	482 (52.2)
Yes	441 (47.8)
Main covariates, n (%)	
eHealth literacy skills access dimension	
Electronic health information for self	
No	208 (22.5)
Yes	715 (77.5)
Electronic test results	
No	455 (49.3)
Yes	468 (50.7)
eHealth literacy skills application dimension	
Electronic communication with doctor or doctor's office	
No	410 (44.4)
Yes	513 (55.6)
Made provider appointments electronically	
No	431 (46.7)
Yes	492 (53.3)

Table 2. Weighted chi-square and 2-tailed *t* tests (test statistics and *P* values) for univariate associations between mHealth app use (dependent variable) and each of the covariates.

Variable	Chi-square (*df*)	*P* value
Sociodemographic factors		
Age (years)	3.77 (921)[a]	<.001
Health insurance	6.28 (1)	.01
Sex at birth	6.25 (1)	.01
Employment status	8.54 (6)	.20
Marital status	18.55 (3)	<.001
Education	78.02 (5)	<.001
Annual household income	34.38 (8)	<.001
Race and ethnicity	11.31 (5)	.046
Census region	5.18 (3)	.16
Main covariates		
eHealth literacy skills: access dimension		
eHealth information for self	95.60 (1)	<.001
Electronic test results	97.48 (1)	<.001
eHealth literacy skills: application dimension		
Electronic communication with doctor or doctor's office	127.87 (1)	<.001
Made provider appointments electronically	81.48 (1)	<.001

[a]*t* test was used for the univariate analysis.

Figure 1. Visualization of weighted 100% stacked bar charts for each of the eHealth literacy skills dimensions (main covariates) against mHealth app use (outcome). A: Electronic health information for self; B: Electronic test results; C: Electronic communication with doctor or doctor's office; D: Made appointments electronically.

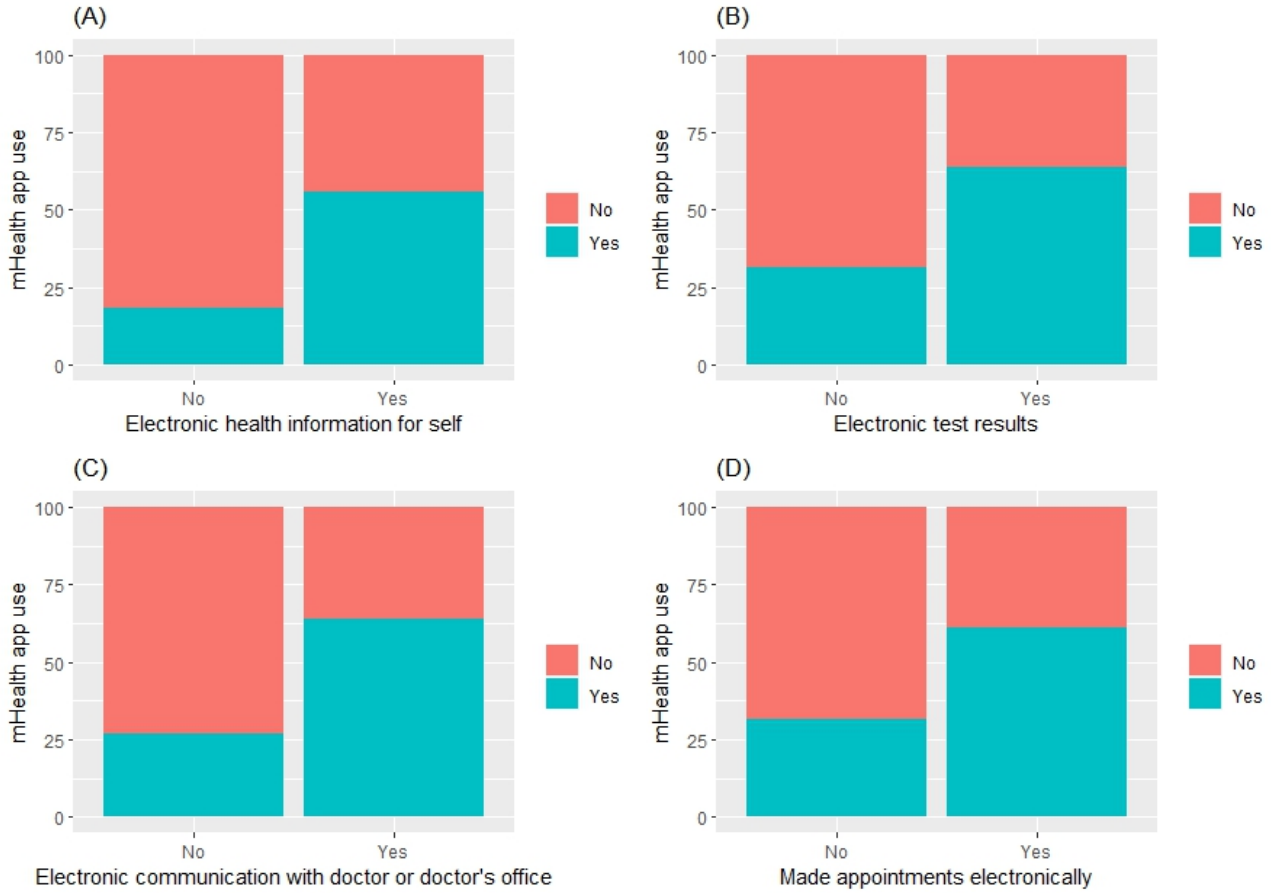


Figure 2. Ising model-weighted network visualization of eLASSO associations (unadjusted by other covariates and with 0.25 penalization factor) between the eHealth literacy skills dimensions (SHI: self-health information; TD: talk to a doctor or doctor's office; TR: test results; MA: made appointments; and mH: mHealth app use). Thicker edges (lines) between nodes (circles) represent stronger associations.

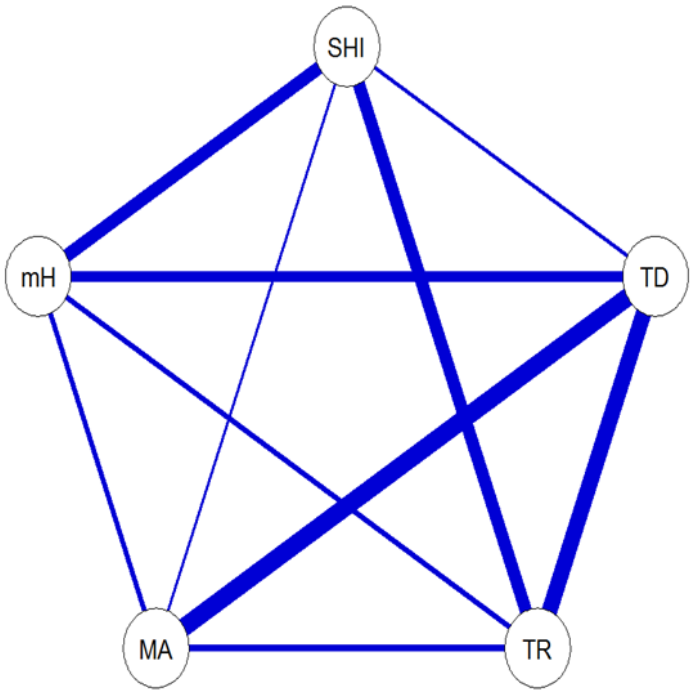


Table 3. Adjusted odds ratios (ORs), corresponding 95% CIs, and *P* values for the multivariable weighted logistic regression model assessing mHealth app use (n=923). The regression model included adjustments for eHealth literacy, age, insurance status, sex, employment and marital status, education, income, race and ethnicity, and census region.

Characteristics	OR (95% CI)	*P* value
Explanatory demographic variables		
Age (years)	0.96 (0.94-0.98)	<.001
Insured	2.25 (0.94-5.38)	.07
Male	0.75 (0.46-1.20)	.22
Employment status		
Employed (reference)	N/A[a]	N/A
Disabled	4.21 (1.28-13.82)	.02
Homemaker	2.10 (0.66-6.70)	.21
Multiple	2.16 (0.92-5.09)	.08
Retired	2.53 (1.14-5.60)	.02
Unemployed	1.12 (0.44-2.87)	.81
Other	0.21 (0.04-1.12)	.07
Marital status		
Married or living as married or with a romantic partner (reference)	N/A	N/A
Separated or divorced	0.67 (0.34-1.33)	.25
Single or never married	0.51 (0.27-0.96)	.04
Widowed	0.19 (0.06-0.57)	.003
Education		
≤11 years (reference)	N/A	N/A
12 years or completed high school	7.77 (2.08-29.01)	.002
Post–high school training other than college	12.75 (3.18-51.17)	<.001
Some college	9.25 (2.60-32.98)	<.001
College graduate	14.01 (3.68-53.26)	<.001
Postgraduate	17.24 (4.09-72.64)	<.001
Annual household income (US $)		
<10,000 (reference)	N/A	N/A
10,000-14,999	1.67 (0.38-7.45)	.50
15,000-19,999	0.81 (0.20-3.18)	.76
20,000-34,999	1.31 (0.41-4.26)	.65
35,000-49,999	2.47 (0.73-8.37)	.15
50,000-74,999	2.27 (0.71-7.27)	.17
75,000-99,999	3.16 (0.90-11.04)	.07
100,000-199,999	2.47 (0.72-8.40)	.15
≥200,000	1.81 (0.37-8.93)	.47
Race and ethnicity		
Black or African American	1.05 (0.51-2.15)	.90
Hispanic	2.61 (1.28-5.33)	.008
Non-Hispanic Asian	0.30 (0.06-1.50)	.14
Non-Hispanic Native Hawaiian or Other Pacific Islander or American Indian or Alaska Native	1.64 (0.17-16.19)	.67
Non-Hispanic White (reference)	N/A	N/A

Characteristics	OR (95% CI)	*P* value
Non-Hispanic multiple races	1.21 (0.34-4.22)	.77
Census region		
South (reference)	N/A	N/A
Midwest	1.42 (0.73-2.75)	.30
Northeast	0.89 (0.47-1.69)	.72
West	1.09 (0.57-2.08)	.80
Intercept	0.03 (0.00-0.19)	<.001
eHealth literacy skills: access dimension		
Electronic health information for self (reference: yes)	3.13 (1.69-5.80)	<.001
Electronic test results (reference: yes)	1.55 (0.87-2.73)	.13
eHealth literacy skills: application dimension		
Electronic communication with a doctor or doctor's office (reference: yes)	2.99 (1.67-5.37)	<.001
Made appointments electronically (reference: yes)	1.53 (0.91-2.58)	.11

[a]N/A: not applicable.

Discussion

Overview

The purpose of this study was to explore the associations between the 2 dimensions, access and application, of eHealth literacy skills and mHealth app use among US adults with obesity. We used the HINTS 2020 data to explore this potential association with a sample of 923 respondents with complete information represented in the data set. We found that the majority of the respondents had health insurance, were female, and were non-Hispanic White, with an average age of 54 years. Also, more than half of the respondents had some level of college or were college graduates. This study highlights the association between eHealth literacy skills for accessing and the application of health information using mHealth apps among people with obesity.

The weighted univariate analyses demonstrated associations between all of the covariates and mHealth app use except employment status and census region. Socioeconomic factors of education and income have been found to be important in the general use of content within digital environments (ie, internet) [44]. However, more specifically to this study, these factors are important in showing the potential relationship they have with mHealth app use among people with obesity. When considering weight management or physical activity interventions using mHealth apps, future interventions should attempt to improve the eHealth literacy of participants by targeting segments of people with obesity identified to be more at risk, such as older individuals with obesity or those in lower income brackets. These initial metrics can be collected through a variety of eHealth literacy assessment tools. The eHealth Literacy Scale, for example, has been studied in diverse languages and populations, and it was designed to convey an estimate of people's eHealth-related skills. Other instruments, such as the eHealth Literacy Questionnaire, were established to support researchers, designers, and the government in evaluating, developing, and applying effective digital health interventions [54]. Previous research identified that patients with adequate eHealth literacy had more ability to seek health information on the internet and find reliable and high-quality information than patients with inadequate eHealth literacy [55,56].

Accessing health information requires active information-seeking skills. Additionally, context and behaviors to gain information are intertwined in this process. Respondents' access to web-based health information (seeking health information for themselves) can be informed by the Lenz search behavior stage. The respondents' access to information requires a search for information from impersonally related sources. There is no indication of the familiarity that respondents have with these resources based on the survey questions. People with obesity who use mHealth devices may exhibit multiple factors in their search for health information, and recent eHealth literacy work supports the nuances involved in seeking health information [57,58].

The information acquired through the information-seeking process impacts an information seeker's decision-making process. Electronic communication with the doctor's office can be the result of gathering enough information to move forward based on their original goal or a stop in the information acquisition process that prompts information seeking through a personal connection through digital communication. The results from this work are not intended to model these variables but demonstrate that Lenz's model, enhanced with recent theories, may help inform studies aimed at understanding active information seeking at the intersection of digital health devices such as mHealth and eHealth literacy skills. Recent models and theories commonly demonstrate that health information–seeking behavior involves the action of seeking out information, irrespective of how or why it is sought [36].

Consistent with Mahmood et al [59], education and age are important sociodemographic factors associated with mHealth app use among people with obesity. As access to health services

increases through the use of telehealth technology embedded within mHealth apps, it is imperative that this population group be able to benefit from this type of health service [60]. People with lower education levels and older individuals experience more limited eHealth literacy skills and lower mHealth app use, further widening the digital divide gap [18,24]. Additionally, when we examine other sociodemographic factors, marginalized ethnic groups such as Hispanic populations may have access to mHealth apps but experience digital divide issues [30,61].

Issues such as use and knowledge as they relate to using mHealth apps can also contribute to the digital divide [62]. Additional attention is needed to focus on these vulnerable populations. Interventions that can attempt to address this issue are the development of apps and health promotion campaigns that are designed to be culturally relevant [63]. Within the realm of health promotion and wellness, mHealth mindfulness approaches have been used for African American populations [64]. Moreover, studies have described the importance of mHealth interventions with phone features that are familiar to the target population group [62]. There should be consideration of acceptability and efficacy during the developmental phases to support the use of mHealth apps. When considering efficacy, simpler solutions in app design and use should be evaluated. For the older population, features such as 1-click access to a dashboard within health apps that are appropriately displayed in size may be appropriate. Future mHealth apps should also consider health education–related features to support users with low eHealth literacy skills [30,65]. The recent COVID-19 pandemic highlighted the continued digital divide and the disparity in health care services for those who lack sufficient digital literacy skills [66]. mHealth apps benefit people with various chronic conditions, including obesity. People with obesity are less likely to benefit from these mHealth apps if they have low eHealth literacy skills.

mHealth apps facilitate access to health information that has increasingly migrated to web-based spaces [67]. More importantly, mHealth devices assist individuals with seeking health information and decision-making regarding their health [68]. mHealth apps are also advantageous to improve access to health information for personal health data management [15,69]. Since we found that the eHealth literacy skills dimension of access for people with obesity is associated with higher odds of using mHealth apps for seeking health information for themselves, health services should reconsider how they disseminate health information to reach higher proportions of the population. Inevitably, the accessibility of web-based health information has changed the way people engage in health decision-making [70]. This is also evident from our network analysis results, which demonstrate the interconnectivity among all elements relating to eHealth literacy skills and mHealth app use among people with obesity, resulting in the need for holistic solutions to enhance mHealth app use and access to health information. Lenz's model primarily focuses on the search process and use of the information; however, future studies should consider the nuanced contextual factors for people with obesity and their use of mHealth-related devices.

Accessing health information through mHealth apps streamlines the application of health information for decision-making. Many people with obesity have additional chronic diseases that can benefit from timely communication with their health care provider [71]. Effective communication is important for reported satisfaction and perceived health management outcomes. Face-to-face communication has been the standard for communication among patients and health care providers. However, there are mixed results on the perceived effectiveness of face-to-face communication versus IT-aided communication such as mHealth devices [72]. Recent studies have found that mHealth apps are viewed as useful by patients for improving communication and the accessibility of health data [73]. Therefore, this constant communication creates the potential for a bidirectional channel of communication among people with obesity and their health care providers. An in-depth content analysis of vaccination apps showed that few apps provide the capability for bidirectional communication among users and health care providers [74]. The challenges of bidirectional communication can be attributed to barriers to data integration. Given the numerous mHealth apps available for download, this creates interoperability challenges for electronic health care record systems [75]. For mHealth apps that are designed to improve physician and consumer communication, transdisciplinary scholarship is necessary to overcome these barriers. More importantly, technical and networking policies must be developed to support and incentivize the ability to improve this type of communication.

This study benefits from the use of a nationally representative sample of noninstitutionalized US adults. This study provides an adjusted analysis of the associations between mHealth app use and eHealth literacy skills among people with obesity. New technologies that require eHealth literacy skills are transforming how we receive health care and access health information, but they also highlight new disparities as they relate to digital health services [30]. However, to address the rise of chronic conditions such as obesity, it is essential to empower patients to engage in their own health management. One promising strategy is using mHealth apps as a complementary tool to manage weight loss and track physical activity [26]. We provide evidence of several significant factors that can be informative when designing inclusive mHealth app-based health intervention studies. Our results also have implications for studies aimed at managing weight loss or tracking the physical activity of people with obesity to assist with mHealth app development and uptake.

Concerning limitations, first, there could be additional confounding variables that are not included in the study, which is limited by the survey design questionnaire. Some of these confounding variables may be related to self-care behaviors or use patterns with mHealth apps [25]. Furthermore, a bias in the survey design includes the assumption that apps are used only on tablets or smartphones, such that only individuals who indicated having a tablet or smartphone were asked within the survey about having or using health or wellness apps. Second, respondents were only asked about access to information within the previous 12 months. There is a possibility that users do not access or seek health information between visits with their doctors on a yearly basis. Nevertheless, many patients with low health literacy are often left dissatisfied and unsure of the information shared by their doctor and seek third-party sources

such as web-based health communities to fill those gaps [76,77]. Also, respondents who report ownership of a tablet, smartphone, or both may also use a computer, but they did not indicate that as part of their response. Third, results from the Ising model visualization show a strong relationship between respondents seeking health information for themselves and mHealth app use. As a result of seeking health information, this may also explain the strong relationship between mHealth app use and talking with one's doctor. Therefore, these correlations may exist because they are measuring the same events. Also, some sociodemographic variables used in this study are correlated (eg, age and retired status), so some multicollinearity may be present. Fourth, with a small sample of uninsured people represented in our sample, the statistical significance for health insurance in our model may have been different with a larger sample of uninsured people. Additionally, the data did not provide a distinction between private and public health insurance, though the information content of such a factor may already be embedded in the income variable. A study aimed at self-monitoring of diet, physical activity, and weight among patients who were underinsured or uninsured demonstrated higher adherence through the use of 2 mHealth-related apps in comparison to a paper group [6]. Also, the sample in this study covers the COVID-19 peak period in early 2020, which may have represented a crest (and potentially a permanent shift in behaviors) in electronic access to health information among people with obesity compared to previous time periods. Lastly, there is a limitation in the HINTS survey questions as they were not designed using a web-based health information–seeking behavior framework, though we were able to detect relevant associations even with this design limitation.

Future directions of this work should consider this model structure for people without obesity. A comparative analysis may identify whether eHealth literacy relevance differs between individuals with obesity and those without obesity. This work also considered mHealth app use, but it did not examine cognitive motivational factors for mHealth use. The identification of motivational barriers and facilitators can be analyzed within the context of psychological motivation frameworks to identify potential intervention targets to leverage in mHealth intervention–based studies. Also, since the COVID-19 pandemic may have brought behavioral changes in the overall population regarding eHealth literacy, a dynamic study that explores those changes over time could highlight whether segments of people with obesity may now experience heightened needs compared to prepandemic stages.

Conclusion

This study estimated the associations between mHealth app use and eHealth literacy skills. Our findings are consistent with previous literature, showing that eHealth literacy skills are associated with accessing digital health information and the application of digital health services. For example, age is negatively associated with mHealth app use among people with obesity, with other sociodemographic factors also showing strong associations. This highlights substantial uneven access to eHealth information among people with obesity, potentially leading to disparities in health outcomes among sociodemographic groups. It is imperative that this phenomenon be further investigated as digital health–related services that involve the use of mHealth apps become more integrated into health care services and aim to reach wider segments of the population. A continued challenge is to engage people with chronic conditions such as obesity to use mHealth apps, especially older individuals with obesity and those with lower educational backgrounds. Our work provides evidence of factors associated with mHealth app use in relation to access and application. This work provides an initial understanding of mHealth app use and eHealth literacy skills among people with obesity, and future studies should identify equitable solutions for people with obesity (as well as other groups) and their use of mHealth apps.

Acknowledgments

This project was funded by the National Institute of Health National's Heart, Lung, and Blood Institute (grant R25HL126145). Dr Thorpe was funded by U54MD000214.

Authors' Contributions

The review concept was designed by GSJ with support from KN and RJTJ. The literature search, screening, analysis, data collection, and data cleaning were conducted by GSJ with support from BDC. LHG provided the data analysis. GSJ, LHG, and BDC wrote the original draft, with review and edits by GSJ, LHG, BDC, RJTJ, and KN. The figures were designed by LHG. Additional review was carried out by GSJ, LHG, RJTJ, and KN.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary figures and table.
[DOCX File , 212 KB - mhealth_v12i1e46656_app1.docx]

References

1. Hill JO, Wyatt HR, Peters JC. Energy balance and obesity. Circulation 2012;126(1):126-132 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.111.087213] [Medline: 22753534]
2. Jehan S, Zizi F, Pandi-Perumal SR, McFarlane SI, Jean-Louis G, Myers AK. Energy imbalance: obesity, associated comorbidities, prevention, management and public health implications. Adv Obes Weight Manag Control 2020;10(5):146-161 [FREE Full text] [Medline: 33305001]
3. Bryan S, Afful J, Carroll MD, Chen TC, Davy O, Fink S, et al. National Health and Nutrition Examination Survey 2017–March 2020 prepandemic data files development of files and prevalence estimates for selected health outcomes. NHSR no. 158. National Health Statistics Reports. 2021. URL: https://stacks.cdc.gov/view/cdc/106273 [accessed 2023-10-27]
4. Ward ZJ, Bleich SN, Long MW, Gortmaker SL. Association of body mass index with health care expenditures in the United States by age and sex. PLoS One 2021;16(3):e0247307 [FREE Full text] [doi: 10.1371/journal.pone.0247307] [Medline: 33760880]
5. Flegal KM, Kit BK, Orpana H, Graubard BI. Association of all-cause mortality with overweight and obesity using standard body mass index categories: a systematic review and meta-analysis. JAMA 2013;309(1):71-82 [FREE Full text] [doi: 10.1001/jama.2012.113905] [Medline: 23280227]
6. Wang J, Cai C, Padhye N, Orlander P, Zare M. A behavioral lifestyle intervention enhanced with multiple-behavior self-monitoring using mobile and connected tools for underserved individuals with type 2 diabetes and comorbid overweight or obesity: pilot comparative effectiveness trial. JMIR Mhealth Uhealth 2018;6(4):e92 [FREE Full text] [doi: 10.2196/mhealth.4478] [Medline: 29636320]
7. De Lorenzo A, Gratteri S, Gualtieri P, Cammarano A, Bertucci P, Di Renzo L. Why primary obesity is a disease? J Transl Med 2019;17(1):169 [FREE Full text] [doi: 10.1186/s12967-019-1919-y] [Medline: 31118060]
8. Agnihothri S, Cui L, Delasay M, Rajan B. The value of mHealth for managing chronic conditions. Health Care Manag Sci 2020;23(2):185-202. [doi: 10.1007/s10729-018-9458-2] [Medline: 30382448]
9. Han M, Rhee SY. Effectiveness of mobile health applications for 5% body weight reduction in obese and overweight adults. J Obes Metab Syndr 2021;30(4):354-364 [FREE Full text] [doi: 10.7570/jomes21014] [Medline: 34853189]
10. Salgado T, Tavares J, Oliveira T. Drivers of mobile health acceptance and use from the patient perspective: survey study and quantitative model development. JMIR Mhealth Uhealth 2020;8(7):e17588 [FREE Full text] [doi: 10.2196/17588] [Medline: 32673249]
11. Mustafa AS, Ali N, Dhillon JS, Alkawsi G, Baashar Y. User engagement and abandonment of mHealth: a cross-sectional survey. Healthcare (Basel) 2022;10(2):221 [FREE Full text] [doi: 10.3390/healthcare10020221] [Medline: 35206837]
12. Amagai S, Pila S, Kaat AJ, Nowinski CJ, Gershon RC. Challenges in participant engagement and retention using mobile health apps: literature review. J Med Internet Res 2022;24(4):e35120 [FREE Full text] [doi: 10.2196/35120] [Medline: 35471414]
13. Okorodudu DE, Bosworth HB, Corsino L. Innovative interventions to promote behavioral change in overweight or obese individuals: a review of the literature. Ann Med 2015;47(3):179-185 [FREE Full text] [doi: 10.3109/07853890.2014.931102] [Medline: 25011006]
14. Cueto V, Wang CJ, Sanders LM. Impact of a mobile app-based health coaching and behavior change program on participant engagement and weight status of overweight and obese children: retrospective cohort study. JMIR Mhealth Uhealth 2019;7(11):e14458 [FREE Full text] [doi: 10.2196/14458] [Medline: 31730041]
15. Bennett GG, Steinberg D, Askew S, Levine E, Foley P, Batch BC, et al. Effectiveness of an app and provider counseling for obesity treatment in primary care. Am J Prev Med 2018;55(6):777-786 [FREE Full text] [doi: 10.1016/j.amepre.2018.07.005] [Medline: 30361140]
16. Meyerowitz-Katz G, Ravi S, Arnolda L, Feng X, Maberly G, Astell-Burt T. Rates of attrition and dropout in app-based interventions for chronic disease: systematic review and meta-analysis. J Med Internet Res 2020;22(9):e20283 [FREE Full text] [doi: 10.2196/20283] [Medline: 32990635]
17. Istepanian R, Jovanov E, Zhang YT. Introduction to the special section on m-health: beyond seamless mobility and global wireless health-care connectivity. IEEE Trans Inf Technol Biomed 2004;8(4):405-414. [doi: 10.1109/titb.2004.840019] [Medline: 15615031]
18. Delva S, Mendez KJW, Cajita M, Koirala B, Shan R, Wongvibulsin S, et al. Efficacy of mobile health for self-management of cardiometabolic risk factors: a theory-guided systematic review. J Cardiovasc Nurs 2021;36(1):34-55 [FREE Full text] [doi: 10.1097/JCN.0000000000000659] [Medline: 32040072]
19. Davidson TM, McGillicuddy J, Mueller M, Brunner-Jackson B, Favella A, Anderson A, et al. Evaluation of an mHealth medication regimen self-management program for African American and Hispanic uncontrolled hypertensives. J Pers Med 2015;5(4):389-405 [FREE Full text] [doi: 10.3390/jpm5040389] [Medline: 26593951]
20. Apps can help keep older folks healthy—but most don't use them. Michigan Medicine-University of Michigan. URL: https://medicine.umich.edu/dept/dgpm/news/archive/202203/apps-can-help-keep-older-folks-healthy-most-dont-use-them [accessed 2023-10-27]
21. Sharpe EE, Karasouli E, Meyer C. Examining factors of engagement with digital interventions for weight management: rapid review. JMIR Res Protoc 2017;6(10):e205 [FREE Full text] [doi: 10.2196/resprot.6059] [Medline: 29061557]

22. Monkman H, Kushniruk AW, Barnett J, Borycki EM, Greiner LE, Sheets D. Are health literacy and eHealth literacy the same or different? Stud Health Technol Inform 2017;245:178-182. [Medline: 29295077]
23. Neter E, Brainin E. eHealth literacy: extending the digital divide to the realm of health information. J Med Internet Res 2012;14(1):e19 [FREE Full text] [doi: 10.2196/jmir.1619] [Medline: 22357448]
24. Kontos E, Blake KD, Chou WYS, Prestin A. Predictors of eHealth usage: insights on the digital divide from the Health Information National Trends Survey 2012. J Med Internet Res 2014;16(7):e172 [FREE Full text] [doi: 10.2196/jmir.3117] [Medline: 25048379]
25. Aljuraiban GS. Use of weight-management mobile phone apps in Saudi Arabia: a web-based survey. JMIR Mhealth Uhealth 2019;7(2):e12692 [FREE Full text] [doi: 10.2196/12692] [Medline: 30794205]
26. Wang Y, Min J, Khuri J, Xue H, Xie B, Kaminsky LA, et al. Effectiveness of mobile health interventions on diabetes and obesity treatment and management: systematic review of systematic reviews. JMIR Mhealth Uhealth 2020;8(4):e15400 [FREE Full text] [doi: 10.2196/15400] [Medline: 32343253]
27. Kreps GL. The relevance of health literacy to mHealth. Inf Serv Use 2017;37(2):123-130 [FREE Full text] [doi: 10.3233/ISU-170828]
28. El Benny M, Kabakian-Khasholian T, El-Jardali F, Bardus M. Application of the eHealth literacy model in digital health interventions: scoping review. J Med Internet Res 2021;23(6):e23473 [FREE Full text] [doi: 10.2196/23473] [Medline: 34081023]
29. Norman CD, Skinner HA. eHealth literacy: essential skills for consumer health in a networked world. J Med Internet Res 2006;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
30. Smith B, Magnani JW. New technologies, new disparities: the intersection of electronic health and digital health literacy. Int J Cardiol 2019;292:280-282 [FREE Full text] [doi: 10.1016/j.ijcard.2019.05.066] [Medline: 31171391]
31. Kayser L, Karnoe A, Furstrand D, Batterham R, Christensen KB, Elsworth G, et al. A multidimensional tool based on the eHealth literacy framework: development and initial validity testing of the eHealth Literacy Questionnaire (eHLQ). J Med Internet Res 2018;20(2):e36 [FREE Full text] [doi: 10.2196/jmir.8371] [Medline: 29434011]
32. Veinot TC, Mitchell H, Ancker JS. Good intentions are not enough: how informatics interventions can worsen inequality. J Am Med Inform Assoc 2018;25(8):1080-1088 [FREE Full text] [doi: 10.1093/jamia/ocy052] [Medline: 29788380]
33. Nelson LA, Pennings JS, Sommer EC, Popescu F, Barkin SL. A 3-item measure of digital health care literacy: development and validation study. JMIR Form Res 2022;6(4):e36043 [FREE Full text] [doi: 10.2196/36043] [Medline: 35486413]
34. Bol N, Helberger N, Weert JCM. Differences in mobile health app use: a source of new digital inequalities? Inf Soc 2018;34(3):183-193 [FREE Full text] [doi: 10.1080/01972243.2018.1438550]
35. Servaes S, Choudhury P, Parikh AK. What is diversity? Pediatr Radiol 2022;52(9):1708-1710 [FREE Full text] [doi: 10.1007/s00247-022-05356-0] [Medline: 35348810]
36. Mirzaei A, Aslani P, Luca EJ, Schneider CR. Predictors of health information-seeking behavior: systematic literature review and network analysis. J Med Internet Res 2021;23(7):e21680 [FREE Full text] [doi: 10.2196/21680] [Medline: 33979776]
37. Chang JE, Lindenfeld Z, Chang VW. Obesity and patient activation: confidence, communication, and information seeking behavior. J Prim Care Community Health 2022;13:21501319221129731 [FREE Full text] [doi: 10.1177/21501319221129731] [Medline: 36222682]
38. Fisher KE, Erdelez S, Mckechnie L. Theories of Information Behavior. New Delhi: Ess Ess Publications; 2008.
39. Zare-Farashbandi F, Lalazaryan A. A review of models and theories of health information seeking behavior. Int J Health Syst Disaster Manage 2014;2(4):193-203 [FREE Full text] [doi: 10.4103/2347-9019.144371]
40. Longo DR, Schubert SL, Wright BA, LeMaster J, Williams CD, Clore JN. Health information seeking, receipt, and use in diabetes self-management. Ann Fam Med 2010;8(4):334-340 [FREE Full text] [doi: 10.1370/afm.1115] [Medline: 20644188]
41. Lenz ER. Information seeking: a component of client decisions and health behavior. ANS Adv Nurs Sci 1984;6(3):59-72. [doi: 10.1097/00012272-198404000-00010] [Medline: 6426379]
42. Chang CC, Huang MH. Antecedents predicting health information seeking: a systematic review and meta-analysis. Int J Inf Manage 2020;54:102115. [doi: 10.1016/j.ijinfomgt.2020.102115]
43. Li J, Theng YL, Foo S. Predictors of online health information seeking behavior: changes between 2002 and 2012. Health Informatics J 2016;22(4):804-814 [FREE Full text] [doi: 10.1177/1460458215595851] [Medline: 26261218]
44. Beauchamp A, Batterham RW, Dodson S, Astbury B, Elsworth GR, McPhee C, et al. Systematic development and implementation of interventions to OPtimise Health Literacy and Access (Ophelia). BMC Public Health 2017;17(1):230 [FREE Full text] [doi: 10.1186/s12889-017-4147-5] [Medline: 28253883]
45. James DCS, Harville C. eHealth literacy, online help-seeking behavior, and willingness to participate in mHealth chronic disease research among African Americans, Florida, 2014-2015. Prev Chronic Dis 2016;13:E156 [FREE Full text] [doi: 10.5888/pcd13.160210] [Medline: 27854421]
46. Health Information National Trends Survey. National Institutes of Health. URL: https://hints.cancer.gov/ [accessed 2023-10-27]
47. Jiang Y, West BT, Barton DL, Harris MR. Acceptance and use of eHealth/mHealth applications for self-management among cancer survivors. Stud Health Technol Inform 2017;245:131-135 [FREE Full text] [Medline: 29295067]

48. Health Information National Trends Survey 5 (HINTS 5): cycle 4 methodology report. National Cancer Institute, Westat. 2020. URL: https://hints.cancer.gov/docs/methodologyreports/HINTS5_Cycle4_MethodologyReport.pdf [accessed 2023-10-27]
49. Xesfingi S, Vozikis A. eHealth literacy: in the quest of the contributing factors. Interact J Med Res 2016;5(2):e16 [FREE Full text] [doi: 10.2196/ijmr.4749] [Medline: 27226146]
50. Richtering SS, Hyun K, Neubeck L, Coorey G, Chalmers J, Usherwood T, et al. eHealth literacy: predictors in a population with moderate-to-high cardiovascular risk. JMIR Hum Factors 2017;4(1):e4 [FREE Full text] [doi: 10.2196/humanfactors.6217] [Medline: 28130203]
51. Bell BA, Onwuegbuzie AJ, Ferron JM, Jiao QG, Hibbard ST, Kromrey JD. Use of design effects and sample weights in complex health survey data: a review of published articles using data from 3 commonly used adolescent health surveys. Am J Public Health 2012;102(7):1399-1405 [FREE Full text] [doi: 10.2105/AJPH.2011.300398] [Medline: 22676502]
52. Population. United States Census Bureau. URL: https://www.census.gov/topics/population.html [accessed 2023-10-27]
53. Shrider EA, Kollar M, Chen F, Semega J. Income and poverty in the United States: 2020. United States Census Bureau. 2021. URL: https://www.census.gov/library/publications/2021/demo/p60-273.html [accessed 2023-10-27]
54. Lee J, Lee EH, Chae D. eHealth literacy instruments: systematic review of measurement properties. J Med Internet Res 2021;23(11):e30644 [FREE Full text] [doi: 10.2196/30644] [Medline: 34779781]
55. Chang FC, Chiu CH, Chen PH, Miao NF, Lee CM, Chiang JT, et al. Relationship between parental and adolescent eHealth literacy and online health information seeking in Taiwan. Cyberpsychol Behav Soc Netw 2015;18(10):618-624. [doi: 10.1089/cyber.2015.0110] [Medline: 26375050]
56. Schulz PJ, Fitzpatrick MA, Hess A, Sudbury-Riley L, Hartung U. Effects of eHealth literacy on general practitioner consultations: a mediation analysis. J Med Internet Res 2017;19(5):e166 [FREE Full text] [doi: 10.2196/jmir.6317] [Medline: 28512081]
57. Wang Z. Analysis of the dissemination of health information in the context of mass communication. J Edu Humanities Soc Sci 2023;13:212-218 [FREE Full text] [doi: 10.54097/ehss.v13i.7896]
58. Sun M, Jiang LC. Interpersonal influences on self-management in the eHealth era: predicting the uses of eHealth tools for self-care in America. Health Soc Care Community 2021;29(2):464-475. [doi: 10.1111/hsc.13107] [Medline: 32716139]
59. Mahmood A, Kedia S, Wyant DK, Ahn S, Bhuyan SS. Use of mobile health applications for health-promoting behavior among individuals with chronic medical conditions. Digit Health 2019;5:2055207619882181 [FREE Full text] [doi: 10.1177/2055207619882181] [Medline: 31656632]
60. Kahan S, Look M, Fitch A. The benefit of telemedicine in obesity care. Obesity (Silver Spring) 2022;30(3):577-586. [doi: 10.1002/oby.23382] [Medline: 35195367]
61. Robbins R, Krebs P, Jagannathan R, Jean-Louis G, Duncan DT. Health app use among US mobile phone users: analysis of trends by chronic disease status. JMIR Mhealth Uhealth 2017;5(12):e197 [FREE Full text] [doi: 10.2196/mhealth.7832] [Medline: 29258981]
62. Kumar D, Hemmige V, Kallen MA, Giordano TP, Arya M. Mobile phones may not bridge the digital divide: a look at mobile phone literacy in an underserved patient population. Cureus 2019;11(2):e4104 [FREE Full text] [doi: 10.7759/cureus.4104] [Medline: 31057998]
63. Martin SS, Feldman DI, Blumenthal RS, Jones SR, Post WS, McKibben RA, et al. mActive: a randomized clinical trial of an automated mHealth intervention for physical activity promotion. J Am Heart Assoc 2015;4(11):e002239 [FREE Full text] [doi: 10.1161/JAHA.115.002239] [Medline: 26553211]
64. Watson-Singleton NN, Pennefather J, Trusty T. Can a culturally-responsive mobile Health (mHealth) application reduce African Americans' stress?: A pilot feasibility study. Curr Psychol 2021;42(2):1434-1443. [doi: 10.1007/s12144-021-01534-9]
65. Pagoto S, Schneider K, Jojic M, DeBiasse M, Mann D. Evidence-based strategies in weight-loss mobile apps. Am J Prev Med 2013;45(5):576-582. [doi: 10.1016/j.amepre.2013.04.025] [Medline: 24139770]
66. Litchfield I, Shukla D, Greenfield S. Impact of COVID-19 on the digital divide: a rapid review. BMJ Open 2021;11(10):e053440 [FREE Full text] [doi: 10.1136/bmjopen-2021-053440] [Medline: 34642200]
67. Park YT. Emerging new era of mobile health technologies. Healthc Inform Res 2016;22(4):253-254 [FREE Full text] [doi: 10.4258/hir.2016.22.4.253] [Medline: 27895955]
68. Alhusseini N, Banta JE, Oh J, Montgomery S. Understanding the use of electronic means to seek personal health information among adults in the United States. Cureus 2020;12(10):e11190. [doi: 10.7759/cureus.11190] [Medline: 33269121]
69. Breland JY, Agha K, Mohankumar R. Adoption and appropriateness of mHealth for weight management in the real world: a qualitative investigation of patient perspectives. JMIR Form Res 2021;5(12):e29916 [FREE Full text] [doi: 10.2196/29916] [Medline: 34889761]
70. Zimmerman MS, Shaw G. Health information seeking behaviour: a concept analysis. Health Info Libr J 2020;37(3):173-191 [FREE Full text] [doi: 10.1111/hir.12287] [Medline: 32052549]
71. Santoro E, Castelnuovo G, Zoppis I, Mauri G, Sicurello F. Social media and mobile applications in chronic disease prevention and management. Front Psychol 2015;6:567 [FREE Full text] [doi: 10.3389/fpsyg.2015.00567] [Medline: 25999884]

72. Mirzaei T, Kashian N. Revisiting effective communication between patients and physicians: cross-sectional questionnaire study comparing text-based electronic versus face-to-face communication. J Med Internet Res 2020;22(5):e16965 [FREE Full text] [doi: 10.2196/16965] [Medline: 32401213]
73. Woods LS, Duff J, Roehrer E, Walker K, Cummings E. Patients' experiences of using a consumer mHealth app for self-management of heart failure: mixed-methods study. JMIR Hum Factors 2019;6(2):e13009 [FREE Full text] [doi: 10.2196/13009] [Medline: 31045504]
74. Shaw G, Nadkarni D, Phann E, Sielaty R, Ledenyi M, Abnowf R, et al. Separating features from functionality in vaccination apps: computational analysis. JMIR Form Res 2022;6(10):e36818 [FREE Full text] [doi: 10.2196/36818] [Medline: 36222791]
75. Kao CK, Liebovitz DM. Consumer mobile health apps: current state, barriers, and future directions. PM R 2017;9(5S):S106-S115. [doi: 10.1016/j.pmrj.2017.02.018] [Medline: 28527495]
76. Rubenstein EL. “Things my doctor never told me”: bridging information gaps in an online community. Proc Am Soc Info Sci Tech 2012;49(1):1-10 [FREE Full text] [doi: 10.1002/meet.14504901126]
77. Zhou J, Fan T. Understanding the factors influencing patient e-health literacy in online health communities (OHCs): a social cognitive theory perspective. Int J Environ Res Public Health 2019;16(14):2455 [FREE Full text] [doi: 10.3390/ijerph16142455] [Medline: 31295924]

Abbreviations

AUC: area under the curve
HINTS: Health Information National Trends Survey
IRB: institutional review board
mHealth: mobile health
OR: odds ratio
ROC: receiver operating characteristic

Edited by L Buis; submitted 22.02.23; peer-reviewed by V Gosselin Boucher, CA Almenara; comments to author 03.07.23; revised version received 01.09.23; accepted 18.09.23; published 10.01.24.

Please cite as:
Shaw Jr G, Castro BA, Gunn LH, Norris K, Thorpe Jr RJ
The Association of eHealth Literacy Skills and mHealth Application Use Among US Adults With Obesity: Analysis of Health Information National Trends Survey Data
JMIR Mhealth Uhealth 2024;12:e46656
URL: https://mhealth.jmir.org/2024/1/e46656
doi:10.2196/46656
PMID:38198196

Original Paper

Physical Activity, Body Composition, and Fitness Variables in Adolescents After Periods of Mandatory, Promoted or Nonmandatory, Nonpromoted Use of Step Tracker Mobile Apps: Randomized Controlled Trial

Adrián Mateo-Orcajada[1*], PhD; Raquel Vaquero-Cristóbal[2*], PhD; Jorge Mota[3*], PhD; Lucía Abenza-Cano[1*], PhD

[1]Facultad de Deporte, Universidad Católica de Murcia, Murcia, Spain

[2]Research Group Movement Sciences and Sport (MS&SPORT), Department of Physical Activity and Sport, Faculty of Sport Sciences, University of Murcia, Murcia, Spain

[3]Faculdade de Desporto, Universidade de Porto, Porto, Portugal

[*]all authors contributed equally

Corresponding Author:
Raquel Vaquero-Cristóbal, PhD
Research Group Movement Sciences and Sport (MS&SPORT), Department of Physical Activity and Sport
Faculty of Sport Sciences
University of Murcia
C. Argentina 19
San Javier
Murcia, 30720
Spain
Phone: 34 868 88 86 84
Email: raquel.vaquero@um.es

Abstract

Background: It is not known whether an intervention made mandatory as a physical education (PE) class assignment and aimed at promoting physical activity (PA) in adolescents can create a healthy walking habit, which would allow further improvements to be achieved after the mandatory and promoted intervention has been completed.

Objective: The aims of this study were to (1) investigate whether, after a period of using a step tracker mobile app made mandatory and promoted as a PE class assignment, adolescents continue to use it when its use is no longer mandatory and promoted; (2) determine whether there are changes in the PA level, body composition, and fitness of adolescents when the use of the app is mandatory and promoted and when it is neither mandatory nor promoted; and (3) analyze whether the covariates maturity status, gender, and specific app used can have an influence.

Methods: A total of 357 students in compulsory secondary education (age: mean 13.92, SD 1.91 y) participated in the study. A randomized controlled trial was conducted consisting of 2 consecutive 10-week interventions. Participants' PA level, body composition, and fitness were measured at baseline (T1), after 10 weeks of mandatory and promoted app use (T2), and after 10 weeks of nonmandatory and nonpromoted app use (T3). Each participant in the experimental group (EG) used 1 of 4 selected step tracker mobile apps after school hours.

Results: The results showed that when the use of the apps was neither mandatory nor promoted as a PE class assignment, only a few adolescents (18/216, 8.3%) continued the walking practice. After the mandatory and promoted intervention period (T1 vs T2), a decrease in the sum of 3 skinfolds (mean difference [MD] 1.679; *P*=.02) as well as improvements in the PA level (MD –0.170; *P*<.001), maximal oxygen uptake (MD –1.006; *P*<.001), countermovement jump test (MD –1.337; *P*=.04), curl-up test (MD –3.791; *P*<.001), and push-up test (MD –1.920; *P*<.001) in the EG were recorded. However, the changes between T1 and T2 were significantly greater in the EG than in the control group only in the PA level and curl-up test. Thus, when comparing the measurements taken between T1 and T3, no significant changes in body composition (*P*=.07) or fitness (*P*=.84) were observed between the EG and the control group. The covariates maturity status, gender, and specific app used showed a significant effect in most of the analyses performed.

Conclusions: A period of mandatory and promoted use of step tracker mobile apps benefited the variables of body composition and fitness in adolescents but did not create a healthy walking habit in this population; therefore, when the use of these apps ceased to be mandatory and promoted, the effects obtained disappeared.

Trial Registration: ClinicalTrials.gov NCT06164041; https://clinicaltrials.gov/study/NCT06164041

(JMIR Mhealth Uhealth 2024;12:e51206) doi:10.2196/51206

KEYWORDS

body composition; detraining; new technologies; physical education subject; physical fitness; youth

Introduction

Background

In recent years, a growing significance has been placed on the engagement of adolescents in physical activity (PA), given the decrease in active time and the increase in sedentary activities and screen time [1]. These behaviors have negatively impacted the health of adolescents by increasing the risk of cardiovascular disease and other associated chronic diseases [2]. This situation has prompted the implementation of new practices that have been shown to be effective in increasing the level of sports practice in adolescents [3,4]. In this context, interventions that incorporate electronic devices have gained relevance [5], given their extensive use during the COVID-19 pandemic and their integration into the daily lives of adolescents [6].

Mobile sports apps have emerged as valuable resources in promoting PA among adolescents [7], and it has been observed that interventions with mobile devices have made it possible to increase moderate-intensity PA and daily step count among users [8]. This has also had a positive impact on the health of the adolescent population because the use of these mobile apps has improved their body composition and fitness levels [9,10], which are fundamental for their subsequent development [11,12]. This is because adolescents who are overweight or obese have a high probability of remaining so in adulthood [13], with the associated high health risk for cardiovascular and respiratory diseases [14]. On the contrary, adequate levels of body fat, within the limits considered healthy [15], as well as an adequate physical fitness level, especially good cardiorespiratory fitness, are indicators of adequate health and serve as preventive factors against various diseases in adulthood [16].

In this regard, mobile apps seem to be effective tools for improving the health status of adolescents and for preventing future health risks. It is crucial to emphasize that the effectiveness of increasing adolescents' PA level through mobile apps was evident only in studies in which the use of these apps was mandatory. Specifically, the promotion of app use as an assignment in physical education (PE) classes played a significant role in achieving positive outcomes [9,17]. Furthermore, it is worth noting that no major differences were found in the effects achieved by the intervention when comparing the different mobile apps used, as long as they were all step trackers [9]; however, the gender of the adolescents was shown to be a determinant factor in the benefits obtained because female adolescents used these apps more often than male adolescents during the mandatory and promoted intervention period, which led to significant differences in the benefits obtained in BMI, corrected calf girth, fat mass, and physical fitness [18]. Therefore, the mandatory use of apps seemed to be effective in this population, although it should be noted that a previous study showed that the first weeks of the intervention were the most effective in the adolescent population due to the novelty of the intervention, but as the intervention progressed, the effects were reduced [19]. This is a relevant aspect because there is a considerable loss of adherence after the first weeks of the intervention [8], which could negatively influence the overall benefits obtained.

In addition to the loss of adherence, there are also periods in the school calendar when PA decreases, such as holiday breaks. These are characterized by the absence of students from school, which makes it difficult to promote the use of this type of intervention as a PE class assignment. This has a particular significance because prior research has revealed a detraining effect, wherein the gains in body composition and fitness achieved during an aerobic intervention period were subsequently lost, leading to a regression to preintervention levels [20].

Therefore, it is essential to verify whether, after a period of mandatory and promoted use of step tracker mobile apps as a PE class assignment, which has shown beneficial effects on body composition and physical fitness in previous research [9], it is possible to create a healthy walking habit in adolescents and to have them continue using the apps when they are neither mandatory nor promoted to try to avoid detraining effects. This would allow us to define strategies to compensate for the loss of adherence and decrease in PA level observed during holiday periods, similar to previous research on flexibility in adolescents [21]. However, no previous research is known to have analyzed whether the effects achieved throughout a period of mandatory and promoted use of step tracker mobile apps are maintained over time when their use is neither mandatory nor promoted as a PE class assignment. Furthermore, previous studies in this area have not analyzed whether the effects of such interventions may depend on the maturity status of adolescents, although numerous studies have shown that the rate of maturation during adolescence varies between individuals [22]. Thus, this factor may condition the changes in body composition and fitness variables in adolescents [23].

Objectives

Therefore, considering the absence of previous research analyzing whether adolescents aged 12 to 16 years continue their walking practice with step tracker mobile apps when their use is no longer mandatory and promoted as a PE class

assignment, as well as the influence that covariates such as maturity status, gender, and specific app used may have on the results, the aims of this study were to (1) investigate whether, after a period of mandatory and promoted use of a step tracker mobile app as a PE class assignment, adolescents continue to use it when its use is no longer mandatory and promoted; (2) determine whether there are changes in PA level, body composition, and fitness of adolescents when the use of the app is mandatory and promoted as a PE class assignment, as well as whether maturity status, gender, and specific app used can have an influence on the results; and (3) analyze whether there are changes in PA level, body composition, and fitness of adolescents when the use of the app is neither mandatory nor promoted as a PE class assignment, as well as whether maturity status, gender, and specific app used can have an influence on the results.

Hypotheses

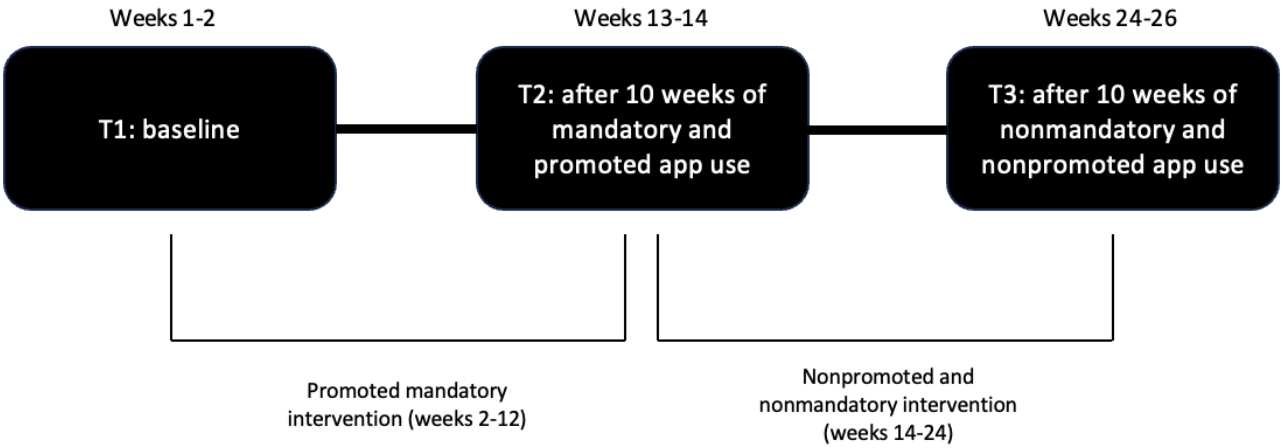
On the basis of the aims of the research and previous research involving technological devices (eg, wearables) or websites, it is hypothesized that adolescents will stop using the mobile app during the period of nonmandatory and nonpromoted use (H1); that there will be significant differences in PA level, body composition, and fitness of adolescents during the mandatory and promoted period, influenced by maturity status and gender but not by specific app used (H2); and that some of the benefits achieved by adolescents during the mandatory and promoted period will be lost after the mandatory and promoted intervention has been completed, with the results being influenced by maturity status and gender but not by specific app used (H3).

Methods

Design

The intervention in this study was carried out by replicating the methodology of previous research [9], the main difference being that this research analyzed what happens to the study variables when the mandatory intervention ends and becomes a nonmandatory, nonpromoted intervention. Our new research design comprised 3 data collection periods (T1: baseline, T2: after 10 weeks of mandatory and promoted app use, and T3: after 10 weeks of nonmandatory and nonpromoted app use) with a total duration of 26 weeks. T1 took place in the first 2 weeks (weeks 1-2); the mandatory intervention with the step tracker mobile apps promoted as a PE class assignment took place in the following 10 weeks (weeks 2-12); T2 took place in the next 2 weeks (weeks 13-14); the use of the step tracker mobile apps was neither mandatory nor promoted during the following 10 weeks (weeks 14-24); and T3 took place in the last 2 weeks of the study (weeks 24-26). Figure 1 shows the timeline of the study. The intervention began on January 9, 2023, and ended on June 23, 2023.

Figure 1. Timeline of the study.



This study was a randomized controlled trial. It followed the CONSORT (Consolidated Standards of Reporting Trials) guidelines [24] and was preregistered at ClinicalTrials.gov (NCT06164041). A convenience sampling method was used to recruit adolescents from accessible educational institutions.

The study used a specific research model (Table 1).

Table 1. Research model.

Variable type	Construct	Groups included	Variables included
Independent	Physical activity level	App use group and control group	Subjective assessment of the level of physical activity
Dependent	Kinanthropometric and body composition variables	App use group and control group	Body mass; height; BMI; sitting height; sum of 3 skinfolds; corrected arm, thigh, and calf girths; waist girth; hips girth; waist to hip ratio; muscle mass; and fat mass
Dependent	Physical fitness variables	App use group and control group	VO_{2max}[a], CMJ[b] test, curl-up test, and push-up test
Covariates	Variables	App use group and control group	Maturity status, gender, and specific app used

[a]VO_{2max}: maximal oxygen uptake.

[b]CMJ: countermovement jump.

Ethical Considerations

This study was approved by the institutional ethics committee of the Catholic University of Murcia (code CE022102) and adhered to the guidelines set forth by the World Medical Association and the Declaration of Helsinki. Adolescents who expressed willingness to participate in the study were required to sign an informed consent form, with both adolescents and their parents acknowledging their understanding of the study aims and procedures.

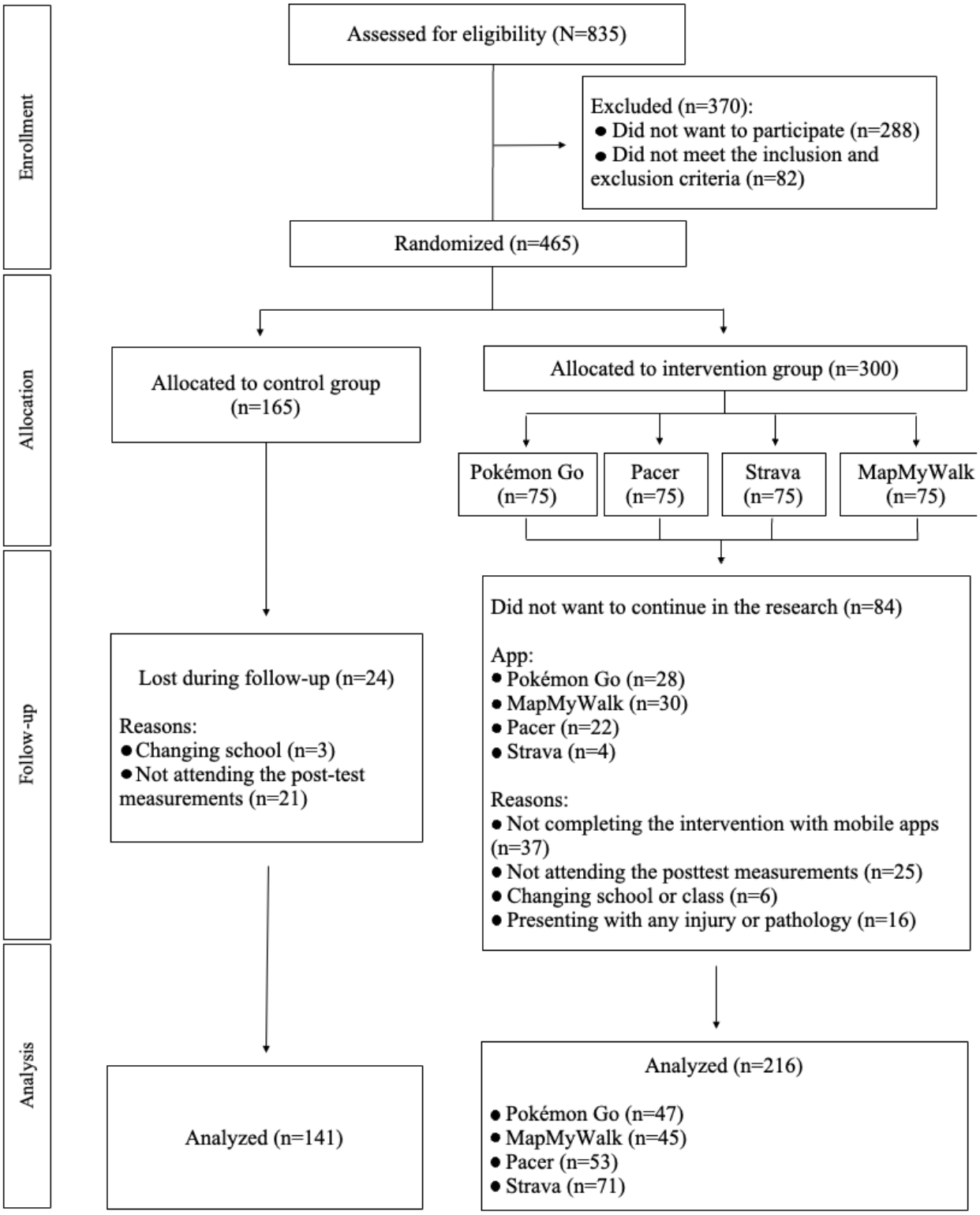
Participants

We recruited participants from 2 compulsory secondary schools located in Murcia. These schools were chosen because of their large student population in secondary education within their respective localities. Initially, the research team contacted the schools to provide a detailed explanation of the study's procedure and objectives. If a particular school declined to participate, the school with the next largest number of students in the locality was approached. Once the school's approval was obtained, the PE department heads were contacted. Subsequently, a face-to-face meeting was arranged with interested students and their parents to discuss the study further.

The minimum sample size necessary for the study was calculated using RStudio software (version 3.15.0; Posit Software PBC) and followed the methodology used in previous studies [25], in which the SD value (0.64) from previous studies that used a similar design with 3 data points to measure changes in PA among adolescents was used [26]. With an estimated error (d) of 0.067 and a CI of 95%, the required sample size was determined to be 350 adolescents.

Figure 2 illustrates the flowchart for the selection of the sample. The final sample comprised 357 adolescents aged between 12 and 16 years. The participants were assigned to the different groups using a cluster randomized design [27]. Group assignment was concealed from the researcher who analyzed each participant's compliance with the inclusion and exclusion criteria. The inclusion criteria for the study were as follows: (1) enrollment in 1 of the selected educational institutions, (2) aged between 12 and 16 years, (3) completion of all questionnaires and physical tests during the 3 measurement periods (T1, T2, and T3), (4) attending the kinanthropometric and body composition assessment sessions, and (5) absence of any pathology or injury that would hinder participation in the tests or measurements conducted. The exclusion criteria were as follows: (1) missing >20% of the mandatory PE sessions throughout the academic year, (2) lack of a mobile phone, (3) failure to meet the minimum mandatory weekly distance requirement in the experimental group (EG) when app use was mandatory and promoted, (4) changing schools or class group during the course of the intervention, (5) starting or ending any form of PA (for reasons unrelated to the study) during the intervention that could alter the PA level being assessed as part of the study, and (6) having presented with any illness during the follow-up period that would have prevented the participant from engaging in their usual PA.

Figure 2. Sample selection flowchart.



Randomization and Blinding

After the recruitment and selection of the adolescents, meetings were held with the teachers to provide them with a clear understanding of the trial's purpose and the randomization process. Parents or legal guardians of the potential participants at each school were notified through a letter that explained the study's objectives and procedures. The principal investigator, along with other uninvolved investigators, carried out the randomization process using a computer-generated random number table. The randomization assigned all students within the same class at each school to the same mobile app group. The classes were randomly assigned to participate as intervention or control classes. A total of 16 classes were finally randomized, of which 11 (69%) were included in the EG and 5 (31%) in the control group (CG). The ratio chosen for the randomized clusters was 2:1 (for every 2 classes included in the EG, 1 was included in the CG) because previous research

with mobile apps has highlighted the lack of adherence to mobile apps, and we wanted to ensure that we had enough participants in the EG to account for a possible high dropout rate (close to 35%) from hindering the extrapolation of the results [28]. The control classes were instructed to continue their regular PE classes, while the intervention was offered to them after the final data collection took place. Baseline measurements were taken before the randomization process. All measurers were blinded to the group to which each individual belonged during the second and third measurements, as well as to the individual's ratings in the previous measurements.

Instruments

The instruments used in this study were the same as those used in previous investigations [9] because these are valid and reliable in the adolescent population.

Questionnaire Measurement

A sociodemographic questionnaire developed ad hoc was administered to obtain data on the age and gender of the participants, their regular PA, and the occurrence of injury or illness, following the pattern of previous studies [23].

PA level was measured using the Physical Activity Questionnaire for Adolescents (PAQ-A) [29]. This questionnaire had been previously validated in Spanish and showed satisfactory reliability, with an intraclass correlation coefficient of 0.71 for the final score [30].

Kinanthropometric and Body Composition Measurement

The anthropometric measurement included 3 basic parameters (body mass, height, and sitting height), 3 skinfold measurements (triceps, thigh, and calf), and 5 girth measurements (arm relaxed, waist, hips, thigh, and calf) [31]. Measurements were performed by anthropometrists (level 3 and level 4) accredited by the International Society for the Advancement of Kinanthropometry [31].

The anthropometric instruments used were the same as those used in previous research [9]: a Harpenden skinfold caliper, a Lufkin W606PM anthropometric tape measure, a Tanita BC418-MA segmental scale, and a Seca 213 stadiometer. All instruments were calibrated before the beginning of each of the measurements (T1, T2, and T3).

The following derived variables were calculated from the anthropometric measurements: BMI, muscle mass [32], fat mass [33], sum of 3 skinfolds (triceps, thigh, and calf), waist to hip ratio (waist girth/hips girth), and corrected girths of the arm (arm relaxed girth – [π × triceps skinfold]), thigh (middle thigh girth – [π × thigh skinfold]), and calf (calf girth – [π × calf skinfold]) [34]. The body composition formulas have been used in previous research [35] and are the ones most often recommended for evaluation in this population [36].

The maturity offset was calculated according to the procedure established by Mirwald et al [37] and using gender-specific formulas: –9.37 + 0.0001882 × ((height – sitting height) × sitting height) – 0.0022 × (age × (height – sitting height)) + 0.005841 × (age × sitting height) – 0.002658 × (age × weight) + 0.07693 × (weight / height). The result of the maturity offset equation is expressed in years from the age at peak height velocity (PHV) when the result is positive and in years to the age at PHV when the result is negative.

The same anthropometrist performed the T1, T2, and T3 measurements on each participant to reduce interevaluator error. The intra- and interevaluator technical errors of measurement [34] were 0.02% and 0.04% for basic measurements, 1.09% and 1.87% for skinfolds, and 0.03% and 0.08% for girths. The correlation coefficients of the anthropometrists with respect to a level 4 expert anthropometrist were 0.96 for basic measurements, 0.91 for skinfolds, and 0.93 for girths.

Physical Fitness Measurements

Cardiorespiratory fitness was evaluated using the 20-meter shuttle run test. The test ends when the adolescent is unable to complete the required distance in the indicated time twice consecutively or when he or she reaches exhaustion. Upon completion of the test, the final speed at which the adolescent concluded the shuttle run was used to calculate their maximal oxygen uptake (VO_{2max}) [38]. This test has high validity and reliability for the determination of VO_{2max} [39].

Lower limb explosive strength was assessed by means of the countermovement jump (CMJ). Adolescents had to perform a 90-degree knee flexion at maximum speed, keeping the back fully straight with hands placed on the hips, followed by a maximal knee extension to jump [40]. The adolescents were required to execute a maximal jump while maintaining their hands on their hips throughout the test. The jump height was determined by measuring the flight height achieved during the jump [40].

For the measurement of abdominal strength and endurance, we used the curl-up test. For the execution to be valid, the adolescents had to keep their feet fully supported on the floor and their arms crossed on the chest, and the trunk flexion had to allow the upper back to be lifted off the floor [41]. The test ended when the time was up (1 min) or when the participant reached exhaustion.

The push-up test was used to evaluate upper body strength. The repetition was valid if the adolescents managed to fully extend their arms and return to the 90-degree position [42]. The adolescents had to perform as many push-ups as possible in 1 minute. The test ended when the time was up (1 min) or when the participant reached exhaustion.

Procedure

The procedure used was also similar to that used in previous research [9], but the difference was that in this study, the intention was to discover how adherence to the intervention changed when the intervention was no longer mandatory or promoted as a PE class assignment and how this affected the variables analyzed. Therefore, unlike previous studies [9], this study comprised 3 measurement periods (T1, T2, and T3), with 2 interventions carried out consecutively. The first one was mandatory and promoted as a PE class assignment, and the second one was neither promoted nor promoted as a PE class assignment.

The data collection process followed the same protocol as in previous research [9,43], with the sociodemographic and PAQ-A questionnaires completed first, followed by the anthropometric measurements. Once these were completed, the fitness tests were explained and performed randomly, twice each, with the 20-meter shuttle run test performed last and only once. The physical test protocol adhered to the guidelines set forth by the National Strength and Conditioning Association, with the aim of minimizing interference between tests and allowing sufficient recovery time from the exertion and metabolic demands of the assessments [44].

To minimize bias in the measurements, they were carried out under the same conditions for all students. The PE class hour was used for the measurements and the adolescents were always measured at the same time and on the same day of the week at T1-T2-T3. The questionnaires were completed in a reserved space in which the adolescents did not have any distractions that could have conditioned their answers. In addition, while the researchers resolved any possible doubts, in no case did they condition the adolescents' responses. For the anthropometric measurements, the air-conditioned locker rooms of the sports pavilion were used to minimize variability due to temperature and humidity fluctuations across the 3 measurement periods. To conduct the physical tests, the indoor sports pavilion at each school was used, which was specifically chosen to eliminate the influence of atmospheric variables that could potentially affect the results and introduce bias.

Mobile App Intervention

Before starting the intervention, 465 adolescents participated in pretest measurements (T1; Figure 2). The mandatory and promoted intervention lasted 10 weeks, during which the adolescents were required to use 1 of the 4 selected apps: Pokémon Go, Pacer, Strava, or MapMyWalk. These apps were selected based on their implementation of a substantial number of behavior change techniques [45] specifically designed to effectively enhance PA level among users. Moreover, they have already been used in previous research with adolescents, with participants demonstrating good adherence [9]. The assignment to each of the app groups was randomized by class group. Thus, initially, an equal number of adolescents was assigned to use each app (Pokémon Go: 75/300, 25%; MapMyWalk: 75/300, 25%; Pacer: 75/300, 25%; and Strava: 75/300, 25%). Of the 465 adolescents, 165 (35.5%) were assigned to the CG.

Before starting the mandatory and promoted intervention, the adolescents were provided with instructions on the proper use of step tracker mobile apps. The aim of the first phase was for students, after receiving instructions on the correct use of the apps, to use them in a manner guided by the PE teachers so that they could become familiar with their use and interface. For this purpose, after randomization, a meeting was held with each of the class groups that were assigned to the EG. In this meeting, the students installed the app corresponding to their class group, and an explanation on the functioning of each was provided to them. Any doubts were resolved by the researchers and the PE teachers. The researchers in charge of explaining how the apps worked were not involved in the measurements or subsequent analysis because they knew which student belonged to each app group and which student belonged to the CG. Once each app had been described and its use explained, a training plan was drawn up to be followed during the period of mandatory and promoted use. During the initial week, the adolescents were instructed to achieve a minimum of 5000 steps or cover a distance of at least 3.19 km each time they used the app. It was established that approximately 1565 steps equals 1 km [46]. This minimum distance was defined to ensure that the adolescents exceeded the sedentary threshold [47]. The initial distance was progressively increased weekly until reaching a distance of 15,520 steps or 8 km each time they used the app. In addition, the researchers followed up with the PE teachers to ensure that the distance was completed by the students every week.

The adolescents were motivated to use the app for a duration of 10 weeks, aiming for a minimum use of 3 times per week. This frequency aligned with the PA recommendations set forth by the World Health Organization [48]. The duration of 10 weeks was justified based on previous research with adolescents, in which a short or moderate duration (6-12 wk) was more effective for producing changes than a longer duration [19], and to be able to adjust it to the duration of the academic year. To encourage the use of the mobile apps during the period of mandatory and promoted use, PE teachers rewarded participation in the study with up to 1 point in the final PE grade for those who completed the study.

After the mandatory and promoted intervention with the mobile apps, posttest 1 measurements were carried out (T2). This was followed by a 10-week period in which the use of the apps was no longer promoted or mandatory as a PE class assignment, after which posttest 2 (T3) measurements were taken. During this period of nonmandatory and nonpromoted use, the adolescents could continue to use the mobile apps voluntarily, just as they would in their daily lives. The adolescents who continued to use the apps were recorded.

In both the mandatory and promoted and the nonmandatory and nonpromoted periods, a researcher who did not participate in the data collection process recorded the distance (in kilometers) and the number of steps taken by each participant after using the mobile apps daily.

A total of 357 adolescents participated in the final measurements (Pokémon Go: n=47, 13.2%; MapMyWalk: n=45, 12.6%; Pacer: n=53, 14.8%; Strava: n=71, 19.9%; and CG: n=141, 39.5%), while 108 adolescents dropped out of the program (Pokémon Go: n=28, 25.9%; MapMyWalk: n=30, 27.8%; Pacer: n=22, 20.4%; Strava: n=4, 3.7%; and CG: n=24, 22.2%; Figure 2). Adolescents who, despite the mandatory and promoted use, did not start using the mobile app were considered to have dropped out, as were those who did not complete at least 25% of the total training volume required because previous research has shown that this is the minimum volume needed to produce significant changes in body composition and fitness variables [43]. Adolescents who completed at least 25% of the training volume required were retained in their respective app groups, those who exceeded 25% of the required training volume but did not complete the entire intervention received up to half a point in the final PE grade, and those who dropped out or did

not complete at least 25% of the training volume did not receive any bonus point in the final PE grade.

Data Analysis

The normality of the data was assessed using the Kolmogorov-Smirnov test, alongside analyses of skewness, kurtosis, and variance. As the variables exhibited a normal distribution, parametric tests were used for their analysis. Three repeated measures ANOVAs were performed. On the first, the group factor was used as the grouping variable; on the second, the time point factor was used; and on the third, the differences in the changes between the CG and EG at the different time points were assessed. In this way, intra- and intergroup differences were determined for each of the study variables. A subsequent Bonferroni analysis made it possible to determine the statistical differences between each of the pairs compared. Three analyses of covariance were also performed to determine the influence of the covariates maturity status, gender, and specific app used on the results obtained for the study variables. Effect size was analyzed using partial eta-squared (η_p^2) and was defined as small (≥0.10), moderate (≥0.30), large (≥0.50), very large (≥0.70) or extremely large (≥0.90). These translate into 0.20, 0.60, 1.20, 2.0 and 4.0 for standardized differences in means [49]. A *P* value <.05 was used to establish statistical significance. The data analysis was performed using SPSS software (version 25.0; IBM Corp).

Results

Overview

Of the 357 adolescents, 186 (52.1%) were male, and 171 (47.9%) were female. Of the 186 male adolescents, 26 (14%) used Pokémon Go, 35 (18.8%) used Strava, 29 (15.6%) used Pacer, 25 (13.4%) used MapMyWalk, and 71 (19.9%) were in the CG. Of the 171 female adolescents, 21 (12.3%) used Pokémon Go, 36 (21.1%) used Strava, 24 (14%) used Pacer, 20 (11.7%) used MapMyWalk, and 70 (40.9%) were in the CG. The mean age of the male adolescents was 13.91 (SD 1.22) years, with a mean maturity offset of 0.20 (SD 1.39) years. Their mean body mass was 55.68 (SD 13.09) kg, and their mean height was 164.59 (SD 10.07) cm. The mean age of the female adolescents was 13.89 (SD 1.21) years, with a mean maturity offset of 1.50 (SD 0.90) years. Their mean body mass was 52.53 (SD 10.92) kg, and their mean height was 158.76 (SD 6.32) cm.

Of the 216 adolescents in the EG during the period of mandatory and promoted use of the app, only 18 (8.3%) continued to use the apps independently during the nonmandatory and nonpromoted period. The average distance walked by these adolescents was 47.69 (SD 23.80; range 5-200) km in the 10 weeks of nonpromoted and nonmandatory use.

Differences in the EG and CG at the Different Study Time Points

Table 2 shows the differences in the measurements taken in the EG and CG at the different study time points (T1, T2, and T3). Significant differences in the PA level were observed exclusively within the EG: it was higher at T2 than at T1 (*P*<.001) but lower at T3 than at T2 (*P*=.03). No differences were found between T1 and T3 (*P*=.47) either in the EG or in the CG in any of the comparisons. In terms of the anthropometric and body composition variables, body mass and height significantly increased in both EG and CG between T1 and T2 (*P*<.001) and T1 and T3 (*P*=.002-.008), but no differences were found between T2 and T3 (*P*=.23-.99). In the sum of 3 skinfolds, the EG showed a significant decrease between T1 and T2 (*P*=.02), but a significant increase was found between T2 and T3 (*P*=.03). All corrected girth (*P*<.001-.049) and muscle mass (*P*<.001-.007) measurements showed significant increases in both groups between the 3 time points, including hips girth (*P*<.001-.03). By contrast, BMI (*P*=.01-.99), sitting height (*P*=.11-.99), fat mass (*P*=.07-.99), and waist girth (*P*=.23-.99) did not show differences in any of the groups in any of the comparisons.

Table 2. Differences in the experimental group (EG) and the control group (CG) during the different study time points (intragroup differences).

Variable and group	T1, mean (SD)	T2, mean (SD)	T3, mean (SD)	Mean difference (T1–T2)	*P* value	Mean difference (T1–T3)	*P* value	Mean difference (T2–T3)	*P* value	*F* test (*df*)	η_p^2
Subjective level of physical activity											
EG	2.62 (0.68)	2.79 (0.59)	2.68 (0.68)	–0.170	<.001	–0.060	.47	0.110	.03	11.208 (1)	0.060
CG	2.72 (0.64)	2.72 (0.73)	2.66 (0.71)	–0.004	.99	0.057	.79	0.061	.74	0.791 (1)	0.004
Body mass (kg)											
EG	55.16 (12.87)	56.06 (12.69)	56.03 (11.63)	–0.901	<.001	–0.873	.004	0.028	.99	24.833 (1)	0.126
CG	52.56 (10.84)	53.51 (10.71)	53.71 (10.72)	–0.946	<.001	–1.148	.002	–0.202	.99	18.259 (1)	0.098
Height (cm)											
EG	162.35 (SD 9.04)	163.11 (8.98)	163.27 (9.74)	–0.760	<.001	–0.915	.008	–0.155	.99	20.640 (1)	0.107
CG	161.02 (8.82)	161.63 (8.77)	162.25 (9.31)	–0.613	<.001	–1.229	.003	–0.616	.23	10.789 (1)	0.059
BMI (kg/m^2)											
EG	20.87 (3.84)	20.98 (3.69)	20.93 (3.69)	–0.107	.09	–0.051	.99	0.055	.53	2.688 (1)	0.015
CG	20.19 (3.34)	20.44 (3.21)	20.26 (3.19)	–0.251	.11	–0.071	.91	0.179	.40	12.041 (1)	0.066
Sitting height (cm)											
EG	84.75 (9.59)	85.54 (4.78)	82.90 (19.43)	–0.799	.98	1.846	.40	2.644	.11	2.184 (1)	0.012
CG	82.85 (12.28)	83.43 (11.18)	83.12 (15.32)	–0.580	.99	–0.263	.99	0.317	.99	0.170 (1)	0.001
Sum of 3 skinfolds (mm)											
EG	52.03 (26.58)	50.35 (24.51)	51.50 (25.40)	1.679	.02	0.530	.99	–1.149	.32	5.599 (1)	0.032
CG	45.05 (24.18)	44.44 (23.30)	45.12 (23.80)	0.607	.99	–0.072	.99	–0.680	.66	0.889 (1)	0.005
Corrected arm girth (cm)											
EG	20.83 (2.77)	21.26 (2.79)	21.48 (2.83)	–0.431	<.001	–0.651	<.001	–0.220	<.001	52.455 (1)	0.235
CG	20.81 (2.75)	21.20 (2.67)	21.49 (2.67)	–0.388	<.001	–0.676	<.001	–0.288	<.001	36.041 (1)	0.174
Corrected thigh girth (cm)											
EG	39.18 (4.78)	40.11 (4.57)	40.19 (4.73)	–0.924	<.001	–1.010	<.001	–0.086	.99	17.816 (1)	0.111
CG	39.43 (5.24)	39.89 (4.22)	40.64 (4.40)	–0.454	.049	–1.201	<.001	–0.747	<.001	14.503 (1)	0.078
Corrected calf girth (cm)											
EG	28.95 (3.55)	29.27 (2.91)	29.35 (2.90)	–0.321	.049	–0.400	.01	–0.078	.45	4.271 (1)	0.024
CG	28.75 (2.75)	29.28 (2.66)	29.37 (2.68)	–0.531	.005	–0.621	.001	–0.090	.53	6.510 (1)	0.037
Waist girth (cm)											

Variable and group	T1, mean (SD)	T2, mean (SD)	T3, mean (SD)	Mean difference (T1–T2)	*P* value	Mean difference (T1–T3)	*P* value	Mean difference (T2–T3)	*P* value	*F* test (*df*)	η_p^2
EG	68.36 (8.84)	68.44 (8.44)	68.48 (9.05)	–0.083	.99	–0.121	.99	–0.038	.99	0.196 (1)	0.001
CG	67.57 (7.13)	67.84 (7.24)	68.01 (7.49)	–0.276	.45	–0.443	.23	–0.167	.99	1.644 (1)	0.010
Hips girth (cm)											
EG	89.24 (9.25)	90.18 (8.74)	90.55 (8.77)	–0.937	<.001	–1.313	<.001	–0.377	.03	21.776 (1)	0.113
CG	86.40 (7.88)	87.67 (7.83)	88.31 (7.96)	–1.270	<.001	–1.906	<.001	–0.636	.001	29.078 (1)	0.145
Waist to hip ratio											
EG	0.77 (0.05)	0.76 (0.05)	0.76 (0.06)	0.007	<.001	0.010	<.001	0.003	.38	16.446 (1)	0.088
CG	0.78 (0.05)	0.77 (0.05)	0.77 (0.06)	0.008	<.001	0.012	<.001	0.004	.41	15.093 (1)	0.081
Muscle mass (kg)											
EG	17.91 (5.05)	18.60 (5.11)	18.85 (5.18)	–0.694	<.001	–0.942	<.001	–0.248	.007	37.598 (1)	0.181
CG	18.38 (4.74)	18.85 (4.48)	19.39 (4.65)	–0.469	<.001	–1.009	<.001	–0.541	<.001	26.987 (1)	0.137
Fat mass (%)											
EG	22.73 (10.23)	22.20 (9.84)	22.31 (9.90)	0.530	.07	0.425	.26	–0.105	.99	2.544 (1)	0.015
CG	20.10 (10.12)	19.79 (9.90)	19.89 (9.75)	0.314	.84	0.210	.99	–0.104	.99	0.620 (1)	0.004
VO_{2max}[a] (ml/kg/min)											
EG	38.03 (4.89)	39.03 (5.71)	38.12 (6.69)	–1.006	<.001	–0.094	.99	0.913	.005	12.772 (1)	0.077Yes
CG	38.76 (5.10)	39.39 (5.10)	39.16 (6.49)	–0.632	.06	–0.404	.89	0.228	.99	2.780 (1)	0.018
CMJ[b] test (cm)											
EG	21.82 (7.53)	23.16 (7.93)	23.19 (8.16)	–1.337	.04	–1.371	.03	–0.034	.99	4.234 (1)	0.023
CG	22.40 (7.01)	22.97 (9.26)	24.56 (8.60)	–0.572	.99	–2.163	.03	–1.591	.03	6.259 (1)	0.034
Curl-up test (repetitions, n)											
EG	20.51 (11.49)	24.31 (10.69)	24.80 (11.28)	–3.791	<.001	–4.282	<.001	–0.490	.99	22.022 (1)	0.115
CG	20.99 (11.10)	22.53 (12.26)	24.07 (11.71)	–1.540	.28	–3.073	.001	–1.533	.16	7.351 (1)	0.041
Push-up test (repetitions, n)											
EG	6.80 (9.41)	8.72 (10.95)	7.93 (10.45)	–1.920	<.001	–1.128	.01	0.793	.34	9.829 (1)	0.061
CG	7.64 (9.24)	8.62 (9.66)	8.36 (10.27)	–0.975	.28	–0.717	.44	0.258	.99	1.868 (1)	0.012

[a]VO_{2max}: maximal oxygen uptake.

[b]CMJ: countermovement jump.

Regarding physical fitness, VO_{2max} significantly increased in the EG between T1 and T2 (*P*<.001), but it significantly decreased between T2 and T3 (*P*=.005). The CG exhibited no discernible differences. The CMJ score significantly increased at T2 compared to T1 (*P*=.04) and remained elevated at T3 compared to T1 (*P*=.03). In the CG, the adolescents showed a higher score at T3 than at T1 (*P*=.003) and T2 (*P*=.03). The curl-up test showed a significant increase in the EG between T1 and T2 (*P*<.001), which remained the same at T3 (*P*<.001), while in the CG, the increase was smaller and was only observed between T3 and T1 (*P*=.001). Finally, in the push-up test, an increase was observed in the score between T1 and T2 (*P*<.001), which remained, although it was less pronounced, at T3 (*P*=.01) only in the EG (Table 2).

As shown in Multimedia Appendix 1, the covariate maturity status was a determinant factor in the differences found in the EG for the variables PA level, body mass, height, sum of 3 skinfolds, corrected girths, hips girth, waist to hip ratio, muscle mass, VO_{2max}, CMJ test, curl-up test, and push-up test between T1 and T2 (*P*<.001-.04); for the variables PA level, sum of 3 skinfolds, corrected arm girth, hips girth, muscle mass, and VO_{2max} between T2 and T3 (*P*<.001-.03); and for the variables height, corrected girths, hips girth, waist to hip ratio, muscle mass, CMJ test, curl-up test, and push-up test between T1 and T3 (*P*<.001-.04). For the CG, significant differences were observed in body mass, height, BMI, corrected girth, hips girth, waist to hip ratio, and muscle mass between T1 and T2 (*P*<.001-.02); in corrected girths, hips girth, muscle mass, and CMJ test between T2 and T3 (*P*<.001-.03); and in height, corrected girths, hips girth, waist to hip ratio, muscle mass, CMJ test, and curl-up test between T1 and T3 (*P*<.001-.01).

The effect of the covariate gender on the study variables is shown in Multimedia Appendix 2. It was a determinant factor in the differences found in the EG in PA level, body mass, height, sum of 3 skinfolds, corrected girths, hips girth, waist to hip ratio, muscle mass, VO_{2max}, CMJ test, curl-up test, and push-up test between T1 and T2 (*P*<.001-.04); in PA level, sum of 3 skinfolds, corrected arm girth, and VO_{2max} between T2 and T3 (*P*<.001-.04); and in height, corrected girths, hips girth, waist to hip ratio, muscle mass, CMJ test, curl-up test, and push-up test between T1 and T3 (*P*<.001-.01). In the CG, this covariate was a determinant factor in the differences found in body mass, height, BMI, corrected girths, hips girth, waist to hip ratio, and muscle mass between T1 and T2 (*P*<.001-.047); in BMI, corrected girths, hips girth, and muscle mass between T2 and T3 (*P*<.001-.002); and in height, corrected girths, hips girth, waist to hip ratio, muscle mass, CMJ test, and curl-up test between T1 and T3 (*P*<.001-.02).

The covariate specific app used (Multimedia Appendix 3) was not shown to be influential either in the differences found in PA level or in anthropometry variables and body composition, although it was influential in the changes in VO_{2max} (T1-T2: *P*=.004; T2-T3: *P*=.007) and the curl-up test (T1-T2: *P*<.001; T1-T3: *P*<.001).

Differences Between the EG and CG in the Study Variables at the Same Time Point During the Research Period

Table 3 shows the differences between the EG and CG in the study variables at the 3 time points (T1, T2, and T3). The differences at the 3 time points were significant in the sum of 3 skinfolds (*P*=.01-.03), in hips girth (*P*=.003-.02), and fat mass (*P*=.02-.03), as well as in the curl-up test at T2 (*P*=.047). The rest of the variables showed no significant differences between the 2 groups at any of the time points.

Table 3. Differences between the experimental group (EG) and control group (CG) at the study time points (intergroup differences).

Variable and time point	EG, mean (SD)	CG, mean (SD)	Mean difference (EG–CG)	*P* value	*F* test (*df*)	η_p^2
Subjective level of physical activity						
T1	2.62 (0.68)	2.72 (0.64)	–0.100	.20	1.664 (1)	0.005
T2	2.79 (0.59)	2.72 (0.73)	0.074	.29	1.104 (1)	0.003
T3	2.68 (0.68)	2.66 (0.71)	0.024	.75	0.749 (1)	0.001
Body mass (kg)						
T1	55.16 (12.87)	52.56 (10.84)	2.594	.05	3.812 (1)	0.011
T2	56.06 (12.69)	53.51 (10.71)	2.549	.05	3.778 (1)	0.011
T3	56.03 (11.63)	53.71 (10.72)	2.319	.06	2.805 (1)	0.010
Height (cm)						
T1	162.35 (9.04)	161.02 (8.82)	1.335	.17	1.852 (1)	0.005
T2	163.11 (8.98)	161.63 (8.77)	1.482	.13	2.313 (1)	0.007
T3	163.27 (9.74)	162.25 (9.31)	1.021	.33	0.949 (1)	0.003
BMI (kg/m^2)						
T1	20.87 (3.84)	20.19 (3.34)	0.682	.09	2.896 (1)	0.008
T2	20.98 (3.69)	20.44 (3.21)	0.538	.16	1.956 (1)	0.006
T3	20.93 (3.69)	20.26 (3.19)	0.662	.09	2.972 (1)	0.009
Sitting height (cm)						
T1	84.75 (9.59)	82.85 (12.28)	1.893	.10	2.661 (1)	0.007
T2	85.54 (4.78)	83.43 (11.18)	2.112	.07	6.001 (1)	0.017
T3	82.90 (19.43)	83.12 (15.32)	–0.215	.91	0.012 (1)	0.001
Sum of 3 skinfolds (mm)						
T1	52.03 (26.58)	45.05 (24.18)	6.979	.01	6.111 (1)	0.018
T2	50.35 (24.51)	44.44 (23.30)	5.908	.03	4.989 (1)	0.014
T3	51.50 (25.40)	45.12 (23.80)	6.377	.02	5.472 (1)	0.016
Corrected arm girth (cm)						
T1	20.83 (2.77)	20.81 (2.75)	0.012	.97	0.002 (1)	0.001
T2	21.26 (2.79)	21.20 (2.67)	0.055	.85	0.034 (1)	0.001
T3	21.48 (2.83)	21.49 (2.67)	–0.013	.97	0.002 (1)	0.001
Corrected thigh girth (cm)						
T1	39.18 (4.78)	39.43 (5.24)	–0.252	.65	0.212 (1)	0.001
T2	40.11 (4.57)	39.89 (4.22)	0.218	.66	0.200 (1)	0.001
T3	40.19 (4.73)	40.64 (4.40)	–0.443	.38	0.765 (1)	0.002
Corrected calf girth (cm)						
T1	28.95 (3.55)	28.75 (2.75)	0.200	.58	0.312 (1)	0.001
T2	29.27 (2.91)	29.28 (2.66)	–0.009	.98	0.001 (1)	0.001
T3	29.35 (2.90)	29.37 (2.68)	–0.021	.95	0.005 (1)	0.001
Waist girth (cm)						
T1	68.36 (8.84)	67.57 (7.13)	0.792	.38	0.769 (1)	0.002
T2	68.44 (8.44)	67.84 (7.24)	0.599	.50	0.464 (1)	0.001
T3	68.48 (9.05)	68.01 (7.49)	0.470	.61	0.255 (1)	0.001
Hips girth (cm)						
T1	89.24 (9.25)	86.40 (7.88)	2.840	.003	8.745 (1)	0.025

Variable and time point	EG, mean (SD)	CG, mean (SD)	Mean difference (EG–CG)	*P* value	*F* test (*df*)	η_p^2
T2	90.18 (8.74)	87.67 (7.83)	2.507	.007	7.380 (1)	0.021
T3	90.55 (8.77)	88.31 (7.96)	2.248	.02	5.835 (1)	0.017
Waist to hip ratio						
T1	0.77 (0.05)	0.78 (0.05)	–0.017	.002	9.392 (1)	0.027
T2	0.76 (0.05)	0.77 (0.05)	–0.016	.005	7.818 (1)	0.022
T3	0.76 (0.06)	0.77 (0.06)	–0.015	.02	5.808 (1)	0.017
Muscle mass (kg)						
T1	17.91 (5.05)	18.38 (4.74)	–0.473	.38	0.759 (1)	0.002
T2	18.60 (5.11)	18.85 (4.48)	–0.247	.65	0.213 (1)	0.001
T3	18.85 (5.18)	19.39 (4.65)	–0.540	.33	0.970 (1)	0.003
Fat mass (%)						
T1	22.73 (10.23)	20.10 (10.12)	2.630	.02	5.505 (1)	0.016
T2	22.20 (9.84)	19.79 (9.90)	2.414	.03	4.947 (1)	0.014
T3	22.31 (9.90)	19.89 (9.75)	2.415	.03	4.973 (1)	0.014
VO_{2max}[a] (ml/kg/min)						
T1	38.03 (4.89)	38.76 (5.10)	–0.734	.20	1.620 (1)	0.005
T2	39.03 (5.71)	39.39 (5.10)	–0.359	.57	0.320 (1)	0.001
T3	38.12 (6.69)	39.16 (6.49)	–1.044	.17	1.855 (1)	0.006
CMJ[b] test (cm)						
T1	21.82 (7.53)	22.40 (7.01)	–0.576	.47	0.529 (1)	0.001
T2	23.16 (7.93)	22.97 (9.26)	0.189	.84	0.043 (1)	0.001
T3	23.19 (8.16)	24.56 (8.60)	–1.368	.13	2.306 (1)	0.006
Curl-up test (repetitions, n)						
T1	20.51 (11.49)	20.99 (11.10)	–0.478	.70	0.146 (1)	0.001
T2	24.31 (10.69)	22.53 (12.26)	1.773	.047	2.011 (1)	0.006
T3	24.80 (11.28)	24.07 (11.71)	0.730	.56	0.334 (1)	0.001
Push-up test (repetitions, n)						
T1	6.80 (9.41)	7.64 (9.24)	–0.844	.44	0.597 (1)	0.002
T2	8.72 (10.95)	8.62 (9.66)	0.101	.93	0.007 (1)	0.001
T3	7.93 (10.45)	8.36 (10.27)	–0.433	.72	0.127 (1)	0.001

[a]VO_{2max}: maximal oxygen uptake.

[b]CMJ: countermovement jump.

Regarding the intergroup differences, it is striking that the differences found in the sum of 3 skinfolds, hips girth, and fat mass were not influenced by the covariate maturity status. In the case of the covariate gender, it could be a determinant factor in the differences found at T1 in PA level, body mass, height, sitting height, corrected calf girth, hips girth, and muscle mass (*P*=.003-.04); at T2 in body mass, height, sitting height, corrected girths, hips girth, muscle mass, VO_{2max}, CMJ test, curl-up test, and push-up test (*P*<.001-.048); and at T3 in body mass, height, corrected calf girth, and muscle mass (*P*=.01-.03). Finally, the covariate specific app used seemed to influence the differences found in the sum of 3 skinfolds, hips girth, waist to hip ratio, and fat mass at T1, T2, and T3 (*P*=.002-.04), in the curl-up test at T2 (*P*=.04) and T3 (*P*=.02), and in VO_{2max} at T3 (*P*=.02; Multimedia Appendix 4).

Table 4 shows the differences in the changes produced between the EG and CG when comparing the different time points (T1–T2, T1–T3, and T2–T3). The results showed that the changes produced in PA level (*P*=.004) and the curl-up test (*P*=.02) were significantly higher in the EG than in the CG between T1 and T2. In addition, the changes in corrected thigh girth (*P*=.003) and muscle mass (*P*=.02) between T2 and T3 were greater in the CG than in the EG. In the rest of the variables, the changes between the EG and the CG at the different time points were not significant.

Table 4. Differences in the changes produced between the experimental group (EG) and control group (CG) when comparing T1–T2, T1–T3, and T2–T3.

Variable	T1: EG–CG	T2: EG–CG	T3: EG–CG	Mean difference (T1–T2)	*P* value	Mean difference (T1–T3)	*P* value	Mean difference (T2–T3)	*P* value
Subjective level of physical activity	–0.092	0.074	0.024	–0.166	.004	–0.116	.08	0.050	.46
Body mass (kg)	2.594	2.549	2.319	0.052	.80	0.273	.52	0.221	.57
Height (cm)	1.335	1.482	1.021	–0.142	.45	0.317	.51	0.459	.31
BMI (kg/m^2)	0.682	0.538	0.662	0.146	.06	0.018	.84	–0.128	.06
Sitting height (cm)	1.893	2.112	–0.215	–0.278	.83	1.114	.20	2.278	.08
Sum of 3 skinfolds (mm)	6.979	5.908	6.377	1.072	.28	0.602	.57	–0.470	.50
Corrected arm girth (cm)	0.012	0.055	–0.013	–0.043	.62	0.025	.81	0.068	.41
Corrected thigh girth (cm)	–0.252	0.218	–0.443	–0.470	.07	0.191	.53	0.661	.003
Corrected calf girth (cm)	0.200	–0.009	–0.021	0.209	.33	0.221	.32	0.012	.89
Waist girth (cm)	0.792	0.599	0.470	0.193	.43	0.322	.32	0.129	.60
Hips girth (cm)	2.840	2.507	2.248	0.333	.20	0.593	.07	0.259	.26
Waist to hip ratio	–0.017	–0.016	–0.015	–0.001	.52	–0.002	.58	–0.001	.84
Muscle mass (kg)	–0.473	–0.247	–0.540	–0.226	.12	0.067	.71	0.292	.02
Fat mass (%)	2.630	2.414	2.415	0.216	.56	0.215	.58	–0.001	.99
VO_{2max}[a] (ml/kg/min)	–0.734	–0.359	–1.044	–0.374	.28	0.311	.53	0.685	.14
CMJ[b] test (cm)	–0.576	0.189	–1.368	–0.765	.37	0.792	.34	1.557	.06
Curl-up test (repetitions, n)	–0.478	1.773	0.730	–2.747	.02	–1.242	.26	1.505	.18
Push-up test (repetitions, n)	–0.844	0.101	–0.433	–0.685	.33	–1.000	.17	–0.314	.71

[a]VO_{2max}: maximal oxygen uptake.

[b]CMJ: countermovement jump.

Regarding the influence of the covariates on the changes found between the EG and CG at the different time points (Multimedia Appendix 5), it was observed that none of the covariates had any influence either on the changes found between T1 and T2 in PA level or on the changes in the muscle mass between T2 and T3. However, changes in the curl-up test were influenced by the covariate gender between T1 and T2 (*P*=.04), just as the changes in the corrected thigh girth were influenced by gender (*P*=.04) and specific app used (*P*=.01) between T2 and T3. The covariate maturity did not have a significant influence on any of the changes (*P*=.09-.97).

Discussion

Summary of the Main Results of the Study

The results of this research show that after the use of the step tracker mobile apps became voluntary and was no longer promoted as a PE class assignment, only a small percentage of adolescents (18/216, 8.3%) continued using them for PA, with minimal training volume. Comparing the changes in the study variables in the EG and CG (intragroup differences) before and after the mandatory and promoted period (T1 vs T2), the EG showed an increase in PA level and fitness variables, with a decrease in the sum of 3 skinfolds. However, at the end of the nonmandatory and nonpromoted period, there was a decrease in adolescents' PA level and VO_{2max}, accompanied by an increase in the sum of 3 skinfolds, compared to the values at the end of the mandatory and promoted period (T2 vs T3). Both EG and CG exhibited increases in corrected arm girth, hips girth, and muscle mass. Finally, when comparing the measurements taken before the start of the mandatory and promoted period and at the end of the nonmandatory and nonpromoted period (T1 vs T3), both groups showed significant increases in body mass, height, corrected girths, hips girth, waist to hip ratio, muscle mass, CMJ test, and curl-up test, as well as an increase in the push-up test only in the EG. These changes were influenced by the covariates maturity status and gender (both of which influenced most of the variables related to PA level, anthropometric measurements, body composition, and fitness in both groups) as well as by the covariate specific app used (which mainly influenced VO_{2max} and the curl-up test).

Regarding intergroup differences, significant differences were noted between the EG and CG at T1, T2, and T3 in sum of 3 skinfolds, hips girth, waist to hip ratio, and fat mass, and at T2 in the curl-up test. While maturity status did not affect intergroup differences, gender and specific app used did have an influence. Despite the intra- and intergroup differences found during both intervention periods, greater changes from T1 to T2 were seen in the EG, particularly in PA level and the curl-up test,

unaffected by the covariates, except for gender in the curl-up test. Conversely, from T2 to T3, the CG exhibited greater changes in corrected thigh girth, being influenced by gender and specific app used, and muscle mass, with no covariate influence. No significant changes were found between T1 and T3, with no influence of the covariates.

Purpose of This Study

The study addresses the challenge of promoting PA among adolescents due to limited school hours and motivation issues [50], which makes it impossible to meet the World Health Organization recommendations [48]. Previous research suggested that mandatory app use promoted as a PE class assignment could enhance PA, body composition, and fitness during the first week of use due to the apps' novelty [9,51]. However, it is unclear whether these effects persist when app use becomes nonmandatory and nonpromoted. For this reason, the study aims to assess whether step tracker mobile apps could encourage PA outside of school hours and establish walking as a healthy habit.

Use of Apps by Adolescents During Mandatory and Promoted and Nonmandatory and Nonpromoted Use Periods

As the results show, during the nonmandatory and nonpromoted intervention period, only a small percentage of adolescents (18/216, 8.3%) used the mobile apps; therefore, the changes achieved during the mandatory and promoted intervention period faded away. These results are similar to those of the study by Slootmaker et al [52], in which the use of wearable devices and websites led to improvements in PA after 3 months of the intervention, although the effects disappeared after 8 months when participants did not use the devices for 5 months. One possible explanation for these results is that PA during adolescence is strongly influenced by intrinsic motivation [53,54] and enjoyment experienced during PA [55]. Knowing this, it is possible that walking with an electronic device is not the most satisfying activity for adolescents; therefore, once the extrinsic component, such as the incentive of a bonus point in the PE grade, is removed, all interest shown initially in the intervention is lost. Thus, if the intention is for this population to use these apps due to the benefits they provide on PA level, body composition, and fitness [9,18,43], their use can be made mandatory and promoted by the school, at least as far as walking for exercise is concerned.

These findings suggest that interventions using step tracker mobile apps, when mandatory and promoted as a PE class assignment, may not establish an independent walking habit among adolescents. The primary goal should be to promote lasting habits because short-term increases in PA, while beneficial for fitness and body composition, lack long-term impact. Perhaps 1 of the main drawbacks that prevents adolescents from adhering to the use of these apps is their mandatory implementation by educational institutions; when their use is not promoted or rewarded, they cease to be effective [56]. This indicates that students are participating in the intervention for the reward (ego orientation), rather than for the benefits it might have on their present and future health (task orientation). Furthermore, it would be necessary to consider whether this type of intervention is less effective because it focuses only on cardiorespiratory improvement, and whether the inclusion of other types of training, such as strength or flexibility, would increase adherence [56] because they are more novel or closer to adolescents' interests.

Considering the results obtained in this study, the first research hypothesis (H1), which proposed that adolescents will stop using the step tracker mobile apps during the period of nonmandatory and nonpromoted use, can be accepted. During the period of nonpromoted and nonmandatory use, only a few of the adolescents (18/216, 8.3%) continued to use the apps, and the distance walked was minimal. This suggests that the return to baseline levels after the period of nonmandatory and nonpromoted use is due to the fact that the adolescents did not continue walking for exercise, which is why the increase in energy expenditure and improvement in physical fitness achieved during the period of mandatory and promoted use was lost during this period.

Effects of Interventions on PA Level Among Adolescents

The EG demonstrated an increase in PA level during the mandatory and promoted intervention, surpassing the CG. However, these benefits were not sustained over time. The findings align with previous research, which demonstrated that the use of mobile apps led to a notable rise in adolescents' PA level [9,17]. Notably, this study adds a new perspective, indicating that the effectiveness of step tracker mobile apps diminishes once their use is nonmandatory and nonpromoted as a PE class assignment. This highlights the importance of enforcing app use for enhancing adolescent PA. Future interventions should consider this because effectiveness may hinge on mandatory use. Future research is needed to promote the nonmandatory use of step tracker mobile apps in adolescents who have not previously used such apps because this would allow us to demonstrate whether it is the lack of mandatory use or the loss of interest in the use of the apps that leads to nonuse.

Effects of Interventions on the Kinanthropometric and Body Composition Variables Among Adolescents

Regarding the kinanthropometric and body composition variables, height and body mass increased significantly in all groups throughout the study, consistent with typical growth patterns during puberty [57]. During PHV, which typically occurs between the age of 11.4 and 12.2 years in female individuals and 13.8 and 14.4 years in male individuals, the height of female individuals and male individuals increases steadily [58,59]. Regarding body mass, previous research has shown similar results [9], and 1 possible explanation for this finding is that the body mass variable does not allow discriminating whether the change produced was due to an improvement in muscle mass or fat mass [60]. In this study, corrected girths and muscle mass also increased significantly in both groups, suggesting potential impacts of maturation rather than app-specific exercises. If the apps used had included strength exercises, it could be speculated that these exercises were the cause of the improvements in girths and muscle mass, as observed in previous research [61]. However, the exclusive use of apps meant for aerobic training makes us consider that

the changes were the consequence of the maturation process of these adolescents, characterized by hormonal changes related to increases in muscle mass [62,63]. These results are corroborated with the inclusion of the covariates maturity status and gender. The maturational state had an influence on the intragroup differences in muscle mass variables but did not influence the intergroup differences at the different time points (EG vs CG at T1, T2, or T3), which establishes the importance of the maturational state in the changes found and also agrees with previous research, which showed that limb girths increase during the maturational process [64]. In addition, it should be noted that the covariate gender also influenced the changes in muscle mass during the different time points, which could be due to the fact that during puberty, muscle development is greater in male individuals than in female individuals due to the higher production of steroid hormones in male individuals, with clear differences between both genders [62].

With respect to the fat variables, the EG experienced a significant decrease in the sum of 3 skinfolds between T1 and T2, but this change reverted to baseline levels at T3. This aligns with previous findings suggesting that the use of step tracker mobile apps can reduce fat mass in adolescents [9]. Increased PA during the mandatory and promoted period likely contributed to this reduction, increasing energy expenditure, as seen in other 10-week aerobic exercise programs in the adolescent population [65], which could have influenced the decrease in fat mass. The main novelty of this study is that when app promotion ceased, fat mass returned to initial levels, echoing findings of detraining studies [20]. This could be because the increase in PA achieved during the period of mandatory and promoted use of the app was lost when the use became nonmandatory and nonpromoted, which could have prevented the adolescents from maintaining their increased energy expenditure. In addition, maturity status and gender did not impact fat mass changes, suggesting the intervention's influence. Nevertheless, the small effect sizes and nonsignificant differences between the groups hint at inconsistency in app-induced changes, possibly explaining the return to baseline values after the intervention. Therefore, future research with the use of step tracker mobile apps for a longer period, which also considers other relevant variables for aerobic training to be effective, such as duration or intensity [66,67], is needed to elucidate their true effect on adolescent body composition.

In hips girth, a significant increase was found in both groups between T1 and T2, as well as T1 and T3. A possible explanation for these results is that hips girth steadily increases during adolescence, ending with a plateau at age 16 years [68], which could explain the increase in this variable in the 2 groups. It would be important for future research to analyze these differences according to the age or ethnicity of the adolescents because these variables significantly influence hips girth [69].

Effects of Interventions on the Physical Fitness Variables Among Adolescents

As for the physical fitness tests, a significant increase in VO_{2max} was only found in the EG between T1 and T2, with a significant decrease observed between T2 and T3. These results are similar to previous research, in which VO_{2max} increased and performance in the 20-meter shuttle run test improved after the period of mobile app use compared to the CG [9,10]. A possible explanation for these results could be that the use of the apps during the mandatory and promoted period, in which an incentive was offered, favored the improvement in the adolescents' physical fitness. However, when its use became nonmandatory and nonpromoted as a PE class assignment, and it was no longer incentivized, it is possible that the adolescents did not walk a sufficient distance or at the intensity necessary to maintain the significant improvements achieved in physical fitness, with these variables significantly influencing VO_{2max} [70]. However, future research analyzing the intensity and volume of adolescents' walks while using step tracker mobile apps is needed to provide more information in this area and to discover whether the maintenance of the benefits obtained is also dependent on these factors.

In the CMJ, curl-up, and push-up tests, it was observed that in the EG, performance improved between T1 and T2 and remained high at T3, while in the CG, the adolescents showed an improvement at T3 compared to T1, with no differences found in the push-up test. These results are similar to previous research in which improvements in the curl-up and push-up tests were found after the mandatory and promoted period of use of mobile apps [9], as well as in specific 6-week aerobic walking programs [71]. Walking has been shown to improve trunk muscle strength [72], which, together with the increases found in this study in muscle mass and corrected girths, could be the reason why the EG showed significant improvements in these fitness tests. However, maturity status, gender, and specific app used should be considered when analyzing changes in physical fitness because they seem to be relevant in the differences found, and future research is needed to learn about the differences in the benefits obtained with the mobile apps as a function of these factors.

In addition, the improvements in the curl-up test in the CG exclusively occurred at T3 compared to T1; this may be solely due to changes in muscle mass caused by the maturation process [62,63], which would be slower in the adolescents in the CG than in the adolescents in the EG who used the app and would favor the improvement in trunk musculature, explaining the absence of differences between T1 and T2 and between T2 and T3 in this group. This was observed with the inclusion of the covariate maturity status because it was shown to influence the differences in the curl-up test of the adolescents in the CG between T1 and T3. In addition, it should be noted that the changes between T1 and T2, when comparing the EG and CG, were significantly greater in the EG, which could be consistent with the fact that walking improves trunk muscle strength [72] because the covariate maturity status did not influence the differences in the changes between the 2 groups. Therefore, the use of step tracker mobile apps could be of interest for improving performance in fitness tests that require trunk strength, allowing for improvements that are superior to those achieved solely due to the maturation process.

One of the unexpected aspects of this study is that the covariate specific app used influenced the results obtained on the adolescents' body composition and fitness. In 1 of the first

studies on the subject, it was found that the step tracker specific app used had almost no influence on the benefits obtained with a compulsory PE class intervention [9]. However, this study has shown that the covariate specific app used can be a determining factor in the changes obtained; therefore, future research should analyze the causes of these differences to try to find the reasons why they occur and whether this could depend on whether adolescents feel more comfortable with a particular app and prefer using this app over another, which would open the door to research that could delve deeper into the specific components included in each app and how they are valued by adolescents.

The second research hypothesis (H2), which proposed that there will be significant differences in adolescents' PA level, body composition, and physical fitness during the mandatory and promoted period, influenced by maturity status and gender but not by specific app used, can be partially accepted. This is because the results showed benefits in the PA level, cardiorespiratory fitness, and fat variables after the mandatory and promoted intervention, although the only significant changes observed between the EG and CG were in abdominal strength and PA level between T1 and T2. Furthermore, as expected, the covariates maturity status and gender influenced the results observed in body composition and fitness, although specific app used also had an influence.

The third research hypothesis (H3), which stated that some of the benefits achieved by the adolescents during the mandatory and promoted period will be lost after the nonmandatory and nonpromoted period, with the results being influenced by maturity status and gender but not by specific app used, can be partially accepted. The benefits obtained during the mandatory and promoted period were lost when adolescents stopped using the apps during the nonmandatory and nonpromoted period. Only the benefits in muscle mass and fitness variables (CMJ, curl-up, and push-up tests) were maintained at T3 compared to T1 in the EG. However, the changes between T1 and T3, when comparing the EG and CG, were not significant in any of the variables analyzed. Again, the covariates maturity status, gender, and specific app used influenced the results.

Limitations of This Study

This study is not without limitations. First, the measurement of PA using a questionnaire has limitations; for example, although some studies show that the PAQ-A can be valid for measuring changes in PA performed by the same group in 2 different time periods [30], others indicate that its validity and reliability are questionable because it does not measure aspects such as frequency or intensity of activity, and nor does it allow for comparisons of changes in PA between 2 groups [73]. Therefore, although the PAQ-A is sometimes the best choice for measuring PA due to its favorable cost-benefit ratio [74,75], its use could affect the results of the study. Future research should include accelerometry, which would also make it possible to differentiate the intensities and the time spent in each activity level, which would make it possible to analyze whether these factors are modulators of change. Second, in future research, it would be appropriate to consider aspects such as the volume and intensity of training performed with step tracker mobile apps because these aspects can influence changes in body composition and fitness, as well as the length of time for which the changes achieved are maintained [76,77]. Third, the nutritional intake of adolescents was not considered in this study, and although previous research has shown that aspects of nutrition, such as adherence to the Mediterranean diet, are not modified with the use of step tracker mobile apps [9], it would be important to consider the amount and type of nutritional intake because these variables can influence changes in body composition [78]. Fourth, only those students who completed at least 25% of the training volume were included in the analyses. This is because previous research has shown that this is the minimum volume needed for differences in adolescent body composition and fitness to begin to occur [43]. In this study, when trying to analyze whether adolescents continued to use the app when it was neither mandatory nor promoted, it was assumed that those who did not use the apps during the mandatory and promoted period would not use them in the nonmandatory and nonpromoted period either. Therefore, this may be a bias because adolescents who did not even start the intervention were eliminated from the analysis, but this was not the aim of the research. Finally, another noteworthy aspect to be considered could be that this type of intervention does not consider the context or the environment of the adolescents; while it places the adolescent at the center of the intervention, it does not consider that other agents in their close environment (family and friends) could be of great relevance for the acquisition of healthy habits [79,80]. Furthermore, another unexplored aspect that should be considered is that at these ages (12-16 y), adolescents practice physical sports activities mainly for their competitive and recreational component [53-55], not for their health benefits, and walking may not be the most motivating and fun activity for this population, which may hinder the establishment of this healthy habit when the performance of these tasks is optional for students, although improvements in health are evident.

Practical Applications

Considering the limitations of this research, a practical application derived from it is that if step tracker mobile apps are intended to be used to increase daily steps and PA level of the adolescent population and are promoted as a PE class assignment, their use must be maintained over time or accompanied by other types of complementary programs that promote the establishment of healthy lifestyle habits [81] that allow the improvements obtained to be maintained. In this regard, previous research has shown that the use of step tracker mobile apps, combined with nutritional programs or training sessions on healthy habits, has a significant effect on improving the health status of adolescents by facilitating their continued use [81,82]. On the contrary, the occasional use of these apps is ineffective; once the period of mandatory and promoted use ends, adolescents who used the apps will return to baseline levels that are similar to those of adolescents who did not use them as a result of detraining, as observed in previous PA promotion programs where young people who undertook aerobic training showed losses in the benefits obtained after a period of detraining [20].

Conclusions

This study is the first to analyze the losses that occur in the positive changes achieved in PA level, body composition, and fitness variables by adolescents aged 12 to 16 years after a period of mandatory and promoted use of step tracker mobile apps when their use becomes nonmandatory and nonpromoted. Our findings suggest that adolescents stop using step tracker mobile apps when their use is neither mandatory nor promoted as a PE class assignment. During the mandatory and promoted period, adolescents in the EG increased their PA level and cardiorespiratory fitness and reduced their fat mass. However, when participation in the step increase program with the apps was neither mandatory nor promoted, adolescents stopped using the apps and discontinued their walking practice; as a consequence, the gains achieved were lost, leading to a regression to baseline levels. The change in the EG compared to the CG was only significant in the curl-up test at the end of the mandatory and promoted intervention but not in the rest of the variables or in the comparison between the other time points. Therefore, this study shows that the use of mobile apps by adolescents aged 12 to 16 years in their free time, promoted as a PE class assignment, did not manage to create a healthy walking habit in this population, which could be a determining factor in fostering independent walking practice in adolescents, which could yield significant health benefits. These results are of relevance for the use of step tracker mobile apps in education because when their use is mandated and promoted as a PE class assignment, they seem effective in reducing fat mass and increasing PA level in adolescents. Special attention should be paid to the covariates maturity status, gender, and specific app used because they might influence the changes achieved during the intervention. Future research along these lines should also analyze the most influential aspects to be considered to achieve independent use and greater adherence of adolescents to step tracker mobile apps, including analyses of their immediate environment and the factors that could be most relevant, to ensure that use is continued after they have already become familiar with the step tracker mobile app, and its use is neither mandatory nor promoted as a PE class assignment.

Acknowledgments

The authors would like to thank the schools, management teams, teachers, students, and research assistants who made it possible to carry out the measurements for this project. This work is part of the doctoral thesis of AMO. AMO's participation in this research is funded by Séneca Foundation – 21409/FPI/20. Fundación Séneca. Región de Murcia (Spain). This manuscript was developed during AMO's international stay at the CIAFEL research group of the Faculdade de Desporto da Universidade de Porto. This research has been financed by grants for knowledge generation projects (Title: fomento de la actividad física diaria recomendada para la salud en adolescentes mediante apps móviles, wearables y una tac gamificada / Call: 2022 / Reference code: PID2022-140245OA-I00).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Effect of the covariate maturity status in the intragroup (T1 vs T2; T1 vs T2 and T2 vs T3) differences.
[DOCX File , 25 KB - mhealth_v12i1e51206_app1.docx]

Multimedia Appendix 2
Effect of the covariate gender in the intragroup (T1 vs T2; T1 vs T2 and T2 vs T3) differences.
[DOCX File , 26 KB - mhealth_v12i1e51206_app2.docx]

Multimedia Appendix 3
Effect of the covariate specific app used in the intragroup (T1 vs T2; T1 vs T2 and T2 vs T3) differences in the experimental group.
[DOCX File , 20 KB - mhealth_v12i1e51206_app3.docx]

Multimedia Appendix 4
Effect of the covariates maturity status, gender, and app used in the intergroup (experimental group vs control group) differences.
[DOCX File , 32 KB - mhealth_v12i1e51206_app4.docx]

Multimedia Appendix 5
Effect of the covariates maturity status, gender, and app used in the differences in the changes produced between the experimental and control groups.
[DOCX File , 24 KB - mhealth_v12i1e51206_app5.docx]

Multimedia Appendix 6
CONSORT e-HEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1184 KB - mhealth_v12i1e51206_app6.pdf]

References

1. Pechtl SM, Kim LP, Jacobsen KH. Physical inactivity and sedentariness: languorous behavior among adolescents in 80 countries. J Adolesc Health 2022 Jun;70(6):950-960. [doi: 10.1016/j.jadohealth.2021.12.017] [Medline: 35177343]
2. Canabrava KL, Amorim PR, Miranda VP, Priore SE, Franceschini SD. Sedentary behavior and cardiovascular risk in children: a systematic review. Rev Bras Med Esporte 2019 Oct;25(5):433. [doi: 10.1590/1517-869220192505168868]
3. Sousa AC, Ferrinho SN, Travassos B. The use of wearable technologies in the assessment of physical activity in preschool- and school-age youth: systematic review. Int J Environ Res Public Health 2023 Feb 15;20(4):3402 [FREE Full text] [doi: 10.3390/ijerph20043402] [Medline: 36834100]
4. Williams WM, Ayres CG. Can active video games improve physical activity in adolescents? A review of RCT. Int J Environ Res Public Health 2020 Jan 20;17(2):669 [FREE Full text] [doi: 10.3390/ijerph17020669] [Medline: 31968642]
5. Böhm B, Karwiese SD, Böhm H, Oberhoffer R. Effects of mobile health including wearable activity trackers to increase physical activity outcomes among healthy children and adolescents: systematic review. JMIR Mhealth Uhealth 2019 Apr 30;7(4):e8298 [FREE Full text] [doi: 10.2196/mhealth.8298] [Medline: 31038460]
6. Serra G, Lo Scalzo L, Giuffrè M, Ferrara P, Corsello G. Smartphone use and addiction during the coronavirus disease 2019 (COVID-19) pandemic: cohort study on 184 Italian children and adolescents. Ital J Pediatr 2021 Jul 02;47(1):150 [FREE Full text] [doi: 10.1186/s13052-021-01102-8] [Medline: 34215311]
7. Lee AM, Chavez S, Bian J, Thompson LA, Gurka MJ, Williamson VG, et al. Efficacy and effectiveness of mobile health technologies for facilitating physical activity in adolescents: scoping review. JMIR Mhealth Uhealth 2019 Feb 12;7(2):e11847 [FREE Full text] [doi: 10.2196/11847] [Medline: 30747716]
8. He Z, Wu H, Yu F, Fu J, Sun S, Huang T, et al. Effects of smartphone-based interventions on physical activity in children and adolescents: systematic review and meta-analysis. JMIR Mhealth Uhealth 2021 Feb 01;9(2):e22601 [FREE Full text] [doi: 10.2196/22601] [Medline: 33522980]
9. Mateo-Orcajada A, Abenza-Cano L, Albaladejo-Saura MD, Vaquero-Cristóbal R. Mandatory after-school use of step tracker apps improves physical activity, body composition and fitness of adolescents. Educ Inf Technol (Dordr) 2023 Jan 24:1-32 [FREE Full text] [doi: 10.1007/s10639-023-11584-0] [Medline: 36714445]
10. Mora-Gonzalez J, Pérez-López IJ, Delgado-Fernández M. The "$in time" gamification project: using a mobile app to improve cardiorespiratory fitness levels of college students. Games Health J 2020 Feb;9(1):37-44 [FREE Full text] [doi: 10.1089/g4h.2019.0001] [Medline: 31545103]
11. Rodríguez G, Moreno LA, Blay MG, Blay VA, Garagorri JM, Sarría A, et al. Body composition in adolescents: measurements and metabolic aspects. Int J Obes Relat Metab Disord 2004 Nov 15;28 Suppl 3(S3):S54-S58. [doi: 10.1038/sj.ijo.0802805] [Medline: 15543220]
12. Ortega FB, Ruiz JR, Castillo MJ, Sjöström M. Physical fitness in childhood and adolescence: a powerful marker of health. Int J Obes (Lond) 2008 Jan 4;32(1):1-11. [doi: 10.1038/sj.ijo.0803774] [Medline: 18043605]
13. Simmonds M, Llewellyn A, Owen CG, Woolacott N. Predicting adult obesity from childhood obesity: a systematic review and meta-analysis. Obes Rev 2016 Feb 23;17(2):95-107 [FREE Full text] [doi: 10.1111/obr.12334] [Medline: 26696565]
14. Minghelli B, Oliveira R, Nunes C. Association of obesity with chronic disease and musculoskeletal factors. Rev Assoc Med Bras (1992) 2015 Aug;61(4):347-354 [FREE Full text] [doi: 10.1590/1806-9282.61.04.347] [Medline: 26466217]
15. Going SB, Lohman TG, Cussler EC, Williams DP, Morrison JA, Horn PS. Percent body fat and chronic disease risk factors in U.S. children and youth. Am J Prev Med 2011 Oct;41(4 Suppl 2):S77-S86. [doi: 10.1016/j.amepre.2011.07.006] [Medline: 21961616]
16. Lang JJ, Larouche R, Tremblay MS. The association between physical fitness and health in a nationally representative sample of Canadian children and youth aged 6 to 17 years. Health Promot Chronic Dis Prev Can 2019 Mar;39(3):104-111 [FREE Full text] [doi: 10.24095/hpcdp.39.3.02] [Medline: 30869473]
17. Gil-Espinosa FJ, Merino-Marbán R, Mayorga-Vega D. Aplicación móvil endomondo para promocionar la actividad física en estudiantes de educación secundaria. Cultura Ciencia y Deporte 2020;15(46):465-473 [FREE Full text] [doi: 10.12800/ccd.v15i46.1597]
18. Mateo-Orcajada A, Vaquero-Cristóbal R, Abenza-Cano L. Gender and academic year as moderators of the efficacy of mobile app interventions to promote physical activity in adolescents: a randomized controlled trial. Humanit Soc Sci Commun 2023 Dec 22;10:980 [FREE Full text] [doi: 10.1057/s41599-023-02502-3]
19. van de Kop JH, van Kernebeek WG, Otten RH, Toussaint HM, Verhoeff AP. School-based physical activity interventions in prevocational adolescents: a systematic review and meta-analyses. J Adolesc Health 2019 Aug;65(2):185-194 [FREE Full text] [doi: 10.1016/j.jadohealth.2019.02.022] [Medline: 31202623]
20. Lo MS, Lin LL, Yao WJ, Ma MC. Training and detraining effects of the resistance vs. endurance program on body composition, body size, and physical performance in young men. J Strength Cond Res 2011 Aug;25(8):2246-2254. [doi: 10.1519/JSC.0b013e3181e8a4be] [Medline: 21747300]

21. Merino-Marban R, Mayorga-Vega D, Fernandez-Rodriguez E, Estrada FV, Viciana J. Effect of a physical education-based stretching programme on sit-and-reach score and its posterior reduction in elementary schoolchildren. Eur Phys Educ Rev 2014 Sep 25;21(1):83-92. [doi: 10.1177/1356336x14550942]
22. Malina RM, Bouchard C. Growth, Maturation, and Physical Activity. Champaign, IL: Human Kinetics; 1991.
23. Albaladejo-Saura M, Vaquero-Cristóbal R, García-Roca JA, Esparza-Ros F. The effect of age, biological maturation and birth quartile in the kinanthropometric and physical fitness differences between male and female adolescent volleyball players. Children (Basel) 2022 Jan 04;9(1):58 [FREE Full text] [doi: 10.3390/children9010058] [Medline: 35053683]
24. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. Int J Surg 2011;9(8):672-677 [FREE Full text] [doi: 10.1016/j.ijsu.2011.09.004] [Medline: 22019563]
25. Kadam P, Bhalerao S. Sample size calculation. Int J Ayurveda Res 2010 Jan;1(1):55-57 [FREE Full text] [doi: 10.4103/0974-7788.59946] [Medline: 20532100]
26. Gauvin-Lepage J, Friedman D, Grilli L, Sufrategui M, De Matteo C, Iverson GL, et al. Effectiveness of an exercise-based active rehabilitation intervention for youth who are slow to recover after concussion. Clin J Sport Med 2020 Sep;30(5):423-432. [doi: 10.1097/JSM.0000000000000634] [Medline: 30095507]
27. Christie J, O'Halloran P, Stevenson M. Planning a cluster randomized controlled trial: methodological issues. Nurs Res 2009;58(2):128-134. [doi: 10.1097/NNR.0b013e3181900cb5] [Medline: 19289934]
28. Egilsson E, Bjarnason R, Njardvik U. Usage and weekly attrition in a smartphone-based health behavior intervention for adolescents: pilot randomized controlled trial. JMIR Form Res 2021 Feb 17;5(2):e21432 [FREE Full text] [doi: 10.2196/21432] [Medline: 33481750]
29. Kowalski KC, Crocker PR, Kowalski NP. Convergent validity of the physical activity questionnaire for adolescents. Pediatr Exerc Sci 1997;9(4):342-352. [doi: 10.1123/pes.9.4.342]
30. Martínez-Gómez D, Martínez-de-Haro V, Pozo T, Welk GJ, Villagra A, Calle ME, et al. Reliability and validity of the PAQ-A questionnaire to assess physical activity in Spanish adolescents. Rev Esp Salud Publica 2009 Jun;83(3):427-439 [FREE Full text] [doi: 10.1590/s1135-57272009000300008] [Medline: 19701574]
31. Esparza-Ros F, Vaquero-Cristóbal R, Marfell-Jones M. International Standars for Anthropometric Assessment. Murcia, Spain: International Society for the Advancement of Kinanthropometry (ISAK); 2019.
32. Poortmans JR, Boisseau N, Moraine JJ, Moreno-Reyes R, Goldman S. Estimation of total-body skeletal muscle mass in children and adolescents. Med Sci Sports Exerc 2005 Feb;37(2):316-322. [doi: 10.1249/01.mss.0000152804.93039.ce] [Medline: 15692329]
33. Slaughter MH, Lohman TG, Boileau RA, Horswill CA, Stillman RJ, Van Loan MD, et al. Skinfold equations for estimation of body fatness in children and youth. Hum Biol 1988 Oct;60(5):709-723. [Medline: 3224965]
34. Esparza-Ros F, Vaquero-Cristóbal R. Antropometría: Fundamentos Para la Aplicación e Interpretación. Madrid, Spain: Aula Magna Key Project McGraw Hill; 2023.
35. Albaladejo-Saura M, Vaquero-Cristóbal R, García-Roca JA, Esparza-Ros F. Influence of biological maturation status on selected anthropometric and physical fitness variables in adolescent male volleyball players. PeerJ 2022;10:e13216 [FREE Full text] [doi: 10.7717/peerj.13216] [Medline: 35402095]
36. Cruz JR, Armesilla MD, Herrero-de-Lucas A, Riaza LM, Pascual CM, Manzañido JP, et al. Body composition assessment in sports medicine. Statement of Spanish group of kinanthropometry of Spanish Federation of Sports Medicine. Version 2010. Archivos de Medicina Del Deporte 2010;27(139):330-344 [FREE Full text]
37. Mirwald RL, G. Baxter-Jones AD, Bailey DA, Beunen GP. An assessment of maturity from anthropometric measurements. Med Sci Sport Exerc 2002;34(4):689-694. [doi: 10.1249/00005768-200204000-00020]
38. Léger LA, Mercier D, Gadoury C, Lambert J. The multistage 20 metre shuttle run test for aerobic fitness. J Sports Sci 1988;6(2):93-101. [doi: 10.1080/02640418808729800] [Medline: 3184250]
39. Tomkinson GR, Lang JJ, Blanchard J, Léger LA, Tremblay MS. The 20-m shuttle run: assessment and interpretation of data in relation to youth aerobic fitness and health. Pediatr Exerc Sci 2019 May 01;31(2):152-163. [doi: 10.1123/pes.2018-0179] [Medline: 30885058]
40. Barker LA, Harry JR, Mercer JA. Relationships between countermovement jump ground reaction forces and jump height, reactive strength index, and jump time. J Strength Cond Res 2018 Jan;32(1):248-254. [doi: 10.1519/JSC.0000000000002160] [Medline: 28746248]
41. Garcia-Pastor T, Salinero JJ, Sanz-Frias D, Pertusa G, Del Coso J. Body fat percentage is more associated with low physical fitness than with sedentarism and diet in male and female adolescents. Physiol Behav 2016 Oct 15;165:166-172. [doi: 10.1016/j.physbeh.2016.07.016] [Medline: 27474415]
42. Castro-Piñero J, Ortega FB, Artero EG, Girela-Rejón MJ, Mora J, Sjöström M, et al. Assessing muscular strength in youth: usefulness of standing long jump as a general index of muscular fitness. J Strength Cond Res 2010 Jul;24(7):1810-1817. [doi: 10.1519/JSC.0b013e3181ddb03d] [Medline: 20555277]
43. Mateo-Orcajada A, Vaquero-Cristóbal R, Abenza-Cano L. Importance of training volume through the use of step trackers apps promoted from the subject of physical education to change body composition, physical fitness and physical activity

in adolescents and the influence of gender. Physiol Behav 2024 Jan 01;273:114402 [FREE Full text] [doi: 10.1016/j.physbeh.2023.114402] [Medline: 37935300]
44. Coburn JW, Malek MH. Manual NSCA: Fundamentos del Entrenamiento Personal. Barcelona, Spain: Paidotribo; 2017.
45. Bondaronek P, Alkhaldi G, Slee A, Hamilton FL, Murray E. Quality of publicly available physical activity apps: review and content analysis. JMIR Mhealth Uhealth 2018 Mar 21;6(3):e53 [FREE Full text] [doi: 10.2196/mhealth.9069] [Medline: 29563080]
46. Morency C, Demers M, Lapierre L. How many steps do you have in reserve?: thoughts and measures about a healthier way to travel. Transp Res Rec 2007 Jan 01;2002(1):1-6. [doi: 10.3141/2002-01]
47. Lubans DR, Plotnikoff RC, Miller A, Scott JJ, Thompson D, Tudor-Locke C. Using pedometers for measuring and increasing physical activity in children and adolescents. Am J Lifestyle Med 2014 Jul 10;9(6):418-427. [doi: 10.1177/1559827614537774]
48. Chaput JP, Willumsen J, Bull F, Chou R, Ekelund U, Firth J, et al. 2020 WHO guidelines on physical activity and sedentary behaviour for children and adolescents aged 5-17 years: summary of the evidence. Int J Behav Nutr Phys Act 2020 Nov 26;17(1):141 [FREE Full text] [doi: 10.1186/s12966-020-01037-z] [Medline: 33239009]
49. Hopkins WG, Marshall SW, Batterham AM, Hanin J. Progressive statistics for studies in sports medicine and exercise science. Med Sci Sports Exerc 2009 Jan;41(1):3-13. [doi: 10.1249/MSS.0b013e31818cb278] [Medline: 19092709]
50. Filkina OM, Kocherova OY, Malyshkina AI, Vorobyeva EA, Dolotova NV. The awareness and attitude of adolescents concerning healthy life-style. Probl Sotsialnoi Gig Zdravookhranenniiai Istor Med 2022 Jan;30(1):33-38. [doi: 10.32687/0869-866X-2022-30-1-33-38] [Medline: 35157377]
51. Mateo-Orcajada A, Vaquero-Cristóbal R, Abenza-Cano L. Mobile application interventions to increase physical activity and their effect on kinanthropometrics, body composition and fitness variables in adolescent aged 12-16 years old: an umbrella review. Child Care Health Dev 2024 Jan;50(1):e13146. [doi: 10.1111/cch.13146] [Medline: 37387258]
52. Slootmaker SM, Chinapaw MJ, Seidell JC, van Mechelen W, Schuit AJ. Accelerometers and internet for physical activity promotion in youth? Feasibility and effectiveness of a minimal intervention [ISRCTN93896459]. Prev Med 2010 Jul;51(1):31-36. [doi: 10.1016/j.ypmed.2010.03.015] [Medline: 20380847]
53. Nogg KA, Vaughn AA, Levy SS, Blashill AJ. Motivation for physical activity among U.S. adolescents: a self-determination theory perspective. Ann Behav Med 2021 Mar 16;55(2):133-143. [doi: 10.1093/abm/kaaa037] [Medline: 32756874]
54. Kalajas-Tilga H, Koka A, Hein V, Tilga H, Raudsepp L. Motivational processes in physical education and objectively measured physical activity among adolescents. J Sport Health Sci 2020 Sep;9(5):462-471 [FREE Full text] [doi: 10.1016/j.jshs.2019.06.001] [Medline: 32928449]
55. Martins J, Costa J, Sarmento H, Marques A, Farias C, Onofre M, et al. Adolescents' perspectives on the barriers and facilitators of physical activity: an updated systematic review of qualitative studies. Int J Environ Res Public Health 2021 May 06;18(9):4954 [FREE Full text] [doi: 10.3390/ijerph18094954] [Medline: 34066596]
56. Mateo-Orcajada A. Promoting aerobic physical activity of adolescents aged 12 to 16 years old during after-school hours using mobile applications: a critical review. Cultura Ciencia y Deporte 2024 Mar 27;19(59):3-15. [doi: 10.12800/ccd.v19i59.2169]
57. Chae HW, Suh I, Kwon AR, Kim YJ, Kim YH, Kang DR, et al. Longitudinal standards for height and height velocity in Korean children and adolescents: the Kangwha study. [corrected]. J Korean Med Sci 2013 Oct;28(10):1512-1517 [FREE Full text] [doi: 10.3346/jkms.2013.28.10.1512] [Medline: 24133358]
58. Beunen G, Malina RM. Growth and physical performance relative to the timing of the adolescent spurt. Exerc Sport Sci Rev 1988;16:503-540. [Medline: 3292266]
59. Beunen GP, Rogol AD, Malina RM. Indicators of biological maturation and secular changes in biological maturation. Food Nutr Bull 2006 Dec;27(4 Suppl Growth Standard):S244-S256. [doi: 10.1177/15648265060274S508] [Medline: 17361661]
60. Micozzi MS, Albanes D. Three limitations of the body mass index. Am J Clin Nutr 1988 Sep;48(3):691-692. [doi: 10.1093/ajcn/48.3.691] [Medline: 3414583]
61. Maestroni L, Read P, Bishop C, Papadopoulos K, Suchomel TJ, Comfort P, et al. The benefits of strength training on musculoskeletal system health: practical applications for interdisciplinary care. Sports Med 2020 Aug;50(8):1431-1450. [doi: 10.1007/s40279-020-01309-5] [Medline: 32564299]
62. Clark PA, Rogol AD. Growth hormones and sex steroid interactions at puberty. Endocrinol Metab Clin North Am 1996 Sep;25(3):665-681. [doi: 10.1016/s0889-8529(05)70346-7] [Medline: 8879992]
63. Handelsman DJ, Hirschberg AL, Bermon S. Circulating testosterone as the hormonal basis of sex differences in athletic performance. Endocr Rev 2018 Oct 01;39(5):803-829 [FREE Full text] [doi: 10.1210/er.2018-00020] [Medline: 30010735]
64. Kryst L, Woronkowicz A, Kowal M, Sobiecki J. Intergenerational changes in limb circumferences in children and adolescents aged 3-18 from Kraków (Poland) from 1983 to 2010. Am J Hum Biol 2018 Sep;30(5):e23165. [doi: 10.1002/ajhb.23165] [Medline: 30129257]
65. Thivel D, Chaput JP, Adamo KB, Goldfield GS. Is energy intake altered by a 10-week aerobic exercise intervention in obese adolescents? Physiol Behav 2014 Aug;135:130-134. [doi: 10.1016/j.physbeh.2014.06.013] [Medline: 24952265]

66. Kelley GA, Kelley KS, Pate RR. Exercise and adiposity in overweight and obese children and adolescents: a systematic review with network meta-analysis of randomised trials. BMJ Open 2019 Nov 11;9(11):e031220 [FREE Full text] [doi: 10.1136/bmjopen-2019-031220] [Medline: 31719081]
67. Alberga AS, Frappier A, Sigal RJ, Prud'homme D, Kenny GP. A review of randomized controlled trials of aerobic exercise training on fitness and cardiometabolic risk factors in obese adolescents. Phys Sportsmed 2013 May;41(2):44-57. [doi: 10.3810/psm.2013.05.2014] [Medline: 23703517]
68. Rönnecke E, Vogel M, Bussler S, Grafe N, Jurkutat A, Schlingmann M, et al. Age- and sex-related percentiles of skinfold thickness, waist and hip circumference, waist-to-hip ratio and waist-to-height ratio: results from a population-based pediatric cohort in Germany (LIFE Child). Obes Facts 2019;12(1):25-39 [FREE Full text] [doi: 10.1159/000494767] [Medline: 30673672]
69. Hammer LD, Wilson DM, Litt IF, Killen JD, Hayward C, Miner B, et al. Impact of pubertal development on body fat distribution among White, Hispanic, and Asian female adolescents. J Pediatr 1991 Jun;118(6):975-980 [FREE Full text] [doi: 10.1016/s0022-3476(05)82223-1] [Medline: 2040937]
70. Scribbans TD, Vecsey S, Hankinson PB, Foster WS, Gurd BJ. The effect of training intensity on VO2max in young healthy adults: a meta-regression and meta-analysis. Int J Exerc Sci 2016 Apr 1;9(2):230-247 [FREE Full text] [Medline: 27182424]
71. Shnayderman I, Katz-Leurer M. An aerobic walking programme versus muscle strengthening programme for chronic low back pain: a randomized controlled trial. Clin Rehabil 2013 Mar;27(3):207-214. [doi: 10.1177/0269215512453353] [Medline: 22850802]
72. Lamoth CJ, Meijer OG, Daffertshofer A, Wuisman PI, Beek PJ. Effects of chronic low back pain on trunk coordination and back muscle activity during walking: changes in motor control. Eur Spine J 2006 Jan;15(1):23-40 [FREE Full text] [doi: 10.1007/s00586-004-0825-y] [Medline: 15864670]
73. Loney T, Standage M, Thompson D, Sebire SJ, Cumming S. Self-report vs. objectively assessed physical activity: which is right for public health? J Phys Act Health 2011 Jan;8(1):62-70. [doi: 10.1123/jpah.8.1.62] [Medline: 21297186]
74. Ainsworth B, Cahalin L, Buman M, Ross R. The current state of physical activity assessment tools. Prog Cardiovasc Dis 2015;57(4):387-395. [doi: 10.1016/j.pcad.2014.10.005] [Medline: 25446555]
75. Nigg CR, Fuchs R, Gerber M, Jekauc D, Koch T, Krell-Roesch J, et al. Assessing physical activity through questionnaires – a consensus of best practices and future directions. Psychol Sport Exerc 2020 Sep;50:101715. [doi: 10.1016/j.psychsport.2020.101715]
76. Albaladejo-Saura M, Vaquero-Cristóbal R, González-Gálvez N, Esparza-Ros F. Relationship between biological maturation, physical fitness, and kinanthropometric variables of young athletes: a systematic review and meta-analysis. Int J Environ Res Public Health 2021 Jan 05;18(1):328 [FREE Full text] [doi: 10.3390/ijerph18010328] [Medline: 33466291]
77. Skinner AM, Vlachopoulos D, Barker AR, Moore SA, Rowlands AV, Soininen S, et al. Physical activity volume and intensity distribution in relation to bone, lean and fat mass in children. Scand J Med Sci Sports 2023 Mar;33(3):267-282 [FREE Full text] [doi: 10.1111/sms.14255] [Medline: 36326758]
78. Del'Arco AP, Previdelli AN, Ferrari G, Fisberg M. Food intake, physical activity and body composition of adolescents and young adults: data from Brazilian Study of Nutrition and Health. BMC Public Health 2021 Jun 12;21(1):1123 [FREE Full text] [doi: 10.1186/s12889-021-11171-3] [Medline: 34118909]
79. Khan SR, Uddin R, Mandic S, Khan A. Parental and peer support are associated with physical activity in adolescents: evidence from 74 countries. Int J Environ Res Public Health 2020 Jun 20;17(12):4435 [FREE Full text] [doi: 10.3390/ijerph17124435] [Medline: 32575699]
80. Laddu D, Paluch AE, LaMonte MJ. The role of the built environment in promoting movement and physical activity across the lifespan: implications for public health. Prog Cardiovasc Dis 2021 Jan;64:33-40. [doi: 10.1016/j.pcad.2020.12.009] [Medline: 33428966]
81. Martínez-Olcina M, Cuestas-Calero BJ, Miralles-Amorós L, Vicente-Martínez M, Sánchez-Sánchez J. Effectiveness of app-based intervention to improve health status of sedentary middle-aged males and females. Int J Environ Res Public Health 2022 May 11;19(10):5857 [FREE Full text] [doi: 10.3390/ijerph19105857] [Medline: 35627392]
82. Schoeppe S, Alley S, Van Lippevelde W, Bray NA, Williams SL, Duncan MJ, et al. Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: a systematic review. Int J Behav Nutr Phys Act 2016 Dec 07;13(1):127 [FREE Full text] [doi: 10.1186/s12966-016-0454-y] [Medline: 27927218]

Abbreviations

CG: control group
CMJ: countermovement jump
CONSORT: Consolidated Standards of Reporting Trials
EG: experimental group
PA: physical activity
PAQ-A: Physical Activity Questionnaire for Adolescents
PE: physical education

PHV: peak height velocity
VO2max: maximal oxygen uptake

Edited by L Buis; submitted 25.07.23; peer-reviewed by N Tottori, S Bonn; comments to author 23.01.24; revised version received 20.02.24; accepted 05.06.24; published 30.07.24.

Please cite as:
Mateo-Orcajada A, Vaquero-Cristóbal R, Mota J, Abenza-Cano L
Physical Activity, Body Composition, and Fitness Variables in Adolescents After Periods of Mandatory, Promoted or Nonmandatory, Nonpromoted Use of Step Tracker Mobile Apps: Randomized Controlled Trial
JMIR Mhealth Uhealth 2024;12:e51206
URL: https://mhealth.jmir.org/2024/1/e51206
doi:10.2196/51206
PMID:

Effects of a Planned Web-Based Educational Intervention Based on the Health Belief Model for Patients With Ischemic Stroke in Promoting Secondary Prevention During the COVID-19 Lockdown in China: Quasi-Experimental Study

Zhuo Liu[1*], PhD; Xin Sun[1*], PhD; Zhen-Ni Guo[1,2], PhD; Ye Sun[1], BD; Yi Yang[1], MD, PhD; Xiuli Yan[1], BD

[1]Stroke Center, Department of Neurology, The First Hospital of Jilin University, Changchun, China

[2]Neuroscience Research Center, Department of Neurology, The First Hospital of Jilin University, Changchun, China

[*]these authors contributed equally

Corresponding Author:
Xiuli Yan, BD
Stroke Center, Department of Neurology, The First Hospital of Jilin University, , Changchun, , China

Related Article:

This is a corrected version. See correction statement: https://mhealth.jmir.org/2024/1/e60953

Abstract

Background: Some common modified vascular risk factors remain poorly controlled among stroke survivors, and educational programs may help improve these conditions.

Objective: This study aimed to evaluate the effect of a planned web-based educational intervention based on the health belief model (HBM) in promoting secondary prevention among patients with ischemic stroke.

Methods: An evaluation-blinded quasi-experimental trial with a historical control group was conducted. Patients admitted from March to June 2020 were assigned to the historical control group, and patients admitted from July to October 2020 were assigned to the intervention group. The control group received routine health management. The intervention group received 6 additional sessions based on the HBM via Tencent Meeting, an audio and video conferencing application, within 3 months after discharge. Sessions were held every 2 weeks, with each session lasting approximately 40 minutes. These sessions were conducted in small groups, with about 8 to 10 people in each group. The primary outcomes were changes in blood pressure (BP), low-density lipoprotein cholesterol (LDL-C), hemoglobin A_{1c} (HbA_{1c}), and the proportion of patients achieving the treatment target. The secondary outcomes were medication adherence, assessed with the Morisky Medicine Adherence Scale (MMAS), and disability, assessed with the modified Rankin scale.

Results: In total, 315 patients experiencing their first-ever stroke were analyzed. More patients in the intervention group had controlled BP (41.9% vs 28.4%; adjusted odds ratio [aOR] 1.93; *P*=.01), LDL-C (83.1% vs 67.7%; aOR 2.66; *P*=.001), and HbA_{1c} (91.9% vs 83.9%; aOR: 3.37; *P*=.04) levels as well as a significant postintervention decrease in the systolic BP (adjusted β −3.94; *P*=.02), LDL-C (adjusted β −0.21; *P*=.008), and HbA_{1c} (adjusted β −0.27; *P*<.001), compared with control groups. Significant between-group differences were observed in medication adherence (79.4% vs 63.2%; aOR 2.31; *P*=.002) but not in favorable functional outcomes.

Conclusions: A web-based education program based on the HBM may be more effective than current methods used to educate patients having strokes on optimal vascular risk factors and medication adherence.

Trial Registration: Chinese Clinical Trial Registry ChiCTR2000040804; https://www.chictr.org.cn/showproj.html?proj=62431

(JMIR Mhealth Uhealth 2024;12:e44463) doi:10.2196/44463

KEYWORDS

health belief model; health education; secondary prevention; stroke; medication adherence; patient education; web-based education; digital intervention; promotion; stroke patients; ischemic; prevention; quasi-experimental study; education; control group; health management; management; systolic blood pressure; blood pressure; effectiveness; medication adherence

Introduction

According to the Global Burden of Disease Study 2019, stroke is the second leading cause of death worldwide [1]. It also presents a substantial challenge to the Chinese population, as it is the leading cause of mortality, long-term disability, and severe disease burden in China. Ischemic stroke accounts for the largest proportion of these cases [2,3]. Meanwhile, the mortality of ischemic stroke is lower than that of hemorrhagic stroke in middle-income countries; therefore, more attention needs to be directed toward patients who experienced ischemic stroke [4]. Data based on the China National Stroke Registries revealed that the stroke recurrence rate was approximately 12.5% within 12 months, which may lead to the deterioration of functional outcomes, decreased quality of life, and even increased mortality [5]. Therefore, it is necessary to achieve secondary prevention of ischemic stroke through lifestyle changes and medication intervention. A qualitative study found that stroke survivors perceived that using preventive medications was more important than modifying lifestyle behaviors [6]. It was previously reported that rational use of these drugs may reduce the risk of recurrent stroke and improve survival [7,8]. However, a systematic review of 22 studies showed that the pooled prevalence of medication nonadherence was 29% [9], and the risk factors remain poorly controlled among stroke survivors [10].

The health belief model (HBM) is widely used to guide health education, aiming to help participants develop, adopt, and practice healthy behaviors [11,12]. It emphasizes the subjective psychological process of individuals, suggesting that individuals with beliefs related to disease and health are more willing to adopt healthy behaviors. This theoretical model guides the health education for secondary prevention of stroke from six aspects: (1) helping patients identify their own risk factors for strokes (perceived susceptibility), (2) understanding the possible adverse consequences of stroke recurrence (perceived severity), (3) becoming aware of barriers (perceived barriers) and (4) benefits of secondary prevention(perceived benefits), (5) receiving information and encouragement through social media (cues to action), and (6) establishing confidence in their ability to engage in disease prevention (self-efficacy). When individuals are provided with these factors, they are more likely to adopt the recommended behaviors for secondary prevention.

In particular, the COVID-19 pandemic has had a negative impact on post–acute stroke care [13]. Care pathways focused on secondary prevention of cardiovascular disease may be disrupted, stroke follow-up may be delayed or reduced, and survivors may have difficulty accessing rehabilitation services [14,15]. China has implemented strict national policies to reduce the risk of exposure to the virus [16]. Staff shortages caused by a large number of medical staff supporting high-risk areas, quarantine efforts, and pathogen testing have led to a reduction in outpatient clinics and medical services available to patients having strokes [17]. Therefore, theory-based and implementable secondary prevention measures become particularly important to improve medication adherence and control of vascular risk factors, such as blood pressure (BP), hemoglobin A_{1c} (HbA_{1c}), and low-density lipoprotein cholesterol (LDL-C) levels, within a reasonable range, for stroke survivors during lockdown.

This study aimed to investigate the effects of planned health education based on the HBM, compared to general health management, on improving medication adherence and controlling vascular risk factors in patients ischemic stroke 3 months after admission during the COVID-19 lockdown.

Methods

Study Design

This study was an assessor-blinded quasi-experimental trial with a historical control group. Consecutive participants admitted to the stroke center of the First Hospital of Jilin University were recruited. The hospital is located in the central part of Jilin Province, northeast China, with a population of 27 million and the highest age-standardized prevalence and incidence rate of stroke [5]. It also has the first province-wide stroke emergency map in China, named the Jilin Province Stroke Emergency Maps [18]. The transportation of patients between cities ensured the diversity of participants in our center. Patients admitted from March to June 2020 were assigned to the historical control group, and patients admitted from July to October 2020 were assigned to the intervention group.

Ethical Considerations

This study was conducted in compliance with the principles of the Declaration of Helsinki. The study protocol was reviewed and approved by the ethical committee of the First Hospital of Jilin University (20K055-001) and was registered in the Chinese Clinical Trial Registry (ChiCTR2000040804). All participants signed an informed consent form.

Participants

Patients diagnosed with a first-ever ischemic stroke [19], confirmed by head computed tomography scanning or magnetic resonance imaging; aged≥18 years; able to understand instructions; and able to use mobile phones proficiently were included. Patients diagnosed with primary mental disorders or other serious organ dysfunction, such as liver and kidney dysfunction, were excluded.

Interventions

Patients in the control group were provided with a health management manual with explanations and discharge advice during hospitalization; they conducted out-of-hospital follow-ups through an online WeChat public group and received monthly telephonic coaching sessions of approximately 15 minutes each, provided mainly by a qualified stroke health manager, as previously published [20]. In response to national policies, we canceled the health education courses held in our center during hospitalization to avoid social gatherings according to hospital rules.

Patients in the intervention group received 6 additional web-based educational sessions based on the HBM via Tencent Meeting, which is an audio and video conferencing application. The sessions were held every 2 weeks between 9 AM and 10 AM, with each session lasting approximately 40 minutes. The

intervention was conducted in small groups, with about 8 to 10 people in each group. A qualified stroke health manager, skilled in health education, facilitated all the sessions. The first session was focused on perceived susceptibility. Approximately 20 minutes were allocated for free discussion to encourage participants to identify and express their own risk factors. The second session focused on perceived severity on the basis of the model. The third session focused on perceived benefits. Participants could ask questions and discuss the possible benefits of their behaviors. The fourth and fifth sessions mainly supplemented cues to actions. Participants were encouraged to put forward questions that they did not fully understand and received prompt explanations. The last session aimed to address patients' perceived barriers to secondary prevention behaviors. Participants worked with others to correct incorrect beliefs and provide strategies to overcome these barriers (Table 1).

Participants received a notification on their mobile phones 1 hour before each session. We also implemented online reminders before each session and ensured that all participants attended the meetings on time. Patients who completed all sessions and participated in 3 or more discussions during the sessions were awarded an electronic version of the "Learning Star" honorary certificate to encourage them to continue to adhere to healthy behaviors.

Table . The education session structure based on the health belief model.

Intervention theme	Intervention targets	Procedure of sessions
Perceived susceptibility	Aware of the severity of the disease, current bad behaviors, and risk factors	Through the "Screen Image Sharing" technology of the Tencent Meeting, the PowerPoint was presented to the patients at the same time as the lecture, describing stroke-related statistics and common risk factors (eg, hypertension, diabetes, and dyslipemia).
Perceived severity	Recognize the possibility and risk of stroke recurrence	Describing the possible physical and psychological impact and economic burden of stroke recurrence through several cases that failed to engage in secondary preventive strategies and encouraging participants to share some feelings of distress, when they experienced their first stroke.
Perceived benefits	Aware of the benefits of changing daily lifestyle and following secondary prevention measures	A patient who had been followed up for 1 year without stroke recurrence and had a better prognosis was invited to share his experiences, particularly some lifestyle changes and daily disease monitoring behaviors after discharge.
Cues to actions	Master the specific ways to change the bad lifestyle	Explaining the specific methods of smoking cessation, weight control, reasonable diet, as well as daily monitoring of blood pressure and glucose, and giving medication guidance.
Cues to actions	Recognize the importance of proper use of drugs	Following the medication regimens and not stopping the medicine on their own. Highlighting the strong correlation between dyslipidemia and recurrence during the course of treatment to correct patients' misconceptions.
Perceived barriers	Address the barriers to perform secondary prevention behaviors	Participants were asked to come up with and discuss beliefs, attitudes, or facts that may prevent them from engaging in behaviors correctly.

Outcome Indicators

The outcome assessor was a full-time physician who was professionally engaged in the follow-up of cerebrovascular disease and was not involved in any of the program implementations. Data on patients' sociodemographic characteristics, complications, and stroke severity as well as BP, HbA_{1c}, and LDL-C results were retrieved from electronic medical records. The above physiological and biochemical indicators were collected twice: within 48 hours of admission and during the face-to-face outpatient clinic follow-up 3 months after stroke onset.

The primary outcomes were control of BP, HbA_{1c}, and LDL-C levels according to the recommendations for treatment targets in the Chinese Stroke Association guidelines [21]. The reference treatment targets for patients with ischemic stroke were as follows: BP<140/90 mmHg, LDL-C<2.6 mmol/L, and HbA_{1c}≤7%. BP of the nonhemiplegic side was measured after the patient rested for 5 minutes in a sitting position using a digital sphygmomanometer (Omron HBP-9020). The mean value of 3 consecutive measurements with 1-minute intervals was taken as the final result. LDL-C concentration and blood levels of HbA_{1c} were measured in fresh venous blood samples.

The outcome of self-reported medication adherence was measured by the Chinese version of the 8-item Morisky Medicine Adherence Scale (MMAS) [22,23,24]. The total score ranges from 0 to 8, with a score ≥6 classified as adherent [25]. We also used the modified Rankin Scale (mRS) to evaluate

clinical outcomes in patients with stroke. mRS is a 7-level scale, ranging from 0 (no symptoms) to 6 (death), for assessing the recovery state of neurological function. Patients with an mRS score of 0-1 were defined as having favorable outcomes.

Statistical Analyses

An a priori sample size calculation was performed for this study based on the outcome "the difference in BP at 3 months after discharge in patients with hypertensive ischemic stroke receiving a comprehensive reminder system intervention [26]." Assuming a 20% drop-out rate, at least 170 patients in each group were included to achieve a power of 0.90 with a 2-sided significance level of 5%. Baseline characteristics were reported as means (SDs), median (IQRs), and proportions, as appropriate. Means or probability and 95% CIs were used to report the differences between groups and were examined using the chi-square test for categorical variables and student 2-tailed *t* test or Mann-Whitney *U* test for continuous variables, respectively. Binary logistic regression was used to adjust for confounding variables and compare the proportion of vascular risk factors that reached the target value between the intervention and control groups. Differences in the levels of vascular risk factors between the 2 groups were compared using multiple linear regression, and clinical baseline parameters were considered as covariates. Two-sided *P* values less than .05 were regarded as statistically significant. Data analysis was performed using IBM SPSS Statistics (version 22.0; IBM Corp).

Results

In total, 340 consecutive patients were included in this study. Figure 1 illustrates the study flow chart; 315 patients were finally analyzed, with 155 in the historical control group and 160 in the intervention group. Table 2 presents the demographic and clinical characteristics of the participants. The mean age of participants was 59.2 (SD 10.5) years (the maximum age was 88 years, and the minimum age was 25 years). There were no significant differences in participants' age, gender, educational level, household income, and health insurance between the 2 groups. Slightly more patients in the control group had comorbidities and were smokers, but this difference was not statistically significant. There were no significant differences between the 2 groups in BP, LDL-C, and HBA_{1c} levels or in the proportion of patients who met treatment targets within 48 hours of admission (Table 3).

Figure 1. Study flow chart.

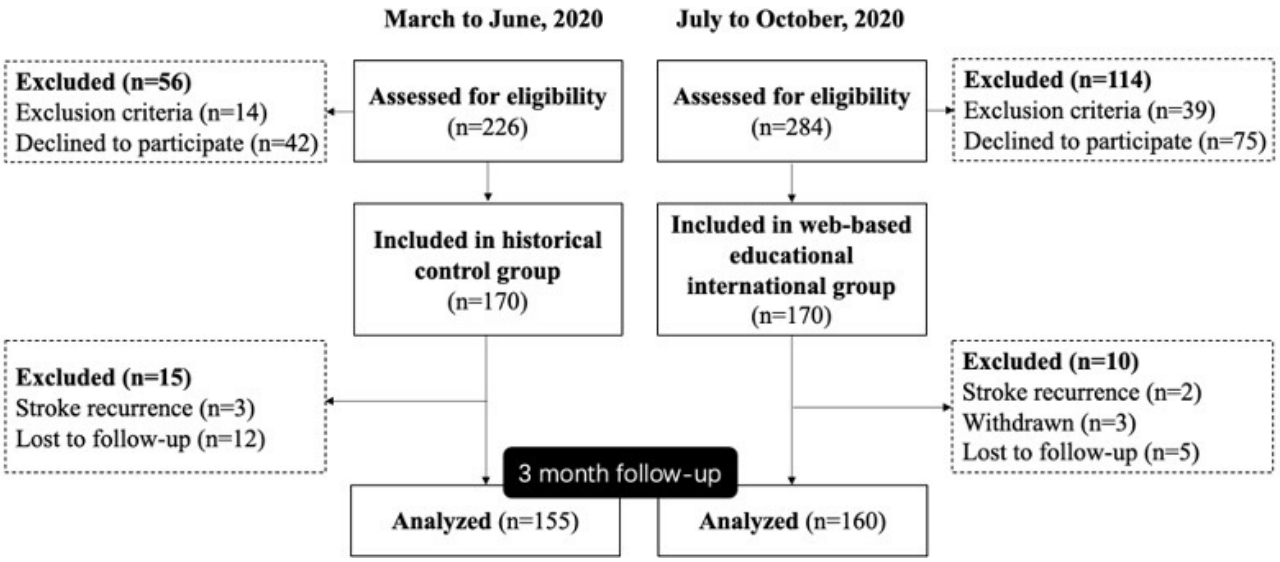


Table . Sociodemographic and clinical characteristics of participants (N=315).

Characteristics		Control group (n=155)	Intervention group (n=160)	*P* value
Age (years), mean (SD)		58.9 (10.3)	59.5 (10.6)	.61
Gender, n (%)				.50
	Male	124 (80)	123 (76.9)	
	Female	31 (20)	37 (23.1)	
Educational level, n (%)				.47
	High school or above	46 (29.7)	41 (25.6)	
	Middle school	41 (26.5)	52 (32.5)	
	Elementary school	68 (43.9)	67 (41.9)	
Monthly per capita family income (RMB[a]), n (%)				.29
	<1000	26 (16.8)	19 (11.9)	
	1000-3000	72 (46.5)	81 (50.6)	
	>3000	57 (36.8)	60 (37.5)	
Health insurance, n (%)				.23
	Yes	142 (91.6)	152 (95)	
	No	13 (8.4)	8 (5)	
Baseline NIHSS[b] score, median (IQR)		2 (0-5)	2 (0-6)	.19
Comorbidities, n (%)				
	Hypertension	112 (72.3)	104 (65)	.17
	Diabetes mellitus	46 (29.7)	40 (25)	.35
	Hyperlipidemia[c]	96 (61.9)	85 (53.1)	.11
	Atrial fibrillation	8 (5.2)	15 (9.4)	.14
Smoking, n (%)		96 (61.9)	91 (56.9)	.36
Alcohol use[d], n (%)		78 (50.3)	88 (55)	.41

[a]1000 RMB is about US $149.

[b]NIHSS: The National Institutes of Health Stroke Scale.

[c]Hyperlipidemia is defined as elevated levels of triglyceride, total cholesterol, low-density lipoprotein cholesterol, and reduced high-density lipoprotein cholesterol levels.

[d]Alcohol use was defined as current or previous habitual consumption of alcohol.

Table . Level of vascular risk factors and proportions of treatment targets achieved at baseline.

Parameter	Control group (n=155)	Intervention group (n=160)	*t* test/χ^2 (*df*)	*P* value
SBP[a] (mmHg), mean (SD)	149.2 (20.5)	148.6 (23.2)	0.266 (310)[b]	.79
DBP[c] (mmHg), mean (SD)	87.0 (11.6)	86.6 (14.4)	0.310 (303)[b]	.76
BP control, n (%)	48 (31.0)	48 (30.0)	0.035 (1)[d]	.85
LDL-C[e] (mmol/L), mean (SD)	2.5 (0.7)	2.6 (0.8)	−1.625 (313)[b]	.11
LDL-C control, n (%)	88 (56.8)	83 (51.9)	0.761 (1)[d]	.38
HbA_{1c}[f] (%), mean (SD)	6.3 (1.4)	6.2 (1.4)	0.306 (313)[b]	.76
HbA_{1c} control, n (%)	126 (81.3)	133 (83.1)	0.181 (1)[d]	.67

[a]SBP: systolic blood pressure.

[b]Groups were compared by *t* test.

[c]DBP: diastolic blood pressure.

[d]Groups were compared by χ^2 test.

[e]LDL-C: low-density lipoprotein cholesterol.

[f]HbA_{1c}: hemoglobin A_{1c}.

After intervention, participants in the intervention group were 1.93 times more likely to have controlled BP (adjusted odds ratio [aOR] 1.93, 95% CI 1.15-3.24; *P*=.012; Table 4); more patients (133/160, 83.1%) had controlled LDL-C compared to those in the control group (105/155, 67.7%; aOR 2.66, 95% CI 1.51-4.69; *P*=.001), and the proportion of patients who had an HbA_{1c} level less than 7.0% at 3 months was 91.9% (147/160) in the intervention group compared with 83.9% (130/155) in the control group (aOR 3.37, 95% CI 1.09 to 10.47; *P*=.04).

Table . Proportions achieving vascular risk factor control, medicine adherent and favorable functional outcome at 3 months.

Variables	Control group (n=155), n (%)	Intervention group (n=160), n (%)	Crude OR[a] (95% CI)	*P* value	Adjusted OR (95% CI)	*P* value
BP[b] control	44 (28.4)	67 (41.9)	1.82 (1.14-2.91)	.01	1.93 (1.15-3.24)[c]	.01
LDL-C[d] control	105 (67.7)	133 (83.1)	2.35 (1.38-4.00)	.002	2.66 (1.51-4.69)[c]	.001
HbA_{1c}[e] control	130 (83.9)	147 (91.9)	2.18 (1.07-4.43)	.03	3.37 (1.09-10.47)[c]	.04
Medicine adherent[f]	98 (63.2)	127 (79.4)	2.24 (1.35-3.70)	.002	2.31 (1.37-3.87)[g]	.002
Favorable functional outcome	97 (62.6)	104 (65)	1.14 (0.72-1.81)	.57	1.22 (0.74-2.02)[h]	.44

[a]OR: odds ratio.

[b]BP: blood pressure.

[c]Adjusted for age, sex, cormorbidities, smoking, alcohol use, and proportion of reaching target value at baseline.

[d]LDL-C: low-density lipoprotein cholesterol.

[e]HbA_{1c}: hemoglobin A_{1c}.

[f]Medicine adherent was assessed with the Morisky Medicine Adherence Scale (MMAS). The MMAS-8 Scale (U.S. Copyright Registration No. TX0008632533), content, name, and trademarks are protected by US copyright and trademark laws. Permission for use of the scale and its coding is required. A license agreement is available from MMAR, LLC., www.moriskyscale.com.

[g]Adjusted for age, sex, cormorbidities, educational level, income, and insurance.

[h]Adjusted for age, sex, cormorbidities, smoking, alcohol use, and the National Institutes of Health Stroke Scale at baseline.

Systolic BP showed a significant reduction at 3 months (adjusted β −3.94, 95% CI −7.18 to −0.60; *P*=.02; Table 5) for the HBM-based planned web-based educational intervention program, but there was no effect on diastolic BP level (adjusted β −2.45, 95% CI −5.06 to 0.17; *P*=.07). There was also a significant postintervention reduction in the mean level of LDL-C (adjusted β −0.21, 95% CI −0.37 to −0.06; *P*=.008) and HbA_{1c} (adjusted β −0.27, 95% CI −0.37 to −0.16; *P*<.001) among the patients in the intervention group compared to those in the comparison group.

Self-reported adherence scores for the intervention and control groups at 3 months were 6.7 (SD 1.5) and 6.2 (SD 1.7), respectively (*P*=.009). The proportion of patients reporting adherence to their medication was 79.4% (127/160) in the

intervention group compared to 63.2% (98/155) in the control group (aOR 2.31, 95% CI 1.37-3.87; *P*=.002). However, there was no significant difference in the favorable functional outcomes between the two groups at 90 days.

Table . Differences of vascular risk factors parameter values between the two groups at 3 months.

Parameter	Control group (n=155)	Intervention group (n=160)	Adjusted β (95% CI)[a]	*P* value
SBP[b] (mmHg), mean (SD)	144.5 (16.7)	140.1 (16.8)	−3.94 (−7.28 to −0.60)	.02
DBP[c] (mmHg), mean (SD)	85.3 (9.9)	82.1 (10.5)	−2.45 (−5.06 to 0.17)	.07
LDL-C[d] (mmol/L), mean (SD)	2.2 (0.8)	2.1 (0.6)	−0.21 (−0.37 to −0.06)	.008
HbA_{1c}[e] (%), mean (SD)	6.2 (1.2)	5.9 (1.0)	−0.27 (−0.37 to −0.16)	<.001

[a]Adjusted for age, sex, cormorbidities, smoking, alcohol use, and baseline value.

[b]SBP: systolic blood pressure.

[c]DBP: diastolic blood pressure.

[d]LDL-C: low-density lipoprotein cholesterol.

[e]HbA_{1c}: hemoglobin A_{1c}.

Discussion

Principal Findings

Our study demonstrated that a planned web-based education program based on the HBM model positively impacted the control of vascular risk factors in patients who had ischemic stroke and improved medication adherence. The theory-guided practice may help enhance the effectiveness of health management. A comprehensive reminder system based on the HBM also improved medication adherence and reduced BP of patients who had a stroke [26]. However, in the SMS4Stroke study, guided by the HBM with behavior change theory, BP did not show a significant change between the 2 groups [27]. A possible reason is that the intervention duration was relatively short at 2 months. The HBM predicts the impact on human behavior from 6 aspects, and participants may change their behavior through health education related to these themes.

Hypertension control is a cornerstone of secondary stroke prevention. Health education based on the HBM increased the BP control compliance rate of participants by 2 times at 3 months, but the level was still suboptimal compared with other studies [26,28]. One of the reasons may be that the cold weather in the northeast and the residents' preference for pickles led to a BP level of 149/87 mmHg at baseline, which was significantly higher than that of patients from Southern and Northern China [29,30]. However, our data are almost consistent with those of the China National Stroke Registry, a multicenter database covering the whole country [31]. Therefore, there is room for improvement in the next period.

Higher LDL-C levels have been associated with stroke recurrence over the past 10 years. Unlike hypertension and hyperglycemia, dyslipidemia may not cause significant physical discomfort for patients. In addition, a lack of knowledge about statins also makes patients likely to ignore adherence to long-term use of this medicine [32,33]. After the web-based education intervention, LDL-C levels decreased to 0.6 mmol/L (95% CI −0.7 to −0.4), which was double the 0.3 mmol/L reported in a 12-week exercise and education program for nondisabling stroke at 6 months follow-up [34]. HbA_{1c} can objectively reflect the average glycemic control over the last 3 months. With the cues to actions we provided, through lifestyle changes and adherence to hypoglycemic agents, the mean HbA_{1c} of patients with diabetes in the intervention group fell within the normal range, and 148/160 (92.5%) patients achieved control goals, which was better than the results of a 3-month poststroke education delivered via 5-minute movies to stroke survivor and caregiver dyads in Pakistan [35]. This is likely because only a quarter of the patients in this study had diabetes, and the patients had better glycemic control at baseline.

The 3-months medication adherence in these 2 groups was higher than that of previous studies [9,36]. However, it is still lower than the 89.0% medication adherence reported in a study of 600 patients conducted in 3 stroke centers across Korea [37]. Only patients who continued to take all prescribed medications were further evaluated for adherence, resulting in the exclusion of patients who reported discontinuing all or some medications, which may explain this difference. A study conducted in 2020 found that the imposed COVID-19 lockdown is unlikely to affect the medication behavior of patients with high preestablished adherence [38]. Therefore, we purposefully established the medication beliefs of patients based on the HBM model, which helps to ensure patients' medication adherence under the normalization of the pandemic.

Although statistically significant improvements were found in both vascular risk factor control and medication adherence in the intervention group compared with the control group, the functional outcomes were not statistically different. This is probably because the neurological condition was not severe in our participants. Multicenters with large samples and multifaceted efforts are required to prove the efficacy of our web-based health education program.

Our study has several limitations. Due to the shortage of medical staff and researchers during the lockdown, this study used a historical control method, which provides less quality evidence compared to randomized controlled trials. Outcomes after intervention were evaluated for 3 months, and long-term effects

need to be further assessed. The study was conducted in a single center, and the included patients may have regional characteristics, having a certain impact on the results. We only investigated overall medication adherence, and the persistence of a specific medication was unknown. The assessment methods of self-report may have led to recall bias; a more objective measure of adherence would be preferable to confirm our findings. However, this study design allowed us to realize the potential benefits of web-based health education under real-life conditions and complement evidence from randomized controlled trials.

Conclusions

This study found that a web-based health education program under the HBM model during the COVID-19 lockdown in autumn 2020 in northeast China, designed to teach patients who experienced a stroke about secondary prevention, may result in improved vascular risk factors and medication adherence. It provides insights into achieving secondary stroke prevention goals, recommended in real-world guidelines. It also serves as a reference for the use of telemedicine to conduct disease prevention under the “new normal” of COVID-19 and the shortage of medical staff.

Acknowledgments

The MMAS-8 Scale (U.S. Copyright Registration No. TX0008632533), content, name, and trademarks are protected by US copyright and trademark laws. Permission for use of the scale and its coding is required. A license agreement is available from MMAR, LLC., www.moriskyscale.com.

This work was supported by the Jilin Province Department of Finance (grant JLSWSRCZX2020-080).

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Authors' Contributions

ZL wrote the original draft and conducted the formal analysis. XS contributed to the writing of the original draft and the study methodology. ZNG contributed to reviewing and editing the manuscript and the study conceptualization. YS conducted data curation. YY contributed to reviewing and editing the manuscript and the study methodology. XY supervised the study and contributed to the study conceptualization, methodology, and funding acquisition.

Conflicts of Interest

None declared.

References

1. Feigin VL, Stark BA, Johnson CO, et al. Global, regional, and national burden of stroke and its risk factors, 1990-2019: a systematic analysis for the global burden of disease study 2019. Lancet Neurol 2021 Oct;20(10):795-820. [doi: 10.1016/S1474-4422(21)00252-0]
2. Ma Q, Li R, Wang L, et al. Temporal trend and attributable risk factors of stroke burden in China, 1990-2019: an analysis for the Global Burden of Disease Study 2019. Lancet Public Health 2021 Dec;6(12):e897-e906. [doi: 10.1016/S2468-2667(21)00228-0] [Medline: 34838196]
3. Zhou M, Wang H, Zeng X, et al. Mortality, morbidity, and risk factors in China and its provinces, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2019 Sep 28;394(10204):1145-1158. [doi: 10.1016/S0140-6736(19)30427-1] [Medline: 31248666]
4. Chen Y, Wright N, Guo Y, et al. Mortality and recurrent vascular events after first incident stroke: a 9-year community-based study of 0.5 million Chinese adults. Lancet Glob Health 2020 Apr;8(4):e580-e590. [doi: 10.1016/S2214-109X(20)30069-3] [Medline: 32199124]
5. Xu J, Zhang X, Jin A, et al. Trends and risk factors associated with stroke recurrence in China, 2007-2018. JAMA Netw Open 2022 Jun 1;5(6):e2216341. [doi: 10.1001/jamanetworkopen.2022.16341] [Medline: 35704318]
6. Lin B, Zhang Z, Thrift AG, et al. Qualitative study of stroke survivors' perceptions of secondary prevention. J Adv Nurs 2022 May;78(5):1377-1388. [doi: 10.1111/jan.15079] [Medline: 34636437]
7. Lank RJ, Lisabeth LD, Levine DA, et al. Ethnic differences in 90-day poststroke medication adherence. Stroke 2019 Jun;50(6):1519-1524. [doi: 10.1161/STROKEAHA.118.024249] [Medline: 31084331]
8. Dalli LL, Kim J, Cadilhac DA, et al. Greater adherence to secondary prevention medications improves survival after stroke or transient ischemic attack: a linked registry study. Stroke 2021 Nov;52(11):3569-3577. [doi: 10.1161/STROKEAHA.120.033133] [Medline: 34315251]
9. Rajahthurai SD, Farrukh MJ, Makmor-Bakry M, et al. Use of complementary and alternative medicine and adherence to medication therapy among stroke patients: a meta-analysis and systematic review. Front Pharmacol 2022 Jun 3;13:870641. [doi: 10.3389/fphar.2022.870641] [Medline: 35721127]

10. Bravata DM, Daggy J, Brosch J, et al. Comparison of risk factor control in the year after discharge for ischemic stroke versus acute myocardial infarction. Stroke 2018 Feb;49(2):296-303. [doi: 10.1161/STROKEAHA.117.017142] [Medline: 29284738]
11. Khorsandi M, Fekrizadeh Z, Roozbahani N. Investigation of the effect of education based on the health belief model on the adoption of hypertension-controlling behaviors in the elderly. Clin Interv Aging 2017 Jan 27;12:233-240. [doi: 10.2147/CIA.S117142] [Medline: 28184154]
12. Hu S, Xiong R, Hu Q, Li Q. Effects of nursing intervention based on health belief model on self-perceived burden, drug compliance, and quality of life of renal transplant recipients. Contrast Media Mol Imaging 2022 May 5;2022:3001780. [doi: 10.1155/2022/3001780] [Medline: 35601567]
13. Markus HS, Brainin M. COVID-19 and stroke-a global world stroke organization perspective. Int J Stroke 2020 Jun;15(4):361-364. [doi: 10.1177/1747493020923472] [Medline: 32310017]
14. Duffy EY, Cainzos-Achirica M, Michos ED. Primary and secondary prevention of cardiovascular disease in the era of the coronavirus pandemic. Circulation 2020 Jun 16;141(24):1943-1945. [doi: 10.1161/CIRCULATIONAHA.120.047194] [Medline: 32310675]
15. Boldrini P, Garcea M, Brichetto G, et al. Living with a disability during the pandemic. "Instant paper from the field" on rehabilitation answers to the COVID-19 emergency. Eur J Phys Rehabil Med 2020 Jun;56(3):331-334. [doi: 10.23736/S1973-9087.20.06373-X] [Medline: 32406226]
16. China CDC novel coronavirus pneumonia prevention guidance for public. The State Council Information Office of the People's Republic of China. 2020. URL: http://www.scio.gov.cn/ztk/dtzt/42313/43142/index.htm [accessed 2022-08-01]
17. Wang X, Ouyang M, Carcel C, et al. Impact of COVID-2019 on stroke services in China: survey from the Chinese Stroke Association. Stroke Vasc Neurol 2020 Dec;5(4):323-330. [doi: 10.1136/svn-2020-000514] [Medline: 32989012]
18. Jin H, Qu Y, Guo ZN, Yan XL, Sun X, Yang Y. Impact of Jilin province Stroke Emergency Maps on acute stroke care improvement in northeast China. Front Neurol 2020 Jul 22;11:734. [doi: 10.3389/fneur.2020.00734] [Medline: 32774322]
19. Powers WJ, Rabinstein AA, Ackerson T, et al. Guidelines for the early management of patients with acute ischemic stroke: 2019 update to the 2018 guidelines for the early management of acute ischemic stroke: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. Stroke 2019 Dec;50(12):e344-e418. [doi: 10.1161/STR.0000000000000211] [Medline: 31662037]
20. Yan X, Liu Z, Guo ZN, et al. Positive influence of stroke health manager on risk factors control and medication adherence after ischemic stroke. Front Neurol 2020 Mar 11;11:168. [doi: 10.3389/fneur.2020.00168] [Medline: 32218766]
21. Chinese Stroke Association. Guidelines for Clinical Management of Cerebrovascular Disorders: People's Medical Publishing House; 2019.
22. Yan J, You LM, Yang Q, et al. Translation and validation of a Chinese version of the 8-item Morisky medication adherence scale in myocardial infarction patients. J Eval Clin Pract 2014 Aug;20(4):311-317. [doi: 10.1111/jep.12125] [Medline: 24813538]
23. Berlowitz DR, Foy CG, Kazis LE, et al. Effect of intensive blood-pressure treatment on patient-reported outcomes. N Engl J Med 2017 Aug 24;377(8):733-744. [doi: 10.1056/NEJMoa1611179] [Medline: 28834483]
24. Bress AP, Bellows BK, King JB, et al. Cost-effectiveness of intensive versus standard blood-pressure control. N Engl J Med 2017 Aug 24;377(8):745-755. [doi: 10.1056/NEJMsa1616035] [Medline: 28834469]
25. Morisky DE, Ang A, Krousel-Wood M, Ward HJ. Predictive validity of a medication adherence measure in an outpatient setting. J Clin Hypertens (Greenwich) 2008 May;10(5):348-354 Retracted inJ Clin Hypertens (Greenwich). 2023 Sep;25(9):889. [doi: 10.1111/jch.14718] [Medline: 37594022]
26. Wan LH, Zhang XP, You LM, Ruan HF, Chen SX. The efficacy of a comprehensive reminder system to improve health behaviors and blood pressure control in hypertensive ischemic stroke patients: a randomized controlled trial. J Cardiovasc Nurs 2018 ;33(6):509-517. [doi: 10.1097/JCN.0000000000000496] [Medline: 29901484]
27. Kamal AK, Shaikh Q, Pasha O, et al. A randomized controlled behavioral intervention trial to improve medication adherence in adult stroke patients with prescription tailored Short Messaging Service (SMS)-SMS4Stroke study. BMC Neurol 2015 Oct 21;15:212. [doi: 10.1186/s12883-015-0471-5] [Medline: 26486857]
28. Gynnild MN, Aakerøy R, Spigset O, et al. Vascular risk factor control and adherence to secondary preventive medication after ischaemic stroke. J Intern Med 2021 Mar;289(3):355-368. [doi: 10.1111/joim.13161] [Medline: 32743852]
29. Yan LL, Gong E, Gu W, et al. Effectiveness of a primary care-based integrated mobile health intervention for stroke management in rural China (SINEMA): a cluster-randomized controlled trial. PLoS Med 2021 Apr;18(4):e1003582. [doi: 10.1371/journal.pmed.1003582] [Medline: 33909607]
30. Ren Y, Shi J, Qiao Y, et al. Epidemiological status quo of hypertension in elderly population in Changchun, China: a cross-sectional study. BMJ Open 2022 Mar 30;12(3):e053086. [doi: 10.1136/bmjopen-2021-053086] [Medline: 35354631]
31. Wang M, Wang CJ, Gu HQ, et al. Sex differences in short-term and long-term outcomes among patients with acute ischemic stroke in China. Stroke 2022 Jul;53(7):2268-2275. [doi: 10.1161/STROKEAHA.121.037121] [Medline: 35130717]
32. Rodrigues JPV, Pereira LB, Zanetti MOB, et al. Use of statins for the secondary prevention of stroke: are we respecting the scientific evidences? J Stroke Cerebrovasc Dis 2020 Aug;29(8):104912. [doi: 10.1016/j.jstrokecerebrovasdis.2020.104912] [Medline: 32689595]

33. Jiang Y, Yang X, Li Z, et al. Persistence of secondary prevention medication and related factors for acute ischemic stroke and transient ischemic attack in China. Neurol Res 2017 Jun;39(6):492-497. [doi: 10.1080/01616412.2017.1312792] [Medline: 28420316]
34. MacKay-Lyons M, Gubitz G, Phillips S, et al. Program of Rehabilitative exercise and education to avert vascular events after non-disabling stroke or transient ischemic attack (PREVENT trial): a randomized controlled trial. Neurorehabil Neural Repair 2022 Feb;36(2):119-130. [doi: 10.1177/15459683211060345] [Medline: 34788569]
35. Kamal A, Khoja A, Usmani B, et al. Effect of 5-minute movies shown via a mobile phone app on risk factors and mortality after stroke in a low- to middle-income country: randomized controlled trial for the stroke caregiver dyad education intervention (Movies4Stroke). JMIR Mhealth Uhealth 2020 Jan 28;8(1):e12113. [doi: 10.2196/12113] [Medline: 32012080]
36. Wang W, Chia GS, Tan IF, et al. Independent predictors of medication adherence among Singaporean patients following an ischaemic stroke or transient ischaemic attack. J Clin Nurs 2017 Dec;26(23-24):5016-5024. [doi: 10.1111/jocn.14001] [Medline: 28793375]
37. Yoo SH, Kim GG, Kim SR, et al. Predictors of long-term medication adherence in stroke survivors: a multicentre, prospective, longitudinal study. J Clin Nurs 2023 Jan;32(1-2):58-70. [doi: 10.1111/jocn.16472] [Medline: 35922959]
38. Dietrich F, Polymeris AA, Verbeek M, et al. Impact of the COVID-19 Lockdown on the adherence of stroke patients to direct oral anticoagulants: a secondary analysis from the MAAESTRO study. J Neurol 2022 Jan;269(1):19-25. [doi: 10.1007/s00415-021-10631-5] [Medline: 34081196]

Abbreviations

aOR: adjusted odds ratio
BP: blood pressure
HbA_{1c}: hemoglobin A_{1c}
HBM: health belief model
LDL-C: low-density lipoprotein cholesterol
MMAS: Morisky Medicine Adherence Scale
mRS: modified Rankin Scale

Edited by L Buis; submitted 20.11.22; peer-reviewed by A Ćirković, G Shakerinejad, J Volkman; revised version received 29.01.24; accepted 26.02.24; published 24.04.24.

Please cite as:
Liu Z, Sun X, Guo ZN, Sun Y, Yang Y, Yan X
Effects of a Planned Web-Based Educational Intervention Based on the Health Belief Model for Patients With Ischemic Stroke in Promoting Secondary Prevention During the COVID-19 Lockdown in China: Quasi-Experimental Study
JMIR Mhealth Uhealth 2024;12:e44463
URL: https://mhealth.jmir.org/2024/1/e44463
doi:10.2196/44463

Original Paper

Barriers and Implications of 5G Technology Adoption for Hospitals in Western China: Integrated Interpretive Structural Modeling and Decision-Making Trial and Evaluation Laboratory Analysis

Linyun Zhou[1,2], MPH; Minghuan Jiang[3], PhD; Ran Duan[4], PhD; Feng Zuo[5], MSc; Zongfang Li[2,6], PhD; Songhua Xu[2], PhD

[1]Department of Health Management, The Second Affiliated Hospital of Xi'an Jiaotong University, Xi'an, China

[2]Institute of Medical Artificial Intelligence, The Second Affiliated Hospital of Xi'an Jiaotong University, Xi'an, China

[3]Department of Pharmacy Administration and Clinical Pharmacy, School of Pharmacy, Xi'an Jiaotong University, Xi'an, China

[4]Information Department, Chongqing General Hospital, Chongqing, China

[5]Information Center, The First Affiliated Hospital of Army Medical University (Southwest Hospital), Chongqing, China

[6]Clinical Research Center for Hepatic & Splenic Diseases of Shaanxi Province, The Second Affiliated Hospital of Xi'an Jiaotong University, Xi'an, China

Corresponding Author:
Songhua Xu, PhD
Institute of Medical Artificial Intelligence
The Second Affiliated Hospital of Xi'an Jiaotong University
No 5 Jianqiang Road
Xi'an, 710016
China
Phone: 86 029 86320798
Email: songhuaxu@126.com

Abstract

Background: 5G technology is gaining traction in Chinese hospitals for its potential to enhance patient care and internal management. However, various barriers hinder its implementation in clinical settings, and studies on their relevance and importance are scarce.

Objective: This study aimed to identify critical barriers hampering the effective implementation of 5G in hospitals in Western China, to identify interaction relationships and priorities of the above-identified barriers, and to assess the intensity of the relationships and cause-and-effect relations between the adoption barriers.

Methods: This paper uses the Delphi expert consultation method to determine key barriers to 5G adoption in Western China hospitals, the interpretive structural modeling to uncover interaction relationships and priorities, and the decision-making trial and evaluation laboratory method to reveal cause-and-effect relationships and their intensity levels.

Results: In total, 14 barriers were determined by literature review and the Delphi method. Among these, "lack of policies on ethics, rights, and responsibilities in core health care scenarios" emerged as the fundamental influencing factor in the entire system, as it was the only factor at the bottom level of the interpretive structural model. Overall, 8 barriers were classified as the "cause group," and 6 as the "effect group" by the decision-making trial and evaluation laboratory method. "High expense" and "organizational barriers within hospitals" were determined as the most significant driving barrier (the highest R–C value of 1.361) and the most critical barrier (the highest R+C value of 4.317), respectively.

Conclusions: Promoting the integration of 5G in hospitals in Western China faces multiple complex and interrelated barriers. The study provides valuable quantitative evidence and a comprehensive approach for regulatory authorities, hospitals, and telecom operators, helping them develop strategic pathways for promoting widespread 5G adoption in health care. It is suggested that the stakeholders cooperate to explore and solve the problems in the 5G medical care era, aiming to achieve the coverage of 5G medical care across the country. To our best knowledge, this study is the first academic exploration systematically analyzing factors resisting 5G integration in Chinese hospitals, and it may give subsequent researchers a solid foundation for further studying the application and development of 5G in health care.

(JMIR Mhealth Uhealth 2024;12:e48842) doi:10.2196/48842

KEYWORDS

5G health care; 5G adoption barriers; 5G adoption strategy; smart health care; Western China hospitals

Introduction

Background

With the advancement of information and communication technology, along with the gradual improvement of China's medical information system construction, China's medical industry is moving away from 1.0 medical informatization to 3.0 medical intelligence [1]. Leveraging the advantages such as ultralow latency, high capacity, high speed, seamless connectivity, high reliability, and low power consumption [2], 5G technology plays an essential role in realizing the interconnection and remote monitoring of medical equipment, patient monitoring, remote consultation, and other telemedicine scenarios. At the same time, 5G technology accelerates the data collection, circulation, analysis, and feedback of various applications in the broad medical and health field. With the advent of 5G technology, medical information can now flow and be shared seamlessly among patients, medical equipment, and hospital information systems. This has paved the way for hospitals to simplify the entire medical treatment and service process, right from prediagnosis to diagnosis and postdiagnosis stages [3].

In recent years, the Chinese government has invested considerably in developing innovative 5G smart hospitals to offer better health care to patients and improve their internal management. For example, in July 2021, a total of 10 departments including the Ministry of Industry and Information Technology, Office of the Central Cyberspace Affairs Commission, National Development and Reform Commission, Ministry of Education, Ministry of Finance, Ministry of Housing and Urban-Rural Development, Ministry of Culture and Tourism, National Health Commission, State-Owned Assets Supervision and Administration Commission of the State Council, and National Energy Administration, jointly released the Sailing Action Plan for 5G Applications (2021-2023). The plan aims to encourage the development of various 5G medical products such as robots, emergency vehicles, medical access gateways, and intelligent medical equipment across the country. The plan also emphasizes the need to strengthen the deployment of 5G medical and health network infrastructure, focusing on improving the coverage of 5G in top-tier national hospitals, disease prevention and control centers, elderly care institutions, and other critical locations. Additionally, the plan aims to build 5G networks and 5G medical edge clouds to enhance in-hospital medical care and telemedicine [4]. In September 2021, the Ministry of Industry and Information Technology, in collaboration with the National Health Commission, released the “Notice on Announcing Pilot Projects for 5G+Medical and Healthcare Applications.” This announcement identified 988 pilot projects aimed at advancing the application of 5G technology in various health care domains, including first aid, telediagnosis, teletreatment, tele-intensive care, traditional Chinese medicine diagnosis and treatment, hospital management, intelligent disease control, health management, and other directions. Among the pilot projects, as many as 611 5G smart medical projects are led by hospitals (general hospitals, emergency centers, and specialized hospitals) [5].

More and more hospitals in China are investing in 5G construction. For instance, Guangdong Provincial People's Hospital put into use the 5G hospital in July 2021, aimed to integrate 5G, big data, artificial intelligence, and other new technologies into various medical scenarios such as treatment, teaching, research, management, and service [6]. Shanghai aims to realize 100% 5G in-depth coverage and 5G typical services for all tertiary hospitals and at least 50% 5G in-depth coverage and 5G typical services for other hospitals by 2023 [7]. Sir Run Run Shaw Hospital Affiliated with Zhejiang University School of Medicine successfully performed a cholecystectomy for a patient from Xinjiang Corps Alar Hospital thousands of miles away by leveraging the robotic arm, which achieved a breakthrough in China's 5G ultraremote robot human liver and gallbladder surgery [8]. The Second Affiliated Hospital of Xi'an Jiaotong University has piloted several scenarios, such as 5G+emergency rescue, 5G+mobile computed tomography, 5G+unmanned aerial vehicle medical delivery, and 5G+integrated remote diagnosis.

However, though 5G technology undoubtedly introduces enormous benefits for hospitals if adequately applied, it has yet to be widely used in many health care scenarios. Hospitals are experiencing various challenges during the actual 5G application process. Different problems are met in the implementation process, including expertise, operation, resource, regulation, and market access factors, as described in the innovation resistance theory (IRT) [9]. At the same time, there are still no systematic studies that have explored the barriers to the adoption of 5G applications in hospitals in Western China. This is particularly important given that technological development in the Eastern region of China is more advanced compared to the Western region. The lag in 5G development in Western China may become another factor that increases the economic imbalance between these 2 regions [10]. Hence, more research is essential for Western China to provide a better understanding of the barriers hindering the adoption of 5G in health care.

Objectives

This study addresses the critical research question below: what are the barriers to implementing 5G in hospitals in Western China? Based on the question, the following research objectives have been formulated: (1) to identify critical barriers hampering the effective implementation of 5G in hospitals in Western China, (2) to identify interaction relationships and priorities of the above-identified barriers, and (3) to assess the intensity of the relationships and cause-and-effect relations between the adoption barriers.

Methods

Ethical Considerations

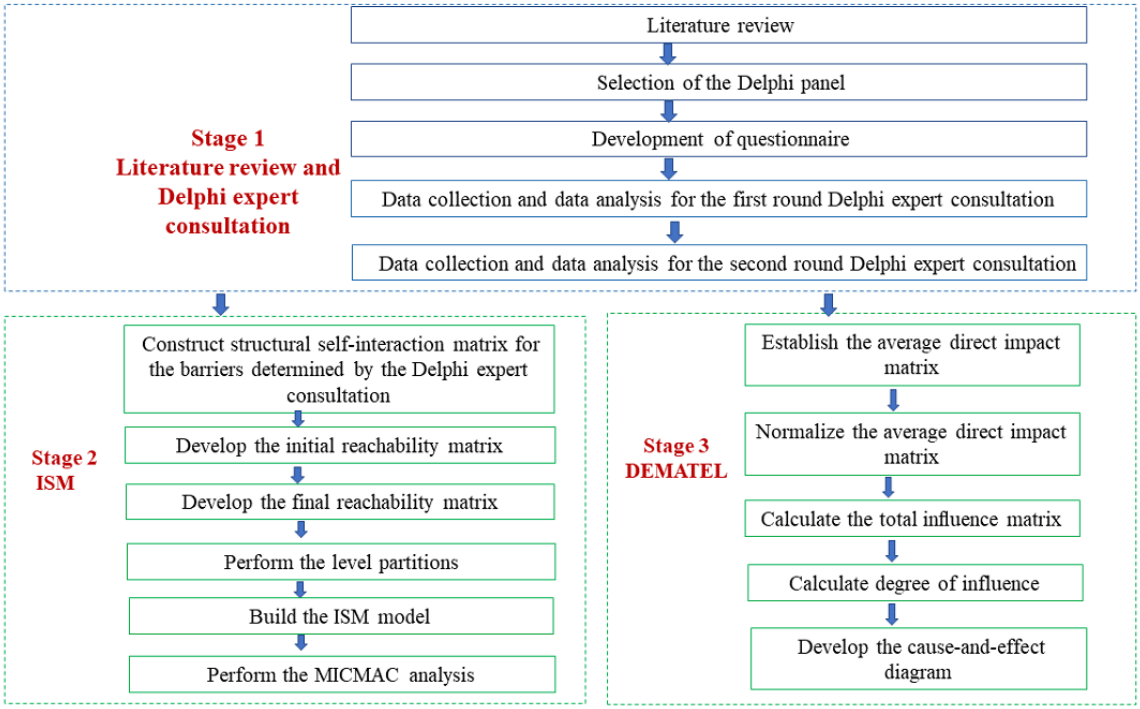
The data were collected through literature review and anonymous questionnaires, which posed no harm to individuals

and did not involve sensitive personal information or commercial interests. Based on the Regulation for Ethical Review of Life Sciences and Medical Research involving human beings issued by the Chinese National Health Commission, Ministry of Education, Ministry of Science and Technology, and State Administration of Traditional Chinese Medicine (Chapter III, Article 32) [11], ethical review was exempted for this study.

Study Design

To address objectives 1-3, a 3-stage hybrid research methodology was proposed by the authors. As shown in Figure 1, a flowchart of the research procedure is conducted. In the first stage, barriers are identified from the existing literature and discussed with experts for further modification and addition using the Delphi technique. The second stage incorporated the interpretive structural modeling (ISM) to obtain a hierarchical structure and interrelationship between the barriers. ISM has a significant advantage because it displays conclusions in the form of a hierarchical topology diagram that is highly intuitive. The hierarchical topology diagram clearly explains the causal relationship and ladder structure among system factors. However, more is needed to determine the intensity of the relationship between factors. It needs to provide the cause-and-effect relation among barriers, which limits the ISM approach [12]. The decision-making trial and evaluation laboratory (DEMATEL) method, on the other hand, can precisely overcome the limitations of the ISM approach. It can determine the strength of influence between variables within the identified structure, providing a deeper understanding of the causal relationship between influencing factors [13]. Therefore, this paper intends to combine Delphi expert consultation, ISM, and DEMATEL to study the hierarchical structure of driving factors and the causal relationship between them. The procedure is explained in more detail in the following section.

Figure 1. Flowchart of research methodology. DEMATEL: decision-making trial and evaluation laboratory; ISM: interpretive structural modeling; MICMAC: Matriced Impacts Corises-multiplication Appliance Classement.



Stage I: Determination of the Barriers Using Literature Review and Delphi Expert Consultation

Overview

This study conducted a literature review from November 20 to 30, 2022, to gather the resistance factors toward 5G development in health care. After that, a 2-round Delphi expert consultation was implemented to refine the factors identified from the above literature review on December 30, 2022, and February 9, 2023. The Delphi technique was developed by the Rand Corporation in 1953 and used as a multistage self-completed questionnaire with individual feedback [14]. It was initially developed as a method for forecasting but has since been widely applied in other areas, including health research [15,16].

Step 1: Literature Review

Web of Science, PubMed, Google Scholar, Chinese government's official websites, and gray literature, including industry reports, were searched by referencing keywords including “5G healthcare,” “5G smart healthcare,” “5G in hospitals,” “5G applications in healthcare,” and “digital health in China.”

Step 2: Selection of the Delphi Panel

To ensure the authoritative scoring results of the consulting expert group on the evaluation indicators, experienced professionals with intermediate or senior titles who are willing to cooperate actively and who are interested in this research were selected. Leaders responsible for informatization work in health authority, heads and frontline employees from the

hospital's information management department, 5G communication suppliers, and scholars in 5G and hospital informatization field were invited to this panel.

Step 3: Development of Questionnaire

To determine the importance of the barriers selected from the literature review, consolidated criteria have been designed in questionnaire format following a 5-point Likert-type scale (5=very important, 4=relatively important, 3=intermediate, 2=unimportant, and 1=very unimportant). To determine the degree of expert authority (Cr), the expert's familiarity with the indicator (Cs) and the judgment basis (Ca) were collected in the questionnaire. The quantitative values for Cs are divided into 5 levels (0.9=very familiar, 0.7=relatively familiar, 0.5=intermediate, 0.3=unfamiliar, and 0.1=very unfamiliar), and the quantitative values for Ca are present in Table 1. The questionnaire is also designed to allow the experts to offer their judgments, with space provided for them to add, remove, and justify their responses.

Table 1. Quantitative values for judgment basis.

Judgment basis (Ca)	Quantitative value of influence degree		
	High	Intermediate	Low
Theoretical analysis	0.3	0.2	0.1
Practical experience	0.5	0.4	0.3
Learn from domestic and foreign peers	0.1	0.1	0.1
Intuition	0.1	0.1	0.1

Step 4: Data Collection

The questionnaires were distributed and collected via the WeChat platform, the most widely and frequently used mobile social media in China, which is profoundly integrated into the daily life of Chinese people [17]. It is often used for distributing and collecting questionnaires.

Step 5: Data Analysis

An analytical stage followed each round of the Delphi questionnaires. The questionnaire recall rate expresses the degree of positivity of the experts. The degree of expert authority Cr can be calculated from the values of Cs and Ca as follows: Cr=(Cs+Ca)/2. The degree of coordination of expert opinions is judged by the coefficients of variation (CVs) and Kendall coefficient of concordance (W). In this study, the barrier screening standard is CV≤0.250. Barriers whose CVs are higher than 0.250 will be modified or deleted. CV is calculated by the mean value and SD. For Kendall coefficient of concordance, the larger the value, the better the coordination of expert opinions. After the analysis, experts' feedback and perspectives will be presented to all participants.

Stage II: Development of Research Framework Using ISM

The ISM method originated from structural modeling and was introduced by Warfield [18] for better decision-making when too many factors or constructs exist. It is a qualitative and interpretive method that involves a mutual learning process that uses the experience of experts to identify the relationship between factors, variables, enablers, and barriers [19,20]. Based on the relationship, an overall multilevel structure is extracted from the complex items. It is very suitable for interdisciplinary research of natural science and social science. The ISM method has been widely used in management and new technology research in different industries.

Referring to the above studies, the basic steps of the ISM method in this study are as follows.

Step 1 involved constructing a “structural self-interaction matrix (SSIM)” for the barriers determined by the Delphi expert consultation. In this step, the symbols “L, M, N, and O” are used to develop a link between the proposed barriers, where L indicates that barrier i has an impact on barrier j, M indicates that barrier j has an impact on barrier i, N indicates that barriers i and j interact with each other, and O indicates that barriers i and j have no interaction with each other.

Step 2 involved converting the SSIM into an “initial reachability matrix (IRM).” In this step, the symbols “L, M, N, and O” are converted into binary elements 0 and 1, and the conversion rules are shown in Table 2.

Step 3 involved checking the transitivity of the IRM to obtain the “final reachability matrix (FRM).” Some new interrelationships between barriers can be established during this step. Transitivity was tested as if barrier A influences barrier B, barrier B influences barrier C, and then, barrier A indirectly influences barrier C.

Step 4 involved performing the level partition through the FRM to get the hierarchy of barriers to plot the ISM. Based on the FRM, a “reachability set,” an “antecedent set,” and an “intersection set” for each barrier were developed.

Step 5 involved building the ISM, checking the model for conceptual inconsistencies, and modifying it accordingly.

Step 6 involved performing the Matriced Impacts Corises-multiplication Appliance Classement (MICMAC) analysis. The driving power (DP) and dependence power (DEP) of the identified barriers based on the FRM were calculated, and the barriers were classified into 4 clusters, known as an autonomous cluster, dependent cluster, linkage cluster, and independent cluster. The details of these 4 clusters are the following:

- Autonomous clusters: the barriers within the autonomous cluster have low DEP and DP. These barriers have no direct

relation with other barriers and can be considered almost isolated from the system.
- Dependent clusters: the barriers in this group do not have robust DP, but their DEP is strong.
- Linkage clusters: the barriers in this cluster are categorized by high DP and DEP. These factors are unstable, so making any changes to them will significantly affect other barriers and may influence them.
- Independent clusters: the barriers within this cluster have high DP and low DEP. These barriers affect other barriers but are less affected.

Table 2. Conversion rule for IRM[a].

(i,j) in SSIM[b]	(i,j) in IRM	(j,i) in IRM
L[c]	1	0
M[d]	0	1
N[e]	1	1
O[f]	0	0

[a]IRM: initial reachability matrix.

[b]SSIM: structural self-interaction matrix.

[c]L indicates that barrier i has an impact on barrier j.

[d]M indicates that barrier j has an impact on barrier i.

[e]N indicates that barriers i and j interact with each other.

[f]O indicates that barriers i and j have no interaction with each other.

Stage III: Identification of Cause-and-Effect Group Using DEMATEL

The DEMATEL approach is a system analysis method based on graph theory and matrix tools. It is used to analyze the cause-effect relationship between factors in complex systems and identify the interaction's intensity [19,21]. The basic steps to carry out DEMATEL analysis are as follows:

Step 1 establishes the average "direct relation matrix." In this step, experts are invited to evaluate each barrier's influence on another using an integer scale. The designed scale has 5 levels, including integers from 0 to 4, where 0 means no impact, 1 means slight impact, 2 means moderate impact, 3 means high impact, and 4 means extremely high impact. Accordingly, the direct influence matrix of each expert is obtained. Then, the average direct relation matrix is obtained by summarizing and averaging all feedback expert data. Given that *k* is the index of experts from a total of *p* experts, *q* is the index of the barriers, and *i* and *j* are the indices for 2 barriers, the decision matrix of each expert is given by  , and then, the direct impact matrix  is given by equation (1).

Step 2 normalizes the average direct relation matrix. The row and maximum value methods are used. The elements of each row in the average direct influence matrix are summed, and then, the maximum value is obtained by comparison. Finally, each element in the average direct relation matrix is divided by the maximum value. The calculation process can be expressed as equation (2):

D=S/x **(2)**

Step 3 calculates the "total influence matrix T" by adding all the direct and indirect effects using equation (3).

Step 4 develops the "cause-and-effect diagram" by adding elements of vector R (row) and vector C (column) using equations (4) and (5), where, R_i is the sum of the row and C_j is the sum of the column of the "total influence matrix." (R_i+C_j) is called the degree of centrality, and (R_i–C_j) is called the degree of cause.

The horizontal and vertical coordinates can be established according to the values of the degrees of centrality and cause. Among them, the degree of centrality is taken as the abscissa, and the degree of cause is taken as the ordinate. Meanwhile, the cause group and effect group are divided according to the positive and negative values of the (R_i–C_j). If the value of (R_i–C_j) is greater than 0, it indicates that this factor has a more significant influence on other factors in the system, and it is classified as a causal factor. If the value of (R_i–C_j) is less than 0, it indicates that other barriers influence this factor greater and attribute it to the outcome factor.

Results

Literature Review

The literature review identified 15 factors influencing the adoption of 5G in health care. Based on the IRT, we divided the barriers to adopting innovation into 5 primary dimensions: expertise, operation, resource, regulation, and market access.

Compared with the unified theory of acceptance and use of technology, technology acceptance model or technology acceptance model 2, and theory of reasoned action, IRT mentioned above has been verified as an effective and significant alternative for researchers who aim to uncover resistance factors in the health care context [22]. The details are listed in Table 3.

Table 3. Barriers influencing adoption of 5G in health care: review of literature.

Barriers	Descriptions
A. Expertise barrier	
A1 Lack of 5G technical talents	5G experts and 5G equipment operators within hospitals are understaffed [23-27].
A2 Insufficient informatization level	The level of informatization construction of different hospitals is uneven. Significant gaps in equipment networking capabilities, medical data collection, and information integration make it challenging to implement and replicate 5G solutions [28,29].
A3 Insufficient security verification	Most of the data in the medical field adopt cross-level and multichannel data collection and analysis methods, including hospital management data and private data such as patient physiology, psychology, and behavior data. The security of the data transmitted through 5G network still needs to be verified [28-35].
B. Operation barrier	
B1 Organizational barriers within hospitals	Not a lot of people understand what 5G is and how it works, and the willingness of traditional hospitals to upgrade and transform 5G networks is relatively low, considering the fact that mature 5G application is mainly concentrated in peripheral medical scenarios such as outpatient guidance and remote consultation [25,28,32,36,37].
B2 Communication obstacles among hospitals	Communication obstacles exist among hospitals, especially among the higher- and lower-level hospitals [24,28,29,38,39].
C. Resource barrier	
C1 High expense	Related equipment and communication costs are high, making it difficult for hospitals to afford [25,30,32,40].
C2 Huge time cost	Installing appropriate equipment and training relevant personnel demand significant time investment [28,40].
C3 Lack of well-trained medical and technical personnel	Existing medical care and technical personnel are insufficient for 5G integration in medical scenarios [33].
C4 Lack of mature compatible equipment and systems	It is difficult for 5G network to integrate with existing equipment and systems [31].
D. Regulation barrier	
D1 Lack of policies related to 5G smart medical integration	Currently, there is no established policy for the integration of 5G smart medical applications [24,28,34].
D2 Lack of ethics, rights, and responsibilities policies	Lack of policies on ethical controversies, rights, and responsibilities related to the application of 5G [28,31,33,34].
D3 Lack of standards for corresponding scenarios	There are many 5G smart medical application scenarios; different scenarios have different requirements for network and technical architecture. At present, there is a lack of 5G application standards corresponding to many medical scenarios [24,29,41-43].
E. Market access barrier	
E1 Lack of unified 5G product standards and listing standards	Emerging 5G smart medical products (such as wearable intelligent terminal equipment and medical instruments) still need unified and perfect listing standards [28,41].
E2 Lack of complete 5G smart medical product system	5G private network equipment and terminal equipment that meet the customized services of smart medical care still need to be further improved [28].
E3 Lack of mature business model	There need to be more mechanisms for cross-field cooperation and mature business models [28].

Delphi Expert Consultation

The Delphi panel in this study comprises 15 members, including practitioners from the health authority, academia, information management departments of the primary, secondary, and tertiary hospitals, and 5G network operators (see Table 4 for panel composition). A 2-round Delphi expert consultation was conducted to explore the views of different experts on the resistance factors toward 5G development in hospitals. The questionnaire was developed based on the literature review.

In the first and second round, we distributed 15 questionnaires each time. In the first round, all the distributed questionnaires were retrieved, while in the second round, 12 questionnaires were collected. The positive coefficient of experts=number of questionnaires returned/number of questionnaires distributed, which can reflect the degree of concern of experts to this study. Thus, in the first round, the positive coefficient of experts was 100%, while in the second round, this coefficient was reduced to 80%.

In the first and second rounds, we obtained data related to the degree of expert authority as follows:

- Degree of familiarity (Cs): 0.670 in the first round and 0.680 in the second round. It shows that the authority of experts in the 2 rounds of consultation is relatively high, and the opinions given are representative to a certain extent.
- Judgment basis (Ca): 0.930 in the first round and 0.920 in the second round.
- Authority coefficient (Cr): 0.800 for both the first and second rounds.

After the first round of expert consultation, the indicator adjustments are as follows, the selection results can be seen in Table 5:

1. Deleted indicators: original C3 (lack of well-trained medical and technical personnel), original C4 (lack of mature compatible equipment and systems), and original E3 (lack of mature business model).
2. Modified indicators: A1 (lack of personnel familiar with 5G within the hospitals) and D2 (lack of policies on ethics, rights, and responsibilities in core health care scenarios).
3. Newly added indicators: B3 (lack of cross-unit resource integration channels) and new C3 (lack of means for hospitals to manage their own 5G networks).

The consultation process of the second round is consistent with the first round. According to the expert's scoring, the CV was calculated, and the W test was carried out. As can be seen in Table 6, none of the CVs for the second round of barriers were higher than 0.250.

In the first and second rounds, we obtained the following data on Kendall coefficient of concordance W test: Kendall coefficient of concordance was 0.195 in the first round and 0.258 in the second round, χ^2_{14}=40.854 in the first round and χ^2_{13}=40.320 in the second round, and *P* value was <.001 for both first and second rounds, which is statistically significant, indicating that the coordination of expert opinions is good. The experts' opinions tended to be unanimous in the second round of consultation, with no modification and new indicators. The final determined barriers can be seen in Figure 2.

Table 4. Basic information of Delphi panelist.

Category	Experts, n	Constituent ratio (%)
Major		
Communication technology	2	13
Computer science and technology	12	80
Health management	1	7
Work experience (years)		
<10	1	7
10-19	10	67
20-29	4	26
Professional title		
Intermediate	1	7
Vice senior	11	73
Senior	3	20
Job description		
Information technology operations management	10	67
Academia	2	13
Administrative management	3	20

Table 5. Selection results of the first round of expert consultation.

	A1	A2	A3	B1	B2	C1	C2	C3	C4	D1	D2	D3	E1	E2	E3
Mean (SD)	3.867 (1.024)	4.267 (0.573)	3.867 (0.806)	4.333 (0.789)	3.200 (0.748)	4.133 (0.884)	3.533 (0.718)	3.267 (1.062)	3.467 (1.087)	3.867 (0.957)	3.467 (0.957)	4.267 (0.573)	4.000 (0.632)	4.133 (0.806)	3.667 (1.247)
CV[a]	0.265	0.134	0.208	0.182	0.234	0.214	0.203	0.325	0.314	0.247	0.276	0.134	0.158	0.195	0.340
Selection criterion (CV≤0.250)	V*[b]	V[c]	V	V	V	V	V	✓[d]	✓	V	V*	V	V	V	✓

[a]CV: coefficient of variation.

[b]The symbol "V*" indicates that the indicator is modified.

[c]The symbol "V" indicates that the indicator is retained.

[d]The symbol "✓"indicates that the indicator is deleted.

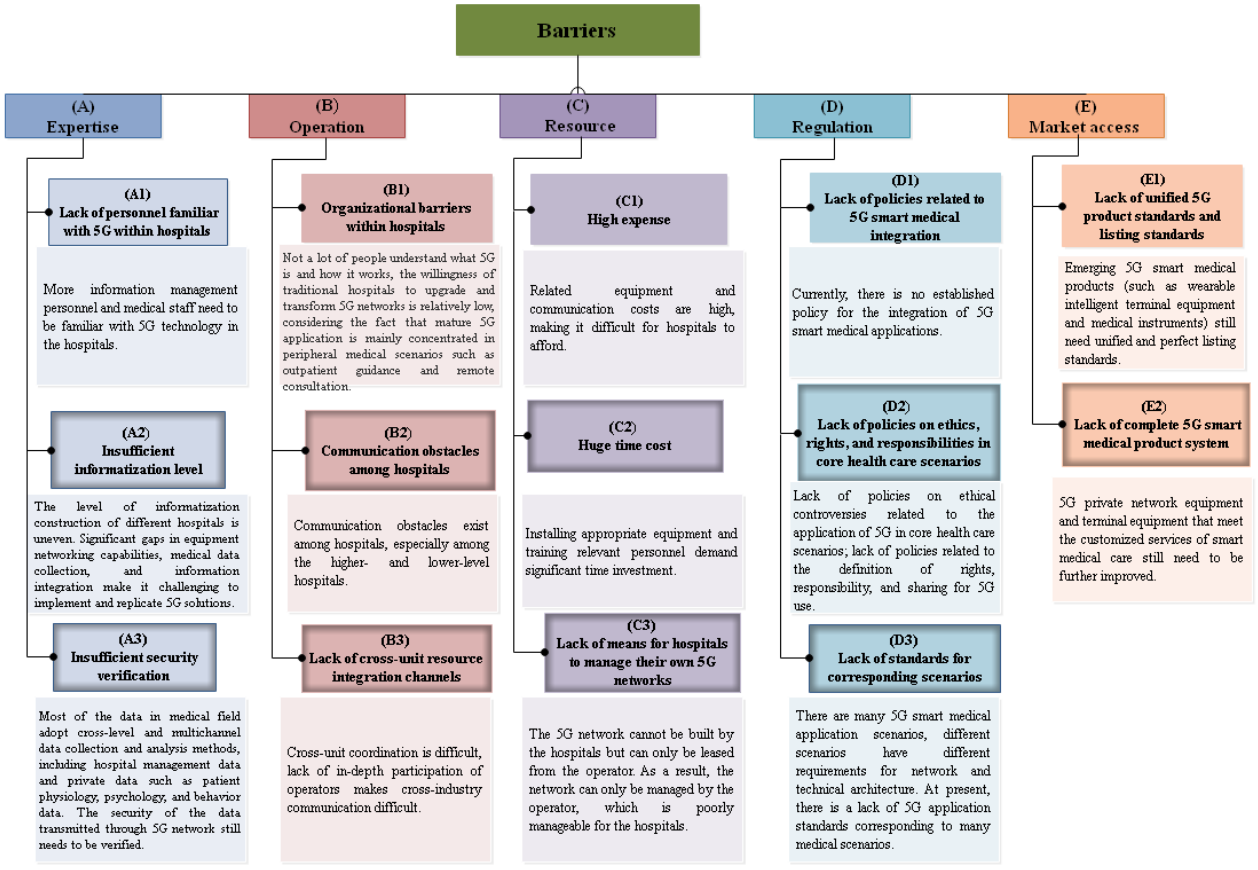
Table 6. Selection results of the second round of expert consultation.

	A1	A2	A3	B1	B2	B3	C1	C2	C3	D1	D2	D3	E1	E2
Mean (SD)	3.917 (0.640)	4.167 (0.553)	3.917 (0.862)	4.750 (0.433)	3.333 (0.745)	4.250 (0.595)	4.167 (0.799)	3.500 (0.866)	3.333 (0.745)	4.167 (0.898)	3.667 (0.850)	4.083 (0.759)	4.000 (0.707)	4.417 (0.759)
CV[a]	0.163	0.133	0.220	0.091	0.224	0.140	0.192	0.247	0.224	0.215	0.232	0.186	0.177	0.172
Selection criterion (CV≤0.250)	V[b]	V	V	V	V	V	V	V	V	V	V	V	V	V

[a]CV: coefficient of variation.

[b]The symbol "V" indicates that the indicator is retained.

Figure 2. Adoption barriers of 5G for hospitals in Western China.



Results of ISM

Establish SSIM

In total, 10 experts from the abovementioned 15 were invited to develop the relationships between different barriers using L, M, N, and O. Due to the nonexistence of a special criterion in the literature of ISM concerning what establishes the majority and as in the methodology of ISM [44], this study adopted a 50% criterion to determine the final relationships between different barriers, such as in a relation, if 5 from 10 experts ticked L, the corresponding column is defined as L. However, to be unbiased, for this kind of relationship, the number of specialists answering by M or N or O should be lesser than or equivalent to 3. Fulfilling both requirements, 47 of 91 cases of relations were included to obtain final results, and the remaining cases were all taken as biased and to be taken for more discussion with experts. To analyze the relations of the remaining barriers, this study proceeded for more discussion with 2 senior experts. To eliminate biases, the selected experts were taken out of the given pool of 10 experts. The 2 senior experts were requested to analyze and discuss the remaining barriers until a consensus was reached. Doing so determined a final relationship between different barriers called SSIM, as shown in Table S1 in Multimedia Appendix 1.

Formation of IRM

Transformation of the SSIM into IRM uses the binary rule shown in Table 2. For example, the entry of A1 and D2 in the SSIM is “O,” which is replaced by “0” for A1 and D2 and “0” for D2 and A1 in the IRM (Table S2 in Multimedia Appendix 1).

Formation of FRM

FRM is formed after checking IRM for transitivity. It was done to represent all indirect connections to maintain the consistency of relationships among the barriers. For example, there is a direct relation between A1 and C2 and C2 and D1, but there is no relation between A1 and D1, as shown in the SSIM. Hence, according to the transitivity rule, there is an indirect relation between A1 and D1 corrected during the formation of FRM. It can be observed in Table S3 in Multimedia Appendix 1, where the relation of A1 and D1 is represented by 1*. All of the asterisk signs represent the indirect relation rectified during the formation of the FRM. The FRM calculates each barrier's DP and DEP. The DP is the summation of the value of all the row elements, while the DEP is the summation of all the column elements corresponding to the respective barrier.

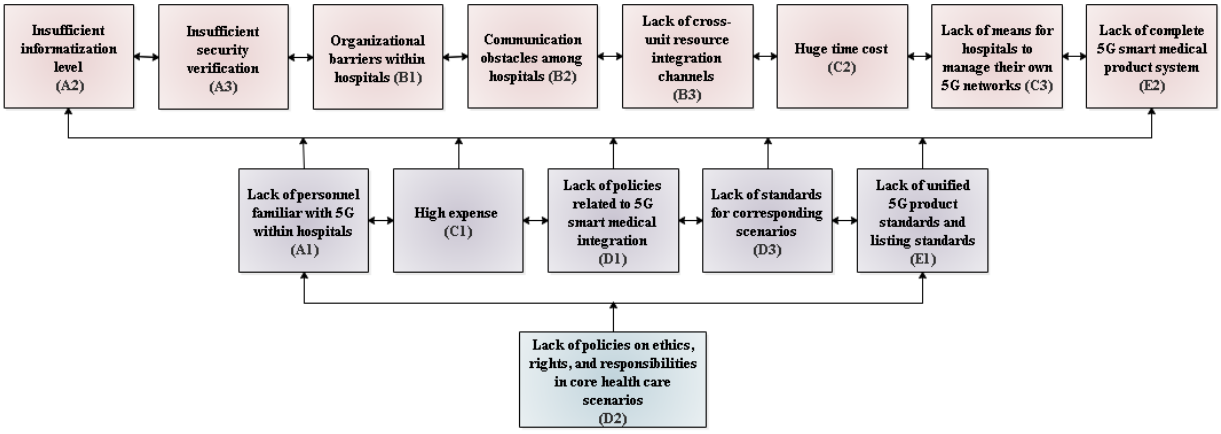
Level Partition

To have a clearer understanding of the relationship between the barriers, a hierarchical structure of the factors is required. Based on FRM, the reachability set, antecedent set, and intersection set for each barrier were developed. Suppose the reachability and intersection set for a specific barrier are identical. In that case, that barrier is deemed at level 1 and assigned the highest position in the ISM hierarchy. After the first iteration, the barriers constituting level 1 are removed, and the previously mentioned procedure is repeated with the remaining barriers until the levels of all barriers have been determined. The results of the different sets and the level iterations are shown in Table S4 in Multimedia Appendix 1.

Formation of ISM

ISM is formulated based on the partition level of barriers. In the first iteration, A2 (insufficient informatization level), A3 (insufficient security verification), B1 (organizational barriers within the hospitals), B2 (communication obstacles among hospitals), B3 (lack of cross-unit resource integration channels), C2 (huge time cost), C3 (lack of means for hospitals to manage their own 5G networks), and E2 (lack of complete 5G smart medical product system) were placed at the top of the ISM. The second iteration resulted in second-level barriers involving A1 (lack of personnel familiar with 5G within hospitals), C1 (high expense), D1 (lack of policies related to 5G smart medical integration), D3 (lack of standards for corresponding scenarios), and E1 (lack of unified 5G product standards and listing standards) placed below the first level. Similarly, in the third iteration, D2 (lack of policies on ethics, rights, and responsibilities in core health care scenarios) was placed below the second level. The developed framework or ISM of barrier adoption is shown in Figure 3.

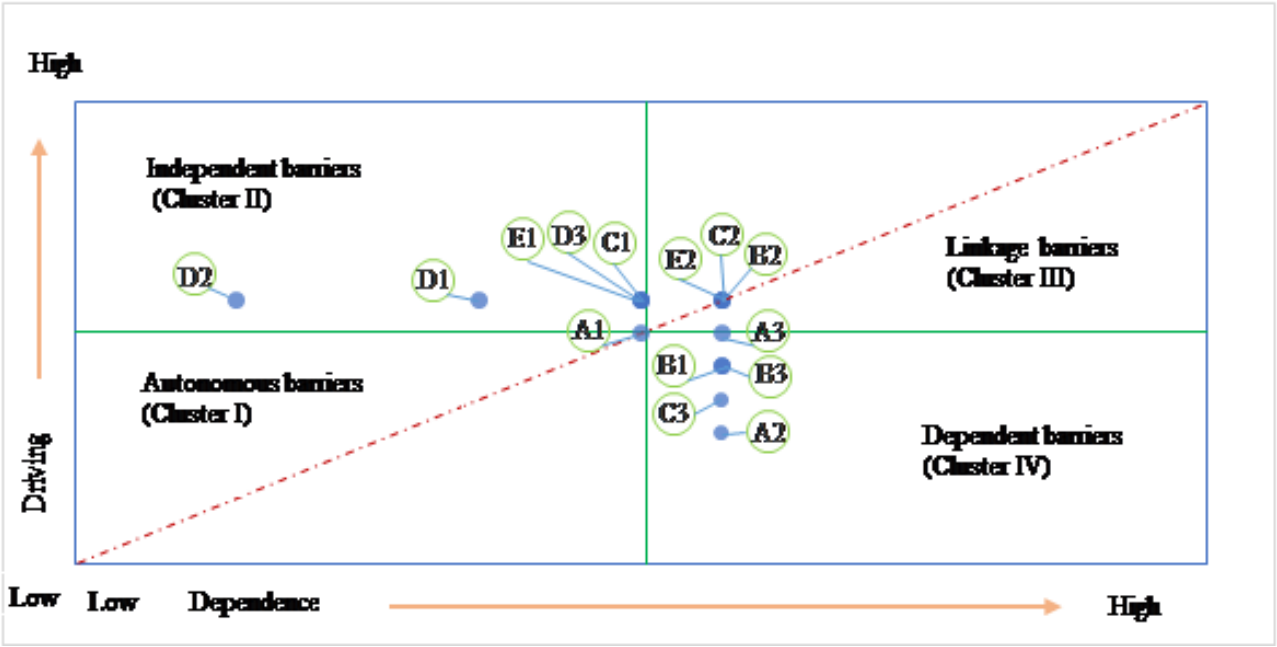
Figure 3. Interpretive structural model.



MICMAC Analysis

The MICMAC analysis is performed to identify barriers' DP and DEP and classify them accordingly. As shown in Figures 3 and 4, A2 (insufficient informatization level), A3 (insufficient security verification), B1 (organizational barriers within hospitals), B3 (lack of cross-unit resource integration channels), and C3 (lack of means for hospitals to manage their own 5G networks) were placed at the top of the ISM and fell under the "dependent" cluster. B2 (communication obstacles among hospitals), C2 (huge time cost), and E2 (lack of complete 5G smart medical product system) were categorized under the "linkage" cluster. The barriers under the linkage cluster were volatile due to high DP and DEP. C1 (high expense), D1 (lack of policies related to 5G smart medical integration), D2 (lack of policies on ethics, rights, and responsibilities in core health care scenarios), D3 (lack of standards for corresponding scenarios), and E1 (lack of unified 5G product standards and listing standards) were placed in the independent cluster. Considering these barriers as drivers of other barriers in the system, hospitals should prioritize them in their decision-making processes. In addition, D2 (lack of policies on ethics, rights, and responsibilities in core health care scenarios) has a relatively high driving force and low dependence force, which reveals that it strongly impacts the whole system as displayed in the ISM.

Figure 4. Driving and dependence diagram.

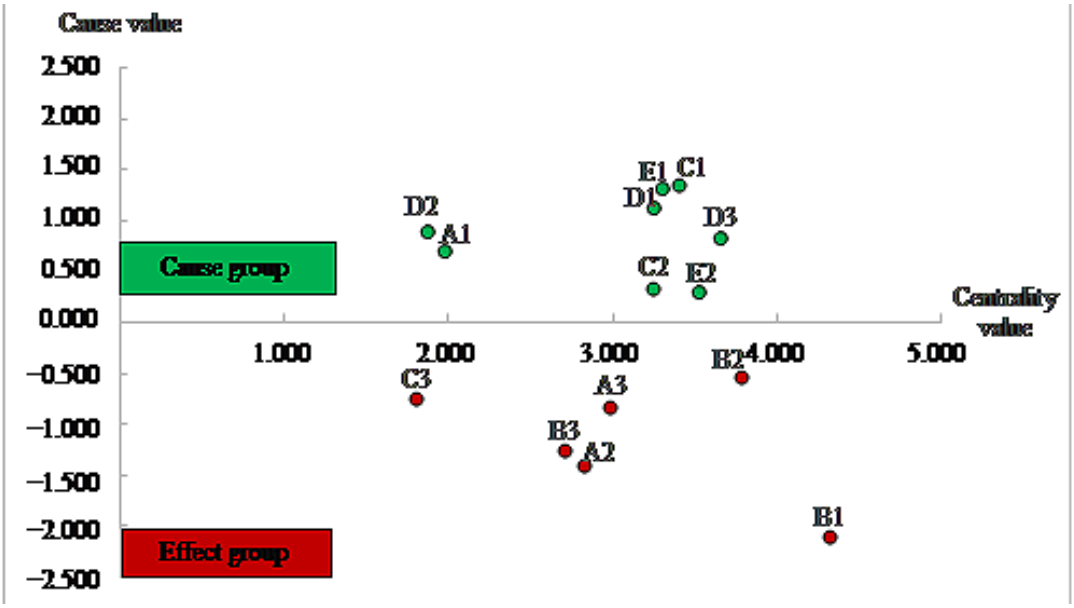


Results of DEMATEL

The abovementioned 10 experts in the ISM scoring process were also invited to participate in the data collection for the DEMATEL analysis. Experts were invited to evaluate each barrier's influence on another using a scale of 0-4. After collecting the direct relation matrix of each expert, the average direct relation matrix (Table S5 in Multimedia Appendix 1) was obtained by summarizing and averaging all feedback expert data. Then, the direct relation matrix was converted into a normalized direct relation matrix (Table S6 in Multimedia Appendix 1) using equation (2). Furthermore, the normalized matrix was converted into a total influence matrix (Table S7 in Multimedia Appendix 1) using equation (3). Finally, the degree of influence was calculated using equations (4) and (5). The cause-effect matrix is shown in Table S8 in Multimedia Appendix 1.

The barriers with an R–C value less than 0 were identified as the effect group, while barriers with an R–C value greater than 0 fell under the cause group. As shown in Figure 5, a total of 8 barriers could be classified in the "cause group," and 6 as the "effect group," in which C1 (high expense), E1 (lack of unified 5G product standards and listing standards), D1 (lack of policies related to 5G smart medical integration), and D2 (lack of policies on ethics, rights, and responsibilities in core health care scenarios) took high priority in the causal group, B1 (organizational barriers within hospitals) and A2 (insufficient informatization level) were the most influenced barriers.

Figure 5. Diagram of cause-effect of barriers.



Sensitivity Analysis

The sensitivity analysis was performed to verify the consistency of the calculated value and validate the stability of professional judgment. As shown in Table 7, a different weighting was assigned to one expert's feedback while keeping equal weightings for the other. Four different total relationship matrices and comparable matrixes were created by multiplying each weight assigned to the experts, the average relationship matrices were then computed, and the cause-effect relationships among the different barriers were established. As shown in Table 8, the same rank order for cause-effect barriers for each expert was obtained, accepting minor rank order variation. Based on Figure 6, the plots of all barriers during 4 iterations of sensitivity analysis are the same as the base rank. Therefore, it is clear that there was no major change in barrier rankings. The sensitivity analysis confirms the robustness of obtained results.

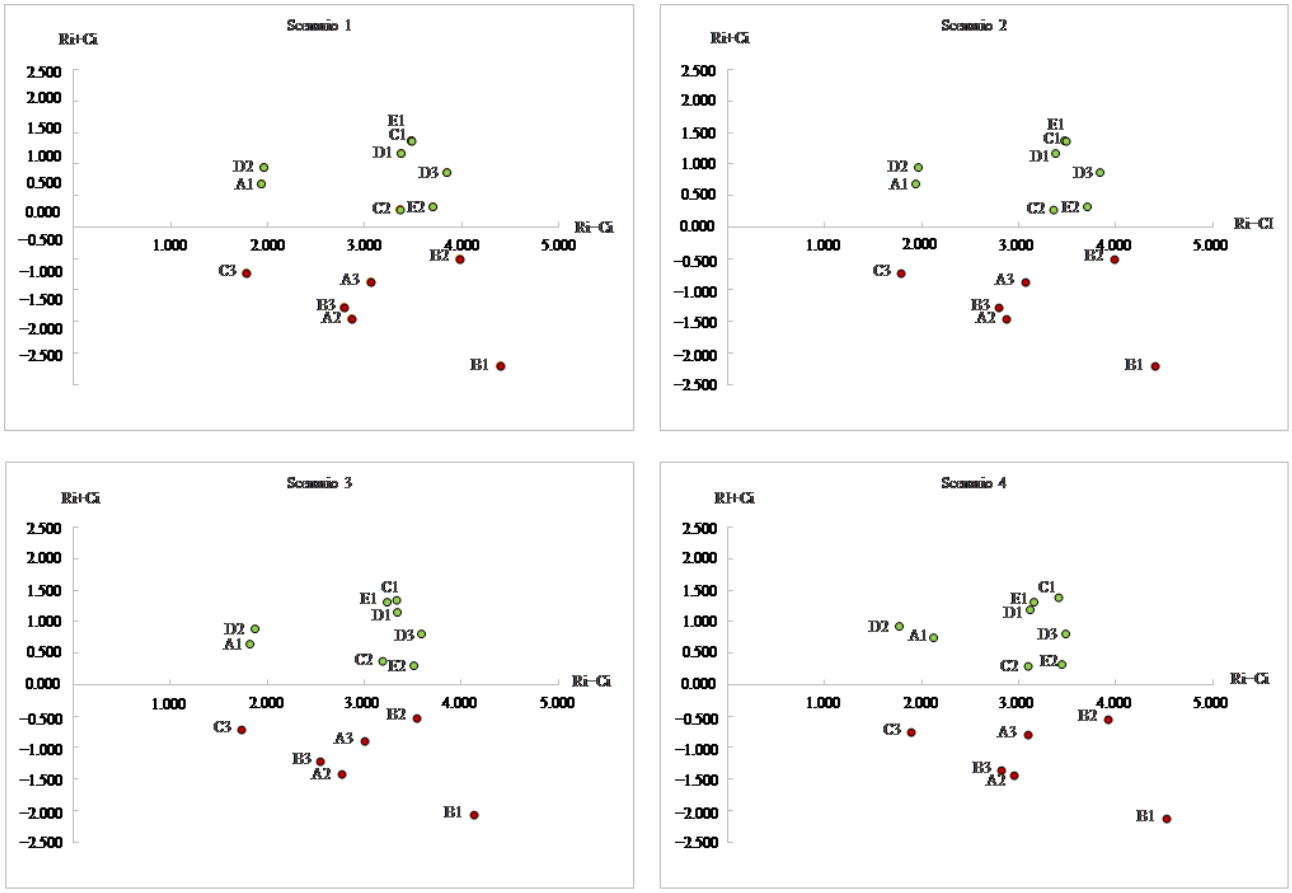
Table 7. Results of sensitivity analysis.

	Scenario 1		Scenario 2		Scenario 3		Scenario 4	
	R+C	R–C	R+C	R–C	R+C	R–C	R+C	R–C
A1	1.939	0.678	1.939	0.678	1.824	0.643	2.115	0.742
A2	2.862	–1.448	2.862	–1.448	2.761	–1.413	2.947	–1.435
A3	3.062	–0.864	3.062	–0.864	3.000	–0.893	3.086	–0.785
B1	4.401	–2.207	4.401	–2.207	4.127	–2.056	4.518	–2.123
B2	3.982	–0.510	3.982	–0.510	3.533	–0.535	3.914	–0.543
B3	2.792	–1.266	2.792	–1.266	2.544	–1.212	2.812	–1.349
C1	3.482	1.372	3.482	1.372	3.329	1.349	3.411	1.383
C2	3.358	0.287	3.358	0.287	3.177	0.387	3.091	0.298
C3	1.780	–0.720	1.780	–0.720	1.724	–0.710	1.879	–0.748
D1	3.379	1.171	3.379	1.171	3.341	1.153	3.120	1.195
D2	1.962	0.955	1.962	0.955	1.876	0.882	1.767	0.921
D3	3.843	0.863	3.843	0.863	3.588	0.811	3.479	0.815
E1	3.474	1.373	3.474	1.373	3.240	1.303	3.162	1.315
E2	3.701	0.316	3.701	0.316	3.504	0.293	3.450	0.315

Table 8. Ranking obtained after sensitivity analysis.

	Inputs for sensitivity analysis			
	Scenario 1	Scenario 2	Scenario 3	Scenario 4
Expert 1	0.4	0.2	0.2	0.2
Expert 2	0.2	0.4	0.2	0.2
Expert 3	0.2	0.2	0.4	0.2
Expert 4	0.2	0.2	0.2	0.4

Figure 6. Sensitivity analysis plots.



Discussion

Principal Results

As shown in Figures 3 and 5, all the barriers in the bottom and middle levels fell into the cause group, indicating those barriers are the primary reasons why 5G technology cannot be adopted in hospitals. In particular, D2 (lack of policies on ethics, rights, and responsibilities in core health care scenarios) was the only factor at the bottom level, implying it is the underlying influencing factor of the whole system. It also has a relatively high driving force and low dependence force based on the MICMAC analysis, which is consistent with the analysis results of DEMATEL as it took high priority in the causal group with an R–C value of 0.900 and an R+C value of 1.875. The finding aligns with that of [28], which highlights that medical and health care fields are closely linked to people's safety. Additionally, ethical considerations and definitions of responsibilities and risks are the most significant obstacles to the evolution and development of 5G in the core areas of health care. It is suggested that the government should strengthen policy supervision to ensure the safe and ethical use of 5G in the medical field. Specific measures need to be implemented to supervise and regulate the application of 5G in the medical field. In addition, further research on relevant policies and norms is required. At the same time, clarification should be made in policy documents, laws, and regulations.

Another significant obstacle in promoting 5G medical applications is the massive capital investment required, as stated in C1, with the highest R–C value of 1.361, indicating its most significant driving force over other barriers. This finding reflects the facts stated by CN-HEALTHCARE [45]. The overall cost of 5G medical treatment includes the construction of 5G communication networks by operators, procurement and maintenance costs for 5G medical information systems and related medical equipment, purchase of medical care equipment, and services for users. The application of 5G in hospitals also

requires installing indoor base stations, software support, and computer room upgrades, all of which incur significant expenses. The high cost of 5G applications limits the financing capabilities of hospitals, especially in the Western China region, where hospitals are generally smaller and have limited funding. As a result, only a few large hospitals with telecom operator support have been able to implement 5G medical care. Meanwhile, smaller hospitals are expected to undertake 5G telemedicine with larger tertiary hospitals. Thus, at this stage, it is recommended that smaller hospitals increase their collaboration with larger hospitals to accelerate the adoption of 5G. At the same time, the government is essential to develop affordable solutions and provide financial support for 5G adoption in health care.

With the highest R+C value of 4.317, B1 (organizational barriers within hospitals) is the most closely related barrier to other factors, indicating that it is the most important barrier to adopting 5G in hospitals in Western China. It also has the lowest R–C value of –2.107, which means it is in the effect group and the most influenced barriers by other factors. These results echo the findings of CN-HEALTHCARE [46], which identified a lack of mature 5G applications in medical scenarios, varying levels of acceptance of new technology, concerns about 5G security, and limitations in human, financial, and material resources as factors influencing attitudes toward 5G adoption among different hospitals. Organizational barriers may be addressed through effective management and leadership, clear communication, and collaboration among different departments.

E1 (lack of unified 5G product standards and listing standards) has the second-highest R–C value of 1.316 and a relatively high R+C value of 3.300. The high R–C value specifies its driving force over the other barriers, while the high R+C value indicates its strong impact on the adoption process. This outcome is akin to the findings of Bruer and Doug [47], who mentioned that 5G standards play a crucial role in hardware infrastructure to software running on top of components, and unified 5G standards help to ensure that a range of devices and equipment can operate in a shared system. Therefore, developing unified 5G product standards and listing standards for corresponding scenarios are essential. It is recommended that the Chinese government, 5G network operators, and hospital administrators accelerate cooperation to establish unified 5G product standards and list standards to jointly promote the large-scale development of 5G in the health care sector.

B2 (communication obstacles among hospitals) is a part of the linkage cluster with the second-highest R+C value of 3.782, indicating it is one of the crucial factors in the whole system. This finding is consistent with that of Wang et al [39], who found that effective communication is critical for the success of 5G adoption in health care. Communication obstacles can lead to misunderstandings and a lack of trust. It is urgent to break through the communication obstacles among hospitals, especially between the higher- and lower-level hospitals, which need to accelerate top-level design, formulate policy documents, and improve relevant legislation and management mechanisms to promote the opening and sharing of 5G medical data and ensure the deep integration of the 5G medical industry.

Concluding Remarks

The paper comprehensively analyzes barriers to 5G adoption in hospitals in Western China. Experts from different stakeholders validated 14 resistance factors. Next, an integrated ISM-DEMATEL approach was applied to model the barriers as a network of factors and alternatives categorized into clusters. All barriers were related to human expertise, resource allocation, operational procedures, laws and regulations, and market access capability. Overall, the study shows that promoting the integration of 5G in hospitals in Western China faces multiple complex and interrelated barriers. It constructs a framework for the main barriers to 5G adoption in the hospital context and provides regulatory authorities, hospital managers, and telecom operators with theoretical and managerial insights into the strategic pathways.

Theoretical and Managerial Implications

- The barrier at the bottom level of the ISM should be emphasized for short-term strategy. The middle-level barriers can be considered for medium- and long-term strategies. The barriers at the top of the ISM can be a long-term strategic focus.
- The effect group can easily be influenced by the cause group, and therefore, managers should give the most attention to causal barriers when implementing 5G practices in hospitals.
- The ranking of cause-effect group barriers can assist regulatory authorities, hospital managers, and telecom operators in developing strategic policy during 5G implementation.
- To overcome these barriers, hospital managers should formulate a 5G adoption strategy that considers the specific needs of the institution and the local market. Specific measures include increasing investment in information infrastructure and human resources, establishing supplier communication channels, and promoting cross-unit resource integration.
- To expand the use of 5G in health care scenarios, it is recommended that the government accelerates the construction of an innovation system consisting of regulators, hospitals, telecom operators, academic researchers, and patient representatives.

Outlook of 5G Health Care

As the infrastructure of intelligent medical care, 5G allows the transmission of vast amounts of data and information, making the information superhighway a reality. Furthermore, with the full deployment of 5G medical care, especially the integration with big data, artificial intelligence, internet, internet of things, and blockchain technology, 5G is expected to bring significant changes to the current medical and health system and promote the evolution of the entire medical ecology, including hospital operation and management. In the long run, 5G health care promotes the sinking of high-quality medical resources and the development of China's “primary health care” and “family doctor” systems. It can improve the population's overall health, reduce medical expenses for ordinary people, and relieve medical insurance burdens. Therefore, no matter how barriers are faced, the benefits of 5G medical care are expected to

outweigh the costs, making it a worthy investment. In the development of 5G medical care, various technical, economic, institutional, interest, and ethical problems will inevitably be encountered. The regulatory authorities, hospitals, telecom operators, and the public must cooperate to explore and solve the problems in the 5G medical care era, aiming to achieve the coverage of 5G medical care across the country.

Limitations

Although this study suggests some useful implications, there are some limitations that could be considered for future research. First, due to the challenges in reaching out to health care professionals and telecom operators from all regions of China, most of the experts invited for this research were from representative cities in Northwest and Southwest China, including Xi'an, Chongqing, and Chengdu. This may not fully represent the entire country, as Eastern China is generally more developed than Western China. Therefore, there is a lack of balance among the groups of participants in this research. In future research, we plan to invite experts from Eastern China as a complement study. Second, the relations established among barriers might be biased because they are selected and analyzed based on expert opinions that are context-dependent and depend on their organization's culture and experience. Third, the outcome of this study is valid for the Chinese health care field and cannot be generalized for other sectors without modifications. It can be extended from the Chinese context to a broader coverage by selecting experts from different countries for benchmarking studies. Finally, only 4 groups of experts, namely, government information department staff, managers from the hospital information technology departments, telecom operators, and scholars, are involved in the research process. Other vital stakeholders, such as patients, can also provide crucial information and insights related to the development of 5G health care.

Acknowledgments

The authors acknowledged the support of the Key Research and Development Projects of Shaanxi Province, China (grant 2021SF-188), and the Youth Fund of the Second Affiliated Hospital of Xi'an Jiaotong University, China (grant YJ(QN)202018). The authors also acknowledged the support of Mr Duwei Dai, Ms Caixia Dong, and Ms Chunyan Zhang during the investigation and data analysis process.

Authors' Contributions

LZ was responsible for the initial concept development, formal analysis, and writing the first draft of the study. MJ contributed to the development of the research methods and edited the writing. RD and FZ were involved in data collection and curation. ZL and SX provided supervision and reviewed the writing. All authors contributed to the final version of the paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interpretive structural modeling model and decision-making trial and evaluation laboratory method operation process.
[DOCX File , 45 KB - mhealth_v12i1e48842_app1.docx]

References

1. Medical Intelligence Industry White Paper. Frost & Sullivan. URL: https://www.cn-healthcare.com/articlewm/20220526/content-1363721.html?prefer_safari=1&prefer_reader_view=12022 [accessed 2022-11-21]
2. Dananjayan S, Raj GM. 5G in healthcare: how fast will be the transformation? Ir J Med Sci 2021;190(2):497-501 [FREE Full text] [doi: 10.1007/s11845-020-02329-w] [Medline: 32737688]
3. White Paper on Smart Healthcare in 5G Era. China Internet Healthcare Industry Alliance. URL: http://news.sohu.com/a/371939775_120099892 [accessed 2022-11-21]
4. Sailing Action Plan for 5G Applications (2021-2023). MIIT. 2021. URL: http://www.gov.cn/zhengce/zhengceku/2021-07/13/content_5624610.htm [accessed 2022-11-21]
5. Notice on Publishing 5G + Medical and Health Application Pilot Projects. MIIT. 2021. URL: http://www.gov.cn/zhengce/zhengceku/2021-09/30/content_5640341.htm [accessed 2022-11-21]
6. The province's first 5G internet hospital goes online, and the integrated smart hospital makes medical care warmer. Guangdong Provincial People's Hospital. URL: https://www.gdghospital.org.cn/NewsHospital/info_itemid_32014.html [accessed 2022-11-22]
7. Shanghai: By 2023, 100% of all tertiary hospitals will achieve 5G deep coverage and 5G typical services online. Shanghai Communications Administration. URL: https://www.cn-healthcare.com/articlewm/20220331/content-1332938.html [accessed 2022-11-22]
8. Run Run Shaw Hospital of Zhejiang University: China's first 5G ultra-remote robot liver and gallbladder surgery completed. Guangming Daily. URL: https://baijiahao.baidu.com/s?id=1758048205578260864&wfr=spider&for=pc [accessed 2023-03-07]

9. Sheth JN, Ram S. Bringing innovation to market: how to break corporate and customer barriers. J Mark 1989;53(2):131-132. [doi: 10.2307/1251421]
10. Zhao Y. Research on the regional development of 5G technology application in China. Electron Compon Inf Technol 2022;6(02):217-218. [doi: 10.19772/j.cnki.2096-4455.2022.2.083]
11. Regulation for Ethical Review of Life Sciences and Medical Research involving Human Beings. Chinese National Health Commission, Ministry of Education, Ministry of Science and Technology, and State Administration of Traditional Chinese Medicine. URL: http://www.nhc.gov.cn/qjjys/s7946/202302/c3374c180dc5489d85f95df5b46afaf5.shtml [accessed 2023-03-28]
12. Yadav VS, Singh AR, Raut RD, Govindarajan UH. Blockchain technology adoption barriers in the Indian agricultural supply chain: an integrated approach. Resour Conserv Recycl 2020;161:104877 [FREE Full text] [doi: 10.1016/j.resconrec.2020.104877]
13. Sharma M, Sehrawat R, Daim T, Shaygan A. Technology assessment: enabling blockchain in hospitality and tourism sectors. Technol Forecast Soc Change 2021;169:120810 [FREE Full text] [doi: 10.1016/j.techfore.2021.120810]
14. McMillan SS, King M, Tully MP. How to use the nominal group and Delphi techniques. Int J Clin Pharm 2016;38(3):655-662 [FREE Full text] [doi: 10.1007/s11096-016-0257-x] [Medline: 26846316]
15. Jorm AF. Using the Delphi expert consensus method in mental health research. Aust N Z J Psychiatry 2015;49(10):887-897 [FREE Full text] [doi: 10.1177/0004867415600891] [Medline: 26296368]
16. Niederberger M, Spranger J. Delphi technique in health sciences: a map. Front Public Health 2020;8:457 [FREE Full text] [doi: 10.3389/fpubh.2020.00457] [Medline: 33072683]
17. Liu L, Wei K, Zhang X, Wen D, Gao L, Lei J. The current status and a new approach for Chinese doctors to obtain medical knowledge using social media: a study of WeChat. Wirel Commun Mob Comput 2018;2018:1-10 [FREE Full text] [doi: 10.1155/2018/2329876]
18. Warfield JN. Developing interconnection matrices in structural modeling. IEEE Trans Syst Man Cybern 1974;SMC-4(1):81-87. [doi: 10.1109/tsmc.1974.5408524]
19. Kumar S, Raut RD, Agrawal N, Cheikhrouhou N, Sharma M, Daim T. Integrated blockchain and internet of things in the food supply chain: adoption barriers. Technovation 2022;118:102589 [FREE Full text] [doi: 10.1016/j.technovation.2022.102589]
20. Malone DW. An introduction to the application of interpretive structural modeling. Proc IEEE 1975;63(3):397-404. [doi: 10.1109/proc.1975.9765]
21. Wu WW, Lee YT. Developing global managers' competencies using the fuzzy DEMATEL method. Expert Syst Appl 2007;32(2):499-507 [FREE Full text] [doi: 10.1016/j.eswa.2005.12.005]
22. Chen PT. Medical big data applications: intertwined effects and effective resource allocation strategies identified through IRA-NRM analysis. Technol Forecast Soc Change 2018;130:150-164 [FREE Full text] [doi: 10.1016/j.techfore.2018.01.033]
23. Zhang J, Lu Q, Shi L. The influence of telemedicine on capacity development in public primary hospitals in China: a scoping review. Clinical eHealth 2022;5:91-99 [FREE Full text] [doi: 10.1016/j.ceh.2022.10.001]
24. Xue L. 5G and public management: opportunities and challenges. Dev Admin 2022;4(1):31-37. [doi: 10.46996/dina.v4i1.6431]
25. Cepal NU. Digital Technologies for a New Future. 2021. URL: https://hdl.handle.net/11362/46817 [accessed 2022-11-25]
26. Chen R, Zheng Y, Xu X, Zhao H, Ren J, Tan HZ. STEM teaching for the internet of things maker course: a teaching model based on the iterative loop. Sustainability 2020;12(14):5758 [FREE Full text] [doi: 10.3390/su12145758]
27. Labrique AB, Wadhwani C, Williams KA, Lamptey P, Hesp C, Luk R, et al. Best practices in scaling digital health in low and middle income countries. Global Health 2018;14(1):103 [FREE Full text] [doi: 10.1186/s12992-018-0424-z] [Medline: 30390686]
28. White Paper on 5G Smart Medical Industry Development. Beijing Telecommunication Technology Development Industry Association. 2022. URL: https://tdia.cn/Home/article/detail/id/446.html [accessed 2022-11-25]
29. 5G technology adds wings to smart healthcare. Xinhua Net. 2022. URL: https://baijiahao.baidu.com/s?id=1646782345941623787&wfr=spider&for=pc [accessed 2022-11-26]
30. Moglia A, Georgiou K, Marinov B, Georgiou E, Berchiolli RN, Satava RM, et al. 5G in healthcare: from COVID-19 to future challenges. IEEE J Biomed Health Inform 2022;26(8):4187-4196 [FREE Full text] [doi: 10.1109/JBHI.2022.3181205] [Medline: 35675255]
31. Guo C, Li H. Application of 5G network combined with AI robots in personalized nursing in China: a literature review. Front Public Health 2022;10:948303 [FREE Full text] [doi: 10.3389/fpubh.2022.948303] [Medline: 36091551]
32. Li G, Lian W, Qu H, Li Z, Zhou Q, Tian J. Improving patient care through the development of a 5G-powered smart hospital. Nat Med 2021;27(6):936-937 [FREE Full text] [doi: 10.1038/s41591-021-01376-9] [Medline: 34017134]
33. Siriwardhana Y, Gür G, Ylianttila M, Liyanage M. The role of 5G for digital healthcare against COVID-19 pandemic: opportunities and challenges. ICT Express 2021;7(2):244-252 [FREE Full text] [doi: 10.1016/j.icte.2020.10.002]
34. Zheng J, Wang Y, Zhang J, Guo W, Yang X, Luo L, et al. 5G ultra-remote robot-assisted laparoscopic surgery in China. Surg Endosc 2020;34(11):5172-5180 [FREE Full text] [doi: 10.1007/s00464-020-07823-x] [Medline: 32700149]
35. Li D. 5G and intelligence medicine-how the next generation of wireless technology will reconstruct healthcare? Precis Clin Med 2019;2(4):205-208 [FREE Full text] [doi: 10.1093/pcmedi/pbz020] [Medline: 31886033]

36. Pifer R. As 5G looms, most hospitals watch from the sidelines. Healthcaredive. 2019. URL: https://www.healthcaredive.com/news/as-5g-looms-most-hospitals-watch-from-the-sidelines/550103/ [accessed 2022-11-28]
37. Ngoc CT, Bigirimana N, Muneene D, Bataringaya JE, Barango P, Eskandar H, et al. Conclusions of the digital health hub of the Transform Africa Summit (2018): strong government leadership and public-private-partnerships are key prerequisites for sustainable scale up of digital health in Africa. BMC Proc 2018;12(Suppl 11):17 [FREE Full text] [doi: 10.1186/s12919-018-0156-3] [Medline: 30540290]
38. Sharma R, Kshetri N. Digital healthcare: historical development, applications, and future research directions. Int J Inf Manage 2020;53:102105 [FREE Full text] [doi: 10.1016/j.ijinfomgt.2020.102105]
39. Wang LW, Chen RC, Zhong NS. Let medicine take on the wings of 5G. Zhonghua Yi Xue Za Zhi 2020;100(16):1201-1204. [doi: 10.3760/cma.j.cn112137-20200107-00038] [Medline: 32344493]
40. Oughton EJ, Frias Z. The cost, coverage and rollout implications of 5G infrastructure in Britain. Telecommun Policy 2018;42(8):636-652 [FREE Full text] [doi: 10.1016/j.telpol.2017.07.009]
41. Qureshi HN, Manalastas M, Ijaz A, Imran A, Liu Y, Al Kalaa MO. Communication requirements in 5G-enabled healthcare applications: review and considerations. Healthcare (Basel) 2022;10(2):293 [FREE Full text] [doi: 10.3390/healthcare10020293] [Medline: 35206907]
42. Zhai Y, Xu X, Chen B, Lu H, Wang Y, Li S, et al. 5G-network-enabled smart ambulance: architecture, application, and evaluation. IEEE Network 2021;35(1):190-196. [doi: 10.1109/mnet.011.2000014]
43. Kumar A, Dhanagopal R, Albreem MA, Le DN. A comprehensive study on the role of advanced technologies in 5G based smart hospital. Alex Eng J 2021;60(6):5527-5536 [FREE Full text] [doi: 10.1016/j.aej.2021.04.016]
44. Usmani MS, Wang J, Ahmad N, Ullah Z, Iqbal M, Ismail M. Establishing a corporate social responsibility implementation model for promoting sustainability in the food sector: a hybrid approach of expert mining and ISM-MICMAC. Environ Sci Pollut Res Int 2022;29(6):8851-8872. [doi: 10.1007/s11356-021-16111-7] [Medline: 34494192]
45. 5G healthcare: pain points, problems, applications and prospects. CN-HEALTHCARE. URL: https://www.sohu.com/a/426797130_139908 [accessed 2023-03-22]
46. Bottom line "5G+medical care" craze: just a need for hospitals or a commercial gimmick? CN-HEALTHCARE. URL: http://tinyurl.com/3v5r72uy [accessed 2023-03-25]
47. Bruer A, Doug B. Mapping the international 5G standards landscape and how it impacts US strategy and policy. Information Technology & Innovation Foundation. URL: https://itif.org/publications/2021/11/08/mapping-international-5g-standards-landscape-and-how-it-impacts-us-strategy/ [accessed 2023-03-25]

Abbreviations

CV: coefficient of variation
DEMATEL: decision-making trial and evaluation laboratory
DEP: dependence power
DP: driving power
FRM: final reachability matrix
IRM: initial reachability matrix
IRT: innovation resistance theory
ISM: interpretive structural modeling
MICMAC: Matriced Impacts Corises-multiplication Appliance Classement
SSIM: structural self-interaction matrix

Edited by L Buis; submitted 10.05.23; peer-reviewed by E Georgiou, C Hudak; comments to author 08.06.23; revised version received 15.06.23; accepted 05.12.23; published 23.01.24.

Please cite as:
Zhou L, Jiang M, Duan R, Zuo F, Li Z, Xu S
Barriers and Implications of 5G Technology Adoption for Hospitals in Western China: Integrated Interpretive Structural Modeling and Decision-Making Trial and Evaluation Laboratory Analysis
JMIR Mhealth Uhealth 2024;12:e48842
URL: https://mhealth.jmir.org/2024/1/e48842
doi:10.2196/48842
PMID:38261368

The Physical and Psychological Effects of Telerehabilitation-Based Exercise for Patients With Nonspecific Low Back Pain: Prospective Randomized Controlled Trial

Weihong Shi[1*], MS; Yuhang Zhang[1*], DPT; Yanyan Bian[2*], PhD; Lixia Chen[1*], MS; Wangshu Yuan[1], MS; Houqiang Zhang[1], MS; Qiyang Feng[1], MS; Huiling Zhang[1], MS; Diana Liu[3], MD; Ye Lin[4], MD

[1]Department of Rehabilitation Medicine, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

[2]Department of Orthopedics, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

[3]Jiakang Zhongzhi Technology Company, Beijing, China

[4]University of Chicago, Chicago, IL, United States

[*]these authors contributed equally

Corresponding Author:
Lixia Chen, MS
Department of Rehabilitation Medicine, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, , Beijing, , China

Abstract

Background: Physical therapy has demonstrated efficacy in managing nonspecific low back pain (NLBP) among patients. Nevertheless, the prevalence of NLBP poses a challenge, as the existing medical infrastructure may be insufficient to care for the large patient population, particularly in geographically remote regions. Telerehabilitation emerges as a promising method to address this concern by offering a method to deliver superior medical care to a greater number of patients with NLBP.

Objective: The purpose of this study is to demonstrate the physical and psychological effectiveness of a user-centered telerehabilitation program, consisting of a smartphone app and integrated sensors, for patients with NLBP.

Methods: This was a single-center, prospective, randomized controlled trial for individuals with NLBP for a duration exceeding 3 months. All participants were assigned randomly to either the telerehabilitation-based exercise group (TBEG) or the outpatient-based exercise group (OBEG). All participants completed a 30-minute regimen of strength and stretching exercises 3 times per week, for a total of 8 weeks, and were required to complete assessment questionnaires at 0, 2, 4, and 8 weeks. The TBEG completed home-based exercises and questionnaires using a telerehabilitation program, while the OBEG completed them in outpatient rehabilitation. The Oswestry Disability Index (ODI) served as the primary outcome measure, assessing physical disability. Secondary outcomes included the Numeric Pain Rating Scale, Fear-Avoidance Beliefs Questionnaire, and 36-item Short-Form Health Survey.

Results: In total, 54 of 129 eligible patients were enrolled and randomly assigned to the study. The completion of all the interventions and assessments in the TBEG and OBEG was 89% (24/27) and 81% (22/27). The findings indicate that no statistical significance was found in the difference of ODI scores between the TBEG and the OBEG at 2 weeks (mean difference −0.91; odds ratio [OR] 0.78, 95% CI −5.96 to 4.14; *P*=.72), 4 weeks (mean difference −3.80; OR 1.33, 95% CI −9.86 to −2.25; *P*=.21), and 8 weeks (mean difference −3.24; OR 0.92, 95% CI −8.65 to 2.17; *P*=.24). The improvement of the ODI in the TBEG (mean −16.42, SD 7.30) and OBEG (mean −13.18, SD 8.48) was higher than 10 after an 8-week intervention. No statistically significant differences were observed between the 2 groups at the 8-week mark regarding the Fear-Avoidance Beliefs Questionnaire (mean difference 8.88; OR 1.04, 95% CI −2.29 to 20.06; *P*=.12) and Numeric Pain Rating Scale (mean difference −0.39; OR 0.44, 95% CI −2.10 to 1.31; *P*=.64). In the subgroup analysis, there was no statistically significant difference in outcomes between the 2 groups.

Conclusions: Telerehabilitation interventions demonstrate comparable therapeutic efficacy for individuals with NLBP when compared to conventional outpatient-based physical therapy, yielding comparable outcomes in pain reduction and improvement in functional limitations.

Trial Registration: Chinese Clinical Trial Registry ChiCTR2300068984; https://www.chictr.org.cn/showproj.html?proj=189852

(JMIR Mhealth Uhealth 2024;12:e56580) doi:10.2196/56580

KEYWORDS

nonspecific low back pain; telerehabilitation; physical therapy; low back pain; back pain; psychological; exercise; randomized controlled trial; efficacy; medical infrastructure; pain intensity; quality of life; health survey; therapeutic; mobile phone

Introduction

Background

Nonspecific low back pain (NLBP) is a broad category of low back pain (LBP) lacking identifiable etiology [1]. The Global Burden of Disease study report indicates that by 2020, the global prevalence of LBP will exceed 619 million, representing a 60% increase since 1990, with projections reaching 843 million by 2050 [2]. Years lived with disability is an index measuring the average duration of life lived with disability due to a disease from onset to death [3]. The Global Burden of Disease study reveals that LBP ranks first among 291 diseases in terms of age-standardized years lived with disability [4]. By 2020, 69 million individuals will experience limited life expectancy due to disability caused by LBP [2].

It is estimated that the annual economic burden for physical therapy of patients with LBP exceeded US $2.41 billion [5]. Nonspecific chronic LBP includes individuals experiencing pain exceeding 3 months, constituting 85% of the overall population with NLBP [6]. Therefore, it is urgent and necessary to explore innovative treatment modalities for individuals with NLBP.

Numerous clinical guidelines recommend incorporating exercise routines into the treatment of individuals with NLBP. This recommendation is based on the substantial pain relief and improved physical function observed in patients with NLBP through exercise interventions. These interventions also tend to have fewer adverse effects compared to pharmaceutical and surgical approaches [7-10].

Under the traditional clinic-based exercise model, patients are required to participate consistently in structured exercise programs. These programs are supervised by qualified physical therapists (PTs) over an extended period. However, the traditional clinic-based exercise model faces challenges because a substantial proportion of patients with NLBP face barriers to completing a structured exercise program due to time constraints, transportation limitations, and geographical challenges; in addition, a large proportion of individuals with NLBP, who live in remote areas, have no access to qualified exercise guidance because the medical resource is undistributed, especially in resource-limited countries [11-13].

As a result, the traditional clinic-based exercise model encounters difficulties in addressing the diverse needs of this patient population [14,15]. Given these circumstances, the telerehabilitation-based exercise model, incorporation of home-based exercises into a telerehabilitation program, emerges as a promising and effective strategy to address the previously mentioned challenges associated with managing NLBP. The telerehabilitation-based exercise model provides patients with a digitalized exercise plan, enabling patients with NLBP to complete their home exercise regimen promptly. In this model, patients' exercise performance is recorded, and any issues encountered are promptly addressed by professional health care providers. Compared to traditional clinic-based exercise models, the telerehabilitation-based model saves time, money, and medical resources.

In recent years, numerous research teams have studied telerehabilitation strategies for individuals with NLBP [16-21]. In the United States, Shebib et al [18] pioneered a comprehensive digital care program encompassing education, sensor-guided physiotherapy, aerobic exercise, and cognitive behavioral therapy tailored for patients with NLBP. Their investigation revealed superior therapeutic outcomes within the digital care program group compared to the control group [19]. In Germany, Toelle et al [19] developed the Kaia app (Kaia Health Corp) specifically designed for patients with NLBP. Results indicated that individuals receiving exercise guidance through the Kaia app exhibited significant pain relief and improvements in physical function compared to the control group [19].

Similarly, in the United Kingdom, Fatoye et al [20] integrated telerehabilitation with the McKenzie exercise approach. Remarkably, the telerehabilitation group achieved therapeutic outcomes equivalent to outpatient rehabilitation. It also demonstrated a lower average medical cost per patient compared to the outpatient group [20]. In resource-limited nations, such as China, the widespread adoption of telerehabilitation is important because of the large patient population and the lack of physical therapy services.

Furthermore, it is worth noting that psychosocial risk factors have a more significant impact on predicting pain-related outcomes in cases of NLBP compared to biomedical factors [22,23]. One widely accepted conceptual framework for understanding how psychosocial factors influence pain-related outcomes is the Fear-Avoidance Model [24]. According to the Fear-Avoidance Model, anxiety, depression, fear, and catastrophizing are risk factors that contribute to pain-related disability [25,26]. Marshall et al [27] found that a significant number of patients with NLBP did not experience improvements in pain intensity or limb disability after receiving professional exercise guidance and participating in weekly exercises. These patients were more susceptible to anxiety and fear [28]. In the traditional clinic-based exercise model, health care professionals, such as PTs, assist patients in correctly understanding pain and addressing their concerns to prevent the occurrence of anxiety and fear [28]. However, it remains unclear whether a telerehabilitation-based exercise model can reduce pain-related fear and anxiety in patients with NLBP.

Objective

The research team used the Healbone Intelligent Rehabilitation System (HIRS), comprised of a smartphone app and integrated sensors. The primary objective is to evaluate the efficacy of this intervention. The program guides and monitors patients with NLBP, as they engage in a structured home-based exercise

regimen. This study measures the program's impact on both the physical and psychological dimensions of NLBP management.

Methods

Ethical Considerations

This study has been approved by the ethics committees of the Peking Union Medical College Hospital (I-23PJ151) and registered in the Chinese Clinical Trial Registry (ChiCTR2300068984).

Trial Design

This study was a single-center, 2-arm, parallel-group, randomized controlled trial (RCT; participant-blinded) with 1:1 RCT, conducted in Peking Union Medical College Hospital, Beijing, China. All patients were assessed on pain, function, quality of life, and fear-belief avoidance. Assessments occurred at 0, 2, 4, and 8 weeks.

Inclusion and Exclusion Criteria

In this study, all participants were recruited from Peking Union Medical College Hospital. Two physicians selected patients with NLBP who met the inclusion and exclusion criteria and referred them to PTs. These patients were thoroughly informed about the purpose, procedures, and potential risks of the trial. Additionally, patients did not participate in any other medical interventions for LBP other than the exercise intervention of this trial until the study's completion. Upon obtaining informed consent, patients were included in the study.

Written informed consents were obtained from all patients. The inclusion and exclusion criteria are shown in Textbox 1.

Textbox 1. Inclusion and exclusion criteria for the study.

Inclusion criteria

- Aged between 18 and 60 years
- Numeric Pain Rating Scale equal to or greater than 3 points
- Oswestry Disability Index equal to or greater than 15 points
- Ongoing pain for at least 3 months
- Able to use a smartphone and complete the exercise protocol independently
- Those who could sign the informed consent independently

Exclusion criteria

- Patients with spinal deformity, spinal structure slip, spinal fracture history, and spinal tumor
- Diagnosed with rheumatoid arthritis and ankylosing spondylitis
- Patients with herniated disk
- Pregnancy
- Patients who receive other treatments before the experiments, including nonsteroid anti-inflammatory drugs or plasters, physical agents therapy, and acupuncture

Sample Size Calculation

The sample size was calculated using PASS 11 (NCSS Corp). Based on the principle of noninferiority RCTs [29] and previous clinical studies [20,30], the mean difference in the Oswestry Disability Index (ODI) between the telerehabilitation-based exercise group (TBEG) and the outpatient-based exercise group (OBEG) was 5, the SD was estimated to be 6 for both groups, and the noninferiority margin for the ODI was 10. A sample size of 38 was required based on a bilateral α=.05 and β=.2, and a sample size of 54 was required to account for a 30% dropout rate.

Blinding and Randomization

In total, 54 participants were assigned in a randomized manner, with equal distribution, to either the TBEG or the OBEG through a platform for randomization. Subsequently, based on the results derived from the platform (eg, C, T, C, T, T, and C), slips of paper labeled with the letters "T" and "C" were placed into sealed, opaque, and identically sized envelopes. After completing the baseline measurements for all participants, the envelopes were sequentially opened to reveal the group assignments. The allocation sequence was prepared by 2 researchers with no involvement in the study using a blocked randomization model.

Intervention

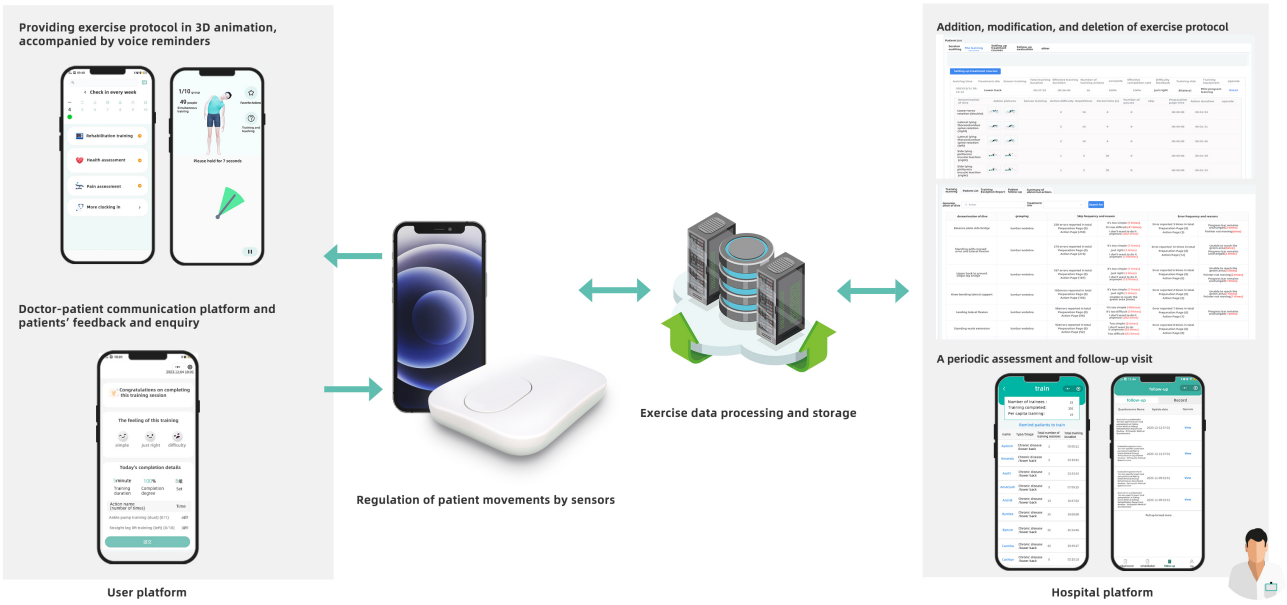
Overview

The exercise plan for patients with NLBP in both TBEG and OBEG was identical. Both consist of muscle strengthening and stretching exercises to increase lumbar stability, coordination, and posture keeping. The detailed exercise plan is shown in Multimedia Appendix 1. Before the initiation of this trial, 2 PTs were trained in three 40-minute sessions.

Telerehabilitation-Based Exercise Group

The HIRS was designed based on a user-centered theory to provide patients with a platform for self-management interventions. Additionally, the HIRS system is made available at no cost to all participants in this study. Figure 1 illustrates the 3 distinct components of HIRS: the physician portal, the user portal, and the transmission portal.

Figure 1. The 3 different parts of the Healbone Intelligent Rehabilitation System. The doctor's portal could be used to create and modify exercises, monitor training progress, and view patient data. The patients could use the user's portal to complete the prescribed exercises, view educational materials, and provide feedback to the physical therapists. Finally, the transmitter portal encrypts and transmits the data collected, ensuring the overall system's integrity.



Prior to the commencement of the experiment, professional medical personnel created specific videos for each training exercise in the rehabilitation program and uploaded them to the HIRS along with detailed instructions. At the onset of the experiment, an app was installed on the smartphones of the patients in the TBEG, through which they registered personal accounts. Subsequently, during the initial session, PTs sent digital exercise training protocols to the patients' personal accounts and educated them on the correct use of the app and sensors for home-based exercises.

Each time the patients engaged in the exercises, they were required to access the app via their personal accounts and calibrate the sensors to accurately perform each exercise within the regimen. Upon initiation of the exercise, the patients were to follow the instructions provided in the video to complete each action in the regimen. If a patient failed to exercise, the system would automatically send a reminder and notify the PTs, who would then contact the patient to ascertain the reason for nonparticipation. Concurrently, if an individual in the TBEG sought advice from the PTs regarding concerns or inquiries related to back pain, the PTs would provide the patient detailed responses to prevent the individual from experiencing fear or anxiety.

Over an 8-week period, all patients in the TBEG were mandated to complete exercise sessions every other day, 3 times a week, with each session lasting 30 minutes. The HIRS transmission portal collected the results of the patients' assessments and automatically recorded their exercise performance, including the duration of each session and the frequency of weekly exercises.

Finally, patients were required to complete digital assessment questionnaires via the app at weeks 0, 2, 4, and 8. The validity of HIRS had been verified by 25 patients before the trial.

Outpatient-Based Exercise Group

In the OBEG, patients underwent a consistent 30-minute exercise regimen every 2 days under the PTs' supervision, with sessions scheduled thrice weekly. Concurrently, during each hospital visit for exercise guidance, the PTs provide face-to-face consultations to address any questions or concerns the patients may have. Furthermore, assessment questionnaires are administered in the outpatient clinic at baseline (week 0), week 2, week 4, and week 8.

Outcome Measures

The ODI, as the primary outcome measure, has been verified for reliability and validity [31]. It is commonly used to evaluate physical function. The minimal clinically important difference (MCID) refers to the smallest change in score that patients perceive as beneficial, irrespective of side effects and costs. Bombardier et al [32] determined that the MCID for the ODI score in patients with NLBP is 5. This indicates that an improvement in the ODI score by at least 5 points after the intervention is considered clinically meaningful for the patient.

In addition, a set of secondary outcome measures was also used. These measures include the Numeric Pain Rating Scale (NPRS) for pain evaluation, the 36-item Short-Form Health Survey (SF-36) for quality of life assessment, and the Fear-Avoidance Beliefs Questionnaire (FABQ) to gauge fear-avoidance beliefs related to work and physical activity [33,34]. Previous studies have determined that the MCID for the FABQ in patients with NLBP is 11 [35], and for the NPRS, it is 2 [36]. However, Grönkvist et al [37] established that the MCID for the 8 dimensions of the SF-36 varies among patients with NLBP [37].

The reliability and validity of the Chinese version of the SF-36 and the FABQ have been confirmed [33,34]. The collection of the primary and secondary outcome measures occurred at weeks 0, 2, 4, and 8. Patients in the TBEG completed all assessments

through the HIRS, while patients in the OBEG completed these assessments in the outpatient clinic, guided by PTs.

Statistical Analysis

The outcomes were analyzed following the intention-to-treat approach, and all participants were analyzed according to the original group assignment. Missing data were handled using multiple imputations by chained equations [38]. Besides, subgroup analysis was conducted following per-protocol analyses in this study.

All the data in this study were analyzed using SPSS (version 23.0; IBM Corp). Demographic data are presented as means (SDs) and numbers (percentages). Descriptive statistics, independent sample 2-tailed *t* tests, and chi-square tests were used to analyze participant characteristics. The normality of distribution for all data was tested by an independent sample *t* test. The results of this study are presented as mean, SD, odds ratio (OR), and 95% CI. The statistical analysis was conducted by a researcher who was blinded and not involved in this study.

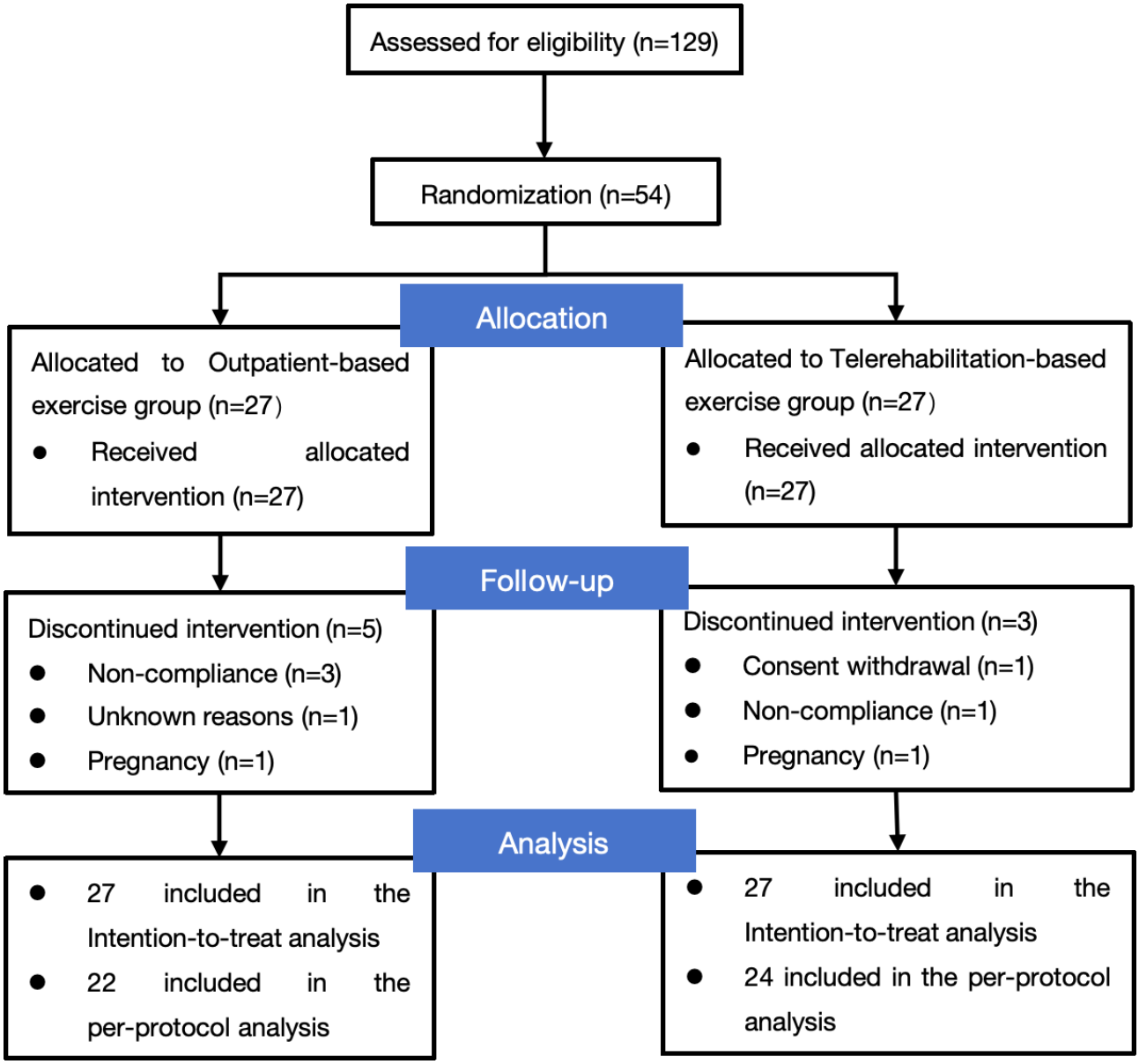
Results

Study Population and Follow-Up

Between March 9, 2023, and November 1, 2023, 129 patients were considered for eligibility. During the initial screening process, 35 patients did not meet the inclusion criteria or met the exclusion criteria. Among the remaining 94 eligible patients, 32 did not consent, and 8 patients withdrew prior to group randomization. Therefore, a total of 54 patients left for the final study. The population was randomly allocated into 2 groups: the TBEG (n=27) and OBEG (n=27), as illustrated in Figure 2.

All patients completed the baseline assessment at week 0 and were asked to complete assessments at weeks 2, 4, and 8. All participants underwent baseline assessments. During the study, 3 patients in the TBEG and 5 patients in the OBEG withdrew due to pregnancy or other reasons. Ultimately, 24 patients in the TBEG and 22 patients in the OBEG completed all treatments and the 8-week follow-up. The completion rates for treatment and assessments in the OBEG were as follows: 96% (26/27) at 2 weeks, 93% (25/27) at 4 weeks, and 81% (22/27) at 8 weeks. In comparison, the completion rates in the TBEG were 100% (27/27) at 2 weeks, 96% (26/27) at 4 weeks, and 89% (24/27) at 8 weeks, as illustrated in Figure 2. Between the 2 groups, patients showed similar clinical and demographic characteristics (Textbox 1).

Figure 2. Participant flowchart.



Primary Outcomes

The baseline ODI scores showed comparable values between the TBEG and the OBEG (Table 1). The mean ODI improvement for the OBEG was −4.70 (SD 9.20) at week 2, −8.40 (SD 10.13) at week 4, and −13.15 (SD 8.48) at week 8. The mean ODI improvement for the TBEG was −5.61 (SD 7.30) at week 2, −11.43 (SD 8.83) at week 4, and −13.70 (SD 7.30) at week 8. After 8 weeks of intervention, both the OBEG and TBEG demonstrated an improvement in ODI scores exceeding 10 points, indicating clinical significance (Table 2).

At the second week, the difference in ODI score changes between the TBEG and OBEG was 0.41 (OR 0.78, 95% CI −0.58 to 1.39); at the fourth week, the difference was −3.80 (OR 1.33, 95% CI −9.86 to 2.25); and at the eighth week, the difference was −3.24 (OR 0.92, 95% CI −8.65 to 2.17; Table 2). Statistical analysis revealed no significant differences between the TBEG and OBEG at weeks 0, 2, 4, and 8 (Figure 3). Following the 8-week intervention, the improvement in ODI scores in the TBEG was noninferior to that in the OBEG.

Table . Demographics and baseline characteristics of all participants.

Characteristics		OBEG[a] (n=27)	TBEG[b] (n=27)	*P* value
Age (years), mean (SD)		38.23 (11.55)	39.11 (10.45)	.77
Height (m), mean (SD)		1.69 (1.07)	1.65 (0.08)	.08
Weight (kg), mean (SD)		63.96 (9.89)	61.85 (10.50)	.46
BMI (kg/m^2), mean (SD)		22.39 (3.19)	22.57 (2.98)	.83
Sex, n (%)				.41
	Male	12 (57)	9 (43)	
	Female	14 (44)	18 (56)	
Sedentary time per day, mean (SD)		6.69 (2.35)	7.08 (2.73)	.59
Pain duration (months), mean (SD)		10.23 (3.57)	10.11 (3.56)	.90
ODI[c], mean (SD)		18.80 (6.57)	20.86 (11.40)	.21
NPRS[d], mean (SD)		5.02 (1.74)	5.42 (3.82)	.27
FABQ[e], mean (SD)		46.47 (14.36)	41.32 (10.60)	.23
SF-36 [f] , mean (SD)				
	Physical functioning	61.85 (12.57)	60.96 (2.54)	.87
	Role-physical	45.37 (6.71)	46.15 (7.88)	.94
	Bodily pain	58.52 (17.26)	55.00 (20.45)	.50
	General health	49.63 (16.35)	46.92 (13.12)	.51
	Vitality	73.70 (12.76)	72.50 (15.44)	.76
	Social functioning	80.96 (16.20)	78.08 (18.45)	.55
	Role-emotional	69.33 (2.25)	56.81 (2.83)	.28
	Mental health	66.30 (15.54)	72.73 (15.87)	.14

[a]OBEG: outpatient-based exercise group.

[b]TBEG: telerehabilitation-based exercise group.

[c]ODI: Oswestry Disability Index.

[d]NPRS: Numeric Pain Rating Scale.

[e]FABQ: Fear-Avoidance Beliefs Questionnaire.

[f]SF-36: 36-item Short-Form Health Survey.

Table . Primary and secondary outcomes for the OBEG[a] and the TBEG[b].

	Two weeks					Four weeks					Eight weeks				
	OBEG (n=27), mean (SD)	TBEG (n=27), mean (SD)	Difference of change between 2 groups		*P* value	OBEG (n=27), mean (SD)	TBEG (n=27), mean (SD)	Difference of change between 2 groups		*P* value	OBEG (n=27), mean (SD)	TBEG (n=27), mean (SD)	Difference of change between 2 groups		*P* value
			Mean	OR[c] (95% CI)				Mean	OR (95% CI)				Mean	OR (95% CI)	
Primary outcome															
ODI[d]	−4.70 (9.20)	−5.61 (7.30)	−0.91	0.78 (−5.96 to 4.14)	.72	−9.93 (10.13)	−13.73 (8.83)	−3.80	1.33 (−9.86 to −2.25)	.21	−13.18 (8.48)	−16.42 (7.30)	−3.24	0.92 (−8.65 to 2.17)	.24
Secondary outcomes															
NPRS[e]	−2.41 (1.95)	−2.00 (1.60)	0.41	0.69 (−0.58 to 1.39)	.41	−3.44 (2.38)	−3.65 (3.77)	−0.21	2.00 (−1.94 to 1.52)	.81	−4.26 (3.90)	−4.65 (2.01)	−0.39	0.44 (−2.10 to 1.31)	.64
FABQ[f]	−4.42 (17.51)	−6.00 (16.21)	−1.58	1.00 (−11.03 to 7.88)	.74	−15.91 (17.87)	−11.60 (16.98)	4.31	0.19 (−7.60 to 16.22)	.47	−40.15 (13.38)	−31.92 (15.07)	8.88	1.04 (−2.29 to 20.06)	.12
SF-36 [g]															
Physical functioning	10.50 (16.75)	13.64 (23.31)	3.14	1.75 (−9.51 to 15.78)	.62	13.26 (15.99)	12.20 (24.72)	−1.06	1.38 (−13.05 to 10.92)	.59	16.35 (16.62)	15.38 (27.81)	−0.96	2.58 (−13.60 to 11.67)	.88
Role-physical	5.00 (32.87)	13.86 (41.89)	8.86	1.26 (−14.52 to 32.25)	.45	19.06 (34.12)	13.64 (40.21)	−6.62	1.20 (−28.57 to 15.34)	.55	19.69 (38.96)	17.27 (47.50)	−0.96	1.41 (−25.30 to 23.38)	.94
Bodily pain	14.71 (21.47)	12.27 (20.45)	−1.73	1.02 (−14.61 to 11.15)	.79	16.30 (21.54)	19.00 (26.03)	2.70	0.88 (−12.01 to 17.40)	.71	25.00 (15.48)	20.77 (26.84)	−4.23	2.03 (−16.84 to 8.38)	.50
General health	9.50 (11.95)	9.32 (14.33)	−0.18	1.71 (−8.24 to 7.88)	.96	13.26 (16.04)	12.80 (14.77)	−0.46	1.10 (−9.22 to 8.30)	.92	12.88 (17.06)	12.50 (11.62)	−0.38	1.00 (−8.69 to 7.92)	.93
Vitality	7.40 (15.50)	4.32 (15.83)	−3.08	1.06 (−12.81 to 6.64)	.53	11.30 (14.71)	6.00 (15.88)	−5.30	1.52 (−15.70 to 5.10)	.31	13.75 (11.23)	10.00 (12.97)	−1.54	1.26 (−9.26 to 6.18)	.69
Social functioning	3.94 (22.00)	9.59 (19.30)	5.66	1.23 (−7.11 to 18.42)	.38	9.39 (22.21)	10.80 (16.06)	1.41	0.27 (−10.04 to 12.85)	.81	14.09 (17.60)	9.39 (21.81)	−4.70	1.44 (−15.37 to 5.98)	.38
Role-emotional	10.35 (14.28)	17.67 (24.08)	7.32	0.75 (−17.38 to 32.02)	.55	32.92 (38.33)	26.47 (18.23)	6.44	0.71 (−20.68 to 33.56)	.64	22.92 (15.53)	30.38 (29.76)	7.46	0.54 (−17.67 to 32.59)	.55
Mental health	7.15 (16.09)	0.95 (15.03)	−6.20	4.80 (−16.41 to 4.02)	.23	23.35 (14.41)	6.44 (23.88)	−6.91	8.25 (−19.94 to 3.01)	.01	14.46 (17.28)	6.50 (15.77)	−7.96	4.20 (−17.53 to 1.06)	.10

[a]OBEG: outpatient-based exercise group.

[b]TBEG: telerehabilitation-based exercise group.

[c]OR: odds ratio.

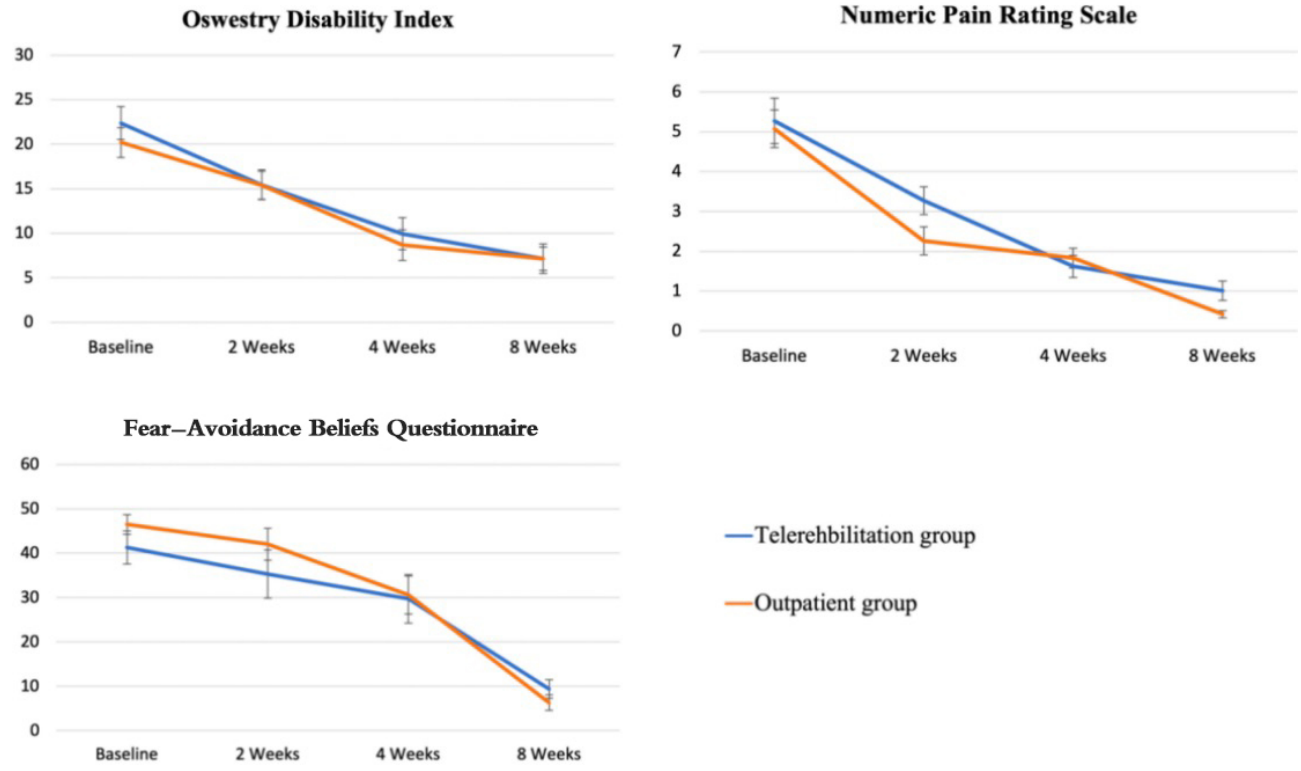
[d]ODI: Oswestry Disability Index.

[e]NPRS: Numeric Pain Rating Scale.

[f]FABQ: Fear-Avoidance Beliefs Questionnaire.

[g]SF-36: 36-item Short-Form Health Survey.

Figure 3. The primary and secondary outcomes Oswestry Disability Index, Numeric Pain Rating Scale, and Fear-Avoidance Beliefs Questionnaire at baseline and 2, 4, and 8 weeks; error bars represent 95% CIs.



Secondary Outcomes

The baseline NPRS scores showed comparable values between the TBEG and the OBEG. Statistical analysis found no significant differences between values for the TBEG and the OBEG at weeks 0, 2, 4, and 8 (Table 1). At week 8, the mean NPRS improvement from baseline was −4.65 (SD 2.01) in OBEG and −4.65 (SD 2.01) in TBEG (Figure 3).

The baseline FABQ scores showed comparable values between the TBEG and the OBEG. Statistical analysis found no significant differences between the values for the TBEG and the OBEG at weeks 0, 2, 4, and 8 (Table 1). At week 8, the mean improvement in FABQ scores from baseline was −40.15 (SD 13.38) in OBEG and −32.48 (SD 15.07) in TBEG (Figure 3).

No statistically significant differences were found in the mean change of the SF-36 scores at weeks 0, 2, 4, and 8 (Tables 1 and 2).

After an 8-week intervention, the NPRS, FABQ, and SF-36 scores in both the TBEG and the OBEG showed significant improvement compared to baseline values. Furthermore, the extent of improvement in NPRS, FABQ, and SF-36 scores in the TBEG was found to be noninferior to that observed in the OBEG.

Subgroup Analysis

After an 8-week intervention, both the OBEG and the TBEG demonstrated clinically significant improvements in the ODI, with a reduction exceeding 10 points, and no statistical differences were found in the changes of the ODI between 2 groups, demonstrating noninferiority (Table 3).

Table . Subgroup analysis of primary and secondary outcomes for the OBEG[a] and the TBEG[b].

	Two weeks					Four weeks					Eight weeks				
	OBEG (n=26), mean (SD)	TBEG (n=27), mean (SD)	Difference of change between 2 groups		*P* value	OBEG (n=25), mean (SD)	TBEG (n=26), mean (SD)	Difference of change between 2 groups		*P* value	OBEG (n=22), mean (SD)	TBEG (n=24), mean (SD)	Difference of change between 2 groups		*P* value
			Mean	OR[c] (95% CI)				Mean	OR (95% CI)				Mean	OR (95% CI)	
Primary outcome															
ODI[d]	−5.11 (9.26)	−5.61 (7.30)	−0.50	0.64 (−5.63 to 4.63)	.84	−9.76 (8.54)	−14.16 (12.14)	−4.40	0.83 (−10.37 to −2.25)	.15	−12.77 (7.04)	−17.48 (11.55)	−4.71	1.05 (−10.49 to 1.08)	.11
Secondary outcomes															
NPRS[e]	−2.42 (1.98)	−2.00 (1.60)	0.42	0.59 (−0.58 to 1.43)	.40	−3.40 (2.31)	−3.72 (3.84)	−0.32	0.29 (−2.12 to 1.52)	.72	−4.27 (1.78)	−4.87 (4.09)	−0.60	0,21 (−2.51 to 1.32)	.53
FABQ[f]	−4.72 (17.81)	−6.00 (16.21)	−1.28	0.85 (−10.91 to 8.35)	.79	−15.38 (19.94)	−12.54 (20.64)	2.84	0.88 (−9.40 to 15.08)	.64	−41.90 (24.10)	−30.91 (15.27)	10.99	1.57 (−1.17 to 23.16)	.08
SF-36 [g]															
Physical functioning	11.05 (16.63)	13.64 (23.31)	2.58	0.47 (−10.36 to 15.53)	.69	11.90 (14.87)	11.88 (24.93)	−0.03	0.94 (−12.60 to 12.54)	.99	14.05 (13.29)	16.09 (29.62)	2.04	0.79 (−12.16 to 16.24)	.77
Role-physical	5.79 (32.50)	13.86 (41.89)	8.07	1.25 (−15.90 to 32.05)	.50	23.57 (35.25)	18.13 (41.49)	−5.45	1.40 (−28.78 to 17.88)	.64	22.14 (38.03)	23.70 (49.78)	1.55	1.41 (−25.60 to 28.70)	.91
Bodily pain	13.68 (21.33)	12.27 (20.45)	−1.41	1.40 (−14.63 to 11.81)	.83	15.48 (21.54)	20.21 (27.84)	4.73	0.90 (−10.40 to 19.87)	.53	23.81 (16.58)	20.87 (29.53)	−2.94	2.03 (−17.70 to 11.82)	.69
General health	10.00 (11.18)	9.32 (14.33)	−0.68	1.50 (−8.90 to 7.53)	.87	12.62 (14.97)	13.75 (5.13)	1.13	0.58 (−7.94 to 10.20)	.80	12.38 (17.57)	12.83 (12.31)	0.45	1.05 (−8.87 to 9.76)	.92
Vitality	8.31 (15.50)	4.32 (15.83)	−4.00	0.83 (−13.82 to 5.83)	.42	11.19 (15.48)	5.00 (19.78)	−6.19	1.57 (−16.98 to 4.60)	.25	12.86 (12.41)	13.70 (15.54)	0.84	1.54 (−7.77 to 9.45)	.85
Social functioning	4.15 (22.)	9.59 (19.30)	5.45	1.01 (−7.66 to 18.56)	.41	6.57 (17.97)	11.25 (19.51)	4.67	0.67 (−6.60 to 15.94)	.41	11.42 (15.15)	9.16 (23.16)	−2.26	0.70 (−14.30 to 9.77)	.71
Role-emotional	10.89 (14.69)	17.67 (24.08)	6.77	0.78 (−18.58 to 32.13)	.59	24.24 (43.56)	31.50 (50.69)	7.26	0.12 (−21.36 to 35.89)	.61	22.00 (36.98)	30.00 (55.03)	8.00	0.56 (−20.81 to 36.81)	.58
Mental health	7.32 (18.19)	0.95 (15.03)	−6.36	1.68 (−16.85 to 4.13)	.23	23.67 (17.28)	15.79 (17.38)	−7.88	2.33 (−18.32 to 2.57)	.14	15.81 (14.18)	8.77 (18.96)	−7.04	2.35 (−17.39 to 3.31)	.18

[a]OBEG: outpatient-based exercise group.

[b]TBEG: telerehabilitation-based exercise group.

[c]OR: odds ratio.

[d]ODI: Oswestry Disability Index.

[e]NPRS: Numeric Pain Rating Scale.

[f]FABQ: Fear-Avoidance Beliefs Questionnaire.

[g]SF-36: 36-item Short-Form Health Survey.

Discussion

Principal Findings

This study was designed to determine the efficacy of the treatment between the TBEG and the OBEG. After an 8-week intervention, the completion rate was 89% (24/27) in the TBEG and 81% (22/27) in the OBEG. The completion rate of exercise was higher in the TBEG compared to the OBEG. In the primary outcomes, there was no statistically significant difference between the TBEG and the OBEG in improving pain-related physical dysfunction, demonstrating noninferiority of telerehabilitation. However, both groups demonstrated an improvement in the ODI score exceeding 10 points, indicating that both telerehabilitation exercises and outpatient exercises have clinical significance in improving the ODI for patients with NLBP. Regarding secondary outcomes, there were no statistically significant differences in the SF-36, NPRS, and FABQ between the groups; the improvements in the SF-36, NPRS, and FABQ surpassed the MCID. This suggests that both TBEG and OBEG interventions have clinical significance in pain relief, reduction in fear-avoidance beliefs, and enhancement of quality of life following an 8-week intervention with similar efficacy.

The Efficacy and Benefits of Telerehabilitation for Patients With NLBP

Compared with previous studies [39-41], this study also demonstrates the efficacy of telerehabilitation for patients with NLBP in the improvement of pain intensity, physical disability, and quality of life.

Additionally, we also found that exercise helped patients in both groups to reduce the impact of pain-related fear on work and daily activity after an 8-week intervention. Figure 2, which visually presents the data, indicates significant improvements in both the NPRS and ODI by week 4. Moreover, the FABQ demonstrates a noticeable reduction by week 8, suggesting a delayed improvement in patients' psychological fear.

Exercise is an important and widely accepted treatment for patients with NLBP [42,43]. To achieve the expected results, it is crucial for patients to consistently follow a prescribed exercise plan for an extended time. However, a study by Palazzo et al [12] found that patients with NLBP face challenges in adhering to home-based exercises. These challenges include factors such as remote locations, difficulties in the exercise program, the patient's attitude toward exercise, and the lack of supervision and follow-up outside of the hospital. Altogether, these factors reduce the effectiveness of the treatment [11,12].

Hence, we have introduced a telerehabilitation system using a smartphone app combined with sensors. This system is designed to offer better monitoring and follow-up beyond usual care. The exercise plan includes stretching and strength exercises, which have been shown to reduce pain and improve physical function [44-46]. The smartphone app uses visual and audio content to enhance patient experience.

Furthermore, the telerehabilitation system allows patients to receive prompt guidance from PTs. PTs can monitor the real-time physical functional status and exercise progression of patients. The exercise routines span over 8 weeks, with sessions occurring every 2 days, 3 times per week. In contrast to patients in the OBEG who need to schedule appointments with PTs in the clinic, those in the TBEG can complete their exercises at home.

In contrast to the OBEG, patients with NLBP in the TBEG exhibited greater flexibility in their exercise scheduling. Within the scope of this study, they consistently adhered to their exercise regimens in a timely manner, which contributed to improved compliance with exercise plans and benefited patients with NLBP. Moreover, the exercise model based on remote rehabilitation can help individuals save more time and expenses related to hospital visits, offering greater convenience compared to outpatient-based exercise models. Simultaneously, the remote rehabilitation–based exercise model provides patients with NLBP with the opportunity to receive qualified exercise guidance. Finally, this remote rehabilitation–based exercise model alleviates the burden on health care institutions and reduces treatment costs.

The telerehabilitation system enables patients with NLBP to adhere to their treatment plans and allows them to manage their health at home with remote supervision. In long-term follow-ups, Hou et al [47] found that patients using telerehabilitation showed more improvement in functional limitations compared to those relying on the traditional in-clinic method. This is especially promising in areas lacking medical accessibility.

Limitations

This study was conducted at 1 medical center to compare telerehabilitation with traditional on-site rehabilitation for patients with NLBP. A total of 54 patients participated in the trial, with 27 in the TBEG and 27 in the OBEG. The exercise routines span over 8 weeks, and assessments were scheduled at weeks 0, 2, 4, and 8.

This was a single-center RCT with a relatively small sample size. To address this limitation, the research team plans to conduct subsequent multicenter RCTs in regions with limited medical resources where patients with NLBP have difficulty accessing professional exercise guidance. Additionally, due to the relatively short follow-up period, future studies will involve

more participants to investigate the effects of remote rehabilitation–based exercise interventions on patients with NLBP over 6 months, 1 year, or even longer durations [47], focusing on adherence, pain relief, and improvement in pain-related physical dysfunction.

This study demonstrates that remote rehabilitation–based exercise training has therapeutic effects on pain relief and improvement in pain-related physical dysfunction in patients with NLBP. However, there is currently limited research analyzing the factors influencing the efficacy of remote rehabilitation in patients with NLBP, and it remains unclear which types of patients with NLBP are more suitable for remote rehabilitation treatment.

Furthermore, the research team plans to validate the effectiveness of the telerehabilitation program through a multicenter clinical trial. These future efforts are designed to study the benefits of telerehabilitation in managing NLBP, providing valuable insights to the medical community.

Conclusions

This study confirms that telerehabilitation and traditional outpatient rehabilitation methods produce comparable outcomes for patients with NLBP. Additionally, telerehabilitation reduces time, cost, and medical resources. It exhibits potential as an alternative for patients lacking access to high-quality rehabilitation services.

Acknowledgments

This study was funded by National High-Level Hospital Clinical Research Funding (2022-PUMCH-B-053).

Authors' Contributions

The inception of the study was facilitated through the collaborative efforts of YB, LC, and WY. YL, YZ, and WS undertook the experimental procedures and were responsible for the composition of the manuscript. Substantial contributions to the analysis and preparation of the manuscript were made by DL, Houqiang Zhang, and Huiling Zhang. QF executed the data analyses, and WS played a crucial role in the analytical process through constructive discussions.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Low back pain remote home-based exercise protocol.
[DOCX File, 2276 KB - mhealth_v12i1e56580_app1.docx]

Checklist 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File, 111 KB - mhealth_v12i1e56580_app2.pdf]

References

1. Maher C, Underwood M, Buchbinder R. Non-specific low back pain. Lancet 2017 Feb 18;389(10070):736-747. [doi: 10.1016/S0140-6736(16)30970-9] [Medline: 27745712]
2. GBD 2021 Low Back Pain Collaborators. Global, regional, and national burden of low back pain, 1990-2020, its attributable risk factors, and projections to 2050: a systematic analysis of the Global Burden of Disease Study 2021. Lancet Rheumatol 2023 Jun;5(6):e316-e329. [doi: 10.1016/S2665-9913(23)00098-X] [Medline: 37273833]
3. Wu A, Dong W, Liu S, et al. The prevalence and years lived with disability caused by low back pain in China, 1990 to 2016: findings from the Global Burden of Disease Study 2016. Pain 2019 Jan;160(1):237-245. [doi: 10.1097/j.pain.0000000000001396] [Medline: 30234697]
4. Hoy D, March L, Brooks P, et al. The global burden of low back pain: estimates from the Global Burden of Disease 2010 study. Ann Rheum Dis 2014 Jun;73(6):968-974. [doi: 10.1136/annrheumdis-2013-204428]
5. Cieza A, Causey K, Kamenov K, Hanson SW, Chatterji S, Vos T. Global estimates of the need for rehabilitation based on the Global Burden of Disease Study 2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet 2021 Dec 19;396(10267):2006-2017. [doi: 10.1016/S0140-6736(20)32340-0] [Medline: 33275908]
6. Vlaeyen JWS, Maher CG, Wiech K, et al. Low back pain. Nat Rev Dis Primers 2018 Dec 13;4(1):52. [doi: 10.1038/s41572-018-0052-1] [Medline: 30546064]
7. Bernstein IA, Malik Q, Carville S, Ward S. Low back pain and sciatica: summary of NICE guidance. BMJ 2017 Jan 6;356:i6748. [doi: 10.1136/bmj.i6748] [Medline: 28062522]
8. Foster NE, Anema JR, Cherkin D, et al. Prevention and treatment of low back pain: evidence, challenges, and promising directions. Lancet 2018 Jun;391(10137):2368-2383. [doi: 10.1016/S0140-6736(18)30489-6]

9. Kreiner DS, Matz P, Bono CM, et al. Guideline summary review: an evidence-based clinical guideline for the diagnosis and treatment of low back pain. Spine J 2020 Jul;20(7):998-1024. [doi: 10.1016/j.spinee.2020.04.006] [Medline: 32333996]
10. Traeger A, Buchbinder R, Harris I, Maher C. Diagnosis and management of low-back pain in primary care. CMAJ 2017 Nov 13;189(45):E1386-E1395. [doi: 10.1503/cmaj.170527] [Medline: 29133540]
11. Hayden JA, Ellis J, Ogilvie R, et al. Some types of exercise are more effective than others in people with chronic low back pain: a network meta-analysis. J Physiother 2021 Oct;67(4):252-262. [doi: 10.1016/j.jphys.2021.09.004] [Medline: 34538747]
12. Palazzo C, Klinger E, Dorner V, et al. Barriers to home-based exercise program adherence with chronic low back pain: patient expectations regarding new technologies. Ann Phys Rehabil Med 2016 Apr;59(2):107-113. [doi: 10.1016/j.rehab.2016.01.009] [Medline: 27050664]
13. Zhou T, Salman D, McGregor A. mHealth apps for the self-management of low back pain: systematic search in app stores and content analysis. JMIR Mhealth Uhealth 2024 Feb 1;12:e53262. [doi: 10.2196/53262] [Medline: 38300700]
14. Bunting JW, Withers TM, Heneghan NR, Greaves CJ. Digital interventions for promoting exercise adherence in chronic musculoskeletal pain: a systematic review and meta-analysis. Physiotherapy 2021 Jun;111:23-30. [doi: 10.1016/j.physio.2020.08.001] [Medline: 33316865]
15. Peek K, Carey M, Mackenzie L, Sanson-Fisher R. Patient adherence to an exercise program for chronic low back pain measured by patient-report, physiotherapist-perception and observational data. Physiother Theory Pract 2019 Dec;35(12):1304-1313. [doi: 10.1080/09593985.2018.1474402] [Medline: 29771180]
16. Marcuzzi A, Nordstoga AL, Bach K, et al. Effect of an artificial intelligence-based self-management app on musculoskeletal health in patients with neck and/or low back pain referred to specialist care: a randomized clinical trial. JAMA Netw Open 2023 Jun 1;6(6):e2320400. [doi: 10.1001/jamanetworkopen.2023.20400] [Medline: 37368401]
17. Øveråas CK, Nilsen TIL, Nicholl BI, et al. Multimorbidity and co-occurring musculoskeletal pain do not modify the effect of the SELFBACK app on low back pain-related disability. BMC Med 2022 Feb 8;20(1):53. [doi: 10.1186/s12916-022-02237-z] [Medline: 35130898]
18. Shebib R, Bailey JF, Smittenaar P, Perez DA, Mecklenburg G, Hunter S. Randomized controlled trial of a 12-week digital care program in improving low back pain. NPJ Digit Med 2019 Jan;2:1. [doi: 10.1038/s41746-018-0076-7] [Medline: 31304351]
19. Toelle TR, Utpadel-Fischler DA, Haas KK, Priebe JA. App-based multidisciplinary back pain treatment versus combined physiotherapy plus online education: a randomized controlled trial. NPJ Digit Med 2019 May;2:34. [doi: 10.1038/s41746-019-0109-x] [Medline: 31304380]
20. Fatoye F, Gebrye T, Fatoye C, et al. The clinical and cost-effectiveness of telerehabilitation for people with nonspecific chronic low back pain: randomized controlled trial. JMIR Mhealth Uhealth 2020 Jun 24;8(6):e15375. [doi: 10.2196/15375] [Medline: 32357128]
21. Itoh N, Mishima H, Yoshida Y, Yoshida M, Oka H, Matsudaira K. Evaluation of the effect of patient education and strengthening exercise therapy using a mobile messaging app on work productivity in Japanese patients with chronic low back pain: open-label, randomized, parallel-group trial. JMIR Mhealth Uhealth 2022 May 16;10(5):e35867. [doi: 10.2196/35867] [Medline: 35576560]
22. Lee H, Hübscher M, Moseley GL, et al. How does pain lead to disability? A systematic review and meta-analysis of mediation studies in people with back and neck pain. Pain 2015 Jun;156(6):988-997. [doi: 10.1097/j.pain.0000000000000146] [Medline: 25760473]
23. Wertli MM, Rasmussen-Barr E, Weiser S, Bachmann LM, Brunner F. The role of fear avoidance beliefs as a prognostic factor for outcome in patients with nonspecific low back pain: a systematic review. Spine J 2014 May 1;14(5):816-836.e4. [doi: 10.1016/j.spinee.2013.09.036] [Medline: 24412032]
24. Vlaeyen JWS, Kole-Snijders AMJ, Boeren RGB, van Eek H. Fear of movement/(re)injury in chronic low back pain and its relation to behavioral performance. Pain 1995 Sep;62(3):363-372. [doi: 10.1016/0304-3959(94)00279-N] [Medline: 8657437]
25. Vlaeyen JWS, Crombez G, Linton SJ. The fear-avoidance model of pain. Pain 2016 Aug;157(8):1588-1589. [doi: 10.1097/j.pain.0000000000000574] [Medline: 27428892]
26. Edwards RR, Dworkin RH, Sullivan MD, Turk DC, Wasan AD. The role of psychosocial processes in the development and maintenance of chronic pain. J Pain 2016 Sep;17(9 Suppl):T70-T92. [doi: 10.1016/j.jpain.2016.01.001] [Medline: 27586832]
27. Marshall PWM, Schabrun S, Knox MF. Physical activity and the mediating effect of fear, depression, anxiety, and catastrophizing on pain related disability in people with chronic low back pain. PLoS One 2017;12(7):e0180788. [doi: 10.1371/journal.pone.0180788] [Medline: 28686644]
28. Marshall A, Joyce CT, Tseng B, et al. Changes in pain self-efficacy, coping skills, and fear-avoidance beliefs in a randomized controlled trial of yoga, physical therapy, and education for chronic low back pain. Pain Med 2022 Apr 8;23(4):834-843. [doi: 10.1093/pm/pnab318] [Medline: 34698869]
29. Le Henanff A, Giraudeau B, Baron G, Ravaud P. Quality of reporting of noninferiority and equivalence randomized trials. JAMA 2006 Mar 8;295(10):1147-1151. [doi: 10.1001/jama.295.10.1147] [Medline: 16522835]

30. Lara-Palomo IC, Gil-Martínez E, Ramírez-García JD, et al. Efficacy of e-Health interventions in patients with chronic low-back pain: a systematic review with meta-analysis. Telemed J E Health 2022 Dec;28(12):1734-1752. [doi: 10.1089/tmj.2021.0599] [Medline: 35532971]
31. Fairbank JC, Pynsent PB. The Oswestry Disability Index. Spine (Phila Pa 1976) 2000 Nov 15;25(22):2940-2952. [doi: 10.1097/00007632-200011150-00017] [Medline: 11074683]
32. Bombardier C, Hayden J, Beaton DE. Minimal clinically important difference. Low back pain: outcome measures. J Rheumatol 2001 Feb;28(2):431-438. [Medline: 11246692]
33. Waddell G, Newton M, Henderson I, Somerville D, Main CJ. A Fear-Avoidance Beliefs Questionnaire (FABQ) and the role of fear-avoidance beliefs in chronic low back pain and disability. Pain 1993 Feb;52(2):157-168. [doi: 10.1016/0304-3959(93)90127-B] [Medline: 8455963]
34. Lam CL, Gandek B, Ren XS, Chan MS. Tests of scaling assumptions and construct validity of the Chinese (HK) version of the SF-36 Health Survey. J Clin Epidemiol 1998 Nov;51(11):1139-1147. [doi: 10.1016/s0895-4356(98)00105-x] [Medline: 9817131]
35. Monticone M, Frigau L, Vernon H, et al. Reliability, responsiveness and minimal clinically important difference of the two Fear Avoidance and Beliefs Questionnaire scales in Italian subjects with chronic low back pain undergoing multidisciplinary rehabilitation. Eur J Phys Rehabil Med 2020 Oct;56(5):32420712. [doi: 10.23736/S1973-9087.20.06158-4]
36. Salaffi F, Stancati A, Silvestri CA, Ciapetti A, Grassi W. Minimal clinically important changes in chronic musculoskeletal pain intensity measured on a numerical rating scale. Eur J Pain 2004 Aug;8(4):283-291. [doi: 10.1016/j.ejpain.2003.09.004] [Medline: 15207508]
37. Grönkvist R, Vixner L, Äng B, Grimby-Ekman A. Measurement error, minimal detectable change, and minimal clinically important difference of the Short Form-36 Health Survey, Hospital Anxiety and Depression Scale, and pain numeric rating scale in patients with chronic pain. J Pain 2024 Sep;25(9):104559. [doi: 10.1016/j.jpain.2024.104559] [Medline: 38734041]
38. Resche-Rigon M, White IR. Multiple imputation by chained equations for systematically and sporadically missing multilevel data. Stat Methods Med Res 2018 Jun;27(6):1634-1649. [doi: 10.1177/0962280216666564] [Medline: 27647809]
39. Moreno-Ligero M, Moral-Munoz JA, Salazar A, Failde I. mHealth intervention for improving pain, quality of life, and functional disability in patients with chronic pain: systematic review. JMIR Mhealth Uhealth 2023 Feb 2;11:e40844. [doi: 10.2196/40844] [Medline: 36729570]
40. Stark C, Cunningham J, Turner P, Johnson MA, Bäcker HC. App-based rehabilitation in back pain, a systematic review. J Pers Med 2022 Sep 22;12(10):1558. [doi: 10.3390/jpm12101558] [Medline: 36294697]
41. de Melo Santana B, Raffin Moura J, de Toledo AM, et al. Efficacy of mHealth interventions for improving the pain and disability of individuals with chronic low back pain: systematic review and meta-analysis. JMIR Mhealth Uhealth 2023 Nov 2;11:e48204. [doi: 10.2196/48204] [Medline: 37962085]
42. Owen PJ, Miller CT, Mundell NL, et al. Which specific modes of exercise training are most effective for treating low back pain? Network meta-analysis. Br J Sports Med 2020 Nov;54(21):1279-1287. [doi: 10.1136/bjsports-2019-100886] [Medline: 31666220]
43. Matheve T, Hodges P, Danneels L. The role of back muscle dysfunctions in chronic low back pain: state-of-the-art and clinical implications. J Clin Med 2023 Aug 24;12(17):5510. [doi: 10.3390/jcm12175510] [Medline: 37685576]
44. Coulombe BJ, Games KE, Neil ER, Eberman LE. Core stability exercise versus general exercise for chronic low back pain. J Athl Train 2017 Jan;52(1):71-72. [doi: 10.4085/1062-6050-51.11.16] [Medline: 27849389]
45. Alanazi MH, Parent EC, Dennett E. Effect of stabilization exercise on back pain, disability and quality of life in adults with scoliosis: a systematic review. Eur J Phys Rehabil Med 2018 Oct;54(5):647-653. [doi: 10.23736/S1973-9087.17.05062-6] [Medline: 29144110]
46. Fernández-Rodríguez R, Álvarez-Bueno C, Cavero-Redondo I, et al. Best exercise options for reducing pain and disability in adults with chronic low back pain: pilates, strength, core-based, and mind-body. A network meta-analysis. J Orthop Sports Phys Ther 2022 Aug;52(8):505-521. [doi: 10.2519/jospt.2022.10671] [Medline: 35722759]
47. Hou J, Yang R, Yang Y, et al. The effectiveness and safety of utilizing mobile phone-based programs for rehabilitation after lumbar spinal surgery: multicenter, prospective randomized controlled trial. JMIR Mhealth Uhealth 2019 Feb 20;7(2):e10201. [doi: 10.2196/10201] [Medline: 30785406]

Abbreviations

FABQ: Fear-Avoidance Beliefs Questionnaire
HIRS: Healbone Intelligent Rehabilitation System
LBP: low back pain
MCID: minimal clinically important difference
NLBP: nonspecific low back pain
NPRS: Numeric Pain Rating Scale
OBEG: outpatient-based exercise group
ODI: Oswestry Disability Index

OR: odds ratio
PT: physical therapist
RCT: randomized controlled trial
SF-36: 36-item Short-Form Health Survey
TBEG: telerehabilitation-based exercise group

Edited by L Buis; submitted 22.01.24; peer-reviewed by F Pereira, SS Naimi; revised version received 22.06.24; accepted 09.07.24; published 06.09.24.

Please cite as:
Shi W, Zhang Y, Bian Y, Chen L, Yuan W, Zhang H, Feng Q, Zhang H, Liu D, Lin Y
The Physical and Psychological Effects of Telerehabilitation-Based Exercise for Patients With Nonspecific Low Back Pain: Prospective Randomized Controlled Trial
JMIR Mhealth Uhealth 2024;12:e56580
URL: https://mhealth.jmir.org/2024/1/e56580
doi:10.2196/56580

Original Paper

Views of Service Users, Their Family or Carers, and Health Care Professionals on Telerehabilitation for People With Neurological Conditions in Ghana: Qualitative Study

Lorna Paul[1*], PhD; Katie Thomson[1*], PhD; Shadrack Osei Asibey[2*], PhD; Marian Brady[1*], PhD; Frederike van Wijck[1*], PhD; Derrick Antwi[2*], MSc; Eric Nkansah Opoku[1*], MSc; Fred Stephen Sarfo[2*], MD

[1]School of Health and Life Sciences, Glasgow Caledonian University, Glasgow, United Kingdom

[2]Department of Psychiatry, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana

[*]all authors contributed equally

Corresponding Author:
Lorna Paul, PhD
School of Health and Life Sciences
Glasgow Caledonian University
Cowcaddens Road
Glasgow, G4 0BA
United Kingdom
Phone: 44 01413318108
Email: Lorna.Paul@gcu.ac.uk

Abstract

Background: Up to 50% of people in low- and middle-income countries do not receive the rehabilitation they require. Telerehabilitation has the potential to improve access to neurorehabilitation services especially in low- and middle-income countries. Although there are reports of the barriers and facilitators to telerehabilitation in such settings, almost all are anecdotal. Furthermore, family or carers have a significant influence on the adoption and success of telerehabilitation, but their views have not been reported.

Objective: This study aimed to investigate the views of service users, their family or carers, and health care professionals (HCPs) on telerehabilitation for people with neurological conditions in Ghana.

Methods: Two focus groups were held at Komfo Anokye Hospital in Kumasi, Ghana: one in person for service users (n=11) and their family or carers (n=9), conducted in the Ghanaian language of Twi, and one hybrid for HCPs (n=18) conducted in English. The mean (SD) age of the service users was 59.8 (8.6) years; 5 users had a stroke and 6 had Parkinson disease. The HCP group consisted of 7 speech and language therapists, 3 physiotherapists, 3 occupational therapists, 3 medical staff, 1 nurse, and 1 industry representative. Focus groups were semi-structured and explored previous experiences of telerehabilitation, perceived benefits and challenges, and solutions to overcome these challenges. Focus groups were audio transcribed, and the service user transcript was translated into English. The resulting transcripts were analyzed using thematic analysis.

Results: Overall, participants were positive about the role of telerehabilitation but recommended hybrid delivery, with in-person rehabilitation in the early stages and telerehabilitation in the later stages. In relation to telerehabilitation in Ghana, there were 3 main themes: benefits, challenges or barriers, and implementation. Benefits included the convenience and lower cost for service users, the higher dose of therapy possible, and increased access for people in remote areas. However, challenges included lack of a stable internet connection, cost of phones and data packages, and low levels of literacy. Implementation issues included cultural relevance, information governance, and the platform used to deliver telerehabilitation, with most participants being familiar with WhatsApp.

Conclusions: Telerehabilitation has the potential to be a useful method of delivering rehabilitation to people with neurological conditions in Ghana, especially in a hybrid rehabilitation model with telerehabilitation augmenting in-person sessions. However, many people were unaware of telerehabilitation, and challenges such as a reliable internet connection, cultural relevance, and costs need to be addressed. Clinical trials of low-cost telerehabilitation interventions contextualized to the specific user group are required.

(JMIR Mhealth Uhealth 2024;12:e49501) doi:10.2196/49501

KEYWORDS

telerehabilitation; low- to middle-income country; LMIC; service user; health care professional; qualitative study; caregiver; neurorehabilitation; barriers; facilitators; eHealth; focus group; thematic analysis

Introduction

The World Health Organization defines rehabilitation as "a set of interventions designed to optimize functioning and reduce disability in individuals with health conditions in interaction with their environment" [1]. They also highlight the significant unmet need for rehabilitation services that is most evident in some low- and middle-income countries (LMICs), where up to 50% of people do not receive the rehabilitation they require [1].

Rehabilitation in Ghana, an LMIC, is provided by the public and private sector and costs between US $5 and $20 per session. Rehabilitation sessions are paid for by the individual or through various national insurance schemes. Even where individuals have insurance, the coverage may be limited and "top-up payments" are required. The duration of neurological rehabilitation can vary between 3 and 12 months depending on the condition, and often, there is a waiting list of 1-2 months.

Numerous barriers to rehabilitation have been cited in LMICs, such as Ghana, including low numbers of therapists, especially in rural locations, services being concentrated in main towns or cities with the majority of the population living rurally, transport issues, costs for appointments, and lack of specialist staff and equipment [2-5].

As well as experiencing barriers to rehabilitation, LMICs are experiencing an increase in life expectancy and greater numbers of people with noncommunicable diseases including neurological conditions [2,6,7]. Although in its early stages, telemedicine is increasingly being explored in LMICs to deliver care to challenging or remote areas. Telerehabilitation, a branch of telemedicine, may provide part of the solution to the increased patient demand coupled with restricted service especially in neurological rehabilitation. Telerehabilitation is defined as the provision of rehabilitation, including physiotherapy, speech and language therapy, and occupational therapy, over distance and, oftentimes, using communication technology [8]. Potential telerehabilitation technologies include telephone and video calls, apps, virtual reality, and robotics [4]. A recent review suggests that the telephone or video is the main media through which telerehabilitation is delivered worldwide [9].

The evidence base for the acceptability, feasibility, and clinical and cost-effectiveness of telerehabilitation for people with neurological conditions is increasing. Several recent systematic and scoping reviews suggest that telerehabilitation improves access to services and is generally well received by patients and therapists, with high adherence to telerehabilitation programs and low adverse events, supporting its safety in practice [10,11]. Evidence of the clinical effectiveness of telerehabilitation for people with neurological conditions is mixed, but overall it is reported to be at least equivalent to standard care [12,13]. There is, however, limited evidence on the cost-effectiveness of telerehabilitation [6,9]. Much of the research in telerehabilitation has been undertaken in high-income countries, with notably fewer studies in LMICs [6,14,15], where the rehabilitation context, as well as the barriers and facilitators to the feasibility, adoption, scalability, and sustainability of telerehabilitation, may be quite different. In terms of neurological conditions, most telerehabilitation research in LMICs has been conducted in stroke [3,14,16].

To influence rehabilitation and improve patient outcomes, technology needs to be adopted into services. There are various models and theories of technology adoption including the Technology Acceptance Model and Self-Determination Theory. Central to these models are the beliefs and attitudes of the users of the technology. In the context of telerehabilitation, only a few previous studies exist on the views and beliefs of therapists and patients in LMICs and beyond [7]; however, these studies tend to be process evaluations of specific telerehabilitation interventions that are being researched and thus are open to selection bias as participants are exposed to the technology under study conditions. Furthermore, although telerehabilitation generally happens at home, there is a recognition that almost no research has sought the views of the patient, carer, or family member—although they have a strong and significant influence on the adoption and success, or otherwise, of telerehabilitation interventions [17].

Telerehabilitation is in its infancy in LMICs, including Ghana; however, to ensure successful adoption, implementation to routine practice, and scalability, it is important to understand the views of health care professionals (HCPs), service users (patients), and their carers. Therefore, the aims of this study were to investigate the views of HCPs involved in neurological rehabilitation, service users with a long-term neurological condition, and the carer or family member of someone with a long-term neurological condition in terms of previous experience of telerehabilitation, the perceived potential benefits, the potential challenges, and possible solutions to overcome these challenges in Ghana.

Methods

Study Design

Focus groups were used as the method of data collection for this qualitative study, and this paper is presented in line with the Consolidated Criteria for Reporting Qualitative Research guidelines. Separate focus groups were conducted: one for service users and carers or family members and another for HCPs and IT industry representatives. Both focus groups were conducted in person in Komfo Anokye Hospital in Kumasi, Ghana, with HCPs given the option of joining remotely via teleconference. The service user focus group was conducted in person to ensure that those who did not have technology or technology skills could contribute, to encourage engagement, and to support those with communication problems. For service users and carers traveling, expenses were covered and both groups were provided with lunch at the end of the focus group. The service user and carer focus group was conducted in Twi,

the local language, and the HCP focus group was conducted in English.

Inclusion Criteria

Inclusion criteria for the stakeholder group were HCPs (physiotherapists, occupational therapists or speech and language therapists, or medical staff) with experience of working in neurological rehabilitation in Ghana or staff with expertise in commissioning or delivering health services (commissioner or industry expert) and able to speak and understand English and attend the focus group either in person or via teleconference (Zoom).

Inclusion criteria for the service users were people with a neurological condition such as stroke, Parkinson disease (PD), or spinal cord injury and able to travel to the venue. Carers or family members had to have experience of caring for someone with a neurological condition and able to travel to the venue.

Recruitment

A convenience sample of service users, who fulfilled the inclusion and exclusion criteria, was identified from two sources: (1) members of a support group for people with PD were telephoned by the research coordinator (SOA) and (2) stroke survivors attending a neurology clinic at Komfo Anokye Hospital 2 weeks before the focus groups were approached by the research coordinator. The study was explained to potential participants, and if they were happy to take part, they were given the details of the focus group date, time, and venue. They were given the option of bringing a carer or family member to the focus group although that was not a requirement. A convenience sample of HCPs was recruited from the professional networks of the research team. They were contacted through email or WhatsApp groups and asked to take part. If they were interested, study information was emailed to them with details of the date, time, and venue of the focus group, or if attending via Zoom, a link was shared. Consent was taken from all participants at the start of the focus group.

Focus Groups

A focus group schedule was prepared for each of the 2 focus groups including main questions and prompts. For both groups, the questions were related to their experiences of telerehabilitation, perceived potential benefits of delivering or receiving rehabilitation via telerehabilitation, perceived potential challenges or difficulties, and suggestions for overcoming these difficulties.

The service user and carer focus group was facilitated by the male research coordinator in the room (SOA). This facilitator had been involved in recruitment of participants, had experience of conducting focus groups, and spoke the local dialect (Twi). The HCP focus group was facilitated by a senior female researcher from the United Kingdom (LP) via Zoom. This facilitator was a physiotherapist, with experience in telerehabilitation research and facilitating focus groups. She did not know any of the participants except those within the research team. The research team was introduced, and participants were provided with background information on the purpose of the study. The service user focus group lasted approximately 1 hour and the HCP group 1.5 hours. Both were audio-recorded and transcribed verbatim in the language in which they were conducted. The transcript of the service user and carer focus group was translated into English by professional transcribers from the University of Energy Sunyani, Ghana, and checked by the research team for accuracy. In addition, research team members made field notes during and after the focus groups.

Analysis

Thematic analysis was the method of data analysis using the 6 phases outlined by Braun and Clark [18], including data familiarization, generation of initial codes, generation of themes, review of themes, definition of themes, and writing the report. Initial coding of the transcripts was completed by one researcher (KT) who had no previous relationship with study participants and checked by a second researcher (LP) who had conducted the HCP focus groups. Themes were derived from the data, and initial themes were presented to the research team (in verbal, written, and diagrammatic form), which were further refined after feedback and discussion. Thematic analysis was undertaken using Microsoft Word (Microsoft Corporation) with documentation shared at each stage ensuring transparent recording of the data analysis process with the research team. Thematic analysis of each focus group took place separately before a final round of analysis integrated these together. Further feedback was provided by the research team before the final themes were confirmed.

Ethical Considerations

Ethical approval was received from the Ethics Committee of Kwame Nkrumah University of Science and Technology, Kumasi, in December 2022 (reference CHRPE/AP/822/22), and all participant data were deidentified. Informed consent was provided by all participants. Travel expenses were covered and lunch was provided for participants who attended in person focus groups.

Results

Characteristics of Participants

A total of 13 service users agreed to take part in the focus group; however, 2 were unable to attend on the day of the focus group, so overall 11 service users (9 male users) took part (Table 1). Service users had a mean (SD) age of 59.8 (8.6) years; 6 lived in an urban location, and 5 in a semiurban location. Six service users had PD, and 5 had had a stroke. The mean (SD) age of those with stroke was 53 (8.3) years, and the mean (SD) time since stroke was 4 (1.6) years. In contrast, the mean (SD) age of those with PD was 65.5 (2.5) years, and they had had the condition for a mean (SD) of 3.8 (2.1) years. In addition, 9 carers took part (4 male carers), and their mean (SD) age was 38.4 (8.6) years. They had a variety of occupations and had been caring for people with PD (n=6) and stroke (n=3) for a mean (SD) of 3.4 (2.2) years (Table 1).

Service users had different prior exposure to telerehabilitation, both synchronous and asynchronous. One had participated in a previous trial of an app that delivered an individualized rehabilitation program remotely supervised by a therapist [14],

some had received rehabilitation plans via WhatsApp to undertake without supervision, and others had no previous experience of telerehabilitation.

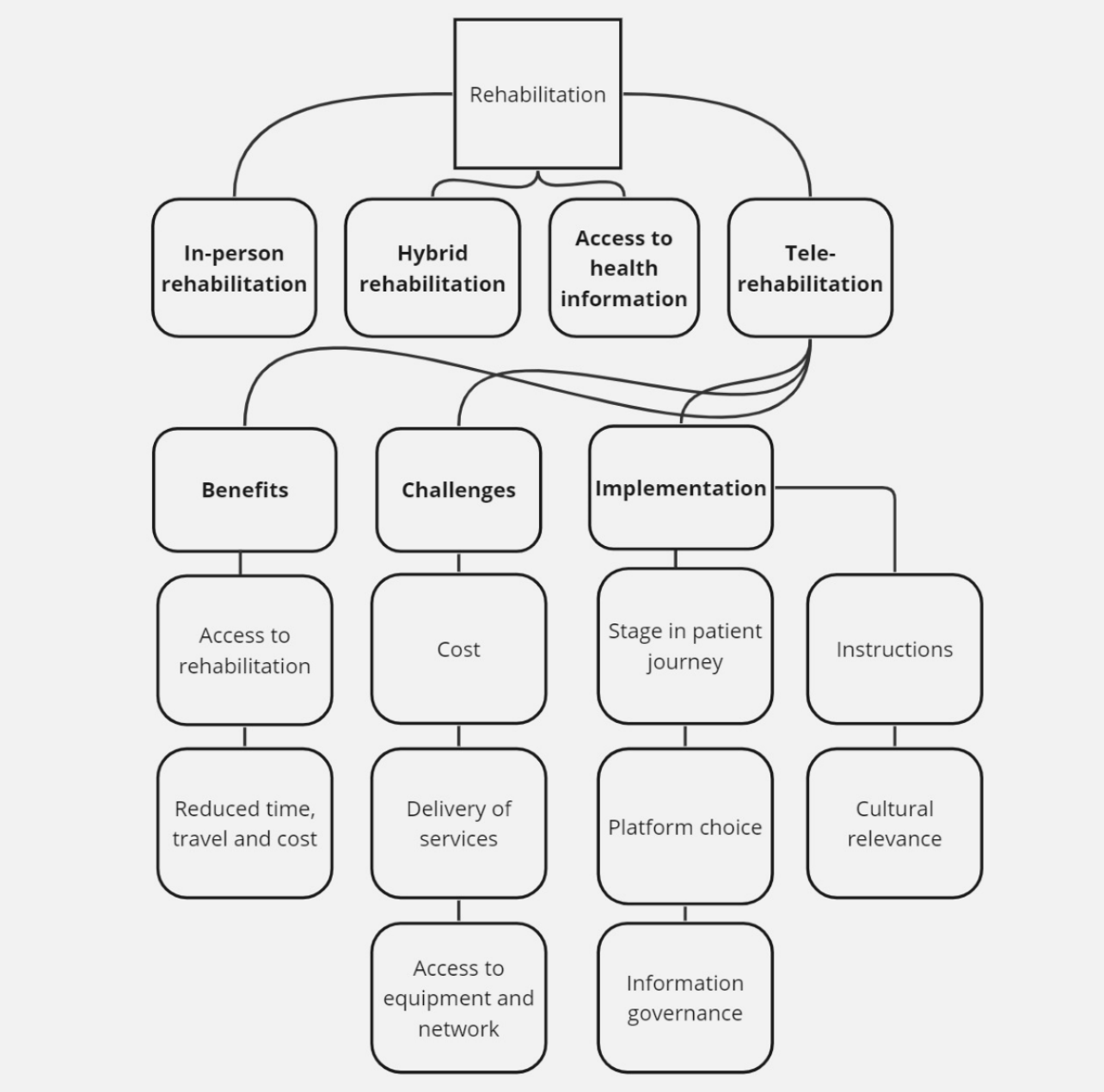
Table 1. Demographic details of service users (n=11) and carers (n=9).

Demographic	Value
Service users (n=11)	
Sex, n	
Male	9
Female	2
Age (years), mean (SD)	59.8 (8.6)
Diagnosis of service user, n (%)	
Stroke	6 (55)
Parkinson disease	5 (45)
Time since diagnosis (years), mean (SD)	6.0 (3.9)
Residence, n (%)	
Urban	6 (55)
Semiurban	5 (45)
Rural	0 (0)
Carer or significant other (n=9)	
Sex, n	
Male	4
Female	5
Age (years), mean (SD)	34.8 (8.6)
Diagnosis of service user, n (%)	
Stroke	3 (33)
Parkinson disease	6 (67)
Length of care (years), mean (SD)	3.3 (2.2)
Relationship to service user, n (%)	
Daughter	4 (44)
Brother	2 (22)
Spouse	1 (11)
Son	1 (11)
Unknown	1 (11)
Occupation, n (%)	
Trader	2 (22)
Driver	2 (22)
Hairdresser	1 (11)
Caterer	1 (11)
Shoemaker	1 (11)
Seamstress	1 (11)
Student	1 (11)
Residence, n (%)	
Urban	5 (56)
Semiurban	4 (44)
Rural	0 (0)

Eighteen stakeholders took part in the focus groups (10 female and 8 male). In total, 7 speech and language therapists, 3 physiotherapists, 3 occupational therapists, 2 neurologists, and 1 specialist physician, 1 nurse, and 1 person from industry participated. Eleven of the participants were aged 30-39 years, 3 were aged 20-29 years, and 3 were aged over 40 years. They had been in their current post for a mean of 4.8 (SD 3.3) years, and all worked in urban locations, most in the 2 main cities of Accra and Kumasi.

Analysis of the focus groups resulted in the following themes and subthemes (Figure 1).

Figure 1. Focus group themes.



Rehabilitation Services

Service users identified different ways of accessing rehabilitation services, in-person services, telerehabilitation, and hybrid (a combination of in-person and remote services). Service users also expressed a need for increased awareness of the differing ways to access rehabilitation services and a need for reliable health information.

In-Person Rehabilitation

Service users valued in-person rehabilitation for the purpose of monitoring their rehabilitation progress and receiving feedback on this.

> *If there is an improvement, the doctor or the health facility that will tell you to either reduce or increase some aspect of your rehabilitation so that the body will recover.* [Service user]

It was however acknowledged that in-person rehabilitation had limitations particularly when there was a large demand for

services, which could result in long queues at the health care facility, or the feedback received was not as hoped.

> *I stopped the rehab because of waking up early to join a queue.* [Service user]
>
> *One physio teasingly told me that my rehab has taken too long which downed my emotion and that made me stop coming for physio.* [Service user]

Hybrid Rehabilitation

Some service users and carers expressed a preference for a hybrid model of rehabilitation (in-person and telerehabilitation), which they had previously found useful.

> *I feel like if we combine the two it will help, both should go hand in hand.* [Service user and carer]
>
> *Both, because sometimes you may miss or do certain procedures wrongly without knowing but going to the hospital you will be able to know the exact procedure, then you can go ahead and practice at home.* [Service user]

Access to Health Information

A need for access to reliable health information was highlighted by both patients and carers with the potential for technology to be used as means of distributing information. Concern was expressed at misinformation, which hindered treatment such as medication use.

> *There should be associations to broadcast to people and to serve as guidance for others to follow.* [Service user and carer]
>
> *I have seen that everything is going on well here, but the information given is very low and because of that most people depend on herbal medicines.* [Service user and carer]

Telerehabilitation

Service users and carers who had experienced telerehabilitation found this to be a helpful way to access services and were happy to recommend this method of service delivery to others.

> *Yes, (I would recommend it) because I have tried and tested and it really helped me.* [Service user and carer]
>
> *We the caregivers also go through the same stress as the patients, so our experiences shows that the tele-rehab is very good.* [Carer]

For others, this was a novel way to access rehabilitation services, which, they thought, would be of benefit, indicating that they would be happy to try this in the future.

> *Gives me hope that even if a person gets any neurological disorder, the telerehab is there to help them get back their strength.* [Service user and carer]

However, there was a lack of awareness of telerehabilitation with service users, suggesting that further awareness raising was needed as well as ensuring that access was available throughout Ghana.

> *Many Ghanaians are not aware of the telerehab and so it should be publicised through TV and other media groups.* [Service user and carer]
>
> *Telerehabilitation should be extended to clinics to enhance easy access. It should not be in the big hospitals alone.* [Service user and carer]

A range of telerehabilitation services were described by service users who had experienced stroke (including those with aphasia), those with PD, and those with tinnitus or balance issues. These services included checking rehabilitation progress or receiving speech and language therapy sessions. In addition, service users sought further follow-up and reminders through telerehabilitation.

HCPs also had a range of experiences using telerehabilitation ranging from no involvement to using telerehabilitation for a range of rehabilitation purposes including teaching exercises, promoting engagement in activities, conducting hearing assessments, sending intervention messages and information, balance re-education, and reviewing videos to direct parent-led rehabilitation.

Benefits of Telerehabilitation

Reduction in Travel, Time, and Associated Costs

A range of benefits were identified by service users, carers, and HCPs, including a reduction in travel, which made rehabilitation more convenient, easier to access, less stressful, accommodated other caring responsibilities, and was more affordable. Less time was also spent waiting for rehabilitation once patients and carers arrived at the rehabilitation unit.

> *It helped because having her come all the way to the clinic and the child who didn't sit and fussy two hours was also time that was solved with tele-therapy.* [HCP]
>
> *It reduces financial costs to the hospital and the stress involved in sitting in a trotro (car).* [Service user and carer]

Increased Access to Rehabilitation

HCPs identified that they were able to access service users who lived some distance away (including visual access to the home environment) increasing access and relevance of the rehabilitation services delivered. This provided the opportunity for further rehabilitation, increasing the intensity, consistency, and adherence to rehabilitation offered.

> *with the in-person you're scheduled to come let's say two or three days a week but with telerehabilitation, even if you want to do it every day as far as the resources are there, you can do it every day. It also allows you to do it more frequently and over a more sustained period of time. That's one of the benefits.* [HCP]
>
> *It can help with intensity because the more we meet them online we can achieve our goals. It is easy to find your therapy with more therapy sessions without having to move from one location to another. You sit in the comfort of your home and have more therapy*

> *sessions within the scheduled period but with a cost-effective system.* [HCP]

HCPs also saw the benefits of using devices such as mobile phones as a repository for information. This provided service users with a reminder of their rehabilitation exercises.

> *Most patients forget the therapy exercise we do for them so probably videoing it and saving for patients can help them do it at their leisure time.* [HCP]

HCPs reported that developing new ways of delivering rehabilitation services encouraged them to be creative in their approach including delivering content in a range of local dialects with the potential of developing new rehabilitation services or using technology to measure changes in service user knowledge for example.

> *I would like to add it makes you creative as a professional because you look for other means of making it fun with telerehabilitation in order to suit your client. It brings out creativity in you.* [HCP]
>
> *I think it will give the opportunity to assess the clients' environment for recommendations for possible adaptations to enhance function because when the person comes in-person, they may describe the home environment and we don't know exactly how it looks but through telerehab with video conference or picture you can see how the environment really looks like for recommendations for the adaptations.* [HCP]

Rehabilitation Progress Using Telerehabilitation

Service users highlighted the progress that they had made with their rehabilitation delivered using telerehabilitation, which included improvements in speech, arm movements, activity levels, and independence in activities.

> *It is through the tele-rehab that I am able to lift my hand today. It really helped me.* [Service user]
>
> *I'm forever grateful for the physios because up till date, they still call me and they helped in so many ways.* [Service user]

Challenges of Telerehabilitation

Challenges of delivering telerehabilitation were identified by service users, carers, and HCPs including access to equipment and availability of a reliable internet connection, the cost of data packages, and challenges to rehabilitation delivery.

Equipment and Network Access

Availability of equipment and the requirement of a reliable network needed for telerehabilitation were highlighted as a barrier. Most service users in the focus group had access to mobile phones (including smartphones that access the internet), but not all did. Challenges with unreliable internet connections meant therapists often prepared alternatives such as printed exercise sheets, sending SMS text messages to promote engagement, watching videos offline, use of images (rather than videos), or switching platforms seeking a better connection. At times, however, the connection was so poor that therapy sessions were abandoned and rescheduled or therapists advised returning to in-person rehabilitation. For some service users, this meant finding other solutions including referral to local therapists to access rehabilitation.

> *Only 35% of our patients said that they own a smartphone and even of that 35% it's not necessarily their own but there is a smartphone in the house. It's not as if the smartphone belongs to them.* [HCP]
>
> *The network decided to fail all of us. When we call again, we can't hear anything. We used other platforms FaceTime, other things, but the network just wouldn't channel us until we had to put the whole therapy to an ending try to refer to someone closer to him.* [HCP]
>
> *In terms of connectivity, we can record the videos, store them, and send them so that when the connectivity improves the person can use the stored videos to be able look back on their exercise.* [HCP]

Financial Cost

Those who did have access to smartphones acknowledged that there was a financial burden of purchasing packages or "bundles" for calls and network access. Some rehabilitation services that required use of additional equipment provided this, but not all did.

> *Most of us are retired so money is hard to come by so it will really help us (if the equipment and data package are provided).* [Service user]

Rehabilitation Delivery

Use of telerehabilitation meant that therapists had to find new ways of delivering therapy services for an individual rather than their usual in-person service that included group therapy, acknowledging that not all service users had the digital literacy skills or ability needed to use technology.

> *because if I am at the physiotherapy department I can supervise maybe 4 or 5 patients simultaneously; this one is doing this, this one is doing that, I can just observe them but in telerehab they have to do one on one supervision so that may also eventually reduce the number of participants or patient they can attend to at a time if it is ongoing supervision they have to do.* [HCP]
>
> *Sometimes they are not very tech savvy so we need to see them in-person. And we have had some patients try it but most are not tech savvy so we haven't been able to expand this to all our clients.* [HCP]
>
> *Another barrier I was thinking about was that if the client has multiple deficits so maybe visual needs and other possible deficits, I think that may impact teletherapy.* [HCP]

Other challenges with delivery of rehabilitation, such as location or timing of therapy sessions, the move toward carer involvement, and the type of therapy session being delivered, were highlighted.

> *Mum and I decided to have teletherapy instead [for the child] because they were coming from far. But each time I book them, I give them a time. But each*

> *time I am ready to have the teletherapy the child might be asleep or would be at a place where it is uncomfortable place to have teletherapy.* [HCP]
>
> *So let me add one more barrier, with the issue of adaptive devices especially if there are no caregivers who really understand or who can be trained on how to retrain the patient on how to use the assistive device it becomes difficult Then unless the patient comes in-person for you to maybe fabricate or measure their assistive device and then train the patient in how to use it.* [HCP]

Implementation of Telerehabilitation

Stage in the Patient Journey

Service users had a range of opinions on what stage in their rehabilitation journey they preferred to receive telerehabilitation, from the initial stages of diagnosis to using telerehabilitation following a period of in-person rehabilitation.

> *Initial stages of my condition.* [Service user]
>
> *At the initial stages, you should visit the hospital for physio then later, join the telerehab.* [Service user and carer]

Platform of Choice

WhatsApp was the preferred platform although Microsoft Teams, Zoom, and FaceTime were also used. Ease of use was important with WhatsApp reported to be familiar with options for low data consumption.

> *Some find zoom cumbersome it's difficult for them to manage their way through zoom and then having meetings with them, but WhatsApp is just like having a call. I just call you; you see me on the video and then whatever we need to, it's easier using that platform compared to the other platforms. I used WhatsApp because I think it is easier over here and with data consumption, you know it has the option for you to select low data mode where the streaming is easier for the patient.* [HCP]

Information Governance

Integrating use of personal technology (such as mobile phones) into therapy did cause concern for information security particularly when sharing media such as videos.

> *I have been thinking about ever since I started working with the clients that I work with. Recently I lost the password to my laptop, and I was thinking there might be a case where a third party has to come in. My phone gets called or any other thing, it's just confidentiality how is it handled? because a third party will have to come in and help me unlock or do something to my phone Videos, I have a lot of videos of the same person doing mostly I delete them, but I have been thinking about situations where I might forget, or something would happen that another person had access to the videos. So sometimes you know a client may feel uncomfortable sharing videos across, because they don't know where that would end so that's one of the barriers I think.* [HCP]

Instructions for Use

Participants felt that further instructions or information was needed to take full advantage of rehabilitation delivered via technology with access to videos requested.

> *I only had to follow the instructions on the phone.* [Service user and carer]
>
> *I feel like the physiotherapists should be involved in the video demonstrations. There should be visual demonstration videos so the patient can see the physio demonstrating the exercise and imitate it correctly.* [Service user and carer]

Cultural Relevance

It was highlighted that many of the current telerehabilitation exercise videos featured White individuals with instructions in English. To increase the cultural relevance of the materials, videos should be in a local dialect featuring Black individuals completing exercises.

> *It should be conducted in our local dialect Twi to be specific.* [Service user and carer]
>
> *Language shouldn't be an issue if we stick to the local dialect. But using English it could be a challenge...another one has to do with the videos where a white person is involved, it makes it difficult for the most clients to understand but it can solve by introducing blacks in such videos.* [HCP]

Discussion

Principal Results

Participants from both focus groups were overall positive in terms of telerehabilitation for people with neurological conditions in Ghana but identified a number of challenges. The conceptual framework for sustainable eHealth in resource-limited countries proposed by Fanta and Pretorius [19], comprising technological, social, economic, and organizational factors, will be used to discuss the findings of the focus groups.

In terms of *technological* factors, one of the main challenges was the lack of a stable internet connection to conduct a telerehabilitation session. This has been identified as one of the main barriers to successful implementation of telerehabilitation in many others studies in LMICs [4,14,15,20], especially in relation to the transfer of images or videos. There was also an appreciation that not all service users had access to a smartphone, as previously reported [5,16]. Interestingly, only 1 therapist raised concerns about the security of service user data particularly when videos of service users completing tasks are sent. In terms of the technology used, this is one of the first studies to report that WhatsApp was the preferred platform for delivering telerehabilitation as it is commonly used in Ghana by both HCPs and service users, so people are familiar with its use and it also has an option for a low data mode. Previous studies of telerehabilitation in LMICs have provided patients

with videos to play in a video player [21] or have used Skype [15].

Social factors were most commonly raised within the focus groups. Participants overall had positive views on telerehabilitation but also discussed the advantages or preferences of in-person rehabilitation especially with regard to the HCP monitoring their progress. Previous research particularly in relation to patients receiving speech and language therapy also reported good satisfaction with telerehabilitation interventions, but many preferred in-person therapy where there was better eye contact between the patient and the therapist and it was easier to understand facial expression [15,22].

Participants in this study however recognized that there were many barriers to service users accessing in-person rehabilitation. Service users also highlighted the lack of awareness of telerehabilitation services in Ghana.

Like previous studies of telerehabilitation generally [11,13,20] and specifically in relation to LMICs [3,23], participants reported a number of benefits to telerehabilitation especially the convenience, requiring less travel time and reduced travel cost. Some HCPs raised that telerehabilitation improved access to rehabilitation services for patients, especially those who stayed a distance from the clinic. In Ghana, up to 43% of people after stroke access herbal medicines [24], as was raised by 1 person in the focus groups, for which there is little or no efficacy data. Improving access to rehabilitation may reduce the reliance on herbal medicine for some patients.

However, there were also some negative aspects. Not all service users had a smartphone, and furthermore, low levels of literacy and digital skills were barriers to implementing telerehabilitation in line with previous papers [3,4,25]. Although there were examples of some apps being used by therapists, they were not felt to be culturally relevant for the Ghanaian context as they were in English rather than the local dialect and tended to have White people (generally Americans or Europeans) demonstrating the activity, which had the potential to reduce engagement of service users. Odetunde et al [21] developed a telerehabilitation video solution for patients with stroke in Nigeria, delivered both in the local language Yoruba and in English, and this was positively received by participants. To promote uptake and adherence, future development of telerehabilitation interventions should consider the local language and other contextual issues [4,10].

In terms of *economic factors*, as discussed above, not all service users had access to a smartphone, and the financial implications of requiring such a phone and data package were a barrier for many people. Sarfo et al [16] reported that only 35% of their respondents owned a smartphone although 80% had a family member who did. A 2020 household survey of information and communication technology use in Ghana reported that 47.9% of the population had a basic phone and 46.1% had a smartphone; however, in rural areas, 61.3% of people had a basic phone and only 28.1% had a smartphone [26]. However, the figures suggest that family or carers may need to be actively involved for telerehabilitation to be implemented, especially in rural areas of Ghana and in other LMICs. On the positive side, telerehabilitation reduced costs and time required to travel to in-person appointments. In terms of clinical effectiveness, telerehabilitation is variable with some evidence in support of telerehabilitation and other evidence suggesting that it is not superior to conventional care [13,27]. Although there are associated costs, even if it is not superior, the convenience, reach of services, and time-saving aspects would support its further development; however, cost-effectiveness analyses of telerehabilitation for neurological conditions generally and specifically related to resource-limited settings are required [6,9].

Organizational factors were generally related to the delivery of the telerehabilitation interventions. Telerehabilitation was felt to be a positive development that could address long waiting times and high demand on services. A number of service users with stroke and PD had used telerehabilitation with examples of telerehabilitation delivery across different allied HCP groups: physiotherapy, occupational therapy, and speech and language therapy. Service users however felt that they would have liked more instructions on how to use telerehabilitation.

HCPs reported using telerehabilitation for undertaking patient assessments, sending information to service users, and receiving videos of patient progress. Neurological rehabilitation often involves intensive therapy with highly repetitive, task-specific exercise to optimize neuroplastic changes in the central nervous system [12]. A novel finding of this study was that therapists reported that telerehabilitation, in this resource-limited setting, allowed more intensive and consistent therapy than would be possible in person and which importantly facilitated increased adherence and improved outcomes. An additional novel finding was that therapists felt that new ways of delivering therapy encouraged them to be more creative in their approach.

The negative aspects considered under organizational factors were the location and timing of telerehabilitation sessions that were arranged in advance; however, when the appointment time came, the service user was not available, or it was not appropriate to complete the session. This often meant that therapists had to have alternative plans in place should that occur. There was also an appreciation that some activities, such as providing assistive devices, required to be done in person.

Strengths and Limitations

This research had a number of strengths. The views of service users and carers, HCPs, and other stakeholders on telerehabilitation in Ghana were sought directly. Aljabri et al [5] recommended that future research should explore the views of HCPs from different disciplines, which we did, including a range of HCPs such as occupational therapy and speech and language therapy, professions seldom included in the telerehabilitation literature in LMICs—perhaps due to their relatively small numbers compared with, for example, physiotherapy. This is also the first study to include participants with PD from LMICs, although there are previous reviews of telerehabilitation in PD but not in an LMIC context [27]. To be as inclusive as possible, the service user and carer focus group was conducted in the local language. The use of teleconferencing for the stakeholder focus group allowed a wide geographical spread of participants from across Ghana. However, this research also had a number of limitations. Although traveling expenses

and refreshments were provided, service users and their carers had to be able to travel to the hospital to take part in the focus group, possibly biasing the sample to a local, urban dwelling and less disabled group. Also, many had had experience of using telerehabilitation, so perhaps they did not represent the views of most people with neurological conditions in Ghana; however, it was important that they were able to share their experiences. Also, none of the study participants resided in a rural setting, thus limiting the transferability of the views captured in this study.

Conclusions

This is the first study to elicit the views of service users, carers, and HCPs of telerehabilitation for people with neurological conditions in a resource-limited setting of Ghana. The focus group findings overall demonstrated that service users, carers, and HCPs had positive views and experiences of telerehabilitation, especially the convenience and lower cost for service users and the consistency and higher intensity of therapy possible, with some negative aspects including lack of a stable internet connection, cost of phones and data packages, and low levels of literacy. Overall, the findings suggest the need for future research of the clinical and cost-effectiveness of lost cost telerehabilitation interventions for people with neurological conditions, taking into account the local context in Ghana and other LMICs. Telerehabilitation in Ghana is currently not covered within the National Insurance system; however, these findings support the development of telerehabilitation in Ghana with suggestions for future implementation and scale.

Acknowledgments

We would like to thank all the participants for their active contribution and professional transcribers from the University of Energy Sunyani Ghana transcribing the focus group recordings. This work was supported by the Academy of Medical Sciences Global Challenges Research Network Grant (Ref GCRFNGR5\1278).

Conflicts of Interest

None declared.

References

1. Rehabilitation competency framework. World Health Organization. 2021. URL: https://www.who.int/teams/noncommunicable-diseases/sensory-functions-disability-and-rehabilitation/rehabilitation-competency-framework [accessed 2024-02-23]
2. Sarfo FS, Adamu S, Awuah D, Ovbiagele B. Tele-neurology in Sub-Saharan Africa: a systematic review of the literature. J Neurol Sci 2017;380:196-199 [FREE Full text] [doi: 10.1016/j.jns.2017.07.037] [Medline: 28870567]
3. Kayola G, Mataa MM, Asukile M, Chishimba L, Chomba M, Mortel D, et al. Stroke rehabilitation in low- and middle-income countries: challenges and opportunities. Am J Phys Med Rehabil 2023;102(2S Suppl 1):S24-S32 [FREE Full text] [doi: 10.1097/PHM.0000000000002128] [Medline: 36634327]
4. Bonnechère B, Kossi O, Mapinduzi J, Panda J, Rintala A, Guidetti S, et al. Mobile health solutions: an opportunity for rehabilitation in low- and middle income countries? Front Public Health 2022;10:1072322 [FREE Full text] [doi: 10.3389/fpubh.2022.1072322] [Medline: 36761328]
5. Aljabri NQ, Bulkeley K, Cusick A. Telerehabilitation in the Middle East North Africa Region: a structured review. Int J Telerehabil 2021;13(2):e6401 [FREE Full text] [doi: 10.5195/ijt.2021.6401] [Medline: 35646235]
6. Solomon RM, Dhakal R, Halpin SJ, Hariharan R, O'Connor RJ, Allsop M, et al. Telerehabilitation for individuals with spinal cord injury in low-and middle-income countries: a systematic review of the literature. Spinal Cord 2022 May;60(5):395-403 [FREE Full text] [doi: 10.1038/s41393-022-00797-8] [Medline: 35411024]
7. Smythe T, Inglis-Jassiem G, Conradie T, Kamalakannan S, Fernandes S, van-Niekerk SM, et al. Access to health care for people with stroke in South Africa: a qualitative study of community perspectives. BMC Health Serv Res 2022;22(1):464 [FREE Full text] [doi: 10.1186/s12913-022-07903-9] [Medline: 35395847]
8. Russell TG. Physical rehabilitation using telemedicine. J Telemed Telecare 2007;13(5):217-220. [doi: 10.1258/135763307781458886] [Medline: 17697506]
9. Stephenson A, Howes S, Murphy PJ, Deutsch JE, Stokes M, Pedlow K, et al. Factors influencing the delivery of telerehabilitation for stroke: a systematic review. PLoS One 2022;17(5):e0265828 [FREE Full text] [doi: 10.1371/journal.pone.0265828] [Medline: 35544471]
10. Moulaei K, Sheikhtaheri A, Nezhad MS, Haghdoost A, Gheysari M, Bahaadinbeigy K. Telerehabilitation for upper limb disabilities: a scoping review on functions, outcomes, and evaluation methods. Arch Public Health 2022;80(1):196 [FREE Full text] [doi: 10.1186/s13690-022-00952-w] [Medline: 35999548]
11. Tchero H, Teguo MT, Lannuzel A, Rusch E. Telerehabilitation for stroke survivors: systematic review and meta-analysis. J Med Internet Res 2018;20(10):e10867 [FREE Full text] [doi: 10.2196/10867] [Medline: 30368437]
12. English C, Ceravolo MG, Dorsch S, Drummond A, Gandhi DB, Green JH, et al. Telehealth for rehabilitation and recovery after stroke: state of the evidence and future directions. Int J Stroke 2022;17(5):487-493. [doi: 10.1177/17474930211062480] [Medline: 34983266]

13. Appleby E, Gill ST, Hayes LK, Walker TL, Walsh M, Kumar S. Effectiveness of telerehabilitation in the management of adults with stroke: a systematic review. PLoS One 2019;14(11):e0225150 [FREE Full text] [doi: 10.1371/journal.pone.0225150] [Medline: 31714924]
14. Sarfo FS, Adusei N, Ampofo M, Kpeme FK, Ovbiagele B. Pilot trial of a tele-rehab intervention to improve outcomes after stroke in Ghana: a feasibility and user satisfaction study. J Neurol Sci 2018;387:94-97 [FREE Full text] [doi: 10.1016/j.jns.2018.01.039] [Medline: 29571880]
15. Jahromi ME, Ahmadian L. Evaluating satisfaction of patients with stutter regarding the tele-speech therapy method and infrastructure. Int J Med Inform 2018;115:128-133. [doi: 10.1016/j.ijmedinf.2018.03.004] [Medline: 29779715]
16. Sarfo FS, Adamu S, Awuah D, Sarfo-Kantanka O, Ovbiagele B. Potential role of tele-rehabilitation to address barriers to implementation of physical therapy among West African stroke survivors: a cross-sectional survey. J Neurol Sci 2017;381:203-208 [FREE Full text] [doi: 10.1016/j.jns.2017.08.3265] [Medline: 28991682]
17. Niknejad N, Ismail W, Bahari M, Nazari B. Understanding telerehabilitation technology to evaluate stakeholders' adoption of telerehabilitation services: a systematic literature review and directions for further research. Arch Phys Med Rehabil 2021;102(7):1390-1403. [doi: 10.1016/j.apmr.2020.12.014] [Medline: 33484693]
18. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
19. Fanta GB, Pretorius L. A conceptual framework for sustainable eHealth implementation in resource-constrained settings. S Afr J Ind Eng 2018;29(3):132-147 [FREE Full text] [doi: 10.7166/29-3-2055]
20. Nizeyimana E, Joseph C, Louw QA. A scoping review of feasibility, cost-effectiveness, access to quality rehabilitation services and impact of telerehabilitation: a review protocol. Digit Health 2022;8:20552076211066708 [FREE Full text] [doi: 10.1177/20552076211066708] [Medline: 35223074]
21. Odetunde MO, Binuyo OT, Maruf FA, Ayenowowon SO, Okonji AM, Odetunde NA, et al. Development and feasibility testing of video home based telerehabilitation for stroke survivors in resource limited settings. Int J Telerehabil 2020;12(2):125-136 [FREE Full text] [doi: 10.5195/ijt.2020.6321] [Medline: 33520100]
22. Shahouzaie N, Arefi MG. Telehealth in speech and language therapy during the COVID-19 pandemic: a systematic review. Disabil Rehabil Assist Technol 2022:1-8. [doi: 10.1080/17483107.2022.2122605] [Medline: 36129435]
23. Selamat SNS, Me RC, Ainuddin HA, Salim MSF, Ramli HR, Romli MH. The application of technological intervention for stroke rehabilitation in Southeast Asia: a scoping review with stakeholders' consultation. Front Public Health 2021;9:783565 [FREE Full text] [doi: 10.3389/fpubh.2021.783565] [Medline: 35198531]
24. Sarfo FS, Ovbiagele B. Quantitative and qualitative analyses of herbal medication use among Ghanaian stroke survivors. J Neurol Sci 2021;427:117540 [FREE Full text] [doi: 10.1016/j.jns.2021.117540] [Medline: 34139450]
25. Fernandes LG, Saragiotto BT. To what extent can telerehabilitation help patients in low- and middle-income countries? Braz J Phys Ther 2021;25(5):481-483 [FREE Full text] [doi: 10.1016/j.bjpt.2020.11.004] [Medline: 33262081]
26. Household survey on ICT in Ghana. National Communications Authority and Ghana Statistical Service. 2020. URL: https://tinyurl.com/mrxhktj7 [accessed 2023-05-25]
27. Özden F. The effect of mobile application-based rehabilitation in patients with Parkinson's disease: a systematic review and meta-analysis. Clin Neurol Neurosurg 2023;225:107579. [doi: 10.1016/j.clineuro.2022.107579] [Medline: 36603336]

Abbreviations

HCP: health care professional
LMIC: low- to middle-income country
PD: Parkinson disease

Edited by L Buis; submitted 31.05.23; peer-reviewed by E Nizeyimana, S Quaglini; comments to author 18.12.23; revised version received 29.01.24; accepted 20.02.24; published 27.03.24.

Please cite as:
Paul L, Thomson K, Asibey SO, Brady M, van Wijck F, Antwi D, Opoku EN, Sarfo FS
Views of Service Users, Their Family or Carers, and Health Care Professionals on Telerehabilitation for People With Neurological Conditions in Ghana: Qualitative Study
JMIR Mhealth Uhealth 2024;12:e49501
URL: https://mhealth.jmir.org/2024/1/e49501
doi:10.2196/49501
PMID:38536218

Review

Rehabilitation Applications Based on Behavioral Therapy for People With Knee Osteoarthritis: Systematic Review

Dian Zhu[1], PhD; Jianan Zhao[1], PhD; Mingxuan Wang[1], BEng; Bochen Cao[1], BEng; Wenhui Zhang[1], BEng; Yunlong Li[2], MArch; Chenqi Zhang[1], PhD; Ting Han[1,3], Prof Dr

[1]School of Design, Shanghai Jiao Tong University, Shanghai, China

[2]Department of Design, Jiangxi Science and Technology Normal University, Shanghai, China

[3]Institute of Medical Robotics, Shanghai Jiao Tong University, Shanghai, China

Corresponding Author:
Ting Han, Prof Dr
School of Design
Shanghai Jiao Tong University
No 800, Dong Chuan Rd
Shanghai, 200240
China
Phone: 86 18901626266
Email: hanting@sjtu.edu.cn

Abstract

Background: The development of digital applications based on behavioral therapies to support patients with knee osteoarthritis (KOA) has attracted increasing attention in the field of rehabilitation. This paper presents a systematic review of research on digital applications based on behavioral therapies for people with KOA.

Objective: This review aims to describe the characteristics of relevant digital applications, with a special focus on the current state of behavioral therapies, digital interaction technologies, and user participation in design. The secondary aim is to summarize intervention outcomes and user evaluations of digital applications.

Methods: A systematic literature search was conducted using the keywords “Knee Osteoarthritis,” “Behavior Therapy,” and “Digitization” in the following databases (from January 2013 to July 2023): Web of Science, Embase, Science Direct, Ovid, and PubMed. The Mixed Methods Assessment Tool (MMAT) was used to assess the quality of evidence. Two researchers independently screened and extracted the data.

Results: A total of 36 studies met the inclusion criteria and were further analyzed. Behavioral change techniques (BCTs) and cognitive behavioral therapy (CBT) were frequently combined when developing digital applications. The most prevalent areas were goals and planning (n=31) and repetition and substitution (n=27), which were frequently used to develop physical activity (PA) goals and adherence. The most prevalent combination strategy was app/website plus SMS text message/telephone/email (n=12), which has tremendous potential. This area of application design offers notable advantages, primarily manifesting in pain mitigation (n=24), reduction of physical dysfunction (n=21), and augmentation of PA levels (n=12). Additionally, when formulating design strategies, it is imperative to consider the perspectives of stakeholders, especially in response to the identified shortcomings in application design elucidated within the study.

Conclusions: The results demonstrate that “goals and planning” and “repetition and substitution” are frequently used to develop PA goals and PA behavior adherence. The most prevalent combination strategy was app/website plus SMS text message/telephone/email, which has tremendous potential. Moreover, incorporating several stakeholders in the design and development stages might enhance user experience, considering the distinct variations in their requirements. To improve the efficacy and availability of digital applications, we have several proposals. First, comprehensive care for patients should be ensured by integrating multiple behavioral therapies that encompass various aspects of the rehabilitation process, such as rehabilitation exercises and status monitoring. Second, therapists could benefit from more precise recommendations by incorporating additional intelligent algorithms to analyze patient data. Third, the implementation scope should be expanded from the home environment to a broader social community rehabilitation setting.

(JMIR Mhealth Uhealth 2024;12:e53798) doi:10.2196/53798

KEYWORDS

knee osteoarthritis; digital application; behavioral therapy; behavior change therapy; cognitive behavioral therapy

Introduction

Knee osteoarthritis (KOA) is a prevalent musculoskeletal disorder that ranks among the primary contributors to disability [1,2]. Possible long-term ramifications encompass diminished levels of physical activity (PA), the development of body dysmorphic disorder, compromised sleep patterns, depressive symptoms, and the onset of disability [3,4]. In recent times, there has been a notable shift in the approach to treating KOA, with a greater emphasis on nonpharmacologic interventions. This change is supported by evidence indicating that nonpharmacologic treatments are more effective in delivering sustained symptom alleviation and in delaying or even preventing functional deterioration [5,6]. The primary nonpharmacological interventions for KOA include educational programs, PA interventions, and weight management strategies [3]. Patient initiation and adherence to these treatments are essential factors for achieving effective symptom control [7]. Traditional nonpharmacological interventions, however, require professional guidance to achieve the desired results, which is associated with high costs and unequal health care resources [8].

Digital health interventions have the potential to offer widespread, cost-effective, readily available, and easily expandable patient education and self-management interventions for individuals with KOA [9-11]. Several research investigations have been carried out to substantiate their efficacy in rehabilitating musculoskeletal problems. For instance, digital health interventions have been found to be successful in decreasing pain, improving functionality, and promoting the self-management of musculoskeletal pain syndromes [12,13]. Significant increases in adherence have also been observed throughout the mid-term follow-up [14]. These systematic evaluations have focused on summarizing various techniques for digital health or intervention effectiveness in relation to health outcomes [15,16]. However, digital interventions do not always provide desirable outcomes. Providing guidance on the ideal dosage required to achieve significant benefits or disclosing the elements of effective digital health treatments is challenging owing to the variations in interventions and the insufficient information in interventions [15]. In recent years, it has been discovered that theory-driven interventions can help organize the content of digital interventions, resulting in improved health outcomes [17-19].

A growing number of studies have used the behavioral psychology theoretical framework in digital format [20]. Compared to generic digital interventions, behavioral therapy-based digital interventions are significantly more effective at relieving pain, improving physical dysfunction, and increasing self-efficacy in patients with KOA [21,22]. Physiotherapists use scalable interventions along with some digital tools to enhance treatment adherence [16,23]. The concept of behavioral therapy (BT) incorporates various therapeutic approaches, including behavioral change techniques (BCTs), dialectical behavioral therapy (DBT), and cognitive behavioral therapy (CBT). It has been used to aid complex intervention designs that include facilitating the adoption of behavior change, promoting behavioral compliance, sustaining behavioral change, and preventing behavioral relapse [24]. Previous studies have employed BCTs in combination with digital interventions among individuals with musculoskeletal pain [25]. These studies have reported the efficacy of such interventions in facilitating the transition of patients from a sedentary lifestyle to an active one [21,22,26]. Multiple studies have demonstrated that the integration of CBT with standard care yielded noteworthy outcomes in the management of KOA. Specifically, the implementation of CBT interventions resulted in a considerable reduction in pain levels and an improvement in insomnia symptoms when compared to the utilization of standard care alone, as indicated by previous investigations [27,28].

Currently, there are evaluations investigating the rehabilitative impacts of digitalization in KOA and highlighting the significance of behavioral theory in some applications [19]. Nevertheless, there is a shortage of thorough exposition of the behavioral theory in digital applications, as well as an absence of an assessment of the suitability of these applications from the patient's point of view. Thus, this review offers a methodical and thorough examination of digital applications rooted in behavioral therapy. It shifts the focus of digital applications from mere practical usability to providing support for behavioral change theories. Additionally, it meticulously analyzes the functional reasoning behind various products, thereby serving as a comprehensive guide for designing future digital interventions. Hence, the objectives of this review are to (1) provide a concise overview of the existing landscape of digital behavioral therapy applications for individuals diagnosed with KOA and examine the potential of digital applications in augmenting the rehabilitation process for KOA patients, and (2) present a comprehensive analysis of the underlying psychological theories, fundamental mechanisms, design methodologies, typical attributes, efficacy of treatment outcomes, and patient preferences pertaining to this particular mode of recovery intervention.

Methods

Registration

This review has been registered in the International Prospective Register of Systematic Reviews (PROSPERO; registration number: CRD42023430716). Furthermore, the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines have been applied (Multimedia Appendix 1) [29].

Search Strategy

Literature searches were conducted in 5 databases: Web of Science, Embase, Science Direct, Ovid, and PubMed. The selection of these databases was based on their provision of comprehensive access to full-text journals and conference

proceedings pertaining to prominent conferences and meetings focused on digital technology and medicine.

To locate relevant articles, we conducted a search by filtering papers based on 3 primary categories of MeSH (Medical Subject Headings) terms: "Knee Osteoarthritis," "Behavioral Therapies," and "Digitization." According to MeSH terminology, "Behavioral Therapy" is divided into "Behavioral Therapy," "Cognitive Behavioral Therapy," and "Dialectical Behavioral Therapy." In this review, in order to understand the categorization of all behavioral therapies, we collected information on the subcategories of these 3 categories related to behavior. In relation to the subject of "digital" content, we gathered relevant material from the report by Safari et al [19] on digital literature, encompassing topics, such as "Telehealth," "Email," "Smartphone," "Computer Systems," "Digital Technologies," and "Mobile Applications," and other forms of digitization. Table 1 illustrates sample search strategies used for the PubMed digital library. The article titles, keywords, and abstracts were searched. Similar search strategies were applied to the remaining 3 databases. Relevant articles published between January 2013 and July 2023 were gathered. We included journal papers and peer-reviewed conference proceedings. Only articles written in English were considered.

Table 1. Literature search strategy.

MeSH[a]	Boolean logic search strings
Knee Osteoarthritis	"Knee Osteoarthritides" OR "Knee Osteoarthritis" OR "Osteoarthritis of Knee" OR "Osteoarthritis of the Knee"
Behavior Therapy	"Behavior Therapies" OR "Behavior Treatment" OR "Conditioning Therapy" OR "Conditioning Therapies" OR "Behavior Change Techniques" OR "Behavior Change Technique" OR "Behavior Modification" OR "Behavior Modifications" OR "Dialectical Behavior Therapies" OR "Cognitive Behavioral Therapies" OR "Cognitive Therapy" OR "Cognitive Behavior Therapy" OR "Cognitive Psychotherapy" OR "Cognition Therapy" OR "Cognitive Behavior Therapies" OR "Cognitive Behavior Therapy"
Digitization	"Telemedicine" OR "Mobile Health" OR "Telehealth" OR "ehealth" OR "mhealth" OR "Email" OR "E-mail" OR "Mobile" OR "Smartphone" OR "smart-phone" OR "smart telephone" OR "Tablet" OR "cell" OR "hand-held" OR "Cell Phone" OR "handheld" OR "Remote Consultation" OR "Teleradiology" OR "Telenursing" OR "Computer Systems" OR "Computer-Assisted Instruction" OR "Internet" OR "web" OR "computer" OR "Digital Technologies" OR "APP" OR "Social Media" OR "Internet-Based Intervention" OR "Mobile Application" OR "Mobile App" OR "Smartphone App" OR "Portable Software Application"

[a]MeSH: Medical Subject Headings.

Eligibility Criteria

The authors DZ and JZ were assisted in the literature search by an experienced librarian well versed in medical database searching. This literature review was guided by the question of how behavioral therapies can be integrated with digital applications in the rehabilitation of patients with KOA. On this basis, we anticipated that this review would (1) generalize and summarize the digital applications used in behavioral therapy, and (2) describe the overall research status and research trends of these digital applications.

Inclusion Criteria

The inclusion criteria were as follows: (1) adult participants (age ≥18 years) with KOA diagnosed by self-reported symptoms or imaging; (2) patients had access to digital applications; (3) any form of intervention or treatment based on the inclusion of at least one behavioral treatment was delivered through any digital application (eg, website or app) within any time frame; and (4) the described interventions were compared to waiting list control (no intervention) or alternative (standard) delivery modalities (eg, face-to-face approaches, classroom-based approaches, and printed materials or handouts), nondigital self-management interventions, and noninteractive digital interventions (eg, web pages with flat copies).

Exclusion Criteria

The exclusion criteria were as follows: (1) patients with KOA were not included; (2) nondigital interventions were assessed; (3) behavioral therapies were not included; (4) research protocols, reviews, conceptual articles, case studies or discussion papers, and conference abstracts; (5) market research; (6) digitization was not designed for the recovery process; (7) text was not written in English; and (8) duplicate reports of the same study from different sources.

Data Extraction

Data relevant to the purpose of the study were extracted independently by the authors MW, WZ, and BC, and any misunderstandings and disagreements were resolved through negotiation. Extracted data included study context details, study population, and digital application details. In more detail, the template included the following categories: (1) basic information (author, year, origin, study population, sample size, presence of a physiotherapy intervention, and duration of the intervention); (2) digital application details (digitalization of behavioral therapy, interactive device function, study outcomes, and application deficiencies); and (3) type of study (randomized controlled trial, cohort experiment, experimental protocol, and qualitative study).

Quality Assessment

The Mixed Methods Assessment Tool (MMAT) was used to evaluate the methodological quality of the included studies [30]. This tool was initially created in 2006 through a comprehensive analysis of systematic evaluations that integrated qualitative and quantitative evidence. In 2018, a revised version of the MMAT was developed by assessing its usefulness, reviewing key assessment tools in the literature, and conducting a modified e-Delphi study involving methodology experts to determine the essential criteria (Multimedia Appendix 2). The MMAT

evaluates the caliber of research employing qualitative, quantitative, and mixed approaches. The primary emphasis is on methodological standards, which encompass 5 fundamental quality criteria for 5 distinct study designs: (1) qualitative, (2) randomized controlled, (3) nonrandomized, (4) quantitative descriptive, and (5) mixed methods.

Results

Overview

Figure 1 provides a summary of the outcomes at various phases of article selection. The search results of the databases are provided in Multimedia Appendix 3. Based on the search strategy, 2975 articles were found initially, and 2507 articles remained after title and abstract screening and removal of duplicates. Next, full-text articles were chosen based on the inclusion and exclusion criteria. We included a total of 131 articles explicitly related to digital behavioral therapy for KOA. Moreover, 4 articles were identified following a manual search of the references for articles that were cited. Final consideration was given to 36 articles for systematic evaluation (Table 2).

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

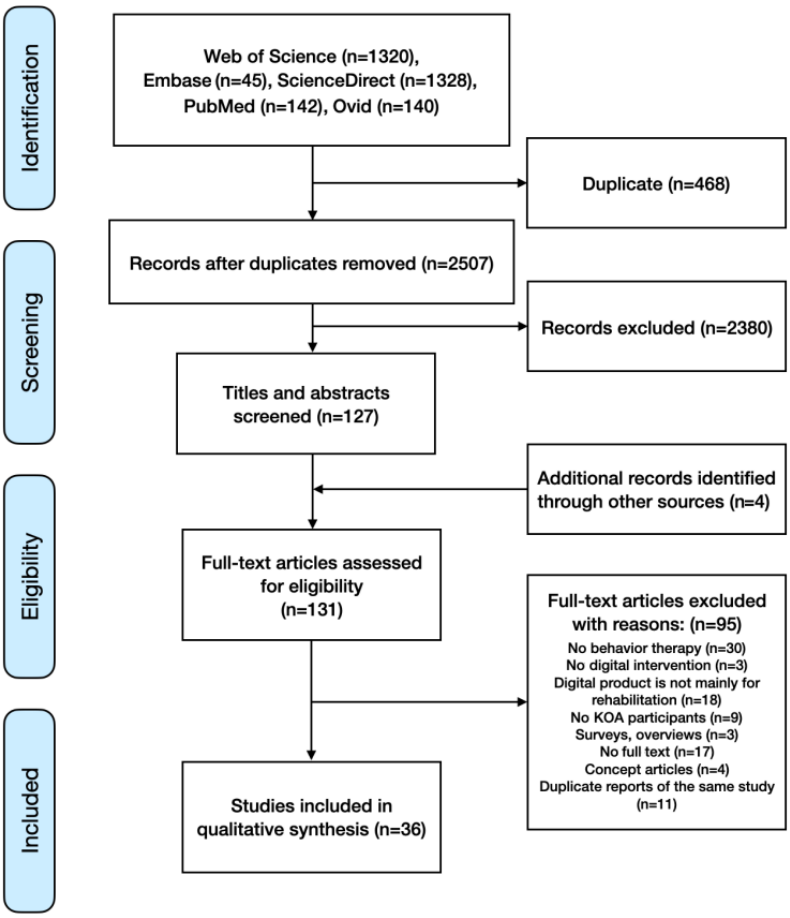


Table 2. Information on the included studies.

Author	Year	Country	Population	Sample size	Physiotherapist	Duration	Digital forms	Experiment	Quality
Bossen et al [31]	2013	Netherlands	KOA[a]/HOA[b]	20	Yes	6-12 weeks	Internet platform + SMS text message/telephone	Nonrandomized pilot study	4
Rini et al [32]	2016	United States	OA[c]	113	No	9-11 weeks	App + "virtual coaching"	Randomized controlled trial	5
Pearson et al [33]	2016	United Kingdom	KOA/HOA	200	No	—[d]	Internet website	Mixed methods research	2
Bennell et al [34]	2017	Australia	KOA	168	Yes	24 weeks	Internet platform + SMS text message/telephone	Randomized clinical trial	5
Bennell et al [11]	2017	Australia	KOA	148	Yes	24 weeks	Internet platform + SMS text message/telephone	Randomized clinical trial	5
Li et al [35]	2017	Canada	KOA	34	Yes	8 weeks	Electronic equipment + telephone	Randomized controlled trial	4
Lawford et al [36]	2018	Australia	KOA	148	Yes	12-36 weeks	PainCOACH software + email	Randomized controlled trial	5
Button et al [37]	2018	United Kingdom	KOA	49	Yes	12 weeks	Online course	Qualitative research	4
Mecklenburg et al [38]	2018	United States	KOA	162	Yes	12 weeks	Hinge Health app + wearable	Randomized controlled trial	5
Kline et al [39]	2019	United States	TKR[e]	100	Yes	—	Online course + wearable	Randomized controlled trial protocol	N/A[f]
Nelligan et al [40]	2019	Australia	KOA	12	No	24 weeks	SMS text message application + messaging interactions	Qualitative research	5
Pelle et al [41]	2019	Netherlands	KOA/HOA	427	No	12-24 weeks	Bart app	Randomized controlled trial	5
Bailey et al [42]	2020	United States	MD[g]	10,264	Yes	9 weeks	Hinge Health app + wearable motion sensors	Cohort study	4
Baker et al [43]	2020	United States	KOA	104	Yes	2 years	Teleconferencing (remote software)	Cohort study	4
Bennell et al [22]	2020	Australia	KOA/obesity	12	No	24 weeks	SMS text message application + messaging interactions	Randomized controlled trial	5
Fitzgibbon et al [44]	2020	United States	OA	203	Yes	8 weeks	F&S! and F&S! Plus + telephone	Comparative effectiveness test	4
Hinman et al [45]	2020	Australia	KOA	165	Yes	24 weeks	Internet website + SMS text message/telephone	Randomized controlled trial	5
Hinman et al [46]	2020	Australia	KOA	394	Yes	12 weeks	Internet website + video consulting	Randomized controlled trial protocol	N/A
Li et al [47]	2020	Canada	KOA	51	Yes	12 weeks	Electronic device + telephone/mail	Randomized controlled trial	4
Nelligan et al [48]	2020	Australia	KOA	16	Yes	—	Internet website + SMS text message/telephone	Qualitative research	2
Dunphy et al [49]	2021	United Kingdom	OA	59	Yes	12 weeks	Internet website	Two-arm parallel randomized controlled trial	5

Author	Year	Country	Population	Sample size	Physiotherapist	Duration	Digital forms	Experiment	Quality
Lindberg et al [50]	2021	Norway	OA	282	Yes	12 weeks	iCBT application + telephone	Randomized controlled trial protocol	N/A
Nelligan et al [21]	2021	Australia	KOA	206	Yes	24 weeks	Internet website + SMS text message/telephone	Randomized controlled trial	5
Pelle et al [51]	2021	Netherlands	KOA/HOA	214	No	26 weeks	Bart app	Randomized controlled trial	4
Rognsvåg et al [52]	2021	United Kingdom	KOA/TKR	4	Yes	—	iCBT application + telephone	Qualitative research	3
Bennell et al [53]	2022	Australia	KOA	88	Yes	24 weeks	Website + remote software	Randomized controlled trial protocol	N/A
Groves-Williams et al [54]	2022	Scotland	KOA	90	No	12-36 weeks	Internet website + SMS text message/telephone	Randomized controlled trial protocol	N/A
Hinman et al [55]	2022	Australia	KOA	182	Yes	14 weeks	App + SMS text message	Randomized controlled trial protocol	N/A
Östlind et al [56]	2022	Sweden	KOA/HOA	20	Yes	12 weeks	App + electronic device	Qualitative research	4
Östlind et al [57]	2022	Sweden	KOA/HOA	160	Yes	12 weeks	App+ electronic device	Randomized controlled trial	3
Whittaker et al [58]	2022	Canada	OA	30	Yes	4 weeks	Video conferencing + wearable + app	Randomized trial	4
Godziuk et al [59]	2023	Canada	KOA	53	Yes	12 weeks	Website + email	Cohort study	3
Lorbeer et al [60]	2023	Germany	KOA	241	Yes	1 year	Teleconferencing (remote software)	Randomized controlled trial	4
Scheer et al [61]	2023	United States	MD	4051	Yes	12 weeks	App + electronic device	Cohort study	4
Truong et al [62]	2023	Canada	OA	16	No	—	Video conferencing + wearable + app	Qualitative research	5
Weber et al [63]	2023	Germany	KOA/HOA	330	Yes	3 weeks	e-Exercise + online physiotherapy	Randomized controlled trial protocol	N/A

[a]KOA: knee osteoarthritis.

[b]HOA: hip osteoarthritis.

[c]OA: osteoarthritis.

[d]Data not available.

[e]TKR: total knee replacement.

[f]N/A: not applicable.

[g]MD: musculoskeletal disorder.

Methodological Quality

Among the included studies, 12 met 100% of the quality assessment criteria, 15 fulfilled 60%-80% of the quality assessment criteria, and 2 met 40% of the quality assessment criteria (Multimedia Appendix 4 [11,21,22,31-63]). The remaining 7 studies could not be evaluated for their quality owing to the absence of results. Nevertheless, the application description portion involved in the studies was highly valuable for analysis.

Digitalization of Behavioral Therapy

Behavior Change Therapy

The majority of digital applications for KOA rehabilitation are based on BCTs. BCTs (achieving objectives, setting goals, restructuring beliefs, and inducing acceptance) are applicable to addressing the central issues of initiating and maintaining PA [64]. The primary categories of BCTs used in the reviewed studies were based on the V1 Taxonomy of Behavior Change by Michie et al [65], which was devised by behavior change researchers [66]. The taxonomy comprises 93 distinct BCTs

organized into 16 hierarchical structures and has been extensively used in the literature on behavior change (Table 3): (1) goals and planning; (2) feedback and monitoring; (3) social support; (4) shaping knowledge; (5) natural consequences; (6) behavioral comparisons; (7) associations; (8) repetition and substitution; (9) outcome comparisons; (10) rewards and threats; (11) regulation; (12) presuppositions; (13) identity; (14) intended consequences; (15) self-confidence; and (16) implicit learning.

Table 3. Behavioral change techniques in digital applications.

Cluster label and component behavioral change techniques	References
1: Goals and planning	
1.1: Goal setting (behavior)	[11,31-37,39-47,49-52,54,58,59,61,62]
1.2: Problem solving/coping planning	[35,39,44,54,63]
1.4: Action planning	[11,21,31-37,43,46,48,52,53,60,63]
1.5: Review of behavior goal(s)	[11,31,33,34,39,43,52]
1.7: Review of outcome goal(s)	[11,34,58,62]
2: Feedback and monitoring	
2.2: Feedback on behavior	[33,42,56,57]
2.3: Self-monitoring of behavior	[31,32,36,39,42,44,47,55,57-59,62]
2.4: Self-monitoring of the outcome of behavior	[33,47,58,61]
2.6: Biofeedback	[11,33-35,37,39,42,47,56-58,61,62]
2.7: Feedback on behavioral outcomes	[32,36]
3: Social support	
3.2: Social support (practical)	[33,59]
3.3: Social support (emotional)	[43,60]
4: Shaping knowledge	
4.1: Instructions on how to perform a behavior	[21,32,42-44,46-50,52-54,56,59,61]
4.2: Antecedents	[32,36,38]
5: Natural consequences	
5.1: Health consequences	[32,36,39,41,42,50,51]
5.4: Self-assessment of affective consequences	[33,43]
5.5: Anticipated regret	[35,39,43,63]
6: Comparison of behavior	
6.1: Modeling of behavior	[42,56]
6.2: Social comparison	[33,53,60]
6.3: Information about others' approval	[32,36,42-44,49,53,58,59,61-63]
7: Associations	
7.1: Prompts/cues	[11,21,22,31,34,37,40,43,45,48,50,54,55,59,60]
8: Repetition and substitution	
8.1: Behavioral rehearsal/practice	[11,21,34,35,37,39-43,46,48,49,51,56,58,59,61-63]
8.6: Generalization of a target behavior	[33,38,42]
8.7: Graded tasks	[31,33,45,49,63]
10: Reward and threat	
10.3: Nonspecific reward	[32,36]
11: Regulation	
11.2: Regulate negative emotions	[33,38,42,43,50,52,59]
12: Antecedents	
12.4: Distraction	[32]
13: Identity	
13.1: Identification of self as a role model	[52]
15: Self-belief	
15.1: Verbal persuasion to boost self-efficacy	[32,36,39]

Cluster label and component behavioral change techniques	References
16: Covert learning	
16.2: Covert conditioning	N/A[a]
16.3: Vicarious reinforcement	[43]

[a]N/A: not applicable.

In 31 studies, objectives and planning were mentioned, including goal setting (behavior), problem solving or coping planning, and reviewing behavioral or outcome goals. Among these factors, goal setting and action planning were shown to be the most prominent components within the area. A total of 26 applications included goal setting, which has a very broad definition in the taxonomy (setting goals defined according to the behavior or outcome to be accomplished) [11,31-37,39-47,49-52,54,58,59,61,62]. Applications created evidence-based, individualized, progressive home exercise plans; promoted increased general PA; and established short-term objectives. Moreover, 16 applications [11,21,31-37,43,46,48,52,53,60,66] contained action planning in which patients were asked or chose to perform activities until their pain tolerance was attained, based on which the patients prescribed their own individual therapeutic actions. Additionally, 5 applications [35,39,44,54,63] addressed problem solving and coping strategies encountered during rehabilitation by other individuals or physiotherapists. Furthermore, 9 applications contained a review of behavioral or outcome objectives [11,31,33,34,39,43,52,58,62], encouraging participants to monitor their progress and assisting them in identifying personal barriers and strategies for overcoming them.

A total of 19 investigations included various forms of feedback and monitoring, such as feedback on behavior, self-monitoring of behavior, biofeedback, self-monitoring of behavioral outcomes, and feedback on behavioral outcomes. Feedback was provided on behavior wherein activities or exercises were recorded on performance metrics through a digital application and discussed by the physiotherapist during follow-up [33,42,56,57]. Among the included studies, 12 involved self-monitoring for managing exercise reminders and records, viewing progress charts, and setting or modifying exercise objectives [31,32,36,39,42,44,47,55,57-59,62]. Moreover, 13 studies offered participants a wearable device with additional features, such as the ability to monitor activity intensity and visualize activity performance over time [11,33-35,37,39,42,47,56-58,61,62]. These features enabled individuals to monitor progress and receive real-time feedback on objective achievement.

A total of 18 studies applied shaping knowledge. They primarily incorporated videos or lectures on osteoarthritis (OA), the effects of PA, self-management, and coping strategies [21,32,42-44,46-50,52-54,56,59,61]. Three of these studies presented information about antecedents via multiple online physical therapy consultations using video phone services [32,36,38].

A limited subset of digital applications employed social support. The programs offered online platforms where individuals could engage in discussions pertaining to joint pain. Four studies documented the beneficial effects of engaging with social organizational structures on the rehabilitation of individuals with KOA [33,43,59,60].

The concept of natural consequences was addressed in 11 investigations [32,33,35,36,39,41-43,50,51,63], and it included information regarding health consequences, monitoring of emotional consequences, and anticipated misgivings. At each online meeting, the interventionist provided the patient with information about the benefits and costs of engaging in or refraining from a particular course of action. In addition, reminders regarding obstacles and facilitators were provided beforehand.

Comparison of behavior was addressed in 16 studies [32,33,36,42-44,49,53,56,58-63]. It was primarily implemented with behavior evidence, social comparisons, and information about the approbation of others. Three studies on group therapy prompted patients to establish a “buddy” system to change their behavior [33,56,59]. In some cases, a physiotherapist was included to provide the patient with assistance or instrumental social support. Twelve studies referred to information about other people’s perceptions of a person’s behavior and whether others would approve or disapprove of any proposed behavioral change to encourage people to decide to set overall goals [32,36,42-44,49,53,58,59,61-63]. For instance, making behavioral decisions was practiced more the following week, along with identifying obstacles to executing the behavior and devising strategies to overcome them.

A total of 16 studies referred to associations, particularly prompts, as reminders for the patient to perform a particular behavior [11,21,22,31,34,37,40,43,45,48,50,54,55,59,60]. The defined frequency, intensity, or duration of the specified behavior, along with a description of at least one context, location, time, and manner, was included. Of those, 12 involved primary distribution by the physiotherapist via short messages or email timed reminders. In addition, 8 studies involved prompts by the application’s included features [21,22,31,40,48,54,55,59].

Repetition and substitution, which involve behavioral practice or rehearsal, generalization of target behaviors, and grading tasks, were the most common components of the applications. In 27 studies, patients were required to rehearse and repeat KOA exercises [11,21,34,35,37,39-43,46,48,49,51,56,58,59,61-63]. Additionally, 1 study described neuromuscular exercises designed to enhance the physical function of the lower extremities, and the targeted behaviors were broken down into daily video bundles sent to patients [59]. Five studies divided the exercises into varying intensities and progressively increased the difficulty until the desired behaviors were achieved [31,33,45,49,63]. Individual progress and the patient’s

perception of the capacity to exercise without aggravating discomfort were taken into account.

Rewards and threats were mentioned in 2 studies [32,36]. Mobile health apps were supplemented with motivation-enhancing techniques, such as praise, encouragement, and material rewards, for the achievement of specific goals.

The primary goal of regulation was to reduce negative emotions in patients. Seven studies trained users to recognize negative thoughts and reactions to them by relaxing mood through thoughts, emotions, and behaviors that affect pain [33,38,42,43,50,52,59]. One study adhered to the practice by revisiting pleasant imagery and distractions from the previous week [32]. In addition, 1 study discussed the potential for novel or alternative pain medications in applied implicit learning [43].

CBT

Complementary and alternative medicine therapy refers to a deliberate, intentional, and organized form of psychotherapy intervention aimed at improving psychological issues by impacting the beliefs and behaviors of patients [67,68]. CBT combines techniques to develop more adaptive cognitions and behaviors, such as psychoeducation, cognitive restructuring, relaxation therapy, and guided imagery (eg, to reduce muscle tension and autonomic arousal), as well as positive thinking training, problem-solving, and stress management [69,70].

Specifically, CBT focuses on reducing pain and distress by altering bodily sensations, catastrophic and contemplative thinking, and maladaptive behaviors, as well as enhancing self-efficacy [71,72]. Four studies addressed common CBT topics, such as catastrophizing, positive coping methods, and anxiety avoidance, through educational interactive modules and internet courses pertaining to behavior change [42,50,52,61]. The remaining 4 studies addressed common barriers to exercise (eg, pain, low confidence, weather, and relapse) and ways to overcome them (eg, increasing confidence through exercise, seeking social support, teaching proper exercise routines and postures, and promoting positive reasoning) in conjunction with programmatic elements of social cognitive theory and goal-setting strategies for exercise behaviors [32,36,43,44].

Interactive Device Function

Digital Presentation Modalities

The emergence of the internet in the health field has drastically altered the medical information available to patients and the manner in which physicians and patients communicate [73]. A significant number of digital applications involving KOA utilize information and communication technology (ICT) to facilitate behavioral therapy. The included studies covered 5 types of digitization (Table 4): (1) app, (2) website, (3) teleconferencing software/remote phone contact, (4) wearable electronic device, and (5) SMS text message/telephone/email.

Table 4. Forms of digital applications.

Digital application type	References
App/website + SMS text message/telephone/email	[11,21,31,34,36,44,48,50,52-54,59]
App/website	[32,33,37,41,46,49,51,63]
Teleconferencing software/remote messaging	[22,40,43,45,55,60]
Wearable electronic device + SMS text message/telephone/email + app/website	[35,38,39,42,47,56,57,61]
Teleconferencing software + wearable electronic device + app/website	[58,62]

Twelve studies adopted the combination of app/website plus SMS text message/telephone/email. Physiotherapists in 2 studies provided verbal and written education or information about OA, benefits of PA or exercise, and strategies to increase adherence [11,44]. In addition, a progressive individualized home exercise program based on scientific evidence was devised, which included several lower extremity exercises and was accessible through an app or website. In 2 studies, weekly emails containing OA-specific content and resources were sent directly to patients. The emails included (1) nutritional advice; (2) an instructional video on exercise; and (3) a video on positive thinking and advice on self-care, motivation, and stress management [36,59]. Nine studies supported general health and wellness behavior change through free website support [11,21,31,34,36,48,53,54,59]. To increase patient compliance, physiotherapists conducted regular telephone counseling sessions to determine if the use of optional sessions should be based on participant preference, confidence, and success in achieving the desired behavior change. In 3 other investigations, the aforementioned functions were implemented in their entirety within a single application [44,50,52].

Eight studies used apps or websites for self-management and coping with arthritis pain through exercise [32,33,37,41,46,49,51,63]. The websites included information on PA or exercise, goal setting, action plans, pacing, medication management, diet, home exercise, understanding pain, pain management, and relaxation modules. One study explained how individuals can input data and view graphical feedback regarding the amount of exercise they have performed, their activity levels, and their mood [33].

Five investigations [22,40,43,45,60] applied teleconferencing software or telemessaging, and 3 programs provided patients with complimentary access to online webinars featuring “expert advice” [43,45,60]. Participants could ask the facilitator queries about nutrition, exercise, or positive thinking. Registered dietitians, registered psychologists, and kinesiologists led the sessions in pairs according to a rotating schedule. Two other studies addressed recommendations for the development of health behavior interventions utilizing only SMS text messaging on mobile phones [22,40].

Eight studies used the combination of wearable electronic devices plus SMS text message/telephone/email plus

app/website to guide participants in setting specific, measurable, achievable, pertinent, and time-bound PA goals [35,38,39,42,47,56,57,61]. Self-monitoring is typically assessed using commercially available wrist-worn wearable activity trackers (such as Fitbit) or other similar devices. These devices collect measures and communicate them through Bluetooth to a smartphone, tablet, or computer application. Subsequently, the application transmits the data to the Fitbit server. Individuals who possess apprehensions over engaging in PA have the option to communicate their concerns via electronic mail to their physiotherapists.

Two studies [58,62] employed teleconferencing plus electronic devices plus apps. The program comprised 3 elements: (1) a 1-time knee boot camp where participants worked at home on their exercise therapy and PA goals; (2) weekly personalized in-home exercise therapy, PA, and tracking where participants received a Fitbit Inspire activity tracker; and (3) weekly physiotherapist-guided exercise therapy and activity action plans via videoconference on Zoom, optional group exercise classes, and exercise therapy and PA goal setting. In the TeleHab app, exercise therapy objective completion levels, target rating of perceived effort, and any associated pain were recorded, and Fitbit data were synchronized with the Fitbit online dashboard.

Design Methodology

Although researchers and developers typically validate their digital applications with end users, it is uncommon for relevant studies to include end-user participation in the design phase. Of the 36 studies in our review, 8 (22%) reported end-user participation in the design phase [31,33,37,40,41,48,50,56], and 6 (75%) of these 8 studies also reported the participation of stakeholders other than end users [33,37,40,41,50,56]. These stakeholders also included caregivers, coaches, physiotherapists, and other individuals who provide care or services to the target population.

Heuristic assessments were used in a study to explore usability of the intervention for patients with KOA. Based on the outcomes of the interviews and heuristic evaluations, the program's time structure was modified to be more flexible. In the most recent iteration, users had the option of repeating modules and adjusting module difficulty. The strategy also addressed improper website design and placement of multiple icons [31]. In addition, a study involved 3 patients with KOA who provided feedback on the prototype to inform the final design. Regarding how participants perceived interventions used outside the study setting, the majority suggested that health professionals, particularly general practitioners or physiotherapists, could deliver interventions to enhance or improve care [48].

Additional stakeholders were included in the qualitative study analysis. One study referred to experimental websites where patients with KOA and physiotherapists provided feedback on prototypes to inform the final design [40,41,50]. In a separate study, the app was also developed in collaboration with physiotherapists, physicians, and patient representatives. Named members of the project team submitted a list of 30 SMART (Specific, Measurable, Achievable, Relevant, and Time-bound) objectives related to OA treatment [41].

Study Outcomes

A total of 23 experimental studies, 7 experimental protocols, and 6 qualitative studies were included in this literature review. The primary outcomes of using digital behavioral therapy for rehabilitation of patients with KOA were (1) pain, (2) symptoms, (3) physical functioning, and (4) PA. Moreover, the secondary outcomes included (1) self-efficacy, (2) usability and user needs, (3) health-related quality of life, (4) satisfaction, (5) negative emotions, (6) quality of sleep, (7) adherence, (8) surgical intent, and (9) understanding of the condition (Table 5).

Table 5. Rehabilitation outcomes for patients with knee osteoarthritis.

Outcome of the intervention	References
Pain	[11,21,22,31,32,34,36,38,41,42,45,46,49-56,59-61,63]
Symptoms	[31,35,36,41,44,47,49-52,60]
Physical function	[11,21,22,31,32,34,36,38,39,41,44-46,49,50,52-55,59,63]
Physical activity	[21,22,31,34,35,39,41,44,46,47,49-51,53-56,58,60,61]
Self-efficacy	[11,21,22,32,36,46,52,55,58-60]
Usability and user requirements	[33,37,40,42,48,49,54,56,58,62]
Satisfaction	[21,42,45,46,55,59,61,63]
Health-related quality of life	[11,21,22,34,36,46,48-51,53-55,59,63]
Negative emotions	[22,32,42,44,52,53,61]
Quality of sleep	[42,61]
Adherence	[22,36,42,43,49,55,58]
Surgical intention	[38,55,59]
Understanding of the condition	[38,51]

A total of 24 studies that aimed to reduce pain in patients with KOA were ultimately identified in this review [11,21,22,31,32,34,36,38,41,42,45,46,49-56,59-61,63]. Moreover, 12 studies reported statistically significant decreases

in pain [11,21,31,32,36,38,41,42,49,59-61]. Of the 12 studies that did not demonstrate improvement, 6 involved randomized controlled trial protocols and 2 involved qualitative studies. In the studies that showed improvement, intervention durations ranged from 9 weeks to 9 months, and in those that did not show positive results, intervention durations ranged from 8 weeks to 6 months.

Overall, 21 studies assessed physical dysfunction [11,21,22,31,32,34,36,38,39,41,44-46,49,50,52-55,59,63], and of these, 6 reported statistically significant improvements [21,34,41,44,45,59]. Among the 15 studies that did not demonstrate improvement, 7 were randomized controlled trial protocols and 2 involved qualitative research. In studies that demonstrated improvement, intervention durations ranged from 3 weeks to 9 months, while in studies that demonstrated no improvement, intervention durations ranged from 4 weeks to 4 months.

A total of 12 studies measured PA outcomes [21,22,31,34,35,39,41,44,46,47,49-51,53-56,58], and 5 of them reported statistically significant improvements [34,41,47,49,60]. Among the 15 studies that did not demonstrate improvement, 6 were randomized controlled trial protocols and 2 involved qualitative research. In studies that demonstrated improvement, intervention durations ranged from 3 weeks to 9 months, while in studies that demonstrated no improvement, intervention durations ranged from 4 weeks to 12 months.

Physical symptoms were assessed in 11 studies [31,35,36,41,44,47,49-52,60], and of these, 4 studies reported statistically significant improvements [31,35,44,49]. In studies that demonstrated improvement, the intervention duration was 2 months, while in studies that demonstrated no improvement, intervention durations ranged from 2 to 9 months.

Self-efficacy was examined in 11 studies [11,21,22,32,36,46,52,55,58-60], and of these, 3 studies reported statistically significant improvements [11,36,58]. A total of 15 studies reported health-related quality of life [11,21,22,34,36,46,48-51,53-55,59,63], and of these, 3 studies reported statistically significant improvements [11,49,59]. Satisfaction was assessed in 8 studies [21,42,45,46,55,59,61,63], and of these, 1 study reported significant improvements [61]. Seven studies reported improvements in self-reported negative affect after the intervention [22,32,42,44,52,53,61]. Adherence was evaluated in 7 studies [22,36,42,43,49,55,58], and of these, 2 studies reported statistically significant improvements [22,58]. Sleep quality was assessed in 2 studies [42,61], and of these, 1 study [61] reported significant improvements. Three studies examined surgical intent [38,55,59]; however, none of the studies reported significant changes in patients' surgical intent. Patients' understanding of their condition was examined in 2 studies [38,51]; however, no improvement was identified.

Application Deficiencies

The design recommendations of users and stakeholders for the app were reported at different stages of the final app study design. Two studies identified the negative emotions associated with the app [31,56]. For instance, Östlind et al [56] found that a wearable activity tracker facilitated PA in various ways and increased the awareness of the optimal number of steps to treat OA symptoms. However, not all participants found the wearable activity tracker to be motivating, and in some cases, if they missed a weekly PA, the app's prompts about PA caused them to feel anxious and frustrated [31].

Two studies examined the efficacy of applying various characteristics. For instance, the study by Lindberg et al [50] demonstrated the combined efficacy of education, exercise therapy, and internet-based CBT, but was unable to distinguish between these interventions individually. Moreover, Dunphy et al [49] suggested that some participants viewed the physiotherapist's participation as positive, customizing the digital program and monitoring their progress. Others described it as restrictive, particularly if the physiotherapist did not understand how the digital program operated.

In terms of how to promote PA, the studies suggested various potential solutions. For instance, setting and achieving a daily step objective would motivate individuals to walk more than usual [56]. In addition, there was a strong desire for support from health care professionals who could monitor and guide progress; reinforce health messages; and offer reassurance, motivation, and encouragement. In the interviews, peer support through online communities (eg, forums and blogs) was also mentioned as a positive feature where people could share their experiences and learn from and support others experiencing joint pain, although users felt that it needed to be supervised to prevent inaccurate and inappropriate posts [37].

Discussion

Overview

This review presents an analysis of 36 papers that examined the use of digital behavioral therapies for patients with KOA. The analysis provides insights into 3 key areas: (1) the effectiveness of digital and behavioral therapies, (2) the role of digital technologies in these treatments, and (3) the importance of involving users and stakeholders throughout the design phase.

Digitalization and Behavioral Therapy

Several digital applications designed to provide behavioral therapy–based approaches for KOA were reviewed. Various theories of behavior modification and cognition were incorporated into these applications to increase their applicability and efficacy. In this review, we discuss behavioral therapy elements that are frequently combined to form applications.

Physiotherapists or applications, goal setting based on the user's current physical status, and BCTs derived from a control theory framework can enhance long-term sustained exercise in patients with KOA [74]. Positive reinforcement of progressive PA is the most essential element of goals and planning [75]. Gradual increases in PA alter the perception that PA is associated with pain and increase confidence in improving PA performance, resulting in favorable physical (eg, physical ability, muscle strength, and joint flexibility) and psychological (eg, self-esteem, pain perception, and anxiety) changes. Repetition and substitution are intended to boost application engagement and health-related behaviors. A "severe" to "extremely severe" program and specific dosages are designed to stimulate strength

gains, resulting in enhanced function. Physiotherapists assisted participants in devising a PA program intended to increase PA, with exercises targeting the hips, knees, and ankles, including sit-to-stand exercises and seated knee extensions. Our findings indicate that the aforementioned 2 BCTs were covered in virtually all of the applications we reviewed.

It has been demonstrated that reminders increase adherence to unsupervised home strengthening exercises [76]. This feature provides personalized behavior change messages to help patients with KOA surmount barriers to exercise participation. The physiotherapist or the application sends reminder messages based on completion to assist and remind the user to reach their exercise objectives. In the mobile app or wearable device, graphical displays were used to monitor workout adherence and provide feedback. These approaches provided the ability to register the completion of weekly workout sessions and send regular messages to encourage weekly workout participation. Despite the prevalence of wearable devices, their effectiveness in enhancing PA has been questioned. These devices frequently incorporate motivational techniques, such as self-monitoring and real-time feedback, but rarely address skills such as action planning and problem solving, which are essential for altering PA behaviors. Some studies also examined common CBT topics, such as catastrophizing, positive coping strategies, and fear avoidance [42]. Specifically, supervised exercise therapy through patient education enables users to access information regarding OA treatment modalities as well as topics such as the advantages of a healthy lifestyle, PA, vitality, and nutrition.

Our analysis revealed that OA digital management applications may be an alternative to traditional therapy and may further assist with the implementation of OA standards in the wider community [77,78]. However, the engagement of a physiotherapist is a vital aspect. Most participants had favorable experiences with their assigned physiotherapist and were motivated by the daily contact and the support and encouragement provided [79].

Insights From Technology

The use of digital technologies in communication by the patient care team can contribute to the enhancement of information flow, facilitation of patient information retention, improvement of information accessibility and portability, customization of information based on individual needs, and provision of tools for patients to actively participate in their health care [75,80]. In our analysis, the process of digitization manifested through a convergence of several technological mediums, including apps and websites, text messaging, phone calls, emails, and wearable gadgets.

New patient self-management programs have emerged in recent years, demonstrating the importance and efficacy of eHealth interventions such as websites and mobile apps. Current topics include PA, exercise, goal setting, action plans, pacing, medication management, nutrition, at-home exercise, pain comprehension, pain management, and relaxation. They may be combined with multimedia technologies to facilitate the sharing of content and with human-computer interface technologies to enhance accessibility.

Most health and medical apps fail to retain users beyond 90 days. This is due to their missing potential for facilitating disease management and provider-patient communication. The development of teleconferencing software or telematic messaging technology is supported by reminders and distinct objectives [81]. The development of text messaging program functionality and message libraries (including message types, message frequency, and program interaction levels) has the potential to increase adherence to unsupervised home exercise.

Wearable electronic devices aid in tracking the user's daily activities and offer continuous visualization support. However, such devices typically carry a greater risk of privacy invasion and social stigmatization. Our findings indicate that none of the studies specifically addressed privacy concerns. Therefore, designers of pertinent systems should be encouraged to consider these factors more thoroughly. In addition, current wearable electronic devices collect limited data, and their primary function is still to provide feedback on physiological signals (steps, consumption, asymmetry, etc), lacking the accumulation of electrical signals unique to KOA. In the future, machine learning can be introduced into ubiquitous devices to predict patient activity in order to improve patient care.

In the reviewed studies, the combination of app/website plus SMS text message/telephone/email was the most common. It has been shown that text messaging programs combined with unsupervised web-based exercise can reduce pain and dysfunction in patients with KOA [22,40]. Although the addition of wearable electronic devices would improve the intervention process by providing more accurate monitoring data, the experience would not be enhanced. However, the cost of ubiquitous electronic devices and the complexity of their operation continue to prevent their widespread adoption.

Insights From the User or Stakeholder Experience

Less than half of the reviewed studies reported structured user or stakeholder participation in the design phase of their systems, according to our analysis. Nonetheless, there is a distinct trend indicating that an increasing number of studies are emphasizing the significance of involving users in the design and development phases (and not just the validation or deployment phases). Specifically, we summarize a set of emerging user or stakeholder involvement design trends in emerging research on the application of digital behavioral therapy for KOA. Based on the interviews conducted for the qualitative study, issues were identified with the digital app user experience of KOA patients [52,56,62].

First, direct patient feedback was not taken into account when developing the content of the apps, and the content was based solely on the knowledge and experience of hospital staff. In future studies, patients should play a more important role in the development of the content of patient-specific apps, evaluating and further optimizing the preferable mode of information for outcomes [26]. Moreover, there is a need for a structured collaborative design involving patients, physicians, and researchers in order to establish multiple collaborative processes based on shared concepts, mutual learning, and respect for diversity and divergent opinions [82].

Second, digital support is regarded as an integral component of OA care. An ideal approach would be one that combines traditional OA care with digital OA care to offer solutions for personalized, comprehensive, straightforward, dependable, and continuous hybrid care [56,81]. The experience of digital applications is divided into 4 subcategories [79]: (1) simplicity of implementation, (2) flexibility in choosing time and location, (3) significance of interaction with health care professionals, and (4) additional motivating factors. Consequently, our research contends that digital applications must incorporate and differentiate between various user experience stages.

The functional requirements of apps that patients, physicians, and researchers deemed most essential, convenient, desirable, and actionable differed significantly. Participants in the studies agreed, despite their differences, that minimum viable products should be electronic, should monitor patients' symptoms and activities, and should include features tailored to factors identified by patients and physicians as well as self-management strategies based on international guidelines. Over the course of the study, participants came to a consensus regarding the order of their functional requirements. Visual symptom mapping, goal setting, exercise programs, daily monitoring, and self-management strategies had the highest priorities.

Future Work

After conducting a thorough study, a set of recommendations were summarized to improve the quality of applications. These recommendations are focused on making the process more efficient and improving future design and development endeavors. Suggestions for future studies are presented below.

Digital Applications that Provide Comprehensive Monitoring of the Entire Treatment Process

To enhance the rehabilitation of patients, it is necessary to thoroughly investigate the efficacy and impact of different behavioral therapies. This will enable the development of more precise application strategies. Multiple behavioral therapies should be integrated to cover all aspects of the rehabilitation process, including rehabilitation exercises, status monitoring, and self-screening, in order to ensure that patients receive comprehensive care. Simultaneously, by promptly understanding the wants of patients, we can bolster their trust in the process of recuperation. This customized service model can enhance patient happiness and confidence, while also facilitating the seamless advancement of the rehabilitation process. Further investigation can enhance our comprehension by examining particular facets of digital health care platforms, such as network effects and the strategic management of platform ecosystem innovation [83].

Data Analysis Requirements of Patients Based on Intelligent Algorithms

Artificial intelligence algorithms are crucial for extracting important insights from vast quantities of data and offering therapists precise recommendations. As an illustration, the algorithm may examine a patient's exercise data while they are undergoing rehabilitation and detect tiny modifications. This helps the therapist in fine-tuning the intensity and substance of the training. Furthermore, through the comparison of data from many patients, the algorithm can investigate the efficacy of different treatments for specific patient groups, thereby establishing a strategic foundation for later treatment decisions. Significantly, these data serve the dual purposes of enhancing the precision and pertinence of treatment and evaluating the impact of different behavioral interventions. In conventional rehabilitation programs, this procedure is frequently based on subjective criteria. By employing a data-driven methodology, we can impartially evaluate the true efficacy of each technique, establishing a strong basis for further research and applications.

Community-Based Rehabilitation Scenarios With Digital Technology Integration

In order to enhance the adoption of digital behavioral therapy in rehabilitation, our objective is to extend its implementation from the domestic setting to the broader social community rehabilitation setting. The community rehabilitation setting offers a platform for patients to engage and assist one another, and fosters patients' motivation and assurance in their recovery process. For instance, an internet-based community can be established to facilitate patients in exchanging their recovery experiences, offering reciprocal motivation and assistance, and deliberating the obstacles and remedies related to the recovery journey. Additionally, offline activities, including rehabilitation lectures, group seminars, and interactive games, can be arranged to augment patients' social interaction and foster a sense of team camaraderie. Strong partnerships can be fostered with medical organizations and rehabilitation professionals in the community to offer comprehensive assistance for patient recovery. This not only enhances patient rehabilitation outcomes and quality of life, but also fosters the integration and optimization of community rehabilitation resources and promotes overall progress in the field of patient recovery.

Limitations

This research has some limitations. Initially, our evaluation thoroughly examined the present state of digital behavioral therapy applications using qualitative analysis. However, it is indisputable that quantitative analysis yields more robust clinical value. Hence, a complete quantitative meta-analysis will be conducted to assess the efficacy of digital behavioral therapy applications. Future research will integrate both qualitative and quantitative analyses to provide a more thorough evaluation. In addition, our search was limited to English-language research publications, and it is possible that there are significant findings in other languages. The search for research papers was restricted to 5 databases (Web of Science, ScienceDirect, PubMed, Ovid, and Embase). As algorithms are added to digital intervention tools, additional computer science databases, such as IEEE and ACM, could be added to increase the comprehensiveness of the assessment. Furthermore, this review was limited by the search criteria employed and the time period during which the papers were published. However, a focus on the last 10 years went a long way in ensuring that this systematic review includes the most recent research.

Conclusion

This review provides an overview of digital behavioral therapy applications for patients with KOA. In this systematic review,

36 studies were examined. The results demonstrate 14 BCTs and show that behavioral cognitive therapies are frequently combined when developing digital applications. The most prevalent areas were “goals and planning” and “repetition and substitution,” which were frequently used to develop PA goals and PA behavior adherence. The most prevalent combination strategy was app/website plus SMS text message/telephone/email, which has tremendous potential. Consequently, this research provided results for digital applications in terms of pain relief, physical function improvement, self-confidence, and improvement in health-related quality of life among patients with KOA. Nevertheless, there was a shortage of evidence indicating enhanced surgical intention, compliance, and disease knowledge. Subsequent quantitative analysis of this phenomenon is required in the future. Moreover, the incorporation of several stakeholders in the design and development stages might enhance the user experience, considering the distinct variations in their requirements. Based on the findings, digital applications should incorporate various stages of user experience and should include a combination of traditional and digital solutions for OA care.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[PDF File (Adobe PDF File), 748 KB - mhealth_v12i1e53798_app1.pdf]

Multimedia Appendix 2
Mixed Methods Appraisal Tool (MMAT), version 2018.
[PDF File (Adobe PDF File), 62 KB - mhealth_v12i1e53798_app2.pdf]

Multimedia Appendix 3
The literature search results.
[PDF File (Adobe PDF File), 82 KB - mhealth_v12i1e53798_app3.pdf]

Multimedia Appendix 4
Methodological quality assessment using the Mixed Methods Appraisal Tool (MMAT).
[DOC File , 67 KB - mhealth_v12i1e53798_app4.doc]

References

1. Uivaraseanu B, Vesa C, Tit D, Abid A, Maghiar O, Maghiar T, et al. Therapeutic approaches in the management of knee osteoarthritis (Review). Exp Ther Med 2022 May 15;23(5):328 [FREE Full text] [doi: 10.3892/etm.2022.11257] [Medline: 35386619]
2. Jormand H, Mohammadi N, Khani Jeihooni A, Afzali Harsini P. Self-care behaviors in older adults suffering from knee osteoarthritis: Application of theory of planned behavior. Front Public Health 2022 Nov 4;10:958614 [FREE Full text] [doi: 10.3389/fpubh.2022.958614] [Medline: 36408046]
3. Sharma L. Osteoarthritis of the Knee. N Engl J Med 2021 Jan 07;384(1):51-59. [doi: 10.1056/NEJMcp1903768] [Medline: 33406330]
4. Veronese N, Honvo G, Bruyère O, Rizzoli R, Barbagallo M, Maggi S, et al. Knee osteoarthritis and adverse health outcomes: an umbrella review of meta-analyses of observational studies. Aging Clin Exp Res 2023 Feb 04;35(2):245-252 [FREE Full text] [doi: 10.1007/s40520-022-02289-4] [Medline: 36331799]
5. Arden NK, Perry TA, Bannuru RR, Bruyère O, Cooper C, Haugen IK, et al. Non-surgical management of knee osteoarthritis: comparison of ESCEO and OARSI 2019 guidelines. Nat Rev Rheumatol 2021 Jan 28;17(1):59-66 [FREE Full text] [doi: 10.1038/s41584-020-00523-9] [Medline: 33116279]
6. Ghomrawi H, Lee J. Commentary on the article risk scoring for time to end-stage knee osteoarthritis: data from the osteoarthritis initiative. Osteoarthritis Cartilage 2020 Aug;28(8):1001-1002 [FREE Full text] [doi: 10.1016/j.joca.2020.03.019] [Medline: 32416219]
7. Dobson F, Bennell K, French S, Nicolson P, Klaasman R, Holden M, et al. Barriers and Facilitators to Exercise Participation in People with Hip and/or Knee Osteoarthritis: Synthesis of the Literature Using Behavior Change Theory. Am J Phys Med Rehabil 2016 May;95(5):372-389. [doi: 10.1097/PHM.0000000000000448] [Medline: 26945211]
8. Kampmeijer R, Pavlova M, Tambor M, Golinowska S, Groot W. The use of e-health and m-health tools in health promotion and primary prevention among older adults: a systematic literature review. BMC Health Serv Res 2016 Sep 05;16 Suppl 5(Suppl 5):290 [FREE Full text] [doi: 10.1186/s12913-016-1522-3] [Medline: 27608677]

9. Wilson R, Chua J, Briggs AM, Abbott JH. The cost-effectiveness of recommended adjunctive interventions for knee osteoarthritis: Results from a computer simulation model. Osteoarthr Cartil Open 2020 Dec;2(4):100123 [FREE Full text] [doi: 10.1016/j.ocarto.2020.100123] [Medline: 36474885]
10. Kostic AM, Leifer VP, Gong Y, Robinson MK, Collins JE, Neogi T, et al. Cost-Effectiveness of Surgical Weight-Loss Interventions for Patients With Knee Osteoarthritis and Class III Obesity. Arthritis Care Res (Hoboken) 2023 Mar 17;75(3):491-500 [FREE Full text] [doi: 10.1002/acr.24967] [Medline: 35657632]
11. Bennell KL, Nelligan R, Dobson F, Rini C, Keefe F, Kasza J, et al. Effectiveness of an Internet-Delivered Exercise and Pain-Coping Skills Training Intervention for Persons With Chronic Knee Pain: A Randomized Trial. Ann Intern Med 2017 Apr 04;166(7):453-462. [doi: 10.7326/M16-1714] [Medline: 28241215]
12. Valentijn PP, Tymchenko L, Jacobson T, Kromann J, Biermann CW, AlMoslemany MA, et al. Digital Health Interventions for Musculoskeletal Pain Conditions: Systematic Review and Meta-analysis of Randomized Controlled Trials. J Med Internet Res 2022 Sep 06;24(9):e37869 [FREE Full text] [doi: 10.2196/37869] [Medline: 36066943]
13. Xie S, Wang Q, Wang L, Wang L, Song K, He C. Effect of Internet-Based Rehabilitation Programs on Improvement of Pain and Physical Function in Patients with Knee Osteoarthritis: Systematic Review and Meta-analysis of Randomized Controlled Trials. J Med Internet Res 2021 Jan 05;23(1):e21542 [FREE Full text] [doi: 10.2196/21542] [Medline: 33399542]
14. Zhang Z, Tian L, He K, Xu L, Wang X, Huang L, et al. Digital Rehabilitation Programs Improve Therapeutic Exercise Adherence for Patients With Musculoskeletal Conditions: A Systematic Review With Meta-Analysis. J Orthop Sports Phys Ther 2022 Nov;52(11):726-739. [doi: 10.2519/jospt.2022.11384] [Medline: 35960507]
15. Hewitt S, Sephton R, Yeowell G. The Effectiveness of Digital Health Interventions in the Management of Musculoskeletal Conditions: Systematic Literature Review. J Med Internet Res 2020 Jun 05;22(6):e15617 [FREE Full text] [doi: 10.2196/15617] [Medline: 32501277]
16. Shah N, Costello K, Mehta A, Kumar D. Applications of Digital Health Technologies in Knee Osteoarthritis: Narrative Review. JMIR Rehabil Assist Technol 2022 Jun 08;9(2):e33489 [FREE Full text] [doi: 10.2196/33489] [Medline: 35675102]
17. Durst J, Roesel I, Sudeck G, Sassenberg K, Krauss I. Effectiveness of Human Versus Computer-Based Instructions for Exercise on Physical Activity-Related Health Competence in Patients with Hip Osteoarthritis: Randomized Noninferiority Crossover Trial. J Med Internet Res 2020 Sep 28;22(9):e18233 [FREE Full text] [doi: 10.2196/18233] [Medline: 32985991]
18. Wang Y, Lombard C, Hussain SM, Harrison C, Kozica S, Brady SRE, et al. Effect of a low-intensity, self-management lifestyle intervention on knee pain in community-based young to middle-aged rural women: a cluster randomised controlled trial. Arthritis Res Ther 2018 Apr 17;20(1):74 [FREE Full text] [doi: 10.1186/s13075-018-1572-5] [Medline: 29665829]
19. Safari R, Jackson J, Sheffield D. Digital Self-Management Interventions for People With Osteoarthritis: Systematic Review With Meta-Analysis. J Med Internet Res 2020 Jul 20;22(7):e15365 [FREE Full text] [doi: 10.2196/15365] [Medline: 32706657]
20. O'moore KA, Newby JM, Andrews G, Hunter DJ, Bennell K, Smith J, et al. Internet Cognitive-Behavioral Therapy for Depression in Older Adults With Knee Osteoarthritis: A Randomized Controlled Trial. Arthritis Care Res (Hoboken) 2018 Jan;70(1):61-70. [doi: 10.1002/acr.23257] [Medline: 28426917]
21. Nelligan RK, Hinman RS, Kasza J, Crofts SJC, Bennell KL. Effects of a Self-directed Web-Based Strengthening Exercise and Physical Activity Program Supported by Automated Text Messages for People With Knee Osteoarthritis: A Randomized Clinical Trial. JAMA Intern Med 2021 Jun 01;181(6):776-785 [FREE Full text] [doi: 10.1001/jamainternmed.2021.0991] [Medline: 33843948]
22. Bennell K, Nelligan RK, Schwartz S, Kasza J, Kimp A, Crofts SJ, et al. Behavior Change Text Messages for Home Exercise Adherence in Knee Osteoarthritis: Randomized Trial. J Med Internet Res 2020 Sep 28;22(9):e21749 [FREE Full text] [doi: 10.2196/21749] [Medline: 32985994]
23. Nelligan RK, Hinman RS, McManus F, Lamb KE, Bennell KL. Moderators of the Effect of a Self-directed Digitally Delivered Exercise Program for People With Knee Osteoarthritis: Exploratory Analysis of a Randomized Controlled Trial. J Med Internet Res 2021 Oct 29;23(10):e30768 [FREE Full text] [doi: 10.2196/30768] [Medline: 34714252]
24. Allegrante JP, Kovar PA, MacKenzie C, Peterson MGE, Gutin B. A walking education program for patients with osteoarthritis of the knee: theory and intervention strategies. Health Educ Q 1993 Sep 04;20(1):63-81. [doi: 10.1177/109019819302000107] [Medline: 8444626]
25. Ravalli S, Roggio F, Lauretta G, Di Rosa M, D'Amico AG, D'agata V, et al. Exploiting real-world data to monitor physical activity in patients with osteoarthritis: the opportunity of digital epidemiology. Heliyon 2022 Feb;8(2):e08991 [FREE Full text] [doi: 10.1016/j.heliyon.2022.e08991] [Medline: 35252602]
26. Timmers T, Janssen L, van der Weegen W, Das D, Marijnissen W, Hannink G, et al. The Effect of an App for Day-to-Day Postoperative Care Education on Patients With Total Knee Replacement: Randomized Controlled Trial. JMIR Mhealth Uhealth 2019 Oct 21;7(10):e15323 [FREE Full text] [doi: 10.2196/15323] [Medline: 31638594]
27. Rini C, Katz AWK, Nwadugbo A, Porter LS, Somers TJ, Keefe FJ. Changes in Identification of Possible Pain Coping Strategies by People with Osteoarthritis who Complete Web-based Pain Coping Skills Training. Int J Behav Med 2021 Aug 10;28(4):488-498 [FREE Full text] [doi: 10.1007/s12529-020-09938-w] [Medline: 33174614]

28. McCurry SM, Zhu W, Von Korff M, Wellman R, Morin CM, Thakral M, et al. Effect of Telephone Cognitive Behavioral Therapy for Insomnia in Older Adults With Osteoarthritis Pain: A Randomized Clinical Trial. JAMA Intern Med 2021 Apr 01;181(4):530-538 [FREE Full text] [doi: 10.1001/jamainternmed.2020.9049] [Medline: 33616613]
29. Moher D, Liberati A, Tetzlaff J, Altman D, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 2009 Aug 18;151(4):264-9, W64 [FREE Full text] [doi: 10.7326/0003-4819-151-4-200908180-00135] [Medline: 19622511]
30. Hong QN, Fàbregues S, Bartlett G, Boardman F, Cargo M, Dagenais P, et al. The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. EFI 2018 Dec 18;34(4):285-291. [doi: 10.3233/EFI-180221]
31. Bossen D, Veenhof C, Van Beek KE, Spreeuwenberg PM, Dekker J, De Bakker DH. Effectiveness of a web-based physical activity intervention in patients with knee and/or hip osteoarthritis: randomized controlled trial. J Med Internet Res 2013 Nov 22;15(11):e257 [FREE Full text] [doi: 10.2196/jmir.2662] [Medline: 24269911]
32. Rini C, Porter LS, Somers TJ, McKee DC, DeVellis RF, Smith M, et al. Automated Internet-based pain coping skills training to manage osteoarthritis pain: a randomized controlled trial. Pain 2015 May;156(5):837-848 [FREE Full text] [doi: 10.1097/j.pain.0000000000000121] [Medline: 25734997]
33. Pearson J, Walsh N, Carter D, Koskela S, Hurley M. Developing a Web-Based Version of An Exercise-Based Rehabilitation Program for People With Chronic Knee and Hip Pain: A Mixed Methods Study. JMIR Res Protoc 2016 May 19;5(2):e67 [FREE Full text] [doi: 10.2196/resprot.5446] [Medline: 27197702]
34. Bennell KL, Campbell PK, Egerton T, Metcalf B, Kasza J, Forbes A, et al. Telephone Coaching to Enhance a Home-Based Physical Activity Program for Knee Osteoarthritis: A Randomized Clinical Trial. Arthritis Care Res (Hoboken) 2017 Jan;69(1):84-94 [FREE Full text] [doi: 10.1002/acr.22915] [Medline: 27111441]
35. Li LC, Sayre EC, Xie H, Clayton C, Feehan LM. A Community-Based Physical Activity Counselling Program for People With Knee Osteoarthritis: Feasibility and Preliminary Efficacy of the Track-OA Study. JMIR Mhealth Uhealth 2017 Jun 26;5(6):e86 [FREE Full text] [doi: 10.2196/mhealth.7863] [Medline: 28652228]
36. Lawford BJ, Hinman RS, Kasza J, Nelligan R, Keefe F, Rini C, et al. Moderators of Effects of Internet-Delivered Exercise and Pain Coping Skills Training for People With Knee Osteoarthritis: Exploratory Analysis of the IMPACT Randomized Controlled Trial. J Med Internet Res 2018 May 09;20(5):e10021 [FREE Full text] [doi: 10.2196/10021] [Medline: 29743149]
37. Button K, Nicholas K, Busse M, Collins M, Spasić I. Integrating self-management support for knee injuries into routine clinical practice: TRAK intervention design and delivery. Musculoskelet Sci Pract 2018 Feb;33:53-60 [FREE Full text] [doi: 10.1016/j.msksp.2017.11.002] [Medline: 29172113]
38. Mecklenburg G, Smittenaar P, Erhart-Hledik JC, Perez DA, Hunter S. Effects of a 12-Week Digital Care Program for Chronic Knee Pain on Pain, Mobility, and Surgery Risk: Randomized Controlled Trial. J Med Internet Res 2018 Apr 25;20(4):e156 [FREE Full text] [doi: 10.2196/jmir.9667] [Medline: 29695370]
39. Kline PW, Melanson EL, Sullivan WJ, Blatchford PJ, Miller MJ, Stevens-Lapsley JE, et al. Improving Physical Activity Through Adjunct Telerehabilitation Following Total Knee Arthroplasty: Randomized Controlled Trial Protocol. Phys Ther 2019 Jan 01;99(1):37-45 [FREE Full text] [doi: 10.1093/ptj/pzy119] [Medline: 30329126]
40. Nelligan RK, Hinman RS, Atkins L, Bennell KL. A Short Message Service Intervention to Support Adherence to Home-Based Strengthening Exercise for People With Knee Osteoarthritis: Intervention Design Applying the Behavior Change Wheel. JMIR Mhealth Uhealth 2019 Oct 18;7(10):e14619 [FREE Full text] [doi: 10.2196/14619] [Medline: 31628786]
41. Pelle T, Bevers K, van der Palen J, van den Hoogen F, van den Ende C. Effect of the dr. Bart application on healthcare use and clinical outcomes in people with osteoarthritis of the knee and/or hip in the Netherlands; a randomized controlled trial. Osteoarthritis Cartilage 2020 Apr;28(4):418-427 [FREE Full text] [doi: 10.1016/j.joca.2020.02.831] [Medline: 32119972]
42. Bailey JF, Agarwal V, Zheng P, Smuck M, Fredericson M, Kennedy DJ, et al. Digital Care for Chronic Musculoskeletal Pain: 10,000 Participant Longitudinal Cohort Study. J Med Internet Res 2020 May 11;22(5):e18250 [FREE Full text] [doi: 10.2196/18250] [Medline: 32208358]
43. Baker K, LaValley MP, Brown C, Felson DT, Ledingham A, Keysor JJ. Efficacy of Computer-Based Telephone Counseling on Long-Term Adherence to Strength Training in Elderly Patients With Knee Osteoarthritis: A Randomized Trial. Arthritis Care Res (Hoboken) 2020 Jul;72(7):982-990 [FREE Full text] [doi: 10.1002/acr.23921] [Medline: 31074576]
44. Fitzgibbon ML, Tussing-Humphreys L, Schiffer L, Smith-Ray R, Marquez DX, DeMott AD, et al. Fit and Strong! Plus: Twelve and eighteen month follow-up results for a comparative effectiveness trial among overweight/obese older adults with osteoarthritis. Prev Med 2020 Dec;141:106267 [FREE Full text] [doi: 10.1016/j.ypmed.2020.106267] [Medline: 33022324]
45. Hinman RS, Campbell PK, Lawford BJ, Briggs AM, Gale J, Bills C, et al. Does telephone-delivered exercise advice and support by physiotherapists improve pain and/or function in people with knee osteoarthritis? Telecare randomised controlled trial. Br J Sports Med 2020 Jul 20;54(13):790-797. [doi: 10.1136/bjsports-2019-101183] [Medline: 31748198]
46. Hinman RS, Kimp AJ, Campbell PK, Russell T, Foster NE, Kasza J, et al. Technology versus tradition: a non-inferiority trial comparing video to face-to-face consultations with a physiotherapist for people with knee osteoarthritis. Protocol for the PEAK randomised controlled trial. BMC Musculoskelet Disord 2020 Aug 07;21(1):522 [FREE Full text] [doi: 10.1186/s12891-020-03523-8] [Medline: 32767989]

47. Li LC, Feehan LM, Xie H, Lu N, Shaw CD, Gromala D, et al. Effects of a 12-Week Multifaceted Wearable-Based Program for People With Knee Osteoarthritis: Randomized Controlled Trial. JMIR Mhealth Uhealth 2020 Jul 03;8(7):e19116 [FREE Full text] [doi: 10.2196/19116] [Medline: 32618578]
48. Nelligan RK, Hinman RS, Teo PL, Bennell KL. Exploring Attitudes and Experiences of People With Knee Osteoarthritis Toward a Self-Directed eHealth Intervention to Support Exercise: Qualitative Study. JMIR Rehabil Assist Technol 2020 Nov 26;7(2):e18860 [FREE Full text] [doi: 10.2196/18860] [Medline: 33242021]
49. Dunphy E, Button K, Hamilton F, Williams J, Spasic I, Murray E. Feasibility randomised controlled trial comparing TRAK-ACL digital rehabilitation intervention plus treatment as usual versus treatment as usual for patients following anterior cruciate ligament reconstruction. BMJ Open Sport Exerc Med 2021 May 06;7(2):e001002 [FREE Full text] [doi: 10.1136/bmjsem-2020-001002] [Medline: 34035951]
50. Lindberg MF, Aamodt A, Badawy M, Bergvad IB, Borchgrevink P, Furnes O, et al. The effectiveness of exercise therapy and education plus cognitive behavioral therapy, alone or in combination with total knee arthroplasty in patients with knee osteoarthritis - study protocol for the MultiKnee trial. BMC Musculoskelet Disord 2021 Dec 20;22(1):1054 [FREE Full text] [doi: 10.1186/s12891-021-04924-z] [Medline: 34930194]
51. Pelle T, van der Palen J, de Graaf F, van den Hoogen FHJ, Bevers K, van den Ende CHM. Use and usability of the dr. Bart app and its relation with health care utilisation and clinical outcomes in people with knee and/or hip osteoarthritis. BMC Health Serv Res 2021 May 10;21(1):444 [FREE Full text] [doi: 10.1186/s12913-021-06440-1] [Medline: 33971861]
52. Rognsvåg T, Lindberg MF, Lerdal A, Stubberud J, Furnes O, Holm I, et al. Development of an internet-delivered cognitive behavioral therapy program for use in combination with exercise therapy and education by patients at increased risk of chronic pain following total knee arthroplasty. BMC Health Serv Res 2021 Oct 25;21(1):1151 [FREE Full text] [doi: 10.1186/s12913-021-07177-7] [Medline: 34696785]
53. Bennell KL, Jones SE, Hinman RS, McManus F, Lamb KE, Quicke JG, et al. Effectiveness of a telehealth physiotherapist-delivered intensive dietary weight loss program combined with exercise in people with knee osteoarthritis and overweight or obesity: study protocol for the POWER randomized controlled trial. BMC Musculoskelet Disord 2022 Jul 30;23(1):733 [FREE Full text] [doi: 10.1186/s12891-022-05685-z] [Medline: 35907828]
54. Groves-Williams D, McHugh GA, Bennell KL, Comer C, Hensor EMA, Conner M, et al. Evaluation of two electronic-rehabilitation programmes for persistent knee pain: protocol for a randomised feasibility trial. BMJ Open 2022 Jun 03;12(6):e063608 [FREE Full text] [doi: 10.1136/bmjopen-2022-063608] [Medline: 36194515]
55. Hinman RS, Nelligan RK, Campbell PK, Kimp AJ, Graham B, Merolli M, et al. Exercise adherence Mobile app for Knee Osteoarthritis: protocol for the MappKO randomised controlled trial. BMC Musculoskelet Disord 2022 Sep 20;23(1):874 [FREE Full text] [doi: 10.1186/s12891-022-05816-6] [Medline: 36127677]
56. Östlind E, Ekvall Hansson E, Eek F, Stigmar K. Experiences of activity monitoring and perceptions of digital support among working individuals with hip and knee osteoarthritis - a focus group study. BMC Public Health 2022 Aug 30;22(1):1641 [FREE Full text] [doi: 10.1186/s12889-022-14065-0] [Medline: 36042425]
57. Östlind E, Eek F, Stigmar K, Sant'Anna A, Hansson EE. Promoting work ability with a wearable activity tracker in working age individuals with hip and/or knee osteoarthritis: a randomized controlled trial. BMC Musculoskelet Disord 2022 Feb 03;23(1):112 [FREE Full text] [doi: 10.1186/s12891-022-05041-1] [Medline: 35114983]
58. Whittaker JL, Truong LK, Silvester-Lee T, Losciale JM, Miciak M, Pajkic A, et al. Feasibility of the SOAR (Stop OsteoARthritis) program. Osteoarthr Cartil Open 2022 Mar;4(1):100239 [FREE Full text] [doi: 10.1016/j.ocarto.2022.100239] [Medline: 36474472]
59. Godziuk K, Prado CM, Quintanilha M, Forhan M. Acceptability and preliminary effectiveness of a single-arm 12-week digital behavioral health intervention in patients with knee osteoarthritis. BMC Musculoskelet Disord 2023 Feb 17;24(1):129 [FREE Full text] [doi: 10.1186/s12891-023-06238-8] [Medline: 36797720]
60. Lorbeer N, Knoll N, Keller J, Domke A, Di Maio S, Armbrecht G, et al. Enhancing physical activity and reducing symptoms of patients with osteoarthritis of the knee: a randomized controlled trial of the PrevOP-Psychological Adherence Program. BMC Musculoskelet Disord 2023 Jul 04;24(1):550 [FREE Full text] [doi: 10.1186/s12891-023-06661-x] [Medline: 37403119]
61. Scheer JK, Costa F, Janela D, Molinos M, Areias AC, Moulder RG, et al. Sleep Disturbance in Musculoskeletal Conditions: Impact of a Digital Care Program. J Pain Res 2023;16:33-46 [FREE Full text] [doi: 10.2147/JPR.S394421] [Medline: 36636267]
62. Truong LK, Mosewich AD, Miciak M, Pajkic A, Silvester-Lee T, Li LC, et al. "I feel I'm leading the charge." Experiences of a virtual physiotherapist-guided knee health program for persons at-risk of osteoarthritis after a sport-related knee injury. Osteoarthr Cartil Open 2023 Mar;5(1):100333 [FREE Full text] [doi: 10.1016/j.ocarto.2022.100333] [Medline: 36685259]
63. Weber F, Müller C, Bahns C, Kopkow C, Färber F, Gellert P, et al. Smartphone-assisted training with education for patients with hip and/or knee osteoarthritis (SmArt-E): study protocol for a multicentre pragmatic randomized controlled trial. BMC Musculoskelet Disord 2023 Mar 23;24(1):221 [FREE Full text] [doi: 10.1186/s12891-023-06255-7] [Medline: 36959595]
64. Bouma AJ, van Wilgen P, Dijkstra A. The barrier-belief approach in the counseling of physical activity. Patient Educ Couns 2015 Feb;98(2):129-136. [doi: 10.1016/j.pec.2014.10.003] [Medline: 25455791]

65. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95 [FREE Full text] [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
66. Wood CE, Hardeman W, Johnston M, Francis J, Abraham C, Michie S. Reporting behaviour change interventions: do the behaviour change technique taxonomy v1, and training in its use, improve the quality of intervention descriptions? Implement Sci 2016 Jun 07;11(1):84 [FREE Full text] [doi: 10.1186/s13012-016-0448-9] [Medline: 27268131]
67. Shepherd N, Parker C. Depression in adults: recognition and management. Clinical Pharmacist 2017;9(4):1-15. [doi: 10.1211/PJ.2017.20202439]
68. DiRenzo D, Finan P. Self-Efficacy and the Role of Non-Pharmacologic Treatment Strategies to Improve Pain and Affect in Arthritis. Curr Treatm Opt Rheumatol 2019 Jun 15;5(2):168-178 [FREE Full text] [doi: 10.1007/s40674-019-00123-z] [Medline: 31832330]
69. Cohen E, Lee YC. A Mechanism-Based Approach to the Management of Osteoarthritis Pain. Curr Osteoporos Rep 2015 Dec 30;13(6):399-406 [FREE Full text] [doi: 10.1007/s11914-015-0291-y] [Medline: 26419467]
70. Broderick JE, Keefe FJ, Schneider S, Junghaenel DU, Bruckenthal P, Schwartz JE, et al. Cognitive behavioral therapy for chronic pain is effective, but for whom? Pain 2016 Sep 25;157(9):2115-2123. [doi: 10.1097/j.pain.0000000000000626] [Medline: 27227692]
71. Evers A, Kraaimaat F, van Riel P, de Jong A. Tailored cognitive-behavioral therapy in early rheumatoid arthritis for patients at risk: a randomized controlled trial. Pain 2002 Nov;100(1-2):141-153. [doi: 10.1016/s0304-3959(02)00274-9] [Medline: 12435467]
72. Dório M, Hunter D, Bowden J, Redman S, Redman A, Dawson G, et al. Training needs and approaches to performance improvement of musculoskeletal physiotherapists coordinating the osteoarthritis chronic care program. Osteoarthritis and Cartilage 2020 Apr;28:S451-S452. [doi: 10.1016/j.joca.2020.02.706]
73. Dreiser R, Schwartz P. AB0602 Influence of internet on relationships between patients and rheumatologists: specific focus on hip and knee osteoarthritis management (survey study). Ann Rheum Dis 2014 Jan 23;72(Suppl 3):A974.2-A9A974. [doi: 10.1136/annrheumdis-2013-eular.2924]
74. Danbjørg D, Villadsen A, Gill E, Rothmann MJ, Clemensen J. Usage of an Exercise App in the Care for People With Osteoarthritis: User-Driven Exploratory Study. JMIR Mhealth Uhealth 2018 Jan 11;6(1):e11 [FREE Full text] [doi: 10.2196/mhealth.7734] [Medline: 29326092]
75. Dunphy E, Gardner EC. Telerehabilitation to Address the Rehabilitation Gap in Anterior Cruciate Ligament Care: Survey of Patients. JMIR Form Res 2020 Sep 18;4(9):e19296 [FREE Full text] [doi: 10.2196/19296] [Medline: 32945776]
76. Willett M, Duda J, Fenton S, Gautrey C, Greig C, Rushton A. Effectiveness of behaviour change techniques in physiotherapy interventions to promote physical activity adherence in lower limb osteoarthritis patients: A systematic review. PLoS One 2019;14(7):e0219482 [FREE Full text] [doi: 10.1371/journal.pone.0219482] [Medline: 31291326]
77. Clarkson P, Vassilev I, Rogers A, Brooks C, Wilson N, Lawson J, et al. Integrating a Web-Based Self-Management Tool (Managing Joint Pain on the Web and Through Resources) for People With Osteoarthritis-Related Joint Pain With a Web-Based Social Network Support Tool (Generating Engagement in Network Involvement): Design, Development, and Early Evaluation. JMIR Form Res 2020 Nov 26;4(11):e18565 [FREE Full text] [doi: 10.2196/18565] [Medline: 33242011]
78. Plinsinga ML, Besomi M, Maclachlan L, Melo L, Robbins S, Lawford BJ, et al. Exploring the Characteristics and Preferences for Online Support Groups: Mixed Method Study. J Med Internet Res 2019 Dec 03;21(12):e15987 [FREE Full text] [doi: 10.2196/15987] [Medline: 31793893]
79. Cronström A, Dahlberg LE, Nero H, Ericson J, Hammarlund CS. 'I would never have done it if it hadn't been digital': a qualitative study on patients' experiences of a digital management programme for hip and knee osteoarthritis in Sweden. BMJ Open 2019 May 24;9(5):e028388 [FREE Full text] [doi: 10.1136/bmjopen-2018-028388] [Medline: 31129601]
80. van Kasteren Y, Freyne J, Hussain MS. Total Knee Replacement and the Effect of Technology on Cocreation for Improved Outcomes and Delivery: Qualitative Multi-Stakeholder Study. J Med Internet Res 2018 Mar 20;20(3):e95 [FREE Full text] [doi: 10.2196/jmir.7541] [Medline: 29559424]
81. Stuhlreyer J, Roder C, Krug F, Zöllner C, Flor H, Klinger R. A digital application and augmented physician rounds reduce postoperative pain and opioid consumption after primary total knee replacement (TKR): a randomized clinical trial. BMC Med 2022 Dec 05;20(1):469 [FREE Full text] [doi: 10.1186/s12916-022-02638-0] [Medline: 36464680]
82. Mrklas KJ, Barber T, Campbell-Scherer D, Green LA, Li LC, Marlett N, et al. Co-Design in the Development of a Mobile Health App for the Management of Knee Osteoarthritis by Patients and Physicians: Qualitative Study. JMIR Mhealth Uhealth 2020 Jul 10;8(7):e17893 [FREE Full text] [doi: 10.2196/17893] [Medline: 32673245]
83. Cenamor J. Use of health self-management platform features: The case of a specialist ehealth app. Technological Forecasting and Social Change 2022 Dec;185:122066. [doi: 10.1016/j.techfore.2022.122066]

Abbreviations

BCT: behavioral change technique

CBT: cognitive behavioral therapy
KOA: knee osteoarthritis
MeSH: Medical Subject Headings
MMAT: Mixed Methods Assessment Tool
OA: osteoarthritis
PA: physical activity

Edited by L Buis; submitted 19.10.23; peer-reviewed by J Zeitlin, TAR Sure; comments to author 22.12.23; revised version received 07.02.24; accepted 14.03.24; published 02.05.24.

Please cite as:
Zhu D, Zhao J, Wang M, Cao B, Zhang W, Li Y, Zhang C, Han T
Rehabilitation Applications Based on Behavioral Therapy for People With Knee Osteoarthritis: Systematic Review
JMIR Mhealth Uhealth 2024;12:e53798
URL: https://mhealth.jmir.org/2024/1/e53798
doi:10.2196/53798
PMID:38696250

Original Paper

Efficacy of a Mobile Health App (eMOTIVA) Regarding Compliance With Cardiac Rehabilitation Guidelines in Patients With Coronary Artery Disease: Randomized Controlled Clinical Trial

Celia Cruz-Cobo[1,2,3], RNC; María Ángeles Bernal-Jiménez[1,2,3], RNC, PhD; Germán Calle[2,4], MD; Livia Luciana Gheorghe[2,4], MD, PhD; Alejandro Gutiérrez-Barrios[2,4], MD, PhD; Dolores Cañadas[2,4], MD; Josep A Tur[5], PhD; Rafael Vázquez-García[2,4], MD, PhD; María José Santi-Cano[1,2,3], MD, PhD

[1]Faculty of Nursing and Physiotherapy, University of Cádiz, Cádiz, Spain

[2]Institute of Biomedical Research and Innovation of Cádiz, Cádiz, Spain

[3]Research Group on Nutrition, Molecular, Pathophysiological and Social Issues, University of Cádiz, Cádiz, Spain

[4]Cardiology Clinical Unit, Puerta del Mar Hospital, Cádiz, Spain

[5]Research Group on Community Nutrition & Oxidative Stress, University of the Balearic Islands, Palma de Mallorca, Spain

Corresponding Author:
María Ángeles Bernal-Jiménez, RNC, PhD
Faculty of Nursing and Physiotherapy
University of Cádiz
Avda. Ana de Viya, 52
Cádiz, 11009
Spain
Phone: 34 956 019042
Email: mariangeles.bernal@uca.es

Abstract

Background: Cardiac rehabilitation is fundamental among patients recovering from a coronary event, and mHealth technology may constitute a useful tool that provides guidelines based on scientific evidence in an entertaining, attractive, and user-friendly format.

Objective: This study aimed to compare the efficacy of an mHealth intervention involving the eMOTIVA app and that of usual care regarding compliance with cardiac rehabilitation guidelines in terms of lifestyle, cardiovascular risk factors, and satisfaction among patients with acute coronary syndrome.

Methods: A randomized controlled clinical trial with a parallel group design was conducted. It included 300 patients (mHealth group, 150; control group, 150) who underwent percutaneous coronary intervention for acute coronary syndrome. Both groups underwent evaluations initially (during hospitalization) and after 3 and 6 months (face-to-face consultations). The eMOTIVA app incorporates a virtual classroom providing audio and video information about a healthy lifestyle, a section for self-recording cardiovascular risk factors, and a section for feedback messages and gamification. The primary outcome variables were (1) adherence to the Mediterranean diet and the frequency of consumption of food; (2) physical activity level, sedentary time, and exercise capacity; (3) smoking cessation and nicotine dependence; (4) level of knowledge about cardiovascular risk factors; and (5) app satisfaction and usability.

Results: The study analyzed 287 patients (mHealth group, 145; control group, 142). Most participants were male (207/300, 69.0%), and the mean age was 62.53 (SD 8.65) years. Significant improvements were observed in the mHealth group compared with the control group at 6 months in terms of (1) adherence to the Mediterranean diet (mean 11.92, SD 1.70 vs 8.92, SD 2.66 points; P<.001) and frequency of eating foods (red meat [≤1/week]: 141/143, 97.9% vs 96/141, 68.1%; industrial pastries [<2/week]: 129/143, 89.6% vs 80/141, 56.8%; oily fish [≥2/week]: 124/143, 86.1% vs 64/141, 41.4%; vegetables [≥2/day]: 130/143, 90.3% vs 78/141, 55.3%; fruit [≥2/day]: 128/143, 88.9% vs 85/141, 60.2%; all P<.001); (2) physical activity (mean 2112.66, SD 1196.67 vs 1372.60, SD 944.62 metabolic equivalents/week; P<.001) and sedentary time (mean 8.38, SD 1.88 vs 9.59, SD 2.09 hours; P<.001); (3) exercise capacity (distance: mean 473.49, SD 102.28 vs 447.25, SD 93.68 meters; P=.04); and (4) level of knowledge (mean 117.85, SD 3.83 vs 111.00, SD 7.11 points; P<.001). App satisfaction was high (mean 42.53, SD 6.38 points), and its usability was excellent (mean 95.60, SD 4.03 points).

Conclusions: With the eMOTIVA app, favorable results were obtained in the intervention group in terms of adherence to the Mediterranean diet, frequency of eating certain foods, physical activity, sedentary time, exercise capacity, knowledge level, systolic blood pressure, heart rate, and blood sugar level. Furthermore, participants reported high app satisfaction and rated its usability as excellent. Thus, this innovative tool is very promising.

Trial Registration: ClinicalTrials.gov NCT05247606; https://clinicaltrials.gov/study/NCT05247606

(JMIR Mhealth Uhealth 2024;12:e55421) doi:10.2196/55421

KEYWORDS

coronary event; coronary heart disease; eHealth; lifestyle; mHealth; mobile health

Introduction

Cardiovascular disease remains the main cause of death worldwide and is responsible for 17.9 million fatalities every year [1]. In Europe, about 4 million deaths occur each year due to cardiovascular diseases. Although significant progress has been made in the diagnosis and treatment of acute coronary syndrome (ACS), nearly half of these deaths are due to ischemic heart disease [2,3]. In Spain, coronary heart disease (mainly acute myocardial infarction [AMI]) remains the leading cause of death, causing 29,068 deaths per year. Thus, reducing the prevalence of ACS is a crucial objective of public health [4,5].

A large amount of evidence has shown that leading a healthy lifestyle and modifying cardiovascular risk factors (CVRFs), such as stopping smoking, consuming a healthy diet, losing weight, achieving a suitable level of physical activity (PA), and adhering to medication, are vital in the prevention of major adverse cardiac and cerebrovascular events and death in people with coronary artery disease (CAD) [6]. However, a third of patients with CAD do not follow advice about eating healthy, doing PA, and stopping smoking [7].

Owing to medical advances, the mean hospital stay of patients after percutaneous coronary intervention (PCI) has decreased greatly in recent years, meaning that less time is available for providing health care education. Health education plays a fundamental role in the process of cardiac rehabilitation (CR) following ACS, as it empowers patients to take control of their health, improve treatment adherence, prevent future cardiovascular events, and enhance their overall quality of life [8,9]. Providing patients with ongoing support after their hospital discharge may be important after ACS. This should include changes in lifestyle, adherence to medication, and psychosocial well-being [10]. Secondary prevention, which focuses on reducing the risk of recurrent cardiovascular events in individuals who have already experienced ACS, plays a crucial role in the comprehensive management and ongoing care of these patients. CR after AMI is of utmost importance for several reasons. It reduces the risk of experiencing another cardiovascular event. Moreover, CR improves cardiovascular health through a structured program of physical exercise, health education, dietary advice, and emotional support designed to improve the quality of life of people who have experienced ACS. These programs help to control blood pressure, reduce stress, and promote healthy lifestyle habits, which contribute to better cardiovascular health. These programs also contribute to functional recovery. After AMI, many people may experience limitations in their physical and functional abilities. CR can help regain muscle strength, endurance, and cardiac function, allowing patients to return to daily activities and work. Psychosocial support is also critical. CR offers emotional and psychological support, which can be instrumental in helping patients cope with anxiety, depression, and stress closely related to coronary heart disease [11]. However, despite its benefits, less than 50% of patients with coronary heart disease who are eligible for a CR program participate in CR after an acute coronary event. This may be due to limited accessibility and availability owing to a lack of facilities and long waiting lists. Patients may also experience logistical and transport barriers that make regular participation in face-to-face CR sessions difficult [12]. The widespread use of information and communication technology via smartphones may make it easier for health care professionals to handle these patients. Mobile health (mHealth) technology can provide evidence-based health care advice in an entertaining, attractive, and user-friendly format, thereby reducing the cost of health care [13]. In some cases, it may be a viable alternative or complementary approach to conventional CR. This modality involves participation in distance rehabilitation programs that encompass essential elements such as remote counseling, social interaction, supervision, and distance education [14].

A recent meta-analysis [15] concluded that mHealth technology has a positive effect on patients who have experienced a coronary event. It analyzed the effectiveness of different kinds of mHealth programs in changing lifestyle, promoting treatment compliance, and controlling modifiable CVRFs. The analysis found improvements in exercise capacity, PA, physical and mental quality of life, and medication adherence. In addition, readmissions for all causes and cardiovascular causes were lower, although no significant improvements were found regarding blood lipids, arterial blood pressure, BMI, and waist circumference (WC). Another meta-analysis analyzed the effects of mHealth interventions on the risk factors of coronary heart disease, showing that they can lead to significant improvements in BMI, WC, blood lipids, diastolic blood pressure (DBP), and levels of depression. However, no improvements were found in systolic blood pressure (SBP) and anxiety [16].

This clinical trial aimed to assess the efficacy of an mHealth intervention based on a mobile phone health (eMOTIVA) app compared with usual care for improving compliance with CR guidelines and evaluate the secondary prevention outcomes in patients who have experienced ACS. The following variables were assessed: improvements in lifestyle (adherence to the Mediterranean diet, frequency of foods consumed, PA, exercise capacity, sedentary time, smoking cessation, and level of

knowledge); control of CVRFs, such as BMI, WC, blood pressure, heart rate (HR), total cholesterol (TC), low-density lipoprotein cholesterol (LDL-C), high-density lipoprotein cholesterol (HDL-C), triglycerides, blood sugar, and HbA_{1c}; and usability and satisfaction with the app.

Methods

Study Design

We conducted a randomized controlled clinical trial with a parallel group design that included 300 patients with CAD who underwent PCI with stent implantation after ACS. The trial was conducted in the Cardiology Service of a public reference hospital in the south of Spain, in which 1500 PCIs are conducted every year.

The trial has been developed and reported in agreement with the CONSORT (Consolidated Standards for Reporting Clinical Trials) checklist (Multimedia Appendix 1) [17]. The trial was registered at ClinicalTrials.gov (NCT05247606). The study protocol has been previously published [18].

Participants

During hospitalization, patients were considered eligible to participate if they had experienced myocardial infarction or angina and undergone revascularization with stent implantation, were younger than 75 years, had a smartphone or tablet with internet access for the duration of the study, and were able to manage the software. Patients were excluded if they had an expected survival of less than 1 year, had a physical disability, had severe heart failure, had a severe psychiatric illness, had dementia, did not speak Spanish, had a congenital heart disease with a rheumatic etiology, or required triple heart bypass surgery.

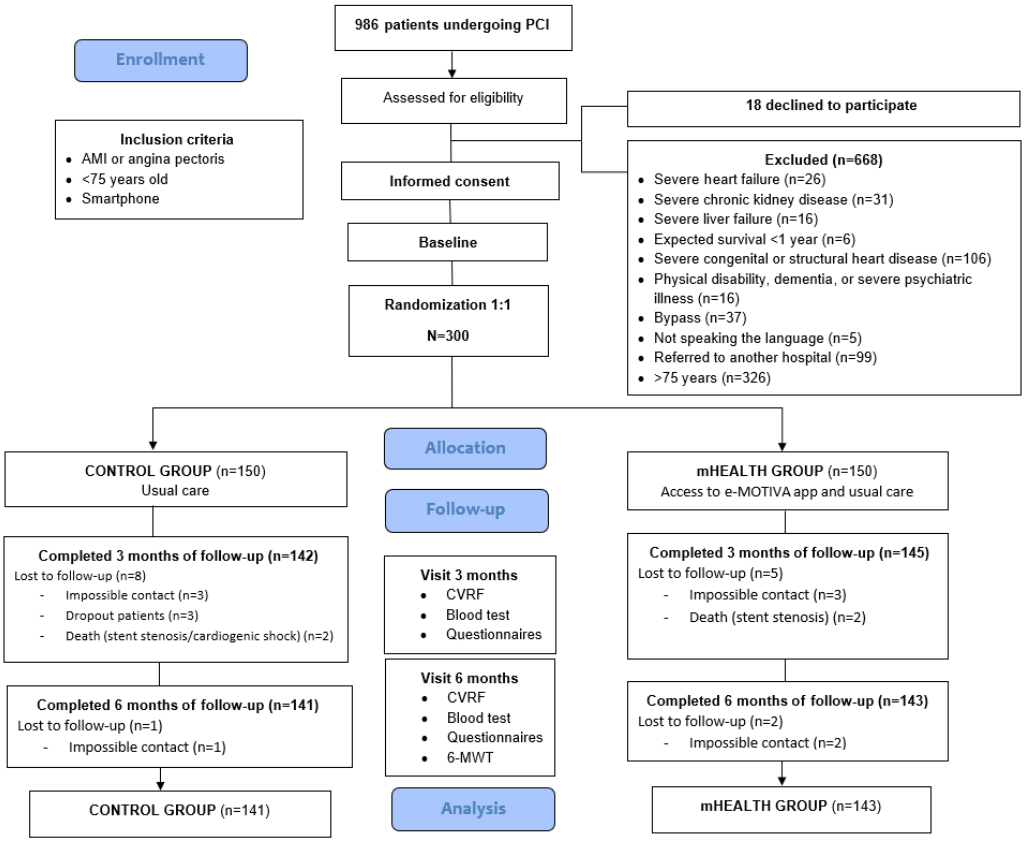
A total of 150 patients were included in each arm. This sample size was considered sufficient to detect a mean effect size (Cohen *d*) of 0.5 [19] with regard to adherence to the Mediterranean diet (mean 8.6, SD 2.0 points) [20], adherence to PA (mean 210.2, SD 221.8 metabolic equivalent (MET)-min/week) [21,22], and a 12% decrease in the prevalence of smokers (prevalence of 21% from the prior pilot study), with a 95% confidence level and a statistical power of 80%.

Recruitment, Randomization, and Blinding

Recruitment took place between February 2022 and February 2023, and the follow-up continued until September 2023. The flow diagram of participants is shown in Figure 1.

Participants meeting the inclusion criteria described above were randomly allocated using a computerized random number generator (1:1) to either the mHealth or control group (usual care). The researchers analyzing the results were blinded to the allocation of the participants.

Figure 1. Flow diagram. 6-MWT: 6-minute walk test; AMI: acute myocardial infarction; CVRF: cardiovascular risk factor; PCI: percutaneous coronary intervention.



Intervention

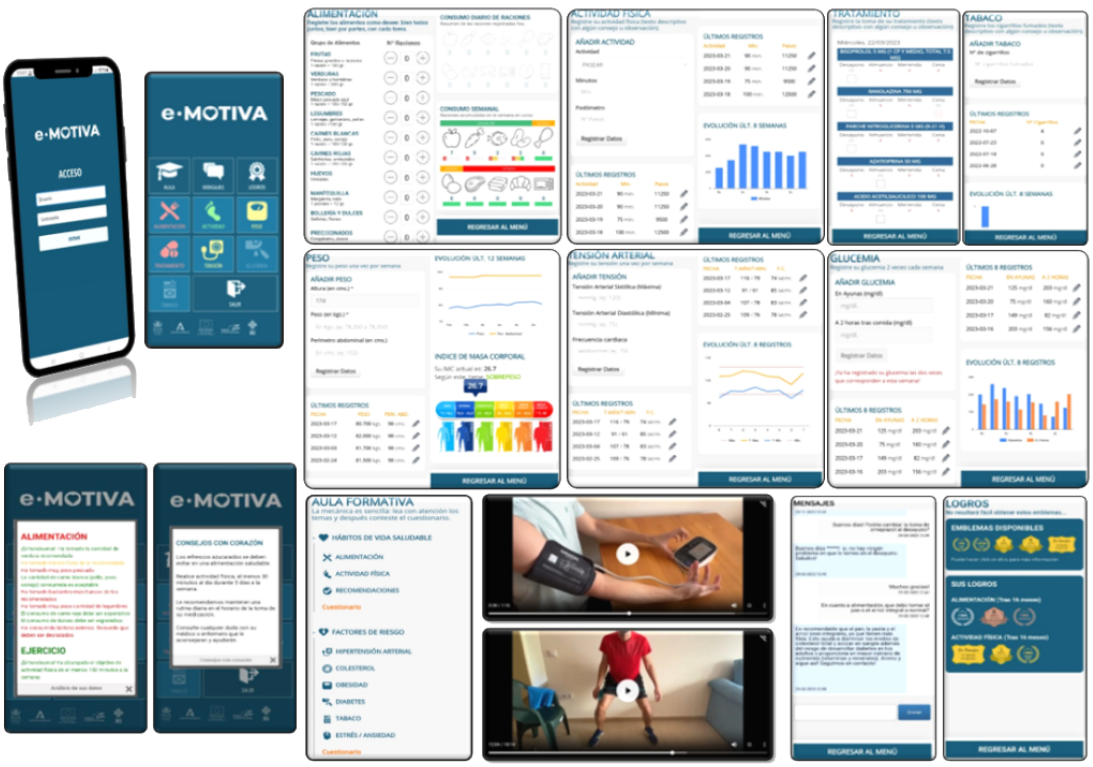
The intervention began while the patient was in the hospital after the coronary event. All participants in the mHealth group had the eMOTIVA app installed on their mobile phones or tablets. The app incorporates a virtual classroom that comprises a space for participation that guides the user using information based on scientific evidence to reach the treatment goals recommended in the clinical practice guidelines and to maintain a healthy lifestyle. This section addresses four cornerstones of the secondary prevention of cardiovascular diseases: (1) healthy lifestyle habits (diet, PA, and recommendations); (2) risk factors (high blood pressure, cholesterol, obesity, diabetes, tobacco use, and stress/anxiety); (3) compliance with the treatment; and (4) goals to be reached regarding diet, PA, body weight, blood pressure, blood sugar, smoking, and treatment. Each section includes online interactive videos (about indoor and outdoor PA, the correct self-measurement of blood pressure and WC, the treatment of cardiac events, and a guided mindfulness relaxation audio). In addition, the classroom provides documents that can be downloaded and printed such as weekly menus and graphics with information (food pyramid, heart health, characteristics and benefits of physical exercise, and recommendations for a healthy lifestyle to stop smoking and control stress). Each section includes a questionnaire that needs to be completed to obtain feedback about the knowledge acquired in the educational sessions. The app includes the use of behavioral strategies to achieve changes in habits through the self-recording of data in the sections related to food consumption, weekly body weight, treatment compliance, PA, smoking, and capillary blood sugar in patients with diabetes.

To motivate the participants to improve and maintain healthy habits, the app includes some functions. First, reminders about healthy habits are generated at random on a pop-up screen once a week. Second, personalized messages are provided according to the user's achievements, and recommendations are adapted to aspects that need to be improved, using information recorded during the previous week. These messages may be green (goal reached), yellow (goal partially reached), or red (goal still to be reached). Furthermore, each icon on the home page of the app appears in the colors according to the goals reached and aspects that need to be improved (Figure 2).

The app uses gamification in the form of achievement icons. Users can obtain different medals if they meet the established recommendations for diet and PA during the months in which they use the app. These systems with fun rewards, such as digital badges obtained for specific objectives, are related to the participation and motivation of users in mHealth interventions, and they encourage an initial and sustained commitment among users to modify CVRFs [23,24]. In addition, gamification can make the intervention more enjoyable, and this is in line with the theory of self-reliance, which assumes that a key part of intrinsic motivation is enjoyment [25]. The app also has fun and colorful warnings and messages, advice, feedback, and self-comparisons through graphics detailing weekly progress (Figure 2).

The app has a messaging section through which the patient can contact health care professionals and resolve any queries (Figure 2). The patients from both groups were evaluated through face-to-face consultations and assessments of medical records at the start and then 3 and 6 months after hospital discharge.

Figure 2. Contents of the eMOTIVA app.



Outcome Variables

The primary outcome measures at the end of the intervention in both groups were changes in behavior regarding (1) healthy diet (adherence to the Mediterranean diet and frequency of eating each food group); (2) level of PA (METs/week and min/week), sedentary time (hours sitting/week), and exercise capacity (6-minute walk test [6-MWT]); (3) smoking cessation in smokers and nicotine dependence; (4) level of knowledge acquired about CVRFs; and (5) app satisfaction and usability.

The secondary outcome measures were (1) BMI and WC; (2) arterial blood pressure and HR; (3) TC, LDL-C, and triglycerides; and (4) HbA_{1c} and blood sugar in patients with diabetes.

Primary Outcomes

Adherence to the Mediterranean diet questionnaire was used (total score: 14 points; <9 points: low adherence, ≥9 points: good adherence) to evaluate diet [20]. The frequency with which each food group was consumed was measured using a food consumption frequency questionnaire (for each food, the participant was required to tick the box indicating the mean frequency of consumption during the last week) [26]. PA time (min/week), intensity (METs/week), and sedentary time (hours seated/week) were analyzed using the International Physical Activity Questionnaire (IPAQ) [27] (light PA: minimum recommended walking of 150 min/week or 495 METs/week, moderate PA: minimum 600 METs/week, and vigorous PA: at least 3000 METs/week). Exercise capacity was measured with the 6-MWT (meters) [28]. Healthy people can walk between 400 and 700 meters in 6 minutes, depending on their age, height, and sex. A greater distance covered is associated with a higher exercise capacity. To this end, a change of 50 meters was established as a clinically significant improvement. A distance below 350 meters is considered a predictor of higher mortality in patients with chronic diseases [29,30]. Smoking cessation was self-reported and nicotine dependence was assessed using the Fagerström test (<4 points: low dependence, 4-7 points: moderate dependence, and 8-10 points: high dependence) [31]. Level of knowledge of CVRFs and a healthy lifestyle were analyzed using a scale validated by the research team in these kinds of patients (maximum score: 120 points). The scale comprises 24 items, each with a score between 1 and 5 points, and respondents were considered to have a high level of knowledge when the correct response was chosen for over 75% of the items (90 points) [32].

Secondary Outcomes

During hospital admission and in the physical follow-up visits, the following measurements were taken: body weight and height to calculate BMI, WC, SBP and DBP, HR, lipid values (TC, HDL-C, LDL-C, and triglycerides), HbA_{1c}, and blood sugar.

Finally, satisfaction with the app was assessed using a specific questionnaire developed by the research team (maximum score: 50 points; a higher score indicates more satisfaction), and the usability of the app was measured using the System Usability Scale (SUS) questionnaire (total score: 100 points; excellent: >80.3 points, good: 68-80.3 points, poor: 51-67 points, and very poor: <51 points) [33].

Statistical Analysis

A descriptive statistical analysis was performed. Continuous variables have been summarized as mean (SD), median, SD, 95% CI, and interquartile interval, depending on the distribution of the values (normal or nonnormal), and categorical variables have been summarized as frequency and percentage. At the end of the 6-month follow-up period, the means of the quantitative primary outcomes of the 2 groups (mHealth and control) were compared using the Student *t*-test (variables with normal distribution) and the Mann-Whitney *U* test (variables with nonnormal distribution). The chi-square or Fisher test was used for the comparison of the proportions of qualitative variables between the 2 groups (mHealth and control). A 2-tailed *P*-value of <.05 was considered statistically significant in all tests. SPSS version 24.0 (IBM Corp) was used for the analyses. The researchers analyzing the results were blinded to the allocation of the participants to each group.

Ethical Considerations

The project was approved by the Costa del Sol Research Ethics Committee, Ministry of Health and Families, Junta de Andalucía and was authorized by the hospital (approval number: 002_jun20_PI-RECAMAR-190. The study complies with Law 14/2007 on Biomedical Research and with the General European Data Protection Regulations, and was conducted following the standards and criteria set out in the latest version of the Helsinki Declaration issued in Fortaleza (Brazil) in October 2013. Moreover, all participants gave their written informed consent to participate in the study.

Concerning the privacy and security of the app, each participant had a private username and password to access the app. Data were stored on a web server and not on a local computer. This web server works with anonymous data and is in Spain to comply with the regulations for the protection of high-level data. The web server performs daily backups of all files, and backups are performed by the software on demand. Thus, the data and program are protected.

Results

Overview

During the recruitment period (February 2022 to February 2023), 986 patients underwent PCI and were evaluated for inclusion in the study. Among the patients evaluated, 668 were excluded for not fulfilling the inclusion and exclusion criteria, and 18 refused to participate. In the end, 300 patients were randomized into either the mHealth or control group (150 in each group). There were 9 dropouts in the control group (8 after 3 months and 1 after 6 months). In the mHealth group, there were 7 dropouts (5 after 3 months and 2 after 6 months) (Figure 1).

The baseline characteristics of the participants are shown in Table 1. Most of the participants were male (207/300, 69.0%), and the mean age was 62.53 (SD 8.65) years. In general, both groups were homogeneous.

Table 1. Patient baseline characteristics.

Characteristic	Total (N=300)	mHealth[a] (n=150)	Control (n=150)	*P* value
Male sex, n (%)	207 (69.0)	103 (68.7)	104 (69.3)	.90
Age (years)				.005
Mean (SD)	62.53 (8.65)	61.13 (8.69)	63.93 (8.41)	
95% CI	61.55-63.51	59.73-62.54	62.57-65.28	
Educational level, n (%)				.11
Primary	128 (44.9)	59 (40.7)	69 (49.3)	
Middle school	117 (41.1)	60 (41.4)	57 (40.7)	
High school	40 (14.0)	26 (17.9)	14 (10.0)	
Employment status, n (%)				.15
Employed	94 (32.1)	55 (37.2)	39 (26.9)	
Unemployed	34 (11.6)	19 (12.8)	15 (10.3)	
Retired	149 (50.9)	68 (45.9)	81 (55.9)	
Occupational disability	16 (5.5)	6 (4.1)	10 (6.9)	
BMI (kg/m^2)				.72
Mean (SD)	28.75 (4.63)	28.84 (4.56)	28.65 (4.70)	
95% CI	28.22-29.27	28.10-29.58	27.89-29.41	
Waist circumference (cm)				.49
Mean (SD)	103.98 (11.49)	104.44 (11.91)	103.46 (11.02)	
95% CI	102.59-105.38	102.45-106.43	101.50-105.42	
Cardiovascular risk factors, n (%)				
Overweight	127 (42.3)	64 (42.7)	63 (42.0)	.31
Obesity	107 (35.7)	52 (34.6)	55 (36.7)	.31
High blood pressure	204 (68.0)	95 (63.3)	109 (72.7)	.08
Diabetes	131 (43.7)	58 (38.7)	73 (48.7)	.08
Dyslipidemia	199 (66.3)	100 (66.7)	99 (66.0)	.90
Smoking	107 (35.7)	56 (37.3)	51 (34.0)	.54
Former smoker	116 (38.7)	54 (36.0)	62 (41.3)	.34
Morbidities[b]	63 (21.1)	25 (16.8)	38 (25.3)	.07
Personal history of CVD[c], n (%)				
Stable angina	48 (16.0)	21 (14.0)	27 (18.0)	.34
Unstable angina	28 (9.3)	10 (6.7)	18 (12.1)	.11
NSTEMI[d]	27 (9.0)	8 (5.3)	19 (12.7)	.02
STEMI[e]	37 (12.3)	15 (10.0)	22 (14.7)	.21
Arrhythmia	17 (5.7)	9 (6.0)	8 (5.3)	.80
Stroke	14 (4.7)	9 (6.0)	5 (3.3)	.27
Peripheral artery disease	5 (1.7)	3 (2.0)	2 (1.3)	.65
LVEF[f] (%)				.55[g]
Mean (SD)	56.79 (10.41)	56.79 (10.10)	56.78 (10.74)	
95% CI	55.57-58.00	55.12-58.46	55.00-58.56	
Reason for catheterization, n (%)				.49
Stable angina	108 (36.0)	54 (36.0)	54 (36.0)	

Characteristic	Total (N=300)	mHealth[a] (n=150)	Control (n=150)	*P* value
Unstable angina	45 (15.0)	19 (12.7)	26 (17.3)	
NSTEMI	59 (19.7)	28 (18.7)	31 (20.7)	
STEMI	88 (29.3)	49 (32.7)	39 (26.0)	
Stents, n				.01[h]
Mean (SD)	2.47 (1.67)	2.26 (1.60)	2.68 (1.70)	
95% CI	2.28-2.66	2.00-2.52	2.40-2.96	
Complete arterial revascularization, n (%)	244 (81.6)	125 (83.9)	119 (79.3)	.30
Discharge treatment, n (%)				
Anticoagulants	30 (10.0)	18 (12.1)	12 (8.0)	.24
Antiplatelets	294 (98.3)	146 (98.0)	148 (98.7)	.64
Antihypertensives	289 (96.7)	140 (94.0)	149 (99.3)	.01
Insulin	43 (14.4)	14 (9.4)	29 (19.3)	.01
Oral antidiabetics	149 (49.8)	66 (44.3)	83 (55.3)	.05
Statins	280 (93.6)	141 (94.6)	139 (92.7)	.48

[a]mHealth: mobile health.

[b]Chronic obstructive pulmonary disease, kidney disease, and obstructive sleep apnea syndrome.

[c]CVD: cardiovascular disease.

[d]NSTEMI: non-ST-segment elevation myocardial infarction.

[e]STEMI: ST-segment elevation acute myocardial infarction.

[f]LVEF: left ventricular ejection fraction.

[g]Mann-Whitney *U* test; median (IQR), mHealth vs control: 58.0 (64.75-52.00) vs 60.0 (65.00-54.00).

[h]Mann-Whitney *U* test; median (IQR), mHealth vs control: 2.00 (3.00-1.00) vs 2.00 (4.00-1.00).

Outcome Variables

Primary Outcomes

The primary outcome variables are shown in Table 2. The score for adherence to the Mediterranean diet was significantly higher in the mHealth group than in the control group after both 3 months (mean 11.63, SD 1.70 points vs mean 9.32, SD 2.55 points; *P*<.001) and 6 months (mean 11.92, SD 1.70 points vs mean 8.92, SD 2.66 points; *P*<.001). The percentage of participants with good adherence to the Mediterranean diet (>9 points) was also significantly higher in the mHealth group than in the control group after 3 months (136/145, 93.8% vs 96/142, 67.6%; *P*<.001) and 6 months (135/143, 94.4% vs 85/141, 60.3%; *P*<.001).

Regarding the frequency of eating food, the consumption of red meat was lower in the mHealth group than in the control group at 3 months (≤1 time/week: 142/145, 97.9% vs 107/142, 75.3%; *P*<.001) and 6 months (≤1 time/week: 141/143, 97.9% vs 96/141, 68.1%; *P*<.001). Moreover, the consumption of industrial pastries was lower in the mHealth group than in the control group at 3 months (<2 times/week: 128/145, 88.2% vs 89/142, 62.7%; *P*<.001) and 6 months (<2 times/week: 129/143, 89.6% vs 80/141, 56.8%; *P*<.001). In addition, the consumption of the following foods was significantly higher in the mHealth group than in the control group: oily fish (≥2 times/week: 116/145, 80.0% vs 68/142, 47.9%; *P*<.001 at 3 months and 124/143, 86.1% vs 64/141, 41.4%; *P*<.001 at 6 months), vegetables (≥2 times/day: 124/145, 85.5% vs 77/142, 54.2%; *P*<.001 at 3 months and 130/143, 90.3% vs 78/141, 55.3%; *P*<.001 at 6 months), fruit (≥2 times/day: 125/145, 86.2% vs 90/142, 63.4%; *P*<.001 at 3 months and 128/143, 88.9% vs 85/141, 60.2%; *P*<.001 at 6 months), and whole-meal cereals (≥1 time/day: 89/145, 61.3% vs 43/142, 30.2%; *P*<.001 at 3 months and 96/143, 66.7% vs 42/141, 29.8%; *P*<.001 at 6 months).

Regarding the time spent doing PA each week (min/week), the mHealth group did significantly more PA than the control group at 3 months (mean 578.10, SD 326.14 min/week vs mean 443.46, SD 278.11 min/week; *P*<.001). Likewise, at 6 months, PA was higher in the mHealth group than in the control group (mean 614.51, SD 332.26 min/week vs mean 408.40, SD 274.49 min/week; *P*<.001). Regarding the intensity of PA (METs/week), the mHealth group performed more intense activity than the control group at 3 months (mean 1991.74, SD 1176.71 METs/week vs mean 1490.48, SD 925.89 METs/week; *P*<.001) and 6 months (mean 2112.66, SD 1196.67 METs/week vs mean 1372.60, SD 944.62 METs/week; *P*<.001). The PA was of moderate intensity in both groups.

The control group had a significantly more sedentary lifestyle than the mHealth group (number of hours seated: mean 9.34, SD 2.13 vs mean 8.57, SD 1.89; *P*=.002 at 3 months and mean 9.59, SD 2.09 vs mean 8.38, SD 1.88; *P*<.001 at 6 months).

Exercise capacity, assessed using the distance covered in meters during the 6-MWT, was significantly higher in the mHealth

group than in the control group (mean 473.49, SD 102.28 meters vs mean 447.25, SD 93.68 meters; *P*=.04).

Regarding smoking cessation, although more participants gave up smoking in the mHealth group than in the control group, the difference was not significant. However, the scores for nicotine dependence at 3 months decreased significantly in the mHealth group compared with the control group (mean 2.30, SD 2.27 points vs mean 4.14, SD 2.96 points; *P*=.03).

The level of knowledge of CVRFs and a healthy lifestyle was significantly higher in the mHealth group than in the control group at both 3 months (mean 116.14, SD 4.23 points vs mean 111.02, SD 6.94 points; *P*<.001) and 6 months (mean 117.85, SD 3.83 points vs mean 111.00, SD 7.11 points; *P*<.001).

Finally, the participants in the mHealth group expressed a high level of satisfaction with the app at 3 months (mean 42.32, SD 5.96 points) and 6 months (mean 42.53, SD 6.38 points), and rated it as excellent (>80.3 points) for usability at 3 months (mean 95.75, SD 4.04 points) and 6 months (mean 95.60, SD 4.03 points).

Table 2. Primary outcome variables at baseline, and 3 and 6 months.

Variable	Total	Mobile health group	Control group	*P* value
Participants, n				—[a]
Baseline	300	150	150	
3 months	287	145	142	
6 months	284	143	141	
Mediterranean diet				
Mediterranean diet adherence (score), mean (SD)				
Baseline	7.85 (2.52)	7.78 (2.62)	7.92 (2.42)	.48[b]
3 months	10.48 (2.45)	11.63 (1.70)	9.32 (2.55)	<.001[b]
6 months	10.43 (2.69)	11.92 (1.70)	8.92 (2.66)	<.001[b]
Good adherence, n (%)				
Baseline	117 (39.3)	58 (38.7)	60 (40.0)	.57
3 months	232 (80.8)	136 (93.8)	96 (67.6)	<.001
6 months	200 (77.5)	135 (94.4)	85 (60.3)	<.001
Food consumption				
Red meat ≤1/week, n (%)				
Baseline	127 (42.5)	63 (42.3)	64 (42.7)	.70
3 months	249 (86.7)	142 (97.9)	107 (75.3)	<.001
6 months	237 (83.2)	141 (97.9)	96 (68.1)	<.001
Blue fish/oily fish ≥2/week, n (%)				
Baseline	120 (40.0)	60 (40.3)	60 (40.0)	.57
3 months	184 (64.1)	116 (80.0)	68 (47.9)	<.001
6 months	186 (66.0)	124 (86.1)	64 (45.4)	<.001
Vegetables ≥2/day, n (%)				
Baseline	98 (32.7)	49 (32.9)	49 (32.6)	.86
3 months	201 (70.0)	124 (85.5)	77 (54.2)	<.001
6 months	208 (73.0)	130 (90.3)	78 (55.3)	<.001
Fruits ≥2/day, n (%)				
Baseline	145 (48.5)	75 (50.4)	70 (46.6)	.30
3 months	215 (74.9)	125 (86.2)	90 (63.4)	<.001
6 months	213 (74.8)	128 (88.9)	85 (60.2)	<.001
Whole grains ≥1/day, n (%)				
Baseline	76 (25.6)	39 (26.4)	37 (24.8)	.76
3 months	132 (46.0)	89 (61.3)	43 (30.2)	<.001
6 months	138 (48.4)	96 (66.7)	42 (29.8)	<.001
Industrial pastry <2/week, n (%)				
Baseline	134 (44.8)	62 (41.7)	72 (47.9)	.55
3 months	217 (75.6)	128 (88.2)	89 (62.7)	<.001
6 months	209 (73.3)	129 (89.6)	80 (56.8)	<.001
Physical activity				
IPAQ[c] (min/week), mean (SD)				
Baseline	387.30 (342.72)	389.81 (355.78)	384.80 (330.33)	.87[b]

Variable	Total	Mobile health group	Control group	*P* value
3 months	511.49 (310.22)	578.10 (326.14)	443.46 (278.11)	<.001[b]
6 months	512.18 (321.44)	614.51 (332.26)	408.40 (274.49)	<.001[b]
IPAQ (METs[d]/week), mean (SD)				
Baseline	1411.48 (1480.98)	1457.28 (1632.15)	1365.68 (1316.49)	.89[b]
3 months	1743.73 (1087.58)	1991.74 (1176.71)	1490.48 (925.89)	<.001[b]
6 months	1745.24 (1139.02)	2112.66 (1196.67)	1372.60 (944.62)	<.001[b]
IPAQ H (sitting/week), mean (SD)				
Baseline	9.64 (2.40)	9.58 (2.44)	9.69 (2.37)	.84[b]
3 months	8.95 (2.04)	8.57 (1.89)	9.34 (2.13)	.002[b]
6 months	8.98 (2.07)	8.38 (1.88)	9.59 (2.09)	<.001[b]
6-MWT[e] (meters), mean (SD)				
6 months	460.75 (98.87)	473.49 (102.28)	447.25 (93.68)	.04
Tobacco				
Smokers, n (%)				
Baseline	107 (35.7)	56 (37.3)	51 (34.0)	.54
3 months	42 (42.0)	20 (37.7)	22 (46.8)	.35
6 months	42 (43.8)	17 (34.7)	25 (53.2)	.06
Smoking cessation, n (%)				
3 months	58 (58.0)	33 (62.3)	25 (53.2)	.35
6 months	54 (56.3)	32 (65.3)	22 (46.8)	.06
Nicotine dependence (Fagerström score), mean (SD)				
Baseline	5.32 (2.77)	5.39 (2.93)	5.24 (2.62)	.77
3 months	3.26 (2.78)	2.30 (2.27)	4.14 (2.96)	.03
6 months	3.05 (2.84)	2.18 (2.37)	3.64 (3.02)	.10
Cardiovascular risk factors				
CVRF[f]knowledge (score), mean (SD)				
Baseline	108.26 (9.34)	108.15 (7.39)	108.37 (10.97)	.40[b]
3 months	113.61 (6.27)	116.14 (4.23)	111.02 (6.94)	<.001[b]
6 months	114.45 (6.64)	117.85 (3.83)	111.00 (7.11)	<.001[b]
App satisfaction (score), mean (SD)				—
3 months	—	42.32 (5.96)	—	
6 months	—	42.53 (6.38)	—	
App usability (score), mean (SD)				—
3 months	—	95.75 (4.04)	—	
6 months	—	95.60 (4.03)	—	

[a]Not applicable.

[b]Mann-Whitney *U* test.

[c]IPAQ: International Physical Activity Questionnaire.

[d]MET: metabolic equivalent.

[e]6-MWT: 6-minute walk test.

[f]CVRF: cardiovascular risk factor.

Secondary Outcomes

The secondary outcome variables are shown in Multimedia Appendix 2. The anthropometric variables (BMI and WC) improved slightly in both groups, with no significant differences between the groups.

SBP was significantly lower in the mHealth group than in the control group at both 3 months (mean 128.96, SD 15.87 mmHg vs mean 133.27, SD 14.85 mmHg; *P*=.01) and 6 months (mean 130.00, SD 21.90 mmHg vs mean 135.78, SD 16.73 mmHg; *P*=.01). However, no significant differences were found in DBP between the groups. HR was significantly lower in the mHealth group than in the control group at 3 months (mean 66.75, SD 8.91 beats/min vs mean 71.93, SD 9.86 beats/min; *P*<.001) but not at 6 months.

The levels of lipid variables (TC, HDL-C, LDL-C, and triglycerides) showed large decreases in both groups, with no significant differences between the groups.

Blood sugar levels were significantly lower in the mHealth group than in the control group at 6 months (mean 101.10, SD 18.57 mg/dL vs mean 115.44, SD 39.46 mg/dL; *P*=.007). However, improvements were not reflected in the HbA_{1c} value.

Discussion

Principal Findings

This clinical trial evaluated the efficacy of an mHealth intervention based on the eMOTIVA app with regard to secondary prevention outcomes in patients who experienced ACS. The following variables were assessed: improvements in lifestyle (adherence to the Mediterranean diet, frequency of consumption of foods, PA, exercise capacity, sedentary time, smoking cessation, and level of knowledge) and control of CVRFs (BMI, WC, blood pressure, HR, TC, LDL-C, HDL-C, triglycerides, blood sugar, and HbA_{1c}). Our results showed that the eMOTIVA app achieved significantly more favorable results in the intervention group compared with the control group in terms of adherence to the Mediterranean diet, frequency of consumption of foods, time and intensity of PA, sedentary time and exercise capacity, level of knowledge about CVRFs, SBP, HR, and blood sugar. Moreover, the participants reported being very satisfied with the app, and they rated its usability as excellent.

Primary Outcome Variables

Healthy Diet

A healthy diet plays a very important role in both the prevention and treatment of CAD. Strong evidence exists about the efficacy of the Mediterranean diet for managing CVRFs for secondary prevention in patients [34,35]. In our trial, adherence to the Mediterranean diet increased significantly in the mHealth group compared with the control group at both 3 and 6 months. Moreover, in the mHealth group, an increase was observed in the consumption of healthy foods, such as fruits, vegetables, whole-meal cereals, and oily fish, and a decrease was observed in the consumption of red meats and industrial pastries. In a previous study that analyzed a cardiac telerehabilitation program with a mobile care monitoring strategy after ACS, significant improvements were noted in adherence to the Mediterranean diet in the intervention group [36]. By contrast, other authors have not reported significant differences between groups for healthy eating with the use of a support program based on text messages for patients with CAD, type 2 diabetes, or both [37]. Given our results, mHealth technology involving an app may be useful for improving eating behavior and maintaining a healthy diet in these patients compared with interventions based on text messages alone. The clinical benefits of these improvements in diet have been reported. For example, studies have stated that eating fish that is rich in omega-3 polyunsaturated fatty acids, such as oily fish, at least once a week is associated with a 16% decrease in the risk of cardiovascular disease [38]. Likewise, an increase in fiber consumption of 7 g/day is associated with a 9% decrease in the risk of cardiovascular disease [39].

Overall PA

PA is a modifiable factor that plays a crucial role in decreasing recurrent coronary events and mortality. The cardiovascular benefits of PA are well known, with recent meta-analyses reporting that it is significantly associated with a decrease in cardiovascular and all-cause mortality in patients with CAD [40-42]. Our results are promising because participants who used the eMOTIVA app performed more PA and were less sedentary. Although PA was self-reported in our trial, an objective test was conducted to measure exercise capacity using the 6-MWT, and participants in the mHealth group were found to have significantly better exercise capacity. Our results are in line with those obtained in other trials in which the effectiveness of mHealth in CAD patients was analyzed [43-46]. Recent meta-analyses have revealed that the use of interactive mobile apps with self-recording and feedback can achieve an increase in the amount of PA performed by participants and an improvement in their functional capacity [15,47].

Tobacco Use

Stopping smoking is one of the most effective secondary prevention measures after experiencing ACS [48]. The EUROASPIRE study [49], which assessed smoking cessation rates in patients with CAD in the whole of Europe and had a follow-up of 2-10 years, stated that individuals who stopped smoking showed a reduction in general mortality of nearly 50%. In our study, although no differences were observed between the groups regarding smoking cessation, nicotine dependence after 3 months, measured by the Fagerström test, was significantly lower in the mHealth group than in the control group. A recent meta-analysis [48] that analyzed smoking cessation and risk factors to continue smoking after ACS concluded that the smoking cessation rate after ACS was 45%. These results are similar to our findings, where we observed that 46.8% (22/51) of participants in the control group stopped smoking, while this figure was higher in the mHealth group (32/56, 65.3%), suggesting that our interactive tool helped participants to maintain the willpower to change, possibly owing to the support and motivation they perceived. Another recent meta-analysis [50] found that telehealth interventions had a significant effect on smoking cessation in patients with CAD.

By contrast, other meta-analyses did not find significant differences in smoking cessation between groups using telehealth interventions, but these interventions did not use interactive tools with recording, feedback, or gamification [47,51,52].

Knowledge of CVRFs

The level of knowledge of CVRFs and a healthy lifestyle in patients is not adequately addressed in trials analyzing the efficacy of mHealth. In our study, the level of knowledge was significantly higher in the mHealth group than in the control group. These results are in line with those obtained by other authors who reported that the use of a social media platform with learning modules significantly increased the knowledge and awareness of CAD [43]. Therefore, interactive and innovative mHealth tools can play a part in increasing the knowledge of a healthy lifestyle. In our study, the virtual classroom incorporated in the app may have been responsible for the observed increase in knowledge.

Secondary Outcome Variables

BMI and WC

A recent meta-analysis that analyzed the efficacy of mHealth for decreasing risk factors related to CAD found significant decreases in both BMI and WC in the intervention group [16]. However, other recent meta-analyses have reported no significant reductions in these anthropometric values with the use of an app [15,47,51,52], which is in agreement with our results. The participants in the mHealth group in our study consumed more vegetables, fruits, whole-meal cereals, and fish, and less red meat and industrial pastries. Moreover, they complied with the recommendation to perform at least 150 minutes of PA per week. Our application, however, was not specifically designed with weight loss in mind, although it did include dietary advice, and losing weight is known to involve more than merely eating healthy food. It is also necessary to limit calorie intake and increase energy expenditure through PA [53].

Blood Pressure and HR

In our study, SBP was significantly lower in the mHealth group than in the control group. Our results are in agreement with the results of other studies that analyzed the use of health care apps in patients with CAD [43,54,55]. This clinical benefit is of note because a meta-analysis [56] concluded that a 10-mmHg decrease in SBP reduces the risks of major cardiovascular events by approximately 20%, CAD by 17%, and all-cause mortality by 13%. However, no significant improvements were found in DBP, possibly due to the intensive drug treatment prescribed after a coronary event that had similar effects on patients in both the mHealth and control groups. On the other hand, the significant decrease in SBP found in our study could be explained by greater compliance with antihypertensive treatment among participants using the app or by greater adherence to the Mediterranean diet and an increase in PA. The recent prevention guidelines for cardiovascular diseases state that lifestyle interventions involving a healthy diet and physical exercise among patients with high blood pressure may be enough to control blood readings and even reduce the amount of medication required to control them [57]. Regarding HR, several trials found that mobile technology did not result in significant differences between groups [24,46,58]. By contrast, Dorje et al [43] reported a significant decrease in HR after 6 months through the use of a WeChat platform. In our study, decreases in HR to below 70 beats/min were found after both 3 and 6 months in the intervention group, but only the decrease at 3 months was significant. The higher HR decrease in the mHealth group compared with the control group may be because the mHealth group performed more PA, which has been shown to be related to a decrease in resting HR [59]. Increases in HR have a direct correlation with cardiovascular events. Several kinds of medications, including beta blockers, have been shown to help with the treatment aim of reducing HR in patients with CAD. Thus, an HR below 80 beats/min and close to 70 beats/min is a treatment goal in hypertensive patients with CAD [60].

Lipids, HbA1c, and Blood Sugar Values

Keeping blood lipid levels under control is a very important aim in the secondary prevention of cardiovascular diseases [54]. A meta-analysis conducted by Cholesterol Treatment Trialists' Collaboration [61] reported that the risk of major vascular events decreased by 21% for each 1 mmol/L reduction in LDL-C achieved with statin treatment. In our study, as in other clinical trials on the efficacy of mHealth in patients with coronary disease, blood lipid values decreased drastically, but no significant differences were found between the groups due to the powerful drug treatment received by all patients after a coronary event [37,46,62]. Likewise, a high blood sugar level is also an important risk factor that can lead to the onset and development of CAD. Diabetes mellitus is an important risk factor for AMI and a common comorbidity among patients hospitalized with AMI (present in approximately 30% of cases) [63]. Our study did not find a significant decrease in HbA_{1c}. However, blood sugar levels decreased significantly in the mHealth group after 6 months. These findings for HbA_{1c} may also be a result of the intensive drug treatment followed by the patients in both the mHealth and control groups.

Satisfaction and Usability

High levels of satisfaction and acceptance of the health care received have been observed to have positive implications for health outcomes and the patient's experience, thereby reducing health care costs and the use of emergency services [64]. In our study, satisfaction after 6 months of using the app reached a mean of 42.53 (SD 6.38) points out of 50 points, which was considered a high level of satisfaction, while the score for usability reached a mean of 95.60 (SD 4.03) points out of 100 points, which was considered to be excellent. The self-recording of PA, diet, and clinical variables along with positive personalized feedback likely contributed to the high level of satisfaction and usability reported by the participants who used the eMOTIVA app. Other studies that have used mHealth interventions with these patients have also reported high levels of usability of 80.4 points out of 100 [36] and 87.3 points out of 100 [65]. These findings highlight the potential of mHealth apps as useful tools for improving recovery and supporting secondary prevention after a coronary event. They are

particularly relevant for populations in which access to a medical center to take part in CR is difficult, either due to living in remote areas or economic reasons.

Limitations

This study has some limitations. First, one of the inclusion criteria was that patients had to have a smartphone. However, the ever-increasing use of these devices in the lives of people globally suggests that this limitation is of little importance. Due to the nature of the study, as in most trials with digital tools, it was impossible for either the patients or health care staff to be blinded. However, the staff analyzing the data were indeed blinded to the group allocation of each participant. Some variables were self-reported by patients (adherence to the Mediterranean diet and PA), which could have resulted in them overestimating their health-promoting behavior. However, the results were confirmed by other variables that were measured by health care professionals, such as exercise capacity assessed using the 6-MWT and blood pressure. Another possible limitation is that patients in the control group were 2 years older than patients in the mHealth group, and the proportion of patients receiving insulin, oral antidiabetics, and antihypertensives was slightly higher in the control group.

Strengths

A strength of the study that stands out is the relatively high number of participants included considering that this was a voluntary intervention study using mHealth, and there were very few dropouts. This might imply that the app was easy to use and that the patients were motivated to change their habits. The use of validated questionnaires specific to this population is another strength. In addition, the hospital where the intervention was conducted is a public reference hospital that treats patients from urban and rural areas. Thus, the sample is representative for the generalization of the results. Finally, the educational sessions and app were designed taking into consideration validated psychological theories. Likewise, our eMOTIVA app included setting objectives, self-monitoring of diet and PA, feedback, and gamification, which are resources that have been shown to improve the results obtained with these mHealth tools [36].

Conclusions

With the use of the eMOTIVA app, favorable results were obtained in the mHealth group compared with the control group in terms of adherence to the Mediterranean diet, frequency of eating certain foods, PA, sedentary time, exercise capacity, level of knowledge of CVRFs, SBP, HR, and blood sugar levels. This trial highlights the potential of mHealth as a complementary or alternative approach to CR programs conducted in medical centers, which are often overburdened. In addition, the participants reported high levels of satisfaction with the app, and it presented excellent usability. Thus, it could be a promising new tool for the CR of patients with CAD in general and for patients who have difficulty attending a health center or hospital in particular.

Acknowledgments

A subsidy was received for the financing of Research and Biomedical Innovation in Health Sciences within the framework of the Integrated Territorial Initiative 2014-2020 for the province of Cádiz. The project was 80% co-financed by the European Union within the framework of the FEDER (European Regional Development Fund) Andalusia Operational Program 2014-2020 and Ministry of Health and Consumption, Junta de Andalucía (reference code: PI-0014-2019). JAT is funded by CIBEROBN (Centro de Investigación Biomédica en Red de la Fisiopatología de la Obesidad y Nutrición; CB12/03/30038)-Instituto de Salud Carlos III, IUNICS (Instituto Universitario de Investigación en Ciencias de la Salud), IDISBA (Fundación Instituto de Investigación Sanitaria Islas Baleares) and co-funded by the European Regional Development Fund.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT e-HEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1111 KB - mhealth_v12i1e55421_app1.pdf]

Multimedia Appendix 2
Secondary outcome variables at baseline, and 3 and 6 months.
[PDF File (Adobe PDF File), 112 KB - mhealth_v12i1e55421_app2.pdf]

References

1. Enfermedades cardiovasculares. World Health Organization. URL: https://www.who.int/es/health-topics/cardiovascular-diseases#tab=tab_1 [accessed 2023-04-19]
2. Ferreira-González I. Epidemiología de la enfermedad coronaria. Revista Española de Cardiología 2014 Feb;67(2):139-144. [doi: 10.1016/j.recesp.2013.10.003]

3. GBD 2019 Diseases and Injuries Collaborators. Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet 2020 Oct 17;396(10258):1204-1222 [FREE Full text] [doi: 10.1016/S0140-6736(20)30925-9] [Medline: 33069326]
4. Hasbani NR, Ligthart S, Brown MR, Heath AS, Bebo A, Ashley KE, et al. American Heart Association's life's simple 7: lifestyle recommendations, polygenic risk, and lifetime risk of coronary heart disease. Circulation 2022 Mar 15;145(11):808-818. [doi: 10.1161/circulationaha.121.053730]
5. Estadística de Defunciones según la Causa de Muerte-Última Nota de prensa. Instituto Nacional de Estadística. URL: https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736176780&menu=ultiDatos&idp=1254735573175 [accessed 2024-03-17]
6. Bae J, Woo S, Lee J, Park S, Kwon SW, Choi SH, et al. mHealth interventions for lifestyle and risk factor modification in coronary heart disease: randomized controlled trial. JMIR Mhealth Uhealth 2021 Sep 24;9(9):e29928 [FREE Full text] [doi: 10.2196/29928] [Medline: 34559058]
7. Chow CK, Ariyarathna N, Islam SMS, Thiagalingam A, Redfern J. mHealth in cardiovascular health care. Heart Lung Circ 2016 Aug;25(8):802-807. [doi: 10.1016/j.hlc.2016.04.009] [Medline: 27262389]
8. Lu C. Analysis of the effectiveness of multi-disciplinary team integrated management combined with full-media health education intervention in patients with coronary heart disease and diabetes mellitus. Biotechnol Genet Eng Rev 2023 Apr 17:1-15. [doi: 10.1080/02648725.2023.2202518] [Medline: 37066841]
9. Liu X, Shi Y, Willis K, Wu C, Johnson M. Health education for patients with acute coronary syndrome and type 2 diabetes mellitus: an umbrella review of systematic reviews and meta-analyses. BMJ Open 2017 Oct 16;7(10):e016857 [FREE Full text] [doi: 10.1136/bmjopen-2017-016857] [Medline: 29042383]
10. Rathod KS, Comer K, Casey-Gillman O, Moore L, Mills G, Ferguson G, et al. Early hospital discharge following PCI for patients with STEMI. J Am Coll Cardiol 2021 Dec 21;78(25):2550-2560 [FREE Full text] [doi: 10.1016/j.jacc.2021.09.1379] [Medline: 34915986]
11. Piepoli MF, Corrà U, Benzer W, Bjarnason-Wehrens B, Dendale P, Gaita D, et al. Secondary prevention through cardiac rehabilitation: from knowledge to implementation. A position paper from the Cardiac Rehabilitation Section of the European Association of Cardiovascular Prevention and Rehabilitation. European Journal of Cardiovascular Prevention & Rehabilitation 2010 Feb;17(1):1-17. [doi: 10.1097/hjr.0b013e3283313592]
12. McDonagh ST, Dalal H, Moore S, Clark CE, Dean SG, Jolly K, et al. Home-based versus centre-based cardiac rehabilitation. Cochrane Database Syst Rev 2023 Oct 27;10(10):CD007130. [doi: 10.1002/14651858.CD007130.pub5] [Medline: 37888805]
13. Aljedaani B, Babar MA. Challenges with developing secure mobile health applications: systematic review. JMIR Mhealth Uhealth 2021 Jun 21;9(6):e15654 [FREE Full text] [doi: 10.2196/15654] [Medline: 34152277]
14. Palmer MJ, Barnard S, Perel P, Free C. Mobile phone-based interventions for improving adherence to medication prescribed for the primary prevention of cardiovascular disease in adults. Cochrane Database Syst Rev 2018 Jun 22;6(6):CD012675 [FREE Full text] [doi: 10.1002/14651858.CD012675.pub2] [Medline: 29932455]
15. Cruz-Cobo C, Bernal-Jiménez M, Vázquez-García R, Santi-Cano MJ. Effectiveness of mHealth interventions in the control of lifestyle and cardiovascular risk factors in patients after a coronary event: systematic review and meta-analysis. JMIR Mhealth Uhealth 2022 Dec 02;10(12):e39593 [FREE Full text] [doi: 10.2196/39593] [Medline: 36459396]
16. Xu Y, Ye H, Zhu Y, Du S, Xu G, Wang Q. The efficacy of mobile health in alleviating risk factors related to the occurrence and development of coronary heart disease: A systematic review and meta-analysis. Clin Cardiol 2021 May 16;44(5):609-619 [FREE Full text] [doi: 10.1002/clc.23596] [Medline: 33724494]
17. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
18. Cruz-Cobo C, Bernal-Jiménez M, Calle-Pérez G, Gheorghe L, Gutiérrez-Barrios A, Cañadas-Pruaño D, et al. Impact of mHealth application on adherence to cardiac rehabilitation guidelines after a coronary event: Randomised controlled clinical trial protocol. Digit Health 2024 Mar 19;10:20552076241234474 [FREE Full text] [doi: 10.1177/20552076241234474] [Medline: 38510574]
19. Cohen J. A power primer. Psychol Bull 1992 Jul;112(1):155-159. [doi: 10.1037//0033-2909.112.1.155] [Medline: 19565683]
20. Martínez-González MA, García-Arellano A, Toledo E, Salas-Salvadó J, Buil-Cosiales P, Corella D, PREDIMED Study Investigators. A 14-item Mediterranean diet assessment tool and obesity indexes among high-risk subjects: the PREDIMED trial. PLoS One 2012;7(8):e43134 [FREE Full text] [doi: 10.1371/journal.pone.0043134] [Medline: 22905215]
21. Elosua R, Marrugat J, Molina L, Pons S, Pujol E. Validation of the Minnesota leisure time physical activity questionnaire in Spanish men. The MARATHOM investigators. Am J Epidemiol 1994 Jun 15;139(12):1197-1209. [doi: 10.1093/oxfordjournals.aje.a116966] [Medline: 8209878]
22. Elosua R, Garcia M, Aguilar A, Molina L, Covas MI, Marrugat J. Validation of the Minnesota leisure time physical activity questionnaire in Spanish women: investigators of the MARATDON group. Med Sci Sports Exerc 2000 Aug;32(8):1431-1437. [doi: 10.1097/00005768-200008000-00011] [Medline: 10949009]
23. Sardi L, Idri A, Fernández-Alemán JL. A systematic review of gamification in e-Health. J Biomed Inform 2017 Jul;71:31-48 [FREE Full text] [doi: 10.1016/j.jbi.2017.05.011] [Medline: 28536062]

24. Gallagher R, Chow C, Parker H, Neubeck L, Celermajer D, Redfern J, et al. The effect of a game-based mobile app 'MyHeartMate' to promote lifestyle change in coronary disease patients: a randomized controlled trial. Eur Heart J Digit Health 2023 Jan;4(1):33-42 [FREE Full text] [doi: 10.1093/ehjdh/ztac069] [Medline: 36743873]
25. Xu L, Li J, Zhang X, Pang Y, Yu T, Lian X, et al. Mobile health-based gamification intervention to increase physical activity participation among patients with coronary heart disease: study protocol of a randomised controlled trial. BMJ Open 2022 Jan 31;12(1):e054623 [FREE Full text] [doi: 10.1136/bmjopen-2021-054623] [Medline: 35105640]
26. Fernández-Ballart JD, Piñol JL, Zazpe I, Corella D, Carrasco P, Toledo E, et al. Relative validity of a semi-quantitative food-frequency questionnaire in an elderly Mediterranean population of Spain. Br J Nutr 2010 Jun;103(12):1808-1816. [doi: 10.1017/S0007114509993837] [Medline: 20102675]
27. Román Viñas B, Ribas Barba L, Ngo J, Serra Majem L. [Validity of the international physical activity questionnaire in the Catalan population (Spain)]. Gac Sanit 2013;27(3):254-257 [FREE Full text] [doi: 10.1016/j.gaceta.2012.05.013] [Medline: 23103093]
28. Bellet RN, Adams L, Morris NR. The 6-minute walk test in outpatient cardiac rehabilitation: validity, reliability and responsiveness--a systematic review. Physiotherapy 2012 Dec;98(4):277-286. [doi: 10.1016/j.physio.2011.11.003] [Medline: 23122432]
29. González MN, Rodríguez NM. Prueba de la marcha de los 6 minutos. Medicina Respiratoria 2016;9(1):15-22 [FREE Full text]
30. Hamilton DM, Haennel RG. Validity and reliability of the 6-minute walk test in a cardiac rehabilitation population. J Cardiopulm Rehabil 2000;20(3):156-164. [doi: 10.1097/00008483-200005000-00003] [Medline: 10860197]
31. Sharma MK, Suman LN, Srivastava K, Suma N, Vishwakarma A. Psychometric properties of Fagerstrom Test of Nicotine Dependence: A systematic review. Ind Psychiatry J 2021;30(2):207-216 [FREE Full text] [doi: 10.4103/ipj.ipj_51_21] [Medline: 35017802]
32. Bernal-Jiménez M, Calle-Pérez G, Gutiérrez-Barrios A, Gheorghe L, Solano-Mulero AM, Trujillo-Garrido N, et al. Design and validation of a scale of knowledge of cardiovascular risk factors and lifestyle after coronary event. J Clin Med 2022 May 14;11(10):2773 [FREE Full text] [doi: 10.3390/jcm11102773] [Medline: 35628900]
33. Hyzy M, Bond R, Mulvenna M, Bai L, Dix A, Leigh S, et al. System usability scale benchmarking for digital health apps: meta-analysis. JMIR Mhealth Uhealth 2022 Aug 18;10(8):e37290 [FREE Full text] [doi: 10.2196/37290] [Medline: 35980732]
34. Quintana-Navarro GM, Alcala-Diaz JF, Lopez-Moreno J, Perez-Corral I, Leon-Acuña A, Torres-Peña JD, et al. Long-term dietary adherence and changes in dietary intake in coronary patients after intervention with a Mediterranean diet or a low-fat diet: the CORDIOPREV randomized trial. Eur J Nutr 2020 Aug;59(5):2099-2110. [doi: 10.1007/s00394-019-02059-5] [Medline: 31342228]
35. Laffond A, Rivera-Picón C, Rodríguez-Muñoz PM, Juárez-Vela R, Ruiz de Viñaspre-Hernández R, Navas-Echazarreta N, et al. Mediterranean diet for primary and secondary prevention of cardiovascular disease and mortality: an updated systematic review. Nutrients 2023 Jul 28;15(15):3356 [FREE Full text] [doi: 10.3390/nu15153356] [Medline: 37571293]
36. Dalli Peydró E, Sanz Sevilla N, Tuzón Segarra MT, Miró Palau V, Sánchez Torrijos J, Cosín Sales J. A randomized controlled clinical trial of cardiac telerehabilitation with a prolonged mobile care monitoring strategy after an acute coronary syndrome. Clin Cardiol 2022 Jan 24;45(1):31-41 [FREE Full text] [doi: 10.1002/clc.23757] [Medline: 34952989]
37. Cheung NW, Redfern J, Thiagalingam A, Hng T, Marschner S, Haider R, SupportMe Investigators. Effect of mobile phone text messaging self-management support for patients with diabetes or coronary heart disease in a chronic disease management program (SupportMe) on blood pressure: pragmatic randomized controlled trial. J Med Internet Res 2023 Jun 16;25:e38275 [FREE Full text] [doi: 10.2196/38275] [Medline: 37327024]
38. Zheng J, Huang T, Yu Y, Hu X, Yang B, Li D. Fish consumption and CHD mortality: an updated meta-analysis of seventeen cohort studies. Public Health Nutr 2012 Apr;15(4):725-737. [doi: 10.1017/S1368980011002254] [Medline: 21914258]
39. Threapleton DE, Greenwood DC, Evans CEL, Cleghorn CL, Nykjaer C, Woodhead C, et al. Dietary fibre intake and risk of cardiovascular disease: systematic review and meta-analysis. BMJ 2013 Dec 19;347(dec19 2):f6879-f6879 [FREE Full text] [doi: 10.1136/bmj.f6879] [Medline: 24355537]
40. Claes J, Buys R, Budts W, Smart N, Cornelissen VA. Longer-term effects of home-based exercise interventions on exercise capacity and physical activity in coronary artery disease patients: A systematic review and meta-analysis. Eur J Prev Cardiol 2017 Feb;24(3):244-256. [doi: 10.1177/2047487316675823] [Medline: 27798366]
41. Cheng W, Zhang Z, Cheng W, Yang C, Diao L, Liu W. Associations of leisure-time physical activity with cardiovascular mortality: A systematic review and meta-analysis of 44 prospective cohort studies. Eur J Prev Cardiol 2018 Nov 30;25(17):1864-1872. [doi: 10.1177/2047487318795194] [Medline: 30157685]
42. Cleven L, Krell-Roesch J, Nigg CR, Woll A. The association between physical activity with incident obesity, coronary heart disease, diabetes and hypertension in adults: a systematic review of longitudinal studies published after 2012. BMC Public Health 2020 May 19;20(1):726 [FREE Full text] [doi: 10.1186/s12889-020-08715-4] [Medline: 32429951]
43. Dorje T, Zhao G, Tso K, Wang J, Chen Y, Tsokey L, et al. Smartphone and social media-based cardiac rehabilitation and secondary prevention in China (SMART-CR/SP): a parallel-group, single-blind, randomised controlled trial. Lancet Digit Health 2019 Nov;1(7):e363-e374 [FREE Full text] [doi: 10.1016/S2589-7500(19)30151-7] [Medline: 33323210]

44. Fang J, Huang B, Xu D, Li J, Au WW. Innovative application of a home-based and remote sensing cardiac rehabilitation protocol in Chinese patients after percutaneous coronary intervention. Telemed J E Health 2019 Apr;25(4):288-293. [doi: 10.1089/tmj.2018.0064] [Medline: 30192210]
45. Piotrowicz E, Pencina MJ, Opolski G, Zareba W, Banach M, Kowalik I, et al. Effects of a 9-week hybrid comprehensive telerehabilitation program on long-term outcomes in patients with heart failure: the telerehabilitation in heart failure patients (TELEREH-HF) randomized clinical trial. JAMA Cardiol 2020 Mar 01;5(3):300-308 [FREE Full text] [doi: 10.1001/jamacardio.2019.5006] [Medline: 31734701]
46. Yudi MB, Clark DJ, Tsang D, Jelinek M, Kalten K, Joshi S, et al. SMARTphone-based, early cardiac REHABilitation in patients with acute coronary syndromes: a randomized controlled trial. Coron Artery Dis 2021 Aug 01;32(5):432-440. [doi: 10.1097/MCA.0000000000000938] [Medline: 32868661]
47. Ramachandran HJ, Jiang Y, Tam WWS, Yeo TJ, Wang W. Effectiveness of home-based cardiac telerehabilitation as an alternative to Phase 2 cardiac rehabilitation of coronary heart disease: a systematic review and meta-analysis. Eur J Prev Cardiol 2022 May 25;29(7):1017-1043 [FREE Full text] [doi: 10.1093/eurjpc/zwab106] [Medline: 34254118]
48. Lovatt S, Wong CW, Holroyd E, Butler R, Phan T, Patwala A, et al. Smoking cessation after acute coronary syndrome: A systematic review and meta-analysis. Int J Clin Pract 2021 Dec;75(12):e14894. [doi: 10.1111/ijcp.14894] [Medline: 34541754]
49. Snaterse M, Deckers JW, Lenzen MJ, Jorstad HT, De Bacquer D, Peters RJG, EUROASPIRE Investigators. Smoking cessation in European patients with coronary heart disease. Results from the EUROASPIRE IV survey: A registry from the European Society of Cardiology. Int J Cardiol 2018 May 01;258:1-6 [FREE Full text] [doi: 10.1016/j.ijcard.2018.01.064] [Medline: 29544918]
50. Turan Kavradim S, Özer Z, Boz İ. Effectiveness of telehealth interventions as a part of secondary prevention in coronary artery disease: a systematic review and meta-analysis. Scand J Caring Sci 2020 Sep 20;34(3):585-603. [doi: 10.1111/scs.12785] [Medline: 31747080]
51. Huang K, Liu W, He D, Huang B, Xiao D, Peng Y, et al. Telehealth interventions versus center-based cardiac rehabilitation of coronary artery disease: A systematic review and meta-analysis. Eur J Prev Cardiol 2015 Aug;22(8):959-971. [doi: 10.1177/2047487314561168] [Medline: 25488550]
52. Jin K, Khonsari S, Gallagher R, Gallagher P, Clark AM, Freedman B, et al. Telehealth interventions for the secondary prevention of coronary heart disease: A systematic review and meta-analysis. Eur J Cardiovasc Nurs 2019 Apr;18(4):260-271. [doi: 10.1177/1474515119826510] [Medline: 30667278]
53. Arnett DK, Blumenthal RS, Albert MA, Buroker AB, Goldberger ZD, Hahn EJ, et al. 2019 ACC/AHA guideline on the primary prevention of cardiovascular disease: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. Circulation 2019 Sep 10;140(11):e596-e646 [FREE Full text] [doi: 10.1161/CIR.0000000000000678] [Medline: 30879355]
54. Skobel E, Knackstedt C, Martinez-Romero A, Salvi D, Vera-Munoz C, Napp A, et al. Internet-based training of coronary artery patients: the Heart Cycle Trial. Heart Vessels 2017 Apr;32(4):408-418 [FREE Full text] [doi: 10.1007/s00380-016-0897-8] [Medline: 27730298]
55. Li Y, Gong Y, Zheng B, Fan F, Yi T, Zheng Y, et al. Effects on adherence to a mobile app-based self-management digital therapeutics among patients with coronary heart disease: pilot randomized controlled trial. JMIR Mhealth Uhealth 2022 Feb 15;10(2):e32251 [FREE Full text] [doi: 10.2196/32251] [Medline: 34906924]
56. Ettehad D, Emdin CA, Kiran A, Anderson SG, Callender T, Emberson J, et al. Blood pressure lowering for prevention of cardiovascular disease and death: a systematic review and meta-analysis. Lancet 2016 Mar 05;387(10022):957-967. [doi: 10.1016/S0140-6736(15)01225-8] [Medline: 26724178]
57. Piepoli MF, Hoes AW, Agewall S, Albus C, Brotons C, Catapano AL, ESC Scientific Document Group. 2016 European Guidelines on cardiovascular disease prevention in clinical practice: The Sixth Joint Task Force of the European Society of Cardiology and Other Societies on Cardiovascular Disease Prevention in Clinical Practice (constituted by representatives of 10 societies and by invited experts) developed with the special contribution of the European Association for Cardiovascular Prevention and Rehabilitation (EACPR). Eur Heart J 2016 Aug 01;37(29):2315-2381 [FREE Full text] [doi: 10.1093/eurheartj/ehw106] [Medline: 27222591]
58. Lunde P, Bye A, Bergland A, Grimsmo J, Jarstad E, Nilsson BB. Long-term follow-up with a smartphone application improves exercise capacity post cardiac rehabilitation: A randomized controlled trial. Eur J Prev Cardiol 2020 Nov;27(16):1782-1792 [FREE Full text] [doi: 10.1177/2047487320905717] [Medline: 32106713]
59. Visseren FLJ, Mach F, Smulders YM, Carballo D, Koskinas KC, Bäck M, ESC National Cardiac Societies, ESC Scientific Document Group. 2021 ESC Guidelines on cardiovascular disease prevention in clinical practice. Eur Heart J 2021 Sep 07;42(34):3227-3337. [doi: 10.1093/eurheartj/ehab484] [Medline: 34458905]
60. Mancia G, Kreutz R, Brunström M, Burnier M, Grassi G, Januszewicz A, et al. 2023 ESH Guidelines for the management of arterial hypertension The Task Force for the management of arterial hypertension of the European Society of Hypertension: Endorsed by the International Society of Hypertension (ISH) and the European Renal Association (ERA). J Hypertens 2023 Dec 01;41(12):1874-2071. [doi: 10.1097/HJH.0000000000003480] [Medline: 37345492]

61. Sniderman A, Thanassoulis G, Couture P, Williams K, Alam A, Furberg CD. Is lower and lower better and better? A re-evaluation of the evidence from the Cholesterol Treatment Trialists' Collaboration meta-analysis for low-density lipoprotein lowering. J Clin Lipidol 2012;6(4):303-309. [doi: 10.1016/j.jacl.2012.05.004] [Medline: 22836065]
62. Widmer RJ, Allison TG, Lennon R, Lopez-Jimenez F, Lerman LO, Lerman A. Digital health intervention during cardiac rehabilitation: A randomized controlled trial. Am Heart J 2017 Jun;188:65-72. [doi: 10.1016/j.ahj.2017.02.016] [Medline: 28577682]
63. Milazzo V, Cosentino N, Genovese S, Campodonico J, Mazza M, De Metrio M, et al. Diabetes mellitus and acute myocardial infarction: impact on short and long-term mortality. Adv Exp Med Biol 2021;1307:153-169. [doi: 10.1007/5584_2020_481] [Medline: 32020518]
64. Hibbard JH, Greene J, Overton V. Patients with lower activation associated with higher costs; delivery systems should know their patients' 'scores'. Health Aff (Millwood) 2013 Feb;32(2):216-222. [doi: 10.1377/hlthaff.2012.1064] [Medline: 23381513]
65. Johnston N, Bodegard J, Jerström S, Åkesson J, Brorsson H, Alfredsson J, et al. Effects of interactive patient smartphone support app on drug adherence and lifestyle changes in myocardial infarction patients: A randomized study. Am Heart J 2016 Aug;178:85-94 [FREE Full text] [doi: 10.1016/j.ahj.2016.05.005] [Medline: 27502855]

Abbreviations

6-MWT: 6-minute walk test
ACS: acute coronary syndrome
AMI: acute myocardial infarction
CAD: coronary artery disease
CR: cardiac rehabilitation
CVRF: cardiovascular risk factor
DBP: diastolic blood pressure
HDL-C: high-density lipoprotein cholesterol
HR: heart rate
LDL-C: low-density lipoprotein cholesterol
MET: metabolic equivalent
PA: physical activity
PCI: percutaneous coronary intervention
SBP: systolic blood pressure
TC: total cholesterol
WC: waist circumference

Edited by L Buis; submitted 13.12.23; peer-reviewed by NJ Haque, TAR Sure, S Lopes; comments to author 01.04.24; revised version received 17.04.24; accepted 28.05.24; published 25.07.24.

Please cite as:
Cruz-Cobo C, Bernal-Jiménez MÁ, Calle G, Gheorghe LL, Gutiérrez-Barrios A, Cañadas D, Tur JA, Vázquez-García R, Santi-Cano MJ
Efficacy of a Mobile Health App (eMOTIVA) Regarding Compliance With Cardiac Rehabilitation Guidelines in Patients With Coronary Artery Disease: Randomized Controlled Clinical Trial
JMIR Mhealth Uhealth 2024;12:e55421
URL: https://mhealth.jmir.org/2024/1/e55421
doi:10.2196/55421
PMID:

Original Paper

Efficacy of the mHealth-Based Exercise Intervention re.flex for Patients With Knee Osteoarthritis: Pilot Randomized Controlled Trial

Valerie Dieter[1,2], MSc; Pia Janssen[1,2], Dr Med; Inga Krauss[1,2], Dr rer soc

[1]Department of Sports Medicine, Medical Clinic, University Hospital Tübingen, Tübingen, Germany

[2]Interfaculty Research Institute for Sports and Physical Activity, Tübingen, Germany

Corresponding Author:
Valerie Dieter, MSc
Department of Sports Medicine
Medical Clinic
University Hospital Tübingen
Hoppe-Seyler-Str. 6
Tübingen, 72076
Germany
Phone: 49 7071 2986477
Email: valerie.dieter@med.uni-tuebingen.de

Abstract

Background: Exercise therapy is recommended by international guidelines as a core treatment for patients with knee osteoarthritis. However, there is a significant gap between recommendations and practice in health care. Digital exercise apps are promising to help solve this undersupply.

Objective: This study aims to evaluate the efficacy of a 12-week fully automated app-based exercise intervention with and without a supporting knee brace on health-related outcomes, performance measures, and adherence in patients with knee osteoarthritis.

Methods: This closed user group trial included participants with moderate to severe unicondylar painful knee osteoarthritis. Randomization was 1:1:2 into an intervention group (IG) with 2 subgroups (app-based training [IG A] and app-based training and a supportive knee brace [IG AB]) and a control group (CG). The intervention included a 12-week home exercise program with 3 sessions per week. Instructions for the exercises were given via the app and monitored using 2 accelerometers placed below and above the affected knee joint. Participants in the CG did not receive any study intervention but were allowed to make use of usual care. Osteoarthritis-specific pain (Knee Injury and Osteoarthritis Outcome Score) was defined as the primary outcome, and secondary outcomes included all other Knee Injury and Osteoarthritis Outcome Score subscales, general health-related quality of life (Veterans RAND 12-item Health Survey), psychological measures (eg, exercise self-efficacy), performance measures (strength and postural control), and the monitoring of adherence and safety. Outcomes were assessed at baseline and after 12 weeks. Intervention effects were calculated using baseline-adjusted analysis of covariance for the joint comparison of IG A and IG AB versus the CG using a per-protocol approach. Subgroup analyses were conducted for each IG separately.

Results: A total of 61 participants were included (IG: n=30, 49%; CG: n=31, 51%; male: n=31, 51%; female: n=30, 49%; mean age 62.9, SD 8.5 years; mean BMI 27.7, SD 4.5 kg/m^2). Analysis revealed statistically significant effects in favor of the IG for pain reduction (P<.001; effect size [ES]=0.76), improvements in physical function (P<.001; ES=0.64), improvements in symptoms (P=.01; ES=0.53), improvements in sport and recreation activities (P=.02; ES=0.47), improvements in knee-related quality of life (P<.001; ES=0.76), and improvements in the physical component of general health-related quality of life (P<.001; ES=0.74). Mean differences ranged from 6.0 to 13.2 points (scale range 0-100). ESs indicated small to medium effects. No effects were found for psychological and performance measures. Participants adhered to 92.5% (899/972) of all scheduled exercise sessions.

Conclusions: Individuals with knee osteoarthritis undergoing a 12-week sensor-assisted app-based exercise intervention with or without an additional knee brace experienced clinically meaningful treatment effects regarding pain relief and improvements in physical function as well as other osteoarthritis-specific concerns compared to controls.

Trial Registration: German Clinical Trials Register (DRKS) DRKS00023269; https://drks.de/search/de/trial/DRKS00023269

(JMIR Mhealth Uhealth 2024;12:e54356) doi:10.2196/54356

KEYWORDS

digital app; mobile health; mHealth; knee osteoarthritis; exercise; knee brace

Introduction

Background

Osteoarthritis is a degenerative joint disease and one of the major contributors to global disabilities [1]. The prevalence rate of osteoarthritis increases with age, with women being more frequently affected than men [1,2]. Almost 30% of the German population in the sixth decade of life have been diagnosed with osteoarthritis [3]. The knee joint is the most commonly affected joint of the lower extremities. With disease progression, knee osteoarthritis is more frequently associated with increasing pain, limitations in physical function [4], and decreased health-related quality of life (HRQoL). National and international guidelines recommend exercise therapy as a nonpharmacological core treatment for patients with knee osteoarthritis [5-8]. Exercise programs have shown to decrease pain and improve physical function [9]. They can include strengthening exercises but also aerobic training, neuromuscular training, balance training, mixed exercise programs, aquatic exercises, or mind-body activities such as tai chi or yoga [5,6,8,10]. Different training settings (individual, group based, and home based) have also shown to be effective, allowing patients to exercise according to their individual preferences [5]. However, it is suggested that most people with knee osteoarthritis need some form of ongoing monitoring or supervision to optimize the clinical benefits of exercise treatment [9,10]. Despite given consensus on the need to recommend exercise with some kind of supervision, there is a considerable discrepancy regarding its implementation in health care. In 2016, <40% of patients with hip, knee, or polyarticular osteoarthritis who were customers of a German statutory health insurance company received a prescription for therapeutic exercise [11], and similar numbers have been described in an international meta-analysis [12]. Therefore, it seems reasonable to explore alternative approaches for people with limited access to therapeutic services [13]. In this regard, digital apps for exercise instructions could be particularly suitable to support patients in doing exercises. A recently conducted questionnaire study with health care professionals revealed very high acceptance of mobile health (mHealth)–based intervention therapies in osteoarthritis treatment. This indicates that they would also recommend or prescribe m-Health exercise interventions [14]. The main advantages of apps are related to their use independent of time and location, making this kind of intervention available for many patients even in rural areas [15-17]. In addition, special app features such as information and advice for guidance, tracking and self-monitoring of health behavior, feedback mechanisms, and reminders via push notifications can be of particular value [18,19]. The integration of accelerometers can additionally support patients in conducting exercises in a correct and safe manner by imitating human supervision. In general, 2 main types of mHealth apps are differentiated: interactive and stand-alone apps. Interactive mHealth apps can be used for communication between patients and health care professionals such as physical therapists [20]. These kinds of apps are frequently used in the context of blended care [21]. In contrast, stand-alone apps do not involve interaction with a health care professional, and patients exercise autonomously [22,23].

Both kinds of apps can provide an added value for patients by supporting them in implementing and maintaining exercise in their life and profiting from associated health benefits. A recent meta-analysis reported short-term improvements in pain relief and quality of life (QoL) in patients with knee osteoarthritis or chronic knee pain following the use of technology-based exercise and physical activity programs [19]. However, only 2 of 12 included randomized controlled trials explicitly used an mHealth app [23,24], and only one of these examined a structured exercise program [23]. Therefore, it seems reasonable to conduct further research specifically to evaluate mHealth-based structured exercise programs for patients with knee osteoarthritis. Consideration should also be given to the type of interaction between patients and health care providers in the app.

In addition to exercise, unloading knee braces for patients with tibiofemoral unicondylar medial knee osteoarthritis can be used with the aim to reduce pain, joint stiffness, and medial compartment loading and enhance joint proprioception and functional stability [6,10,25,26]. These effects may also support the conduction of exercises, and in that case, unloader braces have the potential to serve as a treatment-supporting device. Despite limited evidence of the effectiveness of knee braces [6,8], the German guideline for the treatment of knee osteoarthritis recommends unloader knee braces as a “can do” option [7].

Objectives

Considering that exercise is one of the core treatment options for knee osteoarthritis, as well as the fact that mHealth provides new opportunities to guide home-based exercise, and the potential benefit of unloader braces to support exercise conduction, this study aimed to investigate the efficacy of a 12-week mHealth app–supported exercise intervention (re.flex) with (intervention group [IG] AB) and without (IG A) a corrective knee brace in comparison to a control group (CG) on health-related outcomes in patients with moderate to severe unicondylar knee osteoarthritis. The primary outcome was the joint comparison of the 2 app-based study arms (IG A and IG AB) regarding osteoarthritis-specific pain (Knee Injury and Osteoarthritis Outcome Score [KOOS], pain subscale) versus the CG immediately after the 12-week intervention phase.

Methods

Study Design

This study was conducted as a randomized controlled superiority trial. Study participants were randomly assigned in a 1:1:2 ratio to an IG with 2 subgroup arms (app-based exercise training [IG

A] and app-based exercise training in combination with a supportive knee brace [IG AB]) and a CG. The study is reported following the CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) checklist [27] and the Consensus on Exercise Reporting Template checklist for reporting exercise interventions [28].

Ethical Considerations

Ethics approval was obtained from the ethics committee of the University Hospital Tübingen (550/2020BO). The participants signed a written informed consent and were given a study ID number. Identifiable information was stored on password-protected servers. There was no compensation for participation in the study. Study materials were provided to participants for free. The study was registered in the German Clinical Trials Register (DRKS00023269).

Participants

Participants with knee osteoarthritis were recruited via advertisements in regional newspapers as well as emails sent to the employees of the University Hospital Tübingen and the University of Tübingen. Interested persons were screened for eligibility via phone call. Final inclusion or exclusion took place at the University Hospital Tübingen in the context of the medical examination at baseline. Inclusion and exclusion criteria are described in Textbox 1.

Textbox 1. Inclusion and exclusion criteria for study participants.

Inclusion criteria

- Age of ≥18 years
- Knee osteoarthritis (self-reported according to the wording of the study questionnaire Gesundheit in Deutschland aktuell [29] and verified by the study physician at t0)
- Unicondylar tibiofemoral concerns
- Moderate to severe knee osteoarthritis assessed using the Knee Injury and Osteoarthritis Outcome Score, pain subscale (≤60 points at the time of screening)
- Knee osteoarthritis as the primary location of symptoms
- Access to a tablet or mobile phone with iOS operating system
- Willingness to use the app to exercise
- Willingness to wear a brace while exercising
- Informed consent for study participation

Exclusion criteria

- Scheduled or implanted knee joint replacement
- Osteoarthritis primarily located in the hip joint or a joint other than the knee
- Diffuse knee pain or retro-patellar pain only
- Concerns affecting physical performance in everyday life (measured using the Physical Activity Readiness Questionnaire [30,31] and verified by the study physician at t0)
- Concerns located at the back or lower extremities currently treated by a physician or health professional and other previous surgeries, injuries, or concerns that may impair measures of strength and balance or the exercise intervention itself
- Insufficient German language skills for self-administered questionnaires and app instructions during training

Development of the Intervention

The re.flex intervention (Kineto Tech Rehab SRL) served as the underlying software and hardware for the intervention, including the basic app structure and the biofeedback via an avatar of the moving body part that is regulated by 2 accelerometers to guide and control training exercises. On the basis of this, a 12-week exercise program specifically designed for patients with knee osteoarthritis was developed by a team of experts of the Department of Sports Medicine at the University Hospital Tübingen. This program was then implemented into the app by the software manufacturer. Exercises were selected based on current knee osteoarthritis–specific recommendations from international guidelines [5,6,32] and years of experience by the study team in planning and conducting exercise interventions for patients with hip and knee osteoarthritis [33-38]. To test, analyze, and improve the re.flex knee osteoarthritis intervention during the development process, an iterative design approach [39] was used. In 2 test phases of 2 and 4 weeks, parts of the exercise program, as well as the app handling and usability, were tested by volunteers with knee concerns (data not published). Volunteers involved in the test phases were not included in the randomized controlled trial. In this study, the iOS app version 1.1.38 at the time of intervention completion of the IG was evaluated. During the intervention phase, minor technical bugs were fixed.

App-Guided Exercise Intervention

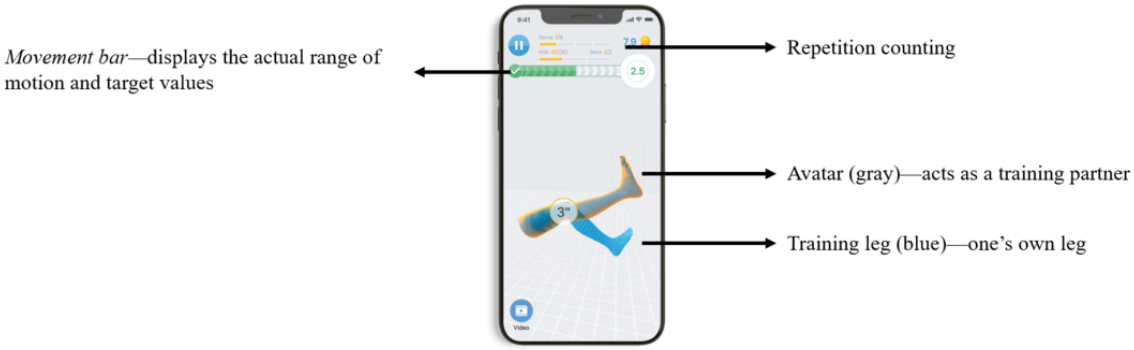
The exercise intervention was a 12-week app-guided home training program specifically designed for patients with knee osteoarthritis.

App Features

The re.flex app can be classified as a fully automated, digital health app including a training app and 2 accelerometers to monitor joint movement. It was used to guide and monitor the 12-week exercise intervention of this study. Sensors were attached proximally and distally to the affected knee joint or to the more affected joint (ie, signal joint) in case of bilateral knee osteoarthritis. They were directly attached to the skin using a hook-and-loop tape (IG A; Figure 1, left) or integrated into the brace (IG AB; Figure 1, right). Before each training session, sensors had to be calibrated by performing a movement task. The app acted as a virtual training partner, providing exercise descriptions and videos as well as setting the number of repetitions and sets of the exercises. Movement execution was monitored by the sensors and visualized via a blue avatar leg in the app interface. The blue avatar had to be aligned with another displayed gray avatar leg that moved according to the recommended movement velocity. A movement bar further visualized the current range of motion of the training leg. This bar served as an orientation on how far the leg should be moved in each direction relative to the starting position. If an exercise was not performed correctly, verbal instructions were given (eg, “extend your knee more”). After each set of exercises with the sensor-equipped leg, patients were called to conduct the set with the other leg as well. However, sets and repetitions of the other leg were performed autonomously and were not monitored in the log files of the app. Another feature of the app was to remind users of upcoming training sessions via push notifications. Figure 2 and the app manual (Multimedia Appendix 1) illustrate the structures and features of the re.flex app. The use of the app and sensors for IG A and IG AB was introduced after randomization at baseline by the pretrained study staff. Patients further received a user manual for software and hardware and log-in data for their personal anonymous and free app user account. The login data did not contain any personal data of the participants but used a fake email address with the patient’s pseudonym and an individual password. Only the authors of this study were able to reidentify them with their personal data.

Figure 1. Re.flex technology directly attached to the lower limb (left) and Sporlastic GmbH GENUDYN OA SMART with re.flex technology (right).

Figure 2. Screenshot with features of the re.flex app.

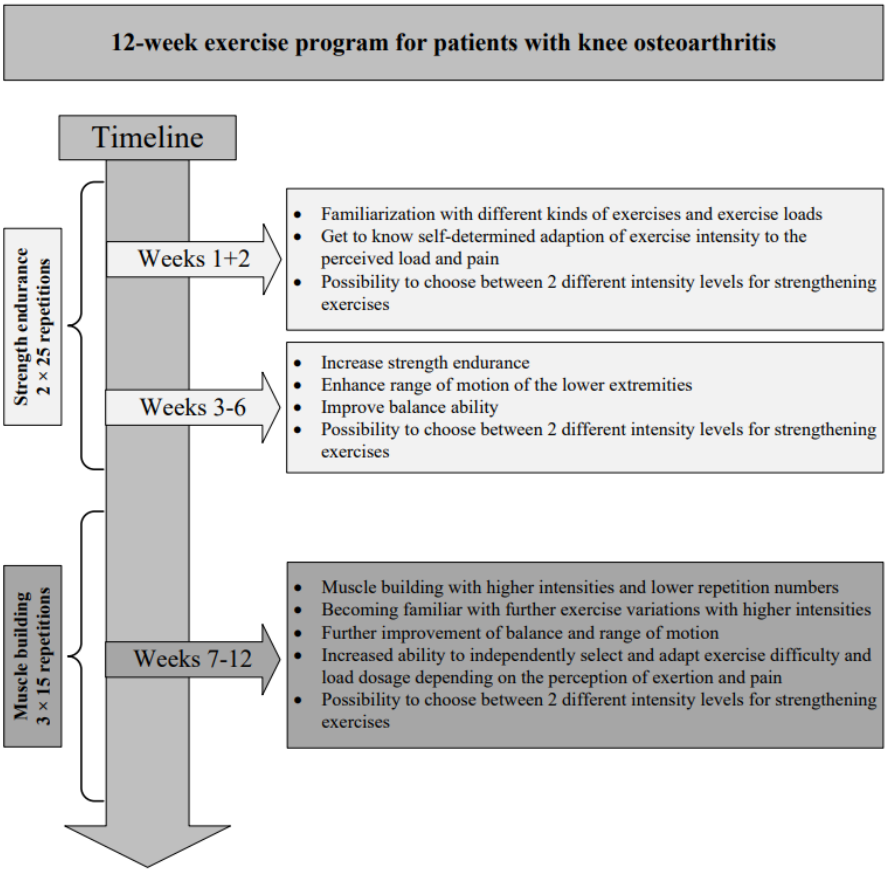


Exercise Program

The progressively designed 12-week program included 3 sessions per week with 5 different exercises and a duration of 25 to 30 minutes each. The exercise poses differed between supine, sitting, and standing. The primary focus of the intervention was to strengthen knee extensors, knee flexors, and hip abductors. Furthermore, exercises aimed for mobilization, muscle stretching, and balance training. The required training material included a chair, a ball or a pillow, and the provided training bands with different resistance levels.

The first 2 weeks focused on familiarization with different kinds of exercises and exercise loads. In this regard, patients were able to adapt exercise intensity self-determinately according to perceived strain and pain, which were assessed after each set of exercises as well as before and after each training session using in-app scales. After the period of familiarization, the exercise sessions of the following 4 weeks were designed to increase strength endurance, enhance the range of motion of the lower extremities, and improve balance ability. From week 7 on, the intervention mainly focused on muscle building. Concurrently, the complexity of the balance tasks was increased accordingly by reducing the sensory input (eg, eyes closed) or modifying the supporting surface (eg, tandem stance). The exercises provided in 2 intensity levels were predefined for each session. An overview of the different phases and objectives within the 12-week exercise program is given in Figure 3. Throughout the intervention phase, users could contact the provider for technical and medical issues using the app messenger service. In the context of the study, this function was supervised by the study personnel.

Figure 3. Objectives of the 12-week exercise program.



Individual Exercise Dosing

At baseline, participants were instructed by the study staff to perform the last 2 to 3 repetitions of each set within a strenuous to very strenuous exertion level. During balance tasks, participants were instructed to perform the task properly at all times while still maintaining the challenge. To ease the fitting of the optimal intensity level, patients could always choose between 2 different intensity levels via an in-app button feature and could further vary the resistance of elastic exercise bands, if applicable. In addition to the intensity specifications, exercises should be performed in a pain-free to low-level pain range. The following instructions were given to the patients if they experienced increased pain during exercising: (1) check exercise performance and correct if necessary, (2) reduce training intensity by selecting an easier exercise variation or reducing the number of repetitions or sets, or (3) skip the exercise. Exercise adaptations in case of increased pain were prioritized versus intensity specifications. The information and guidance for training were provided both orally and written on a fact sheet.

Knee Brace Intervention (Additional to Exercise)

The 2 exercise groups (IG A and IG AB) only differed with respect to the additional use of a corrective knee brace (GENUDYN OA SMART; Sporlastic GmbH) in IG AB. The brace works according to the 3-point principle and exerts pressure onto the unaffected condyle to correct the leg axis.

Thus, it is indicated for patients with unicondylar concerns only. The use of the brace while exercising was mandatory. However, patients were free to use the brace in everyday life as well. Participants were asked to document the wear time of the brace in a paper-and-pencil study diary. The brace was worn at the knee joint affected by osteoarthritis or at the signal joint in case of bilateral knee osteoarthritis. It was fitted by an orthopedic technician during the baseline examination.

CG Arm

Participants on the waiting list did not receive any study intervention or instruction for any change to their normal habits—"Just keep on like before." They were allowed to make use of usual care provided by the treating physician, if applicable. Usual care was defined as any kind of prescribed pharmacological or physical interventions a patient with knee osteoarthritis usually receives when consulting a medical doctor because of knee osteoarthritis. These may include physical therapies such as regular physiotherapy, manual therapy, electrotherapy as well as orthotic devices, and medical prescriptions for pharmacological agents such as nonsteroidal anti-inflammatory drugs (NSAIDs). These reflect the relevant treatment options according to the current national guidelines in Germany [7]. Moreover, participants in this group were informed about the opportunity to make use of the app after the follow-up assessment.

Outcomes

Data collection was conducted at baseline (t0) and at the 3-month follow-up (t3). Medical examinations and the outcome assessments of performance measures before and after the intervention (t3) took place on-site at the University Hospital Tübingen. Patient-reported outcome measures (PROMs) were assessed using web-based questionnaires (Questback GmbH). Questionnaires were activated on the days of assessment (t0 and t3), and study participants were asked to answer promptly. In case of delayed response, participants received a reminder via email.

Patient Characteristics

Age, gender, BMI, medical history (eg, relevant diagnoses and previous injuries or surgeries at the lower extremities or lower back), previous experiences with strengthening exercises or hip or knee exercise groups, and technical affinity were determined at baseline (t0).

Primary Outcome

The primary outcome measure was the joint comparison of the 2 app-based study arms (IG A and IG AB) versus the CG with regard to osteoarthritis-specific pain immediately after the 12-week intervention phase (t3). Pain was determined using the 10-item pain subscale of the KOOS [40,41]. The KOOS is a patient-reported outcome measurement instrument developed to assess the patient's opinion on their knee and associated problems and uses a 5-point Likert scale. It evaluates short-term and long-term consequences of knee injuries and primary osteoarthritis in 5 separately scored subscales. Each subscale is transformed to a scale of 0 to 100 points, with a higher score reflecting a better health status.

Secondary Outcomes

Overview of PROMs

Osteoarthritis-specific symptoms, physical function (activities of daily living), sport and recreation, and knee-related QoL were assessed using the other KOOS subscales. Patient-reported HRQoL was evaluated using the Veterans RAND 12-item Health Survey [42,43]. The Mental Component Score (MCS) and Physical Component Score (PCS) were calculated and used for further analysis. They both can adopt values in the range of 0 to 100 points. Higher scores indicate a better overall HRQoL. Patients' fear of movement was determined using the 11-item German version of the Tampa Scale of Kinesiophobia [44], with a scoring range of 6 to 24 whereby a higher score indicates a greater fear of movement. Physical and sports activity of a typical week, including frequency and time spent on transportation-related cycling and sports, fitness, or recreational activities, were quantified using the European Health Interview Survey–Physical Activity Questionnaire [45]. Exercise-specific self-efficacy was examined using the 9-item multidimensional Self-Efficacy for Exercise Scale [46], which ranges from 0 (not safe at all) to 10 (absolutely safe). The scale was used as a total score and then further divided into 3 subscales: task, coping, and scheduling. Higher scores indicate a higher exercise-specific self-efficacy. Control competence for physical exercising is a subcompetency of the physical activity–related health competence model. It relates to the perceived competence to individually structure and control physical activity in a health-effective way. It is mainly based on action-related knowledge but also requires the ability to sense and interpret body signals (eg, to adjust intensities based on muscle soreness) [47]. Control competence for physical exercising was quantified using 6 items according to Sudeck and Pfeifer [47] and 4 self-constructed items specifically focusing on exercises for the lower limbs [15]. Each item was scored on a 4-point Likert scale ranging from 1 (totally disagree) to 4 (totally agree). The mean value of all items was used for analysis, with higher scores reflecting a higher level of control competence.

Performance Measures

Performance measures included isometric maximum strength measurement of the knee extensors and knee flexors using DAVID strength machines (F200 Leg Extension and F300 Leg Curl; Schupp GmbH & Co. KG). Knee extensor strength was measured at 60° knee flexion, and knee flexion strength was measured at 30° knee flexion. Before testing, participants were instructed to conduct 5 to 8 dynamic concentric repetitions of the target movement at 50% to 60% of maximum force and 2 isometric repetitions at submaximal force in the given test position. Participants were instructed not to provoke an increase in pain level during testing. All measures were taken twice for each leg, and the highest value was used for analysis. Relative values (Newton meters per kilogram of body weight) were reported.

The 30-second chair stand test [48] is an instrument to measure leg strength endurance. Participants were seated with a straight back in the middle of a chair (seat height: 17 inches; participants with a knee angle of <90° received a pad to increase chair height) with hands and arms crossed in front of the upper body.

The feet were completely positioned on the floor. Participants were asked to stand up to full knee extension and then sit back again as many times as possible within 30 seconds. The total number of times the patient did come to a full standing position within the 30 seconds were counted. One complete movement execution was allowed before measurement.

Postural control tests were performed using a plantar pressure mat (zebris GmbH) to evaluate the course of the center of pressure in four different conditions: (1) bipedaled parallel stance with eyes open and (2) eyes closed, (3) bipedaled tandem stance with eyes open and the leg with knee osteoarthritis or signal joint in front and behind, and (4) one-legged stance with eyes open standing on the leg with knee osteoarthritis or signal joint. All tests were conducted in an upright position looking forward with both hands fixed at the superior iliac crest. After one test trial to become familiar with the procedure, conditions 1 and 2 were taken once, and conditions 3 and 4 were performed twice. The lowest value of each condition was used for analysis. The test duration was 10 seconds for conditions 1, 2, and 3 and 6 seconds for condition 4.

Adherence

Sensor- and app-based log files were read out for each exercise session separately and were used to quantify exercise adherence. Intervention finishers were defined as individuals who participated with >50% of overall exercise session adherence and were still active at weeks 11 and 12 of the intervention phase. Overall exercise session adherence was quantified by calculating the percentage of conducted exercise sessions relative to the overall number of prescribed exercise sessions irrespective of the adherence to the prescribed exercise dosage (number of sets and repetitions). Exercise repetition adherence was determined using the number of valid repetitions of all exercises of a session related to the prescribed repetitions. Percentage data were averaged across all exercise sessions (mean and SD). The active training time of an exercise session was calculated by adding all intervals between the time stamps of successive repetitions of an exercise unless the differences were of >60 seconds. If so, these data were not considered to exclude resting times within the active training time. The active training time was averaged across all exercise sessions (mean and SD) and further differentiated for weeks 1 to 6 and 7 to 12. The daily training time was the gross training time. It was calculated as the difference between the first and last repetition on a training day including all breaks, recalibrations, reviewing the exercise instructions, and the training of the other leg. For analysis of the active training time and gross training time, only cases with an exercise repetition adherence of 100% were considered. In addition, cases were excluded for which the gross training time was of >180 minutes as these cases indicate long breaks during the training or split training sessions and, thus, also possibly falsify pain and intensity data. To monitor perceived exercise intensity, participants were asked to rate their perceived overall exertion at the end of each exercise session. Perceived exertion was measured using the rate of perceived exertion scale from 0 (no exertion at all) to 10 (maximum conceivable exertion). Perceived pain before and after each exercise session was measured using the Faces Pain Scale [49]. Instead of numbers, 6 faces with different facial expressions were used to comment on perceived pain. Each face was associated with a textual reference for the pain intensity, and the faces were scored as 0 (no pain), 2 (little pain), 4 (moderate pain), 6 (much pain), 8 (very much pain), and 10 (highest imaginable pain). In the app, the 6 faces of the original version were replaced by standardized emojis of the iOS platform. For perceived exertion and pain analysis, only cases in which the training session was at least started (exercise repetition adherence of >0) were considered.

Concomitant Care

Concomitant pharmacological care (NSAIDs and analgesics) during the study phase was assessed retrospectively at t3.

Safety

Participants included in the study were asked to document all adverse events (AEs) that occurred during the study period. Participants in the IGs were further instructed to interrupt the training program in case of any suspicious symptoms, fatigue, or severe pain during exercising or wearing the knee brace. Mild AEs had to be reported to the responsible study staff within 1 week (via email, phone, or in-app support chat). AEs that required referral to a physician or other health care professional had to be reported immediately. The decision on how to proceed was up to the study physician as well as the study director (sports scientist and physiotherapist) and referred to the options of complete or temporary discontinuation or modification of the training regime. During data analysis, the reports were classified into AEs and serious AEs (SAEs; events related to death, life-threatening illness or injury, or inpatient hospitalization). They were further classified into expected and unexpected events, and the link to the intervention was differentiated into *sure*, *likely*, *possible*, *unlikely*, or *none*. Actions taken were classified into need or no need for immediate medical care (eg, referral to orthopedist, physiotherapy, or medication) and change in intervention modalities (eg, modification, pausing, stopping, or none).

Sample Size

Sample size was calculated based on an a priori power analysis (PASS 2020; NCSS, LLC). The sample size estimation was related to the primary outcome regarding the comparison of the 2 app-guided IGs (independently of the possible supplementary use of a knee brace in IG AB) versus the CG. Furthermore, the sample size estimation was based on the following assumptions: α level of.05, power of β=.8, and a correlation of pretest-posttest values of r=0.5 [50]. Standardized effect sizes (ESs) were used due to a lack of studies with comparable interventions and measures of dispersion. In this regard, required sample sizes to prove ESs between f=0.2 (equal to Cohen d=0.4) and f=0.4 (equal to Cohen d=0.8) were calculated with n=14 for f=0.2 and n=51 for f=0.4. Under the aforementioned assumptions and an expected dropout rate of approximately 15%, 30 participants should be recruited into each group (IG and CG) to verify a medium ES of f=0.2.

Randomization

Before study start, a randomization list was created using computer-generated random numbers (0 and 1) in 7 blocks with 10 slots each. Subsequently, sealed envelopes were prepared containing sequential numbers corresponding to the group

assignment resulting from the randomization list. Randomization into the IG and CG in a 1:1 ratio took place after baseline testing (t0) in order of appointment (eg, the first patient received the envelope with the first lot and so on). Participants allocated to the IG were then again randomized in a 1:1 ratio to 1 of the 2 intervention subgroups, IG A or IG AB, using the aforementioned procedure, including the prestudy preparation of the randomization list and sealed envelopes as outlined previously. The sealed envelopes were handed over by the study personnel. The randomization list and the sealed envelopes were prepared by a person not involved in the conduction, assessment, or data analysis of the study. Participants were not randomized in case of exclusion before completion of the baseline examination at t0.

Blinding

Participants and study personnel responsible for data collection and data analysis were not blinded to the group assignment or type of study intervention.

Statistical Analysis

Baseline characteristics of the IG (IG A and IG AB) and CG study groups at t0 were described using descriptive statistics, with continuous data being presented as mean and SD or median and IQR and categorical variables being presented as absolute numbers and percentages. Unpaired Student 2-tailed *t* tests or Mann-Whitney *U* tests (in case the normal distribution of the data were violated) and Pearson chi-square tests for categorical data were used to compare baseline characteristics in the IG and CG. The primary outcome was evaluated using an analysis of covariance (ANCOVA) with the dependent variable at t3 (KOOS pain subscale at t3), the fixed-effect group (IG and CG), and the covariate (KOOS pain subscale at t0) to adjust for baseline values. Secondary outcomes for the comparison of the IG and CG were handled accordingly. Mann-Whitney *U* tests using the difference t3–t0 of the variable of interest were calculated as a nonparametric alternative. Imputation of missing data was not foreseen. The α level was set to .05. Adjustments for multiple testing were applied for the tandem stance to account for 2 test conditions using Bonferroni correction (α=.025). Data were analyzed as randomized (intention to treat) following the complete case analysis approach. Sensitivity analyses regarding the evaluations of the 5 KOOS subscales (primary analysis and additional explorative analysis) using the last observation carried forward method and the mean imputation method were conducted to replace missing data and, thus, control for any possible bias due to missing values. In case of no significant deviation compared to the complete case analyses, the analyses were continued as initially intended. Between-group ESs were calculated according to Olejnik and Algina [51], with the differences of the adjusted postintervention values divided by the pooled SD at t3 (original data) and interpreted according to Cohen. Thereby, ESs of 0.2 to <0.5 were interpreted as small, ESs of 0.5 to <0.8 were interpreted as medium, and ESs of ≥0.8 were interpreted as large. Data were analyzed using the software packages Microsoft Excel (Microsoft Corp) and SPSS Statistics (IBM Corp).

Subgroup Analysis

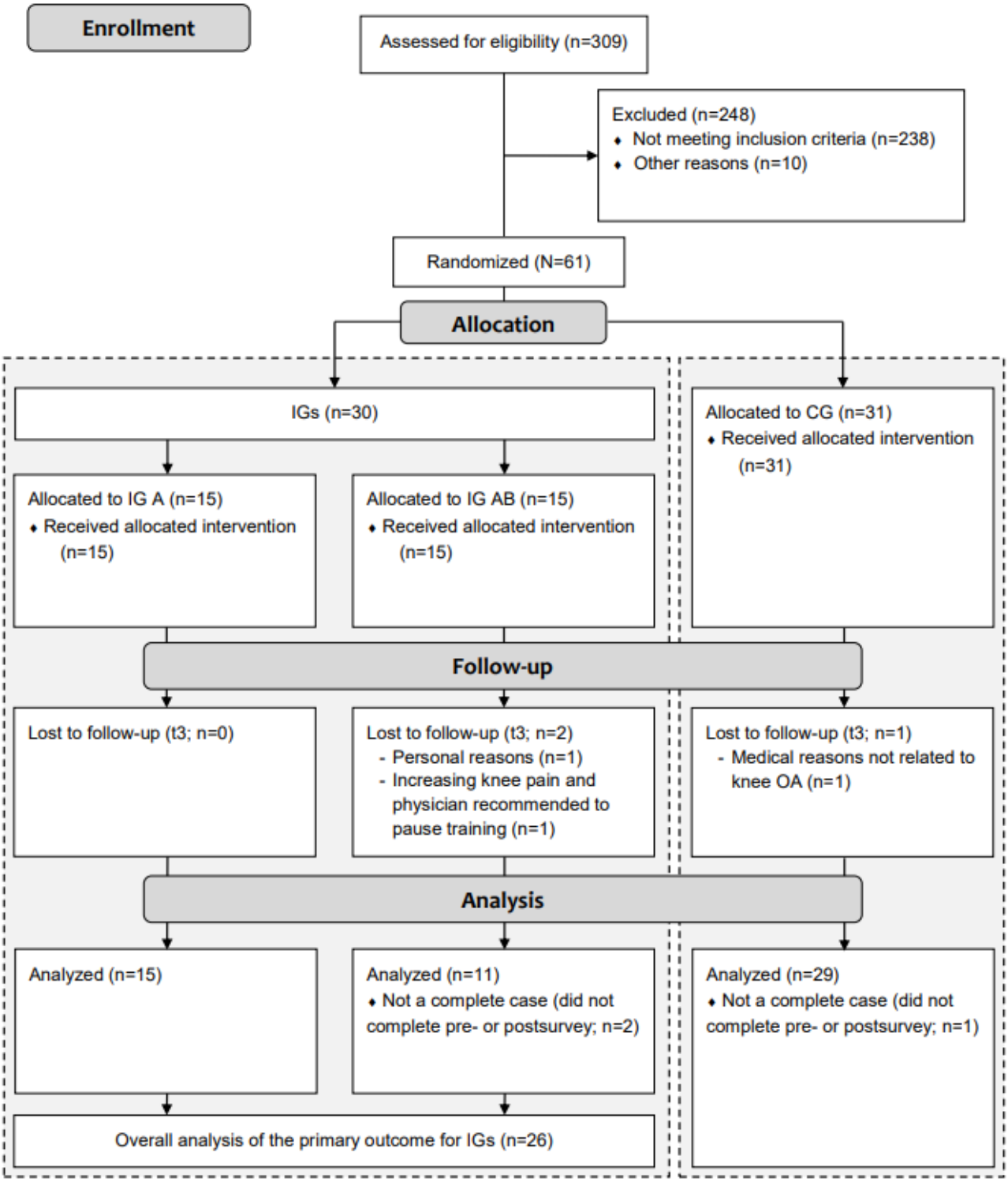
In addition to the aforementioned pooled comparison of the IG versus the CG, explorative separate subgroup comparisons of IG A and IG AB versus the CG as well as comparisons of IG A versus IG AB were conducted for the KOOS at t3. ANCOVA was used according to the procedure described previously.

Results

Participant Flow

Details on participant flow are outlined in Figure 4. Recruitment started on September 21, 2020. All patients were enrolled within 2 weeks, with the first patient included on October 12, 2020. The follow-up after the 12-week intervention period was completed on January 27, 2021. In total, 61 participants were randomized. Thereof, of the 61 participants, 30 (49%) were allocated to the IG (re.flex; n=15, 50% into IG A and IG AB each), and 31 (51%) were assigned to the CG (usual care). Loss to follow-up was 7% (2/30) for the IG and 3% (1/31) for the CG. In addition, 3 complete data sets were excluded from analysis (listwise case exclusion) due to surveys not completed at t0 or t3. Finally, 87% (26/30) of the participants from the IGs and 94% (29/31) of the participants from the CG were considered in the analysis of the primary outcome.

Figure 4. Participant flowchart. CG: control group; IG: intervention group; IG A: app-based training; IG AB: app-based training+brace; OA: osteoarthritis.



Baseline Data

Table 1 reports the sociodemographic and outcome-related baseline values of the participants. At t0, none of the variables showed a statistically significant difference between the IG and CG. Overall, the gender distribution of the participants was balanced (31/61, 51% male and 30/61, 49% female), mean age was 62.9 (SD 8.5) years, and mean BMI was 27.7 (SD 4.5) kg/m^2. However, it has to be noted that the number of male participants was higher in the IG and, on average, the participants in this group were also younger. Most of the participants (58/61, 95%) had never taken part in a hip or knee exercise group before. IG A and IG AB showed statistically significant differences at baseline for the one-legged stance of the signal joint (P=.02), with higher mean values for IG A.

Table 1. Baseline data for the complete case sample in total as well as differentiated according to group assignment (N=61).

Characteristic	Total (N=61)	IG[a] (n=30)	CG[b] (n=31)	*P* value[c]
Gender, n (%)				.16
Men	31 (51)	18 (60)	13 (42)	
Women	30 (49)	12 (40)	18 (58)	
Age (y), mean (SD)	62.9 (8.5)	61.5 (7.5)	64.2 (9.3)	.21
BMI (kg/m^2), mean (SD)	27.7 (4.5)	27.5 (5.0)	27.9 (4.2)	.75
Education, n (%)[d]				.34
Academic education	32 (53)	13 (45)	19 (61)	
Vocational education	27 (45)	15 (52)	12 (39)	
No vocational education	1 (2)	1 (3)	0 (0)	
Employment, n (%)[d]				.44
Employed	30 (50)	16 (55)	14 (45)	
Retired	30 (50)	13 (45)	17 (55)	
Previous experience with exercise therapy, n (%)				.51
Very high	5 (8)	2 (7)	3 (10)	
High	13 (21)	8 (27)	5 (16)	
Moderate	30 (49)	15 (50)	15 (48)	
Low	11 (18)	3 (10)	8 (26)	
Very low	2 (3)	2 (7)	0 (0)	
Previous participation in a hip or knee sports group, n (%)				.54
Yes	3 (5)	2 (7)	1 (3)	
No	58 (95)	28 (93)	30 (97)	
Technical affinity[e], mean (SD)	2.6 (0.6)	2.5 (0.7)	2.7 (0.6)	.25
KOOS[f], mean (SD)				
Pain	53.7 (15.9)	51.0 (15.8)	56.2 (15.8)	.22
Symptoms	56.7 (17.4)	54.3 (17.7)	58.9 (17.1)	.33
Physical function (ADLs[g])	70.4 (17.2)	68.9 (15.6)	71.8 (18.8)	.54
Sport and recreation	33.7 (21.2)	33.9 (21.5)	33.6 (21.3)	.97
QoL[h]	39.0 (15.2)	38.7 (15.7)	39.2 (15.0)	.90
Health-related QoL[i], mean (SD)				
PCS[j]	38.2 (9.3)	37.3 (9.0)	39.0 (9.7)	.51
MCS[k]	55.2 (9.3)	54.6 (10.9)	55.7 (7.9)	.67
Exercise-specific self-efficacy[l]				
Overall, median (IQR)[m]	8.4 (1.9)	8.7 (2.3)	8.3 (1.5)	.79
Task efficacy, median (IQR)[m]	8.3 (2.7)	8.5 (2.8)	8.3 (2.3)	.60
Coping efficacy, mean (SD)	7.7 (1.8)	7.7 (2.2)	7.7 (1.5)	.97
Scheduling efficacy, median (IQR)[m]	9.7 (1.7)	9.5 (2.1)	9.7 (1.2)	.20
Control competence[n], mean (SD)	3.0 (0.7)	3.1 (0.7)	3.0 (0.6)	.50
Fear of movement[o], mean (SD)	10.7 (3.7)	10.9 (3.9)	10.4 (3.5)	.62
Aerobic physical activity[m,p] (minutes per week), median (IQR)	300.0 (425.0)	275.0 (435.0)	345.0 (402.5)	.56

Characteristic	Total (N=61)	IG[a] (n=30)	CG[b] (n=31)	*P* value[c]
Isometric maximum force measurement[q]				
Knee extension (N m/kg), mean (SD)	1.2 (0.5)	1.3 (0.5)	1.1 (0.4)	.08
Knee flexion (N m/kg), median (IQR)[m]	1.02 (0.5)	1.1 (0.5)	1.0 (0.3)	.13
30-second chair stand (repetitions)[m,r], median (IQR)	10.0 (3.0)	10.0 (4.0)	10.0 (3.0)	.18
Postural control—COP[s] path (mm)				
Bipedaled parallel stance (eyes open), median (IQR)[m,t]	47.1 (29.1)	45.5 (31.4)	49.8 (39.5)	.10
Bipedaled parallel stance (eyes closed), median (IQR)[m,u]	86.3 (53.4)	88.5 (50.5)	83.5 (72.5)	.41
Bipedaled tandem stance with signal joint leg in front, mean (SD)[v]	245.6 (91.4)	235.7 (98.0)	256.2 (84.2)	.42
Bipedaled tandem stance with signal joint leg at the back, median (IQR)[m,w]	222.1 (120.1)	213.2 (110.4)	224.1 (128.4)	.64
One-legged stance of signal joint, mean (SD)[x]	184.4 (75.2)	182.5 (78.7)	186.4 (72.8)	.86

[a]IG: intervention group.
[b]CG: control group.
[c]The *P* value related to the comparison of the IG versus the CG.
[d]n=1 missing value.
[e]5-point Likert scale from 1 (*not true at all*) to 5 (*fully true*); n=6 missing values.
[f]KOOS: Knee Injury and Osteoarthritis Outcome Score. Scored from 0 to 100, with higher scores reflecting a better health status; n=6 missing values.
[g]ADL: activity of daily living.
[h]QoL: knee-related quality of life.
[i]Scored from 0 to 100, with higher scores reflecting a better health-related QoL; n=6 missing values.
[j]PCS: Physical Component Score.
[k]MCS: Mental Component Score.
[l]10-point scale from 0 (*not safe at all*) to 10 (*absolutely safe*); n=6 missing values.
[m]In case of nonparametric testing, median and IQR were reported.
[n]4-point Likert scale from 1 (*totally disagree*) to 4 (*totally agree*); n=6 missing values.
[o]Scored from 6 (*no fear*) to 24 (*extreme fear*); n=6 missing values.
[p]n=6 missing values.
[q]n=4 missing values.
[r]Number of counted repetitions; n=4 missing values.
[s]COP: center of pressure.
[t]n=5 missing values.
[u]n=4 missing values.
[v]n=7 missing values.
[w]n=8 missing values.
[x]n=11 missing values.

Primary Outcome

Table 2 and Figure 5 present the primary outcome, the KOOS pain subscale, at 3 months. ANCOVA showed a statistically significant between-group effect ($F_{1,\ 52}$=20.01; *P*<.001; ^{2}=0.278), with greater pain reduction for the IG compared to the CG. The baseline-adjusted mean difference was 13.2 points, demonstrating a medium effect in favor of the IG (ES=0.76).

Table 2. Primary and secondary outcome measures.

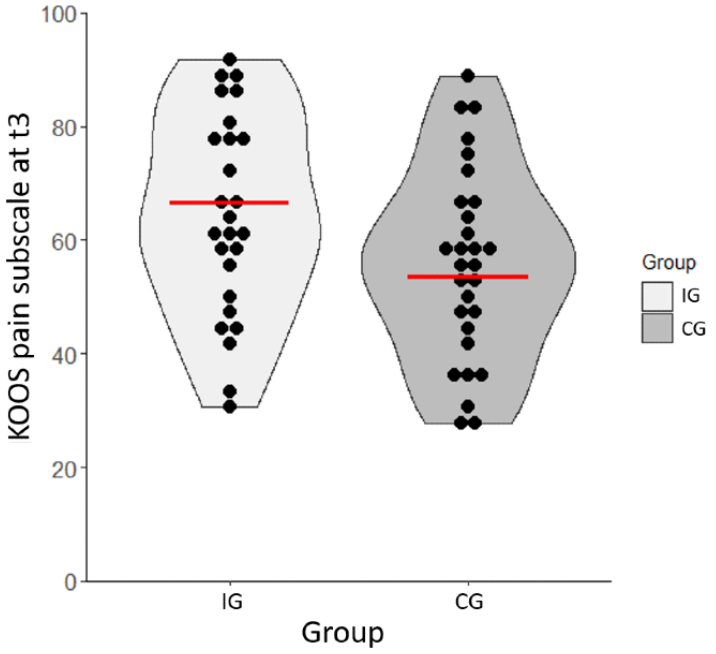
Outcome measure and group	Mean (SEM)		Mean difference (IG[a]–CG[b]; 95% CI)[c]	*P* value	ES[d]
	t0[e]	t3[f]			
Patient-reported outcome measures					
KOOS[g] (score of 0-100; worst to best)					
Pain subscale			13.2 (7.3 to 19.1)	<.001	0.76
IG (n=26)	51.0 (3.1)	66.7 (2.1)			
CG (n=29)	56.2 (2.9)	53.5 (2.0)			
Symptoms subscale			10.0 (2.4 to 17.5)	.01	0.53
IG (n=26)	54.3 (3.5)	65.1 (2.7)			
CG (n=29)	58.9 (3.2)	55.2 (2.6)			
Physical function (ADLs[h]) subscale			12.0 (5.9 to 18.1)	<.001	0.64
IG (n=26)	68.9 (3.1)	79.5 (2.2)			
CG (n=29)	71.8 (3.5)	67.5 (2.1)			
Sport and recreation subscale			10.7 (1.9 to 19.5)	.02	0.47
IG (n=26)	33.8 (4.2)	48.2 (3.2)			
CG (n=29)	33.6 (4.0)	37.5 (3.0)			
QoL[i] subscale			12.5 (6.8 to 18.1)	<.001	0.76
IG (n=26)	38.7 (3.1)	47.6 (2.0)			
CG (n=29)	39.2 (2.8)	35.1 (1.9)			
Health-related QoL (score of 0-100; worst to best)					
PCS[j]			6.0 (2.8 to 9.2)	<.001	0.74
IG (n=26)	37.3 (1.8)	44.0 (1.2)			
CG (n=29)	39.0 (1.8)	38.0 (1.1)			
MCS[k]			−2.6 (−6.2 to 1.0)	.15	—[l]
IG (n=26)	54.6 (2.1)	53.0 (1.3)			
CG (n=29)	55.7 (1.5)	55.6 (1.2)			
Exercise-specific self-efficacy (score of 0-10)					
Overall			0.1[m] (−0.8 to 1.0)	.44	—
IG (n=26)	8.1 (0.4)	7.7 (0.4)[n]			
CG (n=29)	8.2 (0.2)	7.6 (0.3)[n]			
Task efficacy			0.8[m] (−0.2 to 1.8)	.13	—
IG (n=26)	8.1 (0.4)	8.1 (0.4)[n]			
CG (n=29)	7.8 (0.3)	7.3 (0.3)[n]			
Coping efficacy			0.4 (−0.3 to 1.2)	.24	—
IG (n=26)	7.7 (0.4)	7.2 (0.3)			
CG (n=29)	7.7 (0.3)	6.7 (0.3)			
Scheduling efficacy			−0.9[m] (−1.8 to 0.0)	.42	—
IG (n=26)	8.6 (0.4)	7.8 (0.4)[n]			
CG (n=29)	9.2 (0.2)	8.7 (0.2)[n]			

Outcome measure and group	Mean (SEM)		Mean difference (IG[a]–CG[b]; 95% CI)[c]	*P* value	ES[d]
	t0[e]	t3[f]			
Control competence (score of 1-4)			0.2 (−0.0 to 0.3)	.09	—
IG (n=26)	3.1 (0.1)	3.1 (0.1)			
CG (n=29)	3.0 (0.1)	2.9 (0.1)			
Fear of movement (score of 6-24)			−1.6 (−3.3 to 0.1)	.06	—
IG (n=26)	10.9 (0.8)	9.9 (0.6)			
CG (n=29)	10.4 (0.7)	11.5 (0.6)			
Aerobic physical activity (min/wk)			150.5[m] (13.7 to 287.4)	.28	—
IG (n=26)	451.2 (97.6)	388.3 (62.9)[n]			
CG (n=29)	382.2 (66.2)	237.8 (31.6)[n]			
Performance measures					
Muscle strength					
Isometric maximum force—knee extension (N m/kg)			0.0 (−0.1 to 0.2)	.59	—
IG (n=28)	1.3 (0.1)	1.3 (0.1)			
CG (n=29)	1.1 (0.1)	1.3 (0.1)			
Isometric maximum force—knee flexion (N m/kg)			0.2[m] (−0.1 to 0.5)	.40	—
IG (n=28)	1.2 (0.1)	1.2 (0.1)[n]			
CG (n=29)	1.0 (0.1)	1.0 (0.1)[n]			
30-second chair stand test (repetitions)			1.8[m] (0.2 to 3.4)	.23	—
IG (n=28)	10.6 (0.6)	12.3 (0.5)[n]			
CG (n=29)	9.6 (0.5)	10.5 (0.6)[n]			
Postural control					
Bipedaled parallel stance (eyes open; COP[o] path in mm)			2.2[m] (−11.0 to 15.4)	.08	—
IG (n=28)	43.3 (3.3)	57.3 (4.9)[n]			
CG (n=28)	52.3 (4.3)	55.1 (4.4)[n]			
Bipedaled parallel stance (eyes closed; COP path in mm)			−23.2[m] (−53.4 to 7.0)	.22	—
IG (n=28)	89.4 (8.0)	92.2 (9.8)[n]			
CG (n=29)	101.9 (12.5)	115.4 (11.4)[n]			
Bipedaled tandem stance with signal joint in front (COP path in mm)			11.0 (−30.6 to 52.6)	.60[p]	—
IG (n=28)	235.7 (18.5)	263.4 (14.3)			
CG (n=26)	256.2 (16.5)	252.4 (14.9)			
Bipedaled tandem stance with signal joint at the back (COP path in mm)			49.8[m] (−100.9 to 1.3)	.048[p]	—
IG (n=27)	236.6 (18.1)	218.5 (14.1)[n]			
CG (n=26)	248.4 (16.9)	268.3 (21.4)[n]			
One-legged stance of signal joint (COP path in mm)			−6.5 (−30.1 to 17.1)	.58	—
IG (n=26)	182.5 (15.4)	177.1 (8.1)			

Outcome measure and group	Mean (SEM)		Mean difference (IG[a]–CG[b]; 95% CI)[c]	*P* value	ES[d]
	t0[e]	t3[f]			
CG (n=24)	186.4 (14.9)	183.6 (8.5)			

[a]IG: intervention group.
[b]CG: control group.
[c]Reporting the baseline-adjusted means.
[d]ES: effect size; only calculated for significant results.
[e]t0: baseline.
[f]t3: 12 weeks after baseline.
[g]KOOS: Knee Injury and Osteoarthritis Outcome Score.
[h]ADL: activity of daily living.
[i]QoL: quality of life.
[j]PCS: Physical Component Score.
[k]MCS: Mental Component Score.
[l]Not applicable.
[m]Mann-Whitney *U* test comparing the within-group differences t3–t0.
[n]Reporting unadjusted means; prerequisites for analysis of covariance not given.
[o]COP: center of pressure.
[p]Adjusted for multiple testing (*P*<.03).

Figure 5. Knee Injury and Osteoarthritis Outcome Score (KOOS) pain subscale at 3 months (t3) of the intervention group (IG) and control group (CG). Violin plots display data distribution (tails are trimmed to the range of the data) and baseline-adjusted means (red line).



Secondary Outcomes

PROMs and Performance Measures

Secondary outcome measures at 3 months are outlined in Table 2. At t3, statistically significant differences between the IG and the CG were observed for all additional KOOS subscales in favor of the IG (symptoms: $F_{1, 52}$=7.01, *P*=.01, and η^2=0.119; physical function [activities of daily living]: $F_{1, 52}$=15.56, *P*<.001, and η^2=0.230; sport and recreation: $F_{1, 52}$=5.98, *P*=.02, and η^2=0.103; QoL: $F_{1, 52}$=19.87, *P*<.001, and η^2=0.277), as well as for the HRQoL PCS ($F_{1, 52}$=13.94; *P*<.001; ^{2}=0.211). Baseline-adjusted mean differences between the IG and CG were between 10.0 and 12.5 points for the KOOS subscales and 6.0 points for the PCS. Interpreted according to the Cohen *d*, the intervention showed a medium treatment effect for improvements in patients' symptoms, physical function, knee-related QoL, and the physical component of the HRQoL and a small treatment effect for sport and recreation activities. No statistically significant differences between groups were observed for any other secondary patient-reported or performance-related outcome measures.

Subgroup Analyses

The exploratory subgroup analyses between IG A and IG AB demonstrated baseline-adjusted mean differences ranging from 4.7 to 12.1 points for the 5 KOOS subscales in favor of the IG AB group. However, between-group differences were not statistically significant. Subgroup analyses for IG A and IG AB separately versus the CG showed superiority of IG AB versus the CG in all KOOS subscales as well as superiority of the IG A versus the CG for pain, physical function, and QoL. Table S1 in Multimedia Appendix 2 provides more details.

Adherence

Of the 30 participants in the IG, 27 (90%) were defined as intervention finishers. During the intervention phase, one participant dropped out due to personal reasons, and another dropped out because of non–device-related increasing knee pain and a subsequent physician-recommended pausing of the training intervention. One participant ceased the intervention from week 8 onward for unknown reasons. Overall, exercise session adherence for intervention finishers was 92.5% (899/972), indicating the percentage of training sessions that were at least started. In total, 7.5% (73/972) of all training sessions were not performed (adherence=0%). The overall mean exercise repetition adherence was 86.8% (SD 28.9%). The average active training time and gross training time in weeks 1 to 12 (total) was 15.5 (SD 4.1) minutes and 31.7 (SD 17.1) minutes, respectively. On average, participants reported a perceived pain score of 0.9 (SD 1.4) points before and 1.1 (SD 1.5) points after the exercise sessions. This was according to the Faces Pain Scale, equivalent to a slight traction or mild pain. The mean difference from after to before exercise sessions was 0.2 points. After the exercise sessions, the participants reported an average intensity value of 3.5 (SD 1.4) points. According to the rate of perceived exertion scale, this indicates a slight to somewhat exhausting load. Further details on adherence, training time, exertion, and pain outcomes are outlined in Tables 3 and 4.

Table 3. Adherence and training time.

Outcome measure	Weeks 1-12	Weeks 1-6	Weeks 7-12
Overall exercise session adherence (%)[a]	92.5	93.0	92.0
Exercise repetition adherence (%), mean (SD)[a]	86.8 (28.9)	87.6 (28.0)	85.9 (29.8)
Active training time (min), mean (SD)	15.5 (4.1)	17.6 (3.3)[b]	13.3 (3.7)[c]
Gross training time (min), mean (SD)	31.7 (17.1)	36.2 (19.3)[b]	27.3 (13.2)[c]

[a]Data refer to 972 exercise sessions (12 weeks with 3 sessions per week; n=27 [n=3 no intervention finishers]).

[b]Data out of 360 exercise sessions.

[c]Data out of 364 exercise sessions.

Table 4. Perceived exertion and pain outcomes.

Outcome measure	Values, mean (SD)	Values, median (IQR)	Range
Perceived exercise intensity after exercise sessions, mean (SD)[a,b]	3.5 (1.4)	3.0 (1.0)	3.5-9.0
Perceived pain before exercise sessions, mean (SD)[b,c]	0.9 (1.4)	0.0 (2.0)	0.9-6.0
Perceived pain after exercise sessions, mean (SD)[b,c]	1.1 (1.5)	0.0 (2.0)	1.1-6.0

[a]10-point scale from 0 (*no exertion at all*) to 10 (*maximum conceivable exertion*).

[b]Data out of 888 exercise sessions.

[c]10-point scale from 0 (*no pain*) to 10 (*highest imaginable pain*).

Concomitant Care

Concomitant pharmacological care data during the study phase were available for 93% (28/30) of the IG participants and 97% (30/31) of the CG participants. Overall, at t3, a total of 14% (4/28) of the IG participants compared to 10% (3/30) of the CG participants reported a daily intake of NSAIDs or analgesics during the study phase, and 7% (2/28) of the IG participants compared to 7% (2/30) of the CG participants reported a weekly intake of NSAIDs or analgesics during the study phase.

Safety and Technical Issues

No SAEs were reported throughout the intervention phase to the study personnel. In summary, 7 AEs were reported. In total, 4 of the AEs were sure to be intervention related (AEs number 11, 21, 40, and 59), of which 3 (75%) required a modification of the training (AEs number 11, 40, and 59) and 1 (25%) required pausing the training intervention (AE number 21). Medical care was necessary in one case (AE number 40). The remaining 43% (3/7) of the AEs were not intervention related (AEs number 29, 44, and 49). Further details are outlined in Table 5.

Table 5. Adverse events (AEs) throughout the study period.

ID	Group	Harm	Type	Expectation	Link to intervention	MC[a]	CoI[b]
11	IG[c]	Increased pain due to device-initiated overload (repetition count failure and range of motion)	AE	UE[d]	Sure	No	Modification
21	IG	Increased pain and feeling of permanent muscle soreness in legs and arms	AE	EE[e]	Sure	No	Pausing
29	IG	Training interruption due to lumbago	AE	—[f]	No	No	Pausing
40	IG	Increased pain (especially for standing exercises on one leg) and activated OA[g]	AE	UE	Sure	Yes	Modification
44 (DO[h])	CG[i]	Dizziness and personal health problems	AE	—	No	Yes	None
49 (DO)	IG	Fall on knee and subsequently irritated and overloaded knee	AE	—	No	Yes	Stopping
59	IG	Increased pain and knee joint unusually warm	AE	UE	Sure	No	Modification

[a]MC: medical care; need for immediate medical care (yes) or no need for medical care (no).

[b]CoI: change of intervention.

[c]IG: intervention group.

[d]UE: unexpected event.

[e]EE: expected event.

[f]Not applicable or no link.

[g]OA: osteoarthritis.

[h]DO: dropout.

[i]CG: control group.

Technical issues reported by participants when using the app were used for minor technical bug fixes during the study period as well as for adjustments after the study period. In the following paragraph, only the summarized incident report of a participant on a specific topic is considered. On the one hand, bug fixes related to failures of the movement sensors, which, in some cases, did not adequately recognize patients' movements and, thus, led to incorrect counting of exercise repetitions. This was the case for the following exercises: wall slide (n=12), knee extension exercises (seated and supine position; n=7), and hip abduction exercises (seated and standing position; n=5). Other problems included the need for multiple recalibrations during an exercise session (n=17), an incorrect representation of the training leg on the app (twisted and no reaction; n=12), and initial problems connecting the sensors (n=7).

Discussion

Principal Findings

Overview

This pilot study aimed to evaluate the efficacy of a 12-week app-based exercise intervention with or without an additional knee brace on symptoms and function in participants with knee osteoarthritis. The results of the study demonstrated small to medium treatment effects with statistically significant reductions in self-reported osteoarthritis-related pain (primary outcome) and other osteoarthritis-specific concerns (KOOS symptoms, physical function, sports and recreation, and QoL subscales), as well as an increase in the PCS of the general HRQoL after the 12-week app training versus the CG. No intervention effects were found for any other of the secondary outcomes. The intervention showed an excellent adherence rate and no SAEs.

Overview of PROMs

Previous studies have shown superiority of exercise interventions guided by fully automated mobile apps versus control to reduce knee-related pain and improve physical function in patients with knee osteoarthritis [22,23,52]. However, in contrast to our results indicating medium ESs of >0.6 for pain and physical function, Bossen et al [52], who focused their intervention primarily on increasing general physical activity, demonstrated much smaller effects of 0.2 for pain and physical function. Regarding the absolute differences between the baseline-adjusted postmeasures of the IG and CG, substantially higher between-group differences (10.0 to 13.2 points) were reported in this study in comparison to other studies with digital interventions (2.9 to 7.7 points), of which only Mecklenburg et al [23] used a similar sensor-assisted and app-based exercise program [22,23,52]. Uesugi et al [53] did not find a significant between-group effect at all.

From a clinical perspective, within-group differences for pain in the IG of our study exceeded the minimal clinically important improvement (MCII) of 8.7 points/100 as reported for patients with knee osteoarthritis who underwent a 12-week rehabilitation intervention with active and passive therapeutic treatments. Reported cutoff values for physical function (13.4 points/100) could not be reached [54]. However, regarding minimal clinically important differences (MCIDs) between the IG and CG, the ESs of our study are within or above reported thresholds for MCIDs [55]. Nevertheless, there are many different calculation methods for determining MCII and MCID, and these values may also differ between population groups and

interventions, leading to a lack of consensus on which cutoff values should be used. Therefore, future studies should apply an own anchor-based approach to be able to define intervention-related MCID and MCII values for pain and physical function.

At present, nondigital interactions are the gold standard for exercise guidance, and novel interventions should not only be compared between each other but also in reference to the standards. Studies on nondigital interactions have reported small to medium effects in terms of pain reduction and improvement in physical function for delivery modes such as one-on-one treatments, class-based programs, and home-based exercises [9,56]. Verhagen et al [57] recently stated that the estimated effects regarding pain reduction of supervised exercising are very robust and no further intervention studies in this domain would change these findings. This is in contrast to cutting-edge results of an individual patient data meta-analysis on exercise therapy in knee and hip osteoarthritis questioning the clinical importance of reported effects versus those of a CG [58]. Accordingly, it seems reasonable to expand the field of application and investigate alternative, innovative ways of delivering exercise therapy that may further increase our knowledge of the effectiveness and mechanisms of action of new treatment delivery opportunities. Considering this, the app-based instruction appears to be a promising evolution of the currently proven standard therapy, especially as the results of this study indicate clinically important ESs. Another possible reason for the larger effects reported in our study may be related to the fact that we only included patients with at least moderate knee osteoarthritis symptoms. This reduced the potential risk of ceiling effects as the possible range of improvements for patients with early or mild disease-specific symptoms was much lower and baseline pain has been described as a moderator for treatment effectiveness [58].

The exploratory subgroup analyses for the KOOS subscales showed superiority of IG AB versus the CG in all KOOS subscales as well as superiority of IG A versus the CG for pain, physical function, and QoL. The direct comparison of both IGs provided a first indication of the superiority of IG AB versus IG A. However, these findings were not statistically significant. This potential trend of superiority of IG AB must be considered with caution due to the small subgroup sample size and should be verified in a subsequent data analysis with a larger sample.

Benefits of exercise therapy on mental and physical HRQoL in patients with knee osteoarthritis have also been reported in a recent meta-analysis with reported standardized mean differences of 0.52 for the PCS and 0.44 for the MCS [59]. Compared with these values, this study showed even greater improvements for the PCS with an ES of 0.74, yet no improvements in the MCS were observed. However, baseline scores for the MCS already exceeded the US population norm of 50 points [60].

When looking at health-psychological measures, participants in the IG had less fear of movement after the intervention in comparison to participants in the CG, although this finding failed to reach statistical significance. Mean values at baseline already indicated a low level of fear of movement in the population under study. As most participants were recruited via newspapers or newsletters, it seems reasonable that only patients who did not have fear of activity-induced worsening of symptoms would have applied to participate. We also investigated whether the self-efficacy and control competence subcompetencies of the physical activity–related health competence model improved after taking part in the stand-alone mHealth intervention. However, no differences between the IG and CG were observed. In the study population, both measures that included the subscales showed ceiling effects at baseline. Thus, the possibility of change after the intervention phase was limited.

Performance Measures

Strength endurance quantified using the 30-second chair stand test increased by 1.7 repetitions for the IG versus the CG. However, this finding was not statistically significant. Therefore, the results are in contrast to those of a study on the effectiveness of a 6-week internet-based exercise intervention against usual care in patients with knee osteoarthritis reporting a statistically significant between-group effect in favor of the intervention. The participants in the IG improved by an average of 4.5 repetitions. The between-group difference after the intervention was 3.4 repetitions [61].

Change in knee extension strength has been described as a mediator for clinical benefit in patients with knee and hip osteoarthritis [62]. Our results showed no improvement in maximum knee extension strength in the IG. This is in contrast to a meta-analysis including results of 10 studies in which low-intensity resistance training reported short-term ESs with a standardized mean difference of 0.5 when compared to the CG. However, there was also a small group of studies that failed to show significant benefits for strength outcomes. The authors of the meta-analysis [63] hypothesized that one reason for the absence of benefit may be related to the low intensity of these exercise programs as too low intensities cannot trigger sufficient muscle activity to promote neuromotor adaptations and hypertrophy to ultimately generate muscle gains [63]. This could also explain the lack of strength improvements in our study as the analysis of the sensor and app log files after the training sessions revealed only a perceived “slight” to “somewhat exhausting” intensity level, and thus, this tends to fall in the subthreshold exercise dose. In the future, patients should be better educated to enable them to independently adjust their training intensity (eg, by choosing a heavier or lighter exercise variation). At the end of an exercise set, a training-effective exercise load in the range of “strenuous” to “very strenuous” should be achieved.

We did not find consistent superiority of the IG versus the CG for postural control outcomes, and we refrain from discussing this further as measurement instruments, postures, and durations differ across trials and the measures used in our study are not part of the recommended set of performance-based measures to assess physical function in people diagnosed with knee and hip osteoarthritis [64].

Adherence

The success of exercise therapy and the associated improvements in pain, physical function, and QoL are highly dependent on the maximization of the adherence to exercise. Therefore, it is recommended to supervise exercise sessions at least in the initial exercise period to enhance adherence before continuing exercise independently [65]. Various factors (eg, motivational level, physical status, personal goals, self-regulation, and several extrinsic factors) can impact exercise adherence [47,66]. To overcome these barriers, especially for nonsupervised training, the field of digital-based exercising may offer a different approach for guidance and motivation as well as an independence from time and location. The particular group of sensor-based apps can additionally supervise the training execution and provide real-time feedback. The impact of supporting features for barrier management in eHealth and mHealth technologies may also be the reason for the high adherence rates of up to 82% to 91% for digital-based home exercise interventions (all sources, not only sensor based) in patients with knee osteoarthritis [67,68]. These numbers correspond to the adherence rate of 92.5% (899/972) in our study and exceed reported rates of 62% to 75% for nondigital home exercise interventions [69,70]. The results of our study only refer to the short term. As adherence to exercise is critical for the long-term benefit of lifestyle interventions [13], future studies may investigate whether better adherence to app-guided interventions can be sustained over a longer period as well.

Limitations

One limitation of our study is related to potential bias due to missing values of the complete case analyses. According to Jakobsen et al [71], a complete case analysis can be applied up to a threshold of approximately 5% missing values. They further report that missing data can be ignored in the analysis if the impact of missing data on the results is negligible. In our study 9.8% of data were missing but no significant differences were observed for the complete case analyses in comparison to the sensitivity analyses (data not shown) for the KOOS subscales. Another limitation is the lack of blinding of participants because of their obvious group assignment. This may have particularly influenced subjective outcome measures in the CG due to a lack of treatment expectations. A major limitation of the study is related to the sample size of the 2 subgroups, IG A and IG AB. Our study was powered for a joint comparison of IG A and IG AB versus the CG. Therefore, subgroup comparisons of IG A and IG AB versus the CG may lack statistical power and generalizability. Due to the relatively small subgroup sample size, additional data are needed to substantiate or revise the findings of a possible additional treatment effect of wearing a knee brace during exercising.

It should also be mentioned that the study design does not allow for clarification on whether the favorable study results of the IGs were the result of app use or just of the fact that participants exercised more than those in the CG. However, the aim of this trial was not to conduct a comparison of different delivery modes of exercise (eg, supervised in person vs stand-alone app) but to obtain first insights into the efficacy of a sensor-based mHealth intervention in comparison to usual care. Further comparative studies are needed to answer the question of which patients respond best to which type of delivery.

Conclusions

Individuals with knee osteoarthritis undergoing a 12-week sensor-assisted app-based exercise intervention program with or without an additional knee brace experienced positive treatment effects with medical benefits regarding pain relief and improvements in physical function as well as other osteoarthritis-specific concerns compared to those in the CG. Adherence to the exercise intervention was high, and the mobile app can be classified as a safe intervention, with no SAEs being reported. To overcome limitations in the generalizability of the results because of the rather small sample size and the joint comparison of IG A and IG AB, a well-powered trial on the effectiveness of re.flex versus a CG is currently being conducted.

Acknowledgments

The authors want to thank all the study participants. They also highly appreciate the involvement of Georg Haupt, who made a significant contribution to the design of the intervention. They also thank scientific assistants Patrick Riedt, Johannes Goller, Milena Hartwig, and colleague Isabel Schubert for their help with data collection and analysis. The study was funded by Sporlastic GmbH. The funding source was not involved in data collection, data analysis, or the report of the study. Sporlastic GmbH provided the digital app and sensors that were used in the clinical trial. Sporlastic GmbH has a contractional relationship with Kineto Tech Rehab SRL. Kineto Tech Rehab SRL is the manufacturer of the product (sensors and app). Sporlastic GmbH is the exclusive distribution partner for the contract area (Germany). The authors acknowledge support from the Open Access Publication Fund of the University of Tübingen and the company Sporlastic GmbH for the funding of the remaining publication fee. The manuscript was submitted to Sporlastic GmbH for review. Sporlastic GmbH approved the manuscript without any request for modification. The University Hospital Tübingen has a basic right of publication of study data. The study protocol and results will be communicated at scientific congresses and in peer-reviewed scientific journals. Sporlastic GmbH has the right to make the study results public as part of corporate communications.

Authors' Contributions

All authors contributed to the conception and design of the study. IK and PJ were responsible for the acquisition of funding. VD and PJ contributed to the collection of data. PJ was the responsible orthopedist. VD and IK were responsible for data management, data analysis, and interpretation of the data and drafting and revising the paper. All authors read and approved the final manuscript.

Conflicts of Interest

This study was an investigator-initiated trial in cooperation with an industrial partner, the financial supporter of the study. The authors declare no further potential conflict of interest.

Multimedia Appendix 1
Re.flex manual.
[PDF File (Adobe PDF File), 48606 KB - mhealth_v12i1e54356_app1.pdf]

Multimedia Appendix 2
Subgroup analyses for the Knee Injury and Osteoarthritis Outcome Score.
[PDF File (Adobe PDF File), 171 KB - mhealth_v12i1e54356_app2.pdf]

Multimedia Appendix 3
CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) checklist (V 1.6.1).
[PDF File (Adobe PDF File), 68132 KB - mhealth_v12i1e54356_app3.pdf]

References

1. Cross M, Smith E, Hoy D, Nolte S, Ackerman I, Fransen M, et al. The global burden of hip and knee osteoarthritis: estimates from the Global Burden of Disease 2010 study. Ann Rheum Dis 2014 Jul 19;73(7):1323-1330. [doi: 10.1136/annrheumdis-2013-204763] [Medline: 24553908]
2. Vos T, Flaxman AD, Naghavi M, Lozano R, Michaud C, Ezzati M, et al. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 2012 Dec 15;380(9859):2163-2196 [FREE Full text] [doi: 10.1016/S0140-6736(12)61729-2] [Medline: 23245607]
3. Fuchs J, Rabenberg M, Scheidt-Nave C. Prävalenz ausgewählter muskuloskelettaler Erkrankungen. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2013 May 25;56(5-6):678-686. [doi: 10.1007/s00103-013-1687-4] [Medline: 23703486]
4. Zhang W, Doherty M, Peat G, Bierma-Zeinstra MA, Arden NK, Bresnihan B, et al. EULAR evidence-based recommendations for the diagnosis of knee osteoarthritis. Ann Rheum Dis 2010 Mar 17;69(3):483-489. [doi: 10.1136/ard.2009.113100] [Medline: 19762361]
5. Fernandes L, Hagen KB, Bijlsma JW, Andreassen O, Christensen P, Conaghan PG, European League Against Rheumatism (EULAR). EULAR recommendations for the non-pharmacological core management of hip and knee osteoarthritis. Ann Rheum Dis 2013 Jul;72(7):1125-1135 [FREE Full text] [doi: 10.1136/annrheumdis-2012-202745] [Medline: 23595142]
6. McAlindon TE, Bannuru RR, Sullivan MC, Arden NK, Berenbaum F, Bierma-Zeinstra S, et al. OARSI guidelines for the non-surgical management of knee osteoarthritis. Osteoarthritis Cartilage 2014 Mar;22(3):363-388 [FREE Full text] [doi: 10.1016/j.joca.2014.01.003] [Medline: 24462672]
7. Stöve J. S2k-Leitlinie gonarthrose. The Association of Scientific Medical Societies in Germany. 2018. URL: https://register.awmf.org/de/leitlinien/detail/187-050 [accessed 2024-04-29]
8. Bannuru RR, Osani MC, Vaysbrot EE, Arden NK, Bennell K, Bierma-Zeinstra S, et al. OARSI guidelines for the non-surgical management of knee, hip, and polyarticular osteoarthritis. Osteoarthritis Cartilage 2019 Nov;27(11):1578-1589 [FREE Full text] [doi: 10.1016/j.joca.2019.06.011] [Medline: 31278997]
9. Fransen M, McConnell S, Harmer AR, Van der Esch M, Simic M, Bennell KL. Exercise for osteoarthritis of the knee: a Cochrane systematic review. Br J Sports Med 2015 Dec 24;49(24):1554-1557. [doi: 10.1136/bjsports-2015-095424] [Medline: 26405113]
10. Kolasinski SL, Neogi T, Hochberg MC, Oatis C, Guyatt G, Block J, et al. 2019 American College of Rheumatology/Arthritis Foundation guideline for the Management of Osteoarthritis of the Hand, Hip, and Knee. Arthritis Rheumatol 2020 Feb 06;72(2):220-233 [FREE Full text] [doi: 10.1002/art.41142] [Medline: 31908163]
11. Jacobs H, Callhoff J, Albrecht K, Postler A, Saam J, Lange T, et al. Use of physical therapy in patients with osteoarthritis in Germany: an analysis of a linkage of claims and survey data. Arthritis Care Res (Hoboken) 2021 Jul 28;73(7):1013-1022. [doi: 10.1002/acr.24365] [Medline: 32609432]
12. Hagen KB, Smedslund G, Østerås N, Jamtvedt G. Quality of community-based osteoarthritis care: a systematic review and meta-analysis. Arthritis Care Res (Hoboken) 2016 Oct 16;68(10):1443-1452 [FREE Full text] [doi: 10.1002/acr.22891] [Medline: 26991934]
13. Bennell KL, Dobson F, Hinman RS. Exercise in osteoarthritis: moving from prescription to adherence. Best Pract Res Clin Rheumatol 2014 Feb;28(1):93-117. [doi: 10.1016/j.berh.2014.01.009] [Medline: 24792947]
14. Biebl JT, Huber S, Rykala M, Kraft E, Lorenz A. Attitudes and expectations of health care professionals toward app-based therapy in patients with osteoarthritis of the Hip or Knee: questionnaire study. JMIR Mhealth Uhealth 2020 Oct 28;8(10):e21704 [FREE Full text] [doi: 10.2196/21704] [Medline: 33112255]

15. Durst J, Roesel I, Sudeck G, Sassenberg K, Krauss I. Comparison of human versus digital instructions for exercise in patients with hip osteoarthritis: results of a non-inferiority randomized cross-over trial. Osteoarthritis Cartilage 2020 Apr;28:S463. [doi: 10.1016/j.joca.2020.02.725]
16. Bossen D, Veenhof C, Dekker J, de Bakker D. The usability and preliminary effectiveness of a web-based physical activity intervention in patients with knee and/or hip osteoarthritis. BMC Med Inform Decis Mak 2013 May 28;13(1):61 [FREE Full text] [doi: 10.1186/1472-6947-13-61] [Medline: 23714120]
17. Barak A, Klein B, Proudfoot JG. Defining internet-supported therapeutic interventions. Ann Behav Med 2009 Aug 29;38(1):4-17. [doi: 10.1007/s12160-009-9130-7] [Medline: 19787305]
18. Dennison L, Morrison L, Conway G, Yardley L. Opportunities and challenges for smartphone applications in supporting health behavior change: qualitative study. J Med Internet Res 2013 Apr 18;15(4):e86 [FREE Full text] [doi: 10.2196/jmir.2583] [Medline: 23598614]
19. Chen T, Or CK, Chen J. Effects of technology-supported exercise programs on the knee pain, physical function, and quality of life of individuals with knee osteoarthritis and/or chronic knee pain: a systematic review and meta-analysis of randomized controlled trials. J Am Med Inform Assoc 2021 Feb 15;28(2):414-423 [FREE Full text] [doi: 10.1093/jamia/ocaa282] [Medline: 33236109]
20. Jönsson T, Dell'Isola A, Lohmander LS, Wagner P, Cronström A. Comparison of face-to-face vs digital delivery of an osteoarthritis treatment program for hip or knee osteoarthritis. JAMA Netw Open 2022 Nov 01;5(11):e2240126 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.40126] [Medline: 36326763]
21. Weber F, Müller C, Bahns C, Kopkow C, Färber F, Gellert P, et al. Smartphone-assisted training with education for patients with hip and/or knee osteoarthritis (SmArt-E): study protocol for a multicentre pragmatic randomized controlled trial. BMC Musculoskelet Disord 2023 Mar 23;24(1):221 [FREE Full text] [doi: 10.1186/s12891-023-06255-7] [Medline: 36959595]
22. Pelle T, Bevers K, van der Palen J, van den Hoogen FH, van den Ende CH. Effect of the dr. Bart application on healthcare use and clinical outcomes in people with osteoarthritis of the knee and/or hip in the Netherlands; a randomized controlled trial. Osteoarthritis Cartilage 2020 Apr;28(4):418-427 [FREE Full text] [doi: 10.1016/j.joca.2020.02.831] [Medline: 32119972]
23. Mecklenburg G, Smittenaar P, Erhart-Hledik JC, Perez DA, Hunter S. Effects of a 12-week digital care program for chronic knee pain on pain, mobility, and surgery risk: randomized controlled trial. J Med Internet Res 2018 Apr 25;20(4):e156 [FREE Full text] [doi: 10.2196/jmir.9667] [Medline: 29695370]
24. Skrepnik N, Spitzer A, Altman R, Hoekstra J, Stewart J, Toselli R. Assessing the impact of a novel smartphone application compared with standard follow-up on mobility of patients with knee osteoarthritis following treatment with Hylan G-F 20: a randomized controlled trial. JMIR Mhealth Uhealth 2017 May 09;5(5):e64 [FREE Full text] [doi: 10.2196/mhealth.7179] [Medline: 28487266]
25. Beaudreuil J, Bendaya S, Faucher M, Coudeyre E, Ribinik P, Revel M, et al. Clinical practice guidelines for rest orthosis, knee sleeves, and unloading knee braces in knee osteoarthritis. Joint Bone Spine 2009 Dec;76(6):629-636. [doi: 10.1016/j.jbspin.2009.02.002] [Medline: 19467901]
26. Raja K, Dewan N. Efficacy of knee braces and foot orthoses in conservative management of knee osteoarthritis: a systematic review. Am J Phys Med Rehabil 2011 Mar;90(3):247-262. [doi: 10.1097/PHM.0b013e318206386b] [Medline: 21273902]
27. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
28. Slade SC, Dionne CE, Underwood M, Buchbinder R, Beck B, Bennell K, et al. Consensus on exercise reporting template (CERT): modified delphi study. Phys Ther 2016 Oct;96(10):1514-1524. [doi: 10.2522/ptj.20150668] [Medline: 27149962]
29. Fuchs J, Kuhnert R, Scheidt-Nave C. 12-monats-prävalenz von arthrose in deutschland − fact sheet − JoHM 3/2017. Robert Koch-Institut. 2017. URL: https://www.rki.de/DE/Content/Gesundheitsmonitoring/Gesundheitsberichterstattung/GBEDownloadsJ/FactSheets/JoHM_03_2017_Praevalenz_Arthrose.html [accessed 2024-04-29]
30. Physical activity readiness questionnaire - PAR-Q. Canadian Society for Exercise Physiology (CSEP). 2012. URL: http://westpointgrey.org/wp-content/uploads/2014/02/PAR-QXXForm.pdf [accessed 2023-01-10]
31. PAR-Q fragebogen (Deutsche Fassung). Deutsche Gesellschaft für Sportmedizin und Prävention (DGSP). 2012. URL: https://daten2.verwaltungsportal.de/dateien/seitengenerator/leitlinie_vorsorgeuntersuchung_4.10.2007-anlage-1.pdf [accessed 2023-01-10]
32. Garber CE, Blissmer B, Deschenes MR, Franklin BA, Lamonte MJ, Lee IM. American College of Sports Medicine position stand. Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: guidance for prescribing exercise. Med Sci Sports Exerc 2011;43(7):1334-1359. [doi: 10.1249/mss.0b013e318213fefb]
33. Haupt G, Janßen P, Krauß I, Steinhilber B. Das Tübinger Hüftkonzept. Essen, Germany: Verlag Hellblau; 2014:1-176.
34. Krauss I, Mueller G, Haupt G, Steinhilber B, Janssen P, Jentner N, et al. Effectiveness and efficiency of an 11-week exercise intervention for patients with hip or knee osteoarthritis: a protocol for a controlled study in the context of health services research. BMC Public Health 2016 Apr 30;16(1):367 [FREE Full text] [doi: 10.1186/s12889-016-3030-0] [Medline: 27129849]

35. Krauss I, Steinhilber B, Haupt G, Miller R, Grau S, Janssen P. Efficacy of conservative treatment regimes for hip osteoarthritis--evaluation of the therapeutic exercise regime "Hip School": a protocol for a randomised, controlled trial. BMC Musculoskelet Disord 2011 Nov 24;12(1):270 [FREE Full text] [doi: 10.1186/1471-2474-12-270] [Medline: 22114973]
36. Krauß I, Steinhilber B, Haupt G, Miller R, Martus P, Janßen P. Exercise therapy in hip osteoarthritis--a randomized controlled trial. Dtsch Arztebl Int 2014 Sep 01;111(35-36):592-599 [FREE Full text] [doi: 10.3238/arztebl.2014.0592] [Medline: 25249361]
37. Merk J, Horstmann T, Krauss I, Danzinger B, Ziegler C, Belzl H. Sporttherapie für Patienten mit Kniearthrose oder Knieendoprothese - Konzept, Inhalte und Evaluation der Tübinger Kniesportgruppe. Z Physiother 2005;57:262-273 [FREE Full text]
38. Steinhilber B, Haupt G, Miller R, Boeer J, Grau S, Janssen P, et al. Feasibility and efficacy of an 8-week progressive home-based strengthening exercise program in patients with osteoarthritis of the hip and/or total hip joint replacement: a preliminary trial. Clin Rheumatol 2012 Mar 17;31(3):511-519. [doi: 10.1007/s10067-011-1893-0] [Medline: 22086491]
39. Nielsen J. Iterative user-interface design. Computer 1993 Nov;26(11):32-41. [doi: 10.1109/2.241424]
40. Roos EM, Lohmander LS. The Knee injury and Osteoarthritis Outcome Score (KOOS): from joint injury to osteoarthritis. Health Qual Life Outcomes 2003 Nov 03;1:64 [FREE Full text] [doi: 10.1186/1477-7525-1-64] [Medline: 14613558]
41. Kessler S, Lang S, Puhl W, Stöve J. Der knee injury and osteoarthritis outcome score: ein funktionsfragebogen zur outcome-messung in der knieendoprothetik. Z Orthop Ihre Grenzgeb 2003 Jun 24;141(3):277-282. [doi: 10.1055/s-2003-40083] [Medline: 12822074]
42. Iqbal S, Rogers W, Selim A. The Veterans RAND 12 item health survey (VR-12): what it is and how it is used. Boston University. 2007. URL: https://www.bu.edu/sph/files/2015/01/veterans_rand_12_item_health_survey_vr-12_2007.pdf [accessed 2024-03-18]
43. Bucholz I, Kohlmann T. Vergleichende Untersuchung der psychometrischen Eigenschaften des SF-36/SF-12 vs. VR-36/VR-12. Institut für Community Medicine der Universitätsmedizin Greifswald. 2017. URL: https://reha-vffr.de/images/vffrpdf/projekte/2017/VR-Abschlussbericht_vffr205.pdf [accessed 2024-04-29]
44. Rusu AC, Kreddig N, Hallner D, Hülsebusch J, Hasenbring MI. Fear of movement/(Re)injury in low back pain: confirmatory validation of a German version of the Tampa Scale for Kinesiophobia. BMC Musculoskelet Disord 2014 Aug 19;15(1):280 [FREE Full text] [doi: 10.1186/1471-2474-15-280] [Medline: 25138111]
45. Finger JD, Tafforeau J, Gisle L, Oja L, Ziese T, Thelen J, et al. Development of the European Health Interview Survey - physical activity questionnaire (EHIS-PAQ) to monitor physical activity in the European Union. Arch Public Health 2015 Dec 2;73(1):59 [FREE Full text] [doi: 10.1186/s13690-015-0110-z] [Medline: 26634120]
46. Rodgers WM, Wilson PM, Hall CR, Fraser SN, Murray TC. Evidence for a multidimensional self-efficacy for exercise scale. Res Quart Exerc Sport 2008 Jun 01;79(2):222-234. [doi: 10.5641/193250308x13086753543590]
47. Sudeck G, Pfeifer K. Physical activity-related health competence as an integrative objective in exercise therapy and health sports – conception and validation of a short questionnaire. Sportwiss 2016;46:74-87 [FREE Full text] [doi: 10.1007/s12662-016-0405-4]
48. Assessment 30-second chair stand. Centers for Disease Control and Prevention. 2017. URL: https://www.cdc.gov/steadi/pdf/STEADI-Assessment-30Sec-508.pdf [accessed 2023-01-10]
49. Faces pain scale. International Association for the Study of Pain. 2001. URL: https://www.iasp-pain.org/Education/Content.aspx?ItemNumber=1519&navItemNumber=577 [accessed 2023-01-10]
50. Walters SJ, Jacques RM, Dos Anjos Henriques-Cadby IB, Candlish J, Totton N, Xian MT. Sample size estimation for randomised controlled trials with repeated assessment of patient-reported outcomes: what correlation between baseline and follow-up outcomes should we assume? Trials 2019 Sep 13;20(1):566 [FREE Full text] [doi: 10.1186/s13063-019-3671-2] [Medline: 31519202]
51. Olejnik S, Algina J. Measures of effect size for comparative studies: applications, interpretations, and limitations. Contemp Educ Psychol 2000 Jul;25(3):241-286. [doi: 10.1006/ceps.2000.1040] [Medline: 10873373]
52. Bossen D, Veenhof C, Van Beek KE, Spreeuwenberg PM, Dekker J, De Bakker DH. Effectiveness of a web-based physical activity intervention in patients with knee and/or hip osteoarthritis: randomized controlled trial. J Med Internet Res 2013 Nov 22;15(11):e257 [FREE Full text] [doi: 10.2196/jmir.2662] [Medline: 24269911]
53. Uesugi Y, Koyanagi J, Takagi K, Yamaguchi R, Hayashi S, Nishii T. Exercise therapy interventions in patients with hip osteoarthritis: comparison of the effects of DVD and website-based interventions. JMIR Rehabil Assist Technol 2018 May 07;5(1):e10 [FREE Full text] [doi: 10.2196/rehab.8251] [Medline: 29735473]
54. Angst F, Benz T, Lehmann S, Aeschlimann A, Angst J. Multidimensional minimal clinically important differences in knee osteoarthritis after comprehensive rehabilitation: a prospective evaluation from the Bad Zurzach Osteoarthritis Study. RMD Open 2018 Oct 08;4(2):e000685 [FREE Full text] [doi: 10.1136/rmdopen-2018-000685] [Medline: 30402264]
55. Messier SP, Callahan LF, Golightly YM, Keefe F. OARSI clinical trials recommendations: design and conduct of clinical trials of lifestyle diet and exercise interventions for osteoarthritis. Osteoarthritis Cartilage 2015 May;23(5):787-797 [FREE Full text] [doi: 10.1016/j.joca.2015.03.008] [Medline: 25952349]

56. Anwer S, Alghadir A, Brismée JM. Effect of home exercise program in patients with knee osteoarthritis: a systematic review and meta-analysis. J Geriatr Phys Ther 2016;39(1):38-48. [doi: 10.1519/JPT.0000000000000045] [Medline: 25695471]
57. Verhagen A, Ferreira M, Reijneveld-van de Vendel E, Teirlinck C, Runhaar J, van Middelkoop M, et al. Do we need another trial on exercise in patients with knee osteoarthritis?: no new trials on exercise in knee OA. Osteoarthritis Cartilage 2019 Sep;27(9):1266-1269 [FREE Full text] [doi: 10.1016/j.joca.2019.04.020] [Medline: 31220609]
58. Holden M, Hattle M, Runhaar J, Riley R, Healey E, Quicke J, STEER OA Patient Advisory Group, OA Trial Bank Exercise Collaborative. Moderators of the effect of therapeutic exercise for knee and hip osteoarthritis: a systematic review and individual participant data meta-analysis. Lancet Rheumatol 2023 Jul;5(7):e386-e400 [FREE Full text] [doi: 10.1016/S2665-9913(23)00122-4] [Medline: 38251550]
59. Tanaka R, Ozawa J, Kito N, Moriyama H. Does exercise therapy improve the health-related quality of life of people with knee osteoarthritis? A systematic review and meta-analysis of randomized controlled trials. J Phys Ther Sci 2015 Oct;27(10):3309-3314 [FREE Full text] [doi: 10.1589/jpts.27.3309] [Medline: 26644699]
60. Selim AJ, Rogers W, Fleishman JA, Qian SX, Fincke BG, Rothendler JA, et al. Updated U.S. population standard for the Veterans RAND 12-item Health Survey (VR-12). Qual Life Res 2009 Feb 3;18(1):43-52. [doi: 10.1007/s11136-008-9418-2] [Medline: 19051059]
61. Gohir SA, Eek F, Kelly A, Abhishek A, Valdes AM. Effectiveness of internet-based exercises aimed at treating knee osteoarthritis: the iBEAT-OA randomized clinical trial. JAMA Netw Open 2021 Feb 01;4(2):e210012 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.0012] [Medline: 33620447]
62. Runhaar J, Holden MA, Hattle M, Quicke J, Healey EL, van der Windt D, STEER OA Patient Advisory Group, OA Trial Bank Exercise Collaborative. Mechanisms of action of therapeutic exercise for knee and hip OA remain a black box phenomenon: an individual patient data mediation study with the OA trial bank. RMD Open 2023 Aug 28;9(3):e003220 [FREE Full text] [doi: 10.1136/rmdopen-2023-003220] [Medline: 37640513]
63. Zacharias A, Green R, Semciw A, Kingsley M, Pizzari T. Efficacy of rehabilitation programs for improving muscle strength in people with hip or knee osteoarthritis: a systematic review with meta-analysis. Osteoarthritis Cartilage 2014 Nov;22(11):1752-1773 [FREE Full text] [doi: 10.1016/j.joca.2014.07.005] [Medline: 25065642]
64. Dobson F, Hinman RS, Roos EM, Abbott JH, Stratford P, Davis AM, et al. OARSI recommended performance-based tests to assess physical function in people diagnosed with hip or knee osteoarthritis. Osteoarthritis Cartilage 2013 Aug;21(8):1042-1052 [FREE Full text] [doi: 10.1016/j.joca.2013.05.002] [Medline: 23680877]
65. Bennell KL, Hinman RS. A review of the clinical evidence for exercise in osteoarthritis of the hip and knee. J Sci Med Sport 2011 Jan;14(1):4-9. [doi: 10.1016/j.jsams.2010.08.002] [Medline: 20851051]
66. Marks R. Knee osteoarthritis and exercise adherence: a review. Curr Aging Sci 2012 Feb 01;5(1):72-83. [doi: 10.2174/1874612811205010072]
67. Yamamoto Y, Murata Y, Tanaka N, Shigemura T, Maruyama J, Nakane R, et al. Mobile application for home exercise adherence in patients with knee osteoarthritis: a pilot study. Medicine (Baltimore) 2022 Oct 21;101(42):e31181 [FREE Full text] [doi: 10.1097/MD.0000000000031181] [Medline: 36281120]
68. Valenzuela T, Okubo Y, Woodbury A, Lord SR, Delbaere K. Adherence to technology-based exercise programs in older adults: a systematic review. J Geriatr Phys Ther 2018;41(1):49-61. [doi: 10.1519/JPT.0000000000000095] [Medline: 27362526]
69. Ay S, Evcik D, Kutsal YG, Toraman F, Okumu' M, Eyigör S, et al. Compliance to home-based exercise therapy in elderly patients with knee osteoarthritis. Turk J Phys Med Rehab 2016 Dec 19;62(4):323-328. [doi: 10.5606/tftrd.2016.54189]
70. Pisters MF, Veenhof C, de Bakker DH, Schellevis FG, Dekker J. Behavioural graded activity results in better exercise adherence and more physical activity than usual care in people with osteoarthritis: a cluster-randomised trial. J Physiother 2010;56(1):41-47 [FREE Full text] [doi: 10.1016/s1836-9553(10)70053-9] [Medline: 20500136]
71. Jakobsen JC, Gluud C, Wetterslev J, Winkel P. When and how should multiple imputation be used for handling missing data in randomised clinical trials - a practical guide with flowcharts. BMC Med Res Methodol 2017 Dec 06;17(1):162 [FREE Full text] [doi: 10.1186/s12874-017-0442-1] [Medline: 29207961]

Abbreviations

AE: adverse event
ANCOVA: analysis of covariance
CG: control group
CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
ES: effect size
HRQoL: health-related quality of life
IG A: intervention group with app-based exercise training
IG AB: intervention group with app-based exercise training in combination with a supportive knee brace

IG: intervention group
KOOS: Knee Injury and Osteoarthritis Outcome Score
MCID: minimal clinically important difference
MCII: minimal clinically important improvement
MCS: Mental Component Score
mHealth: mobile health
NSAID: nonsteroidal anti-inflammatory drug
PCS: Physical Component Score
PROM: patient-reported outcome measure
QoL: quality of life
SAE: serious adverse event

Edited by L Buis; submitted 07.11.23; peer-reviewed by G Osterhoff, C Jacob, R Eckhoff; comments to author 30.04.24; revised version received 16.05.24; accepted 12.06.24; published 09.09.24.

Please cite as:
Dieter V, Janssen P, Krauss I
Efficacy of the mHealth-Based Exercise Intervention re.flex for Patients With Knee Osteoarthritis: Pilot Randomized Controlled Trial
JMIR Mhealth Uhealth 2024;12:e54356
URL: https://mhealth.jmir.org/2024/1/e54356
doi:10.2196/54356
PMID:39250181

Original Paper

A Dual-Modality Home-Based Cardiac Rehabilitation Program for Adults With Cardiovascular Disease: Single-Arm Remote Clinical Trial

Tim Bilbrey[1], EP, MBA; Jenny Martin[1], APRN, FNP-C; Wen Zhou[1], RN; Changhao Bai[1], PT; Nitin Vaswani[2], MD, MPH, MBA; Rishab Shah[2], MHS; Sara Chokshi[2], DrPH; Xi Chen[1], MD; Satjit Bhusri[3], MD; Samantha Niemi[4]; Hongdao Meng[5], MD, MPH, PhD; Zhen Lei[1], MD, MBA

[1]RecoveryPlus.Health, Inc, New York, NY, United States

[2]Node.Health Foundation, Wilmington, DE, United States

[3]Upper East Side Cardiology PLLC, New York, NY, United States

[4]McCormick School of Engineering, Northwestern University, Chicago, IL, United States

[5]School of Aging Studies, University of South Florida, Tampa, FL, United States

Corresponding Author:
Hongdao Meng, MD, MPH, PhD
School of Aging Studies
University of South Florida
13301 Bruce B Downs Blvd
MHC1300
Tampa, FL, 33612
United States
Phone: 1 18139745504
Email: meng@usf.edu

Abstract

Background: Cardiac rehabilitation (CR) is a safe, effective intervention for individuals with cardiovascular disease (CVD). However, a majority of eligible patients do not complete CR. Growing evidence suggests that home-based cardiac rehabilitation (HBCR) programs are comparable in effectiveness and safety with traditional center-based programs. More research is needed to explore different ways to deliver HBCR programs to patients with CVD.

Objective: We aimed to assess the feasibility and impact of a digital HBCR program (RecoveryPlus.Health) that integrates both telehealth and mHealth modalities on functional exercise capacity, resting heart rate, and quality of life among adults with CVD.

Methods: This 12-week prospective, single-arm remote clinical trial used a within-subject design. We recruited adults with CVD (aged ≥40 years) from the community with a CR-eligible diagnosis (stable angina pectoris, myocardial infarction, and heart failure) between May and August 2023. All enrolled patients referred to the RPH clinic in Roanoke, Texas, were included. The care team provided guideline-concordant CR services to study participants via two modalities: (1) a synchronous telehealth exercise training through videoconferencing; and (2) an asynchronous mobile health (mHealth) coaching app (RPH app). Baseline intake survey, electronic health record, and app log data were used to extract individual characteristics, care processes, and platform engagement data. Feasibility was measured by program completion rate and CR service use. Efficacy was measured by changes in the 6-minute walk test, resting heart rate, and quality of life (12-Item Short-Form Health Survey) before and after the 12-week program. Paired *t* tests were used to examine pre- and postintervention changes in the outcome variables.

Results: In total, 162 met the inclusion criteria and 75 (46.3%) consented and were enrolled (mean age 64, SD 10.30 years; male: n=37, 49%; White: n=46, 61%). Heart failure was the most common diagnosis (37/75, 49%). In total, 62/75 (83%) participants completed the 12-week study and used the telehealth modality with 9.63 (SD 3.33) sessions completed, and 59/75 (79%) used the mHealth modality with 10.97 (SD 11.70) sessions completed. Post intervention, 50/62 (81%) participants' performance in the 6-minute walk test had improved, with an average improvement of 40 (SD 63.39) m (95% CI 25.6-57.1). The average 12-Item Short-Form Health Survey's physical and mental summary scores improved by 2.7 (SD 6.47) points (95% CI 1.1-4.3) and 2.2

(SD 9.09) points (95% CI 0.1-4.5), respectively. There were no changes in resting heart rate and no exercise-related adverse events were reported.

Conclusions: The RecoveryPlus.Health digital HBCR program showed feasibility and efficacy in a group of nationally recruited patients with CVD. The findings add to the evidence that a telehealth and mHealth dual-modality HBCR program may be a promising approach to overcome some of the main barriers to improving CR access in the United States.

Trial Registration: ClinicalTrials.gov NCT05804500; https://clinicaltrials.gov/search?cond=NCT05804500

(JMIR Mhealth Uhealth 2024;12:e59098) doi:10.2196/59098

KEYWORDS

cardiac rehabilitation; telehealth; mHealth; digital health; exercise; quality of life; myocardial infarction; app; application; physical fitness; cardiac rehabilitation; self-management; disease management

Introduction

Cardiovascular disease (CVD) remains the leading cause of morbidity and mortality in the United States, with more than 2 million hospitalizations and 400,000 deaths due to CVD [1,2]. Cardiac rehabilitation (CR) is a grade I guideline-recommended multidisciplinary program intended to reduce the risk of subsequent cardiovascular events and improve quality of life through activities shown to improve physical, psychological, and social functioning, and ultimately reduce associated morbidity among patients with CVD [3,4]. Despite strong evidence regarding its safety and efficacy, CR use rate remains very low, with only about 1 in 4 eligible patients enrolling in CR [5-10]. In addition, disparities in CR uptake with respect to sex, age, race, ethnicity, and geographic locations persist, contributing to the low overall CR participation rate and substantial geographic variations in participation [3,11]. With the initiation of the Million Hearts 2022, a national program aimed at achieving a target of 70% CR participation among eligible patients by 2027 [9], the current 2300 center-based cardiac rehabilitation (CBCR) programs in the United States are unlikely to meet the demand for CR [12,13]. Therefore, new home-based cardiac rehabilitation (HBCR), an alternative to traditional CBCR that allows patients to undergo rehabilitation in their own homes, is urgently needed to expand CR access [3,14].

HBCR alone or in combination (hybrid) with CBCR, has been shown to be a safe and effective option for patients eligible for CR [15-17], with the potential to address some of the key barriers to CR access (eg, time and cost of transportation) [5]. HBCR programs use the same multifaceted approaches (exercise training, education, dietary guidance, and lifestyle modification) as CBCR, providing patients with synchronous or asynchronous delivery of supervised remote exercise training and care management [4,18,19].

While HBCR programs have certain disadvantages compared with CBCR (eg, lack of published clinical standards, lower exercise training intensity, and potential safety concerns for high-risk patients), their key advantages include reduced enrollment delays, expanded capacity, flexible scheduling, minimal travel, and improved access for patients with limited mobility [2,5,20]. The unexpected disruption to CBCR access caused by the COVID-19 pandemic inadvertently accelerated the transition of many in-person CR programs to the home setting and increased the acceptability of HBCR and center-home hybrid CR programs [21]. There is a need for greater engagement in safe and evidence-based CR exercise programming for patients recovering from CVD. Among these new approaches, home-based synchronous remote CR and asynchronous mobile health (mHealth)–based CR modalities, alone or in combination, have the potential to improve access by expanding provider offerings and reduce the time and costs associated with travel to traditional CR centers [22].

The RecoveryPlus.Health digital (RPH-D) HBCR program is an innovative program that combines 2 delivery modalities, a synchronous telehealth modality delivered through videoconferencing by an interdisciplinary care team and an asynchronous mHealth modality delivered through a digital app (RPH app) for individualized exercise therapy. The RPH-D program was developed based on the American Heart Association CR guidelines for delivering tailored, evidence-based CR remotely and on demand.

This study aimed to determine whether the dual-modality RPH-D program is feasible for delivering remote CR to a nationally recruited sample, and to examine its impact on patients' functional exercise capacity (cardiorespiratory endurance), resting heart rate, and quality of life.

Methods

Recruitment

This is a prospective single-arm remote clinical trial using a within-subject design. Between May 1, 2023, and November 30, 2023, a dedicated recruitment website was established to facilitate the sharing of study information for recruitment through 2 primary channels, referrals from cardiology providers in the Dallas-Fort Worth area and through a collaboration with The Mended Hearts, Inc, a national and community-based nonprofit cardiovascular patient support organization. The community partners sent emails to their patients prompting potential participants to visit the study website. Potential participants were screened for eligibility criteria by a secure web-based questionnaire. Patients who passed the initial screening were contacted by the study coordinator by phone to verify eligibility and finalize enrollment.

Participants were included if they were aged 45 years or older; had stable CVD and under medical management; received referral to CR from a provider within 60 days; were able to walk

unassisted; deemed stable with low to moderate risk of a cardiac event; and with a CR eligible diagnosis in previous 12 months as defined by Medicare Part B (stable angina pectoris, myocardial infarction, stable heart failure, and coronary artery bypass graft surgery). Participants were excluded if they were unable to read and speak in English, had a BMI greater than 40 kg/m^2, had hospitalization or significant decline in health, had physical or mental health limitations that prohibit participation in exercise activities, could not use a tablet computer, lack of access to Wi-Fi, or had severe hearing or vision impairment.

Intervention: The RecoveryPlus.Health Digital CR Program

The RPH-D program for patients recovering from a CVD event leverages 2 weekly concurrent remote modalities to engage the study participants: (1) synchronous (telehealth) coaching sessions conducted by an exercise physiologist by Zoom for Healthcare (Zoom Video Communications, Inc); and (2) asynchronous (mHealth) exercise sessions by the RPH app. The telehealth sessions focused on the following core CR services: patient assessment, exercise training, care management, lifestyle counseling, and remote monitoring. The study iPads were preloaded with the RPH app, which offers a library of on-demand exercise videos with varying degrees of difficulty. Wireless heart rate and blood pressure monitors were connected to the iPad by Bluetooth to enable real-time monitoring of patient vital signs and feedback (eg, patient-rated difficulty for each prescribed exercise) and tailor individual patient regimens regularly for safety and fine-tuning the exercise therapy. During the app-based exercise sessions, participants rated the perceived level of difficulty (exertion) of each exercise on a 1-10 scale. If a patient chooses to skip an exercise or exit a session before completing the prescribed exercises, the reasons were documented by the app (Multimedia Appendix 1).

All clinical team members had access to the RPH remote health care provider web portal (HIPAA [Health Insurance Portability and Accountability Act]-complaint) to enter, store, and view participant demographic and clinical information, create fitness assessments, write and update a patient's exercise prescription, and generate reports. The clinicians were trained to use patient feedback along with data derived from heart rate monitors to assess exercise intensity, appropriateness, and adherence to the prescribed exercise program, and, importantly, make adjustments to better personalize the care plan for each patient. The RPH remote CR platform also provided real-time, automated alerts that were sent to the care team for any symptoms or reports of out-of-range heart rate reading to facilitate immediate evaluation to determine the severity of a symptom and which course of action is required for follow-up. In these events, participants were instructed to stop exercising immediately and report any cardiac-related symptoms to their clinician at any time, during or outside of exercise sessions. The clinical care team received and responded to alerts 24 hours a day, 7 days per week.

Outcomes and Measures

The main variables of interest were feasibility and efficacy. Feasibility was measured by program initiation (the proportion of enrolled patients who completed all baseline assessments and completed the initial Zoom teleconferencing session with an exercise physiologist and participation (the proportion of enrolled patients who completed at least 50% or more of prescribed sessions). Feasibility data were captured by the HIPAA-compliant, encrypted, RPH electronic health records. Efficacy was measured by 6-minute walk test (6MWT), resting heart rate, and quality of life. These measures were assessed by the study exercise physiologist and recorded on the RPH electronic health record platform. The 6MWT was measured by the 6WT app, an iOS app validated in 330 volunteers (age range 16-91 years) [23]. Participants were instructed to walk on a flat, hard surface outside of their homes for a period of 6 minutes. The 6MWT is a widely used and well-documented measure of aerobic exercise capacity used to indicate change in fitness [24]. Resting heart rate is positively associated with mortality and is known to decline with regular exercise [25]. Quality of life was measured by the 12-Item Short-Form Health Survey (SF-12, version 2), a validated and commonly used measure of perceived quality of life and functional health among patients with CVD [26]. Individual characteristics included basic sociodemographic variables collected at baseline (age, sex, race, ethnicity, comorbidities, and referring diagnosis).

Sample Size Calculation

Power calculation was based on the primary efficacy measure: 6MWT. At a 5% level of significance, to achieve 85% power in detecting a small to moderate difference of 35 m for the 6MWT before and after the RPH-D program, a sample size of 75 participants was required [24,27]. To account for an estimated 25% attrition rate, a total sample of 100 was set as the enrollment target.

Statistical Analysis

Standard descriptive statistics were used to describe the study sample and compare them by study completion status. For continuous variables, 2-sample *t* test was used, and Fisher exact test was used for categorical variables. To examine the impact of the RPH-D program on the outcome measures, paired *t* tests were used to examine whether there were any within-subject changes post intervention. To quantify the uncertainty surrounding the point estimates of effect sizes, 95% CI was calculated using nonparametric bootstrapping with 5000 repetitions [28]. Statistical analysis was performed in Stata 17 BE (StataCorp).

Ethical Considerations

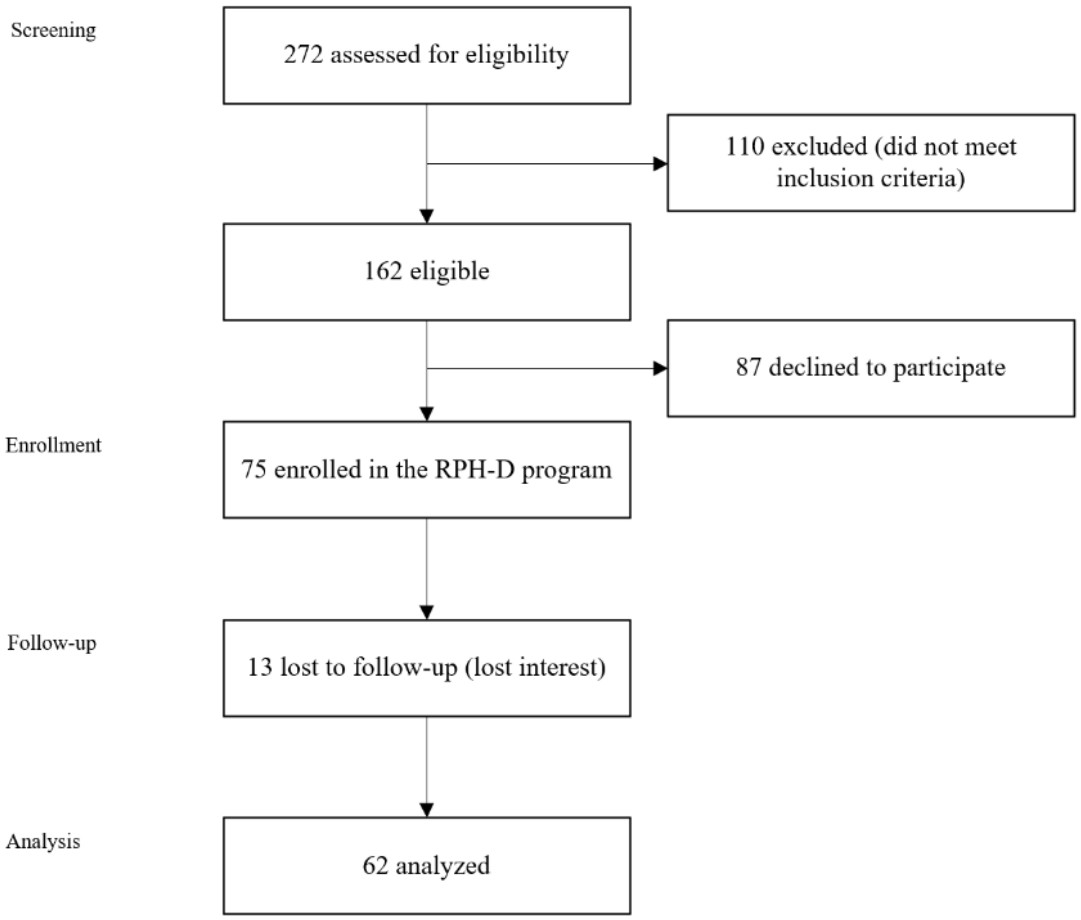
The study protocol was approved by the Advarra Institutional Review Board (Pro00070335) and written informed consent was obtained from all participants.

Results

Overview

Of the 272 individuals screened for eligibility, a total of 162 individuals from 29 states met the study eligibility criteria and 75 (46%) consented to the study and were enrolled in the RPH-D program. In total, 13 participants withdrew from the study and 62 (83%) were included in the analysis (Figure 1).

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) flow diagram for patient enrollment and follow-up. RPH-D: RecoveryPlus.Health Digital.



Participant Characteristics

Table 1 displays the baseline participants' individual characteristics for the overall sample and by participation rate (less than or greater than 50%). Study participants had a mean age of 64.6 (SD 10.0; range: 45-85 years) and 38 (51%) were female (Table 1). Non-Hispanic White participants made up majority of the sample at 46 (61%), followed by 8 (11%) Black American, 5 (7%) Hispanic, and 16 (21%) other or missing race. The top 5 referring diagnoses were chronic heart failure (n=37, 50%), coronary artery bypass grafting (n=9, 15%), valvular surgery (n=9, 15%), acute myocardial infarction (n=6, 10%), and percutaneous coronary intervention (n=6, 10%). Participants who completed 50% or more sessions were older and more likely to have percutaneous coronary intervention as the referring diagnosis. With the exception of age and BMI, there were no significant differences in the characteristics of the participants who completed the study versus those who did not.

A total of 62 out of 75 (83%) participants completed the 12-week study. Out of 62 participants, 50 (81%) completed at least 50% of CR sessions. All participants were prescribed 12 one-on-one videoconferencing (telehealth) sessions (55 minutes per week, 660 minutes total) and 12 app-based exercise (mHealth) sessions (27.8 minutes per week, 334 minutes total) over 12 weeks. A total of 62 out of 75 (83%) participants used the telehealth modality (9.63, SD 3.33 sessions) and 34 of 62 (55%) participants completed all 12 sessions. A total of 59 out of 75 (79%) used the mHealth modality (10.97, SD 11.70 sessions) and 32 out of 62 (52%) completed all 12 sessions. Among those who completed the study, all participants used the telehealth modality and all but 3 participants used the mHealth modality. Participants completed an average of 654.1 (SD 113.6) minutes of telehealth sessions and 421 minutes (SD 306) of mHealth sessions (Table 2). A total of 12 participants visited the emergency room (15 visits total) and none were related to the RPH-D program.

In total, 3 outcome measures were used to examine the impact of the intervention before and after the 12-week intervention (Table 3). Overall, participants who completed the intervention had an average improvement in 6MWT performance of 40 (SD 63.39) meters (95% CI 25.6-57.1). The average SF-12 physical and mental component summary scores improved by 2.7 (SD 6.47, 95% CI 1.1-4.3) and 2.2 (SD 9.09, 95% CI 0.1-4.5) points, respectively. There was a small nonsignificant improvement in average resting heart rates (mean –1.1, SD 9.05, 95% CI –3.4 to 1.1).

Table 1. Baseline characteristics of study participants (N=62).

Variables	Overall (N=62)	<50% completion (n=12)	≥50% completion (n=50)	*P* value
Age (years), mean (SD)	64.6 (10)	56.9 (6.1)	66.4 (9.9)	<.001
Age group (years), n (%)				
45-54	12 (19.4)	4 (33.3)	8 (16)	—[a]
55-64	20 (32.3)	7 (58.3)	13 (26)	—[a]
65-74	19 (30.6)	1 (8.3)	18 (36)	—[a]
75-85	11 (17.7)	0 (0)	11 (22)	—[a]
Sex, n (%)				.75
Male	30 (48.4)	5 (41.7)	25 (50)	—[a]
Female	32 (51.6)	7 (58.3)	25 (50)	—[a]
Race/ethnicity, n (%)				
Non-Hispanic White	37 (59.7)	8 (66.7)	29 (58)	—[a]
Non-Hispanic Black	7 (11.3)	2 (16.7)	5 (10)	—[a]
Hispanic	5 (8.1)	1 (8.3)	4 (8)	—[a]
Other/missing	13 (21)	1 (8.3)	12 (24)	—[a]
BMI (kg/m^2), mean (SD)	29.3 (6.8)	30.4 (8.2)	29 (6.5)	<.001
BMI group (kg/m^2), n (%)				
Normal weight	19 (31.1)	5 (41.7)	14 (28.6)	—[a]
Overweight	19 (31.1)	2 (16.7)	17 (34.7)	—[a]
Obese	23 (37.7)	5 (41.7)	18 (36.7)	—[a]
Referring diagnosis, n (%)				
Coronary artery bypass grafting	9 (14.5)	2 (16.7)	7 (14)	>.99
Chronic heart failure	31 (50)	5 (41.7)	26 (52)	.75
Acute myocardial infarction	6 (9.7)	3 (25)	3 (6)	.08
Valvular surgery	9 (14.5)	2 (16.7)	7 (14)	>.99
Angina pectoris	1 (1.6)	0 (0)	1 (2)	>.99
Percutaneous coronary intervention	6 (9.7)	0 (0)	6 (12)	.59

[a]—: not applicable.

Table 2. Participant engagement with the 2 modalities of remote cardiac rehabilitation (N=62).

Variables	Telehealth modality	Mobile health modality
Number of users, n (%)		
Nonusers	0 (0)	3 (4.8)
Users	62 (100)	60 (95.2)
Completed number of sessions		
Mean (SD)	10.9 (2.3)	12.9 (11.9)
Median (IQR)	12.0 (10-12)	12 (2-19)
Range	4-12	0.0-46.0
Completed total number of minutes		
Mean (SD)	654.1 (113.6)	421 (306)
Median (IQR)	712 (600-720)	391.5 (121.0-636.0)
Range	219-780	24-1118

Table 3. Outcome measures at baseline, 12 weeks, and changes (N=62).

Variables	Baseline	12 weeks	Change, mean (95% CI)
6MWT[a], meters, mean (SD)	422.8 (78.6)	462.8 (77.4)	40.0 (25.6 to 57.1)
Resting heart rate, beats per minute, mean (SD)	70.6 (12.2)	69.5 (10.6)	–1.1 (–3.4 to 1.1)
SF-12[b] physical component summary, points, mean (SD)	41.5 (9)	44.2 (8.9)	2.7 (1.1 to 4.3)
SF-12[b] mental component summary, points, mean (SD)	51.2 (10.8)	53.4 (9.5)	2.2 (0.1 to 4.5)

[a]6MWT: 6-minute walk test.

[b]SF-12: 12-Item Short-Form Health Survey.

Discussion

Principal Findings

In this remote clinical study, the RPH digital home-based CR program that integrated synchronous telehealth and asynchronous mHealth modalities was tested among a group of nationally recruited individuals with cardiovascular disease. Screening, recruitment, and engagement outcomes showed that while enrollment yield was modest, the great majority of consented individuals were able to complete the program. Adherence to both modalities was high, although there were substantial variations in adherence to the mHealth exercise sessions. Overall, the program was feasible and efficacious in improving functional exercise capacity and quality of life.

Home-based versus center-based cardiac rehabilitation continues to receive much attention, with a recent Cochrane review identifying 24 completed trials and at least 14 more registered. While the extent of evidence suggests that these 2 delivery approaches showed comparable effects on total morality and exercise capacity up to 12 months postintervention, with no significant differences in health-related quality of life for up to 24 months [12], evidence regarding the delivery modality for HBCR is less clear. A recent review of asynchronous and synchronous delivery models for HBCR showed that while most studies of HBCR tested asynchronous approaches before 2016, a growing number of HBCR trials started to incorporate synchronous approaches [14]. The findings of this study add to the growing evidence base supporting the feasibility of a concurrent dual-modality remote CR for individuals with CVD. For example, a study in New Zealand among predominantly male patients with CHD found that 82.9% (68/82) of those assigned to the remote CR program completed the 12-week program. A total of 2 small studies in the United States reported varying completion rates. Misra et al [29] tested a digitally delivered remote CR program among 12 patients with atrial fibrillation who underwent catheter ablation in Charlotte, North Carolina. In total, 92% (11/12) were able to complete the 12-week program, with an average of 2.9 exercise sessions per week. Giggins et al [30] reported a completion rate of 72.7% (8/11) for an 8-week web-based remote CR program among patients with CHD. Taken together, these findings support the feasibility of HBCR programs for individuals with CVD.

The broad geographic reach (29 states) of the RPH-D program conducted from a single clinic shows the promise of digital technology-enabled remote CR programs. As geographic disparities in access to traditional center-based CR have been highlighted as one of the main barriers to CR participation [3,31], asynchronous home-based CR programs like the RPH-D program provide greater flexibility for patients with CVD. The findings indicate that the great majority of eligible

participants-initiated CR with high completion rates. The availability of the asynchronous mHealth modality may help in narrowing the digital divide among patients from rural and remote areas [32]. Participation rate (of 50%) did not differ among participants based on sex and race, suggesting that the program may offer women and minority patients an equitable option for HBCR access.

The findings from this study are also in line with those from multiple systematic reviews concluding that eHealth delivery of cardiac rehabilitation increases patient physical activity and compliance [14,33-35]. As suggested by a recent review of remote solutions for CR, the use of multiple devices for monitoring and communication with their health care team, program personalization, and continuous feedback for users all contributed to the program compliance and feasibility found in this study [20]. This review of 19 different studies of remote cardiac rehabilitation programs noted, as did ours, no intervention-related adverse events, suggesting that remote CR is generally safe.

This study contributes to the literature in several ways. To the best of our knowledge, this is the first study that tested a concurrent telehealth and mHealth dual-modality HBCR program in a group of age and sex-diverse adults in the United States. An Australian-based randomized controlled trial with 24-week (12+12) sequential dual modality HBCR program is currently being tested [36]. It was designed to be pragmatic, with care team members providing remote CR service using existing clinical procedures and electronic medical records system for workflow management. Hybrid, remote, and digitally supported CR services appear to be a safe and effective delivery approach for secondary prevention of CVD with growing evidence [4,37-40]. In addition, the encouraging uptake of remote CR services by women and minority patients may offer new ways to reduce disparities in CR care [41], with the potential to greatly increase the number of patients with CVD engaging in evidence-based CR, an intervention shown to positively impact the health, fitness, and quality of life outcomes for these patients [7,42,43]. Our data demonstrate that patients with CVD will maintain engagement with a platform like RPH remote CR that combines synchronous 1:1 telerehabilitation sessions with an EP with asynchronous exercise sessions longer than they would be traditional in-person CR. In addition, the on-demand exercise feature of the program provides an engaging physical activity option outside of the one-on-one teleconferencing sessions, a promising solution to scheduling and scalability challenges of traditional in-person CR programs. Finally, it is encouraging that this study enrolled a gender, culturally, and geographically diverse patient population, all of which has long been recognized as being underserved by CR services [44-46].

Limitations

Several limitations should be considered when interpreting the findings. This single-arm remote clinical trial did not include a control group, so we cannot rule out the possibility that the changes in the patient outcome measures may be due to factors other than the RPH-D program. This design choice was based on clinical and ethical reasons: CR is a well-established clinical approach to manage CVD; thus, a randomized controlled trial would have required either not offering or delaying the active intervention to half of all eligible patients. Even a randomized controlled trial with an active CBCR arm would have required some patients to access center-based CR, which may not be readily available to this patient population. Second, study participants came from a national convenience sample with heterogeneous CR-eligible diagnoses and the sample size was modest. Therefore, the findings should be replicated in larger national studies that would allow for subgroup analyses (eg, by referring diagnoses, age, or sex) to help inform clinical practice. Third, the 12-week follow-up period was limited and cannot offer evidence regarding longer-term adherence and outcomes. In addition, the participation rate for the mHealth modality was limited, as some participants experienced challenges with using the iPad with wireless monitors due to limited digital literacy. Finally, relying on 2 primary recruitment channels (through community partner referral and web-based advertisement) contributed to the under-recruitment of the study. Future studies should broaden the referral network to include primary and specialty care and include an economic evaluation component to generate evidence regarding the value for payers [47]. Devices that enable coverage beyond traditional Wi-Fi technology (such as those based on 5G mobile phone network or satellite network) can further reduce geographic barriers to HBCR. These limitations notwithstanding, this is the first US study to test an innovative dual-modality remote CR program that combined synchronous and asynchronous approaches to deliver a fully remote home-based CR program.

Conclusion

The results of this study show that a dual-modality home-based CR program can be a feasible option to improve access to home-based CR for improving functional exercise capacity after acute CVD. Future research using randomized controlled design (eg, a preference-based RCT) is needed to test the RPH-D CR program in a larger, more diverse pool of patients. If the long-term impact of this program can be confirmed in a multicenter randomized controlled study, a stronger case can be made to implement HBCR programs in health care systems to increase the uptake of CR and bridge the gap between evidence and practice in secondary prevention of CVD.

Acknowledgments

The authors would like to acknowledge The Mended Hearts, Inc, for assistance with member outreach and recruitment. We also wish to acknowledge all study participants and the staff at Node.Health Foundation and RecoveryPlus.Health, for their contributions to this research project.

Authors' Contributions

TB, NV, RS, SC, and ZL contributed to the conception and design of the project. JM, WZ, CB, RS, and SC contributed to the implementation of the program and acquisition of data. RS, SC, and HM drafted the manuscript. TB, JM, SN, SB, HM, and ZL critically revised the manuscript. All authors gave final approval and agreed to take responsibility for the integrity of the data and the accuracy of the work reported.

Conflicts of Interest

TB, JM, WZ, CB, XC, and ZL are employees of RecoveryPlus.Health.

Multimedia Appendix 1
RecoveryPlus.Health App screenshots.
[DOCX File , 754 KB - mhealth_v12i1e59098_app1.docx]

References

1. Ritchey MD, Wall HK, Owens PL, Wright JS. Vital signs: state-level variation in nonfatal and fatal cardiovascular events targeted for prevention by Million Hearts 2022. MMWR Morb Mortal Wkly Rep 2018 Sep 07;67(35):974-982 [FREE Full text] [doi: 10.15585/mmwr.mm6735a3] [Medline: 30188881]
2. Roth GA, Mensah GA, Johnson CO, Addolorato G, Ammirati E, Baddour LM, GBD-NHLBI-JACC Global Burden of Cardiovascular Diseases Writing Group. Global burden of cardiovascular diseases and risk factors, 1990-2019: update from the GBD 2019 study. J Am Coll Cardiol 2020 Dec 22;76(25):2982-3021 [FREE Full text] [doi: 10.1016/j.jacc.2020.11.010] [Medline: 33309175]
3. Beatty AL, Beckie TM, Dodson J, Goldstein CM, Hughes JW, Kraus WE, et al. A new era in cardiac rehabilitation delivery: research gaps, questions, strategies, and priorities. Circulation 2023 Jan 17;147(3):254-266 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.122.061046] [Medline: 36649394]
4. Laranjo L, Lanas F, Sun MC, Chen DA, Hynes L, Imran TF, et al. World Heart Federation roadmap for secondary prevention of cardiovascular disease: 2023 update. Glob Heart 2024;19(1):8 [FREE Full text] [doi: 10.5334/gh.1278] [Medline: 38273995]
5. Thomas RJ, Beatty AL, Beckie TM, Brewer LC, Brown TM, Forman DE, et al. Home-based cardiac rehabilitation: a scientific statement from the American Association of Cardiovascular and Pulmonary Rehabilitation, the American Heart Association, and the American College of Cardiology. Circulation 2019 Jul 02;140(1):e69-e89 [FREE Full text] [doi: 10.1161/CIR.0000000000000663] [Medline: 31082266]
6. Paneroni M, Scalvini S, Corrà U, Lovagnini M, Maestri R, Mazza A, et al. The impact of cardiac rehabilitation on activities of daily life in elderly patients with heart failure. Front Physiol 2021;12:785501 [FREE Full text] [doi: 10.3389/fphys.2021.785501] [Medline: 35069247]
7. Pavy B, Iliou MC, Meurin P, Tabet JY, Corone S, Functional EvaluationCardiac Rehabilitation Working Group of the French Society of Cardiology. Safety of exercise training for cardiac patients: results of the French registry of complications during cardiac rehabilitation. Arch Intern Med 2006 Nov 27;166(21):2329-2334. [doi: 10.1001/archinte.166.21.2329] [Medline: 17130385]
8. Nkonde-Price C, Reynolds K, Najem M, Yang SJ, Batiste C, Cotter T, et al. Comparison of home-based vs center-based cardiac rehabilitation in hospitalization, medication adherence, and risk factor control among patients with cardiovascular disease. JAMA Netw Open 2022 Aug 01;5(8):e2228720 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.28720] [Medline: 36006642]
9. Ritchey MD, Maresh S, McNeely J, Shaffer T, Jackson SL, Keteyian SJ, et al. Tracking cardiac rehabilitation participation and completion among Medicare beneficiaries to inform the efforts of a national initiative. Circ Cardiovasc Qual Outcomes 2020 Jan;13(1):e005902 [FREE Full text] [doi: 10.1161/CIRCOUTCOMES.119.005902] [Medline: 31931615]
10. Park LG, Schopfer DW, Zhang N, Shen H, Whooley MA. Participation in cardiac rehabilitation among patients with heart failure. J Card Fail 2017 May;23(5):427-431 [FREE Full text] [doi: 10.1016/j.cardfail.2017.02.003] [Medline: 28232047]
11. Beatty AL, Truong M, Schopfer DW, Shen H, Bachmann JM, Whooley MA. Geographic variation in cardiac rehabilitation participation in medicare and veterans affairs populations: opportunity for improvement. Circulation 2018;137(18):1899-1908 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.117.029471] [Medline: 29305529]
12. McDonagh ST, Dalal H, Moore S, Clark CE, Dean SG, Jolly K, et al. Home-based versus centre-based cardiac rehabilitation. Cochrane Database Syst Rev 2023;10(10):CD007130. [doi: 10.1002/14651858.CD007130.pub5] [Medline: 37888805]
13. Ades PA, Keteyian SJ, Wright JS, Hamm LF, Lui K, Newlin K, et al. Increasing cardiac rehabilitation participation from 20% to 70%: a road map from the million hearts cardiac rehabilitation collaborative. Mayo Clin Proc 2017;92(2):234-242 [FREE Full text] [doi: 10.1016/j.mayocp.2016.10.014] [Medline: 27855953]
14. Thomas RJ, Petersen CE, Olson TP, Beatty AL, Ding R, Supervia M. Asynchronous and synchronous delivery models for home-based cardiac rehabilitation: A SCIENTIFIC REVIEW. J Cardiopulm Rehabil Prev 2021;41(6):407-412 [FREE Full text] [doi: 10.1097/HCR.0000000000000656] [Medline: 34727559]

15. Imran HM, Baig M, Erqou S, Taveira TH, Shah NR, Morrison A, et al. Home-based cardiac rehabilitation alone and hybrid with center-based cardiac rehabilitation in heart failure: a systematic review and meta-analysis. J Am Heart Assoc 2019;8(16):e012779 [FREE Full text] [doi: 10.1161/JAHA.119.012779] [Medline: 31423874]
16. Piotrowicz E, Pencina MJ, Opolski G, Zareba W, Banach M, Kowalik I, et al. Effects of a 9-week hybrid comprehensive telerehabilitation program on long-term outcomes in patients with heart failure: The telerehabilitation in heart failure patients (TELEREH-HF) randomized clinical trial. JAMA Cardiol 2020;5(3):300-308 [FREE Full text] [doi: 10.1001/jamacardio.2019.5006] [Medline: 31734701]
17. Piotrowicz E, Baranowski R, Bilinska M, Stepnowska M, Piotrowska M, Wójcik A, et al. A new model of home-based telemonitored cardiac rehabilitation in patients with heart failure: effectiveness, quality of life, and adherence. Eur J Heart Fail 2010;12(2):164-171 [FREE Full text] [doi: 10.1093/eurjhf/hfp181] [Medline: 20042423]
18. German CA, Baum SJ, Ferdinand KC, Gulati M, Polonsky TS, Toth PP, et al. Defining preventive cardiology: a clinical practice statement from the American society for preventive cardiology. Am J Prev Cardiol 2022;12:100432 [FREE Full text] [doi: 10.1016/j.ajpc.2022.100432] [Medline: 36425534]
19. Franklin BA, Eijsvogels TMH, Pandey A, Quindry J, Toth PP. Physical activity, cardiorespiratory fitness, and cardiovascular health: a clinical practice statement of the ASPC part i: bioenergetics, contemporary physical activity recommendations, benefits, risks, extreme exercise regimens, potential maladaptations. Am J Prev Cardiol 2022;12:100424 [FREE Full text] [doi: 10.1016/j.ajpc.2022.100424] [Medline: 36281324]
20. Lee KCS, Breznen B, Ukhova A, Koehler F, Martin SS. Virtual healthcare solutions for cardiac rehabilitation: a literature review. Eur Heart J Digit Health 2023;4(2):99-111 [FREE Full text] [doi: 10.1093/ehjdh/ztad005] [Medline: 36974268]
21. Ghisi GLDM, Xu Z, Liu X, Mola A, Gallagher R, Babu AS, et al. Impacts of the COVID-19 pandemic on cardiac rehabilitation delivery around the world. Glob Heart 2021;16(1):43 [FREE Full text] [doi: 10.5334/gh.939] [Medline: 34211829]
22. Subedi N, Rawstorn JC, Gao L, Koorts H, Maddison R. Implementation of telerehabilitation interventions for the self-management of cardiovascular disease: systematic review. JMIR Mhealth Uhealth 2020;8(11):e17957 [FREE Full text] [doi: 10.2196/17957] [Medline: 33245286]
23. Tosic L, Goldberger E, Maldaner N, Sosnova M, Zeitlberger AM, Staartjes VE, et al. Normative data of a smartphone app-based 6-minute walking test, test-retest reliability, and content validity with patient-reported outcome measures. J Neurosurg Spine 2020;33(4):480-489. [doi: 10.3171/2020.3.SPINE2084] [Medline: 32470938]
24. Beatty AL, Schiller NB, Whooley MA. Six-minute walk test as a prognostic tool in stable coronary heart disease: data from the heart and soul study. Arch Intern Med 2012;172(14):1096-1102 [FREE Full text] [doi: 10.1001/archinternmed.2012.2198] [Medline: 22710902]
25. Münzel T, Hahad O, Gori T, Hollmann S, Arnold N, Prochaska JH, et al. Heart rate, mortality, and the relation with clinical and subclinical cardiovascular diseases: results from the gutenberg health study. Clin Res Cardiol 2019;108(12):1313-1323 [FREE Full text] [doi: 10.1007/s00392-019-01466-2] [Medline: 30953178]
26. Müller-Nordhorn J, Roll S, Willich SN. Comparison of the short form (SF)-12 health status instrument with the SF-36 in patients with coronary heart disease. Heart 2004;90(5):523-527 [FREE Full text] [doi: 10.1136/hrt.2003.013995] [Medline: 15084550]
27. Gremeaux V, Troisgros O, Benaïm S, Hannequin A, Laurent Y, Casillas JM, et al. Determining the minimal clinically important difference for the six-minute walk test and the 200-meter fast-walk test during cardiac rehabilitation program in coronary artery disease patients after acute coronary syndrome. Arch Phys Med Rehabil 2011;92(4):611-619 [FREE Full text] [doi: 10.1016/j.apmr.2010.11.023] [Medline: 21440707]
28. Dwivedi AK, Mallawaarachchi I, Alvarado LA. Analysis of small sample size studies using nonparametric bootstrap test with pooled resampling method. Stat Med 2017;36(14):2187-2205. [doi: 10.1002/sim.7263] [Medline: 28276584]
29. Misra S, Niazi K, Swayampakala K, Blackmon A, Lang M, Davenport E, et al. Outcomes of a remote cardiac rehabilitation program for patients undergoing atrial fibrillation ablation: pilot study. JMIR Cardio 2023;7:e49345 [FREE Full text] [doi: 10.2196/49345] [Medline: 38096021]
30. Giggins OM, Doyle J, Smith S, Vavasour G, Moran O, Gavin S, et al. Remotely delivered cardiac rehabilitation exercise for coronary heart disease: nonrandomized feasibility study. JMIR Cardio 2023;7:e40283 [FREE Full text] [doi: 10.2196/40283] [Medline: 36763453]
31. Chen W, Feng Y, Yu M, Zhang Z, Wu J, Liu W, et al. Effects of different rehabilitation modality on cardiopulmonary function in patients with acute coronary syndrome after revascularization. Front Cardiovasc Med 2024;10:1120665 [FREE Full text] [doi: 10.3389/fcvm.2023.1120665] [Medline: 38500679]
32. Sommer CG, Jørgensen LB, Blume B, Møller T, Skou ST, Harrison A, et al. Dropout during a 12-week transitional exercise-based cardiac rehabilitation programme: a mixed-methods prospective cohort study. Eur J Cardiovasc Nurs 2022;21(6):578-586 [FREE Full text] [doi: 10.1093/eurjcn/zvab119] [Medline: 35043160]
33. Hamilton SJ, Mills B, Birch EM, Thompson SC. Smartphones in the secondary prevention of cardiovascular disease: a systematic review. BMC Cardiovasc Disord 2018;18(1):25 [FREE Full text] [doi: 10.1186/s12872-018-0764-x] [Medline: 29415680]

34. Ni R, Liu M, Huang S, Yang J. Effects of eHealth interventions on quality of life and psychological outcomes in cardiac surgery patients: systematic review and meta-analysis. J Med Internet Res 2022;24(8):e40090 [FREE Full text] [doi: 10.2196/40090] [Medline: 35972792]
35. Yu T, Xu H, Sui X, Zhang X, Pang Y, Yu T, et al. Effectiveness of eHealth interventions on moderate-to-vigorous intensity physical activity among patients in cardiac rehabilitation: systematic review and meta-analysis. J Med Internet Res 2023;25:e42845 [FREE Full text] [doi: 10.2196/42845] [Medline: 36989017]
36. Rawstorn JC, Ball K, Oldenburg B, Chow CK, McNaughton SA, Lamb KE, et al. Smartphone cardiac rehabilitation, assisted self-management versus usual care: protocol for a multicenter randomized controlled trial to compare effects and costs among people with coronary heart disease. JMIR Res Protoc 2020;9(1):e15022. [doi: 10.2196/15022] [Medline: 32012103]
37. Dwiputra B, Santoso A, Purwowiyoto BS, Radi B, Ambari AM, Desandri DR, et al. Smartphone-based cardiac rehabilitation program improves functional capacity in coronary heart disease patients: a systematic review and meta-analysis. Glob Heart 2023;18(1):42 [FREE Full text] [doi: 10.5334/gh.1253] [Medline: 37577291]
38. Heizmann A, Chapelle C, Laporte S, Roche F, Hupin D, Le Hello C. Impact of wearable device-based interventions with feedback for increasing daily walking activity and physical capacities in cardiovascular patients: a systematic review and meta-analysis of randomised controlled trials. BMJ Open 2023;13(7):e069966 [FREE Full text] [doi: 10.1136/bmjopen-2022-069966] [Medline: 37433730]
39. Seron P, Oliveros MJ, Marzuca-Nassr GN, Morales G, Román C, Muñoz SR, HYCARET Investigators. Hybrid cardiac rehabilitation program in a low-resource setting: a randomized clinical trial. JAMA Netw Open 2024;7(1):e2350301 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.50301] [Medline: 38194236]
40. Isakadze N, Kim CH, Marvel FA, Ding J, MacFarlane Z, Gao Y, et al. Rationale and design of the mTECH-rehab randomized controlled trial: impact of a mobile technology enabled corrie cardiac rehabilitation program on functional status and cardiovascular health. J Am Heart Assoc 2024;13(2):e030654 [FREE Full text] [doi: 10.1161/JAHA.123.030654] [Medline: 38226511]
41. Wills WB, Athilingam P, Beckie TM. Exercise-based cardiac rehabilitation in women with heart failure: a review of enrollment, adherence, and outcomes. Heart Fail Rev 2023;28(6):1251-1266. [doi: 10.1007/s10741-023-10306-5] [Medline: 37059937]
42. Zhong W, Fu C, Xu L, Sun X, Wang S, He C, et al. Effects of home-based cardiac telerehabilitation programs in patients undergoing percutaneous coronary intervention: a systematic review and meta-analysis. BMC Cardiovasc Disord 2023;23(1):101 [FREE Full text] [doi: 10.1186/s12872-023-03120-2] [Medline: 36814188]
43. Cruz-Cobo C, Bernal-Jiménez M, Vázquez-García R, Santi-Cano MJ. Effectiveness of mHealth interventions in the control of lifestyle and cardiovascular risk factors in patients after a coronary event: systematic review and meta-analysis. JMIR Mhealth Uhealth 2022;10(12):e39593 [FREE Full text] [doi: 10.2196/39593] [Medline: 36459396]
44. Resurrección DM, Motrico E, Rigabert A, Rubio-Valera M, Conejo-Cerón S, Pastor L, et al. Barriers for nonparticipation and dropout of women in cardiac rehabilitation programs: a systematic review. J Womens Health (Larchmt) 2017;26(8):849-859. [doi: 10.1089/jwh.2016.6249] [Medline: 28388314]
45. Valencia HE, Savage PD, Ades PA. Cardiac rehabilitation participation in underserved populations. Minorities, low socioeconomic, and rural residents. J Cardiopulm Rehabil Prev 2011;31(4):203-210. [doi: 10.1097/HCR.0b013e318220a7da] [Medline: 21705915]
46. Zhang L, Sobolev M, Piña IL, Prince DZ, Taub CC. Predictors of cardiac rehabilitation initiation and adherence in a multiracial Urban population. J Cardiopulm Rehabil Prev 2017;37(1):30-38. [doi: 10.1097/HCR.0000000000000226] [Medline: 28005679]
47. Gao L, Maddison R, Rawstorn J, Ball K, Oldenburg B, Chow C, et al. Economic evaluation protocol for a multicentre randomised controlled trial to compare smartphone cardiac rehabilitation, assisted self-management (SCRAM) versus usual care cardiac rehabilitation among people with coronary heart disease. BMJ Open 2020;10(8):e038178 [FREE Full text] [doi: 10.1136/bmjopen-2020-038178] [Medline: 32847918]

Abbreviations

6MWT: 6-minute walk test
CBCR: center-based cardiac rehabilitation
CR: cardiac rehabilitation
CVD: cardiovascular disease
HBCR: home-based cardiac rehabilitation
HIPAA: Health Insurance Portability and Accountability Act
mHealth: mobile health
RPH-D: RecoveryPlus.Health Digital
SF-12: 12-Item Short-Form Health Survey

Edited by L Buis; submitted 08.04.24; peer-reviewed by MJ Santi, Z Wang; comments to author 05.07.24; revised version received 20.07.24; accepted 12.08.24; published 01.10.24.

Please cite as:
Bilbrey T, Martin J, Zhou W, Bai C, Vaswani N, Shah R, Chokshi S, Chen X, Bhusri S, Niemi S, Meng H, Lei Z
A Dual-Modality Home-Based Cardiac Rehabilitation Program for Adults With Cardiovascular Disease: Single-Arm Remote Clinical Trial
JMIR Mhealth Uhealth 2024;12:e59098
URL: https://mhealth.jmir.org/2024/1/e59098
doi:10.2196/59098
PMID:39150858

Original Paper

Patient Engagement in a Mobile App–Based Rehabilitation Program for Total Hip or Knee Arthroplasty: Secondary Data Analysis of a Randomized Controlled Trial

Qingling Wang[1,2], PhD; Regina Lai-Tong Lee[2,3], PhD; Sharyn Hunter[2], PhD; Aiyong Zhu[1], MD; Sally Wai-Chi Chan[4], PhD

[1]School of Nursing and Health Management, Shanghai University of Medicine and Health Sciences, Shanghai, China

[2]School of Nursing and Midwifery, College of Health, Medicine and Wellbeing, University of Newcastle, Newcastle, Australia

[3]The Nethersole School of Nursing, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong, China (Hong Kong)

[4]Tung Wah College, Hong Kong, China (Hong Kong)

Corresponding Author:
Qingling Wang, PhD
School of Nursing and Health Management
Shanghai University of Medicine and Health Sciences
No. 279 Zhouzhu Highway
Pudong New Area
Shanghai, 201318
China
Phone: 86 13918170613
Email: my_wql0719@163.com

Abstract

Background: Health care professionals use mobile apps to support patients' rehabilitation after total hip or knee arthroplasty. Understanding patient engagement in such mobile health interventions can help tailor these interventions to better support patients.

Objective: This study aimed to investigate patient engagement in a mobile app–based arthroplasty rehabilitation program and to investigate the association between patient engagement and their characteristics.

Methods: Data were extracted from a pool of 42 participants in the experimental arm of a randomized controlled trial that used a mobile app (WeChat [Tencent Holdings Limited])–based program to support patients' rehabilitation after total hip or knee arthroplasty. The primary outcomes were the number of days the participants accessed the program and completed recommended rehabilitation tasks. Secondary outcomes included data on the participants' posts on a discussion forum, messages sent by the participants, access to the program components, and reading and sharing the program content. Generalized linear models were used to analyze the association between patient engagement and personal characteristics.

Results: The participants reported in a rehabilitation diary accessing the program on a mean of 5.2 (SD 2) days per week and completing recommended rehabilitation tasks on a mean of 6.5 (SD 0.8) days per week. The majority (31/42, 74%) posted on the discussion forum, with a mean of 18.1 (SD 21.2) posts. Most participants (37/42, 88%) sent messages to health care professionals, with a mean of 14 (SD 15.9) messages. The program components were visited for a total of 525 times. The program content was read 898 times and shared 82 times in total. Generalized linear models showed that both primary outcomes, the number of days the participants accessed the program (B=6.46, 95% CI 1.98-15.35; χ^2_1=11.1, *P*=.001) and the number of days they completed rehabilitation tasks (B=2.65, 95% CI 0.45-5.48; χ^2_1=5.7, *P*=.02), were positively associated with having a high school education or above. In addition, the number of posts on the discussion forum was positively associated with living with family, having a high school education or above, undergoing total knee arthroplasty, having comorbidities, and the score of self-efficacy but was negatively associated with age. The number of messages sent by the participants was positively associated with having a high school education or above, having comorbidities, and the score of self-efficacy.

Conclusions: Patient engagement in mobile arthroplasty rehabilitation is associated with their education level, cohabitation status, age, type of surgery, presence of comorbidities, and sense of self-efficacy. Program developers can consider these characteristics and use strategies, such as family involvement, in the design of mobile arthroplasty rehabilitation programs to enhance patient engagement in such interventions.

Trial Registration: Australian New Zealand Clinical Trials Registry ACTRN12621000867897; https://tinyurl.com/mtdw25fp

(JMIR Mhealth Uhealth 2024;12:e57635) doi:10.2196/57635

KEYWORDS

total hip arthroplasty; total knee arthroplasty; rehabilitation; mobile health; social media application; patient engagement

Introduction

Over 3 million total hip arthroplasty (THA) and total knee arthroplasty (TKA) procedures are performed globally each year [1]. Patients after THA and TKA require rehabilitation to relieve pain, enhance muscle strength, improve mobility, optimize functional outcomes such as walking, and improve health-related quality of life [2,3]. However, patients may have difficulty accessing face-to-face rehabilitation services after arthroplasty due to the geographic distance from rehabilitation facilities or lack of local health resources [4]. This is particularly true in China due to the lack of rehabilitation facilities and qualified health care professionals [5]. With technological advancement, health care professionals can use information and communication technologies, such as mobile apps, to complement or as an alternative to face-to-face services, to support patients' rehabilitation after THA and TKA [6]. Studies have investigated the effectiveness of using mobile apps to support patients' rehabilitation after arthroplasty and reported positive effects of app-based rehabilitation on health outcomes such as pain, range of motion, physical function, and health-related quality of life [7,8]. However, more evidence from robust randomized controlled trials based on physical and psychological outcomes is needed to support the implementation of mobile rehabilitation in arthroplasty populations [9].

Mobile apps used to provide arthroplasty rehabilitation interventions include those specifically designed for target populations [7], commercial medical apps [10,11], and social media apps such as WeChat (Tencent Holdings Limited) [8]. Social media apps are used by many health care providers due to their high penetration, good interactivity, ease of use, and low cost [12]. In China, WeChat is the most popular free social media platform [13] and has been used to provide health interventions, such as weight control and hypertension management [14,15]. We developed a mobile app (WeChat)–based program that was based on theories [16,17] and patients' perceived needs [18] to support patients' rehabilitation in 6 weeks after THA and TKA. A randomized controlled trial was conducted to investigate the effectiveness of the program, and the results showed that the program improved patients' self-efficacy, physical function, and health-related quality of life and reduced levels of anxiety and depression [19]. Study participants reported that the program facilitated their access to health care, encouraged their postoperative recovery, supported the relationships with health care professionals and other patients in the program, and facilitated their learning during postarthroplasty rehabilitation [20]. However, the extent to which patients were engaged in the mobile app–based rehabilitation program and how this engagement was associated with patients' characteristics were not investigated.

It is important to understand patient engagement in mobile health interventions because interventions are more efficient and effective when patients are better engaged [21]. In mobile health, patient engagement is defined as "...the ability for apps to enable collaboration, activation and participation, information-sharing, and decision-making in one's own health" [22]. Patient engagement in mobile health is often indicated by the continued use of an app; the intensity of using a function of the app; information dissemination through the app, such as reading and sharing content; and patients' performance, such as reactions and comments on posts in a social media app [23-25]. Collection and analysis of data on patient engagement will facilitate the understanding of patients' preferences for using mobile health programs, disparities among populations, and barriers to engagement, which will further inform the improvements in the design of mobile health programs as well as strategies to enhance patient engagement in mobile health [24,25].

Patient engagement in mobile arthroplasty rehabilitation has not been extensively explored. Previous studies analyzed usage data (eg, log-in frequency) automatically recorded by commercial medical apps [11] or apps specifically designed for arthroplasty populations [7,26] and reported that 11% to 14% of patients did not adhere to the mobile rehabilitation programs recommended by health care professionals. One reason was that patients could not interact with the program because of technical challenges, such as the need to set up sensors and touch the device multiple times to start a session [7,26]. However, a research gap exists in patient engagement in social media–based health interventions, requiring further studies to explore engagement assessment as well as strategies to promote patient engagement [27]. For example, measuring patient engagement in WeChat-based interventions may differ from that based on specifically designed or commercial medical apps that could track log-in data, as WeChat's privacy policy does not allow researchers to access individual data on log-in frequency and duration [28]. A self-reported rehabilitation diary and counting data (eg, number of posts and messages) may provide information about patient engagement in such social media app–based programs [29].

Furthermore, evidence of the association between patient engagement in mobile health and personal characteristics is inconclusive. For example, a study on a mobile program to support medication adherence reported that age was positively associated with the duration of patients' usage of the program, while gender and education level had no association with either duration or intensity of usage [30]. Another study found that age, education level, family monthly income, and employment status were associated with the usage duration and log-in frequency in a mobile breast cancer support program [31]. In mobile arthroplasty rehabilitation, patient engagement may be

influenced by factors such as age, pain, and limited physical function after surgery [32,33]. In addition, surgical procedures, such as THA and TKA, may also affect patients' use of social media for health purposes [34]. Elucidating the association between patient engagement and their characteristics can guide the development of programs that are delivered through social media platforms to make them acceptable to patients and thus effective in supporting patients' rehabilitation [35].

This study aimed to investigate patient engagement in a social media app–based rehabilitation program for THA and TKA by analyzing data that had been collected in a randomized controlled trial. The objectives were to (1) investigate patient engagement in the 6-week mobile arthroplasty rehabilitation program through data collected from a written rehabilitation diary, data on the patient's performance in the program, and data tracked by the app's background system; and (2) examine the association between patient engagement and their demographic and clinical characteristics.

Methods

Study Participants

This study extracted data from the pool of 43 participants in the experimental arm of a single-center, parallel-group, randomized controlled trial conducted at a university hospital in Shanghai, China. All participants were adults (aged ≥18 years) who (1) were discharged home after a unilateral primary THA or TKA and (2) had access to a 6-week, mobile app–based rehabilitation program after hospital discharge. Patients were excluded from this study if they could not practice recommended rehabilitation exercises due to major complications or serious health conditions such as heart failure or had severe vision impairment.

Description of the Mobile App–Based Rehabilitation Program

A 6-week, theory-underpinned rehabilitation program was provided to study participants through the app WeChat, aiming to enhance their self-efficacy during rehabilitation and improve their rehabilitation practice and outcomes. The program provided exercise demonstration videos that were designed for patients after THA and TKA. The participants were suggested to follow the videos and practice for 1 hour per day and at least 5 days per week. In addition, program components were designed to enhance patients' self-efficacy for rehabilitation, such as setting weekly goals for rehabilitation; scheduling progressive rehabilitation tasks; providing short stories and videos to share the experiences of previous patients; holding a discussion forum that involved health care professionals, patients, and researchers through the app; and providing psychological techniques such as relaxation exercises. To facilitate the participants' learning process, educational materials in the program were designed using visual presentations with short text messages, and a paper booklet on the use of the program was provided in addition to verbal explanations. A detailed description of the program has been published [36].

The mobile app–based rehabilitation program was carried out between May 2021 and January 2022. Educational materials such as demonstration videos were uploaded to the app (WeChat) in advance. The participants could learn at home through WeChat installed on their own mobile devices. To improve participants' engagement in the program, the researchers sent an exercise reminder every Monday, and the participants were encouraged to record their usage of the program and completion of recommended rehabilitation tasks in a written rehabilitation diary.

Patient Engagement in the Program

Patient engagement in the mobile rehabilitation program was evaluated using 3 measures. The first measured the number of days the participants accessed the program and the number of days the participants completed recommended rehabilitation tasks. These data were obtained by reviewing the participants' records in their rehabilitation diary. The total number of days and the number of days per week were calculated for each participant. The second measure counted the number of participants' posts on the discussion forum and the number of messages sent by the participants to health care professionals. The percentages of participants posting on the discussion forum and sending messages were calculated, as well as the number of posts and messages for each participant. The third measure focused on the data from the background system of the app, which tracked how many times the content from the program was read and shared by the participants and how often the program components were visited.

The primary outcomes of this study were the number of days the participants accessed the program and the number of days they completed the recommended rehabilitation tasks as recorded in the paper diary. These outcomes reflected individual engagement in the mobile rehabilitation program. Secondary outcomes focused on engagement in a part of the program, including the number of posts made by the participants on the discussion forum, the number of messages they sent to health care professionals through the app, the total number of visits to the program components, and the total number of times the content was read and shared.

Demographic and Clinical Characteristics

Before starting the mobile rehabilitation program, study participants completed a paper questionnaire that collected their demographic and clinical characteristics. Demographic data included age, gender, BMI, residence, cohabitation status, education level, and employment. Clinical data included the type of surgery; reasons for surgery; comorbidities; long-term use of medication; and health assessment outcomes such as self-efficacy, self-reported physical function, severity of pain, levels of anxiety and depression, and health-related quality of life. Self-efficacy was measured using the Chinese version of the Self-Efficacy for Rehabilitation Outcome Scale, a 12-item scale that measures the patient's belief in their ability to perform physical rehabilitation behaviors after hip or knee surgery [37,38]. Patient-reported physical function was measured using the Chinese version of the Hip Disability and Osteoarthritis Outcome Score Physical Function Short Form for THA and the Knee Injury and Osteoarthritis Outcome Score Physical Function Short Form for TKA [39,40]. The severity of pain was measured using the Numeric Rating Scale for Pain [41]. Levels of anxiety and depression were measured using the Chinese version of the

Hospital Anxiety and Depression Scale [42]. The patient's health-related quality of life was measured using the Chinese version of the 5-level EQ-5D version developed by the EuroQol Group [43].

Statistical Analysis

Data were analyzed using IBM SPSS Statistics (version 28.0; IBM Corp) [44]. Numerical data were described using means with SDs and medians with IQRs. Categorical data were described using frequencies and percentages. In association analyses, data on the number of days accessing the program and the number of days completing rehabilitation tasks were categorized as "equal to or greater than 30 days" and "less than 30 days or missing data" since the participants had been suggested to practice rehabilitation at least 5 days per week (30 days in total) with the support from the program.

As the outcome variables were categorical data (eg, 2 categories of the number of days accessing the program) or counting data (eg, the number of posts and messages), generalized linear models were used to analyze the association between patient engagement and their demographic and clinical characteristics. Pearson correlation analysis was performed for explanatory variables, and pairs of variables with correlation coefficients greater than 0.7 (age and employment as well as comorbidities and long-term use of medication) were discussed within the research team. A decision was made to remove 2 variables, employment and long-term use of medication, from the model, as they might be due to highly correlated variables. A collinearity diagnostic was conducted for the remaining explanatory variables. The results showed that the tolerance of the variables was greater than 0.2 and the variance inflation factor was less than 5, indicating that there was no significant collinearity between the variables. Therefore, all remaining explanatory variables (ie, age, gender, BMI, residence, cohabitation status, education level, type of surgery, reasons for surgery, comorbidities, self-efficacy, self-reported physical function, severity of pain, levels of anxiety and depression, and health-related quality of life) were included in generalized linear models. Considering the coprimary outcomes, the statistical significance was set as an α level of .025.

Ethical Considerations

Ethical approval to conduct this study was obtained from the Human Research Ethics Committee of the University of Newcastle in Australia (reference H-2020-0414) and from the study hospital (reference B2021-096R). All participants signed a written informed consent before they were enrolled in the study. The original approval covers secondary data analysis without additional consent.

Results

Participant Characteristics

As 1 patient dropped out from the study within 1 week after group allocation and did not participate in the mobile rehabilitation program, this study included 42 participants in total. Women comprised 74% (31/42) of the participants, and the age ranged from 33 to 83 (mean 66.7, SD 10.4) years. The participants had undergone THA (23/42, 55%) or TKA (19/42, 45%). Most participants underwent surgery because of osteoarthritis (35/42, 83%), followed by acute fractures (5/42, 12%) and other reasons (2/42, 5%). The majority (30/42, 71%) had chronic diseases, such as hypertension (19/42, 45%) and diabetes mellitus (9/42, 21%), and were taking medication for these diseases. Most participants (39/42, 93%) lived with their families after THA and TKA surgery, and 90% (38/42) of the participants were unemployed or retired at the time of the study. In terms of education levels, 24% (10/42) had completed education at university or college, 31% (13/42) had completed high school education or had equivalent learning experience, 29% (12/42) had completed middle school education, and 17% (7/42) had primary school education or lower.

Patient Engagement in the Mobile App–Based Rehabilitation Program

The majority of study participants (26/42, 62%) returned their completed rehabilitation diary to health care professionals. In the diaries, 65% (17/26) reported accessing the program in 30 or more days, and 96% (25/26) reported completing recommended rehabilitation tasks in 30 or more days. The number of days that the participants accessed the rehabilitation program ranged from 1.3 to 7 (mean 5.2, SD 2; median 5.8, IQR 4.1) days per week, and the number of days that the participants completed recommended rehabilitation tasks ranged from 3.5 to 7 (mean 6.5, SD 0.8; median 7, IQR 0.8) days per week. Table 1 presents descriptive statistics on participant-reported access to the program and completion of rehabilitation tasks, and an exploratory analysis of changes in these data over 6 weeks is provided in Multimedia Appendix 1.

As shown in Table 1, the majority (31/42, 74%) posted questions or comments on the discussion forum in the app. The number of posts ranged from 1 to 70 (mean 18.1, SD 21.2; median 10, IQR 22). Most participants (37/42, 88%) sent messages to health care professionals. The number of messages ranged from 1 to 69 (mean 14, SD 15.9; median 10, IQR 11). The app's background system showed that the participants read the educational materials in the program 898 times, with a mean of 21.4 times per person, and shared these materials 82 times, with a mean of 2 times per person. The most frequently visited program components were related to the rehabilitation after arthroplasty, such as exercise demonstration videos (243/525, 46%), precautions after the surgery (100/525, 19%), and follow-up schedules (84/525, 16%). They are followed by the stories and short videos shared by previous patients (55/525, 11%) and psychological support components (43/525, 8%).

Table 1. Patient engagement in the mobile app–based rehabilitation program for total hip or knee arthroplasty during the 6-week intervention.

Patient engagement in the program	Value, min[a]	Value, max[b]	Value, mean (SD; 95% CI)	Value, median (IQR; 95% CI)
Number of days accessing the program (n=26; d/wk)	1.3	7	5.2 (2; 4.4-6.0)	5.8 (4.1; 4.4-6.9)
Number of days completing rehabilitation tasks (n=26; d/wk)	3.5	7	6.5 (0.8; 6.2-6.8)	7 (0.8; 6.7-7.0)
Number of posts on the discussion forum (n=31; times)	1	70	18.1 (21.2; 10.4-25.9)	10 (22; 5-16)
Number of messages sent by the participants (n=37; times)	1	69	14 (15.9; 8.7-19.3)	10 (11; 5-14)

[a]Min: minimum.

[b]Max: maximum.

Association Between Patient Engagement, Demographic, and Clinical Characteristics

Table 2 shows the association between patient engagement in the mobile app–based rehabilitation program and their demographic and clinical characteristics. The number of days that the participants reported accessing the program was positively associated with having a high school education or above (B=6.46, 95% CI 1.98-15.35; χ^2_1=11.1, *P*=.001). The number of days that the participants reported completing rehabilitation tasks was positively associated with having a high school education or above (B=2.65, 95% CI 0.45-5.48; χ^2_1=5.7, *P*=.02). The number of posts that the participants posted on the discussion forum was negatively associated with age (B=–0.09, 95% CI –0.16 to –0.02; χ^2_1=6.6, *P*=.01) but positively associated with living with family (B=2.74, 95% CI 1.42-4.06; χ^2_1=16.6, *P*<.001), having a high school education or above (B=2.46, 95% CI 1.22-3.69; χ^2_1=15.1, *P*<.001), undergoing TKA surgery (B=1.34, 95% CI 0.35-2.33; χ^2_1=7, *P*=.008), having comorbidities (B=2.13, 95% CI 0.32-3.95; χ^2_1=5.3, *P*=.02), and the score of self-efficacy (B=0.33, 95% CI 0.13-0.53; χ^2_1=10.2, *P*=.001). The number of messages sent by the participants was positively associated with having a high school education or above (B=0.90, 95% CI 0.18-1.62; χ^2_1=6, *P*=.01), having comorbidities (B=1.21, 95% CI 0.20-2.23; χ^2_1=5.5, *P*=.02), and the score of self-efficacy (B=0.22, 95% CI 0.05-0.39; χ^2_1=6.2, *P*=.01).

Table 2. Associations between patient engagement in the mobile app–based rehabilitation program and demographic and clinical characteristics (n=42).

Variables	Number of days accessing the program[a,b]		Number of days completing rehabilitation tasks[a,c]		Number of posts on the discussion forum		Number of messages sent by the participants	
	Chi-square (*df*)	*P* value	Chi-square (*df*)	*P* value	Chi-square (*df*)	*P* value	Chi-square (*df*)	*P* value
Age	0.1 (1)	.72	0.1 (1)	.75	6.6 (1)	.01[d]	4.5 (1)	.03
Gender[a,e]	2.2 (1)	.14	0.3 (1)	.60	0 (1)	.99	0.8 (1)	.38
BMI	0.1 (1)	.75	2.4 (1)	.12	1.2 (1)	.27	0 (1)	.97
Residence[a,f]	0.2 (1)	.68	0.1 (1)	.75	0.1 (1)	.72	0 (1)	.98
Cohabitation status[a,g]	0.5 (1)	.47	0 (1)	.84	16.6 (1)	<.001[d]	3.6 (1)	.06
Education level[a,h]	11.1 (1)	.001[d]	5.7 (1)	.02[d]	15.1 (1)	<.001[d]	6 (1)	.01[d]
Types of surgery[a,i]	1.2 (1)	.28	0.1 (1)	.72	7 (1)	.008[d]	0.3 (1)	.57
Reasons for surgery[a,j]	1.8 (1)	.18	3.6 (1)	.06	0 (1)	.99	4.3 (1)	.04
Comorbidities[a,k]	2.7 (1)	.10	0.3 (1)	.61	5.3 (1)	.02[d]	5.5 (1)	.02[d]
Health outcomes								
SER[l]	0.1 (1)	.75	0 (1)	.99	10.2 (1)	.001[d]	6.2 (1)	.01[d]
HOOS-PS[m] or KOOS-PS[n]	4.7 (1)	.03	0.1 (1)	.71	0.1 (1)	.71	0.5 (1)	.47
NRS[o] Pain	1 (1)	.32	0 (1)	.99	4.2 (1)	.04	4 (1)	.05
HADS[p]-anxiety	4.2 (1)	.04	0 (1)	.94	0 (1)	.94	0.4 (1)	.52
HADS-depression	2.3 (1)	.13	0 (1)	.99	0.8 (1)	.37	4.6 (1)	.03
EQ-5D index	4.5 (1)	.03	2.2 (1)	.14	0.5 (1)	.46	1.4 (1)	.23
EQ-VAS[q]	4.4 (1)	.04	1.4 (1)	.24	3.6 (1)	.06	1.7 (1)	.19

[a]Grouped into categories. Ascending category order was used in data analysis.

[b]Number of days accessing the program: “1=equal to or greater than 30 days” and “2=less than 30 days or missing data.”

[c]Number of days completing rehabilitation tasks: “1=equal to or greater than 30 days” and “2=less than 30 days or missing data.”

[d]Significant at an α level of .025.

[e]Gender: “1=men” and “2=women.”

[f]Residence: “1=in Shanghai” and “2=in other cities.”

[g]Cohabitation status: “1=living with family” and “2=living alone.”

[h]Education level: “1=high school and higher” and “2=middle school and lower.”

[i]Types of surgery: “1=total knee arthroplasty” and “2=total hip arthroplasty.”

[j]Reasons for surgery: “1=osteoarthritis” and “2=other reasons.”

[k]Comorbidities: “1=yes” and “2=none.”

[l]SER: Self-Efficacy for Rehabilitation Outcome Scale.

[m]HOOS-PS: Hip Disability and Osteoarthritis Outcome Score Physical Function Short Form.

[n]KOOS-PS: Knee Injury and Osteoarthritis Outcome Score Physical Function Short Form.

[o]NRS: Numeric Rating Scale.

[p]HADS: Hospital Anxiety and Depression Scale.

[q]VAS: Visual Analogue Scale.

Discussion

Principal Findings

This study investigated patient engagement in a social media app (WeChat)–based arthroplasty rehabilitation program by analyzing data obtained from the app (eg, posts, messages, and reading and sharing of program content) and a written rehabilitation diary (eg, days patients accessed the program and completed recommended rehabilitation tasks). The proportions of patients who accessed the WeChat–based program and who completed recommended rehabilitation tasks are comparable with that reported in previous arthroplasty rehabilitation programs using specifically designed apps or commercial medical apps [7,11,26]. Usage data, such as the number of days per week the patients accessed the program, the average number of posts on the discussion forum, and the average number of messages sent by the patients, are also similar to that in a TKA

rehabilitation program based on a commercial medical app [10]. The previous programs were considered to have high levels of patient acceptance and engagement [10,26]. These findings provide evidence that arthroplasty rehabilitation delivered by WeChat has good patient engagement. Furthermore, the findings from this study suggest that patient engagement in mobile arthroplasty rehabilitation is associated with demographic characteristics such as age, education level, and cohabitation status, as well as clinical characteristics such as type of surgery, presence of comorbidities, and sense of self-efficacy. This adds to the body of evidence about the relationship between patient engagement in mobile arthroplasty rehabilitation and personal characteristics.

Education level was found positively associated with all 4 indicators of patient engagement in this study. This supports the evidence that education levels influence patients' usage of mobile health technologies. A previous study reported that women with higher levels of education had more frequent log-ins and longer usage durations in an app–based breast cancer support program [31]. The authors explained that this was because women with higher levels of education were more receptive to new knowledge and had a greater capacity for understanding health information. In addition to these reasons, patients with higher education levels may be more familiar with the access and use of mobile health technologies, which may have facilitated the engagement of this group of patients in our study. Conversely, patients with lower education levels may need support to facilitate engagement in mobile health care. Program developers could consider using visual presentations, such as videos and pictures, instead of written materials, and plain language instead of medical jargon to improve engagement in mobile health interventions in populations with lower levels of education.

This study found that patients who lived with their families after surgery posted more frequently on the discussion forum held through the app. Because of cultural contexts such as filial piety and kinship, patients after arthroplasty in China tend to live with their families. Family members are usually the primary caregivers of the patients, participating in and influencing the patient's decisions about treatment and care [45]. In our study, family members who lived with the patients may have aided the patients' usage of the mobile app–based rehabilitation program, such as technical assistance, which may have facilitated patient engagement in the program. Patients who have experienced mobile app–based arthroplasty rehabilitation reported that they perceived involving family members would help them better use the programs [20]. Family involvement is also considered important for postarthroplasty recovery in other cultural contexts. One study conducted in 6 European countries recommended the inclusion of family members in health education to fulfill their knowledge expectations and better support patients' postarthroplasty rehabilitation [46]. Therefore, we recommend including family members in mobile arthroplasty rehabilitation programs, such as giving family members access to health information provided in the program and enabling their involvement in discussions. This will help family members understand arthroplasty rehabilitation and support the patients to be more engaged in the rehabilitation program.

In terms of type of surgery, this study found that patients after TKA were more engaged in the discussion forum than patients after THA. It has been reported that THA procedures offer patients better outcomes on pain relief, daily activity performance, expectations met, and overall satisfaction with the surgery than TKA [47]. Approximately 10% (2,754/25,012) of patients undergoing TKA are dissatisfied because of factors such as unmet expectations and persistent pain [48]. Patients after TKA thus may have more concerns and questions than those after THA about their postarthroplasty recovery and need more support during rehabilitation. The discussion forum of the current program enabled patients to obtain timely support from health care professionals and fellow patients undergoing similar procedures. When investigating the effectiveness of the mobile rehabilitation program, this support was confirmed by the patients [20]. Future studies could investigate the specific support required by patients after TKA and solutions to address the requirements so that mobile rehabilitation interventions are tailored for this population.

In our study, patients with comorbidities, such as hypertension and diabetes mellitus, were more engaged in the discussion forum and sending messages. Patients with such chronic diseases may have developed skills to access health information and support and embraced technology to help them understand and manage their chronic conditions [30,49]. These skills may have facilitated their engagement in the current mobile rehabilitation program for arthroplasty. For instance, in a previous study, a patient who underwent THA with systemic lupus erythematosus said that they were familiar with seeking health information about their condition from the internet, but they preferred to use the mobile rehabilitation program that was specifically designed for patients undergoing arthroplasty as they found that the health information provided by health care professionals in the program was more comprehensive and reliable than that available on the internet [20].

Patients' self-efficacy scores were found positively associated with the number of posts they posted on the discussion forum and the number of messages they sent to health care professionals. A systematic review reported that self-efficacy as a predictor of patient engagement in online health interventions was inconclusive, as few studies had explored the relationship [50]. Theoretically, patients with a higher sense of self-efficacy are more confident in accomplishing rehabilitation tasks [16]. These patients may be more inclined to view difficulties arising during rehabilitation as challenges that they should put more effort into or seek support from health care professionals and others [51]. In this study, patients were provided with the opportunity to obtain professional and peer support through the discussion forum held through WeChat and sending messages to health care professionals. This may explain why patients with higher self-efficacy scores were more engaged in these 2 components of the program. Future mobile rehabilitation programs could consider identifying patients with a lower sense of self-efficacy and providing support to facilitate their engagement in the program. The theory of self-efficacy [16] could be used to guide the individualization of mobile rehabilitation interventions.

However, this study found that patient engagement, such as posting on the discussion forum, was negatively associated with age. This is consistent with the findings of a previous systematic review that older age is associated with decreased engagement in mobile technology–based monitoring for TKA [32]. Physiological declines with advancing age, such as decreased visual perception and hearing loss, may hinder using mobile technology–based health interventions in older populations [52]. Low technology proficiency was reported as a barrier to engagement in eHealth in older adults [53], but other researchers have also claimed that advanced age is not a major issue in the use of mobile health technologies if adequate orientation and coaching are provided [54]. Zhu et al [31] found that older women were even more engaged in a mobile app–based breast cancer support program than younger women. The authors believed that user-centered design, technical support, and provision of easy-to-understand information would facilitate older patients' engagement in mobile health interventions [31]. Therefore, age as a predictor of patient engagement in mobile health is inconclusive, and more research is needed to understand the relationship between them and possible mediating factors.

Study Limitations

This study has limitations. Engagement data were extracted from 1 arm of a randomized controlled trial. This may have introduced selection bias in the results as patients who voluntarily participated in the trial might be more familiar with technology and thus better engaged in the program. The results of this study should be interpreted with caution, as the sample size was small, which may have reduced the power of the statistical analysis [55]. Due to the privacy policy of WeChat, the engagement data from the app's background system, such as the times of reading and sharing the program content, are global counts, and the data of each participant cannot be identified. This limits the exploration of individual disparities in program usage. The current program is a 6-week mobile rehabilitation program focusing on acute rehabilitation after arthroplasty, which limits the exploration of how patient engagement in mobile rehabilitation changes with recovery after surgery.

Data collected through the written rehabilitation diary may not fully represent patients' access to and completion of the rehabilitation program. The return rate of the rehabilitation diary was 62% in this study, which is in the middle of the range reported in previous studies where 35% to 86% of patients undergoing THA or TKA returned a written diary to health care professionals [56,57]. However, a large proportion (74% and 88%, respectively) of patients in our study posted and sent messages in the mobile rehabilitation program, suggesting some patients who did not return the diary engaged in the program. App-based electronic diaries have reported high completion rates. For example, a study had 89% to 94% of patients completing the diary [58]. Electronic diaries facilitate the return rate by sending reminders, tracking reports in a real-time way, and collecting data remotely without physical space restriction [59]. However, further investigation is needed on how to incorporate electronic diaries into social media app–based arthroplasty rehabilitation programs. Furthermore, the validity of the written rehabilitation diary may have been reduced if the patients back- and forward-filled in their diaries to meet the expectations of health care professionals [60].

Conclusions

The social media app–based arthroplasty rehabilitation program can support patient engagement. Patient engagement in mobile arthroplasty rehabilitation is associated with their characteristics, such as education level, cohabitation status, age, type of surgery, presence of comorbidities, and sense of self-efficacy. The future design of mobile app–based arthroplasty rehabilitation programs could consider these characteristics and use strategies, such as involving family members, to improve patient engagement with these interventions.

Acknowledgments

We acknowledge all the study participants. We also acknowledge the University of Newcastle in Australia for providing a data storage system and statistical analysis software which were used in this study. This research did not receive any specific grant from funding agencies in the public, commercial, or not-in-profit sectors.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Analysis of changes in patient engagement in the mobile arthroplasty rehabilitation program over 6 weeks after hospital discharge.
[DOCX File , 31 KB - mhealth_v12i1e57635_app1.docx]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 125 KB - mhealth_v12i1e57635_app2.pdf]

References

1. Abdelaal MS, Restrepo C, Sharkey PF. Global perspectives on arthroplasty of hip and knee joints. Orthop Clin North Am 2020;51(2):169-176. [doi: 10.1016/j.ocl.2019.11.003] [Medline: 32138855]
2. Radulovic TN, Lazovic M, Jandric S, Bucma T, Cvjetkovic DD, Manojlovic A. The effects of continued rehabilitation after primary knee replacement. Med Arch 2016;70(2):131-134 [FREE Full text] [doi: 10.5455/medarh.2016.70.131-134] [Medline: 27147789]
3. Moffet H, Collet JP, Shapiro SH, Paradis G, Marquis F, Roy L. Effectiveness of intensive rehabilitation on functional ability and quality of life after first total knee arthroplasty: a single-blind randomized controlled trial. Arch Phys Med Rehabil 2004;85(4):546-556. [doi: 10.1016/j.apmr.2003.08.080] [Medline: 15083429]
4. Hanson HM, Friesen J, Beaupre L, Jasper L, Millington J, Jones CA. Supporting rehabilitation of rural patients receiving total knee arthroplasty through physical activity: perceptions of stakeholder groups. ACR Open Rheumatol 2022;4(10):863-871 [FREE Full text] [doi: 10.1002/acr2.11489] [Medline: 35862257]
5. Ren Y, Guo Q, Li Y, Qiu J, Lv Z, Shao B, et al. The current situation and demands of community rehabilitation medical resources in China [Article in Chinese]. Chin J Rehabil Med 2014;29(8):757-759. [doi: 10.3969/j.issn.1001-1242.2014.08.015]
6. Shukla H, Nair SR, Thakker D. Role of telerehabilitation in patients following total knee arthroplasty: evidence from a systematic literature review and meta-analysis. J Telemed Telecare 2017;23(2):339-346. [doi: 10.1177/1357633X16628996] [Medline: 26843466]
7. Dias Correia F, Nogueira A, Magalhães I, Guimarães J, Moreira M, Barradas I, et al. Digital versus conventional rehabilitation after total hip arthroplasty: a single-center, parallel-group pilot study. JMIR Rehabil Assist Technol 2019;6(1):e14523 [FREE Full text] [doi: 10.2196/14523] [Medline: 31228176]
8. Li W, Song K, Huang T, Chen X, Liu Q, Zhao L, et al. Effect of mobile application rehabilitation guidance on patients after total knee arthroplasty [Article in Chinese]. Chin J Joint Surg (Electronic Edition) 2021 Oct;15(5):547-553. [doi: 10.3877/cma.j.issn.1674-134X.2021.05.005]
9. Wang Q, Lee RLT, Hunter S, Chan SWC. The effectiveness of internet-based telerehabilitation among patients after total joint arthroplasty: an integrative review. Int J Nurs Stud 2021;115:103845. [doi: 10.1016/j.ijnurstu.2020.103845] [Medline: 33360248]
10. Bini SA, Mahajan J. Clinical outcomes of remote asynchronous telerehabilitation are equivalent to traditional therapy following total knee arthroplasty: a randomized control study. J Telemed Telecare 2017;23(2):239-247. [doi: 10.1177/1357633X16634518] [Medline: 26940798]
11. Nelson M, Bourke M, Crossley K, Russell T. Telerehabilitation is non-inferior to usual care following total hip replacement - a randomized controlled non-inferiority trial. Physiotherapy 2020;107:19-27. [doi: 10.1016/j.physio.2019.06.006] [Medline: 32026820]
12. Grajales 3rd FJ, Sheps S, Ho K, Novak-Lauscher H, Eysenbach G. Social media: a review and tutorial of applications in medicine and health care. J Med Internet Res 2014;16(2):e13 [FREE Full text] [doi: 10.2196/jmir.2912] [Medline: 24518354]
13. Monthly usage of major social media platforms in China as of 3rd quarter 2023. Statista. 2024. URL: https://www.statista.com/statistics/250546/leading-social-network-sites-in-china/ [accessed 2024-07-25]
14. Li T, Ding W, Li X, Lin A. Mobile health technology (WeChat) for the hierarchical management of community hypertension: protocol for a cluster randomized controlled trial. Patient Prefer Adherence 2019;13:1339-1352 [FREE Full text] [doi: 10.2147/PPA.S215719] [Medline: 31496664]
15. He C, Wu S, Zhao Y, Li Z, Zhang Y, Le J, et al. Social media-promoted weight loss among an occupational population: cohort study using a wechat mobile phone app-based campaign. J Med Internet Res 2017;19(10):e357 [FREE Full text] [doi: 10.2196/jmir.7861] [Medline: 29061555]
16. Bandura A. Self-efficacy: The Exercise of Control. New York, NY: Freeman; 1997.
17. Illeris KA. A comprehensive understanding of human learning. In: Illeris K, editor. Contemporary Theories of Learning : Learning Theorists in Their Own Words. London, United Kingdom: Routledge; 2009:7-20.
18. Wang Q, Hunter S, Lee RLT, Wang X, Chan SWC. Patients' needs regarding rehabilitation services delivered via mobile applications after arthroplasty: a qualitative study. J Clin Nurs 2022;31(21-22):3178-3189. [doi: 10.1111/jocn.16152] [Medline: 34859523]
19. Wang Q, Hunter S, Lee RLT, Chan SWC. The effectiveness of a mobile application-based programme for rehabilitation after total hip or knee arthroplasty: a randomised controlled trial. Int J Nurs Stud 2023;140:104455 [FREE Full text] [doi: 10.1016/j.ijnurstu.2023.104455] [Medline: 36821950]
20. Wang Q, Lee RLT, Hunter S, Chan SWC. Patients' experiences of using a mobile application-based rehabilitation programme after total hip or knee arthroplasty: a qualitative descriptive study. BMC Nurs 2023;22(1):246 [FREE Full text] [doi: 10.1186/s12912-023-01409-3] [Medline: 37496003]
21. Greene J, Hibbard JH, Sacks R, Overton V, Parrotta CD. When patient activation levels change, health outcomes and costs change, too. Health Aff (Millwood) 2015;34(3):431-437. [doi: 10.1377/hlthaff.2014.0452] [Medline: 25732493]
22. Singh K, Drouin K, Newmark LP, Rozenblum R, Lee J, Landman A, et al. Developing a framework for evaluating the patient engagement, quality, and safety of mobile health applications. Issue Brief (Commonw Fund) 2016;5:1-11. [Medline: 26934758]

23. Lee K, Kwon H, Lee B, Lee G, Lee JH, Park YR, et al. Effect of self-monitoring on long-term patient engagement with mobile health applications. PLoS One 2018;13(7):e0201166 [FREE Full text] [doi: 10.1371/journal.pone.0201166] [Medline: 30048546]
24. Zhang Y, Xia T, Huang L, Yin M, Sun M, Huang J, et al. Factors influencing user engagement of health information disseminated by Chinese provincial Centers for Disease Control and Prevention on WeChat: observational study. JMIR Mhealth Uhealth 2019;7(6):e12245 [FREE Full text] [doi: 10.2196/12245] [Medline: 31250833]
25. Andrade EL, Evans WD, Barrett N, Edberg MC, Cleary SD. Strategies to increase Latino immigrant youth engagement in health promotion using social media: mixed-methods study. JMIR Public Health Surveill 2018;4(4):e71 [FREE Full text] [doi: 10.2196/publichealth.9332] [Medline: 30567689]
26. Correia FD, Nogueira A, Magalhães I, Guimarães J, Moreira M, Barradas I, et al. Medium-term outcomes of digital versus conventional home-based rehabilitation after total knee arthroplasty: prospective, parallel-group feasibility study. JMIR Rehabil Assist Technol 2019;6(1):e13111 [FREE Full text] [doi: 10.2196/13111] [Medline: 30816849]
27. Chen J, Wang Y. Social media use for health purposes: systematic review. J Med Internet Res 2021;23(5):e17917 [FREE Full text] [doi: 10.2196/17917] [Medline: 33978589]
28. WeChat privacy policy. Shenzhen Tencent Computer System Co. Ltd. 2023. URL: https://www.wechat.com/en/privacy_policy.html [accessed 2023-12-18]
29. Aartolahti E, Janhunen M, Katajapuu N, Paloneva J, Pamilo K, Oksanen A, et al. Effectiveness of gamification in knee replacement rehabilitation: protocol for a randomized controlled trial with a qualitative approach. JMIR Res Protoc 2022;11(11):e38434 [FREE Full text] [doi: 10.2196/38434] [Medline: 36441574]
30. Becker S, Brandl C, Meister S, Nagel E, Miron-Shatz T, Mitchell A, et al. Demographic and health related data of users of a mobile application to support drug adherence is associated with usage duration and intensity. PLoS One 2015;10(1):e0116980 [FREE Full text] [doi: 10.1371/journal.pone.0116980] [Medline: 25629939]
31. Zhu H, Chen X, Yang J, Wu Q, Zhu J, Chan SW. Mobile breast cancer e-support program for Chinese women with breast cancer undergoing chemotherapy (part 3): secondary data analysis. JMIR Mhealth Uhealth 2020;8(9):e18896 [FREE Full text] [doi: 10.2196/18896] [Medline: 32936087]
32. Constantinescu D, Pavlis W, Rizzo M, Vanden Berge D, Barnhill S, Hernandez V. The role of commercially available smartphone apps and wearable devices in monitoring patients after total knee arthroplasty: a systematic review. EFORT Open Rev 2022;7(7):481-490 [FREE Full text] [doi: 10.1530/EOR-21-0115] [Medline: 35900191]
33. Hoogland J, Wijnen A, Munsterman T, Gerritsma CL, Dijkstra B, Zijlstra WP, et al. Feasibility and patient experience of a home-based rehabilitation program driven by a tablet app and mobility monitoring for patients after a total hip arthroplasty. JMIR Mhealth Uhealth 2019;7(1):e10342 [FREE Full text] [doi: 10.2196/10342] [Medline: 30702438]
34. Ramkumar PN, Navarro SM, Haeberle HS, Chughtai M, Flynn ME, Mont MA. Social media and total joint arthroplasty: an analysis of patient utilization on instagram. J Arthroplasty 2017;32(9):2694-2700. [doi: 10.1016/j.arth.2017.03.067] [Medline: 28456560]
35. McHugh SM, Riordan F, Kerins C, Curran G, Lewis CC, Presseau J, et al. Understanding tailoring to support the implementation of evidence-based interventions in healthcare: the CUSTOMISE research programme protocol. HRB Open Res 2023;6:7 [FREE Full text] [doi: 10.12688/hrbopenres.13675.1] [Medline: 37361339]
36. Wang Q, Hunter S, Lee RL, Wang X, Chan SW. Mobile rehabilitation support versus usual care in patients after total hip or knee arthroplasty: study protocol for a randomised controlled trial. Trials 2022;23(1):553 [FREE Full text] [doi: 10.1186/s13063-022-06269-x] [Medline: 35804429]
37. Wang H, Xu Y, Hu S, Zhou L, Qian Y, Qian H. Evaluation of the reliability and validity of Chinese version self-efficacy for rehabilitation outcome scale [Article in Chinese]. Chin J Mod Nurs 2014 Jan 26;20(3):268-270. [doi: 10.3760/cma.j.issn.1674-2907.2014.03.007]
38. Waldrop D, Lightsey ORJ, Ethington CA, Woemmel CA, Coke AL. Self-efficacy, optimism, health competence, and recovery from orthopedic surgery. J Couns Psychol 2001;48(2):233-238. [doi: 10.1037//0022-0167.48.2.233]
39. Davis AM, Perruccio AV, Canizares M, Tennant A, Hawker GA, Conaghan PG, et al. The development of a short measure of physical function for hip OA HOOS-Physical Function Shortform (HOOS-PS): an OARSI/OMERACT initiative. Osteoarthritis Cartilage 2008;16(5):551-559 [FREE Full text] [doi: 10.1016/j.joca.2007.12.016] [Medline: 18296074]
40. Perruccio AV, Stefan Lohmander L, Canizares M, Tennant A, Hawker GA, Conaghan PG, et al. The development of a short measure of physical function for knee OA KOOS-Physical Function Shortform (KOOS-PS) - an OARSI/OMERACT initiative. Osteoarthritis Cartilage 2008;16(5):542-550 [FREE Full text] [doi: 10.1016/j.joca.2007.12.014] [Medline: 18294869]
41. Hawker GA, Mian S, Kendzerska T, French M. Measures of adult pain: Visual Analog Scale for Pain (VAS Pain), Numeric Rating Scale for Pain (NRS Pain), McGill Pain Questionnaire (MPQ), Short-Form McGill Pain Questionnaire (SF-MPQ), Chronic Pain Grade Scale (CPGS), Short Form-36 Bodily Pain Scale (SF-36 BPS), and Measure of Intermittent and Constant Psteoarthritis Pain (ICOAP). Arthritis Care Res (Hoboken) 2011;63 Suppl 11:S240-S252 [FREE Full text] [doi: 10.1002/acr.20543] [Medline: 22588748]
42. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand 1983;67(6):361-370. [doi: 10.1111/j.1600-0447.1983.tb09716.x] [Medline: 6880820]

43. EQ-5D-5L user guide,version 3.0. EuroQol Research Foundation. URL: https://euroqol.org/publications/user-guides [accessed 2020-04-22]
44. IBM SPSS statistics for windows, version 28.0. IBM Corp. 2022. URL: https://www.ibm.com/support/pages/node/6445443 [accessed 2024-08-28]
45. Wang L, Geng X, Ji L, Lu G, Lu Q. Treatment decision-making, family influences, and cultural influences of Chinese breast cancer survivors: a qualitative study using an expressive writing method. Support Care Cancer 2020;28(7):3259-3266. [doi: 10.1007/s00520-019-05161-w] [Medline: 31735999]
46. Sigurdardottir AK, Leino-Kilpi H, Charalambous A, Katajisto J, Stark ?, Sourtzi P, et al. Fulfilment of knowledge expectations among family members of patients undergoing arthroplasty: a European perspective. Scand J Caring Sci 2015;29(4):615-624. [doi: 10.1111/scs.12199] [Medline: 25648518]
47. Bourne RB, Chesworth B, Davis A, Mahomed N, Charron K. Comparing patient outcomes after THA and TKA: is there a difference? Clin Orthop Relat Res 2010;468(2):542-546 [FREE Full text] [doi: 10.1007/s11999-009-1046-9] [Medline: 19760472]
48. DeFrance MJ, Scuderi GR. Are 20% of patients actually dissatisfied following total knee arthroplasty? A systematic review of the literature. J Arthroplasty 2023;38(3):594-599. [doi: 10.1016/j.arth.2022.10.011] [Medline: 36252743]
49. Zhao YC, Zhao M, Song S. Online health information seeking among patients with chronic conditions: integrating the health belief model and social support theory. J Med Internet Res 2022;24(11):e42447 [FREE Full text] [doi: 10.2196/42447] [Medline: 36322124]
50. Beatty L, Binnion C. A systematic review of predictors of, and reasons for, adherence to online psychological interventions. Int J Behav Med 2016;23(6):776-794. [doi: 10.1007/s12529-016-9556-9] [Medline: 26957109]
51. Picha KJ, Howell DM. A model to increase rehabilitation adherence to home exercise programmes in patients with varying levels of self-efficacy. Musculoskeletal Care 2018;16(1):233-237. [doi: 10.1002/msc.1194] [Medline: 28401666]
52. Özden F, Sarı Z. The effect of mobile application-based rehabilitation in patients with total knee arthroplasty: a systematic review and meta-analysis. Arch Gerontol Geriatr 2023;113:105058. [doi: 10.1016/j.archger.2023.105058] [Medline: 37172329]
53. Li P, Luo Y, Yu X, Wen J, Mason E, Li W, et al. Patients' perceptions of barriers and facilitators to the adoption of e-hospitals: cross-sectional study in Western China. J Med Internet Res 2020;22(6):e17221 [FREE Full text] [doi: 10.2196/17221] [Medline: 32525483]
54. Carroll JK, Moorhead A, Bond R, LeBlanc WG, Petrella RJ, Fiscella K. Who uses mobile phone health apps and does use matter? A secondary data analytics approach. J Med Internet Res 2017;19(4):e125 [FREE Full text] [doi: 10.2196/jmir.5604] [Medline: 28428170]
55. Serdar CC, Cihan M, Yücel D, Serdar MA. Sample size, power and effect size revisited: simplified and practical approaches in pre-clinical, clinical and laboratory studies. Biochem Med (Zagreb) 2021;31(1):010502. [doi: 10.11613/BM.2021.010502] [Medline: 33380887]
56. Janhunen M, Katajapuu N, Paloneva J, Pamilo K, Oksanen A, Keemu H, et al. Effects of a home-based, exergaming intervention on physical function and pain after total knee replacement in older adults: a randomised controlled trial. BMJ Open Sport Exerc Med 2023;9(1):e001416 [FREE Full text] [doi: 10.1136/bmjsem-2022-001416] [Medline: 36896366]
57. Roberts JL, Pritchard AW, Williams M, Totton N, Morrison V, Din NU, et al. Mixed methods process evaluation of an enhanced community-based rehabilitation intervention for elderly patients with hip fracture. BMJ Open 2018;8(8):e021486 [FREE Full text] [doi: 10.1136/bmjopen-2018-021486] [Medline: 30093516]
58. Rijsbergen M, Niemeyer-van der Kolk T, Rijneveld R, Pinckaers JHFM, Meshcheriakov I, Bouwes Bavinck JN, et al. Mobile e-diary application facilitates the monitoring of patient-reported outcomes and a high treatment adherence for clinical trials in dermatology. J Eur Acad Dermatol Venereol 2020;34(3):633-639 [FREE Full text] [doi: 10.1111/jdv.15872] [Medline: 31419338]
59. Lev-On A, Lowenstein-Barkai H. Viewing diaries in an age of new media: an exploratory analysis of mobile phone app diaries versus paper diaries. Methodol Innov 2019;12(1):205979911984444. [doi: 10.1177/2059799119844442]
60. Frost R, McClurg D, Brady M, Williams B. Optimising the validity and completion of adherence diaries: a multiple case study and randomised crossover trial. Trials 2016;17(1):489 [FREE Full text] [doi: 10.1186/s13063-016-1615-7] [Medline: 27724922]

Abbreviations

THA: total hip arthroplasty
TKA: total knee arthroplasty

Edited by G Badicu; submitted 22.02.24; peer-reviewed by S Liu, J de Batlle; comments to author 01.06.24; revised version received 05.07.24; accepted 07.08.24; published 01.10.24.

Please cite as:
Wang Q, Lee RLT, Hunter S, Zhu A, Chan SWC
Patient Engagement in a Mobile App–Based Rehabilitation Program for Total Hip or Knee Arthroplasty: Secondary Data Analysis of a Randomized Controlled Trial
JMIR Mhealth Uhealth 2024;12:e57635
URL: https://mhealth.jmir.org/2024/1/e57635
doi:10.2196/57635
PMID:

French Version of the User Mobile Application Rating Scale: Adaptation and Validation Study

Ina Saliasi[1*], PhD; Romain Lan[1,2*], PhD; Maryem Rhanoui[1], PhD; Laurie Fraticelli[1], PhD; Stéphane Viennot[1,3], PhD; Delphine Tardivo[2], Prof Dr; Céline Clément[1,4], PhD; Benjamin du Sartz de Vigneulles[1], MS; Sandie Bernard[1], PhD; Adeline Darlington-Bernard[1], MS; Claude Dussart[1,3], Prof Dr; Denis Bourgeois[1,3], Prof Dr; Florence Carrouel[1], Prof Dr

[1]Laboratory Health Systemic Process (P2S), UR4129 Research Unit, University Claude Bernard Lyon 1, University of Lyon, Lyon, France

[2]Anthropologie Bio-Culturelle, Droit, Éthique et Santé Laboratory (ADES, UMR7268), Aix Marseille University, Centre National de la Recherche Scientifique, Etablissement Français du Sang, Marseille, France

[3]Hospices Civils of Lyon, Lyon, France

[4]Laboratory “Interpsy”, UR4432, University of Lorraine, Nancy, France

[*]these authors contributed equally

Corresponding Author:
Florence Carrouel, Prof Dr
Laboratory Health Systemic Process (P2S), UR4129 Research Unit, University Claude Bernard Lyon 1, University of Lyon, , Lyon, , France

Abstract

Background: Managing noncommunicable diseases effectively requires continuous coordination and monitoring, often facilitated by eHealth technologies like mobile health (mHealth) apps. The end-user version of the Mobile Application Rating Scale is a valuable tool for assessing the quality of mHealth apps from the user perspective. However, the absence of a French version restricts its use in French-speaking countries, where the evaluation and regulation of mHealth apps are still lacking, despite the increasing number of apps and their strong relevance in health care.

Objective: This study aims to translate and culturally adapt a French version of the user Mobile Application Rating Scale (uMARS-F) and to test its overall and internal reliability.

Methods: Cross-cultural adaptation and translation followed the universalist approach. The uMARS-F was evaluated as part through a cohort study using the French mHealth app “MonSherpa” (Qare). Participants were French-speaking adults with Apple or Android phones, excluding those with difficulty understanding French, prior app use, or physical limitations. They assessed the app using the uMARS-F twice (T1 and T2) 1 week apart. Scores for each section and overall were assessed for normal distribution using the Shapiro-Wilk test and presented as mean (SD), and potential floor or ceiling effects were calculated accordingly. Overall reliability was evaluated using intraclass correlation coefficients and internal reliability using Cronbach α. Concordance between the 3 subscales (objective quality, subjective quality, and perceived impact), 4 sections, and 26 items at T1 and T2 was evaluated using the paired *t* test (2-tailed) and Pearson correlation.

Results: In total, 167 participants assessed the app at both T1 and T2 (100% compliance). Among them, 49.7% (n=83) were female, and 50.3% (n=84) were male, with a mean age of 43 (SD 16) years. The uMARS-F intraclass correlation coefficients were excellent for objective quality (0.959), excellent for subjective quality (0.993), and moderate for perceived impact (0.624). Cronbach α was good for objective quality (0.881), acceptable for subjective quality (0.701), and excellent for perceived impact (0.936). The paired *t* tests (2-tailed) demonstrated similar scores between the 2 assessments ($P>.05$), and the Pearson correlation coefficient indicated high consistency in each subscale, section, and item ($r>0.76$ and $P<.001$). The reliability and validity of the measures were similar to those found in the original English version as well as in the Spanish, Japanese, Italian, Greek, and Turkish versions that have already been translated and validated.

Conclusions: The uMARS-F is a valid tool for end users to assess the quality of mHealth apps in French-speaking countries. The uMARS-F used in combination with the French version of the Mobile Application Rating Scale could enable health care professionals and public health authorities to identify reliable, high-quality, and valid apps for patients and should be part of French health care education programs.

(JMIR Mhealth Uhealth 2024;12:e63776) doi:10.2196/63776

KEYWORDS

mHealth; mobile health; mobile health apps; eHealth; Mobile Application Rating Scale, user version; mobile apps; quality assessment tool; uMARS

Introduction

Noncommunicable diseases (NCDs), such as cardiovascular diseases, diabetes, cancers, and chronic respiratory conditions, represent a major burden for global health systems and account for 74% of all deaths according to the World Health Organization [1]. The management of NCDs is challenging, requiring complex coordination among various medical stakeholders and regular monitoring of treatments and symptoms outside of hospitalization periods [2].

In the context of new global health care challenges such as demographic changes due to an aging population and a shortage of health care professionals (HCPs), particularly in rural areas, the need to implement out-of-hospital assessment and health monitoring systems to ensure continuous and effective communication between patients and HCPs is critical [3]. eHealth technologies, especially mobile health (mHealth) apps, offer promising solutions to address this need [4,5] and can be applied not only to NCDs but also to health determinants (eg, sleep, diet, exercise, and mental health).

In secondary and tertiary prevention, mHealth apps are revolutionizing the way clinicians and researchers monitor and manage the risk of recurrence or complication [6-9], reducing health inequalities by facilitating access to quality health care [10], improving lifestyle behaviors and chronic condition management, lowering health care costs [11], and increasing patient awareness and autonomy [12,13]. In primary prevention, individuals empower themselves to achieve better health conditions. Despite the growing popularity of mHealth apps and their promising outlook, most of them are unregulated and have not been scientifically evaluated in terms of effectiveness, efficiency, cost, and patient acceptability [14]. The 2016 best practice guidelines by the French National Authority for Health (Haute Autorité de Santé) were an important step toward improving the quality of mHealth apps [15]. These guidelines aimed to provide developers with a framework to ensure apps meet certain standards in terms of functionality, data privacy, and safety but without a clinical purpose. However, the guidelines are nonbinding and do not constitute formal regulation [16]. This means that adherence is voluntary, and there are no legal mechanisms to enforce compliance or monitor apps after launch, which creates gaps in quality assurance. Moreover, while the guidelines outline broad principles, they lack specific evaluation criteria or a standardized certification process, making it difficult for health care providers and users to assess whether an app meets the recommended standards. This creates variability in app quality and can lead to the promotion of apps that may not have been rigorously tested or validated. The absence of mandatory oversight means that some apps might pose risks to users, particularly regarding data security and clinical accuracy.

On the other hand, this lack of evaluation significantly impacts on the difficulties encountered by clinicians and end users face in selecting safe and effective apps [3]. The quality of the information is questionable, and the developer's evaluation is not comprehensive enough to help end users, HCPs, and researchers identify the app's quality [17,18].

Publicly available information, star ratings that may be artificially inflated, or downloads are the most common ways to select a mHealth app rather than validated scientific content [19]. Only a few mHealth apps available on the market have undergone a thorough validation process based on high-level evidence [20]. Appropriate quality and efficacy assessment and assurance are therefore needed both during the development and ongoing use of mHealth apps [21].

To objectively evaluate the validity and functionality of mHealth apps, several standardized scales have been developed for HCPs [22]. One of the most widely used of these assessment tools, the Mobile Application Rating Scale, developed by Stoyanov et al [22], is to date considered as the reference scale for HCPs in the scientific literature. However, this scale requires a level of scientific and clinical expertise and training in mHealth, making it almost impossible for its use by end users. Therefore, the same team of authors developed the end-user version of the Mobile Application Rating Scale (uMARS), which is a valid and objective tool that can be used by end users with different levels of education or by researchers working with end users to evaluate and assess the quality of mHealth apps from end-user perspective [23]. This scale is available in English (uMARS) [23], Spanish (uMARS-S) [24], Italian (uMARS-I) [25], Japanese (uMARS-J) [26], Turkish (uMARS-T) [27], and Greek (uMARS-G) [28]. The cross-cultural translation and validation of the original uMARS into the French language have not yet been carried out, limiting its use in French-speaking countries, despite the fact that 321 millions of people worldwide speak French, representing the fifth most spoken language in the world [29]. The aim of this study was (1) to translate and culturally adapt a French version of the uMARS and (2) to test its overall and internal reliability.

Methods

Study Design

This study followed and applied the universalist approach [30] to translate, cross-culturally adapt, and validate the original version of uMARS [23] into French. The reliability of the French version of the user Mobile Application Rating Scale (uMARS-F) was assessed through a prospective, longitudinal cohort study, adhering to the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines (Multimedia Appendix 1).

Description of the uMARS

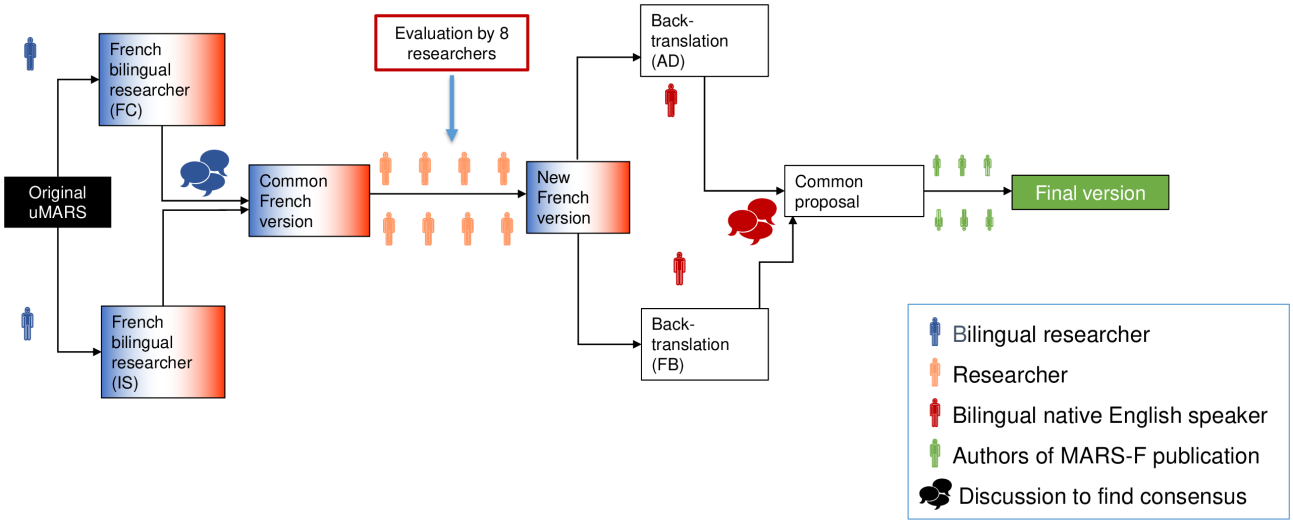
The uMARS is a valid and useful tool designed to allow end users to evaluate the quality of English-language mHealth apps [23]. It consists of 26 items organized into 3 subscales. The first one, the subscale ("objective quality") is divided into 4 sections:

engagement (items 1-5), functionality (items 6-9), aesthetics (items 10-12), and information (items 13-16). The second one, the subjective subscale ("subjective quality") contains 4 items (17-20). The third one, an additional 6-item subscale measuring the app's perceived impact on awareness, knowledge, attitudes, intention to change, help-seeking, and probability of changing the targeted health behavior. Each item of the uMARS is rated on a 5-point Likert scale ranging from 1=poor to 5=excellent. The response "not applicable" (NA) is an additional option included for the information section in case an item may not be applicable. The uMARS scale is both a profile, as each subcategory has a score, and an index, as it provides a final overall score for assessing the quality of mHealth care apps.

Process of Translation and Cultural Adaptation

Figure 1 presents the process of translation and cultural adaptation of the uMARS in the French language. Two native French bilingual researchers (FC and IS) independently translated the original version of the uMARS from English into French. The 2 French translations obtained were compared by the 2 researchers who, in case of discrepancies, discussed to ensure that the meanings were as close as possible to the original English version and proposed a common French version after reaching a consensus. The common version's comprehensibility was then evaluated by 8 researchers (DB, LF, CC, SB, DT, SV, BDSDV, and CD) who, after discussion, agreed on a new consensual French version. This version was blind back-translated in English by 2 native English speakers (AD-B and Felicor Bongolan) independently. Their final common proposal was reviewed and compared with the original English version, which was then read and evaluated by 6 researchers, authors of the French version of the Mobile Application Rating Scale (MARS-F; IS, FC, LF, DB, CD, and DT). Their comments and suggestions were discussed within the research group, and the final version of the uMARS-F was developed.

Figure 1. Translation methodology for uMARS. MARS-F: French version of the Mobile Application Rating Scale; uMARS: user version of the Mobile Application Rating Scale.



Evaluation of Reliability

The reliability of the uMARS-F was evaluated between December 2022 and June 2023 using the French mHealth app "MonSherpa" (Qare). This app, chosen by the research team, is a free psychological support app available in the Google Play and Apple stores, targeting adults (≥18 years) of both sexes.

The sample size for validation of the uMARS-F scale was calculated to ensure sufficient statistical power to assess reliability. Using test-retest reliability with intraclass correlation coefficients (ICCs) in a 2-way random-effects, k-measurement model with consistency, the target was to detect an ICC of 0.75 with 80% power and a 5% significance level, with 2 measures per participant. Based on standard calculations, a minimum of 35 participants was required to achieve this level of reliability.

The recruitment method was based on voluntary participation. Participants were initially enrolled through the established network of research laboratory members, who received an email inviting them to participate. Additionally, social networks such as LinkedIn and Facebook were used to reach a broader audience. A snowball effect was also used, where initial participants were encouraged to refer others who might be interested.

The inclusion criteria were (1) French people aged 18 years and older, (2) persons having access to an Apple or Android phone, and (3) people who did not know the app or had not used it before. The exclusion criteria were people who (1) have difficulty in understanding the French language, (2) already know and used this app personally, and (3) had difficulty using "MonSherpa" for any reason (eg, physical disability).

Eligible participants were asked to download the "MonSherpa" app and use it at least 5 minutes a day for 1 week. On the seventh day, they rated (time 1, T1) this app using a web-based version of the uMARS-F. Then, they were not allowed to use this app for 1 week. Finally, after this washout period, all the participants used this app for only 5 minutes and rated it once again using the same scale (time 2, T2).

Data Analysis

Descriptive Statistics

The distribution of summary scores for each section and for all sections was evaluated for normal distribution using the

Shapiro-Wilk test. For the normal distribution, means and SDs were calculated. Floor or ceiling effects were present when more than 15% of responses were rated as the minimum or maximum scores, respectively [31].

Evaluation of the Validity and Reliability of the uMARS-F

Intraclass Correlation Coefficient

The ICCs were calculated to assess the interrater reliability of the sections and subscales. A random-effects average measures model with absolute agreement was used for this calculation. ICC values were interpreted as follows: less than 0.50 indicated poor reliability, 0.51-0.75 indicated moderate reliability, 0.76-0.89 indicated good reliability, and greater than 0.90 indicated excellent reliability.

Cronbach α

The internal consistency of the uMARS-F questionnaire was evaluated using Cronbach α. A high Cronbach α indicates that the items within the scale are well correlated, suggesting that the scale is reliable. The interpretation of the Cronbach α coefficient was as follows: excellent (≥0.90), good (0.80-0.89), acceptable (0.70-0.79), questionable (0.60-0.69), poor (0.50-0.59), and unacceptable (<0.50) [31].

Pearson Coefficient

The test-retest reliability was evaluated using Pearson *r* coefficients with 95% CIs. The correlation coefficient ranges between –1 and 1, with values closer to 1 indicating a strong positive linear relationship between the subscales, sections, and items, and values closer to –1 indicating a strong negative linear relationship. The significance of the correlations was assessed to determine the strength of the association.

Paired *t* Test

The paired *t* test (2-tailed) was used to compare the means of the 2 times and evaluate the test-retest reliability. This test assesses whether the mean difference between paired observations is significantly different from 0. A *P* value less than .05 indicates that there is a statistically significant difference between the paired means.

Statistical Software

The statistical analysis was carried out using Python (version 3.10; Python Software Foundation), with scipy.stats for statistical calculations and matplotlib along with seaborn for data visualization.

Ethical Considerations

This study did not involve research with access to health data and, therefore, falls outside the scope of the French Jarde law on research involving humans (Law 2012 - 300 of March 5, 2012). Participants gave their consent by completing the web-based questionnaire, and they could withdraw from the study at any time by logging out. Data were completely anonymous, as no identifying data were collected. The platform Claroline Connect from the University Claude Bernard Lyon 1 used for the web-based questionnaire was in accordance with the Regulation EU 2016/679 of the European Parliament and the Council of April 27, 2016, on the protection of individuals with regard to the processing of personal data and on the free movement of such data. Participants received no compensation.

Results

Cross-Cultural Adaptation and Translation Process

Both the conceptual analysis and the translation were considered relevant and appropriate to French culture. No major differences were found between the 2 independent translations of the uMARS into French. The back-translated version of the uMARS-F was equivalent to the original uMARS except for softening changes. The final version of the uMARS-F was produced after review by the authors and mutual agreement on any discrepancies (Multimedia Appendix 2).

Participant Characteristics

In total, 167 participants assessed the app at both time 1 and time 2 (100% compliance). Among them, 49.7% (n=83) were female, and 50.3% (n=84) were male, with a mean age of 43 (SD 16) years. In terms of education level, 21.6% (n=36) of participants did not have the baccalaureate, 25.1% (n=42) had education levels ranging from the baccalaureate to 2 years postbaccalaureate, and 53.3% (n=89) had education levels higher than 2 years postbaccalaureate.

Descriptive Results of the uMARS-F Assessment

All the participants have filled out the uMARS-F questionnaire twice, 7 days apart. They answered all the questions. The mean uMARS-F objective quality score was 4.08 (SD 0.77), whereas the subjective quality score was 2.95 (SD 1.20), and the perceived impact score was 3.42 (SD 0.96). There were no missing values among the responses.

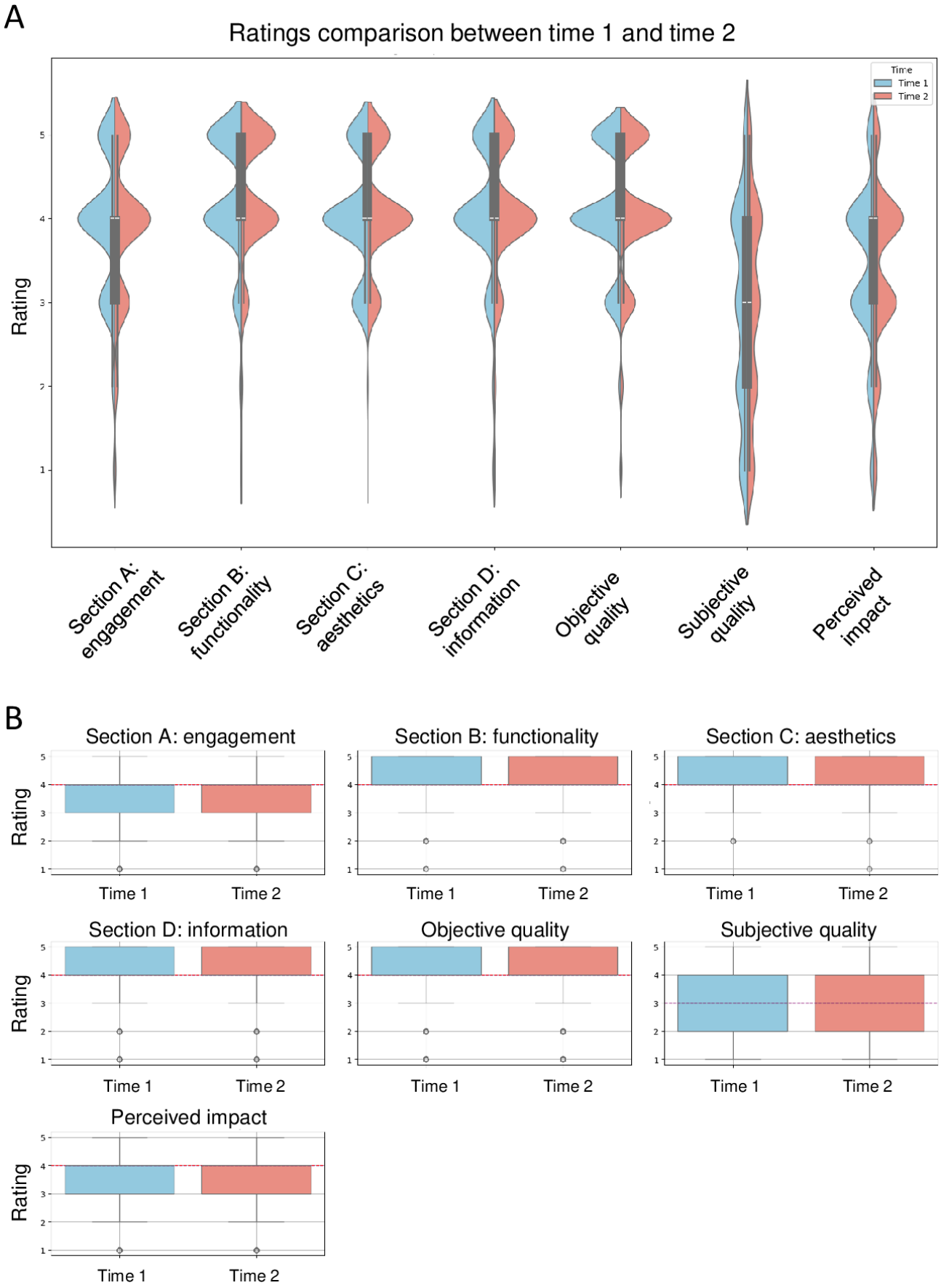
The descriptive analysis of the evaluation is presented in Table 1. A high ceiling effect was observed for the objective quality and its 4 sections (engagement, functionality, aesthetics, and information), suggesting that respondents tended to give higher ratings, indicating overall satisfaction. No floor effect was observed except for the subjective quality (n=26, 15.6%). The Shapiro-Wilk test showed a lack of fit to the normal distribution in subscales and sections.

The distribution of scores across the uMARS-F subscales and sections presented in Figure 2 indicates high consistency between the 2 evaluation times (T1 and T2), as evidenced later by the strong Pearson correlations (all *P*<.001). The objective quality subscale and its sections generally exhibited scores concentrated between 3 and 5, suggesting a favorable evaluation of the app quality. However, the subjective quality subscale showed a wider range of scores from 0 to 5, indicating more variability in user perceptions. For the perceived impact section, scores were predominantly grouped between 3 and 4, reflecting a moderate perceived impact of the app.

Table . Descriptive statistics.

		Ceiling effect, n (%)	Floor effect, n (%)	Shapiro-Wilk, *P* value	Mean (SD)
Objective quality		50 (29.9)	2 (1.2)	<.001	4.08 (0.77)
	Section A (engagement)	38 (22.8)	3 (1.8)	<.001	3.88 (0.87)
	Section B (functionality)	70 (41.9)	1 (0.6)	<.001	4.28 (0.73)
	Section C (aesthetics)	44 (26.3)	0 (0.0)	<.001	4.07 (0.69)
	Section D (information)	47 (28.1)	3 (1.8)	<.001	4.07 (0.80)
Subjective quality		13 (7.8)	26 (15.6)	<.001	2.95 (1.20)
Perceived impact		16 (9.6)	8 (4.8)	<.001	3.42 (0.96)

Figure 2. Descriptive results of the uMARS-F evaluation. (A) Violin plots of the uMARS-F subscales and section scores. (B) Box plot of the uMARS-F subscales and section scores. uMARS-F: French version of the user Mobile Application Rating Scale.



Evaluation of the Validity and Reliability of the uMARS-F

Evaluation of Interrater Reliability Using ICCs and Internal Consistency Using Cronbach α Coefficients

The results for the ICC and Cronbach α are described in Table 2. The ICC was excellent for the objective quality subscale (0.959, 95% CI 0.93-0.98), excellent for the subjective quality subscale (0.993, 95% CI 0.98-1.0), and moderate for the perceived impact (0.624, 95% CI 0.32-0.92). The engagement, aesthetics, and information sections obtained an excellent ICC score, whereas the functionality section obtained a moderate ICC score.

The Cronbach α was good for the objective quality subscale (0.881), acceptable for the subjective quality subscale (0.701), and excellent for the perceived impact subscale (0.936).

Table . Interrater reliability using the intraclass correlation coefficients (ICCs) and internal consistency using the Cronbach α coefficients for the French version of the user Mobile Application Rating Scale.

		ICC (95% CI)	Cronbach α (95% CI)
Objective quality		0.959 (0.93-0.98)	0.881 (0.862-0.899)
	Section A (engagement)	0.975 (0.93-1.0)	0.732 (0.683-0.775)
	Section B (functionality)	0.771 (0.43-0.98)	0.795 (0.757-0.829)
	Section C (aesthetics)	0.914 (0.72-1.0)	0.777 (0.732-0.816)
	Section D (information)	0.926 (0.78-0.99)	0.734 (0.684-0.777)
Subjective quality		0.993 (0.98-1.0)	0.701 (0.646-0.751)
Perceived impact		0.624 (0.32-0.92)	0.936 (0.924-0.946)

Paired t Test and Pearson Coefficient

The paired *t* test (2-tailed) and Pearson coefficient are presented in Table 3. The results from paired *t* tests (2-tailed) indicated no significant differences ($P>.05$) between T1 and T2 between the 20 items and overall subscale scores, highlighting a high test-retest reliability over time in end users' perceptions.

Table . Paired *t* test (2-tailed) and Pearson correlation for the French version of the user Mobile Application Rating Scale.

	Paired *t* test			Pearson correlation	
	Mean score at T1 (SD)	Mean score at T2 (SD)	*P* value	*r*	*P* value
Objective quality	4.06 (0.80)	4.06 (0.80)	.12	0.84	<.001
Section A (engagement)	3.89 (0.86)	3.87 (0.87)	.33	0.84	<.001
1. Entertainment	3.83 (0.79)	3.86 (0.79)	.56	0.78	<.001
2. Interest	3.89 (0.66)	3.89 (0.71)	>.99	0.76	<.001
3. Customization	3.30 (1.01)	3.34 (0.97)	.28	0.88	<.001
4. Interactivity	3.99 (0.76)	3.93 (0.87)	.13	0.81	<.001
5. Target group	4.42 (0.65)	4.33 (0.73)	.05	0.79	<.001
Section B (functionality)	4.27 (0.74)	4.28 (0.73)	.44	0.81	<.001
6. Performance	4.14 (0.77)	4.16 (0.75)	.55	0.87	<.001
7. Ease of use	4.33 (0.67)	4.31 (0.69)	.62	0.76	<.001
8. Navigation	4.21 (0.82)	4.26 (0.81)	.18	0.84	<.001
9. Gestural design	4.40 (0.67)	4.41 (0.64)	.87	0.74	<.001
Section C (aesthetics)	4.05 (0.69)	4.09 (0.69)	.06	0.81	<.001
10. Layout	4.22 (0.63)	4.25 (0.64)	.43	0.81	<.001
11. Graphics	3.86 (0.71)	3.93 (0.70)	.05	0.79	<.001
12. Visual appeal	4.07 (0.68)	4.10 (0.68)	.27	0.81	<.001
Section D (information)	3.95 (0.80)	3.96 (0.77)	.81	0.89	<.001
13. Quality of information	3.95 (0.75)	3.96 (0.73)	.67	0.88	<.001
14. Quantity of information	3.99 (0.75)	3.97 (0.70)	.62	0.79	<.001
15. Visual information	3.93 (0.91)	3.95 (0.89)	.41	0.95	<.001
16. Credibility of source	3.95 (0.80)	3.96 (0.77)	.81	0.89	<.001
Subjective quality	2.94 (1.21)	2.96 (1.20)	.43	0.90	<.001
17. Would you recommend	2.93 (1.06)	2.97 (1.08)	.41	0.81	<.001
18. How many times	3.43 (1.02)	3.43 (1.00)	.89	0.85	<.001
19. Would you pay	1.69 (0.77)	1.74 (0.81)	.15	0.82	<.001
20. Overall (star) rating	3.72 (0.82)	3.69 (0.85)	.51	0.85	<.001

Pearson correlations between T1 and T2 are consistently high (ranging from 0.74 to 0.95), indicating strong positive relationships and high reliability of responses across both time points. These correlations are significant (P<.001), reinforcing the reproducibility of end users' evaluations over time.

Discussion

Principal Results and Comparison With Prior Work

The widespread use of mHealth apps over the last few years has led to a significant revolution in the treatment of lifestyle-related disorders and NCDs, which are now an inevitable part of our modern society [12]. The mHealth apps are an effective health care approach, and their number increased dramatically, accentuated since the COVID-19 crisis [32]. By the end of 2023, there were over 100,000 mHealth apps on the global market, and this number is constantly growing [33]. The French mHealth sector has followed this trend with a growing number of mHealth apps and a high level of interest in their importance for health care over the last 5 years [34]. Therefore, it is crucial to have a standardized and reproductible evaluation scale to identify those that are effective and comply with medical standards [34]. These evaluations must be conducted at 2 main levels: by HCPs, who can assess the quality of the content and its alignment with medical recommendations using the MARS-F;

and by end users using the uMARS-F to evaluate the usability, acceptability, and impact on their engagement in care [7,35,36].

To date, only the MARS-F scale for HCPs [21] is available in French. The aim of this study was to translate, culturally adapt, and validate the French version of the uMARS to enable end users to assess the quality of mHealth apps. For this, a methodology similar to that used for the uMARS-J [26], uMARS-S [24], uMARS-I [25], uMARS-T [27], and uMARS-G [28] scales was followed.

The number of participants completing the questionnaires at T1 and T2 (n=157) was far higher than the calculated minimum sample size (n=35). Consequently, as similar studies have included between 35 [26] and 216 participants [24], it was decided to include all participants and analyze their data to increase the power of the results.

The validation process showed high ceiling effects for objective quality and its 4 sections, with many participants giving high scores, limiting the scale's ability to distinguish top performers. In contrast, subjective quality had a floor effect, indicating dissatisfaction with the personal appeal of the app. The ceiling or floor effects were only observed in the Japanese version [26] and demonstrated no effect.

The reliability of the measures was primarily indicated by the ICC values. The excellent ICC values for most of the sections reflected strong interrater reliability, ensuring that the repeated measures are consistent. The uMARS-F demonstrated higher ICC values for both the objective and subjective quality subscales compared to the English, Spanish, Italian, and Japanese versions [23-26]. The comparison with Turkish and Greek versions was not possible due to unavailable ICC values [27,28].

The higher ICC values observed in the French version of the uMARS compared to the other versions may stem from cultural differences in how users rate apps. Another factor could be the translation and adaptation process of the uMARS into French which can involve, or not, careful attention to linguistic nuances and cultural relevance, contributing to more consistent ratings. This is particularly relevant for the English and Spanish versions, where there are significant linguistic disparities between countries. Additionally, the demographics and characteristics of the study samples (eg, educational levels) may differ across countries, influencing how participants engage with the tool.

These differences have important implications. Higher ICC values in the French version suggest stronger internal consistency, making it a reliable tool for evaluating mHealth apps in French-speaking populations. However, it also highlights the need for localized validation efforts, as the reliability and applicability of the uMARS may vary depending on cultural and linguistic contexts. Future research could explore how these factors impact app evaluations across different regions to ensure the tool is robust and adaptable globally.

Additionally, the results highlight a significant correlation between repeated measures (T1 and T2) using paired *t* test (2-tailed) and Pearson correlation coefficients for all subscales, sections, and items. Paired *t* tests (2-tailed) were uniquely conducted in the Italian version [25] and, similar to this study, showed no statistically significant difference in each answer or group of answers between times 1 and 2 ($P>.05$). Pearson correlation coefficients were evaluated in the Italian [25] and Japanese [26] versions, showing similarly high reliability of responses across both time points (ranging from 0.74 to 0.95).

The validity of the measures is inferred from Cronbach α values, which indicates the internal consistency and the correlation between subscales. High Cronbach α values for objective quality and perceived impact suggested strong internal consistency. The internal consistency across the sections supports the validity of the measures, as it suggests that the subscales are appropriately correlated. For the objective quality subscale, the uMARS-F's Cronbach α (0.881) was similar to that of the original uMARS and other versions, including Italian, Japanese, Spanish, Turkish, and Greek [23-28]. For the subjective quality subscale, the uMARS-F's Cronbach α (0.701) was acceptable and comparable to the original uMARS and Japanese versions [23,26], but higher than the Spanish, Greek, and Turkish versions [24,27,28], though lower than the Italian version [25].

Limitations

This study presents several limitations. First, only 1 mHealth app, MonSherpa, was assessed to validate the study. Further investigations are required to test or retest the uMARS-F on other mHealth apps targeting populations more representative of the general French population. Second, the uMARS-F was developed by native French speakers living in France, and its metric properties must be taken into consideration because French speakers worldwide may have different cultures with different varieties of French depending on their country, and further adaptation may be required [21].

Conclusions

The uMARS-F is a valid tool with adequate metric properties for evaluating the quality of mHealth apps in French-speaking countries. It offers a valuable framework for both developers and researchers to assess and enhance mHealth app quality from an end-user perspective prior to market launch or after the introduction of new functionalities. Therefore, the uMARS-F serves as a cornerstone of the French mHealth field, providing opportunities to identify reliable, high-quality, and valid apps for the benefit of end users.

The integration of the uMARS-F into existing health care systems could significantly enhance mHealth app selection by HCPs and public health authorities, enabling the recommendation of high-quality mHealth apps tailored to patient needs. Moreover, incorporating the u-MARS-F into health care education programs could help future professionals make informed decisions about mHealth tools, potentially improving patient outcomes.

From a regulatory perspective, the uMARS-F could play a key role in establishing standardized quality benchmarks for mHealth apps in French-speaking regions. Developers may also use the tool to align with best practices, ensuring that apps meet both clinical and user-experience standards.

Further research should explore the use of uMARS-F alongside MARS-F in clinical and public health settings to better assess the impact of mHealth apps on patient care. Future studies could also investigate how these tools influence app development, regulation, and long-term health care integration across various French-speaking health care systems.

Acknowledgments

The authors would like to thank Felicor Bongolan for her contribution in the back-translation process and all the participants of this study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) statement—checklist of items.
[DOC File, 84 KB - mhealth_v12i1e63776_app1.doc]

Multimedia Appendix 2
French version of user version of the Mobile Application Rating Scale.
[PDF File, 166 KB - mhealth_v12i1e63776_app2.pdf]

References

1. Noncommunicable diseases. World Health Organization. 2023. URL: https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases [accessed 2024-06-22]
2. GBD 2021 Diseases and Injuries Collaborators. Global incidence, prevalence, years lived with disability (YLDs), disability-adjusted life-years (DALYs), and healthy life expectancy (HALE) for 371 diseases and injuries in 204 countries and territories and 811 subnational locations, 1990-2021: a systematic analysis for the Global Burden of Disease Study 2021. Lancet 2024 May 18;403(10440):2133-2161. [doi: 10.1016/S0140-6736(24)00757-8] [Medline: 38642570]
3. Istepanian RSH. Mobile health (m-Health) in retrospect: the known unknowns. Int J Environ Res Public Health 2022 Mar 22;19(7):3747. [doi: 10.3390/ijerph19073747] [Medline: 35409431]
4. Pandian JD, Gall SL, Kate MP, et al. Prevention of stroke: a global perspective. Lancet 2018 Oct 6;392(10154):1269-1278. [doi: 10.1016/S0140-6736(18)31269-8] [Medline: 30319114]
5. Peiris D, Praveen D, Johnson C, Mogulluru K. Use of mHealth systems and tools for non-communicable diseases in low- and middle-income countries: a systematic review. J Cardiovasc Transl Res 2014 Nov;7(8):677-691. [doi: 10.1007/s12265-014-9581-5] [Medline: 25209729]
6. Amagai S, Pila S, Kaat AJ, Nowinski CJ, Gershon RC. Challenges in participant engagement and retention using mobile health apps: literature review. J Med Internet Res 2022 Apr 26;24(4):e35120. [doi: 10.2196/35120] [Medline: 35471414]
7. Eysenbach G, Group CE. CONSORT-EHEALTH: improving and standardizing evaluation reports of web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126. [doi: 10.2196/jmir.1923] [Medline: 22209829]
8. Huntsman JL, Bulaj G. Health education via “empowerment” digital marketing of consumer products and services: promoting therapeutic benefits of self-care for depression and chronic pain. Front Public Health 2022;10:949518. [doi: 10.3389/fpubh.2022.949518] [Medline: 36703812]
9. Salas-Groves E, Galyean S, Alcorn M, Childress A. Behavior change effectiveness using nutrition apps in people with chronic diseases: scoping review. JMIR Mhealth Uhealth 2023 Jan 13;11:e41235. [doi: 10.2196/41235] [Medline: 36637888]
10. Aljedaani B, Babar MA. Challenges with developing secure mobile health applications: systematic review. JMIR Mhealth Uhealth 2021 Jun 21;9(6):e15654. [doi: 10.2196/15654] [Medline: 34152277]
11. Timmers T, Janssen L, Kool RB, Kremer JA. Educating patients by providing timely information using smartphone and tablet apps: systematic review. J Med Internet Res 2020 Apr 13;22(4):e17342. [doi: 10.2196/17342] [Medline: 32281936]
12. Aguiar M, Trujillo M, Chaves D, Álvarez R, Epelde G. mHealth apps using behavior change techniques to self-report data: systematic review. JMIR Mhealth Uhealth 2022 Sep 9;10(9):e33247. [doi: 10.2196/33247] [Medline: 36083606]
13. Fleming GA, Petrie JR, Bergenstal RM, Holl RW, Peters AL, Heinemann L. Diabetes digital app technology: benefits, challenges, and recommendations. A consensus report by the European Association for the Study of Diabetes (EASD) and the American Diabetes Association (ADA) Diabetes Technology Working Group. Diabetologia 2020 Feb;63(2):229-241. [doi: 10.1007/s00125-019-05034-1] [Medline: 31802144]
14. Llorens-Vernet P, Miró J. Standards for mobile health-related apps: systematic review and development of a guide. JMIR Mhealth Uhealth 2020 Mar 3;8(3):e13057. [doi: 10.2196/13057] [Medline: 32130169]
15. Assessment of apps in the mobile health (mHealth) sector—overview and quality criteria of medical content for referencing digital services in the digital health space and the professional service package. Haute Autorité de Santé. 2021. URL: https:/

/www.has-sante.fr/upload/docs/application/pdf/2021-09/assessment_of_apps_in_the_mobile_health_mhealth_sector._overview_and_quality_criteria_of_medical_content.pdf [accessed 2024-10-09]
16. Della Vecchia C, Leroy T, Bauquier C, et al. Willingness of French general practitioners to prescribe mHealth apps and devices: quantitative study. JMIR Mhealth Uhealth 2022 Feb 11;10(2):e28372. [doi: 10.2196/28372] [Medline: 35147508]
17. Lee M, Lin X, Chai JZQ, Lee ES, Smith H, Tudor Car L. Smartphone apps for point-of-care information summaries: systematic assessment of the quality and content. BMJ Evid Based Med 2023 Oct;28(5):320-327. [doi: 10.1136/bmjebm-2022-112146] [Medline: 36922021]
18. Huang HY, Bashir M. Users' adoption of mental health apps: examining the impact of information cues. JMIR Mhealth Uhealth 2017 Jun 28;5(6):e83. [doi: 10.2196/mhealth.6827] [Medline: 28659256]
19. Singh K, Drouin K, Newmark LP, et al. Many mobile health apps target high-need, high-cost populations, but gaps remain. Health Aff (Millwood) 2016 Dec 1;35(12):2310-2318. [doi: 10.1377/hlthaff.2016.0578] [Medline: 27920321]
20. Giebel GD, Speckemeier C, Schrader NF, et al. Quality assessment of mHealth apps: a scoping review. Front Health Serv 2024;4:1372871. [doi: 10.3389/frhs.2024.1372871] [Medline: 38751854]
21. Saliasi I, Martinon P, Darlington E, et al. Promoting health via mHealth applications using a French version of the Mobile App Rating Scale: adaptation and validation study. JMIR Mhealth Uhealth 2021 Aug 31;9(8):e30480. [doi: 10.2196/30480] [Medline: 34463623]
22. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile App Rating Scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27. [doi: 10.2196/mhealth.3422] [Medline: 25760773]
23. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and validation of the user version of the Mobile Application Rating Scale (uMARS). JMIR Mhealth Uhealth 2016 Jun 10;4(2):e72. [doi: 10.2196/mhealth.5849]
24. Martin-Payo R, Carrasco-Santos S, Cuesta M, Stoyan S, Gonzalez-Mendez X, Fernandez-Alvarez MDM. Spanish adaptation and validation of the user version of the Mobile Application Rating Scale (uMARS). J Am Med Inform Assoc 2021 Nov 25;28(12):2681-2686. [doi: 10.1093/jamia/ocab216] [Medline: 34613400]
25. Morselli S, Sebastianelli A, Domnich A, et al. Translation and validation of the Italian version of the user version of the Mobile Application Rating Scale (uMARS). J Prev Med Hyg 2021 Mar;62(1):E243-E248. [doi: 10.15167/2421-4248/jpmh2021.62.1.1894] [Medline: 34322643]
26. Shinohara Y, Yamamoto K, Ito M, et al. Development and validation of the Japanese version of the uMARS (user version of the Mobile App Rating System). Int J Med Inform 2022 Sep;165:104809. [doi: 10.1016/j.ijmedinf.2022.104809]
27. Calik G, Kartal BB, Stoyanov S, et al. Turkish validation of the user version of the Mobile Application Rating Scale. Turk J Urol 2022 May;48(3):236-242. [doi: 10.5152/tud.2022.21324] [Medline: 35634943]
28. Chasiotis G, Stoyanov SR, Karatzas A, Gravas S. Greek validation of the user version of the Mobile Application Rating Scale (uMARS). J Int Med Res 2023 Mar;51(3):030006052311612. [doi: 10.1177/03000605231161213]
29. La langue française dans le monde (2019-2022). Éditions Gallimard. Organisation internationale de la Francophonie. 2022. URL: https://www.francophonie.org/sites/default/files/2023-03/Rapport-La-langue-francaise-dans-le-monde_VF-2022.pdf [accessed 2024-10-15]
30. Herdman M, Fox-Rushby J, Badia X. A model of equivalence in the cultural adaptation of HRQoL instruments: the universalist approach. Qual Life Res 1998 May;7(4):323-335. [doi: 10.1023/a:1024985930536] [Medline: 9610216]
31. de Vet HCW, Terwee CB, Mokkink LB, Knol DL. Measurement in Medicine: A Practical Guide: Cambridge University Press; 2011. [doi: 10.1017/CBO9780511996214]
32. Yang L, Wu J, Mo X, et al. Changes in mobile health apps usage before and after the COVID-19 outbreak in China: semilongitudinal survey. JMIR Public Health Surveill 2023 Feb 22;9:e40552. [doi: 10.2196/40552] [Medline: 36634256]
33. Topic: mHealth. Statista. URL: https://www.statista.com/topics/2263/mhealth/ [accessed 2024-06-25]
34. Maramba I, Chatterjee A, Newman C. Methods of usability testing in the development of eHealth applications: a scoping review. Int J Med Inform 2019 Jun;126:95-104. [doi: 10.1016/j.ijmedinf.2019.03.018]
35. Holl F, Swoboda W. Key elements for the evaluation of mHealth applications: results from a Delphi survey. Stud Health Technol Inform 2022 May 25;294:895-899. [doi: 10.3233/SHTI220620] [Medline: 35612238]
36. Zhou L, Bao J, Setiawan IMA, Saptono A, Parmanto B. The mHealth App Usability Questionnaire (MAUQ): development and validation study. JMIR Mhealth Uhealth 2019 Apr 11;7(4):e11500. [doi: 10.2196/11500] [Medline: 30973342]

Abbreviations

HCP: health care professional
ICC: intraclass correlation coefficient
MARS-F: French version of the Mobile Application Rating Scale
mHealth: mobile health
NCD: noncommunicable disease
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology
uMARS: user version of the Mobile Application Rating Scale

uMARS-F: French version of the user Mobile Application Rating Scale
uMARS-G: user version of Mobile Application Rating Scale in Greek
uMARS-I: user version of Mobile Application Rating Scale in Italian
uMARS-J: user version of Mobile Application Rating Scale in Japanese
uMARS-S: user version of Mobile Application Rating Scale in Spanish
uMARS-T: user version of Mobile Application Rating Scale in Turkish

Edited by L Buis; submitted 29.06.24; peer-reviewed by F Denis, M Gasmi; revised version received 09.09.24; accepted 12.09.24; published 24.10.24.

Please cite as:
Saliasi I, Lan R, Rhanoui M, Fraticelli L, Viennot S, Tardivo D, Clément C, du Sartz de Vigneulles B, Bernard S, Darlington-Bernard A, Dussart C, Bourgeois D, Carrouel F
French Version of the User Mobile Application Rating Scale: Adaptation and Validation Study
JMIR Mhealth Uhealth 2024;12:e63776
URL: https://mhealth.jmir.org/2024/1/e63776
doi:10.2196/63776

Tailored Prompting to Improve Adherence to Image-Based Dietary Assessment: Mixed Methods Study

Lachlan Lee[1*], MBChB; Rosemary Hall[1*], MBChB, PhD; James Stanley[2*], PhD; Jeremy Krebs[1*], MBChB, MD

[1]Department of Medicine, University of Otago, Wellington, New Zealand

[2]Biostatistics Group, University of Otago, Wellington, New Zealand

[*]all authors contributed equally

Corresponding Author:
Lachlan Lee, MBChB
Department of Medicine, University of Otago, , Wellington, , New Zealand

Abstract

Background: Accurately assessing an individual's diet is vital in the management of personal nutrition and in the study of the effect of diet on health. Despite its importance, the tools available for dietary assessment remain either too imprecise, expensive, or burdensome for clinical or research use. Image-based methods offer a potential new tool to improve the reliability and accessibility of dietary assessment. Though promising, image-based methods are sensitive to adherence, as images cannot be captured from meals that have already been consumed. Adherence to image-based methods may be improved with appropriately timed prompting via text message.

Objective: This study aimed to quantitatively examine the effect of prompt timing on adherence to an image-based dietary record and qualitatively explore the participant experience of dietary assessment in order to inform the design of a novel image-based dietary assessment tool.

Methods: This study used a randomized crossover design to examine the intraindividual effect of 3 prompt settings on the number of images captured in an image-based dietary record. The prompt settings were control, where no prompts were sent; standard, where prompts were sent at 7:15 AM, 11:15 AM, and 5:15 PM for every participant; and tailored, where prompt timing was tailored to habitual meal times for each participant. Participants completed a text-based dietary record at baseline to determine the timing of tailored prompts. Participants were randomized to 1 of 6 study sequences, each with a unique order of the 3 prompt settings, with each 3-day image-based dietary record separated by a washout period of at least 7 days. The qualitative component comprised semistructured interviews and questionnaires exploring the experience of dietary assessment.

Results: A total of 37 people were recruited, and 30 participants (11 male, 19 female; mean age 30, SD 10.8 years), completed all image-based dietary records. The image rate increased by 0.83 images per day in the standard setting compared to control (P=.23) and increased by 1.78 images per day in the tailored setting compared to control (P≤.001). We found that 13/21 (62%) of participants preferred to use the image-based dietary record versus the text-based dietary record but reported method-specific challenges with each method, particularly the inability to record via an image after a meal had been consumed.

Conclusions: Tailored prompting improves adherence to image-based dietary assessment. Future image-based dietary assessment tools should use tailored prompting and offer both image-based and written input options to improve record completeness.

(JMIR Mhealth Uhealth 2024;12:e52074) doi:10.2196/52074

KEYWORDS

dietary assessment; diet; dietary; nutrition; mobile phone apps; image-based dietary assessment; nutritional epidemiology; mHealth; mobile health; app; apps; applications; applications; image; RCT; randomized; controlled trial; controlled trials; cross-over; images; photo; photographs; photos; photograph; assessment; prompt; prompts; nudge; nudges; food; meal; meals; consumption; behaviour change; behavior change

Introduction

Accurate assessment of dietary intake is an essential component in the study of nutrition, energy balance, and interventions for obesity and diabetes. Current tools to measure dietary intake are either too imprecise, or they are precise but too expensive, to provide an accurate assessment for clinical use or large population studies [1].

An increasingly popular tool for measuring dietary intake is image-based dietary assessment. Image-based methods are characterized by images rather than text as the main form of data input and have been reported as a preferred method, particularly among children and adolescents [2]. Image-based

dietary assessment tools, such as Easy Diet Diary, MealLogger, and MyFitnessPal, are increasingly being adopted for clinical, research, and personal use [3-5]. The accuracy of image-based methods is comparable to traditional text-based methods; in a meta-analysis by Ho et al [6], image-based methods underreported energy intake by 20% compared to doubly labeled water, the gold standard for measuring energy intake. For comparison, Burrows et al [7] showed that text-based food records underreported energy intake by 11% to 41% compared to doubly labeled water.

Innovations in phone cameras and computer vision, alongside widespread smartphone ownership, provide an opportunity to advance image-based dietary assessment [8]. For image-based methods to achieve high accuracy, they must improve how users capture the content and portion sizes of the foods they eat. The most basic challenge is that users must remember to record in real time, for the simple reason that they cannot photograph food that has already been eaten. Reminding users to capture images using customized text prompts immediately before a meal has previously been shown to improve adherence to an image-based method [9].

This mixed methods study used a randomized crossover trial to examine if the timing of customized text prompts affected the completeness of dietary recording using an image-based method; we also performed a qualitative study of participant attitudes toward image- and text-based dietary assessment methods and text prompting. This study and its findings will inform the design of a novel dietary assessment application.

Methods

Recruitment

Participants were recruited through a circulating poster on university staff and student email lists, social media sites, and word of mouth. Potential participants contacted the researchers, who then provided information about the study. Individuals were eligible if they met the following inclusion criteria: they (1) were aged ≥18 years, (2) had access to a smartphone with a camera, (3) had internet access on their mobile device, and (4) could speak and read English. There were no additional exclusion criteria.

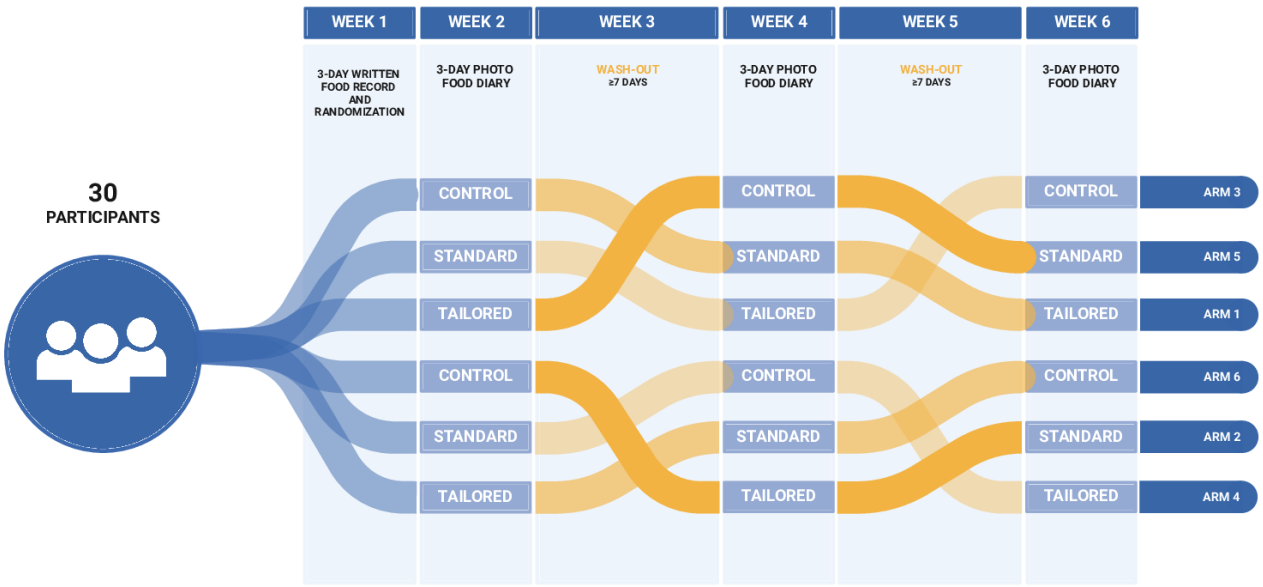
Ethical Considerations

This study was approved by the University of Otago Human Ethics Committee (H21/049). Informed consent was required to participate in the study and was gained at the screening visit. Participants could opt out at any time during the study and were not required to provide a rationale. All participant data were deidentified. Each participant was compensated with a NZ $50 (US $30) supermarket voucher.

Randomized Crossover Trial

The randomized crossover trial compared standard and tailored text prompts to a control period without prompts while completing an image-based dietary record. The study design is shown in Figure 1. Participants attended a screening visit at the research center at the University of Otago, Wellington, New Zealand. After providing written informed consent, eligible participants completed baseline measurements: (1) height in meters using a wall-mounted stadiometer to the nearest 0.5 cm, (2) weight in kilograms to the nearest 0.1 kg using an electronic scale (TBF-300; Tanita Corporation), and (3) resting metabolic rate in kilojoules per day using a ventilated hood (PromethION High-Definition Room Calorimetry System; Sable Systems International). Participants were shown how to use the Easy Diet Diary app (Xyris Software) for the text-based dietary record and image-based dietary record and given the opportunity to ask questions. Easy Diet Diary is free to download from the iOS and Android app stores. Participants then recorded all dietary intake and meal times for 3 days in a text-based dietary record using the text-based features of Easy Diet Diary. Participants were contacted by a researcher if no food was recorded that day to confirm the accuracy of the text-based dietary record. The meal times from this first text-based dietary record were used to calibrate the timing of tailored text prompts for use in the subsequent randomized part of the trial.

Figure 1. Overview of randomized crossover trial design of text prompt settings. Control setting: no prompts received; standard setting: prompts received at 7:15 AM, 11:15 AM, and 5:15 PM; tailored setting: prompts received at times specific to participants' typical meal times as recorded with a text-based dietary record.



After recording the text-based dietary record, participants were randomized into 1 of 6 arms, each arm having a unique order of the 3 text prompt conditions, as described in Figure 1. Participants were reminded to begin their 3-day image-based dietary records using a text message that required confirmation. Each of the image-based dietary records was recorded over 2 weekdays and 1 weekend day. After each 3-day image-based dietary record, participants underwent a washout period of at least 7 days of no recording to mitigate fatigue or training effects. This process was repeated until each participant completed recording under all 3 conditions.

Under the control setting, participants received no prompts. For the standard condition, text prompts were sent at 7:15 AM, 11:15 AM, and 5:15 PM. Data on the typical meal times in New Zealand were unavailable, so these times were chosen to capture participants who might have early meal times. Individually tailored text prompts were determined based on the meal timing provided during the initial text-based dietary record. Each meal time in the text-based dietary record was recorded via the Notes feature of Easy Diet Diary using the following format: "[Time of eating] – [Description of food item consumed]," for example, "1.50 PM – Salmon bagels." Tailored text prompts were set according to the following guidelines: (1) prompts were sent 15 minutes before the earliest recorded eating episode and (2) a prompt for snacks was sent if snack intakes occurred within 1 hour of each other on 2 of 3 days during the text-based dietary record. Both standard and tailored text prompts read, "This is a reminder to record your photo food diary using the Easy Diet Diary app."

Image Rate

The number of images per participant per day was counted across the predefined recording dates. Images generated by the user were automatically uploaded to Easy Diet Diary Connect (Xyris Software) for image rate analysis at the end of each image-based dietary recording period. No analysis or interpretation of nutritional content in the image was performed.

Qualitative Methods

The objective of the qualitative component of the study was to explore the participant experience with each text prompt setting, with the purpose of informing the design of future image-based dietary assessment tools. The qualitative component was composed of interviews and questionnaires. All participants were invited to semistructured open-ended interviews that were conducted one-on-one by a single researcher face-to-face or via video conference (Zoom Video Communications, Inc) after the completion of the image-based dietary records. Interviews lasted for 30 to 60 minutes and focused on (1) the experience of diet monitoring, (2) attitudes toward the Easy Diet Diary app, (3) attitudes toward text prompting, and (4) discussing the participant's "ideal diet monitoring tool." Transcripts of the interviews were coded and thematically analyzed using NVivo Qualitative Data Analysis (QSR International). Additional online questionnaires hosted on Survey Monkey (Momentive Inc) explored the same topics using Likert scales and rankings to identify participant preferences for dietary assessment methods and text prompt settings.

Statistical Analysis

Overview

Statistical analyses were conducted using R (version 4.2; R Foundation for Statistical Computing). Linear mixed effects analysis of the impact of prompt setting on image rate was conducted using the *lme4* package (Bates, Mächler, Bolker, and Walker). Prompt setting and order were fixed effects within the model—order was included to account for uneven sizes in the 6 different orders. Participants were accounted for as random effects (random intercept model), accounting for repeat observations under each prompt setting and correlation of

participant responses across conditions. *P* values for the fixed effect of prompt setting order were obtained using the Wald type 3 test. Q-Q plots were visually inspected and a histogram of the model residuals was checked for homogeneity of variance or any obvious deviations from normality.

Power

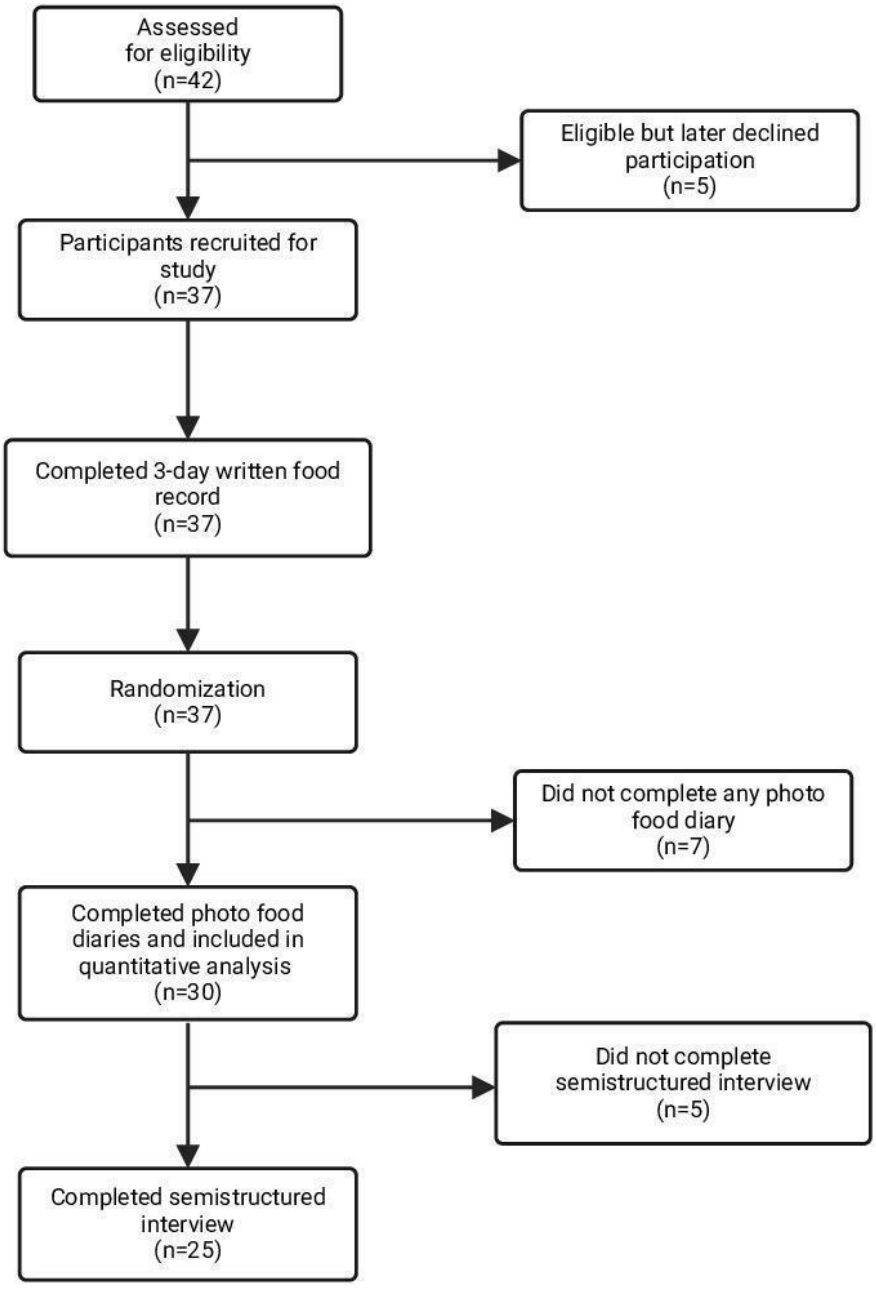
Previous use of image-based dietary records has seen an image rate of 1.4 images per 2.6 meals: a rate of 54% [10]. From this, we estimated under the control setting a mean of 1.6 images per day (54% capture rate applied across 3 meals per day, excluding snacks). With our intervention, we aimed to be able to detect 2.4 images per day (representing 80% coverage) of meals. A sample size of 20 in each of the 3 measurement conditions had over 80% power to detect this difference in rates with a 2-sided α of 5%. The crossover design improved power to detect differences in adherence rates compared to a between-subjects design.

Results

Participants

A total of 42 people were screened for eligibility: 37 met the inclusion criteria and were enrolled, 30 completed the food records, and 25 participated in the qualitative interviews. Seven participants were lost to follow-up and could not be contacted. Five participants did not wish to participate in interviews because of time. A flow diagram of participation is shown in Figure 2.

Figure 2. Flow diagram of participants throughout the study.



The number of participants in each arm and their baseline characteristics are shown in Table 1. The prompt setting order can be seen in Figure 1. Arm population sizes were uneven due to asymmetrical loss to follow-up after randomization; each of

these participants were categorized as lost to follow-up after failing to respond to 3 emails sent to confirm their ongoing participation in the study prior to starting the active part of the intervention. Only those participants completing all 3 image-based dietary records were included in the analysis.

Table . Characteristics of participants who completed the image-based dietary records (n=30).

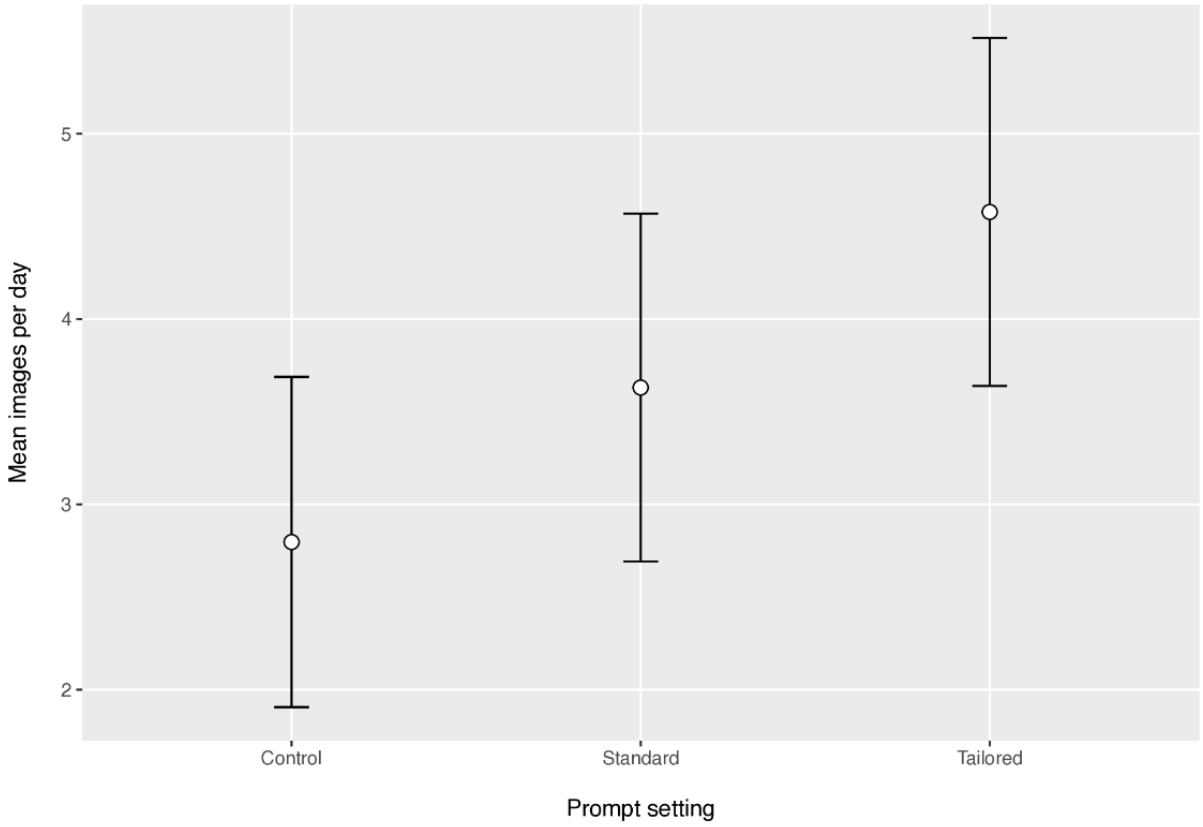
Characteristics		Values
Age (years), mean (SD)		30 (10.8)
BMI (kg/m^2), mean (SD)		24.6 (5.6)
Sex, n (%)		
	Male	11 (37)
	Female	19 (63)
Ethnicity (self-reported), n (%)		
	New Zealand European	15 (50)
	Māori	9 (30)
	Pacific	4 (13)
	Asian	1 (3)
	Other European	1 (3)
Level of education, n (%)		
	Prefer not to say	1 (3)
	No formal education	1 (3)
	High school graduate	3 (10)
	Current tertiary student	17 (57)
	Tertiary graduate	8 (27)
Study arm allocation, n (%) [a]		
	Arm 1	7 (23)
	Arm 2	8 (27)
	Arm 3	4 (13)
	Arm 4	4 (13)
	Arm 5	4 (13)
	Arm 6	3 (10)

[a]Refer to Figure 1 for text prompt setting order by arm.

Image Rate

The image rate, that is, the mean number of images per day recorded on the image-based dietary record, under each prompt setting is shown in Figure 3 with 95% CIs. In the control setting, participants took a mean of 2.8 (95% CI 1.91-3.69) images per day, 3.6 (95% CI 2.69-4.56) images in the standard prompt setting, and 4.6 (95% CI 3.64-5.52) images in the tailored prompt setting. This represents an increase in image rate of 0.83 (95% CI 0.2-2.58) in the standard setting compared to control and an increase of 1.78 (95% CI 1.15-5.51) in the tailored setting compared to control. The order within the sequence of prompt settings had a significant effect *(P*<.001), where the second recording period had 0.82 fewer images per day and the third recording period had 1.19 fewer images per day compared to the first recording.

Figure 3. Interval plot of mean image rate (whiskers represent the 95% CI) under each prompt setting.



Interviews and Questionnaires

Twenty-five participants took part in the interviews (characteristics in Multimedia Appendix 1). Themes that emerged from the interviews were attitudes toward (1) the image-based dietary record, (2) the text-based dietary record, and (3) text prompts. Illustrative quotes relating to these themes can be found in Table 2.

Table . Themes, subthemes, and illustrative quotes from interviews.

Theme and subtheme		Illustrative quote
Attitudes toward text-based dietary records		
	High level of input required	• "… the extra time involved, and finding all the random little bits that you've put in to make a meal, and then trying to put the quantities in that you've eaten… it's just the time, even though it's only a minute or two… it's enough to be a barrier"
	Convenience of retrospective recording	• "Because often it would be like you'd get a text to remind you to take a photo, and then you'd forget straight away and with a photo, you can't eat your food… whereas like, with a written record, that you still know what you've eaten. It could just be like, you can write it the end of the day, or like, whenever you remember" • "As annoying as it was to kind of write everything out, it meant that I could do it at the end of the day and didn't have to remember at the time I was eating. For the most part would write it as I was eating to make sure I was… capturing everything that I was eating, but often actually, by the end of the day, I'd be like, I actually had this snack at this time. So I'll check that in now… it gave me a bit more freedom. But then the picture one was really easy, because you just took a picture… but relied on you to remember at the time you're eating"
	Challenges with the available database and portion input options	• "That's my main complaint… [not having New Zealand brands]… because then you have to manually put it in" • "It would be like in grams or mLs or whatever. And I just kind of put like a handful of something or a scoop of something… kind of tedious trying to be like how many grams would that have been? Or what was it, half a cup or a cup? I guess it's just because I'm not used to it" • "… but in something like Weetbix, it's fine… to quantify how much how much milk I was using in terms of mLs, which is something that I'm not used to measuring… and even Weetbix, because if I had an odd, an odd number of WeetBix, you can only put everything into two sets of Weetbix. So if you had five, you're going to put six or four"
Attitudes toward image-based dietary records		
	Convenience	• "I found photos quite easy, because you just… take a photo and it's done"

Theme and subtheme		Illustrative quote
	Less useful data output	• "… the fact that you just take a photo of it, but it doesn't come up with any information… would it be nice if like, you take a photo and all the information comes out and you can kind of track it" • "… but I don't feel like it did anything really? For like purpose of losing weight. … because I took the picture and didn't tell me like how many calories I was eating… like that cheese and garlic twist, wagyu patty thing… I take picture and think it's not too bad but not realizing it was 1700 calories"
	Difficulty with habit formation	• "Because I wasn't doing it consistently throughout the whole period, it was a lot harder to do than I think would have been"
	Perceived as more accurate than the text-based dietary record	• "[On photo food diaries] … more accurate. You know, this is literally what I'm eating. Whereas doing [the written food record] is sort of like… 'yeah about a cup of raspberries and I don't know what the bun was…' but was that the right one? Or was it like twice as much was this"
	Perceived as less accurate than the text-based dietary record	• "I might've had 12 eggs. You can't really tell in my pictures" • "There was sometimes where that I'd only had about half my muesli in the morning… I was in a rush, but I would have taken a photo of the whole thing. That would have been half the amount of calories"
Attitudes toward text prompts		
	Helpful	• "I think that the reminders definitely helped me to remember to take photos. But I think if it was coming from the app, it would have been less annoying"
	Annoying	• "They were annoying. The first couple were fine. But towards the end, I literally just ignored all of them" • "The days where they were just not at all related to when I usually ate they weren't really helpful at all, because I would just forget again by the time it came to actually take a picture of my food"

Theme 1: Text-Based Dietary Record

Participants reported that the text-based dietary record required a high level of input; this produced detailed food logs with per-item and daily calorie counts, which participants viewed as more useful than the simple image output of the image-based dietary record. Another benefit was the ability to retrospectively record food items; participants often filled in the text-based dietary record at the end of the day—an impossibility with the image-based dietary record. However, participants found completing the text-based dietary record burdensome due to the level of input required: searching and selecting food items, estimating quantities, and reconciling differences between the food item and the database available. The database was often missing food items that were consumed, forcing the participants to input similar but not identical food items. The options available to input portion sizes were also a challenge—participants not knowing the weight of their food in grams, for instance.

Theme 2: Image-Based Dietary Record

Participants preferred to use the image-based dietary record; on a Likert scale with 0 indicating preference for text-based dietary records and 100 indicating preference for image-based dietary records, the average response was 60 (SD 32). Most participants found adhering to the image-based dietary record challenging; 83% of interviewed participants reported difficulty in remembering to capture an image of the entire meal. Most cited simple forgetfulness, though some noted difficulty forming the habit of recording with only 3 days of recording interrupted by

washout periods. Participants reported difficulty with the image-based dietary record if they had partially or entirely consumed the meal before recording, as they could not capture the consumed meal. Participants reported that the images were simple and convenient to capture but provided no calorie counts or dietary information.

Theme 3: Prompts

Attitudes toward the text prompts covered the spectrum from "helpful" to "annoying." Sending prompts via text messaging was reported as an annoyance by all interviewed participants. These notifications, although welcome in the context of text communication, were intrusive as image-based dietary record reminders; "in-app" reminders were preferred. Despite these annoyances, participants generally found the prompts helpful, with tailored prompts viewed as more helpful.

Discussion

Principal Results

This randomized controlled crossover trial demonstrated that text prompts tailored to an individual's typical meal times improved the completeness of image-based dietary records. This suggests that individual tailoring of prompts will improve accuracy of energy intake assessments. Any prompting was better than no prompts, and tailored prompts were the most effective.

These findings are consistent with work shown in Martin et al [9], where standard and customized prompts were compared in a 2-arm study without crossover or a no-prompt arm. Prompt timing could be improved further by longer calibration periods to determine the individual's usual meal time and by recording the time of imaging, which was unavailable using Easy Diet Diary. Automated methods of determining prompt timing may provide pragmatic and scalable methods for tailored prompting [11]. Prompting is not unique to dietary assessment, however, and a diverse range of fields use prompts and study their effectiveness [12]. Previous work has identified that an individual's responsiveness—that is, the likelihood an individual will interact with a prompt or notification—varies based on the characteristics of the person and the prompt [13,14]. Specifically, varying responsiveness to prompts over the course of the day reflects our finding that the timing of a prompt affects the likelihood of interaction. Variable responsiveness poses a particular problem for dietary assessment, where the objective is not only response with a dietary record, but for that response to be accurate. Accuracy of recall has been shown to improve when the recall event is closer to the eating event; that is, recall accuracy decreases with time [15,16]. A prompt must therefore balance the chance of an individual responding to the prompt and the decaying accuracy of the prompted response. Our findings suggest that most individuals in this study were more responsive to dietary assessment prompts closer to meal times, although previous work on prompt responsiveness suggests that this varies from person to person [17]. For some individuals, tailored prompting may reduce the responsiveness of a dietary record. Regular meal times are not the only method to tailor prompts. The use of geolocation to detect likely eating locations, such as restaurants, accelerometers to detect eating or cooking motions, or acoustic sensors to detect chewing noises, among other "context-aware" methods, could also improve prompt timing and appear to be socially acceptable [18-20]. These methods should be selective in their implementation to avoid an excessive number of prompts to participants.

Our findings indicate that prompting tailored to an individual's food intake pattern improves image-based dietary record completeness. An alternate explanation to consider is that the number of prompts, rather than the timing, improves the image rate. This was not supported by these data. Participants in the tailored setting received a mean of 3.1 prompts per day, while all those in the standard setting received 3 prompts. The effect of tailored prompts on image rate therefore should be attributable to the timing, rather than the number, of prompts. The observed increase of 1.78 images per day with tailored prompts (vs control) is important because this represents mean capture of almost 2 meals across a day that would otherwise have been missed and therefore increases accuracy of dietary intake assessment.

Image recording during the second and third instance of the 3-day image-based dietary record decreased by 0.82 and 1.19 images per day, respectively, compared to the first diary in the sequence. This suggests a fatigue effect that was not mitigated by the 7-day washout period and is of concern where ongoing use of a prompted image-based dietary record is desired in a clinical or research setting. Decreasing energy intake over a recording period has previously been identified in national nutrition surveys such as the United Kingdom's National Diet and Nutrition Survey. The average decrease from day 1 to 4 was small at 164 kJ, though this may reflect a true decrease in energy intake under observation [21,22]. This fatigue effect has also been found to apply to prompts themselves, with one study of a diet app finding decreasing responsiveness to push notifications over time [17].

Participant Experience

Participant attitudes toward the image-based dietary record and text-based dietary record contrasted each other. Recording with the image-based dietary record was more convenient at the cost of less useful output in the absence of caloric or macronutrient data. Useful nutritional data can potentially be generated from image-based methods using computer vision–machine learning software, a rapidly advancing field [23]. Text-based dietary records and automated recognition both rely on a representative nutritional database. An incomplete database was highlighted as a barrier to participants providing complete and accurate information. Missing food items or specific brand-name items and inflexible or unfamiliar portion options were challenging. Ensuring that nutritional databases reflect modern and traditional diets, as well as local and culturally relevant foods, should be a priority for current and future dietary assessment tools.

Participants highlighted the convenience of retrospective recording with text-based dietary records, which is not possible with image-based dietary records. The higher number of days with at least 1 entry in the text-based dietary record is suggestive of this retrospective recording, noted in participant interviews and illustrated in Table 2.

Overall, participants reported preference for the image-based dietary record compared to the text-based dietary record, a finding consistent with prior work [24-26]. This preference is notable, as the image-based dietary record did not provide nutritional feedback, and the app was not designed to use images as the primary input method. Participants highlighted that even simple image-based recordings that lack energy or macronutrient intake data provide benefit in the self-assessment of dietary intake, a benefit unique to image-based input that has been identified previously [27,28]. These findings suggest that image-based dietary assessment apps should feature both text and image inputs, providing users the convenience of both simple image-based recording and retrospective text-based recording. Finally, participants preferred tailored prompts and indicated a preference for in-app prompting versus text message prompts.

Strengths

This study had several strengths. The crossover design allowed for intraindividual assessment. Instructing participants to only record images during the image-based dietary records in the absence of automation to provide numeric caloric or macronutrient data means that this information was not available to participants, but this does more closely reflect participant behavior using an automated image-based dietary assessment tool. The diversity of participants included in this study is a strength, with representation of Māori and Pacific views on dietary assessment, which is particularly important in New Zealand.

Limitations

This study had a number of limitations. A 3-day text-based dietary record and rudimentary guidelines were used to generate tailored prompt timing; future studies should use more sophisticated methods—such as machine learning techniques—and expanded timeframes to generate tailored prompt timing. The 3-day time frame for the text-based dietary record and each of the image-based dietary records is brief, and future studies should ideally extend to longer time frames; however, extending these time frames must be balanced against the decreasing accuracy over time. This study placed an artificial dichotomy on the method of dietary assessment. In reality, individuals are free to alternate between image-based and text-based input; a key finding of this study is the importance of offering both, and there are many platforms that provide both methods. We did not directly compare the accuracy of image-based versus text-based input using manual or automated analysis of the images to determine if one method produces more reliable energy intake, macronutrient, or micronutrient estimates, which would affect the interpretation of dietary assessment data from platforms that use both methods.

Conclusions

Individualized tailoring of prompts to match likely meal times improved the completeness of image-based dietary record entries. This is important to improve the accuracy of dietary intake assessments in both clinical and research settings. Although participants preferred the image-based diary, they also identified the need to be able to retrospectively enter data for meals that they had missed recording with an image. The accurate analysis of nutritional composition of a meal from an image requires interpretation of the volume and composition of a meal and integration with nutrient data. Currently available nutritional databases have limitations in the items included, particularly in certain population groups and for culturally specific foods. Therefore, we propose 3 recommendations for future iterations of image-based dietary assessment apps: (1) allowing options for both image-based and written input, (2) publishing and continually updating a nutritional database relevant to the population under study, and (3) the use of tailored in-app prompting for both prospective recording and retrospective recording of missed food items.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Demographics of interviewed participants.
[PDF File, 53 KB - mhealth_v12i1e52074_app1.pdf]

References

1. Irwin ML, Ainsworth BE, Conway JM. Estimation of energy expenditure from physical activity measures: determinants of accuracy. Obes Res 2001 Sep;9(9):517-525. [doi: 10.1038/oby.2001.68] [Medline: 11557832]
2. Boushey CJ, Spoden M, Zhu FM, Delp EJ, Kerr DA. New mobile methods for dietary assessment: review of image-assisted and image-based dietary assessment methods. Proc Nutr Soc 2017 Aug;76(3):283-294. [doi: 10.1017/S0029665116002913] [Medline: 27938425]
3. Simpson A, Gemming L, Baker D, Braakhuis A. Do image-assisted mobile applications improve dietary habits, knowledge, and behaviours in elite athletes? a pilot study. Sports (Basel) 2017 Aug 11;5(3):60. [doi: 10.3390/sports5030060] [Medline: 29910420]
4. Ambrosini GL, Hurworth M, Giglia R, Trapp G, Strauss P. Feasibility of a commercial smartphone application for dietary assessment in epidemiological research and comparison with 24-h dietary recalls. Nutr J 2018 Jan 9;17(1):5. [doi: 10.1186/s12937-018-0315-4] [Medline: 29316930]

5. Teixeira V, Voci SM, Mendes-Netto RS, da Silva DG. The relative validity of a food record using the smartphone application MyFitnessPal. Nutr Diet 2018 Apr;75(2):219-225. [doi: 10.1111/1747-0080.12401] [Medline: 29280547]
6. Ho DKN, Tseng SH, Wu MC, et al. Validity of image-based dietary assessment methods: a systematic review and meta-analysis. Clin Nutr 2020 Oct;39(10):2945-2959. [doi: 10.1016/j.clnu.2020.08.002] [Medline: 32839035]
7. Burrows TL, Ho YY, Rollo ME, Collins CE. Validity of dietary assessment methods when compared to the method of doubly labeled water: a systematic review in adults. Front Endocrinol (Lausanne) 2019;10:850. [doi: 10.3389/fendo.2019.00850] [Medline: 31920966]
8. Smartphone ownership and internet usage continues to climb in emerging economies. Pew Research Center. 2016. URL: https://www.pewresearch.org/global/2016/02/22/smartphone-ownership-and-internet-usage-continues-to-climb-in-emerging-economies/ [accessed 2022-10-14]
9. Martin CK, Correa JB, Han H, et al. Validity of the remote food photography method (RFPM) for estimating energy and nutrient intake in near real-time. Obesity (Silver Spring) 2012 Apr;20(4):891-899. [doi: 10.1038/oby.2011.344] [Medline: 22134199]
10. Fowler LA, Yingling LR, Brooks AT, et al. Digital food records in community-based interventions: mixed-methods pilot study. JMIR Mhealth Uhealth 2018 Jul 17;6(7):e160. [doi: 10.2196/mhealth.9729] [Medline: 30021705]
11. Khanna N, Eicher-Miller HA, Boushey CJ, Gelfand SB, Delp EJ. Temporal dietary patterns using kernel k-means clustering. ISM 2011;2011:375-380. [doi: 10.1109/ISM.2011.68] [Medline: 25258745]
12. Mehrotra A, Pejovic V, Vermeulen J, Hendley R, Musolesi M. My phone and me: understanding people's receptivity to mobile notifications. Presented at: Conference on Human Factors in Computing Systems; May 7 to 12, 2016; San Jose, CA p. 1021-1032. [doi: 10.1145/2858036.2858566]
13. Mehrotra A, Hendley R, Musolesi M. PrefMiner: mining user's preferences for intelligent mobile notification management. Presented at: UbiComp 2016 - the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing; Sep 12 to 16, 2016; Heidelberg, Germany p. 1223-1234. [doi: 10.1145/2971648.2971747]
14. Bidargaddi N, Pituch T, Maaieh H, Short C, Strecher V. Predicting which type of push notification content motivates users to engage in a self-monitoring app. Prev Med Rep 2018 Sep;11:267-273. [doi: 10.1016/j.pmedr.2018.07.004] [Medline: 30109172]
15. Osadchiy T, Poliakov I, Olivier P, Rowland M, Foster E. Progressive 24-hour recall: usability study of short retention intervals in web-based dietary assessment surveys. J Med Internet Res 2020 Feb 3;22(2):e13266. [doi: 10.2196/13266] [Medline: 32012055]
16. Baxter SD, Hitchcock DB, Guinn CH, et al. A validation study concerning the effects of interview content, retention interval, and grade on children's recall accuracy for dietary intake and/or physical activity. J Acad Nutr Diet 2014 Dec;114(12):1902-1914. [doi: 10.1016/j.jand.2014.02.017] [Medline: 24767807]
17. Freyne J, Yin J, Brindal E, Hendrie GA, Berkovsky S, Noakes M. Push notifications in diet apps: influencing engagement times and tasks. Int J Hum-Comput Interact 2017 Oct 3;33(10):833-845. [doi: 10.1080/10447318.2017.1289725]
18. Amft O, Tröster G. Recognition of dietary activity events using on-body sensors. Artif Intell Med 2008 Feb;42(2):121-136. [doi: 10.1016/j.artmed.2007.11.007] [Medline: 18242066]
19. Sen S, Subbaraju V, Misra A, Balan RK. The case for smartwatch-based diet monitoring. 2015 Presented at: 2015 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops); St Louis, MO p. 585-590. [doi: 10.1109/PERCOMW.2015.7134103]
20. Das B, Seelye AM, Thomas BL, Cook DJ, Holder LB, Schmitter-Edgecombe M. Using smart phones for context-aware prompting in smart environments. 2012 Presented at: 2012 IEEE Consumer Communications and Networking Conference; Las Vegas, NV p. 399-403. [doi: 10.1109/CCNC.2012.6181023]
21. Whybrow S, Horgan G, Stubbs RJ. Low-energy reporting and duration of recording period. Eur J Clin Nutr 2008 Sep;62(9):1148-1150. [doi: 10.1038/sj.ejcn.1602826] [Medline: 17538532]
22. Whybrow S, Horgan GW, Macdiarmid JI. Self-reported food intake decreases over recording period in the National Diet and Nutrition Survey. Br J Nutr 2020 Sep 28;124(6):586-590. [doi: 10.1017/S000711452000118X] [Medline: 32234082]
23. Gemming L, Utter J, Ni Mhurchu C. Image-assisted dietary assessment: a systematic review of the evidence. J Acad Nutr Diet 2015 Jan;115(1):64-77. [doi: 10.1016/j.jand.2014.09.015] [Medline: 25441955]
24. Heikkilä L, Vanhala M, Korpelainen R, Tossavainen P. Agreement between an image-based dietary assessment method and a written food diary among adolescents with type 1 diabetes. Nutrients 2021 Apr 16;13(4):1319. [doi: 10.3390/nu13041319] [Medline: 33923638]
25. Wang DH, Kogashiwa M, Kira S. Development of a new instrument for evaluating individuals' dietary intakes. J Am Diet Assoc 2006 Oct;106(10):1588-1593. [doi: 10.1016/j.jada.2006.07.004] [Medline: 17000191]
26. Rollo ME, Ash S, Lyons-Wall P, Russell A. Trial of a mobile phone method for recording dietary intake in adults with type 2 diabetes: evaluation and implications for future applications. J Telemed Telecare 2011;17(6):318-323. [doi: 10.1258/jtt.2011.100906] [Medline: 21844173]
27. Cordeiro F, Bales E, Cherry E, Fogarty J. Rethinking the mobile food journal: exploring opportunities for lightweight photo-based capture. Presented at: 33rd Annual ACM Conference on Human Factors in Computing Systems; Apr 18 to 23, 2015; Seoul, Republic of Korea. [doi: 10.1145/2702123.2702154]

28. Zepeda L, Deal D. Think before you eat: photographic food diaries as intervention tools to change dietary decision making and attitudes. Int J Consumer Studies 2008 Nov;32(6):692-698. [doi: 10.1111/j.1470-6431.2008.00725.x]

Edited by L Buis; submitted 22.08.23; peer-reviewed by E Pustozerov, L Gemming; revised version received 27.11.23; accepted 22.12.23; published 15.04.24.

Please cite as:
Lee L, Hall R, Stanley J, Krebs J
Tailored Prompting to Improve Adherence to Image-Based Dietary Assessment: Mixed Methods Study
JMIR Mhealth Uhealth 2024;12:e52074
URL: https://mhealth.jmir.org/2024/1/e52074
doi:10.2196/52074

Original Paper

Reflective Engagement With a Digital Physical Activity Intervention Among People Living With and Beyond Breast Cancer: Mixed Methods Study

Michael C Robertson[1,2], MPH, PhD; Emily Cox-Martin[3], PhD; Karen Basen-Engquist[4], MPH, PhD; Elizabeth J Lyons[2], MPH, PhD

[1]Department of Family and Preventive Medicine, TSET Health Promotion Research Center, University of Oklahoma Health Sciences Center, Oklahoma City, OK, United States

[2]Department of Nutrition, Metabolism & Rehabilitation Sciences, University of Texas Medical Branch at Galveston, Galveston, TX, United States

[3]VA Puget Sound Health Care System, Tacoma, WA, United States

[4]Department of Health Disparities Research, University of Texas MD Anderson Cancer Center, Houston, TX, United States

Corresponding Author:
Michael C Robertson, MPH, PhD
Department of Family and Preventive Medicine
TSET Health Promotion Research Center
University of Oklahoma Health Sciences Center
655 Research Pkwy #400
Oklahoma City, OK, 73104
United States
Phone: 1 405 271 6872
Email: michael-robertson@ouhsc.edu

Abstract

Background: People living with and beyond breast cancer can face internal barriers to physical activity (eg, fatigue and pain). Digital interventions that promote psychological acceptance and motivation may help this population navigate these barriers. The degree to which individuals (1) adhere to intervention protocols and (2) reflect on and internalize intervention content may predict intervention efficacy.

Objective: The objective of this study was to characterize the nature of reflective processes brought about by an 8-week acceptance- and mindfulness-based physical activity intervention for insufficiently active survivors of breast cancer (n=75). Furthermore, we explored the potential utility of a metric of reflective processes for predicting study outcomes.

Methods: Of the intervention's 8 weekly modules, 7 (88%) included an item that asked participants to reflect on what they found to be most useful. Two coders conducted directed content analysis on participants' written responses. They assessed each comment's depth of reflection using an existing framework (ranging from 0 to 4, with 0=*simple description* and 4=*fundamental change with consideration of social and ethical issues*). The coders identified themes within the various levels of reflection. We fit multiple linear regression models to evaluate whether participants' (1) intervention adherence (ie, number of modules completed) and (2) the mean level of the depth of reflection predicted study outcomes.

Results: Participants were aged on average 57.2 (SD 11.2) years, mostly non-Hispanic White (58/75, 77%), and mostly overweight or obese (54/75, 72%). Of the 407 responses to the item prompting personal reflection, 70 (17.2%) were rated as reflection level 0 (ie, *description*), 247 (60.7%) were level 1 (ie, *reflective description*), 74 (18.2%) were level 2 (ie, *dialogic reflection*), 14 (3.4%) were level 3 (ie, *transformative reflection*), and 2 (0.5%) were level 4 (ie, *critical reflection*). Lower levels of reflection were characterized by the acquisition of knowledge or expressing intentions. Higher levels were characterized by personal insight, commentary on behavior change processes, and a change of perspective. Intervention adherence was associated with increases in self-reported weekly bouts of muscle-strengthening exercise (B=0.26, SE 0.12, 95% CI 0.02-0.50) and decreases in sleep disturbance (B=−1.04, SE 0.50, 95% CI −0.06 to −2.02). The mean level of reflection was associated with increases in psychological acceptance (B=3.42, SE 1.70, 95% CI 0.09-6.75) and motivation for physical activity (ie, integrated regulation: B=0.55, SE 0.25, 95% CI 0.06-1.04).

Conclusions: We identified a useful method for understanding the reflective processes that can occur during digital behavior change interventions serving people living with and beyond breast cancer. Intervention adherence and the depth of reflection

each predicted changes in study outcomes. Deeper reflection on intervention content was associated with beneficial changes in the determinants of sustained behavior change. More research is needed to investigate the relations among digital behavior change intervention use, psychological processes, and intervention efficacy.

(JMIR Mhealth Uhealth 2024;12:e51057) doi:10.2196/51057

KEYWORDS

survivors of cancer; exercise; acceptance and commitment therapy; fatigue; mindfulness; motivation; behavioral sciences

Introduction

Background

There were an estimated 4.1 million people living with and beyond breast cancer in 2022 [1]. This number is estimated to reach nearly 5 million by 2030 [2]. This population can encounter long-term challenges related to health and quality of life, including fatigue, anxiety and depressive symptoms, breast cancer–related lymphedema, metabolic dysregulation, bone loss and osteoporosis, and cancer recurrence [3,4]. Physical activity may protect people living with and beyond breast cancer from these problems [5-7]. However, most people who have been diagnosed with breast cancer do not meet nationally recommended physical activity guidelines [8,9].

Psychotherapy-informed digitally delivered acceptance- and mindfulness-based approaches may facilitate physical activity promotion for people living with and beyond breast cancer. People living with and beyond breast cancer commonly cite uncertainty, frustration, cancer-related fatigue, and pain to be barriers to physical activity [10-13]. Interventions that can help people to better navigate the uncomfortable thoughts and sensations that can act as impediments to physical activity may support their efforts to be more physically active. Authoritative entities recommend acceptance- and mindfulness-based interventions for people living with and beyond cancer; strong evidence supports their efficacy for reducing anxiety and depressive symptoms in this population, and they have been shown to reduce cancer-related fatigue [14-17]. Furthermore, physical activity interventions that simultaneously target both health-related behaviors and quality-of-life issues are more effective at achieving sustained physical activity outcomes than interventions that only promote physical activity for this population [18]. Emerging evidence suggests that acceptance- and mindfulness-based physical activity interventions can be feasible and effective at promoting physical activity [19-21] and that digitally delivered approaches can be acceptable and potentially effective for people living with and beyond cancer [22,23].

The degree to which participants meaningfully engage with and internalize health promotion content may determine the efficacy of digital behavior change interventions (DBCIs). However, achieving high levels of engagement can be a marked challenge for DBCIs [24]. Participants often exhibit low adherence to digital physical activity interventions and poor study retention [25]. Accordingly, researchers have tended to prioritize measuring and optimizing intervention use (eg, optimizing metrics such as the number of modules completed and time spent in an app) [26]. However, using system use data alone as a proxy for engagement is problematic. First, the relationship between these metrics and intervention efficacy is not always straightforward. Although there is sometimes a positive linear relationship between DBCI use and intervention efficacy, this is not always the case [25,27,28]. DBCIs may tend to have certain thresholds of engagement that confer the majority of the benefits, or intervention components may interact with one another. Furthermore, engagement with a DBCI may be expected to taper over time and even be supplanted by engaging in the desired behavior itself [29]. Second, focusing exclusively on quantitative use metrics affords only limited insight into the psychological aspects underpinning participants' experiences with DBCIs. This precludes achieving a full understanding of the DBCI's mechanisms of action. Thus, researchers have called for broadening our conceptualization of DBCI engagement [28].

A growing literature highlights the importance of understanding the psychological processes that occur when participants engage with DBCI content; this includes supplementing system use metrics with additional aspects of DBCI engagement that pertain to affect, attention, interest, immersion, flow, and reflection [26,29,30]. Researchers of human-computer interaction have discussed engagement in similar terms, parsing engagement according to notions of behavioral adherence, behavioral effort (eg, discussing emotions, thoughts, and behaviors in intervention), cognition, and affect [31]. Researchers with an educational perspective have provided similar definitions of engagement with digital systems, noting that the cognitive and affective aspects of engagement are largely neglected [32]. Yang et al [25] have contributed a framework specific to physical activity promotion in mobile health in which they conceptualize engagement as being determined by breadth, depth, interaction, and length of engagement. These broad conceptualizations are useful for helping to orient researchers toward the many and varied facets of engagement. Rather than attempting to establish a universal conceptualization, it may be advantageous to tailor the conceptualizations of engagement based on context [33]. Using qualitative and mixed methods to investigate the nature of participants' engagement with DBCIs may be particularly useful to this end [24,26,29].

Reflection is an essential process for self-improvement and making lasting behavior changes in the context of DBCIs [34]. As applied to health-related education, reflection and reflective processes connotate deliberate critical analysis of knowledge and experiences to achieve a fuller understanding [35]. The process of reflecting on the outcomes of past and ongoing efforts or newly acquired information helps individuals integrate knowledge and skills into practice and overcome persistent barriers to behavior change [35,36]. Reflection may be particularly relevant in the context of DBCIs because it may allow people to gain meaningful insights from their personal

health-related data that can support lasting behavior change [34]. In the education literature, reflective thinking and writing are commonly used strategies to prompt personal reflection, and it is a common practice to evaluate written responses to determine to what degree an individual has considered and applied didactic content in the context of their own lived experiences [37-39]. Fleck and Fitzpatrick [40] present a framework that operationally defines 5 levels of reflection. This framework has been shown to be particularly useful for evaluating the levels of reflection promoted by the design features of DBCIs [34]. In this study, we sought to extend this literature by characterizing the reflective processes engaged in by people living with and beyond cancer who experienced a digital acceptance- and mindfulness-based physical activity intervention.

Objectives

It may be possible to help people living with and beyond breast cancer reframe some of the unpleasant internal sensations that can act as barriers to physical activity (eg, pain, fatigue, and frustration). Understanding the degree to which participants reflect on and internalize digital intervention content targeting these psychological processes may be useful for understanding how best to support this population. This study answers calls to investigate the deeper psychological aspects involved in engagement with DBCIs to provide insight into the interplay among intervention use, individual experiences, and intervention efficacy. The aims of this study were to (1) characterize the nature of reflective processes brought about by a DBCI designed to increase physical activity in insufficiently active people living with and beyond breast cancer and (2) explore the potential utility of a metric of reflective processes in DBCIs within an acceptance- and mindfulness-based physical activity intervention.

Methods

Study Design

We conducted a secondary analysis on data obtained from a 1-group pilot study. The purpose of the parent study was to evaluate the acceptability of an acceptance- and mindfulness-based intervention to increase physical activity in survivors of breast cancer [23]. This study is an investigation of the use of the intervention. Participants were female adults (aged ≥18 y) who had been diagnosed with breast cancer but were not actively preparing for surgery or undergoing chemotherapy or irradiation treatment (n=75). Eligibility criteria included that participants reported engaging in <150 minutes of moderate-intensity aerobic physical activity per week. We recruited participants using a large listserve of individuals who were interested in receiving information about breast cancer–related research studies. Study staff contacted interested individuals via telephone to assess eligibility.

Ethical Considerations

Study procedures were approved by the University of Texas School of Public Health Committee for the Protection of Human Subjects (HSC-SPH-18-1025). All participants provided informed consent for participation.

Intervention Description

The intervention has been described in detail elsewhere [23]. Briefly, the ACTive program was an 8-week DBCI designed to help insufficiently active survivors of breast cancer increase moderate-intensity aerobic and muscle-strengthening physical activity. It was centered on increasing physical activity acceptance (ie, cognitive acceptance and behavioral commitment) and autonomous motivation for physical activity (ie, enjoyment, values, interest, and identification). It was grounded in acceptance and commitment therapy (ACT) principles (ie, values, committed action, acceptance, defusion, and contacting the present moment). It consisted of 8 modules (along with a brief introductory module), each of which was delivered electronically via a weekly REDCap (Research Electronic Data Capture; Vanderbilt University) survey. The modules included didactic content and experiential exercises targeting core ACT principles. They were presented in brief videos, audio files, images, and other documents. Participants were prompted to input information periodically (eg, physical activity levels and reactions and thoughts on experiential exercises). Branching logic within REDCap was used to remind participants of their responses to items from previous modules as well as provide feedback on their responses within modules (eg, an emoji appeared if responses indicated that the participant had met her physical activity goal for the week). The modules contained a repository of additional optional content (eg, muscle-strengthening physical activity videos for survivors of cancer, videos with yoga classes for survivors of breast cancer, and a video on proper walking posture for survivors of cancer).

At the end of modules 1 to 7, an item asked participants, “What is one important, personal take-away point from this session?” This item was included to identify specific aspects of the intervention that were perceived as the most *useful* by participants because usefulness is an important component of acceptability. Participants were presented with an open-text box to provide a reply. Providing participants with the opportunity to reply to open-ended items such as this can help contribute to a more in-depth understanding of participants’ engagement and experience with DBCI content [26].

Measures

Physical Activity Behaviors

The Godin Leisure-Time Exercise Questionnaire was administered to obtain pre- and postintervention estimates of participants’ average weekly leisure time aerobic physical activity over the past month [41,42]. To obtain estimates of participants’ muscle-strengthening physical activity levels, we included an item derived from the Godin Leisure-Time Exercise Questionnaire [43,44]. This item asked, “In a typical week, outside of your job or work around the house, how many days do you do leisure-time physical activities specifically designed to strengthen your muscles such as lifting weights, circuit training, or resistance bands? (Do not include cardio/aerobic types of exercise).” Response options ranged from 0 to 7.

Physical Activity Acceptance

The ACTive program was centered on increasing physical activity acceptance. This was defined as one’s willingness to

experience the negative internal experiences that can sometimes be associated with physical activity, rather than avoiding them. This was operationalized by the Physical Activity Acceptance Questionnaire (PAAQ) [45]. The PAAQ consists of two 5-item subscales. The *cognitive acceptance* subscale measures one's propensity to accept the reality of unpleasant sensations associated with physical activity, whereas the *behavioral commitment* subscale pertains to persisting in committed action despite the occurrence of challenging thoughts or sensations. Responses on items of the PAAQ range from 1=*never true* to 7=*always true*. For scoring, the items on the *cognitive acceptance* subscale are reverse coded, and the items of each subscale are summed (range 5-35). This questionnaire was administered before and after the intervention.

Motivation for Physical Activity

The ACTive program also aimed to increase motivation for physical activity as conceptualized by self-determination theory (SDT). SDT parses motivation conceptually based on the degree to which it is autonomous in nature. SDT posits that changes in more autonomous motivations (eg, integrating a behavior into relevant self-narratives) will yield longer-lasting behavior changes than changes in less autonomous motivations (eg, receiving a *badge* as reinforcement for performance). We operationalized motivation for physical activity before and after the intervention using the Behavioral Regulation for Exercise Questionnaire-3 (BREQ-3). The BREQ-3 consists of six 4-item subscales (*amotivation*, *external regulation*, *introjected regulation*, *identified regulation*, *integrated regulation*, and *intrinsic regulation*). The *intrinsic regulation* subscale captures a highly autonomous form of motivation, defined by the degree to which one engages in a behavior because one finds it inherently interesting or enjoyable. The *integrated regulation* subscale captures another highly autonomous form of motivation, an extrinsic form of motivation defined by the degree to which an individual has fully internalized the reason for action owing to finding it concordant with their values. The *identified regulation* subscale captures a somewhat less autonomous form of motivation (although still relatively autonomous overall), defined by the degree to which an individual consciously values a reason for action. For scoring, the mean scores for each set of items are calculated (range 0-4).

Health-Related Outcomes

For exploratory purposes, we administered measures of quality of life and physical functioning before and after the intervention. To operationalize these constructs, we used the National Institutes of Health–funded Patient-Reported Outcomes Measurement Information System (PROMIS)-29 profile measure (version 2.1) [42]. The PROMIS-29 consists of 8 subscales: *physical function* (eg, "Are you able to run errands and shop?"), *anxiety* (eg, "In the past 7 days...I felt fearful"), *depressive symptoms* (eg, "In the past 7 days...I felt worthless"), *fatigue* (eg, "In the past 7 days...I felt fatigued"), *sleep disturbance* (eg, "In the past 7 days...I had a problem with my sleep"), *ability to participate in social roles and activities* (eg, "I have trouble doing all of my regular leisure activities with others"), *pain interference* (eg, "In the past 7 days...How much did pain interfere with your day to day activities?"), and *pain intensity* (eg, "How would you rate your pain on average?"). All subscales except the *pain intensity* subscale have 4 items and 5-point Likert-type responses ranging from 0 (*not at all*) to 4 (*very much*). The *pain intensity* subscale has 1 item and an 11-point scale ranging from 0 (*no pain*) to 10 (*worst pain imaginable*). Scores are coded, summed, and converted to T-scores such that higher scores indicate more of the concept being measured (eg, range for *physical functioning* subscale: 22.5-57.0).

Reflection Framework

Reflection levels were based on the reflection framework presented by Fleck and Fitzpatrick [40]. This framework emerged from research conducted in the context of the design of digital technologies and human-computer interaction. It defines 5 levels of reflection intended to serve as a resource for thinking about, and designing for, reflection. Reflection level 0 (revisiting) is defined as "Description or statement about events without further elaboration or explanation. Not reflective." Reflection level 1 (reflective description) is defined as "Description including justification or reasons for action or interpretation, but in a reporting or descriptive way. No alternate explanations explored, limited analysis and no change of perspective." Reflection level 2 (dialogic reflection: exploring relationships) is defined as "A different level of thinking about intervention content. Identifying or exploring relationships between relevant concepts. Applying experience or knowledge, providing evidence of cycles of interpreting and questioning, consideration of different explanations, hypotheses and/or other points of view." Reflection level 3 (transformative reflection: fundamental change) is defined as "Revisiting an event or knowledge with intent to re-organize and/or do something differently. Asking fundamental questions and challenging personal assumptions leading to a change in practice or understanding." Finally, reflection level 4 (critical reflection: wider implications) is defined as "Social and ethical issues are taken into consideration. Generally considering the (much wider) picture."

Data Analysis

Qualitative Data Analysis

MCR and EJL conducted 2 phases of directed content analysis on participants' written responses to the open-ended item asking, "What is one important, personal take-away point from this session?" [46]. First, the coders independently rated the reflection level of each individual response by evaluating the response against the reflection framework presented by Fleck and Fitzpatrick [40]. Throughout this process, they produced descriptors to extend and apply the definitions presented by Fleck and Fitzpatrick [40] of the various reflection levels to our study context. The 2 coders first coded all the responses from 1 module independently. Next, they met to discuss the functional definitions and descriptors of reflection levels in our study context and resolve coding discrepancies. The coders then evaluated the rest of the responses independently. After doing so, they met to reconcile discrepancies. The coders created a table, based on the reflection framework presented by Fleck and Fitzpatrick [40] as well as supplemental descriptors and illustrative examples from this study, to further clarify their conception of the various reflective levels as they might be

applied in the context of DBCIs. For the second phase of directed content analysis, MCR sorted responses by reflection level and MCR and EJL coded responses using inductive codes that were informed by principles of behavioral science (eg, behavior change techniques) and ACT [46]. MCR coded all responses first and provided a list of inductive codes to EJL. Next, EJL coded all responses using the list of inductive codes provided by MCR and adding additional codes as needed. The 2 coders then met to discuss codes and reconcile differences.

Quantitative Data Analysis

We computed descriptive statistics for participant demographics, study outcome variables, intervention adherence, and each participant's mean level of reflection (ie, each individual's mean score of the items scored via the qualitative procedures detailed in the previous subsection). We defined intervention adherence as the number of modules that each participant completed; this is a commonly used measure of engagement with DBCIs [26]. We then conducted multiple linear regression analyses with maximum likelihood estimation. The independent variables were (1) the number of modules completed and (2) participants' mean level of reflection. The dependent variables (in separate models) were the follow-up measures of self-reported aerobic and muscle-strengthening physical activity, the PAAQ, the BREQ-3, and the PROMIS-29 subscales described previously. All analyses adjusted for the baseline value of the dependent variable and, given our interest in parsing the possible effects of the breadth versus the depth of engagement, the other independent variable of interest (ie, both intervention adherence and the mean level of reflection were included in all models). All analyses also adjusted for sociodemographic and cancer-related factors that we identified a priori as potentially confounding variables. Specifically, we adjusted for age (years), education level (no bachelor's degree, bachelor's degree, or graduate school), ethnicity (Hispanic or not Hispanic), race (American Indian or Alaska Native or other, Asian, Black or African American, or White), BMI category (underweight, normal, overweight, or obese), time since cancer diagnosis (years), and stage at diagnosis (1, 2, or 3/4). Missing data were handled using full information maximum likelihood. We set our nominal α level to .05 for all analyses. All statistical analyses were performed using R (version 4.0.3; R Foundation for Statistical Computing).

Results

Participant Demographic Characteristics

The average age of the participants was 57.2 (SD 11.2; range 31-78) years. The median time since breast cancer diagnosis was 8.2 (IQR 3-12) years. The participants were relatively well educated, mostly non-Hispanic White (58/75, 77%), and mostly overweight or obese (54/75, 72%; Table 1).

Table 1. Participant characteristics.

Characteristic and category	Participants, n (%)
Education level (n=75)	
High school diploma or GED[a]	0 (0)
Some college	16 (21)
Bachelor's degree	34 (45)
Graduate school degree	25 (33)
Stage of breast cancer at diagnosis (n=71)	
1	32 (45)
2	28 (39)
3	9 (13)
4	2 (3)
Race (n=75)	
American Indian, Alaska Native, or other	1 (1)
Asian	4 (5)
Black or African American	7 (9)
White	63 (84)
Ethnicity (n=74)	
Hispanic	7 (10)
Non-Hispanic	67 (91)
Marital status (n=74)	
Single	11 (15)
Married	54 (73)
Living with significant other	1 (1)
Divorced	5 (7)
Widowed	3 (4)
Employment status (n=68)	
Employed full-time	38 (56)
Employed part-time	10 (15)
Retired	20 (29)
BMI status (n=74)	
Underweight	1 (1)
Normal	19 (26)
Overweight	32 (43)
Obese	22 (30)

[a]GED: General Educational Development Test.

Intervention Adherence

The median number of the 8 modules completed was 8 and the first quartile value was 4 modules (IQR 4). The minimum number of modules completed was 1 (7/75, 9%).

Coding the Levels of Reflection

The coders created a table while going through the process of coding the reflections based on the reflection framework presented by Fleck and Fitzpatrick [40]. We added descriptors and illustrative examples from this study to further clarify our conception of the various reflective levels as they might be applied in the context of DBCIs (Table 2).

Table 2. Levels of reflection with relevant descriptors and illustrative examples.

Reflection level[a] and additional descriptors	Illustrative examples
Reflection level 0. Description: revisiting (description or statement about events without further elaboration or explanation; not reflective)	
Repeating or paraphrasing intervention content	• "That exercise makes you less tired but also helps you sleep Better." [SID[b] 157, 47 years]
Superficial imperative statements	• "[G]et up and move." [SID 95, 49 years]
Surface-level comments	• "I thought the extended mindfulness video was helpful." [SID 100, 37 years]
Platitudes	• "Life is hard sometimes." [SID 99, 68 years]
Not responsive to the prompt	• "Bad weather." [SID 117, 64 years]
Reflection level 1. Reflective description: revisiting with explanation (description including justification or reasons for action or interpretation but in a reporting or descriptive way; no alternate explanations explored, limited analysis, and no change of perspective)	
Elaborating upon intervention content	• "On our sickest days, there is more going right in our bodies than going wrong. I need to appreciate this. And celebrate this." [SID 113, 50 years] • "[T]hat the uncomfortable feel of exercise is actually good for me. I just need to embrace it." [SID 140, 60 years]
Imperative statements (including justification or reasons or descriptive strategy)	• "I need to get moving, so I will feel better." [SID 79, age not given]
Personal insight	• "[T]hat I'm good at putting things off." [SID 134, 60 years] • "I am totally the kid looking out the window—I try to make everything an adventure and to look at the positive mental attitude." [SID 103, 45 years]
Skill building or learning a technique	• "I've learned how to focus away from the chatter in my brain." [SID 99, 68 years] • "[T]hat I can 'pick up' internal barriers which relieves some of the feeling of frustration and feeling like a failure." [SID 98, 53 years] • "I liked the shear stress explanation. I can picture that while I exercise." [SID 100, 37 years]
Reflection level 2. Dialogic reflection: exploring relationships (a different level of thinking about intervention content; identifying or exploring relationships between relevant concepts; applying experience or knowledge, providing evidence of cycles of interpreting and questioning, and consideration of different explanations, hypotheses, and other points of view)	
Applying intervention content to one's own life	• "My old excuse of saying 'I just don't want to exercise,' is not a good enough reason. I know the reasons are that it's uncomfortable and inconvenient, but that's really not true. Also, I know I can adapt and it won't cause pain in my hip with bursitis." [SID 145, 48 years]
Commentary on the nature of the relationships between disparate concepts	• "If I am going to change my fitness habits, I must see how they relate to my values." [SID 124, 42 years]
Taking a different perspective	• "[T]hat I'm being invited, not required to experiment with activity and that I can choose how to do it. I liked the line about experimenting with what we're being told the benefits are in our own bodies. That makes me feel more in control and interested." [SID 137, 53 years]
Applying new skills or knowledge and reflecting on this	• "In the almost 14-minute extended mindfulness exercise, I found that it did relax me even though that wasn't the objective. I kept my eyes closed during the entire exercise. At first, my other thoughts included anxiety over today's election. One of the suggestions was to look at yourself from outside yourself. I find I'm usually able to do that anyway...as it enables me to be more compassionate and understanding of others' pain because of what I've experienced. For me, the only distraction, as I kept my eyes closed, was hearing the/your voice telling me things. I had no problem with pushing away thoughts because the only thing I was seeing was inside my eyes as they were closed. Once opened, everything becomes a distraction." [SID 133, 70 years]
Reflection level 3. Transformative reflection: fundamental change (revisiting an event or knowledge with intent to reorganize and do something differently; asking fundamental questions and challenging personal assumptions leading to a change in practice or understanding)	

Reflection level[a] and additional descriptors	Illustrative examples
Perspective significantly altered	• "[T]hat I and many other people get very wrapped up in goals and disguise them as values. We find that we're not measuring up and we get disappointed in ourselves and give up working at them. Looking at goals as finite and values as infinite and guiding principles that shape goals puts them in a different perspective. It gives me hope that I won't judge myself too harshly if I don't fulfill my goals, that I'll realize that maybe that goal wasn't right for me and didn't fit in well with my values." [SID 137, 53 years]
A change in practice or understanding resulting from personal insight	• "I am stronger than the spoiled, damaged, hurt child inside me. I don't have to listen to her. Just put her plump pouty face in my backpack and carry on." [SID 82, 64 years]
Reflection level 4. Critical reflection: wider implications (social and ethical issues are taken into consideration; generally considering the wider picture)	
A transformative reflection that weaves in broader social or ethical considerations	• "My essential Self is still there...I am alive, the tiny kernel of me, the spark, though almost extinguished, has to be nurtured above all else now...or I am extinguished by the grotesque cancer industry conveyor belt." [SID 110, 74 years]

[a]From Fleck and Fitzpatrick [40].

[b]SID: study identification number.

Levels of Reflection

Overall

There were 407 total responses from the 75 participants over the course of the 7 modules that featured the open-ended item for reflection (the item was not included in module 8). Participants submitted an average of 5.4 (SD 2.2) out of 7 possible responses. Of the 407 responses, we rated 70 (17.2%) as reflection level 0 (ie, *description*), 247 (60.7%) as level 1 (ie, *reflective description*), 74 (18.2%) as level 2 (ie, *dialogic reflection*), 14 (3.4%) as level 3 (ie, *transformative reflection*), and 2 (0.5%) as level 4 (ie, *critical reflection*).

Reflection Level 0 (Description: Revisiting)

Responses that were rated as reflection level 0 were judged to be not reflective and often emphasized one's desire or intention to increase one's physical activity levels, simply repeated or paraphrased subject matter content, or provided commentary on the delivery of the subject matter itself. The themes we identified for reflection level 0 responses related to (1) *making a resolution*, (2) *knowledge of subject matter*, and (3) *appreciation or distaste for intervention content*.

Participants often made short imperative statements concerning a general need to increase physical activity or mindfulness (ie, *making a resolution*), but at this level of reflection, they did not demonstrate sufficient explanation to extend their response beyond merely revisiting intervention content (eg, "Just start!" [study identification number (SID) 148, 64 years]).

The theme related to *knowledge of subject matter* characterized a subset of responses that were descriptive paraphrases of didactic intervention content related to physical activity or ACT principles; for example, a participant wrote, "Even 10 minutes of activity is better than none" (SID 152, 54 years).

Responses that spoke to an *appreciation or distaste for intervention content* generally provided commentary on the intervention content without evidencing deeper reflection:

> *No real huge "ah ha" moments from this since it was all review. I had some problems with technology this time and got booted out twice before saving my answers so fewer long responses this week.* [SID 103, 45 years]

Reflection Level 1 (Reflective Description: Revisiting With Explanation)

Responses that were rated as reflection level 1 were judged to provide more justification than reflection level 0 comments, but this additional substance was generally descriptive in nature and without evidence of a deeper change in perspective. This reflection level made up the majority of responses (247/407, 60.7%). These comments commonly emphasized one's desire or intention to increase one's physical activity levels with some action-oriented or attitudinal elaboration. The themes we identified for reflection level 1 responses related to (1) *making a resolution*, (2) *personal application or action planning*, and (3) *kindness*.

Making a resolution was also a common theme for the level 1 responses. These responses were generally centered on the importance of increasing physical activity levels and provided a more nuanced explanation than the level 0 responses (eg, "[T]hat fitness in itself is a value that I should prioritize for overall happiness and well-being, not just when I want to lose weight and achieve a summer body" [SID 84, 32 years]).

Responses related to *personal application or action planning* tended to apply information from the intervention to the participant's own life. This was often related to engaging in physical activity (eg, "One of the important takeaways for me is the proper way to stand and walk which allows our back to be straight...and assist with our posture" [SID 87, 64 years]). This theme also applied to using acceptance- and mindfulness-based techniques (eg, "Practical Action idea: Pausing to 'unpack' the negative thoughts, physically remove them" [SID 110, 74 years]).

Finally, the *kindness* theme captured responses that included an emphasis on being kind to oneself in the course of gradually

increasing physical activity levels. This theme was also evident in comments that spoke to an appreciation of intervention content that normalized the reality of challenges experienced in the course of engaging in physical activity and cancer survivorship:

> *Life is hard and challenging and that is NORMAL. A rich full life is one that consists of a variety of emotions.* [SID 93, 56 years]

Reflection Level 2 (Dialogic Reflection: Exploring Relationships)

Responses that were rated as reflection level 2 were judged to exhibit reflection on more complex conceptual relationships than responses rated as reflection level 1. These responses often evidenced a change of perspective or a more nuanced consideration of the intervention content and its application to one's own life. Reflection level 2 responses commonly displayed many of the themes presented previously (eg, *making a resolution* and *personal application or action planning*) but were further characterized by the themes of (1) *personal insight* or (2) *discussion of personal facilitators or mechanisms of behavior change*.

The theme of *personal insight* characterized comments that involved elements of personal introspection and reflection on oneself or one's own thought and behavioral patterns:

> *It got me to thinking about what the negative thought might be. Why do I not want to exercise if I know it's good for me? Somewhere way, way deep down I don't think I believe it. That's helpful to realize and good to see in print. Now to figure it out.* [SID 120, 78 years]

Other participants shared candid observations about the difficulties involved with satisfying perceived needs:

> *Need to refocus on what has been important, can be again but with physical limitations like a nagging injury, isolation due to the pandemic, restrictions where I live, too comfortable with aloneness now.* [SID 139, 76 years]

The theme of *discussion of personal facilitators or mechanisms of behavior change* applied to responses that evidenced reflection on what might help increase physical activity or how being more physically active might in turn support a valued aspect of the participant's life:

> *I need to learn how to separate the thought from who I really am. I realize that I am overwhelmed by the "starting point," that place where I am required to overcome the law of inertia in all the things I want or need to do. Once I get started I am generally able to complete the task.* [SID 98, 53 years]

Reflection Level 3 (Transformative Reflection: Fundamental Change)

This level of reflection was relatively rare (14/407, 3.4%). Responses determined to represent reflection level 3 were characterized by substantively deeper reflection than responses determined to represent reflection level 2—additionally marked by evidence suggestive of more profound *personal insight* and a fundamental *change in perspective*. In reflections that attained this level of depth, individuals often intimated that they revised personally held beliefs in ways that were empowering and conducive to enduring change:

> *I think about "the body keeps the score" and how some of the problems and thoughts and feelings I have are programmed into my DNA just like my propensity towards cancer. Some, but not all. I need to look at those moments where I am able to distract myself and notice the little victories. I'll never be perfect—never have a perfect body a perfect kid a perfect life, but there are moments every day where I feel I'm doing it right and I need to notice those moments. Last week i ran 2 miles without stopping. Not much—I used to run 5 without stopping—but that was a great feeling. Last week I carved pumpkins with my son and my neighbor came over and we talked for over an hour. Thoughts of being unlovable and depressed faded during those moments. Also I'm trying to stay away from situations where I feel this loneliness and unlovable-ness... staying off the dating site I paid so much for hurts in one way but since nobody was calling me anyways it's better just to not open the app up.* [SID 104, 50 years]

Participants evidencing this level of reflection often reframed challenges that related to their cancer journey or spoke about appreciating life from a different perspective:

> *I love the point just made: "Even if we fail at our goals, if we act in a way that's consistent with our values we are successful." I didn't do as much walking this week, but I did add strength training this week and it's amazing how much more I felt. And that's not a typo. I don't know how else to describe it than I just felt "more."* [SID 113, 50 years]

Reflection Level 4 (Critical Reflection: Wider Implications)

Responses reaching reflection level 4 were very rare (2/407, 0.5%). These responses were judged to meet the level of intrapersonal depth characteristic of reflection level 3 as well as to incorporate some of the wider social and ethical contexts in which the participants' experiences were situated. The responses spoke to a pressing tension the participants experienced as survivors of cancer in navigating social and community-level factors (1 quote is presented herein, and another quote is presented in Table 2):

> *Dr. Harris' video was great to hear. When I was going through chemo some people would say, "Keep a positive attitude" and I didn't always want to. I wanted to withdraw or rage sometimes. I think it's easier for others to see you happy, it relieves their tension about what is happening to you. It does not mean you are not living a rich and fulfilling life and I was HAPPY to hear him say that.* [SID 120, 78 years]

Intervention Adherence Predicting Outcomes

We observed associations between intervention adherence and change in 2 outcomes of interest (Table 3). Higher intervention adherence was associated with higher weekly bouts of muscle-strengthening physical activity at follow-up, adjusting for baseline levels, age, time since diagnosis, stage at diagnosis, race, ethnicity, education level, BMI category, and mean reflection level (B=0.26, SE 0.12, 95% CI 0.02-0.50; Cohen *d*=2.17). Participants in the fourth quartile of intervention adherence (21/75, 28%) averaged an increase of 0.38 bouts of muscle-strengthening physical activity in a typical week from baseline to follow-up, whereas those in the first quartile (40/75, 53%) averaged an increase of 1.43 bouts in a typical week. Similarly, higher intervention adherence was associated with less sleep disturbance at follow-up, adjusting for baseline levels and other covariates (B=−1.04, SE 0.50, 95% CI −2.02 to −0.06; Cohen *d*=2.08). Participants in the fourth quartile of intervention adherence averaged a decrease of 2.31 in the PROMIS-29 sleep disturbance score from baseline to follow-up, whereas those in the first quartile averaged a decrease of 4.68 in the PROMIS-29 sleep disturbance score. We did not observe statistically significant associations between intervention adherence and other outcomes of interest.

Table 3. Results from multiple linear regression analyses with independent variables regressed on intervention adherence and mean reflection level (n=75).

Outcome variable	Intervention adherence[a]	Mean reflection level[b]
	Estimate (SE; 95% CI)	Estimate (SE; 95% CI)
Self-reported weekly physical activity		
Aerobic moderate to vigorous exercise (min)	11.75 (8.08; −4.09 to 27.59)	−30.84 (31.50; −92.58 to 30.90)
Muscle-strengthening exercise (bouts)	0.26 (0.12; 0.02 to 0.50)	0.11 (0.49; −0.85 to 1.07)
BREQ-3[c]		
Identified regulation	0.01 (0.04; −0.07 to 0.09)	0.06 (0.16; −0.25 to 0.37)
Integrated regulation	0.00 (0.06; −0.12 to 0.12)	0.55 (0.25; 0.06 to 1.04)
Intrinsic regulation	0.01 (0.06; −0.11 to 0.13)	0.28 (0.25; −0.21 to 0.77)
PAAQ[d]		
Cognitive acceptance	−0.07 (0.41; −0.87 to 0.73)	3.42 (1.70; 0.09 to 6.75)
Behavioral commitment	0.07 (0.32; −0.56 to 0.70)	0.93 (1.33; −1.68 to 3.54)
PROMIS-29[e]		
Physical functioning	0.37 (0.41; −0.43 to 1.17)	−1.77 (1.65; −5.00 to 1.46)
Anxiety	0.82 (0.49; −0.14 to 1.78)	−1.19 (2.07; −5.25 to 2.87)
Depressive symptoms	−0.03 (0.41; −0.83 to 0.77)	2.45 (1.72; −0.92 to 5.82)
Fatigue	−0.18 (0.65; −1.45 to 1.09)	0.84 (2.66; −4.37 to 6.05)
Sleep disturbance	−1.04 (0.50; −2.02 to −0.06)	1.16 (2.07; −2.90 to 5.22)
Social roles	−0.28 (0.37; −1.01 to 0.45)	−0.83 (1.54; −3.85 to 2.19)
Pain interference	0.47 (0.60; −0.71 to 1.65)	-0.12 (2.62; −5.26 to 5.02)
Pain intensity	0.11 (0.13; −0.14 to 0.36)	0.09 (0.58; −1.05 to 1.23)

[a]Adjusting for age, time since diagnosis, stage at diagnosis, race, ethnicity, education level, BMI category, mean reflection level, and baseline value of the construct.

[b]Adjusting for age, time since diagnosis, stage at diagnosis, race, ethnicity, education level, BMI category, intervention adherence, and baseline value of the construct.

[c]BREQ-3: Behavioral Regulation for Exercise Questionnaire-3.

[d]PAAQ: Physical Activity Acceptance Questionnaire.

[e]PROMIS: Patient-Reported Outcomes Measurement Information System.

Mean Level of Reflection Predicting Outcomes

We observed associations between participants' mean reflection level and change in 2 outcomes of interest (Table 3). Higher mean reflection level was associated with higher integrated regulation at follow-up, adjusting for baseline levels, age, time since diagnosis, stage at diagnosis, race, ethnicity, education level, BMI category, and intervention adherence (B=0.55, SE 0.25, 95% CI 0.06-1.04; Cohen *d*=2.20). Similarly, higher mean reflection level was associated with higher cognitive acceptance of physical activity at follow-up, adjusting for baseline levels and all other covariates (B=3.42, SE 1.70, 95% CI 0.09-6.75; Cohen *d*=2.01). We did not observe statistically significant

associations between mean reflection level and other outcomes of interest.

Discussion

Principal Findings

This study was a secondary analysis of data from a digitally mediated intervention centered on applying principles and techniques from ACT to help people living with and beyond breast cancer overcome internal barriers to physical activity (eg, pain, fatigue, and frustration). We applied the framework presented by Fleck and Fitzpatrick [40] to gauge the depth of reflection evident in participants' written responses to prompts encouraging reflection on the intervention's weekly modules. We conducted content analysis to characterize participants' responses and multiple linear regression analyses to explore to what extent intervention adherence and participants' mean level of reflection were associated with study outcomes (ie, physical activity behaviors, motivation for physical activity, physical activity acceptance, and health-related quality of life). There were 407 written responses from the 75 participants over the course of 7 modules. Most of the responses were rated as either reflection level 0 (ie, *description*; 70/407, 17.2%) or level 1 (ie, *reflective description*; 247/407, 60.7%). Of the 407 responses, 90 (22.1%) demonstrated evidence of deeper levels of reflection. Intervention adherence was associated with more muscle-strengthening physical activity and better sleep outcomes. Mean reflection level was associated with more integrated motivation for physical activity and higher willingness to experience the full range of sensations that may accompany physical activity.

Comparison With Previous Literature

This study extends previous literature that has highlighted the importance of the cognitive and experiential aspects of DBCI engagement. To date, these aspects of DBCI engagement have largely been inferred from behavioral data (ie, system use metrics). In their review of methodologies for measuring engagement with DBCIs, Short et al [26] suggested that researchers should investigate how intervention content affects the cognitive and experiential aspects of DBCI engagement via the inclusion of open-ended items. We took this approach and found that participants' responses to open-ended items could be evaluated using the reflection framework presented by Fleck and Fitzpatrick [40]. This approach is consistent with research in the education literature that has found reflective writing samples to be amenable to the quantitative assessment of the depth of reflection [35].

Our findings regarding the breakdown of reflection levels evident in this study were consistent with reviews that have evaluated reflective writing in the context of health-related education [39,47]. Even among graduate students, the great majority of reflective processes tend to occur at the descriptive level, and transformative and critical reflections are consistently rare [39,47]. There is limited literature applying these or similar techniques to better understand programs for health education and health promotion [39,47]. In this study, we observed considerable variation in the reflective depth of responses within and between individuals. This suggests that depth of reflection may be modifiable in this context. Indeed, reflection is a common goal of educational interventions, and it seems to be modifiable in other contexts [35]. Providing appropriate scaffolding, fostering collaboration in learning, and using varied exercises to stimulate reflection have been identified as techniques that may increase learners' depth of reflection [35,39]. Time constraints, conflicting values, a lack of feedback, and a lack of trust have been identified as barriers to learning in the context of educational interventions promoting student reflection [35,39]. Future research is needed to identify how DBCIs might be designed to facilitate meaningful reflection and most effectively target the psychosocial determinants of lasting behavior change.

The themes identified in our qualitative analysis provided some insight into the variability of cognitive processes occurring in participants who experienced the digital physical activity intervention. Common themes in the lower levels of reflection generally suggested that participants had acquired new knowledge and endorsed intentions to change their physical activity behaviors (eg, *making a resolution* and *knowledge of subject matter*). These are important antecedents of successful behavior change [48]. However, they are generally not sufficient for realizing physical activity adherence goals; nearly half of the people who indicate that they intend to change their physical activity patterns do not do so [49]. The themes that characterized higher levels of reflective comments are concordant with processes that have been shown to moderate this *intention-behavior gap* [50]. *Discussion of personal facilitators or mechanisms of behavior change* and experiencing *personal insight* or a *fundamental change in perspective* conceptually align with the self-regulatory processes and notions of physical activity identity that tend to moderate the relationship between intention and physical activity adherence [50]. Encouraging personal reflection may be an autonomy-supportive approach to health promotion [34,51,52]. More research is needed to better understand if and how engaging in deep reflection may influence the conscious (eg, affective attitude) and automatic (eg, identity) processes that underlie successful behavior change.

Reflection may be a type of cognitive engagement uniquely suited to supporting knowledge transfer into other domains and contexts [36,53,54]. Schon [55] provides detailed commentary on reflective processes and distinguishes *reflection-on-action* (critical retrospective analysis) from *reflection-in-action* (conscious awareness of real-time behavioral modification). The author emphasizes the primacy of the latter as a determinant of sustained change and suggests that *reflection-on-action* may serve as a prerequisite for implementing change in real time. We found participants' average depth of retrospective reflection to be associated with more integrated regulation for physical activity and cognitive acceptance in the context of physical activity but not with other outcomes of interest. It is encouraging that deeper levels of reflection on intervention content were associated with beneficial changes in motivation and physical activity acceptance, given that these were the theory-informed psychosocial constructs that the DBCI targeted [23]. The cognitive acceptance of physical activity has been shown to be associated with long-term changes in objectively measured physical activity, and integrated regulation is similarly predictive

of physical activity [45,56]. However, in this study, we did not observe evidence that the depth of reflection was associated with physical activity–related outcomes or other outcomes of interest. Simple intervention adherence was associated with increases in muscle-strengthening physical activity and reductions in sleep disturbance. This may be due to the intervention's inclusion of practical resources for muscle-strengthening physical activity and mindfulness exercises. Our findings suggest that designing DBCIs to encourage reflective processes may help target some theory-supported mechanisms of action but may not always be necessary to engender changes in desired end points. More research is needed to investigate whether deep reflection mediates changes in key long-term behavioral and health outcomes and to what degree it should be prioritized in DBCIs.

Considering the depth of reflection may be a useful lens through which to evaluate participants' cognitive engagement with the didactic components of DBCIs. At present, DBCIs are commonly oriented toward maximizing participant engagement as measured by system use. However, rather than simply attempting to maximize DBCI system use, it may be beneficial to target aspects of *effective* engagement [29]. In this study, participants' depth of reflection was associated with 2 key study outcomes. It may be possible to design dynamic interventions that optimize psychological processes such as critical reflection. This endeavor would be supported by the ability to derive metrics that reliably reflect underlying psychological processes from participants' verbal output. We demonstrated that applying the reflection framework presented by Fleck and Fitzpatrick [40] is a promising approach to quantifying participants' qualitative data. Emerging technologies such as large language models may be applied to this end to make this process more expedient and conducive to just-in-time adaptive interventions. Digital health promotion efforts centered on optimizing psychological processes such as critical reflection may supplement or supplant approaches that are narrowly oriented toward maximizing system use.

Reflective writing can serve as a means of self-expression and have therapeutic effects [57]. However, it can also be perceived as burdensome by some individuals. Perski et al [30] emphasized the role played by participants' subjective experiences, characterized by attention, interest, and affect, in DBCI engagement. There are likely important trade-offs that occur in optimizing for cognitive versus affective aspects of DBCI engagement. Participants' comments in this study suggested that although many enjoyed the introspective aspects of the intervention, some did not. This observation is concordant with high ratings of acceptability in the parent study for those who completed the study but a less-than-ideal dropout rate (23.7%) [23]. Achieving deeper levels of reflection may have a dynamic and, in some contexts, diametrically opposed relationship with affective or subjective experiences. Particularly given the importance of affect in physical activity behavior [58], future research should investigate the interrelationships among these aspects of DBCI engagement and how to strike the right balance for different individuals in different contexts.

Limitations

The findings from this study must be considered in the context of its limitations. A small sample size, possible selection bias, and high attrition limit the generalizability of the findings of this study; the analytical sample was a convenience sample that was relatively well educated and had limited racial and ethnic diversity. Short et al [26] highlight that the use of qualitative methods to assess DBCI engagement is inherently limited by a lack of generalizability. The application of the reflection framework presented by Fleck and Fitzpatrick [40] along with directed content analytic methods may facilitate comparison in future studies; however, it is important to note that there may be other dimensions of reflection. We were primarily interested in assessing the vertical dimension of reflection (ie, the depth of reflection), but reflection has also been characterized as having an iterative process-oriented dimension that can be nonlinear and cyclic [39]. Given this prospect and the sometimes cyclical nature of behavior change itself [59], it may be beneficial to longitudinally investigate reflective processes and their bidirectional relationships with physical activity behaviors and determinants. Likewise, there are other ways to define intervention adherence. A limitation of this study is that we did not obtain other metrics that might characterize adherence, such as time spent in the modules. In future studies, obtaining more granular data concerning system use may be advantageous. Given that the study was conducted during the COVID-19 pandemic, history was a threat that may have influenced participants' reflective processes and physical activity–related constructs. Reliance on the self-reported assessment of physical activity patterns also has well-documented limitations. Finally, we conducted multiple statistical tests and, although this investigation was explicitly exploratory in nature, the study findings are prone to an inflated chance of type I error.

Conclusions

In this study, we sought to explore a novel method for understanding critical reflection occurring in a DBCI designed to increase physical activity in insufficiently active people living with and beyond breast cancer. We found the application of qualitative content analysis based on the reflection framework presented by Fleck and Fitzpatrick [40] to be a useful tool for helping to gauge the extent to which participants engaged in reflective processes. Furthermore, we found that deeper reflection levels tended to be associated with changes in the targeted psychosocial constructs. Reflecting on newly acquired information is a critical process in integrating relevant insights for sustained behavior change. Encouraging personal reflection is an autonomy-supportive approach to promoting physical activity, and more research is warranted to investigate this approach in DBCIs serving people living with and beyond cancer.

Acknowledgments

MCR was supported by the National Cancer Institute of the National Institutes of Health (NIH; F31 CA236433). This study was supported in part by the Center for Energy Balance in Cancer Prevention and Survivorship. Participant recruitment was facilitated by the Love Research Army of the Dr Susan Love Research Foundation. This study used the MD Anderson Assessment, Intervention, and Measurement Core, a shared resource supported by the NIH through the MD Anderson Cancer Center support grant (P30 CA016672). The content of this paper is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Conflicts of Interest

None declared.

References

1. Miller KD, Nogueira L, Devasia T, Mariotto AB, Yabroff KR, Jemal A, et al. Cancer treatment and survivorship statistics, 2022. CA Cancer J Clin 2022 Sep 23;72(5):409-436 [FREE Full text] [doi: 10.3322/caac.21731] [Medline: 35736631]
2. Miller KD, Nogueira L, Mariotto AB, Rowland JH, Yabroff KR, Alfano CM, et al. Cancer treatment and survivorship statistics, 2019. CA Cancer J Clin 2019 Sep 11;69(5):363-385 [FREE Full text] [doi: 10.3322/caac.21565] [Medline: 31184787]
3. Bodai BI, Tuso P. Breast cancer survivorship: a comprehensive review of long-term medical issues and lifestyle recommendations. Perm J 2015;19(2):48-79 [FREE Full text] [doi: 10.7812/TPP/14-241] [Medline: 25902343]
4. Nardin S, Mora E, Varughese FM, D'Avanzo F, Vachanaram AR, Rossi V, et al. Breast cancer survivorship, quality of life, and late toxicities. Front Oncol 2020;10:864 [FREE Full text] [doi: 10.3389/fonc.2020.00864] [Medline: 32612947]
5. Campbell KL, Winters-Stone KM, Wiskemann J, May AM, Schwartz AL, Courneya KS, et al. Exercise guidelines for cancer survivors: consensus statement from international multidisciplinary roundtable. Med Sci Sports Exerc 2019 Nov;51(11):2375-2390 [FREE Full text] [doi: 10.1249/MSS.0000000000002116] [Medline: 31626055]
6. Schmitz KH, Campbell AM, Stuiver MM, Pinto BM, Schwartz AL, Morris GS, et al. Exercise is medicine in oncology: engaging clinicians to help patients move through cancer. CA Cancer J Clin 2019 Nov 16;69(6):468-484 [FREE Full text] [doi: 10.3322/caac.21579] [Medline: 31617590]
7. Patel AV, Friedenreich CM, Moore SC, Hayes SC, Silver JK, Campbell KL, et al. American College of Sports Medicine roundtable report on physical activity, sedentary behavior, and cancer prevention and control. Med Sci Sports Exerc 2019 Nov;51(11):2391-2402 [FREE Full text] [doi: 10.1249/MSS.0000000000002117] [Medline: 31626056]
8. Blanchard CM, Courneya KS, Stein K, American Cancer Society's SCS-II. Cancer survivors' adherence to lifestyle behavior recommendations and associations with health-related quality of life: results from the American Cancer Society's SCS-II. J Clin Oncol 2008 May 01;26(13):2198-2204. [doi: 10.1200/JCO.2007.14.6217] [Medline: 18445845]
9. Tarasenko Y, Chen C, Schoenberg N. Self-reported physical activity levels of older cancer survivors: results from the 2014 national health interview survey. J Am Geriatr Soc 2017 Feb 09;65(2):e39-e44. [doi: 10.1111/jgs.14589] [Medline: 27943255]
10. Milne HM, Guilfoyle A, Gordon S, Wallman KE, Courneya KS. Personal accounts of exercise and quality of life from the perspective of breast cancer survivors. Qual Life Res 2007 Nov 24;16(9):1473-1481. [doi: 10.1007/s11136-007-9251-z] [Medline: 17717726]
11. Smith-Turchyn J, Richardson J, Tozer R, McNeely M, Thabane L. Physical activity and breast cancer: a qualitative study on the barriers to and facilitators of exercise promotion from the perspective of health care professionals. Physiother Can 2016 Nov;68(4):383-390 [FREE Full text] [doi: 10.3138/ptc.2015-84] [Medline: 27904238]
12. Smith L, Croker H, Fisher A, Williams K, Wardle J, Beeken RJ. Cancer survivors' attitudes towards and knowledge of physical activity, sources of information, and barriers and facilitators of engagement: a qualitative study. Eur J Cancer Care (Engl) 2017 Jul 30;26(4):e12641 [FREE Full text] [doi: 10.1111/ecc.12641] [Medline: 28135016]
13. Cummins C, Kayes NM, Reeve J, Smith G, MacLeod R, McPherson KM. Navigating physical activity engagement following a diagnosis of cancer: a qualitative exploration. Eur J Cancer Care (Engl) 2017 Jul 07;26(4):e12608. [doi: 10.1111/ecc.12608] [Medline: 27925326]
14. Zhang Q, Zhao H, Zheng Y. Effectiveness of mindfulness-based stress reduction (MBSR) on symptom variables and health-related quality of life in breast cancer patients-a systematic review and meta-analysis. Support Care Cancer 2019 Mar;27(3):771-781. [doi: 10.1007/s00520-018-4570-x] [Medline: 30488223]
15. Haller H, Winkler MM, Klose P, Dobos G, Kümmel S, Cramer H. Mindfulness-based interventions for women with breast cancer: an updated systematic review and meta-analysis. Acta Oncol 2017 Dec;56(12):1665-1676. [doi: 10.1080/0284186X.2017.1342862] [Medline: 28686520]
16. Xie C, Dong B, Wang L, Jing X, Wu Y, Lin L, et al. Mindfulness-based stress reduction can alleviate cancer- related fatigue: a meta-analysis. J Psychosom Res 2020 Mar;130:109916. [doi: 10.1016/j.jpsychores.2019.109916] [Medline: 31927347]
17. Carlson LE, Ismaila N, Addington EL, Asher GN, Atreya C, Balneaves LG, et al. Integrative oncology care of symptoms of anxiety and depression in adults with cancer: society for integrative oncology-ASCO guideline. J Clin Oncol 2023 Oct 01;41(28):4562-4591. [doi: 10.1200/JCO.23.00857] [Medline: 37582238]

18. Sheeran P, Wright CE, Listrom O, Klein WM, Rothman AJ. Which intervention strategies promote the adoption and maintenance of physical activity? Evidence from behavioral trials with cancer survivors. Ann Behav Med 2023 Aug 21;57(9):708-721. [doi: 10.1093/abm/kaad002] [Medline: 37078961]
19. Pears S, Sutton S. Effectiveness of acceptance and commitment therapy (ACT) interventions for promoting physical activity: a systematic review and meta-analysis. Health Psychol Rev 2021 Mar;15(1):159-184. [doi: 10.1080/17437199.2020.1727759] [Medline: 32036768]
20. Manchón J, Quiles MJ, León EM, López-Roig S. Acceptance and commitment therapy on physical activity: a systematic review. J Contextual Behav Sci 2020 Jul;17:135-143 [FREE Full text] [doi: 10.1016/j.jcbs.2020.07.008]
21. Martin EC, Galloway-Williams N, Cox MG, Winett RA. Pilot testing of a mindfulness- and acceptance-based intervention for increasing cardiorespiratory fitness in sedentary adults: a feasibility study. J Contextual Behav Sci 2015 Oct;4(4):237-245 [FREE Full text] [doi: 10.1016/j.jcbs.2015.10.005] [Medline: 27104134]
22. Robertson MC, Cox-Martin E, Liao Y, Flores SA, Shegog R, Markham CM, et al. Acceptance- and mindfulness-based techniques for physical activity promotion in breast cancer survivors: a qualitative study. Support Care Cancer 2022 Jan;30(1):465-473 [FREE Full text] [doi: 10.1007/s00520-021-06428-x] [Medline: 34313858]
23. Robertson MC, Cox-Martin E, Shegog R, Markham CM, Fujimoto K, Durand CP, et al. The acceptability of an electronically delivered acceptance- and mindfulness-based physical activity intervention for survivors of breast cancer: one-group pretest-posttest design. JMIR Cancer 2022 Apr 29;8(2):e31815 [FREE Full text] [doi: 10.2196/31815] [Medline: 35486425]
24. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
25. Yang Y, Boulton E, Todd C. Measurement of adherence to mHealth physical activity interventions and exploration of the factors that affect the adherence: scoping review and proposed framework. J Med Internet Res 2022 Jun 08;24(6):e30817 [FREE Full text] [doi: 10.2196/30817] [Medline: 35675111]
26. Short CE, DeSmet A, Woods C, Williams SL, Maher C, Middelweerd A, et al. Measuring engagement in eHealth and mHealth behavior change interventions: viewpoint of methodologies. J Med Internet Res 2018 Nov 16;20(11):e292 [FREE Full text] [doi: 10.2196/jmir.9397] [Medline: 30446482]
27. Duncan M, Vandelanotte C, Kolt GS, Rosenkranz RR, Caperchione CM, George ES, et al. Effectiveness of a web- and mobile phone-based intervention to promote physical activity and healthy eating in middle-aged males: randomized controlled trial of the ManUp study. J Med Internet Res 2014 Jun 12;16(6):e136 [FREE Full text] [doi: 10.2196/jmir.3107] [Medline: 24927299]
28. Donkin L, Hickie IB, Christensen H, Naismith SL, Neal B, Cockayne NL, et al. Rethinking the dose-response relationship between usage and outcome in an online intervention for depression: randomized controlled trial. J Med Internet Res 2013 Oct 17;15(10):e231 [FREE Full text] [doi: 10.2196/jmir.2771] [Medline: 24135213]
29. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
30. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
31. Bijkerk LE, Oenema A, Geschwind N, Spigt M. Measuring engagement with mental health and behavior change interventions: an integrative review of methods and instruments. Int J Behav Med 2023 Apr;30(2):155-166 [FREE Full text] [doi: 10.1007/s12529-022-10086-6] [Medline: 35578099]
32. Zyngier D. (Re)conceptualising student engagement: doing education not doing time. Teach Teach Educ 2008 Oct;24(7):1765-1776. [doi: 10.1016/j.tate.2007.09.004]
33. Doherty K, Doherty G. Engagement in HCI: conception, theory and measurement. ACM Comput Surv 2018 Nov 19;51(5):1-39. [doi: 10.1145/3234149]
34. Cho J, Xu T, Zimmermann-Niefield A, Voida S. Reflection in theory and reflection in practice: an exploration of the gaps in reflection support among personal informatics apps. In: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. 2022 Presented at: CHI '22; April 29-May 5, 2022; New Orleans, LA p. 1-23 URL: https://dl.acm.org/doi/abs/10.1145/3491102.3501991 [doi: 10.1145/3491102.3501991]
35. Mann K, Gordon J, MacLeod A. Reflection and reflective practice in health professions education: a systematic review. Adv Health Sci Educ Theory Pract 2009 Oct 23;14(4):595-621. [doi: 10.1007/s10459-007-9090-2] [Medline: 18034364]
36. Alt D, Raichel N, Naamati-Schneider L. Higher education students' reflective journal writing and lifelong learning skills: insights from an exploratory sequential study. Front Psychol 2021;12:707168 [FREE Full text] [doi: 10.3389/fpsyg.2021.707168] [Medline: 35069305]
37. Wong FK, Kember D, Chung LY, Yan LY. Assessing the level of student reflection from reflective journals. J Adv Nurs 1995 Jul 28;22(1):48-57. [doi: 10.1046/j.1365-2648.1995.22010048.x] [Medline: 7560535]
38. Plack MM, Driscoll M, Blissett S, McKenna R, Plack TP. A method for assessing reflective journal writing. J Allied Health 2005;34(4):199-208. [Medline: 16529182]

39. Bjerkvik LK, Hilli Y. Reflective writing in undergraduate clinical nursing education: a literature review. Nurse Educ Pract 2019 Feb;35:32-41. [doi: 10.1016/j.nepr.2018.11.013] [Medline: 30660960]
40. Fleck R, Fitzpatrick G. Reflecting on reflection: framing a design landscape. In: Proceedings of the 22nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction. 2010 Presented at: OZCHI '10; November 22-26, 2010; Brisbane, Australia p. 216-223 URL: https://dl.acm.org/doi/10.1145/1952222.1952269 [doi: 10.1145/1952222.1952269]
41. Shephard R. Godin leisure-time exercise questionnaire. Med Sci Sports Exerc 1997;29(Supplement):36-38. [doi: 10.1097/00005768-199706001-00009]
42. Amireault S, Godin G, Lacombe J, Sabiston CM. Validation of the Godin-Shephard leisure-time physical activity questionnaire classification coding system using accelerometer assessment among breast cancer survivors. J Cancer Surviv 2015 Sep;9(3):532-540. [doi: 10.1007/s11764-015-0430-6] [Medline: 25666749]
43. Vallerand JR, Rhodes RE, Walker GJ, Courneya KS. Correlates of meeting the combined and independent aerobic and strength exercise guidelines in hematologic cancer survivors. Int J Behav Nutr Phys Act 2017 Mar 28;14(1):44 [FREE Full text] [doi: 10.1186/s12966-017-0498-7] [Medline: 28351397]
44. Parker NH, Basen-Engquist K, Rubin ML, Li Y, Prakash L, Ngo-Huang A, et al. Factors influencing exercise following pancreatic tumor resection. Ann Surg Oncol 2021 Apr 04;28(4):2299-2309. [doi: 10.1245/s10434-020-09062-9] [Medline: 32886288]
45. Butryn ML, Arigo D, Raggio GA, Kaufman AI, Kerrigan SG, Forman EM. Measuring the ability to tolerate activity-related discomfort: initial validation of the Physical Activity Acceptance Questionnaire (PAAQ). J Phys Act Health 2015 May;12(5):717-716 [FREE Full text] [doi: 10.1123/jpah.2013-0338] [Medline: 25106049]
46. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
47. Dyment JE, O'Connell TS. Assessing the quality of reflection in student journals: a review of the research. Teach High Educ 2011 Feb;16(1):81-97. [doi: 10.1080/13562517.2010.507308]
48. Montaño D, Kasprzyk D. Theory of reasoned action, theory of planned behavior, and the integrated behavioral model. In: Glanz K, Rimer B, Viswanath K, editors. Health Behavior: Theory, Research, and Practice. 5th edition. Hoboken, NJ: Jossey-Bass; 2015:95-124.
49. Feil K, Fritsch J, Rhodes RE. The intention-behaviour gap in physical activity: a systematic review and meta-analysis of the action control framework. Br J Sports Med 2023 Oct;57(19):1265-1271. [doi: 10.1136/bjsports-2022-106640] [Medline: 37460164]
50. Rhodes RE, Dickau L. Moderators of the intention-behaviour relationship in the physical activity domain: a systematic review. Br J Sports Med 2013 Mar 25;47(4):215-225. [doi: 10.1136/bjsports-2011-090411] [Medline: 22278998]
51. Baumer EP, Katz SJ, Freeman JE, Adams P, Gonzales AL, Pollak JP, et al. Prescriptive persuasion and open-ended social awarenessxpanding the design space of mobile health. In: Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work. 2012 Presented at: CSCW '12; February 11-15, 2012; Washington, DC p. 475-484 URL: https://dl.acm.org/doi/10.1145/2145204.2145279 [doi: 10.1145/2145204.2145279]
52. Brynjarsdottir H, Håkansson M, Pierce JE, Baumer EP, DiSalvo CF, Sengers P. Sustainably unpersuaded: how persuasion narrows our vision of sustainability. In: Proceedings of the 2012 SIGCHI Conference on Human Factors in Computing Systems. 2012 Presented at: CHI '12; May 5-10, 2012; Austin, TX p. 947-956 URL: https://dl.acm.org/doi/10.1145/2207676.2208539 [doi: 10.1145/2207676.2208539]
53. Griggs V, Holden R, Lawless A, Rae J. From reflective learning to reflective practice: assessing transfer. Stud High Educ 2016 Oct 14;43(7):1172-1183. [doi: 10.1080/03075079.2016.1232382]
54. Brown T, McCracken M, O'Kane P. 'Don't forget to write': how reflective learning journals can help to facilitate, assess and evaluate training transfer. Hum Resour Dev Int 2011 Sep;14(4):465-481 [FREE Full text] [doi: 10.1080/13678868.2011.601595]
55. Schon DA. The Reflective Practitioner: How Professionals Think in Action. New York, NY: Basic Books; 1984.
56. Teixeira PJ, Carraça EV, Markland D, Silva MN, Ryan RM. Exercise, physical activity, and self-determination theory: a systematic review. Int J Behav Nutr Phys Act 2012 Jun 22;9:78 [FREE Full text] [doi: 10.1186/1479-5868-9-78] [Medline: 22726453]
57. Boud D. Using journal writing to enhance reflective practice. New Dir Adult Contin Educ 2002 Feb 26;2001(90):9-18 [FREE Full text] [doi: 10.1002/ace.16]
58. Stevens CJ, Baldwin AS, Bryan AD, Conner M, Rhodes RE, Williams DM. Affective determinants of physical activity: a conceptual framework and narrative review. Front Psychol 2020;11:568331 [FREE Full text] [doi: 10.3389/fpsyg.2020.568331] [Medline: 33335497]
59. Ritterband LM, Thorndike FP, Cox DJ, Kovatchev BP, Gonder-Frederick LA. A behavior change model for internet interventions. Ann Behav Med 2009 Aug;38(1):18-27 [FREE Full text] [doi: 10.1007/s12160-009-9133-4] [Medline: 19802647]

Abbreviations

ACT: acceptance and commitment therapy
BREQ-3: Behavioral Regulation for Exercise Questionnaire-3
DBCI: digital behavior change intervention
PAAQ: Physical Activity Acceptance Questionnaire
PROMIS: Patient-Reported Outcomes Measurement Information System
REDCap: Research Electronic Data Capture
SID: study identification number
SDT: self-determination theory

Edited by L Buis; submitted 20.07.23; peer-reviewed by N Crane, G Dolatkhah; comments to author 01.09.23; revised version received 21.11.23; accepted 07.12.23; published 09.02.24.

Please cite as:
Robertson MC, Cox-Martin E, Basen-Engquist K, Lyons EJ
Reflective Engagement With a Digital Physical Activity Intervention Among People Living With and Beyond Breast Cancer: Mixed Methods Study
JMIR Mhealth Uhealth 2024;12:e51057
URL: https://mhealth.jmir.org/2024/1/e51057
doi:10.2196/51057
PMID:38335025

Original Paper

Assessing the Efficacy of ChatGPT Versus Human Researchers in Identifying Relevant Studies on mHealth Interventions for Improving Medication Adherence in Patients With Ischemic Stroke When Conducting Systematic Reviews: Comparative Analysis

Suebsarn Ruksakulpiwat[1], RN, MMed, PhD; Lalipat Phianhasin[1], RN, MS, AGPCNP-BC; Chitchanok Benjasirisan[1], RN, MS; Kedong Ding[2], AM; Anuoluwapo Ajibade[3], BS; Ayanesh Kumar[4], MS; Cassie Stewart[5]

[1]Department of Medical Nursing, Faculty of Nursing, Mahidol University, Bangkok, Thailand

[2]Jack, Joseph and Morton Mandel School of Applied Social Sciences, Case Western Reserve University, Cleveland, OH, United States

[3]College of Art and Science, Department of Anthropology, Case Western Reserve University, Cleveland, OH, United States

[4]School of Medicine, Case Western Reserve University, Cleveland, OH, United States

[5]Frances Payne Bolton School of Nursing, Case Western Reserve University, Cleveland, OH, United States

Corresponding Author:
Suebsarn Ruksakulpiwat, RN, MMed, PhD
Department of Medical Nursing, Faculty of Nursing, Mahidol University
2 Wang Lang Road, Siriraj, Bangkok Noi
Bangkok, 10700
Thailand
Phone: 66 984782692
Email: suebsarn25@gmail.com

Related Article:

Comment in: https://mhealth.jmir.org/2024/1/e57978/

Abstract

Background: ChatGPT by OpenAI emerged as a potential tool for researchers, aiding in various aspects of research. One such application was the identification of relevant studies in systematic reviews. However, a comprehensive comparison of the efficacy of relevant study identification between human researchers and ChatGPT has not been conducted.

Objective: This study aims to compare the efficacy of ChatGPT and human researchers in identifying relevant studies on medication adherence improvement using mobile health interventions in patients with ischemic stroke during systematic reviews.

Methods: This study used the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. Four electronic databases, including CINAHL Plus with Full Text, Web of Science, PubMed, and MEDLINE, were searched to identify articles published from inception until 2023 using search terms based on MeSH (Medical Subject Headings) terms generated by human researchers versus ChatGPT. The authors independently screened the titles, abstracts, and full text of the studies identified through separate searches conducted by human researchers and ChatGPT. The comparison encompassed several aspects, including the ability to retrieve relevant studies, accuracy, efficiency, limitations, and challenges associated with each method.

Results: A total of 6 articles identified through search terms generated by human researchers were included in the final analysis, of which 4 (67%) reported improvements in medication adherence after the intervention. However, 33% (2/6) of the included studies did not clearly state whether medication adherence improved after the intervention. A total of 10 studies were included based on search terms generated by ChatGPT, of which 6 (60%) overlapped with studies identified by human researchers. Regarding the impact of mobile health interventions on medication adherence, most included studies (8/10, 80%) based on search terms generated by ChatGPT reported improvements in medication adherence after the intervention. However, 20% (2/10) of the studies did not clearly state whether medication adherence improved after the intervention. The precision in accurately identifying relevant studies was higher in human researchers (0.86) than in ChatGPT (0.77). This is consistent with the percentage of relevance, where human researchers (9.8%) demonstrated a higher percentage of relevance than ChatGPT (3%). However, when considering

the time required for both humans and ChatGPT to identify relevant studies, ChatGPT substantially outperformed human researchers as it took less time to identify relevant studies.

Conclusions: Our comparative analysis highlighted the strengths and limitations of both approaches. Ultimately, the choice between human researchers and ChatGPT depends on the specific requirements and objectives of each review, but the collaborative synergy of both approaches holds the potential to advance evidence-based research and decision-making in the health care field.

(JMIR Mhealth Uhealth 2024;12:e51526) doi:10.2196/51526

KEYWORDS

ChatGPT; systematic reviews; medication adherence; mobile health; mHealth; ischemic stroke; mobile phone

Introduction

Background

Artificial intelligence (AI) is the field of computer science that studies and develops systems that can perform tasks, typically requiring human intelligence, such as reasoning, learning, decision-making, natural language processing (NLP), computer vision, and speech recognition [1]. AI is a rapidly evolving field with applications in various domains, for example, health care, education, business, and entertainment [2]. One of the subfields of AI is NLP, which deals with analyzing and generating natural language texts [3]. Chatbots, a type of NLP system, can interact with humans using natural language, either through text or speech. Chatbots can be used for various purposes, including customer service, entertainment, education, and information retrieval [3]. However, developing chatbots that can engage in natural and coherent conversations with humans is a challenging task that requires advanced NLP techniques and large-scale data.

One of the recent advances in NLP is the development of GPT models, which are neural network models that can generate natural language texts based on a given input or context [4]. GPT models are trained on large corpora of text from various sources, such as books, websites, news articles, and social media posts [4]. GPT models have been used to create chatbots that can generate realistic and diverse responses to human queries or messages [4]. Although GPT models have been developed by various research groups and companies (ie, OpenAI, Google, Facebook, and Microsoft), the first one was introduced by OpenAI in 2019 [5]. Since then, ChatGPT has been improved and refined by researchers and developers, who have applied it to various tasks and scenarios, such as customer service, education, entertainment, and social media [5]. ChatGPT models aim to provide engaging, informative, and coherent dialogues with users across different domains and tasks [4].

ChatGPT has been applied in the medical field in various ways. For instance, in medical practice, it has the ability to help streamline the clinical workflow, enhance diagnostics, and predict disease risk and outcome [6]. For medical education, ChatGPT can be useful in tailoring education and enabling powerful self-learning [6]. In terms of medical research, a previous study reported that ChatGPT has the potential to advance understanding, identify new research questions, and improve data analysis and interpretation [7]. In addition, ChatGPT extends to involve in writing articles through improvement in language and communication of result findings [6]. In particular, in the literature review process, which is time and effort consuming, ChatGPT has a promising advantage because of its potential ability to analyze large amounts of data, particularly in scientific articles [8]. Furthermore, ChatGPT was reported to have the potential to generate effective Boolean queries for systematic review literature searches [9].

Although ChatGPT has several advantages in medical research, it has limitations that could impact the quality of research, particularly in the literature review and search strategies processes. Citation inaccuracies, insufficient references, and references to nonexistent sources were reported as current problems [6]. Moreover, ChatGPT has a limited knowledge period based on the data sets used in ChatGPT training, which limits the reliability of the updated source of the literature review [6]. In a previous study, researchers were advised to consider the potential for incorrect MeSH (Medical Subject Headings) terms and the varying effectiveness of search queries with multiple requests when devising search strategies for a systematic review [9]. However, ChatGPT has a high potential to be used in medical research in the future. Therefore, it is imperative to explore and develop to improve and use it effectively.

Despite the significant benefits and limitations of using ChatGPT, the evaluation of the quality and performance of ChatGPT models in the review process remains unclear. Therefore, this study aims to compare the efficacy of ChatGPT and human researchers in identifying relevant health-related studies, such as research on medication adherence improvement using mobile health (mHealth) interventions in patients with ischemic stroke. The review will use systematic methods to search, select, appraise, and synthesize to address the following questions: (1) How does ChatGPT's performance compare to that of human researchers in terms of accuracy in identifying relevant studies? (2) What challenges and limitations arise from using ChatGPT versus human researchers for identifying relevant studies in systematic reviews? (3) What are the implications of using ChatGPT to enhance the efficiency of systematic reviews? The results of this review will provide crucial insights into the potential of ChatGPT as an innovative tool for conducting systematic reviews.

Objectives

This study aims to compare the efficacy of using ChatGPT and human researchers in identifying relevant studies on medication adherence improvement using mHealth interventions in patients with ischemic stroke during systematic reviews.

Methods

Identify Relevant Studies

In this study, we used the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) [10] guidelines to identify the relevant studies. Overall, 4 electronic databases, including CINAHL Plus with Full Text, Web of Science, PubMed, and MEDLINE, were searched to identify articles published from inception until 2023 on using mHealth interventions for improving medication adherence in patients with ischemic stroke. We used search terms based on MeSH using Boolean phrases generated by human researchers and ChatGPT version 3.5 to identify relevant studies. The reference lists of the included studies, generated by human researchers and ChatGPT, were separately stored and screened in EndNote (EndNote X7 reference management software package). A PRISMA flow diagram was created to present the results of the search and screening process.

Study Selection

The authors independently screened the titles and abstracts of the studies identified through separate searches conducted by human researchers and ChatGPT to determine their relevance. Subsequently, the full text of the selected articles was also assessed to ensure they met the predetermined inclusion criteria. A consistent set of inclusion criteria was applied to ensure that only studies relevant to the review's objective were included. In contrast, the same exclusion criteria were used to eliminate literature unrelated to the review (Textbox 1).

Textbox 1. Study inclusion and exclusion criteria.

Inclusion criteria

- Studies that aimed to use mobile health interventions for improving medication adherence
- Studies that primarily included adults with ischemic stroke or transient ischemic attack (TIA) aged ≥18 years (if the study included other stroke types, such as hemorrhagic stroke, it is acceptable, but the study population must include adults with ischemic stroke or TIA)
- Studies in English
- Studies that were published from inception until 2023

Exclusion criteria

- Studies that included children or adolescents aged <18 years
- Conference proceedings, abstracts, review articles, protocols, dissertations, letters to the editor, brief reports, or statement papers
- Studies that involved animal samples

Data Extraction

A separate summary table for data extraction is presented in Multimedia Appendix 1 [11-20], consisting of the following data for each study: reference, year, country, study design, sample size, target population, intervention and objective, and main findings. This table will be used to compare the included studies obtained through the *Identify relevant studies* phase conducted by human researchers versus ChatGPT. The primary outcome of interest is medication adherence among patients with ischemic stroke. Medication adherence can be measured using various methods, such as drug level measurement, pill count, electronic databases, self-report questionnaires, and electronic monitoring systems [21]. The findings from studies that aimed to use mHealth interventions for improving medication adherence but did not measure medical adherence directly will be evaluated based on how they operationalized medication adherence according to their study design.

Data Analysis

In this study, we will assess the accuracy of both human researchers and ChatGPT in identifying relevant studies from electronic databases by measuring precision. Precision is a performance metric that measures the accuracy of a model's positive predictions. It focuses on the proportion of correctly identified positive instances (true positives) out of all the cases that the model predicted as positive (true positives+false positives) [22]. Precision is calculated using the following formula: precision=true positives/(true positives+false positives).

A high precision value close to 1 indicates that the model has a low rate of false positives. This means that when the model predicts an instance as positive, it will likely be correct. In contrast, a low precision value close to 0 indicates that the model has a high rate of false positives. This means that when the model predicts an instance as positive, it often needs to be corrected [22]. In the context of this study, precision will help evaluate the ability of both human researchers and ChatGPT to accurately identify relevant studies from electronic databases during the systematic review process. We will compare their precision scores to determine which approach yields a higher proportion of true positives and a lower rate of false positives.

In addition, as the human researcher will still need to conduct the screening, eligibility, and inclusion phases, we will also calculate the percentage of relevance using the formula ([true positives/total studies identified from the search]×100). This approach will be chosen to ensure a fair assessment, as relying solely on a formula based on true and false positives (precision) might only reflect human variability and accuracy during the screening, eligibility, and inclusion phases.

Ethical Considerations

This study considers nonhuman research according to the "Self-Assessment form whether an activity is human subject

research which requires ethical approval" recommended by Mahidol University Central Institutional Review Board. Therefore, ethics approval from the research ethics committee was not required.

Results

Search Term

Human Researcher

In the search phase, we used search terms based on MeSH using Boolean operators. The searched topic was related to using mHealth interventions for improving medication adherence in patients with ischemic stroke: (Ischemic Stroke* OR Cryptogenic Ischemic Stroke* OR Cryptogenic Stroke* OR Cryptogenic Embolism Stroke* OR Wake up Stroke* OR Acute Ischemic Stroke* OR Embolic Stroke* OR Cardioembolic Stroke* OR Cardioembolic Stroke* OR Thrombotic Stroke* OR Acute Thrombotic Stroke* OR Lacunar Stroke* OR Lacunar Syndrome* OR Lacunar Infarction* OR Lacunar Infarct*) AND (Medication Adherence OR Medication Nonadherence OR Medication Noncompliance OR Medication Persistence OR Medication Compliance OR Medication Non-Compliance) AND (Tele-Referral* OR Virtual Medicine OR Tele Intensive Care OR Tele ICU OR Mobile Health OR mHealth OR Telehealth OR eHealth OR Remote Consultation OR Teleconsultation* OR Telenursing OR Telepathology OR Teleradiology OR Telerehabilitation* OR Remote Rehabilitation* OR Virtual Rehabilitation*).

ChatGPT

To compare with the search by human researchers, we asked ChatGPT [23] on June 23, 2023, at 1:30 PM EST to provide a search term for conducting a systematic review of the same topic as follows: "Hello ChatGPT, we are researchers and currently conduct a systematic review titled: Using m-health interventions for improving medication adherence in ischemic stroke patients. Can you provide Medical Subject Headings (MeSH) search terms and combine them using Boolean operators for a search process?" The following search terms resulted from ChatGPT, which we used in the search phase and then compared the results with those from human researchers: (Mobile Applications OR Cell Phone OR Smartphone OR Telemedicine OR Text Messaging OR Internet) AND (Medication Adherence OR Patient Compliance OR Medication Systems, Intelligent) AND (Stroke OR Ischemic Attack, Transient OR Cerebrovascular Disorders). The search term (generated by human researchers and ChatGPT) was adjusted according to the database searching requirement before searching, but the original keyword was not changed.

Search Results

We compared the ability of humans and ChatGPT to retrieve all relevant studies. A higher recall indicates a better ability to capture all the relevant literature. Figure 1 shows the flowchart diagram of the selection of included studies based on search terms generated by human researchers. An initial literature search yielded 61 articles, including 30 from PubMed and MEDLINE, 21 from Web of Science, and 10 from CINAHL Plus Full Text. No additional records were found through other sources. After deduplication (n=7 studies), the researchers screened 54 studies, of which 47 (87%) were excluded based on the inclusion and exclusion criteria following the title and abstract screening phase. This left 7 articles for full-text screening, during which 1 article was excluded as it did not include any mHealth-related intervention. Therefore, 6 articles were included in the final analysis. It should be noted that human researchers conducted the identification, screening, eligibility, and inclusion phases.

Figure 1. The flowchart diagram displays the selection method of qualified studies searched by a human researcher. mHealth: mobile health.

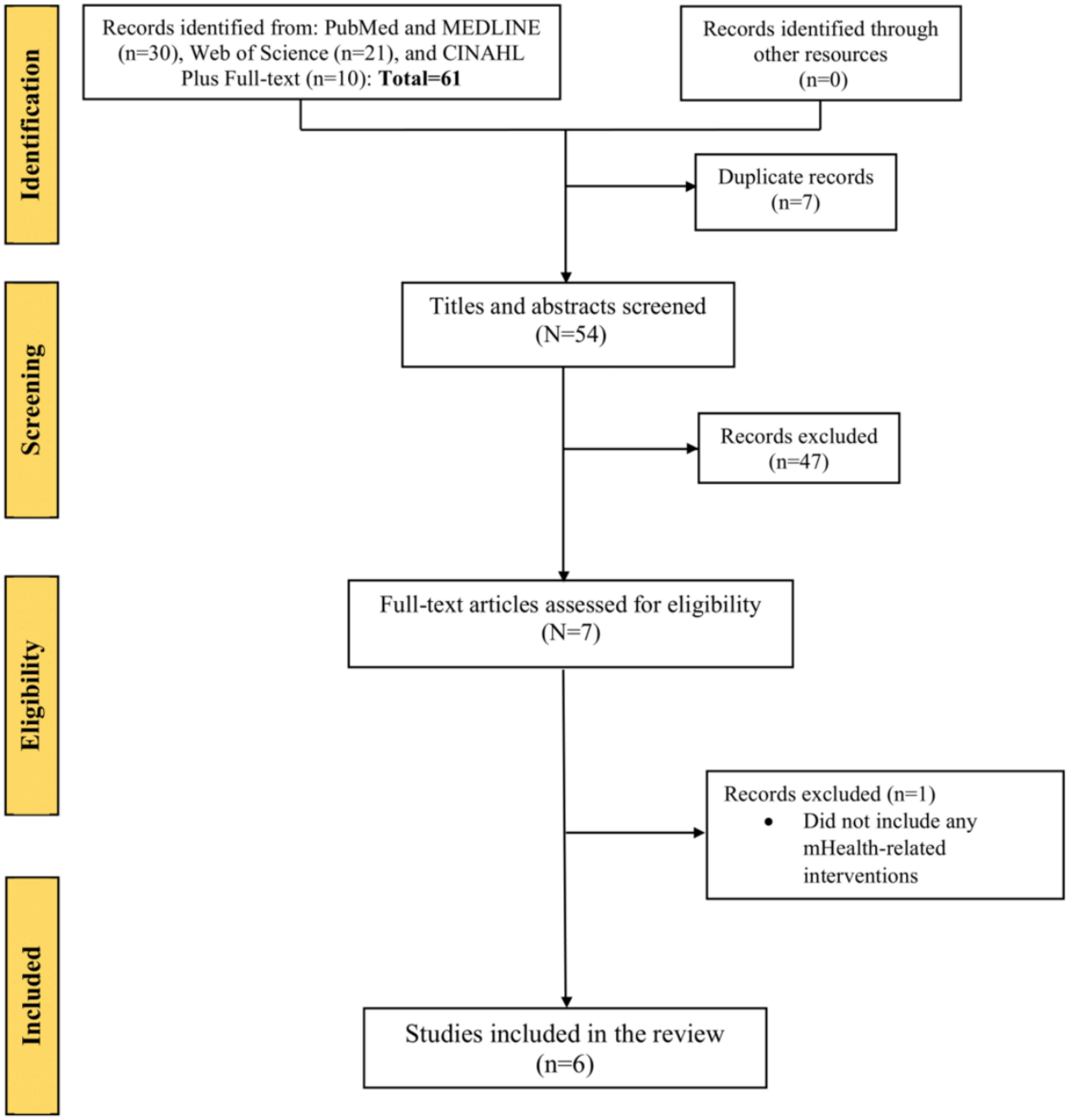
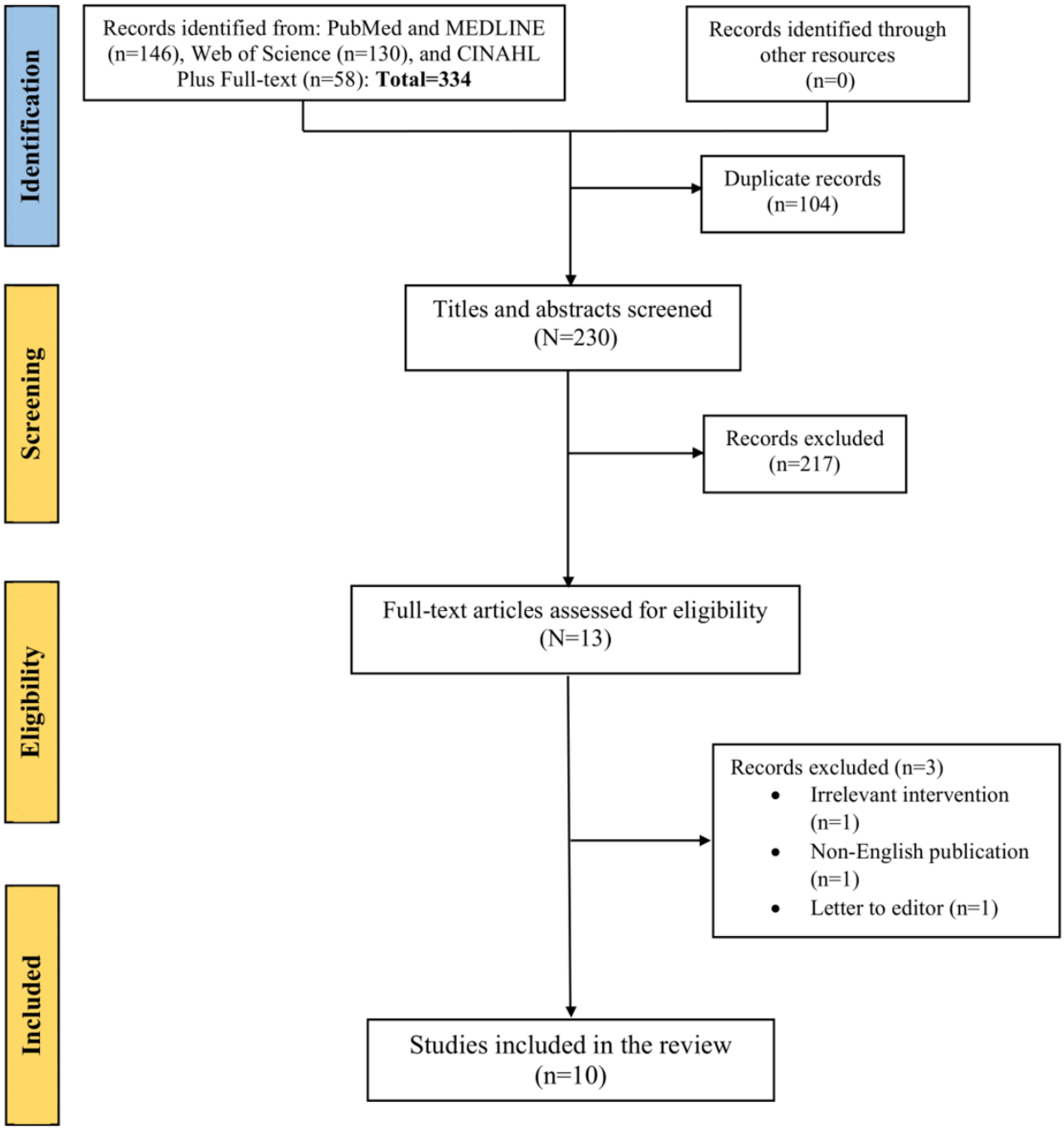


Figure 2 shows the flowchart diagram of the selection of included studies based on search terms generated by ChatGPT. An initial literature search yielded 334 articles, including 146 from PubMed and MEDLINE, 130 from Web of Science, and 58 from CINAHL Plus Full Text. No additional records were found through other sources. After deduplication (n=104 studies), the researchers screened 230 studies, of which 217 (94.3%) were excluded based on the inclusion and exclusion criteria following the title and abstract screening phase. Of the 13 articles that underwent full-text screening, 3 studies were excluded because the intervention was irrelevant (n=1, 33%), the publication was not in English (n=1, 33%), and it was a letter to the editor (n=1, 33%). Finally, 10 articles were included in the final analysis. It should be noted that ChatGPT has been used only in the identification phase. The human researcher conducted the screening, eligibility, and inclusion phases.

Figure 2. The flowchart diagram displays the selection method of qualified studies searched by ChatGPT.



Description of the Included Studies

Studies Included From Human Searches

The analysis included 6 studies obtained from the human search (Multimedia Appendix 1). Most of these studies (3/6, 50%) were published in 2020. Among the countries where the studies were conducted, 50% (3/6) were from China, whereas 17% (1/6) of the studies each originated from Belgium, the Republic of Korea, and Sweden. In terms of study design, of the 6 studies, 3 (50%) were cohort studies, 2 (33%) were randomized controlled trials (RCTs), and 1 (17%) was a non-RCT. The sample sizes varied, with 50% (3/6) of the studies having a sample size ranging from 1 to 300 and the other 50% (3/6) of the studies having a sample size of >300. Regarding the impact of mHealth interventions on medication adherence, most included studies (4/6, 67%) reported improvements in medication adherence after the intervention [11-14]. However, in 33% (2/6) of the included studies, it was not clearly stated whether medication adherence improved after the intervention [15,16].

Studies Included From ChatGPT Searches

A total of 10 studies were obtained from the ChatGPT search, of which 6 (60%) studies overlapped with the human searches (Multimedia Appendix 1). Most of these studies (4/10, 40%) were published in 2020. Among the countries where the studies were conducted, 50% (5/10) of the studies were from China, whereas 10% (1/10) of the studies each originated from Belgium, the Republic of Korea, Sweden, the United States, and Pakistan. In terms of study design, most were RCTs (6/10,

60%), with 30% (3/10) of the studies being cohort studies, and 10% (1/10) of the studies being a non-RCT. The sample sizes varied, with 70% (7/10) of the studies having a sample size ranging from 1 to 300 (70%), and the other 30% (3/10) of the studies having a sample size of >300. Regarding the impact of mHealth interventions on medication adherence, most included studies (8/10, 80%) reported improvements in medication adherence after the intervention [11-14,17-20]. However, in 20% (2/10) of the included studies, it was not clearly stated whether medication adherence improved after the intervention [15,16].

Accuracy

In our study, we used precision as a metric to assess the accuracy of both human researchers and ChatGPT in identifying relevant studies from electronic databases during the systematic review process. By comparing their precision scores, we aimed to determine which approach yielded a higher proportion of true positives (correctly identified relevant studies) and a lower rate of false positives (incorrectly identified irrelevant studies). The precision calculation formula used was as follows: precision=true positives/(true positives+false positives).

Moreover, the human researcher conducted identification, screening, eligibility, and inclusion phases, as illustrated in Figure 1. In contrast, ChatGPT was used only during the identification phase, and the human researcher conducted the screening, eligibility, and inclusion phases, as depicted in Figure 2. Therefore, we also calculated the percentage of relevance using the formula ([true positives/total studies identified from the search]×100). This approach was chosen to ensure a fair assessment, as relying solely on a formula based on true and false positives might only reflect human variability and accuracy during the screening, eligibility, and inclusion phases.

For human researchers, the precision in accurately identifying relevant studies from electronic databases was calculated as 6/(6+1)=0.86, where 6 is the number of studies included in the review (true positive) and 1 (false positive) represents the study that was incorrectly identified as relevant for inclusion in the review (did not include any mHealth-related intervention; Figure 1). This means that out of the studies deemed relevant by human researchers, 86% (6/7) were indeed appropriate for inclusion in the review, whereas 14% (1/7) were falsely identified as relevant. The percentage of relevance for the human researcher was calculated as follows: (true positives/total studies identified from the search)×100=(6/61)×100=9.8%.

Regarding ChatGPT, its precision in accurately identifying relevant studies from electronic databases was calculated as 10/(10+3)=0.77, where 10 is the number of studies included in the review (true positive) and 3 (false positive) represents the studies that were incorrectly identified as relevant for inclusion in the review (irrelevant intervention, non-English publication, and a letter to the editor; Figure 2). This indicates that out of the studies identified by ChatGPT as potentially relevant, 77% (10/13) were indeed relevant and suitable for inclusion in the review, whereas 23% (3/13) were mistakenly identified as relevant. The percentage of relevance for ChatGPT was calculated as follows: (true positives/total studies identified from the search)×100=(10/334)×100=3%.

According to our findings, the precision of human researchers was higher (precision=0.86) compared to ChatGPT (precision=0.77). This is consistent with the percentage of relevance, where human researchers (9.8%) demonstrated a higher percentage of relevance than ChatGPT (3%). These results indicate that human researchers were more effective in identifying relevant studies during the systematic review process. However, it is noteworthy that despite the lower precision and percentage of relevance, ChatGPT's initial search yielded a significantly larger number of studies (n=334) compared to human researchers (n=61), and ultimately resulted in more studies included in the final analysis (n=10 for ChatGPT vs n=6 for human researchers). This suggests that ChatGPT's performance was more efficient in terms of study retrieval and inclusion, although there was a 60% overlap in the studies included between both approaches.

Efficiency

As reported in the Accuracy subsection, human researchers demonstrated higher precision in identifying relevant studies compared to ChatGPT. However, the efficiency and ability of ChatGPT to retrieve relevant studies could still hold value in the systematic review process. When considering the time required for both humans and ChatGPT to identify relevant studies, from the beginning (search term generation) to the outcome (identification of relevant studies before screening), our study found that ChatGPT substantially outperformed human researchers. ChatGPT took approximately 10 minutes, whereas human researchers spent an hour in the search term identification process using MeSH and Boolean operators before obtaining the relevant study.

In our study, we used ChatGPT to generate search terms for conducting the systematic review based on our research topic. This substantially reduced the time and effort required for initial study identification. However, it is important to note that ChatGPT's current capabilities are limited to providing search terms, and human researchers are still required to conduct the screening of titles, abstracts, and full texts of the identified studies, using refined inclusion and exclusion criteria.

Discussion

Principal Findings

According to our findings, the precision of human researchers was higher compared to ChatGPT, indicating that human researchers were more accurate in identifying relevant studies during the systematic review process. Our findings are congruent with a previous study [24], which reports inaccuracies of using ChatGPT in research that requires an in-depth understanding of the literature. Likewise, Zhao et al [25] reported that the factual accuracy of ChatGPT cannot be ensured, although it has massive resources such as Microsoft and Google. In addition, a case study of using ChatGPT to conduct literature searches indicated that ChatGPT does not provide an answer to the queries that researchers ask for [26].

Despite the lower precision of ChatGPT compared to human search, a previous study reported that ChatGPT has more accurate and comprehensive relevance judgments than all other

types of NLP models or techniques [27]. Moreover, our findings show that ChatGPT's initial search yielded a significantly larger number of studies compared to human researchers and ultimately resulted in more studies being included in the final analysis despite its lower precision. This suggests that ChatGPT's performance was more efficient in terms of study retrieval and inclusion, although there was a 60% overlap in the studies included between both approaches. Similarly, a study of ChatGPT's insights on the future of scientific publishing reports it as a valuable resource for initiating discussions [28]. However, a previous study using ChatGPT for retrieval of clinical, radiological information reported that ChatGPT provided only two-thirds of correct responses to questions [29].

Regarding the efficiency issues of using ChatGPT in identifying relevant search terms, the results of this study suggest that ChatGPT can be a useful tool for generating search terms for systematic reviews, as it can save time and effort for human researchers and potentially retrieve more relevant studies. The previous study on the use of ChatGPT Boolean query construction and refinement for systematic review showed that ChatGPT can generate queries with high precision [9]. Therefore, ChatGPT could be a valuable tool, especially for rapid reviews where time is limited and high precision is preferred over high recall [9].

Some researchers may argue that as ChatGPT has lower precision and may generate irrelevant or inaccurate terms, human researchers still need to carefully screen the studies that ChatGPT identified and verify the quality and validity of the evidence [30]. ChatGPT should be used with caution and verification and supplemented with other methods and sources to ensure the validity and rigor of the literature search [9]. Furthermore, ChatGPT's performance may vary depending on the research topic, data availability, and input quality. Thus, future studies are needed to evaluate ChatGPT's generalizability and reliability across different domains and contexts.

Using ChatGPT to generate search terms for systematic reviews raises some ethical questions regarding the quality and validity of the research process. Although ChatGPT may offer some advantages in terms of efficiency and comprehensiveness, it may also introduce some biases and errors that could affect the reliability and reproducibility of the systematic reviews. For example, ChatGPT may generate search terms that are irrelevant to the research topic or too broad or narrow, resulting in either missing or including studies that do not meet the inclusion criteria [31]. Moreover, ChatGPT may generate search terms that are based on its own internal knowledge and information, which may not reflect the current state of the art or the best available evidence in the field [31]. Therefore, human researchers need to carefully evaluate and validate the search terms generated by ChatGPT and document their rationale and methods for using them. In addition, human researchers need to disclose the use of ChatGPT as a tool for generating search terms and report its strengths and limitations and any potential ethical implications in their systematic review reports [31]. This would ensure that the systematic review process is transparent, accountable, and trustworthy and that the results are credible and useful for informing decision-making.

As we embark on a comparative analysis between ChatGPT and human researchers in the pursuit of identifying relevant studies within systematic reviews, particularly focused on mHealth interventions for improving medication adherence in patients with ischemic stroke, it becomes evident that several challenges and limitations underscore the intricate nature of this exploration. These challenges offer insight into the complex interplay between cutting-edge technology and the established domain expertise of human researchers, shaping the landscape in which this study unfolds.

First and foremost, the outcomes of our study are intrinsically linked to the performance of ChatGPT, an AI-driven tool that relies on its current capabilities to generate search terms. As an entity in constant evolution, ChatGPT's performance may undergo shifts over time, potentially influencing the accuracy and efficiency with which it generates relevant search terms. Moreover, replicating the search in subsequent studies is essential due to ChatGPT's intrinsic unpredictability. The lack of such repetition presents challenges in determining whether the observed phenomenon reflects an inherent trait of the model or is simply a random incident.

This dynamic underscores the need to interpret our findings in the context of the tool's state during the study period. Within the realm of medical research, the intricate and evolving nature of terminology poses a formidable challenge. Although ChatGPT exhibits language generation prowess, the intricate nuances of medical terminology—constantly adapting and expanding—could potentially pose challenges to its accurate formulation of search terms. The complexity inherent to medical concepts demands a level of contextual understanding that might be challenging for an AI system.

Another pivotal consideration revolves around the potential biases embedded within ChatGPT's training data. Drawing insights from vast data sets, ChatGPT-generated search terms might inadvertently inherit biases present in the underlying data sources. This potential bias, albeit unintentional, introduces an element of caution when relying solely on AI-generated search terms for systematic reviews. A crucial aspect of our study's execution pertains to refining search terms. Although ChatGPT serves as a catalyst for initial search term generation, human researchers play a pivotal role in the subsequent validation and fine-tuning of these terms. This collaborative process introduces an additional layer of complexity, as human intervention becomes essential to ensure the relevance and accuracy of the generated search terms. Moreover, the resources available and the access to ChatGPT's capabilities could introduce variability in the study's outcomes. Depending on factors such as subscription tiers or institutional resources, the extent of ChatGPT's contributions and, subsequently, its comparative assessment against human researchers may exhibit nuances that warrant consideration. The study's defined scope, focused on mHealth interventions for medication adherence improvement in patients with ischemic stroke, provides a specific lens through which insights are garnered. However, this specificity inherently limits the direct transposability of findings to other medical domains or broader systematic review topics. The nuances of different research contexts might yield distinct results. Language and geographic considerations further amplify the complexity.

The study predominantly engaged with studies in English, potentially omitting valuable research published in other languages or regions. This limitation underscores the need for meticulous attention to language diversity and inclusion in systematic reviews. Human researcher variability introduces a layer of subjectivity into the study. With multiple researchers contributing to search term generation, variations in expertise and individual approaches could impact the study's outcomes. The potential for differing interpretations and formulations of search terms necessitates careful management. Publication bias, a well-known challenge in research, extends its influence into our study's design. Both ChatGPT and human researchers might inadvertently be swayed by publication bias, where certain types of studies are more likely to be published, potentially influencing the pool of studies considered in this review.

External factors beyond the purview of our study could exert unanticipated influence. Variables such as changes in database availability, updates to search algorithms, or shifts in the research landscape might subtly shape the study's design and outcomes, introducing an element of unpredictability. The study's designated time frame for data collection and inclusion introduces potential time constraints and selection bias. Studies published after the search period might be inadvertently omitted, potentially impacting the completeness of the review. Although the study provides valuable insights within its specific scope, the generalizability of findings to other systematic review topics or research questions requires cautious interpretation. The intricate interplay between technology and human expertise forms the cornerstone of our study, emphasizing the necessity for a balanced and nuanced approach when leveraging ChatGPT for systematic reviews.

The Implications of Using ChatGPT to Improve the Efficiency of Systematic Reviews

The integration of ChatGPT into the systematic review process for identifying relevant studies on mHealth interventions holds several noteworthy implications for research methodology, efficiency, and the advancement of evidence-based practices. This section explores the key implications that arise from incorporating ChatGPT as a tool to expedite and enhance the systematic review process.

One of the most immediate and impactful implications of using ChatGPT is its ability to significantly expedite the systematic review process. Traditionally, the generation of search terms for identifying relevant articles is a time-intensive task that requires meticulous crafting and refinement by human researchers. ChatGPT's capacity to swiftly generate search terms offers an innovative solution to this bottleneck, reducing the time invested in this preliminary phase. This acceleration holds the potential to expedite the overall timeline of systematic reviews, enabling researchers to allocate more time to critical appraisal, synthesis, and analysis of selected studies.

The inherent nature of ChatGPT's language generation capabilities allows a more diverse and expansive range of search terms. By tapping into its capacity to comprehend and generate natural language, researchers can explore a broader spectrum of keyword variations and synonyms. This expanded search scope can lead to the inclusion of studies that might have been overlooked using traditional search methods. As a result, the systematic review process becomes more comprehensive, encompassing a wider array of relevant literature.

ChatGPT's ability to generate novel and contextually relevant search terms introduces a valuable avenue for exploratory research and hypothesis generation. Researchers can leverage ChatGPT to identify emerging trends, novel terminologies, or unconventional associations that may inform the direction of their systematic reviews. This capacity to extract insights from the vast expanse of existing literature can potentially lead to the formulation of innovative research questions and avenues for investigation.

Although ChatGPT demonstrates remarkable efficiency in generating search terms, its use necessitates a collaborative approach with human researchers. The synergy between ChatGPT's speed and human researchers' expertise in refining and validating search terms ensures a balanced and accurate outcome. Human researchers play a pivotal role in critically evaluating the generated search terms, refining them to align with the specific objectives of the review, and subsequently verifying the relevance of the identified articles. This collaborative interplay mitigates the risk of introducing erroneous or irrelevant studies into the review process.

In research environments with limited resources, such as time and personnel, ChatGPT offers a solution to address scalability challenges. Its ability to rapidly generate search terms can prove invaluable in scenarios where timely completion of systematic reviews is imperative. Researchers operating within resource-constrained contexts can leverage ChatGPT to conduct preliminary searches efficiently, thus optimizing the allocation of limited resources to subsequent stages of the review.

In summary, the integration of ChatGPT into the systematic review process introduces a transformative approach to enhancing efficiency and enriching the scope of literature exploration. Although its speed and breadth of search terms hold the promise of expediting the review timeline and uncovering hidden associations, the collaborative involvement of human researchers remains pivotal for ensuring accuracy, relevance, and the meticulous execution of subsequent review stages. The strategic use of ChatGPT in conjunction with traditional research practices paves the way for a new era of evidence synthesis and knowledge advancement in the field of health care interventions.

Conclusions

Our study compares the accuracy and efficacy of human researchers and ChatGPT in providing search terms to identify articles during a systematic review on mHealth interventions for improving medication adherence in patients with ischemic stroke. Although human researchers achieved greater precision, ChatGPT's search results exhibited lower accuracy. However, ChatGPT excelled in efficacy, taking less time to generate search terms compared to human researchers, who required more time to identify appropriate search terms. In addition, ChatGPT's search yielded a higher number of articles compared to human researchers. Following exclusions, human researchers were left with 6 articles, and ChatGPT resulted in 10 articles after

screening, 6 (60%) of which overlapped with the findings of human researchers. The use of ChatGPT in creating search terms can significantly accelerate the systematic review process, although human researchers are still essential to carry out the selection process and ensure accuracy.

Acknowledgments

The authors would like to express their gratitude to Mahidol University, Thailand, for covering the article publishing charge. During the preparation of this work, the authors used ChatGPT (version 3.5, 2024; OpenAI [23]) to improve language and readability, which the study group further reviewed and revised.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Summary table.
[DOCX File , 32 KB - mhealth_v12i1e51526_app1.docx]

References

1. Chassignol M, Khoroshavin A, Klimova A, Bilyatdinova A. Artificial Intelligence trends in education: a narrative overview. Procedia Comput Sci 2018;136:16-24. [doi: 10.1016/j.procs.2018.08.233]
2. Bohr A, Memarzadeh K. The rise of artificial intelligence in healthcare applications. In: Bohr A, Memarzadeh K, editors. Artificial Intelligence in healthcarelsevier. Washington, DC: Academic Press; 2020:25-60.
3. Hapke H, Howard C, Lane H. Natural Language Processing in Action: Understanding, Analyzing, and Generating Text With Python. New York, NY: Manning Publications; 2019.
4. Lund BD, Wang T, Mannuru NR, Nie B, Shimray S, Wang Z. ChatGPT and a new academic reality: artificial intelligence-written research papers and the ethics of the large language models in scholarly publishing. J Assoc Inf Sci Technol 2023 Mar 10;74(5):570-581 [FREE Full text] [doi: 10.1002/asi.24750]
5. Why tech insiders are so excited about ChatGPT, a chatbot that answers questions and writes essays. CNBC. 2022. URL: https://www.cnbc.com/2022/12/13/chatgpt-is-a-new-ai-chatbot-that-can-answer-questions-and-write-essays.html [accessed 2024-04-12]
6. Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. Healthcare (Basel) 2023 Mar 19;11(6):887 [FREE Full text] [doi: 10.3390/healthcare11060887] [Medline: 36981544]
7. Ruksakulpiwat S, Kumar A, Ajibade A. Using ChatGPT in medical research: current status and future directions. J Multidiscip Healthc 2023;16:1513-1520 [FREE Full text] [doi: 10.2147/JMDH.S413470] [Medline: 37274428]
8. Dahmen J, Kayaalp ME, Ollivier M, Pareek A, Hirschmann MT, Karlsson J, et al. Artificial intelligence bot ChatGPT in medical research: the potential game changer as a double-edged sword. Knee Surg Sports Traumatol Arthrosc 2023 Apr 21;31(4):1187-1189. [doi: 10.1007/s00167-023-07355-6] [Medline: 36809511]
9. Wang S, Scells H, Koopman B, Zuccon G. Can chatgpt write a good boolean query for systematic review literature search? arXiv preprint arXiv. arXiv. Preprint posted online February 3, 2023 2023. [doi: 10.1145/3539618.3591703]
10. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
11. Kim DY, Kwon H, Nam KW, Lee Y, Kwon HM, Chung YS. Remote management of poststroke patients with a smartphone-based management system integrated in clinical care: prospective, nonrandomized, interventional study. J Med Internet Res 2020 Feb 27;22(2):e15377 [FREE Full text] [doi: 10.2196/15377] [Medline: 32130140]
12. Li DM, Lu XY, Yang PF, Zheng J, Hu HH, Zhou Y, et al. Coordinated patient care via mobile phone-based telemedicine in secondary stroke prevention: a propensity score-matched cohort study. J Nurs Care Qual 2023;38(3):E42-E49. [doi: 10.1097/NCQ.0000000000000693] [Medline: 36827597]
13. Yan LL, Gong E, Gu W, Turner EL, Gallis JA, Zhou Y, et al. Effectiveness of a primary care-based integrated mobile health intervention for stroke management in rural China (SINEMA): a cluster-randomized controlled trial. PLoS Med 2021 Apr 28;18(4):e1003582 [FREE Full text] [doi: 10.1371/journal.pmed.1003582] [Medline: 33909607]
14. Zhang Y, Fan D, Ji H, Qiao S, Li X. Treatment adherence and secondary prevention of ischemic stroke among discharged patients using mobile phone- and WeChat-based improvement services: cohort study. JMIR Mhealth Uhealth 2020 Apr 15;8(4):e16496 [FREE Full text] [doi: 10.2196/16496] [Medline: 32293574]

15. Kamoen O, Maqueda V, Yperzeele L, Pottel H, Cras P, Vanhooren G, et al. Stroke coach: a pilot study of a personal digital coaching program for patients after ischemic stroke. Acta Neurol Belg 2020 Feb 07;120(1):91-97. [doi: 10.1007/s13760-019-01218-z] [Medline: 31701472]
16. Ögren J, Irewall A, Söderström L, Mooe T. Long-term, telephone-based follow-up after stroke and TIA improves risk factors: 36-month results from the randomized controlled NAILED stroke risk factor trial. BMC Neurol 2018 Sep 21;18(1):153 [FREE Full text] [doi: 10.1186/s12883-018-1158-5] [Medline: 30241499]
17. Labovitz DL, Shafner L, Reyes Gil M, Virmani D, Hanina A. Using artificial intelligence to reduce the risk of nonadherence in patients on anticoagulation therapy. Stroke 2017 May;48(5):1416-1419 [FREE Full text] [doi: 10.1161/STROKEAHA.116.016281] [Medline: 28386037]
18. Kamal AK, Shaikh Q, Pasha O, Azam I, Islam M, Memon AA, et al. A randomized controlled behavioral intervention trial to improve medication adherence in adult stroke patients with prescription tailored short messaging service (SMS)-SMS4Stroke study. BMC Neurol 2015 Oct 21;15:212 [FREE Full text] [doi: 10.1186/s12883-015-0471-5] [Medline: 26486857]
19. Wang MY, Shen MJ, Wan LH, Mo MM, Wu Z, Li LL, et al. Effects of a comprehensive reminder system based on the health belief model for patients who have had a stroke on health behaviors, blood pressure, disability, and recurrence from baseline to 6 months: a randomized controlled trial. J Cardiovasc Nurs 2020;35(2):156-164. [doi: 10.1097/JCN.0000000000000631] [Medline: 31904693]
20. Wan L, Zhang X, You L, Ruan H, Chen S. The efficacy of a comprehensive reminder system to improve health behaviors and blood pressure control in hypertensive Ischemic stroke patients: a randomized controlled trial. J Cardiovasc Nurs 2018;33(6):509-517. [doi: 10.1097/JCN.0000000000000496] [Medline: 29901484]
21. Anghel LA, Farcas AM, Oprean RN. An overview of the common methods used to measure treatment adherence. Med Pharm Rep 2019 Apr 22;92(2):117-122 [FREE Full text] [doi: 10.15386/mpr-1201] [Medline: 31086837]
22. Powers DM. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. arXiv. Preprint posted online October 11, 2020 2020 [FREE Full text]
23. ChatGPT (3.5). OpenAI. URL: https://chat.openai.com [accessed 2024-04-05]
24. van Dis EA, Bollen J, Zuidema W, van Rooij R, Bockting CL. ChatGPT: five priorities for research. Nature 2023 Feb 03;614(7947):224-226. [doi: 10.1038/d41586-023-00288-7] [Medline: 36737653]
25. Zhao R, Li X, Chia YK, Ding B, Bing L. Can ChatGPT-like generative models guarantee factual accuracy? On the mistakes of new generation search engines. arXiv. Preprint posted online March 3, 2023 2023 [FREE Full text]
26. McGee RW. Using ChatGPT to conduct literature searches: a case study. J Bus Ethics 2023;95(2):165-178. [doi: 10.13140/RG.2.2.18784.97282]
27. Sun W, Yan L, Ma X, Ren P, Yin D, Ren Z. Is ChatGPT good at search? Investigating large language models as re-ranking agents. arXiv. Preprint posted online April 19, 2023 2023 [FREE Full text] [doi: 10.18653/v1/2023.emnlp-main.923]
28. Hill-Yardin EL, Hutchinson MR, Laycock R, Spencer SJ. A Chat(GPT) about the future of scientific publishing. Brain Behav Immun 2023 May;110:152-154. [doi: 10.1016/j.bbi.2023.02.022] [Medline: 36868432]
29. Wagner MW, Ertl-Wagner BB. Accuracy of information and references using ChatGPT-3 for retrieval of clinical radiological information. Can Assoc Radiol J 2024 Feb 20;75(1):69-73 [FREE Full text] [doi: 10.1177/08465371231171125] [Medline: 37078489]
30. Bozkurt A, Karadeniz A, Baneres D, Guerrero-Roldán AE, Rodríguez ME. Artificial intelligence and reflections from educational landscape: a review of ai studies in half a century. Sustainability 2021 Jan 15;13(2):800. [doi: 10.3390/su13020800]
31. Khlaif ZN. Ethical concerns about using AI-generated text in scientific research. SSRN Journal. Preprint posted online March 20, 2023 2023 [FREE Full text] [doi: 10.2139/ssrn.4387984]

Abbreviations

AI: artificial intelligence
MeSH: Medical Subject Headings
mHealth: mobile health
NLP: natural language processing
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial

Edited by L Buis; submitted 17.08.23; peer-reviewed by S Tran, L Zhu; comments to author 10.02.24; revised version received 11.02.24; accepted 27.03.24; published 06.05.24.

Please cite as:
Ruksakulpiwat S, Phianhasin L, Benjasirisan C, Ding K, Ajibade A, Kumar A, Stewart C
Assessing the Efficacy of ChatGPT Versus Human Researchers in Identifying Relevant Studies on mHealth Interventions for Improving Medication Adherence in Patients With Ischemic Stroke When Conducting Systematic Reviews: Comparative Analysis
JMIR Mhealth Uhealth 2024;12:e51526
URL: https://mhealth.jmir.org/2024/1/e51526
doi:10.2196/51526
PMID:38710069

Original Paper

Remote Inclusion of Vulnerable Users in mHealth Intervention Design: Retrospective Case Analysis

Ingjerd J Straand[1], MSc; Kimberley A Baxter[2,3], PhD; Asbjørn Følstad[4], PhD

[1]Department of Social Work, University of Stavanger, Stavanger, Norway

[2]School of Exercise and Nutrition Sciences, Queensland University of Technology, Brisbane, Australia

[3]Centre for Childhood Nutrition Research, Faculty of Health, Queensland University of Technology, Brisbane, Australia

[4]Department of Sustainable Communication Technologies, Sintef Digital, Oslo, Norway

Corresponding Author:
Ingjerd J Straand, MSc
Department of Social Work
University of Stavanger
Kjell Arholms hus
Stavanger, 4021
Norway
Phone: 47 93222289
Email: ingjerd.j.straand@uis.no

Abstract

Background: Mobile health (mHealth) interventions that promote healthy behaviors or mindsets are a promising avenue to reach vulnerable or at-risk groups. In designing such mHealth interventions, authentic representation of intended participants is essential. The COVID-19 pandemic served as a catalyst for innovation in remote user-centered research methods. The capability of such research methods to effectively engage with vulnerable participants requires inquiry into practice to determine the suitability and appropriateness of these methods.

Objective: In this study, we aimed to explore opportunities and considerations that emerged from involving vulnerable user groups remotely when designing mHealth interventions. Implications and recommendations are presented for researchers and practitioners conducting remote user-centered research with vulnerable populations.

Methods: Remote user-centered research practices from 2 projects involving vulnerable populations in Norway and Australia were examined retrospectively using visual mapping and a reflection-on-action approach. The projects engaged low-income and unemployed groups during the COVID-19 pandemic in user-based evaluation and testing of interactive, web-based mHealth interventions.

Results: Opportunities and considerations were identified as (1) reduced barriers to research inclusion; (2) digital literacy transition; (3) contextualized insights: a window into people's lives; (4) seamless enactment of roles; and (5) increased flexibility for researchers and participants.

Conclusions: Our findings support the capability and suitability of remote user methods to engage with users from vulnerable groups. Remote methods facilitate recruitment, ease the burden of research participation, level out power imbalances, and provide a rich and relevant environment for user-centered evaluation of mHealth interventions. There is a potential for a much more agile research practice. Future research should consider the privacy impacts of increased access to participants' environment via webcams and screen share and how technology mediates participants' action in terms of privacy. The development of support procedures and tools for remote testing of mHealth apps with user participants will be crucial to capitalize on efficiency gains and better protect participants' privacy.

(JMIR Mhealth Uhealth 2024;12:e55548) doi:10.2196/55548

KEYWORDS

user testing; user participation in research; COVID-19; remote testing; intervention design; mobile phone

Introduction

Mobile health (mHealth) interventions, which use mobile technology such as smartphone apps to promote healthy behaviors or mindsets [1], are a promising avenue to reach vulnerable groups [2]. Meaningful user involvement is critical for such interventions [3] to ensure that end user needs and perspectives are adequately represented in the design process [4]. Conducting such feedback and evaluations with users face to face (local testing) [5-7] involves efficiency drawbacks, particularly travel, time, and cost [8]. Researchers and practitioners have thus experimented with remote testing and research [9,10] using both specialized tools (eg, UserTesting and Lookback) and videoconferencing (eg, Zoom, Hangout, and Teams). Studies comparing local and remote research practices have concluded comparable results in the quality of the research output [11]. However, before the COVID-19 pandemic, local testing was the usual practice in research and among practitioners [4,6]. Reasons may include network variance, poor audio or video quality, unfamiliarity with remote technology, and the lack of contextual information or nonverbal cues inherent in remote methods. Local testing, by contrast, removes users from the intended context of use; this is significant for user involvement in the design of mobile solutions such as mHealth interventions.

Traditional research methods tend to involve users from high socioeconomic backgrounds, who are easy to reach and have the means to participate, including resources of time, transport, and social support [12]. Human-computer interaction research calls for adequate reach and engagement with the people affected by the design to ensure an alignment of needs and, ultimately, an effective program [13,14]. This can be challenging when working with community groups who are marginalized or experience social disadvantage, such as racial or ethnic minority groups, individuals who have low income and who are unemployed, people with disabilities [15], or those with gender or sexual diversity [14,16]. This risks diminishing the validity of the findings to the target population and reduces the authenticity of engagement. While mHealth interventions may be particularly relevant for these groups, the suitability of remote practices for user involvement should be explored. More evidence is needed to support the appropriateness and effectiveness of remote user-centered research methods when engaging with vulnerable participants.

Accelerated by the pandemic, remote research and participation tools have become more available and ready-to-hand [17]. Researchers expedited the incorporation of remote methods that allowed for project continuity, highlighting the research community's resilience and researchers' and participants' willingness to experiment with technology. A recent meta-analysis found that one of the most significant effects of the pandemic on user involvement in design was shifting to web-based platforms [11]. At the community level, the increased use of telehealth services across populations to provide continuity of health care and education [18] has increased familiarity and comfort with videoconferencing and other web-based tools.

While the COVID-19 pandemic was a catalyst for innovation and creativity in remote user design methods, now that the pandemic has resolved [19], the opportunity to learn and adopt effective remote methods remains. Conducting meta-research to capture these experiences is important for future research applications. Some examples of such research exist: Hill et al [20] reviewed practical approaches for remote user testing in older adults. Other researchers have compared findings between remote and nonremote methods [21] or discussed specific aspects of the testing, including moderator and observer roles [5,22]. However, few studies have detailed the implementation of user-centered design in mHealth [3] or reflected on the researcher and participant experiences [23] in intervention design targeting vulnerable or diverse population groups.

Thus, the research question addressed by this qualitative and retrospective study is as follows: what are the opportunities and considerations emerging from involving vulnerable user groups remotely in mHealth intervention design? This study will highlight what was learned by adapting to agile remote user involvement during COVID-19 to inform future applications of such involvement with vulnerable user groups. Research practices from 2 projects, which applied remote inclusion of vulnerable population groups to designing and developing mHealth interventions within child health (parental feeding) and social psychology (mindset), were used as cases.

This study is structured as follows: an overview of the research projects and the methodology of this study is provided, followed by case descriptions and lessons learned before the analytical findings and implications are presented.

Methods

Research Context

The 2 research projects in this study used human-centered design (HCD) methodology [24] to design and develop web-based mHealth interventions targeting vulnerable populations. The project aims were to create digital health interventions collaboratively with and for end users and then evaluate these as part of ongoing research. The Responsive Feeding in Tough Times (RFiTT) project in Australia aimed to develop and evaluate a parenting program to promote responsive feeding practices in parents with young children in low-income families. The Career Learning App (CL-APP) project in Norway aimed to design, develop, and test positive psychology intervention apps targeting unemployed adolescents and young adults to promote job-seeking mindset and behaviors. These projects from different contexts have shared characteristics, including transdisciplinary work across design and health and applying an HCD process where users' ideas and feedback were central to the final intervention designs. Both project outcomes were web-based interventions designed for self-administered use on users' mobile phones, and remote user testing was applied with research participants from vulnerable groups.

In the 2 projects, the respective authors (IJS and KAB) developed user-centered design approaches, which were predominantly formative user-based evaluation [4,6] in the form of qualitative, moderated early testing [25] and feedback on

intervention prototypes. This included conducting the posttest analysis of the collected data. From March 2020 to December 2022, a total of 38 sessions were conducted across the 2 research projects. Participants were recruited intentionally with the characteristics of potential end users of the interventions to include their input into the designed outcome.

The projects' remote user engagement timing aligned with different phases of the COVID-19 pandemic. The Norway project experienced acute disruption during user testing (March 2020-April 2021), coinciding with the initial COVID-19 response. In contrast, the Australian project conducted user testing (November 2022-December 2022) during a more stable "living with" COVID-19 phase.

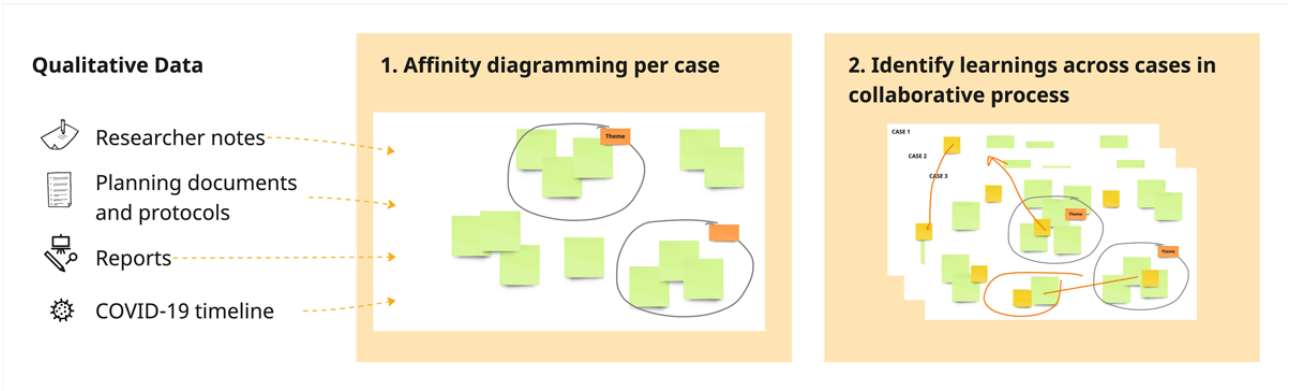
Research Design

Given the unprecedented COVID-19 pandemic during user testing, the retrospective reflection-on-action approach [26,27] was selected to explore the remote research setting. This study's research question and topic were explored [28,29] through "reflecting on action" [26]. Reflection was both internal and in dialogue between the authors and fellow researchers. This approach enabled researchers to reflect on the cases after the upheaval period of the pandemic had receded to uncover knowledge through analyzing and integrating experiences and practices.

We conducted a descriptive and retrospective examination of the research practices and experiences across the 2 projects. This was done through an iterative process using visual mapping (ie, affinity mapping or KJ-method) to sort findings visually [30] in Miro [31]. Affinity mapping builds upon abductive thinking and is commonly used by user experience practitioners [32,33]. This method was selected because of our heterogeneous data set [32] and the need to synthesize ideas from unstructured data. Our data included multiple sources: protocol documents, user test setups and documentation, researcher notes and reflections, postanalysis reports, and photos and screenshots from recordings.

We took a constructivist approach to our analysis, where synthesis and connections are formed through the researchers' critical reflection, and learnings are identified through active engagement and "discussions with the data" [34,35]. Our analysis was conducted stepwise (Figure 1), where we first added our data to the diagrams and started making clusters and groupings of findings relevant to the research question and labeling these on a case-by-case basis. Second, we identified learnings across cases in a collaborative process by regrouping our initial categories and findings of interest into broader categories or constructed themes [34] that represent the opportunities and considerations from different cases.

Figure 1. Stepwise analysis process based on visual diagramming, such as affinity mapping.



Participants

The participants included in this study are considered potentially "vulnerable" due to socioeconomic factors such as unemployment, low income, and economic hardship. *Vulnerability* is viewed as an inclusive term in line with the study by Culén and van der Velden [36], assuming that all users may be "vulnerable" at some point. The 2 research projects had different participant groups and ethical considerations; therefore, we describe them separately below. Table 1 summarizes participants across projects.

Table 1. Description of participants across cases: demography, recruitment channel, and format of the user testing.

	Participants
Case 1	
Target group (inclusion criteria)	• Unemployed or dropped out of school • Aged 18-29 y • Living in Rogaland, Norway • Speaking Norwegian
Recruitment channel	• Invitation via NAV[a] or IPS[b] program • Self-signup on website or via SMS text messaging or mail
Participants	• 12 participants aged 18-27 y; 7 females, 5 males[c]; 2 from ethnic minority groups (immigrant or BIPOC[d])
Format user test	• 3 in-person sessions • 9 remote of which 3 Discord, 5 Zoom, 1 other (Whereby); 2 used mobile phone device to connect to Zoom
Case 2	
Target group (inclusion criteria)	• Unemployed or dropped out of school • Aged 18-29 y • Living in Rogaland, Norway • Speaking Norwegian
Recruitment channel	• Invitation via NAV or IPS program • Self-signup on website or via SMS text messaging or email
Participants	• 13 participants • Aged 18-29 y; 6 female, 7 male[c]; 1 ethnic minority group (immigrant or BIPOC)
Format user test	• 0 in person (not possible) • 13 remote of which 12 Zoom, 1 other (Teams), and 1 used a mobile device to connect to Zoom but switched to computer during session
Case 3	
Target group (inclusion criteria)	• Parent or caregiver of a child aged 6 mo-3 y • Aged >18 y • Self-reported economic hardship
Recruitment channel	• Expression of interest list
Number of participants	• 12 participants • Average age 30 (range 26-36) y; 10 female, 2 male; 9 Australian, 1 Aboriginal Torres Strait Islander, 1 Indian, 1 Cambodian
Format user test	• 12 Zoom • 10 used mobile phone devices to connect to Zoom

[a]NAV: Norwegian Labour and Welfare Administration.

[b]IPS: Individual Placement and Support program.

[c]On the basis of observation, not self-reported.

[d]BIPOC: Black, indigenous, and people of color.

Ethical Considerations

Norway: Participants, Ethical Considerations, and Approval

Participants recruited to the Norwegian project were 18 to 29 years old and either registered as unemployed at the Norwegian Labour and Welfare Administration (NAV) or participating in a regional Individual Placement and Support program. Furthermore, they needed to speak Norwegian because of the in-app language. All participants provided explicit and written consent to participate in the study and were compensated for their time with a gift card of US $30 per session. The study was evaluated and approved by the Norwegian Centre for Research Data (approval number 131074).

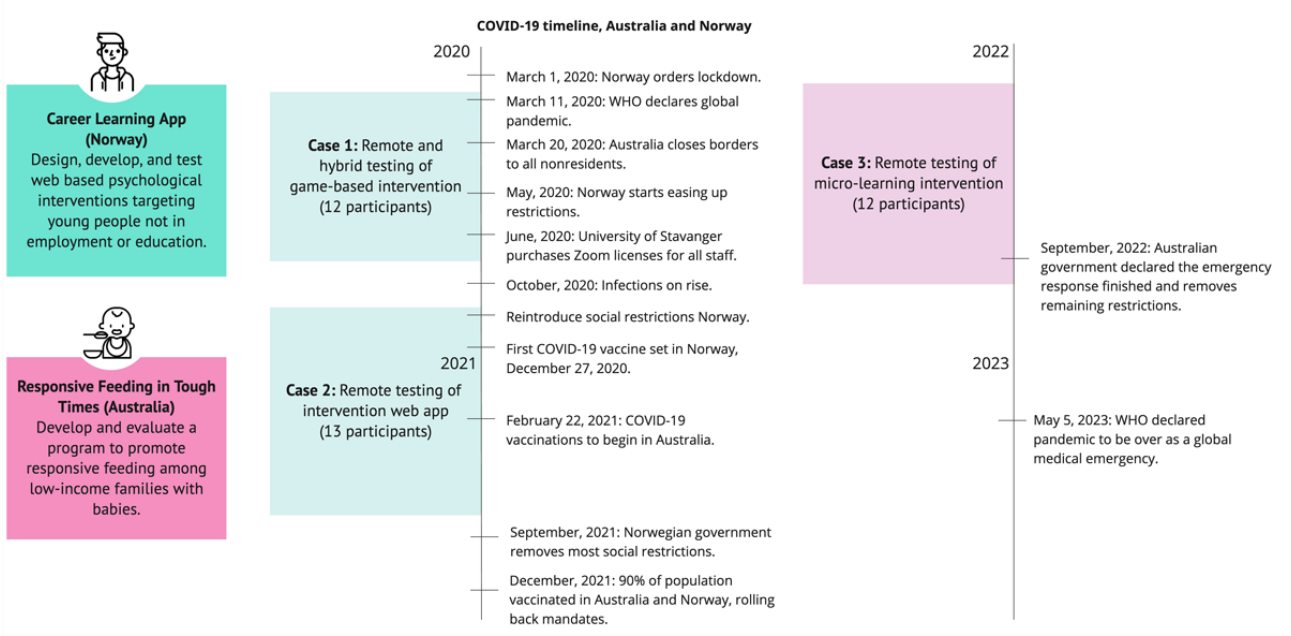
Australia: Participants, Ethical Considerations, and Approval

Participants were recruited Australia-wide and self-identified as experiencing economic hardship during screening. All participants were aged >18 years and caregivers of a child between 6 months and 3 years of age. Individuals were recruited from a pool of potential participants who had previously taken part in a web-based survey and had expressed interest in being contacted about other research activities. Participants were given an electronic gift voucher worth US $18 to thank them for their time. The Children's Health Queensland Hospital and Health Service Human Research Ethics Committee (LNR/21/QCHQ/72314) and the Queensland University of Technology Human Research Ethics Committee (2021000193) approved the study.

Case Descriptions

This section outlines the 2 research projects and details the user involvement protocols. The Norwegian project included 2 instances of user involvement; the Australian project involved 1. Hence, 3 cases are presented across the 2 research projects (Figure 2). Each case is divided into *case description*, *pandemic restrictions*, *test setup*, *participants*, and *case-specific reflections.*

Figure 2. Overview of user involvement conducted in the 2 research projects during the pandemic.



Case 1: Remote and Hybrid Testing of a Game-Based Intervention Concept, Norway

Case Description

The CL-APP project explored an interactive gaming concept to make a positive psychology intervention more engaging and relevant to unemployed young adults. The intervention design explored a 3D-based game. The development work was planned and executed in 3 sprints, with end users involved in formative user testing toward the end of each sprint. Further elaboration of the game concept and user feedback can be found in the study by Straand et al [37], with screenshots provided in Figure 3.

Figure 3. Screenshots from 3D game intervention VitaNova.

Pandemic Situation

On March 12, 2020, the Norwegian government ordered a nationwide lockdown, closing nonessential workplaces, schools, and child care centers. Schools and child care reopened with reduced hours for younger children toward the end of the following month. In May 2020, social and mobility restrictions were eased or replaced with mask mandates and sanitation requirements. However, in September 2020 and October 2020, infections again peaked, and in late October, new restrictions were announced, returning nonessential workers such as university staff to home offices.

Participants

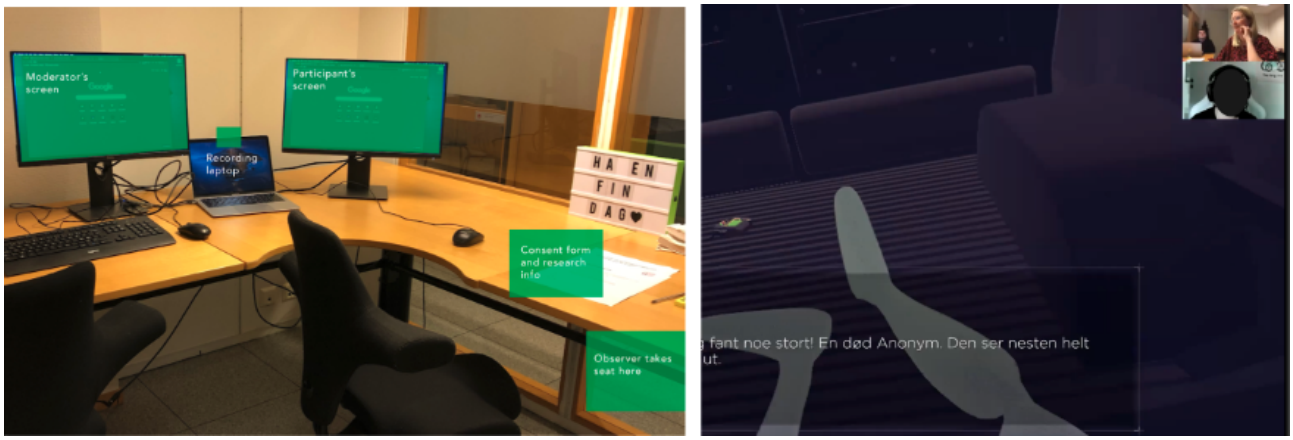
A total of 12 participants (female participants: 7/12, 58%; male participants: 5/12, 42%), aged 18 to 27 years, participated in the study.

Test Setup

In mid-March 2020, amidst pandemic uncertainty and lacking established remote protocols, plans for moderated, in-person usability testing were improvised. Discord, chosen for its familiarity among young gamers, served as the platform for remote testing. The test setup involved several manual operations due to the lack of functionality in Discord, including scheduling, consent, and provision of gift cards. The moderator and observer met a few minutes before and then added the participant to a group call once the participant had logged in to Discord. Despite its suitability for gamers and developers, approximately half of the users encountered startup issues due to unfamiliarity with the software. External software (Apple QuickTime) was used for recording, and this lack of a built-in recorder led to missing audio for some sessions. Platforms designed for usability testing or videoconferencing were rejected at the time from the premise of introducing complexity for the team and the participant users for a relatively short time of need.

As the team transitioned toward testing a functional prototype, concerns over network variance and load time prompted plans for face-to-face testing once restrictions eased in May 2020. A single participant signed up who had been involved in early-stage interviews. The test was conducted with strict sanitation and social distancing. However, with only 1 participant, it had limited value. A subsequent round of testing was planned for October 2020, when COVID-19 restrictions were expected to ease. This time, participants self-enrolled via a website and received SMS text messaging confirmation and reminders. We set up a testing space within the NAV offices. The team adapted its research strategy to allow participants to choose between in-person and remote testing on Zoom on the enrollment website. Remote participants signed digital consent forms and received digital gift cards, while in-person participants completed forms upon arrival and received physical gift cards (see Figure 4 for the hybrid test setup). The team completed tests with 7 participants, with the majority (5/7, 70%) opting for Zoom sessions.

Figure 4. Hybrid test setup: a total of 2 participants used local testing (left), while the remaining users opted for Zoom meetings (right).



Case-Specific Reflections

This case involved improvisation to enable continuity of the research, both with software and tools and with testing procedures. This iteration allowed us to observe how the videoconferencing software impacted the interaction with the participant, creating a new setting for the interaction depending on the software used. In the first rounds of testing using Discord, we all had our camera off. Discord users mostly use illustrations or avatars for their profile pictures and audio-only calls. Thus, although it is possible to share a camera view, none of the sessions using Discord had the participants with camera on; this included our webcams as researchers in the role of moderator and observers. The sessions done via videoconferencing software always had the camera-on mode for the moderator and nearly always for the participants, offering a richer data set for later analysis.

The "hybrid" strategy toward the end of the study meant the moderator and observer were usually in the same room, calling in as 1 user on Zoom. After the first session, it became the established practice for the moderator and observer to join in as individual users; the observer would mute the camera and microphone after a brief introduction at the beginning. This improved the interaction of the session, as the participant did not have to address 2 people. This remote setup allowed the observer to "disappear" into the background, overcoming the issue with the silent notetaker in a face-to-face session.

Some tasks were more challenging to deliver in the remote setup since the test tasks were designed for in-person sessions rather than remote participation. For instance, idea cards were created

that participants could sort according to their preferences. When the testing was on the web, we had to send them a copy of the cards in PDF format; this made the task less engaging and cumbersome. After preliminary user-derived findings, the development of the gaming-based intervention app presented in case 1 was discontinued.

Case 2: Remote Testing of mHealth Intervention Web App Concept, Norway

Case Description

Building on case 1, the CL-APP project redirected the design and development process to a mobile phone web app based on user preferences. The intervention target was foremost to promote a "growth mindset." A growth mindset [38-40] encourages a different interpretation of challenges faced by the young unemployed, normalizing struggles and setbacks to offer a more positive and flexible view of one's intelligence and ability to learn new things. The key objective of engaging with end users was to explore users' motivation to enhance reach and adherence. HCD methods ensured that the intervention was relevant, user-friendly, and motivating (see Figure 5 for screenshots of the app). In this process, researchers collaborated with designers, developers, and stakeholders, including end users, from October 2020 until the app's completion in December 2022. The user testing took place between November 2020 and April 2021.

Figure 5. Intervention web app concept shaped as learning modules.

Pandemic Situation

On October 26, 2020, the Norwegian government announced new health restrictions to reduce social interaction at work and home, strongly recommending that people return to home offices where possible [41].

Participants

Participants were recruited for 4 rounds of testing. A total of 13 participants (female participants: 6/13, 46%; male participants: 7/13, 54%) aged 18 to 29 years participated in the study.

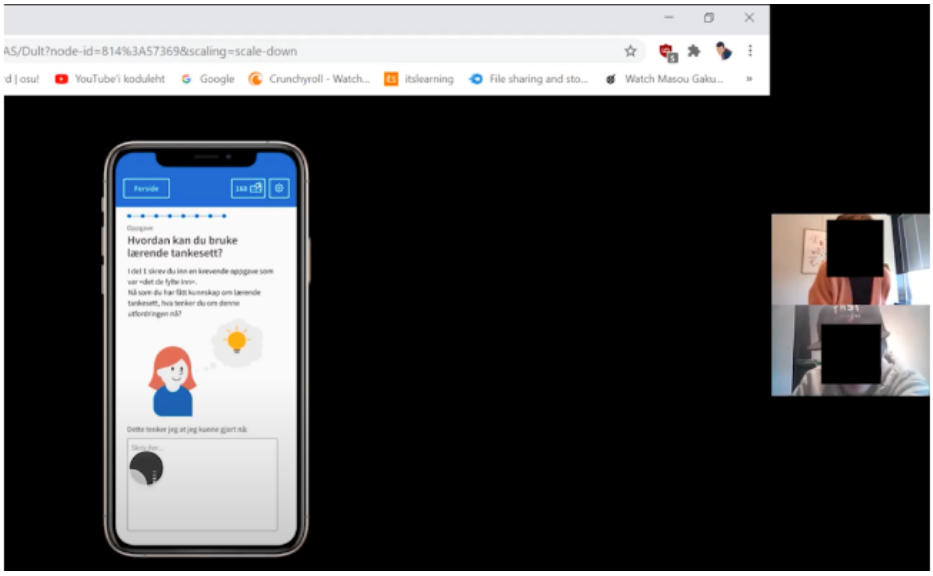
Test Setup

Given the work-from-home directive at the time, the design process, including interaction with end users, was planned remotely via Zoom. During the design process, 4 rounds of testing were performed: the initial test to understand what should be altered in the existing intervention (November: 4 participants) and 3 instances to get feedback on new designs with increasing levels of fidelity as the design progressed (January: 5 participants, February: 2 participants, and March-April: 4 participants). Prototypes were tested using the design tool Figma. The sessions were completed at times that were suitable to the participant. There was 1 session in the evening, but most participants opted for midday sessions (around 11 AM-2 PM).

Case-Specific Reflections

Our main challenge was that the prototype was designed for mobile use, and screen sharing from devices was troublesome in Zoom. Thus, for most of the tests on the new designs, we relied on desktop use and screen sharing from the browser (Figure 6). When a participant dialed in from their phone, the prototype view became unreadable, and we had to ask the participant to switch over to a device with a larger screen. All participants and moderators had their cameras switched on (unprompted). After an initial round of introductions, we continued switching the camera off for the observer or notetaker to reduce their presence in the user-researcher interaction.

Figure 6. Remote test setup in Zoom. The moderator has the camera on; the observer has the camera and microphone muted, therefore hidden from view.



Case 3: Remote Testing of an mHealth Microlearning Concept−Australia

Case Description

The RFiTT research program aimed to develop and evaluate an intervention to promote optimal child-feeding practices among low-income families. The secondary aims were to determine the feasibility, satisfaction, and acceptability of the mode of delivery. Families experiencing socioeconomic disadvantages face challenges feeding their children and following optimal feeding guidelines. The early years are crucial for establishing optimal feeding practices among parents and developing healthy child eating behaviors [42]. Therefore, the target of the intervention was parents or caregivers of children aged 6 to 24 months.

An mHealth digital microlearning concept was developed in response to parent engagement during the project's development phase [43]. Project constraints dictated a technology platform that required no software engineering or development phase and could be generated within a 4- to 6-month time frame. Web-based no-code technologies were researched and piloted to determine a suitable platform.

A learning technology platform (7taps), which used microlearning education, was selected. This platform enabled researchers to create contents that included videos, images, text, and interactions without external input from software engineers or app developers. This platform had a mobile-first design and learning management capability where modules could be delivered with preset timing in customized SMS text messages. Functional prototypes could be created and tested with users using this platform with little to no moderation. A total of 3 test modules that would form part of a microlearning responsive feeding parenting intervention were created (Figure 7).

Figure 7. Eat, Learn, Grow intervention: examples of digital module content and SMS text messaging delivery system.

Pandemic Situation

The RFiTT research program commenced in April 2020. The first case of COVID-19 was confirmed in Australia on January 25, 2020 [43]. On March 18, 2020, the federal government declared a biosecurity emergency, and all Australian States and Territories subsequently implemented lockdown measures [44]. Australia only fully opened its international borders to visitors in February 2022.

During the data collection and engagement phase of RFiTT (2021-2022), recruitment was impacted by the COVID-19 pandemic, and face-to-face data collection attempts were challenging. These recruitment challenges led to experimentation with remote research methods (telephone, web-based survey, noncontact equipment drop-offs) for research activities. By the time of the user testing sessions (November to December 2022), the RFiTT research program had adopted a complete remote research methods approach, and the scope of the population target for the intervention had shifted from a local context (Brisbane, Queensland) to Australia-wide.

Participants

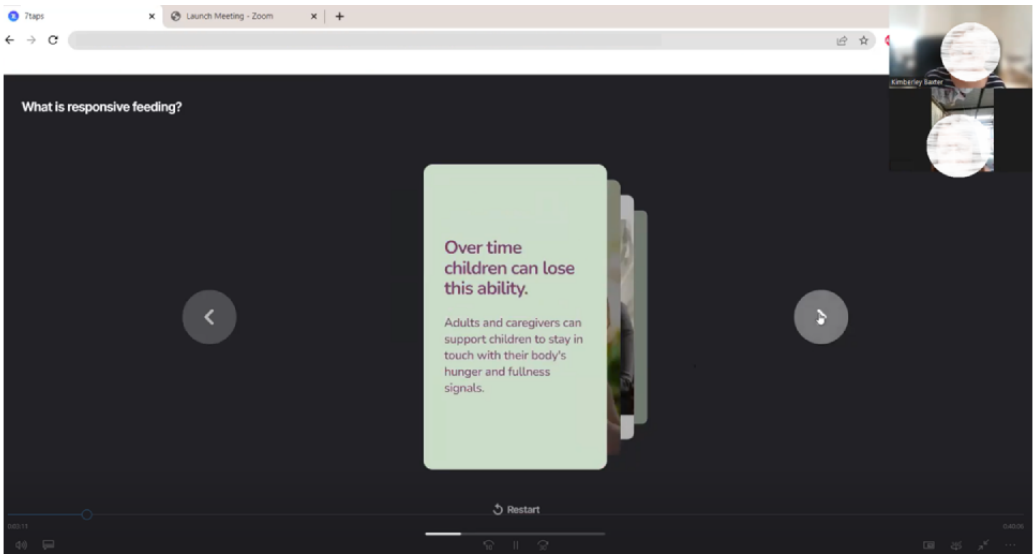
A total of 12 participants tested the prototypes. Of the group, 42% (5/12) had a university degree, and 3 individuals expressed that they had neurodiversity, which impacted their ability to learn and process information (attention-deficit/hyperactivity disorder, dyslexia, and aphantasia). Further details are available in Table 1.

Test Setup

Potential participants were telephoned to invite them to participate in the user testing sessions. Interested participants were sent a digital web link to the Participant Information Statement, a web-based consent form, and a short demographic survey. The web-based form and survey were hosted on REDCap (Research Electronic Data Capture; Vanderbilt University), a secure web application for building and managing web-based surveys [45]. The aim of the testing was twofold: (1) to test the acceptability, readability, and accessibility of 3 examples of microlearning content and (2) to co-design aspects of the content and structure of the intervention.

The sessions were completed at times suitable to the participant, including out-of-hour sessions from November 7, 2022, to December 1, 2022. Most participants joined the session using their mobile phones (10/12, 83%). A total of 3 modules were designed to present different styles of videos, content, and imagery to elicit feedback on the different formats and parents' preferences. The module web links were sent via mobile phone SMS text messaging to participants on the day of the arranged session. Parents viewed the content unmoderated. A Zoom session with the lead researcher (KAB) was arranged on the same day to capture parents' impressions and feedback. All Zoom sessions were video and audio recorded. During the sessions, the researcher shared a preview screen of the digital modules and guided the parent through a talk-out-loud walkthrough of the content (Figure 8). Open-ended questions regarding the usability, accessibility, and satisfaction of the modules were asked. Perspectives from parents were sought on recruitment and retention strategies, the language of key intervention messages, structure, and program timing. The parent intervention was renamed to "*Eat, Learn, Grow*" to reflect parents' feedback.

Figure 8. Screenshot of a remote user testing session conducted on Zoom.



Case-Specific Reflections

Remote inclusion of participants allowed for representation across Australia and of employed parents, who indicated that they would not have been able to participate if the session had been in person. The Zoom platform was effective; no participants had difficulty downloading or using the software. Most of the group (10/12, 83%) used a mobile phone. The mobile phone screen size restricted viewing the content via screen share (see Figure 8). However, it was acceptable, as participants had just engaged with the content. There were minimal connectivity or audio difficulties, but given the participants' home environment, there were interruptions from young children being supervised during sessions. These disruptions did not reduce the effectiveness of the sessions and are a common occurrence in research with parents, where young

children need to accompany parents. The researcher (KAB) is experienced with children and conducting research with families.

The most significant downside noted for the remote sessions in this case was that the parents engaged with the digital modules unmoderated; therefore, researchers did not observe parents interacting with the content for the first time. In the remote user setup, moderated sessions of parents viewing the content on their mobile device were not possible, given that the device needed to be used for videoconference for the feedback session. Moderating the session may have provided helpful information about parents' responses to the digital content. However, given that the platform used (7taps) was a purpose-built learning technology designed for first-time users, this was not the aim of the user testing sessions. This was mitigated by conducting the follow-up Zoom session the same day the digital content was sent to parents.

Results

Overview

We identified opportunities and considerations of conducting remote research with vulnerable users by reflecting on action and through visual diagramming across cases. These are represented as *reduced barriers to research inclusion, digital literacy transition, contextualized insights: a window into people's lives, seamless enactment of roles*, and *increased flexibility for researchers and participants.*

Reduced Barriers to Research Inclusion

During the user testing, the need to quickly design procedures for remote and hybrid research was necessitated by the evolving COVID-19 pandemic. Across the 3 cases, we found that remote research methods effectively engaged the targeted population groups of unemployed young adults and parents experiencing economic hardship. Enabling participants to participate in their home environment removed some systemic barriers to engaging in traditional research. For both groups, there were barriers to meeting face to face beyond the practicalities of travel, time, and capacity. Young people and parent participants displayed increased comfort with digital technologies and remote interactions, facilitating their participation in these research programs.

Remote and agile research methods enabled a broader and more diverse participant pool. RFiTT (case 3) widened the recruitment pool to Australia-wide rather than a small geographical area focus. In the Norway project, the recruitment pool was not widened geographically. However, remote methods enabled continued research during the acute response phase of COVID-19. Toward the end of our testing of the gaming concept (case 1), in-person participation was planned since restrictions had been lifted. However, recruitment was difficult, and participants who did consent failed to attend booked appointments despite a monetary incentive. Through this recruitment period, the population group expressed a high concern about the pandemic to researchers. This experience was confirmed in discussions with stakeholders such as welfare administration staff. Shifting to remote testing via web-conferencing (Zoom) facilitated continued participation.

Remote methods mitigated accessibility barriers and eased participants' potential fear, whether related to the pandemic or the unknown of being involved in a research project. Furthermore, many tests were conducted in the evening to adapt to the needs of parent participants (case 3). Across our populations, catering for continued remote participation was relevant even after restrictions were relaxed and was demonstrated by participants' strong preference for remote methods.

Digital Literacy Transition

Initially, the tools used for remote research were improvised, and methodological planning took an iterative approach. As the pandemic unfolded, users and researchers gained experience with relevant digital technology, reaching greater technology awareness and control. The different time frames in which the case studies were conducted during the COVID-19 pandemic provided a context to explore this trend of what we may refer to as a transition to digital literacy.

Initially, researchers and project stakeholders were reluctant to transition to remote participation (cases 1 and 2), whereas users seemed to prefer remote modes. The preferences of researchers and project stakeholders partly grew out of a desire to conduct the research "as planned" and to use established methods. There was also uncertainty about whether users had the necessary skills to use videoconferencing. Researchers had concerns about the limited opportunity for rapport building through informal conversation before the session started. However, the research team underestimated how digitally literate the participants were. This is unsurprising given the amount of time spent on the internet and the degree of web-based communication and collaboration in both groups across many aspects of life [46,47]. This was coupled with COVID-19 pandemic–driven increases in the use of technology for communication and services, such as telehealth [48-51] and work-from-home needs [52,53].

Contextualized Insights: A Window Into People's Lives

Despite our target participants' familiarity with web-based communication, the rapid adoption of these technologies also required sensitivity in protecting participants' privacy. Contrary to our perception of poorer conversations with the loss of face-to-face conversation, we experienced *more* entry into users' lives than participation at a research site. The recording was done easily as a part of the natural flow of conversation with the participant on the web in Zoom. In contrast, introducing video recording devices into physical meetings is cumbersome and can make people uncomfortable. Furthermore, it was found that the type of software used either increased or reduced the likelihood of data sharing due to its internal logics, customs, or *vibe* [54]. With Discord, it is not customary to use a real profile photo; in most instances, people use an avatar, and it did not feel natural to turn the webcam on. Therefore, this channel collected much less personal information than Zoom. Zoom encourages turning webcams on and recording seamlessly and unobtrusively. The tools used for supporting the research, such as Discord and Zoom and systems for issuing electronic gift cards, required collecting more personal data (such as name, email, phone number, and usernames) than in-person research methods.

Web-based and remote methods were a more natural and relevant environment for the user, revealing more contextual information than expected and providing a temporal window into people's lives. Sometimes, this may include unintended information, such as username, browsing history, or open tabs when participants were screen sharing. The less professional nature of the Zoom session also meant some participants were less formal. In one instance, a participant wore a bath robe, while others had babies crying in the background, pets, or others who entered the conversation. This provided a richer contextual backdrop to who the participants were and sparked informal conversation and trust building. At the same time, this contextualized information from the user tests does introduce privacy concerns.

Seamless Enactment of Roles

It is sometimes necessary to have observers during user testing. For face-to-face sessions, 2-way mirrors or screencasting to another location may be used to enable observation. Additional observers may also be needed in a physical space to take notes; this can be disruptive. The user may feel uncomfortable talking to 2 people, not knowing who to look at when talking and when someone is writing intensively. Remote user sessions may require fewer observers, and they may be less intrusive when they are present.

In the Norway cases, the observer's role as a notetaker was improved by videoconferencing. The observer and interviewer would have the camera on for the start of the testing. Then, after introductions, the observer could mute the camera and microphone and continue taking notes without impacting the session. If the observer wanted to ask follow-up questions, it was easy and natural to either bring the observer back into the conversation or allow the observer to post questions via a chat channel for the interviewer to follow up. With this more silent observer role, there was little disturbance to the flow of the conversation. It was easy and natural to switch roles during the session, which was done in case 2, where the author (IJS) moderated most of the session, and one of the designers ran through the prototype with the participants. In the Australian case, no person other than the author (KAB) was present for the testing.

Increased Flexibility for Researchers and Participants

Remote-only testing was found to be more streamlined and flexible compared with both in-person and hybrid models. Research participation, which is planned to be hybrid (case 1), requires booking and setting up the room. This introduces limitations on the remote research imposed by the physical meetings, such as the timeline and availability of physical space.

Remote testing (cases 2 and 3) allowed for more flexibility; meetings could be conducted in the evenings or during weekends or holidays to accommodate participants, with minor disruption to researchers who could dial in from home but with great benefits to participants. Without the booking and timeline constraints of physical space, sessions could take place over time (case 2), allowing revision and adaptation of design prototypes that could be tested again. This maximized the data collection capacity of the sessions and led to a more agile approach to our research and engagement with participants.

For RFiTT (case 3), most participants (10/12, 83%) engaged with the remote user testing sessions via mobile phone. Participants did not have access to a working computer, and using a mobile device enabled participants to perform essential tasks such as supervising young children. This suggests that flexibility and convenience to do other things may contribute to the preference for remote participation. For CL-APP (cases 1 and 2), nearly all participants connected to the remote testing sessions on their computers (23/25, 92%). Participants had good access to both computers and smartphones. The preference for remote participation in this project was considered to be convenience factors, social anxiety, and COVID-19–related concerns. Further research is needed to verify the reasons for preferring digital and remote engagement with research across different populations.

Discussion

Principal Findings

The global pandemic necessitated the reevaluation of traditional research methodologies, compelling researchers across disciplines to adapt to the changing environment and adopt agile approaches. This study explored opportunities and considerations from involving vulnerable user groups remotely to provide lessons learned for future research. We did this by reflecting on research practices that involved user-centric evaluation of interactive behavioral and psychological intervention designs. A total of five topics emerged from our analysis: (1) reduced barriers to research inclusion; (2) digital literacy transition; (3) contextualized insights: a window into people's lives; (4) seamless enactment of roles; and (5) increased flexibility for researchers and participants.

Across the 3 cases, remote participation contributed to a more accessible inclusion of users in design. The emerging technology on modern mobile phones offers the potential to engage with participants effectively across digital platforms such as Zoom. Mobile smartphones are prolific, and with the declining cost of data [55], remote methods that seamlessly integrate with mobile devices are becoming more accessible and equitable for user engagement. Low-income user groups may have limited access to working laptops or home computers, as was the situation in case 3 of this study. Adequate provision or access to suitable digital devices is important in digital equity and research in vulnerable groups [56].

Remote methods mitigate accessibility barriers such as travel costs and logistical challenges, which may deter participation from vulnerable groups. In countries such as Australia and Norway, with a diverse and "spread out" geographical landscape, this was highly valuable in the intervention development phase, enabling wider recruitment reach. This also has significant implications for scalability and implementation. A broader recruitment scope may make it easier to include more participants who are less represented in research, such as those living in rural areas [14]. Web-based and remote methods of research engagement, such as social media, may facilitate

engagement with vulnerable groups not connected with organizations, workplaces, or other services [12].

With an increased focus on digital health interventions and programs delivered remotely, remote user methods align with the design process of such programs. The benefits of a wider recruitment pool and efficiency gains, such as reduced travel time or inconvenience, were expected from past studies on remote research methods [15,20,57]. In past research, these gains are often contrasted against other shortcomings of being remote [58], such as lack of contextual insight, connection problems, audio or video problems, low digital literacy, and the like. Emerging from the technological leap through the COVID-19 pandemic, these shortcomings are diminishing, while the perception of benefits for researchers and participants is increasing. The research teams' initial reservations were that remote research would be complex for potentially vulnerable user participants and could increase stress or fatigue [59]. There were also reservations that remote participation would not provide rich enough user data; however, in the cases presented here, it was found that this method did provide contextualized insights and increased ecological relevance.

Implications

This study provides insight into the broader learnings from adapting to remote research practices during the COVID-19 pandemic and beyond. From the findings, we have extracted 4 significant implications for future research and practice.

Potential for More Agile Research

Remote research practices may come closer to the ideal of an agile approach to testing ("microtesting"), involving briefer and more frequent evaluation sessions with users. This has also been recommended by other recent publications within mHealth [3,60]. For researchers to take advantage of this potential for mHealth apps and interventions, it will bring mHealth research closer to agile user experience practice [61-64] and continuous testing of minimum viable products or prototypes as a form of hypothesis testing [65]. Remote methods facilitate fast cycle iterations and testing in a research design process of sensemaking through trial and error [66].

Remote Research Increases Ecologic Relevance

The interventions developed in these research programs were designed to be used within the context of users' lives, usually the home. Thus, a remote testing method was more ecologically relevant than a traditional face-to-face user test in an office setting. We evaluated the interventions using remote methods in the user's home and on their devices. This enables contextual inquiry and enhances the representativeness of research findings and the applicability of the designed solution. Screen sharing from a mobile device has also improved [67] compared with during our data collection; this will reduce the problems of remote testing of mHealth interventions, enabling testing and feedback sessions with users in their own contexts and on their devices with direct interaction on the app [68].

Technology Impacts Privacy and Human Action

Our research found that there is a risk of capturing more personal data than planned through the ease of recording and screen sharing when engaging with participants through web-based modes. As the participant joins from home, their home context is recorded, including background information and activity. Screen sharing from the participant's device may enable accidental capturing of on-screen activity, such as open tabs and browsing history, which may be unintentional on the participant's part. Digital ethnographers have highlighted this factor in previous studies [69]. This highlights the need to safeguard participants' privacy, as participants may not fully grasp the need to protect their privacy [70]. Throughout the research, participants became more aware of how to protect their privacy, which is represented by the increasing use of video filters such as blurred backgrounds, muting cameras, or strategically placing the webcam. However, some participants perhaps showed unintended details of their personal lives. Researchers should be aware of the ethical considerations of recording videos of participants in their home environment and take care to protect their privacy. Consenting protocols, which include preparing participants for digital interactions, are essential so that participants are adequately informed and aware. As researchers, we may also incorporate practices from web-based counseling and telehealth. Researchers in telehealth also call for revisiting ethical guidelines and procedures following the "ongoing natural experiment" of the pandemic [71].

Our research suggests that when selecting technologies for remote research, it is necessary to consider their functionality regarding privacy protection and the mediating role of technology [58] on human action [72,73]. For instance, when we choose Zoom, Discord, or any other technology, we should consider the norms of how these technologies are being used in other contexts, how these patterns might influence researchers and participants, and how this may influence the data collected.

Remote Research Leads to User Involvement on Participants' Terms

Researchers were concerned by the limited opportunity that remote methods present for informal conversation and rapport building. This interaction style enables trust building and may make research participation more comfortable and less intimidating. However, we found that remote methods shifted control to participants and offered greater comfort than attending unfamiliar institutional settings for face-to-face sessions. Remote methods have the potential for enhanced anonymity as participants have more control over what they share. This may be particularly pertinent for research that involves sensitive or taboo topics, allowing individuals to feel more at ease sharing their experiences and perspectives [74,75].

It was our experience, during work-from-home COVID-19 mandates, that power imbalances were diminished as both researchers and participants were dialing in from a home setting. Thus, there was a more equal grounding and reduced power differential [76]. This is worth considering for future research, specifically setting up the research so that participants and researchers are in similar settings during interaction. When 2 researchers dial in from the same physical location, that introduces a new imbalance, and future research should consider applying the principle of "one remote, all remote" [77,78] when

there is a need to do hybrid remote research to ensure equal participation.

Limitations

For the cases in this study, participants were involved in design processes to capture their experience with iterative designs and provide feedback on design revisions. This took place at different time points during the pandemic. The original research was not designed to answer the research questions of this study. Instead, this topic *emerged* [28] through practice and through reflecting on practice [26]. Retrospective studies have limitations since they may depend on a review of data not planned for research use [79], and information may be missing. This has been mitigated by the participation of the 2 lead authors who conducted the original research. However, our interpretation may be biased despite taking a critical stance on our reflections and interpretations.

The cases and findings presented spark conceptual development and analytical discussion [80] on remote user design methods. However, there are also limitations regarding participants and to whom the findings are relevant. Across cases, specific inclusion criteria and requirements related to recruitment likely impacted our ability to recruit participants. For instance, in cases 1 and 2, we could not advertise for participants and relied on third parties to share information about the research project with potential participants. There was also a requirement to speak Norwegian fluently due to the in-app language. These factors may have reduced the number of people with minority or immigrant backgrounds who registered for the research in the Norway project. Bearing in mind that the young unemployed are twice as likely as other young people to have come to Norway as migrants, this is a weakness. Both projects called for narrow recruitment strategies to target specific population groups. Findings from this study reflect the experiences of the population groups that were involved and may not be generalizable. Further research should explore the applicability and benefit of remote user methods across other population groups.

Conclusions

The COVID-19 pandemic has reshaped the research landscape in many ways, driving rapid innovation and the adoption of remote research methods. These methods proved crucial in overcoming recruitment challenges and enabling researchers to engage with diverse participant groups across geographical areas. Applying remote methods within hard-to-reach groups reduced participation barriers, facilitated recruitment, and cultivated a more inclusive and comfortable research environment. As researchers and designers navigate the evolving research landscape, the lessons learned underscore the enduring value of remote research methods in promoting user participation in the design of mHealth interventions. Furthermore, they may serve as a reminder to question persistent assumptions about technological competence and access in vulnerable populations.

Acknowledgments

The authors would like to thank the participants who took part in the study and gave their feedback on the different prototypes. Furthermore, the authors thanks researchers at the University of Stavanger (UiS) who contributed to the Career Learning App (CL-APP) project: Venke F Haaland, Espen Sagen, Mari Rege, Hilde Ness Sandvold, Jone Bjørnestad, and Wenche Ten Velden Hegelstad. They would also like to thank the Norwegian Labour and Welfare Administration in Sandnes and Skole- og Jobbresept at the University Hospital of Stavanger. CL-App is funded by The Norwegian Research Council (grant number 296390).

They would also like to thank the researchers at the Queensland University of Technology (QUT) who contributed to the Responsive Feeding in Tough Times (RFiTT) project: Rebecca Byrne, Smita Nambiar, Danielle Gallegos, Jeremy Kerr, Robyn Penny, and Rachel Laws. RFiTT is funded by the Queensland Children's Hospital Foundation thanks to the generosity of Woolworths customers and team members. They further acknowledge the QUT Design Lab for organizing a research seminar that allowed authors 1 and 2 to meet and share ideas in March 2023, which resulted in this study.

Conflicts of Interest

None declared.

References

1. Fedele DA, Cushing CC, Fritz A, Amaro CM, Ortega A. Mobile health interventions for improving health outcomes in youth: a meta-analysis. JAMA Pediatr 2017 May 01;171(5):461-469 [FREE Full text] [doi: 10.1001/jamapediatrics.2017.0042] [Medline: 28319239]
2. Stowell E, Lyson MC, Saksono H, Wurth RC, Jimison H, Parker AG. Designing and evaluating mHealth interventions for vulnerable populations: a systematic review. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018 Presented at: CHI '18; April 21-26, 2018; Montreal, QC p. 1-17 URL: https://dl.acm.org/doi/10.1145/3173574.3173589 [doi: 10.1145/3173574.3173589]
3. Cornet VP, Toscos T, Bolchini D, Rohani Ghahari R, Ahmed R, Daley C, et al. Untold stories in user-centered design of mobile health: practical challenges and strategies learned from the design and evaluation of an app for older adults with heart failure. JMIR Mhealth Uhealth 2020 Jul 21;8(7):e17703 [FREE Full text] [doi: 10.2196/17703] [Medline: 32706745]

4. Lewis JR. Usability: lessons learned … and yet to be learned. Int J Hum Comput Interact 2014 Jun 18;30(9):663-684. [doi: 10.1080/10447318.2014.930311]
5. Alshar'e M, Albadi A, Mustafa M, Tahir N, Al Amri M. A framework of the training module for untrained observers in usability evaluation motivated by COVID-19: enhancing the validity of usability evaluation for children's educational games. Adv Hum Comput Interact 2022 Mar 9;2022:1-11. [doi: 10.1155/2022/7527457]
6. Moran K. Usability testing 101. Nielsen Norman Group. URL: https://www.nngroup.com/articles/usability-testing-101/ [accessed 2023-08-15]
7. Lowgren J, Carroll JM, Hassenzahl M, Erickson T, Blackwell A, Overbeeke K, et al. The Encyclopedia of Human-Computer Interaction. 2nd edition. Aarhus, Denmark: Interaction Design Foundation; 2019.
8. Vasalou A, Ng BD, Wiemer-Hastings P, Oshlyansky L. Human-moderated remote user testing: protocols and applications. In: Proceedings of the 8th ERCIM Workshop 'User Interfaces for All'. 2004 Presented at: UIA '04; June 28-29, 2004; Wien, Austria p. 1-10 URL: https://reed.cs.depaul.edu/peterh/papers/VasalouERCIM2004.pdf
9. Alhadreti O. Assessing academics' perceptions of blackboard usability using SUS and CSUQ: a case study during the COVID-19 pandemic. Int J Hum Comput Interact 2020 Dec 26;37(11):1003-1015. [doi: 10.1080/10447318.2020.1861766]
10. Hill JR, Harrington AB, Adeoye P, Campbell NL, Holden RJ. Going remote-demonstration and evaluation of remote technology delivery and usability assessment with older adults: survey study. JMIR Mhealth Uhealth 2021 Mar 04;9(3):e26702 [FREE Full text] [doi: 10.2196/26702] [Medline: 33606655]
11. Costagliola G, De RM, Fuccella V, Tabari P. The impact of the COVID-19 pandemic on human–computer interaction empirical research. Interact Comput 2023;35(5):578-589 [FREE Full text] [doi: 10.1093/iwc/iwad031]
12. Gallegos D, Durham J, Rutter C, McKechnie R. Working towards the Active Participation of Underrepresented Populations in Research: A Scoping Review and Thematic Synthesis. Health Soc Care Community 2023 Apr 21;2023:1-26. [doi: 10.1155/2023/1312525]
13. Bratteteig T, Wagner I. Unpacking the Notion of Participation in Participatory Design. Comput Support Coop Work 2016 Sep 2;25(6):425-475 [FREE Full text] [doi: 10.1007/s10606-016-9259-4]
14. Ellard-Gray A, Jeffrey NK, Choubak M, Crann SE. Finding the hidden participant: solutions for recruiting hidden, hard-to-reach, and vulnerable populations. Int J Qual Methods 2015 Dec 17;14(5):160940691562142. [doi: 10.1177/1609406915621420]
15. Petrie H, Hamilton F, King N, Pavan P. Remote usability evaluations with disabled people. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2006 Presented at: CHI '06; April 22-27, 2006; Montréal, QC p. 1133-1141 URL: https://dl.acm.org/doi/10.1145/1124772.1124942 [doi: 10.1145/1124772.1124942]
16. Bonevski B, Randell M, Paul C, Chapman K, Twyman L, Bryant J, et al. Reaching the hard-to-reach: a systematic review of strategies for improving health and medical research with socially disadvantaged groups. BMC Med Res Methodol 2014 Mar 25;14(1):42 [FREE Full text] [doi: 10.1186/1471-2288-14-42] [Medline: 24669751]
17. Heidegger M. Being and Time. Hoboken, NJ: Wiley-Blackwell; 1978.
18. Moorhouse BL, Wong KM. The COVID-19 pandemic as a catalyst for teacher pedagogical and technological innovation and development: teachers' perspectives. Asia Pac J Educ 2021 Oct 11;42(sup1):105-120. [doi: 10.1080/02188791.2021.1988511]
19. Statement on the fifteenth meeting of the IHR (2005) emergency committee on the COVID-19 pandemic. World Health Organization. 2023 May 5. URL: https://tinyurl.com/3k786y5e [accessed 2023-08-22]
20. Hill JR, Brown JC, Campbell NL, Holden RJ. Usability-in-place-remote usability testing methods for homebound older adults: rapid literature review. JMIR Form Res 2021 Nov 02;5(11):e26181 [FREE Full text] [doi: 10.2196/26181] [Medline: 34726604]
21. Currie H, Harvey A, Bond R, Magee J, Finlay D. Remote synchronous usability testing of public access defibrillators during social distancing in a pandemic. Sci Rep 2022 Aug 26;12(1):14575. [doi: 10.1038/s41598-022-18873-7] [Medline: 36028561]
22. Lobchuk M, Bathi PR, Ademeyo A, Livingston A. Remote moderator and observer experiences and decision-making during usability testing of a web-based empathy training portal: content analysis. JMIR Form Res 2022 Aug 03;6(8):e35319 [FREE Full text] [doi: 10.2196/35319] [Medline: 35921138]
23. Archibald MM, Ambagtsheer RC, Casey MG, Lawless M. Using Zoom videoconferencing for qualitative data collection: perceptions and experiences of researchers and participants. Int J Qual Methods 2019 Sep 11;18:160940691987459. [doi: 10.1177/1609406919874596]
24. ISO 9241-210:2019: ergonomics of human-system interaction- part 210: human-centred design for interactive systems. International Organization for Standardization. 2019. URL: https://www.iso.org/standard/77520.html [accessed 2023-10-19]
25. Cockton G. Usability evaluation. In: Ahmad A, Blackwell A, Dix A, Schmidt A, Sutcliffe AG, Crabtree A, et al, editors. The Encyclopedia of Human-Computer Interaction. 2nd edition. Aarhus, Denmark: Interaction Design Foundation; 2019.
26. Schon DA. The Reflective Practitioner: How Professionals Think in Action. London, UK: Basic Books; 1984.
27. Palmer AM, Burns S. Reflective Practice in Nursing: The Growth of the Professional Practitione. Hoboken, NJ: Wiley-Blackwell; 1994.

28. Gaver W, Krogh PG, Boucher A, Chatting D. Emergence as a feature of practice-based design research. In: Proceedings of the 2022 ACM Designing Interactive Systems Conference. 2022 Presented at: DIS '22; June 13-17, 2022; Virtual Event p. 517-526 URL: https://dl.acm.org/doi/abs/10.1145/3532106.3533524 [doi: 10.1145/3532106.3533524]
29. Suchman L. Human-Machine Reconfigurations: Plans and Situated Actions. Cambridge, MA: Cambridge University Press; 2006.
30. Pernice K, Krause R. Affinity diagramming for collaboratively sorting UX findings and design ideas. Nielsen Norman Group. URL: https://www.nngroup.com/articles/affinity-diagram/ [accessed 2023-08-22]
31. Miro. URL: http://www.miro.com [accessed 2023-08-10]
32. Scupin R. The KJ method: a technique for analyzing data derived from Japanese ethnology. Hum Organ 1997;56(2):233-237. [doi: 10.17730/humo.56.2.x335923511444655]
33. Fischer C, Gregor S. Forms of reasoning in the design science research process. In: Proceedings of the 6th International Conference on Service-Oriented Perspectives in Design Science Research. 2011 Presented at: DESRIST '11; May 5-6, 2011; Milwaukee, WI p. 17-31 URL: https://link.springer.com/chapter/10.1007/978-3-642-20633-7_2 [doi: 10.1007/978-3-642-20633-7_2]
34. Braun V, Clarke V. Thematic Analysis: A Practical Guide. Thousand Oaks, CA: Sage Publications; 2021.
35. Charmaz K. Constructivist grounded theory. J Posit Psychol 2016 Dec 14;12(3):299-300. [doi: 10.1080/17439760.2016.1262612]
36. Culén AL, van der Velden M. The digital life of vulnerable users: designing with children, patients, and elderly. In: Proceedings of the 4th Scandinavian Conference on Information Systems on Nordic Contributions in IS Research. 2013 Presented at: SCIS '13; August 11-14, 2013; Oslo, Norway p. 53-71 URL: https://link.springer.com/chapter/10.1007/978-3-642-39832-2_4 [doi: 10.1007/978-3-642-39832-2_4]
37. Straand IJ, Følstad A, Bjørnestad JR. Exploring a gaming-based intervention for unemployed young adults: thematic analysis. JMIR Hum Factors 2024 Jan 18;11:e44423 [FREE Full text] [doi: 10.2196/44423] [Medline: 38236624]
38. Dweck CS. Mindset - Updated Edition: Changing The Way You think To Fulfil Your Potential. Paris, France: Hachette Livre; 2017.
39. Dweck CS, Yeager DS. A growth mindset about intelligence. In: Walton GM, Crum AJ, editors. Handbook of Wise Interventions: How Social Psychology Can Help People Change. New York, NY: The Guilford Press; 2020.
40. Yeager DS, Dweck CS. What can be learned from growth mindset controversies? Am Psychol 2020 Dec;75(9):1269-1284 [FREE Full text] [doi: 10.1037/amp0000794] [Medline: 33382294]
41. New national restrictions. The Office of the Prime Minister. 2020 Oct 26. URL: https://www.regjeringen.no/en/historical-archive/solbergs-government/Ministries/smk/Press-releases/2020/new-national-restrictions/id2776995/ [accessed 2023-12-12]
42. Scaglioni S, De Cosmi V, Ciappolino V, Parazzini F, Brambilla P, Agostoni C. Factors influencing children's eating behaviours. Nutrients 2018 May 31;10(6):706 [FREE Full text] [doi: 10.3390/nu10060706] [Medline: 29857549]
43. Baxter KA, Kerr J, Nambiar S, Gallegos D, Penny RA, Laws R, et al. A design thinking-led approach to develop a responsive feeding intervention for Australian families vulnerable to food insecurity: eat, Learn, Grow. Health Expect 2024 Apr 20;27(2):e14051 [FREE Full text] [doi: 10.1111/hex.14051] [Medline: 38642335]
44. COVID-19: a chronology of state and territory government announcements (up until 30 June 2020). Parliament House, Parliament of Australia. URL: https://www.aph.gov.au/About_Parliament/Parliamentary_departments/Parliamentary_Library/pubs/rp/rp2021/Chronologies/COVID-19StateTerritoryGovernmentAnnouncements#_Toc52275790 [accessed 2023-11-01]
45. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, REDCap Consortium. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208 [FREE Full text] [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
46. Lozano-Blasco R, Robres AQ, Sánchez AS. Internet addiction in young adults: a meta-analysis and systematic review. Comput Human Behav 2022 May;130:107201. [doi: 10.1016/j.chb.2022.107201]
47. Bakken IJ, Wenzel HG, Götestam KG, Johansson A, Oren A. Internet addiction among Norwegian adults: a stratified probability sample study. Scand J Psychol 2009 Apr 24;50(2):121-127. [doi: 10.1111/j.1467-9450.2008.00685.x] [Medline: 18826420]
48. Shaver J. The state of telehealth before and after the COVID-19 pandemic. Prim Care 2022 Dec;49(4):517-530 [FREE Full text] [doi: 10.1016/j.pop.2022.04.002] [Medline: 36357058]
49. Anderson JT, Bouchacourt LM, Sussman KL, Bright LF, Wilcox GB. Telehealth adoption during the COVID-19 pandemic: a social media textual and network analysis. Digit Health 2022;8:20552076221090041 [FREE Full text] [doi: 10.1177/20552076221090041] [Medline: 35392254]
50. Nicol TL, Roberts J, Jordan R. Removing regulatory barriers to telehealth before and after COVID-19. Brookings Institution. 2020. URL: https://www.brookings.edu/wp-content/uploads/2020/05/Removing-barriers-to-telehealth-before-and-after-COVID-19_PDF.pdf [accessed 2024-05-19]
51. Appleton R, Williams J, Vera San Juan N, Needle JJ, Schlief M, Jordan H, et al. Implementation, adoption, and perceptions of telemental health during the COVID-19 pandemic: systematic review. J Med Internet Res 2021 Dec 09;23(12):e31746 [FREE Full text] [doi: 10.2196/31746] [Medline: 34709179]

52. Bick A, Blandin A, Mertens K. Work from Home before and after the COVID-19 Outbreak. Am Econ J-Macroecon 2023 Oct 01;15(4):1-39. [doi: 10.1257/mac.20210061]
53. Bick A, Blandin A, Mertens K. Work from home after the COVID-19 outbreak. Federal Reserve Bank of Dallas. 2020 Jul. URL: https://www.dallasfed.org/~/media/documents/research/papers/2020/wp2017r2.pdf [accessed 2024-04-29]
54. Kender K. Tumblr is Queer and Twitter is toxic: speculating about the vibe of social media spaces. In: Proceedings of the 2022 Nordic Human-Computer Interaction Conference. 2022 Presented at: NordiCHI '22; October 8-12, 2022; Aarhus, Denmark p. 1-8 URL: https://dl.acm.org/doi/fullHtml/10.1145/3546155.3547279 [doi: 10.1145/3546155.3547279]
55. Median mobile retail cost per GB in Australia FY 2015-2019, by service type. Statista. 2023. URL: https://www.statista.com/statistics/1099607/australia-mobile-service-median-cost-per-gb-by-service-type/ [accessed 2023-09-24]
56. Lyles CR, Wachter RM, Sarkar U. Focusing on digital health equity. JAMA 2021 Nov 09;326(18):1795-1796. [doi: 10.1001/jama.2021.18459] [Medline: 34677577]
57. Shields MM, McGinnis MN, Selmeczy D. Remote research methods: considerations for work with children. Front Psychol 2021 Oct 28;12:703706 [FREE Full text] [doi: 10.3389/fpsyg.2021.703706] [Medline: 34777090]
58. Frittgen EM, Haltaufderheide J. 'Can you hear me?': communication, relationship and ethics in video-based telepsychiatric consultations. J Med Ethics 2022 Jan 05;48(1):22-30. [doi: 10.1136/medethics-2021-107434] [Medline: 34610979]
59. Anh LE, Whelan E, Umair A. 'You're still on mute'. A study of video conferencing fatigue during the COVID-19 pandemic from a technostress perspective. Behav Inf Technol 2022 Jul 01;42(11):1758-1772. [doi: 10.1080/0144929x.2022.2095304]
60. Buck Harleah G. Does health care technology work for or against older adults? J Gerontol Nurs 2017 Aug;43(8):3-4. [doi: 10.3928/00989134-20170719-01]
61. Gothelf J. Lean UX: Applying Lean Principles to Improve User Experience. Sebastopol, CA: Shroff/O'Reilly; Mar 2013.
62. Knapp J, Zeratsky J, Kowitz B. Sprint: How to Solve Big Problems and Test New Ideas in Just Five Days. New York, NY: Simon & Schuster; 2016.
63. Reis E. The Lean Startup: How Today's Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses. New York, NY: Crown Currency Publishing; 2011.
64. Felin T, Gambardella A, Stern S, Zenger T. Lean startup and the business model: experimentation revisited. Long Range Plann 2020 Aug;53(4):101889 [FREE Full text] [doi: 10.1016/j.lrp.2019.06.002]
65. Aarlien D, Colomo-Palacios R. Lean UX: a systematic literature review. In: Proceedings of the 20th International Conference on Computational Science and Its Applications. 2020 Presented at: ICCSA '20; July 1-4, 2020; Cagliari, Italy p. 500-510 URL: https://link.springer.com/chapter/10.1007/978-3-030-58817-5_37 [doi: 10.1007/978-3-030-58817-5_37]
66. Swann C. Action research and the practice of design. Design Issues 2002 Jan;18(1):49-61 [FREE Full text] [doi: 10.1162/07479360252756287]
67. Sharing your iOS screen from the Zoom desktop client. Zoom Support. 2023 Dec 29. URL: https://support.zoom.com/hc/en/article?id=zm_kb&sysparm_article=KB0066704 [accessed 2024-03-21]
68. Shahil-Feroz A, Yasmin H, Saleem S, Bhutta Z, Seto E. Remote Moderated Usability Testing of a Mobile Phone App for Remote Monitoring of Pregnant Women at High Risk of Preeclampsia in Karachi, Pakistan. Informatics 2023 Oct 17;10(4):79. [doi: 10.3390/informatics10040079]
69. Addeo F, Masullo G. Studying the digital society: digital methods between tradition and innovation in social research. Ital Sociol Rev 2021;11(4S):153. [doi: 10.13136/isr.v11i4S.428]
70. Lee EW, McCloud RF, Viswanath K. Designing effective eHealth interventions for underserved groups: five lessons from a decade of eHealth intervention design and deployment. J Med Internet Res 2022 Jan 07;24(1):e25419 [FREE Full text] [doi: 10.2196/25419] [Medline: 34994700]
71. Kaplan B. Revisiting health information technology ethical, legal, and social issues and evaluation: telehealth/telemedicine and COVID-19. Int J Med Inform 2020 Nov;143:104239 [FREE Full text] [doi: 10.1016/j.ijmedinf.2020.104239] [Medline: 33152653]
72. Verbeek PP. Materializing morality: design ethics and technological mediation. Sci Technol Hum Values 2016 Aug 19;31(3):361-380. [doi: 10.1177/0162243905285847]
73. Goldman SL, Ihde D, Zimmerman ME, Hickman LA. Technology and the lifeworld: from garden to earth. Technol Cult 1991 Oct;32(4):1135. [doi: 10.2307/3106183]
74. Frost J, Vermeulen IE, Beekers N. Anonymity versus privacy: selective information sharing in online cancer communities. J Med Internet Res 2014 May 14;16(5):e126 [FREE Full text] [doi: 10.2196/jmir.2684] [Medline: 24828114]
75. Helou S, Abou-Khalil V, El Helou E, Kiyono K. Factors related to personal health data sharing: data usefulness, sensitivity and anonymity. Stud Health Technol Inform 2021 May 27;281:1051-1055. [doi: 10.3233/SHTI210345] [Medline: 34042839]
76. Goffman E. Stigma: Notes on the Management of Spoiled Identity. New York, NY: Simon & Schuster; 2009.
77. Frisch B, Greene C. What it takes to run a great hybrid meeting. Harvard Business Review. 2021 Jun 3. URL: https://hbr.org/2021/06/what-it-takes-to-run-a-great-hybrid-meeting [accessed 2023-12-07]
78. Bowtell S. How to facilitate a hybrid meeting or workshop. RedAgile. 2023 Mar 13. URL: https://www.redagile.com/post/facilitate-a-hybrid-meeting [accessed 2023-12-07]
79. Talari K, Goyal M. Retrospective studies - utility and caveats. J R Coll Physicians Edinb 2020 Dec;50(4):398-402. [doi: 10.4997/JRCPE.2020.409] [Medline: 33469615]

80. Straand IJ, Jevnaker BH. Leading transformation in an uncertain world: a case for strategic speculative design. In: Proceedings of the 19th European Conference on Management Leadership and Governance. 2023 Presented at: ECMLG '23; November 23-24, 2023; London, UK p. 380-387 URL: https://papers.academic-conferences.org/index.php/ecmlg/article/view/1947/1824 [doi: 10.34190/ecmlg.19.1.1947]

Abbreviations

CL-APP: Career Learning App
HCD: human-centered design
mHealth: mobile health
NAV: Norwegian Labour and Welfare Administration
REDCap: Research Electronic Data Capture
RFiTT: Responsive Feeding in Tough Times

Edited by L Buis; submitted 18.12.23; peer-reviewed by J Weber, G Costagliola, P Tabari, D Singh; comments to author 20.02.24; revised version received 24.03.24; accepted 24.04.24; published 14.06.24.

Please cite as:
Straand IJ, Baxter KA, Følstad A
Remote Inclusion of Vulnerable Users in mHealth Intervention Design: Retrospective Case Analysis
JMIR Mhealth Uhealth 2024;12:e55548
URL: https://mhealth.jmir.org/2024/1/e55548
doi:10.2196/55548
PMID:38875700

Using a Quality-Controlled Dataset From ViSi Mobile Monitoring for Analyzing Posture Patterns of Hospitalized Patients: Retrospective Observational Study

Emily J Huang[1], PhD; Yuexin Chen[1], MS; Clancy J Clark[2], MD

[1]Department of Statistical Sciences, Wake Forest University, 1834 Wake Forest Road, Winston Salem, NC, United States

[2]Department of Surgery, Wake Forest School of Medicine, Winston Salem, NC, United States

Corresponding Author:
Emily J Huang, PhD
Department of Statistical Sciences, Wake Forest University, , 1834 Wake Forest Road, Winston Salem, NC, , United States

Abstract

Background: ViSi Mobile has the capability of monitoring a patient's posture continuously during hospitalization. Analysis of ViSi telemetry data enables researchers and health care providers to quantify an individual patient's movement and investigate collective patterns of many patients. However, erroneous values can exist in routinely collected ViSi telemetry data. Data must be scrutinized to remove erroneous records before statistical analysis.

Objective: The objectives of this study were to (1) develop a data cleaning procedure for a 1-year inpatient ViSi posture dataset, (2) consolidate posture codes into categories, (3) derive concise summary statistics from the continuous monitoring data, and (4) study types of patient posture habits using summary statistics of posture duration and transition frequency.

Methods: This study examined the 2019 inpatient ViSi posture records from Atrium Health Wake Forest Baptist Medical Center. First, 2 types of errors, record overlap and time inconsistency, were identified. An automated procedure was designed to search all records for these errors. A data cleaning procedure removed erroneous records. Second, data preprocessing was conducted. Each patient's categorical time series was simplified by consolidating the 185 ViSi codes into 5 categories (Lying, Reclined, Upright, Unknown, User-defined). A majority vote process was applied to remove bursts of short duration. Third, statistical analysis was conducted. For each patient, summary statistics were generated to measure average time duration of each posture and rate of posture transitions during the whole day and separately during daytime and nighttime. A k-means clustering analysis was performed to divide the patients into subgroups objectively.

Results: The analysis used a sample of 690 patients, with a median of 3 days of extensive ViSi monitoring per patient. The median of posture durations was 10.2 hours/day for Lying, 8.0 hours/day for Reclined, and 2.5 hours/day for Upright. Lying had similar percentages of patients in low and high durations. Reclined showed a decrease in patients for higher durations. Upright had its peak at 0 - 2 hours, with a decrease for higher durations. Scatter plots showed that patients could be divided into several subgroups with different posture habits. This was reinforced by the k-means analysis, which identified an active subgroup and two sedentary ones with different resting styles.

Conclusions: Using a 1-year ViSi dataset from routine inpatient monitoring, we derived summary statistics of posture duration and posture transitions for each patient and analyzed the summary statistics to identify patterns in the patient population. This analysis revealed several types of patient posture habits. Before analysis, we also developed methodology to clean and preprocess routinely collected inpatient ViSi monitoring data, which is a major contribution of this study. The procedure developed for data cleaning and preprocessing can have broad application to other monitoring systems used in hospitals.

(JMIR Mhealth Uhealth 2024;12:e54735) doi:10.2196/54735

KEYWORDS

posture monitoring; ViSi mobile; wearable device; inpatient; quality control; observational study; monitoring data; inpatient monitoring; wearables; posture

Introduction

Monitoring the level of mobilization and its changes over time provides critical health information on hospitalized patients. For example, timing of resuming certain postures after surgery can reflect the pace and quality of recovery [1,2]. A traditional method for tracking posture is direct observation, which can be done by a nurse who directly observes and records the patient's posture manually. However, this is time-consuming and resource-intensive, especially when the observers are tracking multiple patients on a ward [3]. Other limitations of direct

observation are subjective reports and intermittent recordings, which may also suffer from bias and random human error [1,4].

To overcome these limitations, some hospitals utilize monitoring systems that use sensors (eg, accelerometer, infrared, and radio frequency identification sensors) to collect measurements for tracking a patient's posture or movement [1,3,5-7]. The collection of measurements by these sensors is continuous and automated. A trained algorithm in the device converts the collected sensor data to a patient's posture objectively. The output reports the patient's postures and activities instantaneously. These data can be used to derive the amount of time a patient spends on bed rest. Moreover, posture monitoring records before and after surgery could be compared to identify whether a patient is having a smooth recovery or a poor one that requires intervention.

Statistical analyses of large datasets collected from many patients using these monitoring systems can further reveal patterns in patients' condition and behavior that can be useful for medical research and operational guidance. There have been preliminary results in connecting hospital sensor observations and patients' condition statistically. For example, recent accelerometer-based studies found that inpatients with acute illness were highly inactive, spending 93%-98.8% of their stay sedentary [5]. Researchers have also used posture monitoring to capture changes in a patient's mobility over the course of their hospital stay and use it as an indicator of recovery rate. In a study of inpatients aged 60 years or older admitted from the emergency room, Theou et al found that those patients classified as less mobile upon admission experienced an increase in their upright time during their hospital stay, possibly corresponding to an improvement in their health condition [8].

A real-time monitoring system that has demonstrated promise in the hospital setting is ViSi Mobile (Sotera Wireless) [9-12]. This device continuously monitors a patient's vital signs and posture. It can be used to capture walking periods, as well as more subtle movements, such as getting out of bed or changing posture in bed. To the best of our knowledge, there are only 3 published studies analyzing ViSi posture data of hospitalized patients. Restrepo et al tested the accuracy of ViSi posture measurements by comparing them to those from direct observation [3]. This study found that ViSi can accurately classify certain static postures, such as lying down and sitting. However, it can have systematic errors in classifying the activity of walking. In an analysis of 2 randomized trials with patients recovering from abdominal surgery, Rivas et al examined the relationship between self-reported pain score and ViSi-measured mobility [1]. The authors estimated the decreasing rate of change in mobility with increasing pain score, and they also found that lower mobility was associated with more postoperative complications. In a study of ViSi data from noncardiac surgery patients, Turan et al found that increased mobility in the 48 hours postoperation was associated with fewer postoperative complications and shorter length of hospital stay [13].

This study sought to characterize posture habits in a large hospitalized patient population, including both surgery and nonsurgery patients. Another purpose of the study was to conduct quality control of the ViSi posture data because the routine hospital measurements originally collected for operational purposes contain different types of errors that require careful treatment. Overall, the study objectives were to (1) develop a data cleaning procedure for a large inpatient ViSi posture monitoring dataset with a duration of 1 year, (2) consolidate posture codes into categories, (3) derive concise summary statistics from the continuous monitoring data, and (4) study types of patient posture habits using summary statistics of posture duration and transition frequency.

Methods

Ethical Considerations

This study used a deidentified dataset and was approved by the Wake Forest School of Medicine Institutional Review Board (IRB00051033). No patient consent was required. No compensation was provided.

Data Description

Since 2015, real-time posture data have been collected routinely using the ViSi Mobile System for patients on postoperative surgical wards and acute medical inpatient wards at Atrium Health Wake Forest Baptist Medical Center in North Carolina, United States of America. The starting point of ViSi monitoring is at admission to the hospital ward. The ViSi device has a wrist module, upper arm module, and chest module, each of which contains a 3-axis accelerometer [14]. Posture is estimated using data collected by the accelerometer sensors. In addition, other sensors on the ViSi device simultaneously measure blood pressure, heart rate, pulse rate, respiratory rate, and oxygen saturation.

This study focused on the ViSi posture data collected at Atrium Health Wake Forest Baptist Medical Center in 2019 from both surgery and medicine services. The ViSi Mobile system converts accelerometer measurements into 15 categories of posture using a proprietary algorithm developed by Sotera Wireless. Each category is presented by a code (Multimedia Appendix 1). A code can describe a single activity, such as walking (WLK). Some codes may combine more than one position or situation into one posture category. For instance, code U90 for upright can be sitting or standing. Code U45, representing a reclined position, may correspond to sitting in a reclined chair or lying when the bed is in a reclined position. On the other hand, for lying down, multiple codes are used to distinguish different postures, such as supine (LSP), prone (LPR), right side (LRS), and left side (LLS). FALL indicates that a patient may have fallen (this code was not observed in the 2019 ViSi posture dataset). UNK indicates that the patient's posture could not be determined by the proprietary algorithm and was categorized as unknown. The ViSi Mobile system also allows the user (patient, nurse, provider) to self-report the patient's posture. Self-reported posture follows the same code pattern, prefixed by an "S-." For instance, self-reported upright is "S-U90."

The ViSi Mobile system outputs a posture recording every 15 seconds. A timestamp is used to identify the time of a given recording. A recording can be a single posture code (eg, "WLK"), or it can be a permutation of 2 or more codes (eg, "U45 U90") if the patient performed multiple postures during

the 15-second window. In the latter case, the permutation of codes is listed in the order that the postures were performed. For example, "U45 U90" indicates that the patient changed their position from reclined to upright, which is an outcome beyond the original 15 categories. Therefore, the permutation of 2 or more codes can generate many other possible outcomes of a patient's posture recording. In the dataset, there were 185 different outcomes in the posture recordings.

A patient's ViSi posture data are saved in several files. All files for a given patient can be identified by the patient's medical record number (MRN). Each file usually contains data from several ViSi devices because each device needs to be charged daily and replaced with another one for continuous observation [11,12]. Within each file, each posture recording is labeled by its timestamp and the device serial number. The files vary in length, including the number of posture recordings and the number of ViSi devices.

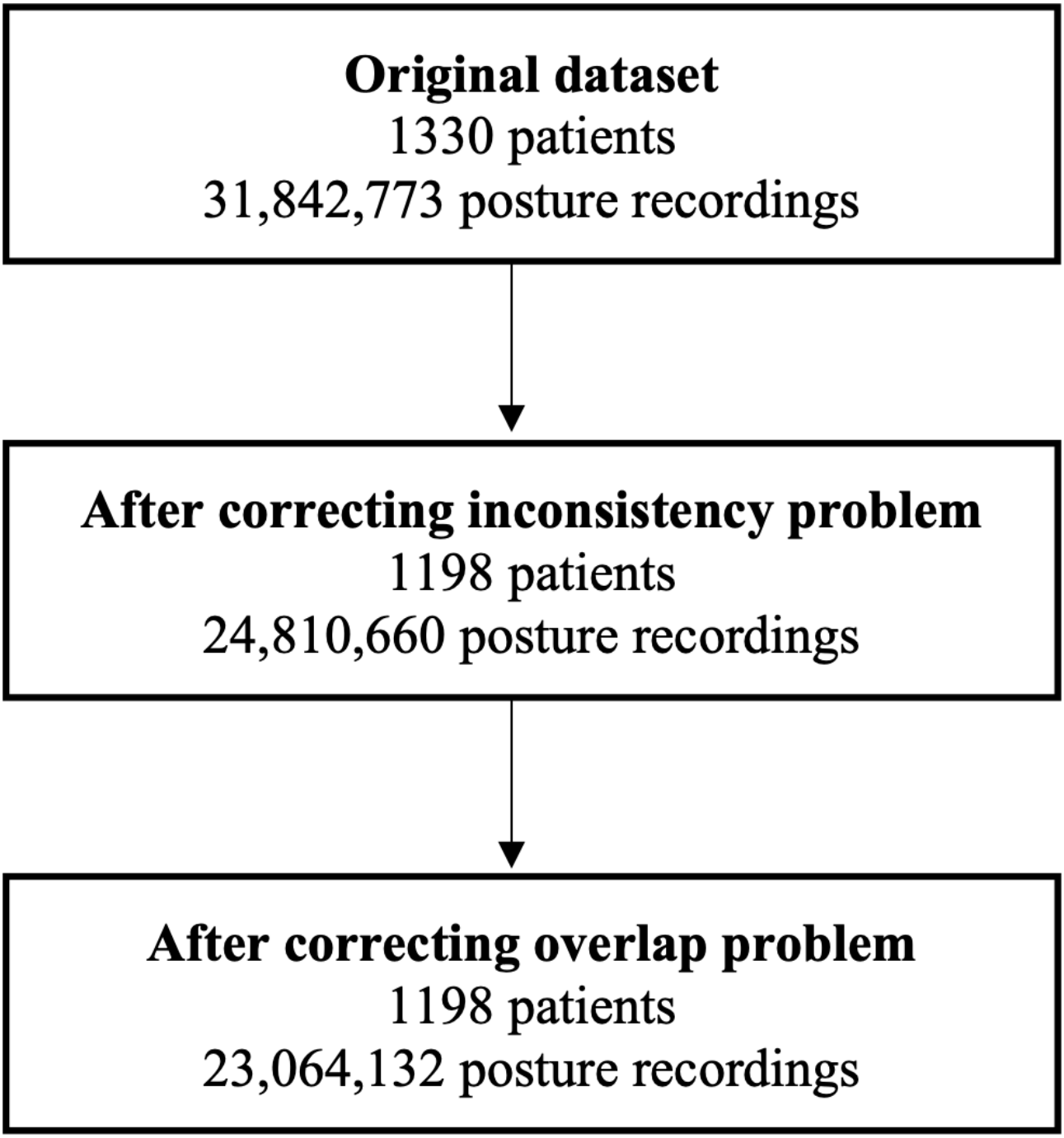
Data Cleaning

The study team conducted a thorough check of the data, which led to the identification of 2 problems. First, "overlap" of posture data was noted for some patients. As described previously, a given patient's records are contained in several files identified by their MRN. For some patients, some data recordings contained in one file overlap with recordings contained in another file. These overlapping recordings have the same timestamps but different readings. They usually come from ViSi devices with different serial numbers. We call this the "overlap problem." About 10% of patients in the dataset exhibited this problem. The overlap period can be long, up to days. Since a patient cannot wear multiple ViSi devices at the same time, overlap should not occur. We speculate that overlap could happen when a recharged ViSi device is reassigned to another patient without updating the MRN. Therefore, when overlap occurs, some data from another patient has likely been misassigned to the given patient.

Second, we identified an inconsistency between the time of hospital stay and the period of ViSi measurements for some patients. For each patient, we compared the ViSi posture data to the admission and discharge times to confirm that the patient's data fell within this time window. However, some patients had ViSi data before their admit time or after their discharge time, which will be referred to as the "inconsistency problem." The mismatch could be large in some cases, up to days. We believe that there are 3 possible reasons for mismatch. One possibility is that the admit and discharge times are not exact. Another is that ViSi data are sometimes collected before admission, resulting in data before the admit time. A third possibility is that the data may have come from another patient. The first 2 situations are benign, while the third situation is problematic and needs to be addressed. Based on the available information, it is not possible to determine which situation has occurred.

The following strategies were used to address these 2 problems. For the overlap problem, all overlapping records were flagged in all files for each individual patient. Data were not immediately discarded, given that one of the data streams could actually be correct. However, following discussion with in-hospital data managers and the device provider, there was no objective method to identify which data stream was correct. Therefore, this study excludes all records flagged as overlapping. A similar approach was used to address the inconsistency problem. Any data before the admission time or after the discharge time was flagged. The statistical analysis in this paper does not include flagged data. Figure 1 shows the sample size before and after we removed the flagged records for the inconsistency and overlap problems. Overall, the 2 procedures excluded 9.9% (132/1330) of patients and 27.6% (8,778,641/31,842,773) of posture recordings.

Figure 1. Flow diagram of data cleaning procedure. Data cleaning included two stages: (1) corrections for the inconsistency problem, and (2) corrections for the overlap problem. We show the number of patients and number of posture recordings remaining after each stage.



Data Preprocessing

Before data analysis, 2 steps of preprocessing were performed to consolidate the possible posture outcomes and to remove short inconsequential fluctuations. First, we consolidated the 185 possible posture outcomes into a more manageable set of 5 groups. These five groups are Upright, Reclined, Lying, User-defined, and Unknown. Compared to the original 185 outcomes, we believe these 5 groups are more easily interpretable and useful to clinicians and analysts. In this consolidated system, each original posture recording (single code or permutation of codes) falls into 1 of the 5 groups. Table 1 provides the definitions of the posture groups, including their connections with the original posture recordings. Essentially, any posture recording that includes a self-reported code goes to the group User-defined. Any recording that includes the UNK code is classified into group Unknown. The group Lying contains posture recordings that are composed of lying codes only (ie, codes beginning with L). Of the remaining recordings, any that has a “WLK” or “U90” is put in the group Upright. The rest of the recordings must contain a “U45” and fall into the group Reclined. The third column of Table 1 shows examples of posture recordings in each group.

Table . The 5 posture groups.[a]

Posture group	Definition	Example recordings in posture group
User-defined	Any posture recording that includes "S - …"	"UNK LSP S-LSP," "UNK LSP S-U45," "UNK S-LLS LSP," "UNK U90 S-U90," "UNK U90 U45 LLS S-U90"
Unknown	• Any posture recording that includes "UNK" (no "S - …"). • Posture recordings that were blank were also classified as Unknown.	"UNK," "UNK LSP"
Upright	Any posture recording including "U90" or "WLK" (no "S - …" or "UNK")	"U90 U45," "U90 U45 WLK," "U90 U45 LLS," "U90 U45 LLS WLK," "U45 LSP WLK"
Reclined	Any posture recording including "U45" (no "S - …," "UNK," "U90," or "WLK")	"U45," "U45 LLS," "U45 LPR," "U45 LPR LRS," "U45 LRS," "U45 LRS LLS," "U45 LSP"
Lying	"LLS," "LRS," "LPR," "LSP," or permutations of these	"LLS," "LPR," "LPR LLS," "LRS," "LRS LLS," "LSP," "LSP LRS"

[a]LLS: lying left side; LPR: lying prone; LRS: lying right side; LSP: lying supine; UNK: unknown; WLK: walking.

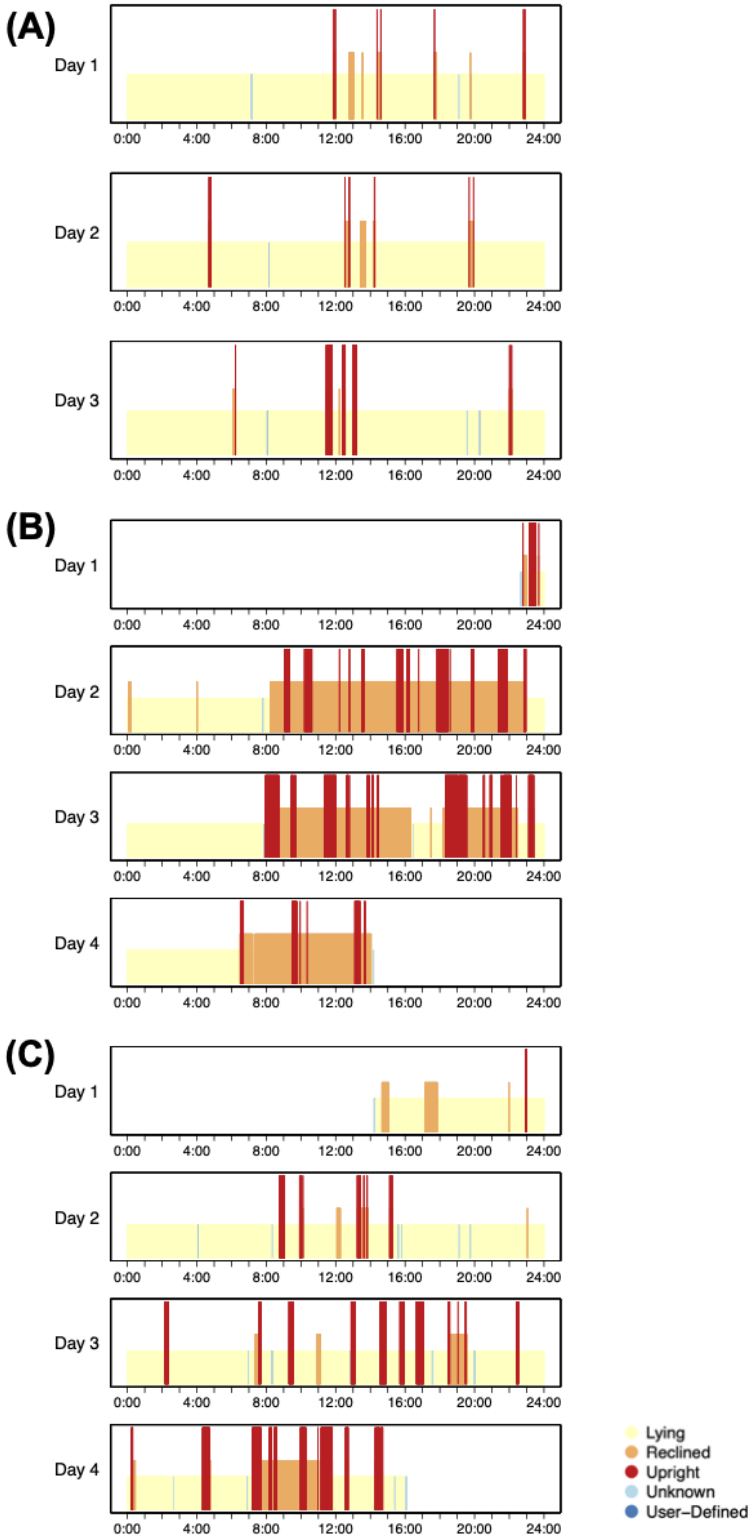
Note that the Upright group combines walking with sitting and standing, rather than treating walking as a separate group. This choice was made because the ViSi device may classify the activity of walking as U90 (the same ViSi code for sitting and standing) instead of WLK, the code for walking [3]. In addition, we set User-defined as its own group, rather than assigning individual user-defined postures to their corresponding categories (eg, "S-LSP" was not assigned to Lying). This was because it is difficult to determine whether the user-defined postures were inputted intentionally. In the rest of the paper, we focus on these 5 posture groups, instead of the 185 posture outcomes. Hereafter, the term "posture" indicates posture group.

After converting the raw posture recordings to posture groups, data smoothing was performed to reduce noise. This was the second step of data preprocessing. Here, noise is defined as an isolated posture of 1 timestamp in between 2 timestamps with the same posture. For example, for a sequence of 3 postures of Lying, Reclined, Lying, the middle one is considered as noise. We removed noise through data smoothing. For each timestamp, we used the following majority vote process to remove the noise. A majority vote was taken, considering the timestamp in question and the 2 adjacent timestamps (ie, 15 seconds before and 15 seconds after). Based on the 3 votes, whichever posture appeared most frequently would be the winner. For example, if there were 2 votes for Lying and 1 vote for Reclined, the posture at the timestamp in question was taken to be Lying. If there was no winner in the majority vote (eg, the three timestamps all had different postures), then the posture at the timestamp in question was left unchanged. This process is analogous to taking a moving average using a sliding window with 3 consecutive timestamps [15], except the mode is taken instead of the average.

Figure 2 presents 3 examples of cleaned and preprocessed posture data, which illustrate how the patterns of posture can vary from 1 patient to another. The first example (Figure 2A) shows long periods of Lying, interspersed with short Reclined and Upright segments. This patient is considered inactive. The second example (Figure 2B) shows a patient who was in the Upright and Reclined postures during the daytime hours and the Lying posture during the nighttime hours. One may infer that this patient was trying to maintain day/night routine behavior during the hospital stay. Figure 2C plots data from a cancer patient postsurgery. The patient spent days 1 and 2 mostly in the Lying posture. Increased activity was observed from day 2 to 3, with more frequent Upright and Reclined periods. On day 4, the patient stayed in Upright and Reclined postures for longer periods than previous days. This shows the patient's process of postsurgery recovery.

The analyses described in the rest of the Methods section used the cleaned and pre-processed data.

Figure 2. Example cleaned and preprocessed data from 3 patients. The short yellow bars correspond to Lying. The taller orange bars correspond to Reclined. The tallest red bars correspond to Upright. The light and dark blue bars correspond to Unknown and User-Defined, respectively.



Summary Statistics

In all, 20 summary statistics were calculated for each patient, using their cleaned and preprocessed data. The purpose was to translate the large amount of data into a concise set of interpretable statistics. The 20 summary statistics can be divided into 3 categories. The first category (data quantity) examines the length of ViSi posture data available for the patient. The second (posture duration) finds the patient's average time spent in each of the 5 postures. The third category (posture transitions) measures the frequency of posture changes in the patient's data.

For data quantity, 2 measures were generated. First, we recorded the total number of days that the patient had any ViSi posture data taken (total days). Total days includes days with complete data (ie, all 24 hours), as well as days with only a short data segment starting from a single recording. Second, we counted

the number of days where the patient had extensive ViSi posture monitoring, defined as at least 22 hours of recordings per day, which we used for further analysis. Hereafter, we refer to these days as analysis days. Since ViSi provides 1 posture recording every 15 seconds, each analysis day contains at least 5280 recordings. The statistics for posture duration and posture transitions, which are described below, were calculated using analysis days only. Thus, any patients who did not have any analysis days have been excluded from this analysis. Based on this criterion, 508 of 1198 patients (42.4%) were excluded. There are 690 patients included in this analysis.

For each patient, the posture duration statistics were defined as the average time spent in each of the 5 posture groups for the whole day (WD), daytime (DT), and nighttime (NT). Thus, there are 15 posture duration statistics from the 5 posture groups for WD, DT, and NT. The unit of measurement for the average time is hours per day. DT is defined as the 12-hour window from 7:00 AM-6:59 PM and NT as the opposite 12-hour window from 7:00 PM-6:59 AM. These definitions match the shift schedule at Atrium Health Wake Forest Baptist. The naming convention for the posture duration statistics is summarized in the first row of Table 2. The WD statistics provide the average number of hours per day that the patient spent in each posture. We calculated the WD statistics using the following procedure. Let k denote the number of analysis days for the patient (k≥1). Consider a posture group (eg, Lying). On each analysis day, we found the proportion of ViSi recordings that fell into the given posture. The average proportion across the k days was calculated, and the resulting value was multiplied by 24 to convert it to the unit of hours per day. This procedure was applied separately to each of the 5 posture groups, yielding a set of 5 WD statistics. Next, we calculated 2 more sets of statistics, one for DT and the other for NT. The purpose was to examine diurnal changes of the patient's posture pattern. To calculate the DT and NT statistics, we applied the same procedure described above to the data collected during the specific 12-hour time block, except changing the conversion factor from 24 to 12. In future work, DT and NT statistics, along with WD statistics, could be used in prediction models for patient outcomes.

Table . List of statistics for posture duration and posture transitions.

	Whole day (WD)	Daytime (DT)	Nighttime (NT)
Posture duration	• Lying-WD • Reclined-WD • Upright-WD • Unknown-WD • UserDefined-WD	• Lying-DT • Reclined-DT • Upright-DT • Unknown-DT • UserDefined-DT	• Lying-NT • Reclined-NT • Upright-NT • Unknown-NT • UserDefined-NT
Posture transitions	FPT-WD[a]	FPT-DT	FPT-NT

[a]FPT: frequency of posture transitions.

The third group, posture transitions, provides the frequency of posture transitions (FPT). We define that a posture transition occurs when a patient's posture group changes from one timestamp to the next. The change is also limited among Lying, Reclined, and Upright only. Based on this definition, there are 6 types of transitions: Lying to Reclined, Lying to Upright, Reclined to Upright, and vice versa. Neither Unknown nor User-defined are considered in posture transitions. This is because whether a physical transition occurs in this case is not clear. The FPT measures the rate of posture transitions including all 6 types described above, with the unit of transitions per hour (tph). As before, we calculated FPT for WD, DT, and NT separately. The naming convention of the statistics is shown in the bottom row of Table 2. The calculation of the FPT-WD statistic is as follows. For each analysis day of the patient, we tallied the number of transitions present and divided this count by the actual number of hours of data in that day. This provided a rate of tph for this specific analysis day. The FPT-WD is the average tph across all k analysis days. The calculation of FPT-DT (or FPT-NT) uses the same procedure, except that the number of transitions and their corresponding hours are from the DT (or NT) window. Overall, 3 statistics are obtained on FPT.

Cluster Analyses

We sought to identify the distinct subtypes of posture habits present in the dataset. For this purpose, a cluster analysis was conducted in which the patients were divided into nonoverlapping clusters, using the posture duration and posture transition statistics. These clusters can be used to identify the different subtypes of posture habits objectively. The cluster analysis was performed using the k-means clustering algorithm [16,17]. The number of clusters, k, must be specified a priori. We tested a range of values for k from 2 to 10 and selected the value of k that maximized the average silhouette width [18,19]. In total, 4 sets of cluster analyses were performed, with different input data. The first set, which we call WD cluster, used the WD statistics (Table 2, column 2). The second set (DT cluster) used the DT statistics (Table 2, column 3). The third set (NT cluster) used the NT statistics (Table 2, column 4). The fourth set (combined cluster) used the combined DT and NT statistics, which doubles the inputted data sample. In the 4 sets, we excluded posture duration statistics corresponding to the User-defined category because patients' values for these statistics were all approximately 0, so they were not useful for differentiating between patients. Before applying the k-means algorithm, each statistic was standardized to have mean 0 and SD 1. For each of the 4 cluster analyses, average silhouette scores were computed across all patients and separately within each of the k clusters.

Analyses were performed using the R programming language (version 4.2.2; R Foundation for Statistical Computing).

Results

Summary Statistics

Table 3 shows the age and sex distributions of the original 1330 patients before data exclusion, and the 690 patients after data exclusion. The age and sex distributions showed little change before versus after data exclusion (Table 3). The number of patients included in the analysis was 690, comprising 407 males and 283 females, with an average age of 60 years.

Table . Age and sex distributions before and after data exclusion.

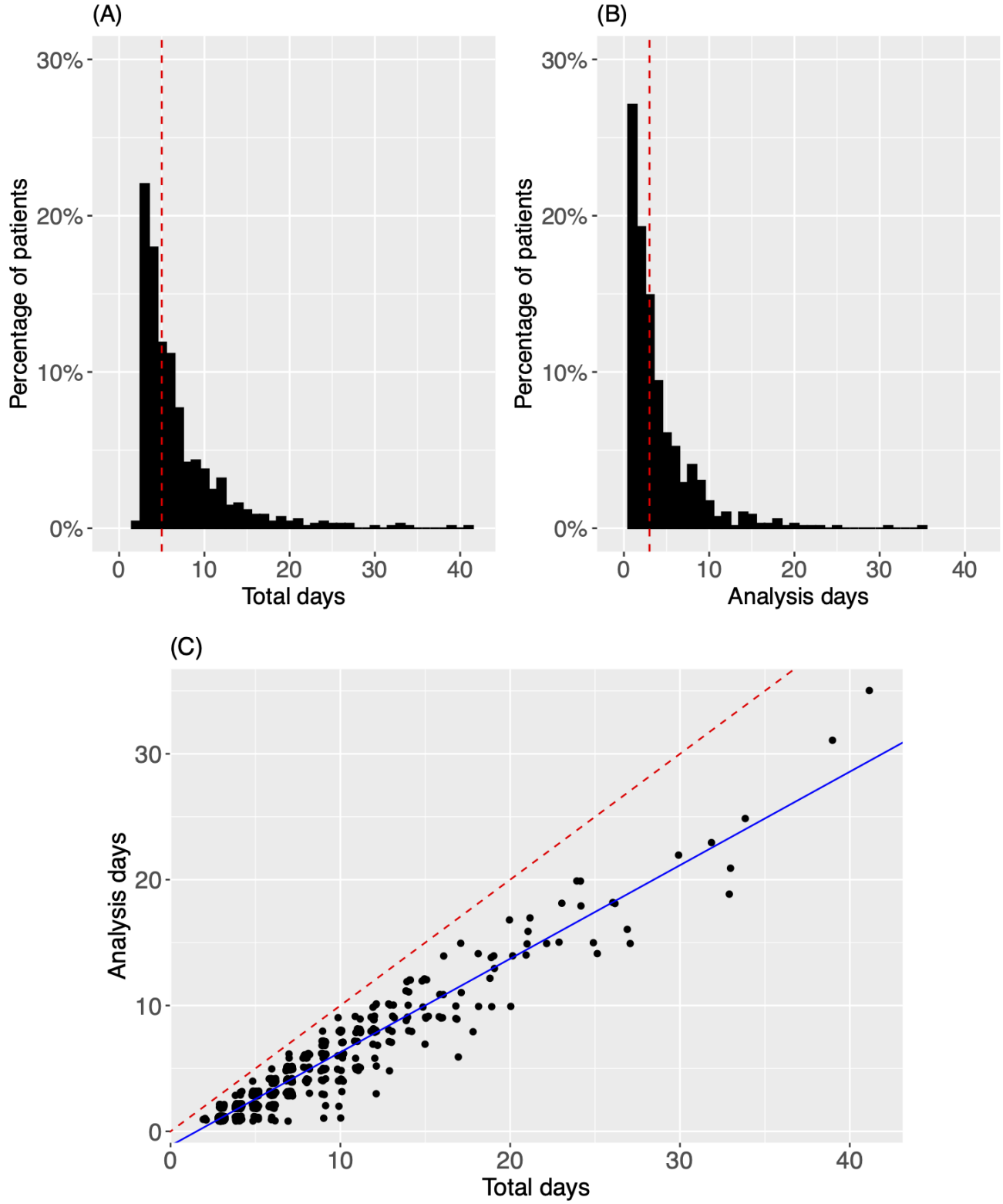
	Before data exclusion (n=1330)	After data exclusion (n=690)
Age [a] (years)		
Mean (SD)	60.5 (15.5)	60.3 (15.0)
Median (IQR)	62.5 (52.0-71.8)	62.4 (52.0-71.4)
Sex		
Male, n (%)	772 (58)	407 (59)
Female, n (%)	557 (41.9)	283 (41)
Unknown, n (%)	1 (0.1)	0 (0)

[a]Age is the subject's age on January 1, 2019.

The univariate distribution of each summary statistic was analyzed for the sample of 690 patients. Figure 3 shows the histograms of total days and analysis days, the summary statistics for data quantity. Both measures exhibit right-skewed distributions, with a higher range for total days than analysis days. The median (red dashed line) was 5 total days per subject and 3 analysis days per subject. The 2 measures are highly correlated with one another (Figure 3C). The scatter plot shows a linear relationship between total days and analysis days, estimated as:

$$\text{AnalysisDays} = -1.14 + 0.74\,\text{TotalDays}$$

Figure 3. Data quantity. (**A**) The histogram for total days per subject. (**B**) The histogram for analysis days. In both (A) and (B), the dotted red line is the median of the distribution. (**C**) A scatter plot of total days versus analysis days. A linear regression fit is overlaid on the points (solid blue). For comparison, the identity line (y=x) is provided (dotted red), which represents the ideal scenario that analysis days=total days. To allow visibility of overlapping points, random noise from a uniform distribution has been added to each point in the horizontal and vertical directions.



As described in the Methods section, only analysis days were considered when calculating the other summary statistics reported hereafter. Approximately 85% (586/690) of patients had between 1 and 7 analysis days, including 27% (187/690) with 1 analysis day, 34% (236/690) with 2 - 3, and 24% (163/690) with 4 - 7 analysis days. The other 15% (104/690) had more than 7 analysis days. The maximum number of analysis days was 35.

Figure 4 presents histograms of the WD statistics for posture duration, including Lying, Upright, Reclined, and Unknown. Table 4 provides the quartiles of the posture duration statistics. The histogram of User-defined is not presented in Figure 4 since

it is concentrated as an isolated peak at near 0 hours/day, consistent with the median of 0.008 hours/day (Table 4). This indicates a very small fraction of User-defined posture among all patients. The distribution of Lying-WD (Figure 4A) is approximately uniform from 2-20 hours near the 9% level, with a decline after 20 hours to 2.5%. There is also a peak from 0-2 hours at 15%. This shows a wide range of time duration spent in Lying across patients. The median for Lying-WD is 10.2 hours/day. Reclined-WD shows a roughly uniform distribution from 0-14 hours at around 12%. The distribution then tapers off on the right side (Figure 4B). The median is 8.0 hours/day (Table 4). In contrast, the distribution of Upright-WD (Figure 4C) is highly right skewed, with a peak at 0 - 2 hours/day accounting for about 45% (305/690) of patients. The interval from 2 - 4 hours/day includes 25% (175/690) of patients. Thus, 70% (480/690) of patients spent between 0 - 4 hours/day Upright. The median is 2.5 hours/day. Unknown-WD (Figure 4D) also shows a right-skewed distribution but with a sharp peak at 0 - 2 hours, accounting for about 80% (548/690) of patients. The median is 0.6 hours/day. This suggests that only a small fraction of patients were characterized in the Unknown category.

Figure 4. Posture duration for whole day. The histograms of the posture duration statistics are shown for whole day. The columns indicate the posture: (A) Lying, (B) Reclined, (C) Upright, and (D) Unknown. User-defined is not shown because this category was rarely observed. In each histogram, the x-axis ranges from 0 to 24 hours, and the y-axis shows the percentage of patients. The red dotted line indicates the median.

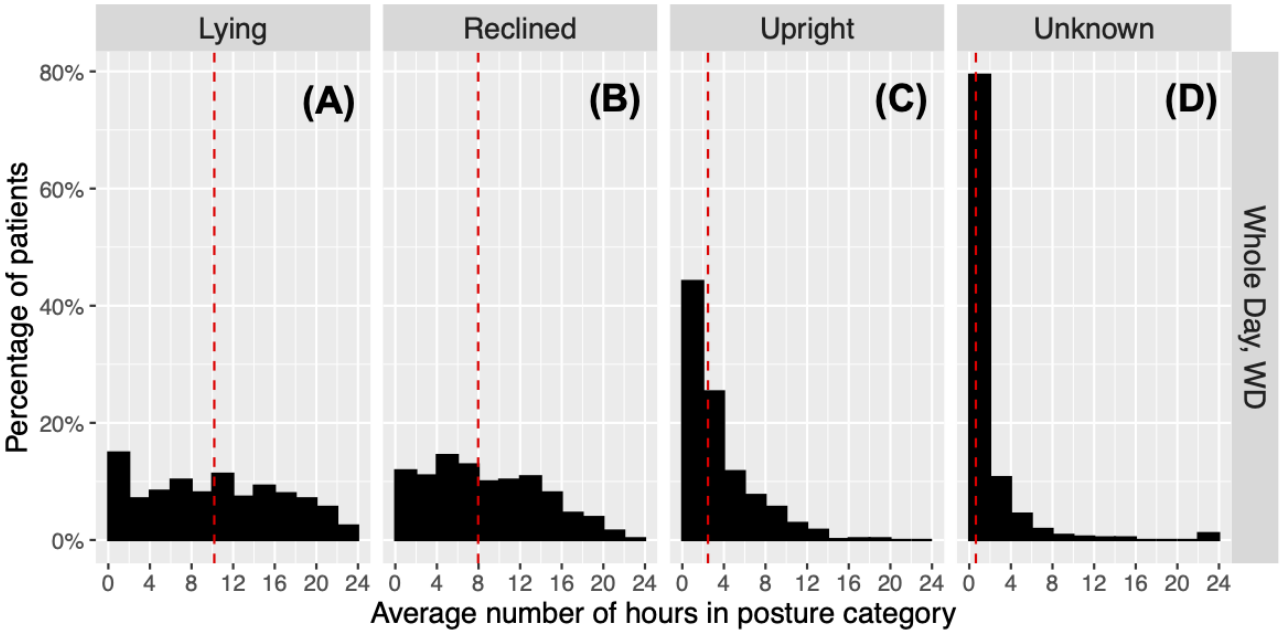


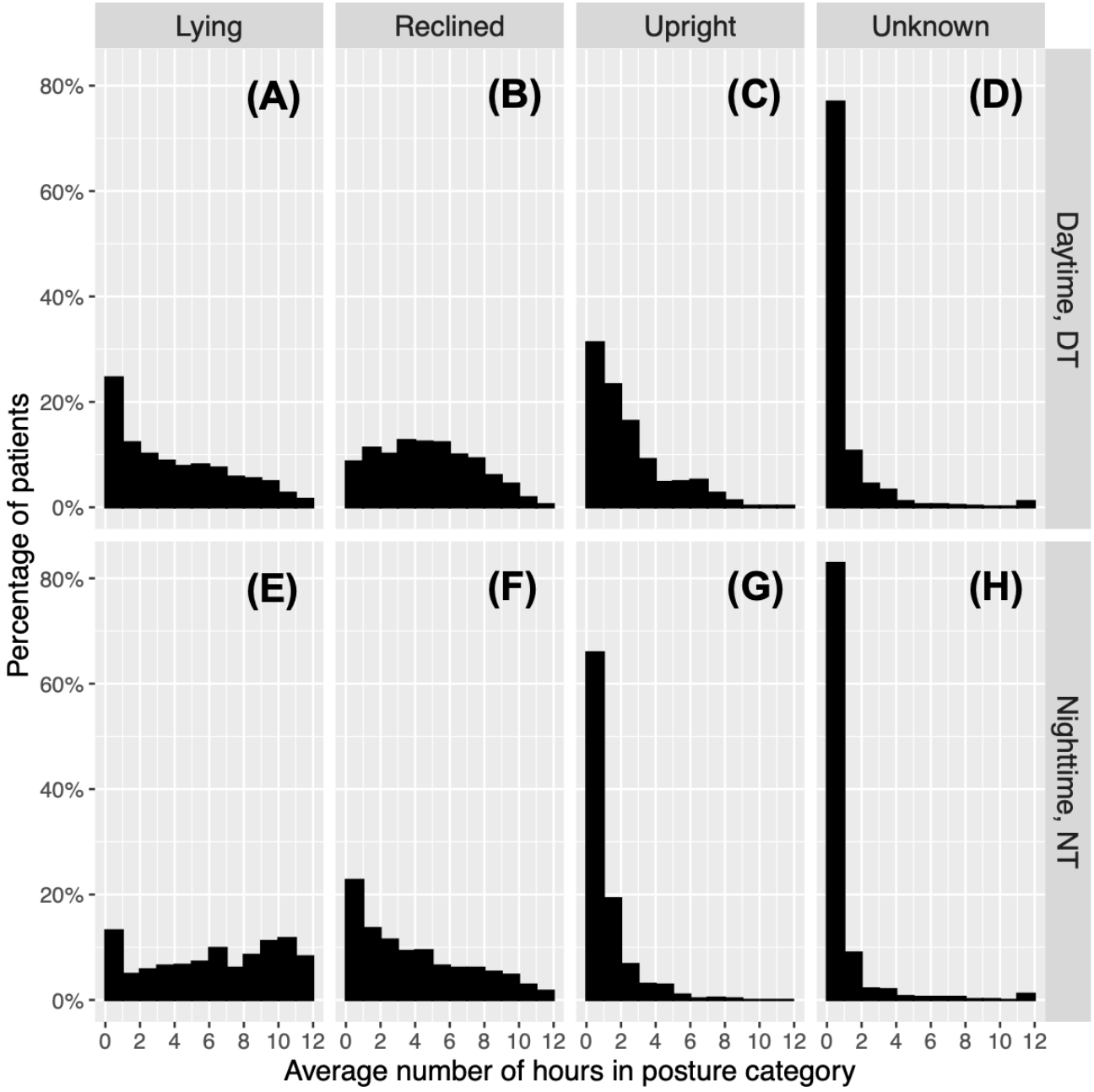
Table . Quartiles for posture duration and posture transition statistics.

	Posture duration					Frequency of posture transitions (transitions/hour)
	Lying (hours/day)	Reclined (hours/day)	Upright (hours/day)	Unknown (hours/day)	User-defined (hours/day)	
Whole day, median (IQR)	10.2 (4.5 - 15.6)	8.0 (4.1 - 12.7)	2.5 (1.1 - 4.7)	0.6 (0.2 - 1.6)	0.008 (0.007 - 0.009)	2.8 (1.9 - 4.3)
Daytime, median (IQR)	3.3 (1.1 - 6.4)	4.5 (2.4 - 6.7)	1.7 (0.8 - 3.4)	0.3 (0.1 - 0.9)	0.004 (0.004 - 0.005)	3.5 (2.3 - 5.3)
Nighttime, median (IQR)	6.6 (3.1 - 9.6)	3.3 (1.1 - 6.3)	0.6 (0.3 - 1.3)	0.2 (0.04 - 0.6)	0.004 (0.003 - 0.004)	2.1 (1.3 - 3.5)

Figure 5 presents histograms of the DT and NT statistics for posture duration. Table 4 shows the quartiles of these statistics. For the Lying posture, both the DT and NT distributions were substantially different from the pattern of the WD distribution, signifying the diurnal changes of this posture. The DT distribution (Figure 5A) peaked at 0 hours at the 25% level, and then decreased monotonically from 0-12 hours. In contrast, the NT histogram (Figure 5E) increased from 1-11 hours. As expected, there is a substantially higher percentage of patients with longer Lying times during NT than during DT. Consequently, the median for NT (6.6 hours) is higher than that for DT (3.3 hours), as shown in Table 4. However, there is a noticeable peak at 0 - 1 hours for NT, comprising about 13% (91/690) of patients. The Reclined distributions showed a diurnal change from roughly symmetric in DT (Figure 5B) to strongly right skewed in NT (Figure 5F). Although they had different shapes in their distribution, the median of DT (4.5 hours) was only slightly higher than that of NT (3.3 hours). The diurnal change could be explained by a certain percentage of patients changing their posture from Reclined in DT to Lying in NT,

but there remained some patients who continued in Reclined from DT to NT. For Upright, the distributions for DT and NT, like that of WD, were right skewed with a peak starting at the left end. The distribution for DT has its peak of 31% at 0 - 1 hours, followed by a slow decrease (Figure 5C). In comparison, the distribution of NT started from a higher peak of 66% at 0 - 1 hours, followed by a sharp decrease (Figure 5G). A noticeable proportion of people spent a substantial amount of time Upright in DT. For example, 18% (123/690) of patients spent between 4 - 8 hours Upright in DT, compared to 5% (32/690) in NT. The medians for DT and NT were 1.7 hours and 0.6 hours, respectively. The Unknown posture did not show a significant diurnal change.

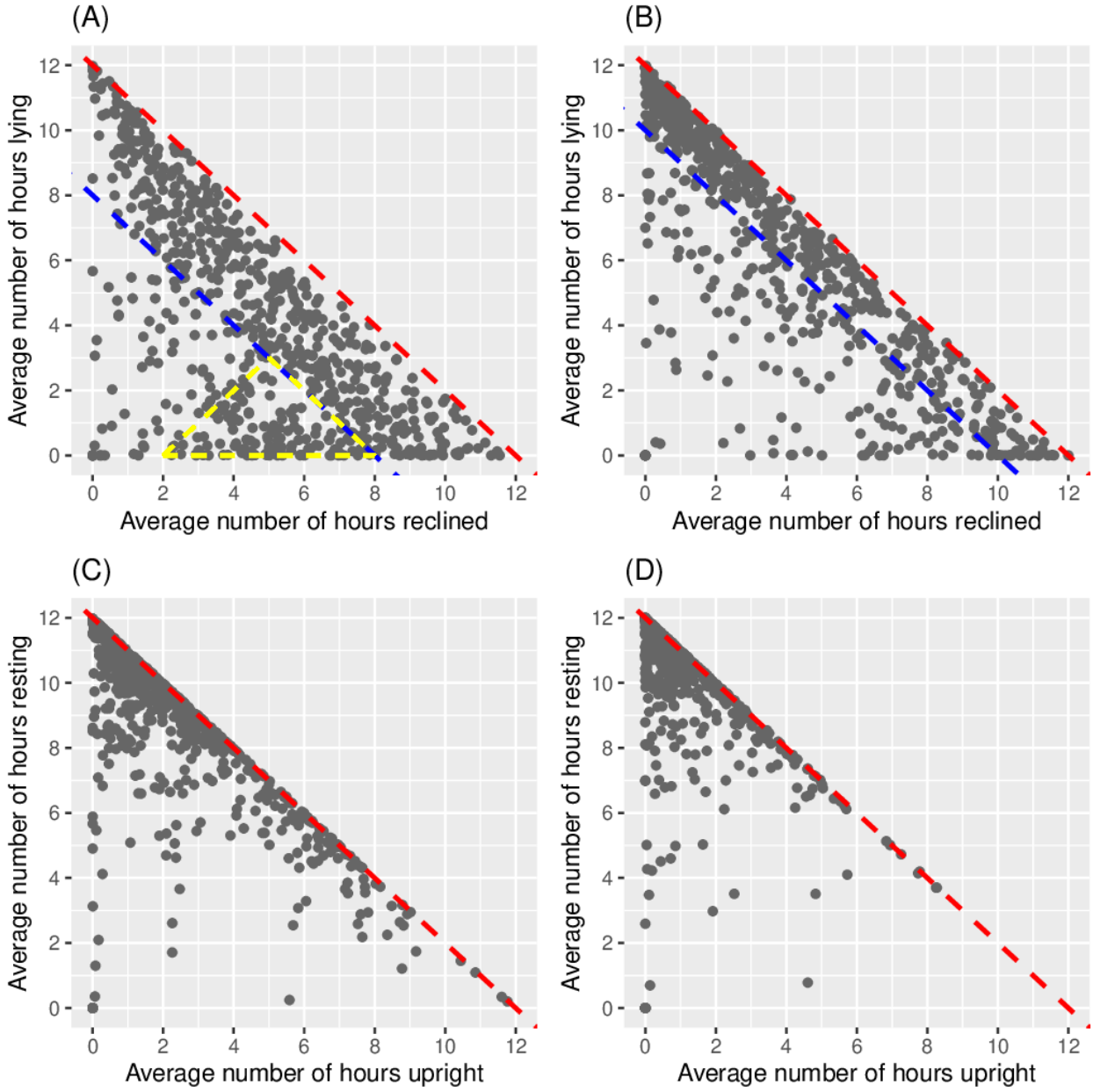
Figure 5. Posture duration for daytime and nighttime. This figure presents histograms of the posture duration statistics for (A-D) DT and (E-H) NT. The rows give the time of day as DT or NT. The columns indicate the posture (Lying, Reclined, Upright, Unknown). For any given row and column pair, the histogram shows the distribution of the average number of hours in the given posture during the given time of day, across the 690 patients. The x-axis of each histogram ranges from 0-12 hours. DT: daytime; NT: nighttime.



Interconnections between the summary statistics were also studied. Figure 6A presents a scatter plot of Reclined-DT versus Lying-DT. All points are contained inside the lower triangle of the square plot because the length of daytime is 12 hours and thus Lying-DT + Reclined-DT ≤ 12. The red dotted line (Lying-DT + Reclined-DT = 12) is the upper boundary of the triangle. Inside the triangle, points are spread out across the domain, but with a concentration at the strip between the upper boundary and lower boundary (blue dotted line) at Lying-DT + Reclined-DT = 8. This strip represents the subgroup that spends at least 8 hours Lying or Reclined during daytime. Another subarea of concentrated samples is the yellow triangle with vertices at (5, 3), (2, 0), and (8, 0). This represents a subgroup that mainly takes the Reclined posture for rest. Their

total resting time is less than those in the strip. They are likely the group with more activities in daytime. Last, the subgroup in the lower left corner is sparser, with only isolated points scattered around. Figure 6B gives the scatter plot of Reclined-NT versus Lying-NT, also with a red dotted line at the upper boundary. It shows a marked change in the relationship between these two postures from daytime to nighttime. Specifically, the strip near the upper boundary of the triangle described in daytime (Figure 6A) narrows during nighttime (Figure 6B), with the lower boundary moving upward to 10 hours (blue dotted line). Most of the study population was concentrated in this strip, and the triangle mentioned previously disappeared. This is consistent with the general behavior that most people rest during the night.

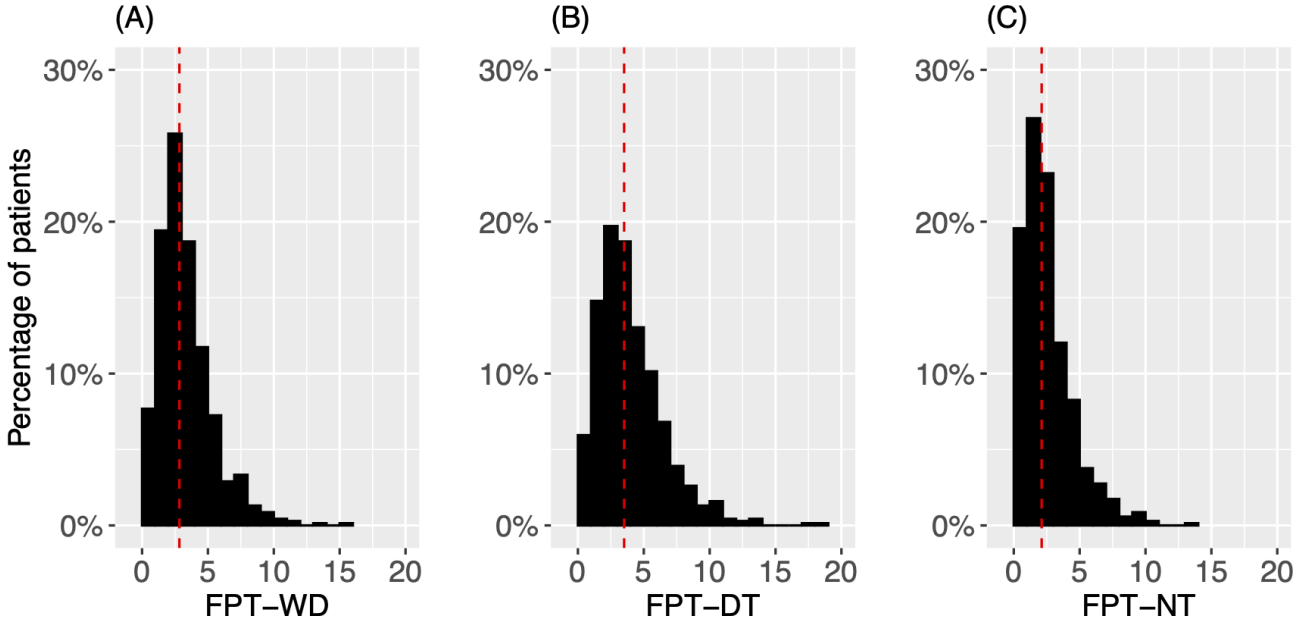
Figure 6. (A) Daytime: Reclined versus Lying. (B) Nighttime: Reclined versus Lying. (C) Daytime: Upright versus Rest. (D) Nighttime: Upright versus Rest. Scatter plots of time spent Lying versus Reclined (first row) and time spent Upright versus in Rest (second row). The left (right) column is for daytime (nighttime). Rest is defined as the sum of Lying and Reclined.



For a more simplified view, Reclined and Lying can be combined into a single category called Rest, defined as the sum of these 2 postures. Figure 6C and D show the relationship between Rest and Upright in DT and NT, respectively. Most of the points are near the upper boundary (red dotted line) because patients spend most of the time in either Rest or Upright postures. In both plots, points are concentrated in the upper left corner of the triangle. The majority are in an Upright posture between 0 - 4 hours and at Rest between 8 - 12 hours. A diurnal change occurs since there is a concentration of points near the upper boundary with Upright between 4 - 8 hours in DT (Figure 6C), while this largely disappears at NT (Figure 6D). In both plots, a small subgroup is well below the upper boundary. These points correspond to patients with higher levels of Unknown, previously shown in the right tail of Figure 5D and H.

Posture transitions were analyzed in addition to posture duration. Figure 7 shows histograms of the FPT for WD, DT, and NT. Table 4 provides the quartiles of each FPT statistic. The distribution of FPT-WD showed a prominent peak at 2 transitions per hour, accounting for 26.7% (184/690) of patients (Figure 7A). The distribution decayed quickly on both sides but with a skew to the right, shown by the long right tail extending to 15 transitions per hour. The distributions of FPT-DT (Figure 7B) and FPT-NT (Figure 7C) were also right-skewed, but showed different characteristics, reflecting a distinct diurnal change of FPT. In comparison to DT, NT had a sharper peak near 27% shifted closer to 0. About 50% (345/690) of people had 1 or 2 transitions per hour in NT. For DT, the peak was at a lower level of 20% and it decayed slower on both sides. About 35% (233/690) of people had 1 or 2 transitions per hour in DT. Moreover, the percentage of patients with higher transition rates (5 - 10 transitions per hour) decreased substantially from DT to NT, demonstrated by the higher weight in the range from 5 to 10 transitions per hour in DT. Both factors contributed to the higher median for FPT-DT compared to FPT-NT (Table 4).

Figure 7. Posture transitions. (A) Histogram for FPT-WD. (B) Histogram for FPT-DT. (C) Histogram for FPT-NT. In each plot, the dotted red line indicates the median for the corresponding statistic. The unit of measurement for the FPT variable is transitions per hour. DT: daytime; FPT: frequency of posture transitions; NT: nighttime; WD: whole day.



Cluster Analyses

Each of the four cluster analyses divided the 690 patients into a small number of clusters. Using the criterion of average silhouette width, the number of clusters k was selected to be 5 for DT, and 4 for WD, NT, and combined. Table 5 gives the cluster averages and number of patients per cluster for WD, DT, and NT. Multimedia Appendix 2 provides the average silhouette scores across all patients and within each cluster, individually for each cluster analysis.

Table . Cluster analyses using the k-means clustering algorithm for WD cluster, DT cluster, and NT cluster. The first k rows for each cluster analysis show the k clusters in order from the largest cluster to the smallest cluster. The last column indicates the number and percentage of patients assigned to the given cluster. The other columns provide the cluster averages for the summary statistics. The last row for each cluster analysis shows the average values across all 690 patients to provide a reference point for comparing the cluster-specific means.

	Lying (hours)	Reclined (hours)	Upright (hours)	Unknown (hours)	Frequency of posture transitions (transitions/hour)	Participants (n=690), n (%)
WD cluster						
W1	16.7	4.3	1.7	1.3	2.5	287 (42)
W2	5.3	14.7	2.9	1.1	2.8	219 (32)
W3	6.3	9.0	7.6	1.1	5.7	161 (23)
W4	3.5	2.9	1.6	16.0	1.3	23 (3)
Average	10.2	8.6	3.5	1.7	3.3	—[a]
DT cluster						
D1	2.2	7.4	1.8	0.6	3.5	241 (35)
D2	7.8	2.4	1.0	0.7	3.0	226 (33)
D3	2.8	5.0	3.6	0.5	8.3	99 (14)
D4	1.0	3.8	6.7	0.6	4.1	95 (14)
D5	1.7	1.5	0.9	7.9	1.1	29 (4)
Average	3.9	4.7	2.4	0.9	4.0	—
NT cluster						
N1	9.2	1.8	0.5	0.4	1.9	358 (52)
N2	2.6	8.0	0.9	0.5	2.3	190 (28)
N3	3.9	4.6	3.1	0.4	5.4	116 (17)
N4	2.2	1.2	0.6	8.0	1.2	26 (4)
Average	6.3	4.0	1.1	0.7	2.6	—

[a]Not applicable.

In the WD cluster (Table 5), the two largest clusters (W1 and W2) spend most of the day in Rest and have low Upright time. W1 prefers the Lying posture for Rest, while W2 prefers Reclined. In both clusters, the average FPT is between 2 - 3 transitions per hour. On the other hand, W3 is the most active, with the highest average Upright time (7.6 hours) and posture transition frequency (5.7 transitions per hour) among the clusters. W4 is the smallest cluster, comprising 3% (23/690) of the patients. Unknown is the dominant category in W4, with 16 hours/day Unknown on average.

The DT cluster (Table 5) identified 5 clusters using the DT statistics. Clusters D1 and D2 spend the daytime hours in Rest. D1 takes Rest mostly through Reclined at 7.4 hours on average, while D2 takes Rest mostly through Lying at 7.8 hours on average. D1 and D2 both have low levels of Upright, at 1.8 and 1.0 hours on average, respectively. D1 and D2 are close in size, accounting for 35% (241/690) and 33% (226/690) of patients, respectively. In comparison, D3 and D4 are more active during DT. D3 has an average Upright time of 3.6 hours, and it prefers Reclined (5.0 hours) over Lying (2.8 hours). D4 has a higher average Upright time (6.7 hours). It also spends more time in Reclined (3.8 hours) than Lying (1.0 hours). D3 and D4 each account for 14% of patients (99/690 and 95/690, respectively). D5 has Unknown as its dominant category and includes 4% (29/690) of patients.

The NT cluster (Table 5) found 4 clusters using the NT statistics. N1 and N2 are the two largest clusters, accounting for 52% (358/690) and 28% (190/690) of the patients, respectively. N1 has Lying as the dominant nighttime posture at 9.2 hours on average, followed by Reclined (1.8 hours). N2 prefers Reclined (8.0 hours) over Lying (2.6 hours) at nighttime. N3 has an average Upright time of 3.1 hours, which is higher than the other clusters. It also has a higher average FPT of 5.4 transitions per hour. N3 accounts for 17% (116/690) of patients. N4 has mostly Unknown posture of 8.0 hours on average.

In the combined cluster (Table 6), the traits of the clusters resemble those from the WD cluster. C1 prefers Lying, C2 favors Reclined, C3 is active in daytime, and C4 is dominated by Unknown. The sizes of C1, C2, C3, and C4 are also similar to the sizes of W1, W2, W3, and W4, respectively. Unlike the WD cluster, the combined cluster analysis revealed further details about changes in the patients' behaviors from daytime to nighttime. For example, cluster C1 spends most of the daytime in Lying (7.0 hours), followed by Reclined (2.9 hours) then Upright (1.3 hours). At nighttime, Lying becomes more

dominant (9.6 hours), with less time spent in Reclined (1.4 hours) and Upright (0.5 hours).

Table . Cluster analyses using the k-means clustering algorithm for the combined cluster.

	Lying (hours)		Reclined (hours)		Upright (hours)		Unknown (hours)		Frequency of posture transitions (transitions/hour)		Participants (n=690), n (%)
	DT[a]	NT[b]	DT	NT	DT	NT	DT	NT	DT	NT	
C1	7	9.6	2.9	1.4	1.3	0.5	0.8	0.5	3.2	1.8	291 (42)
C2	1.6	3.7	7.4	7	2.4	0.8	0.6	0.5	3.6	2.2	235 (34)
C3	1.9	4.2	4.2	4.6	5.2	2.8	0.6	0.4	6.8	5.1	141 (20)
C4	1.4	2.1	1.8	1.1	1.1	0.5	7.7	8.3	1.4	1.2	23 (3)
Average	3.9	6.3	4.7	4	2.4	1.1	0.9	0.7	4	2.6	—[c]

[a]DT: daytime.

[b]NT: nighttime.

[c]Not applicable.

Discussion

Principal Results

This study examined a year-long, routine hospital ViSi posture dataset to study patterns in patient posture habits, with the following results. The dataset required substantial preprocessing before statistical analysis. Two errors in the dataset were identified: overlap and inconsistency. Overlap was highlighted because it mixes data from multiple patients, which can alter the derived summary statistics. This problem could be easily overlooked in analyses of massive amounts of data. Overlap may have resulted from shared use of a ViSi monitoring device between multiple patients and lack of reliable timestamps of device handoff. In data preprocessing, we also consolidated the ViSi posture codes to target the most meaningful postures of Lying, Reclined, and Upright, and removed high frequency noise from the posture recordings. The preprocessing treatments improved the data quality, manageability of the data, and interpretation of results.

Using the preprocessed dataset, summary statistics were generated for each patient and probability distributions of the summary statistics were constructed using all 690 patients. This analysis showed that the 3 main postures (Lying, Reclined, Upright) had distinctive distribution patterns. For the whole day statistics, Lying had a flatter distribution, while Reclined showed a decrease in the longer hours. Upright had its peak at 0 - 2 hours, with a quick monotonic decrease. Thus, Lying and Reclined are more dominant postures. These findings are consistent with known clinical observations of inpatients. Pattern changes from daytime to nighttime were observed for Lying, Reclined, and Upright. The distribution of Lying was decreasing during daytime but relatively flat with a slow increase during nighttime. Although this pattern change may be partly explained by the fact that more patients take longer Lying periods during night, the flat distribution during nighttime still needs explanation. Combined with the similarity of the daytime Lying and nighttime Reclined, we speculate that a certain number of patients take Reclined as their resting posture at night. It is also noted that over one-third of patients show a preferred duration of 3 - 6 hours for Reclined posture during daytime. As expected, Upright shows a sharper decrease from daytime to nighttime.

Last, *k*-means clustering analyses were performed to objectively identify subtypes of patient posture habits. We found that 3 of the 4 analyses each identified 4 unique subtypes. These subtypes included 2 sedentary subgroups with different resting styles, one that prefers Lying and the other Reclined. There is an active subgroup that spends relatively more time Upright and has more frequent posture transitions. The fourth subtype includes a small percentage of patients that have Unknown as the dominant ViSi-measured posture during their hospital stay.

Limitations

This study has the following limitations. First, there may still be some irreversible errors left in the data (eg, incorrect placement of ViSi device on the body, battery failure causing discontinuity), which are beyond the scope of this analysis. The postures from the ViSi device can also differ from the ground-truth postures because the ViSi postures are estimated from accelerometer data. We did our best to mitigate the issue of inaccurate posture recognition. For example, because ViSi may classify walking as U90 instead of WLK, the activities of walking, standing, and sitting were combined into the same category of Upright [3]. The difficulty in classifying walking is possibly because the ViSi algorithm, previously trained from certain data, may not fit individual patients' mobility well. This implies that the ViSi algorithm needs to be calibrated to the individual patient, or more customized algorithms can be used to improve the accuracy of walking (eg, [20]). Second, our study does not consider postoperative pain or postoperative medications, but these two factors could affect a patient's posture after surgery. An area of future work is to examine the effect of postoperative pain and medications on ViSi posture statistics for various types of surgery.

Third, the amount of data excluded due to overlap or inconsistency was substantial, accounting for 9.9% (132/1330) of patients and 27.6% (8,778,641/31,842,773) of posture recordings. Our study took a conservative approach of dropping

suspicious data that exhibited the overlap or inconsistency problems. Although this reduced the available sample size, the number of patients included in the analysis remained high (n=690). We believe that the overlap and inconsistency problems occur randomly instead of systematically for certain types of patients. Thus, the data cleaning procedure, which removed these problems, should not cause bias. The data problems that we faced are associated with data collection, storage, and integration with the medical record. The current ViSi system is designed to provide real-time data for the nurse and clinician, rather than for long-term use. Therefore, data cleaning is an inevitable step for retrospective analyses of routinely collected ViSi data. However, some further improvements of the system could make the data more accessible for retrospective analyses. This could be done by creating standardized data models and specifications for measuring patient movement.

Conclusions

Using ViSi posture data collected from routine inpatient monitoring, we derived summary statistics for each patient and analyzed the summary statistics to find patterns in the patient population. Probability distributions, generated from the large sample size of 690 patients, provided quantitative measures of posture durations and transitions, with respect to their median and interquartile range. We considered both surgery and nonsurgery patients in the dataset to provide a baseline measure of these quantities, which sets a basis to study more specific clinical situations. In future work, we will build on this work to study distributions for patients with specific types of disease, as well as distributions of patients before versus after surgery. After these distributions are established, they can serve as useful reference distributions to measure a future patient's recovery trajectory or condition.

ViSi data collection increases work efficiency in the hospital. For the 690 patients analyzed in this study, the cumulative hours of ViSi posture collection amounted to 90,308 hours taken over 4847 patient-days. As an alternative, direct observation requires a nurse spending 2 hours per patient-day to conduct posture monitoring and documentation at a 30-minute frequency. Therefore, ViSi monitoring saved 9694 hours (4847 × 2) of nurses' effort, which could be devoted to other more pressing clinical responsibilities. This corresponds to a saving of labor cost of US $401,138, based on the median hourly rate of US $41.38 per hour for registered nurses, according to the US Bureau of Labor Statistics [21]. Moreover, the digitized inpatient measurements with automated collection and storage provide a vast and valuable database for retrospective analysis, which has not been fully tapped by clinical researchers.

A major contribution of this study is development of methodology to clean and preprocess routinely collected inpatient ViSi monitoring data. The procedure we used for cleaning the ViSi posture data can have broader application in cleaning other routine hospital datasets. The preprocessing treatments for consolidating posture codes and removing high-frequency noise can also improve manageability and interpretability of ViSi posture data. Thus, this study lays groundwork for future studies on examining the associations of clinical data with posture statistics derived from cleaned, preprocessed ViSi data. In future studies, we plan to study the associations of patients' medical condition and treatment (including diagnosis, surgical procedure, medications, sedatives, and pain level) with posture duration and posture transitions. Based on our experience with ViSi analysis, we believe that examining data quality will be the first step to ensure that clinical variables are accurately linked to the cleaned ViSi data records for each patient.

Data Availability

The data are available upon reasonable request.

Authors' Contributions

EJH and CJC designed the study questions and formed the analysis plan. EJH and YC conducted the data analysis. CJC provided clinical background for the data analysis. EJH and YC prepared the original draft of the manuscript. EJH, YC, and CJC reviewed and edited the manuscript before submission.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Posture codes of ViSi Mobile System.
[PDF File, 61 KB - mhealth_v12i1e54735_app1.pdf]

Multimedia Appendix 2
Table of average silhouette scores for whole day, daytime, nighttime, and combined cluster analyses. For each cluster analysis, the average silhouette score across all patients is shown. In addition, the average silhouette score is provided separately for each cluster.
[PDF File, 77 KB - mhealth_v12i1e54735_app2.pdf]

References

1. Rivas E, Cohen B, Pu X, et al. Pain and opioid consumption and mobilization after surgery: post hoc analysis of two randomized trials. Anesthesiology 2022 Jan 1;136(1):115-126. [doi: 10.1097/ALN.0000000000004037] [Medline: 34780602]
2. Ljungqvist O, Scott M, Fearon KC. Enhanced recovery after surgery: a review. JAMA Surg 2017 Mar 1;152(3):292-298. [doi: 10.1001/jamasurg.2016.4952] [Medline: 28097305]
3. Restrepo A, Saha AK, Khanna AK, Huang E, Clark CJ. Use of a multi-sensor monitoring device in an early post-operative mobilization program. Am Surg 2022 Aug;88(8):1861-1867. [doi: 10.1177/00031348221087196] [Medline: 35430918]
4. Patel V, Orchanian-Cheff A, Wu R. Evaluating the validity and utility of wearable technology for continuously monitoring patients in a hospital setting: systematic review. JMIR mHealth uHealth 2021 Aug 18;9(8):e17411. [doi: 10.2196/17411] [Medline: 34406121]
5. Baldwin C, van Kessel G, Phillips A, Johnston K. Accelerometry shows inpatients with acute medical or surgical conditions spend little time upright and are highly sedentary: systematic review. Phys Ther 2017 Nov 1;97(11):1044-1065. [doi: 10.1093/ptj/pzx076] [Medline: 29077906]
6. Straczkiewicz M, Huang EJ, Onnela JP. A “one-size-fits-most” walking recognition method for smartphones, smartwatches, and wearable accelerometers. NPJ Digit Med 2023 Feb 23;6(1):29. [doi: 10.1038/s41746-022-00745-z] [Medline: 36823348]
7. Dorrell RD, Vermillion SA, Clark CJ. Feasibility of real-time location systems in monitoring recovery after major abdominal surgery. Surg Endosc 2017 Dec;31(12):5457-5462. [doi: 10.1007/s00464-017-5625-7] [Medline: 28593407]
8. Theou O, Kehler DS, Godin J, Mallery K, MacLean MA, Rockwood K. Upright time during hospitalization for older inpatients: a prospective cohort study. Exp Gerontol 2019 Oct 15;126:110681. [doi: 10.1016/j.exger.2019.110681] [Medline: 31382011]
9. Patient Vital Signs Monitoring System. Sotera Digital Health. URL: https://soteradigitalhealth.com [accessed 2023-10-30]
10. Watkins T, Whisman L, Booker P. Nursing assessment of continuous vital sign surveillance to improve patient safety on the medical/surgical unit. J Clin Nurs 2016 Jan;25(1-2):278-281. [doi: 10.1111/jocn.13102] [Medline: 26531215]
11. Weenk M, van Goor H, Frietman B, et al. Continuous monitoring of vital signs using wearable devices on the general ward: pilot study. JMIR mHealth uHealth 2017 Jul 5;5(7):e91. [doi: 10.2196/mhealth.7208] [Medline: 28679490]
12. Weenk M, Koeneman M, van de Belt TH, Engelen L, van Goor H, Bredie SJH. Wireless and continuous monitoring of vital signs in patients at the general ward. Resuscitation 2019 Mar;136:47-53. [doi: 10.1016/j.resuscitation.2019.01.017] [Medline: 30685546]
13. Turan A, Khanna AK, Brooker J, et al. Association between mobilization and composite postoperative complications following major elective surgery. JAMA Surg 2023 Aug 1;158(8):825-830. [doi: 10.1001/jamasurg.2023.1122] [Medline: 37256591]
14. Dhillon MS, McCombie SA, McCombie DB. Towards the prevention of pressure ulcers with a wearable patient posture monitor based on adaptive accelerometer alignment. Annu Int Conf IEEE Eng Med Biol Soc 2012;2012:4513-4516. [doi: 10.1109/EMBC.2012.6346970] [Medline: 23366931]
15. Macaulay FR. Introduction to “the smoothing of time series.”. In: The Smoothing of Time Series: NBER; 1931:17-30 URL: https://www.nber.org/books-and-chapters/smoothing-time-series/introduction-smoothing-time-series [accessed 2023-11-05]
16. Hartigan JA, Wong MA. Algorithm AS 136: A K-Means Clustering Algorithm. Appl Stat 1979;28(1):100. [doi: 10.2307/2346830]
17. James G, Witten D, Hastie T, Tibshirani R. An Introduction to Statistical Learning: With Applications in R: Springer US; 2021. [doi: 10.1007/978-1-0716-1418-1]
18. Rousseeuw PJ. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J Comput Appl Math 1987 Nov;20:53-65. [doi: 10.1016/0377-0427(87)90125-7]
19. Kaufman L, Rousseeuw PJ. Finding Groups in Data: An Introduction to Cluster Analysis: John Wiley & Sons; 2009.
20. Huang EJ, Yan K, Onnela JP. Smartphone-based activity recognition using multistream movelets combining accelerometer and gyroscope data. Sensors (Basel) 2022 Mar 29;22(7):2618. [doi: 10.3390/s22072618] [Medline: 35408232]
21. Occupational Outlook Handbook, Registered Nurses. : U.S. Bureau of Labor Statistics URL: https://www.bls.gov/ooh/healthcare/registered-nurses.htm [accessed 2024-10-23]

Abbreviations

DT: daytime
FPT: frequency of posture transitions
LLS: lying left side
LPR: lying prone
LRS: lying right side
LSP: lying supine
MRN: medical record number
NT: nighttime
tph: transitions per hour

UNK: unknown
WD: whole day
WLK: walking

Edited by L Buis; submitted 21.11.23; peer-reviewed by E Rivas, R Zhang, S Okita; revised version received 27.05.24; accepted 01.10.24; published 06.11.24.

Please cite as:
Huang EJ, Chen Y, Clark CJ
Using a Quality-Controlled Dataset From ViSi Mobile Monitoring for Analyzing Posture Patterns of Hospitalized Patients: Retrospective Observational Study
JMIR Mhealth Uhealth 2024;12:e54735
URL: https://mhealth.jmir.org/2024/1/e54735
doi:10.2196/54735

A Novel Approach for Improving Gait Speed Estimation Using a Single Inertial Measurement Unit Embedded in a Smartphone: Validity and Reliability Study

Pei-An Lee[1,2,3], PhD; Wanting Yu[1], MEng; Junhong Zhou[1,2,3], PhD; Timothy Tsai[1], MPH; Brad Manor[1,2,3], PhD; On-Yee Lo[1,2,3], PhD

[1]Hinda and Arthur Marcus Institute for Aging Research, Hebrew SeniorLife, Boston, MA, United States

[2]Division of Gerontology, Beth Israel Deaconess Medical Center, Boston, MA, United States

[3]Harvard Medical School, Boston, MA, United States

Corresponding Author:
Pei-An Lee, PhD
Hinda and Arthur Marcus Institute for Aging Research, Hebrew SeniorLife, , Boston, MA, , United States

Abstract

Background: Gait speed is a valuable biomarker for mobility and overall health assessment. Existing methods to measure gait speed require expensive equipment or personnel assistance, limiting their use in unsupervised, daily-life conditions. The availability of smartphones equipped with a single inertial measurement unit (IMU) presents a viable and convenient method for measuring gait speed outside of laboratory and clinical settings. Previous works have used the inverted pendulum model to estimate gait speed using a non–smartphone-based IMU attached to the trunk. However, it is unclear whether and how this approach can estimate gait speed using the IMU embedded in a smartphone while being carried in a pants pocket during walking, especially under various walking conditions.

Objective: This study aimed to validate and test the reliability of a smartphone IMU–based gait speed measurement placed in the user's front pants pocket in both healthy young and older adults while walking quietly (ie, normal walking) and walking while conducting a cognitive task (ie, dual-task walking).

Methods: A custom-developed smartphone application (app) was used to record gait data from 12 young adults and 12 older adults during normal and dual-task walking. The validity and reliability of gait speed and step length estimations from the smartphone were compared with the gold standard GAITRite mat. A coefficient-based adjustment based upon a coefficient relative to the original estimation of step length was applied to improve the accuracy of gait speed estimation. The magnitude of error (ie, bias and limits of agreement) between the gait data from the smartphone and the GAITRite mat was calculated for each stride. The Passing-Bablok orthogonal regression model was used to provide agreement (ie, slopes and intercepts) between the smartphone and the GAITRite mat.

Results: The gait speed measured by the smartphone was valid when compared to the GAITRite mat. The original limits of agreement were 0.50 m/s (an ideal value of 0 m/s), and the orthogonal regression analysis indicated a slope of 1.68 (an ideal value of 1) and an intercept of −0.70 (an ideal value of 0). After adjustment, the accuracy of the smartphone-derived gait speed estimation improved, with limits of agreement reduced to 0.34 m/s. The adjusted slope improved to 1.00, with an intercept of 0.03. The test-retest reliability of smartphone-derived gait speed was good to excellent within supervised laboratory settings and unsupervised home conditions. The adjustment coefficients were applicable to a wide range of step lengths and gait speeds.

Conclusions: The inverted pendulum approach is a valid and reliable method for estimating gait speed from a smartphone IMU placed in the pockets of younger and older adults. Adjusting step length by a coefficient derived from the original estimation of step length successfully removed bias and improved the accuracy of gait speed estimation. This novel method has potential applications in various settings and populations, though fine-tuning may be necessary for specific data sets.

(JMIR Mhealth Uhealth 2024;12:e52166) doi:10.2196/52166

KEYWORDS

smartphone app; gait speed; dual-task walking; validity; reliability; mobile phone

Introduction

Gait speed, often considered the sixth vital sign, is a meaningful health indicator that is routinely assessed within clinical and research settings [1-4]. As a biomarker, gait speed is an integrated and reproducible measure of mobility that helps to identify the risk of falls, disability, hospitalization, and even mortality [5,6]. It is also responsive to intervention and is often used to evaluate the effectiveness and progression of rehabilitation programming [3,7]. Within clinical settings, gait speed is most commonly measured by using a stopwatch to record the time taken to walk a short distance [8]. This approach, while validated, requires another person's assistance and is often inaccurate due to human error or bias. Within the laboratory, gait speed is typically measured using motion capture systems or instrumented walkways. This approach increases measurement accuracy yet requires trained personnel and expensive equipment, limiting its accessibility and feasibility for frequent, longitudinal monitoring. Moreover, neither of these clinical or laboratory-based approaches is well-suited for monitoring changes over time via higher-frequency assessments, and neither approach affords the assessment of gait speed assessment under unsupervised conditions during daily life activities. There is thus a need to develop new methods for the low-cost, easy, and accurate measurement of gait speed for use within both clinical and remote, unsupervised settings.

Smartphones, now widely used by younger and older adults and equipped with a high-quality inertial measurement unit (IMU), present a viable and convenient opportunity to objectively measure gait speed outside of laboratory and clinical settings. Previous studies have demonstrated that gait speed can be accurately estimated using multiple IMUs attached to various parts of the body (eg, the trunk and ankles) [9-11]. However, this approach requires specialized equipment that may be difficult to don and doff. Alternatively, studies have attempted to estimate gait speed using a single, research-grade IMU attached to the trunk of the body. In this case, the inverted pendulum model was used to determine gait speed by estimating stride distance and dividing it by stride time, determined by acceleration peaks induced by consecutive heel strikes [12-15]. This approach, while promising, still requires trained personnel for equipment setup and may introduce bias in gait speed estimation, especially at a faster or slower gait speed [13-15]. In this study, we examined the validity and reliability of deriving gait speed from a smartphone IMU with the phone being carried in the individual's front pants pocket when walking.

This study aimed to validate and test the reliability of a smartphone IMU–based gait speed measurement placed in the user's pants pocket in both healthy young and older adults during normal and dual-task walking. The validity of the smartphone-based approach was determined by comparing the estimated step length and gait speed to measurements taken by gold standard instrumentation. Based on initial results, we further examined the use of improving gait speed estimations using a coefficient relative to the estimation of step length as determined by the smartphone. The test-retest reliability of adjusted and unadjusted estimations was examined within both laboratory and unsupervised, real-life home settings. This study is expected to provide valuable insights applicable to a wide range of clinical and everyday gait monitoring scenarios.

Methods

Study Participants

We conducted a design control verification and validation study enrolling younger (aged 20 - 50 years) and older (aged 65 - 90 years) adults between 2016 and 2018. As part of our comprehensive assessment at baseline, we measured key demographic characteristics of the participants, including sex, age (in years), height (in meters), body weight (in kilograms), and ethnicity. We included those who had active Wi-Fi service in their homes and who were able to use the smartphone app by themselves after training. Assistive devices were allowed if participants normally used them when walking. Individuals were excluded if they were hospitalized within the last 6 months; were unable to walk without assistance; self-reported major neuromuscular, cardiovascular, or metabolic disease; had lower-extremity ulcers or amputations; or self-reported pain significantly affecting their gait.

Ethics Approval

This study was approved by the Hebrew SeniorLife Institutional Review Board (approval IRB-2015 - 40).

Smartphone App

Our team previously created an iOS smartphone–based application (app) that uses the phone's IMU sensor to record movements while walking freely at a self-selected speed (ie, normal walking condition) and while walking and concurrently performing a serial-subtraction task (ie, dual-task walking condition) with the phone placed in the user's pants pocket [16]. The app was designed to recreate a commonly used dual-task gait assessment that is typically performed in a laboratory setting. The initial development process involved a detailed analysis of requirements, followed by iterative design and user feedback–driven enhancements, ensuring both user-friendliness and adherence to clinical standards. The app includes a series of instructions for the participants to help ensure assessment reliability. Once the participant presses the "Start" button and places the phone in their preferred front pocket, the app provides auditory instructions for each walking trial via the iPhone speaker, including a randomly generated starting number for the serial-subtraction task for dual-task trials, and cues for the start and end of the trial. These cues trigger the acquisition of 3-axis accelerometer, gyroscope, and magnetometer data at a 100-Hz sampling rate. The data sets are saved on the phone's internal storage and automatically transmitted via Wi-Fi to a remote, cloud-based data server for offline analysis. The validity and reliability of the smartphone app, especially in temporal information (ie, stride time), have been established [16-18]. Additionally, its applicability extends to various populations, including individuals with Parkinson disease and older adults diagnosed with blood cancer [18-21].

Study Procedures

Overview

Each participant completed 2 laboratory visits separated by at least 1 week. Between these visits, participants were asked to complete the walking assessment using the app at home once a day on 3 separate days. Participants were instructed to wear comfortable pants or shorts with front pockets for each visit. Our previous work indicated that stride time estimations were unaffected by pocket tightness [16].

Laboratory Assessments

The same procedures were used on both laboratory visits. Within each visit, participants completed the walking assessment using the app 3 times, that is, 3 pairs of normal and dual-task walking trials. For each walking trial, participants walked around an oval-shaped, 24-m indoor track. The GAITRite mat (CIR Systems, Inc) was placed along one long side of the track. Each trial began with participants standing just behind the beginning of the mat to ensure that the first footfall of each trial was captured by the mat. Participants were instructed to use the app to start and finish each trial. Stride time, step length, and gait speed obtained from the first pass over the GAITRite mat were considered the gold standard and aligned the data sets to enable direct comparison of GAITRite and smartphone app data for each identified step.

Home Assessments

At home, participants were instructed to use the app to complete 1 normal walking and 1 dual-task walking on 3 separate days in between the 2 laboratory visits. The app provided the same instructions as given during the laboratory visit. The participants were instructed to walk continuously along the longest hallway or unobstructed path in their home, making 180-degree turns at each end, throughout the trial. We believe this setup can be effective in remote settings, provided that a flat and consistent walking surface is available, external interference is minimal, and stable Wi-Fi is present.

Data Analysis

Gait Speed Estimation

All data analyses, graphics generation, and statistical analyses were performed using programs developed in-house within MATLAB (R2022b, MathWorks) and SPSS (version 20; SPSS Inc).

For the data obtained from the smartphone, the raw time series 3-axis accelerometer and gyroscope data were transformed from the smartphone local coordinate system to an earth global coordinate system using a quaternion rotation matrix. This transformation ensured that the z-axis of the data aligned with the vertical line of gravity. After the transformation, each z-axis time series was filtered using a Butterworth filter [16,22], chosen for its smooth frequency response and minimal signal distortion, which worked well in our previous published study [16]. The time series data obtained from the accelerometer and gyroscope sensors contained alternating peaks of high and low amplitude that corresponded to the heel-strike and toe-off events [16]. Stride time was defined as the time between 2 consecutive heel strikes of the same foot, which was calculated by determining the number of data points between 2 heel strikes and dividing by 100-Hz sampling frequency. The step length was estimated using a simple inverted pendulum model, which uses the participant's leg length (l) and the change in height of the smartphone's vertical position over each step (h) as the following equation:

$$(1)\text{Step length}=2*2*h*l-h2$$

In particular, the vertical position was derived by doubly integrating the vertical acceleration data (z-axis) and high-pass filtering the outcome using a fourth-order zero-lag Butterworth filter with a 0.11-Hz cutoff frequency to remove any integration drift [12,23]. The fourth order of the filter ensures effective attenuation of frequencies below 0.11 Hz, substantially reducing low-frequency drift caused by noise in the acceleration data. Additionally, the zero-lag design of the Butterworth filter preserves the temporal accuracy of the signal, ensuring precise alignment of the filtered output with the real-time events being measured. Gait speed for each stride was then calculated by dividing stride length (ie, step length multiplied by 2) by the corresponding stride time. Gait variables thus included step length and gait speed from both GAITRite mat and smartphone app data were used for the statistical analyses.

The Step Length Adjustment Process

After calculating the validity of the gait speed and step length derived from the smartphone app and GAITRite (see Statistical Analysis section), we observed a bias between smartphone app and GAITRite measurements for both gait speed and step length across all participants, walking conditions, and laboratory trials. That is, as the step length (average of the smartphone app and GAITRite mat measurement) increased, the step length derived from the app became increasingly greater than the corresponding values derived from the GAITRite mat. This trend was independent of the walking condition and study visit order (Figures 1A and 2A). To address this, we systematically tested various intervals, starting from 0.1, 0.15, 0.2 m, and all the way up to 0.5 m, with a 0.05-m increment, for adjusting the step length. Our analysis revealed that an interval of 0.3 m consistently produced the best validity, evidenced by a slope of 1 and an intercept close to 0 m in the relationship between our adjusting measurements and those from the GAITRite mat. These findings indicate a robust linear correlation and minimal systematic bias. This method was designed for its ability to adjust the step length estimation, aiming to align more closely with the gold standard GAITRite measurements, particularly addressing the identified bias across different step lengths. The intervals were chosen based on a detailed stepwise iterative approach, where each interval was evaluated for accuracy and precision. This approach allowed us to capture the nonlinear relationship between the app-derived and GAITRite-derived step lengths across the range of observed values (Table 1).

$$(2)\text{Adjustedsteplength}=\text{coefficients}*\text{steplength}$$

Figure 1. Bland-Altman plot for gait speed when using the smartphone app–based assessment compared to GAITRite mat before (A) and after (B) adjusting by coefficients relative to the original estimation of step length. The solid lines are the average difference, and the dashed lines are the limit of agreements (SD 1.96).

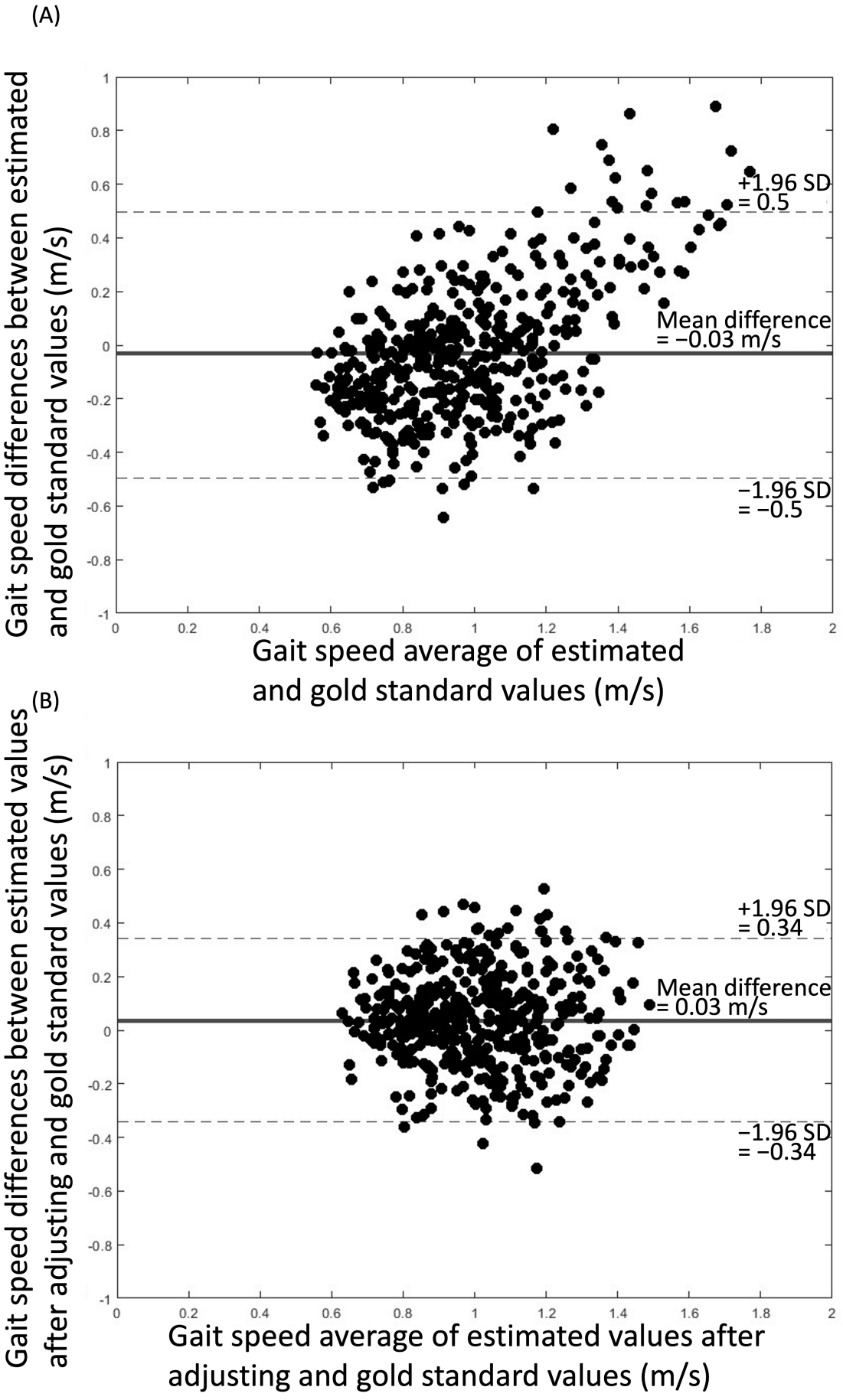


Figure 2. Bland-Altman plot for step length when using the smartphone app–based assessment compared to GAITRite mat before (A) and after (B) adjusting by coefficients relative to the original estimation of step length. The solid lines are the average difference, and the dashed lines are the limit of agreements (SD 1.96).

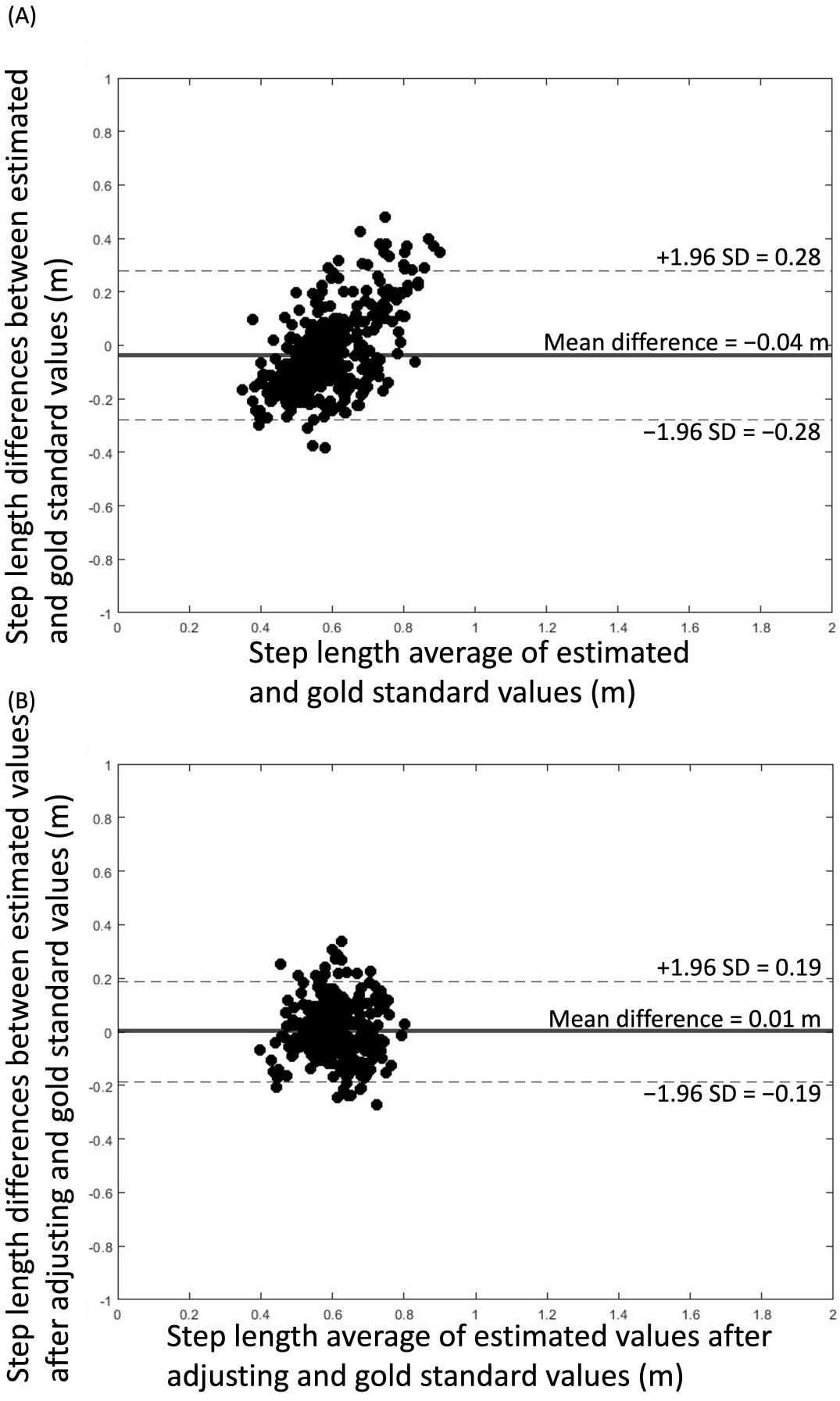


Table . Coefficients that were used to adjust step length and gait speed.

		Step length 0.2 - 0.5 m	Step length 0.5 - 0.8 m	Step length 0.8 - 1.1 m
Coefficients				
	All data points	1.37	1.02	0.74
	Only normal walking data points	1.40	1.04	0.74
	Only dual-task walking data points	1.36	1.01	0.73

Statistical Analysis

The validity of app-derived gait speed and step length was examined by estimating their agreement with corresponding gait data derived from the gold standard GAITRite mat using a Passing-Bablok orthogonal regression model, which is suitable for comparing different measurement methods while acknowledging measurement error [24]. This statistical method was chosen for its robustness in comparing methods with measurement error, providing a more accurate assessment of agreement between different measurement systems. Initial sources and seminal works on the Passing-Bablok regression model highlight its suitability for medical and biomechanical research applications [24]. The magnitude of error between the gait data from the app and the GAITRite mat was calculated for each stride, and a Bland-Altman plot was produced to visualize this error as a function of gait data, which was the average of the gait speed calculated by the app and the GAITRite mat. Test-retest reliability was assessed for all derived gait parameters from both the smartphone app and GAITRite mat using several intraclass correlation coefficients (ICCs). For each of the following two conditions, we computed ICC separately for normal and dual-task walking trials: (1) across trials within each laboratory assessment, and (2) across trials over the 3 home assessments. The unit of interest was the average gait speed in strides derived from each trial of the same condition (ie, normal or dual-task walking; ICC (1, 1)). We considered ICC values greater than 0.75 as excellent reliability, 0.6 - 0.75 as good reliability, 0.4 - 0.6 as fair reliability, and less than 0.4 as poor reliability [25].

Results

Overview

In total, 12 healthy young adults (6 female adults and 6 male adults; age: mean 29.1, SD 4.4 years; height: mean 168.7, SD 13.1 cm; and body mass: mean 74.0, SD 13.9 kg) and 12 healthy older adults (8 female adults and 4 male adults; age: mean 72.0, SD 6.4 years; height: mean 165.6, SD 9.0 cm; and body mass: mean 71.6, SD 16.3 kg) participated in the study.

A total of 24 participants completed 96 passes over the GAITRite mat with the phone placed in their pocket during the laboratory visits. Each session comprised 2 normal walking trials and 2 dual-task walking trials, yielding 4 - 5 strides per trial for each participant. This resulted in a data set comprising a total of 442 strides for direct comparison of gait speed, step length, and stride time estimations between the GAITRite mat and smartphone app approaches. Average normal and dual-task gait speeds for the younger and older groups are presented in Table 2.

Table . Means (SDs) and minimum (min)-maximum (max) of the spatiotemporal parameters during normal and dual-task walking in the younger and older groups.

Tests	Normal walking						Dual-task walking					
	Gait speed (m/s)		Step length (m)		Stride time (s)		Gait speed (m/s)		Step length (m)		Stride time (s)	
	Mean (SD)	Min-max	Mean (SD)	Min-max	Mean (SD)	Min-max	Mean (SD)	Min-max	Mean (SD)	Min-max	Mean (SD)	Min-max
GAITRite												
Younger	1.16 (0.08)	1.06 - 1.35	0.65 (0.04)	0.57 - 0.70	1.15 (0.08)	1.01 - 1.24	1.02 (0.09)	0.94 - 1.23	0.62 (0.03)	0.56 - 0.66	1.23 (0.12)	0.96 - 1.34
Older	1.00 (0.06)	0.94 - 1.18	0.57 (0.02)	0.54 - 0.62	1.18 (0.05)	1.04 - 1.22	0.95 (0.07)	0.81 - 1.12	0.56 (0.02)	0.50 - 0.60	1.22 (0.05)	1.07 - 1.27
Smartphone app: before adjusting												
Younger	1.11 (0.25)	0.90 - 1.77	0.61 (0.11)	0.53 - 0.92	1.12 (0.10)	0.93 - 1.22	0.96 (0.24)	0.82 - 1.66	0.57 (0.12)	0.47 - 0.93	1.22 (0.14)	0.91 - 1.34
Older	0.97 (0.18)	0.58 - 1.24	0.54 (0.10)	0.33 - 0.66	1.15 (0.04)	1.06 - 1.20	0.90 (0.18)	0.54 - 1.18	0.52 (0.09)	0.34 - 0.63	1.18 (0.05)	1.08 - 1.27
Smartphone app: after adjusting												
Younger	1.13 (0.12)	1.01 - 1.36	0.63 (0.02)	0.60 - 0.67	—[a]	—	1.01 (0.14)	0.91 - 1.25	0.60 (0.03)	0.55 - 0.69	—	—
Older	1.04 (0.11)	0.80 - 1.27	0.59 (0.06)	0.45 - 0.67	—	—	1.01 (0.11)	0.75 - 1.20	0.59 (0.05)	0.47 - 0.65	—	—

[a]Not applicable.

Validity and Reliability of the Original Estimated Gait Speed and Step Length

Average normal and dual-task gait speeds within each group, as estimated from the smartphone app, were largely similar to those derived from the GAITRite mat (Table 2). Smartphone-derived gait speed and step length demonstrated good validity as compared with the GAITRite mat. The close-to-zero average magnitude of biases for gait speed and step length was −0.03 m/s and −0.04 m, and the limits of agreement were 0.5 m/s and 0.28 m, respectively (Figures 1A and 2A). Orthogonal regression analysis revealed that the slope and intercept of the gait speed and step length derived from the app were 1.68 (intercept=−0.70) and 2.37 (intercept=−0.86), respectively (Figures 3A and 4A).

Figure 3. Relationship between gait speed as derived from a smartphone app–based assessment and from a GAITRite mat before (A) and after (B) adjusting by coefficients relative to the original estimation of step length. The Passing-Bablok orthogonal best-fit line of these data had a slope of approximately 1 and an intercept of 0 after adjusting.

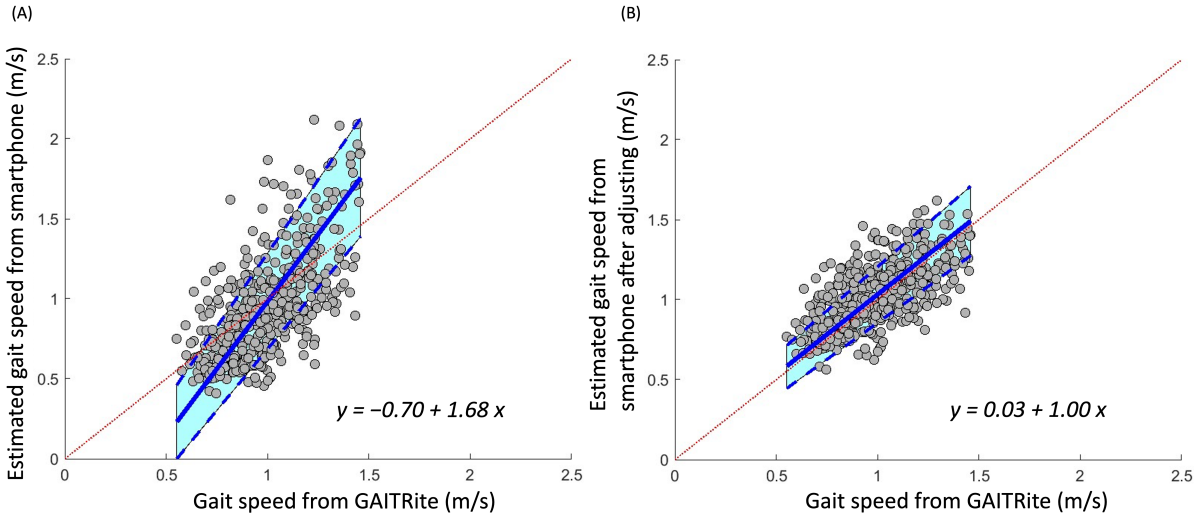
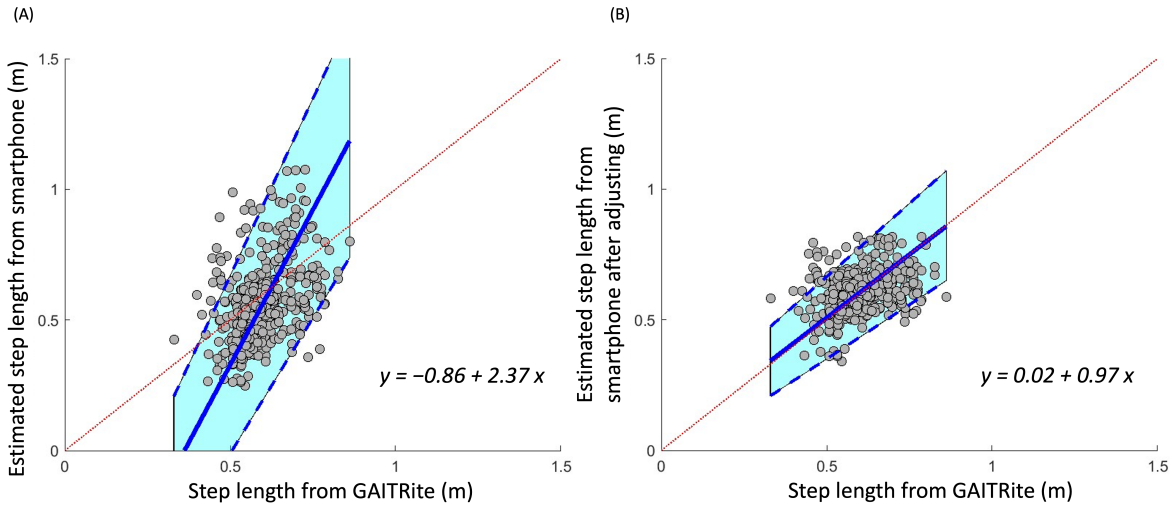


Figure 4. Relationship between step length as derived from a smartphone app–based assessment and from a GAITRite mat before (A) and after (B) adjusting by coefficients relative to the original estimation of step length. The Passing-Bablok orthogonal best-fit line of these data had a slope of approximately 1 and an intercept of 0 after adjusting.



Good to excellent test-retest reliabilities (0.64 - 0.86) were demonstrated for the average app-derived gait speed across trials within each laboratory assessment, as well as over the 3 home assessments, for both normal and dual-task walking in young and older adults (Table 3).

Table . Test-retest reliability of the laboratory- and home-based assessments of average gait speed in strides.

Tests	Normal walking			Dual-task walking		
	ICC[a]	*P* value	95% CI	ICC	*P* value	95% CI
GAITRite mat						
Laboratory assessment within visit	0.94	<.001	0.88-0.97	0.90	<.001	0.81-0.95
Before adjusting: smartphone app						
Laboratory assessment within visit	0.86	<.001	0.73-0.93	0.77	<.001	0.58-0.88
Home assessment	0.64	<.001	0.41-0.81	0.71	<.001	0.49-0.86
After adjusting: smartphone app						
Laboratory assessment within visit	0.72	<.001	0.52-0.85	0.75	<.001	0.56-0.87
Home assessment	0.62	<.001	0.38-0.80	0.76	<.001	0.57-0.89

[a]ICC: intraclass correlation coefficient.

Validity and Reliability of Smartphone-Derived Gait Speed After Adjusting Step Length by a Coefficient Relative to the Original Estimation of Step Length

After iteratively calculating the validities and reliabilities of the different step length zone thresholds as intervals, results indicated that 0.3 as an interval appeared to produce the best validity (Table 4). After adjusting step length with this coefficient relative to the original estimation of step length, the average gait speed in the younger group during normal and dual-task walking was 1.13 (SD 0.12) and 1.01 (SD 0.14) m/s, respectively; and the average gait speed in older group during the 2 conditions was 1.04 (SD 0.11) and 1.01 (SD 0.11) m/s, respectively (Table 2). These adjusted estimations of gait speed and step length derived from the smartphone app demonstrated a considerably smaller bias trend compared to the original estimated data (Figures 1 and 2). For gait speed and step length, mean differences were 0.03 m/s and 0.01 m, and limits of agreement were 0.34 m/s and 0.19 m, respectively (Figures 1B and 2B). Orthogonal regression analysis revealed that the validity of gait speed and step length derived from the app were higher with those measured by the GAITRite mat after adjusting, as indicated by one and close-to-one slopes, which were 1.00 (intercept=0.03) and 0.97 (intercept=0.02), respectively (Figures 3B and 4B).

Table . The slope and intercept for the Passing-Bablok orthogonal best-fit line after adjusting using stepwise iterative approach to define different step length zones by dividing the original estimation of step length using multiple zone thresholds as intervals from 0.1 to 0.5.

Intervals	Slope	Intercept
0.10	0.97	0.23
0.15	0.86	0.16
0.20	0.91	0.12
0.25	1.00	0.04
0.30	1.00	0.03
0.35	1.13	−0.09
0.40	1.23	−0.18
0.45	1.35	−0.28
0.50	1.40	−0.34

In general, the test-retest reliability of adjusted estimations of gait speed was good to excellent (0.62 - 0.76) between laboratory assessments, and between home assessments, for both normal and dual-task walking in the younger and older groups (Table 3).

Discussion

This study indicated that using a smartphone app placed in the front pocket of an individual's pants or shorts can produce valid estimates of gait speed under both normal and dual-task walking conditions in healthy younger and older adults. Moreover, we demonstrated that the accuracy of gait speed estimation using the inverted pendulum model approach can be improved by adjusting the estimated values with coefficients determined by the gold standard measurement of step length. This improved smartphone app–based approach provided high test-retest reliability within a supervised laboratory environment and within the unsupervised home setting.

The current results revealed that the pendulum model approach to calculating gait speed from a single IMU placed in the pocket overestimated step length and therefore gait speed if the step length derived from the GAITRite mat was greater than 0.8 m, yet underestimated both values if the step length was less than 0.5 m. Through visual inspection of published work, it appears that this observed bias in step length estimation is common, even when the IMU is secured to the participant's trunk [13,15]. To our knowledge, however, no one has reported this observation, and no adjustment has been made to correct the bias.

Our work suggests that a simple adjustment of step length for those steps that are relatively short, or relatively long, removes the observed bias and improves the accuracy of gait speed estimation. The current results suggested that the original estimated step length from the smartphone app between 0.5 and 0.8 m may not need to be adjusted as its coefficient was close to 1. Comparing our smartphone app with gold standard measurements, it is evident that in scenarios involving normal or dual-task walking, the app maintains robust performance with small bias across younger and older groups, different walking conditions, and different testing settings (laboratory and home). This implies that the described method can be used accurately within numerous applications. The simple inverted pendulum model requires only 2 pieces of information, namely, the participant's leg length and the change in height of the smartphone's vertical position during each step (from heel strike to contralateral heel strike). We suspect that several factors may contribute to the observed bias. There may be nonlinear relationships between step length and the combinations of changes in the height of the smartphone's vertical position and leg length. This could indicate inherent limitations in the inverted pendulum model. Exploring these factors further may thus offer insights into optimizing the model for different IMU locations and gait patterns, enhancing the applicability of our method across diverse populations and conditions.

The results from this study not only reinforce the reliability of the approach to estimate gait speed under different walking conditions in ambulatory older adults without overt disease but also suggest its applicability to all IMU data. Good to excellent test-retest reliability of results was observed across younger and older groups, different walking conditions (normal walking and dual-task walking), and different testing settings (laboratory and home). This implies that the described method can be used effectively within numerous applications. Moreover, the test-retest reliability observed in this study was noticeably higher than a previous report that used the pendulum model without adjustment [15]. Incorporating a wide range of ages from young to old in this study allowed us to encompass a broader spectrum of gait speeds and step lengths, which enhanced the generalizability of our findings. The observed consistency in the adjustment coefficients across these age groups suggests that the pendulum model approach we used is robust and applicable across a diverse age demographic. This inclusivity in our research approach enriches its relevance and widens its potential impact, thus making our findings more applicable in real-world scenarios where age-related variations in gait are common.

The coefficients established from our cohort may be applicable to other data sets where the populations and the settings of the devices (smartphone), including sampling rate, are similar; yet, the coefficients may need to be fine-tuned for different populations and devices, especially when a gold standard measurement is available. Our coefficients performed well in our data set, which involved a wide range of step length and

gait speed in both healthy young and older populations. In terms of future application to other data sets, this method holds significant potential for various areas. For data sets where a gold standard method is unavailable, our coefficients are worth trying to adjust the original estimated value. If the population (ie, healthy young and healthy older adults) and tool (ie, a smartphone app in the pants pocket) are similar to that tested and used in this study, implementing this type of adjustment is expected to improve the accuracy and reliability of the estimates. For data sets where a gold standard is available, researchers are recommended to develop their own coefficients based on the unique characteristics of each data set, thereby improving the method's precision and applicability to specific contexts.

Further research is required to determine the generalizability of our method to broader populations, especially in cases where a gold standard method is not available. The original step length estimations in this study were classified into 3 categories, which may decrease the test-retest reliability of those step length data close to the margin of categories. To address this, larger data sets are needed to establish a continuous adjustment approach without the need to classify the data into categories. Additionally, refining the inverted pendulum model to better accommodate the nonlinear relationship between step length and smartphone vertical displacement could enhance accuracy. The applicability of the method to individuals with abnormal gait patterns should also be explored. Despite these limitations, this study provided evidence that gait speed can be accurately and reliably estimated with minimum information (leg length) in young and older adults during normal and dual-task walking based upon a single IMU-embedded smartphone placed in the pocket.

Acknowledgments

The authors wish to thank the National Science and Technology Council, Taiwan (112 - 2917-I-564 - 018) and the National Institutes of Health/National Institute on Aging (K01AG075252) for the financial support.

Conflicts of Interest

None declared.

References

1. Hardy SE, Perera S, Roumani YF, Chandler JM, Studenski SA. Improvement in usual gait speed predicts better survival in older adults. J Am Geriatr Soc 2007 Nov;55(11):1727-1734. [doi: 10.1111/j.1532-5415.2007.01413.x] [Medline: 17916121]
2. Purser JL, Weinberger M, Cohen HJ, et al. Walking speed predicts health status and hospital costs for frail elderly male veterans. J Rehabil Res Dev 2005 Jul;42(4):535. [doi: 10.1682/JRRD.2004.07.0087] [Medline: 16320148]
3. Studenski S, Perera S, Patel K, et al. Gait speed and survival in older adults. JAMA 2011 Jan 5;305(1):50-58. [doi: 10.1001/jama.2010.1923] [Medline: 21205966]
4. Schmid A, Duncan PW, Studenski S, et al. Improvements in speed-based gait classifications are meaningful. Stroke 2007 Jul;38(7):2096-2100. [doi: 10.1161/STROKEAHA.106.475921] [Medline: 17510461]
5. Stanaway FF, Gnjidic D, Blyth FM, et al. How fast does the grim reaper walk? Receiver operating characteristics curve analysis in healthy men aged 70 and over. BMJ 2011 Dec 15;343:d7679. [doi: 10.1136/bmj.d7679] [Medline: 22174324]
6. Kraus M, Saller MM, Baumbach SF, et al. Prediction of physical frailty in orthogeriatric patients using sensor insole-based gait analysis and machine learning algorithms: cross-sectional study. JMIR Med Inform 2022 Jan 5;10(1):e32724. [doi: 10.2196/32724] [Medline: 34989684]
7. Buchner DM, Cress ME, Esselman PC, et al. Factors associated with changes in gait speed in older adults. J Gerontol A Biol Sci Med Sci 1996 Nov;51(6):M297-M302. [doi: 10.1093/gerona/51a.6.m297] [Medline: 8914502]
8. O'Loughlin JL, Robitaille Y, Boivin JF, Suissa S. Incidence of and risk factors for falls and injurious falls among the community-dwelling elderly. Am J Epidemiol 1993 Feb 1;137(3):342-354. [doi: 10.1093/oxfordjournals.aje.a116681] [Medline: 8452142]
9. Aminian K, Najafi B, Büla C, Leyvraz PF, Robert P. Spatio-temporal parameters of gait measured by an ambulatory system using miniature gyroscopes. J Biomech 2002 May;35(5):689-699. [doi: 10.1016/s0021-9290(02)00008-8] [Medline: 11955509]
10. Nishiguchi S, Yamada M, Nagai K, et al. Reliability and validity of gait analysis by android-based smartphone. Telemed J E Health 2012 May;18(4):292-296. [doi: 10.1089/tmj.2011.0132] [Medline: 22400972]
11. Takeda R, Tadano S, Todoh M, Morikawa M, Nakayasu M, Yoshinari S. Gait analysis using gravitational acceleration measured by wearable sensors. J Biomech 2009 Feb 9;42(3):223-233. [doi: 10.1016/j.jbiomech.2008.10.027] [Medline: 19121522]
12. Zijlstra W, Hof AL. Displacement of the pelvis during human walking: experimental data and model predictions. Gait Posture 1997 Dec;6(3):249-262. [doi: 10.1016/S0966-6362(97)00021-0]
13. Zijlstra W. Assessment of spatio-temporal parameters during unconstrained walking. Eur J Appl Physiol 2004 Jun;92(1-2):39-44. [doi: 10.1007/s00421-004-1041-5] [Medline: 14985994]

14. Brandes M, Zijlstra W, Heikens S, van Lummel R, Rosenbaum D. Accelerometry based assessment of gait parameters in children. Gait Posture 2006 Dec;24(4):482-486. [doi: 10.1016/j.gaitpost.2005.12.006] [Medline: 16427287]
15. Silsupadol P, Teja K, Lugade V. Reliability and validity of a smartphone-based assessment of gait parameters across walking speed and smartphone locations: body, bag, belt, hand, and pocket. Gait Posture 2017 Oct;58:516-522. [doi: 10.1016/j.gaitpost.2017.09.030] [Medline: 28961548]
16. Manor B, Yu W, Zhu H, et al. Smartphone app-based assessment of gait during normal and dual-task walking: demonstration of validity and reliability. JMIR Mhealth Uhealth 2018 Jan 30;6(1):e36. [doi: 10.2196/mhealth.8815] [Medline: 29382625]
17. Zhou J, Yu W, Zhu H, et al. A novel smartphone app-based assessment of standing postural control: demonstration of reliability and sensitivity to aging and task constraints. Presented at: 2020 IEEE International Conference on E-health Networking, Application & Services; Mar 1-2, 2021; Shenzhen, China. [doi: 10.1109/HEALTHCOM49281.2021.9398972] [Medline: 18954251]
18. Zhou J, Jiang X, Yu W, et al. A smartphone app-based application enabling remote assessments of standing balance during the COVID-19 pandemic and beyond. IEEE Internet Things J 2021 Mar;8(21):15818-15828. [doi: 10.1109/JIOT.2021.3064442]
19. Lee PA, Dumontier C, Yu W, et al. Validity, reliability, and feasibility of smartphone application intended for home measurement of gait speed and other gait dynamics in older adults with blood cancers. Blood 2023 Nov 2;142(1):2354-2354. [doi: 10.1182/blood-2023-174535]
20. Manor B, Zhou J, Lo O, Zhu H, Zafonte R, Travison T, et al. Self-reported head trauma and health in retired NFL players. Innov Aging 2018 Nov;2(1). [doi: 10.1093/geroni/igy023.1898]
21. Su D, Liu Z, Jiang X, et al. Simple smartphone-based assessment of gait characteristics in Parkinson disease: validation study. JMIR Mhealth Uhealth 2021 Feb 19;9(2):e25451. [doi: 10.2196/25451] [Medline: 33605894]
22. Quintero D, Lambert DJ, Villarreal DJ, Gregg RD. Real-time continuous gait phase and speed estimation from a single sensor. Control Technol Appl 2017 Aug;2017:847-852. [doi: 10.1109/CCTA.2017.8062565] [Medline: 30148285]
23. Zijlstra W, Hof AL. Assessment of spatio-temporal gait parameters from trunk accelerations during human walking. Gait Posture 2003 Oct;18(2):1-10. [doi: 10.1016/s0966-6362(02)00190-x] [Medline: 14654202]
24. Passing H. A new biometrical procedure for testing the equality of measurements from two different analytical methods. Application of linear regression procedures for method comparison studies in clinical chemistry, part I. J Clin Chem Clin Biochem 1983 Nov;21(11):709-720. [doi: 10.1515/cclm.1983.21.11.709] [Medline: 6655447]
25. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychol Assess 1994 Dec;6(4):284-290. [doi: 10.1037/1040-3590.6.4.284]

Abbreviations

ICC: intraclass correlation coefficient
IMU: inertial measurement unit

Edited by L Buis; submitted 25.08.23; peer-reviewed by J Ruokolainen, M Kraus, V Lugade; revised version received 07.05.24; accepted 05.06.24; published 13.08.24.

Please cite as:
Lee PA, Yu W, Zhou J, Tsai T, Manor B, Lo OY
A Novel Approach for Improving Gait Speed Estimation Using a Single Inertial Measurement Unit Embedded in a Smartphone: Validity and Reliability Study
JMIR Mhealth Uhealth 2024;12:e52166
URL: https://mhealth.jmir.org/2024/1/e52166
doi:10.2196/52166

Original Paper

Using In-Shoe Inertial Measurement Unit Sensors to Understand Daily-Life Gait Characteristics in Patients With Distal Radius Fractures During 6 Months of Recovery: Cross-Sectional Study

Akiko Yamamoto[1], MD, PhD; Eriku Yamada[1], MD, PhD; Takuya Ibara[2], PhD; Fumiyuki Nihey[3], PhD; Takuma Inai[4], PhD; Kazuya Tsukamoto[1], MD; Tomohiko Waki[1], MD; Toshitaka Yoshii[1], MD, PhD; Yoshiyuki Kobayashi[5], PhD; Kentaro Nakahara[3], PhD; Koji Fujita[2,6], MD, PhD

[1]Department of Orthopaedic and Spinal Surgery, Graduate School of Medical and Dental Sciences, Tokyo Medical and Dental University, Tokyo, Japan

[2]Department of Functional Joint Anatomy, Graduate School of Medical and Dental Sciences, Tokyo Medical and Dental University, Tokyo, Japan

[3]Biometrics Research Laboratories, NEC Corporation, Chiba, Japan

[4]Biomechanics and Exercise Physiology Research Group, Health and Medical Research Institute, Department of Life Science and Technology, National Institute of Advanced Industrial Science and Technology, Kagawa, Japan

[5]Human Augmentation Research Center, National Institute of Advanced Industrial Science and Technology, Tokyo, Japan

[6]Division of Medical Design Innovations, Open Innovation Center, Institute of Research Innovation, Tokyo Medical and Dental University, Tokyo, Japan

Corresponding Author:
Koji Fujita, MD, PhD
Department of Functional Joint Anatomy
Graduate School of Medical and Dental Sciences
Tokyo Medical and Dental University
1-5-45, Yushima
Bunkyo-ku
Tokyo, 113-8519
Japan
Phone: 81 3 5803 5279
Email: fujiorth@tmd.ac.jp

Abstract

Background: A distal radius fracture (DRF) is a common initial fragility fracture among women in their early postmenopausal period, which is associated with an increased risk of subsequent fractures. Gait assessments are valuable for evaluating fracture risk; inertial measurement units (IMUs) have been widely used to assess gait under free-living conditions. However, little is known about long-term changes in patients with DRF, especially concerning daily-life gait. We hypothesized that, in the long term, the daily-life gait parameters in patients with DRF could enable us to reveal future risk factors for falls and fractures.

Objective: This study assessed the spatiotemporal characteristics of patients with DRF at 4 weeks and 6 months of recovery.

Methods: We recruited 16 women in their postmenopausal period with DRF as their first fragility fracture (mean age 62.3, SD 7.0 years) and 28 matched healthy controls (mean age 65.6, SD 8.0 years). Daily-life gait assessments and physical assessments, such as hand grip strength (HGS), were performed using an in-shoe IMU sensor. Participants' results were compared with those of the control group, and their recovery was assessed for 6 months after the fracture.

Results: In the fracture group, at 4 weeks after DRF, lower foot height in the swing phase (P=.049) and higher variability of stride length (P=.03) were observed, which improved gradually. However, the dorsiflexion angle in the fracture group tended to be lower consistently during 6 months (at 4 weeks: P=.06; during 6 months: P=.07). As for the physical assessments, the fracture group showed lower HGS at all time points (at 4 weeks: P<.001; during 6 months: P=.04), despite significant improvement at 6 months (P<.001).

Conclusions: With an in-shoe IMU sensor, we discovered the recovery of spatiotemporal gait characteristics 6 months after DRF surgery without the participants' awareness. The consistently unchanged dorsiflexion angle in the swing phase and lower HGS could be associated with fracture risk, implying the high clinical importance of appropriate interventions for patients with

DRF to prevent future fractures. These results could be applied to a screening tool for evaluating the risk of falls and fractures, which may contribute to constructing a new health care system using wearable devices in the near future.

(JMIR Mhealth Uhealth 2024;12:e55178) doi:10.2196/55178

KEYWORDS

distal radius fracture; gait analysis; daily life; long-term results; gait; sensor; sensors; walk; walking; fracture; fractures; wearable; wearables; recover; rehabilitation; spatiotemporal; inertial measurement; fragility; postmenopausal; menopause; grip; surgery; surgical; orthopedic; postoperative; orthopedics; fall; falls; bone; bones; wrist; radius; radial

Introduction

Gait analysis is useful for predicting future fall risk and reflecting various underlying physiological processes [1]. Quantitative gait characteristics, such as slower gait speed and shorter stride length, are associated with falls, resulting in fragility fractures [2-4]. Recently, inertial measurement units (IMUs) have been widely used to assess gait under free-living conditions owing to their convenience, low cost, small size, and high accuracy [5-7]. The shank and foot are the preferred placements [8,9], and foot kinematics is an important factor related to falls and physical ability [4].

Among fragility fractures resulting from falls, distal radius fractures (DRFs) are one of the most frequent initial fractures in older adults [10]. Many DRFs occur in women in their postmenopausal period, aged <75 years, who are healthy, active, and functionally independent. More than half of these women do not meet the criteria for osteoporosis [11-13]. However, the initial DRF is associated with a greater risk of functional decline [14] and subsequent fractures [15] in all age groups. These changes are more than 5 times higher, even in those aged 50-59 years [16]. This could reflect early changes in frailty [17].

In patients with DRF, lower gait ability was observed in the laboratory, which slightly improved 6 months after the surgery [13,18]. However, these gait assessments were mainly performed for approximately 10 seconds, which may not accurately depict daily-life gait [19]. Further, most studies have only highlighted the consequences of wrist function and pain when investigating the long-term outcomes of DRF, and little is known about the effect of DRF on physical abilities, such as activities of daily living. Therefore, the long-term alterations of daily-life gait characteristics in patients with fractures remain unknown, and this study attempts to bridge this gap.

We previously found out that in-shoe IMU sensors were effective in the assessment of daily-life gait in patients with an initial DRF [7]. We hypothesized that, in the long term, it could enable us to identify future risk factors for secondary fractures by spatiotemporally following daily-life gait parameters using IMU sensors. We aimed to reveal the characteristics of spatiotemporal gait changes during 6 months following DRF.

Methods

Ethical Considerations

This study was approved by the Institutional Review Board of Tokyo Medical and Dental University (M2020-365) and followed the tenets of the Declaration of Helsinki. Written informed consent was provided by all participants. Participation in the study was voluntary, and no compensation was awarded for participation.

Recruitment

In the fracture group, we recruited 16 female patients with DRF who had undergone surgery for their first fragility fracture from 5 general hospitals. We compared their results to those of 28 healthy female volunteers. Women without a history of fragility fractures were recruited as the control group through local media advertisements. The inclusion criteria involved having the ability to walk without any support, no history of lower-extremity injury, and no known neuromuscular disorders or neurophysiological problems that may affect gait. Fragility fractures were defined as those that followed a fall from standing height or less. We excluded patients with DRF due to traffic or industrial accidents or multiorgan injuries. In the fracture group, 6 patients with DRF fell in the house without shoes, and the other 10 patients fell while wearing shoes. Owing to the lack of previous literature on the long-term data of gait in patients with DRF, the sample size estimations were based on the effect size of 0.78 (from the result of hand grip strength [HGS]) [13], with an assumed power of 0.8 and a type I error of 0.05. A sample of 16 participants with fractures was analyzed using G*Power (version 3.1; Heinrich Heine University Duesseldorf) [20].

Daily-Life Gait Assessments

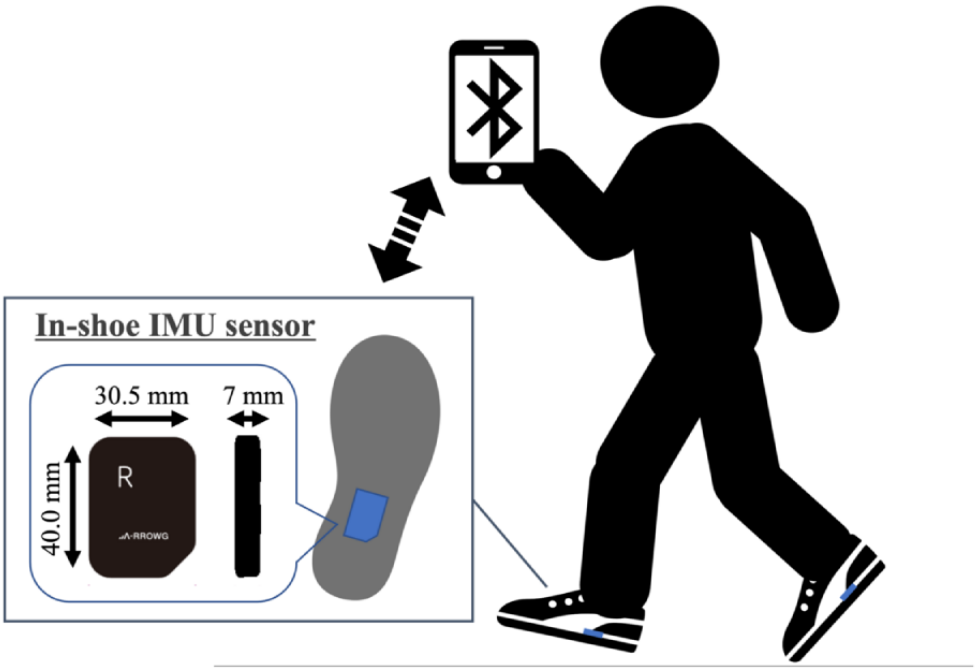
We measured daily-life gait using in-shoe IMU sensors (A-RROWG, NEC Corporation; Figure 1). These sensors are small (40.0 mm × 30.5 mm × 7 mm) and lightweight (11 g), including a 3-axis accelerometer and gyroscope. The IMU sensor in the dedicated insole was placed at the foot arch, and the x-, y-, and z-axes of the IMUs were set along the mediolateral, anteroposterior, and vertical directions, respectively. When a person wearing these sensors walks in a stable straight line over 3 gait cycles between 5 AM and 10 PM, the in-shoe IMU sensor detects that the person is walking based on acceleration in the anteroposterior direction and saves the IMU signals of the next 3 gait cycles as 1 gait measurement [21]. The IMU signals were sampled at a rate of 100 Hz, transferred to a smartphone via Bluetooth, and stored in a specialized app if participants had one with them. If a person did not have a smartphone, the data were uploaded automatically via Bluetooth at 11 PM by keeping the smartphone near the IMU sensors.

From the saved IMU signals, the mean of 7 gait parameters from 3 gait cycles was calculated and stored on a smartphone, as previously described by Fukushi et al [21]. The following 7 parameters were calculated:

1. Gait speed: calculated as stride length (m) divided by stride time (s).
2. Stride length: the distance between the start and end points of the foot trajectory for 1 stride.
3. Dorsiflexion angle: the peak foot angle in the dorsal direction from the ground during the swing phase.
4. Plantarflexion angle: the peak foot angle in the plantar direction from the ground during the swing phase.
5. Foot height: the maximum height of the foot trajectory.
6. Toe-in or toe-out angle: the mean angle of foot adduction or abduction in the direction of the velocity vector during the swing phase.
7. Circumduction: the displacement in the medial-lateral direction during the swing phase.

In addition to these 7 gait parameters, the coefficient of variation (CV), calculated as SD divided by mean multiplied by 100, was used to evaluate the variability.

Figure 1. The in-shoe inertial measurement unit (IMU) sensor. The sensors with accelerometer and gyroscope were set into the dedicated insoles, measured the daily-life gait without the participants' awareness, and saved the obtained data to their smartphone via Bluetooth.



Measurement Protocol

All participants completed a paper-based questionnaire on their general health status, which included fall history in the past year, frequency of stumbling, and fear of falling. Falls at the time of fracture in patients with DRFs were excluded based on the number of falls in the past year.

In the daily gait assessments, we inserted IMU sensors with dedicated insoles into the preferred shoes of both participants' feet. Each participant was provided a smartphone with only the original app for storing gait data. We verified whether participants could walk using the sensors and whether the sensors worked with a smartphone. Participants were instructed to wear shoes with sensors for 2 to 6 weeks. We requested that participants spend their daily lives as usual; therefore, we did not establish a minimum time for wearing or walking with them. The measurements were recorded without the participants' awareness, and they could view their most recent gait data by checking their smartphone. In the fracture group, daily gait assessments in the early postoperative period started 2 weeks after DRF surgery to enable the effects of casting or surgery.

To assess baseline functional ability and frailty, HGS [22] and the Timed Up and Go (TUG) test [23] were performed. HGS was measured in kilograms with a Jamer dynamometer (Sammons Preston). We assessed HGS on the nonfractured side of the fracture group and both sides of the control group. The mean values of 3 measurements were recorded. The time to complete the TUG test was recorded twice: once at the preferred speed and once at the fastest speed. These baseline physical tests were performed 4 weeks after DRF surgery in the fracture group and at the beginning of daily gait measurements in the control group.

To check for long-term functional ability, we asked the fracture group patients to perform the same physical and daily gait assessments again at an outpatient visit 6 months after the surgery.

Data Analysis

The median and CV of each participant were calculated using the obtained gait data. However, the automatically collected data included hills, turns, and straddling, and we used Smirnov-Grubbs analysis for every gait parameter to exclude any outliers from the data. After exclusion, approximately 20-1000 gait measurements were recorded for each participant during the measurement period, reflecting differences in their lifestyles. Even after the exclusion, gait data included various gait types, such as walking exercise and hurried walking; therefore, we excluded participants with less than 50 gait measurements in either foot, based on a previous report [24]. Since we computed the data obtained from both feet together,

we selected participants with 100 or more gait assessments for this study.

Statistical Analysis

Between-group differences were assessed using Student *t* test (2-tailed) for continuous variables in the patient demographics, and the chi-square test for categorical variables. Since the participants' gait data were not normally distributed, they were assessed using nonparametric (ie, Kruskal-Wallis and Steel) tests. The results of the fracture group in the early postoperative and long-term periods were compared using a paired *t* test. Statistical significance was set at $P<.05$.

These analyses were performed using EZR (version 1.55; Division of Haematology, Saitama Medical Center, Jichi Medical University) [25].

Results

Participants' Demographics

There were no significant differences in age, body characteristics, or lifestyle variables between the 2 groups. In the section on questions about falls, patients with fractures demonstrated a significantly higher incidence of falls (Table 1). None of the fracture group members experienced falls or subsequent fractures after the initial DRF.

Table 1. Participants' demographics. *P* values <.05 are considered significant.

Characteristics	Control (n=28)	Fracture (n=16)	*P* value
Age (years), mean (SD)	62.3 (7.0)	65.6 (8.0)	.20[a]
Height (cm), mean (SD)	155.2 (4.3)	154.8 (4.0)	.77[a]
Body weight (kg), mean (SD)	54.4 (8.0)	51.6 (8.5)	.30[a]
BMI (kg/m^2), mean (SD)	22.6 (3.2)	21.5 (3.2)	.28[a]
Hand dominance (right), n (%)	27 (96)	15 (94)	.68[b]
Foot dominance (right), n (%)	23 (82)	14 (88)	.64[b]
Smoking (current and previous), n (%)	5 (18)	5 (31)	.31[b]
Alcohol consumption, n (%)	11 (39)	5 (31)	.59[b]
Comorbidities, n (%)			
Hypertension	8 (29)	3 (19)	.47[b]
Eye disease	1 (4)	1 (6)	.68[b]
Diabetes mellitus	0 (0)	0 (0)	.48[b]
Knee osteoarthritis	2 (7)	0 (0)	.27[b]
Hip osteoarthritis	0 (0)	0 (0)	.48[b]
Rheumatoid arthritis	0 (0)	0 (0)	.48[b]
Number of oral medications, mean (SD)	0.8 (0.03)	0.9 (0.06)	.27[a]
The experience of fall in the past year, n (%)	0 (0)	4 (25)	.006[b]
Number of falls, n			
Once	N/A[c]	1	N/A
twice	N/A	2	N/A
3 times	N/A	1	N/A
The experience of stumbling, n (%)	17 (61)	9 (56)	.77[b]

[a]Independent Student *t* tests were used to compare the groups.

[b]Chi-square test was used for analysis between the groups.

[c]N/A: not applicable.

Spatiotemporal Data in Daily-Life Gait

The results of the spatiotemporal daily-life gait are presented in Table 2. There were no between-group differences in the number of measurements. Compared with the control group, the dorsiflexion angle measured at any time postoperatively in the fracture group tended to be lower and demonstrated no improvement in the postoperative course. Participants in the fracture group demonstrated significant recovery in gait speed, stride length, and plantarflexion angle between 4 weeks and 6

months postoperatively. Among the CV of each gait parameter, only the CV of stride length between the control group and the 4 weeks–postfracture group demonstrated a significant difference.

Table 2. Daily-life spatiotemporal data. *P* values <.05 are considered significant.

Characteristics	Control group (n=28)	Fracture group (n=16)		*P* value			
		4 weeks after surgery	6 months after surgery	Kruskal-Wallis test	Control (at 4 weeks)	Control (at 6 months)	Control (4 weeks to 6 months)
Number of measurements, mean (SD)	479.3 (432.7)	746.6 (468.7)	543.3 (475.8)	.06[a]	.05[b]	.97[b]	.06[c]
Median of each parameter, mean (SD)							
Gait speed (m/s)	1.28 (0.12)	1.22 (0.09)	1.26 (0.10)	.17[a]	.12[b]	.66[b]	.046[c]
Stride length (m)	1.26 (0.12)	1.20 (0.09)	1.24 (0.09)	.26[a]	.23[b]	.94[b]	.001[c]
Dorsiflexion angle (degree)	26.1 (3.83)	22.8 (4.15)	23.1 (3.46)	.03[a]	.06[b]	.07[b]	.24[c]
Plantarflexion angle (degree)	75.0 (6.18)	71.5 (4.23)	72.8 (4.64)	.08[a]	.06[b]	.31[b]	.04[c]
Foot height (cm)	14.0 (1.06)	13.1 (1.35)	13.8 (1.38)	.08[a]	.049[b]	.87[b]	.10[c]
Circumduction (cm)	2.85 (0.85)	3.16 (0.49)	3.17 (0.85)	.12[a]	.09[b]	.30[b]	>.99[c]
Toe-in or toe-out angle (degree)	13.2 (4.63)	13.6 (3.90)	14.1 (3.97)	.93[a]	>.99[b]	.89[b]	.41[c]
CV[d] of each parameter (%), mean (SD)							
Gait speed	15.2 (4.84)	16.3 (2.95)	15.8 (2.96)	.36[a]	.37[b]	.52[b]	.45[c]
Stride length	10.3 (2.89)	12.6 (3.21)	11.2 (2.60)	.05[a]	.03[b]	.57[b]	.15[c]
Dorsiflexion angle	20.7 (6.09)	23.2 (5.48)	22.9 (4.06)	.19[a]	.22[b]	.31[b]	.74[c]
Plantarflexion angle	8.56 (2.74)	10.2 (2.70)	10.1 (3.57)	.14[a]	.12[b]	.34[b]	.65[c]
Foot height	8.01 (2.15)	11.0 (7.05)	9.05 (2.73)	.12[a]	.10[b]	.44[b]	.33[c]
Circumduction	51.3 (12.8)	46.5 (12.0)	46.5 (6.16)	.45[a]	.79[b]	.39[b]	.83[c]
Toe-in or toe-out angle	30.7 (15.3)	37.9 (19.9)	31.7 (13.0)	.51[a]	.44[b]	.89[b]	.15[c]

[a]Kruskal-Wallis test was used to compare the control and fracture groups.

[b]Steel test was used to compare each group.

[c]Paired sample *t* test was used for analysis between the groups.

[d]CV: coefficient of variation.

HGS and Body Balancing Ability

The HGS in the fracture group demonstrated significant recovery between 4 weeks and 6 months after surgery; however, it was significantly lower in the fracture group compared to the control group. In the TUG test, there were no significant differences between the control and fracture groups or at 4 weeks and 6 months postoperatively in the fracture group (Table 3).

Table 3. Physical tests in the control and fracture groups. *P* values <.05 are considered significant.

Characteristics	Control group (n=28)	Fracture group (n=16)		*P* value			
		4 weeks after surgery	6 months after surgery	Kruskal-Wallis test	Control (at 4 weeks)	Control (at 6 months)	Control (4 weeks to 6 months)
Hand grip strength (kg), mean (SD)	23.3 (3.4)	19.1 (2.6)	20.6 (3.1)	<.001[a]	<.001[b]	.04[b]	<.001[c]
TUG[d] test (s), mean (SD)							
Normal speed	8.07 (1.33)	7.53 (0.85)	8.2 (1.28)	.28[a]	.29[b]	.96[b]	.13[c]
Faster speed	6.23 (0.89)	6.09 (0.64)	6.4 (0.95)	.48[a]	.51[b]	.99[b]	.47[c]

[a]Kruskal-Wallis test was used to compare the control and fracture groups.

[b]Steel test was used to compare each group.

[c]Paired sample *t* test was used for analysis between the groups.

[d]TUG: Timed Up and Go.

Discussion

Principal Results

We performed daily-life gait assessments with in-shoe IMU sensors and some physical tests, including HGS, to evaluate the differences in spatiotemporal gait and physical ability between patients with DRF and healthy controls. Moreover, we assessed whether these parameters improved during 6 months following DRF. In daily-life gait assessments, patients with DRF demonstrated a lower foot height and higher CV stride length compared to the control group; however, these differences were no longer present 6 months after DRF surgery, reflecting the improvement in some parameters in the fracture group after 6 months of DRF treatment. On the other hand, the dorsiflexion angle in the fracture group tended to be lower consistently during 6 months. In the physical assessments, patients with DRF at 4 weeks and 6 months after the surgery had significantly lower HGS than those without DRF, even though HGS in patients with DRF improved during 6 months.

Some gait parameters, such as gait speed and stride length, improved in the fracture group during 6 months of daily life. Further, the CV of stride length, which is associated with fall risk [26], was higher in the fracture group at 4 weeks after DRF and did not significantly change after 6 months. With the development of wearable sensors, spatiotemporal gait characteristics and their variability in daily life have been increasingly identified. However, few studies have explored the long-term changes in gait characteristics related to falls and fractures. As daily-life gait is influenced by various factors, such as environmental and psychiatric factors [27], changes in patients with DRF could reflect changes in their lifestyle. Conversely, the risk of subsequent fractures is the highest immediately after the initial fracture [28]. Although no subsequent fractures occurred in this study, further research is warranted to determine the relationship between gait changes and subsequent fractures. The in-shoe IMU sensor must play an important role in further evaluating this relationship.

The dorsiflexion angle in the fracture group remained lower, whereas other parameters in the fracture group improved during 6 months. As for vertebral fractures, which are typical fragility fractures as well as DRF, patients with symptomatic vertebral fractures walked with shorter and wider strides at the time of injury. Although those with vertebral fractures show improvement in stride time and stride length over time, even reaching healthy levels again, their gait pattern and stability persist for 6 months, implying a greater risk of incident disability among these patients [29]. Since gait speed is reported to affect other gait parameters, the persistent lower dorsiflexion angle in the fracture group, despite the improvement in gait speed, might be a characteristic of patients with DRF, indicating that patients cannot fully return to healthy states. The dorsiflexion angle in this study, which means the angle between the ground and the sole of the foot, may depend on the movement of all lower extremities. Kyphosis and flexed hip or knee joints, which are common in older people, are related to foot movement or strength [30]. Although further research is needed to determine the cause of this decrease in dorsiflexion angle, the angle could result in stumbling and falls. Older adults with DRF can be assessed as having a high risk of functional decline, particularly those who have access to a health care facility at an early stage. They should receive appropriate intervention to prevent future falls or fractures along with treatment for the initial fracture. Considering that several previous reports have assessed fall risk using machine learning based on gait data from fallers [31,32], our results could be effective in creating a more precise machine learning model for evaluating the risk of falls. Further research is warranted to explore not only the cost of developing sensors and apps but also intervention methods and the extent of fall reduction achievable. Nonetheless, our findings using in-shoe IMU sensors outside the hospital could be valuable for future screening tools to evaluate the risk of falls and fractures.

As for physical assessments, the HGS in the fracture group significantly improved during 6 months after DRF, which is consistent with a previous report [13]. The increased use of the nonaffected hand with DRF in daily life may improve HGS; however, most studies on DRF have focused on the HGS of the affected side, and little is known about that of the nonaffected side. Generally, HGS is associated with health status, including death, falls, and muscle strength [33]. The improvement in HGS could reflect the improvement in health status, and the lower HGS after 6 months of DRF could be associated with a lower

degree of health status in the fracture group. Contrastingly, there were no differences in the TUG test results, contrary to our previous reports [13,18]. The average TUG test time in both the fracture and control groups in this study was faster compared to previous studies, which may mean that the TUG test, which involves a few steps and seconds in the laboratory, may not reflect the true physical characteristics. Therefore, daily-life gait analysis for a certain period is needed to identify the slight difference between fracture group patients and healthy controls. By using this in-shoe sensor for a certain period, the lower dorsiflexion angle in patients with DRF was revealed, which remained 6 months after the fracture. We would like to further investigate long-term changes in the gait of these patients.

Limitations

This study had some limitations. First, the number of participants was small, which could have affected the power of this study. However, individual changes over 6 months appeared to confirm the statistical outcomes and might not have affected the overall conclusions of our study. Second, we observed progress up to 6 months after the fractures. The HGS on the affected side continued to improve beyond 1 year. Long-term changes in more patients with DRF should be further explored. Third, the participants were all Japanese, who had the habit of taking off their shoes indoors. Considering that 6 patients with DRF fell inside their houses without shoes, the results may not accurately reflect the daily free-living assessments of barefoot individuals.

Conclusions

In summary, we performed a case-control study to investigate the long-term changes in HGS and daily-life gait after DRF. Using an in-shoe IMU sensor, we revealed the recovery of spatiotemporal gait characteristics 6 months after DRF surgery without the participants' awareness. The dorsiflexion angle in the swing phase and HGS were still lower in the fracture group after 6 months, which could be associated with fracture risk. This in-shoe IMU sensor could be useful for evaluating the future fall and fracture risk outside the hospital and for constructing a new health care system related to preventive medicine using wearable devices outside the hospital.

Acknowledgments

This study was supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI (grant 21K17469) and joint research funding from NEC Corporation. We would like to thank Masahiro Fujimoto, Kanako Nakajima, Shoma Kudo, and Gaku Kutsuzawa for collecting data; Hiroshi Kajitani for collecting data and software development; Akimoto Nimura, Hidetoshi Kaburagi, and Takafumi Koyama for critical analysis and proofreading; and Editage for English language editing.

Conflicts of Interest

KF has received joint research funding from NEC Corporation. FN and KN are employees of NEC Corporation.

References

1. Middleton A, Fritz SL, Lusardi M. Walking speed: the functional vital sign. J Aging Phys Act 2015 Apr;23(2):314-322 [FREE Full text] [doi: 10.1123/japa.2013-0236] [Medline: 24812254]
2. Maki BE. Gait changes in older adults: predictors of falls or indicators of fear. J Am Geriatr Soc 1997 Mar;45(3):313-320. [doi: 10.1111/j.1532-5415.1997.tb00946.x] [Medline: 9063277]
3. Job M, Dottor A, Viceconti A, Testa M. Ecological gait as a fall indicator in older adults: a systematic review. Gerontologist 2020 Jul 15;60(5):e395-e412. [doi: 10.1093/geront/gnz113] [Medline: 31504484]
4. Wade FE, Kellaher GK, Pesquera S, Baudendistel ST, Roy A, Clark DJ, et al. Kinematic analysis of speed transitions within walking in younger and older adults. J Biomech 2022 Jun;138:111130 [FREE Full text] [doi: 10.1016/j.jbiomech.2022.111130] [Medline: 35569430]
5. Yamamoto A, Fujita K, Yamada E, Ibara T, Nihey F, Inai T, et al. Gait characteristics in patients with distal radius fracture using an in-shoe inertial measurement system at various gait speeds. Gait Posture 2024 Jan;107:317-323. [doi: 10.1016/j.gaitpost.2023.10.023] [Medline: 37914562]
6. Mobbs RJ, Perring J, Raj SM, Maharaj M, Yoong NKM, Sy LW, et al. Gait metrics analysis utilizing single-point inertial measurement units: a systematic review. Mhealth 2022;8:9 [FREE Full text] [doi: 10.21037/mhealth-21-17] [Medline: 35178440]
7. Yamamoto A, Fujita K, Yamada E, Ibara T, Nihey F, Inai T, et al. Foot characteristics of the daily-life gait in postmenopausal females with distal radius fractures: a cross-sectional study. BMC Musculoskelet Disord 2023 Sep 05;24(1):706 [FREE Full text] [doi: 10.1186/s12891-023-06845-5] [Medline: 37670304]
8. Prasanth H, Caban M, Keller U, Courtine G, Ijspeert A, Vallery H, et al. Wearable sensor-based real-time gait detection: a systematic review. Sensors (Basel) 2021 Apr 13;21(8):1-28 [FREE Full text] [doi: 10.3390/s21082727] [Medline: 33924403]
9. Appelboom G, Taylor BE, Bruce E, Bassile CC, Malakidis C, Yang A, et al. Mobile phone-connected wearable motion sensors to assess postoperative mobilization. JMIR Mhealth Uhealth 2015;3(3):e78 [FREE Full text] [doi: 10.2196/mhealth.3785] [Medline: 26220691]
10. Sontag A, Krege JH. First fractures among postmenopausal women with osteoporosis. J Bone Miner Metab 2010 Jul;28(4):485-488. [doi: 10.1007/s00774-009-0144-9] [Medline: 20052602]

11. Edwards BJ, Song J, Dunlop DD, Fink HA, Cauley JA. Functional decline after incident wrist fractures--Study of Osteoporotic Fractures: prospective cohort study. BMJ 2010 Jul 08;341:c3324 [FREE Full text] [doi: 10.1136/bmj.c3324] [Medline: 20616099]
12. Sarfani S, Scrabeck T, Kearns AE, Berger RA, Kakar S. Clinical efficacy of a fragility care program in distal radius fracture patients. J Hand Surg Am 2014 Apr;39(4):664-669. [doi: 10.1016/j.jhsa.2014.01.009] [Medline: 24576753]
13. Fujita K, Kaburagi H, Nimura A, Miyamoto T, Wakabayashi Y, Seki Y, et al. Lower grip strength and dynamic body balance in women with distal radial fractures. Osteoporos Int 2019 May;30(5):949-956 [FREE Full text] [doi: 10.1007/s00198-018-04816-4] [Medline: 30607458]
14. Vergara I, Vrotsou K, Orive M, Garcia-Gutierrez S, Gonzalez N, Las Hayas C, et al. Wrist fractures and their impact in daily living functionality on elderly people: a prospective cohort study. BMC Geriatr 2016 Jan 14;16:11 [FREE Full text] [doi: 10.1186/s12877-015-0176-z] [Medline: 26762327]
15. Balasubramanian A, Zhang J, Chen L, Wenkert D, Daigle SG, Grauer A, et al. Risk of subsequent fracture after prior fracture among older women. Osteoporos Int 2019 Jan;30(1):79-92 [FREE Full text] [doi: 10.1007/s00198-018-4732-1] [Medline: 30456571]
16. Crandall CJ, Hunt RP, LaCroix AZ, Robbins JA, Wactawski-Wende J, Johnson KC, et al. After the initial fracture in postmenopausal women, where do subsequent fractures occur? EClinicalMedicine 2021 May;35:100826 [FREE Full text] [doi: 10.1016/j.eclinm.2021.100826] [Medline: 34124629]
17. Nellans KW, Kowalski E, Chung KC. The epidemiology of distal radius fractures. Hand Clin 2012 May;28(2):113-125 [FREE Full text] [doi: 10.1016/j.hcl.2012.02.001] [Medline: 22554654]
18. Fujita K, Iijima H, Eguchi R, Kuroiwa T, Sasaki T, Yokoyama Y, et al. Gait analysis of patients with distal radius fracture by using a novel laser Timed Up-and-Go system. Gait Posture 2020 Jul;80:223-227 [FREE Full text] [doi: 10.1016/j.gaitpost.2020.06.005] [Medline: 32540778]
19. Brodie MA, Coppens MJ, Ejupi A, Gschwind YJ, Annegarn J, Schoene D, et al. Comparison between clinical gait and daily-life gait assessments of fall risk in older people. Geriatr Gerontol Int 2017 Nov;17(11):2274-2282. [doi: 10.1111/ggi.12979] [Medline: 28176431]
20. Faul F, Erdfelder E, Buchner A, Lang A. Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. Behav Res Methods 2009 Nov;41(4):1149-1160. [doi: 10.3758/BRM.41.4.1149] [Medline: 19897823]
21. Fukushi K, Huang C, Wang Z, Kajitani H, Nihey F, Nakahara K. On-line algorithms of stride-parameter estimation for in-shoe motion-sensor system. IEEE Sensors J 2022 May 15;22(10):9636-9648. [doi: 10.1109/jsen.2022.3164057]
22. Hillman TE, Nunes QM, Hornby ST, Stanga Z, Neal KR, Rowlands BJ, et al. A practical posture for hand grip dynamometry in the clinical setting. Clin Nutr 2005 Apr;24(2):224-228. [doi: 10.1016/j.clnu.2004.09.013] [Medline: 15784482]
23. Podsiadlo D, Richardson S. J Am Geriatr Soc 1991 Feb;39(2):142-148. [doi: 10.1111/j.1532-5415.1991.tb01616.x] [Medline: 1991946]
24. Kawai H, Obuchi S, Hirayama R, Watanabe Y, Hirano H, Fujiwara Y, et al. Intra-day variation in daily outdoor walking speed among community-dwelling older adults. BMC Geriatr 2021 Jul 08;21(1):417 [FREE Full text] [doi: 10.1186/s12877-021-02349-w] [Medline: 34238238]
25. Kanda Y. Investigation of the freely available easy-to-use software 'EZR' for medical statistics. Bone Marrow Transplant 2013 Mar;48(3):452-458 [FREE Full text] [doi: 10.1038/bmt.2012.244] [Medline: 23208313]
26. Hausdorff JM. Gait variability: methods, modeling and meaning. J Neuroeng Rehabil 2005 Jul 20;2:19 [FREE Full text] [doi: 10.1186/1743-0003-2-19] [Medline: 16033650]
27. Zhong R, Rau PP. A mobile phone-based gait assessment app for the elderly: development and evaluation. JMIR Mhealth Uhealth 2020 Feb 29;8(2):e14453 [FREE Full text] [doi: 10.2196/14453] [Medline: 32452821]
28. Laurs-van Geel TACM, Center JR, Geusens PP, Dinant G, Eisman JA. Clinical fractures cluster in time after initial fracture. Maturitas 2010 Dec;67(4):339-342. [doi: 10.1016/j.maturitas.2010.09.002] [Medline: 20926211]
29. Jacobs E, McCrum C, Senden R, van Rhijn LW, Meijer K, Willems PC. Gait in patients with symptomatic osteoporotic vertebral compression fractures over 6 months of recovery. Aging Clin Exp Res 2020 Feb;32(2):239-246 [FREE Full text] [doi: 10.1007/s40520-019-01203-9] [Medline: 31030421]
30. Sinaki M, Brey RH, Hughes CA, Larson DR, Kaufman KR. Balance disorder and increased risk of falls in osteoporosis and kyphosis: significance of kyphotic posture and muscle strength. Osteoporos Int 2005 Aug;16(8):1004-1010. [doi: 10.1007/s00198-004-1791-2] [Medline: 15549266]
31. Bargiotas I, Wang D, Mantilla J, Quijoux F, Moreau A, Vidal C, et al. Preventing falls: the use of machine learning for the prediction of future falls in individuals without history of fall. J Neurol 2023 Feb;270(2):618-631 [FREE Full text] [doi: 10.1007/s00415-022-11251-3] [Medline: 35817988]
32. Kim J, Bae M, Lee K, Kim J, Hong SG. Explainable artificial intelligence and wearable sensor-based gait analysis to identify patients with osteopenia and sarcopenia in daily life. Biosensors (Basel) 2022 Mar 07;12(3):167 [FREE Full text] [doi: 10.3390/bios12030167] [Medline: 35323437]
33. Leong DP, Teo KK, Rangarajan S, Lopez-Jaramillo P, Avezum A, Orlandini A, et al. Prognostic value of grip strength: findings from the Prospective Urban Rural Epidemiology (PURE) study. Lancet 2015 Jul 18;386(9990):266-273. [doi: 10.1016/S0140-6736(14)62000-6] [Medline: 25982160]

Abbreviations

CV: coefficient of variation
DRF: distal radius fracture
HGS: hand grip strength
IMU: inertial measurement unit
TUG: Timed Up and Go

Edited by L Buis; submitted 05.12.23; peer-reviewed by S Okita, M Nakayama; comments to author 08.01.24; revised version received 29.01.24; accepted 26.02.24; published 20.03.24.

Please cite as:
Yamamoto A, Yamada E, Ibara T, Nihey F, Inai T, Tsukamoto K, Waki T, Yoshii T, Kobayashi Y, Nakahara K, Fujita K
Using In-Shoe Inertial Measurement Unit Sensors to Understand Daily-Life Gait Characteristics in Patients With Distal Radius Fractures During 6 Months of Recovery: Cross-Sectional Study
JMIR Mhealth Uhealth 2024;12:e55178
URL: https://mhealth.jmir.org/2024/1/e55178
doi:10.2196/55178
PMID:38506913

Original Paper

Harnessing Consumer Wearable Digital Biomarkers for Individualized Recognition of Postpartum Depression Using the All of Us Research Program Data Set: Cross-Sectional Study

Eric Hurwitz[1], PhD; Zachary Butzin-Dozier[2], PhD; Hiral Master[3], MPH, PT, PhD; Shawn T O'Neil[1], MS, PhD; Anita Walden[1], MS; Michelle Holko[4], PhD; Rena C Patel[5], MPhil, MPH, MD; Melissa A Haendel[1], PhD

[1]Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

[2]School of Public Health, University of California, Berkeley, Berkeley, CA, United States

[3]Vanderbilt Institute of Clinical and Translational Research, Vanderbilt University Medical Center, Nashville, TN, United States

[4]International Computer Science Institute, Berkeley, CA, United States

[5]Department of Infectious Disease, University of Alabama at Birmingham, Birmingham, AL, United States

Corresponding Author:
Eric Hurwitz, PhD
Department of Genetics
University of North Carolina at Chapel Hill
120 Mason Farm Road
Chapel Hill, NC, 27599
United States
Phone: 1 9198436475
Email: eric_hurwitz@med.unc.edu

Abstract

Background: Postpartum depression (PPD) poses a significant maternal health challenge. The current approach to detecting PPD relies on in-person postpartum visits, which contributes to underdiagnosis. Furthermore, recognizing PPD symptoms can be challenging. Therefore, we explored the potential of using digital biomarkers from consumer wearables for PPD recognition.

Objective: The main goal of this study was to showcase the viability of using machine learning (ML) and digital biomarkers related to heart rate, physical activity, and energy expenditure derived from consumer-grade wearables for the recognition of PPD.

Methods: Using the *All of Us* Research Program Registered Tier v6 data set, we performed computational phenotyping of women with and without PPD following childbirth. Intraindividual ML models were developed using digital biomarkers from Fitbit to discern between prepregnancy, pregnancy, postpartum without depression, and postpartum with depression (ie, PPD diagnosis) periods. Models were built using generalized linear models, random forest, support vector machine, and k-nearest neighbor algorithms and evaluated using the κ statistic and multiclass area under the receiver operating characteristic curve (mAUC) to determine the algorithm with the best performance. The specificity of our individualized ML approach was confirmed in a cohort of women who gave birth and did not experience PPD. Moreover, we assessed the impact of a previous history of depression on model performance. We determined the variable importance for predicting the PPD period using Shapley additive explanations and confirmed the results using a permutation approach. Finally, we compared our individualized ML methodology against a traditional cohort-based ML model for PPD recognition and compared model performance using sensitivity, specificity, precision, recall, and F_1-score.

Results: Patient cohorts of women with valid Fitbit data who gave birth included <20 with PPD and 39 without PPD. Our results demonstrated that intraindividual models using digital biomarkers discerned among prepregnancy, pregnancy, postpartum without depression, and postpartum with depression (ie, PPD diagnosis) periods, with random forest (mAUC=0.85; κ=0.80) models outperforming generalized linear models (mAUC=0.82; κ=0.74), support vector machine (mAUC=0.75; κ=0.72), and k-nearest neighbor (mAUC=0.74; κ=0.62). Model performance decreased in women without PPD, illustrating the method's specificity. Previous depression history did not impact the efficacy of the model for PPD recognition. Moreover, we found that the most predictive biomarker of PPD was calories burned during the basal metabolic rate. Finally, individualized models surpassed the performance of a conventional cohort-based model for PPD detection.

Conclusions: This research establishes consumer wearables as a promising tool for PPD identification and highlights personalized ML approaches, which could transform early disease detection strategies.

(JMIR Mhealth Uhealth 2024;12:e54622) doi:10.2196/54622

KEYWORDS

wearable device; All of Us; postpartum depression; machine learning; Fitbit; mobile phone

Introduction

Background

Postpartum depression (PPD) is the most common complication of childbirth, occurring in approximately 1 in 7 women [1]. PPD can have several implications for women, manifesting in ways such as irritability, mood swings, fatigue, sleep and appetite disturbance, and thoughts of suicide [2]. Undetected PPD has also been shown to have financial implications for affected individuals as it can lead to challenges in maintaining employment or reduced work performance [3]. Furthermore, PPD has been linked to an elevated risk of mood disorders in the child as well as paternal depression [4,5].

Unfortunately, PPD remains significantly underdiagnosed and undertreated, as indicated by the strikingly low treatment rate of only 15% [6]. The current method of diagnosing PPD relies on screening instruments such as the Edinburgh Postnatal Depression Scale (EPDS), Center for Epidemiologic Studies Depression Scale, Patient Health Questionnaire, and Postpartum Depression Screening Scale, where the EPDS is the most commonly used instrument [7]. Often, women also need to undergo blood tests to assess thyroid function as the symptoms of PPD frequently overlap with hyperthyroidism [7]. Due to the challenges in diagnosing PPD, traditional approaches using these screening tools contribute to inadequate screening of women and subsequent underdiagnosis [8,9]. Therefore, the advent of new technologies is greatly needed to enable adequate and, hopefully, earlier detection of PPD.

Digital health tools have been gaining traction in recent years due to the near-ubiquitous ownership of smartphones [10]. Leveraging data passively collected by wearables (ie, digital biomarkers such as the average heart rate [HR], total steps, and calories burned per day) coupled with machine learning (ML) algorithms provides an opportunity to model the relationship between digital biomarkers and a particular disease for early recognition.

Prior Work

Previous studies have demonstrated that ML algorithms using digital biomarkers from smartwatches can predict cardiovascular diseases, infection, diabetes, and mental health conditions [11-14]. For example, one study demonstrated that a wearable device could estimate the changes in the severity of patients with major depressive disorder, where their findings indicated that ML models exclusively using digital biomarkers from wearables achieved moderate performance with correlation coefficients of 0.56 (95% CI 0.39-0.73) and 0.54 (95% CI 0.49-0.59) in the time-split and user-split scenarios, respectively, between model predictions and actual Hamilton Depression Rating Scale scores [15]. Another study recruited individuals with moderate depression for 4 weeks to develop individualized ML models based on digital biomarkers to predict mood. Their findings displayed a correlation between digital biomarkers and depression, as evidenced by high-performing models with a mean absolute error of 0.77 (SD 0.27) points on the 7-point Likert scale, which corresponds to a mean absolute percent error of 27.9% (SD 10.3%) [16]. A study by Wang et al [17] found that students with higher depressive symptoms measured using the 8-item Patient Health Questionnaire were more likely to (1) use their phone at study locations (correlation coefficient [r]=0.39; P<.001) compared to all-day phone use (r=0.28; P=.01), (2) have irregular sleep time (r=0.30; P=.02) and wake time (r=0.27; P=.04) schedules, (3) be stationary for more time (r=0.37; P=.01), and (4) visit fewer places during the day (r=−0.27; P=.02). In addition, students with higher depressive symptoms measured using the 4-item Patient Health Questionnaire scores (1) were around a fewer number of conversations (P=.002), (2) slept for shorter durations (P=.02), (3) fell asleep later (P=.001), (4) woke up later (P=.03), and (5) visited fewer places (P=.003) over the previous 2-week period [17]. Other studies examining the association between digital biomarkers from wearables and depression include those by (1) Moshe et al [18], who demonstrated a negative association between the variability of locations visited and depressive symptoms (β=−.21; P=.04) and a positive association between total sleep time and time in bed and depressive symptoms (β=.24; P=.02); and (2) Rykov et al [19], who showed that a larger variation in nighttime HR between 2 AM and 4 AM (r=0.26; P=.001) and between 4 AM and 6 AM (r=0.18; P=.04) and lower regularity of weekday circadian activity based on steps (r=−0.17; P=.049) were associated with higher severity of depressive symptoms.

Additional research has been conducted related to understanding the relationship between wearable-derived digital biomarkers and PPD. For instance, one study showed that the features most predictive of maternal loneliness, which is commonly associated with PPD, were activity intensity, activity distribution during the day, resting HR, and HR variability [20]. It was also shown that women with milder depression symptoms typically had a larger daily radius of travel compared to those with more severe symptoms (2.7 vs 1.9 miles; P=.04) [21]. Finally, women with depression have been shown to have a lower HR variability (measured using the SD of 24-hour NN intervals, F=6.4; P=.01, and the SD of the averages of NN intervals in 5-minute segments, F=6.04; P=.02) and elevated HR while sleeping (F=5.05; P=.03) compared to women without depression [22].

While these studies highlight a relationship between digital biomarkers and depression or PPD, they suffer from the following limitations: (1) some studies use data in the model that need active patient engagement with partnered mobile apps,

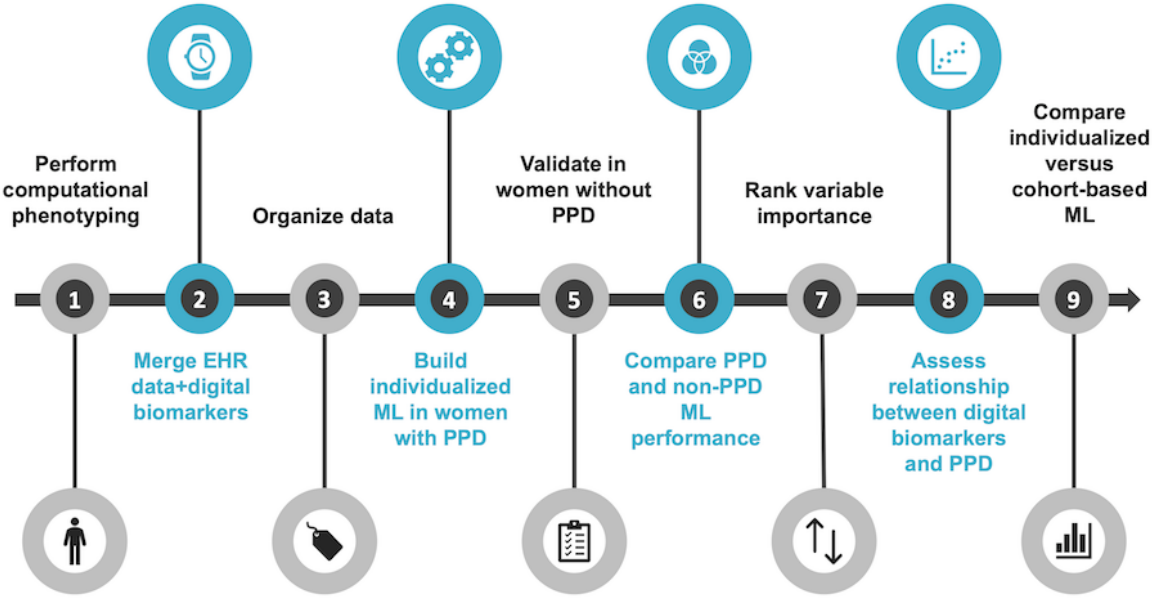
where user retention is known to decrease over time with health-related apps; (2) most studies do not use a predictive framework but rather examine the association between digital biomarkers and depressive symptoms; (3) only one study has developed individualized ML models; (4) most studies analyzing women with PPD have limited time frames and do not capture continuous longitudinal data across different phases of pregnancy; and (5) no studies have developed individualized ML models for women in the postpartum period combining data from wearables and the electronic health record (EHR) [23]. Therefore, a method that provides continuous and personalized monitoring without the need for clinical encounters to enable early detection of mental health disorders, including PPD, is needed.

Goal of This Study

The *All of Us* Research Program (AoURP) is a comprehensive data set that collects several types of health-related data, including surveys, EHRs, physical measurements, and wearable data from Fitbit devices, with an emphasis on patient populations that have been previously underrepresented in biomedical research [24]. Currently, the longitudinal Fitbit data from >15,000 AoURP participants are made available to registered researchers on the *All of Us* Researcher Workbench, providing an opportunity to explore digital biomarkers in a diverse cohort of participants.

It is unknown whether digital biomarkers from consumer wearables can be used to detect PPD. In this study, we combined several orthogonal approaches demonstrating that digital biomarkers can be used for individualized classification of PPD with data collected from Fitbit using the AoURP (Figure 1). This work demonstrated that (1) the integration of data sources, including EHR and wearable data, proves valuable for PPD recognition; (2) using longitudinal and continuous wearable data across various pregnancy phases supports ML model development; and (3) combining these integrated data sources facilitates the creation of individualized ML models, which may outperform cohort-based models. As such, our findings uncovered a novel method for recognizing PPD and serve as a framework that can be leveraged to facilitate early PPD detection. Moreover, the significance of this research underscores the promise of individualized ML models for detecting PPD, which can be applied to other mental health disorders.

Figure 1. An overview of the analysis workflow to evaluate the potential for digital biomarkers in postpartum depression (PPD) recognition. (1) Develop and perform computational phenotyping of PPD and non-PPD cohorts; (2) merge with available digital biomarker data for each woman (heart rate, steps, physical activity, and calories burned); (3) classify each day as 1 of 4 periods (prepregnancy period, pregnancy, postpartum period without depression, or PPD); (4) build and assess individualized ML models testing random forest, generalized linear models, support vector machine, and k-nearest neighbor algorithms; (5) validate the machine learning (ML) approach in women without PPD; (6) compare individualized model performance in women with and without PPD; (7) determine variable importance for PPD recognition; (8) generate Shapley additive explanations dependence plots to assess the relationship between digital biomarkers and PPD; and (9) compare individualized ML models versus a cohort-based model for PPD detection. EHR: electronic health record.



Methods

Data Source and Platform

This study used the AoURP Registered Tier v6 data set. Study analysis was conducted using the AoURP Researcher Workbench cloud platform. All computational phenotyping, data processing, data analysis, and ML algorithms were conducted using R (R Foundation for Statistical Computing). Fitbit data collected in the AoURP adhere to a bring-your-own-device model, wherein participants who contribute their data are already in possession of a Fitbit device. The daily average HR, HR SD, minimum HR, quartile 1 HR, median HR, quartile 3 HR, and maximum HR were calculated using the Fitbit HR level table. The sum of steps was calculated using the Fitbit intraday steps table. Activity calories, calories burned during the basal metabolic rate (calories BMR), calories out, fairly active minutes, lightly active minutes, marginal

calories, sedentary minutes, and very active minutes were taken from the Fitbit activity summary table. Day-level data were calculated for each of the 4 periods: prepregnancy period, pregnancy, postpartum period, and PPD (or PPD equivalent). All digital biomarkers included in this analysis are passively tracked by Fitbit; however, calories BMR is a calculated digital biomarker based on self-reported height, weight, age, and gender [25].

Ethical Considerations

The protocol for the human participant research conducted was reviewed by the institutional review board of the AoURP (protocol 2021-02-TN-001). The institutional review board follows the regulations and guidance of the National Institutes of Health Office for Human Research Protections for all studies, ensuring that the rights and welfare of research participants are overseen and protected uniformly. Participants who contribute data to the AoURP have gone through an informed consent process with the option to withdraw at any time. Privacy is maintained by 1) storing data on protected computers, 2) researchers can't see information to directly identify participants, such as name or social security number, 3) researchers sign a contract they won't try to identify participants. Furthermore, the Researcher Workbench is only accessible to researchers through an institution with a signed Data Use Agreement and to researchers who complete the necessary training. If participants are asked (and decide) to go to an All of Us partner center for physical measurements to give blood, saliva, or urine samples, they are offered a one-time compensation of $25 in the form of cash, a gift card, or an electronic voucher.

In compliance with the Data and Statistics Dissemination Policy of the AoURP, counts of <20 cannot be presented to mitigate the risk of patient reidentification [26]. As the cohort of patients with PPD presented in this analysis comprised <20 patients, percentages were presented as percentage ranges (eg, instead of presenting the data as 53%, they were presented as 50%-55%). Publication of results in this manner has been approved by the AoURP Resource Access Board. Furthermore, race and ethnicity were not reported due to the limited sample size as requested by the AoURP Resource Access Board.

Computational Phenotyping

Identifying Women With PPD

Women with PPD were identified using the following three-fold approach: (1) selecting women with a diagnosis of PPD using the condition data and identifying women with a record of (2) pregnancy or (3) delivery who had been diagnosed with depression or had antidepressant drug exposure during the postpartum period.

The first branch of the 3-fold approach to creating a cohort of women with PPD was conducted using Observational Medical Outcomes Partnership concept IDs in the condition table based on the Observational Health Data Sciences and Informatics initiative in Multimedia Appendix 1 [27,28]. For both the second and third branches of the method, we first identified women with a record of delivery (using condition data) or pregnancy (using the condition and survey tables) based on concept IDs from previously published work in Multimedia Appendix 1. Next, the data were filtered on the earliest record of delivery or pregnancy to capture and analyze digital biomarker data during the prepregnancy period. To estimate the date of pregnancy or delivery (depending on which was available for that individual), the date observed in the EHR from the AoURP was adjusted by adding or subtracting 9 months, which is a typical pregnancy duration [29]. Our next step was to estimate the window of the postpartum period, which was defined as starting from the date of delivery and spanning 24 months after that date, to monitor depressive symptoms [30,31]. Consistent with other EHR computational phenotyping studies of PPD, individuals were also classified as being PPD positive if they had a diagnosis of depression in the condition table or antidepressant drug exposure within the postpartum window [32] (Multimedia Appendix 1). Specific concepts containing the terms *episode*, *remission*, *reactive*, *atypical*, *premenstrual*, *schizoaffective*, and *seasonal* were excluded when identifying individuals with a depression diagnosis as they would not appropriately capture women with a persistent depression during the postpartum period. If a woman in the PPD cohort showed records of depression diagnosis and antidepressant drug exposure, we selected the earliest record to be considered the index date. For women with pregnancy and delivery data available, the index date and data used were based on the delivery record as this provided an elevated level of confidence in defining the postpartum period and, subsequently, whether the depression diagnosis or antidepressant drug exposure occurred during the postpartum period. Finally, the final PPD cohort was generated by selecting unique women from each of the 3 branches of our approach.

Identifying Women Without PPD

Women without PPD were selected as a control group to validate our approach because they experienced the same periods as women in the PPD cohort with the exception of having diagnosed or inferred PPD (see the previous section). Therefore, our modeling approach could be tested in an identical fashion (see more details about ML models in the section titled *Individualized ML Models for Women Without PPD*). To establish a cohort of women without PPD, we applied an identical rationale to that of the second and third branches of our PPD phenotyping, as described previously. Subsequently, women with records indicating PPD or depression diagnosis during the postpartum period from the condition table or any instances of antidepressant drug use from the drug exposure table were excluded.

Data Preparation for Analysis and Individualized ML Models

To prepare the data for analysis and individualized ML models using wearable data, we first merged day-level data from Fitbit (HR, steps, physical activity, and calories burned; see Table S1 in Multimedia Appendix 2 [33-35] for more information on digital biomarkers) for each individual ranging from 2 years before to 30 days after the index date to capture their behavior before, during, and after pregnancy. Previous studies have demonstrated that HR, steps, and activity measurements from Fitbit are fairly accurate and can be used for research purposes [36,37]. The decision to choose measures related to HR instead of resting HR was based on the availability of data and the

consideration of having enough measurements for each individual to train ML models. Digital biomarker data were filtered on days of *compliant* data, which were characterized by (1) at least 10 hours of Fitbit wear time within a day and (2) between 100 and 45,000 steps, as seen in previous studies [38]. Individuals from the PPD cohort were excluded from individualized ML models if they had <50 days of total data.

Statistical Analysis

Assessing Variation in Digital Biomarkers Among Women

The *lme4* and *lmerTest* packages in R were used to construct hierarchical linear regression models aiming to assess the presence of noteworthy differences among women and examine the relationship between each period and digital biomarkers [39,40]. To assess whether there was a significant level of variation in digital biomarkers among individuals, we processed data to calculate the average value of each digital biomarker during each period (eg, average HR during the prepregnancy period, average HR during pregnancy, average HR during the postpartum period, and average HR during PPD) and conducted linear mixed-effects models with person ID as the random effect. One model was built for each digital biomarker, where the digital biomarker served as the outcome variable, the period was considered the independent variable, and person ID was incorporated as a random effect. The presence of significant variability among individuals was evaluated using the performance package at a significance level of .05 [41].

Interrupted Time-Series Analysis, Tukey Honest Significant Differences Test, and Digital Biomarker Directionality Assessment Between Periods

The interrupted time-series analysis (ITSA) was conducted using the *its.analysis* package in R with a significance level of .05 [42]. To compare whether there was a difference in digital biomarkers during different periods before, during, and after pregnancy, in addition to when patients experienced PPD, 4 periods were defined for each individual identified with PPD (prepregnancy period, pregnancy, postpartum period without depression [hereafter referred to as postpartum period], and postpartum period with depression [PPD]). The median duration of each period was 206 (IQR 154.50-313.50) days for the prepregnancy period, 258 (IQR 226-264) days for pregnancy, 42 (IQR 27.5-90) days for the postpartum period, and 42.5 (IQR 40.25-44.75) days for PPD. For each woman, a model was constructed for each digital biomarker, with 250 replications used for bootstrapping, which is a parameter of the *itsa.model()* function. Bootstrapping runs replications of the main model with randomly drawn samples and a trimmed median (10% removed); the *F* value is reported, and a bootstrapped *P* value is derived from it [42]. The dependent variable was the digital biomarker value, the *time* parameter was the date, and the interrupting variable was the period (prepregnancy period, pregnancy, postpartum period, and PPD). The mean and SD were calculated for each digital biomarker during each of the 4 periods for each woman. Furthermore, a Tukey honest significant difference (HSD) test was conducted to assess the statistical significance of the differences in each digital biomarker between each permutation of periods (PPD–prepregnancy period, PPD-pregnancy, PPD–postpartum period, postpartum period–prepregnancy period, postpartum period–pregnancy, and pregnancy–prepregnancy period) within each individual at a significance level of .05 [43]. Next, the percentage of women exhibiting a significant relationship was calculated for each digital biomarker in each group comparison (eg, PPD–prepregnancy period). To determine the overall trend in digital biomarker change between pairs of periods (eg, PPD and prepregnancy period, PPD and pregnancy, and PPD and postpartum period), the average difference across all individuals was computed for each digital biomarker. This average also included nonsignificant differences as they still contributed insights into the directionality of digital biomarkers during those periods even if the differences were not statistically significant. Finally, a 2-sided unpaired *t* test (2-tailed) at a significance level of .05 was conducted to assess the statistical significance of the net difference compared to 0, with positive change defined as an average value of >0 and negative change defined as an average value of <0. The outcomes were visualized in a heat map using the *ggplot2* package in R. Percentages were represented as percentage ranges to preserve patient confidentiality, with the upper value of each range depicted in the heat maps (eg, 62% would fall within the 60%-65% range, and 65% would be displayed in the heat map).

Evaluating Health Care–Seeking Behavior

Health care–seeking behavior was assessed by looking at the number of visits recorded for each woman during the postpartum period (ie, ranging from the date of delivery to 30 days after the index date for each woman). The number of visits was determined by counting the number of rows in the visit occurrence table in the AoURP. We subsequently conducted an unpaired 2-sided Wilcoxon test with a significance level of .05 to determine whether the medians exhibited a significant difference between the PPD and non-PPD cohorts.

We also examined the proportion of women who adhered to the recommendation set by the American College of Obstetricians and Gynecologists, which advised women to attend at least one visit within the initial 6 weeks of the postpartum period. Of note, this guideline was updated in 2018 and now recommends a postpartum visit within the first 3 weeks following delivery [33]. However, we used the pre-2018 guideline in our analysis because the AoURP cohort includes individuals enrolled before 2018. The percentages of women who attended postpartum visits within the first 6 weeks in the PPD and non-PPD cohorts were compared using a 2-proportion *z* test at a significance level of .05. The exact percentage of women in the PPD cohort, in addition to the exact counts used to calculate the percentages, was obfuscated to maintain patient privacy.

Comparing Self-Reported and Gold-Standard Weight Measurements

Weight measurements were queried in AoURP using the measurements table (Observational Medical Outcomes Partnership concept ID 3025315). Self-reported and gold-standard weight measurements were distinguished by referencing the *src_id* column, indicating a physical measurement (self-reported) as opposed to measurements

obtained from an EHR site (gold standard). Subsequently, we identified the self-reported and gold-standard weight measurements with the shortest time interval for each woman. Only measurements taken within a period of <30 days were considered to ensure that the measurements were closely aligned and not too distant. The median and IQR of self-reported and gold-standard measurements were calculated and compared using a paired 2-sided Wilcoxon test at a significance level of .05. This process was repeated in the PPD and non-PPD cohorts.

Comparing Weight Across Periods of Pregnancy

Weights across different periods of pregnancy (prepregnancy period, pregnancy, postpartum period, and PPD [or PPD equivalent for those without PPD]) were computed in the PPD and non-PPD cohorts using linear mixed-effects models in the *lme4* package in R, with weight serving as the outcome variable, period as the independent variable, and person ID as the random effect. The results were evaluated at a significance level of .05. For women in the PPD cohort, the PPD period was used as the reference as it was the period of interest for understanding weight change. Similarly, the PPD-equivalent period was used as the reference for women in the non-PPD cohort. We further calculated the estimated means of weight across periods using the *emmeans* package in R for both the PPD and non-PPD cohorts.

Comparing Weight Retention in the PPD and Non-PPD Cohorts

To assess weight retention among women who experienced PPD compared to those without PPD, we first calculated the median weight of each woman during the prepregnancy period. Second, we identified the weight measurement during the postpartum period that was closest in value to the median prepregnancy weight on an individual basis. Third, the time difference in days was computed between the date of the weight measurement and the onset of pregnancy for each individual. Finally, we determined the median and IQR for the time difference in days mentioned in step 3 (ie, difference in days between the date of the weight measurement during the postpartum time period that was closest in value to the median prepregnancy weight for each individual) and subsequently conducted an unpaired 2-sided Wilcoxon test to assess the difference in medians at a significance level of .05 between women in the PPD and non-PPD cohorts.

Building ML Models

Individualized ML Models for Women in the PPD Cohort

Individualized ML models were developed with the objective of determining the potential of digital biomarkers to differentiate among 4 distinct pregnancy phases: prepregnancy period, pregnancy, postpartum period without depression (ie, postpartum period), and postpartum period with depression (ie, PPD). Specifically, we sought to assess whether we could develop ML models for each woman to make a prediction to classify a day of Fitbit data as falling during the prepregnancy, pregnancy, postpartum, or PPD period based on behavioral and biometric data captured by digital biomarkers on Fitbit. In other words, the models tested whether there was a unique digital signature associated with each period of pregnancy in an individualized manner. Therefore, multinomial models were developed with period as the outcome with all 16 digital biomarkers as the features in the model (see Table S1 in Multimedia Appendix 2 for a list of the digital biomarkers included). Initially, our intention was to examine the model's capacity to discriminate between periods with and without PPD, thereby constructing binomial classification models. However, we recognized the hierarchical nature of the data with repeated measurements (multiple days of data) during the prepregnancy, pregnancy, and postpartum time frames. Consequently, due to the repetitive nature of our data, we opted for constructing multinomial ML models to effectively discern among the 4 identified periods, where the PPD period was treated as both a period and a diagnosis. We were then able to focus on the PPD period by (1) constructing a confusion matrix to assess model performance for the PPD period at an individual level and (2) performing variable importance (see the following *Variable Importance* sections) for the PPD period.

To build intraindividual models, the data were filtered on each woman, where they were considered PPD negative ranging from 2 years before to 15 days before the index date and PPD positive from 14 days before to 30 days after the index date. We selected 14 days preceding the index date as the first day of being positive for PPD because the criteria for diagnosis state that patients must display 5 depressive symptoms lasting 2 weeks [44]. The time frame of 30 days following the index date was chosen because some individuals in the PPD cohort received antidepressant medication on the day of their diagnosis, which can begin to take effect after approximately 4 weeks of use [45]. For each individual, the data were centered and scaled before building models using 3 repeats of 10-fold cross-validation and a tune length of 5 with random forest (RF), generalized linear models (GLMs), support vector machine (SVM), and k-nearest neighbor (KNN) as these algorithms have been used in previous studies assessing depression using wearables [15,46]. Of note, no bootstrapping was performed as part of the individualized ML workflow. Models were built using the *Caret* package in R and evaluated using a combination of the κ statistic and multiclass area under the receiver operating characteristic curve (mAUC), which are standard metrics for classification ML models [47-50]. Model performance for each period was further assessed using a confusion matrix, which calculated sensitivity, specificity, precision, recall, and F_1-score [50].

Comparing Individualized ML Model Performance Between Women With a History of Depression Before or During Pregnancy

To initially ascertain the presence of depression history before or during pregnancy within the PPD cohort, we determined the date of delivery (using condition data) or the date of pregnancy (using condition and survey data) based on the concept IDs detailed in Multimedia Appendix 1. Depending on the available data for each woman, the date of pregnancy was calculated by subtracting 9 months from the date of delivery, whereas the date of delivery was calculated by adding 9 months to the date of pregnancy, representing a standard pregnancy duration [29]. In cases in which both delivery and pregnancy records existed,

priority was given to the date of delivery due to its heightened reliability.

For the evaluation of individualized ML model performance within the PPD cohort concerning women with a history of depression, the cohort was categorized into four subgroups encompassing (1) no previous depression history, (2) depression before pregnancy, (3) depression during pregnancy, and (4) depression both before and during pregnancy. To examine potential disparities in individualized ML model performance, a 2-sided unpaired *t* test was conducted with a significance threshold of .05. This analysis was executed to compare the no-depression-history group with the groups of women exhibiting depression before, during, or both before and during pregnancy. Sensitivity, specificity, precision, recall, and F_1-score metrics were subjected to this statistical comparison process.

Individualized ML Models for Women Without PPD

To construct individualized ML models for women in the non-PPD cohort, we implemented an analogous approach to the one used for women in the PPD cohort, where an ML model was built for each woman with *period* as the multinomial outcome. It is worth noting that women without PPD would not have a fourth period (ie, postpartum period with depression in women with PPD) as they did not experience PPD. To ensure comparability and effectively gauge model performance between women with and without PPD, we created a PPD-equivalent period for the non-PPD cohort mirroring the PPD period. Considering that the median time to diagnose PPD was found to be 83 days following delivery, we ensured uniformity by setting the index date of the PPD-equivalent period at 83 days after delivery. As we established an index date aligned with that of the PPD cohort, the interval of 14 days before the index date was not considered as the PPD-equivalent period for these women because they did not actually experience PPD. The goal was to validate any observed alterations in the PPD cohort by investigating whether there were any changes in the digital signature between the postpartum and PPD-equivalent periods, which should not exist given that these women did not experience PPD. Subsequently, individualized ML models were constructed in a manner akin to those in the PPD cohort using the RF algorithm (as this algorithm yielded optimal results in the PPD cohort) using 3 repetitions of 10-fold cross-validation and a tuning length of 5. Similar to the approach developed for women in the PPD cohort, model performance was evaluated using sensitivity, specificity, precision, recall, and F_1-score [49,50]. Models were not assessed using mAUC or κ as model performance only decreased in the PPD-equivalent period and not in the prepregnancy, pregnancy, or postpartum periods compared to those in the PPD cohort.

Comparing Individualized ML Model Performance for Women in the PPD and Non-PPD Cohorts

For comparing the performance of individualized ML models in the PPD cohort to those in the non-PPD cohort, we performed a 2-sided unpaired *t* test with a significance level of .05.

Variable Importance

Shapley Additive Explanations Approach

We used the RF ML models to generate a ranking of digital biomarkers for each individual as these models had the best performance. Following that, Shapley values were computed for each measurement within each individualized model for the PPD class using the *iml* package in R [51]. To determine the feature ranking within individual models, we computed the average absolute Shapley values across all measurements for each digital biomarker and sorted the rankings from largest to smallest. We then tallied the number of models in which each biomarker ranked among the top 5 most predictive for the PPD class to produce an overall ranking of digital biomarkers. Furthermore, we determined the most predictive feature of PPD by totaling the number of models in which each digital biomarker ranked as the top predictor for the PPD class.

Permutation Approach

To enhance the robustness of our approach, variable importance was also computed using a permutation-based method in the *Caret* package in R [50]. Subsequently, the features were sorted based on the magnitude of values assigned for the variable importance regarding the PPD class. Using a similar methodology as with Shapley additive explanations (SHAP), we tabulated the number of models in which each digital biomarker ranked among the top 5 most predictive for the PPD class, yielding a comprehensive ranking of digital biomarkers. The frequency with which each feature ranked as the foremost predictive digital biomarker was also recorded for the PPD class.

SHAP Dependence Plots

SHAP dependence plots were generated using the *gpplot2* package in R [52]. For each individual, plots were generated by graphing the Shapley value against the corresponding actual value for the digital biomarker. Given that the outcome of the models was multinomial (prepregnancy period, pregnancy, postpartum period, or PPD), 3 separate SHAP dependence plots were generated for each individual using calories BMR data during PPD with one other period (ie, one plot for the prepregnancy and PPD periods [referred to as prepregnancy vs PPD], one plot for pregnancy and PPD [referred to as pregnancy vs PPD], and one plot for the postpartum and PPD periods [referred to as postpartum vs PPD]) to more easily analyze the relationship between calories BMR in a binomial context between PPD and one other period. This process was repeated for women in the non-PPD cohort in a similar fashion to those in the PPD cohort, specifically, PPD-equivalent versus prepregnancy period (prepregnancy vs PPD-equivalent), PPD-equivalent versus pregnancy (pregnancy vs PPD-equivalent), and PPD-equivalent versus postpartum (postpartum vs PPD-equivalent). The Pearson correlation coefficient and its corresponding *P* value were computed at a significance level of .05, followed by calculating the percentages of women with and without a significant correlation. If a significant correlation was observed, we further determined its direction (positive or negative) and calculated the percentages of women with a positive or negative correlation. The overall consensus regarding the relationship was determined by

comparing the percentage of positive and negative correlations for each digital biomarker across all individuals, thereby identifying which direction had a greater rate. In cases in which the proportion of women with a significant correlation was <40%, the direction was not assessed due to the small sample size, which may not be representative of the population.

Building an ML Model for PPD Using a Cohort-Based Approach

For the construction of an ML model that assessed whether a woman had PPD, our focus was on using the PPD and PPD-equivalent periods sourced from both the PPD and non-PPD cohorts. We proceeded to develop a binomial RF classification model in which 75% of individuals from each cohort were designated for the training set and the remaining 25% were assigned to the test set using the *Caret* package in R [50]. To ensure the reliability of model performance assessment, we diligently executed train and test set divisions based on individual person IDs, thereby preventing any overlap of women between the 2 sets that could potentially distort the results [53]. The model's target outcome pertained to a binary classification of whether an individual exhibited PPD relying on all 16 digital biomarkers as input (refer to Table S1 in Multimedia Appendix 2 for a comprehensive description of the digital biomarkers used). The data were normalized through centering and scaling procedures. Notably, repeated cross-validation was omitted due to the presence of repeated measurements stemming from various person IDs. The model's construction integrated a tune length of 5. The models were evaluated using the same κ and area under the receiver operating characteristic curve metrics (not multiclass in this instance as the outcome was binary). Subsequently, a confusion matrix was generated to calculate sensitivity, specificity, precision, recall, and F_1-score [47-50].

Results

Descriptive Statistics

Through computational phenotyping in the AoURP, a patient cohort of women who gave birth with PPD (n<20) and without PPD (n=39) provided valid Fitbit data (Figure 2). The median age in the PPD cohort was 35.60 (IQR 32.83-37.36) years compared to that in the non-PPD cohort, which was 33.60 (IQR 30.72-35.56) years. The median and IQR were calculated for each digital biomarker across all women in the PPD and non-PPD cohorts (Table 1). In both the PPD and non-PPD cohorts, we computed the median number of days with digital biomarker data during the prepregnancy, pregnancy, postpartum, and PPD (or PPD-equivalent) periods and the corresponding IQRs (additional details about the PPD-equivalent period, a similar fourth period for those without PPD, can be found in the *Methods* section; Table 1). Briefly, the digital biomarkers included in this analysis were daily average HR, HR SD, minimum HR, quartile 1 HR, median HR, quartile 3 HR, maximum HR, sum of steps, activity calories, calories BMR, calories out, fairly active minutes, lightly active minutes, marginal calories, sedentary minutes, and very active minutes (see the descriptions in Table S1 in Multimedia Appendix 2).

Figure 2. A schematic of postpartum depression (PPD) computational phenotyping.

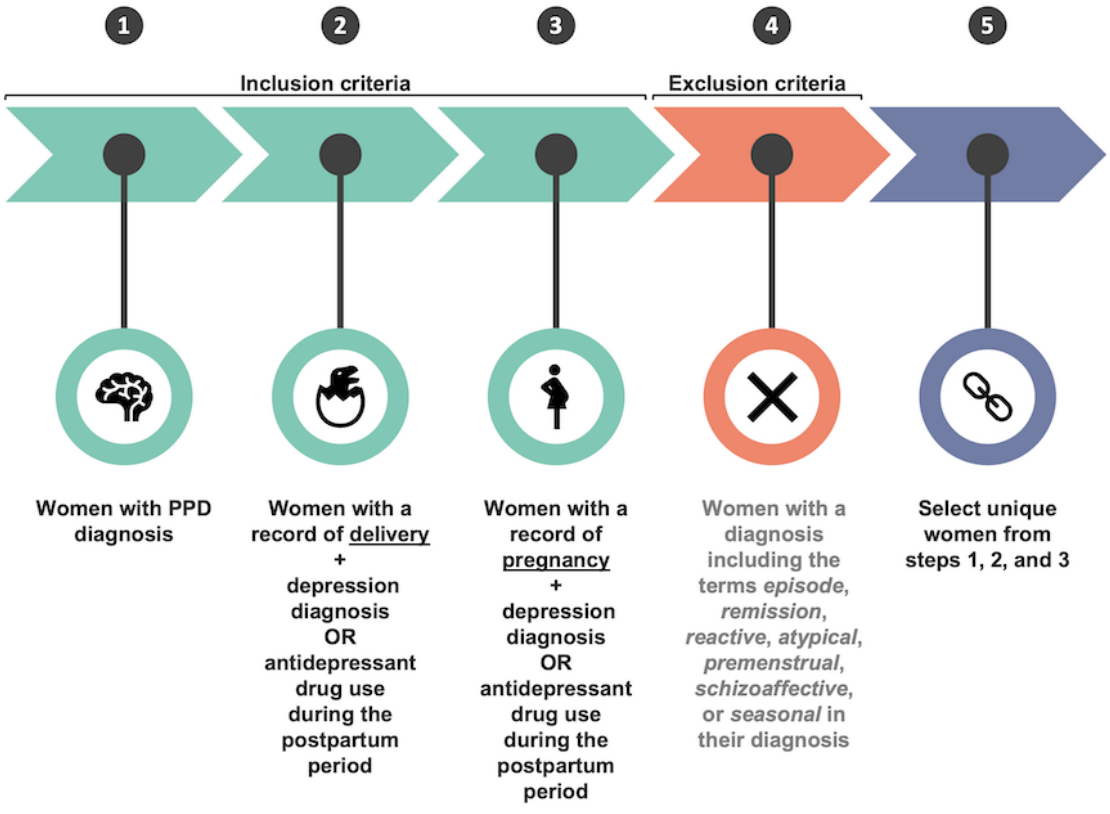


Table 1. Descriptive statistics of the postpartum depression (PPD) and non-PPD patient cohorts in the All of Us Research Program.

Descriptive statistics	PPD (n<20), median (IQR)	Non-PPD (n=39), median (IQR)
Age (y)	35.60 (32.83-37.36)	33.60 (30.72-35.56)
Digital biomarker		
Average HR[a] (bpm)	74.23 (68.36-80.66)	78.31 (72.32-83.97)
HR SD (bpm)	12.18 (10.58-14.05)	12.70 (10.72-15.12)
Minimum HR (bpm)	54.00 (49.00-60.00)	57.00 (52.00-61.00)
Quartile 1 HR (bpm)	64.00 (59.00-71.00)	68.00 (62.00-74.00)
Median HR (bpm)	72.00 (66.00-78.00)	76.00 (70.00-82.00)
Quartile 3 HR (bpm)	81.00 (74.00-88.00)	85.00 (78.00-92.00)
Maximum HR (bpm)	124.00 (117.00-135.00)	127.00 (119.00-141.00)
Sum steps	7567.50 (4884.00-10536.25)	7352.00 (4838.00-10834.00)
Activity calories	989.00 (742.75-1263.00)	964.00 (684.00-1275.00)
Calories burned during BMR[b]	1466.00 (1379.00-1539.00)	1390.00 (1340.00-1496.00)
Calories out	2236.00 (2012.00-2483.25)	2180.00 (1925.00-2465.50)
Fairly active minutes	9.00 (0.00-24.00)	8.00 (0.00-23.00)
Lightly active minutes	245.00 (189.00-315.00)	245.00 (187.50-310.00)
Marginal calories	501.00 (349.00-665.00)	489.00 (322.00-680.00)
Sedentary minutes	646.00 (563.00-741.00)	710.00 (607.00-880.50)
Very active minutes	2.00 (0.00-18.00)	4.00 (0.00-21.00)
Number of days in each period		
Prepregnancy period	206.00 (154.50-313.50)	227.00 (109.50-340.75)
Pregnancy	258.00 (226.00-264.00)	221.00 (129.00-269.50)
Postpartum period	42.00 (27.50-90.00)	72.00 (46.00-82.00)
PPD	42.50 (40.25-44.75)	29.00 (14.50-31.00)

[a]HR: heart rate.

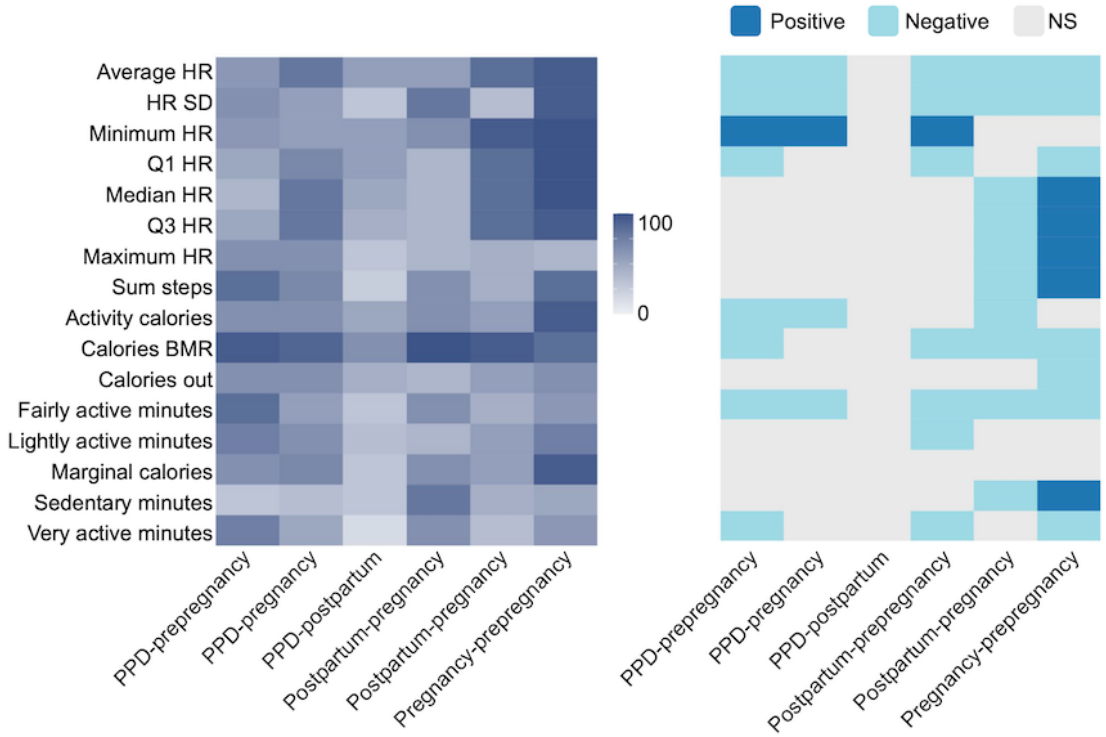
[b]BMR: basal metabolic rate.

Digital Biomarker Comparison Across Periods of Pregnancy Revealed Altered Profiles and Heterogeneity Among Women

Because of the known heterogeneity in depressive symptoms, we hypothesized that variability in digital biomarkers may exist across individuals in the PPD cohort [54]. To test this hypothesis, we conducted linear mixed-effects models for each digital biomarker in women with PPD, where we found that the random effect of person ID was significant (P<.001) for all digital biomarkers, suggesting meaningful variability across individuals (Table S2 in Multimedia Appendix 2). These results, coupled with a smaller cohort sample size, prompted us to perform subsequent analyses using an intraindividual approach.

In women with PPD, we next sought to compare whether there was a difference in digital biomarkers across different periods of pregnancy: prepregnancy period, pregnancy, postpartum period, and PPD (where PPD represents both a period and a diagnosis). Therefore, an intraindividual ITSA and Tukey HSD test were conducted for each digital biomarker. Because of the physiological changes associated with pregnancy, such as increases in blood and stroke volume, in addition to the behavioral fluctuations that occur during PPD, such as a loss of energy and psychomotor retardation, we hypothesized that all digital biomarkers (those related to HR, steps, physical activity, and calories burned) would be altered across the prepregnancy, pregnancy, postpartum, and PPD periods [44,55-57]. ITSA results supported our hypothesis and demonstrated a significant difference in all digital biomarkers across periods in most women with PPD (Table S3 in Multimedia Appendix 2). Consistent with ITSA findings, Tukey HSD results showed that several digital biomarkers were significantly altered between PPD and other periods (prepregnancy, pregnancy, and postpartum periods; Figure 3). We further observed various trends in digital biomarkers between pairs of periods (ie, PPD and prepregnancy period, PPD and pregnancy, and PPD and postpartum period; Figure 3 and Table S4 in Multimedia Appendix 2).

Figure 3. Digital biomarkers vary across different periods of pregnancy among women with postpartum depression (PPD). The percentage of women in the PPD cohort exhibiting a significant difference in digital biomarker values between each pair of periods (left [represented by 0-100]) and the direction of their relationship (right). The x-axis illustrates a comparison of Tukey honest significant differences (HSD) between 2 periods of interest, representing the subtraction of digital biomarker values between the first and second periods. Tukey HSD tests were individually conducted for each woman's data, and the percentage showing a significant relationship was calculated and presented on the heat map. The heat map on the right illustrates the overall relationship between the digital biomarker during the 2 periods of interest among the women who exhibited a significant relationship (as indicated by the percentage shown on the left heat map), with the period listed second serving as the reference. In summary, the findings indicated that digital biomarkers undergo significant alterations across different periods of pregnancy on an individual basis. Calories BMR: calories burned during the basal metabolic rate; HR: heart rate; NS: not significant; Q1: quartile 1; Q3: quartile 3.



Individualized ML Models Effectively Differentiated PPD From Alternative Periods of Pregnancy

Having seen that digital biomarkers were significantly altered across multiple periods of pregnancy in women with PPD, we surmised that individualized multinomial ML models could accurately distinguish between our 4 periods of pregnancy (prepregnancy period, pregnancy, postpartum period, or PPD; Figure 3 and Tables S3 and S4 in Multimedia Appendix 2). Therefore, we sought to assess whether ML models for each woman could accurately classify an unknown day of Fitbit data as falling during the prepregnancy, pregnancy, postpartum, or PPD period based on behavioral and biometric data captured by digital biomarkers on Fitbit. In essence, the models examined whether there existed a distinct digital signature linked to each pregnancy period in an individualized fashion. To probe this hypothesis, intraindividual ML models were generated using RF, GLM, SVM, and KNN to conclude which algorithm would yield the best-performing results. Models were assessed using a combination of the mAUC and κ, which are 2 frequently used metrics [48,58]. After averaging the mAUC for individual models within each algorithm, the results revealed that RF models performed the best, followed by GLM, SVM, and then KNN, with an average mAUC of 0.85, 0.82, 0.75, and 0.74, respectively (Table 2). Assessing models in a similar fashion using another metric, κ, yielded concordant results for RF (0.80), GLM (0.74), SVM (0.72), and KNN (0.62) model performance, suggesting that the RF algorithm had the best performance and should be used going forward (Table 2).

Table 2. Individualized random forest (RF) models exhibited the best performance for multinomial period classification.

Algorithm	mAUC[a], mean (SD)	κ, mean (SD)
Random forest	0.85 (0.09)	0.80 (0.15)
Generalized linear model	0.82 (0.09)	0.74 (0.16)
Support vector machine	0.75 (0.10)	0.72 (0.16)
k-nearest neighbor	0.74 (0.10)	0.62 (0.19)

[a]mAUC: multiclass area under the receiver operating characteristic curve.

As our analysis aimed to assess the potential of digital biomarkers for personalized classification of PPD, we sought to further examine each RF model's performance via a confusion matrix. Thus, the average sensitivity, specificity, precision,

recall, and F_1-score were calculated across all individual models, where the results for the PPD class were 0.79, 0.95, 0.84, 0.79, and 0.81, respectively (Figure S1 in Multimedia Appendix 2). The same metrics for the prepregnancy, pregnancy, and postpartum periods were also calculated (Figure S1 in Multimedia Appendix 2).

To ensure the widespread applicability of these algorithms to a diverse range of women, we did not exclude individuals with a history of depression either before or during pregnancy. Therefore, we sought to determine whether having depression before or during pregnancy impacted individual model performance, specifically for recognizing the PPD class. To answer this question, we computed the average sensitivity, specificity, precision, recall, and F_1-score within the group of women experiencing PPD categorized based on their depression history: (1) no previous history of depression, (2) history before pregnancy, (3) history during pregnancy, or (4) history both before and during pregnancy. Notably, the findings revealed no statistically significant variations in any of these metrics between women with a history of depression during the prepregnancy or pregnancy periods and those without such a history (Figure S2 in Multimedia Appendix 2). Promisingly, this suggests the potential for a forthcoming technology focused on detecting PPD through digital biomarkers to be relevant for women with or without a previous history of depression before or during pregnancy.

Individualized ML Models for PPD Recognition Were Specific

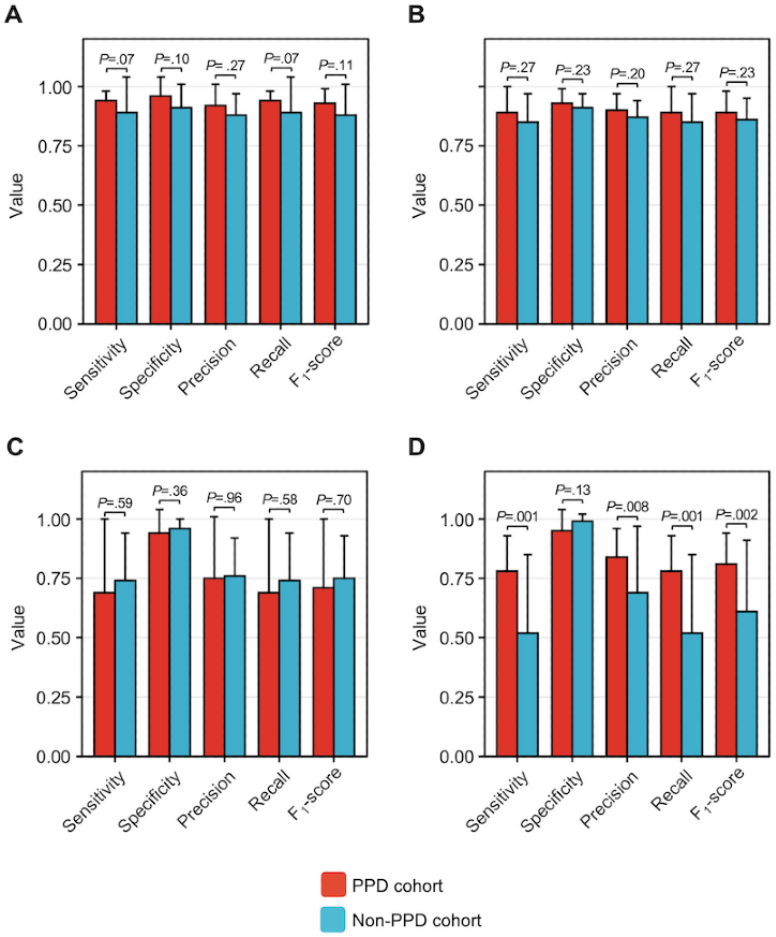
To validate our approach of using digital biomarkers in individualized ML models for PPD detection, we aimed to test our strategy in a cohort of women who had given birth but did not experience PPD. We chose women without PPD as a control group for validation because they experienced the same 3 phases of pregnancy (prepregnancy period, pregnancy, and postpartum period) as women in the PPD cohort with the exception of PPD. Given that women without PPD did not have a distinct PPD-specific period as observed in the PPD cohort, we introduced a fourth time segment in the non-PPD cohort (the PPD-equivalent period). Following the same ML pipeline as for the PPD cohort, individualized RF models were built for women in the non-PPD cohort. If our conjecture held, we anticipated observing elevated model metrics during the prepregnancy and pregnancy periods followed by diminished performance in the postpartum and PPD-equivalent time segments. This expectation arose from the idea that digital biomarkers remain unaltered during the postpartum and PPD-equivalent periods, resulting in the model's inability to differentiate between them.

In line with our hypothesis, the sensitivity, specificity, precision, recall, and F_1-scores substantiated that ML models effectively identified the prepregnancy (0.89, 0.91, 0.88, 0.89, and 0.88, respectively) and pregnancy (0.85, 0.91, 0.87, 0.85, and 0.86, respectively) time intervals through digital biomarkers (Table 3). When compared to model performance in the prepregnancy and pregnancy periods, there was no significant reduction in model performance during the postpartum period (0.74, 0.96, 0.76, 0.74, and 0.75, respectively); however, a noticeable decline in performance was observed during the PPD-equivalent period (0.52, 0.99, 0.69, 0.52, and 0.61, respectively; Table 3). To further assess potential variations in the classification performance between the PPD and PPD-equivalent periods, we carried out a *t* test comparing the average sensitivity, specificity, precision, recall, and F_1-score between the PPD and non-PPD cohorts for these periods. The findings indicated a statistically significant decrease in sensitivity, precision, recall, and F_1-score when predicting the PPD-equivalent period in the non-PPD cohort as opposed to predicting the PPD period in the PPD cohort (Figure 4). On the other hand, specificity remained largely unchanged (Figure 4). The decrease in performance among individualized ML models in the PPD-equivalent period implies that the models were unable to accurately classify the PPD-equivalent period, which was expected as there was no actual distinction between the postpartum and PPD-equivalent periods for these women. Collectively, these outcomes helped demonstrate the specificity of our approach in identifying PPD, reinforcing the agreement that personalized models using digital biomarkers can indeed effectively recognize PPD.

Table 3. Machine learning (ML) models did not accurately detect the postpartum depression (PPD)–equivalent period in women without PPD.

Time period and metric	Value, mean (SD)
Prepregnancy period	
Sensitivity	0.89 (0.15)
Specificity	0.91 (0.10)
Precision	0.88 (0.09)
Recall	0.89 (0.15)
F_1-score	0.88 (0.13)
Pregnancy period	
Sensitivity	0.85 (0.12)
Specificity	0.91 (0.06)
Precision	0.87 (0.07)
Recall	0.85 (0.12)
F_1-score	0.86 (0.09)
Postpartum period	
Sensitivity	0.74 (0.20)
Specificity	0.96 (0.04)
Precision	0.76 (0.16)
Recall	0.74 (0.20)
F_1-score	0.75 (0.18)
PPD-equivalent period	
Sensitivity	0.52 (0.33)
Specificity	0.99 (0.03)
Precision	0.69 (0.28)
Recall	0.52 (0.33)
F_1-score	0.61 (0.30)

Figure 4. Individualized machine learning models for postpartum depression (PPD) recognition outperformed those in women without PPD detecting the PPD-equivalent period. The sensitivity, specificity, precision, recall, and F1-score were calculated across individual random forest models for women in the PPD and non-PPD cohorts for the prepregnancy (A), pregnancy (B), postpartum (C), and PPD or PPD-equivalent periods (D). Individualized model performance was not significantly different regarding sensitivity, specificity, precision, recall, and F1-score for predicting the prepregnancy, pregnancy, or postpartum periods between women in the PPD and non-PPD cohorts. Individualized model performance was reduced for sensitivity, precision, recall, and F1-score, whereas specificity did not differ between the PPD and non-PPD cohorts. Data are expressed as mean and SD.



Calories BMR Was the Most Predictive Digital Biomarker of PPD

To elucidate which digital biomarkers were most predictive of the PPD class, we performed SHAP to explain individual predictions for each digital biomarker across all RF intraindividual models [59]. Features were sorted based on their predictive value for the PPD class within each individual model, and subsequently, the occurrences of each digital biomarker ranking in the top 5 across all intraindividual models were tallied, where an example beeswarm plot for one woman is shown in Figure S3 in Multimedia Appendix 2. This process aimed to identify whether any digital biomarkers consistently played a crucial role in predicting the PPD class. The results showed that the 5 features most frequently ranked in the top 5 were calories BMR, average HR, quartile 1 HR, lightly active minutes, and minimum HR (Table 4). Interestingly, calories BMR ranked in the top 5 features predictive of the PPD class in 95% to 100% of the models and was the number 1 rated digital biomarker in 80% to 85% of the individualized models (Table 4).

To add a layer of robustness to our approach assessing which features were most predictive of the PPD class, the variable importance of each digital biomarker was also calculated using a permutation approach [60]. Consistent with our findings obtained using SHAP, the top 5 digital biomarkers for the PPD class were calories BMR, average HR, quartile 1 HR, minimum HR, and lightly active minutes (Table 4). Calories BMR again ranked in the top 5 digital biomarkers predictive of the PPD class 95% to 100% of the time and ranked number one 95% to 100% of the time (Table 4).

Because of the intriguing observation that calories BMR was highly predictive of PPD across all models, we sought to better understand its relationship with the PPD class in our models using SHAP dependence plots to visualize and calculate the Pearson correlation coefficient between the PPD period and the prepregnancy, pregnancy, or postpartum periods (see Figure S4A-C in Multimedia Appendix 2 for example plots from individual women). Across all individual SHAP dependence plots of calories BMR filtered in the prepregnancy versus PPD periods, our initial observation revealed that 95% to 100% of women exhibited a significant Pearson correlation coefficient (Figure 5A). Of these women, 60% to 65% showed a positive relationship, indicating an elevated level of calories BMR during the PPD period relative to the prepregnancy period (Figure 5A). Compared to the prepregnancy versus PPD period, it was

observed that 75% to 80% and 85% to 90% of individualized SHAP dependence plots of calories BMR during the pregnancy versus PPD and postpartum versus PPD periods exhibited a significant Pearson correlation coefficient, respectively (Figure 5A). Of those, 60% to 65% and 85% to 90% of women during the pregnancy versus PPD and postpartum versus PPD periods demonstrated a negative relationship, respectively, suggesting that a decrease in calories BMR relative to the pregnancy and postpartum periods was predictive of PPD (Figure 5A). SHAP dependence plots were also generated for individualized models of the other top 4 digital biomarkers predictive of PPD (average HR, quartile 1 HR, minimum HR, and lightly active minutes) in the prepregnancy versus PPD, pregnancy versus PPD, and postpartum versus PPD periods (Figure 5A). Notably, during the prepregnancy versus PPD periods, half of the women exhibited a positive relationship in plots of lightly active minutes, indicating an increase in lightly active minutes associated with PPD in those models (Figure 5A). To examine the rise in lightly active minutes relative to other digital biomarkers of physical activity (sedentary minutes, fairly active minutes, and very active minutes), we calculated the ratio of the number of lightly active minutes to each of the 3 other digital biomarkers of physical activity across all individuals. In this case, we observed that the average ratios of lightly active minutes to sedentary minutes, fairly active minutes, and very active minutes were 0.35 (SD 0.49), 17.7 (SD 4.92), and 21.72 (SD 5.11), respectively (Figure S5 in Multimedia Appendix 2).

Table 4. The variable importance rankings demonstrated that calories burned during the basal metabolic rate (calories BMR) were the most predictive digital biomarker of the postpartum depression (PPD) class.

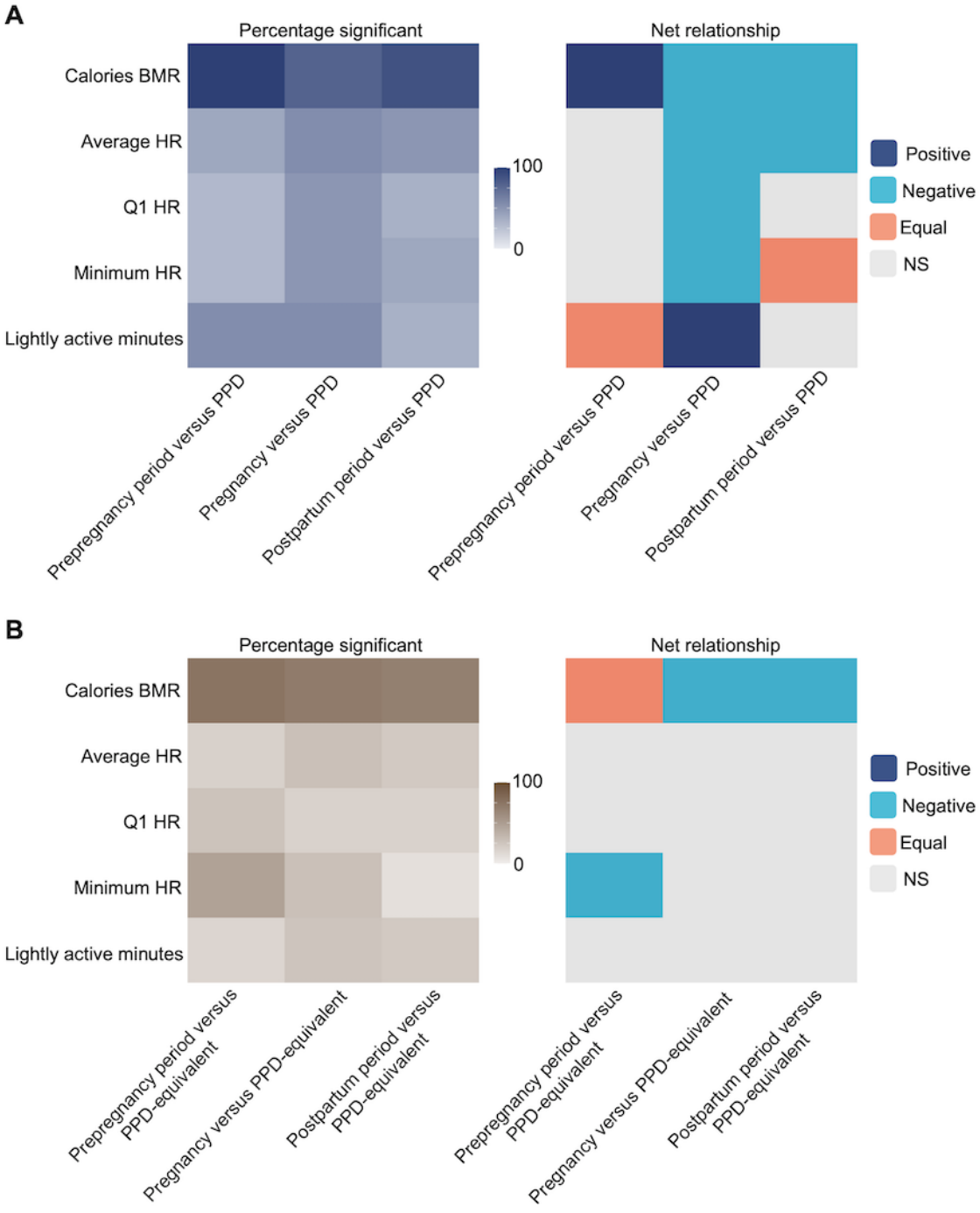
Method and digital biomarker	Percentage ranked top 5	Percentage ranked number 1
SHAP[a]		
Calories BMR	100	85
Average HR[b]	40	0
Quartile 1 HR	40	0
Lightly active minutes	35	10
Minimum HR	35	0
Sedentary minutes	0	10
Sum of steps	0	10
Permutation		
Calories BMR	100	100
Average HR	65	0
Quartile 1 HR	60	0
Minimum HR	50	0
Lightly active minutes	40	0

[a]SHAP: Shapley additive explanations.

[b]HR: heart rate.

For a more comprehensive evaluation of the connection between calories BMR and PPD, we also crafted SHAP dependence plots from individualized ML models for women without PPD. When first assessing the number of women with a significant correlation in SHAP dependence plots of the prepregnancy versus PPD-equivalent, pregnancy versus PPD-equivalent, and postpartum versus PPD-equivalent periods, the results showed that 75% to 80%, 70% to 75%, and 65% to 70% of women had a significant relationship, respectively (Figure 5B). Of those, there was an equal number of women with a positive and negative relationship in the prepregnancy versus PPD-equivalent periods compared to the PPD cohort, where most women (60%-65%) exhibited a positive relationship (Figure 5B). This implies that, among women in the PPD cohort, an escalation in calories BMR corresponds to a higher likelihood of PPD when compared to the prepregnancy period (Figures 5A and 5B). On the other hand, in the non-PPD cohort, there was no uniform pattern of association between calories BMR during the prepregnancy and the PPD-equivalent periods across all women, highlighting the distinctive nature of our observation. During the pregnancy versus PPD-equivalent and postpartum versus PPD-equivalent time frames, 80% to 85% and 75% to 80% of women, respectively, exhibited a significant correlation in SHAP dependence plots between calories BMR and Shapley values (Figure 5B). As anticipated, this follows a similar pattern to women in the PPD cohort (Figure 5A). These findings implied that a reduction in calories BMR compared to the pregnancy or postpartum periods is linked to the PPD (or PPD-equivalent) periods (Figures 5A and 5B).

Figure 5. The direction of digital biomarkers in machine learning models for postpartum depression (PPD) classification was heterogeneous. (A) The percentage of women in the PPD cohort with a significant Pearson correlation (left) and the net relationship (right) for the top 5 overall ranked digital biomarkers for PPD classification. (B) The percentage of women in the non-PPD cohort with a significant Pearson correlation (left) and the net relationship (right) for the top 5 overall ranked digital biomarkers for PPD-equivalent classification. The proportion of women showing a significant Pearson correlation coefficient between Shapley additive explanations (SHAP) values and digital biomarkers varied in both the PPD and non-PPD cohorts. The x-axis illustrates the comparison of 2 periods, with the first period as the reference, whereas the shading indicates the percentage (0%-100%) of women showing a significant relationship (A) or the net relationship among those with a significant relationship (B). SHAP dependence plots were generated for each woman on an individual basis. For instance, the upper left tile in A presenting prepregnancy versus PPD and Calories BMR indicates the percentage of women who showed a significant correlation on SHAP dependence plots of calories burned during the basal metabolic rate (calories BMR) between the prepregnancy and PPD periods. In B, the upper left tile of the heat map for prepregnancy versus PPD and Calories BMR illustrates that most women in the PPD cohort showed a positive relationship (elevated SHAP values with increases in calories BMR, meaning that a higher level of calories BMR was more predictive of PPD than in the prepregnancy period) among women with a significant correlation, as shown in A. Among those showing a significant relationship in SHAP dependence plots during the prepregnancy versus PPD (and prepregnancy versus PPD-equivalent) periods, the correlation pattern for SHAP values and calories BMR differed—most women exhibited a positive correlation in the PPD cohort, whereas there was no uniform pattern among women in the non-PPD cohort. Among women in the pregnancy versus PPD and postpartum versus PPD (and PPD-equivalent) periods, most demonstrated a negative relationship between SHAP values and calories BMR in both the PPD and non-PPD cohorts. HR: heart rate; NS: not significant; Q1: quartile 1.



We further investigated (1) health care–seeking behavior among women in the PPD and non-PPD cohorts; (2) the reliability of self-reported weight (as calories BMR are calculated based on age, sex, height, and weight); (3) the weight difference across each of the 4 periods (prepregnancy period, pregnancy, postpartum period, and PPD [or PPD equivalent]) between women in the PPD and non-PPD cohorts; (4) the relationship among weight, calories BMR, and PPD; and (5) weight retention between women in the PPD and non-PPD cohorts (see more details in Tables S5-S11 in Multimedia Appendix 2).

To showcase the effectiveness of our approach using individualized ML models for PPD detection, we constructed an ML model using conventional techniques. In this endeavor, we harnessed the PPD and PPD-equivalent periods of the PPD and non-PPD cohorts, respectively, enabling an assessment of our individualized approach compared to conventional methods using a binomial model for the classification of individuals with or without PPD. By evaluating model outcomes through metrics such as sensitivity, specificity, precision, recall, and F_1-score, we found that the average performance of the individualized model surpassed that of the cohort-based strategy (Table 5). Specifically, in the individualized approach, we observed sensitivity, specificity, precision, recall, and F_1-score values of 0.78, 0.95, 0.84, 0.78, and 0.81, respectively, in contrast to 0.54, 0.55, 0.49, 0.54, and 0.52, respectively, for the cohort-based approach (Table 5).

Table 5. Individualized machine learning (ML) models outperformed a cohort-based model for postpartum depression (PPD) recognition.

Method and digital biomarker	Value
Sensitivity	
Cohort-based model[a]	0.54
Individualized models, mean (SD)	0.78 (0.15)
Specificity	
Cohort-based model	0.55
Individualized models, mean (SD)	0.95 (0.09)
Precision	
Cohort-based model	0.49
Individualized models, mean (SD)	0.84 (0.12)
Recall	
Cohort-based model	0.54
Individualized models, mean (SD)	0.78 (0.15)
F_1-score	
Cohort-based model	0.52
Individualized models, mean (SD)	0.81 (0.13)

[a]Since the cohort-based ML approach is only to generate 1 model, there is no mean or SD.

Discussion

Principal Findings

In this study, our multifaceted analysis demonstrated that (1) digital biomarkers differed among the prepregnancy, pregnancy, and postpregnancy periods (up to 2 years before pregnancy, pregnancy, postpartum period, and PPD; Figure 3 and Tables S3 and S4 in Multimedia Appendix 2); (2) personalized N-of-1 ML models using digital biomarkers from consumer-grade wearables were able to classify PPD and other periods of pregnancy (Table 2 and Figure S1 in Multimedia Appendix 2); (3) a history of depression before or during pregnancy did not impact individualized ML model performance for PPD recognition (Figure S2 in Multimedia Appendix 2); (4) calories BMR, average HR, quartile 1 HR, lightly active minutes, and minimum HR were the most influential digital biomarkers in predicting the PPD period across all individualized models (Table 4); and (5) individualized ML models for PPD recognition outperformed the traditional cohort-based model approach (Table 5). The results presented in this paper provide a new opportunity for the potential to leverage passively collected digital biomarkers from consumer-grade wearables to facilitate early detection of PPD.

To the best of our knowledge, this is the first study proposing that individualized ML models using passively collected digital biomarkers from consumer-grade wearables can recognize PPD. Moreover, this study is also unique due to (1) integrating EHR and wearable data sources, (2) using longitudinal and continuous wearable data across multiple periods of pregnancy for ML methods, and (3) using individualized ML models for PPD recognition. PPD is most commonly diagnosed using the EPDS, which suffers from the following limitations: (1) postpartum women must attend follow-up visits assessed by care providers for PPD screening, where the rate of postpartum visits is highly variable; (2) using the EPDS only captures the mental health of a woman at a single point in time; and (3) the EPDS uses self-reported symptoms, which may not be representative of a patient's actual mental health status [61-63]. For these reasons,

our approach using passively monitored digital biomarkers from consumer wearable technology may serve as an effective tool for facilitating the detection of PPD in an individualized fashion, especially in nonclinical settings.

Because of the variation in digital biomarkers detected among women across different periods, our limited sample size, and the availability of continuous intraindividual data, our study was geared toward an individualized analytic approach (Table S2 in Multimedia Appendix 2). The observed variability across individuals is consistent with previous studies that have emphasized the heterogeneous nature of depression prompting individualized methodologies [16,54,64-66]. Moreover, ITSA and Tukey HSD results revealed that digital biomarkers were significantly altered among periods within each woman (Figure 3 and Tables S3 and S4 in Multimedia Appendix 2). Overall, there were numerous individual-level alterations, which can be explained by the considerable heterogeneity in depressive symptoms [54]. Collectively, these data suggest that digital biomarkers were significantly different across periods within each person, leading us to believe that individualized ML models would be able to accurately discriminate between PPD and other periods of pregnancy.

Our study also highlights the strength of using individualized N-of-1 ML models using digital biomarkers for identifying PPD. Our findings underscored the models' ability to differentiate between distinct pregnancy phases—namely, prepregnancy, pregnancy, postpartum, and PPD periods (Table 2 and Figure S1 in Multimedia Appendix 2). Notably, our approach's validity was confirmed by the noticeable decrease in model performance during the PPD-equivalent period for the non-PPD cohort compared to the PPD period for the PPD cohort (Table 3 and Figure 4). This demonstrated the distinct behavioral shifts that are observed during the onset of PPD, effectively captured by digital biomarkers [2]. Furthermore, our results did not indicate a significant variation in individualized model performance across the 4 pregnancy periods among women with a history of depression before or during pregnancy (Figure S2 in Multimedia Appendix 2). This accentuated the robust capability of individualized models to differentiate among periods based on the distinct behavioral characteristics and metabolic shifts linked to PPD after pregnancy as opposed to the behavior changes exhibited by each woman before or during pregnancy. This suggests that forthcoming technology centered on detecting PPD through digital biomarkers could have relevance for both individuals with and without a preexisting history of depression before or during pregnancy. Future studies should be conducted in a prospective framework to validate our individualized methodology. We also want to emphasize in future studies the importance of ML algorithms minimizing false positives for PPD detection to prevent unnecessary interventions [67].

Another crucial finding of our study was that the vital digital biomarkers for PPD classification were calories BMR, average HR, quartile 1 HR, minimum HR, and lightly active minutes, where calories BMR was the most predictive feature (Table 4). Therefore, we constructed SHAP dependence plots to enhance our understanding of the relationship between calories BMR and PPD. Plots for the prepregnancy versus PPD periods suggested that an elevated level of calories BMR is predictive of PPD, which is indicative of weight gain in these women (Figure 5A) [2,68]. In plots for the pregnancy versus PPD and postpartum versus PPD periods, the relationship between Shapley and actual values of calories BMR flipped, signifying that an increased number of calories BMR was inversely associated with PPD (Figure 5A). The negative relationship in the context of pregnancy versus PPD can likely be explained by the metabolic changes during pregnancy, resulting in an increased basal metabolic rate [69]. In the context of the postpartum versus PPD periods, we speculate that the negative relationship is because the median duration between the delivery date and PPD diagnosis is 83 days, when patients may not have fully returned to their prepregnancy physiological or behavioral patterns, which can take up to 6 months [70]. As a result, the relationship between Shapley values and actual values of calories BMR may reflect this transitional period and the ongoing postpartum changes experienced by women.

On the other hand, for the prepregnancy versus PPD-equivalent periods for women in the non-PPD cohort, SHAP dependence plots failed to unveil a uniform connection between calories BMR and the PPD-equivalent period, likely due to physiological distinctions, lifestyle changes during pregnancy, and random dissimilarities among women [57,71-73]. However, the comparison of SHAP dependence plots across the pregnancy versus PPD-equivalent and postpartum versus PPD-equivalent periods for women in the non-PPD cohort exhibited a consistent negative correlation, similar to what was observed in the PPD cohort (Figures 5A and 5B). This trend is likely a result of the common occurrence of an increased basal metabolic rate during pregnancy [69]. In the context of the postpartum versus PPD-equivalent periods, our use of an index date set at 83 days after delivery—the median number of days after delivery for PPD diagnosis in the PPD cohort—implies that women likely have not fully returned to their prepregnancy physiological baseline [70]. This aligns with the parallel observation seen during the pregnancy versus PPD-equivalent periods, reaffirming the persisting metabolic effect postpartum.

Calories BMR are calculated using a combination of age, gender, height, and weight. Hence, considering the relative stability of age, gender, and height throughout the 4 phases of pregnancy (prepregnancy period, pregnancy, postpartum period, and PPD), alterations in calories BMR are likely indicative of weight changes [74]. Calories BMR likely rely on self-reported height and weight measurements as most individuals do not have routine access to gold-standard measurements. Previous research has indicated that self-reported weight is generally reliable and accurate, where BMI was correctly determined for 91% of pregnancies using self-reported weight; however, accuracy varied between 70% in women who are underweight and 98% in women who are overweight [75]. In our study, we found no significant difference in self-reported and gold-standard weight measurements (Table S6 in Multimedia Appendix 2). Hence, it is important to acknowledge that an ideal approach would involve a combination of gold-standard weight measurements for accuracy and self-reported weight measurements for feasibility and accessibility to longitudinal data in analyses related to PPD.

Given the known positive relationship between calories BMR and weight, our analyses sought to further examine this relationship in the context of women with PPD. To do so, we leveraged the results from SHAP dependence plots, where we found that the 60% to 65% of women with a positive relationship between calories BMR and PPD relative to the prepregnancy period (ie, a higher value of calories BMR during PPD compared to the prepregnancy period) experienced weight gain (Table S9 in Multimedia Appendix 2). Similarly, the 85% to 90% of women who showed a negative relationship between calories BMR and PPD relative to the postpartum period (ie, a lower value of calories BMR during PPD compared to the postpartum period) showed weight loss (Table S10 in Multimedia Appendix 2). It is worth noting that previous research has suggested that some women experience weight gain during PPD, whereas others experience weight loss. For instance, it has been shown that women with PPD may experience weight gain attributed to emotional overeating [76,77]. Conversely, other studies propose that women grappling with PPD might experience weight loss resulting from skipped meals and overwhelming anxiety [78]. Consistent with the individualized framework presented in our study, we posit that it may be crucial to monitor changes in weight (or calories BMR). For example, although most women in our study exhibited an increase in calories BMR during the PPD period compared to the prepregnancy period, a percentage of women experienced a significant decrease in calories BMR during PPD relative to the prepregnancy period. When assessing the relationship between weight in the prepregnancy and PPD periods in these women, we found that these women did experience weight loss during the PPD period relative to the prepregnancy period. Therefore, we suggest that changes in body weight (or calories BMR), whether positive or negative, could potentially serve as more informative indicators of PPD. Considering the diverse manifestations of depression—some individuals may gain weight, whereas others may lose weight—we advocate for future studies to investigate these changes on an individual level.

Considering previous research indicating that women with PPD encounter challenges in reverting to prepregnancy weight compared to those without PPD, we aimed to assess this phenomenon in our cohorts. Strikingly, our findings did not detect a significant difference in the time taken for women to reach their prepregnancy weight between the PPD and non-PPD cohorts (Table S11 in Multimedia Appendix 2). Due to the association between calories BMR and weight, we aimed to validate this observation using calories BMR data. Interestingly, we observed a consistent pattern of no significant difference in the time taken to return to prepregnancy calories BMR between the PPD and non-PPD cohorts (Table S11 in Multimedia Appendix 2). It is posited that the lack of difference may be because women in the PPD cohort started at a higher average weight during the prepregnancy period compared to those without PPD (Table S8 in Multimedia Appendix 2). We also suspect that the discrepancy in the results between the number of days to return to prepregnancy weight detected via weight measurements and calories BMR was a product of the limited availability of weight data.

In the PPD cohort, the SHAP dependence plots for average HR, quartile 1 HR, and minimum HR during the pregnancy versus PPD periods also demonstrated a negative relationship, indicating that higher values of these digital biomarkers are inversely associated with PPD (Figure 5A). This relationship may be ascribed to the elevated HR commonly observed during pregnancy, which is a physiological response resulting from vascular remodeling for promoting augmented blood flow to the uterus [79-81]. In addition, there was a positive correlation between the increase in lightly active minutes and the occurrence of PPD in the pregnancy versus PPD periods, which may be explained by an inverse relationship between lightly active minutes and fairly active minutes or very active minutes (Figure S5 in Multimedia Appendix 2). Specifically, a higher number of lightly active minutes is concomitant with a decrease in the amount of time spent in fairly active and very active physical activities, aligning with the well-established understanding that reductions in overall physical activity can contribute to an increase in depressive symptoms [82]. In contrast, among women without PPD, a notable correlation was found solely in the prepregnancy versus PPD-equivalent periods concerning minimum HR, where an elevation in minimum HR was linked to the PPD-equivalent period (Figure 5B). Although a subset of women demonstrated a significant correlation in SHAP dependence plots concerning digital biomarkers of average HR, quartile 1 HR, or lightly active minutes across the prepregnancy versus PPD-equivalent, pregnancy versus PPD-equivalent, or postpartum versus PPD-equivalent periods, the overall proportion of women exhibiting such patterns was insufficient to draw definitive conclusions regarding the relationship between digital biomarkers during the prepregnancy, pregnancy, or postpartum periods and the PPD-equivalent period (Figure 5B). We postulate that the contrasting patterns of digital biomarkers between women in the PPD and non-PPD cohorts imply potential differences in these biomarkers for women who eventually experience PPD. Therefore, future studies of great interest may seek to develop ML models during the prepregnancy or pregnancy periods to predict a woman's risk of future PPD onset. These models would allow for the prediction of PPD risk in advance.

Comparison With Prior Work

In general, previous investigations have adhered to conventional ML strategies revolving around the development of a solitary model. In this paradigm, a model is constructed using an extensive patient data set encompassing individuals exhibiting either continuous outcomes (for regression-based models) or categorical outcomes. Subsequently, when a new patient is introduced, the model generates predictions for the patient based on their data and the pre-established model [83]. While this approach carries advantages, it is beset by two primary limitations: (1) reliance on an ample sample size and (2) neglect to accommodate the diverse and heterogeneous spectrum of depressive symptoms [16]. Hence, a captivating domain of exploration has honed in on crafting intraindividual ML models. This advancement tackles the constraints of conventional approaches in 2 ways: first, it sidesteps the need for an extensive sample size given that the model is tailored to a single patient's data, and second, it conscientiously acknowledges the

heterogeneous spectrum of depressive symptoms through a focused evaluation of the unique behaviors exhibited by that specific patient.

The use of individualized models may serve as a superior preference compared to those formulated using cohort-based methodologies. For instance, a cross-sectional study using traditional ML models from Fitbit data from healthy adults to predict depression severity only showed a moderate area under the receiver operating characteristic curve range of 0.51 to 0.66. Moreover, while the results demonstrated commendable specificity (0.98-1), sensitivity exhibited marked inadequacy (0.03-0.13) [19]. Another study aimed to investigate the potential of ML models using digital biomarkers in distinguishing between patients with unipolar and bipolar depression against healthy controls. However, the most successful model exhibited an accuracy rate of 0.73 (73%) and a κ value of 0.44, which does not indicate a notably high-performing model [84,85]. Additional investigations have also been conducted within a cohort-based framework; nevertheless, these studies grapple with a noteworthy drawback—they incorporated patient mood as a predictive feature in their models. Considering that these studies aimed to predict the severity of depression, it is unsurprising that these models exhibited heightened performance levels [13,86].

To effectively underscore the viability of personalized ML models over cohort-based methods, our study directly juxtaposed the performance of both approaches (individualized vs cohort-based ML models) side by side. Notably, our findings vividly showcased the superior performance achieved through the personalized methodology for the PPD class in comparison to conventional techniques with a cohort-based model leveraging digital biomarkers from Fitbit for PPD detection (Table 5). This outcome accentuated that individualized models present an encouraging avenue for crafting ML models aimed at identifying mood disorders.

Limitations

Although this study provides a strong foundation for using digital biomarkers to classify PPD, it is not without limitations. First, this study faced constraints due to the restricted number of patients available, which hindered the implementation of conventional ML techniques. However, due to the limited sample size, we opted for an individualized approach, which not only addressed the small sample size but also provided a means to accommodate the inherent variability among individuals [54]. Second, the process of phenotyping patients with PPD relied on a PPD diagnosis or medication use, which could potentially lack specificity in diagnostic codes and miss undiagnosed cases. Third, our approach assumed a standard pregnancy length of 9 months, which may not always align with individual variations. Fourth, there are several layers of confounding that occur during the different phases of pregnancy that may indirectly influence digital biomarkers and ML models, especially as it relates to PPD classification, such as (1) significant hormonal changes that impact physical and mental states; (2) metabolic changes that occur as a result of pregnancy; (3) increased levels of stress during pregnancy and the postpartum period; (4) modifications to one's lifestyle, such as food consumption during pregnancy and the postpartum period; and (5) alterations in physical activity during the postpartum period as a result of birthing complications [87-91]. This could create a risk of model overfitting on general postpartum features compared to those that are specific to the PPD period. In general, to avoid overfitting of the training data, our method used 3 repetitions of 10-fold cross-validation, a strategy known to reduce overfitting compared to a conventional train-test split [92]. Fifth, this study excluded patients with chronic conditions to mitigate the potential influence of those conditions on digital biomarkers. Sixth, sleep data were absent in the AoURP data set at the time of this analysis using Registered Tier v6 although they might also hold predictive value for PPD. Seventh, there is a possibility of false negatives in the non-PPD cohort given the 1 in 7 prevalence of PPD. We attempted to identify women in the non-PPD cohort with undiagnosed PPD by re-evaluating individual model performance for the PPD-equivalent class, where we counted approximately 10 women in the non-PPD cohort who exhibited results with elevated model performance (sensitivity≥0.78 [the average performance of the PPD class for women in the PPD cohort]). However, estimating the prevalence of undiagnosed PPD is difficult in itself; estimating undiagnosed PPD in EHR data is additionally challenging. We suggest that these exploratory findings require future model development, including true negatives (ie, women who are definitely not experiencing PPD) and false negatives for validation. Eighth, there may be delays in PPD diagnosis due to health care–seeking behavior. However, we observed no significant difference in health care use during the postpartum period between the PPD and non-PPD cohorts along with similar adherence to American College of Obstetricians and Gynecologists postpartum visit guidelines in the PPD and non-PPD cohorts. We posit that delays in diagnosis due to health care–seeking behavior might be minimal (Table S5 in Multimedia Appendix 2). These data suggest that women in the PPD cohort exhibited a fairly normal rate of health care use, which should minimize delay in diagnosing PPD. Furthermore, it may be beneficial for subsequent analyses to account for features such as seasonal variation that may also indirectly influence behavior (which is not possible using the Registered Tier of AoURP) in addition to testing other ML algorithms, such as Extreme Gradient Boosting [93,94].

Conclusions

Overall, the findings of this study suggest that it is feasible to characterize PPD in addition to other periods of pregnancy using passively collected digital biomarkers from consumer-grade wearables. The development of individualized models allows for a personalized approach to capture behavioral differences in the form of digital biomarkers. This research lays a robust foundation for forthcoming applications aimed at enhancing the early detection of PPD, a condition that is often underdiagnosed and undertreated. Moreover, on a broader scale, it indicates the exciting potential for intraindividual ML models to be extended to various health conditions.

Acknowledgments

First, the *All of Us* Research Program would not be possible without the partnership of its participants. The *All of Us* Research Program is supported by the National Institutes of Health Office of the Director: Regional Medical Centers: 1 other transactions (OT) OT2 OD026549, 1 OT2 OD026554, 1 OT2 OD026557, 1 OT2 OD026556, 1 OT2 OD026550, 1 OT2 OD 026552, 1 OT2 OD026553, 1 OT2 OD026548, 1 OT2 OD026551, 1 OT2 OD026555, and IAA#: AOD 16037; Federally Qualified Health Centers: HHSN 263201600085U; Data and Research Center: 5 U2C OD023196; Biobank: 1 U24 OD023121; The Participant Center: U24 OD023176; Participant Technology Systems Center: 1 U24 OD023163; Communications and Engagement: 3 OT2 OD023205 and 3 OT2 OD023206; and Community Partners: 1 OT2 OD025277, 3 OT2 OD025315, 1 OT2 OD025337, and 1 OT2 OD025276). Individual authors were supported by the following funding sources: National Institute of Mental Health R01131542 (RCP), National Human Genome Research Institute RM1HG010860 (EH and MAH), and National Institutes of Health Office of the Director R24OD011883 (EH, STO, and MAH). ChatGPT (GPT-3.5), developed by OpenAI [95], was used to edit some portions of the manuscript by offering synonym suggestions, language enhancements, grammar refinements, and style improvements. It is important to note that all recommendations made by ChatGPT were meticulously reviewed by the author and were not used for generating ideas or content.

Data Availability

The data sets generated during and analyzed during this study are not publicly available due to patient privacy restrictions and guidelines established by the All of Us Research Program that require individuals with access to the platform complete the requisite training, but they are available from the corresponding author on reasonable request.

Authors' Contributions

EH and RCP conceived the study. EH and MAH designed the study. EH analyzed the data. EH, ZBD, HM, STO, and MAH interpreted the results. EH wrote the manuscript draft. MAH provided supervision and acquired funding. All authors reviewed and edited the final manuscript for publication.

Conflicts of Interest

MAH is a founder of Alamya Health. All other authors declare no other conflicts of interest.

Multimedia Appendix 1
Observational Medical Outcomes Partnership concept IDs used in computational phenotyping of postpartum depression (PPD) and non-PPD cohorts.
[XLSX File (Microsoft Excel File), 63 KB - mhealth_v12i1e54622_app1.xlsx]

Multimedia Appendix 2
Supplementary results, tables, and figures.
[DOCX File , 1016 KB - mhealth_v12i1e54622_app2.docx]

References

1. Gaynes BN, Gavin N, Meltzer-Brody S, Lohr KN, Swinson T, Gartlehner G, et al. Perinatal depression: prevalence, screening accuracy, and screening outcomes. Evid Rep Technol Assess (Summ) 2005 Feb(119):1-8. [doi: 10.1037/e439372005-001] [Medline: 15760246]
2. Pearlstein T, Howard M, Salisbury A, Zlotnick C. Postpartum depression. Am J Obstet Gynecol 2009 Apr;200(4):357-364 [FREE Full text] [doi: 10.1016/j.ajog.2008.11.033] [Medline: 19318144]
3. Luca DL, Margiotta C, Staatz C, Garlow E, Christensen A, Zivin K. Financial toll of untreated perinatal mood and anxiety disorders among 2017 births in the United States. Am J Public Health 2020 Jun;110(6):888-896. [doi: 10.2105/AJPH.2020.305619] [Medline: 32298167]
4. Kavanaugh M, Halterman JS, Montes G, Epstein M, Hightower AD, Weitzman M. Maternal depressive symptoms are adversely associated with prevention practices and parenting behaviors for preschool children. Ambul Pediatr 2006;6(1):32-37. [doi: 10.1016/j.ambp.2005.09.002] [Medline: 16443181]
5. Duan Z, Wang Y, Jiang P, Wilson A, Guo Y, Lv Y, et al. Postpartum depression in mothers and fathers: a structural equation model. BMC Pregnancy Childbirth 2020 Sep 15;20(1):537 [FREE Full text] [doi: 10.1186/s12884-020-03228-9] [Medline: 32933502]
6. Hutchens BF, Kearney J. Risk factors for postpartum depression: an umbrella review. J Midwifery Womens Health 2020 Jan 22;65(1):96-108. [doi: 10.1111/jmwh.13067] [Medline: 31970924]
7. Sit DK, Wisner KL. Identification of postpartum depression. Clin Obstet Gynecol 2009 Sep;52(3):456-468 [FREE Full text] [doi: 10.1097/GRF.0b013e3181b5a57c] [Medline: 19661761]

8. Evins GG, Theofrastous JP. Postpartum depression: a review of postpartum screening. Prim Care Update Ob Gyns 1997 Nov;4(6):241-246. [doi: 10.1016/s1068-607x(97)00105-4]
9. Fitelson E, Kim S, Baker AS, Leight K. Treatment of postpartum depression: clinical, psychological and pharmacological options. Int J Womens Health 2010 Dec 30;3:1-14 [FREE Full text] [doi: 10.2147/IJWH.S6938] [Medline: 21339932]
10. Mobile fact sheet. Pew Research Center. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2023-06-07]
11. Avram R, Tison G, Kuhar P, Marcus G, Pletcher M, Olgin JE, et al. Predicting diabetes from photoplethysmography using deep learning. J Am Coll Cardiol 2019 Mar;73(9):16. [doi: 10.1016/s0735-1097(19)33778-7]
12. Mishra T, Wang M, Metwally AA, Bogu GK, Brooks AW, Bahmani A, et al. Pre-symptomatic detection of COVID-19 from smartwatch data. Nat Biomed Eng 2020 Dec 18;4(12):1208-1220 [FREE Full text] [doi: 10.1038/s41551-020-00640-6] [Medline: 33208926]
13. Narziev N, Goh H, Toshnazarov K, Lee SA, Chung KM, Noh Y. STDD: short-term depression detection with passive sensing. Sensors (Basel) 2020 Mar 04;20(5):1396 [FREE Full text] [doi: 10.3390/s20051396] [Medline: 32143358]
14. Steinhubl SR, Mehta RR, Ebner GS, Ballesteros MM, Waalen J, Steinberg G, et al. Rationale and design of a home-based trial using wearable sensors to detect asymptomatic atrial fibrillation in a targeted population: the mHealth Screening To Prevent Strokes (mSToPS) trial. Am Heart J 2016 May;175:77-85 [FREE Full text] [doi: 10.1016/j.ahj.2016.02.011] [Medline: 27179726]
15. Pedrelli P, Fedor S, Ghandeharioun A, Howe E, Ionescu DF, Bhathena D, et al. Monitoring changes in depression severity using wearable and mobile sensors. Front Psychiatry 2020;11:584711 [FREE Full text] [doi: 10.3389/fpsyt.2020.584711] [Medline: 33391050]
16. Shah RV, Grennan G, Zafar-Khan M, Alim F, Dey S, Ramanathan D, et al. Personalized machine learning of depressed mood using wearables. Transl Psychiatry 2021 Jun 09;11(1):338 [FREE Full text] [doi: 10.1038/s41398-021-01445-0] [Medline: 34103481]
17. Wang R, Wang W, daSilva A, Huckins JF, Kelley WM, Heatherton TF, et al. Tracking depression dynamics in college students using mobile phone and wearable sensing. Proc ACM Interact Mob Wearable Ubiquitous Technol 2018 Mar 26;2(1):1-26. [doi: 10.1145/3191775]
18. Moshe I, Terhorst Y, Opoku Asare K, Sander LB, Ferreira D, Baumeister H, et al. Predicting symptoms of depression and anxiety using smartphone and wearable data. Front Psychiatry 2021;12:625247 [FREE Full text] [doi: 10.3389/fpsyt.2021.625247] [Medline: 33584388]
19. Rykov Y, Thach TQ, Bojic I, Christopoulos G, Car J. Digital biomarkers for depression screening with wearable devices: cross-sectional study with machine learning modeling. JMIR Mhealth Uhealth 2021 Oct 25;9(10):e24872 [FREE Full text] [doi: 10.2196/24872] [Medline: 34694233]
20. Sarhaddi F, Azimi I, Niela-Vilen H, Axelin A, Liljeberg P, Rahmani AM. Maternal social loneliness detection using passive sensing through continuous monitoring in everyday settings: longitudinal study. JMIR Form Res 2023 Aug 09;7:e47950 [FREE Full text] [doi: 10.2196/47950] [Medline: 37556183]
21. Faherty LJ, Hantsoo L, Appleby D, Sammel MD, Bennett IM, Wiebe DJ. Movement patterns in women at risk for perinatal depression: use of a mood-monitoring mobile application in pregnancy. J Am Med Inform Assoc 2017 Jul 01;24(4):746-753 [FREE Full text] [doi: 10.1093/jamia/ocx005] [Medline: 28339686]
22. Shea AK, Kamath MV, Fleming A, Streiner DL, Redmond K, Steiner M. The effect of depression on heart rate variability during pregnancy. A naturalistic study. Clin Auton Res 2008 Aug 30;18(4):203-212. [doi: 10.1007/s10286-008-0480-1] [Medline: 18592128]
23. Lee K, Kwon H, Lee B, Lee G, Lee JH, Park YR, et al. Effect of self-monitoring on long-term patient engagement with mobile health applications. PLoS One 2018 Jul 26;13(7):e0201166 [FREE Full text] [doi: 10.1371/journal.pone.0201166] [Medline: 30048546]
24. All of Us Research Program Investigators, Denny JC, Rutter JL, Goldstein DB, Philippakis A, Smoller JW, et al. The "All of Us" research program. N Engl J Med 2019 Aug 15;381(7):668-676 [FREE Full text] [doi: 10.1056/NEJMsr1809937] [Medline: 31412182]
25. Bender CG, Hoffstot JC, Combs BT, Hooshangi S, Cappos J. Measuring the fitness of fitness trackers. In: Proceedings of the 2017 IEEE Sensors Applications Symposium. 2017 Presented at: SAS '17; March 13-15, 2017; Glassboro, NJ p. 1-6 URL: https://ieeexplore.ieee.org/document/7894077 [doi: 10.1109/sas.2017.7894077]
26. Jooma S. Data and statistics dissemination policy. All of Us Research Program. URL: https://support.researchallofus.org/hc/en-us/article_attachments/22307203778836 [accessed 2024-04-05]
27. Hripcsak G, Duke JD, Shah NH, Reich CG, Huser V, Schuemie MJ, et al. Observational health data sciences and informatics (OHDSI): opportunities for observational researchers. Stud Health Technol Inform 2015;216:574-578 [FREE Full text] [Medline: 26262116]
28. Jones SE, Bradwell KR, Chan LE, McMurry JA, Olson-Chen C, Tarleton J, N3C Consortium. Who is pregnant? Defining real-world data-based pregnancy episodes in the National COVID Cohort Collaborative (N3C). JAMIA Open 2023 Oct;6(3):ooad067 [FREE Full text] [doi: 10.1093/jamiaopen/ooad067] [Medline: 37600074]

29. Jukic AM, Baird DD, Weinberg CR, McConnaughey DR, Wilcox AJ. Length of human pregnancy and contributors to its natural variation. Hum Reprod 2013 Oct;28(10):2848-2855 [FREE Full text] [doi: 10.1093/humrep/det297] [Medline: 23922246]
30. Goodman JH. Postpartum depression beyond the early postpartum period. J Obstet Gynecol Neonatal Nurs 2004;33(4):410-420. [doi: 10.1177/0884217504266915] [Medline: 15346666]
31. Horowitz JA, Goodman J. A longitudinal study of maternal postpartum depression symptoms. Res Theory Nurs Pract 2004;18(2-3):149-163. [doi: 10.1891/rtnp.18.2.149.61285] [Medline: 15553344]
32. Amit G, Girshovitz I, Marcus K, Zhang Y, Pathak J, Bar V, et al. Estimation of postpartum depression risk from electronic health records using machine learning. BMC Pregnancy Childbirth 2021 Sep 17;21(1):630 [FREE Full text] [doi: 10.1186/s12884-021-04087-8] [Medline: 34535116]
33. ACOG redesigns postpartum care. American College of Obstetricians and Gynecologists. URL: https://www.acog.org/news/news-releases/2018/04/acog-redesigns-postpartum-care [accessed 2024-02-28]
34. Liberto TL. Screening for depression and help-seeking in postpartum women during well-baby pediatric visits: an integrated review. J Pediatr Health Care 2012 Mar;26(2):109-117. [doi: 10.1016/j.pedhc.2010.06.012] [Medline: 22360930]
35. Herring SJ, Rich-Edwards JW, Oken E, Rifas-Shiman SL, Kleinman KP, Gillman MW. Association of postpartum depression with weight retention 1 year after childbirth. Obesity (Silver Spring) 2008 Jun;16(6):1296-1301 [FREE Full text] [doi: 10.1038/oby.2008.71] [Medline: 18369338]
36. Bent B, Goldstein BA, Kibbe WA, Dunn JP. Investigating sources of inaccuracy in wearable optical heart rate sensors. NPJ Digit Med 2020;3:18 [FREE Full text] [doi: 10.1038/s41746-020-0226-6] [Medline: 32047863]
37. Fuller D, Colwell E, Low J, Orychock K, Tobin MA, Simango B, et al. Reliability and validity of commercially available wearable devices for measuring steps, energy expenditure, and heart rate: systematic review. JMIR Mhealth Uhealth 2020 Sep 08;8(9):e18694 [FREE Full text] [doi: 10.2196/18694] [Medline: 32897239]
38. Master H, Annis J, Huang S, Beckman JA, Ratsimbazafy F, Marginean K, et al. Association of step counts over time with the risk of chronic disease in the All of Us research program. Nat Med 2022 Nov 10;28(11):2301-2308 [FREE Full text] [doi: 10.1038/s41591-022-02012-w] [Medline: 36216933]
39. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Soft 2015;67(1):1-48. [doi: 10.18637/jss.v067.i01]
40. Kuznetsova A, Brockhoff PB, Christensen RH. Package: tests in linear mixed effects models. J Stat Soft 2017;82(13):1-26. [doi: 10.18637/jss.v082.i13]
41. Lüdecke D, Ben-Shachar M, Patil I, Waggoner P, Makowski D. performance: an R package for assessment, comparison and testing of statistical models. J Open Source Softw 2021 Apr;6(60):3139. [doi: 10.21105/joss.03139]
42. English P. The its.analysis R package – modelling short time series data. Social Science Research Network. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3398189 [accessed 2024-04-05]
43. Abdi H, Williams LJ. Tukey's honestly significant difference (HSD). In: Lane DM, editor. Encyclopedia of Research Design. Thousand Oaks, CA: SAGE Publications; 2010:1-5.
44. Mughal S, Azhar Y, Siddiqui W. Postpartum depression. StatPearls. URL: https://www.ncbi.nlm.nih.gov/books/NBK519070/?report=printable [accessed 2024-04-05]
45. Machado-Vieira R, Baumann J, Wheeler-Castillo C, Latov D, Henter ID, Salvadore G, et al. The timing of antidepressant effects: a comparison of diverse pharmacological and somatic treatments. Pharmaceuticals (Basel) 2010 Jan 06;3(1):19-41 [FREE Full text] [doi: 10.3390/ph3010019] [Medline: 27713241]
46. Abd-Alrazaq A, AlSaad R, Aziz S, Ahmed A, Denecke K, Househ M, et al. Wearable artificial intelligence for anxiety and depression: scoping review. J Med Internet Res 2023 Jan 19;25:e42672 [FREE Full text] [doi: 10.2196/42672] [Medline: 36656625]
47. Bradley AP. The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recognit 1997 Jul;30(7):1145-1159. [doi: 10.1016/S0031-3203(96)00142-2]
48. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas 2016 Jul 02;20(1):37-46. [doi: 10.1177/001316446002000104]
49. Hand DJ, Till RJ. A simple generalisation of the area under the ROC curve for multiple class classification problems. Mach Learn 2001;45(2):171-186. [doi: 10.1023/A:1010920819831]
50. Kuhn M. Building predictive models in R using the caret package. J Stat Soft 2008;28(5):1-26. [doi: 10.18637/jss.v028.i05]
51. Molnar C, Casalicchio G, Bischl B. iml: an R package for interpretable machine learning. J Open Source Softw 2018 Jun;3(26):786. [doi: 10.21105/joss.00786]
52. Lundberg SM, Erion GG, Lee SI. Consistent individualized feature attribution for tree ensembles. Preprint posted online on February 12, 2018 [FREE Full text] [doi: 10.5860/choice.189890]
53. Chaibub Neto E, Pratap A, Perumal TM, Tummalacherla M, Snyder P, Bot BM, et al. Detecting the impact of subject characteristics on machine learning-based diagnostic applications. NPJ Digit Med 2019 Oct 11;2(1):99 [FREE Full text] [doi: 10.1038/s41746-019-0178-x] [Medline: 31633058]
54. Belmaker RH, Agam G. Major depressive disorder. N Engl J Med 2008 Jan 03;358(1):55-68. [doi: 10.1056/NEJMra073096] [Medline: 18172175]

55. Doryab A, Min JK, Wiese J, Zimmerman J, Hong JI. Detection of behavior change in people with depression. In: Proceedings of the 28th Conference on Artificial Intelligence. 2014 Presented at: AAAI '14; July 27-31, 2014; Québec City, QC URL: https://cdn.aaai.org/ocs/ws/ws1266/8850-38024-1-PB.pdf [doi: 10.1184/R1/6469988.v1]
56. Heidemann BH, McClure JH. Changes in maternal physiology during pregnancy. BJA CEPD Rev 2003 Jun;3(3):65-68 [FREE Full text] [doi: 10.1093/bjacepd/mkg065]
57. Soma-Pillay P, Nelson-Piercy C, Tolppanen H, Mebazaa A. Physiological changes in pregnancy. Cardiovasc J Afr 2016;27(2):89-94 [FREE Full text] [doi: 10.5830/CVJA-2016-021] [Medline: 27213856]
58. Ling CX, Huang J, Zhang H. AUC: a better measure than accuracy in comparing learning algorithms. In: Proceedings of the 16th Conference of the Canadian Society for Computational Studies of Intelligence. 2003 Presented at: AAI '03; June 11-13, 2003; Halifax, NS p. 329-341 URL: https://link.springer.com/chapter/10.1007/3-540-44886-1_25 [doi: 10.1007/3-540-44886-1_25]
59. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. Preprint posted online on May 22, 2017. [doi: 10.5260/chara.21.2.8]
60. Fisher A, Rudin C, Dominici F. All models are wrong, but many are useful: learning a variable's importance by studying an entire class of prediction models simultaneously. Preprint posted online on January 4, 2018 [FREE Full text] [doi: 10.48550/arXiv.1801.01489]
61. Attanasio LB, Ranchoff BL, Cooper MI, Geissler KH. Postpartum visit attendance in the United States: a systematic review. Womens Health Issues 2022;32(4):369-375 [FREE Full text] [doi: 10.1016/j.whi.2022.02.002] [Medline: 35304034]
62. Guedeney N, Fermanian J, Guelfi JD, Kumar RC. The Edinburgh Postnatal Depression Scale (EPDS) and the detection of major depressive disorders in early postpartum: some concerns about false negatives. J Affect Disord 2000 Dec;61(1-2):107-112. [doi: 10.1016/s0165-0327(99)00186-x] [Medline: 11099748]
63. Levis B, Negeri Z, Sun Y, Benedetti A, Thombs BD, DEPRESsion Screening Data (DEPRESSD) EPDS Group. Accuracy of the Edinburgh Postnatal Depression Scale (EPDS) for screening to detect major depression among pregnant and postpartum women: systematic review and meta-analysis of individual participant data. BMJ 2020 Nov 11;371:m4022 [FREE Full text] [doi: 10.1136/bmj.m4022] [Medline: 33177069]
64. Drysdale AT, Grosenick L, Downar J, Dunlop K, Mansouri F, Meng Y, et al. Resting-state connectivity biomarkers define neurophysiological subtypes of depression. Nat Med 2017 Jan 5;23(1):28-38 [FREE Full text] [doi: 10.1038/nm.4246] [Medline: 27918562]
65. Halaris A. A primary care focus on the diagnosis and treatment of major depressive disorder in adults. J Psychiatr Pract 2011 Sep;17(5):340-350. [doi: 10.1097/01.pra.0000405364.90043.bd] [Medline: 21926529]
66. Institute of Medicine (US) Forum on Drug Discovery, Development, and Translation. Transforming Clinical Research in the United States: Challenges and Opportunities: Workshop Summary. New York, NY: National Academies Press; 2010.
67. Dennis CL, Ross LE. Depressive symptomatology in the immediate postnatal period: identifying maternal characteristics related to true- and false-positive screening scores. Can J Psychiatry 2006 Apr;51(5):265-273. [doi: 10.1177/070674370605100501] [Medline: 16986815]
68. Li S, Yang Z, Yao M, Shen Y, Zhu H, Jiang Y, et al. Exploration for biomarkers of postpartum depression based on metabolomics: a systematic review. J Affect Disord 2022 Nov 15;317:298-306. [doi: 10.1016/j.jad.2022.08.043] [Medline: 36031003]
69. Poppitt SD, Prentice AM, Goldberg GR, Whitehead RG. Energy-sparing strategies to protect human fetal growth. Am J Obstet Gynecol 1994 Jul;171(1):118-125. [doi: 10.1016/s0002-9378(94)70087-7] [Medline: 8030686]
70. Tinius RA, Yoho K, Blankenship MM, Maples JM. Postpartum metabolism: how does it change from pregnancy and what are the potential implications? Int J Womens Health 2021;13:591-599 [FREE Full text] [doi: 10.2147/IJWH.S314469] [Medline: 34168507]
71. Crozier SR, Robinson SM, Borland SE, Godfrey KM, Cooper C, Inskip HM, et al. Do women change their health behaviours in pregnancy? Findings from the Southampton women's survey. Paediatr Perinat Epidemiol 2009 Sep;23(5):446-453 [FREE Full text] [doi: 10.1111/j.1365-3016.2009.01036.x] [Medline: 19689495]
72. Lindqvist M, Lindkvist M, Eurenius E, Persson M, Mogren I. Change of lifestyle habits - motivation and ability reported by pregnant women in northern Sweden. Sex Reprod Healthc 2017 Oct;13:83-90. [doi: 10.1016/j.srhc.2017.07.001] [Medline: 28844363]
73. Zinsser LA, Stoll K, Wieber F, Pehlke-Milde J, Gross MM. Changing behaviour in pregnant women: a scoping review. Midwifery 2020 Jun;85:102680 [FREE Full text] [doi: 10.1016/j.midw.2020.102680] [Medline: 32151875]
74. How does my Fitbit device calculate my daily activity? Fitbit Help Center. URL: https://support.google.com/fitbit/answer/14237111?hl=en#zippy=%2Chow-does-my-fitbit-device-calculate-calories-burned [accessed 2024-02-28]
75. Sharma AJ, Bulkley JE, Stoneburner AB, Dandamudi P, Leo M, Callaghan WM, et al. Bias in self-reported prepregnancy weight across maternal and clinical characteristics. Matern Child Health J 2021 Aug;25(8):1242-1253 [FREE Full text] [doi: 10.1007/s10995-021-03149-9] [Medline: 33929655]
76. Fadzil F, Shamsuddin K, Wan Puteh SE, Mohd Tamil A, Ahmad S, Abdul Hayi NS, et al. Predictors of postpartum weight retention among urban Malaysian mothers: a prospective cohort study. Obes Res Clin Pract 2018 Nov;12(6):493-499. [doi: 10.1016/j.orcp.2018.06.003] [Medline: 29960869]

77. Pettersson CB, Zandian M, Clinton D. Eating disorder symptoms pre- and postpartum. Arch Womens Ment Health 2016 Aug 10;19(4):675-680. [doi: 10.1007/s00737-016-0619-3] [Medline: 26961005]
78. Fowles ER, Stang J, Bryant M, Kim S. Stress, depression, social support, and eating habits reduce diet quality in the first trimester in low-income women: a pilot study. J Acad Nutr Diet 2012 Oct;112(10):1619-1625 [FREE Full text] [doi: 10.1016/j.jand.2012.07.002] [Medline: 23017572]
79. Easterling TR, Benedetti TJ, Schmucker BC, Millard SP. Maternal hemodynamics in normal and preeclamptic pregnancies: a longitudinal study. Obstet Gynecol 1990 Dec;76(6):1061-1069. [Medline: 2234714]
80. Odendaal H, Groenewald C, Myers MM, Fifer WP. Maternal heart rate patterns under resting conditions in late pregnancy. Trends Res 2018 Sep;1(3) [FREE Full text] [doi: 10.15761/TR.1000116] [Medline: 31008442]
81. Thornburg KL, Jacobson SL, Giraud GD, Morton MJ. Hemodynamic changes in pregnancy. Semin Perinatol 2000 Feb;24(1):11-14. [doi: 10.1016/s0146-0005(00)80047-6] [Medline: 10709851]
82. Hu S, Tucker L, Wu C, Yang L. Beneficial effects of exercise on depression and anxiety during the COVID-19 pandemic: a narrative review. Front Psychiatry 2020;11:587557 [FREE Full text] [doi: 10.3389/fpsyt.2020.587557] [Medline: 33329133]
83. Zhou ZH. Machine Learning. Singapore, Singapore: Springer; 2021.
84. Garcia-Ceja E, Riegler M, Jakobsen P, Torresen J, Nordgreen T, Oedegaard KJ, et al. Motor activity based classification of depression in unipolar and bipolar patients. In: Proceedings of the IEEE 31st International Symposium on Computer-Based Medical Systems. 2018 Presented at: CBMS '18; June 18-21, 2018; Karlstad, Sweden p. 316-321 URL: https://ieeexplore.ieee.org/document/8417257 [doi: 10.1109/cbms.2018.00062]
85. Jacobson NC, Weingarden H, Wilhelm S. Digital biomarkers of mood disorders and symptom change. NPJ Digit Med 2019 Feb 1;2(1):3 [FREE Full text] [doi: 10.1038/s41746-019-0078-0] [Medline: 31304353]
86. Opoku Asare K, Moshe I, Terhorst Y, Vega J, Hosio S, Baumeister H, et al. Mood ratings and digital biomarkers from smartphone and wearable data differentiates and predicts depression status: a longitudinal data analysis. Pervasive and Mobile Computing 2022 Jul;83:101621. [doi: 10.1016/j.pmcj.2022.101621]
87. Apostolopoulos M, Hnatiuk JA, Maple JL, Olander EK, Brennan L, van der Pligt P, et al. Influences on physical activity and screen time amongst postpartum women with heightened depressive symptoms: a qualitative study. BMC Pregnancy Childbirth 2021 May 15;21(1):376 [FREE Full text] [doi: 10.1186/s12884-021-03847-w] [Medline: 33992072]
88. Glinoer D. What happens to the normal thyroid during pregnancy? Thyroid 1999 Jul;9(7):631-635. [doi: 10.1089/thy.1999.9.631] [Medline: 10447005]
89. Jardí C, Aparicio E, Bedmar C, Aranda N, Abajo S, March G, The Eclipses Study Group. Food consumption during pregnancy and post-partum. ECLIPSES study. Nutrients 2019 Oct 14;11(10):2447 [FREE Full text] [doi: 10.3390/nu11102447] [Medline: 31615024]
90. La Marca-Ghaemmaghami P, Ehlert U. Stress During Pregnancy. Eur Psychol 2015 Jun 30;20(2):102-119. [doi: 10.1027/1016-9040/a000195]
91. Lain KY, Catalano PM. Metabolic changes in pregnancy. Clin Obstet Gynecol 2007 Dec;50(4):938. [doi: 10.1097/grf.0b013e31815a5494]
92. Hawkins DM. The problem of overfitting. J Chem Inf Comput Sci 2004 Jan;44(1):1-12. [doi: 10.1021/ci0342472] [Medline: 14741005]
93. Spoont MR, Depue RA, Krauss SS. Dimensional measurement of seasonal variation in mood and behavior. Psychiatry Res 1991 Dec;39(3):269-284. [doi: 10.1016/0165-1781(91)90093-5] [Medline: 1798825]
94. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016 Presented at: KDD '16; August 13-17, 2016; San Francisco, CA p. 785-794 URL: https://dl.acm.org/doi/10.1145/2939672.2939785 [doi: 10.1145/2939672.2939785]
95. Home page. OpenAI. URL: https://openai.com/ [accessed 2024-04-05]

Abbreviations

AoURP: *All of Us* Research Program
Calories BMR: calories burned during the basal metabolic rate
EHR: electronic health record
EPDS: Edinburgh Postnatal Depression Scale
GLM: generalized linear model
HR: heart rate
HSD: honest significant difference
ITSA: interrupted time-series analysis
KNN: k-nearest neighbor
mAUC: multiclass area under the receiver operating characteristic curve
ML: machine learning
PPD: postpartum depression
RF: random forest

SHAP: Shapley additive explanations
SVM: support vector machine

Edited by L Buis; submitted 17.11.23; peer-reviewed by L Huang, S Tedesco, C Wang; comments to author 10.01.24; revised version received 06.03.24; accepted 27.03.24; published 02.05.24.

Please cite as:
Hurwitz E, Butzin-Dozier Z, Master H, O'Neil ST, Walden A, Holko M, Patel RC, Haendel MA
Harnessing Consumer Wearable Digital Biomarkers for Individualized Recognition of Postpartum Depression Using the All of Us Research Program Data Set: Cross-Sectional Study
JMIR Mhealth Uhealth 2024;12:e54622
URL: https://mhealth.jmir.org/2024/1/e54622
doi:10.2196/54622
PMID:38696234

Original Paper

Wearable Data From Subjects Playing Super Mario, Taking University Exams, or Performing Physical Exercise Help Detect Acute Mood Disorder Episodes via Self-Supervised Learning: Prospective, Exploratory, Observational Study

Filippo Corponi[1], MSc, MD; Bryan M Li[1,2], MSc; Gerard Anmella[3,4,5,6], MD, PhD; Clàudia Valenzuela-Pascual[3,4,5,6], MSc; Ariadna Mas[3,4,5,6], MSc; Isabella Pacchiarotti[3,4,5,6], MD, PhD; Marc Valentí[3,4,5,6], MD, PhD; Iria Grande[3,4,5,6], MD, PhD; Antoni Benabarre[3,4,5,6], MD, PhD; Marina Garriga[3,4,5,6], MD, PhD; Eduard Vieta[3,4,5,6], MD, PhD; Allan H Young[7], MD, PhD; Stephen M Lawrie[8], MD; Heather C Whalley[8,9], PhD; Diego Hidalgo-Mazzei[3,4,5,6,7*], MD, PhD; Antonio Vergari[1*], PhD

[1]School of Informatics, University of Edinburgh, Edinburgh, United Kingdom

[2]The Alan Turing Institute, London, United Kingdom

[3]Bipolar and Depressive Disorders Unit, Department of Psychiatry and Psychology, Hospital Clínic de Barcelona, Barcelona, Spain

[4]Institut d'Investigacions Biomèdiques August Pi i Sunyer, Barcelona, Spain

[5]Centro de Investigación Biomédica en Red de Salud Mental, Instituto de Salud Carlos III, Madrid, Spain

[6]Departament de Medicina, Facultat de Medicina i Ciències de la Salut, Universitat de Barcelona, Barcelona, Spain

[7]Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom

[8]Division of Psychiatry, Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, United Kingdom

[9]Generation Scotland, Institute for Genetics and Cancer, University of Edinburgh, Edinburgh, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Filippo Corponi, MSc, MD
School of Informatics
University of Edinburgh
Informatics Forum, 10 Crichton St, Newington
Edinburgh, EH89AB
United Kingdom
Phone: 44 131 651 5661
Email: filippo.corponi@ed.ac.uk

Abstract

Background: Personal sensing, leveraging data passively and near-continuously collected with wearables from patients in their ecological environment, is a promising paradigm to monitor mood disorders (MDs), a major determinant of the worldwide disease burden. However, collecting and annotating wearable data is resource intensive. Studies of this kind can thus typically afford to recruit only a few dozen patients. This constitutes one of the major obstacles to applying modern supervised machine learning techniques to MD detection.

Objective: In this paper, we overcame this data bottleneck and advanced the detection of acute MD episodes from wearables' data on the back of recent advances in self-supervised learning (SSL). This approach leverages unlabeled data to learn representations during pretraining, subsequently exploited for a supervised task.

Methods: We collected open access data sets recording with the Empatica E4 wristband spanning different, unrelated to MD monitoring, personal sensing tasks—from emotion recognition in Super Mario players to stress detection in undergraduates—and devised a preprocessing pipeline performing on-/off-body detection, sleep/wake detection, segmentation, and (optionally) feature extraction. With 161 E4-recorded subjects, we introduced E4SelfLearning, the largest-to-date open access collection, and its preprocessing pipeline. We developed a novel E4-tailored transformer (E4mer) architecture, serving as the blueprint for both SSL and fully supervised learning; we assessed whether and under which conditions self-supervised pretraining led to an

improvement over fully supervised baselines (ie, the fully supervised E4mer and pre–deep learning algorithms) in detecting acute MD episodes from recording segments taken in 64 (n=32, 50%, acute, n=32, 50%, stable) patients.

Results: SSL significantly outperformed fully supervised pipelines using either our novel E4mer or extreme gradient boosting (XGBoost): n=3353 (81.23%) against n=3110 (75.35%; E4mer) and n=2973 (72.02%; XGBoost) correctly classified recording segments from a total of 4128 segments. SSL performance was strongly associated with the specific surrogate task used for pretraining, as well as with unlabeled data availability.

Conclusions: We showed that SSL, a paradigm where a model is pretrained on unlabeled data with no need for human annotations before deployment on the supervised target task of interest, helps overcome the annotation bottleneck; the choice of the pretraining surrogate task and the size of unlabeled data for pretraining are key determinants of SSL success. We introduced E4mer, which can be used for SSL, and shared the E4SelfLearning collection, along with its preprocessing pipeline, which can foster and expedite future research into SSL for personal sensing.

(JMIR Mhealth Uhealth 2024;12:e55094) doi:10.2196/55094

KEYWORDS

mood disorder; time-series classification; wearable; personal sensing; deep learning; self-supervised learning; transformer

Introduction

Mood disorders (MDs) are a group of mental health conditions in the *Diagnostic and Statistical Manual, Fifth Edition* (DSM-5) classification system [1]. They are chronic, recurrent disorders featuring disturbances in emotions, energy, and thought, standing out as a leading cause of worldwide disability [2,3] and suicidality [4]. Timely recognition of MD episodes is critical toward better outcomes [5]. However, this is challenging due to generally limited patient insight [6], compounded with the low availability of specialized care for MDs, with rising demand straining current capacity [7,8].

Personal sensing, involving the use of machine learning (ML) to harness data passively and near-continuously collected with wearable devices from patients in their ecological environment, has been attracting interest as a promising paradigm to address this gap [9]. Indeed, some of the core MD clinical features (eg, disturbance in mood and energy levels) translate into changes in physiological parameters measurable with wearable devices [10-12]. A major barrier to the development of clinical decision support systems featuring personal sensing has been the scarcity of labeled data, that is, data with annotations by clinicians about the MD state (eg, diagnosis, disease phase, symptom severity). Collecting and annotating data for personal sensing in MDs is, indeed, an expensive and time-consuming enterprise; thus, studies typically use samples running into only a few dozen patients [13-20].

In this work, we took a different perspective and leveraged *unlabeled* data collected with the Empatica E4 (hereafter E4) wristband [21], a popular research-grade device for personal sensing studies [22], as well as recent advancements in self-supervised learning (SSL) techniques that can learn meaningful representations from such unlabeled data. Specifically, we took advantage of open access data sets that record physiological data with the E4 across different settings but do not address MDs and therefore do not provide information about the mood state of the subjects involved. Although each such data set has only a limited number of subjects, our aggregated and preprocessed data set E4SelfLearning can break the labeled data bottleneck for personal sensing in MDs (Figure 1) [23-33].

Fully supervised systems require vast amounts of data to train, thus limiting their application in different fields, such as health care, where amassing large, high-quality data sets is demanding in terms of time and human resources [34]. Although previous studies on personal sensing for MDs have investigated different tasks, including acute MD episode detection [13-16], regression of a psychometric scale total score [17-19], and, more recently, multitask inference of all items in 2 commonly used psychometric scales [35], they all developed their models in a fully supervised fashion (ie, they were trained on samples for which ground-truth labels were available). As a result, considering that obtaining clinical annotations from patients, especially when on an acute MD episode, is a challenging and expensive enterprise, the sample size is generally modest (eg, N=52 in Côté-Allard et al [15], N=45 in Tazawa et al [13], and N=31 in Pedrelli et al [18]).

SSL, in contrast, is a framework where the model creates proxy supervisory signals within the data themselves, therefore alleviating the annotation bottleneck and allowing us to repurpose existing unlabeled data sets [36]. Specifically, SSL derives supervisory signals from the data themselves, thanks to pretext tasks, which are new supervised challenges, for example, imputing occluded parts of the input data. Through such preparatory pretext tasks, not requiring expert annotation, the model learns useful representations, partial solutions to the downstream target task of interest, for which only a comparatively small amount of annotated data are available [37]. On the back of the great success of SSL in computer vision (CV) [37] and natural language processing (NLP) [38], and with encouraging findings in other health care applications [39], we extended pioneering SSL works on multivariate time series [40-42] to personal sensing in MDs.

In this work, we made the following contributions:

- We gathered 11 open access data sets recording physiological data with an E4 wristband and developed a pipeline for preprocessing such data that performed on-/off-body detection, sleep/wake detection, segmentation, and (optionally) feature extraction. We made the

preprocessing pipeline and the preprocessed data publicly available. This collection (E4SelfLearning), with 161 subjects, is the biggest open access data set to date. We believe that this effort can stimulate future research into SSL with multivariate time-series sensory data by removing 2 barriers, preprocessing and data availability.

- We proposed a novel E4-tailored transformer (E4mer) architecture (Figure 2) [43] and showed that SSL is a viable paradigm, outperforming both fully supervised E4mer and classical machine learning (CML) models using handcrafted features in distinguishing acute MD episodes from clinical stability (euthymia in psychiatric parlance), that is, a time-series (binary) classification task.
- We investigated what makes SSL successful. Specifically, we compared 2 main pretext task designs (ie, masked prediction [MP] and transformation prediction [TP]) [44], and for the best-performing routine, we studied its sensitivity to the unlabeled data availability in ablation analyses. We inspected learned embeddings and showed that they capture meaningful semantics about the underlying context (ie, sleep/wake status) and symptom severity.

Figure 1. A total of ~6254 hours (261 days) of unlabeled recordings from 252 subjects while awake were used for self-supervised pretraining. Unlabeled data comprised a collection of 11 open access data sets, whose aggregation we make publicly available (E4SelfLearning), along with part of the TIMEBASE/INTREPIBD study that was not relevant for the target task under investigation (ie, acute episode vs euthymia classification). Unlabeled data were passed through a model consisting of an encoder and a transform head for self-supervised pretraining; the pretrained encoder block was then retained for the target task, while the transform head was replaced with a new, randomly initialized classification head. *The target task (labeled) training set from the TIMEBASE/INTREPIBD study was also used during self-supervised pretraining. Further details on the data sets used in this study are available in Table S1 in Multimedia Appendix 1. ADARP: Alcohol and Drug Abuse Research Program; PGG-DaLia: PPG Dataset for Motion Compensation and Heart Rate Estimation in Daily Life Activities; TIMEBASE/INTREPIBD: Identifying Digital Biomarkers of Illness Activity in Bipolar Disorder/Identifying Digital Biomarkers of Illness Activity and Treatment Response in Bipolar Disorder; UE4W: Unlabeled Empatica E4 Wristband; WEEE: Wearable Human Energy Expenditure Estimation; WESAD: Wearable Stress and Affect Detection; WESD: Wearable Exam Stress Dataset.

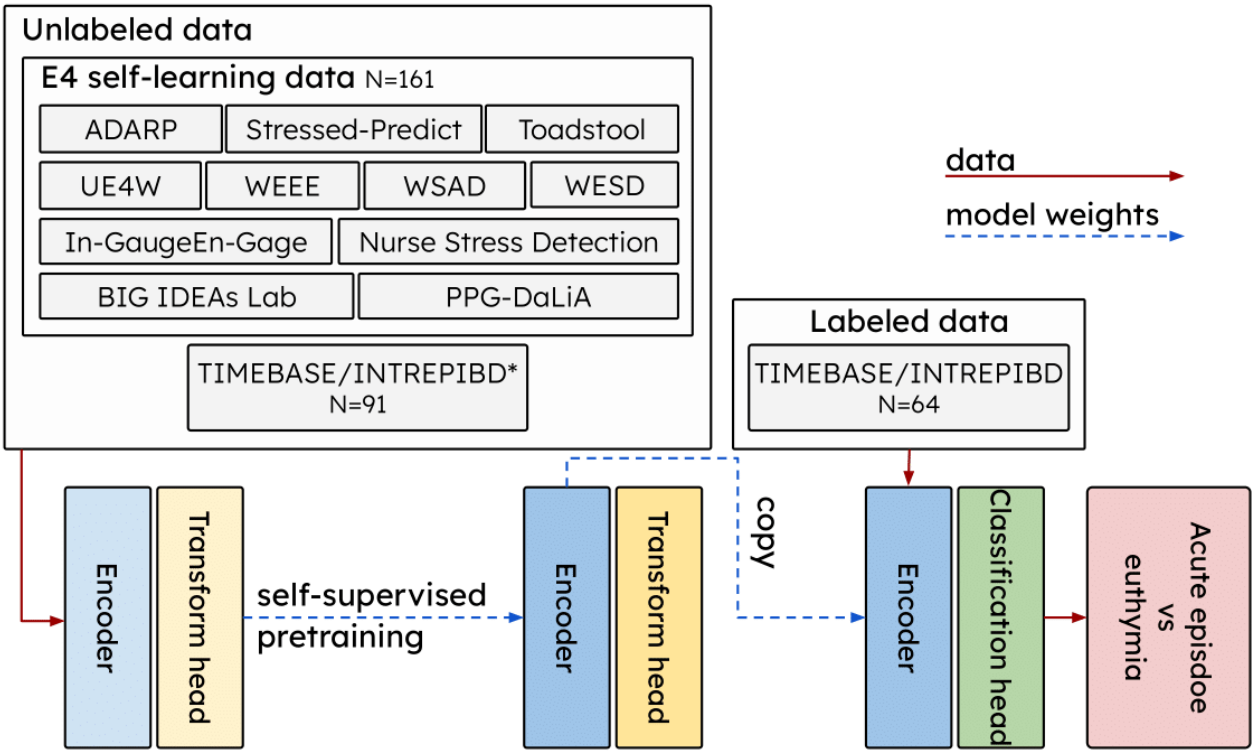
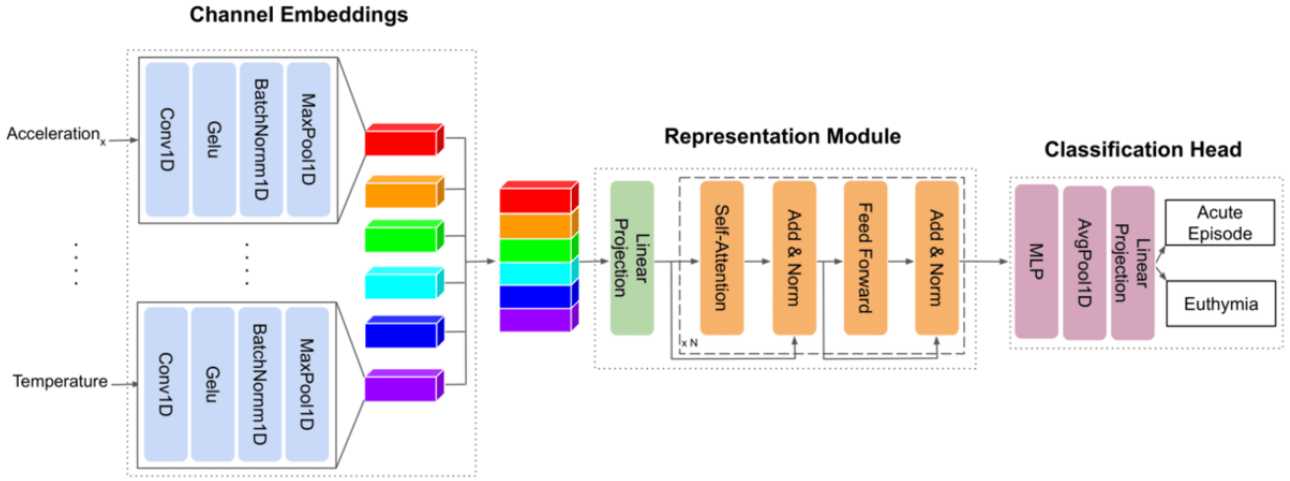


Figure 2. E4mer is a transformer model tailored to the Empatica E4 input data. E4mer consists of 3 sequential modules: (1) channel embeddings set in parallel, 1 for each Empatica E4 raw input channel (ie, acceleration$_x$, acceleration$_y$, acceleration$_z$, BVP, EDA, TEMP), extracting features and mapping channels to tensors of dimensionality (B=batch size, N=time steps, F=number of filters) so that they can be conveniently concatenated along dimension F; (2) RM learning contextual representations of the input time steps within the input segment, thanks to the multihead self-attention mechanism; (3) classification head outputting probabilities for the 2 target classes (ie, acute MD episode and euthymia). SSL models used in our experiments featured the same E4mer architecture described before, where, however, the classification head was replaced with a transform head projecting onto a label space compatible with the pretext task at hand. BVP: blood volume pressure; E4mer: E4-tailored transformer; EDA: electrodermal activity; MD: mood disorder; MLP: multilayer perceptron; RM: representation module; SSL: self-supervised learning; TEMP: temperature.



Methods

Study Sample

The TIMEBASE/INTREPIBD Cohort

Our target task was to distinguish acute MD episodes from euthymia using wearable data. We started from a data set for which we had labeled samples, the TIMEBASE/INTREPIBD (Identifying Digital Biomarkers of Illness Activity in Bipolar Disorder/Identifying Digital Biomarkers of Illness Activity and Treatment Response in Bipolar Disorder) cohort [45]. A detailed description of the data collection campaign was given by Anmella et al [45]. In brief, this was a prospective, exploratory, observational study conducted at the Hospital Clìnic, Barcelona, Spain. Patients with a DSM-5 diagnosis of either major depressive disorder (MDD) or bipolar disorder (BD) were enrolled either in the acute affective episode group (defined according to the "Structured Clinical Interview" for DSM-5 disorder criteria) or in the euthymia group (score≤7 on the Hamilton Depression Rating Scale-17 [46] and the Young Mania Rating Scale [47] for at least 8 weeks [48], as confirmed with weekly ambulatory assessments). The former group had post–acute-phase follow-ups, which were, however, excluded from all analyses presented here. At the time of conducting this study, a total of 64 patients were available for the target task, half in the acute affective episode group and half in the euthymia group. Additionally, an extra 91 subjects (including healthy controls, subjects with schizophrenia, and subjects with a substance abuse disorder), whose status was not relevant to the target task, were available from the TIMEBASE/INTREPIBD cohort for self-supervised pretraining.

Patients were interviewed by a psychiatrist collecting clinical demographics (Table 1 and Table S2 in Multimedia Appendix 1) and were required to wear on their nondominant wrist an E4 wristband until the battery ran out (~48 hours). The E4 records 3D acceleration (sampling rate 32 Hz), blood volume pressure (BVP, sampling rate 64 Hz), electrodermal activity (EDA, sampling rate 4 Hz), heart rate (HR, sampling rate 1 Hz), interbeat interval (IBI, ie, the time between 2 consecutive heart ventricular contractions), and skin temperature (TEMP, sampling rate 1 Hz).

As shown in Table 1, MD episodes clinically lie on a spectrum, with depression on one end and mania on the other; mixed episodes, featuring symptoms from both polarities, are a bridge between the 2 spectrum extremes. In this study, we considered acute MD episodes of any polarity, and similarly, we considered euthymia as a unique class, whether in the context of a BD or an MDD diagnosis. Medication classes administered to the cohort are shown in Table S2 in Multimedia Appendix 1; Bonferroni-corrected chi-square tests found no significant association between treatment status (being on a given drug class or not) and target class (acute affective episode vs euthymia).

Table 1. Clinical-demographic features of the target task (acute affective episode vs euthymia classification) population (N=64).

Features	Acute affective episode group (n=32)	Euthymia group (n=32)
Age (years), means (SD)	50.56 (13.05)	47.22 (16.06)
Females, n (%)	15 (46.9%)	14 (43.8%)
MDE-BD[a]		
Patients, n (%)	9 (28.1)	—[b]
HDRS[c] score, mean (SD)	20.22 (6.34)	—
YMRS[d] score, mean (SD)	2.56 (3.94)	—
MDE-MDD[e]		
Patients, n (%)	7 (21.9)	—
HDRS score, mean (SD)	25.14 (4.78)	—
YMRS score, mean (SD)	1.86 (2.41)	—
ME[f]		
Patients, n (%)	14 (43.8)	—
HDRS score, mean (SD)	5.67 (4.37)	—
YMRS score, mean (SD)	20.13 (6.28)	—
MX[g]		
Patients, n (%)	2 (6.2)	—
HDRS score, mean (SD)	16 (4.24)	—
YMRS score, mean (SD)	13.5 (4.95)	—
BD[h]		
Patients, n (%)	—	26 (81.3)
HDRS score, mean (SD)	—	2.93 (1.73)
YMRS score, mean (SD)	—	1.3 (1.61)
MDD[i]		
Patients, n (%)	—	6 (18.7)
HDRS score, mean (SD)	—	3.14 (1.95)
YMRS score, mean (SD)	—	0.29 (0.76)

[a]MDE-BD: major depressive episode in bipolar disorder.

[b]Not applicable.

[c]HDRS: Hamilton Depression Rating Scale-17.

[d]YMRS: Young Mania Rating Scale.

[e]MDE-MDD: major depressive episode in major depressive disorder.

[f]ME: manic episode.

[g]MX: mixed episode.

[h]BD: bipolar disorder.

[i]HDRS: Hamilton Depression Rating Scale-17.

E4SelfLearning

For self-supervised pretraining, we gathered 11 open access data sets recording with an E4 [23-33]. Although they all used the same hardware, software, and firmware, such data sets could differ substantially for population, recording setting, and task: from students taking exams [29] or attending classes [31] to nurses carrying out their duty [30] and subjects performing different physical activities [28] or playing Super Mario [27]. Subjects that were not part of the target classes from the TIMEBASE/INTREPIBD study were also included in the unlabeled data for SSL.

Data Preprocessing

Our preprocessing encompassed the following sequential stages: on-/off-body detection, sleep/wake detection, segmentation, and (when preparing data for CML models) feature extraction.

During free-living wear, subjects might remove their device or contact with the wrist might be suboptimal. As a result, off-body periods can be erroneously mistaken for periods of sleep or sedentary behavior, due to the shared feature of an absence of movement. Signal discontinuity in biopotentials, such as EDA, due to a lack of skin contact can be reliably leveraged to detect nonwear periods. As shown by Vieluf et al [49] and Nasseri et al [50], we considered measurements less than 0.05 µS as indicative of off-body status. Furthermore, as we noticed occurrences of values greater than the EDA sensor range (ie, 100 µS [51]), as well as instances of TEMP values outside the physiological range (30°C-40°C), we set both to off-body.

As physiological data vary wildly across sleep and wake statuses, we used sleep/wake detection as a form of data cleaning to reduce the variance in the signal and considered only the wake time in our analyses, especially as most publicly available data sets are recorded in wake conditions. We opted for the algorithm developed by Van Hees et al (*Van Hees*) [52], which was reported as the best-performing algorithm in a recent benchmark study on sleep/wake detection (average F_1-score=79.1) [53]. Like most nonproprietary algorithms, Van Hees uses triaxial acceleration and, specifically, relies on a simple heuristic defining sleep with the absence of a change in the arm angle >5° for 5 minutes or more. To accommodate this rule, wherever on-body sampling cycles did not constitute unbroken sequences of at least a 5-minute duration, all the measurements in that period were considered as off-body and discarded from further analysis.

The wake time from each recording was then segmented with a sliding window, whose segment length (ω) and step size (Δω) were set to 512 and 128 seconds, respectively. This approach, also referred to as window slicing [54], is a common form of data augmentation in time-series classification as multiple segments are produced from a single recording, each one marked with the same label, and is common in personal sensing for MDs. Previous relevant works [15,18,55] have defined ω (Δω) based on clinical intuition and convenience concerning the available data. Another work [35] investigating the regression of HDRS and the YMRS items found the optimal ω through tuning, a computationally expensive approach in our setting; however, it showed that ω was not among the most important hyperparameters for the task at hand. Here, we opted for 512 seconds (~8.5 minutes, conveniently a power of 2 for computational efficiency in binary computers), similar to the 5-minute intervals used by Panagiotou et al [55] for training neural autoencoder architectures on anomaly detection by reconstruction error estimation. Our choice was a trade-off between clinical insight and technical constraints. Clinical intuition suggests that too small a value of ω may be ill suited to capture enough information toward acute affective episode versus euthymia discrimination. However, unlabeled data sets used for self-supervised pretraining recorded relatively short sessions (eg 1 hour [26]). As both CML and deep learning models are trained on individual segments and too long a segment length equates to fewer training data points, a 512-second-long segment allowed us to have enough data for developing ML models [55].

Recording segments constituted our basic unit of analysis, and for the target task, segments from the same recording all shared the same ground-truth label (ie, either acute affective episode or euthymia). When fed to deep learning models, segments were channel-wise standardized by subtracting the mean and dividing by the SD. Such statistics were learned from the target task training set or, in the case of SSL, its aggregation with unlabeled data. Acceleration, the BVP, EDA, and TEMP were considered in deep learning models, while the HR and the IBI, as features derived from the BVP through a proprietary algorithm, were excluded from the deep learning experiments shown here (see Multimedia Appendix 1). However, when using CML, handcrafted features were extracted from segments using *FLIRT* [56], a popular open access feature extraction toolkit for the E4. Note that a single row of features per segment was extracted; in other words, the window size parameter in FLIRT was set equal to ω. We used all features available through this package, derived with the *flirt.acc.get_acc_features* (eg, acceleration entropy), *flirt.eda.get_eda_features* (eg, tonic and phasic EDA components), and *flirt.hrv.get_hrv_features* (eg, HR and HR variability measures) functions. As FLIRT does provide built-in functions for TEMP, we also extracted the segment mean (SD) for this channel. Any missing value was handled with mean imputation. The percentage rate of missing values had a range of 0-37.31, with a mean of 10.44 (SD 16.78).

Data Splits and Metrics

In SSL experiments, we split unlabeled data in a ratio of 85:15 into train and validation sets, partitioning recordings across the 2 sets. For the target task, we investigated a time-split scenario, therefore splitting each recording into train, validation, and test sets again in a ratio of 70:15:15 along the recording time, thus testing generalization across future time points. We made sure that segments with overlapping motifs at the border between target task splits (resulting from using a sliding window with $\Delta\omega<\omega$) were confined to 1 split only, thus ultimately producing 18896, 3904, and 4128 segments for the train, validation, and test sets. The target task validation set doubled as a test set for estimating generalization performance on the SSL pretext task. The time-split scenario is common in personal sensing for MDs (eg, [18,35]), and indeed, despite efforts toward learning subject-invariant representations [57,58], cross-subject generalization remains an unsolved challenge, so personal sensing systems typically require access to each subject's physiological data distribution at training time [59].

The target task was a time-series binary classification. As expected in free-living wear, the total wear time and the off-body and wake times varies across subjects (and, as a result, so did the number of segments). Two-tailed *t* tests were performed to verify significant mean differences in off-body and wake times across individuals from the 2 target classes (acute affective episode and euthymia) but yielded a Bonferroni-corrected P value of >.05 (P=.56 for off-body time and P=.82 for wake time). An equal number of segments from each class was extracted for the target task. To that end, we found the pairing of euthymia and acute affective episode recordings that minimized the pairwise difference between the number of segments available per participant; next, within each pair, the first n segments were retained, where n is the number of

segments of the shortest recording in the pair. We optimized models on the target task for segment-level accuracy ($ACC_{segment}$). Second, to provide a subject-level perspective, we reported the subject ACC:

where y_s is the ground-truth mood state of the s-th subject, which is constant across all the s-th subject's recording segments, and  is a majority vote on the s-th subject, corresponding to the majority predicted class across the s-th subject's recording segments.

Machine Learning Models

We developed 2 types of baselines for the target task: (1) an E4-tailored deep learning pipeline inputting raw recording segments (E4mer) and (2) CML models using handcrafted features extracted with FLIRT from recording segments. We then assessed what boost in performance, if any, a self-supervised pretraining phase might deliver, where the SSL models shared the same building blocks as E4mer.

Baseline Models

E4-Tailored Transformer

E4mer is an artificial neural network discriminative classifier modeling the probability of an acute MD episode, given a recording segment. As shown in Figure 2, E4mer has 3 sequential blocks: (1) channel embeddings (CEs) set in parallel, consisting of the same 1D convolutions with a kernel size equal to the channel sampling frequency, followed by Gaussian error linear unit (Gelu) activation, 1D BatchNorm, and 1D MaxPooling using the channel sampling frequency as both kernel size and step size, so each CE output has the same dimensionality and can be conveniently concatenated with the others before being passed onto (2) a transformer [43] representation module (RM), and (3) a multilayer perceptron (MLP) classification head (H_{sl}). The CEs extract features from the input E4 channels and are designed to handle channels sampled at different frequencies; the RM, powered by multihead self-attention, learns contextual representations of the input tokens (timestamps in our case) within a recording segment; lastly, the H_{sl} maps such representations onto a label space appropriate for a binary classification. E4mer was trained to minimize the binary cross-entropy (BCE) loss between acute affective episode/euthymia predictions and the corresponding ground truth.

Classical Machine Learning

We experimented with the following algorithms, given their popularity and state-of-the-art performance in biomedical applications [60], including personal sensing [13,14]: elastic net logistic regression (ENET), K-nearest neighbor (KNN), support vector machine (SVM), and extreme gradient boosting (XGBoost).

Self-Supervised Learning Schemes

SSL schemes rely on devising a pretext task, for which a (relatively) large amount of unlabeled data is available, conducive to learning, during a pretraining phase, representations useful to solve the downstream target task [44]. What defines an SSL paradigm is thus its pretext task, consisting of a process, P, to generate pseudo labels and an objective to guide the pretraining. An SSL model typically consists of (1) an encoder EN(x;θ): X->V, learning a mapping from input views  to a representation vector  , and (2) a transform head $H_{ssl}(\upsilon;\xi)$: V-> Z, projecting the feature embedding into a label space  compatible with the pretext task at hand. When solving the target task, the pretrained encoder EN is retained as a partial solution to the target problem, whereas the pretrained transform head H_{ssl} is discarded and replaced with a new one, H_{sl}. Next, EN's parameter θ may be kept fixed and only H_{sl}'s parameters may be learned on the target task. This approach, often referred to as *linear readout* (LR), amounts to treating EN as a frozen feature extractor. Alternatively, instead of just training a new head, the entire network may be retrained on the target task, initializing EN's parameter θ to the values learned during self-supervised pretraining, a paradigm known as *fine-tuning* (FT). Our SSL models used the same architecture as E4mer, that is, an encoder EN, consisting of convolutional CEs, followed by a transformer RM, and an MLP for the transform head H_{ssl}. The success of SSL methods largely comes from designing appropriate pretext tasks that produce representations useful for the downstream target task. This usually involves domain knowledge of the target task. We investigated how different pretext tasks affected downstream performance, experimenting with 2 popular SSL routines that have shown success in other applications: MP and TP.

Masked Prediction

This family of SSL methods is characterized by training the model to impute data that have been removed or corrupted by P. It relies on the assumption that context can be used to infer some types of missing information in the data if the domain is well modeled. This strategy was popularized by the huge success of bidirectional encoder representations from transformers (BERT) [38] in NLP applications, and 1 of the first adaptations to multivariate time-series classification was proposed by Zerveas et al [41]. Similar to their implementation, for each segment channel, we sampled a Boolean mask where the sequences of 0s and 1s were sampled from geometric distributions with means of l_0 and l_1, respectively, with:

where r is the masking ratio. As shown by Zerveas et al [41], the average length of the 0 sequences (l_m) and the proportion of masked values (r) were set to 3 seconds and 0.15, respectively. Each segment channel was then multiplied by its corresponding mask, effectively setting to 0 some of the channel-recorded measurements, and inputted to a model that was tasked to recover the original channel values. This was done by minimizing the root mean square error (RMSE) between the masked original value x(t, c) and its reconstruction outputted by the network  :

where c and t, respectively, index the channels, and the timestamps of the 0 values in the masks M and |M| are the total number of 0s sampled (ie, the masks' cardinality).

Transformation Prediction

We followed the implementation shown by Wu et al [42], which used SSL for a target task of emotion recognition with E4 recordings. In brief, for each channel, 1 of 6 transformations (ie, identity, Gaussian noise addition, magnitude warping, permutation, time warping, and cropping) was sampled uniformly at random and then applied. The transformed segment was then inputted into a model, which was tasked to guess, for each channel, which of the 6 transformations was applied. This amounted to a multitask, multiclass classification, where the model was trained to minimize channel average categorical cross-entropy (CCE):

where c indexes the channels and j the transformations, $1_{i,j}$ is an indicator taking value 1 when j is the correct transformation for channel c and 0 otherwise, and $p_{c,j}$ denotes the predicted probability that transformation j was applied to channel c. By solving this task, Wu et al [42] argued that the model learns representations robust to disturbances in the magnitude and time domains.

Tuning

A hyperparameter search for all models was carried out with hyperband Bayesian optimization [61]. For the target task, we selected the setting yielding the highest $ACC_{segment}$ in the validation set, whereas in self-supervised pretraining, we selected hyperparameters associated with the lowest relevant loss in the validation pretraining set. Multimedia Appendix 1 shows the hyperparameter search space and the best configuration across all models. Deep learning models were trained with the AdamW optimizer for a maximum of 300 epochs, with a batch size of 256. Moreover, to speed up the training and search procedure, we used an early stopping learning rate scheduler: we reduced the learning rate α_{LR} by a factor of 0.3 if the model did not improve in its validation performance after 10 consecutive epochs, and we terminated the training procedure if the model did not improve after 2 learning rate reductions. Dropout [62] and weight decay were added to prevent overfitting.

Post hoc Analyses

Toward elucidating key contributors to the viability of SSL, in addition to comparing different pretext task designs, we studied how (1) progressively downsampling unlabeled data sets or (2) removing each data set in turn from the unlabeled collection might impact the performance of our best SSL model. Thus, using the most performative self-supervised scheme, we retrained the SSL model from scratch under configurations (1) and (2) and then tested it on the target task. Note that in both settings, the entire target task training set was kept for pretraining; this is because pretraining on the training set can be always performed at no extra cost in terms of data acquisition. Lastly, we conducted statistical tests to better appreciate how the self-supervised E4mer compared against its fully supervised counterpart and the best-performing CML algorithm and how it was affected by different ablations. Based on whether we considered either (1) recording segments or (2) subjects as our basic analysis units, we had 2 different hypotheses. In (1), we used a linear mixed effects (LME) model to analyze the difference in correct class probabilities between the SSL model and each comparator, considering subjects as a random effect. This accounted for the nested structure of the data, where segments were sampled from individual subjects. A fixed effects intercept was included to test a 0 mean difference between the classifiers at the population level. Additionally, as the ML models we implemented, like most state-of-the-art algorithms [63], effectively treat segments as independent and identically distributed, we used a 2-tailed paired *t* test to assess whether a 0 mean difference in the probability assigned to the correct class was 0. In (2), we checked with a 2-tailed paired *t* test whether the between-classifiers mean difference in the $ACC_{segment}$ by subject was different from 0. To account for multiple testing, within both (1) and (2), a Bonferroni correction was applied. The number of tests was 19, that is, 17 different ablation settings plus 2 tests comparing the best baselines (fully supervised E4mer and the best CML) to SSL.

Code Used

Python 3.10 programming language was used where deep learning and CML models were implemented in PyTorch [64] and Scikit-learn [65]/XGBoost [66] respectively, while hyperparameter tuning was performed in both cases with weights and biases [67]. The best hyperparameter setting found during tuning for each model is reported in Multimedia Appendix 1. All deep learning models were trained on a single Nvidia A100 graphical processing unit (GPU).

Ethical Considerations

The TIMEBASE/INTREPIBD study was conducted in accordance with the ethical principles of the Declaration of Helsinki and Good Clinical Practice and the Hospital Clinic Ethics and Research Board (HCB/2021/104). All participants provided written informed consent prior to their inclusion in the study. All data were collected anonymously and stored encrypted in servers complying with the General Data Protection Regulation (GDPR) and the Health Insurance Portability and Accountability Act (HIPAA). Regarding other studies included in this work, we referred to relevant publications.

Results

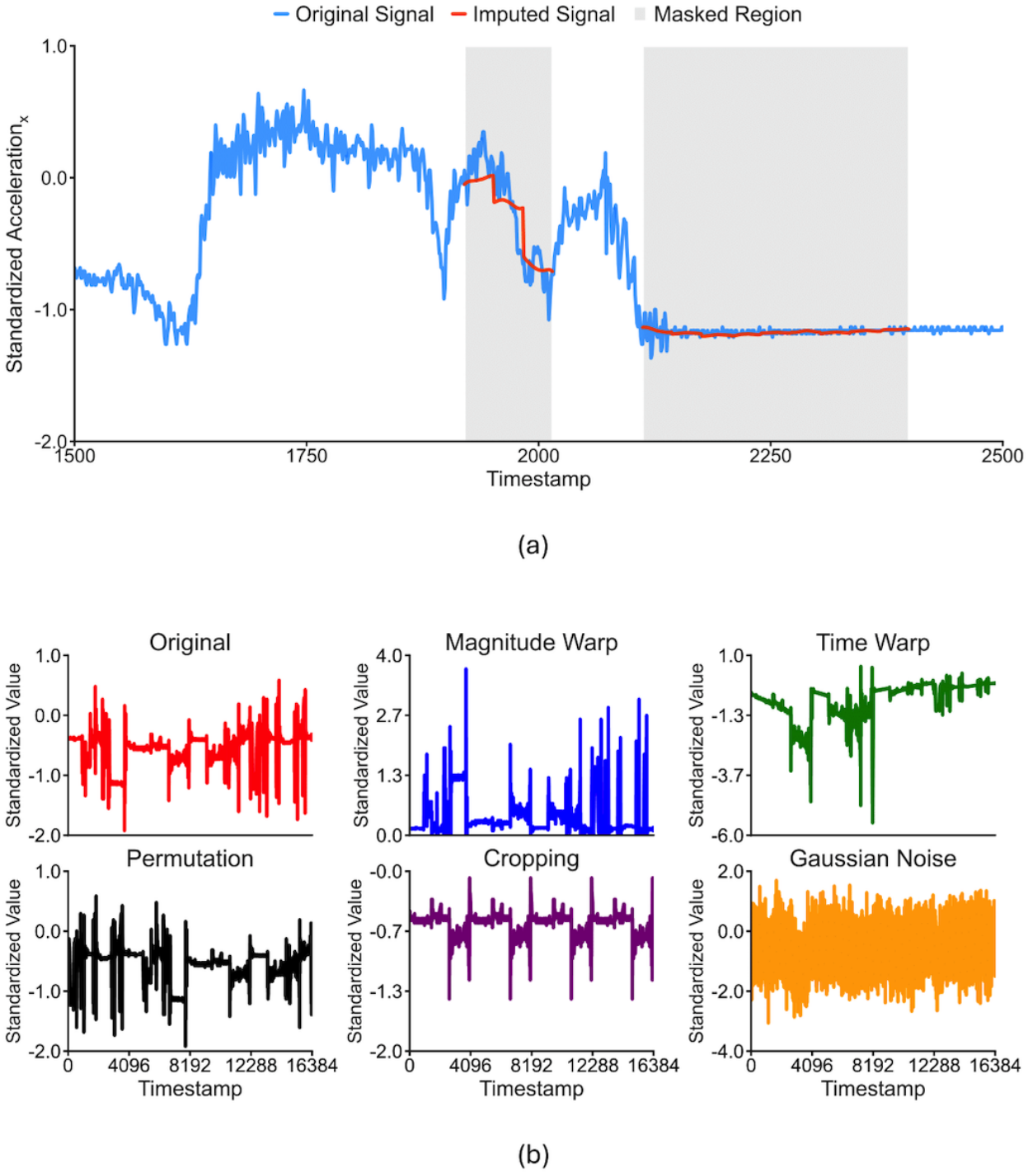
Surrogate Tasks Used in Self-Supervised Pretraining

The same model, using the E4mer architecture (Figure 2), was used across different pretext tasks. Figure 3 illustrates the surrogate tasks we experimented with. In MP (Figure 3a), parts of the input segments were zeroed out by multiplication with a Boolean mask sampled, as shown by Zerveas et al [41], and the model was trained to recover the original input segments. Although the model output entire segments, only the masked values were considered toward the loss computation, that is, the RMSE. The assumption was that the model acquires good representations of the underlying structure of the data when

learning to solve this task. Our best model had an error of 0.1347 on the test set (notice that input segments were channel-wise standardized).

In TP (Figure 3b), 1 transformation was sampled from a set and applied to each channel independently, and the model learned which transformation each channel underwent, minimizing the channel average CCE. We used the same transformations as Wu et al [42], who experimented with an E4 for a downstream task of emotion recognition. The rationale was to encourage robustness against signal disturbances introduced with the transformations. The test loss of the selected model was 0.5000.

Figure 3.



Target Task Performance Comparison

Table 2 illustrates the performance under each model we developed. Although they were all optimized for segment ACC, we also reported subject ACC since in a clinical scenario, a decision needs to be made at the subject level. Note that although ACC was a suitable metric in our use case as data were perfectly balanced, we also provided complementary metrics (precision, recall, F_1-score, and area under the receiver operating characteristic curve [AUROC]), both at the segment and at the patient level. At the subject level, the predicted class was the

result of a majority vote over that subject's segments, while the predicted probabilities under each class were derived by summing segments' predicted probabilities for that subject and normalizing by the corresponding segment number. MP self-supervised pretraining comfortably outperformed end-to-end SSL, while also surpassing other self-supervised approaches.

Table 2. Performance in differentiating an acute MD[a] episode from euthymia across different models.

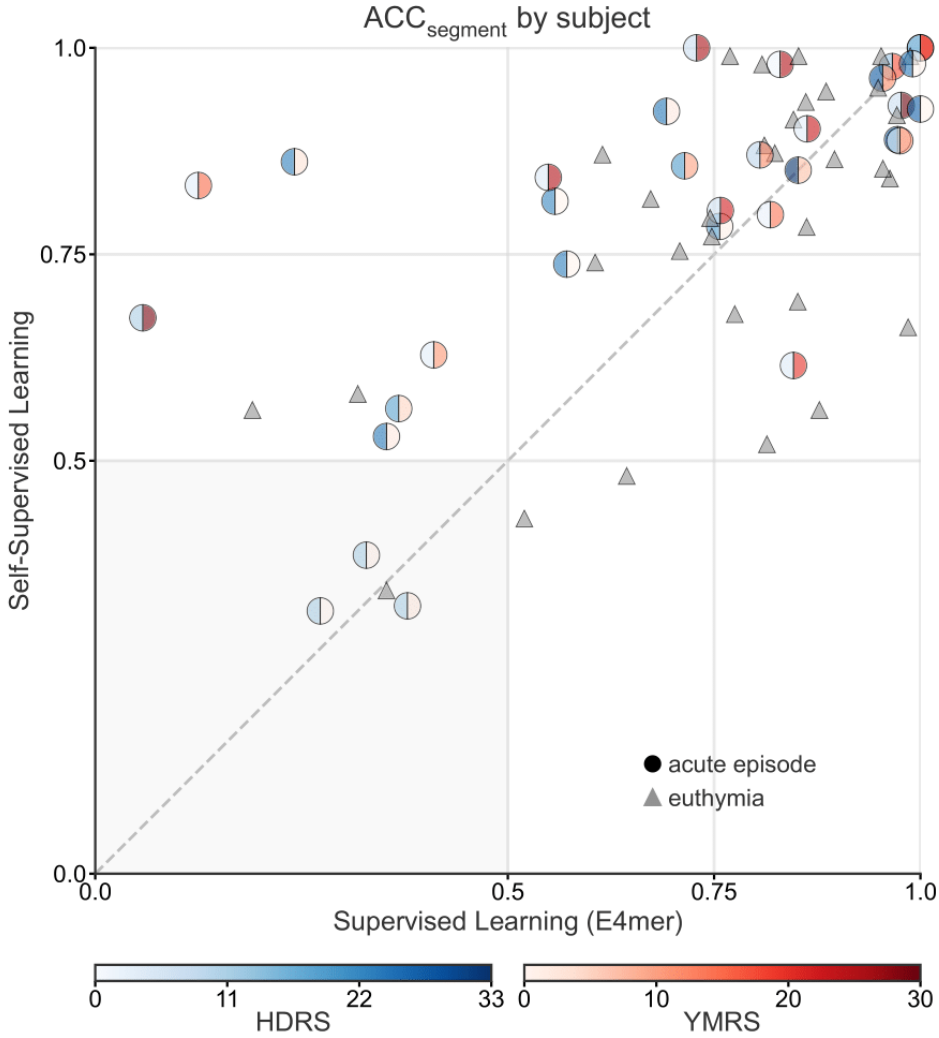
Model	ACC[b]		Precision		Recall		F_1 score		AUROC[c]	
	Segment	Subject	Segment	Subject	Segment	Subject	Segment	Subject	Segment	Subject
SL[d]										
ENET[e]	66.38	71.88	66.22	75	66.86	65.63	66.54	70	72.24	82.25
KNN[f]	70.37	82.81	69.09	80	73.74	81.2	71.34	80.6	73.27	83.26
SVM[g]	71.25	81.25	71.87	80	71.40	77.65	71.63	78.81	73.44	83.21
XGBoost[h]	72.02	82.81	71.33	83	72.11	81.1	71.72	82.03	72.44	83.17
E4mer[i]	75.35	81.25	73.46	80.55	75.34	82.14	74.39	81.33	75.68	82.22
SSL[j]										
MP[k] (LR[l])	77.53	87.5	78.34	88.6	77.41	88	77.87	88.3	78.02	89.2
MP (FT[m])	81.23[n]	90.63[n]	80.91[n]	90.11[n]	82.00[n]	92.87[n]	81.45[n]	91.47[n]	82.02[n]	93.11[n]
TP[o] (LR)	71.16	81.25	72.12	82.44	72.01	82.31	72.06	82.37	71.89	84.12
TP (FT)	75.69	84.38	75.41	82.11	74.79	83.9	75.1	83	75.21	84.23

[a]MD: mood disorder.
[b]ACC: accuracy.
[c]AUROC: area under the receiver operating characteristic curve.
[d]SL: supervised learning.
[e]ENET: elastic net logistic regression.
[f]KNN: K-nearest neighbor.
[g]SVM: support vector machine.
[h]XGBoost: extreme gradient boosting.
[i]E4mer: E4-tailored transformer.
[j]SSL: self-supervised learning
[k]MP: masked prediction.
[l]LR: linear readout.
[m]FT: fine-tuning.
[n]The best results.
[o]TP: transformation prediction.

The E4mer and CML baselines performed to a similar level: although E4mer was superior to XGBoost in terms of $ACC_{segment}$ (75.35 vs 72.02), it was trumped by CML on $ACC_{subject}$ (82.81 vs 81.25). Other CML baselines fared worse than XGBoost. MP pretraining led to a target task performance, substantially higher than the baselines, under both metrics. Although both LR and FT dominated over supervised learning (SL), the latter scored the highest performance with $ACC_{segment}$ and $ACC_{subject}$ of 0.8123 and 0.9063, respectively. However, TP led to only modest improvement over E4mer. Statistical tests comparing the best SSL scheme (ie, MP with FT) against the fully supervised E4mer and XGBoost were significant at both the segment and the subject level. In particular, comparison with E4mer yielded $P_{Bonferroni}$ values of .03 for the LME model and <.001 and .02 for the *t* test at the segment and the subject level, respectively. For XGBoost, $P_{Bonferroni}$ values were .04 for the LME model and <.01 and .01 for the *t* test at the segment and the subject level, respectively.

Comparison of the best SSL with its SL counterpart in terms of $ACC_{segment}$ by subject (Figure 4) suggested that only 2 (3.1%) patients with euthymia were misclassified by SSL but correctly classified by the supervised E4mer. However, SL mispredicted 8 (12.5%) individuals that SSL got right. Patients on an acute MD episode are shown as dots with a color gradient proportional to their total score on the HDRS [46] (left half) and the YMRS [47] (right half), 2 clinician-administered questionnaires tracking depression and mania severity, respectively. Subjects on an acute MD episode misclassified by SL included patients with severe depressive (or manic) symptomatology. Notably, both SSL and SL failed in the case of 4 (6.3%) subjects, including 3 (75%) patients on an acute MD episode with relatively moderate severity.

Figure 4. SSL beats SL by 4 (9.4%) more correctly classified subjects. ACCsegment under SSL and SL (E4mer) within each subject's test segments: subjects in the euthymia group are represented as triangles, while subjects on an acute affective episode are shown as circles with the left half colored in blue and the right half in red, with a gradient proportional to the total sum on the HDRS and the YMRS, respectively. Subjects' position on the x and y axes corresponds to their proportion of recording segments correctly classified by SL and SSL, respectively. Note that a subject's majority vote over their segments is in agreement with the subject's true mood state when the proportion of correctly classified segments from that subject is greater than 0.5. The HDRS and the YMRS range shown on the color bar refer to values scored in the TIMEBASE/INTREPIBD sample, while the total score, in general, range is 0-52 and 0-60, respectively. ACCsegment: segment accuracy; E4mer: E4-tailored transformer; HDRS: Hamilton Depression Rating Scale-17; SL: supervised learning; SSL: self-supervised learning; TIMEBASE/INTREPIBD: Identifying Digital Biomarkers of Illness Activity in Bipolar Disorder/Identifying Digital Biomarkers of Illness Activity and Treatment Response in Bipolar Disorder.



Ablation Analyses and Learned Representations

Tables 3 and 4 show the difference in the target task $ACC_{segment}$ and $ACC_{subject}$ resulting from pretraining the best SSL on parts of the unlabeled data collection and then FT it onto the target task. Ablation analyses showed a positive trend between unlabeled data availability and target task performance, but data set–specific unobserved factors likely played a role. The difference in $ACC_{segment}$ and $ACC_{subject}$ from pretraining on just parts of the entire unlabeled data collection is shown in the tables. An LME model and a 2-tailed paired *t* test assessed whether the mean difference in predicted probabilities for the segment's correct class differed from 0, with the former correcting for subjects as a random effect. A 2-tailed paired *t* test assessed whether the mean difference in the number of correctly classified segments by subject differed from 0. In each test, the comparator was the best-performing self-supervised model. P values are corrected with Bonferroni's method. Note that a majority vote over a subject's segments was used to issue subject-level predictions, and $ACC_{subject}$ was simply the fraction of correct majority votes in the test set. $ACC_{subject}$, therefore, did not consider the proportion of votes over a subject's segments in favor of the subject's correct class but just whether a majority, no matter how small or large, was reached in agreement with the correct class. However, the *t* test (subject) assessed a 0 mean difference in the proportion of votes, within subjects, for the correct class. As shown in Table 3, self-supervised pretraining, preceding FT on the target task, therefore used only a fraction of the total unlabeled collection. A resampling ratio of 0% meant that self-supervised pretraining was performed on the target training set only.

The Pearson correlation coefficient (PCC) between unlabeled data downsampling ratios and the difference in $ACC_{segment}$ and $ACC_{subject}$ was 0.9401 and 0.9449, respectively, indicating a strong dependence between performance and unlabeled data

availability. Similarly, excluding individual data sets from pretraining impacted $ACC_{segment}$ and $ACC_{subject}$ proportionally to their relative size (PCC=–0.8185 and –0.4083, respectively). Notably, however, TIMEBASE/INTREPIBD, despite being collected at the same site as the target task data and making up the largest share of the unlabeled data collection, did not leave the largest dent in performance when excluded from training. Furthermore, excluding some data sets resulted in performance improvement. Differences in $ACC_{segment}$ and $ACC_{subject}$ did not always have the same sign because of the way they were defined. Indeed, it is, for example, possible that the absolute number of correctly classified segments decreased but enough previously misclassified segments within a subject were now correctly classified so that the majority vote for that subject flipped. Statistical analyses showed that the ablation of a single data set was associated with nonsignificantly different performance in terms of correctly classified segments within subjects. At the level of the probability assigned to the correct class for each segment, LME results were significant only for a data set, whereas results were mixed for *t* tests. Stratified resampling gave positive results, but the significance for LME was reached only at lower downsampling ratios.

Lastly, we visualized the representations learned by the encoder, EN, part of our best-performing models to gain further insights. As EN's output had dimensionality (B=number of segments, N=number of timestamps, D=transformer's model dimension), for visualization purposes, we averaged out the D axis and then used Uniform Manifold Approximation and Projection (UMAP) [68], a powerful nonlinear dimensionality reduction technique, to embed the resulting N-dimensional data points into 3 dimensions. The top-left plot of Figure 5 shows the representations learned during self-supervised pretraining with MP. The segments shown are the target task test segments, along with an equal number of segments belonging to the same sessions but taken from the sleep state, which the SSL model was never exposed to during training. Wake and sleep segments have different embeddings, suggesting that the model captured this structure in the physiological data: a Gaussian mixture model, indeed, recovered 2 clusters, one with predominantly sleep segments (n=4081, 82.66%) and the other with the majority of wake segments (n=3272, 95.58%). It should be noted that sleep and wake naturally have quite different semantics with respect to physiological data, and the algorithm we used for sleep/wake differentiation (Van Hees [52]) uses a simple heuristic defining sleep as a sustained lack of significant changes in the acceleration angle. The top-right and bottom plots of Figure 5 illustrate the representations from the SSL model upon FT on the target task. The top-right scatter plot displays the target task test segments, as well as pretraining validation set segments (except for the pretraining segments from the TIMEBASE/INTREPIBD collection). The latter group of segments we assumed as being taken from subjects without an acute MD episode and, arguably, most even without any historical MD diagnosis, since the open access data sets we found did not select for patients with an MD. The plot shows 3 clusters whose composition, as recovered with a Gaussian mixture model, was as follows: (1) n=1464 (79.26%) acute MD episode and n=383 (20.7%) euthymia; (2) n=1120 (74.16%) euthymia and n=390 (25.84%) acute MD episode; and (3) n=7801 (91.01%) unlabeled segments, n=683 (7.96%) euthymia, and n=88 (1.02%) acute MD episode. The bottom plots in Figure 5 show target task segments test segments only (no unlabeled segment), colored with a gradient proportional to symptoms' severity, as assessed with the HDRS [46] and the YMRS [47]. Embeddings would seem to suggest a progression in symptoms' severity across the 2 clusters of segments on the right of the scatter plot.

Table 3. Ablation analyses results: the unlabeled collection was downsampled, stratifying by data sets.

Resampling ratio	80%	60%	40%	20%	0%
$ACC^{a}_{segment}$ difference	–0.23[b]	–2.14[b]	–6.07[b]	–6.35[b]	–7.07[b]
$ACC_{subject}$ difference	–1.57[b]	–1.57[b]	–4.70[b]	–4.70[b]	–7.82[b]
LME[c] *P* value	.09	.07	.06	.05	.04
t Test (segment) *P* value	<.001	<.001	<.001	<.001	<.001
t Test (subject) *P* value	.001	.001	.001	.001	.001

[a]ACC: accuracy.

[b]Deterioration in performance upon retraining on the ablated unlabeled data collection.

[c]LME: linear mixed effects.

Table 4. Ablation analyses results: self-supervised pretraining was conducted, leaving out each data set in turn from the unlabeled collection.

Data set	Relative size	ACC[a]$_{segment}$ difference	ACC$_{subject}$ difference	LME[b] *P* value	*t* Test (segment) *P* value	*t* Test (subject) *P* value
Alcohol and Drug Abuse Research Program (ADARP)	12.34	–2.44[c]	–1.57[c]	.01	<.001	.99
Stress Predict	0.30	–0.21[c]	–1.57[c]	.23	.99	.99
Toadstool	0.04	0.52[d]	–3.13[c]	.99	.99	.99
Unlabeled Empatica E4 Wristband (UE4W)	2.32	–1.93[c]	–4.70[c]	.99	.003	.99
Wearable Human Energy Expenditure Estimation (WEEE)	0.18	1.19[d]	1.57[d]	.99	.90	.99
Wearable Stress and Affect Detection (WESAD)	0.42	–0.51[c]	–1.57[c]	.99	.99	.99
Wearable Exam Stress Dataset (WESD)	0.72	1.90[d]	1.57[d]	.06	.05	.99
In-GaugeEn-Gage	17.55	–4.44[c]	–4.70[c]	.99	<.001	.63
Nurse Stress Detection	11.82	–0.81[c]	–1.57[c]	.99	.99	.99
BIG IDEAs Lab	19.38	–2.09[c]	–1.57[c]	.99	.08	.99
PPG Dataset for Motion Compensation and Heart Rate Estimation in Daily Life Activities (PPG-DaLiA)	0.69	1.93[d]	4.70[d]	.53	<.001	.99
TIMEBASE/INTREPIBD[e]	34.24	–4.32[c]	–3.13[c]	.99	.99	.38

[a]ACC: accuracy.

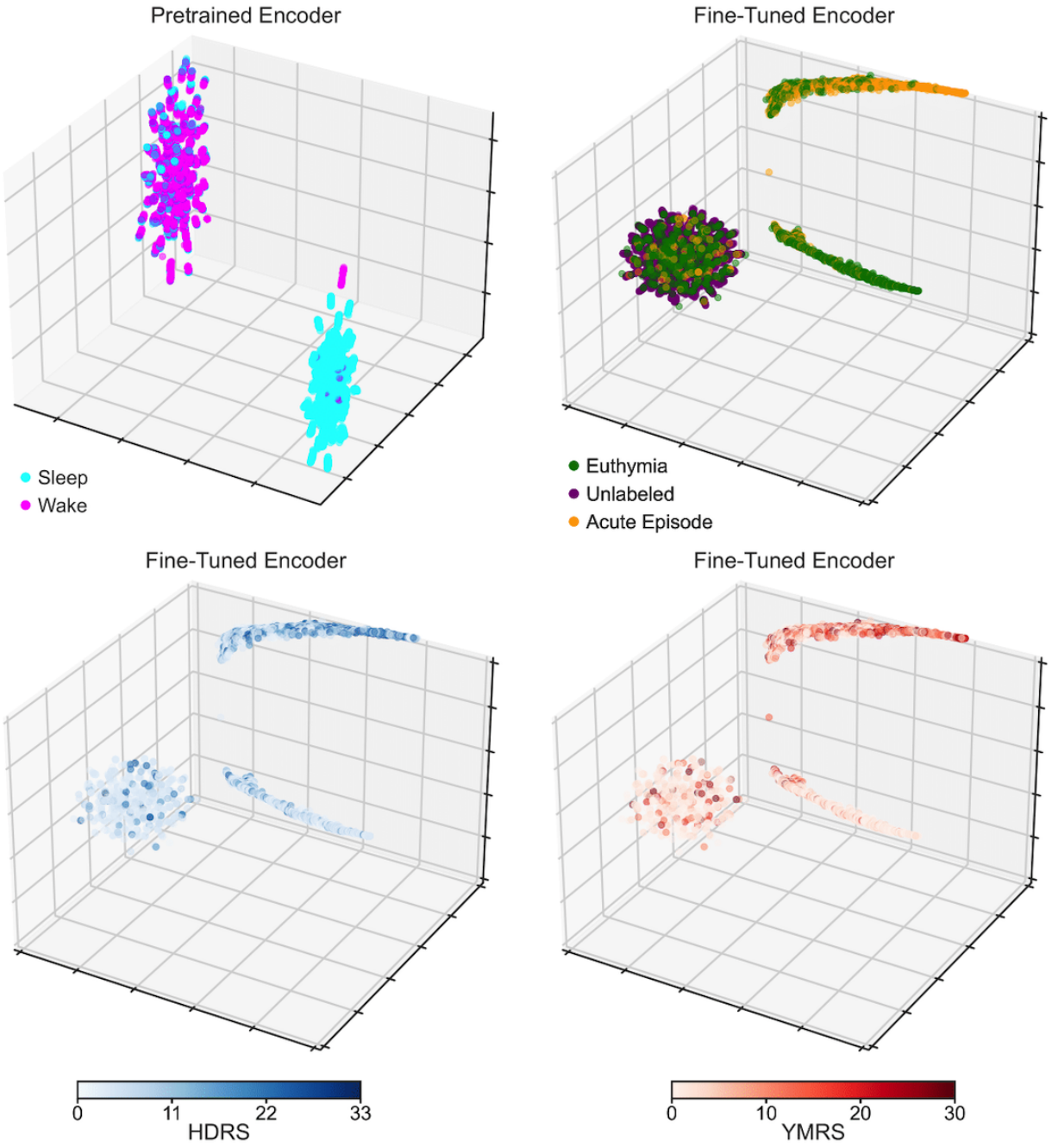
[b]LME: linear mixed effects.

[c]Deterioration in performance upon retraining on the ablated unlabeled data collection.

[d]Improvement in performance upon retraining on the ablated unlabeled data collection.

[e]TIMEBASE/INTREPIBD: Identifying Digital Biomarkers of Illness Activity in Bipolar Disorder/Identifying Digital Biomarkers of Illness Activity and Treatment Response in Bipolar Disorder.

Figure 5. Reassuringly, the learned embeddings seem to have captured meaningful semantics about the underlying context. (Top left) Embeddings from the encoder pretrained on MP map sleep and wake segments to different parts of the latent space. (Top right) Embeddings from the encoder FT on the target task show that segments from the unlabeled open access data sets, which presumably do not contain subjects on an acute MD episode, tend to cluster with part of the segments from patients in euthymia. Embeddings from the fine-tuned encoder show a gradient in symptoms' severity across target task segments, as revealed by (bottom left) the HDRS and (bottom right) the YMRS total score. Note that unlabeled segments are not shown in the bottom left or right plot and that the HDRS and YMRS ranges shown on the color bar refer to values scored in the TIMEBASE/INTREPIBD sample, while the total score range, in general, can be 0-52 and 0-60, respectively. FT: fine-tuning; HDRS: Hamilton Depression Rating Scale-17; MD: mood disorder; MP: masked prediction; TIMEBASE/INTREPIBD: Identifying Digital Biomarkers of Illness Activity in Bipolar Disorder/Identifying Digital Biomarkers of Illness Activity and Treatment Response in Bipolar Disorder; YMRS: Young Mania Rating Scale.



Discussion

Principal Findings

Personal sensing is likely to play a key role in health care supply, creating unprecedented opportunities for patient monitoring and just-in-time adaptive interventions [69]. Toward delivering on this promise, expert annotation is a major obstacle; this is especially the case with MDs, wherein data annotation is particularly challenging and time-consuming, considering the nature of the disorders.

To the best of our knowledge, we are the first to show that SSL is a viable paradigm in personal sensing for MDs, mitigating the annotation bottleneck, thanks to the repurposing of existing unlabeled data collected in settings as different as subjects playing Super Mario [27], taking university exams [29], or performing physical exercise [28].

We took on a straightforward yet fundamental task, that is, to distinguish acute MD episodes from euthymia. Timely recognition of an impending MD episode in someone with a historical MD diagnosis regardless of the episode polarity (depressive, manic, or mixed) may, indeed, enable preemptive interventions and better outcomes [5]. Our results suggest that with a sample size on the order of magnitude that is typical of studies into personal sensing for MDs, a modern deep learning fully supervised pipeline (E4mer) may offer no substantial improvements over simpler CML algorithms (eg, XGBoost), despite higher development and computational costs. However, the accumulation and repurposing of existing unlabeled data sets for an SSL pretraining phase leads to a confident margin of improvement: $ACC_{segment}$ and $ACC_{subject}$ improve by 7.8% and 11.54%, respectively, relative to the fully supervised E4mer, with 6 (9.4%) of 64 more subjects correctly classified.

Our findings further show that careful choice of the pretext task, as well documented in the literature on SSL [40], is key toward learning useful representations for the downstream target task. Unlike MP, improvement, if any at all, from TP was only modest. This is not to say that such a pretext task may in general fail to deliver on acute MD episode versus euthymia differentiation. Indeed, the specific transformations we implemented, borrowed from Wu et al [42], may have been suboptimal for our downstream task, pointing to the importance of domain knowledge (including clinical expertise) in pretext task design. Lastly, although SSL relaxes dependence on large, annotated data sets, our results indicate that its success relies on the size of unlabeled data. Ablation analyses, indeed, showed a positive correlation between target task performance and the size of the corpus available for pretraining. Data set–idiosyncratic factors accounting for the nonperfect correlation between the relative size and impact on target task performance may be present. Speculatively, these may include noise in the data, (dis)similarity of recording conditions, or (ir)relevance for the target task of the representations learned modeling the domain of the unlabeled data set.

Statistical analyses showed that excluding from pretraining any of the individual unlabeled data sets, while keeping all others, is not associated with a significant change in performance on the proportion of correctly classified segments within subjects. The lack of a significant effect in either direction (improvement or deterioration), along with a significantly superior performance of SSL over fully supervised schemes, indicate that pretraining on big data collections leads to higher performance than taking on the target task from scratch. Of importance, adding data sets for pretraining from domains not immediately related to the target task did not undermine the model. Pretraining under progressively lower downsampling ratios lent further support to the importance of data size. This is consistent with the deep learning recipe where the bigger the pretraining corpus, the better the results [70]. Results from tests at the level of segment-predicted probabilities are consistent with this view. Of the data sets comprising less than 1% of the entire unlabeled collection, only 1 reached statistical significance. LME has more flexibility to explain the data since rather than pooling all segments together in a unique (bigger) population, it treats them as embedded within subjects. This explains the lack of statistical significance relative to the *t* tests under various data ablation regimes.

Limitations

We acknowledge the following limitations of this study. We deliberately chose the simplest task that has clinical relevance in personal sensing for MDs since our focus was on SSL; however, we appreciate that a more fine-grained MD description, beyond a simple acute MD episode versus euthymia binary classification, may add further clinical value [35]. As the literature on SSL is expanding at a fast pace, a thorough search of different approaches was beyond the scope of this work. We acknowledge that other pretext tasks can be deployed, and although the architectural choice may have an impact on SSL, we settled for just 1 reasonable, modern model design with a transformer [43] as a workhorse for representation learning. Lastly, given the naturalist design of the study, reflective of the intended use of personal sensing in a clinical setting, we could not exclude the effect of confounders, including medications, on the physiological variables. However, we reported medication classes administered in the cohort and verified a lack of any significant association between target classes (euthymia vs acute MD episode) and being on a given medication class.

Future Directions

As our findings indicate that the choice of the pretext task has a significant impact on target task performance, further efforts should be put into pretext task design. Indeed, although MP is a general-purpose strategy inspired by the great success of BERT [38] in NLP, the literature on SSL [40] suggests that domain knowledge may help tailor the pretext task to the specific use case. A promising approach we did not explore is contrastive learning [71], which, indeed, relies on domain knowledge of how augmented views of the input are created, especially since most experience today is in computer vision and NLP, while physiological multivariate time series are relatively unexplored.

Conclusion

This work shows that SSL is a promising paradigm for mitigating the annotation bottleneck, 1 of the major barriers to the development of artificial intelligence–powered clinical decision support systems using personal sensing to help monitor MDs, thus enabling early intervention. The collection and preprocessing of open access unlabeled data sets that we curated (E4SelfLearning) can foster future research into SSL, therefore advancing the translation of personal sensing into clinical practice.

Acknowledgments

We acknowledge the contribution of all the participants of the study.

FC and BML were supported by the United Kingdom Research and Innovation (UKRI; grant EP/S02431X/1), UKRI Centre for Doctoral Training in Biomedical AI at the University of Edinburgh, School of Informatics. For the purpose of open access, the author applied a Creative Commons Attribution (CC BY) license to any author-accepted manuscript version arising from this submission.

GA was supported by a Rio Hortega 2021 grant (CM21/00017) and an M-AES mobility fellowship (MV22/00058) from the Spanish Ministry of Health financed by the Instituto de Salud Carlos III (ISCIII) and cofinanced by the Fondo Social Europeo Plus (FSE+).

IG thanks the support of the Spanish Ministry of Science and Innovation (MCIN; PI23/00822) integrated into the Plan Nacional de I+D+I and cofinanced by the ISCIII-Subdirección General de Evaluación y confinanciado por la Unión Europea (FEDER, FSE, Next Generation EU/Plan de Recu- peración Transformación y Resiliencia PRTR); the Instituto de Salud Carlos III; the CIBER of Mental Health (CIBERSAM); the Secretaria d'Universitats i Recerca del Departament d'Economia i Coneixement (2021 SGR 01358), CERCA Programme/Generalitat de Catalunya; and the Fundació Clínic per la Recerca Biomèdica (Pons Bartran 2022-FRCB PB1 2022).

AHY's independent research was funded by the National Institute for Health and Care Research (NIHR) Maudsley Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. The views expressed are those of the author and not necessarily those of the NIHR or the Department of Health and Social Care. For the purposes of open access, the author applied a Creative Commons Attribution (CC BY) license to any accepted author manuscript version arising from this submission.

DHM was supported by a Juan Rodés grant (JR18/00021) by the ISCIII. AV was supported by the UNREAL project (EP/Y023838/1) selected by the European Research Council and funded by the UKRI Engineering and Physical Sciences Research Council.

Data Availability

The E4SelfLearning collection is available at Reference [72], and the codebase is available at Reference [73]. Data in deidentified form from the Identifying Digital Biomarkers of Illness Activity in Bipolar Disorder/Identifying Digital Biomarkers of Illness Activity and Treatment Response in Bipolar Disorder study may be made available from the corresponding author upon reasonable request.

Authors' Contributions

FC conceived of the study, proposed the methodology, developed the software codebase for the analyses, prepared the manuscript, and curated data collection. BML contributed to codebase development and manuscript writing. GA, CVP, AM, IP, MV, IGF, AB, and MG collected the data for the TIMEBASE/INTREPIBD (Identifying Digital Biomarkers of Illness Activity in Bipolar Disorder/Identifying Digital Biomarkers of Illness Activity and Treatment Response in Bipolar Disorder) study. EV, AHY, SL, and HW critically reviewed the manuscript and provided feedback on the clinical side. DHM is the coordinator of the TIMEBASE/INTREPIBD study and critically reviewed the manuscript. AV supervised this study and contributed to the study design, methodology development, and manuscript writing.

Conflicts of Interest

GA has received continuing medical education–related honoraria, or consulting fees from the Janssen-Cilag, Lundbeck, Lundbeck/Otsuka, and Angelini, with no financial or other relationship relevant to the subject of this paper. IG has received grants and served as a consultant, advisor, or CME speaker for the following identities: ADAMED, Angelini, Casen Recordati, Esteve, Ferrer, Gedeon Richter, Janssen-Cilag, Lundbeck, Lundbeck-Otsuka, Luye, SEI Healthcare, and Viatris outside the submitted work. She also receives royalties from Oxford University Press, Elsevier, and the Editorial Médica Panamericana. All authors report no financial or other relationship relevant to the subject of this paper.

Multimedia Appendix 1
Supplementary material.
[DOC File , 138 KB - mhealth_v12i1e55094_app1.doc]

References

1. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition. Washington, DC: American Psychiatric Association; 2013.
2. Santomauro DF, Mantilla Herrera AM, Shadid J, Zheng P, Ashbaugh C, Pigott DM, et al. Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. Lancet 2021 Nov;398(10312):1700-1712. [doi: 10.1016/s0140-6736(21)02143-7]

3. Greenberg PE, Fournier A, Sisitsky T, Simes M, Berman R, Koenigsberg SH, et al. The economic burden of adults with major depressive disorder in the United States (2010 and 2018). Pharmacoeconomics 2021 Jun 05;39(6):653-665 [FREE Full text] [doi: 10.1007/s40273-021-01019-4] [Medline: 33950419]
4. Brådvik L. Suicide risk and mental disorders. Int J Environ Res Public Health 2018 Sep 17;15(9):2028 [FREE Full text] [doi: 10.3390/ijerph15092028] [Medline: 30227658]
5. Joyce K, Thompson A, Marwaha S. Is treatment for bipolar disorder more effective earlier in illness course? A comprehensive literature review. Int J Bipolar Disord 2016 Dec 9;4(1):19 [FREE Full text] [doi: 10.1186/s40345-016-0060-6] [Medline: 27613276]
6. Buchman-Wildbaum T, Váradi E, Schmelowszky Á, Griffiths MD, Demetrovics Z, Urbán R. The paradoxical role of insight in mental illness: the experience of stigma and shame in schizophrenia, mood disorders, and anxiety disorders. Arch Psychiatr Nurs 2020 Dec;34(6):449-457 [FREE Full text] [doi: 10.1016/j.apnu.2020.07.009] [Medline: 33280665]
7. Rimmer A. Mental health: staff shortages are causing distressingly long waits for treatment, college warns. BMJ 2021 Oct 07;375:n2439. [doi: 10.1136/bmj.n2439] [Medline: 34620691]
8. Satiani A, Niedermier J, Satiani B, Svendsen DP. Projected workforce of psychiatrists in the United States: a population analysis. Psychiatr Serv 2018 Jun 01;69(6):710-713. [doi: 10.1176/appi.ps.201700344] [Medline: 29540118]
9. Mohr DC, Zhang M, Schueller SM. Personal sensing: understanding mental health using ubiquitous sensors and machine learning. Annu Rev Clin Psychol 2017 May 08;13(1):23-47 [FREE Full text] [doi: 10.1146/annurev-clinpsy-032816-044949] [Medline: 28375728]
10. Faurholt-Jepsen M, Brage S, Kessing LV, Munkholm K. State-related differences in heart rate variability in bipolar disorder. J Psychiatr Res 2017 Jan;84:169-173 [FREE Full text] [doi: 10.1016/j.jpsychires.2016.10.005] [Medline: 27743529]
11. Sarchiapone M, Gramaglia C, Iosue M, Carli V, Mandelli L, Serretti A, et al. The association between electrodermal activity (EDA), depression and suicidal behaviour: a systematic review and narrative synthesis. BMC Psychiatry 2018 Jan 25;18(1):22 [FREE Full text] [doi: 10.1186/s12888-017-1551-4] [Medline: 29370787]
12. Tazawa Y, Wada M, Mitsukura Y, Takamiya A, Kitazawa M, Yoshimura M, et al. Actigraphy for evaluation of mood disorders: a systematic review and meta-analysis. J Affect Disord 2019 Jun 15;253:257-269. [doi: 10.1016/j.jad.2019.04.087] [Medline: 31060012]
13. Tazawa Y, Liang K, Yoshimura M, Kitazawa M, Kaise Y, Takamiya A, et al. Evaluating depression with multimodal wristband-type wearable device: screening and assessing patient severity utilizing machine-learning. Heliyon 2020 Feb;6(2):e03274 [FREE Full text] [doi: 10.1016/j.heliyon.2020.e03274] [Medline: 32055728]
14. Jacobson NC, Weingarden H, Wilhelm S. Digital biomarkers of mood disorders and symptom change. NPJ Digit Med 2019 Feb 01;2(1):3 [FREE Full text] [doi: 10.1038/s41746-019-0078-0] [Medline: 31304353]
15. Cote-Allard U, Jakobsen P, Stautland A, Nordgreen T, Fasmer OB, Oedegaard KJ, et al. Long–short ensemble network for bipolar manic-euthymic state recognition based on wrist-worn sensors. IEEE Pervasive Comput 2022 Apr 1;21(2):20-31. [doi: 10.1109/mprv.2022.3155728]
16. Nguyen D, Chan C, Li AA, Phan D, Lan C. Decision support system for the differentiation of schizophrenia and mood disorders using multiple deep learning models on wearable devices data. Health Informatics J 2022 Nov 01;28(4):14604582221137537 [FREE Full text] [doi: 10.1177/14604582221137537] [Medline: 36317536]
17. Ghandeharioun A, Fedor S, Sangermano L, Ionescu D, Alpert J, Dale C. Objective assessment of depressive symptoms with machine learning and wearable sensors data. 2017 Presented at: ACII 2017: Seventh International Conference on Affective Computing and Intelligent Interaction; October 23-26, 2017; San Antonio, TX. [doi: 10.1109/acii.2017.8273620]
18. Pedrelli P, Fedor S, Ghandeharioun A, Howe E, Ionescu DF, Bhathena D, et al. Monitoring changes in depression severity using wearable and mobile sensors. Front Psychiatry 2020 Dec 18;11:584711 [FREE Full text] [doi: 10.3389/fpsyt.2020.584711] [Medline: 33391050]
19. Lee H, Cho C, Lee T, Jeong J, Yeom JW, Kim S, et al. Prediction of impending mood episode recurrence using real-time digital phenotypes in major depression and bipolar disorders in South Korea: a prospective nationwide cohort study. Psychol Med 2023 Sep;53(12):5636-5644. [doi: 10.1017/S0033291722002847] [Medline: 36146953]
20. Li BM, Corponi F, Anmella G, Mas A, Sanabra M, Hidalgo-Mazzei D, et al. Inferring mood disorder symptoms from multivariate time-series sensory data. 2024 Mar 26 Presented at: NeurIPS 2022 Workshop on Learning from Time Series for Health; December 2, 2022; New Orleans, LA URL: https://openreview.net/forum?id=awjU8fCDZjS [doi: 10.1038/s41398-024-02876-1]
21. E4 wristband technical specifications. Empatica. URL: https://support.empatica.com/hc/en-us/articles/202581999-E4-wristband-technical-specifications [accessed 2024-06-24]
22. Ronca V, Martinez-Levy AC, Vozzi A, Giorgi A, Aricò P, Capotorto R, et al. Wearable technologies for electrodermal and cardiac activity measurements: a comparison between Fitbit Sense, Empatica E4 and Shimmer GSR3. Sensors (Basel) 2023 Jun 23;23(13):5847 [FREE Full text] [doi: 10.3390/s23135847] [Medline: 37447697]
23. Reiss A, Indlekofer I, Schmidt P, Van Laerhoven K. Deep PPG: large-scale heart rate estimation with convolutional neural networks. Sensors (Basel) 2019 Jul 12;19(14):3079 [FREE Full text] [doi: 10.3390/s19143079] [Medline: 31336894]
24. Sah RK, McDonell M, Pendry P, Parent S, Ghasemzadeh H, Cleveland M. ADARP: a multi modal dataset for stress and alcohol relapse quantification in real life setting. 2022 Presented at: 2022 IEEE-EMBS International Conference on Wearable

and Implantable Body Sensor Networks (BSN); September 27-30, 2022; Ioannina Greece. [doi: 10.1109/bsn56160.2022.9928495]
25. Schmidt P, Reiss A, Duerichen R, Marberger C, Van LK. Introducing WESAD, a multimodal dataset for wearable stress and affect detection. 2018 Presented at: 20th ACM International Conference on Multimodal Interaction; October 16-20, 2018; Boulder, CO. [doi: 10.1145/3242969.3242985]
26. Iqbal T, Simpkin AJ, Roshan D, Glynn N, Killilea J, Walsh J, et al. Stress monitoring using wearable sensors: a pilot study and stress-predict dataset. Sensors (Basel) 2022 Oct 24;22(21):8135 [FREE Full text] [doi: 10.3390/s22218135] [Medline: 36365837]
27. Svoren H, Thambawita V, Halvorsen P, Jakobsen P, Garcia-Ceja E, Noori F. Toadstool: a dataset for training emotional intelligent machines playing Super Mario Bros. 2020 Presented at: MMSys '20: 11th ACM Multimedia Systems Conference; June 8-11, 2020; Istanbul Turkey. [doi: 10.1145/3339825.3394939]
28. Gashi S, Min C, Montanari A, Santini S, Kawsar F. A multidevice and multimodal dataset for human energy expenditure estimation using wearable devices. Sci Data 2022 Sep 01;9(1):537 [FREE Full text] [doi: 10.1038/s41597-022-01643-5] [Medline: 36050312]
29. Amin MR, Wickramasuriya D, Faghih R. A wearable exam stress dataset for predicting grades using physiological signals. 2022 Presented at: 2022 IEEE Healthcare Innovations and Point of Care Technologies (HI-POCT); March 10-11, 2022; Houston, TX. [doi: 10.1109/hi-poct54491.2022.9744065]
30. Hosseini S, Gottumukkala R, Katragadda S, Bhupatiraju RT, Ashkar Z, Borst CW, et al. A multimodal sensor dataset for continuous stress detection of nurses in a hospital. Sci Data 2022 Jun 01;9(1):255 [FREE Full text] [doi: 10.1038/s41597-022-01361-y] [Medline: 35650267]
31. Gao N, Marschall M, Burry J, Watkins S, Salim FD. Understanding occupants' behaviour, engagement, emotion, and comfort indoors with heterogeneous sensors and wearables. Sci Data 2022 Jun 02;9(1):261 [FREE Full text] [doi: 10.1038/s41597-022-01347-w] [Medline: 35654857]
32. Hinkle LB, Metsis V. Unlabeled Empatica E4 Wristband Data (UE4W) dataset. Zenodo. 2022 Jul 25. URL: https://zenodo.org/records/6898244 [accessed 2024-06-24]
33. Bent B, Cho PJ, Henriquez M, Wittmann A, Thacker C, Feinglos M, et al. Engineering digital biomarkers of interstitial glucose from noninvasive smartwatches. NPJ Digit Med 2021 Jun 02;4(1):89 [FREE Full text] [doi: 10.1038/s41746-021-00465-w] [Medline: 34079049]
34. Shani C, Zarecki J, Shahaf D. The lean data scientist: recent advances toward overcoming the data bottleneck. Commun ACM 2023 Jan 20;66(2):92-102. [doi: 10.1145/3551635]
35. Corponi F, Li BM, Anmella G, Mas A, Pacchiarotti I, Valentí M, et al. Automated mood disorder symptoms monitoring from multivariate time-series sensory data: getting the full picture beyond a single number. Transl Psychiatry 2024 Mar 26;14(1):161 [FREE Full text] [doi: 10.1038/s41398-024-02876-1] [Medline: 38531865]
36. Rani V, Nabi ST, Kumar M, Mittal A, Kumar K. Self-supervised learning: a succinct review. Arch Comput Methods Eng 2023 Jan 20;30(4):2761-2775 [FREE Full text] [doi: 10.1007/s11831-023-09884-2] [Medline: 36713767]
37. Ohri K, Kumar M. Review on self-supervised image recognition using deep neural networks. Knowl-Based Syst 2021 Jul;224:107090. [doi: 10.1016/j.knosys.2021.107090]
38. Devlin DJ, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. arXiv preprint posted online 2018. [doi: 10.48550/arXiv.1810.04805] 2021. [doi: 10.48550/arXiv.1810.04805]
39. Huang S, Pareek A, Jensen M, Lungren MP, Yeung S, Chaudhari AS. Self-supervised learning for medical image classification: a systematic review and implementation guidelines. NPJ Digit Med 2023 Apr 26;6(1):74 [FREE Full text] [doi: 10.1038/s41746-023-00811-0] [Medline: 37100953]
40. Zhang K, Wen Q, Zhang C, Cai R, Jin M, Liu Y, et al. Self-supervised learning for time series analysis: taxonomy, progress, and prospects. arXiv preprint posted online 2023. [doi: 10.48550/arXiv.2306.10125] 2021. [doi: 10.48550/arXiv.2306.10125]
41. Zerveas G, Jayaraman S, Patel D, Bhamidipaty A, Eickhoff C. A transformer-based framework for multivariate time series representation learning. 2021 Presented at: 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining; August 14-18, 2021; Virtual. [doi: 10.1145/3447548.3467401]
42. Wu Y, Daoudi M, Amad A. Transformer-based self-supervised multimodal representation learning for wearable emotion recognition. IEEE Trans Affective Comput 2024 Jan;15(1):157-172. [doi: 10.1109/taffc.2023.3263907]
43. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A. Attention is all you need. 2017 Presented at: NIPS 2017: Advances in Neural Information Processing Systems 30; December 4-9, 2017; Long Beach, CA.
44. Ericsson L, Gouk H, Loy CC, Hospedales TM. Self-supervised representation learning: introduction, advances, and challenges. IEEE Signal Process Mag 2022 May;39(3):42-62. [doi: 10.1109/msp.2021.3134634]
45. Anmella G, Corponi F, Li BM, Mas A, Sanabra M, Pacchiarotti I, et al. Exploring digital biomarkers of illness activity in mood episodes: hypotheses generating and model development study. JMIR Mhealth Uhealth 2023 May 04;11:e45405 [FREE Full text] [doi: 10.2196/45405] [Medline: 36939345]
46. Hamilton M. A rating scale for depression. J Neurol Neurosurg Psychiatry 1960 Feb 01;23(1):56-62 [FREE Full text] [doi: 10.1136/jnnp.23.1.56] [Medline: 14399272]

47. Young RC, Biggs JT, Ziegler VE, Meyer DA. A rating scale for mania: reliability, validity and sensitivity. Br J Psychiatry 1978 Nov 29;133(5):429-435. [doi: 10.1192/bjp.133.5.429] [Medline: 728692]
48. Tohen M, Frank E, Bowden C, Colom F, Ghaemi N, Yatham L. The International Society of Bipolar Disorders (ISBD) task force on the nomenclature of course and outcome in bipolar disorders. J Affect Disord 2010 Apr;122:S15. [doi: 10.1016/j.jad.2010.01.030]
49. Vieluf S, Amengual-Gual M, Zhang B, El Atrache R, Ufongene C, Jackson MC, et al. Twenty-four-hour patterns in electrodermal activity recordings of patients with and without epileptic seizures. Epilepsia 2021 Apr 23;62(4):960-972. [doi: 10.1111/epi.16843] [Medline: 33619751]
50. Nasseri M, Nurse E, Glasstetter M, Böttcher S, Gregg NM, Laks Nandakumar A, et al. Signal quality and patient experience with wearable devices for epilepsy management. Epilepsia 2020 Nov 04;61 Suppl 1(S1):S25-S35. [doi: 10.1111/epi.16527] [Medline: 32497269]
51. Emaptics user manual. ETH Zurich. URL: https://archive.arch.ethz.ch/esum/downloads/manuals/emaptics.pdf [accessed 2024-06-24]
52. van Hees VT, Sabia S, Anderson KN, Denton SJ, Oliver J, Catt M, et al. A novel, open access method to assess sleep duration using a wrist-worn accelerometer. PLoS One 2015 Nov 16;10(11):e0142533 [FREE Full text] [doi: 10.1371/journal.pone.0142533] [Medline: 26569414]
53. Patterson MR, Nunes AAS, Gerstel D, Pilkar R, Guthrie T, Neishabouri A, et al. 40 years of actigraphy in sleep medicine and current state of the art algorithms. NPJ Digit Med 2023 Mar 24;6(1):51 [FREE Full text] [doi: 10.1038/s41746-023-00802-1] [Medline: 36964203]
54. Cui Z, Chen W, Chen Y. Multi-scale convolutional neural networks for time series classification. arXiv preprint posted online 2016. [doi: 10.48550/arXiv.1603.06995] 2021. [doi: 10.48550/arXiv.1603.06995]
55. 55 PM, Zlatintsi A, Filntisis P, Roumeliotis A, Efthymiou N, Maragos P. A comparative study of autoencoder architectures for mental health analysis using wearable sensors data. 2022 Presented at: EUSIPCO 2022: 30th European Signal Processing Conference; August 29-September 2, 2022; Belgrade, Serbia. [doi: 10.23919/eusipco55093.2022.9909697]
56. Föll S, Maritsch M, Spinola F, Mishra V, Barata F, Kowatsch T, et al. FLIRT: a feature generation toolkit for wearable data. Comput Methods Programs Biomed 2021 Nov;212:106461 [FREE Full text] [doi: 10.1016/j.cmpb.2021.106461] [Medline: 34736174]
57. Özdenizci O, Wang Y, Koike-Akino T, ErdoĞmuŞ D. Learning invariant representations from EEG via adversarial inference. IEEE Access 2020;8:27074-27085 [FREE Full text] [doi: 10.1109/access.2020.2971600] [Medline: 33747669]
58. Cheng JY, Goh H, Dogrusoz K, Tuzel O, Azemi E. Subject-aware contrastive learning for biosignals. arXiv preprint posted online 2020. [doi: 10.48550/arXiv.2007.04871] 2021. [doi: 10.48550/arXiv.2007.04871]
59. Sabry F, Eltaras T, Labda W, Alzoubi K, Malluhi Q. Machine learning for healthcare wearable devices: the big picture. J Healthc Eng 2022;2022:4653923. [doi: 10.1155/2022/4653923] [Medline: 35480146]
60. Strzelecki M, Badura P. Machine learning for biomedical application. Appl Sci 2022 Feb 15;12(4):2022. [doi: 10.3390/app12042022]
61. Li L, Jamieson K, DeSalvo G, Rostamizadeh A, Talwalkar A. Hyperband: a novel bandit-based approach to hyperparameter optimization. J Mach Learn Res 2017;18:6765-6816. [doi: 10.1007/978-1-4899-7687-1_100200]
62. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. J Mach Learn Res 2014;15(56):1929-1958.
63. Ismail Fawaz H, Forestier G, Weber J, Idoumghar L, Muller P. Deep learning for time series classification: a review. Data Min Knowl Disc 2019 Mar 02;33(4):917-963. [doi: 10.1007/s10618-019-00619-1]
64. Paszke P, Gross S, Massa F, Lerer A, Bradbury J, Chanan G. PyTorch: an imperative style, high-performance deep learning library. 2019 Presented at: NeurIPS 2019: 33rd Conference on Neural Information Processing Systems; December 8-14, 2019; Vancouver, Canada p. 8024-8035. [doi: 10.7551/mitpress/11474.003.0014]
65. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. J Mach Learn Res 2011;12(85):2825-2830.
66. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. 2016 Presented at: KDD '16: 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 13-17, 2016; San Francisco, CA. [doi: 10.1145/2939672.2939785]
67. Experiment tracking: the system of record for your model training. Weights & Biases. URL: https://wandb.ai/site/experiment-tracking [accessed 2024-06-24]
68. McInnes L, Healy J, Melville J. UMAP: uniform manifold approximation and projection for dimension reduction. arXiv oreprint posted online 2018. [doi: 10.48550/arXiv.1802.03426] 2021. [doi: 10.48550/arXiv.1802.03426]
69. Birk RH, Samuel G. Digital phenotyping for mental health: reviewing the challenges of using data to monitor and predict mental health problems. Curr Psychiatry Rep 2022 Oct 24;24(10):523-528. [doi: 10.1007/s11920-022-01358-9] [Medline: 36001220]
70. El-Nouby A, Izacard G, Touvron H, Laptev I, Jegou H, Grave E. Are large-scale datasets necessary for self-supervised pre-training. arXiv preprint posted online 2021. [doi: 10.48550/arXiv.2112.10740] 2021. [doi: 10.48550/arXiv.2112.10740]

71. Kumar P, Rawat P, Chauhan S. Contrastive self-supervised learning: review, progress, challenges and future research directions. Int J Multimed Info Retr 2022 Aug 05;11(4):461-488. [doi: 10.1007/s13735-022-00245-6]
72. FcmC / E4SelfLearning. Hugging Face. URL: https://huggingface.co/datasets/FcmC/E4SelfLearning [accessed 2024-07-02]
73. april-tools / e4selflearning. GitHub. URL: https://github.com/april-tools/e4selflearning [accessed 2024-07-02]

Abbreviations

ACC: accuracy
AUROC: area under the receiver operating characteristic curve
BD: bipolar disorder
BVP: blood volume pressure
CCE: categorical cross-entropy
CE: channel embedding
CML: classical machine learning
DSM-5: Diagnostic and Statistical Manual, Fifth Edition
E4mer: E4-tailored transformer
EDA: electrodermal activity
ENET: elastic net logistic regression
FT: fine-tuning
HDRS: Hamilton Depression Rating Scale-17
HR: heart rate
IBI: interbeat interval
KNN: K-nearest neighbor
LME: linear mixed effects
LR: linear readout
MD: mood disorder
MDD: major depressive disorder
ML: machine learning
MLP: multilayer perceptron
MP: masked prediction
NLP: natural language processing
PCC: Pearson correlation coefficient
RM: representation module
RMSE: root mean square error
SL: supervised learning
SSL: self-supervised learning
SVM: support vector machine
TEMP: skin temperature
TIMEBASE/INTREPIBD: Identifying Digital Biomarkers of Illness Activity in Bipolar Disorder/Identifying Digital Biomarkers of Illness Activity and Treatment Response in Bipolar Disorder
TP: transformation prediction
XGBoost: extreme gradient boosting
YMRS: Young Mania Rating Scale

Edited by L Buis; submitted 02.12.23; peer-reviewed by J Zulueta, B Montezano, W Speier; comments to author 29.02.24; revised version received 14.04.24; accepted 24.05.24; published 17.07.24.

Please cite as:
Corponi F, Li BM, Anmella G, Valenzuela-Pascual C, Mas A, Pacchiarotti I, Valentí M, Grande I, Benabarre A, Garriga M, Vieta E, Young AH, Lawrie SM, Whalley HC, Hidalgo-Mazzei D, Vergari A
Wearable Data From Subjects Playing Super Mario, Taking University Exams, or Performing Physical Exercise Help Detect Acute Mood Disorder Episodes via Self-Supervised Learning: Prospective, Exploratory, Observational Study
JMIR Mhealth Uhealth 2024;12:e55094
URL: https://mhealth.jmir.org/2024/1/e55094
doi:10.2196/55094
PMID:39018100

Original Paper

Using Wearables to Study Biopsychosocial Dynamics in Couples Who Cope With a Chronic Health Condition: Ambulatory Assessment Study

Theresa Pauly[1], PhD; Janina Lüscher[2,3], PhD; Lea Olivia Wilhelm[4,5], MSc; Melanie Alexandra Amrein[6], PhD; George Boateng[7], PhD; Tobias Kowatsch[7,8,9], PhD; Elgar Fleisch[7,10], PhD; Guy Bodenmann[11], PhD; Urte Scholz[11,12], PhD

[1]Department of Gerontology, Simon Fraser University, Vancouver, BC, Canada
[2]Swiss Paraplegic Research, Nottwil, Switzerland
[3]Faculty of Health Sciences and Medicine, University of Lucerne, Lucerne, Switzerland
[4]Medical School Berlin, Berlin, Germany
[5]Department of Education and Psychology, Freie Universität Berlin, Berlin, Germany
[6]Department of Health Sciences, Helsana Group, Zurich, Switzerland
[7]Centre for Digital Health Interventions, Department of Management, Technology, and Economics, ETH Zurich, Zurich, Switzerland
[8]Institute for Implementation Science in Health Care, University of Zurich, Zurich, Switzerland
[9]School of Medicine, University of St. Gallen, St.Gallen, Switzerland
[10]Institute of Technology Management, University of St.Gallen, St. Gallen, Switzerland
[11]Department of Psychology, University of Zurich, Zurich, Switzerland
[12]University Research Priority Area Dynamics of Healthy Aging, University of Zurich, Zurich, Switzerland

Corresponding Author:
Theresa Pauly, PhD
Department of Gerontology
Simon Fraser University
515 West Hastings Street
Vancouver, BC, V6B 5K3
Canada
Phone: 1 778 782 7834
Email: theresa_pauly@sfu.ca

Abstract

Background: Technology has become an integral part of our everyday life, and its use to manage and study health is no exception. Romantic partners play a critical role in managing chronic health conditions as they tend to be a primary source of support.

Objective: This study tests the feasibility of using commercial wearables to monitor couples' unique way of communicating and supporting each other and documents the physiological correlates of interpersonal dynamics (ie, heart rate linkage).

Methods: We analyzed 617 audio recordings of 5-minute duration (384 with concurrent heart rate data) and 527 brief self-reports collected from 11 couples in which 1 partner had type II diabetes during the course of their typical daily lives. Audio data were coded by trained raters for social support. The extent to which heart rate fluctuations were linked among couples was quantified using cross-correlations. Random-intercept multilevel models explored whether cross-correlations might differ by social contexts and exchanges.

Results: Sixty percent of audio recordings captured speech between partners and partners reported personal contact with each other in 75% of self-reports. Based on the coding, social support was found in 6% of recordings, whereas at least 1 partner self-reported social support about half the time (53%). Couples, on average, showed small to moderate interconnections in their heart rate fluctuations (r=0.04-0.22). Couples also varied in the extent to which there was lagged linkage, that is, meaning that changes in one partner's heart rate tended to precede changes in the other partner's heart rate. Exploratory analyses showed that heart rate linkage was stronger (1) in rater-coded partner conversations (vs moments of no rater-coded partner conversations: r_{diff}=0.13; P=.03), (2) when partners self-reported interpersonal contact (vs moments of no self-reported interpersonal contact: r_{diff}=0.20; P<.001), and (3) when partners self-reported social support exchanges (vs moments of no self-reported social support exchange: r_{diff}=0.15; P=.004).

Conclusions: Our study provides initial evidence for the utility of using wearables to collect biopsychosocial data in couples managing a chronic health condition in daily life. Specifically, heart rate linkage might play a role in fostering chronic disease management as a couple. Insights from collecting such data could inform future technology interventions to promote healthy lifestyle engagement and adaptive chronic disease management.

International Registered Report Identifier (IRRID): RR2-10.2196/13685

(JMIR Mhealth Uhealth 2024;12:e49576) doi:10.2196/49576

KEYWORDS

couples; wearables; type II diabetes; heart rate; biopsychosocial dynamics; physiological linkage; mobile health; technology; social support; chronic disease; usability; utility; mHealth

Introduction

Coping with a chronic health condition refers to the various psychological, emotional, and behavioral strategies individuals use to manage the challenges, stressors, and lifestyle adjustments associated with living with a long-term health condition [1]. Challenges associated with the maintenance of health-promoting behaviors such as being physically active, following a medication regime, and monitoring one's current health status on top of direct effects of the chronic health condition on cognitive and physical resources (eg, fatigue) can lead to lapses in treatment adherence, which are common [2]. Poor treatment adherence, in turn, results in higher morbidity and mortality rates, as well as increased expenses related to outpatient care and hospitalization for managing diabetes-related complications [3]. Models of chronic disease management acknowledge the important role that close others generally, and romantic partners specifically, play in coping [1,4-6]. Social support refers to how social relationships provide resources and assistance to manage stressors and challenges [7]. These forms of assistance manifest in instrumental ways (ie, practical assistance with a problem or task) and emotional ways (ie, comfort, encouragement, reassurance, and listening empathetically [8]). Social support can facilitate coping by promoting healthy lifestyle choices and helping manage disease-related demands in daily life and has been associated with better disease adjustment and higher quality of life among patients [9,10]. Moreover, in a dyadic context, providing social support has also been conceptualized as a specific form of coping [11]. In laboratory stress paradigms, social support (eg, in the form of physical touch) buffers physiological stress reactivity and speeds up recovery, for example, as assessed by heart rate and salivary cortisol levels [12,13]. This effect is particularly pronounced when the social support comes from a close person, such as a romantic partner [12].

The couple as a unit of study is particularly interesting for researchers studying chronic health conditions because romantic partners tend to be a major source of support and partners' health is closely linked [11,14,15]. For example, individuals are more than 2 times more likely to have diabetes themselves when they are in a relationship with someone who has diabetes [16,17]. This is not surprising as lifestyle risk factors for diabetes such as eating patterns, physical activity, and other health behaviors are often shared in couples [18]. Newer conceptual models of chronic health conditions and coping, including the dyadic regulation connectivity model [5], see couples as a dynamic system, in which dyadic regulation of chronic health conditions occurs in a flexible, dynamic, and complex way adjusting to changing internal contexts (eg, emotional states) as well as external contexts (eg, work demands). Specifically, dyadic regulation involves different network hubs (illness representations hub, coping behaviors hub, and outcomes hub), interconnected through feedback loops and constantly interacting at the person and dyad level. This dynamic process is described as "a 'living' mechanism which is continuously constructed and reconstructed in the mind and the behaviors of the individuals and the couples" [5]. Consequently, the field has called for more advanced methods to be able to study time- and context-sensitive processes shaping health behavior and disease management [19].

New technology including smartphones and smartwatches allows to collect observational and physiological data to observe dynamic phenomena as they unfold in daily life [20]. For example, researchers have equipped couples with audio recorders to capture sound snippets as they go about their usual routines, finding that use of certain speech (eg, positive word use, deep conversation) was associated with better disease adjustment to breast cancer in women [21,22]. Furthermore, health behaviors such as physical activity and medication adherence can be objectively tracked and this information can be used to develop targeted interventions in real time [23]. Researchers have also used sensor data to measure individuals' physiological functioning, for instance, electrodermal activity (EDA), skin temperature, respiratory rate, blood pressure, and heart rate [24,25]. Measures of the autonomic nervous system such as heart rate hold particular significance because they tend to change rapidly in response to shifting contexts (eg, emotions or behavior [26]). This responsiveness, in combination with the relative ease of data collection through wearable devices, makes it an excellent indicator of physiological linkage in couples.

Studies analyzing physiological time series data in couples have found that romantic partners are interconnected in daily fluctuations of heart rate, a phenomenon that has been called "physiological synchrony" or "physiological linkage" [27-29]. Physiological linkage has been associated with central interpersonal outcomes such as trust, empathy, and effective cooperation [28,30-32], and could thus be important for social processes such as supportive interactions in managing a chronic health condition. For example, synchrony in skin conductance levels predicted higher cooperative success in 76 dyads (college students) playing a Prisoner's Dilemma game [30]. In addition, heart rate linkage was increased in 110 college student dyads

when randomized to playing a trust-related game versus a control condition [32]. A recent meta-analysis of 60 published and unpublished experiments also demonstrated that synchrony exhibits a medium-sized effect on prosocial attitudes and behaviors [33]. Thus, physiological linkage might be relevant for coregulation between partners and enable adaptive couple functioning [34]. Yet, no prior research has examined heart rate linkage in the context of everyday social contexts and social support exchanges in individuals with chronic health conditions.

Previous studies on interpersonal dynamics and heart rate linkage in couples have mostly used electrocardiogram data from medical-grade devices such as the BIOPAC collected in laboratory settings [34,35]. However, commercial devices such as smartwatches could be a cost-effective way to collect biopsychosocial data on couples managing chronic health conditions in daily life. Therefore, the objective of this study is to outline the methods we used to gather two types of data: (1) information concerning psychosocial processes, encompassing social contexts and interactions, using observational techniques (audio recordings), and self-report measures; and (2) data on biological processes, specifically heart rate. In doing so, we summarize the feasibility of collecting such data and present descriptive and exploratory findings. Specifically, we examine whether heart rate linkage in couples is higher in moments of personal contact and support.

Methods

Participants and Procedure

We used data of 11 couples who took part in the DYadic MANagement of Diabetes (DYMAND) study (study protocol [36]; description of monitoring system [37]). This study was funded by the Swiss National Science Foundation (CR12I1_166348). Couples were recruited through newspaper advertisement, flyer distribution to diabetes specialists and pharmacies, diabetes forums, and in diabetes departments of hospitals in the German-speaking part of Switzerland from 2019 to 2021. The original recruitment target was 180 couples. However, due to the COVID-19 pandemic and associated restrictions regarding in-person research with vulnerable populations recruitment had to be paused for a considerable duration. This resulted in the current sample size. Eligibility criteria comprised one partner having a medical diagnosis of type II diabetes with prescribed oral drugs and the other partner (without such diagnosis) being willing to participate as well. Participants were excluded if they required insulin injections, inpatient treatment, were working in shiftwork, or had insufficient command of the German language. We collected data from 13 couples, who were living in metropolitan and rural regions in the German-speaking part of Switzerland. Two couples dropped out during the monitoring phase because of time constraints, resulting in a final sample of 11 couples (see Table 1 for descriptives). Participants with type II diabetes were aged 52-80 years (mean 68.9, SD 7.7 years), mostly male (10 males, 1 female), and a majority did not have a higher secondary school degree (8 with lower secondary education, 3 with higher secondary degree). Less than half of the participants with type II diabetes measured their blood sugar levels daily (4/11, 36%) and the average long-term blood sugar concentration was 6.8% (hemoglobin A_{1c} [HbA_{1c}], SD 0.5%; normal HbA_{1c} <5.7% [38]). Most participants reported a positive influence of their partnership on their diabetes management (10/11, 91%). Partners were aged 47-80 (mean 66.7, SD 9.2) years, mostly female (1 male, 10 females), and 1 participant had a higher secondary school degree. Less than half of the participants with type II diabetes (3/11, 27%) and partners (4/11, 36%) were employed, whereas the majority was retired. Couples' average relationship duration was 31.5 (SD 14.6) years. Most couples had a monthly household income above the poverty line for a household of 2 people in Switzerland (>4000 Swiss Francs; 80% [39]).

Table 1. Sample descriptives (N=11 couples).

Variable	Person with type II diabetes	Partner
Age (years), mean (SD)	68.91 (7.71)	66.68 (9.19)
Sex	91% male, 9% female	9% male, 91% female
Education	73% with lower secondary degree	91% with lower secondary degree
Employment status	27% employed	36% employed
Relationship duration (years), mean (SD)	31.58 (14.95)	31.45 (15.01)
Household income (Swiss Francs[a])	2001-4000 (n=1); 4001-6000 (n=2); 6001-8000 (n=4); 8001-10,000 (n=1); and >10,000 (n=3)	4001-6000 (n=3); 6001-8000 (n=2); 8001-10,000 (n=1); >10,000 (n=3); and missing (n=2)
Children, n	No children (n=2), 1 child (n=3), 2 children (n=3), 3 children (n=1), and 4 children (n=2)	No children (n=3), 1 child (n=1), 2 children (n=5), and 3 children (n=2)

[a]1 Swiss Franc=US $1.13.

After a baseline session in which the participants were trained to use the equipment and completed a questionnaire battery, the participants entered a 7-day monitoring phase. An assessment was triggered every hour during a specified time defined by the couples in the morning (eg, 6 AM to 9 AM) and evening (eg, 5 PM to 9 PM) during the week and all day (waking hours, eg, 8 AM to 9 PM) on the weekend. This decision was made to alleviate participant burden by abstaining from data collection during periods when partners are typically at work, thus ensuring that we capture meaningful partner interactions effectively.

Assessments were elicited when partners were close to each other (when the smartwatch Bluetooth system detected a signal strength of the partner's watch greater than –80 dB corresponding to approximately 5 m), and speech was detected. When the 2 conditions of physical closeness and speech were not detected, an assessment was triggered at the end of the hour to ensure sufficient data coverage. The assessment included a 5-minute audio and heart rate recording, followed by a brief self-report questionnaire on the smartphone which asked about partner interactions and health behaviors. Each evening, partners answered a longer survey about their own and their partner's behaviors (data not used in this study). Participants reported that the study app was easy to use (persons with type II diabetes: mean 5.9, SD 0.7; partners: mean 6.0, SD 0.7; 1=completely disagree to 7=fully agree). See the study by Boateng et al [37] for a detailed description of the development and deployment of the smartwatch- and smartphone-monitoring system. After the 7-day monitoring phase, partners returned to the laboratory for an exit session during which feedback on the study was collected, and partners were videotaped while having a 10-minute conversation about their diabetes management. In this exit session, the participants were also allowed to review their audio recordings and delete files before the research team accessed them. None of the couples chose to remove any of the audio recordings. Couples received 100 Swiss Francs as reimbursement for taking part in the study.

Ethical Considerations

Ethics approval was granted by the cantonal ethics committee of the Canton of Zurich, Switzerland (Req-2017_00430), and informed consent was obtained from all participants.

Measures

Audio Recordings

A smartwatch (Polar M600) recorded daily life audio snippets for 5 minutes each. Four trained research assistants used the Social Environment Coding of Sound Inventory [40] to code couples' location, activity, conversation partner, and conversation type. Conversation types included (1) practical: pragmatic conversations focusing on practical daily matters such as making plans and discussing meals; (2) small talk: an interaction without instrumental purpose, involving superficial exchanges of information that have no significant impact or consequence on participants' lives; (3) deep or substantive: conversations with the purpose to exchange thoughts, information, values, and ideas on nonemotional topics such as current events; (4) disclosure: conversations that involve sharing personal feelings or emotions, which may include discussing topics such as the relationship, hopes and dreams, or other deeply meaningful experiences, surpassing the threshold of triviality. Furthermore, we coded for instrumental and emotional social support, following procedures outlined in the study by Wang and Repetti [41]. Instrumental support entailed help with practical problems and tasks, such as assistance with chores or the provision of information to help handle a task-oriented problem (eg, figuring out the fastest driving route to the doctor's office). Emotional support included provisions of comfort, encouragement, advice, or guidance of an emotional nature, for example, listening empathetically to a spouse's frustrations about work. Fifteen percent of recordings were coded by all 4 research assistants to calculate interrater reliability (intraclass correlation [ICC] of 2.1, calculated using SPSS; IBM Corp), which was satisfactory (>0.60 [42]; location: ICC=0.77, activity: ICC=0.67, conversation partners: ICC=0.60-0.74, and support: ICC=0.63), except for conversation types (ICC=0.39 for practical, 0.50 for small talk, 0.24 for deep or substantive, and 0.28 for disclosure).

Heart Rate

The smartwatch also collected heart rate information during the 5-minute audio recordings. The watch did not collect equally spaced recordings (mean spacing=2.7, SD 2.8 seconds). Thus, to match partner heart rate data, we calculated the mean of all heart rate values captured in 5 seconds frames (average n=2.1, SD 1.2, range 1-6 values; if aggregated to 10-second frames, pattern of findings remains the same). This resulted in a total of 55,175 heart rate values. We then aligned the heart rate data of couples by time stamp and kept only segments in which data from both partners were available (n=37,834, 69% heart rate pairings). Missing values in paired data are due to devices eliciting recordings with a small lag in partners or additional recordings that were triggered in just 1 partner. Furthermore, 5-minute units with less than 10 paired heart rate values of partners were deleted (n=6, <1%).

Self-Reported Social Contexts and Exchanges

Subsequently to the audio and heart rate recording, participants reported whether they had any personal contact with their partner in the last 5 minutes. Furthermore, participants with type II diabetes indicated whether they received emotional or practical support from their partner in the last 5 minutes with yes or no: "My partner has supported me emotionally in the last five minutes"/"My partner has supported me practically in the last five minutes." Partners reported emotional and practical support provision with yes or no: "I have been emotionally supportive of my partner in the last five minutes"/"I have been practically supporting my partner in the last five minutes." A social support exchange was coded when provided or received support was indicated by at least 1 partner.

Statistical Analysis

Descriptives were calculated using frequencies, means, and SDs. Heart rate linkage in couples was quantified using cross-correlations [43], calculated with the ccf function in R (Stats Package; R Core Team [44]). The linkage between partners can be in-phase (positive correlation), meaning that fluctuations in partners' heart rates are in the same direction, or antiphase (negative correlation), meaning that couples show opposite patterns of heart rate fluctuations over time. The cross-correlation models the relationship between both partners' heart rate time series data for a given recording. An advantage of the cross-correlation method over traditional correlation techniques is that it calculates the correlation between dyadic time series of heart rate data at a given maximum lag. Specifically, it quantifies the dependence of the heart rate time series data of the person with type II diabetes on its past observations, the partner's concurrent heart rate, and the past or future observations of the partner's heart rate. In this way,

the ccf identifies the maximum cross-correlation for each recording, which could occur at a positive, no, or negative lag. We considered only a lag of ±1 time fragment because we had averaged heart rate recordings across 5-second windows, and cross-correlations can be biased if longer lags are used (see suggestions outlined in the study by Behrens et al [45]). In the presented results, a lag of –1 indicates that changes in the heart rate of the person with type II diabetes tended to be a precursor to changes in the partner's heart rate. A lag of 0 indicates that changes in heart rate tended to co-occur. A lag of +1 indicates that changes in the partner's heart rate tended to be a precursor to changes in the heart rate of the person with type II diabetes. Finally, we explored whether cross-correlations might differ by social contexts and exchanges using simple random-intercept multilevel models (observations nested within individuals nested within couples, controlling for person-level averages; R package lme4 [46]). For these models, cross-correlations were Fisher-Z transformed to normalize the distribution.

Results

Descriptives

We collected 992 audio files from 11 couples, of which 375 captured the same content (ie, they co-occurred in partners; 37.8%). After removing the audio file of 1 partner in these cases, we had a total of 617 audio recordings for analysis (mean 56.1, SD 14.5 per couple; range: 35-83 files). Three-quarters of the audio files contained speech (74%), and participants were mostly at home (80%; 8% in public, 5% in transit, 4% outdoors, and 3% other or unknown). Participants were most frequently watching TV or listening to radio (195/590, 33%), socializing (189/590, 32%), or doing housework (69/590, 12%), and, to a lesser extent, were eating or drinking (32/590, 5%) or physically active (12/590, 2%). In 16% (85/590) of recordings, activities were unknown. Recorded conversations (423/617) mainly included partners talking (372/423, 88%), whereas conversations with friends (77/423, 18%), strangers (17/423, 4%), children (10/423, 2%), and other family members (9/423, 2%) or self-talk (24/423, 6%) occurred less frequently. Concerning conversation types, most conversations were substantive (194/423, 46%), with fewer conversations of practical content (139/423, 33%) or small talk (76/423, 18%; 3% other). Social support was coded in 6% of recorded conversations (27/423), with the most frequent support providers being the partner (16/27, 60%) or a friend (9/27, 33%). The nature of the supportive interaction was 63% (17/27) instrumental and 37% (10/27) emotional (single choice only).

We collected 606 brief self-report questionnaires from partners, of which 79 captured the same situation (completed simultaneously by the partners). Thus, we analyzed 527 self-report questionnaires. In three-quarters of the questionnaires, partners reported having had personal contact with each other (394/527, 75%). Support was self-reported in more than half of the instances (277/527, 53%), with emotional support occurring 84% (233/277) of the time and practical support occurring 82% (227/277) of the time (multiple-choice possible).

Heart Rate Linkage

Dyadic heart rate data were available for 384 of the 617 audio recordings (ie, the 5-minute segments contained heart rate data from both partners). Concerning heart rate linkage, on average, all couples showed small to moderate cross-correlations in their heart rate (range: r=0.04-0.22; see Table 2). However, there was considerable variation in the observed cross-correlation between recordings (SDs 0.26-0.38). This means that couples showed a large variety of cross-correlations between recordings: Each couple sometimes showed a positive cross-correlation (ie, heart rate increases in one partner were linked with heart rate increases in the other partner; see Figure 1 panels A and B for examples), sometimes a cross-correlation close to 0 (ie, changes in heart rate throughout the recording were not systematically linked), and sometimes a negative cross-correlation (eg, heart rate increases in one partner were linked with heart rate decreases in the other partner; see Figure 1 panel C for an example). Histograms of couples' cross-correlations can be found in Figure 2, panel A.

Table 2 also denotes the average of the lags corresponding to each cross-correlation separately by couple. As outlined in the Methods section, the maximum cross-correlation could occur at a lag of +1, 0, or –1 for each recording. As seen in Figure 2 panel B, couples showed large heterogeneity in the occurrence of lags. In some couples, the extent to which heart rate changes in either the person with type II diabetes or the partner tended to precede heart rate changes in the other dyad member was balanced (couples 2, 4, 8, 10, and 11). However, in couples 1 and 7, a large share of recordings featured a negative lag (43%), suggesting that heart rate linkage in the dyad tended to be driven by the partner's heart rate. There was also variation in the extent to which recordings occurred that featured a lag of 0 (indicating same-time linkage with no partner driving the changes)—from every fifth (17%) recording in couple 2 to almost half of recordings in couple 3 (46%).

Table 2. Overview of cross-correlations and their lags by couple.

Couple ID	Number of observations	Cross-correlation, mean (SD)	Lag[a], mean (SD)	Percent negative lag	Percent no lag	Percent positive lag
1	14	0.184 (0.27)	–0.214 (0.80)	43	36	21
2	18	0.086 (0.38)	0.056 (0.94)	39	17	44
3	50	0.133 (0.30)	–0.220 (0.71)	38	46	16
4	44	0.084 (0.29)	0.000 (0.86)	36	27	36
5	22	0.172 (0.32)	0.136 (0.83)	27	32	41
6	53	0.045 (0.35)	0.321 (0.73)	15	38	47
7	47	0.036 (0.37)	–0.106 (0.87)	43	26	32
8	31	0.215 (0.37)	0.000 (0.82)	32	35	32
9	27	0.146 (0.26)	0.185 (0.83)	26	30	44
10	45	0.198 (0.34)	0.067 (0.75)	24	44	31
11	34	0.102 (0.34)	0.059 (0.89)	35	24	41

[a]A negative lag indicates that changes in the heart rate of the person with type II diabetes tend to be a precursor to changes in the partner's heart rate. A lag of 0 indicates that changes in heart rate tended to co-occur. A positive lag indicates that changes in the partner's heart rate tend to be a precursor to changes in the heart rate of the person with type II diabetes.

Figure 1. Three examples of collected dyadic heart rate data across a 5-minute recording. Panel A shows a high positive cross-correlation at a lag of 0 (r=0.91), indicating that partners show linked concurrent increases and decreases in heart rate in the same direction. Panel B shows a high positive cross-correlation at a lag of 1 (r=0.79), indicating that partners show linked increases and decreases in heart rate with heart rate changes in the partner preceding heart rate changes in the individual with type II diabetes by 5 seconds. Panel C shows a high negative cross-correlation at a lag of 0 (r=–0.79), indicating that partners show concurrent increases and decreases in heart rate in opposite directions (eg, one partner shows an increasing heart rate while the other shows decreasing heart rate). T2D: type II diabetes.

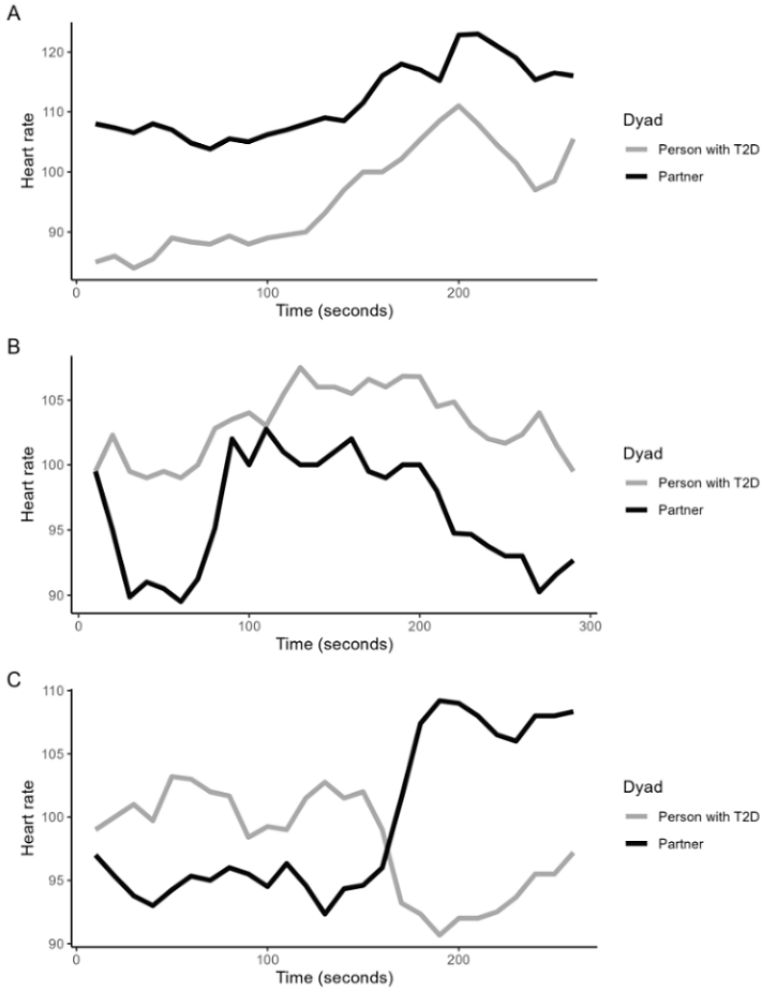
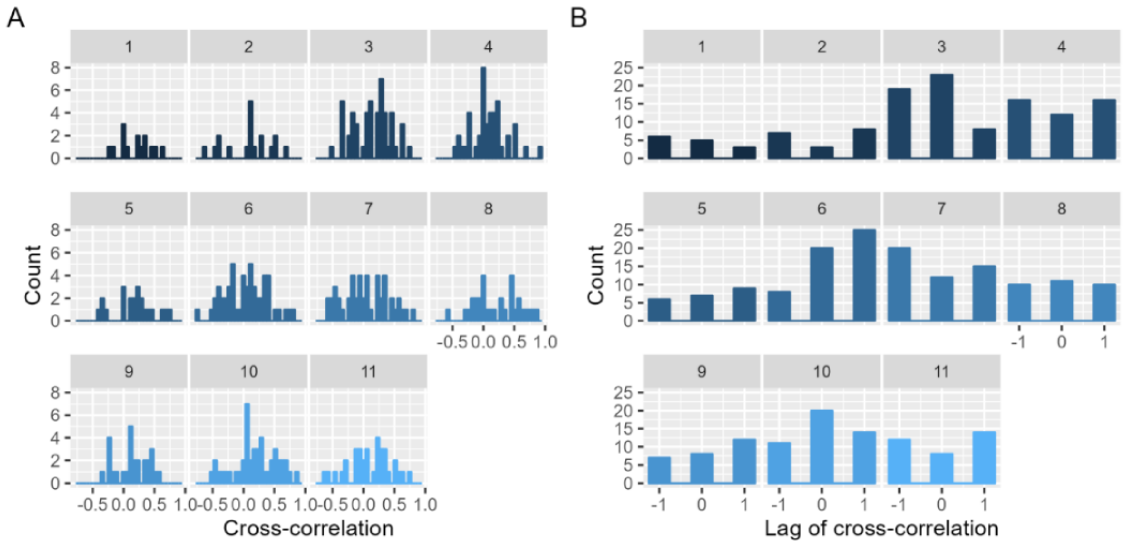


Figure 2. Distributions of cross-correlations (A) and the respective lags of cross-correlations (B) per couple. A lag of –1 indicates that changes in the heart rate of the person with type II diabetes tend to be a precursor to changes in the partner's heart rate. A lag of 0 indicates that changes in heart rate tended to co-occur. A lag of 1 indicates that changes in the partner's heart rate tend to be a precursor to changes in the heart rate of the person with type II diabetes.



Heart Rate Linkage and Social Contexts and Exchanges

Finally, we explored whether heart rate linkage between partners might differ by social contexts and exchanges. Using the rater codings, we found that heart rate linkage was higher in moments when partners were talking with each other (mean r=0.17) than in moments when they were not engaged in conversation (mean r=0.04; b=0.15, SE=0.07; P=.03). Self-report data replicated these findings, showing that heart rate linkage was higher in moments when participants reported having had personal contact with each other (mean r=0.20) than in moments when they reported no such interpersonal contact (mean r<0.01; b=0.22, SE=0.06; P<.001). Furthermore, heart rate linkage was higher when participants self-reported giving or receiving support (mean r=0.22) than in moments of no support (mean r=0.07; b=0.18, SE=0.06; P=.004). We did not examine associations between rater-coded social support and heart rate linkage due to its very low frequency. The findings are displayed in Multimedia Appendix 1.

Discussion

Principal Findings

Most chronic health conditions require major modification of one's lifestyle, including adhering to sometimes complex medication regimes and engaging in health-promoting behaviors. Using new technology to monitor everyday biopsychosocial dynamics in couples coping with chronic health conditions could provide important insights for future intervention development. This study focused on testing a monitoring system that collected audio recordings, brief self-reports, and physiological data (heart rate) to log naturally occurring social interactions and to quantify the extent of physiological linkage in 11 persons with type II diabetes and their partners. Specifically, we used smartwatches to assess the feasibility of using commercially available devices. The smartwatches elicited 5-minute audio and heart rate recordings once an hour (mornings and evenings on weekdays, all day on weekends) when the Bluetooth signal indicated that partners were physically close, and speech was detected or when these 2 conditions were not detected, an assessment was triggered at the end of the hour to ensure sufficient data coverage. Self-reports were triggered consequently on linked smartphones to these recordings. Whereas raters coded social support as relatively infrequent (6% of audio recordings), participants self-reported that social support exchanges took place 53% of the time. On average, couples showed heart rate linkage of small to moderate degree. Notably, there was considerable variation in the extent of heart rate linkage and the occurrence of lags (indicating that one partner tended to drive heart rate changes in the other partner) across recordings within each couple. According to exploratory analyses, heart rate linkage was stronger when raters coded partner conversations and when interpersonal contact or social support exchanges were self-reported by partners. The following outlines important implications and considerations when collecting such data.

Rater-Coded and Self-Reported Social Contexts and Exchanges

Raters coded conversations between partners in 6 out of 10 audio recordings, and couples self-reported personal contact with each other 75% of the time, speaking to a good performance of the triggering system [37]. Concerning social exchanges, raters coded social support in 6% of audio recordings, whereas partners self-reported social support in 53% of self-reports. The low frequency of rater-coded social support dovetails with a previous study showing that support occurred only in about 4% of video data captured during 4 days of healthy couples' daily lives [41].

Self-report measures of social support may be influenced by social desirability bias and gender stereotypes [47]. For example, while self-report data support the gender support-gap hypothesis with men receiving more support from women within couples, naturalistic observational studies show that men provided the same or more everyday support to their partners than women

did [41,48], although the quality of the support might differ [49]. Women might also more frequently seek support from their partners, demonstrating the importance of collecting data on support solicitation in daily life [41]. On the other hand, observational measures are less vulnerable to demand characteristics and can capture supportive interactions that are not explicitly recognized or acknowledged by the individual receiving it [40,41].

However, self-report data capture types of interactions the individual perceives as supportive, whereas such interactions may not be immediately apparent to outside observers unfamiliar with the couple's unique dynamics. Indeed, the same action might be perceived as supportive by one person and as controlling or intrusive by another. The discrepancy between self-reported and rater-coded social support in this study emphasizes the need for a comprehensive approach to assessing social support in future research. For example, in future investigations, researchers could use text message content analysis [50] to capture conversations, expressions of empathy, encouragement, and tangible assistance between the couple. Participants could also be asked to press a button on a smartwatch whenever they are about to provide support, thereby eliciting an audio recording of the real-time support exchange. Furthermore, approaches could be developed using machine learning to automatically detect and capture moments when social support is occurring in daily life. Supplementary subjective measures could then be used to assess individuals' satisfaction with and the quality of support they receive within these objectively captured interactions.

Using Commercial Smartwatches to Monitor Heart Rate Linkage

Over the last decade, researchers have used different physiological parameters to demonstrate physiological linkage in couples, such as breathing rate, EDA, and cortisol [43,51,52] or voice stress [53]. With this project, we build on and extend this work by taking the research out of the laboratory and into daily life, using commercially available devices for heart rate monitoring. We found that everyday heart rate fluctuations in persons with type II diabetes and their partners showed intercorrelations of small to moderate degrees. Using an optical sensor (photoplethysmography) instead of electrical signals (electrocardiogram [ECG]) to estimate heart rate with commercially available devices has the benefit of being relatively unobtrusive, user-friendly, and affordable [54]. Contrasting consumer-grade wearables with the gold standard ECG, initial research overall shows promising findings regarding the accuracy of wrist-worn photoplethysmography devices, although they tend to underestimate heart rate slightly [55]. Specifically, wearables might be less accurate than ECG-based devices when participants are physically active versus resting (about 30% higher absolute error [56]). In our participants, audio recordings were coded for physical activity, which happened relatively rarely (in 2% of cases). However, future research should consider collecting objective movement data via accelerometry as an essential confounder and other covariates (for guidelines, see the study by Nelson et al [55]). Furthermore, future studies could consider integrating multiple physiological markers, such as blood pressure, cortisol, and heart rate variability, and leverage advanced sensor capabilities of wearables (eg, apple sensor kit [57]). This has the benefit of collecting data on markers sensitive to different features of everyday social contexts, for example, cortisol for socioevaluative situations and heart rate variability for prosocial behavior or compassion [58,59]. However, when doing so, it is vital to keep the inherent temporal dynamics of each system in mind. The appropriate sampling rate will differ between fast-acting systems, such as sympathetic activation as indexed by heart rate (seconds) versus hypothalamic-pituitary-adrenal activation as indexed by salivary cortisol (15-20 minutes) and needs to align with the psychosocial context under study [26].

Heart Rate Linkage and Social Contexts and Exchanges

Our exploratory analysis found that heart rate linkage was stronger when raters coded partner conversations and couples self-reported interpersonal contact, compared with moments when raters did not code that partners were engaged in conversations or when couples did not self-report having had personal contact with each other. Findings dovetail with theoretical notions of the dyadic regulation connectivity model, which suggests that any changes in the dyadic system such as interpersonal environments trigger immediate and synchronized changes throughout the network of dynamically linked processes between partners, including their physiology [5]. Other studies have also shown increased physiological linkage when partners are in each other's presence [52,60]. For example, research using a wristwatch to measure EDA in 40 young couples for 1 day showed that EDA linkage was significant only when couples self-reported being together but not apart [61]. When partners are engaged in conversation, they might follow the same narrative stimuli, which has been associated with increased heart rate linkage [62]. Our findings extend past research by focusing on moments when partners were actively interacting with each other instead of just being in physical proximity.

Heart rate linkage was also stronger when social support exchanges were self-reported by partners, as compared with moments when no social support exchanges were reported. Rater-coded social support was relatively infrequent (6%), preventing us from conducting exploratory analyses concerning its association with heart rate linkage. High physiological linkage has been associated with empathy, perspective-taking, positive interpersonal contexts such as feeling understood, appreciated or seeking help, and closeness to the partner [28,52,63]. A recent study investigating partners' cardiovascular markers found significant physiological linkage in heart rate and heart rate variability in 27 young to middle-aged couples while discussing positive and negative aspects of their relationship in the laboratory [34]. In addition, a laboratory study found that synchronicity in younger couples' skin conductance was higher during supportive touch when exposed to a pain stimulus, compared with the partner merely being present [64]. Our findings extend this research by indicating that heart rate fluctuations of couples managing a chronic health condition might link up during supportive, real-life interactions. Coherence between partners, for example, indicated by heart rate linkage, has been theorized to be central to dyadic regulation [5]. Yet, other studies demonstrated high physiological linkage

in couples with marital strain, particularly regarding markers of the stress response system such as cortisol [65]. Potentially, parasympathetic nervous system markers (eg, respiratory sinus arrhythmia) might synchronize more during positive interpersonal contexts. In contrast, markers of the sympathetic nervous system (eg, EDA) might more likely become entrained during negative interpersonal contexts [28,29].

Future studies need to build on and replicate our findings using larger sample sizes, also looking at the underlying mechanisms and moderators of heart rate linkage, such as perspective taking, contagion of emotions, attachment style, relationship quality, and interdependent self-construal [28]. For example, is heart rate linkage stronger when partners try to take each other's perspective and higher in couples with greater relationship satisfaction? Does lagged heart rate linkage accompany emotions (positive and negative) or stress being transferred from one partner to the other? [53] Such a study could use audio recordings to derive indices of vocal quality or pitch indexing emotional arousal such as the fundamental frequency [66]. It is also an open question whether stronger physiological heart rate linkage during social support exchanges is associated with more or less favorable outcomes of supportive interactions [67]. For example, increased skin conductance synchronicity was associated with less self-reported pain intensity during painful thermal stimulation when receiving supportive touch from the partner [64].

Limitations and Future Research

Considering the sample size of 11 couples due to the impact of the COVID-19 pandemic on data collection, findings from this study need to be replicated to test generalizability. Furthermore, our sample comprised couples managing diabetes of mostly older age, long relationship duration, and relatively good diabetic control. Therefore, these couples could represent dyadic systems demonstrating reasonably effective dyadic regulation and thus higher heart rate linkage. It remains to be seen whether biopsychosocial dynamics differ by age and chronic disease type (eg, cancer [21]). In addition, we coded limited information about the nature of everyday social interactions between partners (social support occurrence, type of conversation), and interrater reliabilities were low for some coding categories. The unsatisfactory reliability of rater codings for conversation types prevented us from examining differences in heart rate linkage by type of conversation (eg, deep vs practical). Whether heart rate linkage is tied to the nature and content of partner conversations thus presents an important future direction.

Future studies could also investigate heart rate linkage in the context of support quality and types of support (emotional, instrumental). Given that previous literature has linked conflict to increased physiological linkage, it would also be important to study heart rate synchronization in the context of negative types of social exchanges managing chronic health conditions such as negative social control [68]. Future research should also examine other relevant constructs for couples managing disease, such as dyadic efficacy [69] or shared appraisals of the disease ("we-disease" [70]), and include measures of well-being (eg, happiness, meaning). Eventually, more advanced systems that collect multimodal data from couples managing chronic health conditions could learn when and how to intervene (eg, to nudge partners to provide support) to promote healthy lifestyle engagement and adaptive disease management [61,71]. Increasingly, researchers use machine learning algorithms to design just-in-time adaptive interventions [72,73], although they have seldom been applied in a dyadic context. These interventions can be delivered through digital platforms, making them widely accessible, particularly for individuals living in remote or underserved areas or those with transportation or mobility limitations.

Conclusions

This study explored the feasibility of using commercial wearables to monitor the unique communication and support dynamics between romantic partners, particularly in managing a chronic health condition. First, the rater-coded audio recordings and self-report analysis provided valuable information about the occurrence of social support. Interestingly, while self-reports indicated that partners reported social support approximately half of the time, raters coded social support in only 6% of the recordings. This suggests a potential disparity between partners' self-perception and external observation of supportive behaviors. Second, we demonstrated that couples exhibited small to moderate interconnections in heart rate fluctuations, indicating physiological linkage between partners. Heart rate linkage was stronger when partner conversations were coded, partners self-reported interpersonal contact, and partners self-reported social support exchanges. Findings provide initial evidence on contexts and behaviors that may influence physiological interconnectivity within couples. Overall, the use of wearables for continuous and unobtrusive collection of biopsychosocial data combined with self-report data in real-world settings holds great promise for enabling better and targeted support for individuals managing chronic health conditions and their partners.

Acknowledgments

This study was funded by the Swiss National Science Foundation (CR12I1_166348/1). Theresa Pauly gratefully acknowledges support from the Canada Research Chairs program. We thank Dr Matthias Mehl for his valuable contribution to study design.

Conflicts of Interest

GB, TK, and EF are affiliated with the Centre for Digital Health Interventions (CDHI); a joint initiative of the Institute for Implementation Science in Health Care, University of Zurich; the Department of Management, Technology, and Economics at ETH Zurich; and the Institute of Technology Management and School of Medicine at the University of St. Gallen. CDHI is funded in part by the Swiss health insurer CSS, the Austrian health care provider (and corporate start-up of UNIQA) Mavie Next, and

the Swiss investor MTIP. EF and TK are also cofounders of Pathmate Technologies, a university spin-off company that creates and delivers digital clinical pathways. However, neither CSS, Mavie Next (UNIQA), and MTIP nor Pathmate Technologies were involved in this study. All other authors declare no conflict of interest.

Multimedia Appendix 1
Magnitude of cross-correlation by rater-coded and self-reported social context and exchange. The figure shows that cross-correlations tended to be higher in situations when raters coded that partners were engaged in conversations (A), when participants self-reported a partner interaction (B), and when participants self-reported a social support exchange (C).
[PNG File , 54 KB - mhealth_v12i1e49576_app1.png]

References

1. DeLongis A, O'Brien T. An interpersonal framework for stress and coping: an application to the families of alzheimer's patients. In: Stephens MAP, Crowther JH, Hobfoll SE, Tennenbaum DL, editors. Stress and Coping in Later-Life Families. New York: Hemisphere Publishing Corp; 1990:221-229.
2. Dunbar-Jacob J, Mortimer-Stephens MK. Treatment adherence in chronic disease. J Clin Epidemiol 2001;54 Suppl 1:S57-S60. [doi: 10.1016/s0895-4356(01)00457-7] [Medline: 11750211]
3. Polonsky WH, Henry RR. Poor medication adherence in type 2 diabetes: recognizing the scope of the problem and its key contributors. Patient Prefer Adherence 2016;10:1299-1307 [FREE Full text] [doi: 10.2147/PPA.S106821] [Medline: 27524885]
4. Lyons KS, Lee CS. The theory of dyadic illness management. J Fam Nurs 2018;24(1):8-28. [doi: 10.1177/1074840717745669] [Medline: 29353528]
5. Karademas EC. A new perspective on dyadic regulation in chronic illness: the dyadic regulation connectivity model. Health Psychol Rev 2022;16(1):1-21. [doi: 10.1080/17437199.2021.1874471] [Medline: 33427585]
6. Berg CA, Upchurch R. A developmental-contextual model of couples coping with chronic illness across the adult life span. Psychol Bull 2007;133(6):920-954. [doi: 10.1037/0033-2909.133.6.920] [Medline: 17967089]
7. Schwarzer R, Knoll N. Functional roles of social support within the stress and coping process: a theoretical and empirical overview. Int J Psychol 2007;42(4):243-252. [doi: 10.1080/00207590701396641]
8. Rackow P, Berli C, Lüscher J, Luszczynska A, Scholz U. Emotional or instrumental support? Distinct effects on vigorous exercise and affect. Psychol Sport Exerc 2017;33:66-74. [doi: 10.1016/j.psychsport.2017.07.011]
9. Scholz U, Knoll N, Roigas J, Gralla O. Effects of provision and receipt of social support on adjustment to laparoscopic radical prostatectomy. Anxiety Stress Coping 2008;21(3):227-241. [doi: 10.1080/10615800801983759] [Medline: 18938287]
10. Luszczynska A, Pawlowska I, Cieslak R, Knoll N, Scholz U. Social support and quality of life among lung cancer patients: a systematic review. Psychooncology 2013;22(10):2160-2168. [doi: 10.1002/pon.3218] [Medline: 23097417]
11. Bodenmann G. Dyadic coping: a systemic-transactional view of stress and coping among couples: theory and empirical findings. Eur Rev Appl Psychol 1997;47(2):137-141.
12. Ditzen B, Heinrichs M. Psychobiology of social support: the social dimension of stress buffering. Restor Neurol Neurosci 2014;32(1):149-162. [doi: 10.3233/RNN-139008] [Medline: 23603443]
13. Dreisoerner A, Junker NM, Schlotz W, Heimrich J, Bloemeke S, Ditzen B, et al. Self-soothing touch and being hugged reduce cortisol responses to stress: a randomized controlled trial on stress, physical touch, and social identity. Compr Psychoneuroendocrinol 2021;8:100091 [FREE Full text] [doi: 10.1016/j.cpnec.2021.100091] [Medline: 35757667]
14. Meyler D, Stimpson JP, Peek MK. Health concordance within couples: a systematic review. Soc Sci Med 2007;64(11):2297-2310. [doi: 10.1016/j.socscimed.2007.02.007] [Medline: 17374552]
15. Bodenmann G. Dyadic coping and its significance for marital functioning. In: Revenson TA, Kayser K, Bodenmann G, editors. Couples Coping With Stress: Emerging Perspectives on Dyadic Coping. Washington, DC: American Psychological Association; 2005:33-49.
16. Patel SA, Dhillon PK, Kondal D, Jeemon P, Kahol K, Manimunda SP, et al. Chronic disease concordance within indian households: a cross-sectional study. PLoS Med 2017;14(9):e1002395 [FREE Full text] [doi: 10.1371/journal.pmed.1002395] [Medline: 28961237]
17. Leong A, Rahme E, Dasgupta K. Spousal diabetes as a diabetes risk factor: a systematic review and meta-analysis. BMC Med 2014;12:12 [FREE Full text] [doi: 10.1186/1741-7015-12-12] [Medline: 24460622]
18. Jeong S, Cho SI. Concordance in the health behaviors of couples by age: a cross-sectional study. J Prev Med Public Health 2018;51(1):6-14 [FREE Full text] [doi: 10.3961/jpmph.17.137] [Medline: 29397647]
19. Scholz U. It's time to think about time in health psychology. Appl Psychol Health Well Being 2019;11(2):173-186. [doi: 10.1111/aphw.12156] [Medline: 30972951]
20. Trull TJ, Ebner-Priemer U. Ambulatory assessment. Annu Rev Clin Psychol 2013;9:151-176 [FREE Full text] [doi: 10.1146/annurev-clinpsy-050212-185510] [Medline: 23157450]
21. Robbins ML, Karan A, López AM, Weihs KL. Naturalistically observing noncancer conversations among couples coping with breast cancer. Psychooncology 2018;27(9):2206-2213. [doi: 10.1002/pon.4797] [Medline: 29904980]

22. Robbins ML, Mehl MR, Smith HL, Weihs KL. Linguistic indicators of patient, couple, and family adjustment following breast cancer. Psychooncology 2013;22(7):1501-1508 [FREE Full text] [doi: 10.1002/pon.3161] [Medline: 22887054]
23. Sharma A, Mentz RJ, Granger BB, Heitner JF, Cooper LB, Banerjee D, et al. Utilizing mobile technologies to improve physical activity and medication adherence in patients with heart failure and diabetes mellitus: rationale and design of the TARGET-HF-DM Trial. Am Heart J 2019;211:22-33. [doi: 10.1016/j.ahj.2019.01.007] [Medline: 30831331]
24. Wilhelm FH, Pfaltz MC, Grossman P. Continuous electronic data capture of physiology, behavior and experience in real life: towards ecological momentary assessment of emotion. Interacting With Comput 2006;18(2):171-186. [doi: 10.1016/j.intcom.2005.07.001]
25. Lehmann V, Föll S, Maritsch M, van Weenen E, Kraus M, Lagger S, et al. Noninvasive hypoglycemia detection in people with diabetes using smartwatch data. Diabetes Care 2023;46(5):993-997 [FREE Full text] [doi: 10.2337/dc22-2290] [Medline: 36805169]
26. Marzoratti A, Evans TM. Measurement of interpersonal physiological synchrony in dyads: a review of timing parameters used in the literature. Cogn Affect Behav Neurosci 2022;22(6):1215-1230. [doi: 10.3758/s13415-022-01011-1] [Medline: 35556231]
27. Timmons AC, Margolin G, Saxbe DE. Physiological linkage in couples and its implications for individual and interpersonal functioning: a literature review. J Fam Psychol 2015;29(5):720-731 [FREE Full text] [doi: 10.1037/fam0000115] [Medline: 26147932]
28. Pauly T, Gerstorf D, Wahl HW, Hoppmann CA. A developmental-contextual model of couple synchrony across adulthood and old age. Psychol Aging 2021;36(8):943-956. [doi: 10.1037/pag0000651] [Medline: 34672638]
29. Palumbo RV, Marraccini ME, Weyandt LL, Wilder-Smith O, McGee HA, Liu S, et al. Interpersonal autonomic physiology: a systematic review of the literature. Pers Soc Psychol Rev 2017;21(2):99-141. [doi: 10.1177/1088868316628405] [Medline: 26921410]
30. Behrens F, Snijdewint JA, Moulder RG, Prochazkova E, Sjak-Shie EE, Boker SM, et al. Physiological synchrony is associated with cooperative success in real-life interactions. Sci Rep 2020;10(1):19609 [FREE Full text] [doi: 10.1038/s41598-020-76539-8] [Medline: 33184357]
31. Mønster D, Håkonsson DD, Eskildsen JK, Wallot S. Physiological evidence of interpersonal dynamics in a cooperative production task. Physiol Behav 2016;156:24-34. [doi: 10.1016/j.physbeh.2016.01.004] [Medline: 26773466]
32. Mitkidis P, McGraw JJ, Roepstorff A, Wallot S. Building trust: heart rate synchrony and arousal during joint action increased by public goods game. Physiol Behav 2015;149:101-106. [doi: 10.1016/j.physbeh.2015.05.033] [Medline: 26037635]
33. Rennung M, Göritz AS. Prosocial consequences of interpersonal synchrony: a meta-analysis. Z Psychol 2015;224(3):168-189 [FREE Full text] [doi: 10.1027/2151-2604/a000252] [Medline: 28105388]
34. Coutinho J, Pereira A, Oliveira-Silva P, Meier D, Lourenço V, Tschacher W. When our hearts beat together: cardiac synchrony as an entry point to understand dyadic co-regulation in couples. Psychophysiology 2021;58(3):e13739. [doi: 10.1111/psyp.13739] [Medline: 33355941]
35. Helm JL, Sbarra D, Ferrer E. Assessing cross-partner associations in physiological responses via coupled oscillator models. Emotion 2012;12(4):748-762. [doi: 10.1037/a0025036] [Medline: 21910541]
36. Lüscher J, Kowatsch T, Boateng G, Santhanam P, Bodenmann G, Scholz U. Social support and common dyadic coping in couples' dyadic management of type II diabetes: protocol for an ambulatory assessment application. JMIR Res Protoc 2019;8(10):e13685 [FREE Full text] [doi: 10.2196/13685] [Medline: 31588907]
37. Boateng G, Santhanam P, Fleisch E, Lüscher J, Pauly T, Scholz U, et al. Development, deployment, and evaluation of DyMand—an open-source smartwatch and smartphone system for capturing couples? Dyadic interactions in chronic disease management in daily life. ArXiv :1-25 Preprint posted online on October 11, 2019. [doi: 10.1145/3300061.3343399]
38. Sherwani SI, Khan HA, Ekhzaimy A, Masood A, Sakharkar MK. Significance of HbA1c test in diagnosis and prognosis of diabetic patients. Biomark Insights 2016;11:95-104 [FREE Full text] [doi: 10.4137/BMI.S38440] [Medline: 27398023]
39. Poverty. Swiss Federal Statistical Office. URL: https://www.bfs.admin.ch/bfs/en/home/statistics/economic-social-situation-population/economic-and-social-situation-of-the-population/poverty-deprivation/poverty.html [accessed 2024-06-25]
40. Mehl MR, Pennebaker JW, Crow DM, Dabbs J, Price JH. The electronically activated recorder (EAR): a device for sampling naturalistic daily activities and conversations. Behav Res Methods Instrum Comput 2001;33(4):517-523. [doi: 10.3758/bf03195410] [Medline: 11816455]
41. Wang SW, Repetti RL. Who gives to whom? Testing the support gap hypothesis with naturalistic observations of couple interactions. J Fam Psychol 2016;30(4):492-502. [doi: 10.1037/fam0000196] [Medline: 27010600]
42. Hallgren KA. Computing inter-rater reliability for observational data: an overview and tutorial. Tutor Quant Methods Psychol 2012;8(1):23-34 [FREE Full text] [doi: 10.20982/tqmp.08.1.p023] [Medline: 22833776]
43. McAssey MP, Helm J, Hsieh F, Sbarra DA, Ferrer E. Methodological advances for detecting physiological synchrony during dyadic interactions. Methodology 2013;9(2):41-53. [doi: 10.1027/1614-2241/a000053]
44. R: a language and environment for statistical computing. R Core Team. 2022. URL: https://www.R-project.org/ [accessed 2024-06-25]

45. Behrens F, Moulder RG, Boker SM, Kret ME. Quantifying physiological synchrony through windowed cross-correlation analysis: statistical and theoretical considerations. bioRxiv :1-53 Preprint posted online on August 28, 2020. [doi: 10.1101/2020.08.27.269746]
46. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using. J Stat Soft 2015;67(1):1-48. [doi: 10.18637/jss.v067.i01]
47. Barbee AP, Cunningham MR, Winstead BA, Derlega VJ, Gulley MR, Yankeelov PA, et al. Effects of gender role expectations on the social support process. J Soc Issues 1993;49(3):175-190. [doi: 10.1111/j.1540-4560.1993.tb01175.x]
48. Neff LA, Karney BR. Gender differences in social support: a question of skill or responsiveness? J Pers Soc Psychol 2005;88(1):79-90. [doi: 10.1037/0022-3514.88.1.79] [Medline: 15631576]
49. Bodenmann G, Meuwly N, Germann J, Nussbeck FW, Heinrichs M, Bradbury TN. Effects of stress on the social support provided by men and women in intimate relationships. Psychol Sci 2015;26(10):1584-1594. [doi: 10.1177/0956797615594616] [Medline: 26341561]
50. Brinberg M, Ram N. Do new romantic couples use more similar language over time? Evidence from intensive longitudinal text messages. J Commun 2021;71(3):454-477 [FREE Full text] [doi: 10.1093/joc/jqab012] [Medline: 34335083]
51. Coutinho J, Oliveira-Silva P, Fernandes E, Gonçalves OF, Correia D, Perrone Mc-Govern K, et al. Psychophysiological synchrony during verbal interaction in romantic relationships. Fam Process 2019;58(3):716-733. [doi: 10.1111/famp.12371] [Medline: 29888517]
52. Pauly T, Michalowski VI, Drewelies J, Gerstorf D, Ashe MC, Madden KM, et al. Cortisol synchrony in older couples: daily socioemotional correlates and interpersonal differences. Psychosom Med 2020;82(7):669-677. [doi: 10.1097/PSY.0000000000000838] [Medline: 32868536]
53. Bulling LJ, Hilpert P, Bertschi IC, Ivic A, Bodenmann G. Associations between vocal arousal and dyadic coping during couple interactions after a stress induction. Int J Appl Posit Psychol 2023;8(Suppl 2):187-204 [FREE Full text] [doi: 10.1007/s41042-023-00087-5] [Medline: 38162834]
54. Puterman E, Pauly T, Ruissen G, Nelson B, Faulkner G. Move more, move better: a narrative review of wearable technologies and their application to precision health. Health Psychol 2021;40(11):803-810. [doi: 10.1037/hea0001125] [Medline: 34855416]
55. Nelson BW, Low CA, Jacobson N, Areán P, Torous J, Allen NB. Guidelines for wrist-worn consumer wearable assessment of heart rate in biobehavioral research. NPJ Digit Med 2020;3:90 [FREE Full text] [doi: 10.1038/s41746-020-0297-4] [Medline: 32613085]
56. Bent B, Goldstein BA, Kibbe WA, Dunn JP. Investigating sources of inaccuracy in wearable optical heart rate sensors. NPJ Digit Med 2020;3:18 [FREE Full text] [doi: 10.1038/s41746-020-0226-6] [Medline: 32047863]
57. SensorKit: retrieve data and derived metrics from sensors on an iPhone, or paired Apple Watch. Apple Inc. URL: https://developer.apple.com/documentation/sensorkit/ [accessed 2024-06-25]
58. Di Bello M, Carnevali L, Petrocchi N, Thayer JF, Gilbert P, Ottaviani C. The compassionate vagus: a meta-analysis on the connection between compassion and heart rate variability. Neurosci Biobehav Rev 2020;116:21-30. [doi: 10.1016/j.neubiorev.2020.06.016] [Medline: 32554001]
59. Dickerson SS, Kemeny ME. Acute stressors and cortisol responses: a theoretical integration and synthesis of laboratory research. Psychol Bull 2004;130(3):355-391. [doi: 10.1037/0033-2909.130.3.355] [Medline: 15122924]
60. Saxbe D, Repetti RL. For better or worse? Coregulation of couples' cortisol levels and mood states. J Pers Soc Psychol 2010;98(1):92-103. [doi: 10.1037/a0016959] [Medline: 20053034]
61. Timmons AC, Baucom BR, Han SC, Perrone L, Chaspari T, Narayanan SS, et al. New frontiers in ambulatory assessment: big data methods for capturing couples' emotions, vocalizations, and physiology in daily life. Soc Psychol Pers Sci 2017;8(5):552-563. [doi: 10.1177/1948550617709115]
62. Pérez P, Madsen J, Banellis L, Türker B, Raimondo F, Perlbarg V, et al. Conscious processing of narrative stimuli synchronizes heart rate between individuals. Cell Rep 2021;36(11):109692 [FREE Full text] [doi: 10.1016/j.celrep.2021.109692] [Medline: 34525363]
63. Nelson BW, Laurent SM, Bernstein R, Laurent HK. Perspective-taking influences autonomic attunement between partners during discussion of conflict. J Soc Personal 2017;34(2):139-165. [doi: 10.1177/0265407515626595]
64. Reddan MC, Young H, Falkner J, López-Solà M, Wager TD. Touch and social support influence interpersonal synchrony and pain. Soc Cogn Affect Neurosci 2020;15(10):1064-1075 [FREE Full text] [doi: 10.1093/scan/nsaa048] [Medline: 32301998]
65. Liu S, Rovine MJ, Klein LC, Almeida DM. Synchrony of diurnal cortisol pattern in couples. J Fam Psychol 2013;27(4):579-588 [FREE Full text] [doi: 10.1037/a0033735] [Medline: 23978320]
66. Imel ZE, Barco JS, Brown HJ, Baucom BR, Baer JS, Kircher JC, et al. The association of therapist empathy and synchrony in vocally encoded arousal. J Couns Psychol 2014;61(1):146-153 [FREE Full text] [doi: 10.1037/a0034943] [Medline: 24274679]
67. Zietlow AL, Eckstein M, Hernández C, Nonnenmacher N, Reck C, Schaer M, et al. Dyadic coping and its underlying neuroendocrine mechanisms—implications for stress regulation. Front Psychol 2019;9:2600 [FREE Full text] [doi: 10.3389/fpsyg.2018.02600] [Medline: 30687147]

68. Scholz U, Stadler G, Berli C, Lüscher J, Knoll N. How do people experience and respond to social control from their partner? Three daily diary studies. Front Psychol 2021;11:613546 [FREE Full text] [doi: 10.3389/fpsyg.2020.613546] [Medline: 33519637]
69. Brosseau DC, Braeken J, Carmack CL, Rosberger Z, Körner A. We think we can: development of the dyadic efficacy scale for cancer. J Psychosoc Oncol Res Pract 2021;3(4):e066. [doi: 10.1097/or9.0000000000000066]
70. Badr H, Acitelli LK. Re-thinking dyadic coping in the context of chronic illness. Curr Opin Psychol 2017;13:44-48. [doi: 10.1016/j.copsyc.2016.03.001] [Medline: 28813292]
71. Wulfovich S, Fiordelli M, Rivas H, Concepcion W, Wac K. " I must try harder": design implications for mobile apps and wearables contributing to self-efficacy of patients with chronic conditions. Front Psychol 2019;10:2388 [FREE Full text] [doi: 10.3389/fpsyg.2019.02388] [Medline: 31749733]
72. Oikonomidi T, Ravaud P, LeBeau J, Tran VT. A systematic scoping review of just-in-time, adaptive interventions finds limited automation and incomplete reporting. J Clin Epidemiol 2023;154:108-116. [doi: 10.1016/j.jclinepi.2022.12.006] [Medline: 36521653]
73. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-Time Adaptive Interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. Ann Behav Med 2018;52(6):446-462 [FREE Full text] [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]

Abbreviations

DYMAND: DYadic MANagement of Diabetes
ECG: electrocardiogram
EDA: electrodermal activity
HbA_{1c}: hemoglobin A_{1c}
ICC: intraclass correlation

Edited by L Buis; submitted 07.07.23; peer-reviewed by K Nomura, D Sorkin; comments to author 29.01.24; revised version received 24.04.24; accepted 05.06.24; published 05.08.24.

Please cite as:
Pauly T, Lüscher J, Wilhelm LO, Amrein MA, Boateng G, Kowatsch T, Fleisch E, Bodenmann G, Scholz U
Using Wearables to Study Biopsychosocial Dynamics in Couples Who Cope With a Chronic Health Condition: Ambulatory Assessment Study
JMIR Mhealth Uhealth 2024;12:e49576
URL: https://mhealth.jmir.org/2024/1/e49576
doi:10.2196/49576
PMID:

Original Paper

Sleep During the COVID-19 Pandemic: Longitudinal Observational Study Combining Multisensor Data With Questionnaires

Nguyen Luong[1], MSc; Gloria Mark[2], PhD; Juhi Kulshrestha[1], PhD; Talayeh Aledavood[1], PhD

[1]Department of Computer Science, Aalto University, Espoo, Finland

[2]Informatics Department, University of California, Irvine, Irvine, CA, United States

Corresponding Author:
Nguyen Luong, MSc
Department of Computer Science
Aalto University
Konemiehentie 2
Espoo, 02150
Finland
Phone: 358 0442404485
Email: nguyen.luong@aalto.fi

Abstract

Background: The COVID-19 pandemic prompted various containment strategies, such as work-from-home policies and reduced social contact, which significantly altered people's sleep routines. While previous studies have highlighted the negative impacts of these restrictions on sleep, they often lack a comprehensive perspective that considers other factors, such as seasonal variations and physical activity (PA), which can also influence sleep.

Objective: This study aims to longitudinally examine the detailed changes in sleep patterns among working adults during the COVID-19 pandemic using a combination of repeated questionnaires and high-resolution passive measurements from wearable sensors. We investigate the association between sleep and 5 sets of variables: (1) demographics; (2) sleep-related habits; (3) PA behaviors; and external factors, including (4) pandemic-specific constraints and (5) seasonal variations during the study period.

Methods: We recruited working adults in Finland for a 1-year study (June 2021-June 2022) conducted during the late stage of the COVID-19 pandemic. We collected multisensor data from fitness trackers worn by participants, as well as work and sleep-related measures through monthly questionnaires. Additionally, we used the Stringency Index for Finland at various points in time to estimate the degree of pandemic-related lockdown restrictions during the study period. We applied linear mixed models to examine changes in sleep patterns during this late stage of the pandemic and their association with the 5 sets of variables.

Results: The sleep patterns of 27,350 nights from 112 working adults were analyzed. Stricter pandemic measures were associated with an increase in total sleep time (TST) (β=.003, 95% CI 0.001-0.005; *P*<.001) and a delay in midsleep (MS) (β=.02, 95% CI 0.02-0.03; *P*<.001). Individuals who tend to snooze exhibited greater variability in both TST (β=.15, 95% CI 0.05-0.27; *P*=.006) and MS (β=.17, 95% CI 0.03-0.31; *P*=.01). Occupational differences in sleep pattern were observed, with service staff experiencing longer TST (β=.37, 95% CI 0.14-0.61; *P*=.004) and lower variability in TST (β=–.15, 95% CI –0.27 to –0.05; *P*<.001). Engaging in PA later in the day was associated with longer TST (β=.03, 95% CI 0.02-0.04; *P*<.001) and less variability in TST (β=–.01, 95% CI –0.02 to 0.00; *P*=.02). Higher intradaily variability in rest activity rhythm was associated with shorter TST (β=–.26, 95% CI –0.29 to –0.23; *P*<.001), earlier MS (β=–.29, 95% CI –0.33 to –0.26; *P*<.001), and reduced variability in TST (β=–.16, 95% CI –0.23 to –0.09; *P*<.001).

Conclusions: Our study provided a comprehensive view of the factors affecting sleep patterns during the late stage of the pandemic. As we navigate the future of work after the pandemic, understanding how work arrangements, lifestyle choices, and sleep quality interact will be crucial for optimizing well-being and performance in the workforce.

(JMIR Mhealth Uhealth 2024;12:e53389) doi:10.2196/53389

KEYWORDS

computational social science; digital health; COVID-19; sleep; longitudinal; wearables; surveys; observational study; isolation; sleep patterns; sleep pattern; questionnaires; Finland; fitness trackers; fitness tracker; wearable; sleeping habits; sleeping habit; work from home

Introduction

Background

Sleep is a crucial component of daily life, closely interconnected with all aspects of our routines and overall well-being, including mental health [1,2], physical health [3], and work performance [4,5]. The COVID-19 pandemic profoundly impacted various aspects of daily life, with sleep patterns being a particularly significant area of concern. However, the effects on sleep were often indirect, resulting from changes in daily routines and lifestyle adjustments rather than being a direct consequence of the virus.

In response to the pandemic, outdoor restrictions limited our exposure to natural daylight, a crucial element for regulating circadian rhythms and sleep patterns [6]. Similarly, mobility restrictions altered daily physical activity (PA) patterns. Additionally, workplace restrictions led to work-from-home policies, which resulted in reduced mobility and flexible working hours. While these changes led to more relaxed work schedules, they also blurred the boundaries between professional and personal life. Notably, factors such as daylight exposure, PA, and work routines—each significantly affected by the pandemic—are well-established influences on sleep health [7,8].

Traditional sleep measurements often rely on self-reported methods, such as the Karolinska [9] or Pittsburgh sleep diary [10]. While these methods are effective for tracking day-to-day sleep over short periods, conducting diary studies over longer intervals is generally not feasible due to the cognitive burden on participants. Nonintrusive measurements using smartphones and fitness trackers have recently emerged as a more viable alternative for capturing sleep data over extended periods. While consumer-grade devices may not precisely detect sleep stages, they have shown promising results for measuring sleep onset, duration, and wake-up time. Assessing sleep with these devices has the advantage of capturing data in people's natural living environments, unlike sleep laboratories. Additionally, this method is not subject to memory biases that can occur with survey responses and sleep diaries.

The evolution of mobile health (mHealth) technologies has significantly enhanced traditional sleep monitoring methods, particularly through the use of wearable devices. These devices offer a more accessible and less invasive way to monitor sleep patterns, while also deepening our understanding of sleep-related phenomena. For instance, wearable devices have been used to determine individuals' chronotypes and track their sleep and activity rhythms over extended periods [11,12]. They have also been used to measure sleep alignment between coworkers [13], examine the relationship between sleep and burnout [14], and assess sleep patterns in various populations, including patients with mental disorders [15]. Several studies have confirmed the validity and reliability of wearable devices, demonstrating notable sensitivity compared with the gold-standard polysomnography (PSG). For example, a review of 7 consumer sleep-tracking devices [16] highlighted their high effectiveness in detecting sleep relative to PSG. Similarly, a study [17] evaluated 6 consumer wearable devices and validated their accuracy in assessing sleep timing and duration compared with PSG.

Prior research comparing sleep patterns before and during the pandemic has revealed notable differences. Studies found that following the pandemic's onset, individuals tended to go to bed later [18], slept for longer durations [19], exhibited reduced variability between weekday and weekend sleep [20,21], and experienced increased sleep disturbances or diminished sleep quality [22]. Various factors have been identified as contributing to these disruptions in sleep routines, including decreased PA [23], social isolation [24], increased use of electronic devices [4], and the shift to working from home [13].

While previous studies have focused on the immediate consequences of lockdowns and restriction policies, less attention has been paid to the long-term effects, particularly during the late stages of the pandemic when restrictions began to relax. This phase is crucial for understanding the residual effects of the pandemic on sleep patterns and how quickly individuals revert to their prepandemic sleep habits. The transition to working from home as the default mode has resulted in a less constrained work-life routine, leading to more flexible sleep-wake schedules. Certain demographics may benefit more from these transitions, such as individuals with more flexible routines (eg, research personnel) or those who tend to snooze their alarms after waking, referred to as "snoozers." Additionally, occupation is a known factor influencing sleep patterns, with a classic example being the contrast between shift workers and nonshift workers [25,26]. However, less is understood about the differences between various roles within academia, such as researchers with deadline-driven roles and administrative personnel typically following a 9-to-5 schedule. Therefore, a comprehensive, longitudinal analysis of sleep patterns that includes these variables and extends into the late stages of the pandemic is important.

Objectives

Our study aims to provide a holistic view of how the pandemic has influenced sleep patterns. We evaluate the long-term relationships between sleep patterns, including average and variability in total sleep duration and sleep timing, alongside individuals' characteristics (demographics, occupation, and PA) and external factors (stringency of restriction policies and seasonal variations). Our research utilizes longitudinal data from fitness trackers and questionnaire responses collected from working adults at a Finnish university. This extensive data set enables us to examine shifts in sleep behavior during the later stages of the COVID-19 pandemic, from June 2021 to June 2022. The study's timeframe covers a full annual seasonal cycle, which is crucial for analyzing sleep patterns in Finland, where significant seasonal changes and daylight variations occur due to its northern latitude.

Methods

Study Data

This work used data from the cor:ona (comparison of rhythms: old vs. new) study [27] as part of a 1-year multimodal data set of working adults.

Ethics Approval

The study was approved by the Aalto University Research Ethics Committee (approval number D/536/03.04/2021_COR_ONA).

Participants and Procedures

The cor:ona study recruited 128 full-time employees from a university in Finland for a 1-year investigation into how their daily activities changed during different stages of the COVID-19 pandemic. Throughout the study, participants wore a Polar Ignite fitness tracker (Polar Electro Oy), enabling us to unobtrusively collect various measures related to sleep and PA. In addition, participants completed an initial baseline questionnaire, an exit questionnaire, and a shorter version of the baseline questionnaire each month. The monthly questionnaires asked for information about their daily routines, work, and sleep quality over the past month. The detailed recruitment procedure and participants' demographics were described in a previous study [27].

Fitness Tracker Data

Sleep Measures

The fitness trackers measured bedtime (defined as the recorded time when a person fell asleep), waketime (defined as the recorded time when a person woke up), and interruption duration (defined as the total time in seconds spent awake between sleep start and end times) for each day. A sleep period was defined as the longest sleep episode for each day. Sleep patterns were measured using 4 metrics: (1) total sleep time (TST), which measured the time a person spent asleep, calculated as the duration from bedtime minus the interruption duration; (2) midsleep (MS), the midpoint between bedtime and waketime, which was used to measure sleep timing and computed as (bedtime + TST)/2. Additionally, we proposed 2 other metrics to measure sleep regularity: (3) TST variability, computed as the SD of TST during weekdays (Sunday night to Thursday night); and (4) MS variability, computed as the SD of MS during weekdays. We focused exclusively on weekdays due to the expected differences between weekday and weekend sleep patterns. The Niimpy behavioral data analysis toolbox [28] was used for extracting sleep measurements.

Physical Activity Measures

The fitness tracker recorded the number of steps taken each hour, which were then summed to provide a daily step count. To comprehensively account for daily PA patterns, including their timing and distribution, we introduced 2 additional metrics: midstep and intradaily variability (IV) [29]. These metrics are designed to capture the timing and dispersion of PAs throughout the day. Specifically, midstep represents the hour of the day when half of the total number of steps is achieved, analogous to MS in the context of PA. By contrast, IV quantifies the fragmentation of the activity-rest rhythm and is measured as follows:

where N=24 is the total number of samples within each day; X_i is the i measurement sampled at P=60-minute interval; and  is the average value of all samples in a day. Low IV indicates less fragmented activity-rest rhythm, whereas high IV could imply daytime naps or nighttime awakenings.

External Data

Seasonal data were collected from the World Weather Online developer application programming interface [30]. Given the significant variation in day length in Finland during the study (up to 13 hours), day length was used as a proxy for seasonal variables. The choice of day length as a proxy was motivated by Friborg et al [31]. The study compared 2 geographically distinct locations with substantial differences in day length variability: Ghana and Norway. Although no noticeable seasonal effects of day length were observed in Ghanaians, Norwegians showed a delay in both bedtime and waketime during summer weekdays, though sleep duration remained relatively unaffected.

We also utilized the Stringency Index (SI) [32], a composite measure ranging from 0 to 100, to assess daily COVID-19 restriction policies. Higher values on this index indicate more stringent COVID-19 restrictions, including measures such as school and workplace closures, the cancellation of public events, and the enforcement of stay-at-home orders. This index allows for standardized comparisons of policy responses across different countries or regions, as well as changes within the same region over time.

Questionnaire Data

Upon entering the study, participants completed a baseline questionnaire that collected basic background information, including age, gender, chronotype, occupation, and origin, among others. Chronotype was assessed using the reduced Morningness-Eveningness Questionnaire (MEQ) [33], with higher scores indicating a morning type and lower scores indicating an evening type. For the origin-related question, participants chose from 3 options: Finland, Europe (excluding Finland), or outside of Europe. Participants indicating they were from Finland were classified as Finnish, while those selecting other options were described as having a "migrant background." Regarding occupation, participants specified whether they were academic or service staff. The term "academic staff" refers to individuals involved in academic and research activities within the organization, while "service staff" includes those in roles such as human resources and other administrative or support functions. Participants were determined as a snoozer if they answered "yes" to the following question: "Snoozing can be considered as choosing to go back to sleep after an alarm has awakened you intending to wake up later; setting the alarm earlier than when you intend to wake up; or setting multiple alarms with the intent to not wake up on the first alarm. Do you currently consider yourself a snoozer using this definition?," as adapted from [34].

For the analysis of snoozer characteristics, we used the 2-item Patient Health Questionnaire (PHQ-2) [35] and the short form of the Pittsburgh Sleep Quality Index (PSQI) [10], averaging the values collected from the monthly questionnaires. Additionally, the short form of the Positive and Negative Affect Schedule (PANAS-SF) [36] was used in the initial baseline questionnaire.

Data Exclusion and Preprocessing

Sleep data were restricted to the period from July 1, 2021, to May 31, 2022. Because of our rolling recruitment process, which started in mid-June 2021 and ended in June 2022, we excluded data from June of both years. This exclusion was necessary because we lacked complete data for these months, and including partial data could have introduced bias. A standard filter, adopted from [37], was applied to remove outliers in TST (TST<3 hours and TST>13 hours). Participants with fewer than 30 recorded nights due to dropout or technical issues were excluded. For gender-related analysis, nonbinary participants (n=1) were excluded to preserve their privacy. To maintain the interpretability of the relationships between sleep patterns and the examined variables, we chose not to normalize the dependent and independent variables.

Statistical Analysis

We used a logistic regression model to examine factors predicting snoozing behavior. Using snoozing behavior as the dependent variable, and to replicate the findings from [34], we included the same set of independent variables: age, gender, step count, TST, BIG-5 personality traits (openness, conscientiousness, extraversion, agreeableness, and neuroticism), PANAS-SF, PHQ-2, PSQI, and MEQ. To further investigate the potential confounding effects of chronotype (measured by MEQ) on the relationship between personality traits and snoozing behavior, we conducted a Baron and Kenny [38] mediation analysis.

Given the nature of our data set, which included repeated sleep measurements for each participant, we used mixed effects linear models [39] to analyze how sleep patterns and their regularity evolve over time. The models included TST, MS, and the variability of TST and MS as dependent variables. For models with variability of TST and MS as dependent variables, the numerical independent variables were averaged across weekdays. We adopted a sequential modeling strategy, building 3 distinct models for each dependent variable. Model 1 included basic characteristics such as chronotype, age, gender, origin, occupation, and parenting cohabitation status (number of children in the household). Model 2 extended model 1 by adjusting for external factors such as the stringency of restrictions and day length. Finally, model 3 built on model 2 by incorporating PA metrics, including step count, midstep, and IV. This approach allows for the exploration of the unique contributions of each new set of variables beyond those accounted for in the previous model. All models included hierarchical random effects for the study participants to account for repeated measurements. The models are formulated as follows:

$$\text{Model 1: } Y_{ij} = \beta_0 + \beta_1 x_{ij1} + \beta_2 xij_2 + \cdots + \beta_2 xij_7 + \beta_8 x_{ij8} + u_j + {}_{ij}$$

$$\text{Model 2: } Y_{ij} = \beta_0 + \beta_1 x_{ij1} + \beta_2 xij_2 + \cdots + \beta_9 xij_9 + \beta_{10} x_{ij10} + u_j + {}_{ij}$$

$$\text{Model 3: } Y_{ij} = \beta_0 + \beta_1 x_{ij1} + \beta_2 xij_2 + \cdots + \beta_{11} xij_{11} + \beta_{12} x_{ij12} + \beta_{13} x_{ij13} + u_j + {}_{ij}$$

where the independent variables are x_{ij1}=age, x_{ij2}=gender, x_{ij3}=number of children, x_{ij4}=origin, x_{ij5}=occupation, x_{ij6}=MEQ, x_{ij7}=snoozer, x_{ij8}=free day, x_{ij9}=Stringency Index, x_{ij10}=day length, x_{ij11}=steps(×1000), x_{ij12}=midstep, and x_{ij13}=step entropy.

95% CIs were reported using bootstrapping. The performance of the model was compared using the likelihood ratio test (LRT) to ensure model parsimony. All statistical analyses were performed using R software (version 3.6.1; R Foundation) [40]. Linear mixed models were tested using the lme4 package [41], and *P* values for these models were calculated using the lmerTest package [42].

Results

Data Summary

In total, 112 users and 27,350 nights were included in the TST and MS analyses. The models for the variability of TST and MS used the weekday SD of both measures, which included 3682 observations. The average age of participants was 39.5 (SD 9.9) years. Of these 112 participants, 49 were academic staff and 63 were service staff. Figure 1 presents the average values of the 4 sleep metrics—TST, MS, and their corresponding SDs—for each participant included in the analysis. Figure 2 illustrates the sleep patterns over time for 2 participants: 1 with low variability and 1 with high variability in their sleep patterns.

Figure 1. TST, MS, and their SDs of participants included in the analysis. Each dot represents the participant's mean value for the corresponding metrics. MS: midsleep; TST: total sleep time.

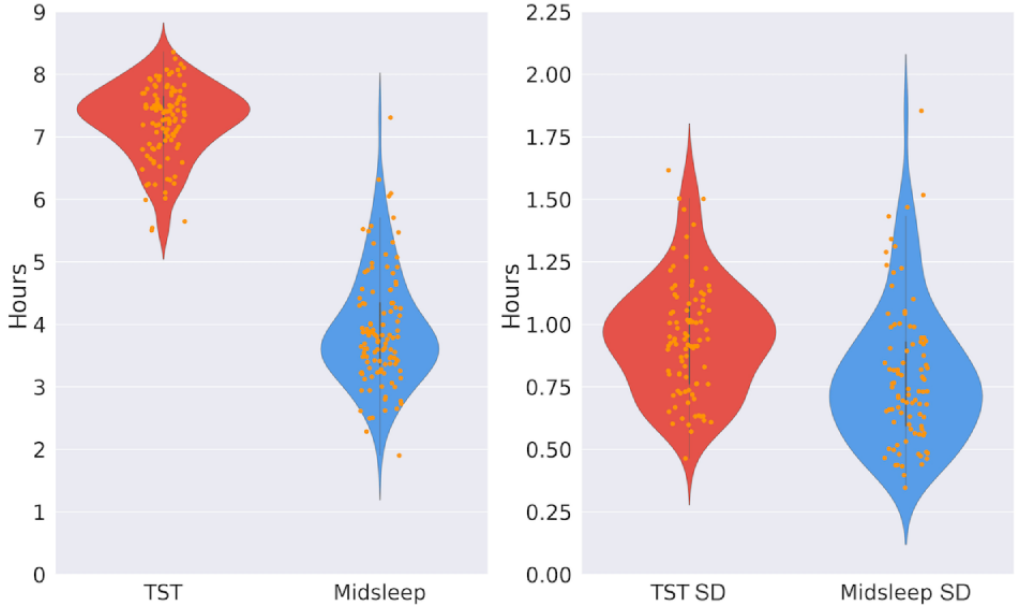
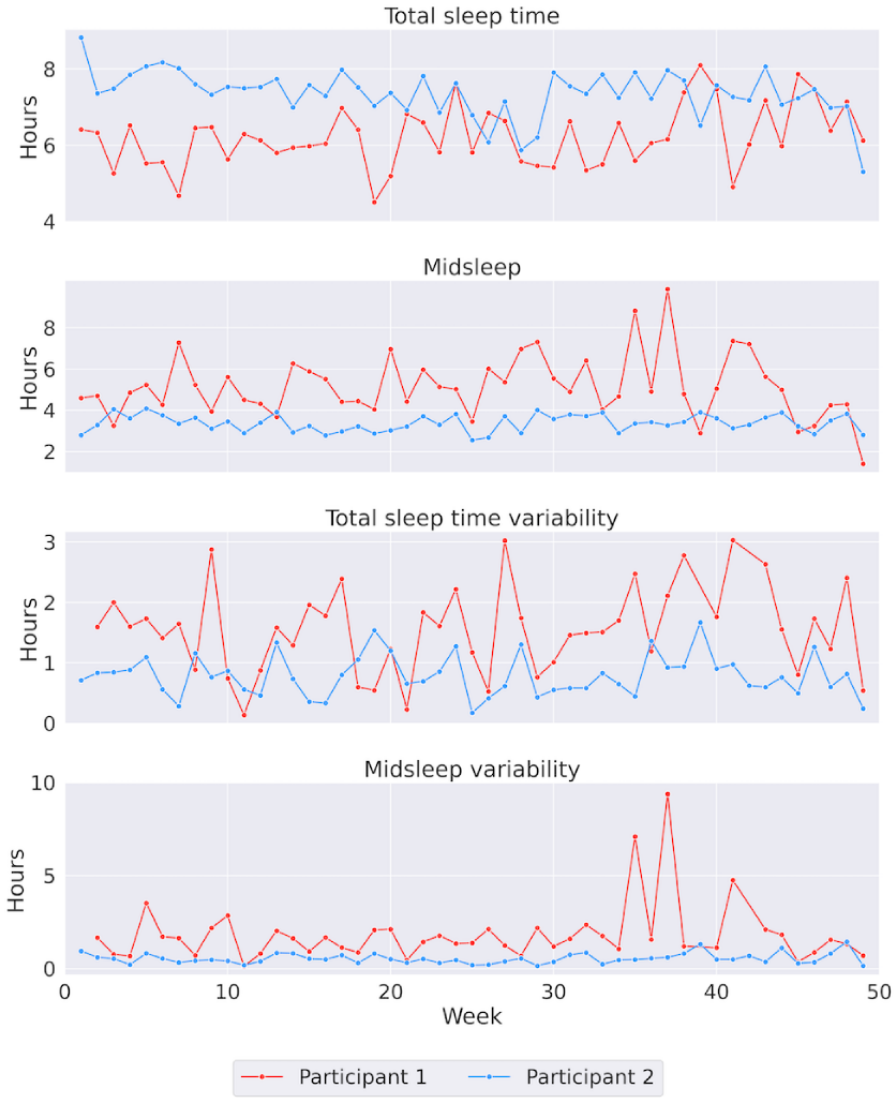


Figure 2. Sleep data over time from two participants. Participant 1 (red line) demonstrates shorter, later, and more variable sleep compared to Participant 2 (blue line).



Total Sleep Time

We begin by investigating the factors that influence TST using the 3 linear mixed models described earlier. Table 1 presents the results of these models predicting TST. For improved interpretability, the rate of change in TST is expressed as the estimate of the predictors multiplied by 60 minutes. In the full model (model 3), an increase in age by 1 year was associated with a 1.2-minute decrease in TST (95% CI –1.8 to –0.6; *P*=.008). Regarding gender, males were found to sleep 20.4 minutes less than females (95% CI –33.0 to –7.8; *P*<.001). Comparing occupations, service staff were found to sleep 22.2 minutes more than academic staff (95% CI 8.4-36.6; *P*=.004). A detailed monthly breakdown of the variations in sleep pattern measurements across different occupations is provided in Multimedia Appendix 1. After adjusting for day length and the SI, an additional hour of day length was associated with a 0.60-minute decrease in TST (95% CI –0.72 to –0.36; *P*<.001). Conversely, a 1-point increase in the SI offset this decrease by 0.18 minutes (95% CI 0.06-0.30; *P*<.001). In the full model, which included PA, a 1-unit increase in IV was associated with a 15.6-minute decrease in TST (95% CI –17.5 to –13.8; *P*<.001). Moreover, an additional hour in midstep was associated with a 1.8-minute increase in TST (95% CI 1.2-2.4; *P*<.001).

Table 1. Estimates of fixed effects from the linear mixed effects model predicting TST[a].

Predictors	Model 1[b]			Model 2[c]			Model 3[d]		
	Estimated	CI	*P* value[e]	Estimated	CI	*P* value[e]	Estimated	CI	*P* value[e]
Age	–0.02	–0.03 to –0.01	*.002*[f]	–0.02	–0.03 to –0.01	*.002*[f]	–0.02	–0.03 to –0.01	*.008*[f]
Gender (male)	–0.34	–0.55 to –0.14	*<.001*[g]	–0.34	–0.55 to –0.14	*<.001*[g]	–0.34	–0.55 to –0.13	*<.001*[g]
Number of children	0.04	–0.07 to 0.15	.51	0.04	–0.07 to 0.15	.51	0.03	–0.07 to 0.14	.56
Origin (migrant background)	0.01	–0.27 to 0.27	.97	0.01	–0.27 to 0.27	.97	0.01	–0.24 to 0.24	.93
Occupation (service)	0.36	0.11 to 0.60	*.01*[h]	0.36	0.11 to 0.60	*.01*[h]	0.37	0.14 to 0.61	*.004*[f]
MEQ	–0.01	–0.04 to 0.02	.59	–0.01	–0.04 to 0.02	.60	0.30	–0.20 to 0.22	.79
Snoozer (Yes)	–0.2	–0.44 to 0.05	.11	–0.2	–0.44 to 0.05	.11	–0.18	–0.44 to 0.05	.14
Free day (Yes)	0.10	0.07 to 0.13	*<.001*[g]	0.10	0.07 to 0.13	*<.001*[g]	0.08	0.06 to 0.11	*<.001*[g]
Stringency Index	—[i]	—	—	0.005	0.003 to 0.007	*<.001*[g]	0.003	0.001 to 0.005	*<.001*[g]
Day length	—	—	—	–0.01	–0.01 to –0.01	*<.001*[g]	–0.01	–0.012 to –0.006	*<.001*[g]
Steps (×1000)	—	—	—	—	—	—	–0.01	–0.01 to 0.01	<.001[g]
Midsteps	—	—	—	—	—	—	0.03	0.02 to 0.04	*<.001*[g]
Intradaily variability	—	—	—	—	—	—	–0.26	–0.29 to –0.23	*<.001*[g]

[a]The σ^2 values for models 1-3 were 1.13, 1.13, and 1.1, respectively. The intraclass correlation coefficient values for models 1-3 were 0.19, 0.19, and 0.19, respectively. The marginal R^2/conditional R^2 values for models 1-3 were 0.055/0.234, 0.057/0.235, and 0.069/0.245, respectively. The Akaike information criterion values for models 1-3 were 81,208.16, 81,154.98, and 80,783.26, respectively.

[b]Includes demographic and occupational variables.

[c]Includes model 1 + restriction and seasonal factors.

[d]Includes model 2 + physical activity influences.

[e]Italicized values denote significance.

[f]*P*<.01.

[g]*P*<.001.

[h]*P*<.05.

[i]Not available.

The marginal R^2 values represent the proportion of variance explained by the fixed effects, while the conditional R^2 values indicate the proportion of variance accounted for by both fixed and random effects. The increase in both R^2 values suggests that more complex models, particularly model 3, explained a greater proportion of the variance in the dependent variable. The LRT between models 1 and 2 indicated that model 2 was a significantly better fit (χ^2_2=57.17; *P*<.001). Additionally, the LRT between models 2 and 3 showed that model 3 provided a significantly improved fit (χ^2_3=377.72; *P*<.001). The performance of the full model (model 3) was further supported by the Akaike information criterion (AIC), which was lowest for model 3 (AIC 80,783.26), indicating that it offered the most optimal fit for the data.

Midsleep

Using the same approach, we developed 3 linear mixed models to assess the associations between the same set of predictors and MS. The results are presented in Table 2. To enhance interpretability, the rate of change in MS is measured as the estimate of the predictors multiplied by 60 minutes. Across all 3 models, chronotype (MEQ) (*P*<.001) and sleep on a free day (*P*<.001) consistently emerged as significant factors. In the full model (model 3), a 1-point increase in the MEQ was associated with an 8.4-minute decrease in MS (95% CI –10.8 to –5.4; *P*<.001). Sleep on a free day occurred 11.4 minutes later (95% CI 9.6-12.6; *P*<.001) compared with a workday. After adjusting for season and restriction policies, MS was delayed by 0.6 minutes (95% CI 0.6-1.2; *P*<.001) for each additional hour of day length. A 1-point increase in the SI was associated with a 1.2-minute increase in MS (95% CI 1.2-1.8; *P*<.001). In the full model, which included PA variables, a 1-unit increase in IV was linked to a 17.4-minute earlier MS (95% CI –19.8 to –15.6; *P*<.001). Similarly, an increase in step count was associated with a 0.6-minute earlier MS (95% CI –1.2 to 0.0; *P*=.04).

The LRT between models 1 and 2 indicated that model 2 was a better fit (χ^2_2=443.70; *P*<.001). Additionally, the LRT between models 2 and 3 showed that model 3 provided a significantly improved fit (χ^2_3=291.63; *P*<.001). The AIC value for model 3 was also the lowest (AIC 86,315.145), indicating that it provided the best fit for the data.

Table 2. Estimates of fixed effects from the linear mixed effects models predicting MS[a].

Predictors	Model 1[b]			Model 2[c]			Model 3[d]		
	Estimated	CI	*P* value[e]	Estimated	CI	*P* value[e]	Estimated	CI	*P* value[e]
Age	–0.01	–0.02 to 0.01	.38	–0.01	–0.02 to 0.01	.46	–0.006	–0.02 to 0.01	.56
Gender (male)	0.13	–0.16 to 0.41	.39	0.12	–0.17 to 0.40	.43	0.11	–0.19 to 0.38	.49
Number of children	–0.1[f]	–0.33 to –0.02	*.02*[f]	–0.18	–0.33 to –0.02	*.02*[f]	–0.14	–0.30 to 0.02	.08
Origin (migrant background)	0.19	–0.18 to 0.55	.32	0.18	–0.20 to 0.53	.34	0.09	–0.29 to 0.02	.63
Occupation (service)	–0.17	–0.51 to 0.17	.33	–0.18	–0.52 to 0.15	.29	–0.21	–0.55 to 0.14	.23
MEQ	–0.14	–0.18 to –0.09	*<.001*[g]	–0.14	–0.18 to –0.09	*<.001*[g]	–0.14	–0.18 to –0.09	*<.001*[g]
Snoozer (Yes)	0.27	–0.06 to 0.61	.09	0.29	–0.04 to 0.63	.08	0.32	–0.05 to 0.66	.08
Free day (Yes)	0.21	0.18 to 0.24	*<.001*[g]	0.21	0.18 to 0.24	*<.001*[g]	0.19	0.16 to 0.21	*<.001*[g]
Stringency Index	—[h]	—	—	0.02	0.02 to 0.03	*<.001*[g]	0.02	0.02 to 0.03	*<.001*[g]
Day length	—	—	—	0.00	0.00 to 0.01	*.048*[f]	0.01	0.00 to 0.01	*.002*[i]
Steps (×1000)	—	—	—	—	—	—	–0.01	–0.02 to –0.01	*<.001*[g]
Midsteps	—	—	—	—	—	—	0.00	–0.00 to 0.01	.43
Intradaily variability	—	—	—	—	—	—	–0.29	–0.33 to –0.26	*<.001*[f]

[a]The σ^2 values for models 1-3 were 1.38, 1.36, and 1.36, respectively. The intraclass correlation coefficient values for models 1-3 were 0.26, 0.26, and 0.27, respectively. The marginal R^2/conditional R^2 values for models 1-3 were 0.168/0.389, 0.178/0.400, and 0.179/0.400, respectively. The Akaike information criterion values for models 1-3 were 86,990.188, 86,573.060, and 86,315.145, respectively.

[b]Includes demographic and occupational variables.

[c]Includes model 1 + restriction and seasonal factors.

[d]Includes model 2 + physical activity influences.

[e]Italicized values denote significance.

[f]*P*<.05.

[g]*P*<.001.

[h]Not available.

[i]*P*<.01.

Total Sleep Time Variability

Table 3 presents the factors predicting the variability in TST. Across the 3 models, age (*P*=.01), number of children (*P*=.03), occupation (*P*<.001), and snoozing behavior (*P*=.006) emerged as significant factors. In the final model (model 3), each additional year of age was associated with a 0.01-unit increase in TST variability (95% CI 0.00-0.01; *P*=.01). Notably, participants with snoozing habits exhibited higher TST variability, increasing by 0.15 units (95% CI 0.05-0.27; *P*=.006). Each additional child was associated with a 0.06-unit reduction in TST variability (95% CI –0.11 to –0.00; *P*=.03). Service staff also demonstrated lower TST variability, with a reduction of 0.15 units compared with academic staff (95% CI –0.27 to –0.05; *P*<.001). When accounting for PA, a decrease of 1 hour in midsteps was correlated with a 0.01-unit increase in TST variability (95% CI –0.02 to –0.00; *P*=.03), while a 1-unit increase in IV was associated with a 0.16-unit decrease in TST variability (95% CI –0.23 to –0.09; *P*=.03). The LRT indicated that model 2 did not provide an improvement over the baseline model (χ^2_2=4.78; *P*=.09). However, model 3 demonstrated better performance compared with the baseline model (χ^2_5=31.95; *P*<.001).

Table 3. Estimates of fixed effects from the linear mixed effects model predicting TST variability[a].

Predictors	Model 1[b]			Model 2[c]			Model 3[d]		
	Estimated	CI	*P* value[e]	Estimated	CI	*P* value[e]	Estimated	CI	*P* value[e]
Age	0.01	0.00 to 0.01	*.01*[f]	0.01	0.00 to 0.01	*.01*[f]	0.01	0.00 to 0.01	*.01*[f]
Gender (male)	0.11	0.01 to 0.21	*.038*[g]	0.11	0.01 to 0.21	*.03*[g]	0.10	0.00 to 0.21	.06
Number of children	–0.05	–0.11 to 0.00	.056	–0.05	–0.11 to 0.00	.052	–0.06	–0.11 to –0.00	*.01*[g]
Origin (migrant background)	–0.03	–0.15 to 0.10	.56	–0.03	–0.15 to 0.09	.54	–0.03	–0.15 to 0.08	.56
Occupation (service)	–0.17	–0.28 to –0.05	*.004*[g]	–0.17	–0.28 to –0.05	*.004*[g]	–0.15	–0.27 to –0.05	*<.001*[h]
MEQ	0	–0.01 to 0.02	.55	0	–0.01 to 0.02	.57	0	–0.01 to 0.02	.60
Snoozer (yes)	0.18	0.07 to 0.30	*.002* [g]	0.18	0.07 to 0.30	*.002*[f]	0.15	0.05 to 0.27	*.006*[f]
Daylength	__[i]	—	—	0	–0.00 to 0.01	.16	0	–0.00 to 0.01	.08
Stringency Index	—	—	—	0	–0.00 to 0.00	.10	0	–0.00 to 0.01	.17
Steps (×1000)	—	—	—	—	—	—	–0.01	–0.01 to 0.00	.007[g]
Midsteps	—	—	—	—	—	—	–0.01	–0.02 to –0.00	*.02*[f]
Intradaily variability	—	—	—	—	—	—	–0.16	–0.23 to –0.09	*.001*[h]

[a]The σ^2 values for models 1-3 were 0.24, 0.24, and 0.24, respectively. The intraclass correlation coefficient values for models 1-3 were 0.14, 0.14, and 0.14, respectively. The marginal R^2/conditional R^2 values for models 1-3 were 0.059/0.195, 0.060/0.194, and 0.068/0.200, respectively. The Akaike information criterion values for models 1-3 were 5458.745, 5457.957, and 5436.793, respectively.

[b]Includes demographic and occupational variables.

[c]Includes model 1 + restriction and seasonal factors.

[d]Includes model 2 + physical activity influences.

[f]*P*<.05.

[g]*P*<.01.

[h]*P*<.001.

[i]Not available.

[e]Italicized values denote significance.

Midsleep Variability

Table 4 presents the factors predicting the variability of MS. Across the 3 models, the number of children (*P*=.004), snoozing behavior (*P*=.01), midsteps (*P*=.008), and IV (*P*=.001) emerged as significant factors. For each additional child, MS variability was reduced by 0.10 units (95% CI –0.16 to –0.03; *P*=.004). In all models, being a snoozer correlated with increased MS variability. Specifically, snoozers experienced a 0.17-unit increase in MS variability compared with nonsnoozers (95% CI 0.03-0.31; *P*=.01). To better understand the characteristics of snoozers, we conducted an analysis based on Mattingly et al's study [34]. Interestingly, our results revealed that age (*P*=.02) and chronotype (*P*=.002) were significant factors in predicting snoozing behavior. The full results are detailed in Multimedia Appendix 2.

In the full model, including PA variables, midsteps also became significant. Each hour increase in midsteps was associated with a 0.02-unit decrease in MS variability (95% CI –0.04 to –0.00; *P*=.008). However, the more complex models did not show a significant improvement over the baseline model, as indicated by the LRT (model 2: χ^2_2=1.00; *P*=.60/model 3: χ^2_5=10.17; *P*=.07).

Table 4. Estimates of fixed effects from the linear mixed effects model predicting MS variability[a].

Predictors	Model 1[b]			Model 2[c]			Model 3[d]		
	Estimated	CI	*P* value[e]	Estimated	CI	*P* value[e]	Estimated	CI	*P* value[e]
Age	0.00	–0.00 to 0.01	.41	0.00	–0.00 to 0.01	.41	0.00	–0.00 to 0.01	.55
Gender (male)	0.10	–0.03 to 0.23	.12	0.10	–0.03 to 0.23	.12	0.09	–0.04 to 0.22	.17
Number of children	–0.09	–0.16 to –0.02	*.01*[f]	–0.09[f]	–0.16 to –0.02	*.01*[f]	–0.10	–0.16 to –0.03	*.004*[g]
Origin (migrant background)	–0.05	–0.19 to 0.11	.49	–0.05	–0.19 to 0.11	.49	–0.05	–0.20 to 0.09	.45
Occupation (service)	–0.12	–0.26 to 0.02	.11	–0.12	–0.25 to 0.02	.11	–0.11	–0.25 to 0.02	.10
MEQ	0.00	–0.02 to 0.02	.96	0.00	–0.02 to 0.02	.98	0.00	–0.02 to 0.02	.94
Snoozer (yes)	0.20	0.06 to 0.35	*.006*[g]	0.20	0.06 to 0.35	*.006*[g]	0.17	0.03 to 0.31	*.01*[f]
Daylength	—[h]	—	—	0.00	–0.00 to 0.01	.34	0.00	–0.00 to 0.01	.26
Stringency Index	—	—	—	0.00	–0.00 to 0.00	.94	0.00	–0.00 to 0.00	.89
Steps (×1000)	—	—	—	—	—	—	0.00	–0.01 to 0.01	.41
Midsteps	—	—	—	—	—	—	–0.02	–0.03 to –0.00	*.008*[g]
Intradaily variability	—	—	—	—	—	—	–0.09	–0.21 to –02	09

[a]The σ^2 values for models 1-3 were 0.59, 0.59, and 0.59, respectively. The intraclass correlation coefficient values for models 1-3 were 0.09, 0.09, and 0.09, respectively. The marginal R^2/conditional R^2 values for models 1-3 were 0.034/0.120, 0.034/0.120, and 0.038/0.122, respectively. The Akaike information criterion values for models 1-3 were 8679.371, 8682.369, and 8679.197, respectively.

[b]Includes demographic and occupational variables.

[c]Includes model 1 + restriction and seasonal factors.

[d]Includes model 2 + physical activity influences.

[e]Italicized values denote significance.

[f]$P<.05$.

[g]$P<.01$.

[h]Not available.

Discussion

Principal Findings

In this study, we used a year-long longitudinal data set from 112 working adults and identified several significant relationships between changes in sleep over time and various factors, including restriction policies, seasonal changes, PA, and sociodemographics. We found that more stringent restrictions were associated with increased TST and delayed MS. Additionally, seasonal factors played a notable role: increased day length was linked to reduced TST and delayed MS. Changes in work arrangements, particularly the shift to remote work, directly impacted individuals based on their occupations and sleep patterns. Academic personnel, with more flexible schedules, slept less and exhibited greater variability in TST compared with service personnel, who had more structured work schedules. Additionally, individuals identified as “snoozers” had more flexible sleep schedules with greater variability in both TST and MS compared with nonsnoozers. Moreover, activity patterns played a significant role: exercising later in the day was associated with longer TST and reduced variability in both TST and MS. To contextualize our findings within the broader scope of sleep during the pandemic, the following section details our results and compares them with previous studies.

Demographic Factors

Previous research has highlighted several epidemiological factors affecting sleep patterns, notably, age, gender, and chronotype. Consistent with previous studies, we found that older individuals tend to sleep less [43,44]. However, our findings reveal a correlation between older age and increased TST variability, which contrasts with prior results [45]. The

variance in the observed correlations may be due to our study using objective sleep measures, while [45] relied on self-reported data. Additionally, we found no significant association between MS variability and age. Regarding gender differences, our study shows that males tend to have shorter and less consistent TST compared with females. While the shorter TST among males is well-documented [46,47], evidence regarding gender disparity in TST variability is inconsistent. For instance, an actigraphy study on a middle-aged cohort found that females exhibited greater TST variability than males [48]. Conversely, a survey-based study on university students [49] reported no gender differences in TST variability. Additionally, our study observes that parental duties significantly impact sleep patterns. Parents typically exhibited earlier sleep times and more consistent TST and MS than nonparents. The underlying reasons for these observations remain uncertain, but one hypothesis is that parents' sleep/wake schedules are more stable due to the need to synchronize their sleep patterns with those of their children. While the specific relationship between parenting and sleep pattern variability has not been extensively studied, research on cohabitation suggests that living with others can influence sleep patterns by reducing variability in sleep timing and duration [50,51]. This context highlights how factors related to shared living arrangements, such as parenting, can contribute to greater sleep pattern regularity.

Snoozing Behavior

We observed higher variability in TST and MS among individuals identified as "snoozers." Interestingly, younger individuals and those with an evening chronotype are more likely to be "snoozers," suggesting an interplay between age, chronotype, and snoozing habits. The natural sleep-wake patterns associated with an individual's chronotype may influence their tendency to snooze alarms. Morning types, who wake up earlier, might not feel the need to snooze as much because their schedules align better with societal norms, in contrast to evening types.

Clinically, snoozing can be linked to prolonged sleep inertia, a state of reduced alertness upon waking [52]. Morning types (with high MEQ scores) may be less prone to snoozing and thus avoid significant sleep inertia, potentially leading to better alertness and performance. Conversely, evening types who snooze might experience greater sleep inertia, which could present additional challenges, such as managing increased work demands.

Occupational Factors

We found that academic staff have shorter and more variable TST compared with service staff, and also exhibit greater variability in MS. The flexible and deadline-driven nature of academic schedules may contribute significantly to these irregular sleep patterns. As academics frequently adjust their schedules to meet project deadlines or prepare lectures, the dynamic nature of their workload can disrupt regular sleep patterns. Additionally, the intellectual and creative demands of academic work often extend beyond the traditional 9-5 workday, further contributing to irregular sleep schedules.

Nonetheless, it is noteworthy that increased variability in sleep patterns might impact overall health and well-being. For instance, studies using actigraphy have found that higher TST variability is associated with an increase in depressive symptoms [53,54]. These implications become even more significant in the context of the COVID-19 pandemic. The shift to remote working and learning may introduce greater flexibility for academic personnel. Although this flexibility allows for more control over schedules, it may also blur the boundaries between work and personal life, potentially leading to longer work hours and more irregular sleep patterns.

Restriction Policies

The influence of lockdown measures during the pandemic on sleep patterns is well documented, with increased TST and later MS observed during lockdown periods [18,21,22]. Our findings further reinforce previous evidence at a more granular scale. In a more detailed analysis using the SI to measure lockdown severity, Ong et al [55] found that a higher SI was correlated with later and more variable MS. Contrary to Ong et al's findings, our study did not find a correlation between the SI and the variability of MS. However, it is important to note the methodological differences between our studies: while Ong et al [55] conducted their correlation measurements on a monthly basis, our analysis was performed at a weekly level. These differences in granularity may account for the contrasting results.

Although not closely examined in this study, we postulate that the side effects of restriction policies might significantly impact sleep. Prolonged periods of staying at home could induce stress, potentially increasing the prevalence of insomnia [56]. Furthermore, loneliness due to self-isolation could further worsen sleep quality [57]. Despite these adverse effects, restriction policies have also had positive aspects. The shift to remote work persists, as postpandemic workplace policies increasingly encourage hybrid and remote work [58]. This change allows for greater flexibility in daily schedules, potentially leading to improved and longer sleep.

Seasonality

Seasonal factors, such as day length, have been shown to influence sleep patterns, including sleep duration and timing [59,60]. Longer daylight hours during summer may encourage longer waking periods, while shorter days in winter can disrupt melatonin production, potentially leading to extended sleep duration. Additionally, these seasonal shifts align with changes in social schedules, such as holidays, which can further affect regular sleep routines. In southern Finland, where day length can vary by up to 13 hours between summer and winter, these influences might be more pronounced. It is possible that reduced exposure to natural daylight, due to limited mobility during the pandemic, could have altered the effect of day length on sleep.

Physical Activity

The connection between PA and sleep has been extensively studied [61-63]. While regular PA is generally recommended for promoting good sleep, it is crucial to recognize that PA is a multifaceted behavior with various elements—such as duration, timing, and intensity—that can each influence sleep differently [61]. Therefore, investigations into the relationship

between sleep and PA should consider these diverse aspects of PA.

When considering the timing of PA and its effect on sleep, our findings indicate that engaging in PA later in the day is associated with longer TST and reduced variability in both TST and MS. This supports previous research, such as a review by Youngstedt et al [63], which suggested that exercising later in the day can be beneficial for sleep. Similarly, a survey study found that engaging in light- to moderate-intensity workouts early in the evening may have beneficial effects on sleep [64]. The impact of PA's intensity on sleep could potentially modify the effects of its timing. Sleep hygiene guidelines suggest that vigorous exercise late at night may increase arousal and subsequently impair sleep quality [65]. However, recent research challenges this convention. For instance, Myllymäki et al [66] conducted a study under controlled laboratory conditions and found that exercise performed 4 hours before bedtime did not disturb sleep. Furthermore, a review by Stutz et al [67] suggested that evening exercise does not necessarily adversely impact sleep, although exercising less than an hour before bedtime could potentially disrupt sleep.

In addition to the volume and timing of PA, we found that the fragmentation of activity rhythms, measured by IV, significantly predicted sleep patterns. Our finding of a negative association between IV and TST reinforces previous research [68], which suggests that greater fragmentation in daily PA is linked to shorter sleep duration. Additionally, the novel associations between IV and MS, as well as the variability of TST, contribute new insights into the study of activity rhythms and sleep patterns.

By leveraging longitudinal data from fitness trackers, our study highlights the potential of mHealth to offer deeper insights into behavioral health patterns, especially regarding how lifestyle changes during the pandemic have impacted sleep. This integration of mHealth approaches in sleep research exemplifies how technological advancements can enhance our understanding and interventions in public health.

Limitations

This work has several unavoidable limitations. First, the absence of baseline data from the prepandemic period limits our study, preventing a comparison of sleep patterns and quality before and during the later stages of the pandemic. Second, the study was conducted among university staff, leading to a nonrepresentative sample that may introduce bias and result in a limited sample size. The relatively small sample size may have contributed to the wide CIs observed, indicating that the precision of our estimates could be improved. Consequently, our findings should be interpreted with caution, especially when generalizing to a broader population. Third, although we attempted to control for all known factors affecting sleep, there may still be unaddressed confounding variables. Fourth, we used consumer-grade wearables for data collection, which, despite their accessibility, may not provide the same accuracy and reliability as professional-grade equipment. Fifth, recall bias in self-reported measures is an inherent challenge. However, we addressed this issue by using validated questionnaires and conducting monthly data collection to minimize recall intervals. Finally, the study's geographical limitation restricts the generalizability of our findings to other cultural or social contexts.

Future Directions

One possible future research direction is to further investigate the relationship between snoozing behavior and specific demographics, such as age, to identify potential causative factors. For example, a case-control study could be conducted to compare individuals who frequently snooze with those who rarely or never do, across various age groups. This approach would enable a detailed examination of how snoozing behavior varies with age, while controlling for potential confounding variables.

Conclusions

Our study, through a holistic approach, provided insights into the changes in sleep patterns and PA levels among working adults during the late stages of the COVID-19 pandemic. The flexible working hours during the pandemic led to corresponding flexibility in sleep patterns in certain occupations and sleep traits, particularly among individuals who self-identified as snoozers. Our findings underscore the significant impact of lifestyle habits on sleep health, particularly during unprecedented times like a global pandemic. Moving forward, it is essential to further investigate changes in sleep patterns across diverse populations. Such research will help inform workplace policies in the postpandemic era, considering the potential benefits and challenges of remote work. One notable advantage to consider is the increased amount of sleep that workers may experience, potentially enhancing overall efficiency and productivity. As we navigate the future of work, understanding the interplay between work arrangements, lifestyle choices, and sleep quality will be essential for promoting optimal well-being and performance in the workforce.

Acknowledgments

TA acknowledges the support of the MAGICS infrastructure for running the cor:ona study. We acknowledge the computational resources provided by the Aalto Science-IT project.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Additional analyses.
[DOCX File , 8 KB - mhealth_v12i1e53389_app1.docx]

Multimedia Appendix 2
Supplementary Tables S1 and S2.
[DOCX File , 10 KB - mhealth_v12i1e53389_app2.docx]

References

1. Surridge-David M, MacLean A, Coulter ME, Knowles JB. Mood change following an acute delay of sleep. Psychiatry Res 1987 Oct;22(2):149-158. [doi: 10.1016/0165-1781(87)90102-8] [Medline: 3685222]
2. Benca RM, Okawa M, Uchiyama M, Ozaki S, Nakajima T, Shibui K, et al. Sleep and mood disorders. Sleep Med Rev 1997 Nov;1(1):45-56. [doi: 10.1016/s1087-0792(97)90005-8] [Medline: 15310523]
3. Benham G. Sleep: an important factor in stress - health models. Stress and Health 2010 Aug 05;26(3):204-214. [doi: 10.1002/smi.1304]
4. Hafner M, Stepanek M, Taylor J, Troxel W, van SC. Why sleep matters? The economic costs of insufficient sleep. Rand Health Q 2017;6(4):11. [doi: 10.7249/rb9962]
5. Baldwin DC, Daugherty SR. Sleep deprivation and fatigue in residency training: results of a national survey of first- and second-year residents. Sleep 2004 Mar 15;27(2):217-223. [doi: 10.1093/sleep/27.2.217] [Medline: 15124713]
6. Wright K, McHill A, Birks B, Griffin B, Rusterholz T, Chinoy E. Entrainment of the human circadian clock to the natural light-dark cycle. Current Biology 2013 Aug 19;23(16):1554-1558 [FREE Full text] [doi: 10.1016/j.cub.2013.06.039] [Medline: 23910656]
7. Kredlow MA, Capozzoli MC, Hearon BA, Calkins AW, Otto MW. The effects of physical activity on sleep: a meta-analytic review. J Behav Med 2015 Jun;38(3):427-449. [doi: 10.1007/s10865-015-9617-6] [Medline: 25596964]
8. Kamstra MJ, Kramer LA, Levi MD. Losing sleep at the market: the daylight saving anomaly. American Economic Review 2000 Sep 01;90(4):1005-1011. [doi: 10.1257/aer.90.4.1005]
9. Akerstedt T, Hume K, Minors D, Waterhouse J. The subjective meaning of good sleep, an intraindividual approach using the Karolinska Sleep Diary. Percept Mot Skills 1994 Aug 04;79(1 Pt 1):287-296. [doi: 10.2466/pms.1994.79.1.287] [Medline: 7991323]
10. Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res 1989 May;28(2):193-213. [doi: 10.1016/0165-1781(89)90047-4] [Medline: 2748771]
11. Aledavood T, Lehmann S, Saramäki J. Social network differences of chronotypes identified from mobile phone data. EPJ Data Sci 2018 Oct 24;7(1):46-46. [doi: 10.1140/epjds/s13688-018-0174-4]
12. Aledavood T, Kivimäki I, Lehmann S, Saramäki J. Quantifying daily rhythms with non-negative matrix factorization applied to mobile phone data. Sci Rep 2022 Apr 01;12(1):5544 [FREE Full text] [doi: 10.1038/s41598-022-09273-y] [Medline: 35365710]
13. Breideband T, Martinez GJ, Talkad Sukumar P, Caruso M, D'Mello S, Striegel AD, et al. Sleep patterns and sleep alignment in remote teams during COVID-19. Proc ACM Hum-Comput Interact 2022 Nov 11;6(CSCW2):1-31. [doi: 10.1145/3555217]
14. Grossi NR, Batinic B, Moharitsch S. Sleep and health: examining the relation of sleep to burnout and well-being using a consumer fitness tracker. Health Technol 2021 Oct 05;11(6):1247-1257. [doi: 10.1007/s12553-021-00603-0]
15. Aledavood T, Torous J, Triana Hoyos AM, Naslund JA, Onnela J, Keshavan M. Smartphone-based tracking of sleep in depression, anxiety, and psychotic disorders. Curr Psychiatry Rep 2019 Jun 4;21(7):49-49 [FREE Full text] [doi: 10.1007/s11920-019-1043-y] [Medline: 31161412]
16. Chinoy ED, Cuellar JA, Huwa KE, Jameson JT, Watson CH, Bessman SC, et al. Performance of seven consumer sleep-tracking devices compared with polysomnography. Sleep 2021 May 14;44(5):zsaa291 [FREE Full text] [doi: 10.1093/sleep/zsaa291] [Medline: 33378539]
17. Miller DJ, Sargent C, Roach GD. A validation of six wearable devices for estimating sleep, heart rate and heart rate variability in healthy adults. Sensors (Basel) 2022 Aug 22;22(16):6317 [FREE Full text] [doi: 10.3390/s22166317] [Medline: 36016077]
18. Sinha M, Pande B, Sinha R. Impact of COVID-19 lockdown on sleep-wake schedule and associated lifestyle related behavior: a national survey. J Public Health Res 2020 Jul 28;9(3):1826 [FREE Full text] [doi: 10.4081/jphr.2020.1826] [Medline: 32874967]
19. Lee P, Marek J, Nálevka P. Sleep pattern in the US and 16 European countries during the COVID-19 outbreak using crowdsourced smartphone data. Eur J Public Health 2021 Feb 01;31(1):23-30 [FREE Full text] [doi: 10.1093/eurpub/ckaa208] [Medline: 33169169]
20. Capodilupo ER, Miller DJ. Changes in health promoting behavior during COVID-19 physical distancing: utilizing wearable technology to examine trends in sleep, activity, and cardiovascular indicators of health. PLoS One 2021 Aug 12;16(8):e0256063 [FREE Full text] [doi: 10.1371/journal.pone.0256063] [Medline: 34383837]

21. Rezaei N, Grandner MA. Changes in sleep duration, timing, and variability during the COVID-19 pandemic: large-scale Fitbit data from 6 major US cities. Sleep Health 2021 Jun;7(3):303-313 [FREE Full text] [doi: 10.1016/j.sleh.2021.02.008] [Medline: 33771534]
22. Cellini N, Canale N, Mioni G, Costa S. Changes in sleep pattern, sense of time and digital media use during COVID-19 lockdown in Italy. J Sleep Res 2020 Aug;29(4):e13074 [FREE Full text] [doi: 10.1111/jsr.13074] [Medline: 32410272]
23. Diniz T, Christofaro D, Tebar W, Cucato GG, Botero JP, Correia MA, et al. Reduction of physical activity levels during the COVID-19 pandemic might negatively disturb sleep pattern. Front Psychol 2020;11:586157 [FREE Full text] [doi: 10.3389/fpsyg.2020.586157] [Medline: 33424702]
24. Escobar-Córdoba F, Ramírez-Ortiz J, Fontecha-Hernández J. Effects of social isolation on sleep during the COVID-19 pandemic. Sleep Sci 2021 Nov 30;14(Spec 1):86-93 [FREE Full text] [doi: 10.5935/1984-0063.20200097] [Medline: 34917279]
25. Hulsegge G, Loef B, van Kerkhof LW, Roenneberg T, van der Beek AJ, Proper KI. Shift work, sleep disturbances and social jetlag in healthcare workers. J Sleep Res 2019 Aug 05;28(4):e12802. [doi: 10.1111/jsr.12802] [Medline: 30520209]
26. Ganesan S, Magee M, Stone JE, Mulhall MD, Collins A, Howard ME, et al. The impact of shift work on sleep, alertness and performance in healthcare workers. Sci Rep 2019 Mar 15;9(1):4635 [FREE Full text] [doi: 10.1038/s41598-019-40914-x] [Medline: 30874565]
27. Luong N, Barnett I, Aledavood T. The impact of the COVID-19 pandemic on daily rhythms. J Am Med Inform Assoc 2023 Nov 17;30(12):1943-1953 [FREE Full text] [doi: 10.1093/jamia/ocad140] [Medline: 37550242]
28. Ikäheimonen A, Triana AM, Luong N, Ziaei A, Rantaharju J, Darst R, et al. Niimpy: a toolbox for behavioral data analysis. SoftwareX 2023 Jul;23:101472. [doi: 10.1016/j.softx.2023.101472]
29. Witting W, Kwa I, Eikelenboom P, Mirmiran M, Swaab D. Alterations in the circadian rest-activity rhythm in aging and Alzheimer's disease. Biol Psychiatry 1990 Mar 15;27(6):563-572. [doi: 10.1016/0006-3223(90)90523-5] [Medline: 2322616]
30. Premium Weather API. World Weather Online. URL: https://www.worldweatheronline.com/weather-api/api/ [accessed 2024-05-15]
31. Friborg O, Bjorvatn B, Amponsah B, Pallesen S. Associations between seasonal variations in day length (photoperiod), sleep timing, sleep quality and mood: a comparison between Ghana (5°) and Norway (69°). Journal of Sleep Research 2011 Nov 10;21(2):176-184. [doi: 10.1111/j.1365-2869.2011.00982.x] [Medline: 22074234]
32. Hale T, Angrist N, Goldszmidt R, Kira B, Petherick A, Phillips T, et al. A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker). Nat Hum Behav 2021 Apr 08;5(4):529-538. [doi: 10.1038/s41562-021-01079-8] [Medline: 33686204]
33. Adan A, Almirall H. Personal Individ Differ 1991;12(3):241-253. [doi: 10.1016/0191-8869(91)90110-w]
34. Mattingly SM, Martinez G, Young J, Cain MK, Striegel A. Snoozing: an examination of a common method of waking. Sleep 2022 Oct 10;45(10):zsac184. [doi: 10.1093/sleep/zsac184] [Medline: 35951011]
35. Kroenke K, Spitzer RL, Williams JBW. The Patient Health Questionnaire-2: validity of a two-item depression screener. Med Care 2003 Nov;41(11):1284-1292. [doi: 10.1097/01.MLR.0000093487.78664.3C] [Medline: 14583691]
36. Thompson ER. Development and validation of an internationally reliable Short-Form of the Positive and Negative Affect Schedule (PANAS). Journal of Cross-Cultural Psychology 2016 Jul 26;38(2):227-242. [doi: 10.1177/0022022106297301]
37. Roenneberg T, Allebrandt KV, Merrow M, Vetter C. Social jetlag and obesity. Curr Biol 2012 May 22;22(10):939-943 [FREE Full text] [doi: 10.1016/j.cub.2012.03.038] [Medline: 22578422]
38. Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. J Pers Soc Psychol 1986 Dec;51(6):1173-1182. [doi: 10.1037//0022-3514.51.6.1173] [Medline: 3806354]
39. Raudenbush S, Bryk AS. Hierarchical Linear Models: Applications and Data Analysis Methods. Thousand Oaks, CA: Sage Publications; 2002.
40. R: a language and environment for statistical computing. R Project/R Foundation. URL: http://www.r-project.org/ [accessed 2024-05-15]
41. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using. J. Stat. Soft 2015;67(1):1-48. [doi: 10.18637/jss.v067.i01]
42. Kuznetsova A, Brockhoff PB, Christensen RHB. Package: tests in linear mixed effects models. J Stat Soft 2017;82(13):1-26. [doi: 10.18637/jss.v082.i13]
43. Ohayon MM, Carskadon MA, Guilleminault C, Vitiello MV. Meta-analysis of quantitative sleep parameters from childhood to old age in healthy individuals: developing normative sleep values across the human lifespan. Sleep 2004 Nov 01;27(7):1255-1273. [doi: 10.1093/sleep/27.7.1255] [Medline: 15586779]
44. Jonasdottir SS, Minor K, Lehmann S. Gender differences in nighttime sleep patterns and variability across the adult lifespan: a global-scale wearables study. Sleep 2021 Feb 12;44(2):zsaa169. [doi: 10.1093/sleep/zsaa169] [Medline: 32886772]
45. Dillon HR, Lichstein KL, Dautovich ND, Taylor DJ, Riedel BW, Bush AJ. Variability in self-reported normal sleep across the adult age span. J Gerontol B Psychol Sci Soc Sci 2015 Jan 13;70(1):46-56 [FREE Full text] [doi: 10.1093/geronb/gbu035] [Medline: 24829303]

46. Ong JL, Tandi J, Patanaik A, Lo JC, Chee MWL. Large-scale data from wearables reveal regional disparities in sleep patterns that persist across age and sex. Sci Rep 2019 Mar 04;9(1):3415 [FREE Full text] [doi: 10.1038/s41598-019-40156-x] [Medline: 30833649]
47. Tonetti L, Fabbri M, Natale V. Sex difference in sleep‐time preference and sleep need: a cross‐sectional survey among Italian pre‐adolescents, adolescents, and adults. Chronobiology International 2009 Jul 07;25(5):745-759. [doi: 10.1080/07420520802394191] [Medline: 18780201]
48. Mezick EJ, Matthews KA, Hall M, Kamarck TW, Buysse DJ, Owens JF, et al. Intra-individual variability in sleep duration and fragmentation: associations with stress. Psychoneuroendocrinology 2009 Oct;34(9):1346-1354 [FREE Full text] [doi: 10.1016/j.psyneuen.2009.04.005] [Medline: 19450933]
49. Roane BM, Seifer R, Sharkey KM, Van Reen E, Bond TLY, Raffray T, et al. What role does sleep play in weight gain in the first semester of university? Behav Sleep Med 2015 Aug 12;13(6):491-505 [FREE Full text] [doi: 10.1080/15402002.2014.940109] [Medline: 25115969]
50. Minors D, Atkinson G, Bent N, Rabbitt P, Waterhouse J. The effects of age upon some aspects of lifestyle and implications for studies on circadian rhythmicity. Age Ageing 1998 Jan;27(1):67-72. [doi: 10.1093/ageing/27.1.67] [Medline: 9504369]
51. Bliwise DL, Ansari FP, Straight L, Parker KP. Age changes in timing and 24-hour distribution of self-reported sleep. The American Journal of Geriatric Psychiatry 2005 Dec;13(12):1077-1082. [doi: 10.1097/00019442-200512000-00007]
52. Ogawa K, Kaizuma-Ueyama E, Hayashi M. Effects of using a snooze alarm on sleep inertia after morning awakening. J Physiol Anthropol 2022 Dec 31;41(1):43 [FREE Full text] [doi: 10.1186/s40101-022-00317-w] [Medline: 36587230]
53. Michael Vanderlind W, Beevers CG, Sherman SM, Trujillo LT, McGeary JE, Matthews MD, et al. Sleep and sadness: exploring the relation among sleep, cognitive control, and depressive symptoms in young adults. Sleep Medicine 2014 Jan;15(1):144-149. [doi: 10.1016/j.sleep.2013.10.006]
54. Lemola S, Ledermann T, Friedman EM. Variability of sleep duration is related to subjective sleep quality and subjective well-being: an actigraphy study. PLoS One 2013 Aug 14;8(8):e71292 [FREE Full text] [doi: 10.1371/journal.pone.0071292] [Medline: 23967186]
55. Ong JL, Lau T, Karsikas M, Kinnunen H, Chee MWL. A longitudinal analysis of COVID-19 lockdown stringency on sleep and resting heart rate measures across 20 countries. Sci Rep 2021 Jul 13;11(1):14413 [FREE Full text] [doi: 10.1038/s41598-021-93924-z] [Medline: 34257380]
56. Lin S, Lan Y, Hsia P, Kao CM, Tsou H, Lin Y. Internet searches for "insomnia" and "suicide" mediated by stay-at-home behaviors in 45 countries during the first 12 months of the COVID-19 pandemic. J Affect Disord 2023 Mar 15;325:119-126 [FREE Full text] [doi: 10.1016/j.jad.2022.12.139] [Medline: 36621674]
57. Benson JA, McSorley VE, Hawkley LC, Lauderdale DS. Associations of loneliness and social isolation with actigraph and self-reported sleep quality in a national sample of older adults. Sleep 2021 Jan 21;44(1):zsaa140 [FREE Full text] [doi: 10.1093/sleep/zsaa140] [Medline: 32691067]
58. Peters SE, Dennerlein JT, Wagner GR, Sorensen G. Work and worker health in the post-pandemic world: a public health perspective. The Lancet Public Health 2022 Feb;7(2):e188-e194. [doi: 10.1016/s2468-2667(21)00259-0]
59. Lehnkering H, Siegmund R. Influence of chronotype, season, and sex of subject on sleep behavior of young adults. Chronobiol Int 2007 Jul 07;24(5):875-888. [doi: 10.1080/07420520701648259] [Medline: 17994343]
60. Mattingly SM, Grover T, Martinez GJ, Aledavood T, Robles-Granda P, Nies K, et al. The effects of seasons and weather on sleep patterns measured through longitudinal multimodal sensing. npj Digit. Med 2021 Apr 28;4(1):76. [doi: 10.1038/s41746-021-00435-2]
61. Irish LA, Kline CE, Gunn HE, Buysse DJ, Hall MH. The role of sleep hygiene in promoting public health: a review of empirical evidence. Sleep Med Rev 2015 Aug;22:23-36. [doi: 10.1016/j.smrv.2014.10.001] [Medline: 25454674]
62. Kubitz KA, Landers DM, Petruzzello SJ, Han M. The effects of acute and chronic exercise on sleep. A meta-analytic review. Sports Med 1996 Apr;21(4):277-291. [doi: 10.2165/00007256-199621040-00004] [Medline: 8726346]
63. Youngstedt SD, O'Connor PJ, Dishman RK. The effects of acute exercise on sleep: a quantitative synthesis. Sleep 1997 Mar;20(3):203-214. [doi: 10.1093/sleep/20.3.203] [Medline: 9178916]
64. Vuori I, Urponen H, Hasan J, Partinen M. Epidemiology of exercise effects on sleep. Acta Physiol Scand Suppl 1988;574:3-7. [Medline: 3245463]
65. American Academy of Sleep Medicine. The international classification of sleep disorders: diagnostic and coding manual. Ann Intern Med 1991 Sep 01;115(5):413-413. [doi: 10.7326/0003-4819-115-5-413_1]
66. Myllymäki T, Kyröläinen H, Savolainen K, Hokka L, Jakonen R, Juuti T, et al. Effects of vigorous late-night exercise on sleep quality and cardiac autonomic activity. J Sleep Res 2011 Mar;20(1 Pt 2):146-153. [doi: 10.1111/j.1365-2869.2010.00874.x] [Medline: 20673290]
67. Stutz J, Eiholzer R, Spengler CM. Effects of evening exercise on sleep in healthy participants: a systematic review and meta-analysis. Sports Med 2019 Feb 29;49(2):269-287. [doi: 10.1007/s40279-018-1015-0] [Medline: 30374942]
68. Luik AI, Zuurbier LA, Hofman A, Van Someren EJW, Tiemeier H. Stability and fragmentation of the activity rhythm across the sleep-wake cycle: the importance of age, lifestyle, and mental health. Chronobiology International 2013 Aug 23;30(10):1223-1230. [doi: 10.3109/07420528.2013.813528] [Medline: 23971909]

Abbreviations

AIC: Akaike information criterion
IV: intradaily variability
LRT: likelihood ratio test
MEQ: Morningness-Eveningness Questionnaire
MS: midsleep
PA: physical activity
PANAS-SF: short form of the Positive and Negative Affect Schedule
PHQ-2: 2-item Patient Health Questionnaire
PSG: polysomnography
PSQI: Pittsburgh Sleep Quality Index
SI: Stringency Index
TST: total sleep time

Edited by L Buis; submitted 05.10.23; peer-reviewed by S Šćepanović, S Mary, S Mitra; comments to author 24.02.24; revised version received 19.04.24; accepted 08.07.24; published 03.09.24.

Please cite as:
Luong N, Mark G, Kulshrestha J, Aledavood T
Sleep During the COVID-19 Pandemic: Longitudinal Observational Study Combining Multisensor Data With Questionnaires
JMIR Mhealth Uhealth 2024;12:e53389
URL: https://mhealth.jmir.org/2024/1/e53389
doi:10.2196/53389
PMID:39226100

Original Paper

Controlled and Real-Life Investigation of Optical Tracking Sensors in Smart Glasses for Monitoring Eating Behavior Using Deep Learning: Cross-Sectional Study

Simon Stankoski[1], MSc; Ivana Kiprijanovska[1], MSc; Martin Gjoreski[2], PhD; Filip Panchevski[1], BSc; Borjan Sazdov[1], BSc; Bojan Sofronievski[1], MSc; Andrew Cleal[1], BSc; Mohsen Fatoorechi[1], PhD; Charles Nduka[1], MBBS, MA, MD; Hristijan Gjoreski[1,3], PhD

[1]Emteq Ltd., Brighton, United Kingdom

[2]Faculty of Informatics, Università della Svizzera Italiana, Lugano, Switzerland

[3]Faculty of Electrical Engineering and Information Technologies, Saints Cyril and Methodius University in Skopje, Skopje, the Former Yugoslav Republic of Macedonia

Corresponding Author:
Simon Stankoski, MSc
Emteq Ltd.
Science Park Square
Brighton, BN1 9SB
United Kingdom
Phone: 44 1273 769251
Email: simon.stankoski@emteqlabs.com

Abstract

Background: The increasing prevalence of obesity necessitates innovative approaches to better understand this health crisis, particularly given its strong connection to chronic diseases such as diabetes, cancer, and cardiovascular conditions. Monitoring dietary behavior is crucial for designing effective interventions that help decrease obesity prevalence and promote healthy lifestyles. However, traditional dietary tracking methods are limited by participant burden and recall bias. Exploring microlevel eating activities, such as meal duration and chewing frequency, in addition to eating episodes, is crucial due to their substantial relation to obesity and disease risk.

Objective: The primary objective of the study was to develop an accurate and noninvasive system for automatically monitoring eating and chewing activities using sensor-equipped smart glasses. The system distinguishes chewing from other facial activities, such as speaking and teeth clenching. The secondary objective was to evaluate the system's performance on unseen test users using a combination of laboratory-controlled and real-life user studies. Unlike state-of-the-art studies that focus on detecting full eating episodes, our approach provides a more granular analysis by specifically detecting chewing segments within each eating episode.

Methods: The study uses OCO optical sensors embedded in smart glasses to monitor facial muscle activations related to eating and chewing activities. The sensors measure relative movements on the skin's surface in 2 dimensions (X and Y). Data from these sensors are analyzed using deep learning (DL) to distinguish chewing from other facial activities. To address the temporal dependence between chewing events in real life, we integrate a hidden Markov model as an additional component that analyzes the output from the DL model.

Results: Statistical tests of mean sensor activations revealed statistically significant differences across all 6 comparison pairs (P<.001) involving 2 sensors (cheeks and temple) and 3 facial activities (eating, clenching, and speaking). These results demonstrate the sensitivity of the sensor data. Furthermore, the convolutional long short-term memory model, which is a combination of convolutional and long short-term memory neural networks, emerged as the best-performing DL model for chewing detection. In controlled laboratory settings, the model achieved an F_1-score of 0.91, demonstrating robust performance. In real-life scenarios, the system demonstrated high precision (0.95) and recall (0.82) for detecting eating segments. The chewing rates and the number of chews evaluated in the real-life study showed consistency with expected real-life eating behaviors.

Conclusions: The study represents a substantial advancement in dietary monitoring and health technology. By providing a reliable and noninvasive method for tracking eating behavior, it has the potential to revolutionize how dietary data are collected

and used. This could lead to more effective health interventions and a better understanding of the factors influencing eating habits and their health implications.

(JMIR Mhealth Uhealth 2024;12:e59469) doi:10.2196/59469

KEYWORDS

chewing detection; eating detection; smart glasses; automatic dietary monitoring; eating behavior

Introduction

Background

Obesity is a public health issue that leads to chronic diseases [1], including diabetes [2], cancer [3], and cardiovascular diseases [4]. In the United Kingdom, the obesity levels increased from 15% in 1993 to 28% in 2019 [5]. Similarly, in the United States, the obesity levels increased from 14.5% in 1970s to 39.6% [6]. Furthermore, poor diet was estimated to have contributed to 11 million deaths globally in 2017 [7].

Given these alarming statistics, gaining insight into people's dietary habits is crucial for designing effective interventions aimed at promoting a healthy lifestyle. Dietary behavior tracking includes a spectrum of approaches ranging from manual to highly automated methods. At the most manual end, traditional food diaries require users to write down manually or digitally every item they eat or drink. The most commonly used manual tools to assess dietary intake and eating behaviors are 24-hour recalls, food records (food diaries), and food frequency questionnaires [8,9]. Major limitations of these methods include participant burden and recall or memory bias [10], which can lead to under- and overreporting of dietary intake. Digital tools and apps (eg, MyFitnessPal [11]) simplify the manual input process and integrate nutritional data, yet they require active user engagement, and in some cases nutrition knowledge to estimate calorie intake from precooked meals. A visual and less structured alternative is photographing meals, which offers an alternative way to recall and review dietary choices, sometimes shared with a dietitian for professional advice.

Related Works

The existing research indicates a growing interest in developing automated tools for monitoring eating activities. Regarding the tools and studies closest to our study, several studies have explored monitoring eating activities based on sensor-enabled glasses. Most of these studies are focused on detecting eating or drinking episodes [12-16] and are performed in controlled environments [17,18]. Only one study has explored a more complicated scenario than the typical eating or noneating detection [18] by exploring the detection of chewing events using eyeglasses equipped with electromyography sensors in a study involving 10 participants both in controlled and in real-life conditions. Compared to the existing work, we present the first study to use smart glasses with integrated optical surface tracking sensors and deep learning (DL) to accurately identify both eating and chewing events, assessed both in controlled laboratory settings and through real-life trials, thus addressing research gaps and proving its efficacy in natural environments.

Objective

This study aimed to develop and evaluate a novel, noninvasive system for automatically monitoring eating behavior by detecting eating and chewing activities. The system aims to enhance the accuracy and ease of tracking eating behaviors, addressing the limitations of self-reporting by providing precise, objective data.

The study provides a comprehensive evaluation of the proposed method using a combination of laboratory-controlled and real-life user studies, ensuring robust and noninvasive way to distinguish chewing activity from other activities, such as speaking, teeth clenching, grinding, smiling, frowning, brow raise, and winking.

The real-life data collection and analysis addresses a substantial gap in previous research and allows for the evaluation of the system's performance in natural settings, providing insights into its practical application and adaptability.

Methods

Terminology

Throughout this study, several terms related to eating behaviors are used. To ensure clarity and consistency in their use, the following definitions are provided:

- Bite: the act of placing food into the mouth, chewing it, and then swallowing it as part of the eating process.
- Chew: a masticatory cycle involving the grinding or crushing of food with the teeth, preparing it for swallowing.
- Chewing: the overall process of breaking down food with the teeth.
- Chewing rate: the frequency of masticatory cycles (chews) per unit of time, measured in chews per second.
- Eating segment: a continuous period during which the participant consumes food without interruption, encompassing consecutive bites and chewing cycles without pauses between bites. Thus, one eating segment can include one or several bites and chewing events.

Smart Glasses and Data Collection Setup

Overview

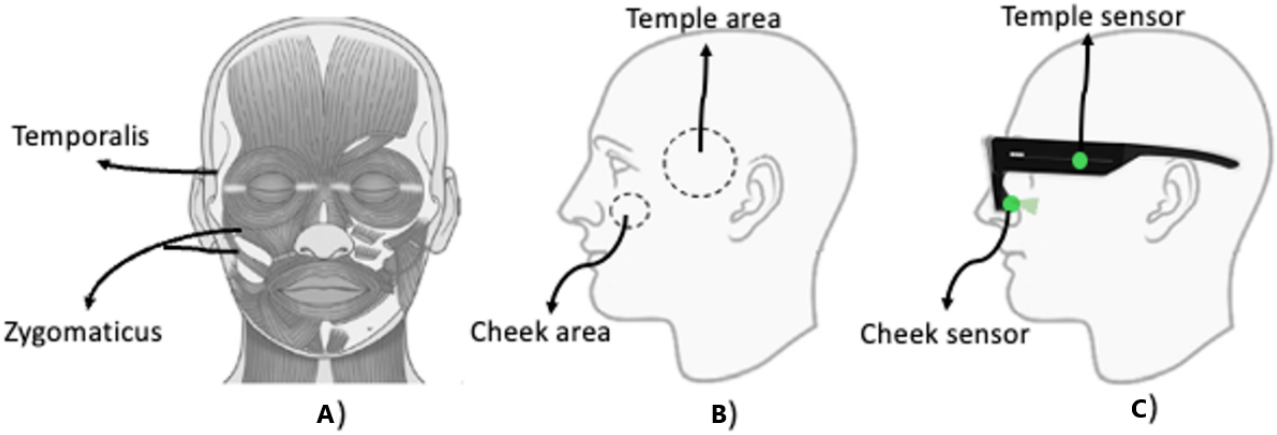
In this section, we describe our data collection setup, providing insights into the configuration and sensors of the used smart glasses. In addition, we describe the methodologies used for data collection in both controlled laboratory settings and real-life scenarios.

In contrast to the methods that require manual input, in this study, we propose an approach to automatic monitoring of eating behavior by monitoring facial muscle activations using optical

sensors incorporated in smart glasses frame. The approach offers real-time feedback that can be integrated with mobile health apps, allowing users to monitor their dietary habits seamlessly. The data collected can be used to personalize dietary recommendations, support weight management programs, and contribute to research in nutritional epidemiology. Ultimately, the goal is to empower individuals with actionable insights to improve their eating habits and promote long-term health and well-being.

The proposed system is depicted in Figure 1: (A) facial muscles associated with chewing that we aim to monitor; (B) the areas of skin that are monitored by the system; and (C) OCO optical sensors embedded in smart glasses. One of the muscles associated with the chewing activity is the temporalis muscle. The temporalis muscle is near the temple and extends downward in a direction toward the mouth. It controls movement of the lower jaw (eg, opening and closing of the mouth). This area is monitored by the OCO *temple* sensor in the glasses. Other muscles that are activated during chewing are the cheek muscles such as zygomaticus major and minor. This area is monitored by the OCO *cheek* sensor in the glasses. Our approach is based on the assumption that chewing activates multiple facial muscles, which causes the facial skin to move in a parallel direction relative to the sensors embedded within a glasses frame. These movements of the facial skin in the X-Y plane are monitored by our novel patented optical tracking sensors—OCO. The optical sensor data are then analyzed using DL to distinguish chewing activity from other activities that cause facial skin movements, such as speaking, teeth clenching, smiling, frowning, winking, and similar.

Figure 1. (A) The 2 types of facial muscles related to chewing (temporalis and zygomaticus); (B) monitored skin areas by the smart glasses; and (C) the placement of OCO sensors within the glasses frame.



OCOsense Smart Glasses and OCO Sensors Data

The OCOsense smart glasses integrate 6 optical tracking—OCO sensors [19], 3 proximity sensors, a 9-axis inertial measurement unit, an altimeter, and dual speech detection microphones. The OCO sensors use optomyography, an optical noncontact methodology, to measure skin movement in 2 dimensions resulting from underlying myogenic activity. They consist of an optical surface tracking sensor that measure relative movements on the skin's surface in 2 dimensions (*X and Y* dimensions). These sensors operate accurately within a range of 4 to 30 mm without requiring direct skin contact [19]. Positioned within the glasses frame, their focus lies on monitoring skin movement over specific facial muscle groups, including the frontalis and corrugator muscles on both sides of the forehead, the zygomaticus major and minor muscles on the left and right sides of the cheeks, the orbicularis muscles around each eye, and the left and right temples.

The eating activity activates two types of facial muscles that we can monitor with the glasses: (1) the temporalis muscle, which is near the temple, and controls movement of the lower jaw (opening and closing of the mouth), and (2) zygomaticus major and minor, which are located in the cheek area, and are activated during the chewing activity. Therefore, in this paper, we primarily focus on data collected from the cheek and temple OCO sensors (marked with green rectangles in Figure S1 in Multimedia Appendix 1), as these areas are more relevant to eating activity, compared with the rest of the sensors available in the glasses (marked with red rectangles in Figure S1 in Multimedia Appendix 1). A corresponding sensor data are presented in Figures S2 and S3 in Multimedia Appendix 1.

Data Collection Methodology

For development and evaluation of our method we collected 2 data sets. The first data set was collected in laboratory environment, while the second data set was collected in-the-wild. The laboratory data enabled us to establish a foundational understanding of eating behaviors under controlled conditions. However, evaluating the method on real-life data allows for assessing its generalization capability and adaptability to diverse and unpredictable environments.

To be more precise, we used the laboratory data sets to:

- Perform a statistical analysis comparing measurements obtained during 3 activities (eating or chewing, speaking, and teeth clenching) from both temple and cheek sensors, assessing skin movement along both the X and Y axes. This analysis is based on data from 28 participants for whom we have both eating and noneating labeled data.
- Develop and evaluate DL models for chewing detection. We compared the performance of four DL architectures. For the best-performing DL method, we conducted a more detailed analysis, including the impact of individual sensors

on the performance of DL models (eg, temple, cheek, temple + cheek, left temple + left cheek, and right temple + right cheek), and the impact of segmentation window size on the performance the DL models, varying the window size between 2 and 15 seconds.

We used the real-life data set to evaluate chewing detection and eating segments detection methods using data collected in-the-wild. Summary of the collected data sets is presented in Table 1.

Participant recruitment involved booking a time for data collection through social media announcements and completing a Google form to confirm eligibility. Eligible participants were required to be in good health; with no history of eating disorders; and without dietary restrictions, allergies, or intolerances. In addition, participants with conditions affecting facial muscle activation, such as stroke or facial palsy, or any other conditions impacting normal and symmetrical chewing and swallowing were excluded. An important inclusion criterion was that participants have proper glasses fit to ensure accurate detection of skin movements by the sensors.

Table 1. Summary of collected data sets.

Data set	Participants, n	Median duration	Total duration
Eating (laboratory)	28	9 min 49 s	369 min 15 s
Noneating (laboratory)	126 (same 28+98 new)	11 min 50 s	1601 min 47 s
Real life	8	907 min	7163 min

Study Procedures

Laboratory-Based Data Collection (Controlled Environment)

In the laboratory-based experiments, we collected two data sets:

1. Eating data set: the participants engaged in a full meal, providing them with the freedom to choose from a diverse range of food options, including:
   - Crispy or hard foods: apples, carrots, nuts, crisps, and crackers
   - Creamy or soft foods: porridge, banana, yogurt, fruit salad, and green salad
   - Chewy foods: breakfast bars; pop-tart; toast, bagel, or croissant; and biscuits

In addition, they were allowed to eat with or without utensils, based on their preference. There were no time constraints for completing the meal. Participants ate their meals in a laboratory setting designed to simulate a natural dining environment. They consumed their meals alongside the researchers, which helped create a more relaxed and realistic atmosphere. Despite the laboratory setting, participants were encouraged to consume their meals in a natural manner, simulating real-life conditions. This allowed for varied behaviors, for example, some participants used their phones during meals and others engaged in conversations. During the data collection, the participants were continuously video recorded, providing synchronized data between the video recording and sensor data. This enabled us to label each chewing segment later manually. For the eating activity, we annotated all segments where the participants had food in their mouth. Two researchers independently coded the bites, ensuring reliability and validity through cross-verification. The availability of video data allowed for accurate annotation, as each segment was reviewed by at least 2 researchers to confirm the presence of food in the participants' mouths.

2. Noneating data set: the data collection was performed in a controlled laboratory setting, where participants were instructed on the activities they should perform. First, participants performed a subset of activities associated with facial muscle engagement. This category includes brushing teeth, engaging in conversation, reading aloud, and diverse expressions of bruxism, encompassing teeth clenching, grinding, and tapping. In addition, we incorporated various facial expressions and gestures, such as smiling, frowning, winking, and similar, to capture a diverse range of facial movements. Moreover, we included a variety of activities that do not specifically rely on facial muscle engagement. These include hygiene-related activities such as handwashing and dishwashing, routine activities such as walking and sitting in a chair, and physical activities such as jogging and stair climbing.

Real-Life Data Collection (Uncontrolled Environment)

In the real-life setting, the participants were instructed to wear the OCOsense smart glasses continuously for a minimum of 8 hours a day over a span of 2 days. The participants were allowed to follow their daily routines without any imposed limitations during this period. This enabled the capture of eating behaviors in various settings such as home, workplace, and other public spaces. In addition, there were no restrictions placed on participants regarding their food choices or other diet-related decision. For the data collection procedure, we developed an application that collects data from the glasses and enables the participants to annotate when engaged in eating activities. More specifically, they were asked to press a button when they start eating and press it again when they finish eating. A researcher monitored the number of labeled eating events per day per participant. In instances where participants forgot to press the start or end buttons, they were asked to note the approximate times of their eating sessions. These cases were then manually analyzed by a researcher using the sensor data from the glasses to provide precise labels for the eating start and end times. These labeled segments served as the ground truth for subsequent experiments. The annotations collected with this approach result in whole data segments labeled as eating, yet these segments may also include a range of activities beyond eating itself, such as engaging in conversation or pausing briefly between bites, which typically occur during regular real-life meals.

Statistical Analysis

Data Preprocessing

To perform statistical comparison, the following data preprocessing steps were applied to the sensor data:

1. Calculation of the vector magnitude for each sensor: as the OCO sensors measure skin movement in 2 dimensions (*X* and *Y*), the vector magnitude was calculated for each sensor ( ).
2. Combination of processed sensor signals values from the left and right sensors: the vector magnitude value from the left cheek sensor was added to the vector magnitude value from the right cheek sensor, and the same was done for the temple sensors. This resulted in the creation of 2 signals, one representing the total cheek movement (left+right), and one representing the total temple movement (left+right).
3. Smoothing of the resulting signals: the resulting cheek and temple signals were smoothed using a rolling median filter with a window size of 15 samples (0.3 s) to reduce the effects of noise on the signals.

Hypothesis Testing

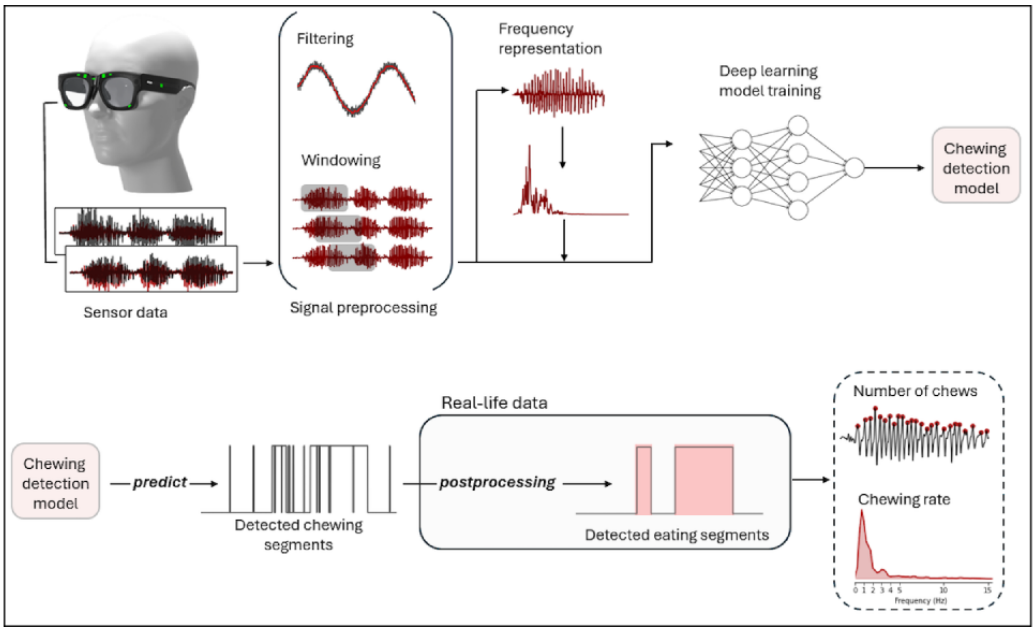
Hypothesis testing was conducted using the Wilcoxon signed-rank test, a nonparametric alternative to the paired 2-tailed *t* test. This test evaluates the distribution of differences between related paired samples to ascertain whether they originate from the same distribution. The null hypothesis is that the samples derive from the same distribution. To account for multiple comparisons, *P* values were adjusted using the Bonferroni correction method (α=.05).

Chewing Detection Methodology

Overview

This section describes the method used in this study for automatic chewing and eating segment detection. Initially, the sensor data undergoes preprocessing, including filtering and segmentation into windows. Then, both the filtered signals and their frequency representations are used as input to DL models, which classify the windows into chewing or nonchewing. To enhance the accuracy in real-life scenarios, we introduce a supplementary model—hidden Markov model (HMM). This integration enables the grouping of chewing predictions and the construction of coherent eating segments. Finally, we calculate the number of chews and the chewing rate for the detected eating segments. The block diagram of the pipeline is shown in Figure 2.

Figure 2. Overview of the developed method for detection of chewing and eating segments estimation.



Signal Preprocessing

The method uses data from the 4 OCO sensors—left and right temple and left and right cheek. Let O_{RC}, O_{LC}, O_{RT}, *and* O_{LT} denote these sensors in the specified sequence. The set of sensors can be represented as S = {O_{RC}, O_{LC}, O_{RT}, O_{LT}}, where each sensor S^i reads data . in the time interval of *T* from timestamps t_1 to t_n. The main objective is defined as:

1. Partitioning *T* into partially overlapping windows of equal size $W = \{W_1,W_2,...,W_n\}$ and assuming a target activity set $Y = \{Y_1,Y_2,...,Y_n\}$
2. Assigning each window W_i a target label Y_j from the target label set $Y = \{Y_1,Y_2,...,Y_n\}$ and training a classifier accordingly

First, to remove the noise from the data, a fifth order median filter was applied to each sensor channel within the sensor set *S*. This filter was proven to effectively remove noise while preserving essential signal features in our previous studies on expression recognition using the same type of sensor [20]. Following the medial filter, the next step in the process involved determining the appropriate window size for data segmentation. We experimented with various window sizes ranging between 1 and 15 seconds. Once the sensor set *S* was segmented into

windows (W), the next step was to enhance the information carried by the input signals further. To achieve this, we used Fourier transformation for each sensor within the segmented windows. This transformation allowed us to convert the time-domain signals into their frequency representations, thereby extracting additional features from the data. The Fourier transformation process provided valuable insights into the frequency components present in the sensor data, which could be crucial for detecting subtle patterns associated with chewing activity.

Chewing Detection With DL Models

In this study, we used 4 distinct DL models based on convolutional neural networks (CNNs) for the purpose of chewing detection. We focused on DL architectures commonly used for wearable sensor data, such as CNN 1D [21,22], CNN 2D [21], attention model [23], and convolutional long short-term memory (ConvLSTM) [24]. By using these common architectures, we aim to demonstrate the baseline accuracy achievable with existing methods. This serves as a foundation upon which further improvements can be made. Specifically, developing DL architectures tailored to the unique specifications of the glasses and the specific use-case of detecting eating activity could potentially enhance accuracy beyond the baseline results established in this study.

An overview of the architectures and their associated hyperparameters is as follows:

1. CNN 2D: Our initial model adopts a standard CNN [21], crafted to extract hierarchical spatial features from input data. The feature extraction module consists of 3 consecutive convolutional layers, each followed by group normalization and max-pooling layers. Extracted features are then passed through 2 fully connected layers, each containing 128 neurons, connecting to the output nodes.
2. ConvLSTM: Expanding on CNN's foundation, the ConvLSTM model [24] introduces a temporal dimension to our analysis. It shares the same convolutional layers with the CNN 2D architecture and integrates 2 LSTM layers, each featuring 128 hidden units. This modification allows the model to effectively capture sequential patterns and dependencies within the data.
3. Attention model: Incorporating insights from attention mechanisms, the attention model [25] comprises 4 convolutional layers with 64 feature maps followed by 2 LSTM layers, each with 128 hidden units [23], and an attention layer. The attention layers allow the model to prioritize relevant information during the learning process.
4. CNN 1D with statistical features: The last model incorporates a 1D CNN architecture [22], enhanced with statistical features. It consists of a single convolutional layer with 256 filters followed by a max-pooling layer. The resulting features are then flattened and fused with statistical features extracted from filtered sensor data, including mean, variance, and absolute sum. The joint vector is then processed through a fully connected layer with 1024 neurons capturing both spatial and statistical characteristics.

The determination of architecture parameters, such the kernel size in the convolutional layers, output size of CNN layers, LSTM units, and fully connected units, was guided by a pragmatic approach focused on achieving a balance between model's ability to capture complex data patterns and model's complexity. These parameters were fine-tuned on the validation set to optimize performance.

Each model was trained for 100 epochs with a batch size set at 256. Prior the beginning of the learning process, we used orthogonal weight initialization for both weights and biases, aiming to enhance the stability and effectiveness of neural network training. Cross-entropy loss was used as the objective function for training. Furthermore, all the models were trained using the Adam optimizer with an initial learning rate of 1e-3. To avoid overfitting as well as to reduce the training time, early stopping, monitoring validation F_1-macro score with patience of 15 epochs was applied. In the end, the optimal weights were selected based on the epoch with the highest validation F_1-macro score.

Detection of Eating Segments

In the initial phase of our eating detection system, we use a DL model to detect chewing moments at a window-level granularity. By incorporating the temporal dependence between the detected chews, we aim to enable our system to identify not only individual chewing instances but also to discern when eating segments occur within real-life data. This allows us to effectively mitigate the occurrence of short false-positive predictions and consolidate densely clustered chewing instances into coherent eating segments. By doing so, we anticipate a more robust and precise analysis of dietary patterns.

To address the temporal dependence between chewing events in real life, we integrate HMM as a supplementary model that analyzes the detected chews from the DL model. The HMM was initialized and trained as described in the study by Stankoski et al [26]. This process is visually illustrated in Figure S4 in Multimedia Appendix 1.

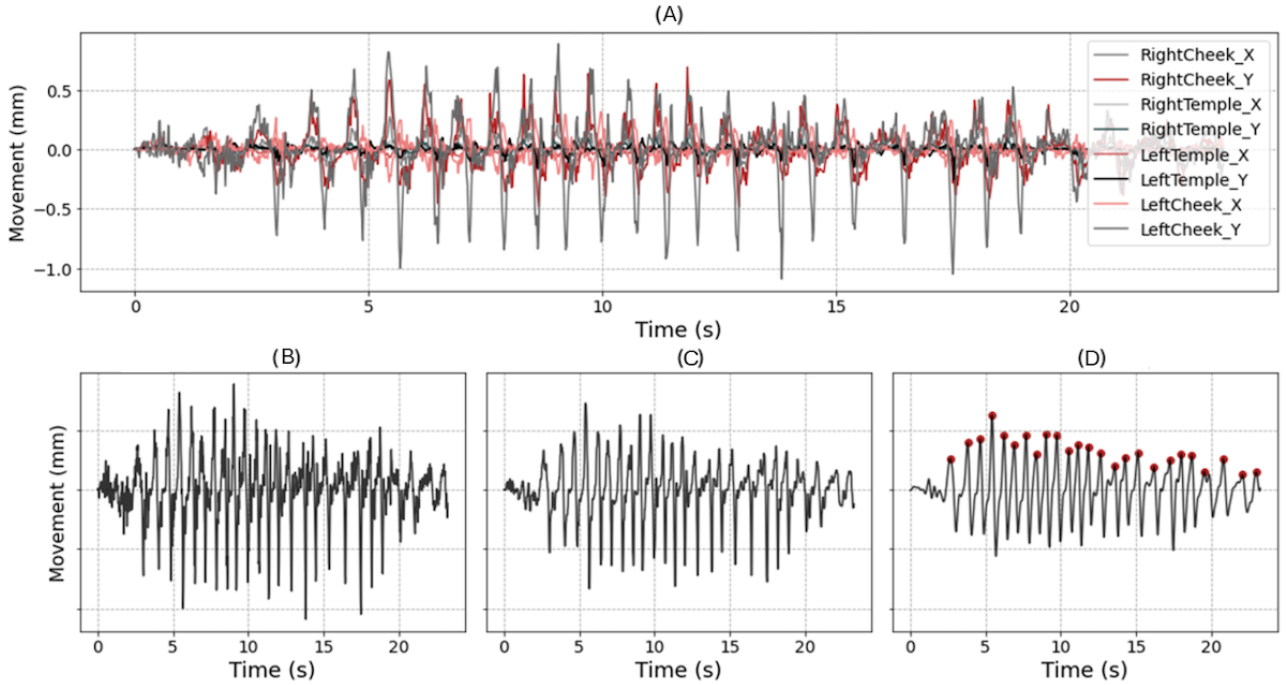
Detection of Number of Chews and Chewing Rate Estimation

After the detection of eating segments, to determine the number of chews in an eating segment, we additionally analyze and process the signals from the sensors. Figure 3 presents a visual representation of the data processing steps used in the detection of chews within a randomly selected eating segment from the data set. The initial step involves identifying the signal with the highest root mean square value. Subsequently, we use a 2-step filtering process to enhance the selected signal. First, a median filter with a kernel size of 5 is applied, followed by a second-order bandpass filter within the frequency range of 0.5 to 3 Hz. For the calculation of the number of chews, we used an existing peak detection algorithm (using SciPy [27]) on the processed signal. This involves configuring the threshold and distance parameters to identify relevant peaks in the signal. Furthermore, peaks with insufficient prominence are excluded from the final set.

Each retained peak after this step is considered as a separate chew in the signal. The parameter values used in the filtering and peak detection processes were determined empirically.

Following the detection of chews within eating segments, we extend the analysis to estimate the chewing rate. To achieve this, we use the same signal with the highest root mean square value. This signal is subjected to further analysis through Fourier transformation to compute its frequency spectrum. By examining the resulting spectrum, we identify the most substantial frequency component, which corresponds to the dominant chewing frequency within the examined eating segment.

Figure 3. Processing steps for detecting the number of chews in an eating segment: (A) cheek and temple signals; (B) selection of signal with highest root mean square (RMS) value; (C) filtering the chosen signal; (D) detection of peaks and chews in the filtered signal.



Evaluation Setup

To evaluate the effectiveness of the models, we used the Leave-One-Group-Out cross-validation technique. This involved dividing the initial data set into N separate groups, where the data from a single participant is present in only one subset. Each model is trained on combined data from *N*-2 subsets, leaving one subset to be used as validation data set and a second subset for testing the final model. Thus, all the models are person-independent, that is, the experimental results demonstrate the model's accuracy on unseen test users.

Regarding evaluation metrics, we used recall, precision, and F_1-score. Recall indicates the proportion of actual chewing segments correctly identified by the model, while precision denotes the proportion of identified chewing segments that are truly chewing segments. The F_1-score is the harmonic mean of the recall and the precision—which is more balanced metric compared with accuracy especially in unbalanced data sets where one of the classes is more frequent. The reported metrics reflect the models' ability to detect chewing at a window level, and they are calculated as follows:

(1)

(2)

(3)

In the equations (1) to (3), *TP* represents true positives, *TN* represents true negatives, *FP* represents false positives, and *FN* represents false negatives. In the context of chewing detection, these metrics can be interpreted as follows:

- *TP* indicates the number of windows from the chewing class correctly classified as chewing.
- *FP* indicates the number of windows from the nonchewing class incorrectly classified as chewing.
- *FN* indicates the number of windows from the chewing class incorrectly classified as nonchewing.

In addition, for the evaluation of eating detection in the real-life scenario, we used custom metric to provide deeper insights into the models' performance within eating segments. The metric was defined to analyze the number of eating segments that are correctly identified based on the frequency of positive chewing predictions within each eating segment:

- Detected eating segments: the number of eating segments where at least 50% of the instances (windows) are correctly identified as eating.

Ethical Considerations

To ensure ethical compliance, ethics approval was obtained from the London—Riverside Research Ethics Committee on July 15, 2022 (ref: 22/LIO/0415). After a detailed explanation of the experimental procedure, all participants provided written

informed consent before participating in the study. The consent forms addressed the use of their data. To protect participant privacy, all data were deidentified. The participants who took part in the laboratory sessions were compensated with US $26.5, while those involved in the real-life study received US $26.5 per day for their participation. The experiment was conducted following institutional ethical provisions and the Declaration of Helsinki.

Results

Overview

In this section, we present the results from the experiments. The *Statistical Analysis of OCO Sensors for Facial Muscle Movements* section presents the outcomes of the statistical analysis, focusing on the ability of OCO sensors in detecting facial muscle movements during various activities, including eating. The *Laboratory-Based Data Set DL Experiments* section assesses the performance of different DL models, sensor combinations, and window sizes for chewing detection in a controlled laboratory data set. Finally, the *Real-Life Data Set Experiments—Chewing and Eating Segments Detection* section presents the results obtained with the real-life data set and evaluate the performance of the method for detection of eating segments.

Statistical Analysis of OCO Sensors for Facial Muscle Movements

To evaluate the ability of the OCO sensors to detect facial muscle movements during different activities, we first conducted a statistical analysis. Our focus was on comparing measurements obtained from both temple and cheek sensors, assessing skin movement along both the *x* and *y* axes. In this context, we focused on comparing facial muscle movements during the activities of eating or chewing, speaking, and teeth clenching. The selection of these activities was based on the potential similarity in facial muscle activation patterns. For example, Figure 4 presents 6 graphs. The top row measures movements from sensors placed over the zygomaticus major muscle (cheek area) and the bottom row from sensors positioned on the temples. Each column of graphs represents 1 of the 3 activities being measured (eating, speaking, and clenching). The horizontal axis of each graph represents time in seconds, and the vertical axis shows the magnitude of skin movement in millimeters. By comparing these graphs, we can assess the differences and similarities in facial muscle activation patterns during the 3 activities.

We calculated mean movements measured from the cheek and temple OCO sensors for each participant during eating or chewing, speaking, and teeth clenching. The mean values were calculated over all data points corresponding to each activity, resulting in n=28 (number of participants present in both the eating and noneating laboratory data set) tuples, with each tuple comprising 3 values representing the mean cheek or temple movement for eating or chewing, speaking, and teeth clenching.

Figure 5 shows the mean cheek (left plot) and temple (right plot) movements during different activities, presented on the x-axis, and the results from the Wilcoxon signed-rank (paired) test with Bonferroni correction (α=.05).

For the cheek OCO sensors, we can observe an increased movement during eating (median value 0.113 mm) compared with relatively lower values observed during speaking (median value 0.036 mm) and the teeth clenching (median value 0.008 mm). The results from the statistical test further indicate significant differences in cheek movements between speaking and eating (P<.001), eating and teeth clenching (P<.001), as well as speaking and teeth clenching (P<.001).

Similarly, for the temple OCO sensors, a notable increase in movement with a median value of 0.027 mm during eating is observed, compared with 0.008 mm during speaking and 0.002 during the teeth clenching. The statistical tests affirm the significance of these differences, demonstrating that mean temple movements differ significantly between speaking and eating (P<.001), eating and the teeth clenching (P<.001), as well as speaking and the teeth clenching (P<.001).

These findings highlight the potential sensitivity of the cheek and temple OCO sensors in capturing distinct patterns and subtle variations in facial muscle activation across different activities.

Figure 4. Sensor signals from the sensors on the right cheek and temple recorded during eating, speaking, and teeth clenching activity performed by one participant.

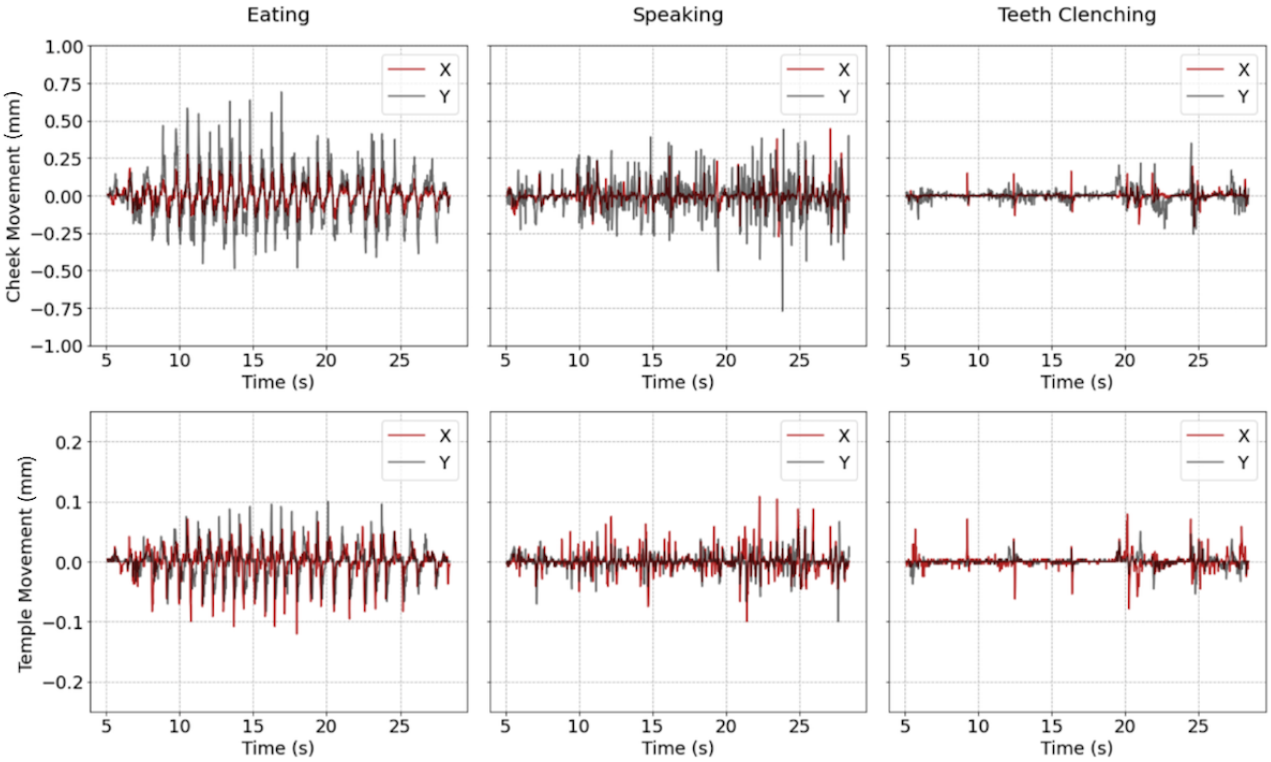
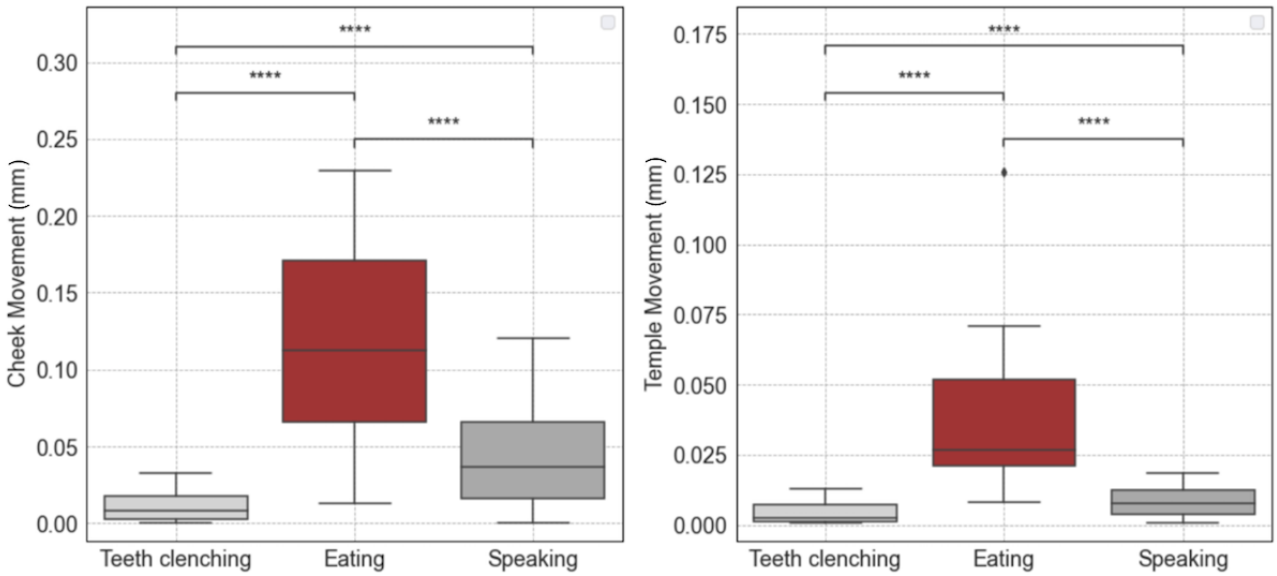


Figure 5. Wilcoxon signed-rank (paired) test with Bonferroni correction for comparing mean cheek and temple movements during activity pairs (n=28): clenching versus eating, clenching versus speaking, and eating versus speaking. Statistical significance annotations: *If P ∈ {.05, .01); **if P ∈ {.01, .001); ***if P ∈ {.001, .0001); and ****if P≤.001.



Laboratory-Based Data Set DL Experiments

In this section, we present the sample characteristics of the data set, the results of the experiments for chewing detection, conducted on the laboratory-based data set, offering insights into the results achieved across various DL architectures, sensor combinations, and window sizes.

Sample Characteristics

The laboratory-based data set consists of 2 subsets, one for eating activities and another for noneating-related activities. In the controlled eating data set, we gathered data from a cohort of 28 participants, comprising 13 (46%) males and 15 (54%) females, with an average age of 25.6 (SD 9.1) years. The data set comprises a total of 6.1 hours of recorded data. The noneating data set includes data from the same 28 participants, along with an additional 98 participants (n=48, 49% males and n=50, 51% females) with an average age of 23.3 (SD 6.4) years. Each participant contributed data for various activities, totaling 26.7 hours of recorded data. In summary, the data set comprises 126 participants and spans a combined total of 32.8 hours of recorded data.

DL Models for Chewing Detection

In this section, we present a comparison of various DL architectures used for the task of chewing detection. Table 2 provides a summary of the performance metrics, including F_1-score, recall, and precision for the chewing class, for each architecture.

The results show that all architectures demonstrated strong results, indicating that the sensor data provided from the glasses is informative for the chewing detection task. ConvLSTM demonstrated the highest F_1-score of 0.91, precision of 0.92, and recall of 0.89. CNN 2D also performed well with balanced metrics, achieving a slightly lower precision of 0.90, recall of 0.90, and F_1-score of 0.90. In contrast, the attention model displayed moderate performance with precision, recall, and F_1-score of 0.89, 0.90, and 0.89, respectively. The CNN 1D architecture, despite exhibiting a high precision of 0.90, fell short in recall at 0.86, resulting in a lower overall F_1-score of 0.88.

The confusion matrices for the evaluated models are presented in Figure 6. They provide additional insights into the models' behavior. Notably, the ConvLSTM model also demonstrated a lower number of false positives (FPs), totaling 1749 instances. This number is approximately 20% lower than that of the second-best model, CNN 2D, which recorded 2089 FP instances.

Figure S5 in Multimedia Appendix 1 shows the FP rates for various noneating activities detected by the ConvLSTM model. Socializing has the highest rate (0.72%), followed by reading (0.28%). Both involve speaking, leading to confusion with eating due to similar facial movements. The overall FP rate is 2%, also shown in Figure 6 (0.02 in confusion matrix D).

Table 3 provides a comparison of complexity and resource metrics for the evaluated architectures focusing on network parameters, computational complexity expressed as the number of floating-point operations per second (FLOPS) during a forward pass and the size of the model. CNN 1D with 2.4 gigaFLOPS and the 270-kB model size is the smallest model making it suitable for embedded applications in the future. CNN 2D, although larger, offers a balanced trade-off between performance and model size. The attention model, despite having fewer parameters than CNN 2D, has the highest computational complexity of 80.25 gigaFLOPS. ConvLSTM demonstrates a balance between accuracy and resource requirements.

Considering the results, ConvLSTM emerged as the preferred choice for the chewing detection tasks based on the model accuracy and computational complexity and resources needed, thus this architecture was used in the subsequent experiments.

Table 2. Performance metrics of different deep learning architectures for chewing detection. Precision, recall, and F1-score are calculated for the eating class.

DL architecture	Precision	Recall	F_1-score
CNN[a] 1D	0.9	0.86	0.88
CNN 2D	0.9	0.9	0.9
Attention model	0.89	0.9	0.89
ConvLSTM [b]	*0.92* [c]	*0.89*	*0.91*

[a]CNN: convolutional neural network.

[b]ConvLSTM: convolutional long short-term memory.

[c]Best performing algorithm.

Figure 6. Confusion matrices for the evaluated deep learning architectures: (A) convolutional neural network (CNN) 1D; (B) CNN 2D; (C) attention model; (D) convolutional long short-term memory (ConvLSTM).

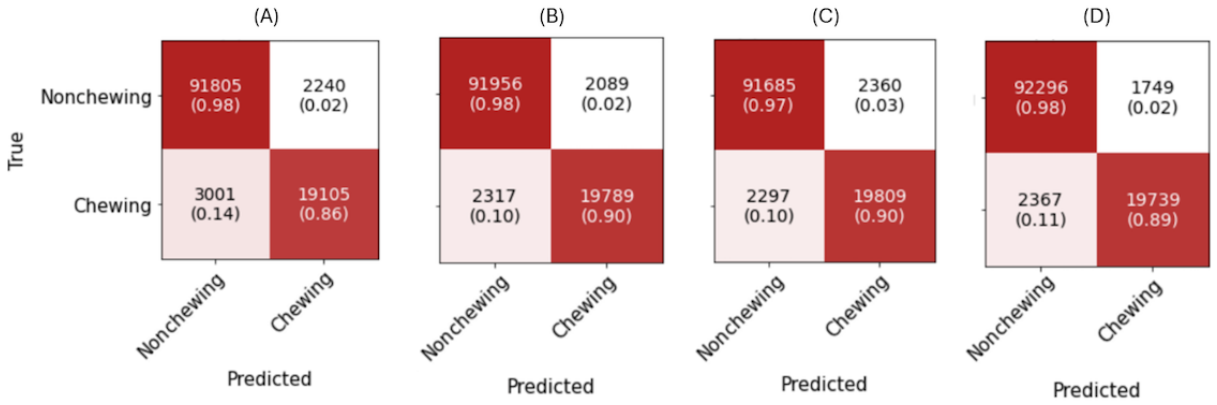


Table 3. Complexity and resource metrics for the evaluated deep learning (DL) architectures.

DL architecture	Total parameters	Computational complexity (GFLOPS[a])	Model size (MB)
CNN[b] 1D	43,366	2.41	0.27
CNN 2D	1,208,578	20.92	4.88
Attention model	407,620	80.25	2.68
ConvLSTM[c]	997,890	31.42	4.03

[a]GFLOPS: giga floating-point operations per second.

[b]CNN: convolutional neural network.

[c]ConvLSTM: convolutional long short-term memory.

Impact of Individual Sensors on the Performance of Chewing Detection Models

In this section, we present the results from the analysis of the impact of individual sensors on the performance of the chewing detection models. Having identified the ConvLSTM architecture as the best-performing architecture among the models that we evaluated in the previous experiments, we proceeded with this architecture for a series of experiments encompassing various sensor combinations. The tested sensor combinations included temple, cheek, temple and cheek, as well as the left versus right side. The results from these experiments are presented in Table 4.

From Table 4 it can be observed that the cheek sensor outperforms the temple sensor. Specifically, in detecting chewing segments, the cheek sensor achieves recall of 0.88, which is 4 percentage points higher than the recall achieved by the model trained with temple sensor data (0.84).

The performance of the model trained with data from the cheek sensors can be attributed to the role of the cheek region in eating activities, predominantly chewing. The sensors are adept at capturing the specific circular movements of the cheek area during such activities, which produce distinct signal pattern associated with eating. The ability to capture these specific patterns results in the model's high precision in distinguishing eating episodes, thus enhancing recall rates.

Although the temple muscle is uniquely activated during chewing activity, the results show that the activation measured by the sensor is not very high across all people. However, if we combine the temple and the cheek sensors, we can see that the recall is improved by 1 percentage point. This shows that the temple sensor data provides additional information to the model.

In addition, we explored the performance of the models by using only one side of the temple and cheek sensors. On the basis of the results, we can see that the combination with the sensors measuring the right temple and cheek achieves recall of 0.89, which is 3 percentage points higher than the recall achieved by the model trained with left temple and cheek sensor data. This might be expected because most people prefer to chew the food on one side of their mouth [27,28] and the activation of the muscles is higher, which results in higher values in the sensor data.

Table 4. Performance metrics of convolutional long short-term memory for chewing detection with multiple combinations of sensor data. Precision, recall, and F1-score are calculated for the eating class.

Sensor combination	Precision	Recall	F_1-score
Temple	0.83	0.84	0.83
Cheek	0.92	0.88	0.90
Temple+cheek	*0.92* [a]	*0.89*	*0.91*
Left temple+left cheek	0.9	0.86	0.88
Right temple+right cheek	0.89	0.89	0.89

[a]The selected combination shows the best results based on the F_1-score.

Window Size Impact on the Performance of Chewing Detection Models

This section presents the results of the analysis of how window size influences the performance of the chewing detection models. For this purpose, a series of experiments were conducted, exploring various window sizes that extend beyond the default 4-second window size used in the previous experiments. For this analysis, we used a consistent 1-second window slide, with the aim to prevent delays in prediction changes and to ensure that the model will be able to promptly detect eating-related movements. The results from the experiments are presented in Table 5.

The performance of the ConvLSTM architecture demonstrated a noticeable enhancement with the increase in window size in terms of precision, recall, and F_1-score. More specifically, as the window size extends from 2 to 10 seconds, we consistently observe improvements in results. However, upon reaching a 15-second window, we observe saturation in performance metrics, where the obtained results remain consistent with those achieved at the 10-second window. This is probably because

longer windows might include nonchewing data, leading the model to misclassify entire instances as noneating.

Although, among the window sizes of 6- and 10-second improvement can be observed, we decided to proceed with the 4-second window. This decision was based on its advantage in processing fewer data compared with the 6- and 10-second modes, leading to a reduced computational demand and potentially lower energy use.

Table 5. Performance metrics of convolutional long short-term memory for chewing detection with various window sizes. Precision, recall, and F1-score are calculated for the eating class.

Window size	Precision	Recall	F_1-score
2 seconds	0.90	0.86	0.88
4 seconds	0.92	0.89	0.91
6 seconds	0.93	0.92	0.92
10 seconds	*0.93* [a]	*0.94*	*0.93*
15 seconds	0.92	0.94	0.93

[a]Best performing result.

Real-Life Data Set Experiments: Chewing and Eating Segments Detection

Overview

To assess the effectiveness of our chewing detection and eating segments detection methods using data collected in-the-wild, we conducted a series of experiments. In the first subsection, we present the sample size of the data set. Then, in the second subsection, we present the results of the chewing detection method using real-life data. Next, in the third subsection, evaluation of the eating segment detection is presented. In the last subsection, we show the estimation of the chewing characteristics.

Sample Characteristics

The real-life setting data collection involved 8 participants (5 males and 3 females; average age 30.8, SD 12.4 years). Each participant wore the glasses for a minimum of 8 hours per day over 2 days, resulting in 16 hours of recorded data per participant and a total of 128 hours of recorded data.

Chewing Detection Evaluation Using Real-Life Data

This evaluation allows us to explore whether a model trained with seminaturalistic behavior data collected in a laboratory setting can perform well on a real-life data from unseen participants. Figure 7 presents the results obtained on the real-life data set at a window level using the model for chewing detection. The classification report is shown in Table 6. It shows that the model achieved precision of 0.95, recall of 0.82, and an F_1-score of 0.88 for the eating class. The accuracy of this model was 98%.

We derived the probability density function of the model's probability outputs. The resulting graph, depicted in Figure 8, reveals a bimodal distribution, exhibiting one smaller peak near a probability of 0.2 and a larger, more substantial peak beginning at approximately 0.8 probability. The prominence of the second peak starting from a higher probability threshold signifies the model's strong confidence in identifying chewing activity, within the labeled eating segments. In addition, the predictions around the first peak can be interpreted as instances where the model is relatively certain that chewing is not occurring within the eating-labeled segments.

These results are in line with our expectations and understanding of the real-life data set. As previously described, the ground truth of the real-life data set contains only the information when eating segments took place. When evaluating the chewing detection method on data set where eating segments are labeled, the presence of false negatives can be attributed to the nature of the data set. Eating segments may encompass various activities beyond just chewing, such as talking, short breaks between bites, holding food, and similar. Therefore, segments labeled as "eating" may indeed involve nonchewing activities.

Figure 7. Confusion matrix for the chewing detection model evaluated on the real-life data set on window level.

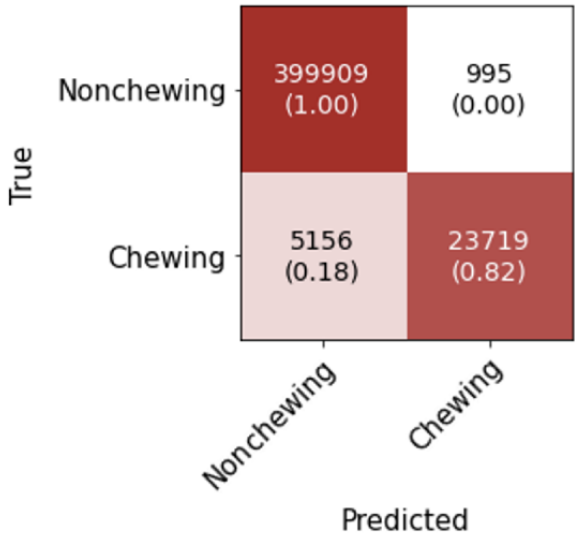
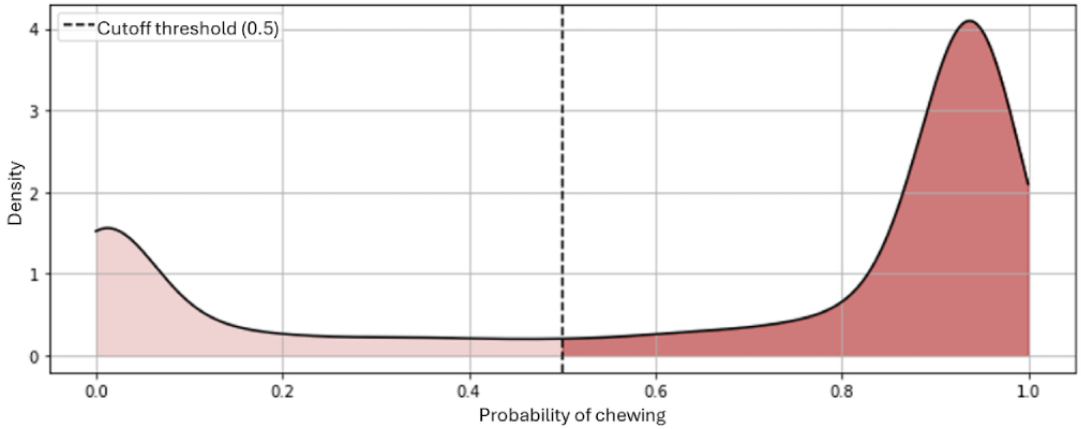


Table 6. Classification report for the chewing detection model evaluated on the real-life data set on window level.

Class	Precision	Recall	F_1-score
Noneating	0.99	1.00	0.99
Eating	0.95	0.82	0.88
Macroaverage	0.97	0.91	0.94

Figure 8. Probability density function of the model's output probabilities for the chewing-labeled instances.



Evaluation of the Method for Eating Segment Detection

As previously described, the ground truth for the real-life data set contains information when the eating segments took place. This means that the annotated eating segments may contain short breaks between bites, conversations, food preparation, and similar. Because of this, we evaluated the eating segments detection based on the temporal information of the chewing detection algorithm as described in the *Detection of Eating Segments* section.

The results obtained on a segment level are shown in Table 7. This table contains the total number of eating segments, number of detected eating segments, and falsely detected eating segments for each participant. An eating segment is considered as detected if >50% of the instances in the labeled segment are predicted as chewing. The result of this evaluation shows that from total of 74 eating segments labeled by the participants, we can accurately detect 71 eating segments. The number of the falsely detected eating segments is relatively low for all participants, having total of 7 false detections.

Furthermore, we extended our analysis of the real-life data set to explore the suitability of the sensor data obtained from the smart glasses in-the-wild for capturing more detailed eating-related metrics, beyond only detecting instances of eating. In particular, we aimed to quantify the number of chews and the chewing rate within eating segments, although this method was not subjected to formal evaluation, mainly because of the lack of ground truth in the real-life data set.

Table 7. Evaluation of eating segment (ES) detection on the real-life data set, including total number of ES, number of true detected ES, number of false detected ES, and mean duration of falsely detected ES per participant.

Participant ID	Total ES	True detected ES	False detected ES	Mean duration of falsely detected ES, SD (min)
1	5	4	0	—[a]
2	12	12	1	1.77 (0)
3	20	20	0	—
4	7	6	2	0.72 (0.5)
5	5	4	3	0.54 (0.2)
6	15	15	0	—
7	4	4	1	0.46 (0)
8	5	5	0	—
Total	74	71	7	—

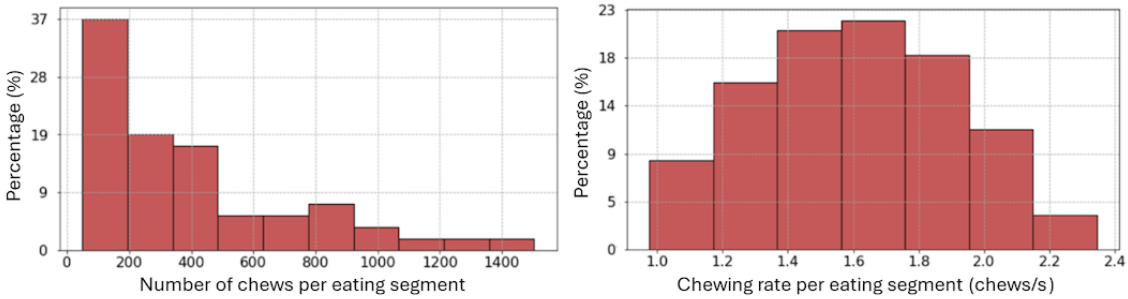
[a]Not available.

Estimation of Chewing Rate and Number of Chews Using Real-Life Data

After using the previously described methodology for calculating the number of chews on a window level for the eating segments, the resulting values ranged from 48 to 1505. The distribution of these values, as depicted in Figure 9A, indicates that participants have recorded both short snacks and long-duration meals, reflecting the diversity of eating behaviors captured in the data set.

Similarly, upon applying the approach for chewing rate estimation on a window-level for the eating segments, the derived values ranged between 0.8 and 2.3 chews/s. Notably, these values align with expectations observed in real-life eating scenarios [28]. Figure 9B depicts the distribution of the chewing rate values. Figure S6 in Multimedia Appendix 1 presents the mean chewing rate and the total number of chews for all eating segments across all participants in the data set.

Figure 9. Distribution of estimated number of chews and chewing rates per eating segment in the real-life data set.



Discussion

Principal Findings

While most people know that modifying eating behavior is key to sustained weight management, improving what is not measured is difficult. Traditional methods of eating assessment typically summarize dietary measures at the hour, day, week, or even year level [10]. Although these can clarify high-level relationships between eating behavior and disease risk, important short-term patterns are not measured and cannot be explored with the traditional methods. The ability to explore microlevel eating activities—such as meal microstructure (eg, meal duration and chewing frequency) [29] and food choices [30]—is important because they play an important role on food selection, dietary intake, and ultimately, obesity and disease risk [31-33].

In this study, we explored the potential of optical tracking sensors integrated into smart glasses for detection of eating, focusing on chewing activity as a critical component of dietary monitoring. Table 8 compares our study with existing glasses-based eating and chewing monitoring systems, highlighting differences in study goals, sensor types, number of participants, study setups, and performance metrics. It is important to note that the performance metrics may not be directly comparable due to differences in sensors, devices, population, and evaluation setup. Unlike most studies, which primarily monitor eating and drinking behaviors, our approach provides a more granular analysis by specifically detecting chewing activities. Only 2 other studies, including Zhang and Amft [18], have focused on chewing as an activity, and among these, only Zhang and Amft [18] and our study have conducted evaluations in real-world settings. Our results are slightly better, likely because electromyography sensors used by Zhang and Amft [18] require skin contact, whereas our optical sensor-based tracking does not, making our estimates more robust and less intrusive.

Accurate estimates of chewing can further be used for calculating eating rate. Such granularity offers better insights

for nutritional management. For example, eating rate has gained interest over recent years, as studies suggest a link between eating quickly and being overweight [31]. Other studies suggest that faster eating rate is associated with higher BMI and higher energy intake [32]. Studies also suggest that eating rate is independently associated with insulin resistance [33], which might be explained by the rapid entrance of glucose into the circulation at the beginning of the meal [34].

Our investigation spanned controlled laboratory settings to the real-life environment, providing a robust assessment of the technology's effectiveness and practicality. The system leverages contactless optical tracking technology—OCO, to monitor facial muscle activation related. These activations are further processed by a DL model for detection of eating and chewing segments. On the basis of the results of the experiments where we evaluated various DL architectures, ConvLSTM model was selected as the best-performing model for identifying chewing events in our eating detection experiments. Regarding the real-life experiments, our method for chewing detection and eating segments detection, validated on data collected in-the-wild with 8 participants, demonstrates promising results. The model achieved high precision (0.95) and recall (0.82) for the eating class, with an F_1-score of 0.88 at a window level. However, false negatives in chewing detection can be attributed to the diverse activities encompassed within eating segments beyond just chewing, such as talking or short breaks between bites. Evaluating eating segments detection based on temporal information from the chewing detection algorithm revealed accurate detection of 71 out of 74 labeled eating segments, with only 7 false detections across participants.

Regarding the sensor positioning, based on the results in Table 4 and the data from 128 participants, we observed that the cheek sensor outperforms the temple sensor, achieving a recall of 0.88 compared with 0.84 for detecting chewing segments. The addition of temple sensor data shows a modest 1 percentage point improvement in recall, indicating its supplementary role in enhancing overall performance.

Regarding the segmentation window size, as we increased the window size from 2 to 10 seconds, the ConvLSTM model's precision, recall, and F_1-score improved. However, at a 10-second window, performance plateaued, likely due to the inclusion of irrelevant nonchewing data. Despite better results at 6 and 10 seconds, we chose a 4-second window for its lower computational load and energy consumption. This aspect is crucial for deployment on mobile or wearable devices, where processing power and battery life are limited. Moreover, the use of a 4-second window has proven to offer stable performance that ensures the model is sufficiently fast to adapt to changes in eating behavior without substantial delays, due to the 1-second sliding segment.

Table 8. Comparison with glasses-based eating and chewing monitoring systems.

Study	Goal	Sensors	Participants, n	Setup	Performance
Bedri et al [13], 2020	Eating and drinking	IMU[a], proximity, and camera	18 (laboratory) and 5 (real-world)	Laboratory and real world	F_1-score: 0.89
Shin et al [14], 2022	Eating	Piezoelectric and IMU	30	Real-world	F_1-score: 0.92
Bello et al [15], 2023	Facial expressions and eating and drinking	IMU, pressure, microphone, force, and piezoelectric	10	Real-world	F_1-score: 0.86 (expressions); 0.94 (eating or drinking)
Farooq and Sazonov [16], 2016	Eating and physical activity	Piezoelectric and IMU	10	Laboratory	F_1-score: 99% (eating vs activity)
Chung et al [17], 2017	Chewing and 5 other activities	Load cells	10	Laboratory	F_1-score: 94%
Zhang and Amft [18], 2018	Chewing and eating	EMG[b] sensors	10	Laboratory and real-world	Precision or recall: 95% (laboratory) and 78% (real-world)
This study	Chewing and eating	OCO optical sensors	128 (laboratory) and 8 (real-world)	Laboratory and real-world	F_1-score: 0.91 (laboratory) and 0.88 (real-world)

[a]IMU: inertial measurement unit.

[b]EMG: electromyography.

Limitations and Future Work

Regarding limitations, while our study included data from >100 participants, most of the data were collected in controlled setup. In contrast, our real-world data came from 8 participants observed for 2 days. To strengthen these findings, a broader and more prolonged study is needed.

Regarding the technical aspects of the system, the technology should be assessed across various demographics and age groups to ensure its generalizability. Factors such as the shape of a participant's head or nose might alter sensor position and data quality, and participants' adherence to wearing the device properly must be considered. Developing personalized models that adapt to individual eating patterns and preferences could improve the system's accuracy and user acceptance. Machine learning algorithms that learn and adapt to each user's unique behaviors over time could provide personalized and more accurate monitoring.

In terms of system application domains, combining the eating detection system with nutritional analysis tools could provide a comprehensive solution for monitoring not only eating behaviors but also dietary intake and nutritional quality, offering more actionable insights for health interventions. Furthermore, investigating the long-term impact of using such monitoring systems on health outcomes, including weight management, metabolic health, and behavior change, could provide valuable evidence for the efficacy of these technologies in promoting healthy eating habits and preventing chronic diseases.

Conclusions

Our study demonstrates the efficacy and feasibility of using optical tracking sensors integrated into smart glasses, particularly with the OCO technology, for noninvasive monitoring of eating behaviors, with a focus on chewing detection and eating segment detection. Through rigorous experimentation on data from 128 in-laboratory participants and 8 real-world participants, we determined that the proposed approach can accurately detect chewing activity in both laboratory and real-life scenario, highlighting the promising potential of this system for dietary monitoring applications.

Acknowledgments

This study was funded by emteq labs and by the Swiss National Science Foundation (project XAI-PAC, grant Z00P2_216405).

Conflicts of Interest

All authors, except for MG, are employed by Emteq Limited. MG has previously worked as a consultant for Emteq Limited, though he did not receive funding for this study. CN is a founder of Emteq Limited.

Multimedia Appendix 1
2D and 3D plots illustrating sensor data during eating, step-by-step predictions of the proposed method, false positive analyses categorized by activity type, and an extended review of related work on wearable technology for eating detection.
[DOCX File , 1000 KB - mhealth_v12i1e59469_app1.docx]

References

1. Neuhouser ML. The importance of healthy dietary patterns in chronic disease prevention. Nutr Res 2019 Oct;70:3-6 [FREE Full text] [doi: 10.1016/j.nutres.2018.06.002] [Medline: 30077352]
2. Jannasch F, Kröger J, Schulze MB. Dietary patterns and type 2 diabetes: a systematic literature review and meta-analysis of prospective studies. J Nutr 2017 Jun;147(6):1174-1182 [FREE Full text] [doi: 10.3945/jn.116.242552] [Medline: 28424256]
3. Zhang Y, Song M, Chan AT, Schernhammer ES, Wolpin BM, Stampfer MJ, et al. Unrestrained eating behavior and risk of digestive system cancers: a prospective cohort study. Am J Clin Nutr 2021 Nov 08;114(5):1612-1624 [FREE Full text] [doi: 10.1093/ajcn/nqab235] [Medline: 34293086]
4. Tourlouki E, Matalas AL, Panagiotakos DB. Dietary habits and cardiovascular disease risk in middle-aged and elderly populations: a review of evidence. Clin Interv Aging 2009;4:319-330 [FREE Full text] [doi: 10.2147/cia.s5697] [Medline: 19696896]
5. Baker C. Obesity statistics. House of Commons Library. 2023 Jan 12. URL: https://researchbriefings.files.parliament.uk/documents/SN03336/SN03336.pdf [accessed 2023-02-28]
6. Hales CM, Fryar CD, Carroll MD, Freedman DS, Ogden CL. Trends in obesity and severe obesity prevalence in US youth and adults by sex and age, 2007-2008 to 2015-2016. JAMA 2018 Apr 24;319(16):1723-1725 [FREE Full text] [doi: 10.1001/jama.2018.3060] [Medline: 29570750]
7. GBD 2017 Diet Collaborators. Health effects of dietary risks in 195 countries, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2019 May 11;393(10184):1958-1972 [FREE Full text] [doi: 10.1016/S0140-6736(19)30041-8] [Medline: 30954305]
8. Willett WC, Sampson L, Stampfer MJ, Rosner B, Bain C, Witschi J, et al. Reproducibility and validity of a semiquantitative food frequency questionnaire. Am J Epidemiol 1985 Jul;122(1):51-65. [doi: 10.1093/oxfordjournals.aje.a114086] [Medline: 4014201]
9. Willett W. Nutritional Epidemiology. Oxford, UK: Oxford Academic Press; Nov 07, 2012.
10. Dao MC, Subar AF, Warthon-Medina M, Cade JE, Burrows T, Golley RK, et al. Dietary assessment toolkits: an overview. Public Health Nutr 2018 Nov 15;22(3):404-418. [doi: 10.1017/s1368980018002951]
11. Evans D. MyFitnessPal. Br J Sports Med 2016 Jan 27;51(14):1101-1102. [doi: 10.1136/bjsports-2015-095538]
12. Saphala A, Zhang R, Amft O. Proximity-based eating event detection in smart eyeglasses with expert and data models. In: Proceedings of the 2022 ACM International Symposium on Wearable Computers. 2022 Presented at: ISWC '22; September 11-15, 2022; Cambridge, UK. [doi: 10.1145/3544794.3558476]

13. Bedri A, Li D, Khurana R, Bhuwalka K, Goel M. FitByte: automatic diet monitoring in unconstrained situations using multimodal sensing on eyeglasses. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. 2020 Presented at: CHI '20; April 25-30, 2020; Honolulu, HI. [doi: 10.1145/3313831.3376869]
14. Shin J, Lee S, Gong T, Yoon H, Roh H, Bianchi A, et al. MyDJ: sensing food intakes with an attachable on your eyeglass frame. In: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. 2022 Presented at: CHI '22; April 29-May 5, 2022; New Orleans, LA. [doi: 10.1145/3491102.3502041]
15. Bello H, Suh S, Zhou B, Lukowicz P. MeciFace: mechanomyography and inertial fusion-based glasses for edge real-time recognition of facial and eating activities. arXiv Preprint posted online on June 19, 2023. [doi: 10.48550/arXiv.2306.13674]
16. Farooq M, Sazonov E. A novel wearable device for food intake and physical activity recognition. Sensors 2016 Jul 11;16(7):1067. [doi: 10.3390/S16071067]
17. Chung J, Chung J, Oh W, Yoo Y, Lee WG, Bang H. A glasses-type wearable device for monitoring the patterns of food intake and facial activity. Sci Rep 2017 Jan 30;7:41690 [FREE Full text] [doi: 10.1038/srep41690] [Medline: 28134303]
18. Zhang R, Amft O. Monitoring chewing and eating in free-living using smart eyeglasses. IEEE J Biomed Health Inform 2018 Jan;22(1):23-32. [doi: 10.1109/JBHI.2017.2698523] [Medline: 28463209]
19. Archer J, Mavridou I, Stankoski S, Broulidakis MJ, Cleal A, Walas P, et al. OCOsense™ smart glasses for analyzing facial expressions using optomyographic sensors. IEEE Pervasive Comput 2023;22(3):18-26. [doi: 10.1109/mprv.2023.3276471]
20. Kiprijanovska I, Stankoski S, Broulidakis MJ, Archer J, Fatoorechi M, Gjoreski M, et al. Towards smart glasses for facial expression recognition using OMG and machine learning. Sci Rep 2023 Sep 25;13(1):16043 [FREE Full text] [doi: 10.1038/s41598-023-43135-5] [Medline: 37749176]
21. Yang JB, Nguyen MN, San PP, Li XL, Krishnaswamy S. Deep convolutional neural networks on multichannel time series for human activity recognition. In: Proceedings of the 24th International Conference on Artificial Intelligence. 2015 Presented at: IJCAI'15; July 25-31, 2015; Buenos Aires, Argentina. [doi: 10.5555/2832747.2832806]
22. Ignatov A. Real-time human activity recognition from accelerometer data using convolutional neural networks. Appl Soft Comput 2018 Jan;62:915-922. [doi: 10.1016/j.asoc.2017.09.027]
23. Ordóñez FJ, Roggen D. Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition. Sensors (Basel) 2016 Jan 18;16(1):115 [FREE Full text] [doi: 10.3390/s16010115] [Medline: 26797612]
24. Morales FJ, Roggen D. Deep convolutional feature transfer across mobile activity recognition domains, sensor modalities and locations. In: Proceedings of the 2016 ACM International Symposium on Wearable Computers. 2016 Presented at: ISWC '16; September 12-16, 2016; Heidelberg, Germany. [doi: 10.1145/2971763.2971764]
25. Murahari VS, Plötz T. On attention models for human activity recognition. arXiv Preprint posted online on May 19, 2018. [doi: 10.1145/3267242.3267287]
26. Stankoski S, Jordan M, Gjoreski H, Luštrek M. Smartwatch-based eating detection: data selection for machine learning from imbalanced data with imperfect labels. Sensors (Basel) 2021 Mar 09;21(5):1902 [FREE Full text] [doi: 10.3390/s21051902] [Medline: 33803121]
27. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods 2020 Mar;17(3):261-272 [FREE Full text] [doi: 10.1038/s41592-019-0686-2] [Medline: 32015543]
28. Martinez-Gomis J, Lujan-Climent M, Palau S, Bizar J, Salsench J, Peraire M. Relationship between chewing side preference and handedness and lateral asymmetry of peripheral factors. Arch Oral Biol 2009 Feb;54(2):101-107. [doi: 10.1016/j.archoralbio.2008.09.006] [Medline: 18947820]
29. Doulah A, Farooq M, Yang X, Parton J, McCrory MA, Higgins JA, et al. Meal microstructure characterization from sensor-based food intake detection. Front Nutr 2017 Jul 17;4:31 [FREE Full text] [doi: 10.3389/fnut.2017.00031] [Medline: 28770206]
30. Gemming L, Ni Mhurchu C. Dietary under-reporting: what foods and which meals are typically under-reported? Eur J Clin Nutr 2016 May;70(5):640-641. [doi: 10.1038/ejcn.2015.204] [Medline: 26669571]
31. Maruyama K, Sato S, Ohira T, Maeda K, Noda H, Kubota Y, et al. The joint impact on being overweight of self reported behaviours of eating quickly and eating until full: cross sectional survey. BMJ 2008 Oct 21;337:a2002 [FREE Full text] [doi: 10.1136/bmj.a2002] [Medline: 18940848]
32. Simon SL, Pan Z, Marden T, Zhou W, Ghosh T, Hossain D, et al. Body mass index and variability in meal duration and association with rate of eating. Front Nutr 2022 Jul 26;9:941001 [FREE Full text] [doi: 10.3389/fnut.2022.941001] [Medline: 35958246]
33. Otsuka R, Tamakoshi K, Yatsuya H, Wada K, Matsushita K, OuYang P, et al. Eating fast leads to insulin resistance: findings in middle-aged Japanese men and women. Prev Med 2008 Feb;46(2):154-159. [doi: 10.1016/j.ypmed.2007.07.031] [Medline: 17822753]
34. Argyrakopoulou G, Simati S, Dimitriadis G, Kokkinos A. How important is eating rate in the physiological response to food intake, control of body weight, and glycemia? Nutrients 2020 Jun 10;12(6):1734 [FREE Full text] [doi: 10.3390/nu12061734] [Medline: 32532002]

Abbreviations

CNN: convolutional neural network
ConvLSTM: convolutional long short-term memory
DL: deep learning
FLOPS: floating-point operations per second
FP: false positive
HMM: hidden Markov model
LSTM: long short-term memory

Edited by L Buis; submitted 12.04.24; peer-reviewed by M Anandhalli, B Bell, A Jitpattanakul; comments to author 22.05.24; revised version received 12.07.24; accepted 27.08.24; published 26.09.24.

Please cite as:
Stankoski S, Kiprijanovska I, Gjoreski M, Panchevski F, Sazdov B, Sofronievski B, Cleal A, Fatoorechi M, Nduka C, Gjoreski H Controlled and Real-Life Investigation of Optical Tracking Sensors in Smart Glasses for Monitoring Eating Behavior Using Deep Learning: Cross-Sectional Study
JMIR Mhealth Uhealth 2024;12:e59469
URL: https://mhealth.jmir.org/2024/1/e59469
doi:10.2196/59469
PMID:39325528

Validity of a Consumer-Based Wearable to Measure Clinical Parameters in Patients With Chronic Obstructive Pulmonary Disease and Healthy Controls: Observational Study

Fien Hermans[1,2], MSc; Eva Arents[1], MSc; Astrid Blondeel[2], PhD; Wim Janssens[3,4], MD, PhD; Nina Cardinaels[5], MD; Patrick Calders[1], PhD; Thierry Troosters[2], PhD; Eric Derom[5], MD, PhD; Heleen Demeyer[1,2], PhD

[1]Department of Rehabilitation Sciences, Ghent University, Corneel Heymanslaan 10, Entrance 46, Ghent, Belgium

[2]Department of Rehabilitation Sciences, KU Leuven, Leuven, Belgium

[3]Department of Chronic Diseases, Metabolism and Aging (CHROMETA) - BREATHE, KU Leuven, Leuven, Belgium

[4]Clinical Department of Respiratory Diseases, University Hospitals Leuven, Leuven, Belgium

[5]Clinical Department of Respiratory Medicine, Ghent University Hospital, Ghent, Belgium

Corresponding Author:
Heleen Demeyer, PhD
Department of Rehabilitation Sciences, Ghent University, , Corneel Heymanslaan 10, Entrance 46, Ghent, , Belgium

Abstract

Background: Consumer-based wearables are becoming more popular and provide opportunities to track individual's clinical parameters remotely. However, literature about their criterion and known-groups validity is scarce.

Objective: This study aimed to assess the validity of the Fitbit Charge 4, a wrist-worn consumer-based wearable, to measure clinical parameters (ie, daily step count, resting heart rate [RHR], heart rate variability [HRV], respiratory rate [RR], and oxygen saturation) in patients with chronic obstructive pulmonary disease (COPD) and healthy controls in free-living conditions in Belgium by comparing it with medical-grade devices.

Methods: Participants wore the Fitbit Charge 4 along with three medical-grade devices: (1) Dynaport MoveMonitor for 7 days, retrieving daily step count; (2) Polar H10 for 5 days, retrieving RHR, HRV, and RR; and (3) Nonin WristOX$_2$ 3150 for 4 nights, retrieving oxygen saturation. Criterion validity was assessed by investigating the agreement between day-by-day measures of the Fitbit Charge 4 and the corresponding reference devices. Known-groups validity was assessed by comparing patients with COPD and healthy controls.

Results: Data of 30 patients with COPD and 25 age- and gender-matched healthy controls resulted in good agreement between the Fitbit Charge 4 and the corresponding reference device for measuring daily step count (intraclass correlation coefficient [ICC$_{2,1}$]=0.79 and ICC$_{2,1}$=0.85, respectively), RHR (ICC$_{2,1}$=0.80 and ICC$_{2,1}$=0.79, respectively), and RR (ICC$_{2,1}$=0.84 and ICC$_{2,1}$=0.77, respectively). The agreement for HRV was moderate (healthy controls: ICC$_{2,1}$=0.69) to strong (COPD: ICC$_{2,1}$=0.87). The agreement in measuring oxygen saturation in patients with COPD was poor (ICC$_{2,1}$=0.32). The Fitbit device overestimated the daily step count and underestimated HRV in both groups. While RHR and RR were overestimated in healthy controls, no difference was observed in patients with COPD. Oxygen saturation was overestimated in patients with COPD. The Fitbit Charge 4 detected significant differences in daily step count, RHR, and RR between patients with COPD and healthy controls, similar to those identified by the reference devices, supporting known-groups validity.

Conclusions: Although the Fitbit Charge 4 shows mainly moderate to good agreement, measures of clinical parameters deviated from the reference devices, indicating that monitoring patients remotely and interpreting parameters requires caution. Differences in clinical parameters between patients with COPD and healthy controls that were measured by the reference devices were all detected by the Fitbit Charge 4.

(JMIR Mhealth Uhealth 2024;12:e56027) doi:10.2196/56027

KEYWORDS

chronic obstructive pulmonary disease; COPD; wearable; Fitbit; clinical parameters; physical activity; validity; observational study; wrist-worn wearable; heart rate; heart rate variability; respiratory rate; oxygen saturation; devices; monitoring

Introduction

In recent years, consumer-based wearables have become increasingly popular due to their relatively low cost, ease of use, and ability to provide real-time feedback on several clinical parameters such as heart rate and step count. In Europe, 26% of 45 - to 74-year-olds used a wearable device in 2022 [1]. Alongside their rising popularity, wearable technologies also advanced significantly. Contemporary generation wrist-worn wearables use photoplethysmography technology, which measures the volumetric variations of blood circulation via an infrared light [2]. This technology enables to measure multiple parameters of autonomic function (ie, resting heart rate [RHR], heart rate variability [HRV], and respiratory rate [RR]) and oxygen saturation (SpO_2) [3].

Chronic obstructive pulmonary disease (COPD) is the third leading cause of death worldwide [4]. Although COPD is a disease characterized by respiratory symptoms and exercise intolerance related to abnormalities of the airways, the alveoli, or both, COPD seems also to alter the autonomic nervous system [5]. The autonomic nervous system adjusts heart rate, blood pressure, and RR in response to internal and external stimuli [6]. Patients with COPD show an elevated RHR, a reduced HRV and an increased RR compared to healthy controls [5,7]. Besides the autonomic function, SpO_2 is a clinical parameter that differs between patients with COPD and healthy controls, with patients with COPD having a lower SpO_2 due to alterations in gas exchange and experiencing a sudden drop in SpO_2 during exercise and when experiencing an acute exacerbation [4,8]. Both parameters of autonomic function and SpO_2 are prognostic markers of mortality and thus important to monitor [5,8,9].

Wearables potentially enable continuous monitoring of clinical parameters remotely and unobtrusively over a long period of time [10]. Whereas manufacturers specify that consumer-based wearables are not intended for medical purposes as they do not qualify as medical devices, continuously monitoring these parameters would provide information on the management of the patients' health at home, and investigate the effectiveness of interventions based on measures outside clinical visits in an easy way [11]. Moreover, given that these clinical parameters are linked to the worsening of health in various chronic diseases, it is tempting to actively monitor these parameters remotely, as this may lead to earlier detection of patients' deterioration or provide an explanation for reduced engagement in physical activity [12-16].

Previous literature showed that Fitbit wearables are valid devices for monitoring daily step counts in healthy individuals and can be used to monitor patterns of physical activity in patients with COPD [17-19]. However, literature in the healthy population on the validity of Fitbit measuring other clinical parameters is scarce, and no data are available for patients with COPD [18,20-22]. These studies used various devices, including medical-grade devices (such as the ActiGraph GT3X+ and Dynaport MoveMonitor for daily step count, and the Polar H7 for heart rate), as well as gold-standard measurements (such as ECG Holter monitoring for heart rate), to assess the validity of Fitbit wearables.

The aim of this observational study was to investigate the criterion validity and the known-groups validity of a consumer-based wearable for monitoring physical activity (ie, daily step count) and parameters of autonomic function (ie, RHR, HRV, and RR) in patients with COPD and a reference population consisting of healthy age-matched controls. Furthermore, the criterion validity and known-groups validity of this consumer-based wearable for monitoring SpO_2 was examined in patients with COPD.

Methods

Population and Design

This observational study was nested in a randomized controlled trial investigating the long-term effects of a telecoaching intervention in patients with COPD (NCT04139200). All patients with COPD included via Ghent University Hospital (Ghent, Belgium) and examined between April 2022 and June 2023 were enrolled in this substudy, ensuring representation of the entire patient cohort at this site. Patients aged 40 years and older with a smoking history of at least 10 pack years and with a clinical diagnosis of COPD (confirmed by spirometry [Tiffeneau-index<70%]) but no history of exacerbations in the past month were eligible to participate. Patients were excluded if they had orthopedic or other problems preventing them from improving physical activity, had undergone lung transplantation, were involved in a multidisciplinary rehabilitation program, or were unable to learn to work with electronic devices.

In addition, healthy controls were recruited between November 2022 and August 2023. These participants were 50 - 80 years old, had never smoked or had stopped smoking more than 20 years before inclusion, had a Tiffeneau-index ≥70%, had no chronic health problems or orthopedic problems preventing them from being physically active, and did not participate in a rehabilitation program.

Ethical Considerations

This study consisted of a single clinical visit and a follow-up period of 7 days. Both studies were approved by the ethical committee of Ghent University Hospital (BC-10267 [COPD] and ONZ-2022 - 0387 [healthy controls]). All participants signed the informed consent prior to data collection and were assigned a study ID number. No compensation was provided for participation in this study.

Clinical Assessments

At the clinical visit, the following assessments were performed in both groups: (1) sociodemographic and clinical data (age, sex, height, weight, smoking history, and medication intake); (2) postbronchodilator spirometry according to European Respiratory Society (ERS) – American Thoracic Society (ATS) guidelines [23]; (3) functional exercise capacity based on the best of two 6-minute walk tests following ERS-ATS guidelines [24]; and (4) Composite Autonomic Symptom Score (COMPASS) 31 questionnaire questioning autonomic nervous system symptoms [25]. In patients with COPD, the modified Medical Research Council scale (mMRC) questioning dyspnea and COPD Assessment Test (CAT) questioning health status were collected additionally [26,27].

During the 7-day wearing period following the clinical visit, participants were asked to wear a wrist-worn wearable, the Fitbit Charge 4, along with 3 approved medical-grade devices. During this wearing period, participants completed a daily diary including the time of waking up and going to bed, intake of medication, and consumption of beverages influencing the heart rate (eg, coffee, tea, alcohol, and energy drinks).

Wearable

The Fitbit Charge 4 (Fitbit Inc) is a triaxial consumer-based wearable worn on the nondominant wrist of the participant. This device records daily step count (accelerometry), as well as RHR, HRV, RR, and SpO_2 (photoplethysmography). These variables were extracted as day-by-day outcomes from the Fitbit platform, as calculated by the proprietary algorithms. Fitbit defines RHR as the heart rate while in a relaxed state during both sleep and being awake. HRV and RR are defined as the variation of duration between heartbeats, expressed as the root mean square of successive differences (RMSSD) in milliseconds (ms), and the number of breaths per minute during the night, respectively, when more than 3 hours of continuous sleep are recorded. Transcutaneous SpO_2 was defined as the average hemoglobin SpO_2 during the night (>3 hours of continuous sleep). Participants were instructed to wear the device for 7 consecutive days (24 hours per day). The device had a battery life of up to 7 days, but participants were advised to charge the battery when the battery level dropped below 10% and when they were not performing any activity.

Reference Devices

Physical Activity—Dynaport MoveMonitor

The Dynaport MoveMonitor (DAM; McRoberts) is a triaxial accelerometer validated to objectively measure physical activity in patients with COPD and healthy controls [28,29]. Participants were instructed to wear the monitor at the lower back during waking hours for 7 consecutive days, except for bathing and water activities, according to current recommendations [30]. Days with a wearing time lower than 8 hours were excluded for further analyses [30]. Wearing time and daily step count were extracted from the monitor for further analysis.

Cardiac Autonomic Function—Polar H10

The Polar H10 sensor chest strap (Polar Electro Oy) is a validated device to capture heart rate and HRV in healthy subjects [31]. This sensor was moistened before being applied below the chest muscles of the participants, as described by the manufacturer. To collect heart rate data for 100 hours, participants wore a corresponding Polar Ignite 2 watch (at the preferred wrist), which was used as a Polar H10 data logger. They were instructed to wear these devices for 5 consecutive days (24 hours per day), except for bathing and water activities, based on recommendations and the maximum recording capacity of 100 hours. Data were recorded using a 1-second time interval. Heart rate and beat-to-beat RR intervals (time between 2 successive R-waves of the QRS signal on the electrocardiogram) were extracted from the device using the Polar Flow web service.

To obtain the RHR, the average heart rate during the night (actual sleep time, ie, starting 1 hour later than the reported time of going to bed and ending 1 hour earlier than the reported time of waking up) was calculated to best match Fitbit's definition. Nights with 3 hours of sleep or less were excluded for further analyses.

The beat-to-beat RR intervals (time between 2 successive R-waves of the QRS signal on the electrocardiogram) were transferred to Kubios HRV Software (version 4.0.1, Kuopio) to analyze HRV (expressed as RMSSD in ms) and RR [32]. The actual sleep time as reported in the diary was entered into the software. Nights with 3 hours of sleep or less were excluded for further analyses to best fit with Fitbit's definition.

SpO_2 (COPD Only)—Nonin WristOX$_2$ 3150

The wearable finger pulse oximeter Nonin $WristOX_2$ 3150 (Nonin Medical Inc) was used as a reference device for measuring SpO_2, as it has become a commonly used device for home monitoring of patients with COPD and has a high accuracy (±2%) according to the manufacturer [33]. Patients with COPD were instructed to wear the device on the index finger of their dominant hand for 4 consecutive nights (limited battery life of 48 hours using continuous measurements, allowing up to 12 hours of monitoring per night). Data were recorded using a 4-second time interval. Data were stored on the internal memory of the device and downloaded after the 7-day follow-up period using the nVISION software (version 6.5.1.2, Nonin Medical Inc). The average SpO_2 during the night measured by Nonin was calculated after removing impossible values. Nights with 3 hours of sleep (as judged by the Fitbit device) or less were excluded for further analyses to best fit with Fitbit's definition.

Statistical Analyses

The sample size of this substudy was chosen to align with prior research within this field [18]. All statistical analyses were performed using the SAS statistical package (version 9.4, SAS Institute). Data are presented as mean (SD) or median (IQR), as appropriate after testing for normality using the Shapiro-Wilk test. Statistical significance was set at $P<.05$ for all analyses.

Criterion validity was investigated by comparing day-by-day data obtained by the Fitbit Charge 4 with the reference devices by use of a 2-tailed paired *t* test, Bland-Altman plots, and intraclass correlation coefficients ($ICC_{2,1}$). The cutoffs that were used to interpret the findings were ICC<0.50 as "poor," ICC=0.50 - 0.75 as "moderate," ICC=0.75 - 0.90 as "good," and ICC>0.90 as "excellent" [34]. These analyses were performed for patients with COPD and healthy controls separately. To evaluate whether the Fitbit Charge 4 is able to pick up day-by-day fluctuations, the delta (day minus day-1) of each measured clinical parameter was calculated on consecutive days. The agreement between the Fitbit Charge 4 and the corresponding reference device with regard to day-by-day fluctuations was determined via Pearson correlation. The correlation was interpreted using the cutoffs $r<0.30$ classed as "no correlation," $r=0.30$ - 0.50 as "weak correlation," $r=0.50$ - 0.70 as "moderate correlation," $r=0.70$ - 0.90 as "strong correlation," and $r>0.90$ as "very strong correlation" [35]. Next, known-groups validity was assessed by investigating

whether the differences between patients with COPD and healthy controls were picked up by the reference device and the Fitbit Charge 4 in the same way. For this, the outcomes were compared between patients with COPD and healthy controls for both devices using an unpaired *t* test. The differences between both groups were examined by an interaction effect based on a mixed model analysis. As a sensitivity analysis, all analyses were performed with the exclusion of patients taking beta-blockers.

Results

Patient Characteristics

In total, 32 patients with COPD and 26 age- and gender-matched healthy controls were included in this study. Overall, valid data of 30 patients with COPD (1 patient dropped out and 1 patient was not willing to wear the additional devices) and 25 healthy controls (1 participant was excluded because spirometry displayed an obstructive syndrome) were obtained, but across the clinical parameters, there is a variation in participants included in the analyses. This variation can be attributed to various reasons, such as technical problems (the Fitbit not capturing clinical parameters at night, sudden interruptions in the measurement by the reference device, battery issues), as well as participants forgetting to wear the (reference) device. Baseline characteristics are shown in Table 1.

Table . Characteristics of all participants included in the analyses.

	Patients with COPD[a] (n=30)	Healthy controls (n=25)	*P* value
Age (years), mean (SD)	70 (7)	68 (7)	.28
Sex (male), n (%)	23 (77)	20 (80)	.77
BMI (kg/m^2), mean (SD)	27 (6)	28 (4)	.86
Forced expiratory volume in the first second (% predicted), mean (SD)	54 (15)	109 (15)	<.001
Severity of COPD, n (%)			
I-II	2 (7)-15 (50)	N/A[b]	N/A
III-IV	12 (40)-1 (3)	N/A	N/A
Use of β-blockers, n (%)	8 (27)	4 (16)	.34
Current smokers, n (%)	1 (3)	0 (0)	.36
6MWD[c] (m), mean (SD)	482 (95)	618 (81)	<.001
6MWD (% predicted), mean (SD)	78 (14)	97 (10)	<.001
COMPASS 31[d] (score), median (IQR)	11 (4 - 21)	5 (2-12)	.005
mMRC[e] (score), median (IQR)	2 (1-2)	N/A	N/A
CAT[f] (score), mean (SD)	15 (6)	N/A	N/A

[a]COPD: chronic obstructive pulmonary disease.

[b]N/A: not applicable.

[c]6MWD: 6-minute walking distance.

[d]COMPASS 31: Composite Autonomic Symptom Score 31 questionnaire (0 - 100); a higher score indicated more symptoms of autonomic dysfunction.

[e]mMRC: modified Medical Research Council dyspnea scale (0 - 4); a higher score indicated more dyspnea.

[f]CAT: COPD Assessment Test (0 - 40); a higher score indicated a worse health status.

Criterion Validity

Physical Activity

Daily step count was analyzed based on 199 (min-max 4 to 7 days per patient, n=30) and 157 (4 to 7 days per participant, n=24) overlapping data points for patients with COPD and healthy controls, respectively. As expected, patients with COPD were less active compared to healthy controls (Table 2).

Table . Average step count, resting heart rate, heart rate variability, and respiratory rate in patients with COPD[a] and healthy controls.

	Patients with COPD (n=30), mean (SD)	Healthy controls (n=25), mean (SD)	Difference, mean (SD)	*P* value[b]
Daily steps (steps/day)				
Dynaport MoveMonitor	5625 (3618)	7933 (4263)	−2308 (3915)	<.001
Fitbit Charge 4	7423 (4325)	9055 (4956)	−1632 (4613)	.001
Resting heart rate (beats/min)				
Polar H10	70 (8)	60 (10)	10 (9)	<.001
Fitbit Charge 4	70 (8)	64 (10)	6 (9)	<.001
Heart rate variability (ms)				
Polar H10	27 (19)	33 (21)	−6 (20)	.16
Fitbit Charge 4	24 (14)	26 (13)	−2 (14)	.53
Respiratory rate (breaths/min)				
Polar H10	16 (2)	14 (2)	2 (2)	<.001
Fitbit Charge 4	16 (3)	15 (2)	1 (2)	.01

[a]COPD: chronic obstructive pulmonary disease.

[b]*P* values based on unpaired *t* test.

In both groups, the mean step count measured by the Fitbit Charge 4 was significantly higher compared to DAM (mean, SD; COPD: Δ+1798, SD 2070 steps/day; *P*<.001 and healthy controls: Δ+1122, SD 2297 steps/day; *P*<.001). However, in both groups, a good agreement between the devices was found (COPD: $ICC_{2,1}$=0.79; 95% CI 0.36 - 0.90 and healthy controls: $ICC_{2,1}$=0.85; 95% CI 0.74 - 0.91). These findings are supported by the Bland-Altman plots depicted in Figure 1A and 1B. Concurrent validity with associated scatterplots can be found in Multimedia Appendix 1.

Figure 1. Bland-Altman plots with mean and 95% CI for Fitbit Charge 4 compared to the reference devices. (**A, B**) Daily steps (steps/day) measured by Fitbit Charge 4 and DAM (COPD: n=30 and healthy controls: n=24). (**C, D**) Resting heart rate (beats/min) measured by Fitbit Charge 4 and Polar H10 (COPD: n=25 and healthy controls: n=20). (**E, F**) Heart rate variability (RMSSD in ms) measured by Fitbit Charge 4 and Polar H10 (COPD: n=22 and healthy controls: n=16). (**G, H**) Respiratory rate (breaths/min) measured by Fitbit Charge 4 and Polar H10 (COPD: n=21 and healthy controls: n=17). (**I**) Oxygen saturation (%) measured by Fitbit Charge 4 and Nonin WristOX$_2$ 3150 (COPD: n=19). Large open dots represent the mean individual outcome per patient. Small dots represent daily data. Mean and 95% CI are calculated based on average data. COPD: chronic obstructive pulmonary disease; DAM: Dynaport MoveMonitor; RMSSD: root mean square of successive differences.

Figure 2A shows the scatterplot of the day-by-day fluctuations in daily step count measured with the Fitbit Charge 4 and the DAM. A strong to very strong association was observed between both devices in patients with COPD (r=0.71) and healthy controls (r=0.91), respectively.

Figure 2. Scatterplots of the day-by-day fluctuations in daily step count, parameters of autonomic function, and SpO_2 in patients with COPD and healthy controls. (**A**) Day-by-day fluctuations in daily step count (COPD: 169 data points and healthy controls: 133 data points). (**B**) Day-by-day fluctuations in resting heart rate (COPD: 42 data points and healthy controls: 39 data points). (**C**) Day-by-day fluctuations in heart rate variability (COPD: 27 data points and healthy controls: 28 data points). (**D**) Day-by-day fluctuations in respiratory rate (COPD: 24 data points and healthy controls: 30 data points). (**E**) Day-by-day fluctuations in oxygen saturation (COPD: 42 data points). Patients with COPD are depicted in the closed dots and healthy controls are depicted in the open dots. An identity line is displayed on the scatterplots. COPD: chronic obstructive pulmonary disease; DAM: Dynaport MoveMonitor; HRV: heart rate variability; RHR: resting heart rate; RR: respiratory rate; SpO_2: oxygen saturation.

Resting Heart Rate

RHR was analyzed based on 66 (min-max 1 to 4 days per patient, n=25) and 59 (1 to 4 days per healthy control, n=20) overlapping data points. RHR was higher in patients with COPD compared to healthy controls (Table 2).

In the COPD group, RHR measured by the Fitbit showed no difference compared to Polar H10 (Δ−0.3, SD 5 beats/min; *P*=.67), whereas Fitbit significantly overestimated RHR in the healthy control group (Δ+4, SD 6 beats/min; *P*<.001). In both groups, Fitbit Charge 4 and Polar H10 showed a good agreement for assessing RHR (COPD: $ICC_{2,1}$=0.80; 95% CI 0.70 - 0.87 and healthy controls: $ICC_{2,1}$=0.79; 95% CI 0.50 - 0.90). The findings are depicted in Bland-Altman plots (Figure 1C and 1D). A strong association between the devices was observed in both groups (Multimedia Appendix 1).

A weak negative association was observed in patients with COPD for picking up day-by-day fluctuations in RHR (*r*=−0.35; Figure 2B).

Heart Rate Variability

HRV analyses are based on 49 (min-max 1 to 3 days per patient, n=22) and 44 (1 to 4 days per participant, n=16) overlapping data points for patients with COPD and healthy controls, respectively. No difference in HRV was observed between patients with COPD and healthy controls (Table 2).

The Fitbit Charge 4 significantly underestimated HRV in both groups (respectively in COPD and healthy controls: Δ−3, SD 8 ms; *P*=.03 and Δ−7, SD 13 ms; *P*=.001). In patients with COPD, a good agreement was observed between both devices ($ICC_{2,1}$=0.87; 95% CI 0.77 - 0.93). Whereas in the healthy control group, a moderate agreement was found for measuring HRV ($ICC_{2,1}$=0.69; 95% CI 0.43 - 0.83). Bland-Altman plots (Figure 1E and 1F) represent these results. Scatterplots showing the concurrent validity can be found in Multimedia Appendix 1.

In patients with COPD, a strong association was found for picking up day-by-day fluctuations in HRV (*r*=0.83), whereas healthy controls exhibited a moderate association (*r*=0.64; Figure 2C).

Respiratory Rate

RR was analyzed based on 45 (min-max 1 to 5 days per patient, n=21) and 47 (1 to 4 days per healthy control, n=17) overlapping data points. RR was higher in patients with COPD compared to healthy controls (Table 2).

In patients with COPD, no difference was detected in RR measured by Fitbit Charge 4 compared to Polar H10 (Δ+0.3, SD 1 breaths/min; *P*=.25). However, in the healthy control group, the Fitbit Charge 4 significantly overestimated the RR (Δ+1, SD 1 breaths/min; *P*<.001). A good agreement was observed in patients with COPD, as well as healthy controls (COPD: $ICC_{2,1}$=0.84; 95% CI 0.72 - 0.91 and healthy controls: $ICC_{2,1}$=0.77; 95% CI 0.44 - 0.89). These findings are supported by the Bland-Altman plots depicted in Figure 1G and 1H. In both groups, a strong association was found (Multimedia Appendix 1).

A weak association was observed in both groups for picking up day-by-day fluctuations in RR (COPD: *r*=0.47 and healthy controls: *r*=0.50; Figure 2D).

Oxygen Saturation

SpO_2 was analyzed based on 61 (min-max 1 to 5 days per patient, n=19) overlapping data points in patients with COPD. Fitbit Charge 4 significantly overestimated SpO_2 (Δ+2, SD 2%; *P*<.001). A poor agreement between both devices was found ($ICC_{2,1}$=0.32; 95% CI −0.10 - 0.65). The Bland-Altman analysis is shown in Figure 1I.

Figure 2E displays the scatterplot of the day-by-day fluctuations measured with the Fitbit Charge 4 and the Nonin $WristOX_2$ 3150, showing a weak association for SpO_2 in patients with COPD (*r*=0.40).

Known-Groups Validity

The significant differences between patients with COPD and healthy controls in daily steps, RHR, and RR identified by the reference devices were picked up in a similar way by the Fitbit Charge 4 (see Table 2). The difference between both groups is significantly smaller when assessing RHR or RR using the Fitbit Charge 4 compared to the reference device (*P*<.05).

Sensitivity Analysis

Excluding participants on stable doses of beta-blockers had minimal impact on the results. The agreement between the Fitbit Charge 4 and its corresponding reference devices remained unchanged. The exclusion of participants taking beta-blockers had no effect on the known-groups validity (Multimedia Appendix 2).

Discussion

Principal Findings

This study, which aimed to investigate the validity of a consumer-based wearable, the Fitbit Charge 4, in patients with COPD and healthy controls found mixed results for criterion validity. First, the Fitbit Charge 4 significantly overestimated daily step count and significantly underestimated HRV in patients with COPD and healthy controls. In patients with COPD, RHR and RR are not different between the Fitbit Charge 4 and the Polar H10, but both parameters were overestimated in the healthy control group. The Fitbit Charge 4 significantly overestimated SpO_2 in patients with COPD. Second, the agreement between the Fitbit Charge 4 and the corresponding reference devices is moderate to good for monitoring most clinical parameters (ie, daily steps, RHR, HRV, and RR) but poor for tracking SpO_2. The Fitbit is able to pick up day-by-day fluctuations in daily step count and HRV but lacks accuracy to pick up the small day-by-day fluctuations in RHR, RR, and SpO_2. The known-groups validity of the Fitbit Charge 4 is good. All expected differences between patients with COPD and age- and gender-matched healthy controls are picked up by the wearable.

Our results are consistent with previous research in older adults in free-living conditions showing that the daily step count

measured by the Fitbit Charge, a wrist-worn wearable, is highly correlated with the daily step count measured by a validated accelerometer, but the Fitbit Charge significantly overestimates the daily step count [21,36,37]. Previous studies conducted in other patient cohorts within our department showed that the wrist-worn Fitbit device significantly overestimated the daily step count in healthy individuals and cancer survivors, whereas it did not in people with Parkinson disease [38,39]. Blondeel et al [19] concluded that the Fitbit Alta, also a wrist-worn wearable, did not significantly overestimate step count in patients with COPD, but did in healthy controls. This corresponds well with our findings, although in our sample, the wearable also overestimated daily step count in patients with COPD. This could potentially be explained by the inclusion of patients with a better functional exercise tolerance (6-minute walking distance of 482 m vs 454 m) in this study.

A few studies examined the validity of Fitbit measuring the heart rate during sleep among healthy adults in different situations (ie, home environment and laboratory-based setting). One study in healthy adolescents showed that Fitbit significantly underestimated the heart rate by 0.9 beats/min on average as compared to ECG, while another study in healthy adults found no difference between Fitbit and the reference device [40,41]. These findings are consistent with our results in the COPD group but are inconsistent with our results showing that Fitbit significantly overestimates RHR by 4 beats/min in the healthy control group. The discrepancy in results can possibly be attributed to the difference in how the definition is applied, with Fitbit defining RHR as the heart rate while in a relaxed state during both sleep and being awake, whereas we calculated RHR from Polar for sleep time values only because we do not have exact information on how RHR is calculated by Fitbit.

To the best of our knowledge, the validity of Fitbit measuring HRV, RR, and SpO_2 has not been investigated so far. However, some research has been performed regarding other wearables also using photoplethysmography technology (eg, Apple watch, Garmin, and Polar), showing a large range for measuring HRV (ICC ranging from 0.24 to 0.99) depending on the wearable used in a laboratory setting in healthy adults [42]. These results are consistent with our findings in healthy controls ($ICC_{2,1}$=0.69). Interestingly, we found a higher agreement in patients with COPD ($ICC_{2,1}$=0.87), but no previous research has been conducted on the validity of wearables measuring HRV among this population.

Data on the validity of a wrist-worn wearable estimating RR based on photoplethysmography have not been published so far. Existing studies have only focused on the validity of devices measuring this parameter using different technologies (ie, wearable biosensors and wearable pressure sensors) [43,44]. This difference in technology makes it challenging to compare these results.

Several studies indicate that the Apple Watch Series 6 shows a good agreement for measuring SpO_2 compared to the gold-standard (ie, arterial blood gas analyses) or reference devices (ie, oximeters) in patients with lung diseases (ICCs ranging from 0.90 to 0.94), as well as healthy individuals (mean bias 1.7, SD 2.1%) [45-47]. These results are in contrast with our findings achieving an ICC of 0.32. The results of Garmin concerning the measurements of SpO_2 in healthy individuals vary widely, with ICCs ranging from 0.28 to 0.55, which is more in line with our results [48,49]. In our study, the SpO_2 was measured during sleep, where the contact between the watch and the skin was maybe occasionally suboptimal due to unintentionally lying on the wearable, leading to divergent measurements through photoplethysmography. This is in contrast to the other studies where the measurement of SpO_2 was conducted under controlled settings over a short period of time (maximum 10 minutes) while being awake.

Regarding the known-groups validity, the differences in daily step count (lower in COPD), RHR (higher in COPD), and RR (higher in COPD) were in line with previous research [7,9,19]. We could not confirm a difference in HRV, which was shown to be lower in COPD in previous research (COPD and healthy controls: 27, SD 19 ms and 33, SD 21 ms; *P*=.16 [this study] vs 11, SD 3 ms and 19, SD 7 ms; *P*=.002, respectively) [50]. This study confirms that previously laboratory-based findings are now also observed in free-living situations.

It is remarkable that our findings indicate that Fitbit estimates certain clinical parameters, such as RHR, HRV, and RR, better in patients with COPD as compared to healthy controls. In this regard, we speculate that faster aging of the skin and skin thinning, a typical feature in patients with COPD on inhaled and systemic steroids, improve reflection of lights for measuring clinical parameters through photoplethysmography compared to slightly thicker skins in healthy controls as the distance of the infrared light to measure volumetric changes in the blood is reduced [51-53].

Clinical Implications

While our results suggest that Fitbit can measure certain clinical parameters better in patients with COPD than in healthy controls, it is not the best choice if one aims to measure important parameters of autonomic function (ie, RHR and RR) as well as SpO_2 in patients with COPD, as the wearable fails to pick up day-by-day fluctuations. Nevertheless, Fitbit can be used for commercial purposes and well-being monitoring, including data on daily step counts and HRV. Advanced and accurate (medical) devices are more appropriate for remotely monitoring clinical parameters. However, these results are only based on a 5 - to 7-day assessment in a stable situation. Longer time series of data also including larger fluctuations in the clinical parameters (eg, when patients experience a deterioration in their health) are needed to confirm these findings.

Strengths and Limitations

This study offers valuable insights into the validity of a popular wearable device to estimate various clinical parameters during a free-living situation. Our approach involved studying both patients with COPD, as well as age- and gender-matched healthy controls, aiming to establish evidence in both populations. Even though 4 of the 25 (16%) control participants were using beta-blockers, they can be considered as a representative sample of the general population [54]. Nonetheless, this study has some limitations. First, it should be mentioned that Fitbit operates as

a black box, using proprietary algorithms of which its details are not disclosed. This makes it challenging to establish an agreement with reference devices. Second, technical issues arose, preventing the collection of certain data. In some cases, Fitbit failed to measure parameters of autonomic function or SpO_2, rendering a comparison with reference devices impossible. Third, validated medical-grade devices that show good agreement with gold-standard methods were used in this study to assess the validity of the Fitbit Charge 4 in free-living conditions. Unfortunately, gold-standard devices (ie, video recording or manual step counting [physical activity] and ECG Holter monitoring [HRV]) are restricted to laboratory-based measurements. However, it is important to acknowledge that the reference devices may still involve measurement inaccuracies. Given that the measurements were conducted in daily life, the results have a high level of generalizability. Fourth, a selection bias within our healthy controls is difficult to avoid, as probably the more motivated individuals are more prone to participate in these kinds of studies. However, the clinical parameters in our control group align with those reported in the general population [55,56].

Conclusions

Both in patients with COPD and healthy controls, measures of clinical parameters collected by the commercially Fitbit Charge 4 showed moderate to good agreement with the reference devices. However, these measures deviated significantly. In patients with COPD, the Fitbit Charge 4 is accurate in measuring RHR and RR. The wearable lacks accuracy to pick up day-by-day fluctuations in RHR, RR, and SpO_2; hence, the Fitbit Charge 4 should be used with caution when information on clinical parameters is collected over a short period of time.

Acknowledgments

The authors would like to thank the colleagues from the pulmonary function testing at Ghent University Hospital for their assistance with the clinical assessments. They would also like to thank Dr Nathalie Michels for providing access to the Kubios software and for sharing her expertise on its usage. This research was funded by the Flemish Research Foundation (FWO project G0C0720N). AB is a predoctoral research fellow of the Flemish Research Foundation (FWO-Flanders 1194320N). WJ is a senior clinical research fellow of the Flemish Research Foundation (FWO-Flanders). HD is a postdoctoral research fellow of the Flemish Research Foundation (FWO-Flanders 12H7517N).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Concurrent validity between the Fitbit Charge 4 and the corresponding reference device for measuring daily step count, resting heart rate, heart rate variability, and respiratory rate.
[DOCX File, 358 KB - mhealth_v12i1e56027_app1.docx]

Multimedia Appendix 2
Sensitivity analysis—criterion and known-groups validity of the Fitbit Charge 4 in participants not taking beta-blockers.
[DOCX File, 21 KB - mhealth_v12i1e56027_app2.docx]

References

1. Growing importance of internet-connected devices. Eurostat. URL: https://ec.europa.eu/eurostat/web/products-eurostat-news/w/ddn-20230829-1 [accessed 2023-08-25]
2. Tamura T, Maeda Y, Sekine M, Yoshida M. Wearable photoplethysmographic sensors—past and present. Electronics (Basel) 2014;3(2):282-302. [doi: 10.3390/electronics3020282]
3. Castaneda D, Esparza A, Ghamari M, Soltanpur C, Nazeran H. A review on wearable photoplethysmography sensors and their potential future applications in health care. Int J Biosens Bioelectron 2018;4(4):195-202. [doi: 10.15406/ijbsbe.2018.04.00125] [Medline: 30906922]
4. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: 2024 report. Global Initiative for Chronic Obstructive Lung Disease. 2024. URL: https://goldcopd.org/2024-gold-report/ [accessed 2023-08-02]
5. van Gestel AJR, Steier J. Autonomic dysfunction in patients with chronic obstructive pulmonary disease (COPD). J Thorac Dis 2010 Dec;2(4):215-222. [doi: 10.3978/j.issn.2072-1439.2010.02.04.5] [Medline: 22263050]
6. Wehrwein EA, Orer HS, Barman SM. Overview of the anatomy, physiology, and pharmacology of the autonomic nervous system. Compr Physiol 2016 Jun 13;6(3):1239-1278. [doi: 10.1002/cphy.c150037] [Medline: 27347892]
7. Yentes JM, Fallahtafti F, Denton W, Rennard SI. COPD patients have a restricted breathing pattern that persists with increased metabolic demands. COPD 2020 Jun;17(3):245-252. [doi: 10.1080/15412555.2020.1750578] [Medline: 32301362]
8. Kent BD, Mitchell PD, McNicholas WT. Hypoxemia in patients with COPD: cause, effects, and disease progression. Int J Chron Obstruct Pulmon Dis 2011;6:199-208. [doi: 10.2147/COPD.S10611] [Medline: 21660297]

9. Jensen MT, Marott JL, Lange P, et al. Resting heart rate is a predictor of mortality in COPD. Eur Respir J 2013 Aug;42(2):341-349. [doi: 10.1183/09031936.00072212] [Medline: 23143550]
10. Buekers J, Arbillaga-Etxarri A, Gimeno-Santos E, et al. Heart rate and oxygen uptake kinetics obtained from continuous measurements with wearable devices during outdoor walks of patients with COPD. Digit Health 2023;9:20552076231162989. [doi: 10.1177/20552076231162989] [Medline: 36937691]
11. Fitbit important safety and product and services information. Fitbit device documentation. URL: https://www.fitbit.com/global/us/legal/safety-instructions [accessed 2023-11-08]
12. Nolan J, Flapan AD, Capewell S, MacDonald TM, Neilson JM, Ewing DJ. Decreased cardiac parasympathetic activity in chronic heart failure and its relation to left ventricular function. Br Heart J 1992 Jun;67(6):482-485. [doi: 10.1136/hrt.67.6.482] [Medline: 1622699]
13. De Pablo-Fernandez E, Tur C, Revesz T, Lees AJ, Holton JL, Warner TT. Association of autonomic dysfunction with disease progression and survival in Parkinson disease. JAMA Neurol 2017 Aug 1;74(8):970-976. [doi: 10.1001/jamaneurol.2017.1125] [Medline: 28655059]
14. Lanza GA, Guido V, Galeazzi MM, et al. Prognostic role of heart rate variability in patients with a recent acute myocardial infarction. Am J Cardiol 1998 Dec;82(11):1323-1328. [doi: 10.1016/S0002-9149(98)00635-3]
15. Raju S, Woo H, Fawzy A, et al. Decreased cardiac autonomic function is associated with higher exacerbation risk and symptom burden in chronic obstructive pulmonary disease. Chronic Obstr Pulm Dis 2023 Jul 26;10(3):328-334. [doi: 10.15326/jcopdf.2023.0410] [Medline: 37267605]
16. Hurst JR, Donaldson GC, Quint JK, Goldring JJ, Patel AR, Wedzicha JA. Domiciliary pulse-oximetry at exacerbation of chronic obstructive pulmonary disease: prospective pilot study. BMC Pulm Med 2010 Dec;10(1):52. [doi: 10.1186/1471-2466-10-52]
17. Paul SS, Tiedemann A, Hassett LM, et al. Validity of the Fitbit activity tracker for measuring steps in community-dwelling older adults. BMJ Open Sport Exerc Med 2015;1(1):e000013. [doi: 10.1136/bmjsem-2015-000013] [Medline: 27900119]
18. Fuller D, Colwell E, Low J, et al. Reliability and validity of commercially available wearable devices for measuring steps, energy expenditure, and heart rate: systematic review. JMIR Mhealth Uhealth 2020 Sep 8;8(9):e18694. [doi: 10.2196/18694] [Medline: 32897239]
19. Blondeel A, Demeyer H, Janssens W, Troosters T. Accuracy of consumer-based activity trackers as measuring tool and coaching device in patients with COPD and healthy controls. PLoS One 2020;15(8):e0236676. [doi: 10.1371/journal.pone.0236676] [Medline: 32750073]
20. Nissen M, Slim S, Jäger K, et al. Heart rate measurement accuracy of Fitbit Charge 4 and Samsung Galaxy Watch Active2: device evaluation study. JMIR Form Res 2022 Mar 1;6(3):e33635. [doi: 10.2196/33635] [Medline: 35230250]
21. Bai Y, Hibbing P, Mantis C, Welk GJ. Comparative evaluation of heart rate-based monitors: Apple Watch vs Fitbit Charge HR. J Sports Sci 2018 Aug;36(15):1734-1741. [doi: 10.1080/02640414.2017.1412235] [Medline: 29210326]
22. Benedetto S, Caldato C, Bazzan E, Greenwood DC, Pensabene V, Actis P. Assessment of the Fitbit Charge 2 for monitoring heart rate. PLoS One 2018;13(2):e0192691. [doi: 10.1371/journal.pone.0192691] [Medline: 29489850]
23. Graham BL, Steenbruggen I, Miller MR, et al. Standardization of spirometry 2019 update. An official American Thoracic Society and European Respiratory Society technical statement. Am J Respir Crit Care Med 2019 Oct 15;200(8):e70-e88. [doi: 10.1164/rccm.201908-1590ST] [Medline: 31613151]
24. Holland AE, Spruit MA, Troosters T, et al. An official European Respiratory Society/American Thoracic Society technical standard: field walking tests in chronic respiratory disease. Eur Respir J 2014 Dec;44(6):1428-1446. [doi: 10.1183/09031936.00150314] [Medline: 25359355]
25. Sletten DM, Suarez GA, Low PA, Mandrekar J, Singer W. COMPASS 31: a refined and abbreviated Composite Autonomic Symptom Score. Mayo Clin Proc 2012 Dec;87(12):1196-1201. [doi: 10.1016/j.mayocp.2012.10.013] [Medline: 23218087]
26. Mahler DA, Wells CK. Evaluation of clinical methods for rating dyspnea. Chest 1988 Mar;93(3):580-586. [doi: 10.1378/chest.93.3.580] [Medline: 3342669]
27. Jones PW, Harding G, Berry P, Wiklund I, Chen WH, Kline Leidy N. Development and first validation of the COPD assessment test. Eur Respir J 2009 Sep;34(3):648-654. [doi: 10.1183/09031936.00102509] [Medline: 19720809]
28. Van Remoortel H, Raste Y, Louvaris Z, et al. Validity of six activity monitors in chronic obstructive pulmonary disease: a comparison with indirect calorimetry. PLoS One 2012;7(6):e39198. [doi: 10.1371/journal.pone.0039198] [Medline: 22745715]
29. Rabinovich RA, Louvaris Z, Raste Y, et al. Validity of physical activity monitors during daily life in patients with COPD. Eur Respir J 2013 Nov;42(5):1205-1215. [doi: 10.1183/09031936.00134312] [Medline: 23397303]
30. Demeyer H, Mohan D, Burtin C, et al. Objectively measured physical activity in patients with COPD: recommendations from an international task force on physical activity. Chronic Obstr Pulm Dis 2021 Oct 28;8(4):528-550. [doi: 10.15326/jcopdf.2021.0213] [Medline: 34433239]
31. Gilgen-Ammann R, Schweizer T, Wyss T. RR interval signal quality of a heart rate monitor and an ECG Holter at rest and during exercise. Eur J Appl Physiol 2019 Jul;119(7):1525-1532. [doi: 10.1007/s00421-019-04142-5] [Medline: 31004219]
32. Tarvainen MP, Niskanen JP, Lipponen JA, Ranta-Aho PO, Karjalainen PA. Kubios HRV--heart rate variability analysis software. Comput Methods Programs Biomed 2014;113(1):210-220. [doi: 10.1016/j.cmpb.2013.07.024] [Medline: 24054542]

33. Buekers J, Theunis J, De Boever P, et al. Wearable finger pulse oximetry for continuous oxygen saturation measurements during daily home routines of patients with chronic obstructive pulmonary disease (COPD) over one week: observational study. JMIR Mhealth Uhealth 2019 Jun 6;7(6):e12866. [doi: 10.2196/12866] [Medline: 31199331]
34. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016 Jun;15(2):155-163. [doi: 10.1016/j.jcm.2016.02.012] [Medline: 27330520]
35. Mukaka MM. Statistics corner: a guide to appropriate use of correlation coefficient in medical research. Malawi Med J 2012 Sep;24(3):69-71. [Medline: 23638278]
36. Tedesco S, Sica M, Ancillao A, Timmons S, Barton J, O'Flynn B. Validity evaluation of the Fitbit Charge2 and the Garmin vivosmart HR+ in free-living environments in an older adult cohort. JMIR Mhealth Uhealth 2019 Jun 19;7(6):e13084. [doi: 10.2196/13084] [Medline: 31219048]
37. Farina N, Lowry RG. The validity of consumer-level activity monitors in healthy older adults in free-living conditions. J Aging Phys Act 2018 Jan 1;26(1):128-135. [doi: 10.1123/japa.2016-0344] [Medline: 28595019]
38. De Groef A, Asnong A, Blondeel A, et al. Accuracy of consumer-based activity trackers as measuring tool and coaching device in breast and colorectal cancer survivors. Support Care Cancer 2023 Sep 28;31(10):596. [doi: 10.1007/s00520-023-08061-2] [Medline: 37768403]
39. Ginis P, Goris M, De Groef A, et al. Validation of commercial activity trackers in everyday life of people with Parkinson's disease. Sensors (Basel) 2023 Apr 21;23(8):4156. [doi: 10.3390/s23084156] [Medline: 37112496]
40. de Zambotti M, Baker FC, Willoughby AR, et al. Measures of sleep and cardiac functioning during sleep using a multi-sensory commercially-available wristband in adolescents. Physiol Behav 2016 May 1;158:143-149. [doi: 10.1016/j.physbeh.2016.03.006] [Medline: 26969518]
41. Haghayegh S, Khoshnevis S, Smolensky MH, Diller KR. Accuracy of PurePulse photoplethysmography technology of Fitbit Charge 2 for assessment of heart rate during sleep. Chronobiol Int 2019 Jul;36(7):927-933. [doi: 10.1080/07420528.2019.1596947] [Medline: 30990098]
42. Miller DJ, Sargent C, Roach GD. A validation of six wearable devices for estimating sleep, heart rate and heart rate variability in healthy adults. Sensors (Basel) 2022 Aug 22;22(16):6317. [doi: 10.3390/s22166317] [Medline: 36016077]
43. Breteler MJMM, Huizinga E, van Loon K, et al. Reliability of wireless monitoring using a wearable patch sensor in high-risk surgical patients at a step-down unit in the Netherlands: a clinical validation study. BMJ Open 2018 Feb 27;8(2):e020162. [doi: 10.1136/bmjopen-2017-020162] [Medline: 29487076]
44. Takahashi Y, Okura K, Minakata S, et al. Accuracy of heart rate and respiratory rate measurements using two types of wearable devices. Prog Rehabil Med 2022;7:20220016. [doi: 10.2490/prm.20220016] [Medline: 35434406]
45. Pipek LZ, Nascimento RFV, Acencio MMP, Teixeira LR. Comparison of SpO2 and heart rate values on Apple Watch and conventional commercial oximeters devices in patients with lung disease. Sci Rep 2021 Sep 23;11(1):18901. [doi: 10.1038/s41598-021-98453-3] [Medline: 34556765]
46. Arslan B, Sener K, Guven R, et al. Accuracy of the Apple Watch in measuring oxygen saturation: comparison with pulse oximetry and ABG. Ir J Med Sci 2024 Feb;193(1):477-483. [doi: 10.1007/s11845-023-03456-w] [Medline: 37440093]
47. Hearn EL, Byford J, Wolfe C, et al. Measuring arterial oxygen saturation using wearable devices under varying conditions. Aerosp Med Hum Perform 2023 Jan 1;94(1):42-47. [doi: 10.3357/AMHP.6078.2023] [Medline: 36757230]
48. Hermand E, Coll C, Richalet JP, Lhuissier FJ. Accuracy and reliability of pulse O2 saturation measured by a wrist-worn oximeter. Int J Sports Med 2021 Dec;42(14):1268-1273. [doi: 10.1055/a-1337-2790] [Medline: 34000752]
49. Schiefer LM, Treff G, Treff F, et al. Validity of peripheral oxygen saturation measurements with the Garmin Fēnix® 5X Plus wearable device at 4559 m. Sensors (Basel) 2021 Sep 23;21(19):6363. [doi: 10.3390/s21196363] [Medline: 34640680]
50. Carvalho TD, Pastre CM, de Godoy MF, et al. Fractal correlation property of heart rate variability in chronic obstructive pulmonary disease. Int J Chron Obstruct Pulmon Dis 2011 Jan 6;6:23-28. [doi: 10.2147/COPD.S15099] [Medline: 21311690]
51. Stanbury RM, Graham EM. Systemic corticosteroid therapy--side effects and their management. Br J Ophthalmol 1998 Jun;82(6):704-708. [doi: 10.1136/bjo.82.6.704] [Medline: 9797677]
52. Maclay JD, McAllister DA, Rabinovich R, et al. Systemic elastin degradation in chronic obstructive pulmonary disease. Thorax 2012 Jul;67(7):606-612. [doi: 10.1136/thoraxjnl-2011-200949] [Medline: 22374923]
53. Fine J, Branan KL, Rodriguez AJ, et al. Sources of inaccuracy in photoplethysmography for continuous cardiovascular monitoring. Biosensors (Basel) 2021 Apr 16;11(4):126. [doi: 10.3390/bios11040126] [Medline: 33923469]
54. Rouette J, McDonald EG, Schuster T, Brophy JM, Azoulay L. Treatment and prescribing trends of antihypertensive drugs in 2.7 million UK primary care patients over 31 years: a population-based cohort study. BMJ Open 2022 Jun 10;12(6):e057510. [doi: 10.1136/bmjopen-2021-057510] [Medline: 35688595]
55. Ballin M, Nordström P, Niklasson J, et al. Daily step count and incident diabetes in community-dwelling 70-year-olds: a prospective cohort study. BMC Public Health 2020 Nov 30;20(1):1830. [doi: 10.1186/s12889-020-09929-2] [Medline: 33256704]
56. Ogliari G, Mahinrad S, Stott DJ, et al. Resting heart rate, heart rate variability and functional decline in old age. CMAJ 2015 Oct 20;187(15):E442-E449. [doi: 10.1503/cmaj.150462] [Medline: 26323697]

Abbreviations

ATS: American Thoracic Society
CAT: COPD Assessment Test
COMPASS: Composite Autonomic Symptom Score
COPD: chronic obstructive pulmonary disease
DAM: Dynaport MoveMonitor
ERS: European Respiratory Society
HRV: heart rate variability
ICC: intraclass correlation coefficient
mMRC: modified Medical Research Council
RHR: resting heart rate
RMSSD: root mean square of successive differences
RR: respiratory rate
SpO_2: oxygen saturation

Edited by L Buis; submitted 03.01.24; peer-reviewed by C Shah, J Edwards, M Fraser; revised version received 28.06.24; accepted 29.08.24; published 06.11.24.

Please cite as:
Hermans F, Arents E, Blondeel A, Janssens W, Cardinaels N, Calders P, Troosters T, Derom E, Demeyer H
Validity of a Consumer-Based Wearable to Measure Clinical Parameters in Patients With Chronic Obstructive Pulmonary Disease and Healthy Controls: Observational Study
JMIR Mhealth Uhealth 2024;12:e56027
URL: https://mhealth.jmir.org/2024/1/e56027
doi:10.2196/56027

Original Paper

Effectiveness of Telemonitoring in Reducing Hospitalization and Associated Costs for Patients With Heart Failure in Finland: Nonrandomized Pre-Post Telemonitoring Study

Jorma Kokkonen[1], MD, PhD; Pirjo Mustonen[2], MD, PhD; Eija Heikkilä[3], PhD; Riikka-Leena Leskelä[3], PhD; Paula Pennanen[3], MSc; Kati Krühn[4], MSc; Arto Jalkanen[5], MSc; Jussi-Pekka Laakso[6], MSc; Jari Kempers[7], PhD; Sami Väisänen[8], PhD; Paulus Torkki[9], PhD

[1]Tampere Heart Hospital, Tampere, Finland
[2]The Wellbeing Services County of Southwest Finland, Turku, Finland
[3]Nordic Healthcare Group, Helsinki, Finland
[4]Roche Diagnostics (Schweiz) AG, Zug, Switzerland
[5]The Wellbeing Services County of South Savo, Mikkeli, Finland
[6]Ceneron Advisor Oy, Helsinki, Finland
[7]European Health Economics Oy, Jyväskylä, Finland
[8]Roche Diagnostics Oy, Espoo, Finland
[9]Department of Public Health, Faculty of Medicine, University of Helsinki, Helsinki, Finland

Corresponding Author:
Jorma Kokkonen, MD, PhD
Tampere Heart Hospital
Elämänaukio 1
Tampere, 33520
Finland
Phone: 358 50 5981078
Email: sydanjopi@gmail.com

Abstract

Background: Many patients with chronic heart failure (HF) experience a reduced health status, leading to readmission after hospitalization despite receiving conventional care. Telemonitoring approaches aim to improve the early detection of HF decompensations and prevent readmissions. However, knowledge about the impact of telemonitoring on preventing readmissions and related costs remains scarce.

Objective: This study assessed the effectiveness of adding a telemonitoring solution to the standard of care (SOC) for the prevention of hospitalization and related costs in patients with HF in Finland.

Methods: We performed a nonrandomized pre-post telemonitoring study to estimate health care costs and resource use during 6 months on SOC followed by 6 months on SOC with a novel telemonitoring solution. The telemonitoring solution consisted of a digital platform for patient-reported symptoms and daily weight and blood pressure measurements, automatically generated alerts triggering phone calls with secondary care nurses, and rapid response to alerts by treating physicians. Telemonitoring solution data were linked to patient register data on primary care, secondary care, and hospitalization. The patient register of the Southern Savonia Social and Health Care Authority (Essote) was used. Eligible patients had at least 1 hospital admission within the last 12 months and self-reported New York Heart Association class II-IV from the central hospital in the Southern Savonia region.

Results: Out of 50 recruited patients with HF, 43 completed the study and were included in the analysis. The hospitalization-related cost decreased (49%; *P*=.03) from €2189 (95% CI €1384-€2994; a currency exchange rate of EUR €1=US $1.10589 is applicable) during SOC to €1114 (95% CI €425-€1803) during telemonitoring. The number of patients with at least 1 hospitalization due to HF was reduced by 70% (*P*=.002) from 20 (47%) out of 43patients during SOC to 6 (14%) out of 43 patients in telemonitoring. The estimated mean total health care cost per patient was €3124 (95% CI €2212-€4036) during SOC and €2104 (95% CI €1313-€2895) during telemonitoring, resulting in a 33% reduction (*P*=.07) in costs with telemonitoring.

Conclusions: The results suggest that the telemonitoring solution can reduce hospital-related costs for patients with HF with a recent hospital admission.

(JMIR Mhealth Uhealth 2024;12:e51841) doi:10.2196/51841

KEYWORDS

cost; Finland; heart failure; hospital; resource use; telemonitoring

Introduction

The prevalence of heart failure (HF) and related costs is increasing worldwide due to an aging population [1]. The estimated prevalence of HF in the adult population is 1% to 2%, increasing to 10% in older adults aged 70 years or older [2,3]. HF often leads to gradual or acute changes in HF symptoms (decompensation) that require repeated and prolonged hospitalization [4]. Hospital admission is a strong predictor of further hospital admission: 20% to 25% of patients with HF are rehospitalized within 1 month and approximately 50% within 5 months of discharge [5]. Decompensation requiring hospitalization is also linked to increased mortality. A European registry study following patients for 1 year after hospitalization reported mortality rates of 24% for acute HF and 6.4% for chronic HF [6]. Hospitalization accounted for around 80% of HF health care costs [1].

An early return to the hospital following discharge may be a result of incomplete inpatient treatment and poor coordination and planning of follow-up care. Even for patients with regular follow-up care, however, the signs of decompensation may not occur during cardiology visits. Patients often contact clinics when symptoms are at an advanced stage [7]. Self-monitoring of symptoms, such as increased blood pressure, weight gain, or other health status-related symptoms, is particularly important in HF management [4]. Self-monitoring requires patients to be motivated to measure symptoms associated with HF and to have access to clinical advice when symptoms appear [8].

Remote monitoring aims to improve monitoring of patients' health status and is defined as a part of telehealth [9]. A basic level of remote monitoring involves regular and structured telephone support provided by health care professionals (HCPs) to discuss symptoms, self-monitoring measurements, lifestyle, and drug therapy. Structured telephone support can reduce HF-related hospitalization but does not seem to have an impact on the all-cause hospitalization of patients with HF [10]. Remote monitoring solutions are noninvasive stand-alone systems in which patient data on biometric measurements (such as body weight, blood pressure, and heart rate) and reported symptoms are frequently transmitted to HCPs through a secure digital system. HCPs manually review the data on digital platforms, which may also include integrated automated alerts, and necessary action is taken to optimize treatment.

The effect of noninvasive telemonitoring has been compared to the standard of care (SOC) in several studies, primarily through randomized trials. Some studies found telemonitoring had a beneficial impact on reducing hospitalization [11], while others did not find any effect [12,13]. However, a recent meta-analysis, encompassing 91 randomized trials and observational studies, revealed that noninvasive telemonitoring reduced all-cause mortality by 16%, first hospitalization by 19%, and total HF hospitalizations by 15%. When comparing telemonitoring studies and developing optimal telemonitoring approaches, it is crucial to consider various determinants, including the telemonitoring intervention models, health care systems, and the characteristics of the population with HF in the studies [14].

There are only a few international studies that have explored the cost-effectiveness of telemonitoring compared to the SOC [15-18].

The objective of this nonrandomized pre-post intervention study in patients with HF with a recent (<12 month) hospitalization was to assess the effectiveness of adding a telemonitoring solution to SOC on hospitalizations and related costs in the Finnish health care system. The study compared hospitalization occurrence and related costs with SOC and following the introduction of a telemonitoring solution. Secondary outcomes included hospital admissions and total health care costs.

Methods

Study Design

The nonrandomized pre-post intervention study was performed in Southern Savonia, Finland. During the 12-month study period, patients were treated with SOC for the first 6 months and then with a telemonitoring solution in addition to SOC for the next 6 months. The primary outcome was hospitalization-related costs during 6 months with SOC versus telemonitoring. Secondary outcomes included the number of patients with at least 1 hospital admission due to HF or a cardiovascular cause other than HF emergency care visits and primary care or cardiology (secondary care) calls and visits. Health care costs for secondary outcomes included the total health care costs of primary care, secondary care (for cardiology), emergency visits, and phone calls. The study was designed to demonstrate the effectiveness of remote monitoring within the Finnish health care system. The costs of the telemonitoring service itself were not analyzed.

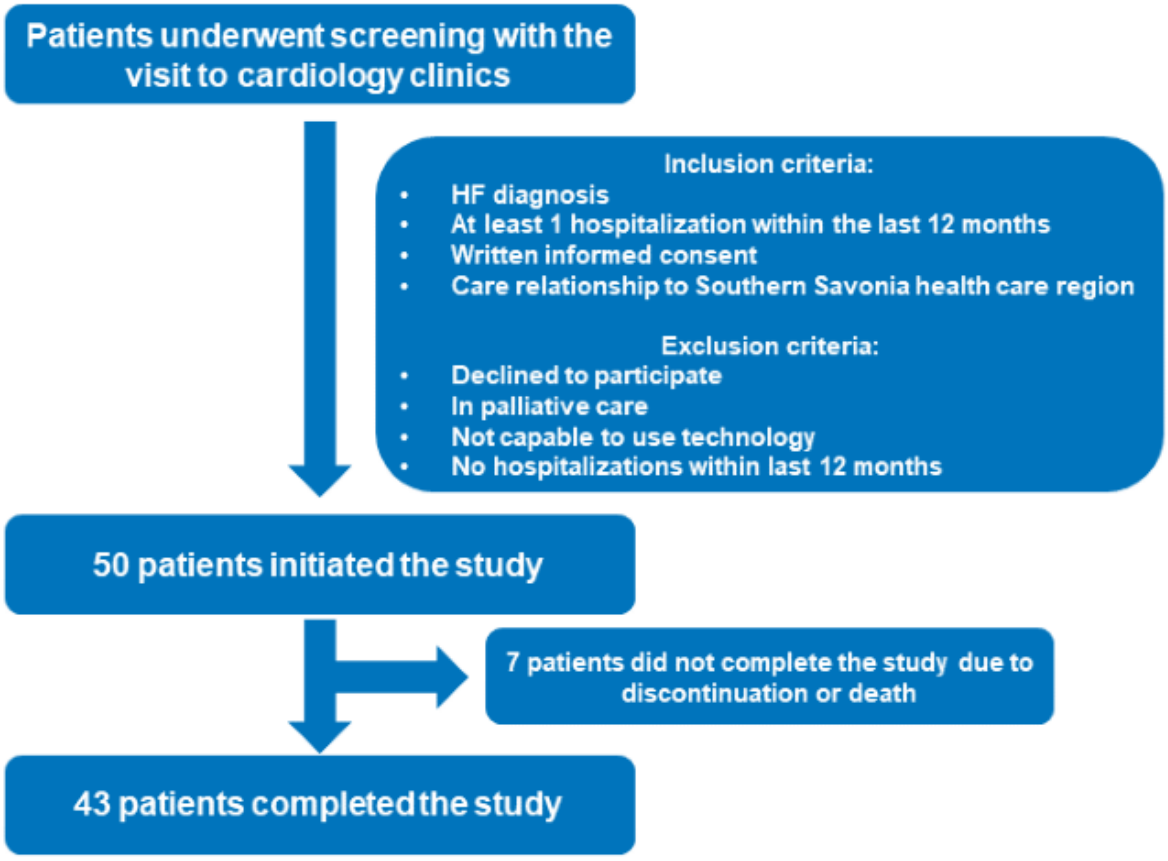
Health care resource use was collected for each patient during SOC and telemonitoring from the patient register of the Southern Savonia Social and Health Care Authority (Essote). The data was pseudonymized by the register holder. The Health Care Authority is responsible for all social and health care services for the population of approximately 100,000 inhabitants in Southern Savonia, Finland.

Study Patients

Patients were recruited from Mikkeli Central Hospital in Finland's Southern Savonia region. Patients with an HF

diagnosis confirmed by a cardiologist, at least 1 hospital admission in the 12 months preceding study initiation, and self-reported New York Heart Association (NYHA) class II-IV were eligible for the study (Figure 1).

Figure 1. Flowchart for patient selection for the study.



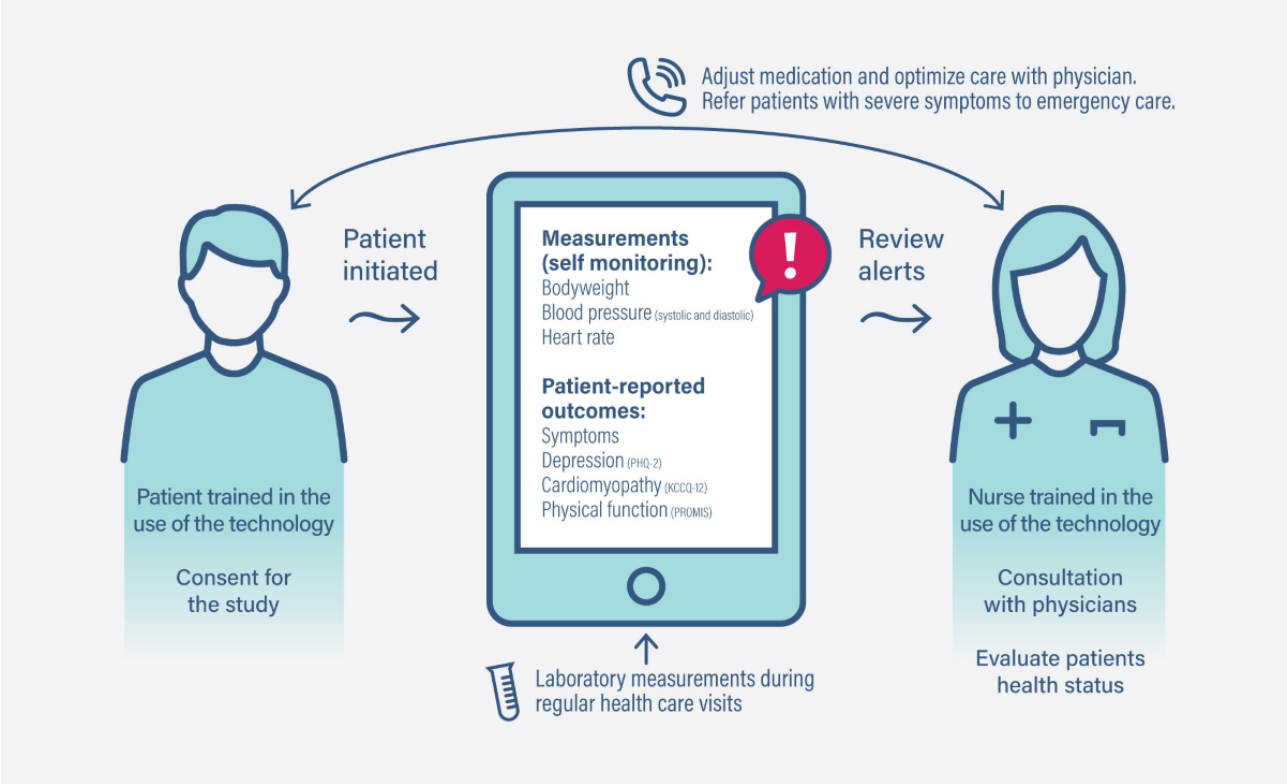
The inclusion criteria also stated that patients must be able to manage the telemonitoring devices and digital platform used in the study. Palliative care was an exclusion criterion.

Study Procedures

The SOC, during the first 6 months, included regular cardiology appointments and laboratory tests planned by a cardiologist for each patient with HF according to local care guidelines for HF treatment. Nurses followed up with patients through phone calls, depending on the state of HF. After inpatient stays, the cardiologist or internist at the hospital made an individual plan for the follow-up of patients posthospitalization. During the follow-up period, patients measured their weight and blood pressure at home, and nurses followed up with patients through phone calls to discuss their health status and measurement results.

Telemonitoring was added on top of SOC during the next 6 months and consisted of a digital platform, home measurement devices, and nurses monitoring patients through the digital platform. The digital platform used Veta Health's remote patient monitoring platform (Veta Health Inc), customized for the study. Patients used their smartphones, handheld devices, or personal computers to access the digital platform. Patients measured their weight daily with a digital scale (Omron Corporation) and their blood pressure with a digital blood pressure measuring device (Omron M7000 Intelli IT) and transferred the measurements into the digital platform (Figure 2). The digital platform also included symptom-related questions.

Figure 2. A schematic presentation of the remote patient management model. KCCQ-12: Kansas City Cardiomyopathy Questionnaire-12; PHQ2: Patient Health Questionnaire-2; PROMIS: Patient Reported Outcomes Measurement Information System.



The digital platform automatically compared patients' body weight and HF symptom answers against preset thresholds and generated semiurgent or urgent alerts predicting HF worsening (Table S1 in Multimedia Appendix 1). Depending on the alert type, the digital platform either advised the patient to contact a nurse or a nurse to contact the patient to validate the health status. Nurses had access to alerts on working days. If needed, nurses referred a patient to a cardiologist to optimize HF care or medication. The treating cardiologist reacted to the nurses' referrals within 24 hours. For urgent alerts, the digital platform advised patients to go to emergency care. Nurses also provided technical support for patients as required. The digital platform collected blood pressure data and laboratory results from regular health care visits, not for the alert algorithm but to allow nurses to evaluate the patient's health status.

Health Care Resource Use and Costs

Patients' health care costs and resource use were estimated from the Essote patient register and consisted of public primary care, secondary care (cardiology unit), emergency visits, hospitalizations (cardiology and internal medicine; primary care), and phone calls for primary and secondary care (cardiology). A unique personal identification number for each resident in Finland connected the digital platform data and patient register data. The *International Classification of Diseases, Tenth Revision* (ICD-10) diagnosis code registered for each health care event was used to separate hospitalizations for HF (ICD-10 code I50) from hospitalizations for a cardiovascular cause other than HF (ICD-10 codes I10, I25, I42, I46, I48, I49, and I70). All-cause hospitalizations included hospitalizations with any diagnosis. The costs of health care use were calculated using Essote diagnosis-related group prices.

Statistical Analysis

The study analysis included only patients who completed the study. Patients who died during the study or discontinued the study were excluded.

Patient demographics and NYHA class were summarized as n (%) of patients per category or median (IQR). During the SOC and telemonitoring periods, n (%) of patients with at least 1 hospitalization and the mean number of inpatient days per patient (95% CI) were reported. The mean number (95% CI) of visits per patient (primary, secondary, and emergency) and the mean number of calls (primary and secondary care) per patient were also reported for each period. Mean health care costs per patient were reported for each period. The normal distribution of each variable was assessed through visual inspection and the Shapiro-Wilk test. For data found to be nonnormally distributed, differences between SOC and telemonitoring periods were tested using the Wilcoxon signed rank test, and a value of $P<.05$ was considered statistically significant. The Pearson chi-square test with Yates correction was used for testing the difference between SOC and telemonitoring periods (a binary variable) in the number of patients with at least 1 hospitalization.

Ethical Considerations

The ethics committee of the Northern Savonia Hospital District approved the study protocol (1401/2020). The study followed good clinical practice following the Declaration of Helsinki and the laws and regulations applicable in Finland. Patients gave written consent upon recruitment to the study. Participation in the study was voluntary and no financial compensation was awarded for participation.

Results

Study Population and Patient Characteristics

Between December 15, 2020, and March 24, 2021, a total of 50 patients with HF were recruited from the Mikkeli Central Hospital. A total of 7 patients did not complete the study due to their deaths or withdrawals from it. All 43 (86%) patients who completed the 12-month study period were included in the analysis. During the telemonitoring period, 20% (9/43) of the daily weight and blood pressure measurements were missing.

The median age of patients was 73 (IQR 66-80) years, 74% (37/50) were male, and 60% (30/50) of patients had NYHA classes III-IV (Table 1).

Table 1. Patient demographics and disease characteristics (n=50).

Characteristic	Value
Age (years), median (IQR)	73 (66-80)
Sex, n (%)	
Male	37 (74)
Female	13 (26)
NYHA[a] score, n (%)	
NYHA class II	20 (40)
NYHA class III-IV	30 (60)
Systolic blood pressure (mm Hg), mean (SD)	118 (18)
Heart rate (beats/min), mean (SD)	72 (11)
BMI (kg/m^2), mean (SD)	27 (6)
Serum creatinine (µmol/L), mean (SD)	125 (44)
Pro–B-type natriuretic peptide (ng/L), median (IQR)	3122 (1590-5598)
Left ventricular ejection fraction (%), mean (SD)	37 (11)
Etiology of heart failure, n (%)	
Ischemic cardiomyopathy	15 (30)
Idiopathic dilated cardiomyopathy	13 (26)
Hypertensive cardiomyopathy	12 (24)
Tachycardia cardiomyopathy	<5
Cytostatic cardiomyopathy	<5
Valvular cardiomyopathy	<5
Genetic cardiomyopathy	<5
Medical history, n (%)	
Hypertension	24 (48)
Diabetes	14 (28)
Coronary heart disease	17 (34)
Myocardial infarction	6 (12)
Atrial fibrillation	31 (62)
Valvular heart disease	<5
Medication at recruitment, n (%)	
Diuretic	46 (92)
Digitalis	6 (12)
β-blocker	48 (96)
Mineralocorticoid antagonist	35 (70)
ACE[b]-inhibitor	7 (14)
Angiotensin receptor-blocker	10 (20)
Valsartan-sacubitril	30 (60)
SGLT2[c]-inhibitor	<5
Statin	26 (52)
Anticoagulant	42 (84)
ASA[d] or clopidogrel	11 (22)

[a]NYHA: New York Heart Association.

[b]ACE: angiotensin-converting enzyme.

[c]SGLT2: sodium-glucose cotransporter-2 .

[d]ASA: acetylsalicylic acid.

Health Care Resource Use

Significantly fewer patients (6 patients in telemonitoring vs 20 patients in SOC; *P*=.002) had an HF hospitalization during the telemonitoring versus SOC period. The number of inpatient days per patient due to HF decreased by 48% during the telemonitoring period (mean 1.2, 95% CI 0.1-2.3 days vs 2.3, 95% CI 1-3.6 days with SOC; *P*=.17). The number of emergency care visits decreased significantly during the telemonitoring period by 44% (mean 0.7, 95% CI 0.4-1 vs mean 1.3, 95% CI 0.9-1.7 with SOC; *P*=.006). Patients with HF made significantly more phone calls to secondary care during the telemonitoring period (mean 8.3, 95% CI 6.6-10 vs mean 2, 95% CI 1.3-2.7 with SOC; 318% increase; *P*.001) and had significantly more primary care visits (mean 4, 95% CI 2.2-5.8 vs mean 2.8, 95% CI 1.7-3.9; 44% increase; *P*=.02; Table 2).

Table 2. Use of health care per patient in standard of care (SOC) or telemonitoring solution for a 6-month period. Statistics were calculated with the Wilcoxon signed rank test or the Pearson chi-square test with Yates correction for the binary variables.

Variable	SOC (n=43)	Telemonitoring (n=43)	Absolute change (relative change; %)	*P* value for difference
All-cause hospitalizations				
Patients with ≥1 event, n (%)	23 (53)	14 (33)	–9 (–39)	.08
Inpatient days, mean (95% CI)	2.9 (1.6)	1.7 (1.2)	–1.2 (–41)	.20
Hospitalizations for cardiovascular cause other than HF[a]				
Patients with ≥1 event, n (%)	6 (14)	<5 (~10)	<5 (~–30)	.70
Inpatient days, mean (95% CI)	0.2 (0.3)	0.02 (0.05)	–0.16 (–88)	.40
Hospitalizations for HF				
Patients with ≥1 event, n (%)	20 (47)	6 (14)	–14 (–70)	.002
Mean inpatient days, days (95% CI)	2.3 (1.3)	1.2 (1.1)	–1.1 (–48)	.17
Mean number of emergency care visit for cardiovascular cause, n (95% CI)	1.3 (0.4)	0.7 (0.3)	–0.6 (–44)	.006
Mean number of phone calls to primary care, n (95% CI)	3.9 (1.3)	2.5 (0.9)	–1.4 (–36)	.01
Mean number of phone calls to secondary care, n (95% CI)	2 (0.7)	8.3 (1.7)	6.3 (+318)	<.001
Mean number of all-cause primary care visits, n (95% CI)	2.8 (1.1)	4 (1.8)	1.2 (+44)	.02
Mean number of secondary care visits (cardiology), n (95% CI)	1.8 (0.5)	1.9 (0.6)	0.1 (+8)	.80

[a]HF: heart failure.

Health Care Costs

Mean hospitalization costs per patient decreased significantly by 49% during the telemonitoring period (mean €1114 vs €2189 with SOC; *P*=.03; a currency exchange rate of EUR €1=US $1.10589 is applicable), while total health care costs decreased by 33% (mean €2104 vs €3124 with SOC; *P*=.07; Table 3).

The cost of emergency care visits was also significantly lower in the telemonitoring period (mean €209 vs €347 with SOC; 40% decrease; *P*=.009), and mean costs per patient for phone calls to secondary care increased significantly (mean €268 vs €114 with SOC; 134% increase; *P*.001) in the telemonitoring period (Table 3).

Table 3. Estimated mean direct health care cost per patient in standard of care (SOC) and in telemonitoring solution, respectively, during a 6-month period (2021). A currency exchange rate of EUR €1=US $1.10589 is applicable. Statistics were calculated with the Wilcoxon signed rank test.

Cost category	SOC (n=43), mean cost (€; 95% CI)	Telemonitoring (n=43), mean cost (€; 95% CI)	Absolute change (€; relative change in mean cost; %)	*P* value for difference
Hospitalizations[a]	2189 (805)	1114 (689)	–1075 (–49)	.03
Primary care visits	75 (37)	101 (74)	27 (+36)	.30
Secondary care visits (cardiology)	288 (77)	337 (87)	50 (+17)	.20
Emergency care visits	347 (122)	209 (99)	–137 (–40)	.009
Phone calls to primary care	112 (43)	74 (35)	–38 (–34)	.02
Phone calls to secondary care	114 (42)	268 (68)	153 (+134)	<.001
Total cost	3124 (912)	2104 (791)	–1020 (–33)	.07

[a]Hospitalization from cardiology and internal medicine ward and from primary care.

Discussion

Principal Findings

In this pre-post study of a novel telemonitoring solution for patients with HF in Finland, hospitalization costs were significantly lower during the 6-month telemonitoring period versus the SOC period (from €2189 per patient during SOC to €1114 during telemonitoring). The number of hospitalized patients was significantly lower during telemonitoring (from 20 during SOC to 6 patients), and the mean length of stay decreased from 2.3 days to 1.2 days (not statistically significant). The number of emergency visits and associated costs were also significantly lower during telemonitoring. By contrast, patients with HF had significantly more primary care visits and phone calls to secondary care nurses during telemonitoring versus SOC; however, total health care costs were 33% lower than during SOC (not statistically significant).

The reduction in inpatient days due to HF during telemonitoring was not statistically significant, most probably due to the low number of patients with HF in the study. While secondary care phone calls increased significantly, a substantial part of these were for technical help in using the digital platform at the start of the telemonitoring period. The number of primary care phone calls, on the other hand, decreased during the telemonitoring period, which may be because secondary care nurses were monitoring patients.

Comparison With Previous Work

A comprehensive meta-analysis of recently published studies on telemonitoring has provided evidence that telemonitoring is beneficial in reducing mortality and hospitalizations in patients with HF. However, individual studies show both beneficial and neutral effects of telemonitoring when compared to SOC [11,13,19-21]. The main objective of the studies was to investigate whether remote monitoring can improve the detection of early signs of decompensation and decrease hospitalization and mortality. The variability of the results may be due to differences in the health care system, telemonitoring model, population with HF, and follow-up durations [9,14].

This study found that 47% (20/43) of patients were hospitalized due to HF during the SOC period, versus 14% (6/43) during the telemonitoring period in the Southern Savonia region of Finland. Vuorinen et al [21] conducted a telemonitoring study in Finland from 2010 to 2012. They showed that only 28% of patients were hospitalized in the SOC group and 17% in the remote monitoring group during a 6-month follow-up period. The inclusion criteria for the population with HF included NYHA II-IV but did not require a recent hospitalization, which could explain the lower HF hospitalization risk compared with this study. Our data align with previous studies showing that nearly half of patients with HF are rehospitalized within 6 months after discharge [21,22]. However, further studies using a similar population with HF are needed to confirm these findings.

Similar to this study, Vuorinen et al [21] also found a nonstatistically significant decrease in inpatient days with telemonitoring (mean 0.7 vs 1.4 days with SOC). The significant reduction in hospitalization-related costs and the number of patients hospitalized due to HF in this study support the idea that telemonitoring reduces hospitalizations.

A Spanish telemonitoring trial (n=117) had similar findings to this study. In this trial, 50% of the patients were hospitalized in the SOC group versus 28% in the telemonitoring group over a 6-month follow-up period. The patients were enrolled in the study upon hospitalization [19]. Thus, the results from this study support our findings on the benefits of telemonitoring in reducing hospitalizations. However, a large Better Effectiveness After Transition–Heart Failure (BEAT-HF) trial (n=1437) conducted in California could not see a reduction in readmissions in patients with HF in a telemonitoring group compared to SOC during a 6-month follow-up period [20]. The BEAT-HF trial's limitations were that patients were recruited from academic medical centers, which may restrict the generalizability of the results, as most patients with HF do not receive care in academic medical centers. Upon receiving alerts, nurses advised patients to contact the physicians, or nurses called the physicians, but physicians were not directly involved with the interventions. Thus, the monitoring may not have affected care in practice. In this study and the Spanish study, nurses and treating physicians collaborated upon receiving alerts, which may have increased the benefits of telemonitoring. For example, in this study, physicians reacted to patient alerts within 24 hours.

There are several studies on the effect of telemonitoring on hospitalization and mortality, but only a few studies on costs.

This study estimated the health care cost related to resource use using real-word data and showed that health care costs were 33% (€2104/€3124) lower during telemonitoring versus SOC. As expected, most of the cost reductions originated from reduced hospitalizations. A cost-effectiveness study in a Danish telemonitoring trial (n=274) reported similar results by showing that telemonitoring reduced total health care costs by 35% versus SOC with a 1-year follow-up [16]. In a Spanish telemonitoring trial, Comin-Colet et al [15] found a total cost reduction of 45% with telemonitoring versus SOC, with 178 patients and a 6-month follow-up. A Belgian Telemonitoring in the Management of Heart Failure (TEMA-HF) study (n=160) found a 27% cost reduction (not statistically significant) with telemonitoring versus SOC during a 6-month follow-up [23]. The German Heart Failure II trial (TIM-HF2) showed an 18% reduction in annual costs per patient in the telemonitoring group compared to the SOC group during a 1-year follow-up [18]. These studies support our conclusion that telemonitoring may result in substantial cost savings in HF care. To justify reimbursement for telemonitoring, studies are needed on the cost-effectiveness of large-scale telemonitoring for decision makers. Furthermore, a cost-effective telemonitoring model applicable to different health care systems and settings needs to be developed.

Strengths and Limitations

A strength of this study was the cost analysis, which included both HF-related health care costs as well as other costs accrued during the follow-up period. The study had some limitations. The study was conducted in a single region, the Southern Savonia region, which may limit the generalizability of the results. However, the study population is representative of the region, as all patients are directed to the same central hospital where recruitment was done. A randomized controlled trial design was not feasible due to the limited number of suitable patients with HF for remote monitoring. No patients with HF were included from other health care districts, as divergent monitoring practices could potentially bias the analysis results. Patients were not randomized, and patients needed to be able to use the digital platform, which may have resulted in a possible selection bias. As the pre-post design uses a historical control group (ie, patients on SOC in the period before starting telemonitoring), the underlying assumption in the analysis, given the deteriorating nature of HF, is that health care use in the absence of telemonitoring would remain at least at the same level as during SOC. Follow-up with telemonitoring was limited to 6 months, and it is unclear how use of health care services would develop beyond this period. Finally, due to the small patient numbers, the absence of a control group, and the 6-month follow-up period, it was not feasible to conduct mortality analyses.

The following must be considered when generalizing our results and applying our telemonitoring solution to other health care systems: our telemonitoring solution was applied to a patient population with a high risk of readmission due to a recent hospital admission and NYHA class II-IV. Other patient characteristics considered were the mean age (73 years), male proportion (37/50, 74%), proportion of patients in NYHA class II-IV (30/50, 60%), and proportion of patients with at least 1 admission within 6 months (20/50, 47%). These patient characteristics were comparable to those of other reported telemonitoring study populations [11,13,18,19].

Conclusions

In conclusion, our results suggest that the novel telemonitoring solution can help reduce hospital admissions and hospitalization costs as well as total health care costs in a population with HF with a recent hospital admission in the past 12 months.

Acknowledgments

The authors would like to thank Saara Liukkonen, Kiia Pätsi, and Jenni Tikkala for their valuable contributions.

Conflicts of Interest

PM, JK, and J-PL are consultants for Roche Diagnostics. R-LL, EH, and PP receive consulting fees paid to their employer by Roche Diagnostics. KK and SV are employees of Roche Diagnostics. AJ is an employee of Essote.

Multimedia Appendix 1
Definition of the alert triggers.
[DOCX File , 16 KB - mhealth_v12i1e51841_app1.docx]

References

1. Heidenreich PA, Albert NM, Allen LA, Bluemke DA, Butler J, Fonarow GC, et al. Forecasting the impact of heart failure in the United States: a policy statement from the American Heart Association. Circ Heart Fail 2013 May;6(3):606-619 [FREE Full text] [doi: 10.1161/HHF.0b013e318291329a] [Medline: 23616602]
2. Huusko J, Kurki S, Toppila I, Purmonen T, Lassenius M, Gullberg E, et al. Heart failure in Finland: clinical characteristics, mortality, and healthcare resource use. ESC Heart Fail 2019 Aug;6(4):603-612 [FREE Full text] [doi: 10.1002/ehf2.12443] [Medline: 31054212]
3. Roger VL. Epidemiology of heart failure. Circ Res 2013 Aug 30;113(6):646-659 [FREE Full text] [doi: 10.1161/CIRCRESAHA.113.300268] [Medline: 23989710]

4. McDonagh TA, Metra M, Adamo M, Gardner RS, Baumbach A, Böhm M, et al. 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure. Eur Heart J 2021 Sep 21;42(36):3599-3726 [FREE Full text] [doi: 10.1093/eurheartj/ehab368] [Medline: 34447992]
5. Desai AS, Stevenson LW. Rehospitalization for heart failure: predict or prevent? Circulation 2012 Jul 24;126(4):501-506 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.112.125435] [Medline: 22825412]
6. Crespo-Leiro MG, Anker SD, Maggioni AP, Coats AJ, Filippatos G, Ruschitzka F, et al. European Society of Cardiology Heart Failure Long-Term Registry (ESC-HF-LT): 1-year follow-up outcomes and differences across regions. Eur J Heart Fail 2016 Jun;18(6):613-625 [FREE Full text] [doi: 10.1002/ejhf.566] [Medline: 27324686]
7. Schiff GD, Fung S, Speroff T, McNutt RA. Decompensated heart failure: symptoms, patterns of onset, and contributing factors. Am J Med 2003 Jun 01;114(8):625-630. [doi: 10.1016/s0002-9343(03)00132-3] [Medline: 12798449]
8. Clark AM, Freydberg CN, McAlister FA, Tsuyuki RT, Armstrong PW, Strain LA. Patient and informal caregivers' knowledge of heart failure: necessary but insufficient for effective self-care. Eur J Heart Fail 2009 Jun;11(6):617-621 [FREE Full text] [doi: 10.1093/eurjhf/hfp058] [Medline: 19414477]
9. Singhal A, Cowie MR. Digital health: implications for heart failure management. Card Fail Rev 2021 Mar;7:e08 [FREE Full text] [doi: 10.15420/cfr.2020.28] [Medline: 34035953]
10. Inglis SC, Clark RA, Dierckx R, Prieto-Merino D, Cleland JGF. Structured telephone support or non-invasive telemonitoring for patients with heart failure. Cochrane Database Syst Rev 2015 Oct 31;2015(10):CD007228 [FREE Full text] [doi: 10.1002/14651858.CD007228.pub3] [Medline: 26517969]
11. Koehler F, Koehler K, Deckwart O, Prescher S, Wegscheider K, Kirwan BA, et al. Efficacy of telemedical interventional management in patients with heart failure (TIM-HF2): a randomised, controlled, parallel-group, unmasked trial. Lancet 2018 Sep 22;392(10152):1047-1057. [doi: 10.1016/S0140-6736(18)31880-4] [Medline: 30153985]
12. Völler H, Bindl D, Nagels K, Hofmann R, Vettorazzi E, Wegscheider K, et al. The first year of noninvasive remote telemonitoring in chronic heart failure is not cost saving but improves quality of life: The randomized controlled CardioBBEAT trial. Telemed J E Health 2022 Mar 21;28(11):1613-1622 [FREE Full text] [doi: 10.1089/tmj.2022.0021] [Medline: 35325562]
13. Galinier M, Roubille F, Berdague P, Brierre G, Cantie P, Dary P, et al. Telemonitoring versus standard care in heart failure: a randomised multicentre trial. Eur J Heart Fail 2020 Jun;22(6):985-994 [FREE Full text] [doi: 10.1002/ejhf.1906] [Medline: 32438483]
14. Scholte NTB, Gürgöze MT, Aydin D, Theuns DAMJ, Manintveld OC, Ronner E, et al. Telemonitoring for heart failure: a meta-analysis. Eur Heart J 2023 Aug 14;44(31):2911-2926 [FREE Full text] [doi: 10.1093/eurheartj/ehad280] [Medline: 37216272]
15. Comín-Colet J, Enjuanes C, Verdú-Rotellar JM, Linas A, Ruiz-Rodriguez P, González-Robledo G, et al. Impact on clinical events and healthcare costs of adding telemedicine to multidisciplinary disease management programmes for heart failure: Results of a randomized controlled trial. J Telemed Telecare 2016 Jul;22(5):282-295. [doi: 10.1177/1357633X15600583] [Medline: 26350543]
16. Vestergaard AS, Hansen L, Sørensen SS, Jensen MB, Ehlers LH. Is telehealthcare for heart failure patients cost-effective? an economic evaluation alongside the Danish TeleCare North heart failure trial. BMJ Open 2020 Jan 27;10(1):e031670 [FREE Full text] [doi: 10.1136/bmjopen-2019-031670] [Medline: 31992604]
17. Jiang X, Yao J, You JH. Telemonitoring versus usual care for elderly patients with heart failure discharged from the hospital in the united states: cost-effectiveness analysis. JMIR Mhealth Uhealth 2020 Jul 06;8(7):e17846 [FREE Full text] [doi: 10.2196/17846] [Medline: 32407288]
18. Sydow H, Prescher S, Koehler F, Koehler K, Dorenkamp M, Spethmann S, et al. Cost-effectiveness of noninvasive telemedical interventional management in patients with heart failure: health economic analysis of the TIM-HF2 trial. Clin Res Cardiol 2022 Nov;111(11):1231-1244 [FREE Full text] [doi: 10.1007/s00392-021-01980-2] [Medline: 34894273]
19. Jiménez-Marrero S, Yun S, Cainzos-Achirica M, Enjuanes C, Garay A, Farre N, et al. Impact of telemedicine on the clinical outcomes and healthcare costs of patients with chronic heart failure and mid-range or preserved ejection fraction managed in a multidisciplinary chronic heart failure programme: a sub-analysis of the iCOR randomized trial. J Telemed Telecare 2020;26(1-2):64-72 [FREE Full text] [doi: 10.1177/1357633X18796439] [Medline: 30193564]
20. Ong MK, Romano PS, Edgington S, Aronow HU, Auerbach AD, Black JT, et al. Effectiveness of remote patient monitoring after discharge of hospitalized patients with heart failure: the Better Effectiveness After Transition -- Heart Failure (BEAT-HF) randomized clinical trial. JAMA Intern Med 2016 Mar;176(3):310-318 [FREE Full text] [doi: 10.1001/jamainternmed.2015.7712] [Medline: 26857383]
21. Vuorinen AL, Leppänen J, Kaijanranta H, Kulju M, Heliö T, van Gils M, et al. Use of home telemonitoring to support multidisciplinary care of heart failure patients in Finland: randomized controlled trial. J Med Internet Res 2014 Dec 11;16(12):e282 [FREE Full text] [doi: 10.2196/jmir.3651] [Medline: 25498992]
22. Butler J, Marti C, Pina I, DeFilippi C. Scope of heart failure hospitalization. Congest Heart Fail 2012;18 Suppl 1(Suppl 1):S1-S4 [FREE Full text] [doi: 10.1111/j.1751-7133.2012.00305.x] [Medline: 22891800]

23. Frederix I, Vanderlinden L, Verboven AS, Welten M, Wouters D, De Keulenaer G, et al. Long-term impact of a six-month telemedical care programme on mortality, heart failure readmissions and healthcare costs in patients with chronic heart failure. J Telemed Telecare 2019 Jun;25(5):286-293. [doi: 10.1177/1357633X18774632] [Medline: 29742959]

Abbreviations

BEAT-HF: Better Effectiveness After Transition–Heart Failure
HCP: health care professional
HF: heart failure
ICD-10: International Classification of Diseases, Tenth Revision
NYHA: New York Heart Association
SOC: standard of care
TEMA-HF: Telemonitoring in the Management of Heart Failure
TIM-HF2: German Heart Failure II trial

Edited by L Buis; submitted 16.08.23; peer-reviewed by S Spethmann, N Farre; comments to author 29.09.23; revised version received 24.11.23; accepted 11.12.23; published 07.02.24.

Please cite as:
Kokkonen J, Mustonen P, Heikkilä E, Leskelä RL, Pennanen P, Krühn K, Jalkanen A, Laakso JP, Kempers J, Väisänen S, Torkki P
Effectiveness of Telemonitoring in Reducing Hospitalization and Associated Costs for Patients With Heart Failure in Finland: Nonrandomized Pre-Post Telemonitoring Study
JMIR Mhealth Uhealth 2024;12:e51841
URL: https://mhealth.jmir.org/2024/1/e51841
doi:10.2196/51841
PMID:38324366

Original Paper

Economic Evaluation of a Web Application Implemented in Primary Care for the Treatment of Depression in Patients With Type 2 Diabetes Mellitus: Multicenter Randomized Controlled Trial

Esperanza Varela-Moreno[1,2,3,4], PhD; Maria Teresa Anarte-Ortiz[3,4], PhD; Francisco Jodar-Sanchez[5,6], PhD; Azucena Garcia-Palacios[7,8], PhD; Alicia Monreal-Bartolomé[9,10], PhD; Margalida Gili[11,12], PhD; Javier García-Campayo[2,9], PhD; Fermin Mayoral-Cleries[4,13], PhD

[1]Research and Innovation Unit, Costa del Sol University Hospital, Marbella, Spain

[2]Network for Research on Chronicity, Primary Care, and Health Promotion (RICAPPS), Málaga, Spain

[3]Department of Personality, Assessment and Psychological Treatment, Faculty of Psychology, University of Málaga, Málaga, Spain

[4]Biomedical Research Institute of Malaga, Malaga, Spain

[5]Department of Applied Economics, Faculty of Economics and Business Administration, University of Malaga, Malaga, Spain

[6]Pharmacoeconomics: Clinical and Economic Evaluation of Medications and Palliative Care, Malaga, Spain

[7]Network Biomedical Research Center. Physiopathology Obesity and Nutrition (CIBERobn), Carlos III Health Institute, Madrid, Spain

[8]Department of Clinical and Basic Psychology and Biopsychology, Faculty of Health Sciences, University Jaume I, Castellon, Spain

[9]Institute of Health Research of Aragon, Zaragoza, Spain

[10]Department of Psychology and Sociology, University of Zaragoza, Zaragoza, Spain

[11]Research Network on Preventive Activities and Health Promotion in Primary Health Care (RedIAPP), Madrid, Spain

[12]Institut Universitari d'Investigació en Ciències de la Salut, University Institute for Research in Health Sciences (IUNICS)- Palma Health Research Institute (IDISPA), University of the Balearic Islands, Palma, Spain

[13]Mental Health Clinical Management Unit, University Regional Hospital of Malaga, Malaga, Spain

Corresponding Author:
Francisco Jodar-Sanchez, PhD
Department of Applied Economics
Faculty of Economics and Business Administration
University of Malaga
C. El Ejido, 6
Malaga, 29071
Spain
Phone: 34 951976620
Email: fjodar@uma.es

Abstract

Background: Depressive disorder and type 2 diabetes mellitus (T2DM) are prevalent in primary care (PC). Pharmacological treatment, despite controversy, is commonly chosen due to resource limitations and difficulties in accessing face-to-face interventions. Depression significantly impacts various aspects of a person's life, affecting adherence to medical prescriptions and glycemic control and leading to future complications and increased health care costs. To address these challenges, information and communication technologies (eg, eHealth) have been introduced, showing promise in improving treatment continuity and accessibility. However, while eHealth programs have demonstrated effectiveness in alleviating depressive symptoms, evidence regarding glycemic control remains inconclusive. This randomized controlled trial aimed to test the efficacy of a low-intensity psychological intervention via a web app for mild-moderate depressive symptoms in individuals with T2DM compared with treatment as usual (TAU) in PC.

Objective: This study aimed to analyze the cost-effectiveness and cost-utility of a web-based psychological intervention to treat depressive symptomatology in people with T2DM compared with TAU in a PC setting.

Methods: A multicenter randomized controlled trial was conducted with 49 patients with T2DM, depressive symptoms of moderate severity, and glycosylated hemoglobin (HbA_{1c}) of 7.47% in PC settings. Patients were randomized to TAU (n=27) or a web-based psychological treatment group (n=22). This web-based treatment consisted of cognitive behavioral therapy,

improvement of diabetes self-care behaviors, and mindfulness. Cost-effectiveness analysis for the improvement of depressive symptomatology was conducted based on reductions in 3, 5, or 50 points on the Patient Health Questionnaire–9 (PHQ-9). The efficacy of diabetes control was estimated based on a 0.5% reduction in HbA_{1c} levels. Follow-up was performed at 3 and 6 months. The cost-utility analysis was performed based on quality-adjusted life years.

Results: Efficacy analysis showed that the web-based treatment program was more effective in improving depressive symptoms than TAU but showed only a slight improvement in HbA_{1c}. Incremental cost-effectiveness ratios of 186.76 for a 3-point reduction in PHQ-9 and 206.31 for reductions of 5 and 50 percentage points were obtained. In contrast, the incremental cost-effectiveness ratio for improving HbA_{1c} levels amounted to €1510.90 (€1=US $1.18 in 2018) per participant. The incremental cost-utility ratio resulted in €4119.33 per quality-adjusted life year gained.

Conclusions: The intervention, using web-based modules incorporating cognitive behavioral therapy tools, diabetes self-care promotion, and mindfulness, effectively reduced depressive symptoms and enhanced glycemic control in patients with T2DM. Notably, it demonstrated clinical efficacy and economic efficiency. This supports the idea that eHealth interventions not only benefit patients clinically but also offer cost-effectiveness for health care systems. The study emphasizes the importance of including specific modules to enhance diabetes self-care behaviors in future web-based psychological interventions, emphasizing personalization and adaptation for this population.

Trial Registration: ClinicalTrials.gov NCT03426709; https://clinicaltrials.gov/study/NCT03426709

International Registered Report Identifier (IRRID): RR2-10.1186/S12888-019-2037-3

(JMIR Mhealth Uhealth 2024;12:e55483) doi:10.2196/55483

KEYWORDS

depression; depressive; type 2; diabetes; diabetic; type 2 diabetes mellitus; eHealth; web-based intervention; efficacy; economic evaluation; cost-effectiveness; cost-utility; randomized controlled trial; RCT; randomized; controlled trial; controlled trials; cost; costs; economic; economics; web based; internet based; CBT; psychotherapy; cognitive behavioral therapy; cognitive behavioral therapy; mental health

Introduction

Depressive disorder and type 2 diabetes mellitus (T2DM) are highly prevalent pathologies, mainly treated in primary care (PC) [1-5]. Their comorbidity has been reported in the previous literature in numerous studies [6-12]. However, the diagnosis of depression in PC is very low and only half of the diagnosed patients receive adequate care [1,2]. Pharmacological treatment remains the alternative of choice despite controversial results [13] and the existence of more effective psychological interventions preferred by patients [14,15]. This situation is due to the lack of available resources [16], the difficulty in accessing face-to-face interventions, and the shortage of professionals [17]. The identification and treatment of depression are essential components in the comprehensive treatment of diabetes since the presence of depressive symptoms in these patients not only affects mood but also has repercussions on all aspects of the person's life and influences adherence to medical prescriptions [18-20], glycemic control, and the number of future complications [21,22]. All these have an impact on increased health care costs [18,23], so it is necessary to offer comprehensive treatment aimed at improving both psychological and medical outcomes [24].

Information and communication technologies have recently been incorporated into the health care setting (eHealth) as an alternative to improve the continuity and accessibility of treatment, offering economic advantages. In a recent systematic review [25], the authors identified the main randomized controlled trials published to date on eHealth psychological interventions for depression in adults with diabetes. These programs contributed to improving depressive symptomatology [26-32], but there is still no evidence in improvement in glycemic control. This study aimed to evaluate the efficacy of a low-intensity psychological intervention for the treatment of mild-moderate depressive symptomatology specifically in people with T2DM using a web application compared with treatment as usual (TAU) implemented in PC.

Methods

Design

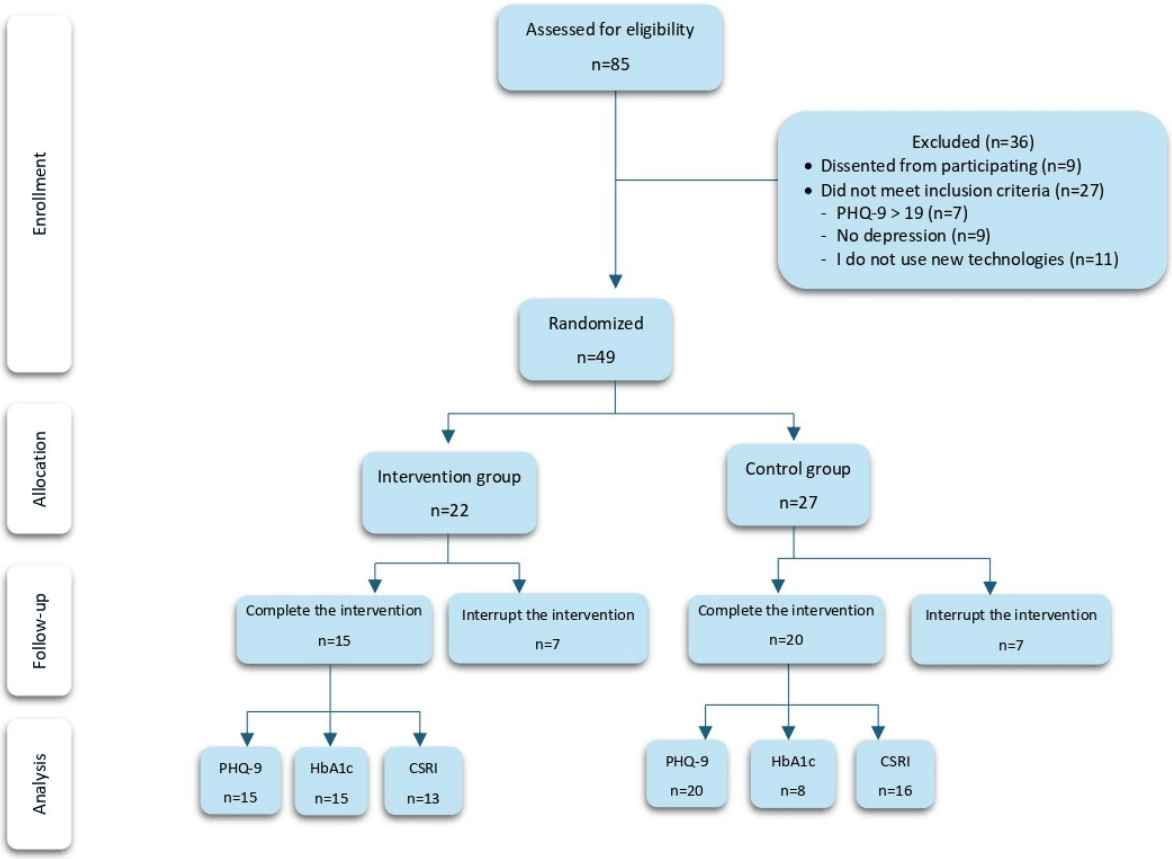
A multicenter randomized controlled trial was conducted. Participants were recruited by family physicians from PC centers in 3 Spanish regions: Andalusia, Aragon, and the Balearic Islands and were randomly assigned to either the intervention group (IG), which included a web-based treatment program, or the control group (CG) that only received TAU.

Study Procedure

A total of 49 patients with T2DM and depression were recruited. Figure 1 presents the flow diagram of the randomization and patient assessment process from patient selection by PC physicians to 6-month follow-up. Each participant, after being identified by his or her PC physician and signing the informed consent form, was referred to the study investigator for baseline assessment. If they met the criteria, randomization was performed independently by a person outside the study. The inclusion period was from December 1, 2018, to December 31, 2019. Inclusion criteria were adults older than 18 years old, diagnosis of depression (*Diagnostic and Statistical Manual of Mental Disorders fifth edition* [*DMS-5*] criteria) with mild or moderate severity (≤19 points on the Patient Health

Questionnaire-9 [PHQ-9]), diagnosis of T2DM (American Diabetes Association), and having an internet connection.

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) flow diagram. CSRI: Client Service Receipt Inventory; HbA1c: glycosylated hemoglobin; PHQ-9: Patient Health Questionnaire–9.



Ethical Considerations

The research study was approved by the Human Research Ethics Committee of the Regional Health Authority of Aragon (CEICA; PI16/0259) and was designed in accordance with the ethical standards laid down in the Declaration of Helsinki and its later amendments. The research study followed the CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) guidelines (Multimedia Appendix 1). All participants signed informed consent forms after being informed of the nature, objectives, potential benefits, and risks associated with participation in the study, as well as the confidentiality of the data collected.

Interventions

The intervention was based on the cognitive behavioral therapy (CBT) and mindfulness self-help program applied through the internet for the treatment of depression “Smiling is fun” [33], which consists of different therapeutic modules. For this study, in addition to the modules previously designed for “Smiling is fun” [33], a specific module for the promotion of diabetes self-care behaviors, a welcome module, and 2 face-to-face sessions were included. The efficacy of this program has been demonstrated in previous studies [33-36], and the study protocol [37] has been published elsewhere. The CG received only TAU, which refers to treatment administered from PC to patients with depression and T2DM.

Study Measures

The PHQ-9 Questionnaire

Depressive symptomatology was assessed using the PHQ-9 [38], Spanish version [39]. The sum of the responses indicates different levels of depression (0-27). In general, scores above 5 suggest the presence of depressive symptomatology. It presents good internal consistency with a Cronbach α coefficient of 0.89.

Glycosylated Hemoglobin

Glycosylated hemoglobin (HbA_{1c}) obtained from blood samples was used as an outcome measure of T2DM control, indicating the average blood glucose level during the last 2 or 3 months [40].

The 12-Item Short Form Survey Health Questionnaire

To assess the quality of life, the 12-Item Short Form Survey (SF-12) [41], Spanish version [42] of this instrument, was used, which provides a profile of health status and consists of 8 dimensions ranging from 0 (the worst health status for that

dimension) to 100 (the best health status). Thus, the higher the score, the higher the perceived quality of life. For this study, the quality-adjusted life years (QALYs) were calculated from the scores of this questionnaire.

Consumption of Health Resources and Cost Procedure

The consumption of health services and the social impact were recorded using the Spanish version of the Client Service Receipt Inventory [43] in a self-reported manner by the patient or using the electronic medical record. Data were collected on those resources derived exclusively from depressive symptomatology and T2DM used during the 6 months before and 6 months after the start of the intervention.

Economic Evaluation

Cost Analysis

The economic evaluation was carried out from a social perspective, including indirect and direct health care costs expressed in euros for the year 2018. Direct health care costs were estimated based on the use of health care services according to Andalusian public prices [44] (Table 1). The costs of developing and maintaining the web page were calculated using the equivalent annual cost method. The intervention program included two 25-minute face-to-face sessions with a psychologist, and this cost was estimated according to the minimum interprofessional salary for a psychologist in Spain [45]. The costs of lost productivity (indirect costs) were estimated using the human capital approach, multiplying the minimum interprofessional wage by the number of days of sick leave [46].

Table 1. Direct health care costs (financial year 2018; values in €[a]).

Health resources	IG[b] (€), mean (SD)	CG[c] (€), mean (SD)
Medical PC[d]	248.08 (364.24)	248.59 (398.31)
Nurse PC	32.29 (82.74)	23.14 (42.73)
Medical PC home	N/A[e]	23.98 (95.92)
Nurse PC home	N/A	5.64 (22.57)
Endocrinology	21.75 (54.26)	21.40 (46.00)
Mental health	30.15 (80.83)	7.13 (28.53)
Emergencies	55.47 (73.04)	27.04 (78.45)
Hospitalizations	N/A	N/A
Total health care cost	387.74 (508.91)	356.94 (500.49)

[a]€1=US $1.18 in 2018.

[b]IG: intervention group.

[c]CG: control group.

[d]PC: primary care.

[e]N/A: not applicable.

Efficacy, Cost-Effectiveness, and Cost-Utility

The outcome measures to assess the efficacy of the psychological web-based treatment program were improvements in depressive symptomatology and HbA_{1c}. Improvement in depressive symptomatology was estimated based on an improvement in PHQ-9 scores of 3, 5, and 50 percentage points. Efficacy for diabetes control was estimated based on a 0.5% reduction in HbA_{1c}. These criteria were determined based on previous studies [28,31,47-51] in the absence of a gold standard criterion.

For each group, cost, incremental cost, efficacy, incremental efficacy, and dominance were calculated, and in the absence of dominance, the results were expressed in terms of incremental cost-effectiveness ratio (ICER).

The unit of measurement for the cost-utility analysis was QALYs, estimated from the SF-12 questionnaire. In addition to QALY, QALY gain was calculated by considering baseline imbalances in utility levels [52]. Similarly, as in the cost-effectiveness analysis (CEA), cost, incremental cost, QALY, incremental QALY, and dominance were calculated for each group, and in the absence of dominance, the results were expressed in terms of incremental cost-utility ratio (ICUR).

The results of the base case were presented in the cost-effectiveness plane that allows us to outline 4 possible scenarios or outcomes after economic evaluation, corresponding to each one of the quadrants of this plan: (1) southeast quadrant—the evaluated technology or new treatment is more effective and less costly than the reference treatment; (2) upper left quadrant—the evaluated technology is less effective and more costly than the reference treatment; (3) southwest quadrant—the evaluated technology is less effective and less costly than the reference treatment; and (4) northeast quadrant—the evaluated technology provides a gain in effectiveness in exchange for equally higher costs.

Sensitivity Analysis

A univariate sensitivity analysis was performed to analyze the robustness of the results obtained after cost-effectiveness and cost-utility analysis, assuming a variability of ±20% in the main components of the economic evaluation: health costs, direct costs associated with the intervention, indirect costs, effectiveness of depression (PHQ-9), effectiveness of diabetes (HbA_{1c}), and QALYs. These results were presented using a tornado diagram.

Statistical Analysis

Data collection was performed with Excel (Microsoft Corp), and statistical analysis was performed with SPSS (version 20; IBM Corp), licensed by the University of Malaga. A descriptive analysis of categorical and qualitative variables was performed, using frequencies and proportions. Means and SDs were calculated for quantitative variables. Comparisons between groups were performed using χ^2 tests and the McNemar test for qualitative variables and the 2-tailed Student *t* test or Mann-Whitney *U* test bilaterally for quantitative variables. In all cases, statistical significance corresponded to a P value of <.05.

Results

Baseline Characteristics of the Sample

The sociodemographic and clinical characteristics are shown in Table 2; there were no significant differences between the groups before the intervention. At the baseline assessment, moderate depressive symptoms (total score between 10 and 14 on the PHQ-9) were observed in both the IG (mean 12.71, SD 3.60) and the CG (mean 11.81, SD 3.14; P=.37). Regarding the treatment for depressive symptoms (N=49), only 12% (n=6) of the patients had attended mental health consultations, 31% (n=15) of the participants were using antidepressants, and 24% (n=12) of the participants were prescribed anxiolytics. In terms of biomedical variables or diabetes control, the IG had an HbA_{1c} level of 7.16%, while the CG had a baseline HbA_{1c} level of 7.78% (P=.29; P=.51). As for the treatment for diabetes control, 45% (n=22) of the participants were using oral antidiabetic drugs and 12% (n=6) of the participants were using insulin.

Table 2. Baseline characteristics of the sample.[a]

Characteristics	Intervention group (n=22)	Control group (n=27)
Sociodemographic		
Age (years), mean (SD)	58.27 (8.01)	56.52 (9.04)
Female, n (%)	16 (73)	12 (44)
Married or in a relationship, n (%)	15 (68)	18 (67)
Educational level, n (%)		
No qualification	6 (27)	6 (22)
High school	12 (54)	11 (41)
College qualification or more	4 (18)	10 (37)
Employment status, n (%)		
Employed	6 (27)	11 (41)
Unemployed	7 (32)	10 (37)
Retired	5 (23)	4 (15)
Depressive symptoms: PHQ-9[b] (range 5-27), mean (SD)	12.71 (3.60)	11.81 (3.14)
HbA_{1c}[c] (%)[d]	7.16	7.78

[a]Intervention and control groups did not significantly differ (P>.05 in all cases) on any of the sociodemographic or clinical baseline characteristics. Values >5 on the PHQ-9 indicate the presence of depressive symptoms.

[b]PHQ-9: Patient Health Questionnaire-9.

[c]HbA_{1c}: glycosylated hemoglobin.

[d]Percentage of glycosylated hemoglobin, absolute values do not apply.

After 6 months from the start of the study, of the total number of patients included, 71% (25/35 participants) of patients with T2DM completed follow-up. The study dropout rate was very similar between both groups, being 36% (8/22) for IG versus 33% (9/27) for CG, with no statistically significant differences. Among the patients in IG (n=22) who dropped out, 14% (n=3) stated that they dropped out because of "lack of time," 4% (n=1) reported that they dropped out because they did not like the web program, and 18% (n=4) dropped out for "other reasons." With respect to the number of therapeutic modules completed, 41% (n=9) of the patients completed all modules, 18% (4/22) did not complete any module, and 4% (n=1) completed only 1 module. No statistically significant differences in sociodemographic and clinical characteristics were found between patients who dropped out and those who completed follow-up, as well as between those who completed all modules and those who did not.

Economic Evaluation

Cost Analysis

The mean health care costs associated with PC consultations (medical and nursing consultations in person and at home), endocrinology, mental health, emergency room visits, and hospital admissions were very similar between the 2 groups, being €387.75 (€1=US $1.18 in 2018) in the IG and €356.94 for the CG, with no statistically significant differences between the groups. The development of the web platform had a total cost of €130,123. However, considering that the complete duration of the treatment, including follow-up, is a total of 6 months, the cost of developing the web platform per patient would be €65.06. On the other hand, the implementation of the intervention had associated personnel costs for the development of 2 face-to-face sessions, estimated at €51.10 per patient. The cost associated with productivity losses for the CG or group that received only the TAU from PC was €61.67 per patient. The total costs for the IG were €503.91 per patient and for the CG €418.61 per patient (Table 3).

Table 3. Estimated total direct and indirect costs for each study group.

	Intervention group (€[a]), mean (SD)	Control group (€), mean (SD)
Health care costs	387.75 (508.91)	356.94 (500.49)
Web development costs	65.06 (0.00)	N/A[b]
Face-to-face psychological intervention	51.10 (0.00)	N/A
Productivity losses	N/A	61.67 (246.67)
Total costs	503.91 (508.91)	418.61 (567.42)

[a] €1=US $1.18 in 2018.

[b]N/A: not applicable.

CEA as an Outcome Measure for the Improvement of Depressive Symptomatology

The web-based psychological intervention proved to be more effective than the TAU in all scenarios analyzed: reduction of 3 points (76.92% vs 31.25%), 5 points, and 50 percentage points in PHQ-9 (53.85% vs 12.50%). The results of the CEA show that this web-based psychological treatment program can be considered an efficient intervention for the control of depressive symptomatology in people with T2DM. In addition, although the IG had a slightly higher mean cost per patient than the CG (€503.91 vs €418.61) mainly due to the costs of the development of the intervention itself, the ICER was €186.76 per patient with no symptoms (for the 3-point scenario) and €206.31 per patient with no symptoms for the 5- and 50-point scenarios (Table 4). In the cost-effectiveness plane, the ICER for the 3 criteria was in the northeast quadrant (upper right), indicating that the web-based intervention was more effective than TAU from PC and at a higher cost (Figure 2).

Table 4. Results of the cost-effectiveness analysis of the web-based psychological intervention program to treat depressive symptomatology versus usual treatment in primary care.

	Cost, €[a], (95% CI)	Efficacy, % (95% CI)	CER[b] (€)
3-point reduction in the PHQ-9[c]			
Psychological web-based intervention	503.91 (266.65 to 794.84)	76.92 (54.00 to 99.80)	655.08
TAU[d]	418.61 (179.75 to 725.94)	31.25 (8.50 to 54.00)	1339.55
Incremental	85.30 (–325.51 to 496.11)	45.67 (13.40 to 77.90)	186.76
5-point reduction in the PHQ-9			
Psychological web-based intervention	503.91 (266.65 to 794.84)	53.85 (26.70 to 80.90)	935.83
TAU	418.61 (179.75 to 725.94)	12.50 (–3.70 to 28.70)	3348.87
Incremental	85.30 (–325.51 to 496.11)	41.35 (9.80 to 72.90)	206.31
50 percentage point reduction PHQ-9			
Psychological web-based intervention	503.91 (266.65 to 794.84)	53.85 (26.70 to 80.90)	935.83
TAU	418.61 (179.75 to 725.94)	12.50 (–3.70 to 28.70)	3348.87
Incremental	85.30 (–325.51 to 496.11)	41.35 (9.80 to 72.90)	206.31
0.5 reduction in HbA_{1c}[e]			
Psychological web-based intervention	481.59 (231.05 to 801.47)	8.33 (–7.30 to 24.00)	5779.13
TAU	571.53 (179.44 to 1177.61)	14.29 (–11.60 to 40.20)	4000.70
Incremental	–89.93 (–773.18 to 593.31)	–5.95 (–36.20 to 24.30)	1510.90

[a] €1=US $1.18 in 2018.

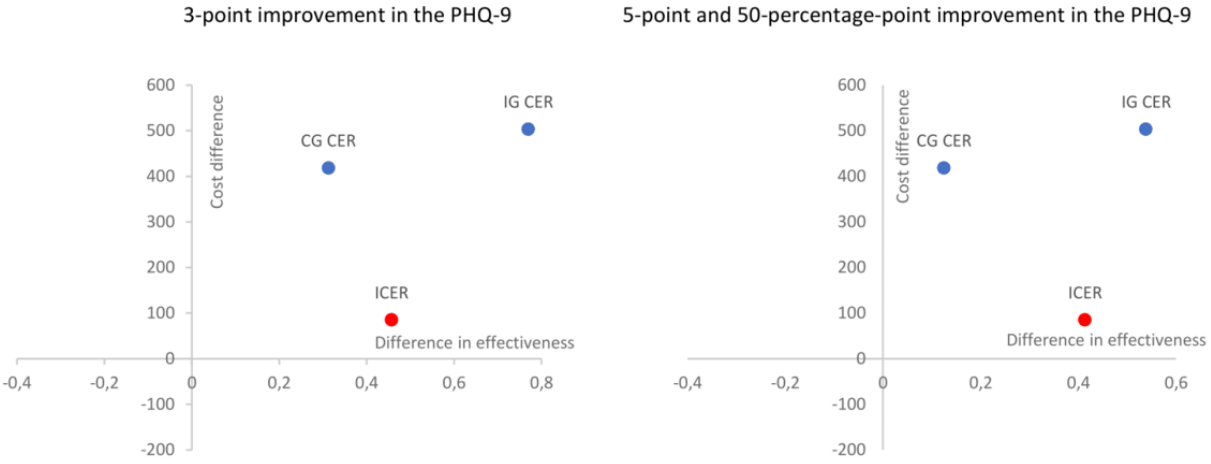
[b]CER: cost-effectiveness ratio.

[c]PHQ-9: Patient Health Questionnaire-9.

[d]TAU: treatment as usual.

[e]HbA_{1c}: glycosylated hemoglobin.

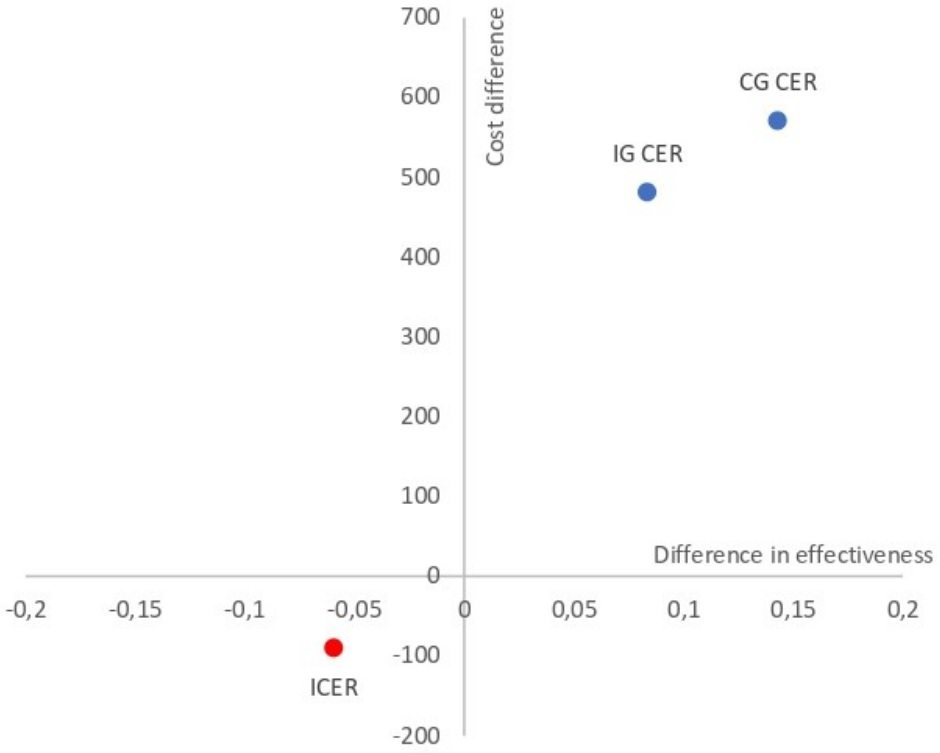
Figure 2. Cost-effectiveness plan for the improvement of depressive symptomatology in patients with type 2 diabetes mellitus after web-based psychological intervention. CG CER: control group cost-effectiveness ratio; ICER: incremental cost-effectiveness ratio; IG CER: intervention group cost-effectiveness ratio; PHQ-9: Patient Health Questionnaire-9.



CEA as an Outcome Measure of Improvement in HbA1c Levels

The web-based psychological intervention to improve HbA_{1c} levels in people with T2DM is reduced in IG compared with CG (8.33% vs 14.29%). Furthermore, although IG had a slightly lower cost per patient than CG (€481.59 vs €571.53), the ICER was €1510.90 (Table 4). In the cost-effectiveness plane, the ICER is in the southwest quadrant (lower left), indicating that the web-based intervention was less effective and less costly than the TAU from PC (Figure 3).

Figure 3. Cost-effectiveness plan for the improvement of HbA1c control in patients with type 2 diabetes mellitus after web-based psychological intervention. CG CER: control group cost-effectiveness ratio; ICER: incremental cost-effectiveness ratio; IG CER: intervention group cost-effectiveness ratio.



Cost-Utility Analysis

Both study groups started from a baseline level of utility that was practically similar (QALY 0.5898 vs 0.5935). The ICUR was €4119.33 per QALY. Therefore, the ICUR shows that the group that received the web-based psychological intervention program from PC achieved greater health gains (incremental QALY of 0.0207) and a higher cost compared to the group that only received TAU. Regarding the effect on QALY gained, this was higher in the IG (0.0321) than in the CG (0.0095). The ICUR resulted in 4174.92 per QALY gained. Therefore, the CUA shows that the group receiving the web-based psychological intervention program from PC achieved greater health gains (incremental QALY gain of 0.0225) and a higher cost compared to the group receiving only TAU (Table 5). Figure 4 shows the results of the cost-utility plane using QALY and QALY gained as the benefit outcome measure. ICUR is in the northwest quadrant (lower left), indicating that the web-based psychological intervention is more costly but contributes to increased QALYs.

Table 5. Results of the cost-utility analysis of the web-based psychological intervention program for the improvement of depressive symptomatology versus usual treatment.

	Cost, €[a] (95% CI	Gain (95% CI)	CUR[b], €
QALYs[c]			
Psychological web-based intervention	503.91 (266.65 to 794.84	0.3270 (0.2920 to 0.3623	1541.10
TAU[d]	418.61 (179.75 to 725.94	0.3063 (0.2840 to 0.3278	1366.78
Incremental	85.30 (–325.51 to 496.11	0.0207 (–0.0229 to 0.0643	4119.33
QALYs gained			
Psychological web-based intervention	503.91 (266.65 to 794.84	0.0321 (0.0133 to 0.0512	15718.80
TAU	418.61 (179.75 to 725.94	0.0095 (–0.0036 to 0.0234	43955.57
Incremental	85.30 (–325.51 to 496.11	0.0225 (–0.0023 to 0.0474	3785.36

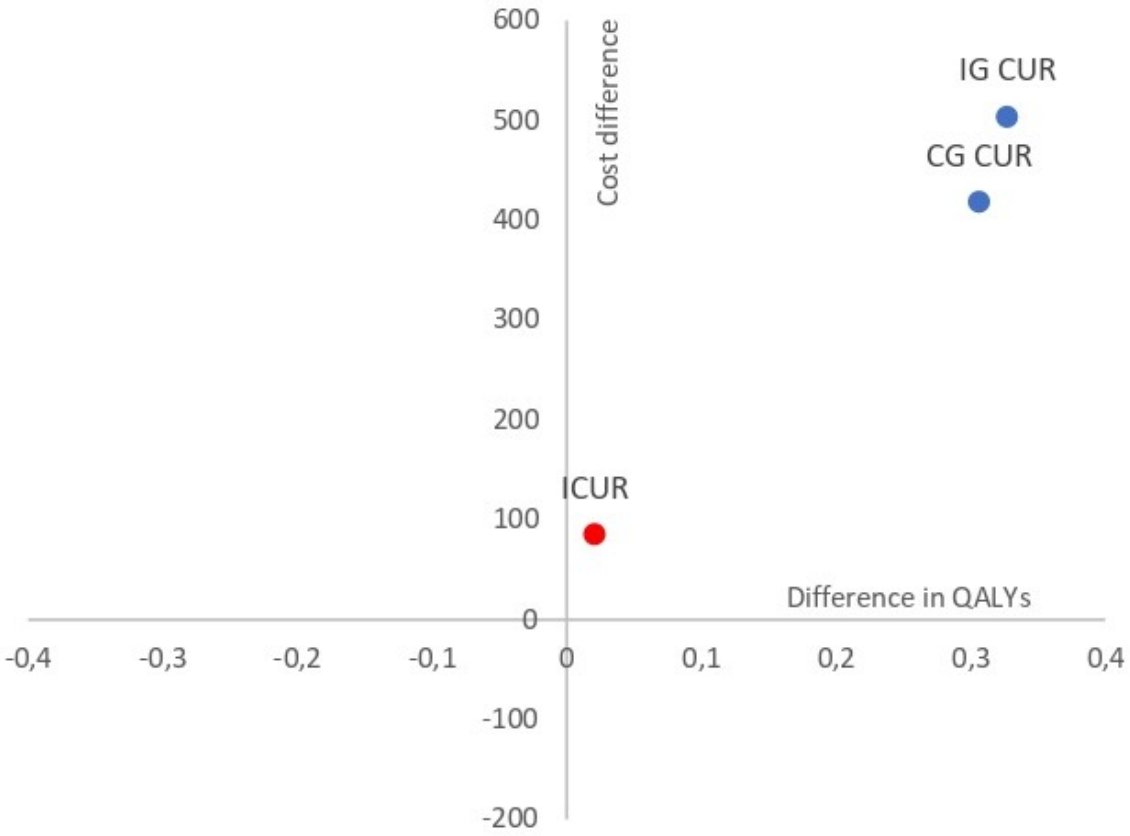
[a] €1=US $1.18 in 2018.

[b]CUR: cost-utility ratio.

[c]QALYs: quality-adjusted life year.

[d]TAU: treatment as usual.

Figure 4. Cost-utility per QALY of the web-based psychological intervention program for the improvement of depressive symptomatology in people with type 2 diabetes mellitus. CG CUR: control group cost-utility ratio; ICUR: incremental cost-utility ratio; IG CUR: intervention group cost-utility ratio; QALYs: quality-adjusted life years.

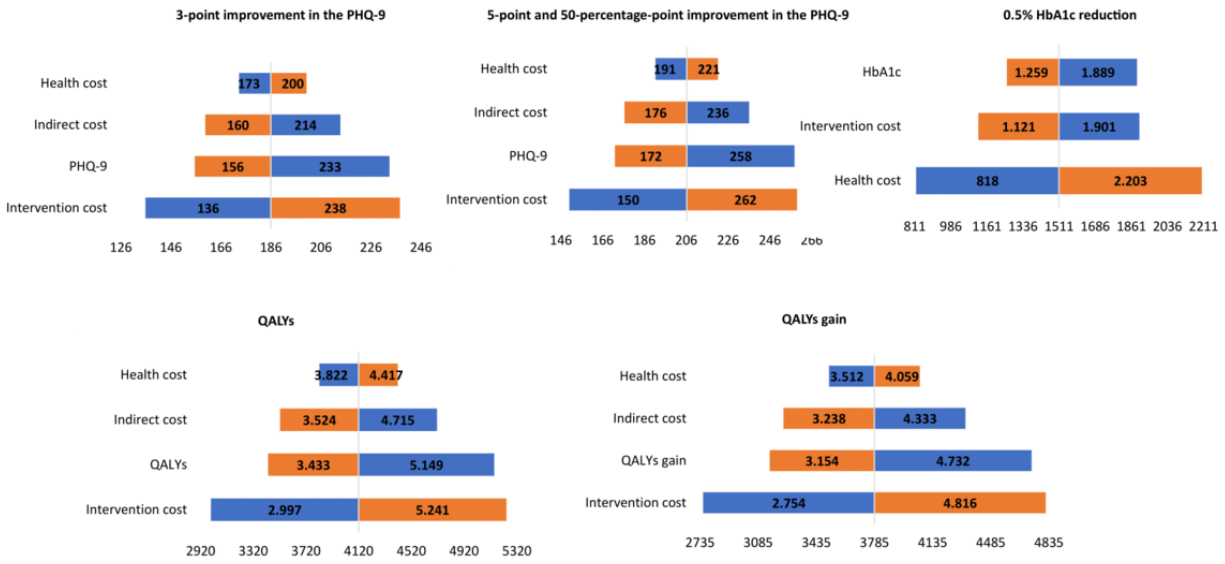


Sensitivity Analysis

Sensitivity analysis again showed that the ICER was cost-effective and that the results obtained were robust with respect to the base case. The variable that generated the most uncertainty in almost all cases was the costs of the intervention. However, taking a reduction in HbA_{1c} as a reference, the variable that generated the most uncertainty was health care

costs. These results can be seen in Figure 5. Taking as a reference the results obtained based on a reduction in the PHQ-9 as a criterion for improving depressive symptomatology, the variable that generated the most uncertainty was the costs of the intervention, as shown in Figure 5. The results were obtained based on a reduction of 3 points on the PHQ-9 as a criterion for improving depressive symptomatology, and the variable that generated the most uncertainty was the costs of the intervention, with oscillations of between €136 per patient with no depressive symptomatology and €238 per patient with no depressive symptomatology, considering the best- and worst-case scenarios. In criterion B and C, reduction of 5 points and 50 percentage points on the PHQ-9. The variable that generated the most uncertainty was also the cost of the intervention, with oscillations of between €150 per patient without depressive symptomatology and €262 per patient without depressive symptomatology.

Figure 5. Sensitivity analysis for a web-based psychological intervention program. HbA1c: glycosylated hemoglobin; PHQ-9: Patient Health Questionnaire-9; QALYs: quality-adjusted life years.



The results of the sensitivity analysis for the ICUR, using QALYs as a measure of benefit, were also robust to the base case. Figure 5 shows how variations are associated with oscillations between €2997 per QALY and €5241 per QALY for patients with no depressive symptomatology. On the other hand, variations in health care costs ranged from €3822 per QALY to €4417 per QALY for patients with no depressive symptomatology. The results of this analysis for the ICUR using the QALY gain as a measure of benefit obtained results are similar to those described in the case of QALYs. These results show that they are cost-utility (Figure 5).

Discussion

Principal Findings

This study performed the economic evaluation of a psychological intervention based on a web-based treatment with CBT and mindfulness to reduce depressive symptomatology and improve diabetes control. Although this type of intervention has been analyzed in other studies [26,28,30,31,53-56], this is the first study developed specifically for people with T2DM, not including patients with other types of diabetes (eg, patients with type 1 diabetes mellitus together with patients with T2DM) who, therefore, have different treatments and characteristics. It also includes a specific module aimed at promoting diabetes self-care behaviors and implemented directly by the Spanish PC health system.

The results of the CEA show that the intervention proved to be more effective and more costly in reducing the symptoms of depression compared to the TAU group. These results are to be expected after analyzing novel interventions developed using information and communication technologies and are in line with previous studies designed for the general population [36,57-61]. For a population with diabetes, only the study presented by Nobis et al [62] published results of CEA and CUA of a similar intervention, obtaining results that are equally in line with those obtained by this study.

Finally, the CUA results show that the intervention improved quality of life, measured in terms of QALYs, compared with TAU in PC. These results, with ICUR of €4119 per QALY and €3785 per QALY gained, are below the threshold value of between €22,000 and €25,000 per QALY estimated in Spain [63,64] as the maximum willingness to pay per QALY gained. Compared with previous studies, our results are very promising. Warmerdam et al [65] found that the additional cost of obtaining a QALY after the use of web-based CBT was €22,609 compared with the waiting list group, which is much higher than the results obtained in this study. Nobis et al [62] concluded that treatment response was not reflected in QALY gains, and the extra cost of obtaining a QALY amounted to €14,000. Romero-Sanchiz

et al [36] have been the only economic analysis study on a web-based psychological intervention performed so far in Spain in the PC setting in depression and showed a reduction in the cost of €11,390.

This study has limitations. First, difficulties were encountered in obtaining blood samples and determining HbA_{1c} during the follow-up phase. These difficulties were generated by the COVID-19 confinement. However, despite this circumstance, the results obtained are similar to those reported in previous studies [28-31,53,54] on face-to-face psychological interventions. Second, the cost analysis did not include the consumption associated with drugs for depression or diabetes. However, the reported medication costs were equally distributed between the 2 groups, so they would not make much difference to the results of the economic evaluation [36]. Finally, there were limitations in defining the criteria for the efficacy of the intervention because there was no gold standard for improvements in depressive symptomatology or HbA_{1c}.

Although the results are promising, more studies are needed because the evidence on the cost-effectiveness of these interventions is scarce. Therefore, studies are needed to replicate this intervention, increasing the sample size to be able to compare and generalize the results obtained. Likewise, the current challenge is directed toward the design of web-based psychological interventions in a personalized way according to the pathology to be treated to achieve better results and better adherence to them [66,67].

Conclusions

This intervention, devised to reduce depressive symptoms and enhance glycemic control among individuals with T2DM through a series of online web-based modules incorporating CBT tools, promotion of diabetes self-care behaviors, and mindfulness, has proven to be both effective and efficient. This effectiveness extends to both clinical outcomes and economic considerations. Consequently, these eHealth interventions not only facilitate clinical improvements for patients but also demonstrate cost-effectiveness for health care systems.

The findings of this study underscore the significance of incorporating a specific module focused on enhancing diabetes self-care behaviors. This element is deemed indispensable for the future development of web-based psychological interventions, ensuring their personalization and adaptation to the unique needs of individuals within this diabetic population. Moreover, the study supports the notion that interventions targeting the improvement of depressive symptoms in individuals with diabetes, while essential, may not be adequate in isolation for optimizing the overall clinical management of diabetes [25,30,68-71].

In summary, this intervention not only directly benefits patients by improving their mental and physical health but also brings efficiency and adaptability to health care settings, highlighting its importance in the effective management of diabetes in diverse health care settings.

Acknowledgments

The authors thank all the health care staff of the primary care centers who participated in this study for their outstanding collaboration in the recruitment and referral of participants. They also thank all the researchers who participated in the fieldwork. AM-B has a Margarita Salas postdoctoral fellowship at the University of Zaragoza, funded by the NextGeneration EU. This study has been funded by Carlos III Health Institute through the project PI19/01131 and was cofunded by the European Union. AMB had an FI predoctoral contract awarded by the Carlos III Health Institute (FI17/00180) at the Institute for Health Research of Aragon to conduct this study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) checklist.
[PDF File (Adobe PDF File), 1156 KB - mhealth_v12i1e55483_app1.pdf]

References

1. Craven MA, Bland R. Depression in primary care: current and future challenges. Can J Psychiatry 2013;58(8):442-448. [doi: 10.1177/070674371305800802] [Medline: 23972105]
2. Mitchell AJ, Vaze A, Rao S. Clinical diagnosis of depression in primary care: a meta-analysis. Lancet 2009;374(9690):609-619. [doi: 10.1016/S0140-6736(09)60879-5] [Medline: 19640579]
3. Fernández A, Pinto-Meza A, Bellón JA, Roura-Poch P, Haro JM, Autonell J, et al. Is major depression adequately diagnosed and treated by general practitioners? Results from an epidemiological study. Gen Hosp Psychiatry 2010;32(2):201-209. [doi: 10.1016/j.genhosppsych.2009.11.015] [Medline: 20302995]
4. Saeedi P, Petersohn I, Salpea P, Malanda B, Karuranga S, Unwin N, et al. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: results from the International Diabetes Federation Diabetes Atlas, 9 edition. Diabetes Res Clin Pract 2019;157:107843 [FREE Full text] [doi: 10.1016/j.diabres.2019.107843] [Medline: 31518657]

5. Khan MAB, Hashim MJ, King JK, Govender RD, Mustafa H, Kaabi JA. Epidemiology of type 2 diabetes—global burden of disease and forecasted trends. J Epidemiol Glob Health 2020;10(1):107-111 [FREE Full text] [doi: 10.2991/jegh.k.191028.001] [Medline: 32175717]
6. Anderson RJ, Freedland KE, Clouse RE, Lustman PJ. The prevalence of comorbid depression in adults with diabetes: a meta-analysis. Diabetes Care 2001;24(6):1069-1078 [FREE Full text] [doi: 10.2337/diacare.24.6.1069] [Medline: 11375373]
7. Chapman DP, Perry GS, Strine TW. The vital link between chronic disease and depressive disorders. Prev Chronic Dis 2005;2(1):A14 [FREE Full text] [Medline: 15670467]
8. Chen PC, Chan YT, Chen HF, Ko MC, Li CY. Population-based cohort analyses of the bidirectional relationship between type 2 diabetes and depression. Diabetes Care 2013;36(2):376-382 [FREE Full text] [doi: 10.2337/dc12-0473] [Medline: 23150281]
9. Mezuk B, Eaton WW, Albrecht S, Golden SH. Depression and type 2 diabetes over the lifespan: a meta-analysis. Diabetes Care 2008;31(12):2383-2390 [FREE Full text] [doi: 10.2337/dc08-0985] [Medline: 19033418]
10. Khaledi M, Haghighatdoost F, Feizi A, Aminorroaya A. The prevalence of comorbid depression in patients with type 2 diabetes: an updated systematic review and meta-analysis on huge number of observational studies. Acta Diabetol 2019;56(6):631-650. [doi: 10.1007/s00592-019-01295-9] [Medline: 30903433]
11. Farooqi A, Gillies C, Sathanapally H, Abner S, Seidu S, Davies MJ, et al. A systematic review and meta-analysis to compare the prevalence of depression between people with and without type 1 and type 2 diabetes. Prim Care Diabetes 2022;16(1):1-10. [doi: 10.1016/j.pcd.2021.11.001] [Medline: 34810141]
12. Lindekilde N, Scheuer SH, Rutters F, Knudsen L, Lasgaard M, Rubin KH, et al. Prevalence of type 2 diabetes in psychiatric disorders: an umbrella review with meta-analysis of 245 observational studies from 32 systematic reviews. Diabetologia 2022;65(3):440-456 [FREE Full text] [doi: 10.1007/s00125-021-05609-x] [Medline: 34841451]
13. Lustman PJ, Griffith LS, Clouse RE, Freedland KE, Eisen SA, Rubin EH, et al. Effects of nortriptyline on depression and glycemic control in diabetes: results of a double-blind, placebo-controlled trial. Psychosom Med 1997;59(3):241-250. [doi: 10.1097/00006842-199705000-00007] [Medline: 9178335]
14. Hofmann SG, Asnaani A, Vonk IJJ, Sawyer AT, Fang A. The efficacy of cognitive behavioral therapy: a review of meta-analyses. Cognit Ther Res 2012;36(5):427-440 [FREE Full text] [doi: 10.1007/s10608-012-9476-1] [Medline: 23459093]
15. Cuijpers P, van Straten A, Warmerdam L, Andersson G. Psychotherapy versus the combination of psychotherapy and pharmacotherapy in the treatment of depression: a meta-analysis. Depress Anxiety 2009;26(3):279-288 [FREE Full text] [doi: 10.1002/da.20519] [Medline: 19031487]
16. Egede LE, Ellis C. Diabetes and depression: global perspectives. Diabetes Res Clin Pract 2010;87(3):302-312 [FREE Full text] [doi: 10.1016/j.diabres.2010.01.024] [Medline: 20181405]
17. Lehtinen V, Riikonen E, Lahtinen E. Promotion of mental health on the European agenda. Julkari. 1997. URL: https://www.julkari.fi/bitstream/handle/10024/75859/Promotion.pdf?sequenc [accessed 2024-04-16]
18. Ciechanowski PS, Katon WJ, Russo JE. Depression and diabetes: impact of depressive symptoms on adherence, function, and costs. Arch Intern Med 2000;160(21):3278-3285 [FREE Full text] [doi: 10.1001/archinte.160.21.3278] [Medline: 11088090]
19. Osborn CY, Egede LE. The relationship between depressive symptoms and medication nonadherence in type 2 diabetes: the role of social support. Gen Hosp Psychiatry 2012;34(3):249-253 [FREE Full text] [doi: 10.1016/j.genhosppsych.2012.01.015] [Medline: 22401705]
20. Zuberi SI, Syed EU, Bhatti JA. Association of depression with treatment outcomes in type 2 diabetes mellitus: a cross-sectional study from Karachi, Pakistan. BMC Psychiatry 2011;11:27 [FREE Full text] [doi: 10.1186/1471-244X-11-27] [Medline: 21324127]
21. Lin EHB, Katon W, Von Korff M, Rutter C, Simon GE, Oliver M, et al. Relationship of depression and diabetes self-care, medication adherence, and preventive care. Diabetes Care 2004;27(9):2154-2160 [FREE Full text] [doi: 10.2337/diacare.27.9.2154] [Medline: 15333477]
22. Lustman PJ, Anderson RJ, Freedland KE, de Groot M, Carney RM, Clouse RE. Depression and poor glycemic control: a meta-analytic review of the literature. Diabetes Care 2000;23(7):934-942. [doi: 10.2337/diacare.23.7.934] [Medline: 10895843]
23. Egede LE, Zheng D, Simpson K. Comorbid depression is associated with increased health care use and expenditures in individuals with diabetes. Diabetes Care 2002;25(3):464-470 [FREE Full text] [doi: 10.2337/diacare.25.3.464] [Medline: 11874931]
24. Petrak F, Herpertz S. Treatment of depression in diabetes: an update. Curr Opin Psychiatry 2009;22(2):211-217. [doi: 10.1097/YCO.0b013e3283207b45] [Medline: 19553878]
25. Varela-Moreno E, Soler MC, Guzmán-Parra J, Jódar-Sánchez F, Mayoral-Cleries F, Anarte-Ortíz MT. Effectiveness of eHealth-based psychological interventions for depression treatment in patients with type 1 or type 2 diabetes mellitus: a systematic review. Front Psychol 2021;12:746217 [FREE Full text] [doi: 10.3389/fpsyg.2021.746217] [Medline: 35173644]

26. Clarke J, Sanatkar S, Baldwin PA, Fletcher S, Gunn J, Wilhelm K, et al. A web-based cognitive behavior therapy intervention to improve social and occupational functioning in adults with type 2 diabetes (The SpringboarD Trial): randomized controlled trial. J Med Internet Res 2019;21(5):e12246 [FREE Full text] [doi: 10.2196/12246] [Medline: 31115345]
27. Egede LE, Walker RJ, Payne EH, Knapp RG, Acierno R, Frueh BC. Effect of psychotherapy for depression via home telehealth on glycemic control in adults with type 2 diabetes: subgroup analysis of a randomized clinical trial. J Telemed Telecare 2018;24(9):596-602. [doi: 10.1177/1357633X17730419] [Medline: 28945160]
28. Naik AD, Hundt NE, Vaughan EM, Petersen NJ, Zeno D, Kunik ME, et al. Effect of telephone-delivered collaborative goal setting and behavioral activation vs enhanced usual care for depression among adults with uncontrolled diabetes: a randomized clinical trial. JAMA Netw Open 2019;2(8):e198634 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.8634] [Medline: 31390035]
29. Newby J, Robins L, Wilhelm K, Smith J, Fletcher T, Gillis I, et al. Web-based cognitive behavior therapy for depression in people with diabetes mellitus: a randomized controlled trial. J Med Internet Res 2017;19(5):e157 [FREE Full text] [doi: 10.2196/jmir.7274] [Medline: 28506956]
30. Nobis S, Lehr D, Ebert DD, Baumeister H, Snoek F, Riper H, et al. Efficacy of a web-based intervention with mobile phone support in treating depressive symptoms in adults with type 1 and type 2 diabetes: a randomized controlled trial. Diabetes Care 2015;38(5):776-783 [FREE Full text] [doi: 10.2337/dc14-1728] [Medline: 25710923]
31. Piette JD, Richardson C, Himle J, Duffy S, Torres T, Vogel M, et al. A randomized trial of telephonic counseling plus walking for depressed diabetes patients. Med Care 2011;49(7):641-648 [FREE Full text] [doi: 10.1097/MLR.0b013e318215d0c9] [Medline: 21478777]
32. van Bastelaar KMP, Pouwer F, Cuijpers P, Riper H, Snoek FJ. Web-based depression treatment for type 1 and type 2 diabetic patients: a randomized, controlled trial. Diabetes Care 2011;34(2):320-325 [FREE Full text] [doi: 10.2337/dc10-1248] [Medline: 21216855]
33. López-del-Hoyo Y, Olivan B, Luciano JV, Mayoral F, Roca M, Gili M, et al. Low intensity vs. self-guided internet-delivered psychotherapy for major depression: a multicenter, controlled, randomized study. BMC Psychiatry 2013;13:21 [FREE Full text] [doi: 10.1186/1471-244X-13-21] [Medline: 23312003]
34. Mira A, Soler C, Alda M, Baños R, Castilla D, Castro A, et al. Exploring the relationship between the acceptability of an internet-based intervention for depression in primary care and clinical outcomes: secondary analysis of a randomized controlled trial. Front Psychiatry 2019;10:325 [FREE Full text] [doi: 10.3389/fpsyt.2019.00325] [Medline: 31133899]
35. Montero-Marín J, Araya R, Pérez-Yus MC, Mayoral F, Gili M, Botella C, et al. An internet-based intervention for depression in primary care in Spain: a randomized controlled trial. J Med Internet Res 2016;18(8):e231 [FREE Full text] [doi: 10.2196/jmir.5695] [Medline: 27565118]
36. Romero-Sanchiz P, Nogueira-Arjona R, García-Ruiz A, Luciano JV, Campayo JG, Gili M, et al. Economic evaluation of a guided and unguided internet-based CBT intervention for major depression: results from a multi-center, three-armed randomized controlled trial conducted in primary care. PLoS One 2017;12(2):e0172741 [FREE Full text] [doi: 10.1371/journal.pone.0172741] [Medline: 28241025]
37. Monreal-Bartolomé A, Barceló-Soler A, Castro A, Pérez-Ara M, Gili M, Mayoral F, et al. Efficacy of a blended low-intensity internet-delivered psychological programme in patients with multimorbidity in primary care: study protocol for a randomized controlled trial. BMC Psychiatry 2019;19(1):66 [FREE Full text] [doi: 10.1186/s12888-019-2037-3] [Medline: 30744610]
38. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
39. Diez-Quevedo C, Rangil T, Sanchez-Planell L, Kroenke K, Spitzer RL. Validation and utility of the patient health questionnaire in diagnosing mental disorders in 1003 general hospital Spanish inpatients. Psychosom Med 2001;63(4):679-686. [doi: 10.1097/00006842-200107000-00021] [Medline: 11485122]
40. American Diabetes Association Professional Practice Committee. 6. Glycemic targets: standards of medical care in diabetes-2022. Diabetes Care 2022;45(Suppl 1):S83-S96 [FREE Full text] [doi: 10.2337/dc22-S006] [Medline: 34964868]
41. Ware J, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med Care 1996;34(3):220-233. [doi: 10.1097/00005650-199603000-00003] [Medline: 8628042]
42. Alonso J, Regidor E, Barrio G, Prieto L, Rodríguez C, de la Fuente L. Population reference values of the Spanish version of the health questionnaire SF-36. Med Clin (Barc) 1998;111(11):410-416. [Medline: 9834913]
43. Vázquez-Barquero JL, Gaite L, Cuesta MJ, García UE, Knapp M, Beecham J. Spanish version of the CSRI: a mental health cost evaluation interview. Arch Neurobiol (Madr) 1997;60(2):171-184.
44. BOJA 210, of October 27, 2005. General provisions. Junta of Andalusia. URL: https://www.juntadeandalucia.es/boja/2005/210/28 [accessed 2024-04-16]
45. Regulations on public prices. Andalusian Health Service. 2005. URL: https://www.sspa.juntadeandalucia.es/servicioandaluzdesalud/sites/default/files/sincfiles/wsas-media-sas_normativa_mediafile/2019/orden2005_precios_publicos.pdf
46. Spain in figures 2017. INE. URL: https://www.ine.es/prodyser/espa_cifras/2017/files/assets/common/downloads/publication.pdf [accessed 2024-04-16]

47. Cramer H, Salisbury C, Conrad J, Eldred J, Araya R. Group cognitive behavioural therapy for women with depression: pilot and feasibility study for a randomised controlled trial using mixed methods. BMC Psychiatry 2011;11:82 [FREE Full text] [doi: 10.1186/1471-244X-11-82] [Medline: 21569488]
48. Richards DA, Hill JJ, Gask L, Lovell K, Chew-Graham C, Bower P, et al. Clinical effectiveness of collaborative care for depression in UK primary care (CADET): cluster randomised controlled trial. BMJ 2013;347:f4913 [FREE Full text] [doi: 10.1136/bmj.f4913] [Medline: 23959152]
49. Rojas G, Martínez P, Guajardo V, Campos S, Herrera P, Vöhringer PA, et al. A collaborative, computer-assisted, psycho-educational intervention for depressed patients with chronic disease at primary care: protocol for a cluster randomized controlled trial. BMC Psychiatry 2021;21(1):418 [FREE Full text] [doi: 10.1186/s12888-021-03380-2] [Medline: 34419010]
50. Scazufca M, de Paula Couto MCP, Henrique MG, Mendes AV, Matijasevich A, Pereda PC, et al. Pilot study of a two-arm non-randomized controlled cluster trial of a psychosocial intervention to improve late life depression in socioeconomically deprived areas of São Paulo, Brazil (PROACTIVE): feasibility study of a psychosocial intervention for late life depression in São Paulo. BMC Public Health 2019;19(1):1152 [FREE Full text] [doi: 10.1186/s12889-019-7495-5] [Medline: 31438903]
51. Scazufca M, Nakamura CA, Peters TJ, Henrique MG, Seabra A, La Rotta EG, et al. A collaborative care psychosocial intervention to improve late life depression in socioeconomically deprived areas of Guarulhos, Brazil: the PROACTIVE cluster randomised controlled trial protocol. Trials 2020;21(1):914 [FREE Full text] [doi: 10.1186/s13063-020-04826-w] [Medline: 33153482]
52. Manca A, Hawkins N, Sculpher MJ. Estimating mean QALYs in trial-based cost-effectiveness analysis: the importance of controlling for baseline utility. Health Econ 2005;14(5):487-496. [doi: 10.1002/hec.944] [Medline: 15497198]
53. van Bastelaar K, Cuijpers P, Pouwer F, Riper H, Snoek FJ. Development and reach of a web-based cognitive behavioural therapy programme to reduce symptoms of depression and diabetes-specific distress. Patient Educ Couns 2011;84(1):49-55. [doi: 10.1016/j.pec.2010.06.013] [Medline: 20619577]
54. Ebert DD, Nobis S, Lehr D, Baumeister H, Riper H, Auerbach RP, et al. The 6-month effectiveness of internet-based guided self-help for depression in adults with type 1 and 2 diabetes mellitus. Diabet Med 2017;34(1):99-107. [doi: 10.1111/dme.13173] [Medline: 27334444]
55. Baldwin PA, Sanatkar S, Clarke J, Fletcher S, Gunn J, Wilhelm K, et al. A web-based mental health intervention to improve social and occupational functioning in adults with type 2 diabetes (the springboard trial): 12-month outcomes of a randomized controlled trial. J Med Internet Res 2020;22(12):e16729 [FREE Full text] [doi: 10.2196/16729] [Medline: 33258790]
56. van Bastelaar KMP, Pouwer F, Cuijpers P, Riper H, Twisk JWR, Snoek FJ. Is a severe clinical profile an effect modifier in a web-based depression treatment for adults with type 1 or type 2 diabetes? Secondary analyses from a randomized controlled trial. J Med Internet Res 2012;14(1):e2 [FREE Full text] [doi: 10.2196/jmir.1657] [Medline: 22262728]
57. Geraedts AS, van Dongen JM, Kleiboer AM, Wiezer NM, van Mechelen W, Cuijpers P, et al. Economic evaluation of a web-based guided self-help intervention for employees with depressive symptoms: results of a randomized controlled trial. J Occup Environ Med 2015;57(6):666-675. [doi: 10.1097/JOM.0000000000000423] [Medline: 25741795]
58. Gerhards SAH, de Graaf LE, Jacobs LE, Severens JL, Huibers MJH, Arntz A, et al. Economic evaluation of online computerised cognitive-behavioural therapy without support for depression in primary care: randomised trial. Br J Psychiatry 2010;196(4):310-318 [FREE Full text] [doi: 10.1192/bjp.bp.109.065748] [Medline: 20357309]
59. Hollinghurst S, Peters TJ, Kaur S, Wiles N, Lewis G, Kessler D. Cost-effectiveness of therapist-delivered online cognitive-behavioural therapy for depression: randomised controlled trial. Br J Psychiatry 2010;197(4):297-304 [FREE Full text] [doi: 10.1192/bjp.bp.109.073080] [Medline: 20884953]
60. McCrone P, Knapp M, Proudfoot J, Ryden C, Cavanagh K, Shapiro DA, et al. Cost-effectiveness of computerised cognitive-behavioural therapy for anxiety and depression in primary care: randomised controlled trial. Br J Psychiatry 2004;185:55-62 [FREE Full text] [doi: 10.1192/bjp.185.1.55] [Medline: 15231556]
61. Jagayat JK, Kumar A, Shao Y, Pannu A, Patel C, Shirazi A, et al. Incorporating a stepped care approach into internet-based cognitive behavioral therapy for depression: randomized controlled trial. JMIR Ment Health 2024;11:e51704 [FREE Full text] [doi: 10.2196/51704] [Medline: 38173167]
62. Nobis S, Ebert DD, Lehr D, Smit F, Buntrock C, Berking M, et al. Web-based intervention for depressive symptoms in adults with types 1 and 2 diabetes mellitus: a health economic evaluation. Br J Psychiatry 2018;212(4):199-206 [FREE Full text] [doi: 10.1192/bjp.2018.10] [Medline: 30071909]
63. Sacristán JA, Oliva J, Campillo-Artero C, Puig-Junoy J, Pinto-Prades JL, Dilla T, et al. ¿Qué es una intervención sanitaria eficiente en España en 2020? [What is an efficient health intervention in Spain in 2020?]. Gac Sanit 2020;34(2):189-193 [FREE Full text] [doi: 10.1016/j.gaceta.2019.06.007] [Medline: 31558385]
64. Vallejo-Torres L, García-Lorenzo B, Serrano-Aguilar P. Estimating a cost-effectiveness threshold for the Spanish NHS. Health Econ 2018;27(4):746-761. [doi: 10.1002/hec.3633] [Medline: 29282798]
65. Warmerdam L, Smit F, van Straten A, Riper H, Cuijpers P. Cost-utility and cost-effectiveness of internet-based treatment for adults with depressive symptoms: randomized trial. J Med Internet Res 2010;12(5):e53 [FREE Full text] [doi: 10.2196/jmir.1436] [Medline: 21169166]
66. Jahedi F, Henman PWF, Ryan JC. Personalization in digital psychological interventions for young adults. Int J Human-Computer Interact 2022:1-11. [doi: 10.1080/10447318.2022.2158261]

67. Ebert DD, Van Daele T, Nordgreen T, Karekla M, Compare A, Zarbo C, et al. Internet and mobile-based psychotherapy: applications, efficacy and potential for improving mental health care in Europe. Eur Psychol 2018;23(2):167-187 [FREE Full text] [doi: 10.1027/1016-9040/a000318]
68. Padilla-Ruiz M, Morales-Suárez-Varela M, Rivas-Ruiz F, Alcaide J, Varela-Moreno E, Zarcos-Pedrinaci I, et al. Influence of diagnostic delay on survival rates for patients with colorectal cancer. Int J Environ Res Public Health 2022;19(6):3626 [FREE Full text] [doi: 10.3390/ijerph19063626] [Medline: 35329320]
69. Markowitz SM, Gonzalez JS, Wilkinson JL, Safren SA. A review of treating depression in diabetes: emerging findings. Psychosomatics 2011;52(1):1-18 [FREE Full text] [doi: 10.1016/j.psym.2010.11.007] [Medline: 21300190]
70. Petrak F, Baumeister H, Skinner TC, Brown A, Holt RIG. Depression and diabetes: treatment and health-care delivery. Lancet Diabetes Endocrinol 2015;3(6):472-485. [doi: 10.1016/S2213-8587(15)00045-5] [Medline: 25995125]
71. van der Feltz-Cornelis CM, Nuyen J, Stoop C, Chan J, Jacobson AM, Katon W, et al. Effect of interventions for major depressive disorder and significant depressive symptoms in patients with diabetes mellitus: a systematic review and meta-analysis. Gen Hosp Psychiatry 2010;32(4):380-395. [doi: 10.1016/j.genhosppsych.2010.03.011] [Medline: 20633742]

Abbreviations

CBT: cognitive behavioral therapy
CEA: cost-effectiveness analysis
CG: control group
CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
CUA: cost-utility analysis
DSM-5: Diagnostic and Statistical Manual of Mental Disorders, fifth edition
HbA1c: glycosylated hemoglobin
ICER: incremental cost-effectiveness ratio
ICUR: incremental cost-utility ratio
IG: intervention group
PC: primary care
PHQ-9: Patient Health Questionnaire-9
QALY: quality-adjusted life year
SF-12: 12-Item Short Form Survey
TAU: treatment as usual
T2DM: type 2 diabetes mellitus

Edited by L Buis; submitted 14.12.23; peer-reviewed by S Mueller-Mielitz, D Singh, TAR Sure; comments to author 24.01.24; revised version received 04.03.24; accepted 14.03.24; published 16.05.24.

Please cite as:
Varela-Moreno E, Anarte-Ortiz MT, Jodar-Sanchez F, Garcia-Palacios A, Monreal-Bartolomé A, Gili M, García-Campayo J, Mayoral-Cleries F
Economic Evaluation of a Web Application Implemented in Primary Care for the Treatment of Depression in Patients With Type 2 Diabetes Mellitus: Multicenter Randomized Controlled Trial
JMIR Mhealth Uhealth 2024;12:e55483
URL: https://mhealth.jmir.org/2024/1/e55483
doi:10.2196/55483
PMID:38754101

Corrigenda and Addenda

Correction: Mobile Apps for COVID-19 Detection and Diagnosis for Future Pandemic Control: Multidimensional Systematic Review

Mehdi Gheisari[1,2], PhD; Mustafa Ghaderzadeh[3*], PhD; Huxiong Li[1*], PhD; Tania Taami[4], PhD; Christian Fernández-Campusano[5], PhD; Hamidreza Sadeghsalehi[6], PhD; Aaqif Afzaal Abbasi[7], PhD

[1]Institute of Artificial Intelligence, Shaoxing University, Shaoxing, China

[2]Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, India

[3]School of Nursing and Health Sciences of Boukan, Urmia University of Medical Sciences, Urmia, Iran

[4]Florida State University, Tallahassee, FL, United States

[5]Departamento de Ingeniería Eléctrica, Facultad de Ingeniería, Universidad de Santiago de Chile, Santiago, Chile

[6]Department of Neuroscience, Faculty of Advanced Technologies in Medicine, Tehran, Iran

[7]Department of Earth and Marine Sciences, University of Palermo, Palermo, Italy

[*]these authors contributed equally

Corresponding Author:
Mustafa Ghaderzadeh, PhD
School of Nursing and Health Sciences of Boukan
Urmia University of Medical Sciences
Kurdistan Blv Boukan
Urmia, 5951715161
Iran
Phone: 98 9129378390
Email: Mustaf.ghaderzadeh@sbmu.ac.ir

Related Article:

Correction of: https://mhealth.jmir.org/2024/1/e44406

(JMIR Mhealth Uhealth 2024;12:e58810) doi:10.2196/58810

In “Mobile Apps for COVID-19 Detection and Diagnosis for Future Pandemic Control: Multidimensional Systematic Review” ([JMIR Mhealth Uhealth 2024;12:e44406]) the authors noted one error.

In the original manuscript, the second affiliation for Mehdi Gheisari appeared as follows:

> *Department of Cognitive Computing, Institute of Computer Science and Engineering, Saveetha School of Engineering, Chennai, India*

This has been corrected to:

> *Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, India*

The correction will appear in the online version of the paper on the JMIR Publications website on April 8, 2024, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 25.03.24; this is a non–peer-reviewed article;accepted 28.03.24; published 08.04.24.

Please cite as:
Gheisari M, Ghaderzadeh M, Li H, Taami T, Fernández-Campusano C, Sadeghsalehi H, Afzaal Abbasi A
Correction: Mobile Apps for COVID-19 Detection and Diagnosis for Future Pandemic Control: Multidimensional Systematic Review
JMIR Mhealth Uhealth 2024;12:e58810
URL: https://mhealth.jmir.org/2024/1/e58810
doi:10.2196/58810
PMID:38588531

Corrigenda and Addenda

Correction: The Effect of an mHealth Self-Monitoring Intervention (MI-BP) on Blood Pressure Among Black Individuals With Uncontrolled Hypertension: Randomized Controlled Trial

Lorraine R Buis[1], PhD; Junhan Kim[2], MS; Ananda Sen[1,3], PhD; Dongru Chen[1], MS; Katee Dawood[4], MBA; Reema Kadri[1], MLiS; Rachelle Muladore[4], BS; Melissa Plegue[5], MA; Caroline R Richardson[1,6], MD; Zora Djuric[1], PhD; Candace McNaughton[7], MD; David Hutton[8], PhD; Lionel P Robert[2], PhD; Sun Young Park[2], PhD; Phillip Levy[4], MPH, MD

[1]Department of Family Medicine, University of Michigan, Ann Arbor, MI, United States
[2]School of Information, University of Michigan, Ann Arbor, MI, United States
[3]Department of Biostatistics, University of Michigan, Ann Arbor, MI, United States
[4]Department of Emergency Medicine and Integrative Biosciences Center, Wayne State University, Detroit, MI, United States
[5]Department of Pediatrics, University of Michigan, Ann Arbor, MI, United States
[6]Department of Family Medicine, Brown University, Providence, RI, United States
[7]Department of Medicine, Sunnybrook Research Institute, University of Toronto, Toronto, ON, Canada
[8]School of Public Health, University of Michigan, Ann Arbor, MI, United States

Corresponding Author:
Lorraine R Buis, PhD
Department of Family Medicine
University of Michigan
1018 Fuller St.
Ann Arbor, MI, 48104
United States
Phone: 1 734 998 7120
Fax: 1 734 998 7345
Email: buisl@umich.edu

Related Article:

Correction of: https://mhealth.jmir.org/2024/1/e57863

(JMIR Mhealth Uhealth 2024;12:e64632) doi:10.2196/64632

In "The Effect of an mHealth Self-Monitoring Intervention (MI-BP) on Blood Pressure Among Black Individuals With Uncontrolled Hypertension: Randomized Controlled Trial" (JMIR Mhealth Uhealth 2024;12:e57863) the authors noted an incorrect refernce [23]:

> *Marcus BH, Selby VC, Niaura RS, Rossi JS. Self-efficacy and the stages of exercise behavior change. Res Q Exerc Sport. Mar 1992;63(1):60-66.*

The reference has been revised to:

> *Sallis J, Pinski R, Grossman R, Patterson T, Nader P. The development of self-efficacy scales for healthrelated diet and exercise behaviors. Health Educ Res. 1988;3(3):283-292.*

The correction will appear in the online version of the paper on the JMIR Publications website on August 6, 2024, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Reference

23. Sallis J, Pinski R, Grossman R, Patterson T, Nader P. The development of self-efficacy scales for healthrelated diet and exercise behaviors. Health Educ Res 1988;3(3):283-292. [doi: 10.1093/her/3.3.283]

Submitted 22.07.24; this is a non–peer-reviewed article;accepted 22.07.24; published 06.08.24.

Please cite as:
Buis LR, Kim J, Sen A, Chen D, Dawood K, Kadri R, Muladore R, Plegue M, Richardson CR, Djuric Z, McNaughton C, Hutton D, Robert LP, Park SY, Levy P
Correction: The Effect of an mHealth Self-Monitoring Intervention (MI-BP) on Blood Pressure Among Black Individuals With Uncontrolled Hypertension: Randomized Controlled Trial
JMIR Mhealth Uhealth 2024;12:e64632
URL: https://mhealth.jmir.org/2024/1/e64632
doi:10.2196/64632
PMID:

Evaluating the Sensitivity of Wearable Devices in Posttranscatheter Aortic Valve Implantation Functional Assessment

Jinghui An*, MD; Fengwu Shi*, MM; Huajun Wang*, MM; Hang Zhang, MM; Su Liu, MD

Department of Cardiac Surgery, the Second Hospital of Hebei Medical University, No 215, Heping West Road, Shijiazhuang, China

*these authors contributed equally

Corresponding Author:
Su Liu, MD
Department of Cardiac Surgery, the Second Hospital of Hebei Medical University, , No 215, Heping West Road, Shijiazhuang, , China

Related Article:

https://mhealth.jmir.org/2024/1/e53964

(JMIR Mhealth Uhealth 2024;12:e65277) doi:10.2196/65277

KEYWORDS

aortic valve; implantation functional; wearable devices

I read with great interest the recent article by Eerdekens et al [1] titled "Cardiac Health Assessment Using a Wearable Device Before and After Transcatheter Aortic Valve Implantation: Prospective Study," published in *JMIR mHealth and uHealth*. The study presented an innovative approach by using a wearable device to assess cardiac health outcomes before and after transcatheter aortic valve implantation (TAVI), addressing an important gap in the objective evaluation of functional improvement post TAVI.

While the findings of the study were significant, particularly the introduction of the Cardiac Energy Expenditure Slope (CEES) as a potential metric for assessing cardiovascular efficiency, there were some critical considerations that merited discussion.

First, the study's conclusion that wearable device parameters, such as step count and total activity time, did not significantly change post TAVI, raised questions about the sensitivity of these devices in capturing subtle improvements in daily activity. It is well established that older populations, especially those undergoing TAVI, may not exhibit dramatic changes in physical activity due to a combination of frailty, preexisting comorbidities, and lifestyle factors [2-5]. However, the lack of significant change in these parameters might also reflect limitations in the wearable device's ability to capture variations in physical activity that are clinically meaningful but subtle. For instance, improvements in quality of life and functional capacity may have manifested in ways that were not fully captured by metrics like step count alone.

Second, the study's reliance on a 3-month follow-up period to assess post-TAVI outcomes could be seen as a limitation. While the authors argued that most patients reach full capacity by this time, it is possible that some patients might show delayed improvements in physical activity and cardiovascular efficiency. Extending the follow-up period to 6 months or even a year could have provided a more comprehensive view of the long-term impact of TAVI on patient activity levels and cardiac health.

Additionally, the introduction of CEES as a novel metric was intriguing, yet its application and utility needed further validation in larger, diverse cohorts. The metric's correlation with traditional measures of functional improvement, such as the 6-minute walking test, and its predictive value in long-term outcomes post TAVI, should be explored in future studies. This would help establish CEES as a reliable tool in both clinical practice and research settings.

In conclusion, while the study by Eerdekens et al [1] contributed valuable insights into the use of wearable devices for cardiac health assessment, it also highlighted the need for further research to optimize these tools for older, comorbid populations. Continued exploration into novel metrics like CEES and longer follow-up periods could enhance our understanding of post-TAVI recovery and guide personalized treatment strategies.

Conflicts of Interest

None declared.

Editorial Notice

The corresponding author of "Cardiac Health Assessment Using a Wearable Device Before and After Transcatheter Aortic Valve Implantation: Prospective Study" declined to respond to this letter.

References

1. Eerdekens R, Zelis J, Ter Horst H, et al. Cardiac health assessment using a wearable device before and after transcatheter aortic valve implantation: prospective study. JMIR mHealth uHealth 2024 Jun 3;12:e53964. [doi: 10.2196/53964] [Medline: 38832585]
2. Tsevat J, Dawson NV. Health values of hospitalized patients 80 years or older. JAMA 1998 Feb 4;279(5):371. [doi: 10.1001/jama.279.5.371]
3. Kim CA, Rasania SP, Afilalo J, Popma JJ, Lipsitz LA, Kim DH. Functional status and quality of life after transcatheter aortic valve replacement: a systematic review. Ann Intern Med 2014 Feb 18;160(4):243-254. [doi: 10.7326/M13-1316] [Medline: 24727842]
4. Afilalo J, Lauck S, Kim DH, et al. Frailty in older adults undergoing aortic valve replacement. J Am Coll Cardiol 2017 Aug;70(6):689-700. [doi: 10.1016/j.jacc.2017.06.024] [Medline: 28693934]
5. Green P, Arnold SV, Cohen DJ, et al. Relation of frailty to outcomes after transcatheter aortic valve replacement (from the PARTNER trial). Am J Cardiol 2015 Jul 15;116(2):264-269. [doi: 10.1016/j.amjcard.2015.03.061] [Medline: 25963221]

Abbreviations

CEES: Cardiac Energy Expenditure Slope
TAVI: transcatheter aortic valve implantation

Edited by K Prairie, TDA Cardoso; submitted 11.08.24; this is a non–peer-reviewed article;accepted 20.08.24; published 08.11.24.

Please cite as:
An J, Shi F, Wang H, Zhang H, Liu S
Evaluating the Sensitivity of Wearable Devices in Posttranscatheter Aortic Valve Implantation Functional Assessment
JMIR Mhealth Uhealth 2024;12:e65277
URL: https://mhealth.jmir.org/2024/1/e65277
doi:10.2196/65277

Correction: Effects of a Planned Web-Based Educational Intervention Based on the Health Belief Model for Patients With Ischemic Stroke in Promoting Secondary Prevention During the COVID-19 Lockdown in China: Quasi-Experimental Study

Zhuo Liu[1*], PhD; Xin Sun[1*], PhD; Zhen-ni Guo[1,2], PhD; Ye Sun[1], BD; Yi Yang[1], MD, PhD; Xiuli Yan[1], BD

[1]Stroke Center, Department of Neurology, The First Hospital of Jilin University, Changchun, China

[2]Neuroscience Research Center, Department of Neurology, The First Hospital of Jilin University, Changchun, China

[*]these authors contributed equally

Corresponding Author:
Xiuli Yan, BD
Stroke Center, Department of Neurology, The First Hospital of Jilin University, , Changchun, , China

Related Article:

Correction of: https://mhealth.jmir.org/2024/1/e44463

(JMIR Mhealth Uhealth 2024;12:e60953) doi:10.2196/60953

In "Effects of a Planned Web-Based Educational Intervention Based on the Health Belief Model for Patients With Ischemic Stroke in Promoting Secondary Prevention During the COVID-19 Lockdown in China: Quasi-Experimental Study" (JMIR Mhealth Uhealth 2024;12:e44463) the authors noted one error.

In the "Ethical Considerations" section, the following sentence:

> *The study protocol was reviewed and approved by the ethical committee of the First Hospital of Jilin University (20K056-001) and was registered in the Chinese Clinical Trial Registry (ChiCTR2000040804).*

Has been changed to read as follows:

> *The study protocol was reviewed and approved by the ethical committee of the First Hospital of Jilin University (20K055-001) and was registered in the Chinese Clinical Trial Registry (ChiCTR2000040804).*

The correction will appear in the online version of the paper on the JMIR Publications website on June 28, 2024, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Conflicts of Interest

None declared.

Submitted 27.05.24; this is a non–peer-reviewed article;accepted 12.06.24; published 28.06.24.

Please cite as:
Liu Z, Sun X, Guo ZN, Sun Y, Yang Y, Yan X
Correction: Effects of a Planned Web-Based Educational Intervention Based on the Health Belief Model for Patients With Ischemic Stroke in Promoting Secondary Prevention During the COVID-19 Lockdown in China: Quasi-Experimental Study
JMIR Mhealth Uhealth 2024;12:e60953
URL: https://mhealth.jmir.org/2024/1/e60953
doi:10.2196/60953

Correction: The Importance of Activating Factors in Physical Activity Interventions for Older Adults Using Information and Communication Technologies: Systematic Review

Ellen Bentlage, MSc; John Jnr Nyamadi, BSc; Rosemary Dubbeldam, PhD

Department of Movement Science, Institute of Sport and Exercise Sciences, University of Münster, Münster, Germany

Corresponding Author:
Ellen Bentlage, MSc
Department of Movement Science, Institute of Sport and Exercise Sciences, University of Münster, , Münster, , Germany

Related Article:

https://mhealth.jmir.org/2023/1/e42968

(JMIR Mhealth Uhealth 2024;12:e55486) doi:10.2196/55486

In "The Importance of Activating Factors in Physical Activity Interventions for Older Adults Using Information and Communication Technologies: Systematic Review" (JMIR Mhealth Uhealth 2023; 11(1): e42968) the authors noted some errors].

In the "Conclusions" section of the Abstract, the following phrase:

> *So far, only a limited number of available BCTs (21/99, 21%) have been integrated.*

Has been changed to:

> *So far, only a limited number of available BCTs (21/102, 21%) have been integrated.*

In the "Introduction" subsection "Activation Using Information and Communication Technology," the following sentence:

> *Dugas and colleagues [12] added 2 more categories (ie, gamification and personalization).*

Has been changed to:

> *Dugas and colleagues [12] added 2 more categories (ie, personalization and gamification), including 9 BCTs in total.*

In the "Methods" subsection "Selection of Studies and Data Extraction," the line:

> *Additionally, the two categories suggested by Dugas and colleagues [12]—gamification and personalization—were incorporated as categories 17 and 18.*

Has been changed to:

> *Additionally, the two categories suggested by Dugas and colleagues [12]—personalization and gamification—were incorporated as categories 17 and 18.*

In Table 1, under the "Skills (ability)" column, the following text:

- *Category 7: Repetition and substitution*

Has been changed to:

- *Category 8: Repetition and substitution*

Elsewhere in Table 1, under the "Knowledge (awareness)" column, the following content:

- *Category 5: Natural consequences*
- *Category 9: Comparison of outcomes*
- *Category 11: Regulation*
- *Category 14: Scheduled consequences*
- *Category 16: Covert learning*

Has been changed to:

- *Category 5: Natural consequences*
- *Category 9: Comparison of outcomes*
- *Category 11: Regulation*

In Table 1, under the "Motivation (triggers)" column, the following text:

- *Category 1: Goals and planning*
- *Category 2: Feedback and monitoring*
- *Category 3: Social support*
- *Category 8: Associations*
- *Category 10: Reward and thread*
- *Category 12: Antecedents*
- *Category 17: Gamefication*
- *Category 18: Personalization*

Has been changed to:

- *Category 1: Goals and planning*
- *Category 2: Feedback and monitoring*
- *Category 3: Social support*
- *Category 7: Associations*
- *Category 10: Reward and thread*
- *Category 12: Antecedents*
- *Category 14: Scheduled consequences*
- *Category 16: Covert learning*
- *Category 17: Personalization*
- *Category 18: Gamification*

In the "Results" subsection "Delivering Activation Factors", the following sentence:

> *Of the 22 available BCTs [11,12] that target skills, the aforementioned 5 (23%) were used among all included articles.*

Has been changed to:

> *Of the 23 available BCTs [11,12] that target skills, the aforementioned 6 (23%) were used among all included articles.*

Further in the same section, the following paragraphs:

> *Others were prompts and cues (4 intervention groups), social comparison (1 intervention group), and problem-solving discussions for finding ways to overcome barriers (1 intervention group). Of the 64 available BCTs [11,12] that target motivation, 12 (19%; inclusive of the 7 aforementioned BCTs) were used among all included articles.*
>
> *Altogether, of the 99 potential BCTs [11,12], 21 (21%) were integrated in all articles.*

Have been changed to:

> *Others were prompts and cues (4 intervention groups), adjustment of intervention content to the performance (4 intervention groups), action planning (2 intervention groups) and problem-solving discussions for finding ways to overcome barriers (1 intervention group). Of the 66 available BCTs [11,12] that target motivation, 11 were used among all included articles.*
>
> *Altogether, of the 102 potential BCTs [11,12], 21 (21%) were integrated in all articles.*

Within Textbox 1, the "Skills" subheading which previously appeared as:

Skills

- *Promoted in 19 of 20 interventions*
- *Five BCTs were used; main behavior change techniques (BCTs) were the instruction of optimal behavior performance and the demonstration of behavior*

Has been changed to:

Skills

- *Promoted in 19 of 20 interventions*
- *Six BCTs were used; main behavior change techniques (BCTs) were the instruction of optimal behavior performance and the demonstration of behavior*

Further within Textbox 1, the "Motivation" subheading previously read as:

Motivation

- *Promoted in 17 of 20 interventions*
- *Twelve BCTs were used; main BCTs were self-monitoring, feedback on behavior, social support, and goal setting*

And will now appear as:

Motivation

- *Promoted in 17 of 20 interventions*
- *Eleven BCTs were used; main BCTs were self-monitoring, feedback on behavior, social support, and goal setting*

Within the "Discussion" subsection "Delivering Activation Factors," the sentence:

> *To target participants' motivation, 12 BCTs, such as self-monitoring, feedback on behavior, social support, and goal setting, were used.*

Was changed to:

> *To target participants' motivation, 11 BCTs, such as self-monitoring, feedback on behavior, social support, adjusting intervention content to the performance and goal setting, were used.*

The "Discussion" subsection "Conclusions" includes the claim that:

> *Although a broad variety of BCTs were used in the articles, they were limited to about 21% (21/99) of available BCTs.*

This has been adjusted to appear as:

> *Although a broad variety of BCTs were used in the articles, they were limited to about 21% (21/102) of available BCTs.*

In Multimedia Appendix 1, the rows "Li, 2020," "Mansson, 2020," "Rowley, 2019," and "Van Dyck, 2016" have been amended from:

> *18 Personalization*

And now appear as:

> *17.4 Adjusting intervention content to performance*

Finally, Multimedia Appendix 2 has been changed to match the main manuscript in the ways listed below.

The "Skills" column has been changed from:

Skills (30 BCTs)

5 BCTs

To read as:

Skills (23 BCTs)

6 BCTs

The "Knowledge" column has been changed from:

Knowledge (27 BCTs)

4 BCTs

To read as:

Knowledge (13 BCTs)

4 BCTs

The "Motivation" column has been changed from:

Motivation (50 BCTs)

12 BCTs

To appear as:

Motivation (66 BCTs)

11 BCTs

The correction will appear in the online version of the paper on the JMIR Publications website on February 1, 2024, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Multimedia Appendix 1
Study characteristics.
[DOCX File, 67 KB - mhealth_v12i1e55486_app1.docx]

Multimedia Appendix 2
Behaviour change techniques overview corrected.
[DOCX File, 21 KB - mhealth_v12i1e55486_app2.docx]

Submitted 14.12.23; this is a non–peer-reviewed article;accepted 14.12.23; published 01.02.24.

Please cite as:
Bentlage E, Nyamadi JJ, Dubbeldam R
Correction: The Importance of Activating Factors in Physical Activity Interventions for Older Adults Using Information and Communication Technologies: Systematic Review
JMIR Mhealth Uhealth 2024;12:e55486
URL: https://mhealth.jmir.org/2024/1/e55486
doi:10.2196/55486

Corrigenda and Addenda

Correction: Mobile App Intervention of a Randomized Controlled Trial for Patients With Obesity and Those Who Are Overweight in General Practice: User Engagement Analysis Quantitative Study

Vera Helen Buss[1*], PhD; Margo Barr[1*], PhD; Sharon M Parker[1*], MPH; Alamgir Kabir[1*], PhD; Annie Y S Lau[2*], PhD; Siaw-Teng Liaw[3*], PhD; Nigel Stocks[4*], MD; Mark F Harris[1*], MD

[1]Centre for Primary Health Care and Equity, University of New South Wales, Sydney, Australia

[2]Australian Institute of Health Innovation, Macquarie University, Sydney, Australia

[3]School of Population Health, University of New South Wales, Sydney, Australia

[4]Adelaide Medical School, University of Adelaide, Adelaide, Australia

[*]all authors contributed equally

Corresponding Author:
Margo Barr, PhD
Centre for Primary Health Care and Equity
University of New South Wales
AGSM Building
High Street, Kensington Campus
Sydney, 2052
Australia
Phone: 61 290656041
Email: Margo.barr@unsw.edu.au

Related Article:

Correction of: https://mhealth.jmir.org/2024/1/e45942

(JMIR Mhealth Uhealth 2024;12:e58507) doi:10.2196/58507

In "Mobile App Intervention of a Randomized Controlled Trial for Patients With Obesity and Those Who Are Overweight in General Practice: User Engagement Analysis Quantitative Study" (JMIR Mhealth Uhealth 2024;12:e45942), the authors noted two errors. The following corrections have been made:

Under "Methods", the third paragraph of the "Outcome Measures" subsection has been changed from:

> *Specifically, the Health Literacy Questionnaire domain 8 questions were the following [11]: please indicate how difficult or easy the following tasks are for you now: (1) find information about health problems; (2) find health information from several different places; (3) get information about health so you are up to date with the best information; (4) get health information in words you understand; and (5) get health information by yourself. There is a 5-point response option scale for each question (cannot do or always difficult, usually difficult, sometimes difficult, usually easy or always easy). The scores are reported as averages for the domain (with a range between 1 and 5) with high scores representing higher health literacy.*

And now reads as follows:

> *The Health Literacy Questionnaire domain 8 questions were used [11]:(1) find information about health problems; (2) find health information from several …. ; (3) get information about health so you…; (4) get health information in words you…; and (5) get health information by yourself. There is a 5-point response option scale for each question (cannot do or always difficult, usually difficult, sometimes difficult, usually easy or always easy). The scores are reported as averages for the domain (with a range between 1 and 5) with high scores representing higher health literacy.*

Additionally, in Table 1 the definition of the "Health literary" row has been changed from:

> *Health literacy, specifically the self-reported ability to find good quality health information, according to domain 8 of the Health Literacy Questionnaire [9], at baseline and 6-month follow-up.*

To reference a different source, this now appears as:

> *Health literacy, specifically the self-reported ability to find good quality health information, according to domain 8 of the Health Literacy Questionnaire [11], at baseline and 6-month follow-up.*

The correction will appear in the online version of the paper on the JMIR Publications website on April 2, 2024, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 18.03.24; this is a non–peer-reviewed article;accepted 22.03.24; published 02.04.24.

Please cite as:
Buss VH, Barr M, Parker SM, Kabir A, Lau AYS, Liaw ST, Stocks N, Harris MF
Correction: Mobile App Intervention of a Randomized Controlled Trial for Patients With Obesity and Those Who Are Overweight in General Practice: User Engagement Analysis Quantitative Study
JMIR Mhealth Uhealth 2024;12:e58507
URL: https://mhealth.jmir.org/2024/1/e58507
doi:10.2196/58507
PMID:38564771

Corrigenda and Addenda

Table Correction: A Smartphone Food Record App Developed for the Dutch National Food Consumption Survey: Relative Validity Study

Marga Ocké[1*], PhD; Ceciel Simone Dinnissen[1*], MSc; Coline van den Bogaard[1], BSc; Marja Beukers[1], BSc; José Drijvers[1], BSc; Eline Sanderman-Nawijn[1], MSc; Caroline van Rossum[1], PhD; Ido Toxopeus[1], PhD

National Institute for Public Health and the Environment, Bilthoven, Netherlands

[*]these authors contributed equally

Corresponding Author:
Ceciel Simone Dinnissen, MSc
National Institute for Public Health and the Environment
Antonie van Leeuwenhoeklaan 9
Bilthoven, 3721 MA
Netherlands
Phone: 31 088 689 8989
Email: ceciel.dinnissen@rivm.nl

Related Article:

Correction of: https://mhealth.jmir.org/2024/1/e50196

(JMIR Mhealth Uhealth 2024;12:e59530) doi:10.2196/59530

In "A Smartphone Food Record App Developed for the Dutch National Food Consumption Survey: Relative Validity Study" (JMIR Mhealth Uhealth 2024;12:e50196) the authors noted the following errors in Tables 2 and 5:

Due to an unintended line alignment issue, all values in the *DitEetIk! app food record* and *GloboDiet 24-hour dietary recall (g/d)* columns of Table 2 were incorrect for the following rows: *Meat*, *Eggs*, *Nuts*, *Milk (products)*, *Cheese*, *Bread*, *Cereal products*, *Potatoes*, *Drinks*, *Sandwich spreads*, *Snacks*, *Sauces*, and *Other.*

Table 2 has been corrected as follows:

Table 2. The mean, SD, median, and IQR of consumption of food groups[a] as assessed using the DitEetIk! app and 24-hour dietary recalls for the same day and their correlation for the 211 participants with plausible energy intakes.

Food group	DitEetIk! app food record (g/d)		GloboDiet 24-hour dietary recall (g/d)		Wilcoxon signed rank test *P* value[b]	Spearman correlation coefficient
	Values, mean (SD)	Values, median (IQR)	Values, mean (SD)	Values, median (IQR)		
Vegetables	163 (200)	117 (31-226)	160 (144)	130 (50-240)	.13	0.76
Fruit	128 (186)	83 (0-188)	140 (146)	130 (0-217)	.005	0.79
Added fats	16 (16)	12 (3-24)	19 (15)	17 (6-29)	.001	0.54
Meat	103 (112)	73 (23-135)	92 (83)	75 (33-120)	.10	0.70
Eggs	17 (37)	0 (0-0)	17 (34)	0 (0-13)	.44	0.76
Nuts	15 (30)	0 (0-20)	15 (30)	0 (0-20)	.73	0.84
Milk (products)	264 (263)	219 (16-391)	288 (248)	252 (80-423)	.02	0.80
Cheese	33 (36)	30 (0-56)	39 (44)	31 (0-62)	.006	0.76
Bread	146 (113)	126 (70-199)	138 (88)	132 (70-180)	.95	0.85
Cereal products	67 (133)	6 (0-88)	74 (106)	20 (0-119)	.01	0.80
Potatoes	72 (119)	0 (0-128)	66 (104)	0 (0-120)	.96	0.88
Drinks	1888 (956)	1836 (1275-2311)	2097 (889)	1963 (1582-2539)	<.001	0.68
Sandwich spreads	15 (27)	0 (0-20)	12 (23)	0 (0-15)	.05	0.88
Snacks	91 (119)	52 (15-118)	83 (89)	56 (14-126)	.41	0.88
Sauces	21 (37)	2 (0-26)	33 (38)	22 (0-57)	<.001	0.60
Other	13 (52)	0 (0-10)	5 (12)	0 (0-5)	<.001	0.50

[a]Food groups are Wheel of Five food groups—main groups [23]. The food groups *Fish*, *Legumes*, and *Soups* were excluded as the 75th percentile was 0 for both methods. Table 3 provides more information on these food groups.

[b]Wilcoxon signed rank test (normal approximation) of the differences between intake assessed using the DitEetIk! app and the GloboDiet 24-hour dietary recalls for the same day.

Similarly, due to an unintended line alignment issue, all values in the *DitEetIk! app evaluation study (g/d)* and *DNFCS 2019-2021 (g/d)* columns of Table 5 were incorrect for the following rows: *Meat*, *Eggs*, *Nuts*, *Milk and milk products*, *Cheese*, *Bread*, *Cereal products*, *Potatoes*, *Drinks*, *Sandwich spreads*, *Soups*, *Snacks*, *Sauces*, and *Other*.

Table 5 has been corrected as follows:

Table 5. Comparison of consumption of food groups assessed using the GloboDiet 24-hour dietary recalls in the DitEetIk! app evaluation study and the first interview in the Dutch National Food Consumption Survey (DNFCS) 2019 to 2021 for a matched group of participants (n=211).

Food group[a]	DitEetIk! app evaluation study (g/d)		DNFCS 2019-2021 (g/d)		*P* value[b]
	Values, mean (SD)	Values, median (IQR)	Values, mean (SD)	Values, median (IQR)	
Vegetables	160 (144)	130 (50-240)	155 (140)	125 (53-217)	>.99
Fruit	140 (146)	130 (0-217)	124 (135)	108 (0-195)	.51
Added fats	19 (15)	17 (6-29)	22 (20)	18 (8-32)	.006
Fish	17 (57)	0 (0-0)	15 (44)	0 (0-0)	.91
Legumes	4 (20)	0 (0-0)	8 (36)	0 (0-0)	.09
Meat	92 (83)	75 (33-120)	88 (80)	77 (29-116)	.76
Eggs	17 (34)	0 (0-13)	16 (32)	0 (0-13)	.85
Nuts	15 (30)	0 (0-20)	19 (45)	0 (0-22)	.24
Milk and milk products	288 (248)	252 (80-423)	332 (267)	282 (150-484)	.03
Cheese	39 (44)	31 (0-62)	38 (39)	30 (0-62)	.85
Bread	138 (88)	132 (70-180)	117 (80)	105 (60-169)	.03
Cereal products	74 (106)	20 (0-119)	79 (108)	30 (0-122)	.57
Potatoes	66 (104)	0 (0-120)	69 (93)	0 (0-140)	.26
Drinks	2097 (889)	1963 (1582-2539)	2132 (914)	1958 (1468-2608)	.63
Sandwich spreads	12 (23)	0 (0-15)	18 (29)	0 (0-23)	.03
Soups	12 (46)	0 (0-0)	17 (66)	0 (0-0)	.57
Snacks	83 (89)	56 (14-126)	71 (79)	41 (10-114)	.24
Sauces	33 (38)	22 (0-57)	29 (44)	11 (0-36)	.14
Other	5 (12)	0 (0-5)	6 (16)	0 (0-5)	.75

[a]Food groups are Wheel of Five food groups [23].

[b]Wilcoxon signed rank test (normal approximation) of the differences between intake assessed using GloboDiet 24-hour dietary recalls in the DitEetIk! app evaluation study and the first interview with adults in the DNFCS 2019 to 2021.

The correction will appear in the online version of the paper on the JMIR Publications website on April 26, 2024, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 15.04.24; this is a non–peer-reviewed article;accepted 19.04.24; published 26.04.24.

Please cite as:
Ocké M, Dinnissen CS, van den Bogaard C, Beukers M, Drijvers J, Sanderman-Nawijn E, van Rossum C, Toxopeus I
Table Correction: A Smartphone Food Record App Developed for the Dutch National Food Consumption Survey: Relative Validity Study
JMIR Mhealth Uhealth 2024;12:e59530
URL: https://mhealth.jmir.org/2024/1/e59530
doi:10.2196/59530
PMID:38669673

Original Paper

Digital Phenotyping of Geriatric Depression Using a Community-Based Digital Mental Health Monitoring Platform for Socially Vulnerable Older Adults and Their Community Caregivers: 6-Week Living Lab Single-Arm Pilot Study

Sunmi Song[1,2,3], PhD; YoungBin Seo[4,5], MS; SeoYeon Hwang[3,5], BS; Hae-Young Kim[3,4,5,6], PhD, DDS; Junesun Kim[1,2,3,5], PT, PhD

[1]Department of Health and Environmental Science, Undergraduate School, Korea University, Seoul, Republic of Korea

[2]Department of Physical Therapy, College of Health Science, Korea University, Seoul, Republic of Korea

[3]Department of Public Health Sciences, Graduate School, Korea University, Seoul, Republic of Korea

[4]Department of Healthcare Sciences, Graduate School, Korea University, Seoul, Republic of Korea

[5]BK21FOUR: L-HOPE Program for Community-Based Total Learning Health Systems, College of Health Science, Korea University, Seoul, Republic of Korea

[6]Department of Health Policy and Management, College of Health Science, Korea University, Seoul, Republic of Korea

Corresponding Author:
Junesun Kim, PT, PhD
Department of Health and Environmental Science
Undergraduate School
Korea University
661 B-Hana Science Building
Seongbuk-gu
Seoul, 02841
Republic of Korea
Phone: 82 2 3290 5689
Email: junokim@korea.ac.kr

Abstract

Background: Despite the increasing need for digital services to support geriatric mental health, the development and implementation of digital mental health care systems for older adults have been hindered by a lack of studies involving socially vulnerable older adult users and their caregivers in natural living environments.

Objective: This study aims to determine whether digital sensing data on heart rate variability, sleep quality, and physical activity can predict same-day or next-day depressive symptoms among socially vulnerable older adults in their everyday living environments. In addition, this study tested the feasibility of a digital mental health monitoring platform designed to inform older adult users and their community caregivers about day-to-day changes in the health status of older adults.

Methods: A single-arm, nonrandomized living lab pilot study was conducted with socially vulnerable older adults (n=25), their community caregivers (n=16), and a managerial social worker over a 6-week period during and after the COVID-19 pandemic. Depressive symptoms were assessed daily using the 9-item Patient Health Questionnaire via scripted verbal conversations with a mobile chatbot. Digital biomarkers for depression, including heart rate variability, sleep, and physical activity, were measured using a wearable sensor (Fitbit Sense) that was worn continuously, except during charging times. Daily individualized feedback, using traffic signal signs, on the health status of older adult users regarding stress, sleep, physical activity, and health emergency status was displayed on a mobile app for the users and on a web application for their community caregivers. Multilevel modeling was used to examine whether the digital biomarkers predicted same-day or next-day depressive symptoms. Study staff conducted pre- and postsurveys in person at the homes of older adult users to monitor changes in depressive symptoms, sleep quality, and system usability.

Results: Among the 31 older adult participants, 25 provided data for the living lab and 24 provided data for the pre-post test analysis. The multilevel modeling results showed that increases in daily sleep fragmentation (P=.003) and sleep efficiency (P=.001) compared with one's average were associated with an increased risk of daily depressive symptoms in older adults. The pre-post

test results indicated improvements in depressive symptoms (*P*=.048) and sleep quality (*P*=.02), but not in the system usability (*P*=.18).

Conclusions: The findings suggest that wearable sensors assessing sleep quality may be utilized to predict daily fluctuations in depressive symptoms among socially vulnerable older adults. The results also imply that receiving individualized health feedback and sharing it with community caregivers may help improve the mental health of older adults. However, additional in-person training may be necessary to enhance usability.

Trial Registration: ClinicalTrials.gov NCT06270121; https://clinicaltrials.gov/study/NCT06270121

(JMIR Mhealth Uhealth 2024;12:e55842) doi:10.2196/55842

KEYWORDS

depression; monitoring system; IoT; AI; wearable device; digital mental health phenotyping; living lab; senior care; Internet of Things; artificial intelligence

Introduction

Over the past 2 decades, there has been a notable increase in geriatric depression and other psychiatric disorders, coinciding with a global rise in life expectancy and population aging [1-4]. Since 2009, South Korea has had the highest suicide rate for older adults among the Organisation for Economic Co-operation and Development (OECD) countries. This is attributed to elevated levels of geriatric depression, economic poverty, and social isolation resulting from the rapid nuclearization of the family [5]. With the rapid aging of the global population, caring for older family members with mental disorders has become an overwhelming task for younger generations [6]. Despite the increasing role of community services, there are significant budget shortages in local communities and government health departments, as well as a lack of skilled geriatric labor to meet the needs of these older adults with mental disorders.

The digitalization of mental health screening and intervention is expected to provide innovative solutions to the challenges in mental health care for older adults. For example, digital phenotyping can facilitate the early detection of depression and help reduce the high rate of undiagnosed depression (50%) among older adults [7]. Digital phenotyping of mental health is defined as the "moment-to-moment quantification of the individual-level human phenotype of mental health status in real-life contexts using data collected from personal digital devices" [8]. Recent studies have suggested the potential of digital phenotyping for depressive symptoms using ecological momentary assessments, including self-reports of depressive mood [9]. Beyond momentary self-reports, digital sensing technologies enable unobtrusive passive sensing of depressive symptoms through smartphone apps, wearable sensors, and the Internet of Things. The Internet of Things, a ubiquitous network of interconnected devices, facilitates seamless data collection and intelligent monitoring and management to ensure users' health and safety [10-13]. However, previous studies have primarily examined digital phenotypes of depressive symptoms in young adults or small groups of patients with depression. This has created challenges in applying digital phenotyping technologies to older adults [3].

The underrepresentation of older adults in digital mental health care research is due to several barriers hindering their participation in studies involving novel digital technologies. First, older adults often have sensory and cognitive impairments that necessitate the use of different design principles than those effective for young or middle-aged adults [14-16]. For example, older adults generally prefer displays with simple layouts and multimodal command functions, such as voice commands in addition to touch screens. Second, implementing digital health care services for older adults should involve both family and community caregivers, who often do not live with the older adult, as well as support from multiple community health care institutions [17,18]. When learning to use new technology, older adults require repeated in-person assistance and educational materials tailored to low digital literacy levels [14,15]. Digital health care services are most likely to benefit older adults when they can connect them to the necessary health care services within the community. Third, the living environments of older adults often hinder their use of digital mental health care services, as a significant proportion (40%-60%) may not have an internet connection or access to personal computers and other mobile devices [19,20].

These functional, social, and environmental barriers underscore the necessity for proof-of-concept and feasibility trials for geriatric mental health care services among older adults and community caregivers in their natural living environments, utilizing living labs [3,18]. Living labs are defined as "user-centered, open innovation ecosystems based on a systematic user co-creation approach, integrating research and innovation processes in real-life communities and settings to create sustainable impact" [21]. Living labs are essential for designing solutions tailored to the needs of older adults. Digital mental health services will not be acceptable or sustainable unless they are designed to be compatible with the cognitive and physical capacities of older adults, as well as their natural living environments, and the working environments of community caregivers. This compatibility can be ensured through living lab testing [22,23].

This study aims to address the gap in current geriatric health literature by testing the feasibility of a digital mental health monitoring platform. This platform could be integrated with existing community senior care services, aiming to prevent and detect early signs of mental health decline in socially isolated older adults. We conducted a single-arm, nonrandomized living lab pilot study involving 25 socially vulnerable older adults. These individuals received personalized daily health monitoring in their natural living environments, both during and after the

COVID-19 pandemic. Additionally, the monitoring results of the older adult participants were shared with their community caregivers (n=16) and a managerial social worker at a community senior welfare center (hereafter referred to as the "community center"). This information was utilized by caregivers during their regular in-person senior caregiving services and emergency responses. The study assessed pre- and postintervention changes in mental health indicators (eg, depressive symptoms and sleep quality) and the usability of the monitoring platform among older adults. This evaluation aimed to test both the health-enhancing effects and the usability of the platform. Additionally, beyond feasibility testing, the research investigated whether utilizing digital biomarkers of geriatric depression detected by a wearable sensor could predict daily fluctuations in depressive symptoms among older adults.

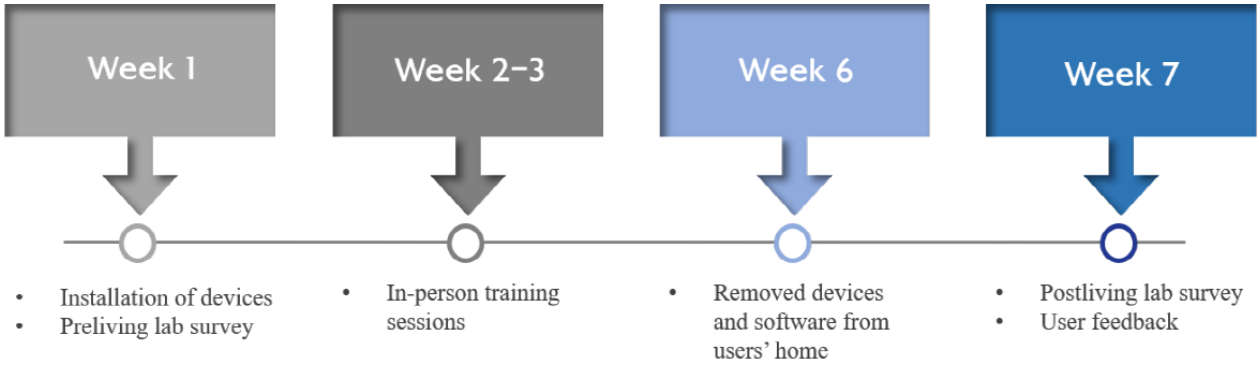
Methods

Overview

This single-arm, nonrandomized living lab pilot study was undertaken as part of a broader research endeavor aimed at developing a sustainable digital health monitoring platform. The platform is designed to be integrated with a community-based public senior care service, with the goal of enhancing the mental and physical well-being of older adults, both during and following the COVID-19 pandemic. Previously, we conducted formative research using surveys with older adults (n=99) and focused group interviews involving older adults (n=16), community caregivers (n=12), and social workers (n=3) at the community center. The objective was to identify the primary health concerns among older adults, understand the workflow of community caregiving services, and determine the necessary structures and features for the health monitoring service platform. The present digital health monitoring platform incorporates a smartphone chatbot, a smartwatch (Fitbit Sense; Google Inc.), and a motion-sensing camera (Azure Kinect SDK 1.3.0; Microsoft Corporation) installed in the home environment. This platform is designed to enhance the self-care capabilities of older adults by delivering daily updates on their mental and physical health status compared with their baseline averages established during the initial week of the living lab. Additionally, the platform aims to bolster the social support network of older adults by sharing their daily health status information and health emergency alerts with their community caregivers. These caregivers offer regular in-person senior caregiving services to older adult users through the community center. Furthermore, the information was shared with the managerial social worker at the community center to provide assistance in case of any health emergencies during and after the COVID-19 pandemic. Before the pilot study, we conducted an informal prepilot test involving study staff (n=6) to ensure the overall functionality of mobile apps and processing algorithms. The results from the motion-sensing camera, which was integrated with experimental sessions outside the living lab, are reported separately from this study [24]. Trained research assistants, along with a community caregiver, visited older adult participants in their homes to obtain informed consent and install a mobile app on the users' smartphones and digital devices. Participants were requested to complete surveys at home both before and after the living lab activities, aimed at gauging the usability of the platform and assessing their mental and physical health status. The primary randomized controlled trials featuring both intervention and comparison groups have been registered (registered with Clinicaltrials.gov; registration number NCT06270121); however, this pilot study did not include a control group.

Figure 1 illustrates the timeline of the living lab study procedure, conducted from September 2022 to August 2023. The living lab spanned 6 weeks, commencing with a 1-week adaptation period. Older adults were instructed to launch the smartphone app in the morning, which would prompt them to engage with a chatbot inquiring about their well-being and presenting 2 questions regarding daily depressive symptoms. Following interaction with the chatbot, participants could access their daily health status information, including stress levels (measured via the high frequency [HF] measure of heart rate variability [HRV]), sleep quality (indicated by total sleep time and sleep fragmentation), and physical activity (PA) levels (tracked by step count) from the previous day. This information was compared with their own average during the initial week of the living lab. Voice recordings, daily health status updates, and real-time emergency alerts of the older adult participants were transmitted to their corresponding community caregivers and the managerial social worker through the web or app interface. Throughout the duration of the living lab, participants were instructed to wear a smartwatch continuously, except during battery charging periods.

Figure 1. Timeline of the 6-week pilot living lab procedure.



Trained research assistants conducted surveys at the homes of older adults, both before and after the living lab, to assess significant changes in pilot trial outcomes, such as depressive symptoms, sleep quality, and system usability. During the

posttest survey, participants were also queried about the frequency and types of functions they utilized on the platform during the living lab period, and any additional feedback to enhance the platform for the main trials.

To facilitate participants' acclimatization to the digital monitoring app and devices, and to mitigate the risk of missing data, a member of the research staff visited participants in their homes for 2-3 additional in-person training sessions. These sessions focused on guiding participants on how to use the verbal surveys (Textbox 1) and digital devices effectively, as well as how to check their daily health status. These training sessions were conducted during the second and third weeks of the living lab. If a participant's data were missed for 3 consecutive days, study staff promptly contacted the participant via phone to emphasize the significance of responding to verbal surveys or wearing the smartwatch. Additionally, they offered assistance in resolving any technical issues hindering the older adult's participation. Furthermore, community caregivers played a vital role in assisting older adults in adapting to the platform's usage.

Textbox 1. The methods for assessing daily depressive symptoms via the mobile app chatbot.

The 9-Item Patient Health Questionnaire on depressive symptoms

- Do you feel down, depressed, or hopeless today?
- Do you feel little interest or pleasure in doing things today?
- Have you had trouble falling or staying asleep, or sleeping too much today?
- Are you feeling tired or having little energy today?
- Have you experienced poor appetite or overeating today?
- Do you experience trouble concentrating on things today?
- Are you feeling bad about yourself or that you are a failure or have let yourself or your family down?
- Are you moving or speaking so slowly that other people could have noticed?

5 items on daily greetings (recommended by community caregivers during the formative research)

- Did you sleep well last night?
- Have you eaten your meal?
- How are you feeling? Are you feeling pain in any part of your body?
- What are you planning to do today?
- Do you need to go to the hospital today?

An example daily voice survey scenario combining 1 randomly selected greeting item and 2 randomly selected items from the 9-Item Patient Health Questionnaire

Good morning, Ma'am!

[Greeting item] Did you sleep well last night? (Recording)

[PHQ-9 #1] Do you feel down, depressed, or hopeless today? (Recording)

[PHQ-9 #2] Are you moving or speaking so slowly that other people could have noticed? (Recording)

Thank you! Have a good day.

Ethics Approval

The study protocol received approval from the Institutional Review Board of Korea University (approval number KUIRB-2021-0324-02).

Recruitment

Adults older than 65 years, along with their community caregivers, were recruited from a community center in Seoul, South Korea. A meeting was convened with the community caregivers and managers at the community center to elucidate the study's objectives and procedures, solicit participation as community caregivers, and seek assistance in recruiting older adult participants. Based on the formative research findings, community caregivers can be described as predominantly middle-aged women (mean age 58.04, SD 3.17 years), with a gender composition of women only. On average, they possessed 1-2 years of work experience. Each caregiver was responsible for providing in-person caregiving services to multiple older adults, with a maximum caseload of 16 individuals. Caregivers who expressed interest in participating in the study with their older adult service recipients explained the study protocol to the older adult during their subsequent regular in-person visits. If both the older adult and their community caregiver were interested in participating, the study staff arranged a home visit to obtain informed consent from the older adult, accompanied by their caregiver.

The inclusion criteria encompassed older adults receiving in-person senior care services due to socioeconomic vulnerability, particularly those living alone with low income. Additionally, participants were required to use a Samsung

Galaxy smartphone, as the mobile app was exclusively developed for Android smartphones, which are widely utilized by older adults in South Korea. Exclusion criteria encompassed cognitive and functional impairments that could impede study participation (such as hearing loss), as well as individuals residing with others in the same household. This exclusion was due to technical challenges associated with motion-sensing camera detection, which were pertinent to the broader study. Community caregivers offering in-person public senior care services at the community center were deemed eligible to participate in the study. A managerial social worker responsible for overseeing the in-person senior care service at the community center was also recruited to evaluate the website of the monitoring platform.

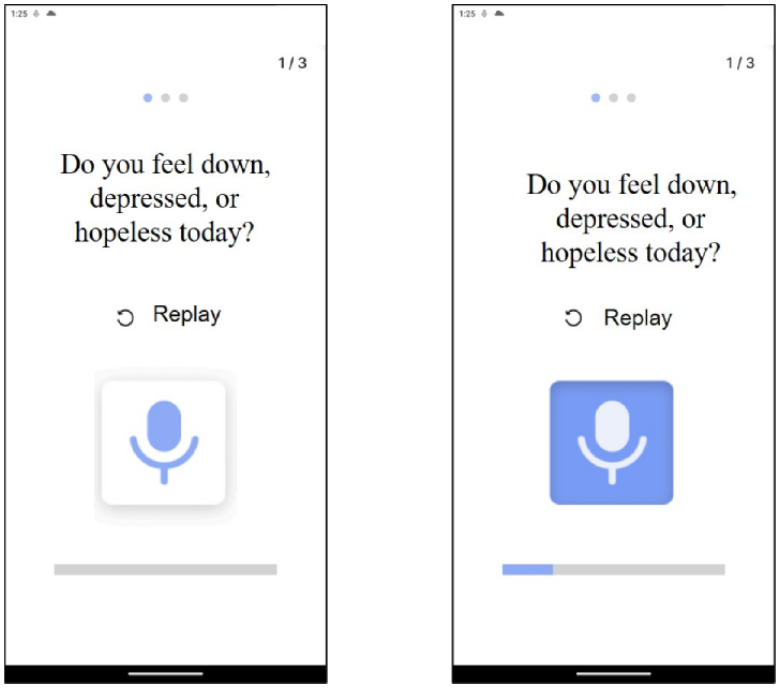
Of the 39 older adults initially recruited for the study, 8 declined participation, primarily citing busyness and concerns about the long-term use of the platform as reasons. Consequently, the study proceeded with the consented participation of the remaining 31 older adults.

Patient Health Questionnaire-9

The 9-item Patient Health Questionnaire-9 (PHQ-9) was used to evaluate daily depressive symptoms through the smartphone app chatbot, reflecting insights from the formative research of the larger study. This decision was informed by older adult users' challenges with touch screen usage and their preference for verbal communication features over visual displays. The PHQ-9 is a widely recognized tool utilized to identify mild and clinical depressive symptoms in nonpsychiatric settings [25,26]. One item from the PHQ-9 questionnaire, specifically concerning suicidal thoughts ("Have you had thoughts that you would be better off dead, or of hurting yourself?"), was omitted due to the potential risk of eliciting negative thoughts, particularly among individuals who may be at risk for depression.

In addition to the PHQ-9 items, we created a chatbot script for daily greetings and safety checks, incorporating 5 commonly used questions recommended by community caregivers. Older adult participants were instructed to open the mobile app every morning, triggering an automated conversation with the chatbot (upon the first daily app launch). Voice recording was activated for 30 seconds after each question, and the recorded file was instantly uploaded to the website accessible to community caregivers (Figure 2). These voice recording files were monitored during the living lab and coded later to identify the presence of daily depressive symptoms (1=depressive symptoms indicated on at least one PHQ-9 item, 0=no reports of depressive symptoms) by 2 research assistants (YBS and SYH) independently. Any discrepancies were resolved through discussions among the 2 coders and an experienced supervisor.

Figure 2. Mobile app function for collecting daily depressive symptoms; the chatbot voice asked 2 randomly selected PHQ-9 items and older adult participants' answers were automatically recorded. The blue bar and the blue microphone icon indicate that voice recording has been activated. PHQ-9: 9-item Patient Health Questionnaire.



Smartwatch-Derived Measures

Daily Sleep Quality

Sleep quality was evaluated using several metrics, including total sleep time, sleep fragmentation index, and the number of long fragmentation episodes. To validate our sleep calculation algorithm, we examined baseline data on participants' usual sleep time and wake time using items from the Pittsburg Sleep Quality Index (PSQI). Our algorithm commenced at 6 PM and concluded at noon the following day to encompass all sleep periods. Additional adjustments were implemented for 1 older adult whose day and night cycles had reversed, allowing for the capture of daytime sleep periods (n=1). The Fitbit algorithm categorized the activity level for each minute the older adult wore the sensor as either asleep or awake. Furthermore, the classification data were prescreened using the following 7

criteria to improve the accuracy of actigraphy-assessed sleep detection based on previous studies [27,28]: (1) if the previous 4 minutes were categorized as awake, the initial minute of the sleep period was adjusted to sleep; (2) if the previous 10 minutes were identified as awake, the subsequent 3 minutes were adjusted to awake; if more than (3) 15 minutes before and (4) after a sleep period lasting less than 6 minutes were classified as awake, then the period of less than 6 minutes of sleep was adjusted to awake; (5) if more than 20 minutes before and after a sleep period lasting less than 10 minutes were categorized as awake, the period of less than 10 minutes of sleep was adjusted to awake. The total time in bed was computed as the duration between the initiation and cessation of the sleep cycle; (6) the sleep onset time was determined as the first time block featuring at least 10 minutes of uninterrupted sleep; and (7) the sleep offset time was determined as the final 10 minutes of uninterrupted sleep before rising from bed. The total sleep time was computed by aggregating the minutes spent asleep from the sleep onset to the sleep offset. The sleep fragmentation index was determined by dividing the number of times the participant awakened for more than 1 minute by the total sleep time [27]. Sleep efficiency was calculated by dividing the total sleep time by the duration of time spent in bed [27].

Heart Rate Variability

HRV was evaluated using both time and frequency domain indicators. These indicators were computed every 5 minutes, around the clock, using the Python-based (Python Foundation) open-source program code Aura-healthcare [29]. The code converted heart rate data to R-R intervals, which represent the time elapsed between 2 successive R-waves in the QRS signal on the electrocardiogram. These intervals are known to be influenced by the activity of the sinus node and autonomic nerve stimulation [30]. Next, the code was used to compute the time domain indicators, which included the SD of the N-N intervals, the normalized or filtered R-R intervals, and the root mean square of successive differences. Additionally, frequency domain indicators were derived, such as the ratio of low frequency (LF) to HF, and HF after applying fast Fourier transformation. This transformation categorized the power of heart rate into HF, LF, and very-low-frequency components. The characteristics and reliability of HRV assessments using Fitbit have been documented in a previous study involving a diverse population [31].

Physical Activity Indices

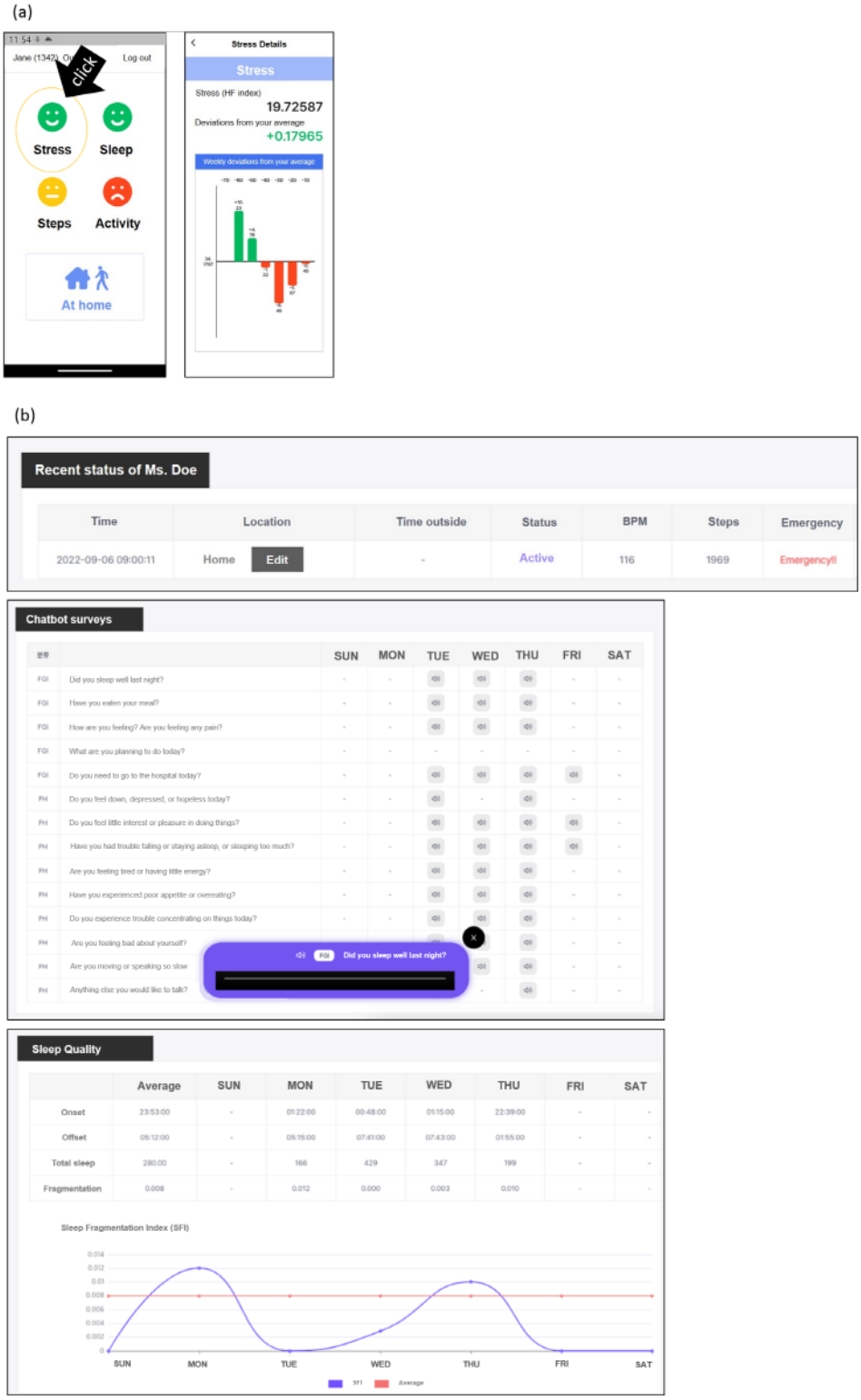
PA levels were evaluated using Fitbit's indicators for steps taken and minutes spent engaging in light, moderate, and intense activity each day throughout the living lab participation. Fitbit uses a 3-axis accelerometer to track steps and PA by analyzing the frequency, duration, intensity, and patterns of movements. Previous studies have confirmed the accuracy of Fitbit activity data [32-34].

Provision of Individualized Daily Health Status Feedback Via the Mobile App

The mobile app designed for older adults and their community caregivers delivered personalized daily health status feedback on stress, sleep, steps, and activity. This feedback was presented using traffic signal colors and accompanied by detailed information, as depicted in Figure 3A. Using the traffic signal colors, a green face within the health status for stress, steps, and activity domain indicated a value higher than 1 SD of the user's average in that domain. A yellow face denoted a value within +1 SD or –1 SD of the average of the health status domain score. A red face indicated a value lower than 1 SD of the average level. The ranges for the green and red faces were inverted for sleep (as measured by sleep fragmentation) compared with the other measurable categories, as a higher value indicates more fragmented sleep. Clicking each face on the initial page (Figure 3A on the left side) advanced the user to the next page, where details (Figure 3A on the right side) were provided regarding the differences between today's value in the domain compared with the average during the first week. Additionally, a weekly graph was included to illustrate the pattern of changes in the weekly values.

The web application designed for the community caregiver and the managerial social worker (Figure 3B) facilitated centralized monitoring by community caregivers, allowing for swift emergency responses to older adults during and after the COVID-19 pandemic. Additionally, it enabled the sharing of daily health status updates. The emergency signal would be triggered if an older adult remained inactive for over 8 hours and/or if their heart rate fell below 30 bpm or rose above 140 bpm.

Figure 3. Example of customized daily health status feedback on stress, sleep, steps, and activity using traffic signal colors and detailed information via the mobile app for (A) the older adult users and (B) the web application for the community caregivers and the managerial social worker.



Pre- and Postsurvey Measures

Geriatric Depression Scale

The 15-item version of the Geriatric Depression Scale (GDS) [35,36] was used to evaluate levels of depressive symptoms in older adults during the pre- and postliving lab surveys. Scores ranged from 0 to 15, with cutoff scores of ≥5 and ≥10 indicating a risk for mild and severe depression, respectively.

Pittsburgh Sleep Quality Index

Subjective sleep quality was evaluated using the Pittsburgh Sleep Quality Index (PSQI) in the pre- and postliving lab surveys [37,38]. The PSQI comprises 19 items and assesses 7 components of sleep over the past month. The total PSQI score was utilized to indicate the overall level of subjective sleep quality, with a cutoff score of 5 or more indicating a risk for sleep disorders.

System Usability Scale

The *System Usability Scale* (SUS) was used to gauge the older adults' experience levels and perceived difficulty in using the digital monitoring platform during the pre- and postliving lab surveys [39]. The scale includes questions regarding the participant's frequency of digital technology usage and their perception of its ease of use, among others. Responses are rated on a 5-point scale ranging from 0=strongly disagree to 4=strongly agree.

Many older adult participants encountered challenges in identifying digital technology to base their answers on for this scale during the preliving lab survey. This survey was conducted on the day when the digital health monitoring platform was installed. To assist participants in answering the survey, examples of digital devices such as smartphones or kiosks at hospitals were provided. Postliving lab assessments of the SUS were conducted based on the participants' experiences with the developed health monitoring platform.

In addition to the SUS, participants were queried about the number of days per week they utilized the platform and the number of functions they accessed on the platform during the postliving lab survey. Furthermore, at the posttest, participants were requested to offer qualitative feedback aimed at enhancing the platform for future trials. This feedback was transcribed verbatim by a research assistant.

Covariates

Covariates were selected a priori, drawing from existing literature that outlines demographic and health risk factors associated with daily depressive symptoms among older adults [7,40]. Alongside baseline depression levels, demographic and health factors known to be correlated with depression (eg, age, sex, and chronic health conditions) were utilized as covariates.

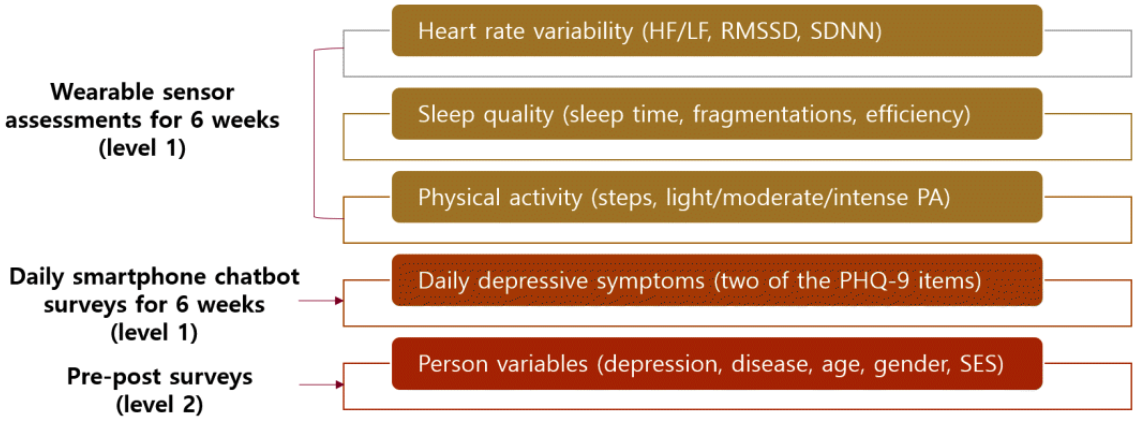
Statistical Analysis

After analyzing missing data and assessing the distributions of key variables, descriptive analyses were performed. Group comparisons between depressed and nondepressed older adults were conducted using *t* tests for continuous variables (eg, age, BMI, PSQI total score, daily depressive symptoms, wearable sensor assessments of HRV, sleep, and PA indicators), chi-square tests for categorical variables (eg, sex, education, income, chronic disease, sleep disorder categories), and Fisher exact tests for smoking status. Subsequently, bivariate correlations were explored between baseline depression and person mean variables (daily depressive symptoms and wearable sensor assessments) over 5 weeks, excluding the initial week for adaptation.

To identify digital biomarkers for daily depressive symptoms in older adults, we used multilevel modeling (MLM) analysis. This method allowed us to scrutinize the daily shifts in depressive symptoms, as well as the smartwatch-detected metrics of sleep quality, HRV, and activity levels over time. We utilized SAS PROC MIXED (SAS Institute) for this analysis. Applying MLM analysis to daily depression and smartwatch data facilitates the identification of antecedents and correlates of daily depressive symptoms. It also allows for the use of participants as their own controls [41]. In the multilevel models, the level 1 equation investigated whether daily sleep, HRV, and PA indices predicted the likelihood of reporting any depressive symptoms on the same day and the following day. To assess the primary impact of smartwatch indices on daily depressive symptoms, the main effect models for each of the sleep/HRV/activity model indices were utilized as level 1 predictors, with daily depressive symptoms on the same day or the next day serving as level 1 outcome variables. In all MLM analyses, the level 2 equations incorporated personal characteristics, including participants' demographic and health covariates. Daily measures were person mean centered, while level 2 covariates were grand mean centered. This approach allows for the interpretation of estimates as the probability of daily depressive symptoms when there are deviations in the predicting digital biomarker variable from the participant's own average across the 5-week living lab period. The multilevel data structure of this study is illustrated in Figure 4. A type I error rate of 0.05 was established for statistical analyses, except for the MLM analyses, where Bonferroni and Holm corrections for multiple testing were applied, resulting in a type I error rate of 0.005.

The pre- and posttest results were assessed using a paired *t* test for variables demonstrating normal distributions, such as system usability. However, because of their skewed distributions, depressive symptoms and sleep quality were analyzed using the Wilcoxon signed rank test. To delve deeper into the age moderation effects on the pre- and posttest results for usability, repeated-measure ANOVA was applied to examine potential differences in system usability between the oldest group (>75 years old) and the rest (65-74 years old).

Figure 4. Multilevel data structure of this study with wearable sensor and daily depressive symptom chatbot survey data as level 1 and pre- and posttest survey data as level 2. HF: high frequency; LF: low frequency; PA: physical activity; PHQ-9: 9-item Patient Health Questionnaire; RMSSD: root mean square of successive differences; SDNN: standard deviation of N-N intervals; SES: socioeconomic status.



Power Analysis

Power analysis for multilevel model design is recognized for its complexity. Previous studies have revealed a compensatory relationship between the number of participants and the number of observations per participant [42,43]. We gathered 807 days of observational data from 25 older adults, averaging 32.28 days per participant (range 8-40 days). The variability in the number of available days stemmed primarily from 2 reasons: participants' nonadherence to data collection protocols (eg, forgetting to wear the Fitbit) and the scheduling preferences for installing and removing digital devices from their homes (eg, 1 participant needed to schedule device removal 1 week later than others because of personal reasons). A recent simulation study proposed that a sample size of 25 combined with continuous data collection over 30 days yields acceptable performance levels in terms of parameter estimations [42].

Data Exclusion and Missing Data

Out of the initial 31 adults who participated, data from 6 participants were excluded from the study. They dropped out after the installation of sensing devices due to experiencing inconvenience and difficulties in using mobile apps and digital devices. Among the initial group, 25 older adults provided their responses to daily verbal surveys and wearable sensor data. These data were included in the multilevel analyses, focusing on the primary outcome of daily depressive symptoms. Of 808 days of assessments (with a mean of 32.32 days per participant), there were 186 days (23% of total days) with missing daily verbal survey data, primarily as a result of participants forgetting to respond or no responses being recorded. Additionally, there were 269 days (33.3%) with missing HRV data, 440 days (54.5%) with missing sleep measures, and 298 days (36.9%) with missing steps data. These missing data were primarily attributed to participants forgetting to wear a smartwatch during the day or night, or experiencing issues with charging the device. Previous studies on missing data analysis suggest that multilevel analyses using full information maximum likelihood estimation methods are generally robust against estimation biases arising from data with partial missingness [44]. Regarding pre-post test effects, 1 participant did not complete the posttest survey as a result of long-term travel for family matters, resulting in data from 24 older adults being included in the analyses.

Results

Descriptive Overview

The demographic and health characteristics of the older adult living lab participants are detailed in Table 1. The participants exhibited characteristics typical of older adults, with a mean age of 76.40 (SD 4.23) years. Additionally, they displayed social vulnerabilities concerning education and income levels. Among the participants, there were more women (n=19) than men (n=6). On average, participants reported having 4 chronic disease conditions, with 3 (12%) of the 25 participants reporting depressive disorders and 10 (40%) reporting arthritis or diabetes. The average level of sleep quality was poor, with 22 (88%) participants at high risk for sleep disorders.

Compared with individuals with a low risk of depression at baseline, those with a high risk were more likely to report daily depressive symptoms via verbal surveys. Additionally, they tended to sleep longer but with lower efficiency and were less likely to engage in light, moderate, and intense PA (*P* values ranged from <.001 to.02).

Figure 5 illustrates the 24-hour profile of heart rate on both total days and days when older adults reported depressive symptoms. The patterns of heart rate on days with depressive symptoms seemed to exhibit more variability throughout the day compared with total days.

Table 1. Characteristics of older adult living lab participants.

Characteristics	Total (n=25)	Baseline depressive symptoms[a]		*P* value
		No (n=15)	Yes (n=10)	
Demographic characteristics				
Age (years), mean (SD)	76.40 (4.23)	76.40 (4.42)	76.40 (4.17)	>.99
Sex: women, n (%)	19 (76)	11 (73)	8 (80)	.70
Education, n (%)				.51
Elementary school or less	14 (56)	7 (47)	7 (70)	
Middle school	2 (8)	2 (13)	0 (0)	
High school	5 (20)	3 (20)	2 (20)	
College or more	4 (16)	3 (20)	1 (10)	
Monthly income (Korean won[b]), n (%)				.95
Less than 500,000	4 (16)	2 (13)	2 (20)	
500,001-1,000,000	16 (64)	10 (67)	6 (60)	
1,000,001-1,500,000	2 (8)	1 (7)	1 (10)	
1,500,001 or more	3 (12)	2 (13)	1 (10)	
Health characteristics				
Smoking, n (%)				.18
Current smoking	1 (4)	0 (0)	1 (10)	
Past smoking	6 (24)	6 (40)	0 (0)	
Never	18 (72)	9 (60)	9 (90)	
BMI (kg/m^2), mean (SD)	24.56 (3.94)	24.65 (2.44)	24.43 (5.68)	.89
Number of chronic diseases, mean (SD)	4.20 (1.71)	4.20 (1.37)	4.20 (2.20)	>.99
Arthritis, n (%)	10 (40)	5 (33)	5 (50)	.49
Cardiovascular disease, n (%)	18 (72)	12 (80)	6 (60)	.27
Diabetes, n (%)	10 (40)	7 (47)	3 (30)	.40
Depression, n (%)	3 (12)	2 (13)	1 (10)	.80
Dementia, n (%)	1 (4)	0 (0)	1 (10)	.21
Baseline study variables				
PSQI[c] global sleep quality (0-21 points[d]), median (IQR)	8.00 (6.00-10.00)	7.00 (6.00-9.00)	9.00 (6.75-11.25)	.22
Sleep disorders, n (%) of PSQI>5	22 (88)	12 (80)	10 (100)	.13
Digital assessments, person mean (SD)				
Daily PHQ symptoms, 0-1	0.37 (0.48)	0.31 (0.46)	0.48 (0.50)	<.001
Heart rate variability, mean (SD)				
Low frequency/high frequency	5.94 (1.33)	5.89 (1.30)	6.00 (1.36)	.34
High frequency	26.02 (37.14)	23.75 (20.57)	28.71 (49.99)	.14
SD of the N-N intervals	103.01 (41.89)	108.81 (44.73)	96.17 (37.22)	<.001
Root mean square of successive differences	9.32 (3.87)	9.13 (3.58)	9.55 (4.19)	.21
Sleep, mean (SD)				
Total sleep time	319.63 (163.08)	297.98 (166.73)	345.12 (155.30)	.004
Sleep fragmentation index	0.02 (0.02)	0.02 (0.02)	0.02 (0.02)	.61
Sleep efficiency	65.98 (20.86)	68.31 (21.18)	63.24 (20.20)	.02
Physical activity, mean (SD)				

Characteristics	Total (n=25)	Baseline depressive symptoms[a]		*P* value
		No (n=15)	Yes (n=10)	
Steps	3207.11 (3901.80)	3512.54 (4024.61)	2839.53 (3724.16)	.05
Light physical activity	130.46 (101.55)	148.81 (107.43)	104.19 (86.19)	<.001
Moderate physical activity	14.05 (19.34)	15.75 (20.09)	11.62 (17.98)	.02
Intense physical activity	21.89 (27.85)	24.40 (26.02)	18.30 (29.98)	.02

[a]The presence or absence of depressive symptoms was determined based on a score of 5 or higher on the Geriatric Depression Scale.

[b]US $1=1344 Korean won.

[c]PSQI: Pittsburg Sleep Quality Index.

[d]Higher scores indicate worse sleep quality.

Figure 5. The average 24-hour profiles of heart rate on (A) total days and (B) days with depressive symptoms.

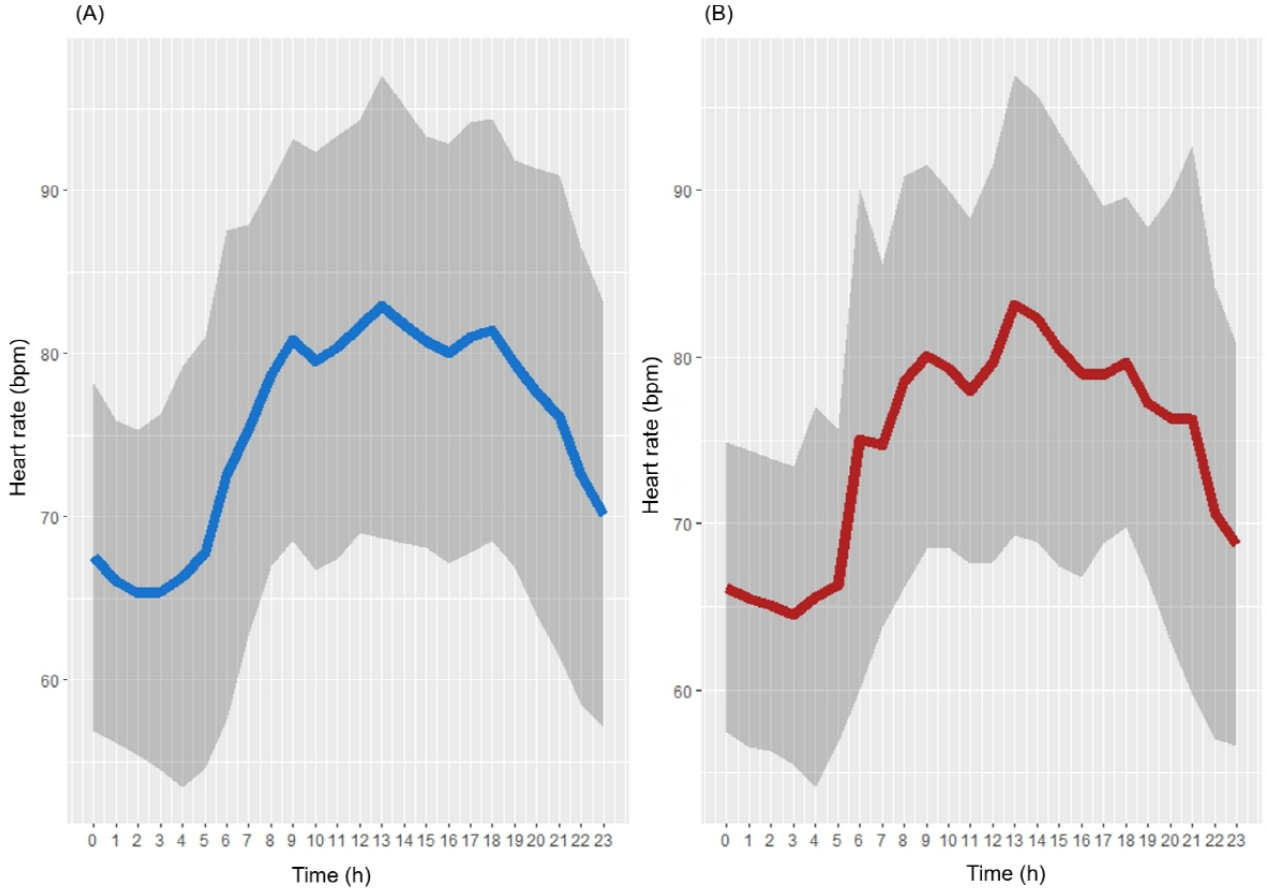
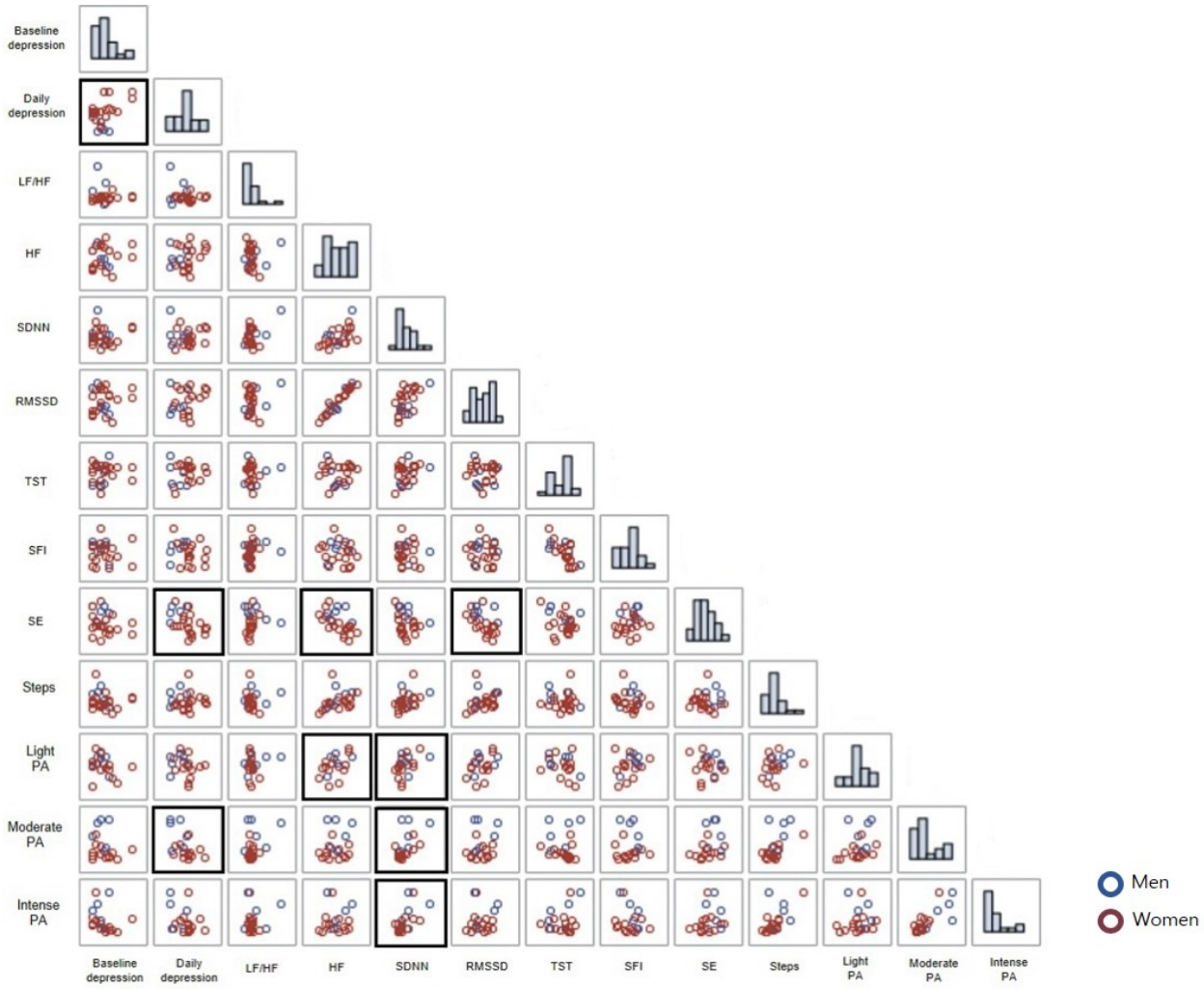


Figure 6 displays the results of bivariate correlation analysis among baseline depression measures; the person mean score of daily depressive symptoms; and the smartwatch measures of HRV, sleep quality, and PA. Additionally, distributions of the variables are presented on the diagonals. As anticipated, baseline depressive symptoms were significantly correlated with the person mean variable of daily depressive symptoms (r=0.43, P=.03). Interestingly, the person mean of daily depressive symptoms exhibited a negative association with daily sleep efficiency (r=–0.47, P=.02) and the amount of moderate PA (r=–0.50, P=.02). Furthermore, HRV, sleep, and PA indicators exhibited correlations with each other. Specifically, person mean variables of daily HF and root mean square of successive differences were associated with daily sleep efficiency (r=0.58, P=.003 and r=–0.51, P=.009, respectively). Additionally, the person mean levels of daily HRV indicators were associated with the person mean levels of daily PA. HF exhibited a positive correlation with light PA (r=0.49, P=.03), whereas SD of the N-N intervals was positively associated with light (r=0.45, P=.05), moderate (r=0.54, P=.01), and intense PA (r=0.59, P=.006).

Figure 6. Bidirectional correlations among baseline depression; person mean of daily depression; and digital markers of heart rate variability indicators, sleep indicators, and physical activity measures (bold box indicates significance at $P<.05$; red dots represent women, blue dots represent men. HF: high frequency; LF/HF: low frequency-to-high frequency power ratio; PA: physical activity; RMSSD: root mean square of successive differences; SDNN: standard deviation of N-N intervals; SE: sleep efficiency; SFI: sleep fragmentation index; TST: total sleep time.



Primary Outcomes

As depicted in Table 2, the MLM results indicated that both the daily sleep fragmentation index and sleep efficiency significantly predicted the occurrence of daily depressive symptoms measured on the subsequent day, even after adjusting for baseline depression, age, sex, and chronic disease conditions (odds ratio 2.066, 95% CI 1.252-3.411; P=.003 for daily sleep fragmentation and odds ratio 0.972, 95% CI 0.955-0.989; P=.001 for daily sleep efficiency). In essence, on days following more fragmented sleep or lower sleep efficiency compared with their own average, older adult participants were more inclined to report depressive symptoms via the chatbot survey. Notably, the effects of daily sleep fragmentation and efficiency on daily depressive symptoms remained significant even after applying Bonferroni-Holm corrections for multiple testing. However, none of the daily assessments in HRV, sleep quality, or PA predicted the occurrence of depressive symptoms the following day.

Table 2. The concurrent and lagged effect models of daily digital indicators on daily depressive symptoms[a,b,c].

Digital predictors	Same-day depressive symptoms		Next-day depressive symptoms	
	Odds ratio (95% CI)	*P* value	Odds ratio (95% CI)	*P* value
Heart rate variability				
High frequency/low frequency	0.975 (0.790-1.204)	.82	0.929 (0.751-1.148)	.49
High frequency	0.987 (0.617-1.579)	.96	0.978 (0.632-1.513)	.92
SD of the N-N intervals	0.998 (0.990-1.005)	.57	1.001 (0.994-1.008)	.84
Root mean square of successive differences	0.954 (0.866-1.051)	.34	0.998 (0.924-1.078)	.96
Sleep				
Total sleep time	0.999 (0.996-1.001)	.16	1.001 (0.999-1.003)	.35
Sleep fragmentation index	2.066 (1.252-3.411)	.003[c]	1.149 (0.750-1.759)	.77
Sleep efficiency (%)	0.972 (0.955-0.989)	.001[c]	0.999 (0.983-1.016)	.93
Physical activity				
Steps	1.000 (1.000-1.000)	.28	1.000 (1.000-1.000)	.21
Light physical activity	0.999(0.996-1.002)	.48	1.002 (0.998-1.005)	.33
Moderate physical activity	0.992 (0.971-1.014)	.48	0.985 (0.964-1.006)	.16
Intense physical activity	0.998 (0.985-1.011)	.73	1.002 (0.989-1.015)	.75

[a]All analyses were adjusted for age, sex, chronic disease conditions, and baseline depression.

[b]All digital predictors were person mean centered so that daily values represent deviations from the individuals' own means across the participation days.

[c]The Bonferroni-Holm correction suggests that *P* values indicating significance should be <.005, with 11 multiple tests for each of the 2 outcome variables.

Pre- and Posttest Results

As presented in Table 3, the pre- and posttest results showed significant decreases in depressive symptoms (pre-post difference –1.000, 95% CI –2.000 to 0.000, W=143.500, z=–1.976, P=.048) and improvements in sleep quality (pre-post difference –1.500, 95% CI –3.000 to 0.000, W=165.000, z=–2.252, P=.02) after the 6-week living lab than before. However, there were no significant changes in the levels of usability (pre-post difference –6.354, 95% CI –15.969 to 3.261, t_{23}=–1.376, P=.18). Age moderation was also not significant in the pre- and posttest changes in usability ($F_{1,22}$=437.682, P=.20). At the posttest survey, participants reported using the monitoring app a mean of 6.68 (SD 1.44) days per week and utilizing 1.58 functions (SD 1.91) out of a possible 5. These functions included recording verbal surveys for daily depressive symptoms and checking individualized health feedback for stress, steps, physical functions, and sleep time.

In the posttest survey, older adults expressed satisfaction with the chatbot survey, particularly appreciating its inquiries about their daily lives and health conditions. They felt cared for and safe, knowing that their voice message would be delivered to their community caregiver daily. However, older adults also voiced frustrations with smartwatch malfunctions. They experienced technical issues during the living lab period as a result of difficulties in regularly charging the smartwatch, accidentally turning off the smartphone's Bluetooth connection, and overloading of the central server.

Table 3. Changes in depressive symptoms, sleep quality, and usability levels from pre- to posttests.

Variables	Pretest	Posttest	*P* value
Depressive symptoms (Geriatric Depression Scale), median (IQR)[a]	3.000 (1.000-6.000)	1.500 (0.250-5.000)	.048
Sleep quality (Pittsburg Sleep Quality Index), median (IQR)[a]	8.000 (6.000-10.000)	6.000 (4.000-7.750)	.02
Usability (System Usability Scale), mean (SD)[b]	53.333 (24.524)	59.688 (19.620)	.18

[a]Wilcoxon signed rank test was used to analyze depressive symptoms and sleep quality due to their skewed distributions.

[b]Paired *t* test was used to evaluate usability based on a normal distribution of the usability scores.

Discussion

Principal Findings

This study explored the viability of digitally phenotyping depressive symptoms and implementing continuous digital mental health monitoring among a small cohort of socially vulnerable older adults (n=25) in their everyday living settings. By conducting daily verbal assessments of depressive symptoms via a chatbot and monitoring daily fluctuations in digital biomarkers using a smartwatch, this study indicates that 2 digital measures of daily sleep quality, namely, sleep fragmentation and sleep efficiency, predicted the occurrence of daily depressive symptoms on the subsequent day. These associations remained significant even after adjusting for potential confounders such as baseline depression, age, gender, and chronic disease conditions. To the best of our knowledge, this study represents the first attempt to explore the digital phenotyping of depressive symptoms in socially vulnerable older adults within their own living environments over an extended duration. Furthermore, the pre- and posttest findings revealed that while depressive symptoms and sleep issues improved following the 6-week platform utilization alongside community caregiver support, no significant enhancement was observed in system usability. Although the current findings are preliminary and lack a comparison group, they suggest the potential health advantages of integrating a digital health monitoring platform for older adults in conjunction with in-person community senior care services.

Comparison With Prior Work

Through the development of analytic algorithms for continuous sensing of sleep time, fragmentation, and efficiency, this study unveiled that daily alterations in sleep fragmentation and efficiency during the night, relative to an individual's average, were predictive of daily depressive symptoms among older adults on the subsequent day. These findings are consistent with prior studies that have demonstrated a significant correlation between actigraphy-assessed wake time after sleep onset, sleep efficiency, and 1-time survey assessments of depressive symptoms among adults with a history of clinical depression [12,40,45]. A meta-analysis of 38 studies incorporating actigraphy assessments similarly found significant disparities in longer wake time after sleep onset between patients with depression and healthy controls [46]. Our findings contribute to the existing literature by investigating daily variations in depressive symptoms among socially isolated older adults over an extended period. Moreover, this study examines whether within-person fluctuations in sleep quality forecast ongoing changes in depressive symptoms across consecutive days.

The current findings revealed that neither actigraphy-assessed HRV nor PA measures significantly predicted daily depressive symptoms. Prior research has demonstrated varied associations between HRV and depressive symptoms, depending on factors such as the diagnosis or severity of depression and the cardiovascular health of the population under study. Indeed, a meta-analysis of 21 studies highlighted significant distinctions between individuals with depression and healthy controls [47]. Additionally, a previous investigation involving patients with depression documented a negative correlation between HRV (eg, root mean square of successive differences) and cognitive symptoms of depression (eg, rumination) on the same day [48]. However, such relationships between HRV and depressive symptoms were not observed in adults without clinical depression or those with cardiovascular health issues. For instance, depressive symptoms were not significantly correlated with electrocardiogram-measured HRV in healthy adults [49] or adults at risk of coronary artery disease [50]. The majority of older adults (18/25, 72%) in this study had cardiovascular disease conditions, a prevalence higher than the national average in South Korea (40.36%) and the global prevalence (31.0%-70%) for adults over 70 years [51-53]. Future research with a larger sample of older adults is warranted to investigate the associations between depressive symptoms and HRV among older adults with diverse chronic disease conditions.

Regarding the association between depressive symptoms and PA, in contrast to previous findings, daily fluctuations in PA did not predict daily depressive symptoms in this study. Indeed, a meta-analysis comprising 42 studies utilizing actigraphy or pedometer assessments of PA demonstrated significant associations between average PA levels and depressive symptom severity among adults, regardless of clinical depression status [54]. Additionally, intervention programs aimed at increasing PA were associated with lower levels of depressive symptoms compared with control groups in adult populations without clinical depression [55]. There could be several potential explanations for the lack of association between daily PA and daily depressive symptoms observed in our study. It is plausible that long-term patterns of PA, rather than day-to-day fluctuations, may better explain changes in depressive symptoms. Consistent with this hypothesis, the present findings demonstrated a significant negative correlation between the average levels of depressive symptoms and the average levels of moderate PA across individuals. Future research is warranted to explore potential factors that might elucidate the relationship between daily depressive symptoms and PA. This could include investigating cumulative patterns of PA, the impact of specific types of PA (such as group exercise), or external circumstances that may impede physical activities (such as the COVID-19 pandemic).

Finally, the pre- and posttest findings indicated that older adults exhibited enhancements in depressive symptoms and sleep quality following their utilization of the platform alongside their community caregivers. It is important to note that the participating older adults had previously been receiving in-person senior caregiving services, which entailed regular phone calls and weekly in-person visits for safety checks, as well as assistance with hospital visits if necessary. However, the utilization of the present digital monitoring platform may have empowered community caregivers to optimize the timing of service provision based on when older adults were most in need. For example, an older adult user received red lights on sleep and physical activity via the user app when they had poor sleep and skipped exercise due to a flare-up of chronic disease conditions. Their community caregiver received the same daily feedback and voice recordings (from the respective user) about health issues via the caregiver app. Subsequently, the caregiver

phoned or visited each participant to check on the negative health changes for that day. The current findings align with recent studies demonstrating the health-improving effects of digital health care services for older adults. These effects were observed when mental health professionals provided in-person services [56,57] or when community health workers with intensive training [58,59] were connected to the digital service. However, this study makes a unique contribution to the literature by presenting preliminary findings suggesting that community caregivers without a health care specialty or intensive training in mental health care can potentially enhance the mental health of socially vulnerable older adults through the utilization of a digital health monitoring platform.

Contrary to expectations, the results of the usability test in this study suggest that older adults experienced difficulties in using the mental health monitoring system, and these challenges did not improve over time. These findings underscore the importance of offering adequate in-person assistance, including group lessons, in-person training sessions, and troubleshooting visits for older adults when introducing a new digital health care service.

Limitations

Our findings should be interpreted with caution due to several limitations. First, the sample size was limited to 25 older adults, which may not have provided sufficient power to detect small effect sizes of associations between daily depressive symptoms and sensor-based daily indicators, as well as changes in outcome measures pre- to posttest. The current results may not be generalizable to other older adult populations or other countries due to our deliberate selection of older adults with social vulnerabilities, including those living alone with low income. These individuals may experience greater challenges in accessing digital technologies compared with the overall older adult population [60]. We used MLM to investigate the concurrent digital biomarkers of daily depressive symptoms. However, future studies may benefit from exploring alternative analytic approaches such as machine learning to classify older adults into high- and low-risk groups for depression. Additionally, it is important to acknowledge that the strict implementation of COVID-19–related social restriction policies in 2022 may have influenced the findings of this study.

Conclusions

This study investigated the feasibility of mental and physical health monitoring platforms for socially vulnerable older adults and their community caregivers in their everyday living environments. Additionally, it explored whether passively sensed measurements of HRV, sleep, and PA predicted daily fluctuations of depressive symptoms in older adults. The findings indicate that older adults successfully utilized the monitoring platform throughout the 6-week study to monitor their daily health status. This information was also shared with community caregivers to enhance the existing senior care service. Additionally, the MLM results revealed same-day associations between daily sleep quality indicators (sleep fragmentation index and sleep efficiency from the previous night) and daily depressive symptoms. The pre- and posttest results indicate that older adults exhibited enhancements in depressive symptoms and sleep quality following the utilization of the monitoring platform, which was integrated with their existing community care services. These findings offer preliminary support for the digital phenotyping of geriatric depressive symptoms using sleep measures obtained from a wearable sensor. We also propose a potential service delivery model for developing a hybrid senior care service that integrates online and offline components. This model leverages existing community-based senior care services to facilitate the early detection and prevention of mental health declines in socially vulnerable older adults.

Acknowledgments

We are very grateful for the support of the staff at the Seong-buk City Senior Welfare Center, who assisted with the recruitment of participants and in-person data collection during the pandemic period. This study was supported by a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number H21C0572); the Basic Science Research Program through the National Research Foundation of Korea (NRF), funded by the Ministry of Education (grant number NRF2020R1A6A3A0110041011); and the Pioneer Research Center Program through the National Research Foundation of Korea, funded by the Ministry of Science, ICT & Future Planning (grant number 2022M3C1A3081294).

Conflicts of Interest

JK and SS developed the health monitoring platform software in collaboration with M2S+ Co. and owned the intellectual property rights. The Department of Health and Welfare of the Korean Government provided funding for the development of this platform and owned the right to utilize the developed platform as a public service for older adults.

References

1. The old-age to working-age ratio will nearly quadruple in Korea by 2060. OECD Library. Paris, France: OECD Publishing; 2022. URL: https://tinyurl.com/axapwskj [accessed 2024-06-09]
2. Kim GE, Jo M, Shin Y. Increased prevalence of depression in South Korea from 2002 to 2013. Sci Rep 2020 Oct 12;10(1):16979. [doi: 10.1038/s41598-020-74119-4] [Medline: 33046758]

3. Fortuna KL, Torous J, Depp CA, Jimenez DE, Areán PA, Walker R, et al. A future research agenda for digital geriatric mental healthcare. Am J Geriatr Psychiatry 2019 Nov;27(11):1277-1285 [FREE Full text] [doi: 10.1016/j.jagp.2019.05.013] [Medline: 31196619]
4. World Health Organization (WHO). Mental health of older adults. WHO. Geneva, Switzerland: WHO; 2023. URL: https://www.who.int/news-room/fact-sheets/detail/mental-health-of-older-adults [accessed 2023-12-12]
5. OECD. Old-age suicide rate is very high in Korea but has been declining. In: OECD Reviews of Pension Systems: Korea. Paris, France: OECD Publishing; 2022.
6. World Health Organization (WHO). Aging and health. WHO. Geneva, Switzerland: WHO; 2023. URL: https://www.who.int/news-room/fact-sheets/detail/ageing-and-health [accessed 2023-12-12]
7. Zenebe Y, Akele B, W/Selassie M, Necho M. Prevalence and determinants of depression among old age: a systematic review and meta-analysis. Ann Gen Psychiatry 2021 Dec 18;20(1):55 [FREE Full text] [doi: 10.1186/s12991-021-00375-x] [Medline: 34922595]
8. Kamath J, Leon Barriera R, Jain N, Keisari E, Wang B. Digital phenotyping in depression diagnostics: integrating psychiatric and engineering perspectives. World J Psychiatry 2022 Mar 19;12(3):393-409 [FREE Full text] [doi: 10.5498/wjp.v12.i3.393] [Medline: 35433319]
9. Sano A, Taylor S, McHill AW, Phillips AJ, Barger LK, Klerman E, et al. Identifying objective physiological markers and modifiable behaviors for self-reported stress and mental health status using wearable sensors and mobile phones: observational study. J Med Internet Res 2018 Jun 08;20(6):e210 [FREE Full text] [doi: 10.2196/jmir.9410] [Medline: 29884610]
10. Onnela J, Rauch SL. Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. Neuropsychopharmacology 2016 Jun;41(7):1691-1696 [FREE Full text] [doi: 10.1038/npp.2016.7] [Medline: 26818126]
11. BinDhim NF, Shaman AM, Trevena L, Basyouni MH, Pont LG, Alhawassi TM. Depression screening via a smartphone app: cross-country user characteristics and feasibility. J Am Med Inform Assoc 2015 Jan 17;22(1):29-34 [FREE Full text] [doi: 10.1136/amiajnl-2014-002840] [Medline: 25326599]
12. Difrancesco S, Lamers F, Riese H, Merikangas KR, Beekman ATF, van Hemert AM, et al. Sleep, circadian rhythm, and physical activity patterns in depressive and anxiety disorders: a 2-week ambulatory assessment study. Depress Anxiety 2019 Oct 26;36(10):975-986 [FREE Full text] [doi: 10.1002/da.22949] [Medline: 31348850]
13. Sorri K, Mustafee N, Seppänen M. Revisiting IoT definitions: a framework towards comprehensive use. Technological Forecasting and Social Change 2022 Jun;179:121623. [doi: 10.1016/j.techfore.2022.121623]
14. Wildenbos GA, Peute L, Jaspers M. Aging barriers influencing mobile health usability for older adults: a literature based framework (MOLD-US). Int J Med Inform 2018 Jun;114:66-75. [doi: 10.1016/j.ijmedinf.2018.03.012] [Medline: 29673606]
15. Xie B, Charness N, Fingerman K, Kaye J, Kim MT, Khurshid A. When going digital becomes a necessity: ensuring older adults' needs for information, services, and social inclusion during COVID-19. J Aging Soc Policy 2020;32(4-5):460-470 [FREE Full text] [doi: 10.1080/08959420.2020.1771237] [Medline: 32507061]
16. Chen K, Chan AHS. Gerontechnology acceptance by elderly Hong Kong Chinese: a senior technology acceptance model (STAM). Ergonomics 2014;57(5):635-652. [doi: 10.1080/00140139.2014.895855] [Medline: 24655221]
17. Willems SH, Rao J, Bhambere S, Patel D, Biggins Y, Guite JW. Digital solutions to alleviate the burden on health systems during a public health care crisis: COVID-19 as an opportunity. JMIR Mhealth Uhealth 2021 Jun 11;9(6):e25021 [FREE Full text] [doi: 10.2196/25021] [Medline: 34033575]
18. Chen C, Ding S, Wang J. Digital health for aging populations. Nat Med 2023 Jul 18;29(7):1623-1630. [doi: 10.1038/s41591-023-02391-8] [Medline: 37464029]
19. Hunsaker A, Hargittai E. A review of Internet use among older adults. New Media & Society 2018 Jul 16;20(10):3937-3954. [doi: 10.1177/1461444818787348]
20. Delello JA, McWhorter RR. Reducing the digital divide: connecting older adults to iPad technology. J Appl Gerontol 2017 Jan;36(1):3-28. [doi: 10.1177/0733464815589985] [Medline: 26084479]
21. European Network of Living Labs (ENoLL). What Are Living Labs? ENoLL. 2023. URL: https://enoll.org/about-us/ [accessed 2023-12-12]
22. Verloo H, Lorette A, Rosselet Amoussou J, Gillès de Pélichy E, Matos Queirós A, von Gunten A, et al. Using living labs to explore needs and solutions for older adults with dementia: scoping review. JMIR Aging 2021 Aug 19;4(3):e29031 [FREE Full text] [doi: 10.2196/29031] [Medline: 34420916]
23. Korman M, Weiss PL, Kizony R. Living Labs: overview of ecological approaches for health promotion and rehabilitation. Disabil Rehabil 2016 Jul 03;38(7):613-619. [doi: 10.3109/09638288.2015.1059494] [Medline: 26138020]
24. Jo S, Song S, Kim J, Song C. Agreement between Azure Kinect and marker-based motion analysis during functional movements: a feasibility study. Sensors (Basel) 2022 Dec 14;22(24):9819 [FREE Full text] [doi: 10.3390/s22249819] [Medline: 36560187]
25. Martin A, Rief W, Klaiberg A, Braehler E. Validity of the Brief Patient Health Questionnaire Mood Scale (PHQ-9) in the general population. Gen Hosp Psychiatry 2006;28(1):71-77. [doi: 10.1016/j.genhosppsych.2005.07.003] [Medline: 16377369]
26. Manea L, Gilbody S, McMillan D. A diagnostic meta-analysis of the Patient Health Questionnaire-9 (PHQ-9) algorithm scoring method as a screen for depression. Gen Hosp Psychiatry 2015 Jan;37(1):67-75. [doi: 10.1016/j.genhosppsych.2014.09.009] [Medline: 25439733]

27. Fekedulegn D, Andrew ME, Shi M, Violanti JM, Knox S, Innes KE. Actigraphy-based assessment of sleep parameters. Ann Work Expo Health 2020 Apr 30;64(4):350-367 [FREE Full text] [doi: 10.1093/annweh/wxaa007] [Medline: 32053169]
28. Webster JB, Kripke DF, Messin S, Mullaney DJ, Wyborney G. An activity-based sleep monitor system for ambulatory use. Sleep 1982;5(4):389-399. [doi: 10.1093/sleep/5.4.389] [Medline: 7163726]
29. hrv-analysis documentation. Github. URL: https://aura-healthcare.github.io/hrv-analysis [accessed 2024-05-05]
30. Pappano AJ, Wier WG. Cardiovascular Physiology: Cardiovascular Physiology--E-Book. Edinburgh, Scotland: Elsevier Health Sciences; 2018.
31. Natarajan A, Pantelopoulos A, Emir-Farinas H, Natarajan P. Heart rate variability with photoplethysmography in 8 million individuals: a cross-sectional study. Lancet Digital Health 2020 Dec;2(12):e650-e657. [doi: 10.1016/s2589-7500(20)30246-6]
32. Feehan LM, Geldman J, Sayre EC, Park C, Ezzat AM, Yoo JY, et al. Accuracy of Fitbit devices: systematic review and narrative syntheses of quantitative data. JMIR Mhealth Uhealth 2018 Aug 09;6(8):e10527 [FREE Full text] [doi: 10.2196/10527] [Medline: 30093371]
33. Redenius N, Kim Y, Byun W. Concurrent validity of the Fitbit for assessing sedentary behavior and moderate-to-vigorous physical activity. BMC Med Res Methodol 2019 Feb 07;19(1):29 [FREE Full text] [doi: 10.1186/s12874-019-0668-1] [Medline: 30732582]
34. Sushames A, Edwards A, Thompson F, McDermott R, Gebel K. Validity and reliability of Fitbit Flex for step count, moderate to vigorous physical activity and activity energy expenditure. PLoS One 2016;11(9):e0161224 [FREE Full text] [doi: 10.1371/journal.pone.0161224] [Medline: 27589592]
35. Yesavage JA, Brink TL, Rose TL, Lum O, Huang V, Adey M, et al. Development and validation of a geriatric depression screening scale: a preliminary report. J Psychiatr Res 1982;17(1):37-49. [doi: 10.1016/0022-3956(82)90033-4] [Medline: 7183759]
36. Jung IK, Kwak DI, Shin DK, Lee MS, Lee HS, Kim JY. A reliability and validity study of Geriatric Depression Scale. J Korean Neuropsychiatr Assoc 1997;36(1):103-112 [FREE Full text]
37. Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res 1989 May;28(2):193-213. [doi: 10.1016/0165-1781(89)90047-4] [Medline: 2748771]
38. Sohn SI, Kim DH, Lee MY, Cho YW. The reliability and validity of the Korean version of the Pittsburgh Sleep Quality Index. Sleep Breath 2012 Sep;16(3):803-812. [doi: 10.1007/s11325-011-0579-9] [Medline: 21901299]
39. Holden RJ. A Simplified System Usability Scale (SUS) for Cognitively Impaired and Older Adults. Proceedings of the International Symposium on Human Factors and Ergonomics in Health Care 2020 Sep 16;9(1):180-182 [FREE Full text] [doi: 10.1177/2327857920091021]
40. Kim H, Lee S, Lee S, Hong S, Kang H, Kim N. Depression prediction by using ecological momentary assessment, Actiwatch data, and machine learning: observational study on older adults living alone. JMIR Mhealth Uhealth 2019 Oct 16;7(10):e14149 [FREE Full text] [doi: 10.2196/14149] [Medline: 31621642]
41. Raudenbush SW, Bryk AS. Hierarchical Linear Models: Applications and Data Analysis Methods. Thousand Oaks, CA: Sage Publications; 200.
42. Hecht M, Zitzmann S. Sample size recommendations for continuous-time models: compensating shorter time series with larger numbers of persons and vice versa. Structural Equation Modeling: A Multidisciplinary Journal 2020 Jul 24;28(2):229-236. [doi: 10.1080/10705511.2020.1779069]
43. Martens MJ, Logan BR. A unified approach to sample size and power determination for testing parameters in generalized linear and time-to-event regression models. Stat Med 2021 Feb 28;40(5):1121-1132 [FREE Full text] [doi: 10.1002/sim.8823] [Medline: 33210321]
44. Graham JW. Missing data analysis: making it work in the real world. Annu Rev Psychol 2009 Jan;60(1):549-576. [doi: 10.1146/annurev.psych.58.110405.085530] [Medline: 18652544]
45. Robillard R, Naismith SL, Smith KL, Rogers NL, White D, Terpening Z, et al. Sleep-wake cycle in young and older persons with a lifetime history of mood disorders. PLoS One 2014 Feb 25;9(2):e87763 [FREE Full text] [doi: 10.1371/journal.pone.0087763] [Medline: 24586290]
46. Tazawa Y, Wada M, Mitsukura Y, Takamiya A, Kitazawa M, Yoshimura M, et al. Actigraphy for evaluation of mood disorders: a systematic review and meta-analysis. J Affect Disord 2019 Jun 15;253:257-269. [doi: 10.1016/j.jad.2019.04.087] [Medline: 31060012]
47. Koch C, Wilhelm M, Salzmann S, Rief W, Euteneuer F. A meta-analysis of heart rate variability in major depression. Psychol Med 2019 Jun 26;49(12):1948-1957. [doi: 10.1017/s0033291719001351]
48. Ottaviani C, Shahabi L, Tarvainen M, Cook I, Abrams M, Shapiro D. Cognitive, behavioral, and autonomic correlates of mind wandering and perseverative cognition in major depression. Front Neurosci 2014 Jan 05;8:433 [FREE Full text] [doi: 10.3389/fnins.2014.00433] [Medline: 25601824]
49. Schwerdtfeger A, Friedrich-Mai P. Social interaction moderates the relationship between depressive mood and heart rate variability: evidence from an ambulatory monitoring study. Health Psychol 2009 Jul;28(4):501-509. [doi: 10.1037/a0014664] [Medline: 19594275]

50. Bhattacharyya M, Whitehead D, Rakhit R, Steptoe A. Depressed mood, positive affect, and heart rate variability in patients with suspected coronary artery disease. Psychosom Med 2008 Nov;70(9):1020-1027. [doi: 10.1097/PSY.0b013e318189afcc] [Medline: 18941130]
51. Aïdoud A, Gana W, Poitau F, Debacq C, Leroy V, Nkodo J, et al. High Prevalence of Geriatric Conditions Among Older Adults With Cardiovascular Disease. JAHA 2023 Jan 17;12(2):e026850. [doi: 10.1161/jaha.122.026850]
52. Oh MS, Jeong MH. Sex differences in cardiovascular disease risk factors among Korean adults. Korean J Med 2020 Aug;95(4):266-275. [doi: 10.3904/kjm.2020.95.4.266]
53. Arnold AM, Psaty BM, Kuller LH, Burke GL, Manolio TA, Fried LP, et al. Incidence of cardiovascular disease in older Americans: the Cardiovascular Health Study. J Am Geriatr Soc 2005 Feb 26;53(2):211-218. [doi: 10.1111/j.1532-5415.2005.53105.x] [Medline: 15673343]
54. Gianfredi V, Blandi L, Cacitti S, Minelli M, Signorelli C, Amerio A, et al. Depression and objectively measured physical activity: a systematic review and meta-analysis. Int J Environ Res Public Health 2020 May 25;17(10):3738 [FREE Full text] [doi: 10.3390/ijerph17103738] [Medline: 32466242]
55. Rebar AL, Stanton R, Geard D, Short C, Duncan MJ, Vandelanotte C. A meta-meta-analysis of the effect of physical activity on depression and anxiety in non-clinical adult populations. Health Psychol Rev 2015 Jul 03;9(3):366-378. [doi: 10.1080/17437199.2015.1022901] [Medline: 25739893]
56. Hong S, Lee S, Song K, Kim M, Kim Y, Kim H, et al. A nurse-led mHealth intervention to alleviate depressive symptoms in older adults living alone in the community: a quasi-experimental study. Int J Nurs Stud 2023 Feb;138:104431 [FREE Full text] [doi: 10.1016/j.ijnurstu.2022.104431] [Medline: 36630872]
57. Peng R, Guo Y, Zhang C, Li X, Huang J, Chen X, et al. Internet-delivered psychological interventions for older adults with depression: a scoping review. Geriatr Nurs 2024 Jan;55:97-104. [doi: 10.1016/j.gerinurse.2023.10.026] [Medline: 37976561]
58. Richards D, Timulak L, Doherty G, Sharry J, Colla A, Joyce C, et al. Internet-delivered treatment: its potential as a low-intensity community intervention for adults with symptoms of depression: protocol for a randomized controlled trial. BMC Psychiatry 2014 May 21;14:147 [FREE Full text] [doi: 10.1186/1471-244X-14-147] [Medline: 24886179]
59. Scazufca M, Nakamura CA, Seward N, Moreno-Agostino D, van de Ven P, Hollingworth W, et al. A task-shared, collaborative care psychosocial intervention for improving depressive symptomatology among older adults in a socioeconomically deprived area of Brazil (PROACTIVE): a pragmatic, two-arm, parallel-group, cluster-randomised controlled trial. Lancet Healthy Longev 2022 Oct;3(10):e690-e702 [FREE Full text] [doi: 10.1016/S2666-7568(22)00194-5] [Medline: 36202129]
60. Seifert A, Cotten SR, Xie B. A double burden of exclusion? Digital and social exclusion of older adults in times of COVID-19. J Gerontol B Psychol Sci Soc Sci 2021 Feb 17;76(3):e99-e103 [FREE Full text] [doi: 10.1093/geronb/gbaa098] [Medline: 32672332]

Abbreviations

GDS: Geriatric Depression Scale
HF: high frequency
HRV: heart rate variability
LF: low frequency
MLM: multilevel modeling
OECD: Organisation for Economic Co-operation and Development
PA: physical activity
PHQ-9: 9-item Patient Health Questionnaire
PSQI: Pittsburg Sleep Quality Index
SUS: System Usability Scale

Edited by L Buis; submitted 29.12.23; peer-reviewed by E Sheldrake; comments to author 15.03.24; revised version received 03.05.24; accepted 23.05.24; published 17.06.24.

Please cite as:
Song S, Seo Y, Hwang S, Kim HY, Kim J
Digital Phenotyping of Geriatric Depression Using a Community-Based Digital Mental Health Monitoring Platform for Socially Vulnerable Older Adults and Their Community Caregivers: 6-Week Living Lab Single-Arm Pilot Study
JMIR Mhealth Uhealth 2024;12:e55842
URL: https://mhealth.jmir.org/2024/1/e55842
doi:10.2196/55842
PMID:38885033

Original Paper

Auxiliary Diagnosis of Children With Attention-Deficit/Hyperactivity Disorder Using Eye-Tracking and Digital Biomarkers: Case-Control Study

Zhongling Liu[1*], BMed; Jinkai Li[2*], BEng; Yuanyuan Zhang[1], PhD; Dan Wu[1], PhD; Yanyan Huo[1], MM; Jianxin Yang[2], MEng; Musen Zhang[2], MEng; Chuanfei Dong[1], MM; Luhui Jiang[3]; Ruohan Sun[3]; Ruoyin Zhou[3]; Fei Li[4], Prof Dr Med; Xiaodan Yu[5], Prof Dr Med; Daqian Zhu[6], Prof Dr Med; Yao Guo[2], Prof Dr; Jinjin Chen[1], Prof Dr Med

[1]Child Health Care Medical Division, Shanghai Children's Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, China

[2]Institute of Medical Robotics, School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai, China

[3]School of Medicine, Shanghai Jiao Tong University, Shanghai, China

[4]Department of Developmental and Behavioural Pediatric & Child Primary Care, Xinhua Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, China

[5]Department of Developmental and Behavioral Pediatrics, Shanghai Children's Medical Center, School of Medicine, Shanghai Jiao Tong University, Shanghai, China

[6]Department of Psychiatry, Children's Hospital of Fudan University, Shanghai, China

[*]these authors contributed equally

Corresponding Author:
Jinjin Chen, Prof Dr Med
Child Health Care Medical Division
Shanghai Children's Hospital
School of Medicine, Shanghai Jiao Tong University
355 Luding Road
Putuo District
Shanghai, 200062
China
Phone: 86 52976152
Email: chenjj@shchildren.com.cn

Abstract

Background: Attention-deficit/hyperactivity disorder (ADHD) is a common neurodevelopmental disorder in school-aged children. The lack of objective biomarkers for ADHD often results in missed diagnoses or misdiagnoses, which lead to inappropriate or delayed interventions. Eye-tracking technology provides an objective method to assess children's neuropsychological behavior.

Objective: The aim of this study was to develop an objective and reliable auxiliary diagnostic system for ADHD using eye-tracking technology. This system would be valuable for screening for ADHD in schools and communities and may help identify objective biomarkers for the clinical diagnosis of ADHD.

Methods: We conducted a case-control study of children with ADHD and typically developing (TD) children. We designed an eye-tracking assessment paradigm based on the core cognitive deficits of ADHD and extracted various digital biomarkers that represented participant behaviors. These biomarkers and developmental patterns were compared between the ADHD and TD groups. Machine learning (ML) was implemented to validate the ability of the extracted eye-tracking biomarkers to predict ADHD. The performance of the ML models was evaluated using 5-fold cross-validation.

Results: We recruited 216 participants, of whom 94 (43.5%) were children with ADHD and 122 (56.5%) were TD children. The ADHD group showed significantly poorer performance (for accuracy and completion time) than the TD group in the prosaccade, antisaccade, and delayed saccade tasks. In addition, there were substantial group differences in digital biomarkers, such as pupil diameter fluctuation, regularity of gaze trajectory, and fixations on unrelated areas. Although the accuracy and task completion speed of the ADHD group increased over time, their eye-movement patterns remained irregular. The TD group with children aged 5 to 6 years outperformed the ADHD group with children aged 9 to 10 years, and this difference remained relatively stable over time, which indicated that the ADHD group followed a unique developmental pattern. The ML model was effective in discriminating the groups, achieving an area under the curve of 0.965 and an accuracy of 0.908.

Conclusions: The eye-tracking biomarkers proposed in this study effectively identified differences in various aspects of eye-movement patterns between the ADHD and TD groups. In addition, the ML model constructed using these digital biomarkers achieved high accuracy and reliability in identifying ADHD. Our system can facilitate early screening for ADHD in schools and communities and provide clinicians with objective biomarkers as a reference.

(JMIR Mhealth Uhealth 2024;12:e58927) doi:10.2196/58927

KEYWORDS

attention deficit disorder with hyperactivity; eye-tracking; auxiliary diagnosis; digital biomarker; antisaccade; machine learning

Introduction

Background

Attention-deficit/hyperactivity disorder (ADHD) is a common neurodevelopmental disorder in school-aged children, characterized by deficits in attention, hyperactivity, and impulsivity. Globally, the estimated prevalence of ADHD in children and adolescents is approximately 5.29% [1]; in China, the prevalence is approximately 6.4% [2]. People with ADHD typically exhibit deficiencies in various cognitive domains, and these symptoms can persist into adolescence and adulthood, which can result in academic underachievement and societal issues, such as substance abuse and violence [3]. Therefore, early identification, diagnosis, and intervention for ADHD are essential.

Despite recent advances, the diagnosis of ADHD relies heavily on subjective judgments based on the observations of children's behavior. Consequently, this can lead to both over and underdiagnosis, as well as inappropriate treatments. Therefore, there is an urgent need to develop methods to identify reliable ADHD biomarkers. Furthermore, given that poor academic performance is the most common concern of individuals with ADHD, it is crucial that we improve awareness and understanding of ADHD among parents and teachers to ensure timely identification of ADHD. However, on the one hand, most nonmedical professionals cannot be expected to gain specialized medical expertise, and on the other hand, physicians cannot frequently visit campuses to aid in ADHD assessment. This situation has resulted in delays in diagnosing children with ADHD. Developing mobile screening equipment will enable on-campus ADHD screening to facilitate timely identification and diagnosis of ADHD.

Eye-tracking technology is particularly suited for the assessment and diagnosis of ADHD because it offers an objective measurement of children's neuropsychological behavior. Studies have shown that there is a significant overlap between the neural networks responsible for attention and those responsible for eye-movement control [4]. Children with ADHD experience difficulties with spatial perception and visual-motor integration [5], and these neurophysiological features associated with ADHD can be identified using eye-tracking assessments. In addition, children with ADHD often find lengthy and complex assessments challenging, particularly if they are required to wear additional equipment. Eye-tracking technology surpasses other neurophysiological techniques in its ability to record the neuropsychological activity of participants in a more natural setting [6]. This leads to better cooperation of children during assessments and higher reliability and generalizability of results.

Recent advances in computational psychiatry have enabled the extraction of eye-tracking metrics to discern behavioral alterations in children with ADHD [7-9]. These metrics encompass various aspects of visual attention, such as fixation duration, saccade velocity, and gaze entropy [10-12], which may serve as digital biomarkers for neurodevelopmental disorders [13,14]. By analyzing the temporal and spatial characteristics of eye movements, computational models can capture differences in visual behaviors between ADHD and typically developing (TD) children. Machine learning (ML) techniques have emerged as powerful tools for processing and interpreting large amounts of eye-tracking data [15-17]. Training ML models on labeled eye-tracking metrics has allowed the construction of robust and accurate classifiers to identify whether individuals belong to an ADHD or a TD group. Precise eye-tracking measurements and digital biomarkers hold great promise as objective and automated screening tools for ADHD, which will facilitate the development of early intervention strategies and improve the clinical outcomes of affected children [7,18,19]. Moreover, the evolution of mobile eye-tracking technology and devices, coupled with portable computing sources, such as smartphones and tablets, will allow the implementation of eye-tracking assessments in various scenarios and thus address the need for ADHD screening in the community [20-22].

Related Work

Neuroimaging studies have shown that children with ADHD have multidimensional brain function abnormalities. The impairment of inhibitory control is a fundamental factor contributing to cognitive and executive functioning deficiencies in individuals with ADHD [23]. However, these individuals also have motor coordination difficulties, poorer spatial perception [24-26], reduced auditory sensitivity, and problems with attentional integration of audiovisual stimuli [27].

Recently, there has been a growing interest in exploring the use of eye-tracking technology to study the neurophysiological features of ADHD. A meta-analysis of the various behavioral tests developed over the last 5 decades to evaluate eye movement and cognitive control [28] revealed that eye-tracking evaluations of children with ADHD yielded the most reliable and consistent outcomes when eliminating bias. Most of these tests focused on saccade, which is one of the most crucial type of eye movement. Children with ADHD perform significantly worse than TD children across all tasks, with greater variability for each metric in the antisaccade task [29].

To ensure that the screening method is appropriate for children with ADHD, we must use a paradigm that is brief and simple

to perform yet capable of highlighting cognitive deficits. In addition, the extracted eye-movement metrics should be able to comprehensively characterize children's task performance. Several recent studies have used eye tracking to explore the characteristics of ADHD. Lemel et al [30] incorporated spoken-word recognition accuracy, gaze duration, and the number of transitions in response to a phonological competitor to analyze spoken-word processing in adverse listening conditions in individuals with ADHD. However, this paradigm was complex and required word recognition and was thus more suited to adult patients. Another study used a paradigm to assess children's working memory; however, the task took 30 minutes to complete [31], which is not conducive to task completion in children with ADHD. Siqueiros et al [32] used the antisaccade task, which is a simple and reliable paradigm that suits children. However, only directional errors and expected eye movements were assessed; moreover, the paradigm was not sufficiently comprehensive to assess children's task performance.

Objectives

Studies conducted to date have provided valuable insight into automatic screening approaches for ADHD in children using eye-tracking devices. However, these studies have drawbacks that have hindered the development of a more robust and accurate auxiliary diagnostic system. For example, the paradigms were too time-consuming or complex for clinical ADHD screening, and the extracted metrics were not sufficiently comprehensive. ML models used in previous studies have typically achieved only modest accuracy and sensitivity, which limits clinical applicability. Furthermore, small sample sizes have limited the robustness of the results.

To address the aforementioned challenges, we aimed to develop an accurate and reliable auxiliary diagnostic system for ADHD in children using eye-tracking technology. Specifically, the objectives of this study were as follows:

1. To design an eye-tracking assessment paradigm that is easy to implement and can identify differences in eye-movement patterns between children with ADHD and TD children.
2. To extract effective eye-tracking metrics as digital biomarkers that quantitatively represent various aspects of eye-movement behaviors and use these biomarkers to construct and validate ML models to enable automatic screening of children for ADHD.
3. To achieve high accuracy and reliability of the ML model using a large dataset, which will facilitate early screening for ADHD and timely intervention for children with ADHD and thus contribute to improving the effectiveness of the health care system.

Methods

Participants

To ensure the representativeness of the ADHD and TD groups in this case-control study, we recruited participants from hospitals and schools separately. Children with ADHD were recruited from an outpatient clinic at a public pediatric hospital in Shanghai, China, whereas TD children were recruited from 2 general public elementary schools in Shanghai (one from an urban area and another from a suburban area). The children were divided into 3 age groups: group 1 (5-6 years), group 2 (7-8 years), and group 3 (9-10 years).

The inclusion criteria for the ADHD group were children in grades 1 to 3 with a clinical diagnosis of ADHD who were not currently receiving treatment. The inclusion criteria for the TD group were children in grades 1 to 3 with a negative assessment on the Swanson, Nolan, and Pelham Rating Scale (SNAP-IV) [33].

The exclusion criteria were children with a full-scale score of <75 on the Wechsler Intelligence Scale for Children; children who had a history of severe traumatic brain injury, neurological disorders, severe physical illnesses, and psychiatric illnesses (eg, mood disorders and schizophrenia); and those unable to undergo eye-tracking examinations.

From December 2022 to April 2023, a total of 100 children with a clinical diagnosis of ADHD were recruited. Of these, 4 participants with a history of severe traumatic brain injury, neurological disorders, and other severe physical and psychiatric disorders and 2 participants who were unable to tolerate the eye-tracking assessment were excluded. This resulted in 94 participants in the ADHD group.

A total of 150 children were randomly selected as the TD group. Of these, 15 children refused to participate in the program. In addition, 2 children with a history of severe traumatic brain injury, neurological disorders, and other severe physical and psychiatric disorders and 11 children who were considered to have ADHD after the interviews and evaluations were excluded. Finally, 122 children were included in the study as the TD control group.

All personnel involved in administering the assessments in this study were full-time child health practitioners who had been working in child health care for more than 3 years. Standardized survey administration training was provided before the tests were administered.

Ethical Considerations

Before the assessment began, the purpose of the project was explained to the children and their guardians, and written informed consent was obtained from the guardians. All participants could withdraw at any stage of the study. Interviews were then conducted with the guardians to gather data on the basic conditions of the children. Children who fulfilled the inclusion and exclusion criteria were formally enrolled in the study and underwent the SNAP-IV and eye-tracking assessments. All data will be stored in a deidentified form. No participants will receive any benefit from participating in this study, but they will receive a booklet reporting the results of the assessments involved in this study as a souvenir.

The study protocol and informed consent form were approved by the Shanghai Children's Hospital Institutional Review Board (2022R126-F01).

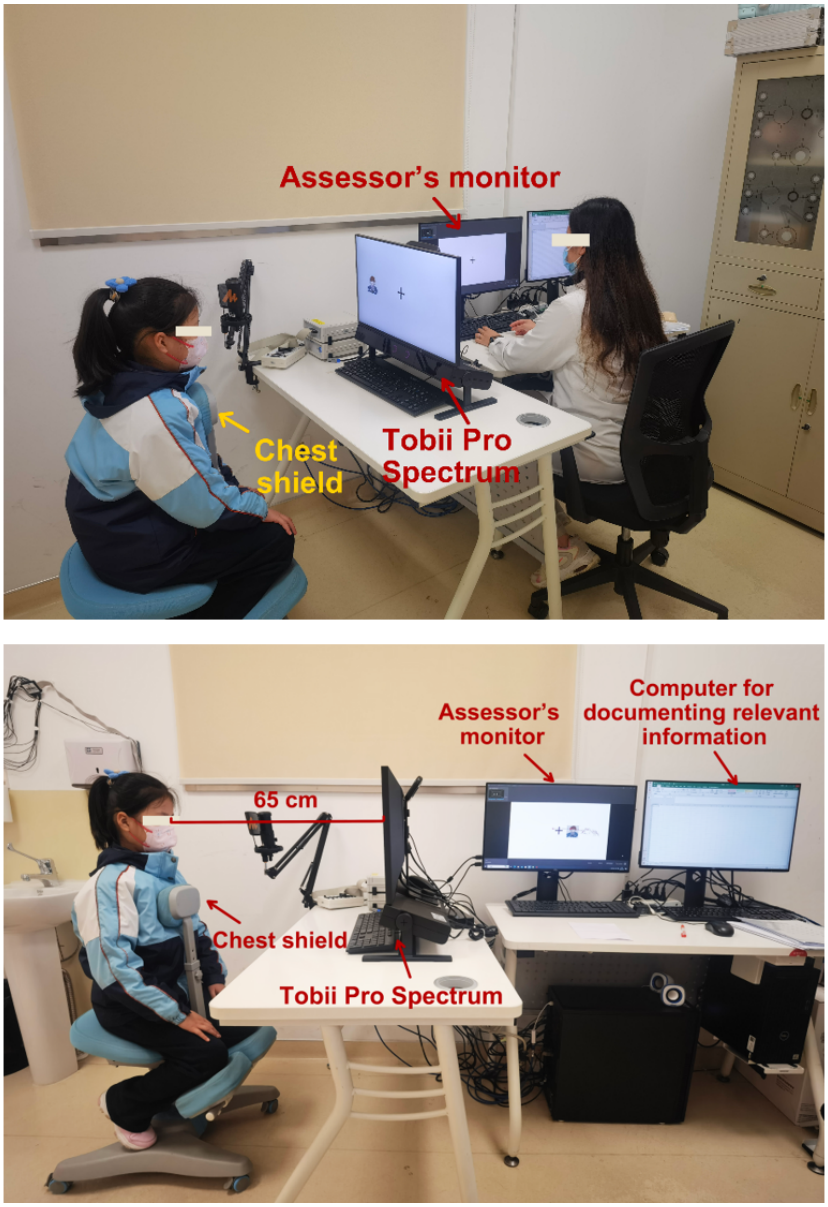
Paradigm Design

Overview

Eye movements were recorded at a sampling rate of 1200 Hz using the Tobii Pro Spectrum eye tracker (Tobii Pro AB), a screen-based eye tracker that captures eye movements and pupillary responses. Visual stimuli were presented at a screen response rate of <5 milliseconds on a 24-inch monitor with a resolution of 1920×1080 pixels (16:9 ratio). The Tobii Pro Lab software (version 1.194; Tobii Pro AB) was used to set up the experiment.

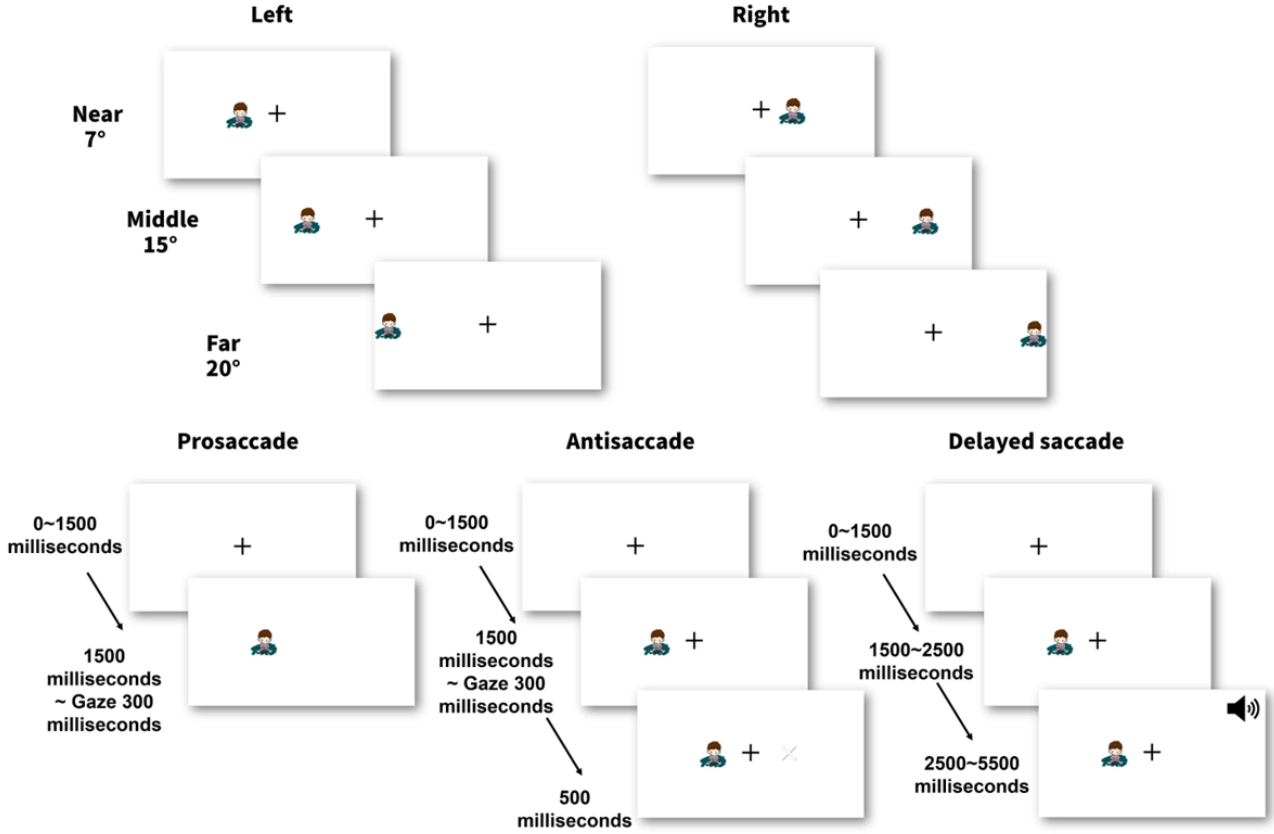
The assessment procedure was performed in a quiet room with only 1 overhead light source (Figure 1). Participants were seated in a special seat with a chest shield to limit upper body movement and help stabilize the head. The cushion was adjusted to ensure that the center of the screen was at the same level as the participant's head. The participant was seated in a position in which they were unable to observe the assessor's screen or operations to minimize distractions. Participants maintained a distance of 65 cm from the screen and began the formal assessment following a 5-point calibration. Before each task, a prompt screen appeared, and the assessor provided detailed instructions to ensure that the participant fully understood the task content before proceeding with formal testing.

Figure 1. Eye-tracking assessment scenario settings.



During the assessment, participants were asked to complete 3 saccade tasks sequentially (Figure 2): prosaccade, antisaccade, and delayed saccade. The stimulus was 5 cm high and 5 cm wide and randomly appeared on the left or right side of the screen. There was a central fixation cross in the middle of the screen, and the stimuli were set at 7°, 15°, and 20° away from the central cross for the different eccentricities. For each trial, a stimulus would randomly appear twice at one of the aforementioned 6 positions.

Figure 2. The eye-tracking assessment paradigm.



Prosaccade Task

Prosaccade, also known as reflexive saccade or visually guided saccade, is an abrupt eye movement triggered by the sudden appearance of a stimulus [34]. It is primarily induced by exogenous stimuli and serves as a baseline measure. In the prosaccade task, participants were instructed to initially fixate on the central fixation cross. After 1500 milliseconds, a stimulus appeared randomly in one of the aforementioned 6 positions. Participants were required to quickly shift their gaze toward the stimulus. Once participants fixated on the stimulus area (SA) for more than 300 milliseconds, the next trial was started automatically.

Antisaccade Task

In the antisaccade task, participants were required to first fixate on the central fixation cross. After 1500 milliseconds, 1 stimulus appeared randomly in one of the 6 aforementioned positions. Participants were required to quickly shift their gaze to the target area (TA), which was the location symmetrically opposite to the stimulus relative to the central fixation cross. Upon maintaining fixation at the TA for more than 300 milliseconds, a white feedback cross automatically appeared at the TA position to indicate success before proceeding to the next trial. If the participant decided to abandon the trial, the assessor pressed the space bar to skip the trial, and a white cross was displayed at the TA before moving on to the next trial. Previous studies have used a paradigm in which the central fixation cross disappears when the stimulus is presented [28]. However, this can make accurately localizing the TA more challenging, which may result in children being unable to complete the task. Therefore, in this study, the central cross was retained to assist participants in locating the TA.

Delayed Saccade Task

The delayed saccade task, based on the go–no-go paradigm [35], was adapted to the cognitive abilities of children with ADHD. This task not only directly assesses inhibition but also requires participants to combine auditory discrimination and visuomotor modulation. Thus, the task assesses the multisensory integration and coordination capacity of individuals with ADHD. During the task, participants were instructed to fixate on the central fixation cross. After 1500 milliseconds, 1 stimulus appeared randomly in one of the 6 aforementioned positions. Participants were asked to maintain fixation on the central cross until they heard a sound cue after 1000 milliseconds, after which they were required to shift their gaze toward the SA as fast as possible. Then, after another 3000 milliseconds, the next trial was started automatically.

For each saccade task, there were 12 formal trials (2 trials for each position). Before the formal test, practice trials were provided, where stimuli were presented randomly in the 6 positions, to allow participants to familiarize themselves with the task.

Area of Interest Division Across Tasks

To quantify the eye movements made during the different tasks, we divided the area viewed by participants into different areas (Figure 3): the TA, the SA, the center area (CA), the unrelated area (UA), the proper-side area (PSA), and the wrong-side area (WSA). The TA represented the area that participants were

required to fixate on, and the SA represented the area of the stimulus. For the delayed saccade task, we further divided TA into TA during the proper period (TA-P) and TA during the wrong period (TA-W) to represent the TA area in the proper or wrong time periods, respectively (Figure 4). The TA and SA were the same in the prosaccade and delayed saccade tasks, whereas in the antisaccade task, they were horizontally symmetrical. The CA represented a 5 cm × 5 cm area around the central fixation cross. The UA was unrelated to the task requirements and expected to attract minimal attention during the tasks. The PSA and WSA were defined for the antisaccade task only and represented the proper and wrong areas, respectively, besides the CA.

Figure 3. Illustration of the division of areas for extracting area-based eye-tracking metrics. CA: center area; PSA: proper-side area; SA: stimulus area; TA: target area; UA: unrelated area; WSA: wrong-side area.

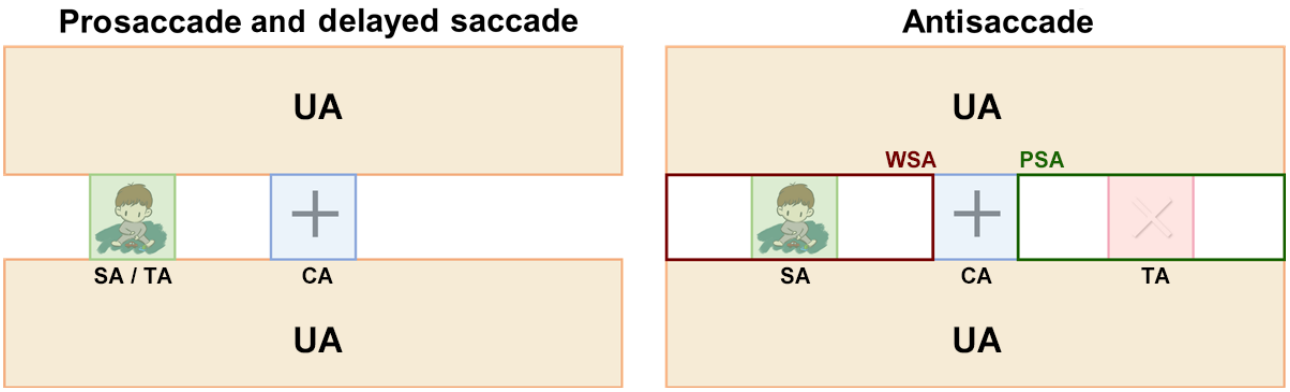
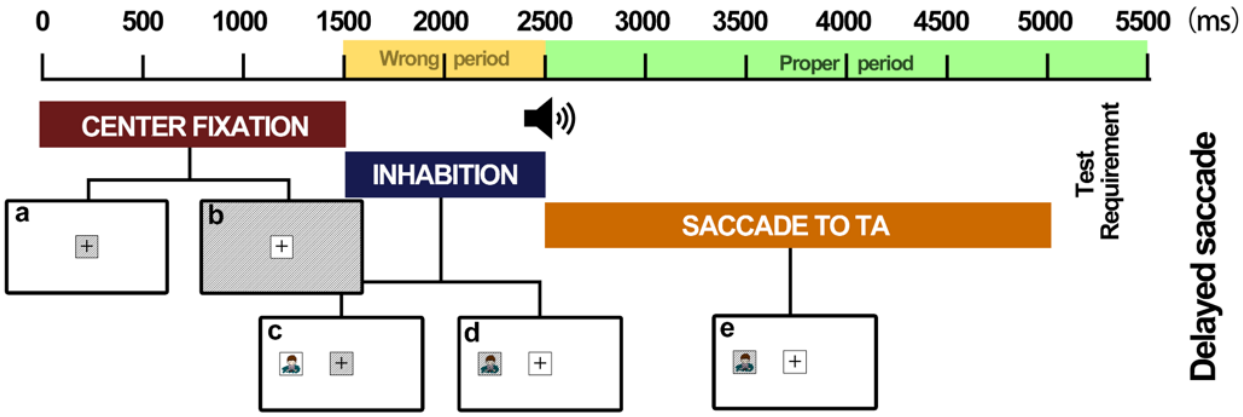


Figure 4. The different completion statuses in the delayed saccade task. From 0 to 1500 milliseconds, participants were asked to gaze at the center area (shaded area in a). If fixation fell into the shaded area in b, this indicated the occurrence of an intrusive saccade. From 1500 to 2500 milliseconds, participants were asked to maintain their fixation on the center area (shaded area in c) until they heard the cue. Thus, if fixation fell into the shaded area in d during this period, this was defined as a target area during the wrong period fixation (ie, saccade to the target area (TA) but during the wrong period). At 2500 milliseconds, the sound cue was presented, and participants were required to fixate on the TA (shaded area in e) as fast as possible. Fixation on the shaded area after 2500 milliseconds was defined as a target area during the proper period fixation (ie, saccade to the TA during the proper period).



Extraction of Digital Biomarkers

Overview

On the basis of the eye-tracking paradigm, we calculated 28 digital biomarkers from the raw data recorded by the eye tracker. These biomarkers quantitatively reflect various behaviors of participants during the task, which were divided into 5 categories: general metrics (8/28, 29%), pupil-based metrics (4/28, 14%), area-based metrics (11/28, 39%), search-based metrics (3/28, 11%), and entropy-based metrics (2/28, 7%). For each assessment trial, we recorded 4 trial attributes (ie, task: prosaccades, antisaccades, and delayed saccades, target side: left and right, target eccentricity: 7°, 15°, and 20°, and trial order: first and second) and 6 participant attributes (ie, name, ID, category [ADHD and TD], sex [male and female], age, and age group). Table 1 summarizes these biomarkers in terms of category, symbol, description, and task.

Table 1. Descriptions of the digital biomarkers.

Category and symbol	Description	Task
General metrics		
$N_{Fix.}$	Total number of fixations	All[a]
$N_{Sac.}$	Total number of saccades	All
T_{Total}	Total duration of the trial	All
$T_{Fix.\ Avg.}$	Average fixation duration	All
$T_{Sac.\ Avg.}$	Average saccade duration	All
$V_{Sac.\ Avg.}$	Average saccade velocity	All
$V_{Sac.\ Peak}$	Peak value of saccade velocity	All
$A_{Sac.\ Avg.}$	Average saccade amplitude	All
Pupil-based metrics		
$D_{Pupil\ Avg.}$	Average pupil diameter	All
$D_{Pupil\ Max.}$	Maximum pupil diameter	All
$D_{Pupil\ Min.}$	Minimum pupil diameter	All
$D_{Pupil\ Sd.}$	SD of pupil diameter	All
Area-based metrics		
$B_{TA\ Fix.}$	Boolean value to signify the occurrence of fixations in the TA[b] (TA-P[c] for the delayed saccade task)	All
$L_{TA\ Fix.}$	Fixation latency of the TA (TA-P for the delayed saccade task)	All
$N_{UA\ Fix.}$	Number of fixations in the UA[d]	P[e] and A[f]
$N_{TA\ Fix.}$	Number of fixations in the TA for the whole period	D[g]
$N_{TA\text{-}P\ Fix.}$	Number of fixations in the TA for the proper period	D
$N_{TA\text{-}W\ Fix.}$	Number of fixations in the TA for the wrong period	D
$N_{SA\ Fix.}$	Number of fixations in the SA[h]	A
$B_{PSA\ Fix.}$	Boolean value to signify the occurrence of fixations in the PSA[i]	A
$B_{WSA\ Fix.}$	Boolean value to signify the occurrence of fixations in the WSA[j]	A
$B_{PSA\ Fix.\ 1st}$	Boolean value to signify if the first fixation located in the PSA	A
$B_{Intrusive\ Sac.}$	Boolean value to signify the occurrence of intrusive saccade during the center fixation period	D
Search-based metrics		
B_{Search}	Boolean value to signify the occurrence of the search behavior	A
N_{Search}	Number of search behavior occurrences	A
T_{Search}	Total duration of search behavior	A
Entropy-based metrics		
SGE_{norm}	Normalized stationary gaze entropy	All
GTE_{norm}	Normalized gaze transition entropy	All

[a]All: all tasks, including prosaccade, antisaccade, and delayed saccade tasks.

[b]TA: target area.

[c]TA-P: target area during the proper period in the delayed saccade task.

[d]UA: unrelated area.

[e]P: prosaccade task.

[f]A: antisaccade task.

[g]D: delayed saccade task.

[h]SA: stimulus area.

[i]PSA: proper-side area.

[j]WSA: wrong-side area.

General Metrics

Human eye-movement patterns can be divided into fixations, saccades, and pursuits [36], of which the former 2 patterns are the focus of our paradigm. Using the Tobii Pro Lab software, we extracted the fixations and saccades of participants in chronological order from the raw gaze data. Subsequently, we calculated the total number of fixations ($N_{Fix.}$) and saccades ($N_{Sac.}$) and their average durations ($T_{Fix.\,Avg.}$ and $T_{Sac.\,Avg.}$), which reflects participants' holistic visual behavior. The velocity and amplitude of saccades were automatically recorded by the software. We calculated the average and peak saccade velocity ($V_{Sac.\,Avg.}$ and $V_{Sac.\,Peak}$) and the average saccade amplitude ($A_{Sac.\,Avg.}$) for each trial. These values reflect the scanning and information retrieval process, respectively. In addition, the total time taken for each trial (T_{Total}) was recorded.

Pupil-Based Metrics

Pupil size is a crucial physiological measure that reflects autonomic nervous system activity, cognitive load, and emotional arousal. It has been applied extensively to various research fields [37-40]. The eye tracker continuously recorded participants' pupil diameter during each trial. We preprocessed the raw data and extracted pupil-based metrics following 5 steps (Textbox 1) [41].

Textbox 1. Preprocessed raw data and extracted pupil-based metrics.

Step 1: We removed samples labeled by the eye tracker as "invalid" and pupil diameters that fell outside the feasible range of 1.5 to 9.0 mm.

Step 2: We calculated pupil dilation speed to remove samples with a disproportionately large change in pupil size, which was usually caused by blinks or system errors. Because of the inconsistent sampling intervals, pupil diameter changes were not directly comparable between adjacent samples. Therefore, we calculated the normalized dilation speed between samples using the formula:

$$s_i = \max\left(\left| \frac{(p_i - p_{i-1})}{(t_i - t_{i-1})} \right|, \left| \frac{(p_{i+1} - p_i)}{(t_{i+1} - t_i)} \right| \right), \textbf{(1)}$$

where p_i and t_i are the pupil diameter sequence and timestamp sequence, respectively. To detect outliers in the dilation speed sequence (s_i), we calculated the threshold, T, using the median absolute deviation (MAD):

$$MAD = \text{median}\left(\left| s_i - \text{median}\left(s_i \right) \right| \right), \textbf{(2)}$$

$$T = \text{median}\left(s_i \right) + n \cdot MAD, \textbf{(3)}$$

where the scalar n was chosen as 1.5. Samples with an s_i larger than T were removed as outliers. Because the eye tracker simultaneously collected data from both the left and right pupils, we performed steps 1 and 2 for each pupil separately.

Step 3: We excluded samples in which data of 1 pupil was missing and calculated the mean data sequence of the left and right pupil diameters.

Step 4: Because of nonuniform sampling and the presence of noise, we used a size 20 sliding window to resample and smooth the data sequence at 500 Hz. This involved an exponential moving average based on the timestamp and skipped data gaps ≥50 milliseconds.

Step 5: Following the above preprocessing steps, we obtained a valid, uniform, and smooth sequence of pupil diameter data. We then calculated the average ($D_{Pupil\,Avg.}$), maximum ($D_{Pupil\,Max.}$), minimum ($D_{Pupil\,Min.}$), and SD ($D_{Pupil\,Sd.}$) pupil diameter values of the sequence for each trial, which reflect various aspects of the pupil state of participants.

Area-Based Metrics

We extracted a range of metrics according to the area of interest (AOI) divisions. A Boolean value for fixation incidence ($B_{TA\,Fix.}$) was recorded to signify the completion of the task by detecting whether the TA (or TA-P for the delayed saccade task) contained any fixations. The latency of the first fixation in the TA (or TA-P) was recorded as the fixation latency ($L_{TA\,Fix.}$). The number of fixations was counted for the SA (only in the antisaccade task), UA (in the prosaccade and antisaccade tasks), TA-P (only in the delayed saccade task), and TA-W (only in the delayed saccade task), which were denoted as $N_{SA\,Fix.}$, $N_{UA\,Fix.}$, $N_{TA\text{-}P\,Fix.}$, and $N_{TA\text{-}W\,Fix.}$, respectively. For the delayed saccade task, fixations outside of the CA during the center fixation period were defined as intrusive saccades and thus recorded as a Boolean value ($B_{Intrusive\,Sac.}$). For the antisaccade task, if fixations were detected in the PSA ($B_{PSA\,Fix.}$) or WSA ($B_{WSA\,Fix.}$), these were recorded as Boolean values. We also used a Boolean metric to signify that the first fixation that occurred after the stimulus appeared was located in the PSA ($B_{PSA\,Fix.\,1st}$).

Search-Based Metrics

During the antisaccade task, participants may have had difficulty determining the correct fixation position, which may have led to a series of consecutive fixations around the TA before finally reaching the TA. In practice, we detected fixations in the surrounding area outside the TA and within a distance of $1.5 \cdot L_{TA}$ from the TA center, where L_{TA} is the length of the TA edge. Therefore, the consecutive sequences of ≥2 detected fixations were extracted as search behaviors. For each antisaccade trial, we recorded the following search-based metrics: the occurrence of search behaviors (B_{Search}), the number of search behaviors (N_{Search}), and their total duration (T_{Search}).

Successful antisaccade trials required both a reversed saccade as well as an accurate landing position. Therefore, these metrics based on search behavior represent participants' vision control and distance perception abilities.

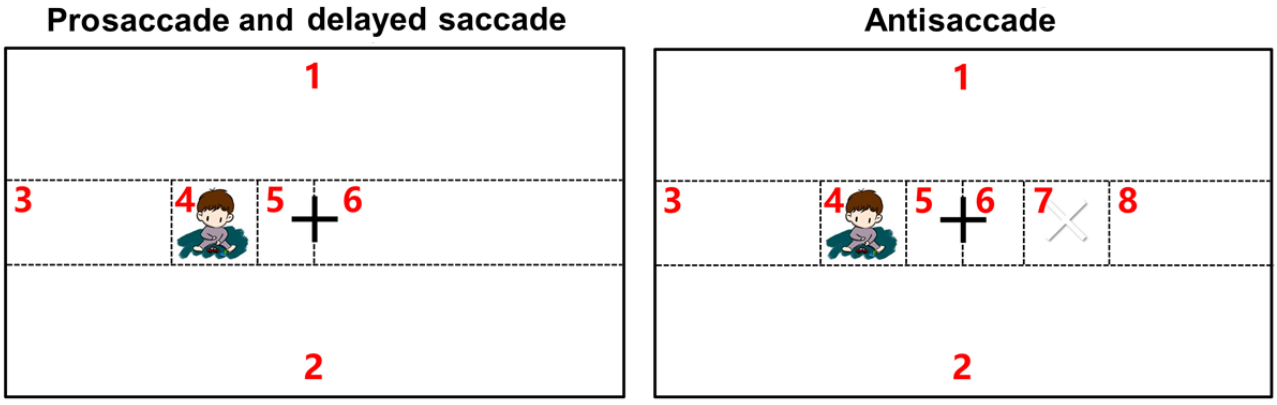
Entropy-Based Metrics

Entropy in information theory [42] suggests that gaze entropy reflects the degree of uncertainty or predictability exhibited by the human eye during visual exploration. Thus, gaze entropy can provide valuable insight into the cognitive processes involved in visual perception and attention. There are 2 types of gaze entropy: stationary gaze entropy (SGE) and gaze transition entropy (GTE) [43]. SGE evaluates the spatial distribution of fixations, with a higher value indicating a more dispersed eye-movement pattern [44]. GTE focuses on the randomness of eye movements between fixations and reflects the flexibility and complexity of the scanning pattern.

As shown in Figure 5, the images were divided into n different areas, which served as the individual state spaces of a discrete system. We calculated the proportion of fixations located in each area, denoted as p_i for the i-th area, which formed the approximate probability distribution of the states [45,46]. On the basis of the entropy equation by Shannon [42], SGE was calculated as follows:

$$SGE = -\text{sum}_i\,(\,p_i \cdot \log_2 p_i\,). \textbf{ (4)}$$

Figure 5. Division of areas for the calculation of gaze entropy metrics. It should be noted that the areas here are different from those for the area-based metrics shown in Figure 3.



Applying the first-order Markov transition matrix [47], we derived $p(j|i)$ from the fixation sequence, which represented the conditional probability of a gaze transitioning from the i-th to the j-th area. Then, GTE was computed based on the conditional entropy equation [47,48] as follows:

$$GTE = -\text{sum}_i\,(\,p_i \cdot \text{sum}_j\,(\,p(j|i) \cdot \log_2 p(j|i)\,)\,). \textbf{ (5)}$$

The maximum entropy of a system is determined by the number of available state spaces, which occurs when they are equally distributed [49]. To enable a comparison between different tasks, we used the corresponding maximum value, $H_{max} = \log_2 n$, to normalize the computed SGE and GTE into a range from 0 to 1:

$$SGE_{norm} = SGE / \log_2 n, \textbf{ (6)}$$

$$GTE_{norm} = GTE / \log_2 n. \textbf{ (7)}$$

As introduced earlier, n represents the number of areas, where n=6 for the prosaccade and delayed saccade tasks, and n=8 for the antisaccade task.

Statistical Analysis

We reviewed and uniformly numbered basic information and scale data. After eliminating data with incomplete information, data were entered in duplicate using the Chinese version of EpiData 3.1 (The EpiData Association), and Excel (version 2019; Microsoft Corp) was used to clean and organize the data.

The Tobii Pro Lab software was used to analyze basic eye-movement metrics and export data. Participants with >80% valid data were included in the analysis. Python (version 3.8) was used to extract the eye-tracking metrics.

All data were tested for normality and homogeneity of variance. Samples conforming to a normal or approximately normal distribution are represented as means and SDs, and nonnormally distributed data are described as means and 95% CIs. Count data are expressed as n (%), and differences between groups were calculated using the chi-square test. For visual harmonization, 4 valid digits were retained for the eye-tracking metrics. We used independent samples 2-tailed t tests to compare normally distributed data between the 2 groups. To compare nonnormally distributed data between the 2 groups, we used the Wilcoxon Mann-Whitney U test, and the Kruskal-Wallis test was used to compare among multiple groups. Paired comparisons for significant multiple-group comparisons were performed using the Bonferroni method. A 2-sided $P<.05$ was considered statistically significant.

ML Analysis

Overview

To validate the effectiveness of the proposed digital biomarkers, we conducted an ML analysis of the eye-tracking metrics to classify the ADHD and TD groups. First, we preprocessed the extracted metrics to meet the requirements of ML analysis and sequentially performed variable filtering, model construction, and model evaluation to verify the effectiveness of the extracted biomarkers. To ensure the reliability and generalizability of the model, we applied 5-fold cross-validation.

Data Preprocessing

The eye-tracking metrics were subdivided into multiple variables according to trial attributes (ie, task, target eccentricity, target side, and trial order). For each metric, we performed an average calculation for the target side and trial order, while maintaining different values for different task types and target eccentricities. For example, the metric $N_{Fix.}$ was obtained from the prosaccade, antisaccade, and delayed saccade tasks with 7°, 15°, and 20° target eccentricities, respectively, which were subdivided into 9 variables as follows: ${}^{P7}N_{Fix.}$, ${}^{P15}N_{Fix.}$, ${}^{P20}N_{Fix.}$, ${}^{A7}N_{Fix.}$, ${}^{A15}N_{Fix.}$, ${}^{A20}N_{Fix.}$, ${}^{D7}N_{Fix.}$, ${}^{D15}N_{Fix.}$, and ${}^{D20}N_{Fix}$. This ensured that the variability of the metrics would be reasonably preserved. The preprocessing resulted in 183 eye-tracking variables, and each participant became 1 data point for the ML analysis.

Model Construction

Before model training, we performed filtering to remove redundant variables and enhance computational efficiency. Variables that were significantly different between groups, compared using the Mann-Whitney *U* test, were retained.

To predict the categories of participants, we used the extreme gradient boosting (XGBoost) algorithm as the classification model. XGBoost is an advanced implementation of the gradient boosting decision tree framework, which sequentially builds an ensemble of decision trees to refine the prediction. The learning process minimizes the gradient of the loss function, thereby enhancing the model's performance. The XGBoost algorithm applies regularization techniques to efficiently boost the model and has thus demonstrated superior performance than the conventional gradient boosting decision tree framework in similar studies [50,51]. We implemented the XGBoost model in Python (version 3.8) using the packages *xgboost* (version 2.0.1) and *scikit-learn* (version 1.3.0). The hyperparameter settings of the model are listed in Multimedia Appendix 1, which are mainly the default values without adjustment to objectively illustrate the model's performance.

Model Evaluation

The 5-fold cross-validation method with 500 repeats was applied to evaluate classification performance. The model was trained with 173 samples and tested with 43 samples for each fold. To evaluate the models, we used the receiver operating characteristic (ROC) curve and the area under the ROC curve (AUC), which consider the trade-off between the true positive rate and false positive rate at various classification thresholds and provide a holistic assessment of the model's classification performance. We used the evaluation metrics of accuracy, sensitivity, specificity, precision, and F_1-score to quantify classification performance.

Variable Importance

When training the XGBoost model, the split gain was calculated at each node of the decision tree, which indicated the contribution of variables to the model. After the training process, the split gain was aggregated for each variable among all the decision trees to provide a comprehensive measure of the variable's relative importance in the classification of ADHD or TD groups.

Results

Characteristics of the Participants

A total of 216 participants (n=122, 56.5% in the TD group and n=94, 43.5% in the ADHD group) were enrolled in the study (Table 2). Overall, there was no significant difference in age (t_{214}=–0.30; P=.76); full-scale IQ (t_{214}=1.14; P=.25); or verbal IQ (t_{214}=0.03; P=.98) between the TD and ADHD groups. However, the ADHD group scored significantly lower than the TD group for performance IQ (t_{214}=2.08; P=.04). On the SNAP-IV, children in the TD group scored within the normal range, whereas the ADHD group scored significantly higher than the TD group on all 3 core symptoms (all P<.001).

Table 2. The basic information of the participants.

Variables	TD[a] (n=122)	ADHD[b] (n=94)	*t* test or chi-square test (*df*)[c]	*P* value
Sex, n (%)			37.28 (1)	<.001
Male	61 (50)	84 (89.4)		
Female	61 (50)	10 (10.6)		
Age (y), mean (SD)	7.18 (1.19)	7.24 (1.39)	–0.30 (214)	.76
Age group, n (%)			0.63 (2)	.73
Group 1 (5-6 y)	44 (36.1)	36 (38.3)		
Group 2 (7-8 y)	45 (36.9)	37 (39.4)		
Group 3 (9-10 y)	33 (27)	21 (22.3)		
IQ, mean (SD)				
Verbal IQ	97.36 (12.51)	97.41 (12.90)	0.03 (214)	.98
Performance IQ	103.02 (13.25)	99.01 (15.02)	2.08 (214)	.04
Full-scale IQ	100.06 (12.44)	98.11 (12.43)	1.14 (214)	.25
SNAP-IV [d], mean (SD)				
Inattentive	0.63 (0.25)	15.09 (0.75)	–199.20 (214)	<.001
Hyperactivity or impulsive	0.50 (0.30)	11.59 (0.82)	–137.43 (214)	<.001
Oppositional defiant	0.36 (0.12)	7.66 (0.56)	–118.54 (214)	<.001

[a]TD: typically developing.

[b]ADHD: attention-deficit/hyperactivity disorder.

[c]*t*-tests were used for variables presenting means and standard deviations (Age, IQ, and SNAP-IV scores), and chi-square tests were used for variables presenting numbers and percentages (Sex and Age group).

[d]SNAP-IV: Swanson, Nolan, and Pelham Rating Scale.

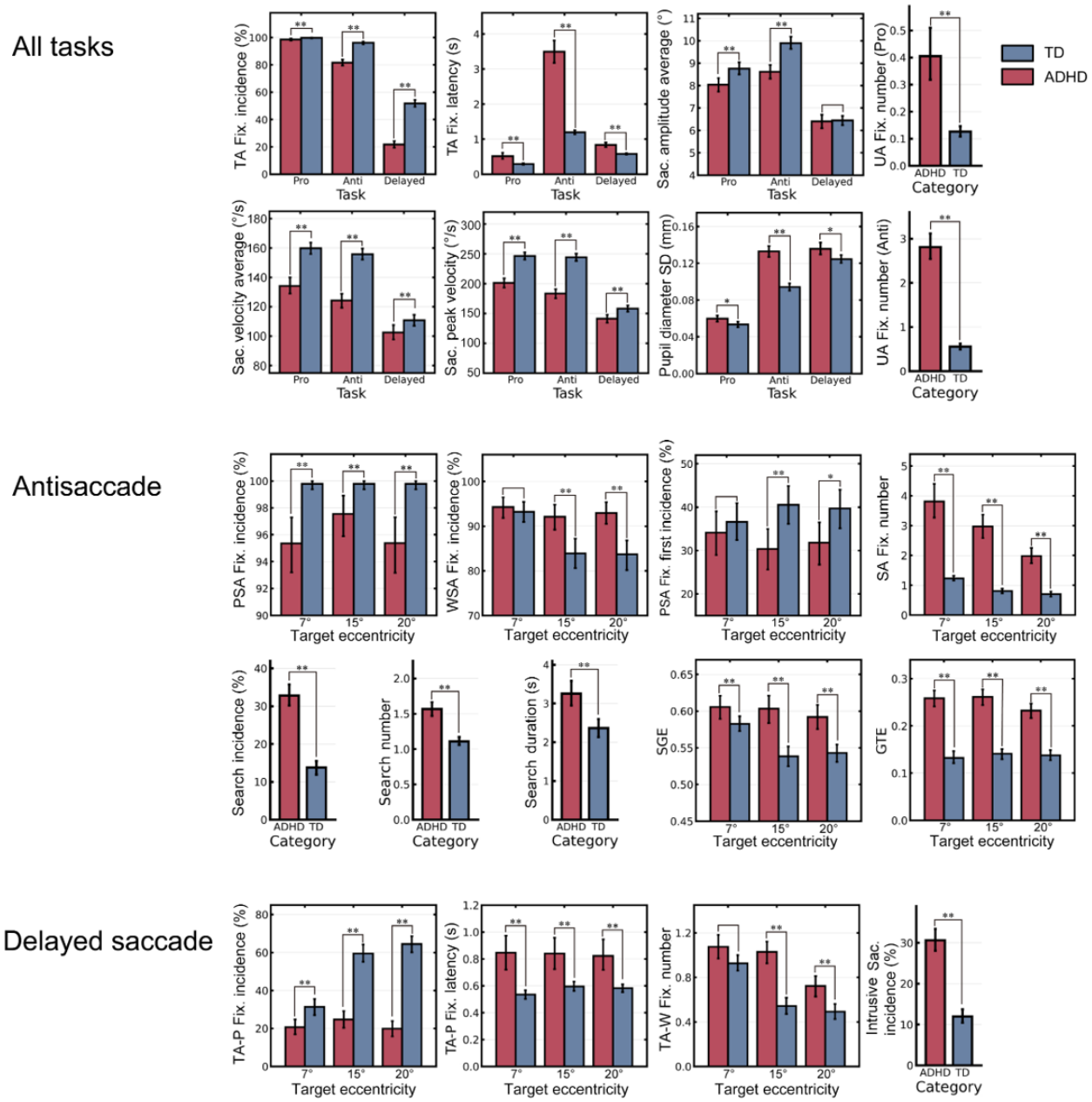
Comparison of Digital Biomarkers Between the ADHD and TD Groups

Eye-Tracking Metrics Across the 3 Tasks

The analysis of the biomarkers identified for all 3 tasks (Figure 6; Multimedia Appendices 2 and 3) showed that for completion, there were significant differences in TA fixation incidence (calculated based on $B_{TA\ Fix.}$) and $L_{TA\ Fix.}$ between the ADHD and TD groups for all 3 tasks (both P<.001). $A_{Sac.\ Avg.}$ of the ADHD group was significantly smaller than that of the TD group in the prosaccade and antisaccade tasks (all P<.001), whereas $V_{Sac.\ Avg}$ and $V_{Sac.\ Peak}$ of the ADHD group was significantly slower than those of the TD group for all tasks (all P<.001). $D_{Pupil\ Sd.}$ of the ADHD group was significantly greater than that of the TD group for all tasks (P=.03 for the prosaccade task, P<.001 for the antisaccade task, and P=.02 for the delayed saccade task).

In terms of attention control, in both the prosaccade and antisaccade tasks, more irrelevant fixations (ie, $N_{UA\ Fix.}$) occurred in the ADHD group than in the TD group (all P<.001). In addition, the ADHD group fixated more frequently on the UA during the antisaccade task than in the prosaccade task.

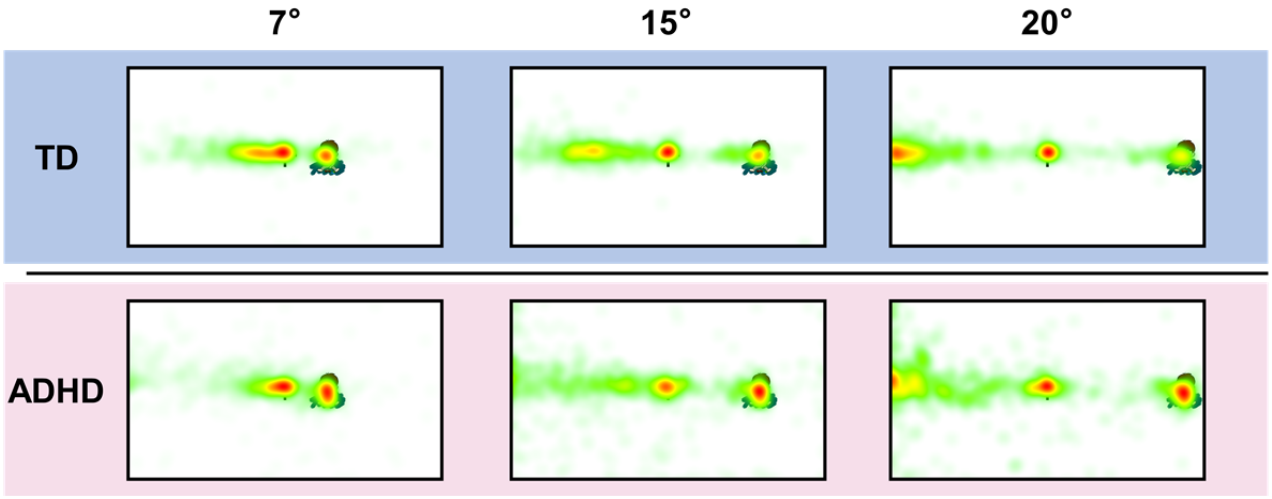
Figure 6. Comparisons of eye-tracking metrics between the attention-deficit/hyperactivity disorder (ADHD) and typically developing (TD) groups. Results of the corresponding data analyses are presented in Multimedia Appendices 3 and 4. *P<.05, **P<.01. Fix.; fixation; GTE: gaze transition entropy; PSA: proper-side area; SA: stimulus area; Sac.: saccade; SGE: stationary gaze entropy; TA: target area; UA: unrelated area; WSA: wrong-side area.



Eye-Tracking Metrics of the Antisaccade Task

The heat maps (Figure 7) of the analysis of the different target eccentricities (Multimedia Appendix 4) revealed that the TD group's fixations were concentrated along the horizontal position where the SA and TA were located, whereas the ADHD group's fixations were more widespread. Moreover, the TD group was more accurate than the ADHD group in fixating on the TA, whereas the ADHD group showed more erroneous localization deviations in both the 7° and 15° trials. Interestingly, in the 20° trial, we noted that the fixation concentration of the ADHD group deviated from the stimulus: there was a longitudinal distribution of fixations along the edge of the correct side of the screen, which suggested that the ADHD group did not localize fixation according to the logic of symmetry; rather, they relied purely on the edge of the screen to assist in their fixation positioning.

Figure 7. Heat maps of fixations of the typically developing (TD) and attention-deficit/hyperactivity disorder (ADHD) groups for stimuli of different target eccentricities in the antisaccade task.



As shown in Figure 6 and Multimedia Appendix 4, the ADHD group had more WSA fixations (calculated from $B_{WSA\ Fix.}$) and fewer PSA fixations (calculated from $B_{PSA\ Fix.}$) than the TD group (all P<.001). Among the 3 eccentricities, the number of WSA fixations during the 15° and 20° trials were significantly different between the groups (U=81,316 for 15°, U=80,812 for 20°, all P<.001), whereas in the 7° trials, both groups showed a higher number of WSA fixations (U=87,841, P=.52) than PSA fixations. However, the TD group had more PSA fixations in the 7° trials and a higher incidence of the first fixation in the PSA (calculated from $B_{PSA\ Fix.\ 1st}$) than the ADHD group (all P<.001).

Comparisons of search incidence (calculated from B_{Search}), N_{Search}, and T_{Search} between the ADHD and TD groups showed that the ADHD group was significantly higher than the TD group for all 3 metrics (P<.001, P<.001, and P=.008, respectively). Both SGE and GTE were significantly higher in the ADHD group than in the TD group (all P<.001).

Eye-Tracking Metrics in the Delayed Saccade Task

As shown in Figure 6 and Multimedia Appendix 4, TA-P fixation incidence (calculated from $B_{TA\ Fix.}$) and $L_{TA\ Fix.}$ were significantly different between the 2 groups at all eccentricities. Moreover, the TD group had a lower $N_{TA\text{-}W\ Fix.}$ than the ADHD group (all P<.001).

As the stimulus eccentricity increased from the center point, only the TD group showed an improvement in performance. The TD group showed a lower $N_{TA\text{-}W\ Fix}$ when the eccentricity was 15° than when the eccentricity was 7°, whereas the decrease in $N_{TA\text{-}W\ Fix}$ in the ADHD group from an eccentricity of 15° to 20° was more gradual than that in the TD group.

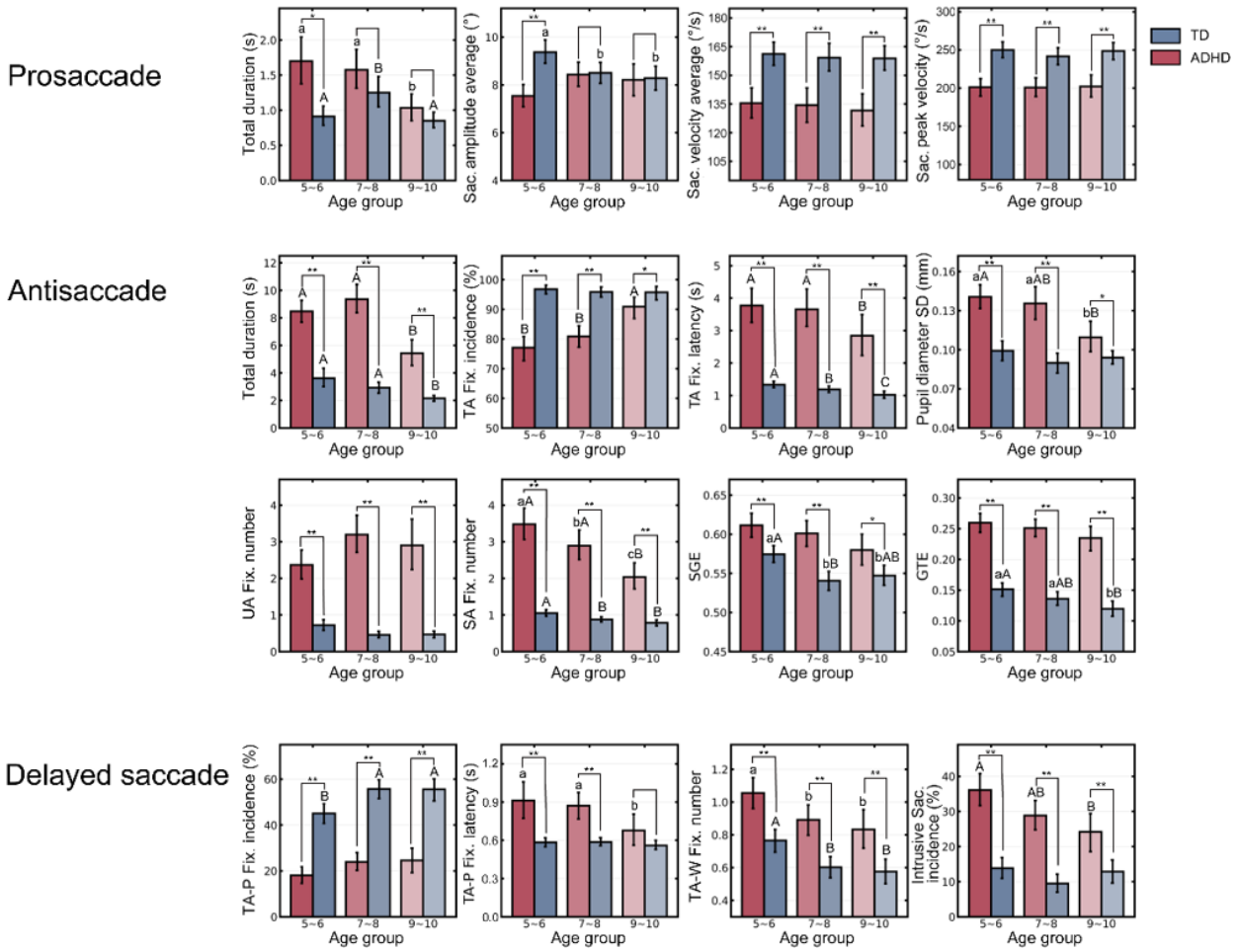
The assessment of intrusive saccades for stability of eye movements showed that the ADHD group had more intrusive saccades (calculated from $B_{Intrusive\ \ Sac.}$) and less stable eye-movement patterns than the TD group (P<.001).

Comparisons of Digital Biomarkers Among Age Groups

We discovered that several digital biomarkers showed consistent changes with age (Figure 8; Multimedia Appendices 5 and 6). In the prosaccade task, the overall T_{Total} of both groups showed a decreasing trend with age (P=.02 for ADHD, P<.001 for TD). In addition, an age-related decrease in $A_{Sac.\ Avg.}$ was observed in the TD group only (P=.007), whereas $V_{Sac.\ Avg.}$ and $V_{Sac.\ Peak}$ remained stable in both groups (P=.71 for $V_{Sac.\ Avg.}$ and P=.46 for $V_{Sac.\ Peak}$). In the antisaccade task, both the TD and ADHD groups showed an increasing trend for accuracy (P<.001 for ADHD, P=.63 for TD) and efficiency (P<.001 for ADHD and TD) in completing the task. In fact, the ADHD group showed significantly greater improvement than the TD group (P<.001). The ADHD group also exhibited a propensity for $D_{Pupil\ Sd.}$ to decrease with age (P<.001). Across all age groups, the ADHD group had a higher $N_{UA\ Fix.}$ than the TD group (P<.001), and this did not significantly improve with age; although the $N_{SA\ Fix.}$ significantly dropped with age (P<.001). We also found that there was a greater tendency for SGE and GTE to decline with age in the TD group than in the ADHD group (P<.001 for SGE and P=.001 for GTE).

The TA-P fixation incidence (P=.06) did not significantly differ with age in the ADHD group for the delayed saccade task. This was true despite the ADHD group showing improvements in $L_{TA\ Fix.}$ (P=.01), $N_{TA\text{-}W\ Fix.}$ (P=.005), and intrusive saccade incidence (calculated from $B_{Intrusive\ Sac.}$; P=.003) with age.

Figure 8. Comparisons of eye-tracking metrics among age groups. Letters above the bars indicate the results of the post hoc tests using Bonferroni correction among different age groups in the attention-deficit/hyperactivity disorder (ADHD) and typically developing (TD) groups. Lower case letters indicate $P<.05$; upper case letters indicate $P<.01$. *$P<.05$, **$P<.01$. Fix.; fixation; GTE: gaze transition entropy; SA: stimulus area; Sac.: saccade; SGE: stationary gaze entropy; TA: target area; UA: unrelated area.



ML Analysis With the Proposed Digital Biomarkers

The evaluation metrics (AUC, accuracy, sensitivity, specificity, precision, and F_1-score) are reported as means (95% CIs). The XGBoost model trained on the eye-tracking variables achieved an AUC of 0.965 (0.964-0.966), an accuracy of 0.908 (0.907-0.910), a sensitivity of 0.877 (0.874-0.880), a specificity of 0.932 (0.930-0.934), a precision of 0.913 (0.910-0.915), and an F_1-score of 0.892 (0.890-0.894). The averaged ROC curve is shown in Figure 9, which illustrates the effectiveness of the proposed digital biomarkers for discriminating the ADHD and TD groups. The 10 most important variables for the model are reported with their scores in Figure 10.

Figure 9. Receiver operating characteristic curve of the classification model. AUC: area under the receiver operating characteristic curve.

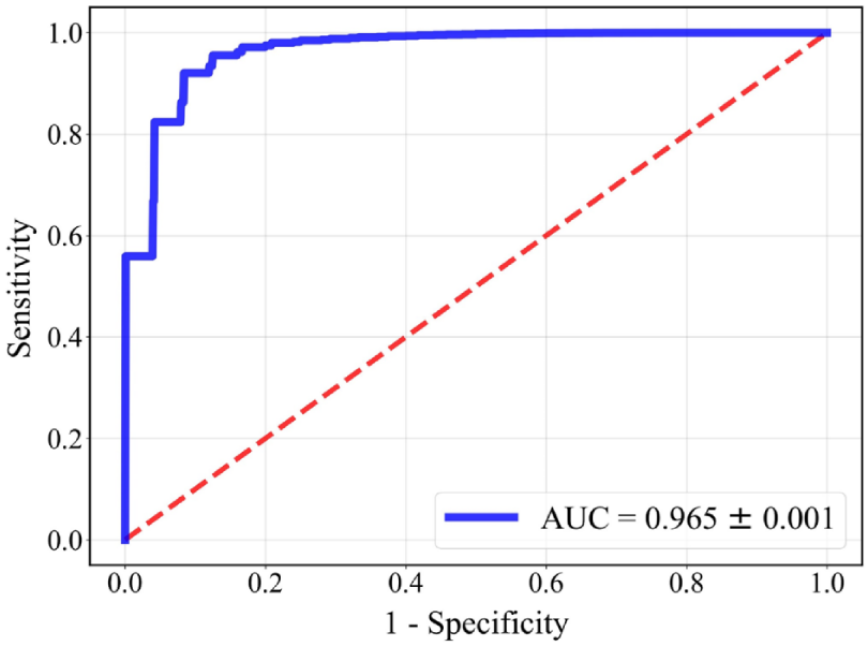
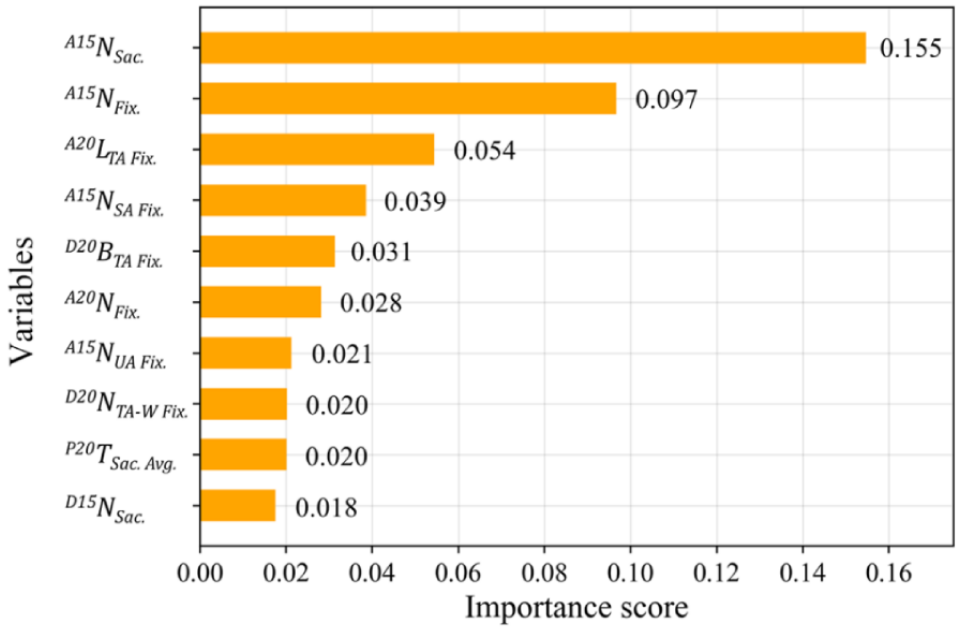


Figure 10. Importance scores of the top 10 most important variables. $N_{Sac.}$: total number of saccades; $N_{Fix.}$: total number of fixations; $L_{TA\ Fix.}$: fixation latency of the TA; $N_{SA\ Fix.}$: number of fixations in the stimulus area; $B_{TA\ Fix.}$: Boolean value to signify the occurrence of fixations in the TA (TA-P for the delayed-saccade task); $N_{UA\ Fix.}$: number of fixations in the UA; $N_{TA\text{-}W\ Fix.}$: number of fixations in the TA for the wrong period; $T_{Sac.\ Avg.}$: average of saccade duration.



Discussion

Principal Findings

Clinical Behavioral Performance

The 3 saccade tasks consistently showed that the performance of the ADHD group was poorer than that of the TD group, which suggests that the paradigm serves as a reliable and objective measure of cognitive and executive functioning. Furthermore, the ADHD group exhibited a pattern of amelioration with aging, whereas the TD group showed consistent performance across the different age groups. This may be because TD individuals had already achieved a higher cognitive skill level and a relatively stable state of corresponding biomarkers than ADHD individuals of the same age. Therefore, despite the ADHD group showing a faster rate of improvement, they performed significantly worse than TD individuals across all age groups. This finding demonstrates distinct developmental eye-movement patterns associated with ADHD.

Attention and Inhibitory Control

The ADHD group exhibited a significant lag in the ability to inhibit stimuli, which was characterized by poorer performance than the TD group on tasks with a weaker perceptual load.

Previous studies have confirmed that human visual features are divided into 3 regions–the foveal region at a viewing angle of 2.5° from the gaze point has the highest visual sensitivity, followed by the parafoveal region from 2.5° to 4.2°, and the peripheral region from 4.2° to 9.2° has the lowest visual sensitivity [52]. In this study, the 7° eccentricity stimulus was closest to the central cross and within the peripheral region, whereas the other 2 stimulus types were located outside the peripheral region. Thus, the task of inhibiting the 15° and 20° eccentricity stimuli was a low perceptual load task, which was relatively easy for the TD group. However, the performance of the PSA first incidence showed that the ADHD group had poorer inhibitory control for the lower perceptual load task of inhibiting stimuli that were located outside of the peripheral region (ie, the 15° and 20° eccentricity stimuli, as shown in Figure 6). This confirms the existence of up-down attention control impairment in individuals with ADHD [53] and emphasizes that children with ADHD may be more prone to distraction in low perceptual load environments because of a higher central threshold of response to perceptual load [54]. This finding also corroborates previous reports that individuals with ADHD are more sensitive to stimuli located in peripheral regions.

Furthermore, although individuals with ADHD had difficulty suppressing the sudden appearance of distracting stimuli, they also had a longer completion time than the TD group for the prosaccade task with a single instruction. This may be attributed to the low load of the prosaccade task, which may not have elicited sufficient cognitive arousal in the ADHD group, leading to poorer task performance. In addition, in the delayed saccade tasks that involved sequential instructions (ie, “do not look at the stimulus until you hear the cue, and then quickly look at the stimulus”), the weak task-switching ability of the ADHD group may have also prolonged fixation latency.

Organizing and Planning

In the antisaccade task, the ADHD group exhibited significantly lower TA fixation incidence and longer $L_{TA\ Fix.}$ compared with the TD group (Figure 6). This suggests that most children in the ADHD group were unable to accurately localize the TA, and those who succeeded took longer. On the basis of the heat map and UA fixation (Figures 6 and 7), the ADHD group exhibited greater fixation deviation and more frequent search behaviors.

In addition, the ADHD group had much higher SGE and GTE than the TD group for overall eye-movement trajectory, which indicated that they exhibited more eye-movement pattern shifts and spatial dispersion of fixations. This suggests that patients with ADHD favor an irregular search pattern and lack forethought when organizing and coordinating eye movements during symmetrical localization, resulting in prolonged search time to accurately locate the target. Furthermore, the positive correlation between SGE and GTE in the ADHD group supports the impact of top-down interference on visual scanning in ADHD [43].

The TD group followed a significant declining trend in SGE (P<.001) and GTE (P=.001) with age, whereas the ADHD group maintained high entropy values. We also observed that the frequent UA fixation in the ADHD group did not improve with age. These findings suggest that with age, the TD group better localized the landing point, which led to a more regular eye-movement trajectory. In contrast, the irregular eye-movement pattern of the ADHD group was exhibited across all age groups.

Eye-Movement Coordination With Age

Previous studies have mainly focused on age-related changes in the general population by comparing individuals among different age groups. However, few studies have examined variations in eye movement among younger individuals with ADHD and TD individuals. A recent study evaluating the performance of visually guided horizontal prosaccades in healthy people aged 3 years to >80 years found that peak saccade velocity increases until the age of 6 years, after which it remains relatively stable until 10 years of age [55]. The results of our prosaccade task similarly demonstrated that $V_{Sac.\ Avg.}$ and $V_{Sac.\ Peak}$ remained stable from ages 5 to 10 years in both the ADHD and TD groups, which indicates that the developmental pattern of saccade velocity is similar across both groups.

We also discovered that the ADHD group was more likely to experience intrusive saccades during the central fixation stage. The percentage of intrusive saccades decreased with age in the ADHD group, whereas that in the TD group remained at a well-performing and stable level across age groups. This further highlights the overall impairment in eye-movement control in the ADHD group.

The TD group showed a consistently higher $V_{Sac.\ Avg.}$ than the ADHD group. However, it showed a decreasing trend with age for $A_{Sac.\ Avg}$ than the ADHD group. In addition to speed, accurate localization is also required to successfully perform the prosaccade task. With age, children may modulate their eye movements to a lower speed for greater controllability, rather than simply sweeping their eyes rapidly toward the target, and thus, increase task efficiency.

Previous research has reported that the cerebellum is a crucial hub of the motor network that interacts with the executive control circuits of the frontoparietal lobe, which are involved in inhibition and stimulus response [56]. Furthermore, studies have demonstrated reduced volume and under activation of the cerebellum in individuals with ADHD [57], which suggests that impairment of the cerebellum contributes to poor control and coordination of eye movements in patients with ADHD.

Variations in Pupil Diameter and Cognitive Stress

It is well-established that when humans encounter stressful situations, they dilate their pupils to improve vision [58]. Previous research using eye-tracking technology has also revealed that when people are engaged in an active coping task, their pupils enlarge significantly. These findings suggest that a larger pupil diameter is linked to higher cognitive load while preparing for challenging tasks [58]. According to previous research examining the relationship between pupil diameter and attention, there is an inverted U–shaped pattern between pupil diameter and attentional performance; that is, when pupil diameter becomes too small or large, error rates are higher and response times are slower [59]. In our study, we discovered that

for all tasks, children with ADHD displayed greater pupil diameter variation than TD individuals. This finding supports the theory that excessively large or small pupil diameter is an indicator of inattentiveness when completing tasks requiring active responses. Alternatively, executive function deficiencies at the functional level of the brain and inefficient brain network connectivity in the ADHD group may account for the higher cognitive load when responding to complex task demands [60].

ML Analysis

For the classification of ADHD and TD children, the ML model achieved an AUC of 0.965 and an accuracy of 0.908, which demonstrates promise for the model to serve as an automated screening tool for ADHD children. Moreover, the high performance of the model highlights the effectiveness of the paradigm and its ability to extract digital eye-tracking biomarkers. In a previous study focused on screening for ADHD using eye-tracking and ML methods, Lev et al [18] conducted continuous performance tests in 66 participants (33 adult patients with ADHD and 33 healthy controls) and used eye-movement metrics during the tests to classify patients and controls. They applied a regression model to combine the relative gaze durations of 4 AOIs as the diagnostic scale and achieved an AUC of 0.826. Das and Khanna [19] extracted pupil size dynamics features as an objective biomarker and trained 5 types of commonly used classification models to detect ADHD. Using the data of 50 participants (28 patients with ADHD and 22 healthy controls) and 10-fold cross-validation, they attained an AUC of 0.856. Deng et al [61] built an eye-tracking ML classifier for ADHD using the natural reading paradigm; however, the model was difficult to interpret, and the classification performance (AUC of 0.646) was not as high as the performance achieved by our model.

Compared with previous work, we recruited a larger number of participants (ie, 94 ADHD and 122 TD individuals), obtained higher evaluation metrics, and achieved better classification performance for children with ADHD. Moreover, we extracted a larger variety of eye-tracking metrics and provided a more comprehensive description of participants' eye-movement behaviors. These advantages emphasize the effectiveness, reliability, and potential practical applications of the model. Furthermore, our findings offer valuable insight into the field of ADHD diagnosis using ML.

Because we plan to extend our findings using portable eye-tracking devices in the future, we validated the performance of our model at lower sampling frequencies using external samples. Results demonstrated that the model adapted well to low-sampling rate data, which further confirmed its high generalizability and applicability to portable devices (Multimedia Appendix 7).

Advantages of the Study

First, we used eye-tracking technology in a natural and straightforward assessment setting, which enabled direct visual and on-screen interactions without complicated rules or restrictions on head motion. Unlike the paradigms used in previous studies, our approach did not require participants to wear additional equipment [7] or make additional keystrokes [18]. In addition, our method avoided interference from other environments and devices, facilitated children's participation, and minimized inaccuracies in eye-movement measurement due to excessive head movement.

Second, our paradigm allowed a more comprehensive exploration of children's cognitive skills. In addition to testing attentional and inhibitory ability, our paradigm included audiovisual integration, which has been shown to be effective in evaluating children with ADHD.

Third, we provided a more comprehensive scheme for extracting digital eye-tracking biomarkers by expanding the evaluation system of classical paradigms. The presentation of stimuli was further divided into defined areas of fixation for quantitative analyses; moreover, behaviors, such as search behaviors that are typically observed in the clinic, were quantified alongside numerous metrics based on the AOI, such as fixation duration, saccade velocity and amplitude, and pupil diameter change. This enabled the extraction of more detailed eye-movement metrics during different saccade tests than those used in previous studies [8,18,19,62] while ensuring that the extracted digital biomarkers were interpretable and objectively reflected cognitive deficits. As a result, we were able to provide a practical and thorough description of children's performance in completing the various tasks.

In addition, we applied ML modeling using the extracted digital biomarkers and achieved promising results, which confirmed that these biomarkers are highly valuable for the future development of screening and auxiliary diagnostic tools. We also investigated age-related developmental patterns of eye movement in addition to simple eye-movement metrics in children with ADHD in a larger, more trustworthy, and more representative dataset than previous research. In terms of practical applications, the implementation of the paradigm is straightforward, and the 7-minute duration of the assessment is suitable for children with ADHD. These features will increase the likelihood that the assessment can be completed successfully by children with ADHD. Taken together, we have provided a reliable and practical solution for auxiliary diagnosis and screening for ADHD at the primary care level.

Limitations

Although our sample size was larger than previous studies, we only recruited from 1 city in China. Therefore, the representativeness of the sample can be improved. There was also a sex imbalance between the TD and ADHD groups. The TD group had a 1:1 sex ratio, whereas the ADHD group had considerably more boys (84/94, 89%) than girls (10/94, 11%). Although this was attributed to morbidity bias, our modeling would have benefited from a balanced sex ratio, especially because previous studies have reported sex differences in comorbidity and cognitive impairment in children with ADHD [63]. To verify the effect of sex on the validity of the model, 2 models were trained using data from boys and girls separately and validated using data from the opposite sex. Results showed that the 2 models performed well, which suggested that sex does not significantly affect modeling and that the current model trained with predominantly male data can also be applied to predict female participants' behavior (Multimedia Appendix

8). Nevertheless, we must consider recruiting participants with various presentations (eg, different geography and balanced sex ratios) in future studies to ensure the generalizability of the ML model. Finally, our eye-tracking instrument is heavy and difficult to carry. Thus, in the future, a portable eye-tracking instrument could be used.

Conclusions and Future Research

We successfully adapted eye-tracking technology for clinical use as a tool for auxiliary diagnosis and campus and community screening for ADHD. The system includes standard paradigms and a reliable digital biomarker extraction process. We validated the use of digital biomarkers to build robust ML models. In addition, the entire assessment process was conducted in a natural setting without the need for extra equipment to be worn by participants. The assessment is also brief and simple, which makes it particularly suitable for clinical applications and ensures completion of the assessment.

For the next steps of our research, we plan to further expand the sample size and implement multicenter data collection using the proposed paradigm and digital biomarker extraction scheme. We aim to build a robust ML model and externally validate classifiers to improve their predictive accuracy and stability. This will ensure that the auxiliary diagnosis model can be effectively applied to real clinical scenarios and improve primary care–level screening and diagnosis of ADHD.

Acknowledgments

This study was financially supported by the Joint Project on Emerging Frontier Technologies for Municipal Hospitals of Shanghai Hospital Development Center (grant SHDC12022114), Exemplary Research Physician Training Program on Innovation and Transformation Capabilities Special Project on Transformation of Innovation Results of Medical-Enterprise Integration of SHDC (grant SHDC2022CRD012), National Natural and Science Foundation of China (grant 62203296), Shanghai Pujiang Program (grant 22PJ1405500), Shanghai Jiao Tong University Medical-Industrial Crossroads Center Project, National Program on Key Research Program (grant 2022YFC2705203), Shanghai 2022 “Science and Technology Innovation Action Plan” Technical Standards Project (grant 22DZ2203500), the Three-year Action Plan for Strengthening the Construction of Shanghai Public Health System (2023-2025) Outstanding Young Talents (grant GWVI-11.2-YQ23), Shanghai Pilot Program for Basic Research—Shanghai Jiao Tong University (grant 21TQ1400203), the Science and Technology Commission of Shanghai Municipality (grant 20DZ2220400), the National Natural Science Foundation of China (grant 81703235), and the Fundamental Research Funds for the Central Universities (grant YG2022QN100).

The authors thank Tianbao Chen, Ruoya Qian, and Ruihan Wang from Shanghai Jiao Tong University School of Medicine for their help with data processing.

They thank Sarina Iwabuchi, PhD, from Liwen Bianji (Edanz), for editing the English text of a draft of this paper.

Authors' Contributions

ZL performed the data collection and data analysis of the clinical part, and wrote the paper; JL performed the construction of the biomarker extraction scheme and machine learning modeling, and wrote the paper; YZ, DW, and YH were involved in the analysis of the clinical data; JY and MZ were involved in the feature extraction and machine learning modeling; CD, LJ, RS, and RZ implemented the patient interviews and assessments; FL, XY, and DZ reviewed and revised the paper; YG and JC set the topic, critically revised the article, and funded the study.
YG and JC contributed equally as cocorresponding authors.
All authors reviewed the final paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Hyperparameters of XGBoost model.
[DOCX File , 13 KB - mhealth_v12i1e58927_app1.docx]

Multimedia Appendix 2
Detailed description of digital biomarkers illustrated in result figures.
[DOCX File , 20 KB - mhealth_v12i1e58927_app2.docx]

Multimedia Appendix 3
Differences in eye-movement metrics between attention-deficit/hyperactivity disorder and typically developing groups.
[DOCX File , 26 KB - mhealth_v12i1e58927_app3.docx]

Multimedia Appendix 4
Differences in eye-movement metrics between attention-deficit/hyperactivity disorder and typically developing groups in the antisaccade and delayed saccade tasks with different target eccentricities.
[DOCX File , 21 KB - mhealth_v12i1e58927_app4.docx]

Multimedia Appendix 5
Differences in eye-movement metrics between attention-deficit/hyperactivity disorder and typically developing groups for different age groups.
[DOCX File , 26 KB - mhealth_v12i1e58927_app5.docx]

Multimedia Appendix 6
Differences in eye-movement metrics between different age groups for attention-deficit/hyperactivity disorder and typically developing groups.
[DOCX File , 22 KB - mhealth_v12i1e58927_app6.docx]

Multimedia Appendix 7
Validation of the impact of down-sampling on model training.
[DOCX File , 19 KB - mhealth_v12i1e58927_app7.docx]

Multimedia Appendix 8
Validation of the effect of sex differences on model training.
[DOCX File , 16 KB - mhealth_v12i1e58927_app8.docx]

References

1. Posner J, Polanczyk GV, Sonuga-Barke E. Attention-deficit hyperactivity disorder. Lancet 2020 Feb 08;395(10222):450-462 [FREE Full text] [doi: 10.1016/S0140-6736(19)33004-1] [Medline: 31982036]
2. Li F, Cui Y, Li Y, Guo L, Ke X, Liu J, et al. Prevalence of mental disorders in school children and adolescents in China: diagnostic data from detailed clinical assessments of 17,524 individuals. J Child Psychol Psychiatry 2022 Jan 21;63(1):34-46. [doi: 10.1111/jcpp.13445] [Medline: 34019305]
3. Wright N, Moldavsky M, Schneider J, Chakrabarti I, Coates J, Daley D, et al. Practitioner review: pathways to care for ADHD - a systematic review of barriers and facilitators. J Child Psychol Psychiatry 2015 Jun;56(6):598-617 [FREE Full text] [doi: 10.1111/jcpp.12398] [Medline: 25706049]
4. Toba MN, Malkinson TS, Howells H, Mackie MA, Spagna A. Same, same but different? A multi-method review of the processes underlying executive control. Neuropsychol Rev 2023 Mar 27;34(2):418-454. [doi: 10.1007/s11065-023-09577-4] [Medline: 36967445]
5. Sutton GP, Barchard KA, Bello DT, Thaler NS, Ringdahl E, Mayfield J, et al. Beery-Buktenica developmental test of visual-motor integration performance in children with traumatic brain injury and attention-deficit/hyperactivity disorder. Psychol Assess 2011 Sep;23(3):805-809. [doi: 10.1037/a0023370] [Medline: 21875221]
6. Nag A, Haber N, Voss C, Tamura S, Daniels J, Ma J, et al. Toward continuous social phenotyping: analyzing gaze patterns in an emotion recognition task for children with autism through wearable smart glasses. J Med Internet Res 2020 Apr 22;22(4):e13810 [FREE Full text] [doi: 10.2196/13810] [Medline: 32319961]
7. Merzon L, Pettersson K, Aronen ET, Huhdanpää H, Seesjärvi E, Henriksson L, et al. Eye movement behavior in a real-world virtual reality task reveals ADHD in children. Sci Rep 2022 Nov 24;12(1):20308 [FREE Full text] [doi: 10.1038/s41598-022-24552-4] [Medline: 36434040]
8. Lee DY, Shin Y, Park RW, Cho SM, Han S, Yoon C, et al. Use of eye tracking to improve the identification of attention-deficit/hyperactivity disorder in children. Sci Rep 2023 Sep 02;13(1):14469 [FREE Full text] [doi: 10.1038/s41598-023-41654-9] [Medline: 37660094]
9. Chen X, Wang S, Yang X, Yu C, Ni F, Yang J, et al. Utilizing artificial intelligence-based eye tracking technology for screening ADHD symptoms in children. Front Psychiatry 2023 Nov 14;14:1260031 [FREE Full text] [doi: 10.3389/fpsyt.2023.1260031] [Medline: 38034916]
10. Mahanama B, Jayawardana Y, Rengarajan S, Jayawardena G, Chukoskie L, Snider J, et al. Eye movement and pupil measures: a review. Front Comput Sci 2022 Jan 11;3:733531 [FREE Full text] [doi: 10.3389/fcomp.2021.733531]
11. Lim JZ, Mountstephens J, Teo J. Eye-tracking feature extraction for biometric machine learning. Front Neurorobot 2021 Feb 1;15:796895 [FREE Full text] [doi: 10.3389/fnbot.2021.796895] [Medline: 35177973]
12. Salthouse TA, Ellis CL. Determinants of eye-fixation duration. Am J Psychol 1980 Jun;93(2):207. [doi: 10.2307/1422228]
13. Zhao Z, Tang H, Zhang X, Qu X, Hu X, Lu J. Classification of children with autism and typical development using eye-tracking data from face-to-face conversations: machine learning model development and performance evaluation. J Med Internet Res 2021 Aug 26;23(8):e29328 [FREE Full text] [doi: 10.2196/29328] [Medline: 34435957]

14. Cilia F, Carette R, Elbattah M, Dequen G, Guérin JL, Bosche J, et al. Computer-aided screening of autism spectrum disorder: eye-tracking study using data visualization and deep learning. JMIR Hum Factors 2021 Oct 25;8(4):e27706 [FREE Full text] [doi: 10.2196/27706] [Medline: 34694238]
15. Zemblys R, Niehorster DC, Komogortsev O, Holmqvist K. Using machine learning to detect events in eye-tracking data. Behav Res Methods 2018 Mar;50(1):160-181. [doi: 10.3758/s13428-017-0860-3] [Medline: 28233250]
16. Kollias KF, Syriopoulou-Delli CK, Sarigiannidis P, Fragulis GF. The contribution of machine learning and eye-tracking technology in autism spectrum disorder research: a systematic review. Electronics 2021 Nov 30;10(23):2982. [doi: 10.3390/electronics10232982]
17. Krol M, Krol M. A novel approach to studying strategic decisions with eye-tracking and machine learning. Judgm Decis Mak 2023 Jan 01;12(6):596-609. [doi: 10.1017/S1930297500006720]
18. Lev A, Braw Y, Elbaum T, Wagner M, Rassovsky Y. Eye tracking during a continuous performance test: utility for assessing ADHD patients. J Atten Disord 2022 Jan;26(2):245-255. [doi: 10.1177/1087054720972786] [Medline: 33238787]
19. Das W, Khanna S. A robust machine learning based framework for the automated detection of ADHD using pupillometric biomarkers and time series analysis. Sci Rep 2021 Aug 12;11(1):16370 [FREE Full text] [doi: 10.1038/s41598-021-95673-5] [Medline: 34385511]
20. Mandal S, Kumar GP, Saini M, Satija U, Kumar Y. A unified deep learning framework for smartphone-enabled ADHD detection. IEEE Trans Instrum Meas 2024;73:1-11. [doi: 10.1109/tim.2023.3338713]
21. Krafka K, Khosla A, Kellnhofer P, Kannan H, Bhandarkar S, Matusik W. Eye tracking for everyone. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. 2016 Presented at: CVPR '16; June 27-30, 2016; Las Vegas, NV p. 2176-2184 URL: https://ieeexplore.ieee.org/document/7780608 [doi: 10.1109/cvpr.2016.239]
22. Li J, Yang J, Liu Y, Li Z, Yang GZ, Guo Y. EasyGaze3D: towards effective and flexible 3D gaze estimation from a single RGB camera. In: Proceedings of the 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems. 2023 Presented at: IROS '23; October 1-5, 2023; Detroit, MI p. 6537-6543 URL: https://ieeexplore.ieee.org/abstract/document/10342361 [doi: 10.1109/iros55552.2023.10342361]
23. Roberts W, Fillmore MT, Milich R. Separating automatic and intentional inhibitory mechanisms of attention in adults with attention-deficit/hyperactivity disorder. J Abnorm Psychol 2011 Feb;120(1):223-233 [FREE Full text] [doi: 10.1037/a0021408] [Medline: 21058752]
24. Menon V, D'Esposito M. The role of PFC networks in cognitive control and executive function. Neuropsychopharmacology 2022 Jan;47(1):90-103 [FREE Full text] [doi: 10.1038/s41386-021-01152-w] [Medline: 34408276]
25. Fettes P, Schulze L, Downar J. Cortico-striatal-thalamic loop circuits of the orbitofrontal cortex: promising therapeutic targets in psychiatric illness. Front Syst Neurosci 2017;11:25 [FREE Full text] [doi: 10.3389/fnsys.2017.00025] [Medline: 28496402]
26. Bai S, Liu W, Guan Y. The visuospatial and sensorimotor functions of posterior parietal cortex in drawing tasks: a review. Front Aging Neurosci 2021;13:717002 [FREE Full text] [doi: 10.3389/fnagi.2021.717002] [Medline: 34720989]
27. Tien YM, Chen VC, Lo TS, Hsu CF, Gossop M, Huang KY. Deficits in auditory sensory discrimination among children with attention-deficit/hyperactivity disorder. Eur Child Adolesc Psychiatry 2019 May;28(5):645-653. [doi: 10.1007/s00787-018-1228-7] [Medline: 30229307]
28. Maron DN, Bowe SJ, Spencer-Smith M, Mellahn OJ, Perrykkad K, Bellgrove MA, et al. Oculomotor deficits in attention deficit hyperactivity disorder (ADHD): a systematic review and comprehensive meta-analysis. Neurosci Biobehav Rev 2021 Dec;131:1198-1213. [doi: 10.1016/j.neubiorev.2021.10.012] [Medline: 34655657]
29. Klein CH, Raschke A, Brandenbusch A. Development of pro- and antisaccades in children with attention-deficit hyperactivity disorder (ADHD) and healthy controls. Psychophysiology 2003 Jan;40(1):17-28. [doi: 10.1111/1469-8986.00003] [Medline: 12756978]
30. Lemel R, Shalev L, Nitsan G, Ben-David BM. Listen up! ADHD slows spoken-word processing in adverse listening conditions: evidence from eye movements. Res Dev Disabil 2023 Feb;133:104401 [FREE Full text] [doi: 10.1016/j.ridd.2022.104401] [Medline: 36577332]
31. Rojas-Líbano D, Wainstein G, Carrasco X, Aboitiz F, Crossley N, Ossandón T. A pupil size, eye-tracking and neuropsychological dataset from ADHD children during a cognitive task. Sci Data 2019 Apr 11;6(1):25 [FREE Full text] [doi: 10.1038/s41597-019-0037-2] [Medline: 30975993]
32. Siqueiros Sanchez M, Falck-Ytter T, Kennedy DP, Bölte S, Lichtenstein P, D'Onofrio BM, et al. Volitional eye movement control and ADHD traits: a twin study. J Child Psychol Psychiatry 2020 Dec;61(12):1309-1316 [FREE Full text] [doi: 10.1111/jcpp.13210] [Medline: 32020616]
33. Gau SS, Lin CH, Hu FC, Shang CY, Swanson JM, Liu YC, et al. Psychometric properties of the Chinese version of the Swanson, Nolan, and Pelham, Version IV scale-teacher form. J Pediatr Psychol 2009 Sep;34(8):850-861. [doi: 10.1093/jpepsy/jsn133] [Medline: 19074488]
34. Parton A, Nachev P, Hodgson TL, Mort D, Thomas D, Ordidge R, et al. Role of the human supplementary eye field in the control of saccadic eye movements. Neuropsychologia 2007 Mar 14;45(5):997-1008 [FREE Full text] [doi: 10.1016/j.neuropsychologia.2006.09.007] [Medline: 17069864]

35. Panagiotidi M, Overton PG, Stafford T. Attention-deficit hyperactivity disorder-like traits and distractibility in the visual periphery. Perception 2017 Jun;46(6):665-678 [FREE Full text] [doi: 10.1177/0301006616681313] [Medline: 27895292]
36. Purves D, Augustine GJ, Fitzpatrick D, Lawrence CK, Lamantia AS, McNamara JO, et al. Neuroscience. 2nd edition. Sunderland, MA: Sinauer Associates; 2001.
37. Cherng YG, Baird T, Chen JT, Wang CA. Background luminance effects on pupil size associated with emotion and saccade preparation. Sci Rep 2020 Sep 24;10(1):15718 [FREE Full text] [doi: 10.1038/s41598-020-72954-z] [Medline: 32973283]
38. Bradley MM, Miccoli L, Escrig MA, Lang PJ. The pupil as a measure of emotional arousal and autonomic activation. Psychophysiology 2008 Jul;45(4):602-607 [FREE Full text] [doi: 10.1111/j.1469-8986.2008.00654.x] [Medline: 18282202]
39. Jerčić P, Sennersten C, Lindley C. Modeling cognitive load and physiological arousal through pupil diameter and heart rate. Multimed Tools Appl 2018 Sep 3;79(5-6):3145-3159. [doi: 10.1007/S11042-018-6518-Z]
40. Guo Y, Freer D, Deligianni F, Yang GZ. Eye-tracking for performance evaluation and workload estimation in space telerobotic training. IEEE Trans Human Mach Syst 2022 Feb;52(1):1-11. [doi: 10.1109/thms.2021.3107519]
41. Kret ME, Sjak-Shie EE. Preprocessing pupil size data: guidelines and code. Behav Res Methods 2019 Jun 10;51(3):1336-1342 [FREE Full text] [doi: 10.3758/s13428-018-1075-y] [Medline: 29992408]
42. Shannon CE. A mathematical theory of communication. Bell Syst Tech J 1948;27(3):379-423 [FREE Full text] [doi: 10.1002/J.1538-7305.1948.tb01338.X]
43. Shiferaw B, Downey L, Crewther D. A review of gaze entropy as a measure of visual scanning efficiency. Neurosci Biobehav Rev 2019 Jan;96:353-366. [doi: 10.1016/j.neubiorev.2018.12.007] [Medline: 30621861]
44. Holmqvist K, Andersson R. Eye Tracking: A Guide to Methods, Paradigms, and Measures: A Comprehensive Guide to Methods, Paradigms, and Measures. Burlington, MA: CreateSpace Independent Publishing; 2018.
45. Batty M, Morphet R, Masucci P, Stanilov K. Entropy, complexity, and spatial information. J Geogr Syst 2014 Sep 24;16(4):363-385 [FREE Full text] [doi: 10.1007/s10109-014-0202-2] [Medline: 25309123]
46. Gao J, Liu F, Zhang J, Hu J, Cao Y. Information entropy as a basic building block of complexity theory. Entropy 2013 Aug 29;15(9):3396-3418. [doi: 10.3390/E15093396]
47. Ciuperca G, Girardin V. On the estimation of the entropy rate of finite Markov chains. Academia. 2005. URL: https://www.academia.edu/5736693/On_the_estimation_of_the_entropy_rate_of_finite_Markov_chains [accessed 2024-04-29]
48. Cover TM, Thomas JA. Elements of Information Theory. Hoboken, NJ: Wiley-Interscience; 2005.
49. Krejtz K, Duchowski A, Szmidt T, Krejtz I, Perilli FG, Pires A, et al. Gaze transition entropy. ACM Trans Appl Percept 2015 Dec 10;13(1):1-20. [doi: 10.1145/2834121]
50. Li J, Liu S, Hu Y, Zhu L, Mao Y, Liu J. Predicting mortality in intensive care unit patients with heart failure using an interpretable machine learning model: retrospective cohort study. J Med Internet Res 2022 Aug 09;24(8):e38082 [FREE Full text] [doi: 10.2196/38082] [Medline: 35943767]
51. Jacobson NC, Weingarden H, Wilhelm S. Digital biomarkers of mood disorders and symptom change. NPJ Digit Med 2019;2:3 [FREE Full text] [doi: 10.1038/s41746-019-0078-0] [Medline: 31304353]
52. Polyak SL. The Retina. Oxford, UK: The University of Chicago Press; 1941.
53. Friedman-Hill SR, Wagman MR, Gex SE, Pine DS, Leibenluft E, Ungerleider LG. What does distractibility in ADHD reveal about mechanisms for top-down attentional control? Cognition 2010 Apr;115(1):93-103. [doi: 10.1016/j.cognition.2009.11.013] [Medline: 20096409]
54. Forster S, Lavie N. High perceptual load makes everybody equal: eliminating individual differences in distractibility with load. Psychol Sci 2007 May;18(5):377-381. [doi: 10.1111/j.1467-9280.2007.01908.x] [Medline: 17576274]
55. Irving EL, Steinbach MJ, Lillakas L, Babu RJ, Hutchings N. Horizontal saccade dynamics across the human life span. Invest Ophthalmol Vis Sci 2006 Jun;47(6):2478-2484. [doi: 10.1167/iovs.05-1311] [Medline: 16723459]
56. Faraone SV, Asherson P, Banaschewski T, Biederman J, Buitelaar JK, Ramos-Quiroga JA, et al. Attention-deficit/hyperactivity disorder. Nat Rev Dis Primers 2015 Aug 06;1(1):15020. [doi: 10.1038/nrdp.2015.20] [Medline: 27189265]
57. Saad JF, Griffiths KR, Korgaonkar MS. A systematic review of imaging studies in the combined and inattentive subtypes of attention deficit hyperactivity disorder. Front Integr Neurosci 2020;14:31 [FREE Full text] [doi: 10.3389/fnint.2020.00031] [Medline: 32670028]
58. Sege CT, Bradley MM, Lang PJ. Motivated action: pupil diameter during active coping. Biol Psychol 2020 May;153:107885. [doi: 10.1016/j.biopsycho.2020.107885] [Medline: 32278595]
59. van den Brink RL, Murphy PR, Nieuwenhuis S. Pupil diameter tracks lapses of attention. PLoS One 2016;11(10):e0165274 [FREE Full text] [doi: 10.1371/journal.pone.0165274] [Medline: 27768778]
60. Gossé LK, Bell SW, Hosseini SM. Functional near-infrared spectroscopy in developmental psychiatry: a review of attention deficit hyperactivity disorder. Eur Arch Psychiatry Clin Neurosci 2022 Mar;272(2):273-290 [FREE Full text] [doi: 10.1007/s00406-021-01288-2] [Medline: 34185132]
61. Deng S, Prasse P, Reich DR, Dziemian S, Stegenwallner-Schütz M, Krakowczyk D, et al. Detection of ADHD based on eye movements during natural viewing. In: Proceedings of the 2022 European Conference on Machine Learning and Knowledge Discovery in Databases. 2022 Presented at: ECML PKDD '22; September 19-23, 2022; Grenoble, France p. 403-418 URL: https://link.springer.com/chapter/10.1007/978-3-031-26422-1_25

62. Rivera KP, Pizarro CD, Dueñas AJ, Rodríguez JJ, Figueroa CE, Aizpuru AP, et al. Comparation of machine learning algorithms for ADHD detection with eye tracking. In: Proceedings of the 2023 Conference on XLVI Mexican Conference on Biomedical Engineering. 2023 Presented at: CNIB '23; November 2-4, 2023; Villahermosa Tabasco, México p. 3-13 URL: https://link.springer.com/chapter/10.1007/978-3-031-46933-6_1 [doi: 10.1007/978-3-031-46933-6_1]
63. Greven CU, Richards JS, Buitelaar JK. Sex differences in ADHD. In: Banaschewski T, Coghill D, Zuddas A, editors. Oxford Textbook of Attention Deficit Hyperactivity Disorder. Oxford, UK: Oxford University Press; 2018:154-160.

Abbreviations

ADHD: attention-deficit/hyperactivity disorder
AOI: area of interest
AUC: area under the receiver operating characteristic curve
CA: center area
GTE: gaze transition entropy
ML: machine learning
PSA: proper-side area
ROC: receiver operating characteristic
SA: stimulus area
SGE: stationary gaze entropy
SNAP-IV: Swanson, Nolan, and Pelham Rating Scale
TA: target area
TA-P: target area during the proper period
TA-W: target area during the wrong period
TD: typically developing
UA: unrelated area
WSA: wrong-side area
XGBoost: extreme gradient boosting

Edited by L Buis; submitted 29.03.24; peer-reviewed by J Jiang, WB Lee, MM Jankovic; comments to author 09.07.24; revised version received 30.08.24; accepted 18.10.24; published 29.11.24.

Please cite as:
Liu Z, Li J, Zhang Y, Wu D, Huo Y, Yang J, Zhang M, Dong C, Jiang L, Sun R, Zhou R, Li F, Yu X, Zhu D, Guo Y, Chen J
Auxiliary Diagnosis of Children With Attention-Deficit/Hyperactivity Disorder Using Eye-Tracking and Digital Biomarkers: Case-Control Study
JMIR Mhealth Uhealth 2024;12:e58927
URL: https://mhealth.jmir.org/2024/1/e58927
doi:10.2196/58927
PMID:39477792

Trial Participants' Perceptions of the Impact of Ecological Momentary Assessment on Smoking Behaviors: Qualitative Analysis

Elizabeth R Stevens[1], MPH, PhD; Rina Li[1]; Grace Xiang[1], BA; Rachel Wisniewski[1], BA; Sidney Rojas[1], BA; Katherine O'Connor[1], MS; Olivia Wilker[1], MPH; Mahathi Vojjala[1,2], MPH; Omar El-Shahawy[1,2], MPH, MD, PhD; Scott E Sherman[1], MPH, MD

[1]Department of Population Health, NYU Grossman School of Medicine, New York, NY, United States

[2]School of Global Public Health, New York University, New York, NY, United States

Corresponding Author:
Elizabeth R Stevens, MPH, PhD
Department of Population Health, NYU Grossman School of Medicine, , New York, NY, , United States

Abstract

Background: Ecological momentary assessment (EMA) is an increasingly used tool for data collection in behavioral research, including smoking cessation studies. As previous addiction research suggests, EMA has the potential to elicit cue reactivity by triggering craving and increasing behavioral awareness. However, there has been limited evaluation of its potential influence on behavior.

Objective: By examining the perspectives of research participants enrolled in a tobacco treatment intervention trial, this qualitative analysis aims to understand the potential impact that EMA use may have had on smoking behaviors that may not have otherwise been captured through other study measures.

Methods: We performed a qualitative analysis of in-depth interviews with participants enrolled in a pilot randomized controlled trial of a tobacco treatment intervention that used SMS text messaging to collect EMA data on smoking behaviors. In the pilot randomized controlled trial, combustible cigarette and e-cigarette use and smoking-related cravings were measured as part of an EMA protocol, in which SMS text messaging served as a smoking diary. SMS text messaging was intended for data collection only and not designed to serve as part of the intervention. After a baseline assessment, participants were asked to record daily nicotine use for 12 weeks by responding to text message prompts that they received 4 times per day. Participants were prompted to share their experiences with the EMA text messaging component of the trial but were not directly asked about the influence of EMA on their behaviors. Transcripts were coded according to the principles of the framework for applied research. The codes were then examined, summarized, and grouped into themes based on the principles of grounded theory.

Results: Interviews were analyzed for 26 participants. The themes developed from the analysis suggested the potential for EMA, in the form of an SMS text messaging smoking diary, to influence participants' smoking behaviors. The perceived impacts of EMA text messaging on smoking behaviors were polarized; some participants emphasized the positive impacts of text messages on their efforts to reduce smoking, while others stressed the ways that text messaging negatively impacted their smoking reduction efforts. These contrasting experiences were captured by themes reflecting the positive impacts on smoking behaviors, including increased awareness of smoking behaviors and a sense of accountability, and the negative impacts on emotions and smoking behaviors, including provoking a sense of guilt and triggering smoking behaviors.

Conclusions: The collection of EMA smoking behavior data via SMS text messaging may influence the behaviors and perceptions of participants in tobacco treatment interventions. More research is needed to determine the magnitude of impact and mechanisms, to account for the potential effects of EMA. A broader discussion of the unintended effects introduced by EMA use is warranted among the research community.

(JMIR Mhealth Uhealth 2024;12:e52122) doi:10.2196/52122

KEYWORDS

smoking behavior; ecological momentary assessment; bias; behavioral impact; smoking; smoker; qualitative analysis; pilot study; tool; data collection; tobacco; text message; accountability; mHealth; mobile health; message; trigger; cigarette

Introduction

Ecological momentary assessment (EMA) is a data collection method that is increasingly being used in health and behavioral sciences [1,2]. EMA has been shown to be a useful tool for measuring behaviors associated with substance use [3]. Consequently, there has been a strong interest in the use of EMA in smoking cessation studies [4]. However, as indicated in previous addiction research, EMA has the potential to elicit cue reactivity by triggering craving and increasing behavioral awareness [5-8]. As such, there is potential for the use of EMA to create an assessment effect and inadvertently influence behaviors in some settings.

Assessment effects refer to the phenomenon in which the outcome of interest (eg, a behavior) is modified simply by assessing it, arising due to the assessment method or its interaction with the intervention [9,10]. The frequent prompts for self-reporting that are often required by EMA might inadvertently influence participants' behaviors by, among other mechanisms, increasing their self-awareness, altering an emotional response, or serving as reminders of the behavior [2,8,11,12]. Depending on the behavior of interest, these EMA consequences may differentially impact outcomes. These assessment effects can have a significant impact on the interpretation of trial results but are rarely considered in trial design [13].

Despite its potential importance to the interpretation of research results, there are a limited number of studies investigating the potential for EMA to produce an assessment effect. Within the research that does exist, there have been mixed results reported, with some studies reporting no impact on behavior [11,14,15] and others indicating that EMA likely has an impact [2,5-8,11,12,14]. For this reason, it is important to understand if and how the use of EMA data collection in smoking research could impact smoking behaviors.

To better understand the potential impact of using EMA to measure smoking behaviors, we performed a qualitative analysis of in-depth interviews with participants in a randomized controlled trial (RCT) pilot study of a smoking intervention that used SMS text messaging–based EMA as a data collection strategy [16]. By examining the perspectives of the trial participants, this qualitative analysis aims to understand the potential impact that the use of EMA may have had on smoking behaviors that may not have otherwise been captured through other means of data collection.

Methods

Study Design

For the purposes of assessing the acceptability of a smoking intervention and determining points for potential program improvement, we performed a qualitative analysis of in-depth interviews conducted with participants of an RCT pilot study that compared the effectiveness of behavioral counseling and the use of e-cigarettes on smoking outcomes to that of behavioral counseling and nicotine replacement therapy (NRT).

Ethical Considerations

The interviews and analysis procedures were reported in accordance with the COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist (Checklist 1), as applicable [17]. The study protocol was approved by the New York University Langone Health Institutional Review Board (approval number i20-00839), and written documentation of informed consent was received prior to starting data collection. Participants were provided with a US $20 incentive for their participation. Participant data were maintained on a secure server. After transcription, all participant data were deidentified prior to data analysis.

Setting and Participants

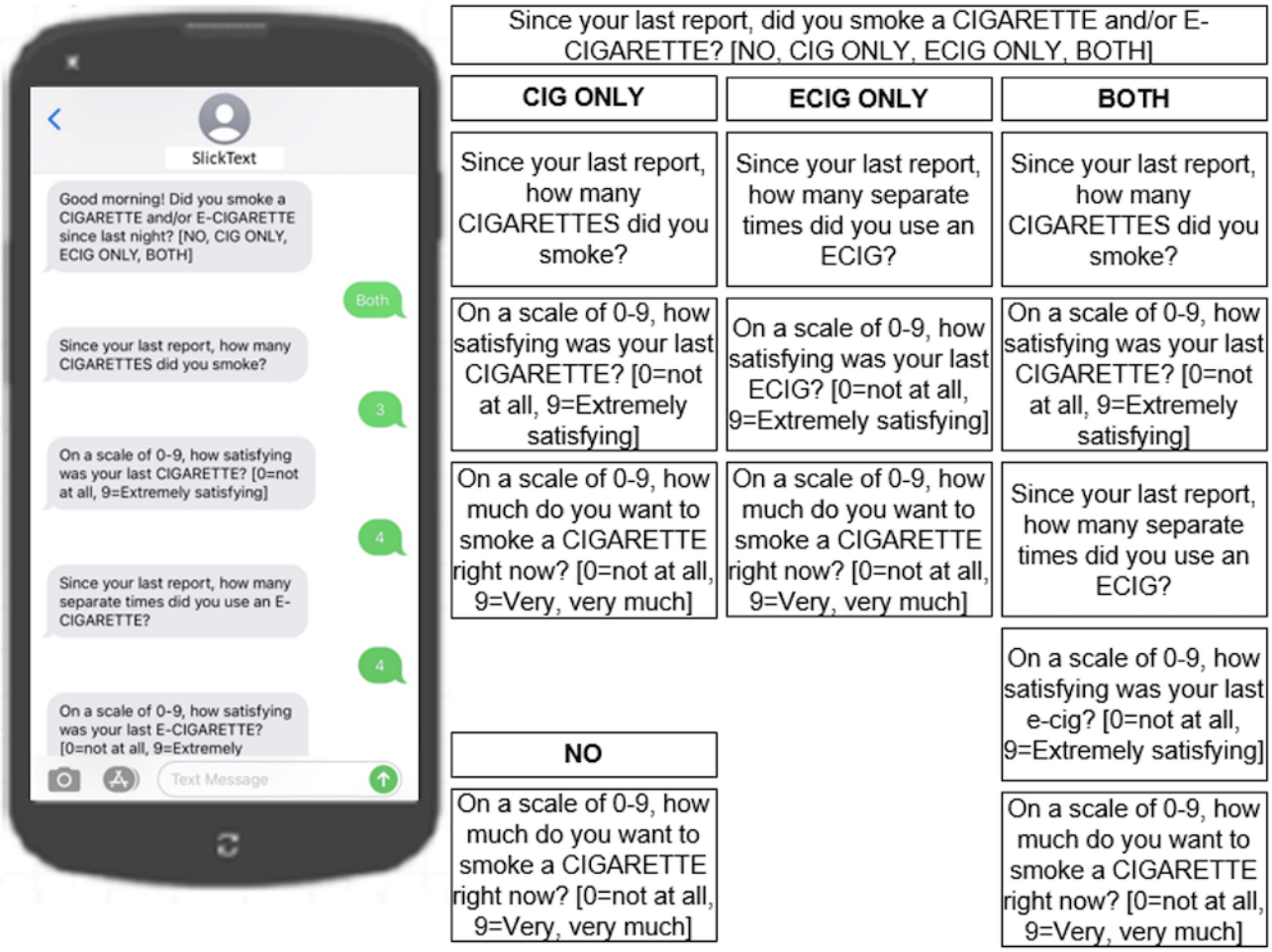
Interview participants were recruited upon completion of the intervention phase of the RCT at the 12-week follow-up study visit. All RCT participants were invited to participate in an in-depth interview to discuss their experiences with the intervention and other aspects of the trial. Interview recruitment ended once thematic saturation was reached. Interviews lasted approximately 30 minutes and were performed between April 2021 and November 2022.

The pilot RCT was performed to determine the feasibility and acceptability of an e-cigarette–based smoking intervention and to compare the effectiveness of counseling and e-cigarette use on smoking outcomes to that of counseling and NRT [16]. Text message–based EMA data collection was used to record smoking patterns. Patients from the electronic health record system of New York University Langone Health—a private hospital system serving New York, New Jersey, and Connecticut—were recruited as RCT participants. The RCT participant sample was initially restricted to patients with a diagnosis of chronic obstructive pulmonary disease, but the scope was later expanded to include patients with a diagnosis of coronary artery disease, peripheral artery disease, or asthma. In addition, to be eligible, RCT participants were required to smoke ≥4 days per week, with at least 5 cigarettes smoked on the days that participants did smoke; be motivated to quit smoking; and possess a phone with SMS text messaging capabilities. A total of 121 participants were recruited into the pilot RCT.

RCT EMA Protocol

In the pilot RCT, combustible cigarette use and e-cigarette use, as well as smoking-related cravings, were measured as part of an EMA protocol, in which SMS text messaging served as a smoking diary. The SMS text messaging was intended for data collection only and not designed to serve as part of the intervention. After a baseline assessment, participants were asked to record their daily nicotine use for 12 weeks by responding to text message prompts that they received 4 times per day. The coverage design prompted participants to provide brief check-in reports via text message over the course of each day, wherein they were asked to report combustible cigarette and e-cigarette use based on their study arm (Figure 1). Responses to 1-item measures of cigarette craving and satisfaction were also collected from each report.

Figure 1. Sample of ecological momentary assessment texting prompts. CIG: cigarette; ECIG: e-cigarette.



Data Collection

Semistructured, in-depth telephone interviews, which lasted approximately 30 minutes, were conducted by research staff (OW and MV). Interviews were audio-recorded, transcribed, and imported into Dedoose software (SocioCultural Research Consultants LLC) for qualitative data analyses.

The interview guide covered topics that were designed for the following goal: gaining a deeper understanding of the participants' experiences, intervention satisfaction, attitudes toward e-cigarettes, and intentions to quit. The interviews were intended to assist with the further adaptation of the e-cigarette smoking intervention and behavioral counseling manual used in the RCT. The interview guide was developed based on the pilot RCT procedures to provide additional insights about the barriers and facilitators of e-cigarette use and how to refine the current approach to enhance program retention and outcomes. Interviews covered topics such as program aspects that the participants liked or disliked, features of the intervention that should be modified, participants' experiences with using e-cigarettes or NRT, intentions of using e-cigarettes after the intervention, and whether participants' health symptoms interfered with their ability to engage in the intervention. The participants were prompted to share their opinions on the EMA texting component of the trial but were not directly asked about the influence of the EMA on their behaviors. Participants were prompted to discuss the texting in the interview, as follows: "During the program, you answered questions over text on daily basis—What was that experience like for you?" This was followed by probes, including "How may texting have affected your overall experience with the program?" A copy of the interview script is available in Multimedia Appendix 1.

Data Analysis

Interview transcripts were coded by using procedures that were designed to ensure thoroughness and reliability. We used Dedoose software to manage the data and coded the data according to the principles of the framework for applied research [18], which consists of the following 5-stage process: familiarization, identifying themes, indexing, charting, and interpretation. Codes were primarily developed a priori based on intervention components and the quality improvement goals of this study. Additional codes were developed by reviewing a random sample of interviews and via discussion with the coding team. The general development of themes arose from the data, using the principles of grounded theory [19]. To enhance reliability, 2 researchers took part in the coding and analysis process for each interview. Prior to full coding, a random sample of interviews was double coded, and intercoder reliability was assessed based on percent agreement (range 79.5%-87.9%). Afterward, all disagreements were discussed as a group to improve concordance among coders. All interviews were then independently coded by 5 coders (GX, KO, RL, RW, and SR) who worked in various pair combinations; each pair met with the other coders and a coauthor (ERS) to resolve discrepancies.

When coding was completed, the quotations for each code were examined, summarized, and grouped together into themes.

Results

Participants

Interviews were performed with a total of 27 participants. Due to an audio malfunction, 1 interview was not included in the analysis. As such, 14 participants were in the e-cigarette study arm and 12 were in the NRT study arm. The average age of the participants included in the analysis was 57.1 (range 28-74) years; 54% (14/26) of participants identified as female, 46% (12/26) were White, 31% (8/26) were Black, 15% (4/26) were Hispanic, and 4% (1/26) were Asian. The majority (17/26, 65%) of participants had completed at least some college; 35% had a high school education or less. Around two-thirds (17/26, 65%) of participants had a diagnosis of chronic obstructive pulmonary disease, 15% (4/26) of participants were diagnosed with coronary artery disease or peripheral artery disease, and 19% (5/26) of participants were diagnosed with asthma.

Themes

Overview of Themes

Without being directly prompted, 18 of the 26 participants described EMA impacting their behaviors or emotions, and several themes indicating a potential, inadvertent influence of EMA on smoking behaviors emerged. The perceived impacts of EMA texting on smoking behaviors were polarized; some participants emphasized the positive impacts of the text messages on their efforts to reduce smoking, while others stressed the ways in which the texting negatively impacted their smoking reduction efforts. These contrasting experiences were captured by 2 major themes and their subthemes, reflecting the positive impacts on smoking behaviors, including increased awareness of smoking behaviors and a sense of accountability, and the negative impacts on emotions and smoking behaviors, including provoking a sense of guilt and triggering smoking behaviors (Table 1).

Table . Major themes and subthemes.

Themes and subthemes	Frequency, n
EMA[a] texting has a positive impact on smoking behaviors	16
EMA texting serves as a source of accountability	13
Anticipation of the next text message serves as a deterrent to impulsive smoking	6
Texting prompts increase awareness of smoking habits	16
Reminders of the goal to reduce cigarette smoking	11
Check-ins serve as markers of progress made toward quitting	5
Negative impact of EMA texting on emotions and smoking behaviors	7
Repeated text messages asking about smoking behaviors produce negative emotions	4
Text messages inquiring about cigarette use may have a triggering effect	5

[a]EMA: ecological momentary assessment.

Perceived Positive Impacts of EMA Texting Procedures

Overview of Positive Impacts

A major theme that arose was the perceived positive impacts of the EMA methods on efforts to reduce smoking, as many (n=16) participants perceived the SMS text messaging smoking diary as an important component of the tobacco treatment intervention and their experience during their efforts to reduce their cigarette smoking. Participants described the EMA text messages as helpful because "you could use [them] for yourself as a tool" (Participant E09), and while stating that "the text messages are a pain in the butt," a participant thought that "they were very useful.…A useful pain in the butt" (Participant E04). Within this overarching theme—the positive influence of EMA—four subthemes that highlighted the potential roles of EMA text messages emerged: EMA text messages (1) increase awareness of smoking habits, (2) serve as reminders of smoking goals, (3) mark progress made, and (4) provide a sense of accountability.

Texting Prompts Increase Participants' Awareness of Their Smoking Habits

Many (n=16) participants noted that the EMA text messages helped them with their smoking behaviors because the act of recording the number of cigarettes smoked increased their awareness of their smoking behaviors. This theme—increased smoking habit awareness—encompassed the following three layers: (1) awareness of the number of cigarettes smoked and smoking habits, (2) reminders of the goal to reduce cigarette smoking, and (3) markers of progress made.

The predictable and repeated EMA text message requests for participants to report the number of cigarettes smoked prompted reflection on their smoking habits. By engaging in the conscious effort of quantifying the number of cigarettes smoked, the participants heightened their self-awareness and gained a clearer

understanding of the frequency of their cigarette use. One participant remarked:

> *[The text message would say] "…have you smoked? When was the last time you smoked your last cigarette?" So, it helped me to be aware of how many cigarettes I was smoking per day.* [Participant N24]

With regard to the act of recording cigarette use, a participant noted:

> *…you're so aware with the…text messages…of just how often you smoke.…Because with the text, I was physically writing it and seeing it.* [Participant E27]

Reminders of the Goal to Reduce Smoking

For some participants (n=11), the EMA text messages also served as reminders of their goal to reduce or quit smoking and heightened their sense of purpose and determination. One participant said, "It [the text messaging] does remind you what you're supposed to be focused on" (Participant N20), and in this way, the messages served as frequent reminders of participants' intentions to change their smoking behaviors. The messages also invigorated their commitment to achieving these goals; one participant said:

> *It [the text messages] kept me going. It kept me wanting to quit, you know, and to keep doing it, to keep at the program.* [Participant N36]

The text messages also kept participants feeling involved:

> *…[questions] like, "how many cigarettes have you smoked?"…kind of keep you involved in it instead of letting it go on the way side. Like it would kind of keep reminding you that…this is what you're working on, you know?* [Participant N50]

Markers of Progress Made Toward Smoking Reduction

The EMA text messages served as markers of progress made toward reducing smoking (n=5). By regularly quantifying and reporting their smoking behaviors, participants saw tangible evidence and took note of the accomplishments resulting from their efforts. By tracking progress over time, participants observed patterns of improvement, which reinforced their motivation to continue cutting back on smoking. When describing why they found the text messages useful, one participant said:

> *…it's like a progress type thing. So, I enjoyed the texts.…It kind of gave me a reminder [of my progress] because as I went on, I had less and less craving.* [Participant N58]

EMA Texting Serves as a Source of Accountability for Progress Toward Reducing Cigarette Smoking

Many participants (n=13) believed that the EMA text messages had a positive effect on their sense of accountability in their efforts to reduce smoking. Knowing that they would receive inquiries about their smoking habits increased participants' mindfulness in reducing their smoking. One participant stated:

> *I loved it [the text messaging] because it kept me…accountable…because it kept asking the questions over and over again. In the beginning, it was like a little stressing because I was like, "Oh, my God, these messages, I don't want to deal with it." But, it kept me accountable. And it was good.* [Participant N36]

Similarly, another participant remarked:

> *I had to answer for all the cigarettes I smoked today and hold myself accountable. I thought that was ingenious.* [Participant N06]

In some instances (n=6), the text messages served as deterrents to impulsive smoking. One participant shared:

> *[There] were a couple of times when they [cigarettes] were right on hand…I'm upstairs, I'm getting ready to light and then boom [sic], oh, "did you smoke today? How many times did you smoke since we last communicated?"* [Participant N31]

Similarly, the anticipation of upcoming EMA messages served as a motivator for refraining from smoking or delaying the next cigarette. The participant went on to describe their morning routine:

> *…while you're trying to wake up and organize yourself, you pick up the cigarette…[but] you know you're going to get a text at 9 o'clock saying, "did you smoke?" right? So that notion pops in your head, so you don't smoke right away.* [Participant N31]

Having the knowledge that there was a regular time for reporting their smoking activities encouraged participants to make more deliberate choices regarding their smoking behaviors. One participant discussed how they used EMA check-ins to consciously reduce their cigarette consumption during the time leading up to the EMA prompt:

> *I knew at a certain time, I was going to get this text. So, when I went to the text, I wanted to have everything in line. I wanted it to be right. So therefore, I would only smoke four cigarettes because that was the allotted cigarettes that I was supposed to smoke at the time. During that time, I would only smoke three cigarettes because I had cut down to three during that period of time and I looked forward to doing it.* [Participant N04]

EMA Text Messages Can Negatively Impact Participants' Emotions and Smoking Behaviors

Overview of Negative Impacts

A second major theme that emerged from the data was the potential negative effects that EMA texting could have on participants. Although many participants reported experiencing positive effects of EMA on their smoking behaviors, some participants (n=7) reflected on the potential negative impacts of the EMA text messages. This theme—the negative effects of EMA—was further distilled into the following two subthemes: the potential roles of EMA in (1) producing negative emotions and (2) triggering cigarette cravings.

Repeated Text Messages Asking About Smoking Behaviors Produce Negative Emotions

Some participants (n=4) mentioned feeling "bad" when they had to admit to smoking or experiencing relapses via the text messages. However, some respondents admitted that the guilt experienced as a result of the text messages, while being a negative emotion, reinforced their determination and prompted them to make renewed efforts, with one participant stating that the text messages "helped me because when I wrote how…I smoked a cigarette…I kind of felt bad" (Participant N58). Moreover, other participants emphasized the negative emotions and guilt experienced when a spotlight was focused on their perceived failures. One participant described the experience by saying, "being able to tell about my progress, or lack thereof, I would have felt bad if I had a slip up" (Participant N21).

Text Messages Inquiring About Cigarette Use May Have a Triggering Effect

Several participants (n=5) expressed their desire to remove the texting component of the program due to its triggering effect on their smoking. Some participants reported that the text messages acted as triggers for cigarette use, as the text messages reminded them of smoking, thereby eliciting an urge to smoke. One participant said that the EMA "was a reminder, actually," and "…with the reminder came in the struggles" (Participant N20). Notably, another participant said:

> *…[the text messages] happened so often, and you knew they were coming, and they started to almost act like a trigger because you sometimes weren't even thinking about [smoking], but then they would ask you about cigarettes and suddenly you're thinking about it…they were more triggers to smoke than to prevent smoking.* [Participant N31]

Discussion

Principal Findings

This study indicates that smoking intervention participants perceive the collection of EMA smoking behavior data via SMS text messaging as a potential influence on their smoking behaviors. The themes developed from the analysis revealed that EMA, in the form of an SMS text messaging smoking diary, may be perceived as a source of accountability for smoking reduction but may also be a trigger for cigarette use among some people. The results of this study emphasize the need to examine the potential influence of EMA data collection techniques on participants' behaviors within smoking interventions, as well as in other behavioral research.

The perceived impacts of EMA on smoking behaviors are consistent with previously made observations that the act of receiving EMA prompts can increase behavioral awareness and act as a trigger for craving [6-8], in addition to altering participants' moods [2,12]. This suggests that EMA for data collection purposes has the potential to unintentionally create an assessment and intervention effects in itself. Although a lack of EMA impact on behavior has been reported in some studies [11,14,15], these studies may be limited by the choice of measures used. As seen in suicide research, EMA prompts have been observed to have an effect on some measures, such as mood, but a minimal effect on other measures, such as suicidal ideation [2]. Therefore, when designing a study, it is important for researchers to reflect on the various factors that may influence the behaviors of interest and consider the potential effect that EMA may have on these factors, in addition to the primary outcomes of interest. Due to its potential effect on participants, investigators should consider and discuss the potential for a behavioral influence to be introduced into a study through the use of EMA data collection.

There is a need for further investigation into the ways that and the degree to which EMA affects participants. Within the EMA literature, there is a general lack of discussion around the effects of EMA on participants' behaviors. When seeking to improve EMA methods, focus is often placed on participant retention and the validity of the data collection method [20-23], with little to no discussion on the potential behavioral impact of EMA. Indeed, when discussing strengths and limitations of EMA, a large portion of the EMA study literature discusses and reports measures of adherence to and reliability of EMA data collection [20-23], with few studies exploring the potential limitation of EMA in which the data collection itself may affect behaviors of interest [2,5-8,11,12,14,15]. A potential effect from EMA may influence the interpretation of the results; therefore, investigators ought to be encouraged to report considerations related to EMA when designing and publishing a study. Future research may benefit from randomizing a subset of participants to receive one EMA modality (eg, texting) while observing behaviors among all participants with another measurement modality (eg, Bluetooth e-cigarette monitor).

When the potential for EMA to influence study outcomes is identified, less obtrusive EMA methods could be considered, when available. The participants in this study expressed the omnipresent awareness and anticipation of the SMS text messaging–based EMA. This awareness altered participants' behaviors and resulted in negative emotions that likely would not have been emphasized had the EMA not been used or had been subtler. There are numerous types of EMA strategies used in smoking research [4]. Future intervention research studies could consider less frequent SMS text messaging or EMA data collection methods outside of SMS text messaging that may have a more minor impact on smoking behaviors, such as the use of biosensors [24], Bluetooth-enabled devices [25], or puff counters [26].

This study had a few limitations. First, the interview guide was not designed to investigate the impact of EMA on participants' smoking behaviors. Therefore, further details on the effects of EMA were not deeply explored, limiting the scope of this analysis. The unprompted nature of the participants' observations of behavioral impact, however, strengthens the conclusion that the SMS text messaging–based EMA had a meaningful impact on the trial participants. Second, the sample of interview participants was not randomized, and the interview was not required; rather, it was offered to all participants sequentially as an optional component. This potentially introduced selection bias, as those with stronger opinions on the program may have been more likely to participate. Third, as EMA data were collected as part of a smoking reduction trial,

it is difficult to completely disentangle the effects of the intervention on changes in behavior from the effects of EMA. Finally, the impact of EMA on behavior change was based on the self-reported perceptions of interview participants, and behavior changes were not directly observed. Therefore, this study cannot be used as conclusive evidence that the EMA had a significant impact on smoking behaviors, and further research is needed.

Conclusion

The collection of EMA smoking behavior data via SMS text messaging may influence the behaviors and perceptions of participants in smoking interventions. More research is needed to determine the magnitude of impact and mechanisms, to account for the potential effects of EMA on behavior change. Furthermore, a broader discussion of the behavioral influence introduced by the use of EMA may be warranted among the EMA research community.

Acknowledgments

Funding was received from internal funds provided by NYU Grossman School of Medicine (principal investigator: SES).

Data Availability

The data sets generated during and/or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

ERS conceived the study, contributed to methods development, participated in data collection and analysis, and led manuscript writing. RL contributed to methods development, participated in data analysis, and contributed to manuscript writing. GX and RW contributed to data collection and analysis and provided manuscript revisions. SR and KO contributed to data analysis and provided manuscript revisions. OW contributed to methods development and data collection and provided manuscript revisions. MV contributed to data collection and provided manuscript revisions. OES contributed to methods development and provided manuscript revisions. SES obtained funding, contributed to methods development, and provided manuscript revisions. All authors reviewed and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Study in-depth interview guide.
[DOCX File, 28 KB - mhealth_v12i1e52122_app1.docx]

Checklist 1
COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist.
[DOCX File, 17 KB - mhealth_v12i1e52122_app2.docx]

References

1. Perski O, Keller J, Kale D, et al. Understanding health behaviours in context: a systematic review and meta-analysis of ecological momentary assessment studies of five key health behaviours. Health Psychol Rev 2022 Dec;16(4):576-601. [doi: 10.1080/17437199.2022.2112258] [Medline: 35975950]
2. Kivelä L, van der Does WAJ, Riese H, Antypa N. Don't miss the moment: a systematic review of ecological momentary assessment in suicide research. Front Digit Health 2022 May 6;4:876595. [doi: 10.3389/fdgth.2022.876595] [Medline: 35601888]
3. Wolfson M, Wagoner KG, Song EY, Pankratz M, Azagba S. Methods for population research on substance use and consequences. In: Nader MA, Hurd YL, editors. Substance Use Disorders: From Etiology to Treatment: Springer International Publishing; 2020:1-29. [doi: 10.1007/978-3-030-33679-0]
4. Vinci C, Haslam A, Lam CY, Kumar S, Wetter DW. The use of ambulatory assessment in smoking cessation. Addict Behav 2018 Aug;83:18-24. [doi: 10.1016/j.addbeh.2018.01.018] [Medline: 29398067]
5. Hufford MR, Shields AL, Shiffman S, Paty J, Balabanis M. Reactivity to ecological momentary assessment: an example using undergraduate problem drinkers. Psychol Addict Behav 2002 Sep;16(3):205-211. [Medline: 12236455]
6. Collins RL, Morsheimer ET, Shiffman S, Paty JA, Gnys M, Papandonatos GD. Ecological momentary assessment in a behavioral drinking moderation training program. Exp Clin Psychopharmacol 1998 Aug;6(3):306-315. [doi: 10.1037//1064-1297.6.3.306] [Medline: 9725114]

7. Betts JM, Dowd AN, Forney M, Hetelekides E, Tiffany ST. A meta-analysis of cue reactivity in tobacco cigarette smokers. Nicotine Tob Res 2021 Jan 22;23(2):249-258. [doi: 10.1093/ntr/ntaa147] [Medline: 32772094]
8. Litt MD, Cooney NL, Morse P. Ecological momentary assessment (EMA) with treated alcoholics: methodological problems and potential solutions. Health Psychol 1998 Jan;17(1):48-52. [doi: 10.1037//0278-6133.17.1.48] [Medline: 9459069]
9. Campbell DT. Factors relevant to the validity of experiments in social settings. Psychol Bull 1957 Jul;54(4):297-312. [doi: 10.1037/h0040950] [Medline: 13465924]
10. Solomon RL. An extension of control group design. Psychol Bull 1949 Mar;46(2):137-150. [doi: 10.1037/h0062958] [Medline: 18116724]
11. McCarthy DE, Minami H, Yeh VM, Bold KW. An experimental investigation of reactivity to ecological momentary assessment frequency among adults trying to quit smoking. Addiction 2015 Oct;110(10):1549-1560. [doi: 10.1111/add.12996] [Medline: 26011583]
12. Dewey D, McDonald MK, Brown WJ, Boyd SJ, Bunnell BE, Schuldberg D. The impact of ecological momentary assessment on posttraumatic stress symptom trajectory. Psychiatry Res 2015 Dec 15;230(2):300-303. [doi: 10.1016/j.psychres.2015.09.009] [Medline: 26381184]
13. Song MK, Ward SE. Assessment effects in educational and psychosocial intervention trials: an important but often-overlooked problem. Res Nurs Health 2015 Jun;38(3):241-247. [doi: 10.1002/nur.21651] [Medline: 25728502]
14. Stone AA, Broderick JE, Schwartz JE, Shiffman S, Litcher-Kelly L, Calvanese P. Intensive momentary reporting of pain with an electronic diary: reactivity, compliance, and patient satisfaction. Pain 2003 Jul;104(1-2):343-351. [doi: 10.1016/s0304-3959(03)00040-x] [Medline: 12855344]
15. Cajita MI, Rathbun SL, Shiffman S, et al. Examining reactivity to intensive longitudinal ecological momentary assessment: 12-month prospective study. Eat Weight Disord 2023 Feb 27;28(1):26. [doi: 10.1007/s40519-023-01556-1] [Medline: 36849665]
16. Stevens ER, Lei L, Cleland CM, et al. Electronic cigarettes as a harm reduction strategy among patients with COPD: protocol for an open-label two arm randomized controlled pilot trial. Addict Sci Clin Pract 2022 Jan 6;17(1):2. [doi: 10.1186/s13722-021-00284-0] [Medline: 34991693]
17. Tong A, Sainsbury P, Craig J. Consolidated Criteria for Reporting Qualitative Research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
18. Ritchie J, Spencer L. Qualitative data analysis for applied policy research. In: Bryman A, Burgess B, editors. Analyzing Qualitative Data: Routledge; 1994:187-208. [doi: 10.4324/9780203413081]
19. Strauss AL, Corbin JM. Basics of Qualitative Research: Grounded Theory Procedures and Techniques: Sage Publications Inc; 1990. ISBN: 0803932502.
20. Jones A, Remmerswaal D, Verveer I, et al. Compliance with ecological momentary assessment protocols in substance users: a meta-analysis. Addiction 2019 Apr;114(4):609-619. [doi: 10.1111/add.14503] [Medline: 30461120]
21. Stone AA, Schneider S, Smyth JM. Evaluation of pressing issues in ecological momentary assessment. Annu Rev Clin Psychol 2023 May 9;19:107-131. [doi: 10.1146/annurev-clinpsy-080921-083128] [Medline: 36475718]
22. Stinson L, Liu Y, Dallery J. Ecological momentary assessment: a systematic review of validity research. Perspect Behav Sci 2022 May 6;45(2):469-493. [doi: 10.1007/s40614-022-00339-w] [Medline: 35719870]
23. Wrzus C, Neubauer AB. Ecological momentary assessment: a meta-analysis on designs, samples, and compliance across research fields. Assessment 2023 Apr;30(3):825-846. [doi: 10.1177/10731911211067538] [Medline: 35016567]
24. Colombo D, Fernández-Álvarez J, Patané A, et al. Current state and future directions of technology-based ecological momentary assessment and intervention for major depressive disorder: a systematic review. J Clin Med 2019 Apr 5;8(4):465. [doi: 10.3390/jcm8040465] [Medline: 30959828]
25. Li Z, Benowitz-Fredericks C, Ling PM, Cohen JE, Thrul J. Assessing young adults' ENDS use via ecological momentary assessment and a smart Bluetooth enabled ENDS device. Nicotine Tob Res 2021 May 4;23(5):842-848. [doi: 10.1093/ntr/ntaa205] [Medline: 33031497]
26. Yingst J, Foulds J, Veldheer S, et al. Measurement of electronic cigarette frequency of use among smokers participating in a randomized controlled trial. Nicotine Tob Res 2020 Apr 21;22(5):699-704. [doi: 10.1093/ntr/nty233] [Medline: 30365024]

Abbreviations

COREQ: Consolidated Criteria for Reporting Qualitative Research
EMA: ecological momentary assessment
NRT: nicotine replacement therapy
RCT: randomized controlled trial

Edited by L Buis; submitted 23.08.23; peer-reviewed by N Baharom, W Evans; revised version received 15.11.23; accepted 22.11.23; published 16.01.24.

Please cite as:
Stevens ER, Li R, Xiang G, Wisniewski R, Rojas S, O'Connor K, Wilker O, Vojjala M, El-Shahawy O, Sherman SE
Trial Participants' Perceptions of the Impact of Ecological Momentary Assessment on Smoking Behaviors: Qualitative Analysis
JMIR Mhealth Uhealth 2024;12:e52122
URL: https://mhealth.jmir.org/2024/1/e52122
doi:10.2196/52122

Original Paper

Investigating Receptivity and Affect Using Machine Learning: Ecological Momentary Assessment and Wearable Sensing Study

Zachary D King[1], MS; Han Yu[1], MS; Thomas Vaessen[2,3], PhD; Inez Myin-Germeys[2], PhD; Akane Sano[1], PhD

[1]Department of Electrical and Computer Engineering, Rice University, Houston, TX, United States

[2]Center For Contextual Psychiatry, Katholieke Universiteit Leuven, Leuven, Belgium

[3]Department of Psychology, Health & Technology, University of Twente, Enschede, Netherlands

Corresponding Author:
Zachary D King, MS
Department of Electrical and Computer Engineering
Rice University
6100 Main St
Houston, TX, 77005
United States
Phone: 1 713 348 0000
Email: zdk2@rice.edu

Abstract

Background: As mobile health (mHealth) studies become increasingly productive owing to the advancements in wearable and mobile sensor technology, our ability to monitor and model human behavior will be constrained by participant receptivity. Many health constructs are dependent on subjective responses, and without such responses, researchers are left with little to no ground truth to accompany our ever-growing biobehavioral data. This issue can significantly impact the quality of a study, particularly for populations known to exhibit lower compliance rates. To address this challenge, researchers have proposed innovative approaches that use machine learning (ML) and sensor data to modify the timing and delivery of surveys. However, an overarching concern is the potential introduction of biases or unintended influences on participants' responses when implementing new survey delivery methods.

Objective: This study aims to demonstrate the potential impact of an ML-based ecological momentary assessment (EMA) delivery system (using receptivity as the predictor variable) on the participants' reported emotional state. We examine the factors that affect participants' receptivity to EMAs in a 10-day wearable and EMA–based emotional state–sensing mHealth study. We study the physiological relationships indicative of receptivity and affect while also analyzing the interaction between the 2 constructs.

Methods: We collected data from 45 healthy participants wearing 2 devices measuring electrodermal activity, accelerometer, electrocardiography, and skin temperature while answering 10 EMAs daily, containing questions about perceived mood. Owing to the nature of our constructs, we can only obtain ground truth measures for both affect and receptivity during responses. Therefore, we used unsupervised and supervised ML methods to infer affect when a participant did not respond. Our unsupervised method used *k*-means clustering to determine the relationship between physiology and receptivity and then inferred the emotional state during nonresponses. For the supervised learning method, we primarily used random forest and neural networks to predict the affect of unlabeled data points as well as receptivity.

Results: Our findings showed that using a receptivity model to trigger EMAs decreased the reported negative affect by >3 points or 0.29 SDs in our self-reported affect measure, scored between 13 and 91. The findings also showed a bimodal distribution of our predicted affect during nonresponses. This indicates that this system initiates EMAs more commonly during states of higher positive emotions.

Conclusions: Our results showed a clear relationship between affect and receptivity. This relationship can affect the efficacy of an mHealth study, particularly those that use an ML algorithm to trigger EMAs. Therefore, we propose that future work should focus on a smart trigger that promotes EMA receptivity without influencing affect during sampled time points.

(JMIR Mhealth Uhealth 2024;12:e46347) doi:10.2196/46347

KEYWORDS

mobile health; mHealth; affect inference; study design; ecological momentary assessment; EMA; just-in-time adaptive interventions; JITAIs; receptivity; mobile phone

Introduction

User Engagement in Mobile Health Systems

Mobile health (mHealth) technologies continue to grow within the health care sector and are imperative for precision medicine initiatives. mHealth can provide beneficial interactions between health care providers and patients outside clinical settings. An engaged and responsive user base in any mHealth system is vital for maximizing the knowledge that researchers and providers acquire. Mental health research mainly depends on active users because investigators rely on participant survey responses to establish ground truth. Researchers can only adequately interpret the relationships between physiology and psychological state with a population that is compliant with sensors and surveys. Evaluating a health construct is only possible with highly receptive participants in mHealth studies.

Here, we discuss 2 forms of interaction between participants and mHealth systems: ecological momentary assessments (EMAs) and just-in-time interventions (JITIs). EMAs gather in situ data from users in real time. EMAs are commonly used in mHealth studies, as they allow researchers to prompt participants regularly throughout the day [1]. In the case of mHealth studies focusing on psychological states, EMAs enable users to report their momentary symptoms or context in a natural environment, often using smartphones, because of their accessibility. JITI is a method that allows investigators to send interventions as needed. The just-in-time adaptive intervention (JITAI) uses incoming information (physiological, contextual, or psychological markers) as context to determine when an intervention is required [2]. Researchers have been working on enhancing the efficiency of these interactions. As mentioned previously, this effort is crucial because ineffective interactions in an mHealth study can have significant effects on outcomes. Failing to collect EMA responses may impede researchers' ability to identify real-world measures of health behaviors, and without participants receiving or engaging in JITIs, researchers may find it challenging to measure the efficacy of the intervention.

Improving EMA Receptivity

To enhance compliance with EMAs and JITIs, it is imperative to gain a comprehensive understanding of the factors that influence participant adherence. Ho and Intille [3] described 11 factors that influence a person's interruptability (willingness to follow through if notified or interrupted). These factors encompass contextual aspects, such as social engagement, ongoing activities, future schedule, and emotional state, as well as message-related attributes, including frequency, complexity, modality, and utility.

Currently, many researchers have reduced interruptability by altering message-related attributes, often involving strategies such as reducing the complexity or frequency of an EMA or increasing the incentives for a response [4,5]. Reducing the size of the instrument relieves some of the burden associated with answering an EMA [6]. This is done by excluding redundant questions or by choosing a less complex instrument. The Perceived Stress Scale [7] was initially a 14-item question set. However, after some statistical analysis, researchers found that a 10-item instrument was sufficient for measuring stress. Another factor affecting receptivity is the frequency at which users are sampled. In 2 separate reviews, researchers demonstrated conflicting findings regarding the effects of frequency on EMA compliance [8,9]. These conflicting results can be attributed to the author's focus on differing populations and the many other factors that play a role in EMA compliance. The third method for improving receptivity rates is to increase the incentives based on EMA compliance. However, this method can be costly and seen as exploitative, especially when dealing with susceptible populations.

An emerging method for improving receptivity rates is the use of machine learning (ML). This can be achieved by using wearable data to predict the likelihood of a response, which can help deliver EMAs that mitigate interruptability. Mishra et al [10] used ML models built from previously collected data to improve the receptivity of a JITAI by contacting users at points where they are more likely to be receptive. The study showed a difference of >38% in receptivity rates between an ML-based static model (using previously collected data) and a control model (using a set schedule) to distribute EMAs. Mishra et al [11] built a model for predicting the optimal time to send an EMA. Their results demonstrated that a model built from contextual cues such as activity, audio, conversation, and location could significantly outperform a baseline model (prediction based on the proportion of responded EMAs). Researchers have also shown that contextual cues, including location [12,13], personality traits [14,15], physical activity [14,16], and time of day [17], influence participants' willingness to respond to regular surveys. Together, these methods can predict and respond to the unobserved contextual aspects of an interruption, thus offering a more holistic approach to addressing participant engagement. However, a system that reacts to these contextual aspects may have unintentional effects on the response of the user. For instance, emotional state is an underlying factor that affects receptivity. A model designed to initiate EMAs when a participant is most likely to respond favors prompting users when experiencing positive emotions. Consequently, this approach could influence the reported emotional state during each prompt, potentially making it challenging to collect subjective responses during negative emotions. Understanding the influence of ML-based EMA triggers on these underlying receptivity factors allows us to incorporate additional variables into an algorithm. Integrating predicted affect into the decision-making of an ML-based EMA trigger will ensure that participants receive prompts across a broad spectrum of emotions.

Relationship Between Affect and EMA Receptivity

Clark and Watson [18] described how positive and negative affect (NA) can influence participation in activities of daily living. Their results show differences in the expected mean across many social activities, with reported positive affect (PA) having more significance in differentiating the 2 groups. Similarly, research has also demonstrated a negative relationship between students' emotional state and academic achievement [19,20]. Although none of these studies demonstrate the relationship between affect and EMA receptivity during mHealth studies, they all demonstrate the effect of emotional state on a participant's general ability to engage in normal activities of daily living.

Several authors have examined the effect of emotional state on EMA adherence by using the preceding response as a gauge of affect during instances of nonresponse. Murray et al [21] conducted a study (N=261) demonstrating that NA and stress reduce the chance of a response during the next prompt. Other researchers have expanded on this by examining various contextual cues within an EMA that precede instances of nonresponse. The authors found that variables such as medication use, activity, battery life, and being away from home negatively impacted the compliance of the following EMA [22,23]. This work contributes to understanding how affect can influence participants' response behavior but falls short of providing real-time explanations for the absence of responses. Alternatively, real-time explanations for receptivity can be derived through passive sensing and ML. Leveraging these explanations allows for delivering EMAs at moments of heightened receptivity, guided by current contextual and physiological factors.

Objectives and Hypothesis

This study aims to analyze the relationship between participant EMA receptivity and affect in a 10-day wearable and EMA–based affect-sensing study (N=45). We hypothesize that a relationship exists between EMA receptivity and affect in mental health–related mHealth studies. We can establish the relationship between emotions when participants respond. However, to investigate this connection during nonresponses, we need to infer affect when a participant fails to provide a response. Therefore, we implemented ML models for identifying receptive time points and predicting emotional states. This allowed us to determine whether there was a statistically significant difference in emotions between responses and nonresponses. If this relationship exists and the likelihood of a response is dependent on emotional state, it would bias the outcome of an ML-based EMA delivery mechanism.

Methods

Ethical Considerations

Ethics approval was granted by the *Sociaal-Maatschappelijke Etische Commissie* of Katholieke Universitei Leuven (G-2018 09 1339) [24]. Informed consents were obtained from the participants. All data was de-identified prior to analysis.

Data Collection

This study included 45 healthy adult participants from Leuven, Belgium [24]. The average age of the participants was 24.5 (SD 3) years and ranged from 19 to 35 years. In total, 84% (38/45) of the participants were female. The participants were recruited via flyers distributed to areas around Leuven.

The study lasted for 10 days. The participants wore a sensor suite (Figure 1), including a chest patch with 2 electrodes for gathering electrocardiography (ECG) at 256 Hz and a wristband for electrodermal activity (EDA) at 256 Hz, skin temperature at 1 Hz, and accelerometer at 32 Hz. Participants were allowed to remove the device while they slept and were asked to remove the devices while bathing or participating in rigorous activities. The sensors had a battery life that surpassed the duration of the study, and the data were recorded on the device on an SD card.

Figure 1. (A) Chest patch for gathering electrocardiogram and accelerometer and (B) wristband for gathering electrodermal activity, skin temperature, and accelerometer.

Participants were given a research phone, and 10 EMAs were sent to the participants daily at random time points between 15 and 90 minutes apart. EMAs were initiated via text messages, and the participants had a specific amount of time to respond to the survey attached to the text message before it closed. The EMAs contained a question set to assess mood [25] in 3 languages: English, Belgian, and French. In total, there were 13 questions, including 9 negative (worried, stressed, anxious, annoyed, down, restless, tense, under pressure, and ashamed) and 4 positive (relaxed, cheerful, confident, and in control) affect-related questions. The questions were prefaced with the phrase "At the moment, I feel...," followed by a rating scale for each emotion, ranging from 1 (not at all) to 7 (very much). The participants were given €0.5 (US $0.54) for each EMA they responded to.

EMA Analysis

Our EMA question set was scored by adding the numerical interpretation of the 9 negative responses to the inverse (1 is 7 and 7 is 1) of the positive questions. The range of possible scores was between 13 and 91, with higher scores indicating more negative emotions. Owing to the low variance in reported positive and NA, we used a composite score of both positive and NA.

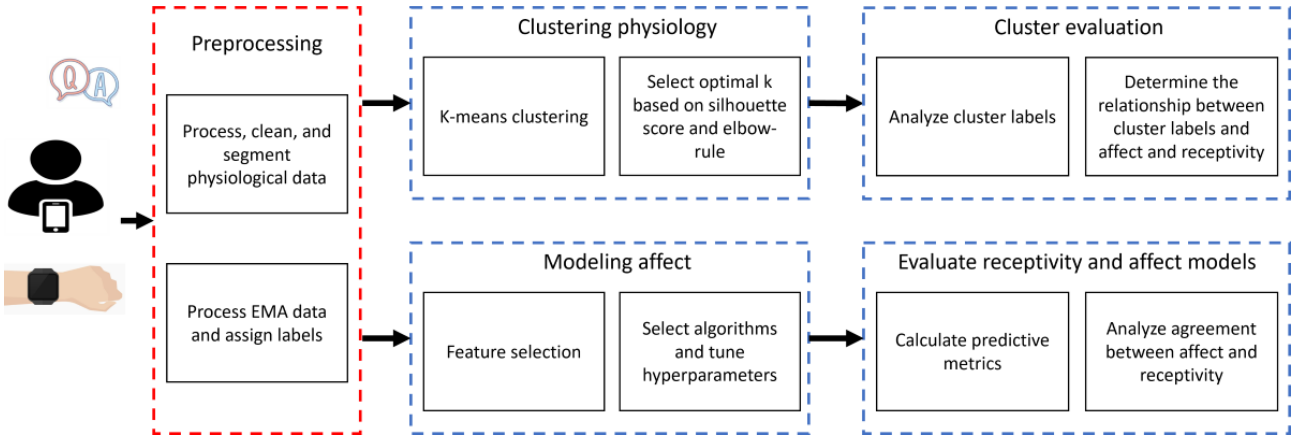
We also analyzed the participants' response time (time between the notification and onset of EMA) and the response rate to EMA. We then investigated the potential for loss of engagement over time, which may lead to reduced participant receptivity. The lack of engagement may impede our capacity to discern the underlying causes of nonresponsiveness, particularly when assessing the relationship between affect and receptivity.

EMA Receptivity and Affect Detection Models

Overview

In the following sections, we discuss the sequential methodology, which encompasses the collection of raw signal data, the subsequent data processing and feature extraction, and the design of ML models for inferring 2 constructs—receptivity and affect. This framework is shown in Figure 2.

Figure 2. Methodology used from the raw signals to our evaluation of the relationship between affect and receptivity. EMA: ecological momentary assessment.



We began by processing our 4 sets of time series data: skin temperature, ECG, EDA, and accelerometer. Once we processed the data, we segmented them and attached labels to each segment based on the conditions explained in the *EMA Receptivity Labels* section. Next, we built and tested multiple ML algorithms to infer EMA receptivity and affect and verified the results using several statistical techniques.

Preprocessing

Time Series Processing

We began by extracting all the data from the 4 time series data sets. Table 1 shows the features computed for the 4 sets of the data. We used IQR to process skin temperature to remove outliers. We used *biosppy* [26] for the ECG to process the data and extract the R peaks. *Biosppy* uses a bandpass filter with frequencies of 3 Hz and 45 Hz, a sampling rate of 256, and the Hamilton segmentation algorithm to extract R peaks. We then validated the R peaks using an algorithm by Hovsepian et al [27]; this algorithm uses the criterion beat difference based on the maximum expected difference for a beat and the minimal artifact difference. We then used heart rate variability analysis to extract heart rate and heart rate variability features such as number of pairs of successive normal-to-normal intervals that differ by more than 20 ms and root mean square of successive differences between normal heartbeats [28]. We also obtained some frequency- and geometric-based features. For EDA, we used the method proposed by Taylor et al [29] to process and extract the statistical and wavelet features. Finally, for accelerometer, we smoothed the signal by using a fourth-order 10-Hz low-pass Butterworth filter and obtaining an average, and then, we used a package from the study by Simon [30] to extract step features. The features we extracted and the information on how those features were calculated are shown in Table 1.

Table 1. Features from our 3 raw sources and definitions of the features that are less commonly used.

Signal	Features	Description	Prior work
ST[a]	Mean, median, mode, minimum, range, root mean square, zero cross, kurtosis, skew, and IQR (25th percentile and 75th percentile)	Zero cross here is based on the number of times ST crosses over the mean ST. Kurtosis measures the extremity of the data in the segment, and skew is the measure of asymmetry.	[31]
ECG[b]	Mean, median, mode, minimum, range, root mean square, zero cross, kurtosis, skew, IQR (25th percentile and 75th percentile), RMSSD[c], CVSD[d], CVNNI[e] SDNN[f], NNI50[g], NNI20[h], PNNI50[i], PNNI20[j], LF[k], VLF[l], HF[m], high/low-frequency ratio	Normal to Normal or RR[n] interval indicates time between heartbeats. NNI20 or NNI50 refers to the number of successive intervals that differ by more than 20 or 50 ms. "P" indicates the proportion of NNI20 or NNI50 in the segment. RMSSD is the root mean square of successive differences between heartbeats. CVNNI and CVSD are the coefficients of variation SDNN/mean and RMSSD/mean, respectively. Our frequency domain features are based on how much of the signal lies between 0.003 and 0.04 Hz (VLF), 0.04 and 0.15 Hz (LF), and 0.15 and 0.40 Hz (HF).	[32-35]
Electrodermal activity	*Wavelet:* maximum, mean, SD, median, and above zero (1-second and half-second wavelet); *raw*: amplitude, maximum, minimum, and mean; and *filtered*: amplitude, maximum, minimum, and average	A 1-second and a half-second window were used for wavelet features. Features were calculated for both the first and second derivatives of each window size.	[31,36-39]

[a]ST: skin temperature.

[b]ECG: electrocardiography.

[c]RMSSD: root mean square of successive differences between normal heartbeats.

[d]CVSD: coefficient of variation of differences between adjacent normal-to-normal intervals.

[e]CVNNI: coefficient of variation of the normal-to-normal intervals.

[f]SDNN: SD of the normal-to-normal intervals.

[g]NNI50: number of pairs of adjacent normal-to-normal intervals differing by more than 50 ms.

[h]NNI20: number of pairs of adjacent normal-to-normal intervals differing by more than 20 ms.

[i]PNNI50: percentage of pairs of adjacent normal-to-normal intervals differing by more than 50 ms.

[j]PNNI20: percentage of pairs of adjacent normal-to-normal intervals differing by more than 20 ms.

[k]LF: low frequency.

[l]VLF: very low frequency.

[m]HF: high frequency.

[n]RR: R-peak to R-peak.

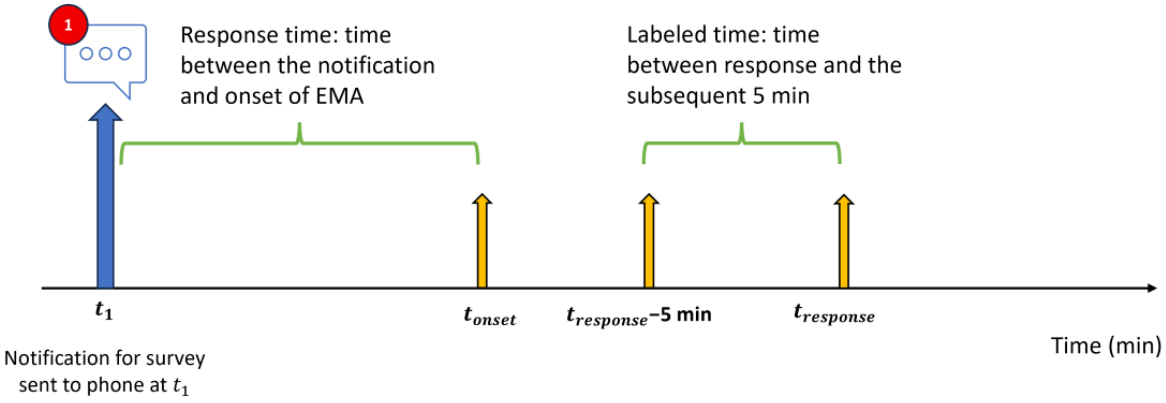
Segmentation

We segmented the data into 1-minute windows with a 30-second overlap. We then calculated the statistical features for each of the sensors, excluding steps. For each of these windows, we calculated historic features. To do so, we elongated each of the windows by 5, 30, and 60 minutes and then extracted the features with the extended window size (ie, for each 1-min window, we have not only the features from the 1 min but also the features going back to these 4 time frames).

EMA Receptivity Labels

Labels for receptivity were based on whether the user responded to the EMA and were assigned to segments based on whether it was within a specified time of the scheduled notification. By expanding the window of labeled data, we can increase the size of the labeled data set (pseudolabeling). However, as this window increases, so does the distance between some of our time points and the corresponding label. We tested windows that are 5, 30, 60, and 120 minutes long. For instance, for the 5-minute window, if an EMA was sent at midnight, the segments that fell between 11:55 AM and midnight would be labeled "responded" if they did respond and "no response" if they did not. We applied the same method for the affect labels (Figure 3). We ultimately chose 30-minute windows owing to the balance between the size of the training set and the labeled points being relatively close in terms of time to the actual response (or nonresponse).

Figure 3. Representation of our labeled segments for the 5-minute window. It also demonstrates how we calculate response time (time between notification [t1]) and the start of the ecological momentary assessment (EMA; tonset).



Affect Labels

Previous research on monitoring and tracking emotional states using wearables has commonly used binary or categorical affect measures to detect emotions [40]. These psychological instruments often feature well-defined categorical score representations, which make it easier to distinguish between emotions. The distribution of the reported composite affect scores made defining an adequate categorization of the labels difficult. Most participants reported positive emotional states, which complicated setting an appropriate cutoff. Setting the cutoff at a high value would result in an imbalanced set of labels, whereas selecting a lower value would create a balanced data set but lack logical consistency. For instance, choosing a cutoff of 26 to distinguish between positive and negative emotions would lead to a balanced data set. However, the range of possible responses was between 13 and 91, so a response would be considered negative even if the participant indicated relatively positive or neutral emotions.

In response to these challenges, we used recorded composite affect values as our class labels and designed our ML algorithms as a regression problem. Although this method of affect inference is less commonly found in the literature, it prevents the need for arbitrary data classification. Given that the data exhibited an inherent imbalance, with less frequent occurrences of negative emotional states, using regression may still affect our ability to predict these less common negative emotions.

Analysis of the Relationship Between Features and Receptivity and Affect

We examined the significance of each feature in terms of its ability to predict affect and receptivity. To do so, we conducted a repeated measures ANOVA test to assess how well each feature is related to the response class labels. In addition, we used a linear mixed model (LMM) to investigate the relationship between features and affect scores. We used an LMM because we worked with constant labels instead of converting the affect score into binary or categorical values, as done for receptivity. The dependent variable (affect score) was regressed on the fixed effect variable (features), while accounting for random effects (participant ID). These tests help identify any features or signals that may have significance in determining receptivity or affect.

Receptivity and Affect Model Design and Hyperparameter Tuning

We designed ML models to infer EMA receptivity and affect. A wide variety of ML algorithms are used in affect and receptivity prediction including random forest (RF) [31,32,39], support vector machine [33,34,39], logistic regression, *k*-nearest neighbors [30], neural network (NN; long short-term memory, recurrent NN, convolutional NN, etc) [31,39], and naive Bayes [39,41]. On the basis of our sensor data, initial tests, and drawing inspiration from previous studies, especially those by Mishra et al [10,11]. We selected (1) RF for predicting emotional state and receptivity, (2) an NN for predicting emotional state, and (3) a baseline model. This baseline model serves as a benchmark for evaluating whether our models outperform random chance, whereas the NN algorithm was introduced as a possible improvement on existing model implementations. Unlike the research mentioned previously, we used physiological data rather than contextual data. These signals are sampled at higher frequencies compared with contextual data and allow the extraction of more fine-grained features, making NNs more feasible. We designed personalized models to infer the receptivity and effect of EMA.

To optimize our personalized model, we selected hyperparameters using the grid search method for each participant, explicitly using the *GridSearchCV* method defined in *scikit-learn*. This method uses an exhaustive search method (ie, testing each user-defined parameter permutation). The hyperparameters tested included the number of estimators, maximum depth of the estimator, minimum number of samples per leaf, minimum number of samples for split, and maximum number of features that can be used for the split. Using training and validation sets, we selected the parameters and then applied the optimal model to our test set. The optimal set of hyperparameters differed for each participant, although the most common optimal hyperparameters chosen included 60 estimators, maximum depth=3, minimum sample leaf=2, minimum sample split=2, and maximum features=square root of the number of features.

Our NN model was structured to use 3 densely connected layers using a rectified linear unit activation function at each layer. The output dimension of each layer was 256, 128, and 64, and the output layer was a densely connected layer with 2 output

dimensions. The reasoning for an output layer of 2 is to define a CI for our regression model.

The baseline model was built by predicting random output based on the distribution of the class labels in the training set (ie, if 10/100, 10% of the labels were nonresponses and 90/100, 90% were responses, the model would predict nonresponses 10/100, 10% of the time). We can determine the expected outputs for this model; our true positive rate should be equal to *Pr(response in the training set)* × *Pr(response in the test set)*. The more evenly the class labels are distributed, the worse the performance of the model. For the affect regression models, we used a normal sampling method with the mean and SD based on the training set class labels.

As there are more labeled responses compared with nonresponses, we considered this imbalance in the receptivity prediction model, weighting the classes based on the distribution in our training set. All models were built using the Python packages *scikit-learn* [42] or *Tensorflow* [43].

Model Uncertainty

To determine the relationship between affect and receptivity, we must use predictions to infer the emotional state of our participants during nonresponses. As affect is a complex and difficult-to-predict construct, we need a method for filtering our predictions based on some level of confidence. Therefore, we introduced a method for calculating uncertainty for regression using an NN.

Determining a confidence value for a regression model is difficult compared with a binary or categorical model. We can use a custom loss function in our NN to estimate epistemic and aleatoric uncertainty for our regression model, where epistemic uncertainty is based on our ability to predict our class labels with the available data (affected by lack of knowledge or data), and the aleatoric uncertainty is affected by randomness, which is unknown or unmeasured in the model [44].

Our affect prediction model outputs are 2D rather than a single predicted output. The first output is the predicted affect, $\mu(x)$, and the second output, $\ln(\sigma(x))$, is the predicted variance (the log allows us to take the exponent to ensure a positive value for σ). Both μ and σ are functions of our training set x.

The loss function L is shown in the equation (custom loss function for measuring model uncertainty) and is derived from the mean square error (MSE) calculation and the maximum likelihood of a normal Gaussian distribution [45]. The numerator of this equation is identical to the MSE loss function, where $\mu(x)$ is the predicted output of our model. Unlike the MSE loss function, we continuously update not only our predicted output μ but also the predicted variance σ. The σ output of our model is based on error; the sigma value increases to account for higher error and decreases to account for lower error. This σ value can be used as an uncertainty or error metric. Although it is still a predicted value, it should align with how confident the model is in the $\sigma(x)$ output. The σ value plays a crucial role in assessing the confidence of our affect predictions, given that we use predicted affect to infer emotional states during nonresponses. Consequently, to illustrate the relationship between the predicted sigma value and model uncertainty, we performed a mixed effect model analysis using affect scores and the predicted sigma values and tested whether greater uncertainty will occur in emotional states that are less frequently represented and when the testing error is larger. As uncertainty is a measure of the model's confidence in its predictions, we can reasonably assume that predictions associated with larger testing errors would correspond to higher levels of uncertainty.

Model Evaluation

For cross-validation, we used a personalized random train-test split cross-validation method. We randomly split the data into training and testing sets using the response label (whether they responded to the EMA or not) to stratify the split. Responses and nonresponses can encompass multiple segments; by grouping them together, we avoid splitting up segments from a single response or nonresponse. As our response labels are unbalanced, we want to ensure that our training, validation, and test sets have a relatively even number of responses and nonresponses. For the purpose of fairness, we excluded 3 participants who had a single nonresponse from our receptivity results.

We first normalized the training and test sets independently of one another based on the participant. In total, we obtained approximately 230 features from the sensor signals. We reduced our feature set using principal component analysis. Our implemented principal component analysis was set such that the number of produced components explained 99% of the variance (48 features). This method was used for each model, excluding the RF model, in which the original normalized data were used as the input.

Analysis of the Relationship Between Affect and Receptivity

Overview

We conducted two different analyses to understand the relationship between affect and receptivity better:

1. To infer emotional state during nonresponses, we clustered the physiological data and then examined the makeup of the clusters. By doing so, we can assume the emotional state of different clusters and unlabeled data points.
2. For EMAs the participants did not respond to, we used the affect prediction model described in the previous section to infer the emotional state at the time of a nonresponse. With these newly predicted affect scores, we can analyze the differences in the emotional state during a response and nonresponse.

Cluster Evaluation

We used the most significant features (based on correlation) when predicting receptivity for our clustering analysis. To determine the optimal clustering method, we tested several clustering methods, including hierarchical and *k*-means clustering, with a maximum number of iterations of 300. We then calculated the silhouette score across all clusters using receptivity as our ground truth and selected our best-performing set of hyperparameters. On the basis of the cluster distribution,

we analyzed the difference in the perceived emotional state of the participants. We calculated the average NA, PA, and receptivity rates in the clusters for each participant and then characterized the clusters based on receptivity rates (high receptive and low receptive clusters). Next, using repeated measure ANOVA, we demonstrated the statistical difference between affect and the clusters. Given that the clusters were created from physiological data, we know that the data points within each cluster are physiologically similar; therefore, we inferred that they would also exhibit similar psychological states. This allowed us to assign affect scores to nonresponsive data points within each cluster based on the labeled data points within that cluster. Unlike affect prediction, we used the raw NA and PA values in our evaluation as the clustering was performed independently of affect scores; therefore, the lack of variance in responses did not affect the output of the clustering. These results gave us a sense of participants' perceived emotional state during nonresponses. We also investigated differences in receptivity in 2 clusters using the chi-square test.

Analysis of the Receptivity and Affect Relationship

Ideally, we would show the interaction between affect and receptivity using the data collected. However, because nonresponses do not have a corresponding affect score, we designed and implemented our models for receptivity and emotional state.

After generating predictions for our test data set, we assessed the agreement (using Cohen κ) and correlation (using the point biserial method) between receptivity and predicted affect, leveraging true labels at time points when affect measures were reported. A high level of agreement or correlation would suggest a strong relationship between these 2 constructs, thereby highlighting the potential influence each construct would have on an ML algorithm to predict the other construct. We then examined the disparities between the predicted affect during nonresponses and the reported affect during responses. By doing so, we can establish the extent to which emotional state influences receptivity. Substantial disparities in affect between responses and nonresponses suggest that participants' emotional states impact their receptivity. Consequently, a model designed to predict receptivity would indirectly include emotional state as a determinant of a participant's receptiveness. However, it is essential to acknowledge that some of these variations could be attributed to model error. As a result, we also compared the predicted and reported affects during responses to investigate the significance of the model error. We then calculated and visualized the cumulative distribution of these 3 sets of values to illustrate the influence of affect on receptivity and the associated model error.

Finally, we investigated the potential effects that an ML-based receptivity algorithm would have on reported affect, influencing the outcome of the study. On the basis of our receptivity model, we can estimate the difference in the reported perceived emotional state between our true findings and predicted affect during time points that would initiate an EMA.

Results

In the following sections, we discuss the results of our study, particularly the methods of evaluation that were discussed in the previous section.

EMA Analysis: Affect and Receptivity

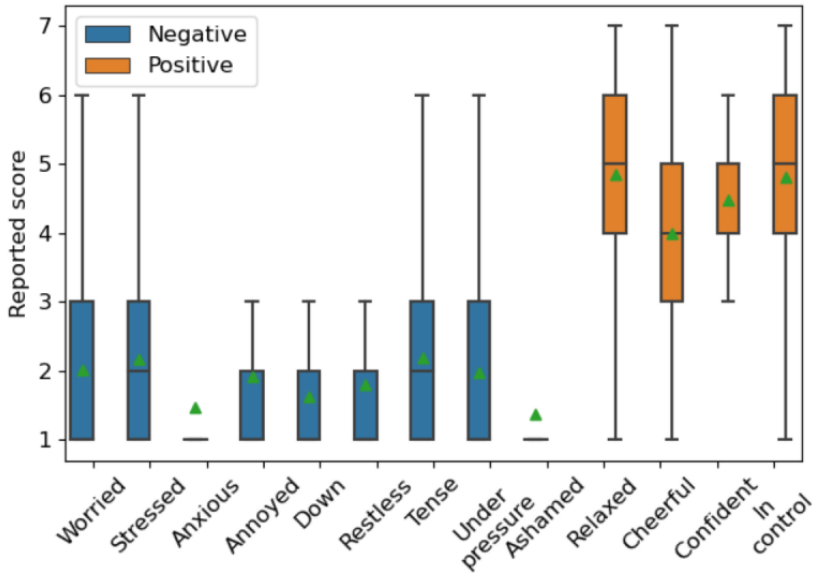
The distribution of EMA responses is shown in Figure 4. Although participants rarely indicated high negative emotions, this trend is evident in Figure 5, illustrating a box plot of composite scores for each participant. Participants' average and median reported affect were <26, meaning that, on average, the participant responded to each question with a relatively low score of 2 (on a scale between 1 and 7, where 1 indicates high positive emotion and 7 shows high negative emotion). We also investigated participants' emotional states as the study advanced and observed minimal to no variations based on their duration of enrollment or time of day.

On average, participants responded with a 4.5 for PA questions and a 1.8 for NA questions. This disparity in affect intensity was consistent with previous research [23]. There was a slight difference in the reported affect between male and female participants. On average, female participants responded with a 1.9 (SD 1.08) for the NA questions and 4.5 (SD 1.3) for the PA questions, whereas male participants responded with a 1.8 (SD 0.9) for the NA questions and 4.7 (SD 1.0) for the PA questions.

Of the 3885 notifications sent to the 45 participants, there were 3066 (78.92%) responses. As the study persisted, there was little to no drop-off in receptivity rates over time. This finding helped confirm that loss of engagement was not a contributing factor to receptivity. Most studies have stated that the quality receptivity rate is at 80%. The range of response time (time between notification and initiation of the EMA; Figure 3) was between 0.5 seconds and 306 seconds. Participants responded to the notification on average in 20.9 seconds and had a median response time of 8.7 seconds. There were no responses after 306 seconds of a notification. The reason for this fast response time is that participants were allowed 90 seconds to begin the survey, after which the survey would no longer be accessible (we had a few responses after the 90-second restriction owing to software or design issues). This restriction makes it challenging to relate response times to participant affect, as has been done by other researchers.

We found that none of the mood responses were strongly correlated with the time to respond. Across each question, we did not obtain a correlation coefficient >0.03 (all correlations indicated significant confidence; $P<.05$). This low correlation coefficient indicates that the participant's mood had little to do with how long it took the participant to initiate the EMA. Although considering the limit we put on the response time, this relationship might be difficult to assume.

Figure 4. Question set: includes the 13 questions used to measure affect with their mean, SD, and correlation to the final affect score. For each question, participants were asked to rate the degree they were experiencing each emotion. These 13 questions can be split into positive affect (orange) and negative affect (blue).



Analysis of Features

The features that we found to be the most significantly related to receptivity were ECG low frequency (1 min, momentary: $F_{2,54}$=6.7; P<.001) and very low–frequency features (1 min, momentary: $F_{2,54}$=4.7; P=.02 and 60 min: $F_{2,54}$=4.1; P<.001); EDA mean ($F_{2,54}$=10.2; P<.001) and median ($F_{2,54}$=15.4; P<.001); number of pairs of adjacent NN intervals differing by more than 20 ms in the 5 and 60 minute windows; percentage of pairs of adjacent NN intervals differing by more than 50 ms in the 30 minute window ($F_{2,54}$=11.2; P<.001); and maximum ($F_{2,54}$=6.3; P<.001), minimum ($F_{2,54}$=3.6; P=.009), and absolute maximum ($F_{2,54}$=6.6; P=.002) of the first and second derivatives for EDA. These results show that ECG and EDA-related features were best at differentiating between responses and nonresponses compared with features derived from accelerometer and skin temperature.

When running the LMMs to determine the relationship between features and emotional state, we found a nonsignificant relationship between affect scores and steps or sleep features. However, heart rate was significant when predicting emotional state, particularly negative emotion. This LMM showed a significant positive relationship between heart rate and affect (β=.007; P<.04). This underscores the significance of heart rate as a predictor of emotional state, although it does not necessarily imply that steps and sleep features lack importance in this context.

Receptivity and Affect Models

After processing, cleaning, and filtering out segments with confounding values, we obtained 1368 responses with usable physiological data. As our class labels were expanded to include segments 30 minutes before the point of response (pseudo labeling), we ended up with 13,477 data points for determining affect and 17,254 data points for predicting response.

Model Performance

Table 2 shows the results of our receptivity (binary) and affect (regression) models. On the basis of these results, there was little difference between the RF and NN models, although we used the NN models to demonstrate the relationship between affect and receptivity in the following section.

Table 2. Model results for predicting receptivity (binary) and affect (regression).

Model	Accuracy, mean (SD)	Precision, mean (SD)	Recall, mean (SD)	F_1-score, mean (SD)	Root mean square error (SD)
Baseline	0.73 (0.001)	0.83 (0.002)	0.84 (0.002)	0.83 (0.002)	11.1 (4.3)
Neural network	0.84 (0.19)	0.82 (0.006)	0.85 (0.10)	0.86 (0.20)	7.3 (2.7)
Random forest	0.83 (0.11)	0.82 (0.15)	0.94 (0.10)	0.87 (0.12)	7.5 (3.1)

Analyzing Uncertainty in the Affect Model

Figure 6A shows the relationship between the calculated sigma value (uncertainty) and the reported affect scores. Uncertainty should follow a pattern where class labels that are more represented in the training set should have lower uncertainty. Conversely, values that are less represented in the data set should have larger uncertainty. As can be seen, Figure 6A σ values are smaller when the reported emotional state is more positive. As shown in Figure 5, most respondents indicated relatively low composite scores, with a few participants reporting an affect score >40. We also observed a statistically significant relationship between sigma and affect scores, as shown in Figure 6A, using a mixed effect model. In this model, we accounted

for the random effect associated with participants, as indicated by the mixed linear model results (intercept: 7.090; P<.001 and affect score: 0.002; P=.046).

Figure 6B shows the relationship between σ and the testing error; in particular, σ values were larger when the model was further from the ground truth. This relationship shows that our σ value is an accurate representation of model uncertainty. On the basis of Figure 6, we can say that the σ value we calculated is related in some way to uncertainty. Figure 6B shows that most responses indicating an affect score of <39 had a σ of <6. Therefore, we chose 6 as the cutoff for uncertainty. This cutoff filters out many of the predictions that are more likely to have higher errors because we cannot look at errors during nonresponses, as we have no affect label.

Figure 5. Box plot of perceived emotional state, minimum is 13 (negative) and maximum is 91 (positive). The average perceived emotional state is 26.42, denoted by the blue horizontal line.

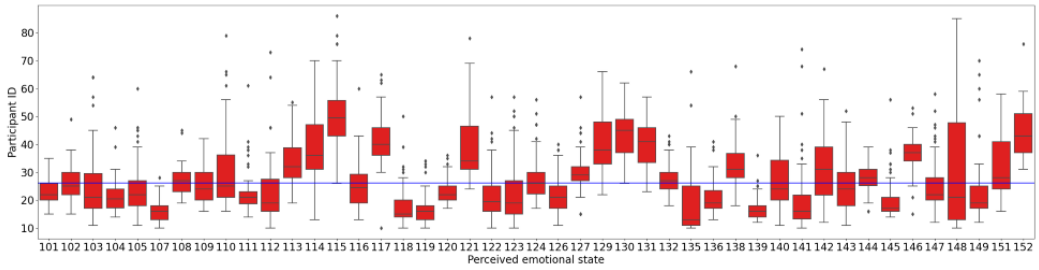
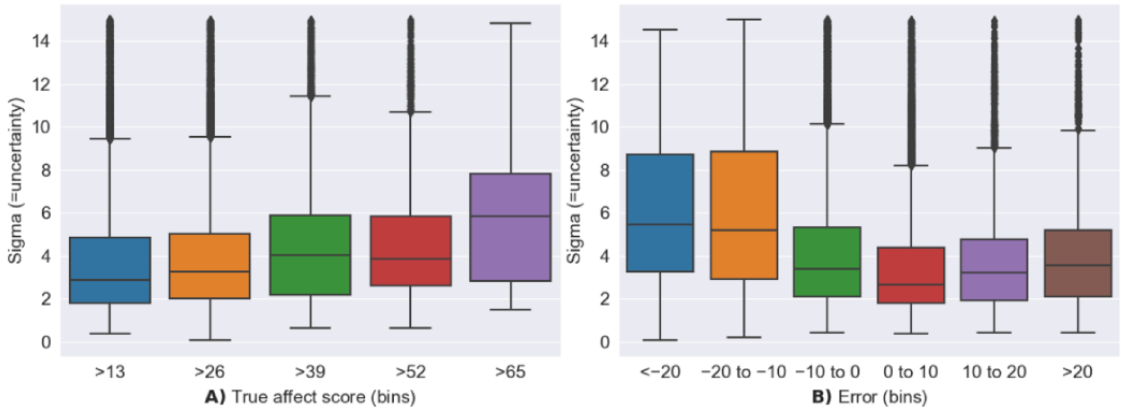


Figure 6. Box plot showing the σ value for (A) true labels and (B) predicted error.



Receptivity and Affect Analyses

Cluster Analysis

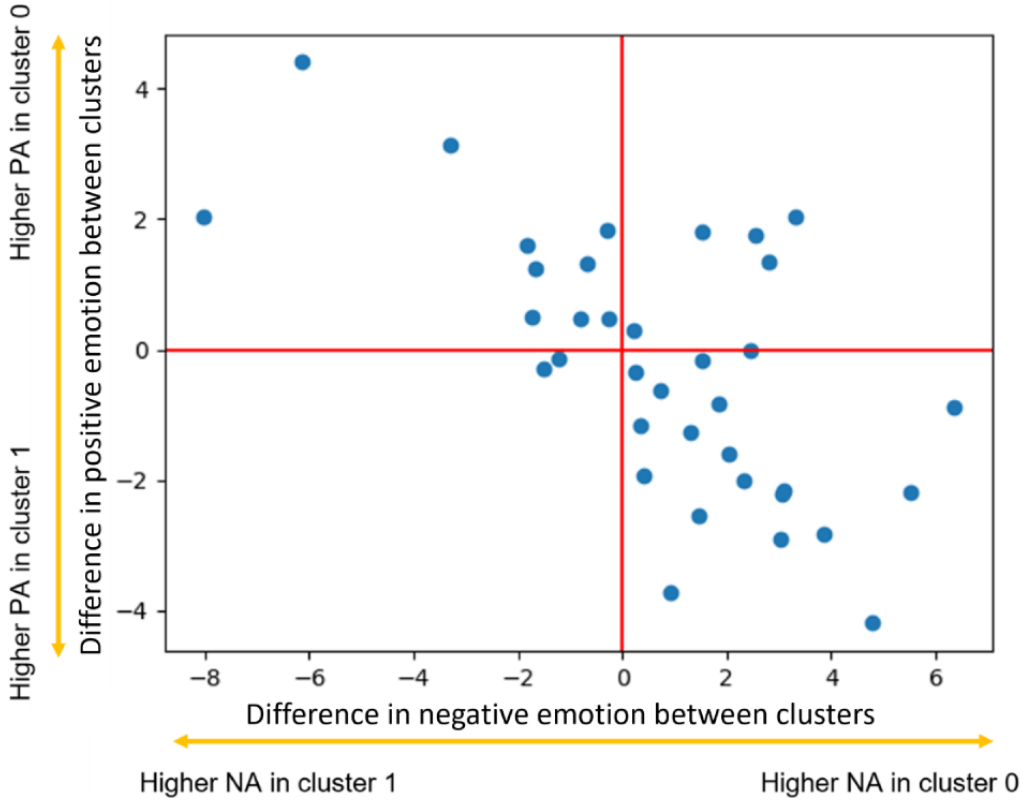
On the basis of the "elbow rule" of silhouette scores, we chose *k*-means as our clustering method with 2 clusters. We found that the distribution of receptivity was somewhat different between clusters. Cluster 0 contained a higher density of responses, with just <15% nonresponses, whereas cluster 1 had a higher density of nonresponses of just >21%. We first analyzed the overall affect scores in the 2 clusters, where we found the average reported affect score in cluster 1 to be >3 points higher than the average reported affect in cluster 0 (repeated measure ANOVA, F_2=23.16; P<.001). The receptivity rates and average reported affect scores for the 2 clusters are shown in Table 3. We also found that the distribution of receptivity was different between the 2 clusters using the chi-square test of independence (χ^2_2=898.8; P<.001). These results indicate distinctions between response and affect across the cluster labels. Considering that the cluster with a higher density of nonresponses (cluster 1) also had a higher average affect score (higher scores indicate more intense negative emotions or lower positive emotions), we can assume that there was a relationship between EMA receptivity and reported affect.

Figure 7 shows a scatter plot of the difference in perceived PA between the 2 clusters and the difference in perceived NA between the 2 clusters for each participant. The results show that participants' perceived emotion was more negative regarding lower PA and higher NA in cluster 1 compared with their perceived emotional state in cluster 0. As stated earlier, cluster 1 contains a higher percentage of nonresponses compared with cluster 0, indicating that cluster 1 is a better representation of a nonresponse. Therefore, it appears that there is a relationship between negative perceived emotional state and receptivity. Using the cluster labels as groups, we calculated the F test statistic using an ANOVA test for each feature. The features that separated the 2 clusters were mostly calculated from the ECG signal, including the minimum heart rate, low or very low frequency, mean heart rate, coefficient of variation of the NN intervals, coefficient of variation of differences between adjacent NN intervals, high frequency, and maximum heart rate (in order of F_1-score). Features obtained from the EDA, accelerometer, and body temperature did not return significant P values when calculating the F test statistic.

Table 3. Receptivity rates and average reported affect scores in each cluster.

Cluster number	Receptivity rate	Reported affect score, mean (SD)
Cluster 0	0.85	24.3 (4.7)
Cluster 1	0.78	27.3 (4.9)

Figure 7. Each point represents a participant, where the x-axis denotes the difference between average negative affect (NA) of cluster 0 and 1, whereas the y-axis represents the difference between average positive affect (PA) of clusters 0 and 1.



Relationship and Analysis Between Receptivity and Affect

Figure 8 shows the cumulative distribution of reported affect scores for responses and predicted affect scores for responses and nonresponses. On the basis of this figure, there is a clear difference between the predicted affect during nonresponses and our true affect scores. Although this could be a model error, we also predicted affect scores during these responses and found that our model consistently predicted lower affect values (higher PA).

There was a fair amount of agreement between our affect and our binary response model, with a Cohen κ score of 0.33 and a correlation of 0.44. When our model predicted a response, 77.42% (22,761/29,399) of the segments were during times when the affect model predicted PA. Only 69.72% (7760/11,131) of the predicted nonresponses reported PA. This indicates that the predicted response is negatively related to affect (ie, responses are associated with PA, whereas nonresponses are associated with NA). The reason determining the relationship between our constructs is important is because this bias can, and as we show, affect the overall outcome of a study. For instance, the average predicted affect score for times that we predicted as low likelihood for a response was a full 1.5 (SD 1.35) or 2.01 points higher than the average predicted affect for points predicted to be of high likelihood for a response. When observing only the segments where we misclassified a response (ie, we had a true affect, but the response was misclassified as a nonresponse), we found that the average affect score dipped slightly from 26.1 (predicted nonresponse) to 25 (predicted response). This difference in affect between responses and nonresponses is evidence that our receptivity model is indirectly based on affect. The SD of the affect score also decreased from 11.1 (true labels) to 9.8 (true affect and predicted response) during responses.

The average predicted affect score for a nonresponse was 30.9 (SD 11.2), and the average affect score for a response was 29.3 (SD 10.7; true) and 27.7 (SD 8.9; predicted). The predicted affect scores during nonresponses were higher than the reported and predicted affect scores during responses. Given that our average testing error was −1.6, we could also assume that the predicted affect during these nonresponses could be more negative than the true predictions. The distribution of these scores is shown in Figure 9. In Figure 9, all 3 groups' affect scores peaked at around 20 to 25; this is probably owing to the large number of reported affect scores in this range. However, nonresponses had a second peak at an affect score of 40. This bimodal distribution could indicate that our affect distribution

during nonresponse was affected by ≥2 factors. Some nonresponses may not be affected by their affect but perhaps by their daily life activities (seeing a movie, spending time with family, showering, etc). In contrast, the second peak indicates that NA is related to nonresponses.

Figure 8. Cumulative distribution of predicted and actual affect scores for responses and nonresponses.

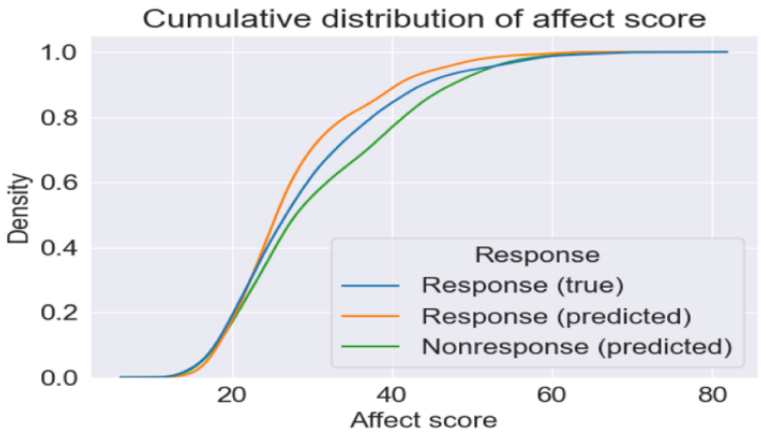
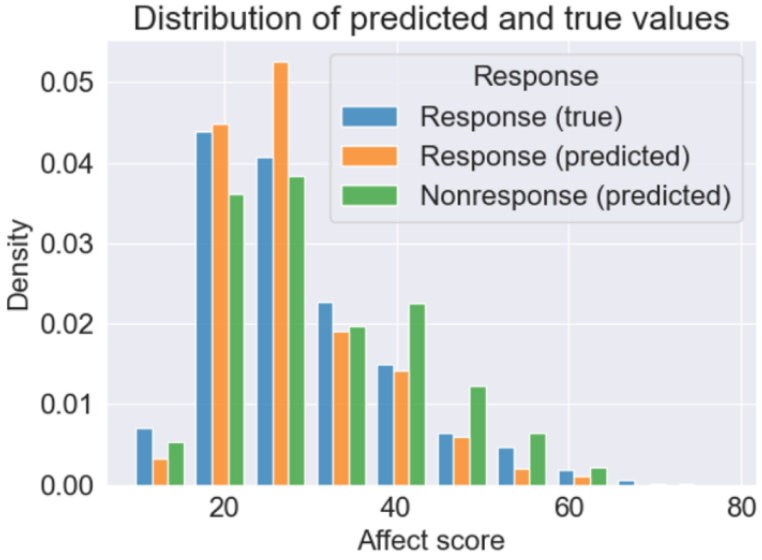


Figure 9. Distribution of predicted and true affect scores for responses and nonresponses. Density is specific to the response and nonresponse.



Discussion

In this section, we discuss the outcome of our study, particularly the relationship between emotional state and receptivity, what that means, how it affects our results, and how we might implement a receptivity model that removes this bias. We also mention the limitations of this study.

Principal Findings

This study aimed to understand how ML models used to improve participant receptivity can affect the outcome of a study. Although we focus on emotional state in this study, we feel as if there are many health constructs and outcomes that can be affected by these receptivity models. Improving receptivity is not a new concept, but in the realm of mHealth, it is an emerging problem. The factors influencing study adherence have been analyzed and discussed in depth in previous research. One such scope is in medication adherence. Researchers have found many factors that influence medication adherence, from social, therapeutic, patient-related, and disease-related factors [46]. However, few have examined the momentary factors that affect adherence to medication or a health construct, and few have had the ability to do so without wearable sensors and momentary assessments.

Our findings using supervised learning and clustering indicate a clear relationship between emotional state and user receptivity. The clustering method demonstrated clear differences in affect between a highly receptive cluster and a less receptive cluster. The results of supervised learning demonstrate that users experience more negative emotions during nonreceptive time points. Although our results showed promise for a model dedicated to predicting response, we also showed the biases inherent in such a model. Ideally, we would want a receptivity model that is completely independent of emotion. Otherwise, we are influencing the participant's responses.

Our results demonstrate that an mHealth study implementing a receptivity trigger based purely on the likelihood of responding (a model that triggers EMAs and JITIs using predicted receptivity) will bias the participant's response. In this case, the model would initiate an EMA or JITI during times of more positive emotions, thereby decreasing the overall affect score

for the EMA and possibly sending the JITI during times when the intended construct was not being met. As our ability to predict binary affect is limited with this data set, we believe that using the affect regression and ground truth labels for responses will return the most realistic representation of affect during nonresponses.

Comparison With Prior Work

Our findings are consistent with those of the previous studies. Many prompt-level studies [21-23] found a relationship between nonresponses and higher levels of NA in previous prompts. Although these results can provide insight into what makes a participant less compliant with EMAs, they do not offer a reasonable method for using this information in real-time decision-making. Using ML, wearable sensors, and contextual cues allows researchers to predict noncompliance components and distribute EMAs accordingly.

Consistently, our models either surpassed or achieved equivalent performance compared with previous research efforts. We achieved F_1-scores ranging from 0.83 to 0.87 when predicting receptivity. In contrast, Künzler et al [14] reported F_1-scores of approximately 0.4 while relying solely on contextual features. It is important to note that these results are not directly comparable, as contextual data lack the granularity of the data collected in our study.

Regarding affect prediction, our results present a unique challenge for comparison because we used regression in our predictions, unlike most researchers who typically use binary or categorical labels for emotion recognition. We chose not to convert our ground truth data into binary or categorical labels because of the inherent ambiguity in setting the thresholds and the limited variance in user responses. The effectiveness of affect prediction can vary significantly depending on the specific construct of interest and the sensors and signals available. Schmidt et al [40] conducted a review and reported an emotion recognition accuracy ranging from 40% to 95% using wearable sensors and signals. In terms of regression analysis, Tuarob et al [47] achieved nearly identical root MSE scores (PA=7.37; NA=7.40) when forecasting positive and NA scores from Positive and Negative Affect Scale using RF regression and previously collected questionnaire data.

Limitations

In this section, we address the limitations of our study, which can be categorized as limitations in our population, study design, data collection, and affect prediction models.

The major concern of our study population is that our results may be specific to this cohort. The study population was very receptive, even with 10 EMAs sent daily. This could be difficult for other researchers to implement, as the frequency and complexity of the EMA are fairly burdensome. Although we believe that the relationship between affect and receptivity would extend to other studies, it is important to note that our population was relatively small (N=45), predominately young (age 24.5 y), and had a higher representation of female participants (38/45, 84%). Consequently, our results may be specific to our cohort and EMA question set, but previous studies analyzing medication adherence and prompt-level relationships between EMAs and nonresponses indicate that the effect of emotional state on receptivity is common across multiple populations [21-23]. Further research is needed to explore the extent of this relationship between different emotional states and receptivity across multiple populations.

One limitation of the study design is that we cannot examine how loss of engagement over time affects the relationship between emotional state and receptivity. There was little to no drop-off in receptivity rates as our study progressed. This may have been because of the relatively short time frame in which the participants were enrolled. As a result, it is difficult to explore the effect emotional state would have on EMAs in the latter part of a study when participants can be more fatigued and less engaged. In future work, we intend to study a population for an extended period to analyze how emotional state affects participant response rates later in the study. Ideally, this will allow us to see the rate at which responses decay, the causes, and how we might combat it. Furthermore, we believe that a measure of this decay in engagement could be included in our ML-based decision-making for delivering EMAs that mitigate study fatigue, similar to how we would use model uncertainty to diversify emotional response.

Another potential study design limitation is that the app and research phone were shared with participants. Carrying 2 phones, especially one dedicated solely to responding to EMAs, can be burdensome for participants. In addition, the app designed for EMA distribution requires further usability evaluation. In future work, we aim to develop an app that can seamlessly integrate into users' devices and assess its ease of use.

The data gathered in this study were limited to physiological features and user-defined responses. Although the physiological features make up a large portion of what researchers consider important for predicting psychological constructs, the data set lacks sampling contextual data. Certain contextual information is imperative for recognizing emotions and improving EMA response rates that cannot be obtained using physiology, such as social context. The social context can help infer the participant's emotional state and willingness to respond to an EMA or JITAI.

Similarly, by incorporating more psychological and environmental cues (personality traits, working hours, etc), we can better understand what to expect from our participants regarding receptivity and affect before the start of the study. Using these prestudy measurements, we could assess the type of participants enrolled. Specifically, what will be their needs regarding receiving and responding to EMAs. This will help us develop and personalize our ML models for affect and receptivity.

The last significant limitation of our study is the use of predicted affect labels in determining the relationship between emotional state and receptivity. We can never collect reported affect during nonresponses for this or any data set. We attempt to reduce this limitation by using uncertainty to filter out less-confident predictions. Nevertheless, the predicted affect is only as good as our models. The only way to overcome this limitation is to improve the affect models. Although some may argue that the quality of our models needs to be more robust to claim a

relationship between affect and receptivity, the effects of emotional state on engagement in social and daily life activities are well documented and consistent with our conclusion.

Conclusions

This paper presents the possibilities for bias in ML models to trigger surveys and interventions for participants in mHealth studies. Our results show a clear relationship between emotional state and user EMA receptivity. By designing an mHealth study using a "trigger" to improve participant response, it is imperative to consider some biases that may arise, in this case, affect. Participants were more likely to respond to an EMA during positive emotional states. If we distribute those EMAs to times when they are more likely to respond, we would further be biasing our participants' recorded emotional state. Although this may not be a significant problem for less responsive populations, for the general population, this could change researchers' perception of the participant's perceived emotional state. In this study, we did not examine other constructs that might be a factor of receptivity because affect is the focal point of this study. For this objective, we are collecting both subjective and physiological data. Although this may be broad, it can be applied to any construct, particularly the intended construct of an mHealth study.

The pitfall of any mHealth study, particularly those involving psychological concepts, is the dependency on subjective user responses. The sampling rate of subjective responses will always be less than that of the physiological sensors and even some contextual cues. As our feature set became increasingly comprehensive, our labeled data remained relatively sparse. Considering that our proposed trigger considers factors beyond receptivity, it would likely have lower receptivity rates compared with triggers solely based on receptivity. However, the importance of even a minimal increase in a user's adherence or engagement in a study can drastically improve researchers' understanding of the health construct.

The models discussed in this paper have mostly proposed single-objective optimization functions that try to optimize based on whether the model considers that a user will respond to an EMA. In future work, we will propose a multiobjective optimization function for triggering EMAs and JITAIs based on the likelihood of responding and an active-learning measurement of the health construct. This multiobjective function would base the timing of the EMAs on 2 separate objectives: receptivity and model uncertainty. By initiating EMAs or JITAIs based on these 2 objectives, we can obtain an expected response that is more diverse in terms of affect. We hope that the work presented in this paper can be used to further enhance communication and the ability to gain knowledge from participants.

Conflicts of Interest

AS has received travel reimbursement or honorarium payments from Leuven Mindgate, American Epilepsy Society, the Institute of Electrical and Electronics Engineers, and Apple. AS has also received research support from Microsoft, Sony Corporation, NEC Corporation, Pola Chemicals, and Meta and consulting fees from Gideon Health and Suntory Global Innovation Center. All other authors declare no other conflicts of interest.

References

1. Arthur AS. Ecological momentary assessment (ema) in behavorial medicine. Ann Behav Med 1994;16(3):199-202 [FREE Full text] [doi: 10.1093/abm/16.3.199]
2. Margaret M, Stephen SI, Jennifer SB. Embedded assessment: overcoming barriers to early detection with pervasive computing. In: Proceedings of the 3rd International Conference on Pervasive Computing. 2005 Presented at: PERVASIVE '05; May 8-13, 2005; Munich, Germany p. 333-346 URL: https://link.springer.com/chapter/10.1007/11428572_20 [doi: 10.1007/11428572_20]
3. Ho J, Intille SS. Using context-aware computing to reduce the perceived burden of interruptions from mobile devices. In: Proceedings of the 2005 SIGCHI Conference on Human Factors in Computing Systems. 2005 Presented at: CHI '05; April 2-7, 2005; Portland, OR p. 909-918 URL: https://dl.acm.org/doi/10.1145/1054972.1055100 [doi: 10.1145/1054972.1055100]
4. Intille S, Haynes C, Maniar D, Ponnada A, Manjourides J. µEMA: microinteraction-based ecological momentary assessment (EMA) using a smartwatch. Proc ACM Int Conf Ubiquitous Comput 2016 Sep;2016:1124-1128 [FREE Full text] [doi: 10.1145/2971648.2971717] [Medline: 30238088]
5. King ZD, Moskowitz J, Egilmez B, Zhang S, Zhang L, Bass M, et al. micro-stress EMA: a passive sensing framework for detecting in-the-wild stress in pregnant mothers. Proc ACM Interact Mob Wearable Ubiquitous Technol 2019 Sep 09;3(3):1-22 [FREE Full text] [doi: 10.1145/3351249] [Medline: 32432212]
6. Ponnada A, Haynes C, Maniar D, Manjourides J, Intille S. Microinteraction ecological momentary assessment response rates: effect of microinteractions or the smartwatch? Proc ACM Interact Mob Wearable Ubiquitous Technol 2017 Sep 11;1(3):1-16 [FREE Full text] [doi: 10.1145/3130957] [Medline: 30198012]
7. Sheldon C. Perceived stress in a probability sample of the united states. In: Spacapan S, Oskamp S, editors. The Social Psychology of Health. Thousand Oaks, CA: Sage Publications; 1988:31-67.
8. Wen CK, Schneider S, Stone AA, Spruijt-Metz D. Compliance with mobile ecological momentary assessment protocols in children and adolescents: a systematic review and meta-analysis. J Med Internet Res 2017 Apr 26;19(4):e132 [FREE Full text] [doi: 10.2196/jmir.6641] [Medline: 28446418]

9. Jones A, Remmerswaal D, Verveer I, Robinson E, Franken IH, Wen CK, et al. Compliance with ecological momentary assessment protocols in substance users: a meta-analysis. Addiction 2019 Apr 21;114(4):609-619 [FREE Full text] [doi: 10.1111/add.14503] [Medline: 30461120]
10. Mishra V, Künzler F, Kramer J, Fleisch E, Kowatsch T, Kotz D. Detecting receptivity for mHealth interventions in the natural environment. Proc ACM Interact Mob Wearable Ubiquitous Technol 2021 Jun 24;5(2):1-24 [FREE Full text] [doi: 10.1145/3463492] [Medline: 34926979]
11. Mishra V, Lowens BM, Lord SE, Caine KE, Kotz D. Investigating contextual cues as indicators for ema delivery. In: Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers. 2017 Presented at: UbiComp '17; September 11-15, 2017; Maui, HI p. 935-940 URL: https://dl.acm.org/doi/10.1145/3123024.3124571 [doi: 10.1145/3123024.3124571]
12. Morrison LG, Hargood C, Pejovic V, Geraghty AW, Lloyd S, Goodman N, et al. The effect of timing and frequency of push notifications on usage of a smartphone-based stress management intervention: an exploratory trial. PLoS One 2017 Jan 3;12(1):e0169162 [FREE Full text] [doi: 10.1371/journal.pone.0169162] [Medline: 28046034]
13. Pielot M, Cardoso B, Katevas K, Serrà J, Matic A, Oliver N. Beyond interruptibility: predicting opportune moments to engage mobile phone users. Proc ACM Interact Mob Wearable Ubiquitous Technol 2017 Sep 11;1(3):1-25 [FREE Full text] [doi: 10.1145/3130956]
14. Künzler F, Mishra V, Kramer J, Kotz D, Fleisch E, Kowatsch T. Exploring the sstate-of-receptivity for mHealth Interventions. Proc ACM Interact Mob Wearable Ubiquitous Technol 2019 Dec 14;3(4):1-27 [FREE Full text] [doi: 10.1145/3369805] [Medline: 36159209]
15. Abhinav M, Veljko P, Jo V, Robert H, Mirco M. My phone and me: understanding people's receptivity to mobile notifications. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. 2016 Presented at: CHI '16; May 7-12, 2016; San Jose, CA p. 1012-1032 URL: https://dl.acm.org/doi/10.1145/2858036.2858566 [doi: 10.1145/2858036.2858566]
16. Abhinav M, Mirco M, Robert H, Veljko P. Designing content-driven intelligent notification mechanisms for mobile applications. In: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous. 2015 Presented at: UbiComp '15; September 7-11, 2015; Osaka, Japan p. 813-824 URL: https://dl.acm.org/doi/10.1145/2750858.2807544 [doi: 10.1145/2750858.2807544]
17. Bidargaddi N, Almirall D, Murphy S, Nahum-Shani I, Kovalcik M, Pituch T, et al. To prompt or not to prompt? A microrandomized trial of time-varying push notifications to increase proximal engagement with a mobile health app. JMIR Mhealth Uhealth 2018 Nov 29;6(11):e10123 [FREE Full text] [doi: 10.2196/10123] [Medline: 30497999]
18. Clark LA, Watson D. Mood and the mundane: relations between daily life events and self-reported mood. J Pers Soc Psychol 1988;54(2):296-308. [doi: 10.1037/0022-3514.54.2.296]
19. Duchesne S, Vitaro F, Larose S, Tremblay RE. Trajectories of anxiety during elementary-school years and the prediction of high school noncompletion. J Youth Adolesc 2007 Sep 19;37(9):1134-1146. [doi: 10.1007/s10964-007-9224-0]
20. Valiente C, Swanson J, Eisenberg N. Linking students' emotions and academic achievement: when and why emotions matter. Child Dev Perspect 2012 Jun 15;6(2):129-135 [FREE Full text] [doi: 10.1111/j.1750-8606.2011.00192.x] [Medline: 23115577]
21. Murray AL, Brown R, Zhu X, Speyer LG, Yang Y, Xiao Z, et al. Prompt-level predictors of compliance in an ecological momentary assessment study of young adults' mental health. J Affect Disord 2023 Feb 01;322:125-131 [FREE Full text] [doi: 10.1016/j.jad.2022.11.014] [Medline: 36372127]
22. Ponnada A, Li J, Wang S, Wang WL, Do B, Dunton GF, et al. Contextual biases in microinteraction ecological momentary assessment (μEMA) non-response. Proc ACM Interact Mob Wearable Ubiquitous Technol 2022 Mar 29;6(1):1-24 [FREE Full text] [doi: 10.1145/3517259]
23. Rintala A, Wampers M, Myin-Germeys I, Viechtbauer W. Momentary predictors of compliance in studies using the experience sampling method. Psychiatry Res 2020 Feb 29;286:112896. [doi: 10.1016/j.psychres.2020.112896] [Medline: 32146247]
24. De Calheiros Velozo J, Vaessen T, Lafit G, Claes S, Myin-Germeys I. Is daily-life stress reactivity a measure of stress recovery? An investigation of laboratory and daily-life stress. Stress Health 2023 Aug 20;39(3):638-650. [doi: 10.1002/smi.3213] [Medline: 36521434]
25. Myin-Germeys I, van Os J, Schwartz JE, Stone AA, Delespaul PA. Emotional reactivity to daily life stress in psychosis. Arch Gen Psychiatry 2001 Dec 01;58(12):1137-1144. [doi: 10.1001/archpsyc.58.12.1137] [Medline: 11735842]
26. Carlos C, Ana PA, André L, Filipe C, Hugo S, Ana F. BioSPPy: Biosignal processing in Python, 2015? Online; accessed 2015.
27. Hovsepian K, al'Absi M, Ertin E, Kamarck T, Nakajima M, Kumar S. cStress: towards a gold standard for continuous stress assessment in the mobile environment. Proc ACM Int Conf Ubiquitous Comput 2015 Sep;2015:493-504 [FREE Full text] [doi: 10.1145/2750858.2807526] [Medline: 26543926]
28. Pichot V, Roche F, Celle S, Barthélémy JC, Chouchou F. HRVanalysis: a free software for analyzing cardiac autonomic activity. Front Physiol 2016 Nov 22;7:557 [FREE Full text] [doi: 10.3389/fphys.2016.00557] [Medline: 27920726]

29. Taylor S, Jaques N, Chen W, Fedor S, Sano A, Picard R. Automatic identification of artifacts in electrodermal activity data. Annu Int Conf IEEE Eng Med Biol Soc 2015;2015:1934-1937 [FREE Full text] [doi: 10.1109/EMBC.2015.7318762] [Medline: 26736662]
30. Simon H. Sensor motion documentation. Read the Docs. 2018 Nov 18. URL: https://readthedocs.org/projects/sensormotion/downloads/pdf/latest/ [accessed 2023-12-11]
31. Lam H, Tri N, Thu N, Susanna P, Pekka S. Stressnas: affect state and stress detection using neural architecture search. In: Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers. 2021 Presented at: UbiComp/ISWC '21; September 21-26, 2021; Virtual Event p. 121-125 URL: https://dl.acm.org/doi/10.1145/3460418.3479320 [doi: 10.1145/3460418.3479320]
32. Ha U, Madani S, Adib F. WiStress: contactless stress monitoring using wireless signals. Proc ACM Interact Mob Wearable Ubiquitous Technol 2021 Sep 14;5(3):1-37 [FREE Full text] [doi: 10.1145/3478121]
33. Cheng H, Yun-jin Y, Xue-song Y. An emotion recognition system based on physiological signals obtained by wearable sensors. In: Yang C, Virk GS, Yang H, editors. Wearable Sensors and Robots. Singapore, Singapore: Springer; 2017:15-25.
34. Hu L, Yang J, Chen M, Qian Y, Rodrigues JJ. SCAI-SVSC: smart clothing for effective interaction with a sustainable vital sign collection. Future Gener Comput Syst 2018 Sep;86:329-338. [doi: 10.1016/j.future.2018.03.042]
35. Kanlaya R, Michiko O, Makoto M. Emotion monitoring from physiological signals for service robots in the living space. In: Proceedings of the 2010 International Conference on Control, Automation and Systems. 2010 Presented at: ICCAS '10; October 27-30, 2010; Gyeonggi-do, South Korea p. 580-583 URL: https://ieeexplore.ieee.org/document/5669914 [doi: 10.1109/iccas.2010.5669914]
36. Nalepa GJ, Kutt K, Giżycka B, Jemioło P, Bobek S. Analysis and use of the emotional context with wearable devices for games and intelligent assistants. Sensors (Basel) 2019 May 31;19(11):2509 [FREE Full text] [doi: 10.3390/s19112509] [Medline: 31159317]
37. Ragot M, Martin N, Em S, Pallamin N, Diverrez JM. Emotion recognition using physiological signals: laboratory vs. wearable sensors. In: Proceedings of the AHFE 2017 International Conference on Advances in Human Factors and Wearable Technologies and Game Design. 2017 Presented at: AFHE '17; July 17-21, 2017; Los Angeles, CA p. 15-22 URL: https://link.springer.com/chapter/10.1007/978-3-319-60639-2_2 [doi: 10.1007/978-3-319-60639-2_2]
38. Sano A, Picard RW. Stress recognition using wearable sensors and mobile phones. In: Proceedings of the 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction. 2013 Presented at: ACII '13; September 2-5, 2013; Geneva, Switzerland p. 671-676 URL: https://ieeexplore.ieee.org/abstract/document/6681508 [doi: 10.1109/acii.2013.117]
39. Bobo Z, Zhu W, Zhiwen Y, Bin G. Emotionsense: emotion recognition based on wearable wristband. In: Proceedings of the 2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation. 2018 Presented at: SmartWorld '18; October 7-11, 2018; Guangzhou, China p. 346-355 URL: https://ieeexplore.ieee.org/document/8560069 [doi: 10.1109/smartworld.2018.00091]
40. Schmidt P, Reiss A, Dürichen R, Laerhoven KV. Wearable-based affect recognition-a review. Sensors (Basel) 2019 Sep 20;19(19):4079 [FREE Full text] [doi: 10.3390/s19194079] [Medline: 31547220]
41. Zhang J, Yin Z, Chen P, Nichele S. Emotion recognition using multi-modal data and machine learning techniques: a tutorial and review. Inf Fusion 2020 Jul;59:103-126. [doi: 10.1016/j.inffus.2020.01.011]
42. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. J Mach Learn Res 2011;12:2825-2830 [FREE Full text]
43. Martín A, Ashish A, Paul B, Eugene B, Zhifeng C, Craig C, et al. Tensorflow: large-scale machine learning on heterogeneous distributed systems. arXiv. Preprint posted online March 14, 2016 2016 [FREE Full text] [doi: 10.48550/arXiv.1603.04467]
44. Kiureghian AD, Ditlevsen O. Aleatory or epistemic? Does it matter? Struct Saf 2009 Mar;31(2):105-112. [doi: 10.1016/j.strusafe.2008.06.020]
45. Valdenegro-Toro M, Mori DS. A deeper look into aleatoric and epistemic uncertainty disentanglement. In: Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2022 Presented at: CVPRW '22; June 18-24, 2022; New Orleans, LA p. 1508-1516 URL: https://www.computer.org/csdl/proceedings-article/cvprw/2022/873900b508/1G56BW7QpEs [doi: 10.1109/cvprw56347.2022.00157]
46. Gast A, Mathes T. Medication adherence influencing factors-an (updated) overview of systematic reviews. Syst Rev 2019 May 10;8(1):112 [FREE Full text] [doi: 10.1186/s13643-019-1014-8] [Medline: 31077247]
47. Tuarob S, Tucker CS, Kumara S, Giles CL, Pincus AL, Conroy DE, et al. How are you feeling?: a personalized methodology for predicting mental states from temporally observable physical and behavioral information. J Biomed Inform 2017 Apr;68:1-19 [FREE Full text] [doi: 10.1016/j.jbi.2017.02.010] [Medline: 28213145]

Abbreviations

ECG: electrocardiography

EDA: electrodermal activity
EMA: ecological momentary assessment
JITAI: just-in-time adaptive intervention
JITI: just-in-time intervention
LMM: linear mixed model
mHealth: mobile health
ML: machine learning
MSE: mean square error
NA: negative affect
NN: neural network
PA: positive affect
RF: random forest

Edited by L Buis; submitted 08.02.23; peer-reviewed by R Pryss, S Schneider, S Jeong; comments to author 24.08.23; revised version received 01.11.23; accepted 21.11.23; published 07.02.24.

Please cite as:
King ZD, Yu H, Vaessen T, Myin-Germeys I, Sano A
Investigating Receptivity and Affect Using Machine Learning: Ecological Momentary Assessment and Wearable Sensing Study
JMIR Mhealth Uhealth 2024;12:e46347
URL: https://mhealth.jmir.org/2024/1/e46347
doi:10.2196/46347
PMID:38324358

Original Paper

Ecological Momentary Assessment of Alcohol Marketing Exposure, Alcohol Use, and Purchases Among University Students: Prospective Cohort Study

Min Jin Zhang[1], MPH; Tzu Tsun Luk[1], PhD; Sai Yin Ho[2], PhD; Man Ping Wang[1], PhD; Tai Hing Lam[2], MD; Yee Tak Derek Cheung[1], PhD

[1]School of Nursing, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China (Hong Kong)

[2]School of Public Health, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China (Hong Kong)

Corresponding Author:
Yee Tak Derek Cheung, PhD
School of Nursing
Li Ka Shing Faculty of Medicine
The University of Hong Kong
5/F, Academic Building 3 Sassoon Road, Pokfulam
Hong Kong
China (Hong Kong)
Phone: 852 39176652
Fax: 852 28726079
Email: derekcheung@hku.hk

Abstract

Background: The relationships between alcohol marketing exposure, alcohol use, and purchase have been widely studied. However, prospective studies examining the causal relationships in real-world settings using mobile health tools are limited.

Objective: We used ecological momentary assessment (EMA) to examine both the within-person– and between-person–level effects of alcohol marketing exposure on any alcohol use, amount of alcohol use, any alcohol purchase, and frequency of alcohol purchase among university students.

Methods: From January to June 2020, we conducted a prospective cohort study via EMA among university students in Hong Kong who reported current drinking. Over 14 consecutive days, each participant completed 5 fixed-interval, signal-contingent EMAs daily via a smartphone app. Each EMA asked about the number and types of alcohol marketing exposures, the amount and types of alcohol used, and whether any alcohol was purchased, all within the past 3 hours. We used 2-part models, including multilevel logistic regressions and multilevel gamma regressions, to examine if the number of alcohol marketing exposure was associated with subsequent alcohol use and alcohol purchase.

Results: A total of 49 students participated, with 33% (16/49) being male. The mean age was 22.6 (SD 2.6) years. They completed 2360 EMAs (completion rate: 2360/3430, 68.8%). Participants reported exposure to alcohol marketing in 5.9% (140/2360), alcohol use in 6.1% (145/2360), and alcohol purchase in 2.4% (56/2360) of all the EMAs. At the between-person level, exposure to more alcohol marketing predicted a higher likelihood of alcohol use (adjusted odd ratio [AOR]=3.51, 95% CI 1.29-9.54) and a higher likelihood of alcohol purchase (AOR=4.59, 95% CI 1.46-14.49) the following day. Exposure to more alcohol marketing did not increase the amount of alcohol use or frequency of alcohol purchases the following day in participants who used or purchased alcohol. At the within-person level, exposure to more alcohol marketing was not associated with a higher likelihood of alcohol use, amount of alcohol use, higher likelihood of alcohol purchase, or frequency of alcohol purchases the following day (all P_s>.05). Each additional exposure to alcohol marketing within 1 week predicted an increase of 0.85 alcoholic drinks consumed in the following week (adjusted B=0.85, 95% CI 0.09-1.61). On days of reporting alcohol use, the 3 measures for alcohol marketing receptivity were not associated with more alcohol use or purchase (all P_s>.05).

Conclusions: By using EMA, we provided the first evidence for the effect of alcohol marketing exposure on initiating alcohol use and purchase in current-drinking university students. Our findings provide evidence of the regulation of alcohol marketing for the reduction of alcohol use and purchase among young adults.

(JMIR Mhealth Uhealth 2024;12:e60052) doi:10.2196/60052

KEYWORDS

alcohol marketing; drinking; ecological momentary assessment; health behaviors; young adults; mobile phone

Introduction

Alcohol use is a leading risk factor for premature death and disability worldwide and has no safe level of consumption [1]. Among individuals aged 15-29 years, alcohol use is the single biggest risk factor for violence and death due to external causes and plays a major role in mental health problems such as depression and anxiety [2]. According to the social-ecological model, the associated factors of alcohol use are nested in 5 levels—individual, interpersonal, organizational, community, and policy [3]. Alcohol marketing is a major community- and policy-level factor through advertising, promotion, sponsorships, and point-of-sale display in retail [4].

The relationship between alcohol marketing and drinking behaviors has been extensively studied, particularly in Western countries. Alcohol marketing influences drinking behaviors through short-term effects [5], where repeated exposure increases familiarity and leads to more consumption and purchases, and long-term effects, where marketing normalizes alcohol use by increasing perceived social approval and emphasizing positive consequences [6]. Two systematic reviews, including 25 cross-sectional and longitudinal studies, consistently indicated a positive association between alcohol marketing exposure and intentions to drink, as well as binge or hazardous drinking [7,8]. These studies also found dose-response relationships between marketing exposure and the initiation and frequency of drinking [7,8]. However, these studies may underestimate the true impact of alcohol marketing as they often focus on advertising, neglecting other forms of promotion. Another systematic review extended the analysis to various forms of alcohol marketing, including advertising on mixed media, points of sales, and so forth, and these were positively associated with frequency and quantity of alcohol use [9]. However, using self-reported marketing exposure in the previous week or month had a large recall bias. Also, cross-sectional studies failed to establish causality. The alcohol industry argues that marketing regulations are ineffective, claiming insufficient evidence that marketing influences behavior and asserting that it only affects brand selection and market share [10].

To better understand the causal relationship between alcohol marketing and drinking behavior, several randomized controlled trials (RCTs) have been conducted [11-17]. A meta-analysis of 7 RCTs in young adults found that exposure to alcohol marketing within a 1- to 90-minute timeframe led to a small but significant increase in alcohol use (effect size: 0.20, 95% CI 0.05-0.34) [18]. However, these RCTs, which were all laboratory based with low ecological validity, often failed to account for prior exposure to alcohol marketing and involved only television marketing. This highlights the need for studies that accurately measure alcohol marketing exposure and drinking behavior in real time and explore the prospective associations between marketing exposure and drinking behavior in real-world settings.

Ecological momentary assessment (EMA) is a repeated real-time data collection method in the natural environment and allows the modeling of temporal dynamics [19]. EMA facilitates monitoring of behavioral processes in their natural context, which can minimize recall bias [20] and avoid reverse causality [21]. Our searches on PubMed and Web of Science up to July 2023, using relevant keywords of “ecological momentary assessment,” “alcohol marketing,” “alcohol advertisement,” “drinking,” “alcohol consumption,” and “alcohol use,” found only 2 studies that used EMA to assess the association between alcohol marketing exposure and normalization of alcohol use in adolescents [22,23]. One study found that more alcohol marketing exposure was associated with higher perceived social approval and popularity of alcohol use [22]. Another study found that higher adolescent perceptions of social approval and popularity of alcohol use were associated with more receptivity toward marketing exposures [23]. We found no other EMA-based studies examining the effect of alcohol marketing directly on alcohol use and purchase in adolescents and young adults.

Hong Kong has a low alcohol consumption compared to most Western countries and Asian countries such as Japan (8.0 L per capita), with a per capita consumption of 2.9 L [24,25]. A recent study found that about 30% of Hong Kong adults drank alcohol in the past 12 months [26], and Hong Kong drinkers consume 9.7 L per capita, nearly two-thirds of the per capita consumption in Japanese drinkers (14.1 L) [24]. The Hong Kong government eliminated the tax on beer and wine in 2008, which coincided with an increase in both the frequency of alcohol consumption among ever-drinkers and the prevalence of new drinkers [27,28]. This policy, combined with the high accessibility of alcohol due to no restrictions on the time and place of alcohol sales, as well as the high density of alcohol outlets, has led to an increase in alcohol consumption in the recent decade. The prevalence of past 30-day alcohol use in Hong Kong students aged 19 years and older increased from 24.2% in 2008 [29] to 33.8% in 2020 [30]. Young adults aged 18-35 years have the highest proportion of binge drinking and alcohol abuse compared to older age groups and are at high risk of various alcohol-related harms [26-28,31]. Alcohol use is also significantly associated with substance use among young adults [32,33]. A recent cross-sectional study among young adults in Hong Kong found that 71.6% were exposed to traditional marketing and 53% to social media marketing, with exposure positively associated with past-month alcohol consumption [34]. Despite this, Hong Kong imposes minimal restrictions on alcohol marketing. The only policy is a ban on alcohol advertisement on domestic free television from 4 PM to 8:30 PM [35]. Thus, Hong Kong young adults are frequently exposed to alcohol marketing in both offline and internet-based channels [36,37].

Given the extensive alcohol marketing and the surge in drinking among young adults, we aimed to examine the association between alcohol marketing and drinking behaviors to inform

effective public health interventions. We used EMA to examine the effects of alcohol marketing exposure on drinking behaviors among university students. We hypothesized that first, more exposure to alcohol marketing in a day would predict any alcohol use, any alcohol purchase, increased alcohol use, and more frequent alcohol purchases the following day. This association was examined at both the within-person and between-person levels to account for intra- and interindividual variability in responses to marketing exposure, respectively. Second, more exposure to alcohol marketing in the first week would predict any alcohol use, any alcohol purchase, increased alcohol use, and more frequent alcohol purchases the following week. Finally, positive alcohol marketing receptivity would predict more alcohol use and a higher frequency of alcohol purchases the following day.

Methods

Study Design

This prospective study included 49 university students who participated in a 2-arm RCT (allocation ratio 1:1) to examine the discrepancy in reporting alcohol marketing exposure between EMA and conventional retrospective survey. The participants in the intervention group who completed 2 weeks of EMAs were included in this analysis.

Procedures

We recruited participants from January to June 2020 via mass emails with a link to an internet-based anonymous survey of all undergraduate and postgraduate students in the University of Hong Kong (HKU). Potential participants interested in this project can complete the internet-based enrollment form by clicking the provided link in the mass emails. The enrollment form included questions to screen the eligibility. Students who were (1) Hong Kong residents, (2) aged 18 to 35 years, (3) enrolled in an undergraduate or postgraduate program, (4) had a smartphone with internet access, (5) drank any alcohol in the past 30 days, (6) able to read and write Chinese, and (7) stayed in Hong Kong throughout the study were eligible. A research assistant contacted the eligible participants; provided a brief overview of the study via telephone; and scheduled a face-to-face enrollment session with them in the HKU campus to obtain written consent, conduct the baseline survey, and install the EMA app on their smartphones.

EMA Operation

Participants allocated to the intervention group were instructed to install and set up a free EMA app on their smartphones during the enrollment session. This setup included imputing a personal identification number (the last 5 digits of their phone number) and specifying the date and time they wanted to receive the first EMA. The app "HKU alcohol study" (HKU, Hong Kong Special Administrative Region of the People's Republic of China) was developed specifically for this study. The app featured automated notifications, a customizable scheduling system, and intuitive data entry interfaces allowing participants to select options without manual text input. Alcohol marketing exposure is typically discrete. Hence, fixed-interval and signal-contingent prompts can reliably capture all exposures and behaviors, enabling comprehensive data collection on the cumulative effects of these exposures on drinking behaviors [38]. Additionally, signal-contingent EMAs are the predominant method for prompting in EMA studies, with over 77% (81/105) of EMA studies using this approach [39], as it mitigates the underreporting of alcohol events often observed in event-contingent EMAs [40]. Participants were required to complete 5 fixed-interval, signal-contingent [41] EMAs each day for 14 consecutive days, with a fixed time interval of 3 hours between each EMA. The app notified participants with a pop-up message, reminding them to use the app and complete the EMA within 5 minutes of receiving the prompt. If they did not respond, 2 additional prompts would be sent within the next 10 minutes. If they still did not respond, the corresponding EMA episode would be treated as nonresponse. The completed EMA data were immediately uploaded to the HKU server.

All participants were awarded HK $100 (US $1=HK $7.8) shopping vouchers after completing the baseline and follow-up questionnaires. Additionally, participants who completed 3 EMAs within each day would be rewarded with a HK $15 shopping voucher. An additional HK $10 shopping voucher was further provided to participants who completed all 5 EMAs within each day.

Ethical Considerations

Ethical approval was obtained from the institutional review board of the University of Hong Kong/Hong Kong Authority Hong Kong West Cluster (UW-19-155). All participants were informed that the collected data will be kept strictly confidential and used solely for research purposes. All data were deidentified and no personal information was disclosed in this paper. Participants' contact information was securely stored on the server located in HKU, encrypted with a password accessible only to the principal investigator and authorized research staff. Participants could be rewarded up to HK $450 for shopping vouchers in total as compensation.

Measures

Alcohol Marketing Exposure

In each EMA, participants reported any exposure to alcohol marketing in the past 3 hours, except for the first episode, which enquired the exposure "since the last survey" (yes or no). Alcohol marketing includes all forms of promotional material or activities for alcoholic products; alcohol brands; or the culture of alcohol consumption such as beer festivals, wine fairs, and media of new alcoholic beverages or bars. If they reported exposure, they were asked about the number of exposures (an exposure could be counted multiple times if the same marketing was seen in various places) and the sources of exposure (including restaurants, YouTube, apps, social media, web banners, bus stations, metro stations, ads on public transport, point-of-sale at retail stores, television, newspapers, magazines, radios, and others). The number of exposures to alcohol marketing was aggregated daily, bi-daily, weekly, and 2 weeks for each participant.

Marketing Receptivity Toward Alcohol Marketing

For each reported exposure to alcohol marketing, participants rated their receptivity toward the marketing content using 3 items: "I like the alcohol marketing," "I think the alcohol marketing was innovative," and "I think the alcohol marketing was attractive" (1=strongly disagree and 5=strongly agree). For participants who reported exposure to alcohol marketing once a day, the daily marketing receptivity score for each item was the score rated by the participant for that single exposure. For participants with multiple exposures in a day, the daily score for each receptivity item was calculated by averaging the scores for all exposures on that day. These average scores were recorded as a binary variable (1 to 3: disagree and >3: agree). The "number of positive receptivity" was calculated by summing the binary variables for attractiveness, innovation, and likeability, resulting in a score from 0 to 3, indicating the number of these items the participant agreed with. Indicating "agree" to any item of marketing receptivity was treated as "any positive alcohol marketing receptivity" (yes or no) for that day.

Alcohol Use

In each EMA, participants reported any alcohol use in the past 3 hours, except for the first episode, which enquired about alcohol use "since the last survey" (yes or no). If they responded to alcohol use, they were asked the type of alcohol (including beer, alcopops, wine, spirits, cocktails, rice wine, Chinese spirits, Japanese sake, and others) and the number of drinks consumed. One drink is equivalent to one 330-mL bottle of beer or alcopops, one 125-mL glass of wine, one 22-mL shot of spirits or cocktails, one 180-mL glass of rice wine, or one 20-mL glass of Chinese spirits or Japanese sake. The number of drinks was aggregated daily, bi-daily, weekly, and 2 weeks for each participant. Participants who consumed at least 1 drink for the time periods (1) within the past day, (2) within the past 2 days, and (3) within the past week were classified as having had alcohol (yes or no) use for those time periods.

Alcohol Purchase

In each EMA episode, participants reported any purchase of alcohol in the past 3 hours, except for the first episode, which enquired about alcohol purchase "since the last survey" (yes or no). We did not ask about the quantity or frequency of alcohol purchased within each EMA episode. If they responded "yes," it was treated as 1 instance of an alcohol purchase. The frequency of alcohol purchase was then aggregated daily, bi-daily, weekly, and 2 weeks for each participant. Participants who purchased alcohol at least once for the time periods (1) within the past day, (2) within the past 2 days, and (3) within the past week were classified as having purchased alcohol (yes or no) for those time periods.

Baseline Characteristics

Sex, age, study program, smoking status, age of initiating alcohol use, age of purchasing alcohol for the first time, and Alcohol Use Disorder Identification Test (AUDIT) were assessed at baseline. AUDIT is a 10-item scale (0 to 7=low-risk, 8 to 15=increasing risk, 16 to 19=harmful risk, and ≥20=probable dependence) [42,43]. AUDIT has been validated in Chinese [44]; the Cronbach α was 0.79 in this study.

Statistical Analyses

Baseline demographic variables, smoking status, age of initiating alcohol use, age of purchasing alcohol for the first time, and AUDIT were described using percentage or mean as appropriate.

To test the first hypothesis, we used a 2-part modeling approach to address the zero-inflated and right skewed alcohol use and purchase data. First, we used multilevel logistic regressions with a random intercept to assess the prospective association between daily alcohol marketing exposure and the following day's outcomes of (1) any alcohol use and (2) any alcohol purchase. Second, for days with reported alcohol use or purchase, multilevel gamma regression models assessed the association between daily alcohol marketing exposure and the following day's outcomes of (1) amount of alcohol used and (2) frequency of alcohol purchase. Analyses included within-person and between-person effects, with predictor variables centered using the *xtcenter* command in Stata (StataCorp).

To test the second hypothesis, we used multivariable logistic and linear regressions to assess the effect of alcohol marketing exposure in the first week on the following week's outcomes. Logistic regression assessed the association between alcohol marketing exposure with any alcohol use and any alcohol purchase, while linear regression assessed the association between alcohol marketing exposure with amount of alcohol use and frequency of purchases. Since participants were assessed for only 2 weeks, multilevel regression was not applied.

To test the third hypothesis, we used 2-part modeling approaches. First, multilevel logistic regressions were used to assess the association between alcohol marketing receptivity in a day (binary responses in the 3 separate items of receptivity and a total number of positive receptivity) and the following day's outcomes. Second, for days with reported alcohol use or purchase, multilevel gamma regression models were used to assess the association between alcohol marketing receptivity in a day (binary responses in the 3 separate items of receptivity and a total number of positive receptivity) with the following day's amount of alcohol used and frequency of purchases. Sex, age, and AUDIT were adjusted for in all regressions as younger male drinkers were more likely to be exposed to alcohol marketing and drink more alcohol [45,46].

Three sensitivity analyses with the same 4 outcomes were used to supplement our result interpretation. First, in line with prior findings suggesting the effect of marketing exposure might fade out within 1.5 days [47], we performed a sensitivity analysis to evaluate the impact of a 2-day cumulative exposure on the outcomes on the subsequent day. Second, we used multivariable logistic and linear regressions to assess the cross-sectional association between alcohol marketing exposure and the 4 outcomes over 2 weeks. Third, regarding the prospective associations between alcohol marketing receptivity with alcohol use and purchase, we conducted 2 additional analyses treating the 3 separate items of alcohol marketing receptivity as continuous variables and EMAs with no marketing exposure as 0 or 3. All analyses were done using Stata (version 16.0).

Results

Sample Description

This study enrolled 51 participants in the EMA group. A total of 2 participants were unable to complete the 2-week EMAs due to failure to install the EMA app and receive prompts; hence, EMA data from 49 participants were analyzed. Table 1 shows that 33% (16/49) of participants were male. The mean age was 22.6 (SD 2.6) years. About 74% (36/49) had a bachelor's degree and 18% (9/49) were current smokers. The mean age of initiating alcohol use and mean age of purchasing alcohol for the first time were 16 (SD 3.3) and 18.2 (SD 1.1) years, respectively. The mean AUDIT score was 5.8 (SD 4.5).

Table 1. Demographic characteristics, drinking behavior and EMA[a] completion rate of participants (N=49).

Characteristics	Values
Age (years), mean (SD)	22.6 (2.6)
Male, n (%)	16 (33)
Study program, n (%)	
Bachelor's	36 (74)
Master's or above	13 (27)
Age of initiating alcohol use (years), mean (SD)	16.0 (3.3)
Age of purchasing alcohol for the first time (years), mean (SD)	18.2 (1.1)
Smoking in the past 30 days	9 (18)
AUDIT[b], mean (SD)	5.8 (4.5)
Low-risk drinking (1-7), n (%)	36 (74)
Increasing risk (8-15), n (%)	10 (20)
Harmful risk (16-19), n (%)	2 (4)
Probable dependence (≥20), n (%)	1 (2)
EMA completion rate, n (%)	
Below 50%	7 (14)
50%-74.9%	19 (39)
75% or above	23 (47)

[a]EMA: ecological momentary assessment.

[b]AUDIT: Alcohol Use Disorders Identification Test. Total score ranged from 0 to 40, higher scores indicating higher level of alcohol dependence.

EMA Completion Rate and Description

A total of 49 participants completed 2360 EMAs upon the 3430 prompts from the app (completion rate: 2360/3430, 68.8%). Table 2 shows that 37 (76%) of the 49 participants reported 173 exposures to alcohol marketing from various sources—restaurants (52/173, 30.1%), internet (49/173, 28.3%), public transportation (25/173, 14.5%), point-of-sale retail locations (19/173, 11%), television (14/173, 8.1%), and other places (14/173, 8.1%). On average, each participant reported 4.7 (SD 3.7) times of exposure to alcohol marketing within the 2-week EMA period. Among all alcohol marketing exposures, 38.2% (66/173) were rated as being liked, 32.9% (57/173) as innovative, and 35.2% (61/173) as attractive.

Table 2. Description of EMA[a] data for alcohol marketing exposure, alcohol use, and alcohol purchase.

Variables	Values
Total episodes of exposure to alcohol marketing, n	140
Sources of alcohol marketing exposure (n= 173), n (%)	
Restaurant	52 (30.1)
Internet (YouTube, app, social media, web banners, etc)	49 (28.3)
Transportation (bus stations, metro stations, ads on public transport, etc)	25 (14.5)
Point-of-sale at retail stores	19 (11.0)
Television	14 (8.1)
Others (newspapers, magazines, radios, etc)	14 (8.1)
Marketing receptivity toward alcohol marketing in those exposure (n=173), n (%)	
I agree that I liked the alcohol marketing	66 (38.2)
I agree that the alcohol marketing is innovative	57 (32.9)
I agree that the alcohol marketing is attractive	61 (35.2)
Total episodes of alcohol use, n	145
Type of alcohol use (n=145), n (%)	
Beer	51 (35.2)
Wine	37 (25.5)
Alcopops	22 (15.2)
Cocktails	14 (9.7)
Spirits	14 (9.7)
Japanese sake	12 (8.3)
Total episodes of alcohol purchase, n	56

[a]EMA: ecological momentary assessment.

Table 2 also shows that 39 (80%) participants reported 145 episodes of alcohol use including beer (51/145, 35.2%), wine (37/145, 25.5%), alcopops (22/145, 15.2%), cocktails (14/145, 9.7%), spirits (14/145, 9.7%), and Japanese sake (12/145, 8.3%). On average, each participant reported 3.7 (SD 3.0) episodes of alcohol use and consumed 7.5 (SD 12.1) drinks within the 2-week EMA period. A total of 26 (53%) participants reported 56 episodes of purchasing alcohol within the 2-week EMA period.

Association of Alcohol Marketing Exposure With Alcohol Use and Purchase

Table 3 shows that at the between-person level, exposure to more alcohol marketing predicted a higher likelihood of alcohol use (adjusted odd ratio [AOR]=3.51, 95% CI 1.29-9.54; *P*=.01) and a higher likelihood of alcohol purchase the following day (AOR=4.59, 95% CI 1.46-14.49; *P*=.01). At the within-person level, exposure to more alcohol marketing was not associated with the likelihood of alcohol use, amount of alcohol use, the likelihood of alcohol purchase, or frequency of alcohol purchase the following day (all P_s>.05). The results at both the between-person and within-person levels remained robust in the 2-day cumulative exposure models.

Table 3. Prospective associations of alcohol marketing exposure with alcohol use and alcohol purchase on the following day (N=49).

Predictors	Outcomes			
	Any alcohol use or purchase the following day		Amount of alcohol use or purchase the following day	
	Adjusted OR[a] per exposure (95% CI)	*P* value	Adjusted exp (B) (95% CI)	*P* value
Alcohol use				
Number of exposures to alcohol marketing within a day[b]				
Within-person	0.92 (0.64-1.31)[c]	.63	1.16 (0.98-1.38)[d]	.09
Between-person	3.51 (1.29-9.54)[c]	.01	1.89 (0.92-3.88)[d]	.08
Number of exposures to alcohol marketing over the past 2 days[e]				
Within-person	0.86 (0.65-1.13)[f]	.27	1.13 (0.98-1.30)[g]	.07
Between-person	2.42 (1.45-4.03)[f]	.001	1.44 (0.97-2.13)[g]	.07
Alcohol purchase				
Number of exposures to alcohol marketing within a day[h]				
Within-person	1.15 (0.75-1.76)[i]	.53	0.93 (0.85-1.02)[j]	.10
Between-person	4.59 (1.46-14.49)[i]	.01	1.15 (0.96-1.39)[j]	.14
Number of exposures to alcohol marketing over the past 2 days[k]				
Within-person	1.11 (0.79-1.54)[l]	.56	1.01 (0.95-1.08)[m]	.77
Between-person	1.93 (0.99-3.76)[l]	.05	1.02 (0.90-1.16)[m]	.75

[a]OR: odds ratio.

[b]The exposure is the number of alcohol marketing exposures within a day. The outcomes are any alcohol use the following day and the amount of alcohol use the following day.

[c]Multilevel logistic regression adjusted for sex, age, and baseline AUDIT (Alcohol Use Disorder Identification Test; number of observations=562).

[d]Multilevel gamma regression adjusted for sex, age, and baseline AUDIT, exclude daily alcohol use=0 (number of observations=97).

[e]The exposure is the number of alcohol marketing exposures over the past 2 days. The outcomes are any alcohol use on the subsequent day after the 2-day cumulative exposure and the amount of alcohol use on the subsequent day after the 2-day cumulative exposure.

[f]Multilevel logistic regression adjusted for sex, age, and baseline AUDIT (number of observations=532).

[g]Multilevel gamma regression adjusted for sex, age, and baseline AUDIT, exclude daily alcohol use=0 (number of observations=90).

[h]The exposure is the number of alcohol marketing exposures within a day. The outcomes are any alcohol purchase the following day and the frequency of alcohol purchases the following day.

[i]Multilevel logistic regression adjusted for sex, age, and baseline AUDIT (number of observations=562).

[j]Multilevel gamma regression adjusted for sex, age, and baseline AUDIT, exclude daily alcohol purchase=0 (number of observations=45).

[k]The exposure is the number of alcohol marketing exposures over the past 2 days. The outcomes are any alcohol purchase on the subsequent day after the 2-day cumulative exposure and the frequency of alcohol purchases on the subsequent day after the 2-day cumulative exposure.

[l]Multilevel logistic regression adjusted for sex, age, and baseline AUDIT (number of observations=532).

[m]Multilevel gamma regression adjusted for sex, age, and baseline AUDIT, exclude daily alcohol purchase=0 (number of observations=42).

Within the 2-week EMA period, 12 participants reported 0 exposure to alcohol marketing. A total of 10 participants reported no alcohol use. A total of 23 participants reported no alcohol purchase. Summing up the 3 key indicators, we identified 4 participants who did not report any exposure to alcohol marketing, alcohol use, and alcohol purchase. Therefore, we excluded these participants and ran the 2-part models of examining the prospective associations between alcohol marketing exposure with alcohol use and alcohol purchase. The results (Multimedia Appendix 1) were still consistent with Table 3.

Table 4 shows each additional exposure to alcohol marketing within 1 week predicted an increase of 0.85 alcoholic drinks consumed in the following week (adjusted B=0.85; *P*=.03). Although each additional exposure to alcohol marketing over a week was marginally associated with increased likelihood of alcohol use (AOR=1.62; *P*=.054), it was not significantly associated with likelihood of purchasing alcohol or frequency of alcohol purchases (both P_s>.05, see Table 4) in the following week. Multimedia Appendix 2 shows that exposure to more marketing exposure over 2 weeks was associated with a larger amount of alcohol use (adjusted B=0.90; *P*=.02) and more frequent alcohol purchases (adjusted B=0.14; *P*=.01).

Table 4. Prospective associations of alcohol marketing exposure with alcohol use and alcohol purchase in the following week (N=49).

Predictors	Outcomes							
	Any alcohol use the following week[a]		Amount of alcohol use the following week[b]		Any alcohol purchase the following week[a]		Frequency of alcohol purchases the following week[b]	
	Adjusted OR[c] per exposure (95% CI)	*P* value	Adjusted B per exposure (95% CI)	*P* value	Adjusted OR per exposure (95% CI)	*P* value	Adjusted B per exposure (95% CI)	*P* value
Number of exposures to alcohol marketing within a week	1.62 (0.99-2.65)	.054	0.85 (0.09-1.61)	.03	0.98 (0.71-1.36)	.92	0.00 (–0.18 to 0.18)	.98

[a]Multivariable logistic regression adjusted for sex, age, and baseline Alcohol Use Disorders Identification Test (number of observations=46).

[b]Multivariable linear regression adjusted for sex, age, and baseline Alcohol Use Disorders Identification Test (number of observations=46).

[c]OR: odds ratio.

Association of Alcohol Marketing Receptivity With Alcohol Use and Purchase

Table 5 shows that on days of liking alcohol marketing, when perceiving the marketing as innovative or attractive, they were more likely to use and purchase alcohol the following day compared to days where they reported no marketing exposure, but the results were not significant (all P_s>.05). On days of reporting alcohol use, the 3 measures for alcohol marketing receptivity were not associated with more alcohol use or purchase (all P_s>.05). Our sensitivity analysis by treating alcohol marketing receptivity as a continuous variable showed similar results as the main analysis (Multimedia Appendix 3).

Table 5. Prospective association of alcohol marketing receptivity with alcohol use and alcohol purchase on the following day (N=49).

Predictors	Any alcohol use the following day[a]			Amount of alcohol use the following day[b]		Any alcohol purchase the following day[c]			Frequency of alcohol purchases the following day[d]	
	Yes, n/N (%)	Adjusted OR[e] (95% CI)	*P* value	Adjusted exp (B) (95% CI)	*P* value	Yes, n/N (%)	Adjusted OR (95% CI)	*P* value	Adjusted exp (B) (95% CI)	*P* value
I like the alcohol marketing										
No exposure	78/481 (16.2)	Ref[f]	N/A[g]	N/A	N/A	33/481 (6.9)	Ref	N/A	N/A	N/A
No	8/59 (13.6)	0.65 (0.27-1.56)	.34	1.04 (0.58-1.88)	.89	5/59 (8.5)	1.13 (0.40-3.19)	.82	0.94 (0.77-1.14)	.52
Yes	13/44 (29.6)	1.41 (0.63-3.15)	.40	1.03 (0.67-1.59)	.90	7/44 (15.9)	1.94 (0.73-5.17)	.19	0.89 (0.75-1.06)	.20
The alcohol marketing was innovative										
No exposure	78/482 (16.2)	Ref	N/A	N/A	N/A	33/482 (6.9)	Ref	N/A	N/A	N/A
No	9/61 (14.8)	0.67 (0.30-1.54)	.34	0.91 (0.52-1.57)	.73	6/61 (9.9)	1.29 (0.49-3.41)	.61	0.92 (0.77-1.11)	.39
Yes	12/40 (30.0)	1.58 (0.70-3.59)	.23	1.14 (0.71-1.75)	.64	6/40 (15.0)	1.85 (0.66-5.16)	.24	0.90 (0.76-1.08)	.26
The alcohol marketing was attractive										
No exposure	81/494 (16.4)	Ref	N/A	N/A	N/A	34/494 (6.9)	Ref	N/A	N/A	N/A
No	8/47 (17.0)	0.75 (0.31-1.86)	.54	1.08 (0.60-1.92)	.81	6/47 (12.8)	1.58 (0.57-4.35)	.38	1.08 (0.60-1.92)	.81
Yes	10/43 (23.3)	1.00 (0.42-2.28)	.99	1.12 (0.69-1.83)	.64	5/43 (11.6)	1.33 (0.45-3.90)	.61	1.12 (0.69-1.83)	.64
Positive alcohol marketing receptivity (score range 0-3)										
No exposure	81/494 (16.4)	Ref	N/A	N/A	N/A	34/494 (6.9)	Ref	N/A	N/A	N/A
0	5/35 (14.3)	0.57 (0.19-1.74)	.33	1.28 (0.61-2.67)	.52	3/35 (8.6)	1.08 (0.29-4.00)	.91	0.93 (0.72-1.20)	.60
1	3/18 (16.7)	1.05 (0.26-4.19)	.88	0.77 (0.32-1.80)	.55	2/18 (11.1)	1.36 (0.27-6.78)	.71	0.97 (0.73-1.29)	.84
2	2/12 (16.7)	0.64 (0.12-3.34)	.60	0.63 (0.24-1.68)	.36	2/12 (16.7)	1.80 (0.33-9.90)	.50	0.93 (0.69-1.24)	.61
3	8/25 (32.0)	1.40 (0.50-3.87)	.59	1.31 (0.77-2.34)	.32	4/25 (16.0)	1.92 (0.56-6.56)	.30	0.87 (0.70-1.08)	.21
Any positive alcohol marketing receptivity										
No exposure	81/494 (16.4)	Ref	N/A	N/A	N/A	34/494 (6.9)	Ref	N/A	N/A	N/A
No	5/35 (14.3)	0.56 (0.18-1.70)	.31	1.23 (0.59-2.61)	.58	3/35 (8.6)	1.07 (0.29-3.99)	.92	0.94 (0.73-1.21)	.63
Yes	13/55 (23.6)	1.10 (0.51-2.37)	.81	1.07 (0.69-1.64)	.77	8/55 (14.6)	1.7 (0.69-4.22)	.25	0.91 (0.78-1.07)	.26

[a]Multilevel logistic regression adjusted for sex, age, and baseline Alcohol Use Disorder Identification Test (AUDIT), daily alcohol use=0 versus daily alcohol use>0.

[b]Multilevel gamma regression adjusted for sex, age, and baseline AUDIT, exclude daily alcohol use=0.

[c]Multilevel logistic regression adjusted for sex, age, and baseline AUDIT, daily alcohol purchase=0 versus daily alcohol purchase>0.

[d]Multilevel gamma regression adjusted for sex, age, and baseline AUDIT, exclude daily alcohol purchase=0.

[e]OR: odds ratio.

[f]Ref: reference group.

[g]N/A: not applicable.

Discussion

Principal Findings

Our findings provide the first evidence that increased exposure to alcohol marketing within a single day or 2 days predicted higher likelihoods of alcohol use and purchase the following days, implying the direct effect of marketing on alcohol use and purchases. Alcohol marketing exposure in a week predicted a larger amount of alcohol use the following week, implying the accumulative effect of marketing on consumption level. Sensitivity analyses by excluding 4 participants with 0 reports of the key indicators supported the robustness of the results.

Our finding is the first to support that increased exposure to alcohol marketing was associated with a higher likelihood of alcohol use within the following day and subsequent day after 2-day cumulative exposure. This extends the evidence from previous experimental RCTs [11-17], which lacked ecological validity and only examined the immediate effects 30 minutes after the exposure. Such a direct effect is consistent with the "mere exposure effect" [5], whereby drinkers might be stimulated to initiate drinking due to greater familiarity with alcohol products due to more recent exposure to marketing. Therefore, to reduce the likelihood of alcohol use in young adults, alcohol control policies in reducing the exposure to alcohol marketing and avoiding the glamorization of drinking behavior should be implemented.

Exposure to alcohol marketing might stimulate alcohol use, but our findings showed that the exposure did not significantly increase the amount of alcohol use on the following day. We showed that 74% (36/49) of our participants have low AUDIT scores; hence, most were light drinkers and nondaily drinkers, and the marketing effect on alcohol consumption in 1 day may be small. However, we found a positive effect of weekly alcohol marketing exposure and amount of alcohol use the following week (adjusted B=0.85). Thus, the effect of accumulative alcohol marketing exposures on the amount of alcohol use assessed in the week was more detectable in this group with low alcohol consumption.

The lack of significant association between daily variations in alcohol marketing exposure and the amount of alcohol use can be attributed to the low variability in individual exposure to alcohol marketing, with only 5.9% (140/2360) of EMA episodes reporting such exposure. This limited variability in individual exposure levels results in insufficient statistical power to detect significant within-person effects. Besides, this study was conducted with the onset of the COVID-19 pandemic in Hong Kong. During this period, the government implemented numerous restrictive policies such as bar closures and dining restrictions. These restrictions likely reduced participants' opportunities to encounter varying levels of alcohol marketing exposure, thus reducing the detectable within-person effects.

We found a positive effect of alcohol marketing exposure within a day on alcohol purchase the following day, but not on the frequency of alcohol purchase. The null association between alcohol marketing exposure in a day and the frequency of alcohol purchase the following day can be attributed to several factors. First, our EMAs did not ask for and analyze the quantity of alcohol purchased. Karaoke bars and pubs in Hong Kong often target students with discounted party packages and fixed-price "all-you-can-drink" nights, promoting larger 1-time purchases. Second, university students often consume alcohol obtained from others in social settings. A recent study has shown that students living in residence halls are twice as likely to have binge drinking compared to those living with family [48]. Third, financial constraints may limit university students' ability to purchase alcohol frequently within a short period. Besides, our cross-sectional analysis (Multimedia Appendix 2) found that exposure to more alcohol marketing over a 2-week EMA period was associated with a higher frequency of alcohol purchases during the same period, which supported the long-term effect of alcohol marketing on purchases. Finally, reverse causation of the alcohol purchase and exposure to alcohol marketing was likely to happen.

Contrasting with previous studies that alcohol marketing receptivity might increase the frequency and amount of alcohol use [49,50], our analysis did not show sufficient evidence to support similar results at day-level. The discrepancy may be attributed to only a small number of participants who were exposed to alcohol marketing being eligible to respond to receptivity questions. In the 2360 completed EMA episodes, only 173 episodes of exposure to alcohol marketing and the corresponding marketing receptivity were assessed. It might limit the statistical power to confirm the association between marketing receptivity and alcohol use from both the main and sensitivity analyses. Further studies with a larger sample size to examine the effect of alcohol marketing receptivity and alcohol use and purchase are warranted.

Our study had a few limitations. First, the overall completion rate of EMA was about 68.8% (2360/3430), which was generally lower than previous EMA studies, which was about 76.4% on average (4 to 5 prompts per day) [51]. Future EMA studies may use a combination of event-contingent and signal-contingent prompts to capture more alcohol marketing exposure, thereby increasing the completion rate [52]. Second, the study sample was not a representative sample of all university students or young adults. Third, due to the time constraint in each EMA, we did not assess the type, quantity, and venue of alcohol purchases. Furthermore, the small sample size is unable to assess the association between different sources of alcohol marketing exposure and drinking behaviors. Our power analysis estimated the power of the between-person effect of daily marketing exposure on the likelihood of alcohol use in our study was about 22.5% (95% CI 19.9%-25.1%). Future studies with larger, more representative samples are warranted to evaluate the effects of various types of marketing exposure on drinking behavior.

Conclusions

In conclusion, by using EMA, our study showed the direct effect of alcohol marketing exposure on initiating alcohol use and alcohol purchase in current-drinking university students in a real-world environment, refuting the claim by the alcohol industry that the marketing is only for brand promotion. Our findings provide evidence of regulating alcohol marketing for the reduction of alcohol use and purchase in young adults.

Acknowledgments

The authors are grateful to all the project coordinators, including Titan Mak and Bonnie Wu, for their effort to recruit participants and administrative duties. This study was supported by the Seed Fund for Basic Research of the University of Hong Kong (201711159257). The funder has no roles in study design, collection, management, analysis, and interpretation of data.

Data Availability

The data sets analyzed during this study are not publicly available due to privacy and ethical restrictions but are available from the corresponding author on reasonable request

Authors' Contributions

YTDC obtained funding for, conceptualized, and administered this study. MJZ and YTDC contributed to the methodology. MJZ conducted the formal analysis and investigation and wrote the original draft. All authors contributed to reviewing and editing the paper. All authors have approved the final paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Sensitivity analysis for prospective associations of alcohol marketing exposure with alcohol use and alcohol purchase on the following day in participants excluding those who reported 0 incidences of alcohol marketing exposure, alcohol drinking, and alcohol purchase (n=45).
[DOCX File , 22 KB - mhealth_v12i1e60052_app1.docx]

Multimedia Appendix 2
Association of alcohol marketing exposure with alcohol use and alcohol purchase over 2-week ecological momentary assessment period (N=49).
[DOCX File , 18 KB - mhealth_v12i1e60052_app2.docx]

Multimedia Appendix 3
Sensitivity analysis for prospective association of alcohol marketing receptivity with alcohol use and alcohol purchase on the following day.
[DOCX File , 20 KB - mhealth_v12i1e60052_app3.docx]

References

1. No level of alcohol consumption is safe for our health. World Health Organization. 2023. URL: https://www.who.int/europe/news/item/04-01-2023-no-level-of-alcohol-consumption-is-safe-for-our-health [accessed 2024-01-31]
2. How alcohol harms young people and what can you do about it. World Health Organization. 2015. URL: https://apps.who.int/iris/bitstream/handle/10665/208205/WPR_2015_DNH_002_eng.pdf?sequence=1&isAllowed=y [accessed 2023-01-31]
3. Bronfenbrenner U. Ecological models of human development. In: International Encyclopedia of Education. Oxford: Elsevier; 1994:37-43.
4. Global strategy to reduce the harmful use of alcohol. World Health Organization. 2010. URL: https://iris.who.int/bitstream/handle/10665/44395/9789241599931_eng.pdf?sequence=1 [accessed 2024-01-31]
5. Zajonc RB. Attitudinal effects of mere exposure. J Pers Soc Psychol 1968;9(2, Pt.2):1-27. [doi: 10.1037/h0025848]
6. Borsari B, Carey KB. Descriptive and injunctive norms in college drinking: a meta-analytic integration. J Stud Alcohol 2003 May;64(3):331-341 [FREE Full text] [doi: 10.15288/jsa.2003.64.331] [Medline: 12817821]
7. Anderson P, de Bruijn A, Angus K, Gordon R, Hastings G. Impact of alcohol advertising and media exposure on adolescent alcohol use: a systematic review of longitudinal studies. Alcohol Alcohol 2009;44(3):229-243. [doi: 10.1093/alcalc/agn115] [Medline: 19144976]
8. Jernigan D, Noel J, Landon J, Thornton N, Lobstein T. Alcohol marketing and youth alcohol consumption: a systematic review of longitudinal studies published since 2008. Addiction 2017 Jan;112 Suppl 1:7-20. [doi: 10.1111/add.13591] [Medline: 27565582]

9. Finan LJ, Lipperman-Kreda S, Grube JW, Balassone A, Kaner E. Alcohol marketing and adolescent and young adult alcohol use behaviors: a systematic review of cross-sectional studies. J Stud Alcohol Drugs Suppl 2020 Mar(19):42-56 [FREE Full text] [doi: 10.15288/jsads.2020.s19.42] [Medline: 32079561]
10. Savell E, Fooks G, Gilmore AB. How does the alcohol industry attempt to influence marketing regulations? A systematic review. Addiction 2016 Jan;111(1):18-32 [FREE Full text] [doi: 10.1111/add.13048] [Medline: 26173765]
11. Kohn PM, Smart RG. Wine, women, suspiciousness and advertising. J Stud Alcohol 1987 Mar;48(2):161-166. [doi: 10.15288/jsa.1987.48.161] [Medline: 3560952]
12. Engels RCME, Hermans R, van Baaren RB, Hollenstein T, Bot SM. Alcohol portrayal on television affects actual drinking behaviour. Alcohol Alcohol 2009;44(3):244-249. [doi: 10.1093/alcalc/agp003] [Medline: 19237443]
13. Kohn PM, Smart RG. The impact of television advertising on alcohol consumption: an experiment. J Stud Alcohol 1984 Jul;45(4):295-301. [doi: 10.15288/jsa.1984.45.295] [Medline: 6482432]
14. Koordeman R, Anschutz DJ, Engels RCME. Exposure to alcohol commercials in movie theaters affects actual alcohol consumption in young adult high weekly drinkers: an experimental study. Am J Addict 2011;20(3):285-291. [doi: 10.1111/j.1521-0391.2011.00134.x] [Medline: 21477057]
15. Koordeman R, Anschutz DJ, Engels RCME. The effect of alcohol advertising on immediate alcohol consumption in college students: an experimental study. Alcohol Clin Exp Res 2012 May;36(5):874-880. [doi: 10.1111/j.1530-0277.2011.01655.x] [Medline: 22017281]
16. Sobell LC, Sobell MB, Riley DM, Klajner F, Leo GI, Pavan D, et al. Effect of television programming and advertising on alcohol consumption in normal drinkers. J Stud Alcohol 1986 Jul;47(4):333-340 [FREE Full text] [doi: 10.15288/jsa.1986.47.333] [Medline: 3747533]
17. Wilks J, Vardanega AT, Callan VJ. Effect of television advertising of alcohol on alcohol consumption and intentions to drive. Drug Alcohol Rev 1992;11(1):15-21. [doi: 10.1080/09595239200185031] [Medline: 16840063]
18. Stautz K, Brown KG, King SE, Shemilt I, Marteau TM. Immediate effects of alcohol marketing communications and media portrayals on consumption and cognition: a systematic review and meta-analysis of experimental studies. BMC Public Health 2016 Jun 09;16:465 [FREE Full text] [doi: 10.1186/s12889-016-3116-8] [Medline: 27278656]
19. Shiffman S. Conceptualizing analyses of ecological momentary assessment data. Nicotine Tob Res 2014 May;16 Suppl 2(Suppl 2):S76-S87 [FREE Full text] [doi: 10.1093/ntr/ntt195] [Medline: 24323571]
20. Poulton A, Pan J, Bruns LR, Sinnott RO, Hester R. Assessment of alcohol intake: retrospective measures versus a smartphone application. Addict Behav 2018 Aug;83:35-41. [doi: 10.1016/j.addbeh.2017.11.003] [Medline: 29128148]
21. Stone A, Shiffman S, Atienza A, Nebeling L. The Science of Real-time Data Capture: Self-reports in Health Research. England: Oxford University Press; 2007.
22. Martino SC, Kovalchik SA, Collins RL, Becker KM, Shadel WG, D'Amico EJ. Ecological momentary assessment of the association between exposure to alcohol advertising and early adolescents' beliefs about alcohol. J Adolesc Health 2016 Jan;58(1):85-91 [FREE Full text] [doi: 10.1016/j.jadohealth.2015.08.010] [Medline: 26480846]
23. Collins RL, Martino SC, Kovalchik SA, D'Amico EJ, Shadel WG, Becker KM, et al. Exposure to alcohol advertising and adolescents' drinking beliefs: role of message interpretation. Health Psychol 2017 Sep;36(9):890-897 [FREE Full text] [doi: 10.1037/hea0000521] [Medline: 28639822]
24. Global status report on alcohol and health 2018. World Health Organization. 2019. URL: https://who.int/publications/i/item/9789241565639 [accessed 2024-08-12]
25. Alcohol consumption per capita in Hong Kong. Department of Health. 2022. URL: https://www.change4health.gov.hk/en/alcohol_aware/figures/alcohol_consumption/index2.html [accessed 2024-06-01]
26. Yu J, Sumerlin TS, Goggins WB, Dong D, Chung RY, Kim JH. First- and second-hand alcohol-related harms among urban Chinese: a population-based study from Hong Kong. Drug Alcohol Rev 2022 Jan;41(1):208-220. [doi: 10.1111/dar.13339] [Medline: 34184790]
27. Chung VCH, Yip BHK, Griffiths SM, Yu ELM, Kim JH, Tam WWS, et al. The impact of cutting alcohol duties on drinking patterns in Hong Kong. Alcohol Alcohol 2013;48(6):720-728. [doi: 10.1093/alcalc/agt065] [Medline: 23825091]
28. Kim JH, Lee S, Chow J, Lau J, Tsang A, Choi J, et al. Prevalence and the factors associated with binge drinking, alcohol abuse, and alcohol dependence: a population-based study of Chinese adults in Hong Kong. Alcohol Alcohol 2008;43(3):360-370. [doi: 10.1093/alcalc/agm181] [Medline: 18230698]
29. 2008/09 Survey of drug use among students. Narcotics Division Security Bureau HKSAR. 2009. URL: https://www.nd.gov.hk/pdf/survey_drug_use/2008-2009/Appendix.pdf [accessed 2024-02-03]
30. 2020/21 Survey of drug use among students. Narcotics Division Security Bureau HKSAR. 2022. URL: https://www.nd.gov.hk/pdf/survey_drug_use/2020-2021/Appendix.pdf [accessed 2024-02-03]
31. Kim JH, Lee S, Chan KWC, Lau J, Tsang A, Griffiths SM. A population-based study on the prevalence and correlates of drinking and driving in Hong Kong. Accid Anal Prev 2010 Jul;42(4):994-1002. [doi: 10.1016/j.aap.2009.12.001] [Medline: 20441805]
32. Lau JTF, Kim JH, Tsui HY. Prevalence, health outcomes, and patterns of psychotropic substance use in a Chinese population in Hong Kong: a population-based study. Subst Use Misuse 2005;40(2):187-209. [doi: 10.1081/ja-200048454] [Medline: 15770884]

33. Abdullah ASM, Fielding R, Hedley AJ. Patterns of cigarette smoking, alcohol use and other substance use among Chinese university students in Hong Kong. Am J Addict 2002;11(3):235-246. [doi: 10.1080/10550490290088018] [Medline: 12202016]
34. Chan RHW, Dong D, Yu J, Kim JH. Who is being targeted by alcohol social media marketing? A study of Chinese young adults in Hong Kong. Drug Alcohol Rev 2024. [doi: 10.1111/dar.13892] [Medline: 38951715]
35. Authority C. Generic code of practice on television advertising standards. Communications Authority. 2022. URL: https://www.coms-auth.hk/filemanager/common/policies_regulations/cop/code_tvad_e.pdf [accessed 2024-11-22]
36. Chan RHW, Wong TY, Dong D, Kim JH. Alcohol social media marketing in Hong Kong: a content analysis of Facebook posts. J Adolesc Health 2023 Sep;73(3):461-469. [doi: 10.1016/j.jadohealth.2023.05.007] [Medline: 37389523]
37. Lo SKH, Louie JCY. Food and beverage advertising in Hong Kong mass transit railway stations. Public Health Nutr 2020 Oct;23(14):2563-2570 [FREE Full text] [doi: 10.1017/S136898002000083X] [Medline: 32524925]
38. Shiffman S. Ecological momentary assessment (EMA) in studies of substance use. Psychol Assess 2009 Dec;21(4):486-497 [FREE Full text] [doi: 10.1037/a0017074] [Medline: 19947783]
39. Williams MT, Lewthwaite H, Fraysse F, Gajewska A, Ignatavicius J, Ferrar K. Compliance with mobile ecological momentary assessment of self-reported health-related behaviors and psychological constructs in adults: systematic review and meta-analysis. J Med Internet Res 2021 Mar 03;23(3):e17023 [FREE Full text] [doi: 10.2196/17023] [Medline: 33656451]
40. Yang C, Linas B, Kirk G, Bollinger R, Chang L, Chander G, et al. Feasibility and acceptability of smartphone-based ecological momentary assessment of alcohol use among African American men who have sex with men in Baltimore. JMIR Mhealth Uhealth 2015 Jun 17;3(2):e67 [FREE Full text] [doi: 10.2196/mhealth.4344] [Medline: 26085078]
41. Shiffman S, Stone AA. Ecological momentary assessment: a new tool for behavioral medicine research. In: Technology and Methods in Behavioral Medicine. East Sussex, England: Psychology Press; 1998:117-131.
42. Babor T, Grant M. From clinical research to secondary prevention: international collaboration in the development of the alcohol disorders identification test (AUDIT). Alcohol Health Res World 1989;13(4):371-374.
43. Saunders JB, Aasland OG, Babor TF, de la Fuente JR, Grant M. Development of the alcohol use disorders identification test (AUDIT): WHO collaborative project on early detection of persons with harmful alcohol consumption--II. Addiction 1993;88(6):791-804. [doi: 10.1111/j.1360-0443.1993.tb02093.x] [Medline: 8329970]
44. Li Q, Babor TF, Hao W, Chen X. The Chinese translations of Alcohol Use Disorders Identification Test (AUDIT) in China: a systematic review. Alcohol Alcohol 2011;46(4):416-423 [FREE Full text] [doi: 10.1093/alcalc/agr012] [Medline: 21467046]
45. Lillard DR, Molloy E, Zan H. Television and magazine alcohol advertising: exposure and trends by sex and age. J Stud Alcohol Drugs 2018 Nov;79(6):881-892 [FREE Full text] [doi: 10.15288/jsad.2018.79.881] [Medline: 30573019]
46. Jones SC, Magee CA. Exposure to alcohol advertising and alcohol consumption among Australian adolescents. Alcohol Alcohol 2011;46(5):630-637. [doi: 10.1093/alcalc/agr080] [Medline: 21733835]
47. Martino SC, Setodji CM, Collins RL, D'Amico EJ, Shadel WG, Tolpadi A, et al. Persistence of shifts in beliefs associated with exposure to alcohol advertising among adolescents. J Stud Alcohol Drugs 2018 May;79(3):399-407 [FREE Full text] [doi: 10.15288/jsad.2018.79.399] [Medline: 29885147]
48. Kim JH, Chan KWC, Chow JKW, Fung KP, Fong BYF, Cheuk KK, et al. University binge drinking patterns and changes in patterns of alcohol consumption among chinese undergraduates in a Hong Kong university. J Am Coll Health 2009;58(3):255-265. [doi: 10.1080/07448480903295318] [Medline: 19959440]
49. McClure AC, Stoolmiller M, Tanski SE, Engels RCME, Sargent JD. Alcohol marketing receptivity, marketing-specific cognitions, and underage binge drinking. Alcohol Clin Exp Res 2013 Jan;37 Suppl 1(Suppl 1):E404-E413 [FREE Full text] [doi: 10.1111/j.1530-0277.2012.01932.x] [Medline: 23256927]
50. Henriksen L, Feighery EC, Schleicher NC, Fortmann SP. Receptivity to alcohol marketing predicts initiation of alcohol use. J Adolesc Health 2008 Jan;42(1):28-35 [FREE Full text] [doi: 10.1016/j.jadohealth.2007.07.005] [Medline: 18155027]
51. Jones A, Remmerswaal D, Verveer I, Robinson E, Franken IHA, Wen CKF, et al. Compliance with ecological momentary assessment protocols in substance users: a meta-analysis. Addiction 2019 Apr;114(4):609-619 [FREE Full text] [doi: 10.1111/add.14503] [Medline: 30461120]
52. Shadel WG, Martino SC, Setodji C, Scharf D. Momentary effects of exposure to prosmoking media on college students' future smoking risk. Health Psychol 2012 Jul;31(4):460-466 [FREE Full text] [doi: 10.1037/a0027291] [Medline: 22353027]

Abbreviations

AOR: adjusted odd ratio
AUDIT: Alcohol Use Disorder Identification Test
EMA: ecological momentary assessment
HKU: University of Hong Kong
RCT: randomized controlled trial

Edited by L Buis; submitted 30.04.24; peer-reviewed by S Martino, J Kim; comments to author 26.05.24; revised version received 14.06.24; accepted 08.08.24; published 03.09.24.

Please cite as:
Zhang MJ, Luk TT, Ho SY, Wang MP, Lam TH, Cheung YTD
Ecological Momentary Assessment of Alcohol Marketing Exposure, Alcohol Use, and Purchases Among University Students: Prospective Cohort Study
JMIR Mhealth Uhealth 2024;12:e60052
URL: https://mhealth.jmir.org/2024/1/e60052
doi:10.2196/60052
PMID:39226102

Original Paper

An Ecological Momentary Assessment Approach of Environmental Triggers in the Role of Daily Affect, Rumination, and Movement Patterns in Early Alcohol Use Among Healthy Adolescents: Exploratory Study

Maren Prignitz[1*], MSc; Stella Guldner[2*], PhD; Stephan Johann Lehmler[3], MA; Pascal-M Aggensteiner[2], PhD; Frauke Nees[2,3], PhD; IMAC-Mind Consortium[4*]

[1]Institute of Cognitive and Clinical Neuroscience, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany

[2]Institute of Child and Adolescent Psychiatry and Psychotherapy, Central Institute of Mental Health, Medical Faculty Mannheim, University of Heidelberg, Mannheim, Germany

[3]Institute of Medical Psychology and Medical Sociology, University Medical Center Schleswig Holstein, Kiel University, Kiel, Germany

[4]see Acknowledgments

[*]these authors contributed equally

Corresponding Author:
Maren Prignitz, MSc
Institute of Cognitive and Clinical Neuroscience
Central Institute of Mental Health, Medical Faculty Mannheim
Heidelberg University
J 5
Mannheim, 68159
Germany
Phone: 49 62117036313
Email: maren.prignitz@zi-mannheim.de

Abstract

Background: Adolescence is a period characterized by an increased susceptibility to developing risky alcohol consumption habits. This susceptibility can be influenced by social and situational factors encountered in daily life, which, in conjunction with emotions and thoughts, contribute to behavioral patterns related to alcohol use even in the early stages of alcohol experimentation, when initial experiences with alcohol are formed, and regular consumption is still evolving.

Objective: This study aimed to examine the association between detailed behavioral and movement patterns, along with emotional and cognitive factors, and the early onset of alcohol use in the everyday lives of adolescents.

Methods: A total of 65 healthy adolescents (33 male, twenty-nine 14-year-olds, and thirty-six 16-year-olds) underwent mobile-based ecological momentary assessments on alcohol (once a day at 9 AM, assessing alcohol use the day before), positive and negative affect, craving, rumination, and social context (6 prompts/day at 9 AM, 11 AM, 2 PM, 4 PM, 6 PM and 8 PM), type of day (weekdays or weekends, with weekend including Fridays, Saturdays, and Sundays), and using geospatial measures (specifically roaming entropy and number and type of trigger points for alcohol use met) over 14 days. After adjusting for a compliance rate of at least 50%, 52 participants (26 male and twenty-four 14-year-olds) were included in the analyses.

Results: Generalized linear multilevel models revealed that higher positive affect (b=0.685, P=.007), higher rumination (b=0.586, P=.02), and a larger movement radius (roaming entropy) (b=8.126, P=.02) were positively associated with alcohol use on the same day. However, social context (b=–0.076, P=.90), negative affect (b=–0.077, P=.80), or potential trigger points (all P>.05) did not show significant associations. Alcohol use varied depending on the type of day, with more alcohol use on weekends (b=1.082, P<.001) and age (t_{50}=–2.910, P=.005), with 16-year-olds (mean 1.61, SD 1.66) reporting more days of alcohol consumption than 14-year-olds (mean 0.548, SD 0.72).

Conclusions: Our findings support previously identified factors as significant contributors to very early and low levels of alcohol consumption through fine-grained analysis of daily behaviors. These factors include positive affect, rumination, weekend days,

and age. In addition, we emphasize that exploratory environmental movement behavior (roaming entropy) is also significantly associated with adolescent alcohol use, highlighting its importance as an additional factor.

(JMIR Mhealth Uhealth 2024;12:e53401) doi:10.2196/53401

KEYWORDS

alcohol use; adolescence; affect; rumination; ecological momentary assessment; geospatial measures

Introduction

Adolescence (spanning from ages 10 to 19 years, according to the World Health Organization [1]) represents a sensitive period characterized by rapid developmental transitions across biological, neurophysiological, psychological, and social domains [2]. These changes are important for the development of healthy behaviors in daily life [3] but can, inversely, also pose a risk for the development of unhealthy behavioral patterns. Such behaviors include the use of alcohol, which is usually initiated in midadolescence (ages 14 to 17 years) [4,5] and peaks in late adolescence around the age of 21 years [6-9]. Alcohol represents the most commonly used and abused substance during adolescence (eg, in Germany [10]) while also accounting for the most pervasive and devastating effects on life quality, general well-being, and neurodevelopmental processes [11-14]. Understanding the factors influencing very early adolescent alcohol use is crucial, as earlier initiation of alcohol consumption during this period is associated with heightened alcohol use in late adolescence and an elevated risk of developing alcohol use disorder in young adulthood [15,16]. Furthermore, alcohol use during adolescence can adversely affect structural and functional neural development [14,17], as well as various aspects of adolescent health and well-being [18,19], particularly when it involves risky behaviors such as binge drinking [13]. Therefore, elucidating the determinants of very early adolescent alcohol use, when potentially harmful patterns may be emerging for the first time, is paramount.

A Theoretical Approach: The Incentive Sensitization Theory

Previous frameworks stem from findings on inter-individual differences in neurophysiological, genetic, personality, social, and environmental factors and their interaction in contribution to adolescent alcohol use [20-23]. The Incentive Sensitization Theory (IST) [24] proposes alcohol as a potentially rewarding stimulus that is liked at first (ie, its use is positively reinforced), for instance, due to its positive impact on mood and evolves into a so-called wanting in the course of addiction development, where its use is negatively reinforced for instance through the amelioration of negative mood states or craving [24]. Greater incentive salience is assigned to alcohol-related cues (eg, the favorite bar or the favorite alcoholic drink), which results in greater attention to these cues compared with other rewarding cues or conditions [25]. Importantly, these cues can then trigger the urge to drink alcohol [26]. Especially adolescents show hyperreactive responses to alcohol-related rewarding cues [7,27]. This model represents a theoretical approach to explaining how alcohol use transitions from early use to habitual use over time. However, specifically in the very early periods of alcohol use, that is, the onset and first experiences, information on these processes is rather scarce. Recently, through advanced methodologies, it is becoming possible to receive much more fine-grained information at the intraindividual level and in daily life to explain when and why adolescents consume alcohol.

Daily Life Experiences Influence Alcohol Use in Adolescence

Methods to map daily life experiences, trajectories, and movements in daily life are increasingly available. These encompass ecological momentary assessments (EMAs) and GPS tracking (geospatial measures), enabling the concurrent investigation of interindividual mechanisms identified as risk factors for alcohol use (eg, IST) alongside behavioral patterns. These patterns include movements within the daily environment and exposure to specific environmental triggers (eg, nightclubs, youth gatherings, or social contexts such as being in close proximity to peers). This combination is an important addition to traditional diary methods as a resource for gaining information on individuals' health status over time. Due to the potential to capture momentary states multiple times throughout the day, EMA methodology allows to obtain a high-resolution representation of self-reports, including emotions and thoughts as well as daily behavior and information on contexts, including social situations. Its intuitive usability further ensures greater attractiveness and, consequently, increased compliance and ecological validity compared with retrospective assessments as well as paper and pencil applications [28,29]. This also becomes significant in the realm of health research, as it provides more sensitive data regarding critical daily triggers for symptoms. Consequently, it aids in deciding on the most adaptive interventions. Previous studies using EMA have demonstrated that in adolescent and young adult alcohol users, contextual factors [30], such as the presence of peers or being in specific locations like bars or restaurants [31], along with mood-related factors, including heightened negative affect [32-34], are highly influential in initiating alcohol use. In addition, smartphone technology allows us to assess geospatial movement patterns as proxies for exploratory behaviors and environmental context factors in which specific behaviors occur (ie, alcohol-associated trigger points) and combine this with information gathered from using EMA. This provides the possibility to investigate and identify triggers with not only respect to individual feelings or thoughts but also specific contexts, including information on where (eg, private or public environments) or when (eg, time of the day, day in the week) adolescents consume alcohol [35]. These triggers can also encompass specific situational cues that are associated with alcohol use [36-38]. This information can be gained through exploratory movement patterns, which are estimated, for example, through roaming entropy, an index of "the variability in an individuals' physical location over the course of a day" [39]. Even in the early period of alcohol use,

these data could be beneficial to explore how potential trigger points develop over time. In terms of alcohol use, geospatial measures have successfully been used in app-based alcohol use disorder treatment in adults by sending alerts if patients are near alcohol-related trigger points that are risky places for lapse, relapse, or craving [40-43]. Furthermore, studies have demonstrated that neighborhood disadvantages are associated with increased substance use among adolescents [44], and adolescents tend to consume alcohol in proximity to locations where it is available for purchase [45]. However, environmental exploratory movement patterns, such as roaming entropy, have not been considered in these studies. Thus, together with EMA data, geospatial measures might allow additional insight into adolescent drinking behavior and driving factors and can add to our understanding of transitions from nonrisky to risky alcohol use, and thus inform the development of early prevention approaches, for example, in the form of ecological momentary interventions (EMI). Therefore, this study aimed to delineate the influence of daily variability in affect, craving, rumination, and presence of others (social context), together with exposure to potential environmental alcohol triggers and the characterization of environmental movement patterns on early adolescent daily alcohol use.

Methods

Sample

Participants were included in the study if they were either 14 or 16 years old at the time of enrollment, aiming to capture early drinking patterns typically initiated around the age of 14 years, and the period when alcohol consumption becomes more frequent and prevalent, around age 16 years [4]. This approach facilitated a better understanding and comparison of these developmental stages. In addition, participants had to be fluent German speakers, right-handed, safe to perform magnetic resonance imaging tasks (eg, no metal in the body or claustrophobia), and without any present history of psychiatric or somatic disorders [46]. Out of 72 recruited participants, 70 participants (twenty-nine 14-year-olds, thirty-six 16-year-olds, 49% female) underwent the entire assessment battery. This included a series of questionnaires (eg, sociodemographic information of participants and their parents, general alcohol consumption, stress, mindfulness, experience with meditation techniques, emotion regulation, personality, quality of life, and rumination), behavioral tasks, neuropsychological tasks in an magnetic resonance imaging scanner, and EMA. Participants received a reimbursement of €50 (approximately US $58) for their participation in the entire study. For the present purpose, only the EMA was of interest, and 65 participants underwent the entire EMA assessment.

Instruments

Daily Life Experiences and Geospatial Measures

Daily life experiences were assessed by an EMA. The implementation of the EMA was done by movisensXS (movisens GmbH) [47], which allows the integration of relevant items and additional information like geospatial measures. Within the EMA assessment, participants were asked to carry along a study phone (Nokia 5 with Android version 7.1.1 Nougat) over 14 days [48]. This period was selected due to the relatively low frequency of drinking behavior observed in a healthy sample of adolescents, aiming to increase the variance in drinking occasions. A total of 6 prompts per day were presented on fixed time points (at 9 AM, 11 AM, 2 PM, 4 PM, 6 PM, and 8 PM) with 2 reminders at 5-minute intervals [48]. We thus asked participants to answer a total of 43 items regarding affect, event appraisal, craving, social context, self-esteem, and rumination within a 15-minute time window. Answering one prompt took about 2 minutes. For the current purpose, we concentrated on affect (5 items for positive affect, 5 items for negative affect, eg, “Right now I feel scared”; response scale: 1=“not at all” to 7=“very much” [49], based on Positive and Negative Affect Schedule [50]), craving (6 items, eg, “Drinking would make things seem perfect”; response scale: 1=“I totally agree” to 7=“I totally disagree,” based on ACQ [51]), social context (1 item: “Are you currently alone?” [Yes/no]) and rumination (4 items, eg, “The moment before the beep, I thought about my feelings,” reversed, response scale: 1=“I totally agree” to 7=“I totally disagree,” similar to [52]). We did not include event appraisal or self-esteem in our study. This is based on assumptions that event appraisal for good and bad experiences is reflected in positive and negative affect [53] and that fluctuations in self-esteem are rather discussed in relation to depression [54]. For positive affect, negative affect, craving, and rumination, a mean score was and for social context, a sum score was calculated for each day within each participant, with higher scores indicating higher expression of each construct (for social context, higher scores indicating more time spent alone within a day). In order to increase compliance with EMA measurements, we offered participants an additional monetary incentive (cinema voucher) if they responded to 90% or more of the EMA prompts. As compliance is central to the quality of EMA data [55], only subjects with a compliance rate of 50% or more were included in the data analyses [55], leaving a sample of 52 participants (twenty-four 14-year-olds and 26 female).

In addition to the psychometric instruments of the EMA, we gathered the GPS location of each participant on a per-minute basis, allowing us to build detailed daily movement patterns. While GPS was not linked directly to individual EMA assessments, places, or drinking events, it still enables a characterization of daily activity of and environmental influences on the participants. In addition, participants provided information on places they visited frequently during the EMA assessment (what or who was visited and where was the place or address).

Alcohol Use

Our main alcohol outcome was assessed within the EMA once a day, within the 9 AM prompt. Participants were asked about their alcohol use the day before (1 item: “Did you drink alcohol yesterday?” [Yes/No]). This method was used to capture every occurrence of alcohol use the day before, since the last prompt of EMA reached participants at 8 PM, and it is conceivable that participants used alcohol even after this time point [56]. If participants missed the 9 AM prompt, we coded alcohol use for the day before as missing data. In addition, as a control, we

included the sum score (range 0 to 40) of the 10-item self-report Alcohol Use Disorder Identification Test (AUDIT) [57] to check whether general alcohol use is generally associated with daily alcohol patterns.

Data Analysis

Geospatial Measures Preprocessing

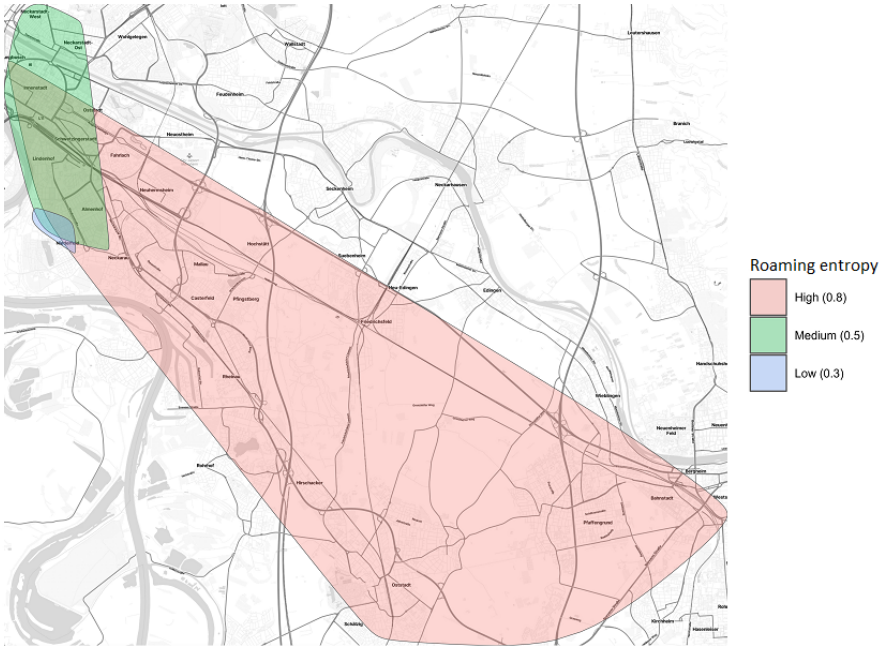
To use geospatial measures in our model, we preprocessed GPS information. In particular, we calculated 2 measures based on individual trajectories: Roaming Entropy as a marker of environmental exploratory movement and the contact with potential alcohol trigger points, that is, locations where alcohol use might be especially salient. All analyses were done using R (R Foundation of Statistical Computing) [58].

Roaming entropy (RE) is an information-theoretic measure used to quantify the variability of a movement trajectory. Specifically, it calculates the entropy of a trajectory based on unique places visited [55]. Places are defined as distinct latitude-longitude-pairs, rounded to the fourth decimal place [39,59], providing an accuracy of approximately 11 m around the participant, which corresponds to the size of a small house or street section.

For each participant and each day separately, Roaming Entropy was calculated using the following formula:

Here, p_i is the probability (estimated by the relative within-day frequency) that the participant is at location i, while n is the total number of unique locations visited during that day. RE can take values from zero to one, with zero meaning a participant did not move at all during the day and one meaning that every place visited was unique [39]. Figure 1 shows exemplary movement patterns with high, medium, and low RE.

Figure 1. Abstract examples of 3 participants categorized by their roaming entropy within this study, depicting high (red), medium (green), and low (blue) levels. While roaming entropy primarily quantifies movement variability rather than area, it frequently aligns with larger areas covered.



Please note, the map tiles are by Stamen Design, under CC BY 4.0. Data by OpenStreetMap, under ODbL.

To calculate the contact of each participant with trigger points (TPs), we had to use additional data sources to add semantic meaning to locations. We used geolocalized information from "OpenStreetMap" (OpenStreetMap Stiftung) [60] and the "Kulturatlas Mannheim" (Stadt Mannheim) [61] to identify bars, clubs, parks, and other potential places of interest. These places were selected based on insights from previous research indicating that adolescents typically consume alcohol in places where it is legally permissible in Germany (such as private homes, friends' homes, bars, or restaurants for 16-year-olds for specific alcoholic beverages) and that alcohol consumption among adolescents often occurs in social settings (such as clubs, meeting spots, etc). In addition, we used information provided by participants regarding their frequently visited places during the assessment period to distinguish between places of residence, study, or social interactions, such as meeting with peers.

Using these data sources, we defined categories of potential TPs. To the best of our knowledge, there is currently no standardized framework for classifying potential TPs for adolescent alcohol use based on geospatial features. Therefore, we had to create the categories by considering the available tagged locations in the data and the information provided by participants. We defined categories of TPs related to distinct social situations likely associated with alcohol use in adolescents. The five defined categories were "nightlife," "culture," "leisure," "peers," and "meeting spots," with most of these categories standardized across the entire sample (all locations grouped under each category can be found in Table S1 in Multimedia Appendix 1). Figure 2 shows a view of the city of Mannheim, Germany, with all places categorized

accordingly. Please note, the map tiles are by Stamen Design, under CC BY 4.0. Data by OpenStreetMap, under ODbL.

For each subject, we tracked how often, during one day, they would get in contact with each of the TP categories. Contact was defined by physical closeness (within 15 m) with places belonging to a category. Based only on GPS trajectories, without further detailed information on the locality and participant's direction of view, any measurement of closeness can only be a rough estimate. In this context, the radius of 15 m was chosen as a reasonable tradeoff between accuracy and precision. Small changes in the radius did not affect our model's results. Continuous contact (eg, many minutes or hours at the same location) was not accumulated over time but rather counted as one contact. It was possible to be in contact with multiple types of TPs at the same time.

Figure 2. Trigger point categories within city of Mannheim, Germany.

Effects of Daily Life Experiences and Geospatial Information on Daily Alcohol Use

For a comprehensive examination of daily alcohol use in relation to daily life experiences and geospatial information, we used binomial generalized multilevel models with logit link function. These models are used to conduct logistic regression analyses of alcohol use as a binary outcome for each day and subject. For this purpose, the alcohol item, which was assessed on the following day, was shifted to the day before, and thus, the alcohol use of the respective day was related to the corresponding life experiences of that day. Alcohol use was predicted then by daily life experiences as independent variables assessed by EMA: positive affect (PA), negative affect (NA), craving, rumination, and social context. In addition, we used the geospatial measures TPs and RE of each day as independent variables, as well as a binary classification of assessment day into weekday (Monday, Tuesday, Wednesday, and Thursday) or day of weekend (Friday, Saturday, and Sunday) and personal level variables age and sex. To account for intraindividual correlation, we used a generalized linear multilevel model approach with a fixed slope and a random intercept for each participant, with participant as random effect and independent variables (PA, NA, craving, rumination, social context, TP, RE, weekday, age, and sex) as fixed effects.

Not all participants responded to every EMA prompt on each day. Requiring participants to answer all prompts to be included in the analyses is an open discussion in the EMA research, as it would always result in a significant reduction of the sample size [55]. Similarly, in our study, requiring participants to answer all 6 prompts would significantly reduce the sample size, making it too stringent as a cutoff criterion. Conversely, setting a cutoff too low could potentially introduce bias control challenges, particularly if nonresponse is related to drinking behavior. Therefore, we calculated a model that strikes a balance between the number of answered prompts (here, 4 prompts per day) and the number of participants, including all participants in the analysis. This approach aligns with previous work [55]. In addition, we calculated 5 additional models with varying levels of prompt completion (1 prompt/day, 2 prompts/day, 3 prompts/day, 5 prompts/day, and 6 prompts/day) to assess the stability of the results and the impact of participant compliance (more details in Figure S1 in Multimedia Appendix 1). For our main model, we evaluate the overall significance by comparing the marginal or null-model (model with only random effects) deviance to the conditional model (model containing random and fixed effects) deviance by chi-square tests. The significance of random effects is tested using the Wald test.

All analyses were done with R [58] using the lme4 package [62] and a significance threshold of $P<.05$. To check for overall effects of sex, age, and the COVID-19 pandemic restrictions, we used *t* tests for independent measures to compare the overall reported days of alcohol use within the 14-day assessment period. Since the data acquisition period coincided with the COVID-19 pandemic in Germany [63], we decided to check for these potential differences in our sample. We also evaluated

the correlation of the AUDIT sum score in the sample with the number of drinking days during the EMA assessment and tested whether compliance rate had an influence on all assessed variables with *t* tests.

Ethical Considerations

Within our subproject of the IMAC-Mind consortium [64], participants were recruited through advertisements in schools in Mannheim, Germany, through social media, and with the help of the local registration office of Mannheim. Both participants and their primary caretaker gave written informed consent before study participation. The ethics committee of Medical Faculty Mannheim, Heidelberg University, approved the study (2007-024-N-MA).

Results

General Sample Information

There was no significant difference in alcohol use between participants doing the EMA under COVID-19 restrictions compared with participants doing the EMA before COVID-19 occurred (t_{50}=–0.372, *P*=.71), and no significant differences in alcohol use due to sex (t_{50}=0.199, *P*=.85). However, there was a significant effect for age (t_{50}=–2.910, *P*=.005), with 16-year-olds (mean 1.61, SD 1.66) reporting significantly more alcohol drinking days than 14-year-olds (mean 0.548, SD 0.72). We noted a significant positive correlation (*r*=0.49, 95% CI 0.26 to 0.68; *P*<.001) between reported alcohol days within the EMA and AUDIT sum score. Additionally, there was a significant sex difference in rumination (t_{50}=–2.513, *P*=.02), with males having lower scores (mean 1.56, SD 0.513) than females (mean 2.08, SD 0.907) and a significant difference between participants doing the EMA under COVID-19 restrictions compared with participants doing EMA before COVID-19 occurred in negative affect (t_{50}=–2.091, *P*=.04) with participants doing EMA before COVID-19 having lower scores (mean 0.89, SD 0.67) than participants doing EMA under COVID-19 restrictions (mean 1.26, SD 0.55) and social context (t_{50}=2.027, *P*=.048) with participants doing EMA before COVID-19 having higher scores (mean 0.549, SD 0.215) than participants doing EMA under COVID-19 restrictions (mean 0.438, SD 0.162). All other variables were comparable between males and females and between 14- and 16-year-olds (all *P*>.05; more details in Table 1 and Table 2, and Figure S2 in Multimedia Appendix 1). Furthermore, no significant differences were observed in any of the variables between participants who did not meet the inclusion criteria of compliance (>50%) and those who did meet these criteria (all *P*>.05).

Table 1. Distribution of general sample characteristics within the total sample and separated by age.

Variable	Total, n	14-year-olds, n	16-year-olds, n
Sex			
Male	26	12	14
Female	26	12	14
EMA [a]			
Before the COVID-19 pandemic	15	7	8
During the COVID-19 pandemic	37	17	20
Current activity or job			
Student	50	24	26
Other activity or job	2	0	2
School type			
Middle school[b]	2	0	2
Comprehensive school[c]	3	0	3
Academic high school[d]	43	23	19
Other	2	1	2
Highest graduation of the father			
Certificate of secondary education[e]	8	3	5
Middle school[b]	3	3	0
Qualification for access to higher education[f]	24	11	13
University degree	16	7	9
Other	1	0	1
Highest graduation of the mother			
Certificate of secondary education[e]	2	0	2
Middle school[b]	9	5	4
Qualification for access to higher education[f]	28	14	14
University degree	12	4	8
Other	1	1	0

[a]EMA: ecological momentary assessment.

[b]Corresponds to “Realschule” in the German educational system.

[c]Corresponds to “Gesamtschule” in the German educational system.

[d]Corresponds to “Gymnasium” in the German educational system.

[e]Corresponds to “Hauptschulabschluss” in the German educational system.

[f]Corresponds to “Abitur/Fachabitur” in the German educational system.

Table 2. General information about ecological momentary assessment for the total sample and separated by age.

	Total, mean (SD)	14-years-old, mean (SD)	16-years-old, mean (SD)
Answered prompts (max[a] possible: 84)	65.192 (14.712)	62.7 (14.3)	67.3 (15.0)
Compliance rate	0.797 (0.140)	0.771 (0.139)	0.818 (0.140)
Positive affect	3.70 (0.803)	3.66 (0.803)	3.74 (0.816)
Negative affect	1.15 (0.607)	1.08 (0.534)	1.22 (0.666)
Craving	1.50 (0.765)	1.28 (0.495)	1.68 (0.905)
Rumination	1.82 (0.774)	1.66 (0.475)	1.96 (0.947)
Social context	0.470 (0.184)	0.493 (0.178)	0.450 (0.190)
Alcohol use days	1.115 (1.409)	0.542 (0.721)	1.61 (1.66)

[a]Max: maximum.

Effect of Daily Life Experiences and Geospatial Measures on Daily Alcohol Use

An overview of the model statistics can be found in Table 3. The model comprises a total of 587 observation days, which is the sum of the days on which participants completed at least 4 prompts per day. There is high variability between participants, with a range of 4 days included in the model for the participant with the fewest days and 14 days for the participant with the most days (mean observation days per participant: 11.29, SD 3.15). Comparing the deviance of the conditional or full model (with fixed and random effects) to the marginal or null model (containing only fixed effects), we see an overall significant difference. The pseudo R^2 of the generalized linear mixed models for both conditional and marginal models show weak to moderate fit. While some predictors remained stable, there was variability across models with different amounts of prompts included (Table S2 in Multimedia Appendix 1 shows detailed results of models with 1, 2, 3, 5, or 6 prompts included). In the model analyzing 4 prompts per day (more details in Table 3), predictors of alcohol use were positive affect (b=0.685, P=.007), rumination (b=0.586, P=.02), weekend day (b=1.082, P<.001), and RE (b=8.126, P=.02). All other predictors did not reach significance in this model.

Table 3. Detailed outcome of generalized linear mixed model for daily life experience and geospatial variables (roaming entropy and potential trigger points) on drinking behavior for at least 4 prompts per day.

Parameters	Values
4 prompts	
Participants, n	52
Total observation days[a], n	587
Overall model design	
Marginal (pseudo) R^2	0.244
Conditional (pseudo) R^2	0.405
Deviance	377.40
Null model	332.83
df	14
Chi-square	44.571
P value	<.001
Predictor	
Intercept	
b	–3.519
SE	1.910
z score	–1.843
P value	.07
Positive affect [b]	
b	0.685
SE	0.253
z score	2.703
P value	.007
Negative affect	
b	–0.077
SE	0.303
z score	–0.252
P value	.80
Craving	
b	0.348
SE	0.189
z score	1.844
P value	.07
Rumination [b]	
b	0.586
SE	0.241
z score	2.434
P value	.02
Social context	
b	–0.076
SE	0.584
z score	–0.131

Parameters	Values
P value	.90
Weekend [b]	
b	1.082
SE	0.305
z score	3.534
P value	<.001
Roaming entropy [b]	
b	8.126
SE	3.525
z score	2.305
P value	.02
Culture (trigger)	
b	0.015
SE	0.542
z score	0.027
P value	.98
Nightlife (trigger)	
b	0.040
SE	0.305
z score	0.132
P value	.90
Leisure (trigger)	
b	0.030
SE	0.383
z score	0.078
P value	.94
Meeting spots (trigger)	
b	–0.081
SE	0.210
z score	–0.385
P value	.70
Peers (trigger)	
b	–0.005
SE	0.460
z score	–0.010
P value	.99
Age	
b	0.721
SE	0.457
z score	1.579
P value	.11
Sex	

Parameters	Values
b	–0.295
SE	0.451
z score	–0.654
P value	.51

[a]Number of days with at least 4 answered prompts per day summarized over all participants.

[b]Predictors are significant at $P<.05$ or below.

Discussion

Interpretation of Results

The aim of this study was to examine the influence of daily life experiences, such as affect, craving, rumination, and social context, together with the exposure to potential environmental alcohol triggers (TP) and environmental movement patterns (RE) in contributing to daily alcohol use in healthy adolescents in the very early period of alcohol experience. Our exploratory results show that daily variations in positive affect, rumination, and RE can predict drinking behavior in healthy 14- and 16-year-old adolescents. We identified core concepts that might represent a pattern associated with a higher risk for the development of risky alcohol use.

Interestingly, positive affect (not negative affect) was significantly positively associated with alcohol use on that day in our sample, with robust effects across varying levels of compliance rates (more details in Table S2 in Multimedia Appendix 1). From previous research concerning affective states on alcohol use (stress reactivity theory [5]), it would be expected that alcohol use would follow negative affect, representing a coping strategy to deal with unwanted and aversive emotions [65,66]. In contrast, our results indicate that during early alcohol use, with a low level of frequency, that is, seldom use, adolescents use alcohol when experiencing positive emotions during the day, suggesting it might be used to enhance positive emotions. This is in line with previous work on alcohol-drinking motives in adolescents [18,67]. In addition, adolescents often underestimate the risks of alcohol use and have positive attitudes toward alcohol [68]. However, it is important to note that our data does not allow for the derivation of directionality between positive affect and alcohol use. This limitation stems from the retrospective assessment of alcohol use in this study, conducted without further situational or contextual information. It is plausible that either affect may change as a result of alcohol use, or that situational features prompt both positive affect and alcohol use (eg, a celebratory event, more details in Table S3 in Multimedia Appendix 1), or that adolescents experiencing positive emotions are more likely socialize and thus consume alcohol with peers. Future studies could address this by adopting a more detailed assessment of alcohol use, including information on when, where, and in what context alcohol was consumed, or by using event-related sampling schemes to specifically collect data when alcohol is used.

Furthermore, another central factor in association with alcohol use was rumination, with higher rumination being associated with a higher probability of alcohol use on that day. This association was rather weak compared with the other predictors, and rumination ratings were generally low, but the results are still robust across differing compliance thresholds (more details in Table S2 in Multimedia Appendix 1). Thus, our data suggests that already in early, as of yet non-hazardous alcohol-using adolescents, rumination is a factor that contributes to momentary alcohol use. As we outlined above, we cannot infer directionality from our data. However, it can be speculated that even at early use, alcohol might be used to relieve negative thoughts, posing a risk factor for the development of risky drinking patterns. This is in accordance with a body of work investigating rumination as a predisposing factor of alcohol use in patients with alcohol use disorder [69] and also in adult [70] and adolescent [71,72] regular drinkers. Furthermore, our results on rumination and positive affect might be cautiously interpreted in regard to the IST [72], representing the rather prominent “liking” mechanism of early alcohol use (experiencing positive affect) [73,74] and a risk potential already at early stages for a “wanting” mechanisms (relieving rumination), which might later generalize from momentary to more stable patterns, as more habitual alcohol use develops.

Our results regarding environmental exploratory movement patterns, such as RE and potential alcohol trigger points (TPs), suggest that, unlike individuals with alcohol use disorder or heavy drinkers [31,42], adolescents in the earlier stages of alcohol use may not yet have established stable environmental cues or locations that consistently trigger alcohol use, neither on weekdays nor on weekends (more details in Table S4 in Multimedia Appendix 1). Furthermore, places potentially associated with alcohol use were visited similarly during both weekdays and weekends (more details in Figure S3 in Multimedia Appendix 1). However, it is important to consider that we used a “standardized” categorization of trigger points across the entire sample. Future studies could benefit from assessing more personalized potential trigger points for adolescents by directly asking them where they typically consume alcohol and associating these locations with EMA-based information such as affect or rumination.

Nonetheless, our results suggest that these TPs might indeed be developing, as indicated by higher RE. RE emerged as a consistently robust predictor of alcohol use across various models (more details in Table S2 in Multimedia Appendix 1). It could be speculated that this association is driven by the fact that RE may lead to increased exposure to potential alcohol contexts, such as alcohol advertisements, bars, drinking opportunities, or social contexts. However, there may be other factors contributing to this relationship. For instance, higher RE could be linked to the motivation to drink, such as prosocial

drinking motives [67,75], or it could reflect a decrease in daily structure and reduced parental supervision [76,77] or a lack of scheduled activities [78-80], such as regular participation in sports clubs. In addition, it is possible that RE varies substantially within individuals from day to day due to planned behaviors related to alcohol consumption on drinking days compared to nondrinking days [81]. Exploring the relationship between RE and these factors influencing alcohol use warrants further investigation in future studies, potentially using event-related sampling schemes to align movement patterns with alcohol use patterns. In addition, in order to learn more about the development of alcohol use disorders, it would be beneficial to examine whether a similar association with alcohol use remains in adolescents with alcohol use disorders.

Finally, our results indicate that alcohol use also seems to be dependent on adolescents' age and day of the week. These findings corroborate previous work showing alcohol use increases with age [82] and that adolescents mainly use alcohol on weekends [83,84] when there are usually no obligations to be fulfilled, but there is instead leisure time to spend with peers. Importantly, our results showing adolescents drink more on weekend days might reflect a particular behavioral pattern of adolescence that can change over time (less drinking, but also during the week) or become more pronounced (development towards alcohol addiction).

Our results have implications for prevention approaches, which should sensitize adolescents to alcohol use and its consequences and risks at an early stage before alcohol use manifests in hazardous or risky patterns. At this stage, adolescents might be less aware of the risks concerning alcohol, such as a gradual beginning that might cascade into problematic alcohol use [85]. A relevant concept of prevention might be mindfulness, which has already been successfully used in alcohol addiction treatment [86,87]. However, our results on trait attention regulation, as a relevant part of mindfulness [88], do not show a relevant influence on daily alcohol use (more details in Table S5 in Multimedia Appendix 1). It can be assumed that both mindfulness and attention regulation might be situation-dependent (eg, someone is stressed or in a hurry vs relaxed) and thus may also have a situation-dependent influence on alcohol use.

Limitations

Several limitations of this study should be noted. First, we assessed alcohol use only once a day in the morning, inquiring about alcohol use on the previous day. Consequently, we cannot infer the directionality underlying our findings nor establish any causal relationship. In addition, we are unable to link specific alcohol use episodes to momentary ratings or motives (EMA), social contexts, event appraisal, fluctuations in self-esteem, environmental locations (geospatial information), or the amount of alcohol consumed, which would provide additional crucial information for understanding the occurrence of alcohol-related harm [89]. Future studies could benefit from using participant-initiated assessments (event-related sampling), such as when drinking episodes commence or when participants report a specific context or emotional state [31]. This would provide insights into how the appraisal of occurring events and daily fluctuations in self-esteem contribute to and interact with emotional states and environmental exploratory movement behaviors during certain alcohol consumption occasions. This, in turn, would offer a comprehensive understanding of additional factors that contribute to the initiation of alcohol use in adolescents on a daily basis and help to investigate potential comorbidities with alcohol use behaviors, such as depression, where self-esteem is discussed as a potential critical factor as well [54]. This could also help in further examining alcohol-related coping strategies and their interaction with individual factors.

Second, our study partially coincided with the COVID-19 pandemic in Germany [63], so the generalizability of our results is limited due to COVID-19 restrictions regarding environmental and social factors [90]. Nevertheless, we identify factors influencing alcohol use even under these extremely restricted circumstances. Thus, our results might be underestimated compared with times when access to alcohol and peer communities is easier.

Furthermore, we lacked information on the socioeconomic status of participants and their families. We attempted to estimate this based on the type of school attended by the adolescents (more details in Table S6 in Multimedia Appendix 1), but future studies may benefit from directly inquiring about socioeconomic status, such as family income, as this factor may also influence adolescents' drinking behavior [91]. In addition, geospatial information like GPS shows a certain inaccuracy, especially in a dense urban environment like Mannheim, which leads to ambiguity as to what a participant is currently seeing or interacting with. This is exacerbated by the reliance on publicly available crowdsourced data sources, which correctness and relevance are not guaranteed, especially in view of temporal gaps between the EMA study and the data processing or temporal development within the assessment phase.

Future Directions

Our results provide a first indication of how emotional and other alcohol-related factors might be related to adolescent activity profiles, that is, RE, to predict very early adolescent alcohol use. However, some of the questions raised remain unanswered. Additional EMA studies are warranted to validate the associations identified herein, including within other, more diverse (adolescent) populations, and to assess whether completing an EMA could potentially influence subsequent behavior, such as through postassessment questionnaires [92]. Once these findings are corroborated, low-threshold smartphone or app-based EMIs or ecological momentary prevention approaches could be derived from them. These interventions may assist adolescents in the early stages of alcohol use and potentially prevent the progression to risky or even abusive alcohol consumption during this vulnerable phase by bolstering resilience factors, such as psychoeducation or mindfulness [93].

Geospatial data might contribute to our understanding of developing risky alcohol use in addition to the aforementioned trigger points; in fact, internal contexts (ie, emotional trigger states) might be identified by linking emotional experience or changes in emotional states with geolocation (construction of emomaps [94]). This might be an important step for future

research and would enable a link between environmental causes and behavior, such as alcohol use [95]. Building such emotional maps could be achieved by combining GPS positioning with EMA assessment as in our study, thereby enabling the exploration of the interaction of different alcohol theories with the real-life experiences of participants.

Furthermore, we observed reduced effects in certain variables when applying a stricter threshold for EMA participation, possibly due to fewer observations (more details in Table S2 in Multimedia Appendix 1). Future approaches could potentially enhance participant compliance by offering incentives, a common challenge in EMA studies [48,96], such as reducing the number of prompts per day. Our model comparison indicates that, with a fixed time-based sampling method, an optimal balance for prompts per day exists, ensuring adequate momentary data collection for statistical analysis and comparison. Study designs could be further strengthened by using a longitudinal design and a more age-diverse sample so that the effects on alcohol development could be observed over the period of adolescence.

Conclusion

Through the use of an EMA and geospatial approach, we identified daily life positive affect, rumination, weekend days, and environmental exploratory movement behaviors, as indicated by roaming entropy, as prominent factors associated with very early adolescent drinking behavior during the initial stages of drinking pattern development. Specifically, exploring such movement patterns could serve as an important avenue for future research, offering insights into the development of trigger points from early alcohol use to potentially hazardous use or use disorders.

While environmental factors may not yet play a significant role in the early stages of alcohol use, they are recognized as critical mechanisms in later stages of alcohol use. This understanding can inform the development of future research and may be used in prevention approaches during the early stage of alcohol use.

Acknowledgments

This study is supported by the Federal Ministry of Education and Research Germany (BMBF 01GL1745B, subproject 2), the Deutsche Forschungsgemeinschaft (FL 156/44-1, NE 1383/15-1 and BA 2088/7-1, CoviDrug).

With the public-funded research project IMAC-Mind: Improving Mental Health and Reducing Addiction in Childhood and Adolescence through Mindfulness: Mechanisms, Prevention and Treatment (2017-2023), the Federal Ministry of Education and Research contributes to improving the prevention and treatment of children and adolescents with substance use disorders and associated mental disorders. The project coordination is realized by the German Center of Addiction Research in Childhood and Adolescence at the University Medical Center Hamburg–Eppendorf. The consortium comprises seven projects in Germany. The Principal investigators are Rainer Thomasius (Coordinator, University Medical Center Hamburg–Eppendorf), Tobias Banaschewski and Herta Flor (both Central Institute for Mental Health, Mannheim), Johannes Kornhuber (Friedrich-Alexander-Universität Erlangen–Nürnberg), Michael Klein (Catholic University of Applied Sciences, Cologne), Olaf Reis (University Medicine of Rostock), Tanja Legenbauer (Ruhr-University Bochum), and Antonia Zapf (University Medical Center Hamburg–Eppendorf). Further members of the consortium are Nicolas Arnaud, Austermann Maria, Christiane Baldus, Anne Daubmann, Léa Josette Laurenz, Sabrina Kunze, Sophie Luise Schiller, and Michael Supplieth (University Medical Center Hamburg–Eppendorf), Frauke Nees, Karl Gottfried, Stella Guldner, Sabina Millenet, and Maren Prignitz (Central Institute for Mental Health, Mannheim), Bernd Lenz, Peter Fasching, Matthias Beckmann, Hartmut Heinrich, Verena Nadine Buchholz, Anna Eichler, Lothar Häberle, Christiane Mühle, Patrick Stelzl, and Bernhard Volz (Friedrich-Alexander-University Erlangen–Nürnberg), Katharina Ise and Diana Moesgen (Catholic University of Applied Sciences, Cologne), Lucie Waedel (University Medicine of Rostock), and Martin Holtmann, Regina Herdering, Carina Maria Huhn, and Laura Mokros (Ruhr-University Bochum). For more information, please visit our home page [64].

This study also received funding from the European Union. Complementary funding was received by UK Research and Innovation (UKRI) under the UK government's Horizon Europe funding guarantee (10041392 and 10038599). Views and opinions expressed are, however, those of the authors only and do not necessarily reflect those of the European Union, the European Health and Digital Executive Agency (HADEA), or UKRI. Neither the European Union, HADEA, nor UKRI can be held responsible for them.

Data Availability

The anonymized datasets generated during and/or analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1

An Ecological Momentary Assessment Approach of Environmental Triggers in the Role of Daily Affect, Rumination, and Movement Patterns in Early Alcohol Use Among Healthy Adolescents: Exploratory Study - Supplement.
[DOCX File , 542 KB - mhealth_v12i1e53401_app1.docx]

References

1. World Health Organization. Adolescent Health. 2024. URL: https://www.who.int/health-topics/adolescent-health/#tab=tab_1 [accessed 2024-04-09]
2. Conrod PJ, Nikolaou K. Annual research review: on the developmental neuropsychology of substance use disorders. J Child Psychol Psychiatry 2016;57(3):371-394. [doi: 10.1111/jcpp.12516] [Medline: 26889898]
3. Carlo G, Padilla‐Walker L. Adolescents' prosocial behaviors through a multidimensional and multicultural Lens. Child Dev Perspectives 2020;14(4):265-272. [doi: 10.1111/cdep.12391]
4. Konrad K, König J. Biopsychologische Veränderungen. In: Lohaus A, editor. Entwicklungspsychologie des Jugendalters. Berlin, Heidelberg: Springer Berlin Heidelberg; 2018:1-22.
5. Jordan CJ, Andersen SL. Sensitive periods of substance abuse: early risk for the transition to dependence. Dev Cogn Neurosci 2017;25:29-44 [FREE Full text] [doi: 10.1016/j.dcn.2016.10.004] [Medline: 27840157]
6. Ernst M, Pine DS, Hardin M. Triadic model of the neurobiology of motivated behavior in adolescence. Psychol Med 2006;36(3):299-312 [FREE Full text] [doi: 10.1017/S0033291705005891] [Medline: 16472412]
7. Blakemore SJ, Robbins TW. Decision-making in the adolescent brain. Nat Neurosci 2012;15(9):1184-1191. [doi: 10.1038/nn.3177] [Medline: 22929913]
8. Shadur JM, Lejuez CW. Adolescent substance use and comorbid psychopathology: emotion regulation deficits as a transdiagnostic risk factor. Curr Addict Rep 2015;2(4):354-363 [FREE Full text] [doi: 10.1007/s40429-015-0070-y] [Medline: 26889402]
9. Johnston LD, O'Malley PM, Bachman JG. Monitoring the future: national results on adolescent drug use: overview of key findings. FOC 2003;1(2):213-234. [doi: 10.1176/foc.1.2.213]
10. Zeiher J, Lange C, Starker A, Lampert T, Kuntz B. Tabak- und Alkoholkonsum bei 11- bis 17-Jährigen in Deutschland - Querschnittergebnisse aus KiGGS Welle 2 und Trends. Journal of Health Monitoring 2018;3(2):23-44 [FREE Full text] [doi: 10.17886/RKI-GBE-2018-071] [Medline: 35586374]
11. Foster JH, Powell JE, Marshall EJ, Peters TJ. Quality of life in alcohol-dependent subjects--a review. Qual Life Res 1999;8(3):255-261. [doi: 10.1023/a:1008802711478] [Medline: 10472156]
12. Sæther SMM, Knapstad M, Askeland KG, Skogen JC. Alcohol consumption, life satisfaction and mental health among norwegian college and university students. Addict Behav Rep 2019;10:100216 [FREE Full text] [doi: 10.1016/j.abrep.2019.100216] [Medline: 31692685]
13. Skala K, Walter H. Adolescence and Alcohol: a review of the literature. Neuropsychiatr 2013;27(4):202-211. [doi: 10.1007/s40211-013-0066-6] [Medline: 23839238]
14. Infante MA, Eberson SC, Zhang Y, Brumback T, Brown SA, Colrain IM, et al. Adolescent binge drinking is associated with accelerated decline of gray matter volume. Cereb Cortex 2022;32(12):2611-2620 [FREE Full text] [doi: 10.1093/cercor/bhab368] [Medline: 34729592]
15. Tharp-Taylor S, Haviland A, D'Amico EJ. Victimization from mental and physical bullying and substance use in early adolescence. Addict Behav 2009 Jun;34(6-7):561-567 [FREE Full text] [doi: 10.1016/j.addbeh.2009.03.012] [Medline: 19398162]
16. McCambridge J, McAlaney J, Rowe R. Adult consequences of late adolescent alcohol consumption: a systematic review of cohort studies. PLoS Med 2011;8(2):e1000413 [FREE Full text] [doi: 10.1371/journal.pmed.1000413] [Medline: 21346802]
17. Squeglia LM, Tapert SF, Sullivan EV, Jacobus J, Meloy MJ, Rohlfing T, et al. Brain development in heavy-drinking adolescents. Am J Psychiatry 2015;172(6):531-542 [FREE Full text] [doi: 10.1176/appi.ajp.2015.14101249] [Medline: 25982660]
18. Kuntsche E, Knibbe R, Gmel G, Engels R. Why do young people drink? A review of drinking motives. Clin Psychol Rev 2005;25(7):841-861. [doi: 10.1016/j.cpr.2005.06.002] [Medline: 16095785]
19. Diego MA, Field TM, Sanders CE. Academic performance, popularity, and depression predict adolescent substance use. Adolescence 2003;38(149):35-42 [FREE Full text] [Medline: 12803452]
20. Nees F, Tzschoppe J, Patrick CJ, Vollstädt-Klein S, Steiner S, Poustka L, et al. Determinants of early alcohol use in healthy adolescents: the differential contribution of neuroimaging and psychological factors. Neuropsychopharmacology 2012;37(4):986-995 [FREE Full text] [doi: 10.1038/npp.2011.282] [Medline: 22113088]
21. Heinrich A, Schumann G, Flor H, Nees F. Identification of key items regarding personality, environment, and life events to assess risk and resilience factors for harmful alcohol drinking in adolescents. Alcohol Alcohol 2016;51(6):710-715. [doi: 10.1093/alcalc/agw012] [Medline: 27001996]
22. Whelan R, Watts R, Orr CA, Althoff RR, Artiges E, Banaschewski T, et al. Neuropsychosocial profiles of current and future adolescent alcohol misusers. Nature 2014;512(7513):185-189 [FREE Full text] [doi: 10.1038/nature13402] [Medline: 25043041]

23. Freichel R, Pfirrmann J, Cousjin J, de Jong P, Franken I, Banaschewski T, et al. Drinking motives, personality traits and life stressors-identifying pathways to harmful alcohol use in adolescence using a panel network approach. Addiction 2023;118(10):1908-1919. [doi: 10.1111/add.16231] [Medline: 37157052]
24. Robinson TE, Berridge KC. The neural basis of drug craving: an incentive-sensitization theory of addiction. Brain Res Brain Res Rev 1993;18(3):247-291 [FREE Full text] [doi: 10.1016/0165-0173(93)90013-p] [Medline: 8401595]
25. Robinson TE, Berridge KC. Review. The incentive sensitization theory of addiction: some current issues. Philos Trans R Soc Lond B Biol Sci 2008;363(1507):3137-3146 [FREE Full text] [doi: 10.1098/rstb.2008.0093] [Medline: 18640920]
26. Thomas SE, Drobes DJ, Deas D. Alcohol cue reactivity in alcohol-dependent adolescents. J Stud Alcohol 2005;66(3):354-360. [doi: 10.15288/jsa.2005.66.354] [Medline: 16047524]
27. Urošević S, Collins P, Muetzel R, Schissel A, Lim KO, Luciana M. Effects of reward sensitivity and regional brain volumes on substance use initiation in adolescence. Soc Cogn Affect Neurosci 2015;10(1):106-113 [FREE Full text] [doi: 10.1093/scan/nsu022] [Medline: 24526186]
28. Russell MA, Gajos JM. Annual research review: ecological momentary assessment studies in child psychology and psychiatry. J Child Psychol Psychiatry 2020;61(3):376-394 [FREE Full text] [doi: 10.1111/jcpp.13204] [Medline: 31997358]
29. Trull TJ, Ebner-Priemer U. Ambulatory assessment. Annu Rev Clin Psychol 2013;9:151-176 [FREE Full text] [doi: 10.1146/annurev-clinpsy-050212-185510] [Medline: 23157450]
30. Comulada WS, Swendeman D, Wu N. Cell phone-based ecological momentary assessment of substance use context for latino youth in outpatient treatment: who, what, when and where. Drug Alcohol Depend 2016;167:207-213 [FREE Full text] [doi: 10.1016/j.drugalcdep.2016.08.623] [Medline: 27590744]
31. Fischer AM, King AC, Cao D, Fridberg DJ. Drinking context, alcohol use, and subjective responses during binge drinking episodes measured by high-resolution ecological momentary assessment (HR-EMA). Psychol Addict Behav 2023;37(2):258-266 [FREE Full text] [doi: 10.1037/adb0000887] [Medline: 36326673]
32. Yang JJ, Lin HC, Ou TS, Tong Z, Li R, Piper ME, et al. The situational contexts and subjective effects of co-use of electronic cigarettes and alcohol among college students: an ecological momentary assessment (EMA) study. Drug Alcohol Dependence 2022;239:109594 [FREE Full text] [doi: 10.1016/j.drugalcdep.2022.109594] [Medline: 35988530]
33. Stevenson BL, Blevins CE, Marsh E, Feltus S, Stein M, Abrantes AM. An ecological momentary assessment of mood, coping and alcohol use among emerging adults in psychiatric treatment. Am J Drug Alcohol Abuse 2020;46(5):651-658. [doi: 10.1080/00952990.2020.1783672] [Medline: 32851900]
34. Kashdan TB, Ferssizidis P, Collins RL, Muraven M. Emotion differentiation as resilience against excessive alcohol use: an ecological momentary assessment in underage social drinkers. Psychol Sci 2010;21(9):1341-1347. [doi: 10.1177/0956797610379863] [Medline: 20696854]
35. Goncy EA, Mrug S. Where and when adolescents use tobacco, alcohol, and marijuana: comparisons by age, gender, and race. J Stud Alcohol Drugs 2013;74(2):288-300 [FREE Full text] [doi: 10.15288/jsad.2013.74.288] [Medline: 23384377]
36. Carter BL, Tiffany ST. Meta ‐ analysis of cue ‐ reactivity in addiction research. Addiction 2002;94(3):327-340. [doi: 10.1046/j.1360-0443.1999.9433273.x]
37. Drummond DC. What does cue ‐ reactivity have to offer clinical research? Addiction 2002;95(8s2):129-144. [doi: 10.1046/j.1360-0443.95.8s2.2.x]
38. Schacht JP, Anton RF, Myrick H. Functional neuroimaging studies of alcohol cue reactivity: a quantitative meta-analysis and systematic review. Addict Biol 2013;18(1):121-133 [FREE Full text] [doi: 10.1111/j.1369-1600.2012.00464.x] [Medline: 22574861]
39. Heller AS, Shi TC, Ezie CEC, Reneau TR, Baez LM, Gibbons CJ, et al. Association between real-world experiential diversity and positive affect relates to hippocampal-striatal functional connectivity. Nat Neurosci 2020;23(7):800-804 [FREE Full text] [doi: 10.1038/s41593-020-0636-4] [Medline: 32424287]
40. Dulin PL, Gonzalez VM, Campbell K. Results of a pilot test of a self-administered smartphone-based treatment system for alcohol use disorders: usability and early outcomes. Subst Abus 2014;35(2):168-175 [FREE Full text] [doi: 10.1080/08897077.2013.821437] [Medline: 24821354]
41. Giroux D, Bacon S, King DK, Dulin P, Gonzalez V. Examining perceptions of a smartphone-based intervention system for alcohol use disorders. Telemed J E Health 2014;20(10):923-929 [FREE Full text] [doi: 10.1089/tmj.2013.0222] [Medline: 25243480]
42. McTavish FM, Chih MY, Shah D, Gustafson DH. How patients recovering from alcoholism use a smartphone intervention. J Dual Diagn 2012;8(4):294-304 [FREE Full text] [doi: 10.1080/15504263.2012.723312] [Medline: 23316127]
43. Gustafson DH, McTavish FM, Chih M, Atwood AK, Johnson RA, Boyle MG, et al. A smartphone application to support recovery from alcoholism: a randomized clinical trial. JAMA Psychiatry 2014;71(5):566-572 [FREE Full text] [doi: 10.1001/jamapsychiatry.2013.4642] [Medline: 24671165]
44. Mennis J, Mason M, Light J, Rusby J, Westling E, Way T, et al. Does substance use moderate the association of neighborhood disadvantage with perceived stress and safety in the activity spaces of urban youth? Drug Alcohol Depend 2016;165:288-292. [doi: 10.1016/j.drugalcdep.2016.06.019] [Medline: 27372218]

45. Byrnes HF, Miller BA, Morrison CN, Wiebe DJ, Woychik M, Wiehe SE. Association of environmental indicators with teen alcohol use and problem behavior: teens' observations vs. objectively-measured indicators. Health Place 2017;43:151-157 [FREE Full text] [doi: 10.1016/j.healthplace.2016.12.004] [Medline: 28061392]
46. Prignitz M, Guldner S, Nees F. Jugendliches Alkoholkonsumverhalten während der COVID-19-Pandemie und die Bedeutung von Achtsamkeit. SUCHT 2021;67(6):287-297. [doi: 10.1024/0939-5911/a000734]
47. Movisens GmbH. MovisensXS. 2023. URL: https://www.movisens.com/en/products/movisensxs/ [accessed 2023-08-09]
48. Wen CKF, Schneider S, Stone AA, Spruijt-Metz D. Compliance with mobile ecological momentary assessment protocols in children and adolescents: a systematic review and meta-analysis. J Med Internet Res 2017;19(4):e132 [FREE Full text] [doi: 10.2196/jmir.6641] [Medline: 28446418]
49. Leonhardt A, Könen T, Dirk J, Schmiedek F. How differentiated do children experience affect? An investigation of the within- and between-person structure of children's affect. Psychol Assess 2016;28(5):575-585. [doi: 10.1037/pas0000195] [Medline: 26280488]
50. Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: the PANAS scales. J Pers Soc Psychol 1988;54(6):1063-1070. [doi: 10.1037//0022-3514.54.6.1063] [Medline: 3397865]
51. Raabe A, Grüsser SM, Wessa M, Podschus J, Flor H. The assessment of craving: psychometric properties, factor structure and a revised version of the alcohol craving questionnaire (ACQ). Addiction 2005;100(2):227-234. [doi: 10.1111/j.1360-0443.2005.00960.x] [Medline: 15679752]
52. Rosenkranz T, Takano K, Watkins ER, Ehring T. Assessing repetitive negative thinking in daily life: development of an ecological momentary assessment paradigm. PLoS One 2020;15(4):e0231783 [FREE Full text] [doi: 10.1371/journal.pone.0231783] [Medline: 32310979]
53. Grosse Rueschkamp JM, Kuppens P, Riediger M, Blanke ES, Brose A. Higher well-being is related to reduced affective reactivity to positive events in daily life. Emotion 2020;20(3):376-390. [doi: 10.1037/emo0000557] [Medline: 30550304]
54. Crowe E, Daly M, Delaney L, Carroll S, Malone KM. The intra-day dynamics of affect, self-esteem, tiredness, and suicidality in major depression. Psychiatry Res 2019;279:98-108. [doi: 10.1016/j.psychres.2018.02.032] [Medline: 29661498]
55. Trull TJ, Ebner-Priemer UW. Ambulatory assessment in psychopathology research: a review of recommended reporting guidelines and current practices. J Abnorm Psychol 2020;129(1):56-63. [doi: 10.1037/abn0000473] [Medline: 31868388]
56. Liang W, Chikritzhs T. Weekly and daily cycle of alcohol use among the U.S. general population. Injury 2015;46(5):898-901. [doi: 10.1016/j.injury.2015.01.029] [Medline: 25661106]
57. Babor TE, Grant M. From clinical research to secondary prevention: international collaboration in the development of the alcohol disorders identification test (AUDIT). Alcohol Health and Research World 1989;13:4 [FREE Full text]
58. R. The R Foundation. 2023. URL: https://www.r-project.org/ [accessed 2023-10-02]
59. Freund J, Brandmaier AM, Lewejohann L, Kirste I, Kritzler M, Krüger A, et al. Emergence of individuality in genetically identical mice. Science 2013;340(6133):756-759. [doi: 10.1126/science.1235294] [Medline: 23661762]
60. OpenStreetMap Stiftung. OpenStreetMap. 2023. URL: https://www.openstreetmap.de/ [accessed 2023-09-01]
61. Stadt Mannheim. Kulturatlas Mannheim: Fachbereich Geoinformation und Stadtplanung. 2023. URL: https://www.gis-mannheim.de/mannheim/index.php?service=cultural_mapping [accessed 2023-09-07]
62. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using. J. Stat. Soft 2015;67(1):1-48. [doi: 10.18637/jss.v067.i01]
63. Schilling J, Buda S, Fischer M, Goerlitz L, Grote U, Haas W. Retrospektive Phaseneinteilung der COVID-19-Pandemie in Deutschland bis Februar 2021. Lexington, MA: Robert Koch-Institut; 2021.
64. Thomasius R. IMAC-Mind. 2023. URL: https://www.imac-mind.de/ [accessed 2023-10-02]
65. Duif M, Thewissen V, Wouters S, Lechner L, Jacobs N. Associations between affect and alcohol consumption in adults: an ecological momentary assessment study. Am J Drug Alcohol Abuse 2020;46(1):88-97. [doi: 10.1080/00952990.2019.1635606] [Medline: 31430201]
66. Simons RM, Hahn AM, Simons JS, Murase H. Emotion dysregulation and peer drinking norms uniquely predict alcohol-related problems via motives. Drug Alcohol Depend 2017;177:54-58. [doi: 10.1016/j.drugalcdep.2017.03.019] [Medline: 28558272]
67. Sjödin L, Larm P, Karlsson P, Livingston M, Raninen J. Drinking motives and their associations with alcohol use among adolescents in Sweden. Nordisk Alkohol Nark 2021;38(3):256-269 [FREE Full text] [doi: 10.1177/1455072520985974] [Medline: 35310616]
68. Cleveland MJ, Feinberg ME, Jones DE. Predicting alcohol use across adolescence: relative strength of individual, family, peer, and contextual risk and protective factors. Psychol Addict Behav 2012;26(4):703-713 [FREE Full text] [doi: 10.1037/a0027583] [Medline: 22390336]
69. Caselli G, Ferretti C, Leoni M, Rebecchi D, Rovetto F, Spada MM. Rumination as a predictor of drinking behaviour in alcohol abusers: a prospective study. Addiction 2010;105(6):1041-1048. [doi: 10.1111/j.1360-0443.2010.02912.x] [Medline: 20331550]
70. Devynck F, Rousseau A, Romo L. Does repetitive negative thinking influence alcohol use? a systematic review of the literature. Front Psychol 2019;10:1482 [FREE Full text] [doi: 10.3389/fpsyg.2019.01482] [Medline: 31333536]

71. Willem L, Bijttebier P, Claes L, Raes F. Rumination subtypes in relation to problematic substance use in adolescence. Personality and Individual Differences 2011;50(5):695-699. [doi: 10.1016/j.paid.2010.12.020]
72. Nolen-Hoeksema S, Stice E, Wade E, Bohon C. Reciprocal relations between rumination and bulimic, substance abuse, and depressive symptoms in female adolescents. J Abnorm Psychol 2007;116(1):198-207. [doi: 10.1037/0021-843X.116.1.198] [Medline: 17324030]
73. Arulkadacham LJ, Richardson B, Staiger PK, Kambouropoulos N, O'Donnell RL, Ling M. Dissociation between wanting and liking for alcohol and caffeine: a test of the incentive sensitisation theory. J Psychopharmacol 2017;31(7):927-933. [doi: 10.1177/0269881117711711] [Medline: 28627332]
74. Everitt BJ, Robbins TW. Neural systems of reinforcement for drug addiction: from actions to habits to compulsion. Nat Neurosci 2005;8(11):1481-1489. [doi: 10.1038/nn1579] [Medline: 16251991]
75. Schelleman-Offermans K, Kuntsche E, Knibbe RA. Associations between drinking motives and changes in adolescents' alcohol consumption: a full cross-lagged panel study. Addiction 2011;106(7):1270-1278. [doi: 10.1111/j.1360-0443.2011.03423.x] [Medline: 21375645]
76. Barnes GM, Reifman AS, Farrell MP, Dintcheff BA. The effects of parenting on the development of adolescent alcohol misuse: a six ‐ wave latent growth model. J of Marriage and Family 2004;62(1):175-186. [doi: 10.1111/j.1741-3737.2000.00175.x]
77. DiClemente RJ, Wingood GM, Crosby R, Sionean C, Cobb BK, Harrington K, et al. Parental monitoring: association with adolescents' risk behaviors. Pediatrics 2001;107(6):1363-1368. [doi: 10.1542/peds.107.6.1363] [Medline: 11389258]
78. Albertos A, Koning I, Benítez E, De Irala J. Adolescents' alcohol use: does the type of leisure activity matter? A cross-national study. Int J Environ Res Public Health 2021;18(21):11477 [FREE Full text] [doi: 10.3390/ijerph182111477] [Medline: 34769992]
79. Goldstein AL, Wall AM, Wekerle C, Krank M. The impact of perceived reinforcement from alcohol and involvement in leisure activities on adolescent alcohol use. Journal of Child & Adolescent Substance Abuse 2013;22(4):340-363. [doi: 10.1080/1067828x.2012.735190]
80. Prieto-Damm B, de la Rosa PA, Lopez-Del Burgo C, Calatrava M, Osorio A, Albertos A, et al. Leisure activities and alcohol consumption among adolescents from Peru and El Salvador. Drug Alcohol Depend 2019;199:27-34. [doi: 10.1016/j.drugalcdep.2019.03.002] [Medline: 30981046]
81. Kyrrestad H, Mabille G, Adolfsen F, Koposov R, Martinussen M. Gender differences in alcohol onset and drinking frequency in adolescents: an application of the theory of planned behavior. Drugs: Education, Prevention and Policy 2020;29(1):21-31. [doi: 10.1080/09687637.2020.1865271]
82. Brown SA, McGue M, Maggs J, Schulenberg J, Hingson R, Swartzwelder S, et al. Underage alcohol use: summary of developmental processes and mechanisms: ages 16-20. Alcohol Res Health 2009;32(1):41-52 [FREE Full text] [Medline: 23104446]
83. Finlay AK, Ram N, Maggs JL, Caldwell LL. Leisure activities, the social weekend, and alcohol use: evidence from a daily study of first-year college students. J Stud Alcohol Drugs 2012;73(2):250-259 [FREE Full text] [doi: 10.15288/jsad.2012.73.250] [Medline: 22333332]
84. Kuntsche E, Cooper ML. Drinking to have fun and to get drunk: motives as predictors of weekend drinking over and above usual drinking habits. Drug Alcohol Depend 2010;110(3):259-262. [doi: 10.1016/j.drugalcdep.2010.02.021] [Medline: 20363080]
85. Wittchen HU, Behrendt S, Höfler M, Perkonigg A, Lieb R, Bühringer G, et al. What are the high risk periods for incident substance use and transitions to abuse and dependence? Implications for early intervention and prevention. Int J Methods Psychiatr Res 2008;17 Suppl 1(Suppl 1):S16-S29 [FREE Full text] [doi: 10.1002/mpr.254] [Medline: 18543359]
86. Garland EL, Froeliger B, Howard MO. Mindfulness training targets neurocognitive mechanisms of addiction at the attention-appraisal-emotion interface. Front Psychiatry 2014;4:173 [FREE Full text] [doi: 10.3389/fpsyt.2013.00173] [Medline: 24454293]
87. Garland EL, Howard MO. Mindfulness-based treatment of addiction: current state of the field and envisioning the next wave of research. Addict Sci Clin Pract 2018;13(1):14 [FREE Full text] [doi: 10.1186/s13722-018-0115-3] [Medline: 29669599]
88. Mehling WE, Price C, Daubenmier JJ, Acree M, Bartmess E, Stewart A. The multidimensional assessment of interoceptive awareness (MAIA). PLoS One 2012;7(11):e48230 [FREE Full text] [doi: 10.1371/journal.pone.0048230] [Medline: 23133619]
89. Antai D, Lopez GB, Antai J, Anthony DS. Alcohol drinking patterns and differences in alcohol-related harm: a population-based study of the United States. Biomed Res Int 2014;2014:853410 [FREE Full text] [doi: 10.1155/2014/853410] [Medline: 25057502]
90. Bundesregierung Deutschland. Coronavirus in Deutschland. 2021. URL: https://www.bundesregierung.de/breg-de [accessed 2021-06-16]
91. Lemstra M, Bennett NR, Neudorf C, Kunst A, Nannapaneni U, Warren LM, et al. A meta-analysis of marijuana and alcohol use by socio-economic status in adolescents aged 10-15 years. Can J Public Health 2008;99(3):172-177 [FREE Full text] [doi: 10.1007/BF03405467] [Medline: 18615935]

92. Ebner-Priemer U. Ambulantes Psychophysiologisches Monitoring in der Psychiatrischen Forschung. Chancen und Möglichkeiten am Beispiel der Borderline-Persönlichkeitsstörung. Freiburg: Albert-Ludwigs-Universität; 2003.
93. Szeto EH, Schoenmakers TM, van de Mheen D, Snelleman M, Waters AJ. Associations between dispositional mindfulness, craving, and drinking in alcohol-dependent patients: an ecological momentary assessment study. Psychol Addict Behav 2019;33(5):431-441. [doi: 10.1037/adb0000473] [Medline: 31294578]
94. Streich B, Zeile P, Höffken S, Exner JP. Menschen als "Smarte Sensoren"? Neue Möglichkeiten für die Stadtplanung. In: Junkernheinrich M, Ziegler K, editors. Räume im Wandel: Empirie und Politik. Wiesbaden: Springer VS; 2013:51-72.
95. Schumann G, Andreassen OA, Banaschewski T, Calhoun VD, Clinton N, Desrivieres S, et al. Addressing global environmental challenges to mental health using population neuroscience: a review. JAMA Psychiatry 2023;80(10):1066-1074. [doi: 10.1001/jamapsychiatry.2023.2996] [Medline: 37610741]
96. Jones A, Remmerswaal D, Verveer I, Robinson E, Franken IHA, Wen CKF, et al. Compliance with ecological momentary assessment protocols in substance users: a meta-analysis. Addiction 2019;114(4):609-619 [FREE Full text] [doi: 10.1111/add.14503] [Medline: 30461120]

Abbreviations

AUDIT: Alcohol Use Disorder Identification Test
EMA: ecological momentary assessment
EMI: ecological momentary intervention
IST: Incentive Sensitization Theory
NA: negative affect
PA: positive affect
RE: roaming entropy
TP: trigger point

Edited by L Buis; submitted 05.10.23; peer-reviewed by M Amato, M Gansner, K Cummins, M DeJonckheere; comments to author 10.01.24; revised version received 07.05.24; accepted 08.08.24; published 10.12.24.

Please cite as:
Prignitz M, Guldner S, Lehmler SJ, Aggensteiner PM, Nees F, IMAC-Mind Consortium
An Ecological Momentary Assessment Approach of Environmental Triggers in the Role of Daily Affect, Rumination, and Movement Patterns in Early Alcohol Use Among Healthy Adolescents: Exploratory Study
JMIR Mhealth Uhealth 2024;12:e53401
URL: https://mhealth.jmir.org/2024/1/e53401
doi:10.2196/53401
PMID:

Original Paper

Continuous Monitoring of Heart Rate Variability and Respiration for the Remote Diagnosis of Chronic Obstructive Pulmonary Disease: Prospective Observational Study

Xiaolan Chen[1,2*], MMS; Han Zhang[2*], PhD; Zhiwen Li[3*], PhD; Shuang Liu[4], MMS; Yuqi Zhou[4], PhD

[1]Guangdong Basic Research Center of Excellence for Structure and Fundamental Interactions of Matter, Guangdong Provincial Key Laboratory of Quantum Engineering and Quantum Materials, School of Physics, South China Normal University, Guangzhou, China

[2]Guangdong Provincial Engineering Technology Research Center of Cardiovascular Individual Medicine and Big Data, School of Electronic and Information Engineering, South China Normal University, Foshan, China

[3]Key Laboratory of Reproductive Health National Health Commission of the People's Republic of China, Institute of Reproductive and Child Health, Peking University, Beijing, China

[4]Department of Pulmonary and Critical Care Medicine, Guangdong Provincial Technology Research Center of Chronic Obstructive Pulmonary Disease Rehabilitation, Third Affiliated Hospital of Sun Yat-sen University, Guangzhou, China

[*]these authors contributed equally

Corresponding Author:
Yuqi Zhou, PhD
Department of Pulmonary and Critical Care Medicine
Guangdong Provincial Technology Research Center of Chronic Obstructive Pulmonary Disease Rehabilitation
Third Affiliated Hospital of Sun Yat-sen University
No 600 Tianhe Road
Tianhe District
Guangzhou, 510630
China
Phone: 86 13533489943
Email: zhouyuqi@mail.sysu.edu.cn

Abstract

Background: Conventional daytime monitoring in a single day may be influenced by factors such as motion artifacts and emotions, and continuous monitoring of nighttime heart rate variability (HRV) and respiration to assist in chronic obstructive pulmonary disease (COPD) diagnosis has not been reported yet.

Objective: The aim of this study was to explore and compare the effects of continuously monitored HRV, heart rate (HR), and respiration during night sleep on the remote diagnosis of COPD.

Methods: We recruited patients with different severities of COPD and healthy controls between January 2021 and November 2022. Vital signs such as HRV, HR, and respiration were recorded using noncontact bed sensors from 10 PM to 8 AM of the following day, and the recordings of each patient lasted for at least 30 days. We obtained statistical means of HRV, HR, and respiration over time periods of 7, 14, and 30 days by continuous monitoring. Additionally, the effects that the statistical means of HRV, HR, and respiration had on COPD diagnosis were evaluated at different times of recordings.

Results: In this study, 146 individuals were enrolled: 37 patients with COPD in the case group and 109 participants in the control group. The median number of continuous night-sleep monitoring days per person was 56.5 (IQR 32.0-113.0) days. Using the features regarding the statistical means of HRV, HR, and respiration over 1, 7, 14, and 30 days, binary logistic regression classification of COPD yielded an accuracy, Youden index, and area under the receiver operating characteristic curve of 0.958, 0.904, and 0.989, respectively. The classification performance for COPD diagnosis was directionally proportional to the monitoring duration of vital signs at night. The importance of the features for diagnosis was determined by the statistical means of respiration, HRV, and HR, which followed the order of respiration > HRV > HR. Specifically, the statistical means of the duration of respiration rate faster than 21 times/min (RRF), high frequency band power of 0.15-0.40 Hz (HF), and respiration rate (RR) were identified as the top 3 most significant features for classification, corresponding to cutoff values of 0.1 minute, 1316.3 nU, and 16.3 times/min, respectively.

Conclusions: Continuous monitoring of nocturnal vital signs has significant potential for the remote diagnosis of COPD. As the duration of night-sleep monitoring increased from 1 to 30 days, the statistical means of HRV, HR, and respiration showed a better reflection of an individual's health condition compared to monitoring the vital signs in a single day or night, and better was the classification performance for COPD diagnosis. Further, the statistical means of RRF, HF, and RR are crucial features for diagnosing COPD, demonstrating the importance of monitoring HRV and respiration during night sleep.

(JMIR Mhealth Uhealth 2024;12:e56226) doi:10.2196/56226

KEYWORDS

continuous monitoring; chronic obstructive pulmonary disease; COPD diagnosis; prospective study; ROC curve; heart rate variability; respiratory rate; heart rate; noncontact bed sensors

Introduction

Chronic obstructive pulmonary disease (COPD) is a prevalent chronic respiratory disease characterized by persistent airflow restriction, with main symptoms including chronic cough, expectoration, and dyspnea that may not be apparent in the early stages of the disease [1]. In recent years, there has been an increasing trend in the prevalence rate of COPD, and large numbers of cases and deaths have been reported. A cross-sectional study conducted in 2007 among 20,245 adults from 7 provinces in China revealed a COPD prevalence rate of 8.2% among individuals older than 40 years [1]. In 2018, data from 2 large epidemiological surveys in China revealed a prevalence rate of ~13.6% among individuals older than 40 years [2,3]. According to the 2019 Global Burden of Disease Study, the global mortality due to COPD reached 3.3 million individuals, with China alone accounting for over 1 million deaths, representing approximately one-third of the total [4]. Due to dyspnea and other symptoms, the daily activities and exercise ability of patients with COPD are greatly limited, which not only seriously affects their quality of life but also imposes a substantial economic burden on the families and society [5,6].

According to the guidelines for the diagnosis and treatment of COPD in China, the current clinical diagnosis methods of COPD primarily include the pulmonary function test, chest imaging examination, pulse oxygen saturation monitoring, arterial blood gas analysis, electrocardiogram, and echocardiography. Among these methods, the pulmonary function test is the gold standard for diagnosing COPD by assessing the degree of airflow limitation and evaluating disease severity. If the ratio of the forced expiratory volume in the first one second to the forced vital capacity of the lungs is less than 70% after inhalation of the bronchodilator, one will be considered to have persistent airflow limitation and can be further diagnosed in conjunction with the clinical symptoms, signs, and history of exposure to risk factors [7].

Patients with COPD commonly exhibit autonomic nervous dysfunction, and the heart rate variability (HRV) serves as a valuable indicator for assessing autonomic nervous function by reflecting the tension and balance between the sympathetic and parasympathetic nerves [8]. Previous studies [9,10] have compared the resting heart rate (HR) and respiratory rate (RR) monitored during daytime between patients with COPD and healthy individuals, and the results suggested that HR and RR can be used as critical indicators for the auxiliary diagnosis of COPD.

In recent years, the rapid development of sensing technology has facilitated the employment of wearable devices for vital signs monitoring [11]. Pioneer studies have demonstrated that the vital signs (including HRV, HR, and respiration) monitored using wearable sensors are identical to those recorded by medical devices [12]. Thus, wearable sensors–aided continuous monitoring is applicable to patients with chronic disease at home [13,14]. Long-term continuous monitoring with wearable sensors during night sleep can effectively record the vital signs of patients [15,16]. Specifically, Bellos et al [17] utilized wearable devices and external devices from the CHRONIOUS platform system to record vital signs such as blood pressure and blood glucose levels and other short-term information of 30 patients with COPD. A mixed classifier was utilized after performing data fusion to classify the severity of COPD among patients, and it yielded an accuracy of 94%. The principal component analysis–based classification using the daytime-recorded vital signs, including HR, respiration, and blood pressure from 8 healthy individuals and 47 patients with COPD, was examined to assess the severity of COPD, and this yielded an accuracy of 88% [18]. In addition to using HR and respiration, Rahman et al [19] considered 30-minute recorded daytime HRV as a feature by using smartwatches to aid in the diagnosis of COPD, yielding a classification accuracy around 80%. By analogy, with daytime-recorded HRV and HR by using wearable devices, Tiwari et al [20] attempted to predict the acute exacerbation of COPD and verified the feasibility of remote sensing–based early warning. The aforementioned studies were based on wearable device–based vital signs recording during daytime in a single day. However, wearable devices have limitations for recording vital signs during night sleep. In addition, studies on diagnosis of COPD based on long-term monitored vital signs (over weeks or even months) have not been reported yet.

Based on the above concerns, we propose to explore the potential of long-term monitored vital signs (including HR, HRV, and respiration) during night sleep to assist in the remote diagnosis of COPD. To the best of our knowledge, this is the first prospective study based on the continuous monitoring of nocturnal vital signs for the remote diagnosis of COPD. Specifically, the HR, HRV, and respiration in patients with COPD and healthy individuals were recorded with bed sensors from 10 PM to 8 AM on the forthcoming day, and the continuous monitoring of each patient lasted no less than 30 days. With the continuously monitored vital signs, the statistical means of HR, HRV, and respiration over the time periods of 7, 14, and 30 days were extracted as features for the diagnosis of COPD.

Methods

Study Sites and Participants

A prospective observational study was conducted to recruit case and control individuals from January 2021 to November 2022. The case group comprised of patients with confirmed COPD, while the control group consisted of individuals without any definite diagnosis, and both these groups were recruited from the Guangdong Provincial Technology Research Center of COPD Rehabilitation, the Third Affiliated Hospital of Sun Yat-sen University. Specifically, the case group was derived from the outpatient, emergency, and inpatient departments of this hospital, and the control group consisted of individuals recruited by the Guangdong Provincial Technology Research Center of COPD Rehabilitation. Pulmonary function examination was performed either at or prior to the enrollment of the case group. According to the 2022 Global Initiative for Chronic Obstructive Lung Disease (GOLD) diagnostic criteria, if the ratio of the forced expiratory volume in the first one second to the forced vital capacity of the lungs is less than 70% after inhalation of the bronchodilator, one will be considered to have persistent airflow limitation [21]. This is the gold standard for COPD diagnosis and the only inclusion criterion for case groups. The exclusion criteria for the case and control groups were as follows:

1. Participants younger than 45 years
2. Individuals exhibiting poor compliance, dropouts, or loss to follow-up, including those in coma, in shock, with multiple organ failure, or with a history of epilepsy or any other condition that hinders continued participation in the monitoring study; also excluded were individuals who did not utilize monitoring equipment and those with missing lung function indicators
3. Patients with a history of severe lung diseases such as lung cancer, tuberculosis, severe pulmonary heart disease, and any other malignant tumor disease; individuals with severe cardiovascular diseases, including myocardial infarction, heart failure, arrhythmia, hypertension, and coronary heart disease, and patients with diabetes
4. Individuals with irregular schedules.

Ethics Approval

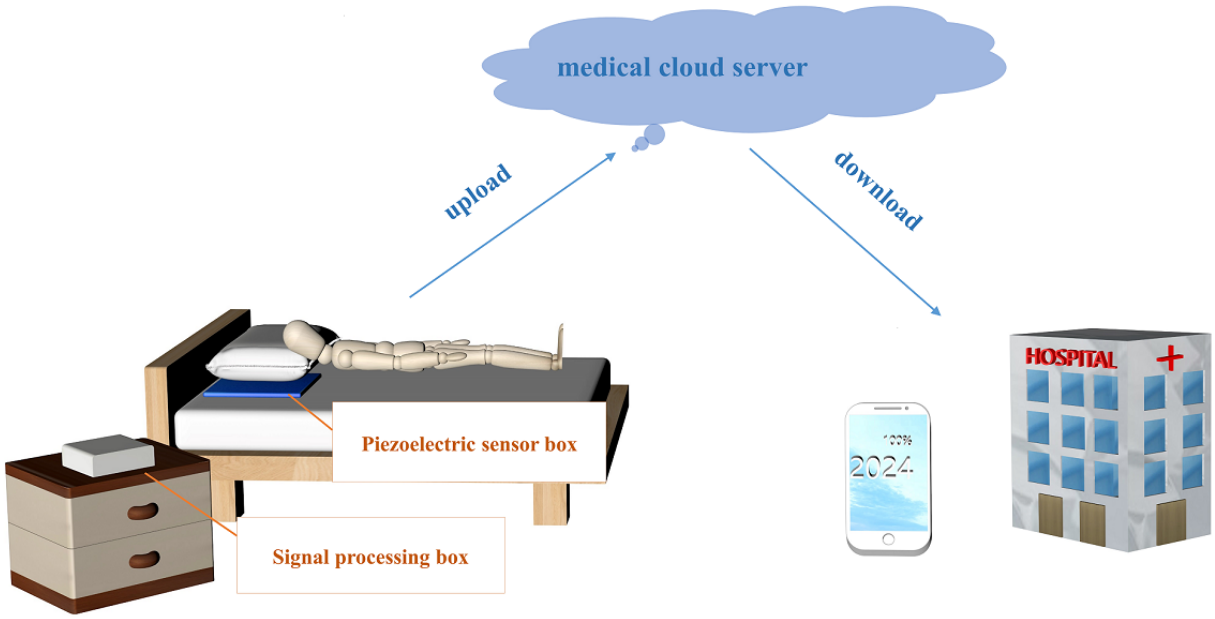
This study obtained the written informed consent from all participants, and the study protocol was reviewed by the medical ethics committee of the Third Affiliated Hospital of Sun Yat-sen University (approval [2021]02-019-01).

Data Collection and Processing

We obtained the pertinent data of individuals in the case group through the hospital information management system, including demographic characteristics, disease history, pulmonary function test data, and acute exacerbations. We also recorded the demographic characteristics and the disease history of participants in the control group. Piezoelectric sensor–based HR and respiratory recorders (model: WSM-LN-03, medical device; China Food and Drug Administration; 20182071130) were distributed to the participants and placed under the pillow to continuously record the HR, respiration, and HRV during night sleep from 10 PM to 8 AM the next day. Prior to the examination, all participants passed the standardized training on how to use the vital sign monitoring device. The recorded vital signs were manually uploaded to the medical cloud server for further data analysis. The architecture of the proposed study is shown in Figure 1. Due to the long duration of the experiment, there may be cases of inevitable data loss. Therefore, specific exclusion criteria with regarding to the data on the medical cloud server were as follows:

1. Number of artifact motions during night sleep ≥600
2. Sleep duration less than 5 hours or longer than 9 hours
3. Data loss in 1 night sleep ≥1 hour
4. Outliers such as HR or RR is 0
5. Data obtained 7 days before and during cases of acute exacerbation of COPD

Figure 1. Architecture of the proposed study.



Principle of Device and Feature Description

The distributed vital signs monitoring device was deployed under the pillow. With piezoelectric sensors, the analogy signals representing cardiac and respiratory activities are recorded with a sampling frequency of 1 kHz. Specifically designed filter banks can detect the heartbeat interval and respiration per minute during night sleep, and the averaged HR (beat/min) and respiration (times/min) of a whole night can be obtained [22-24]. In addition, by taking advantage of the heartbeat interval, both time and frequency domain HRV parameters can be calculated, as illustrated in Table 1 [25]. Specifically, HRV contains 7 times/frequency domain features, including standard deviation of heartbeat interval (SDNN), total power of heartbeat interval (TP), low frequency of TP (LF), high frequency of TP (HF), low frequency/high frequency (LF/HF), very low frequency of TP (VLF), and ultralow frequency of TP (ULF). The features of respiration also contain the cumulative duration of RR faster than 21 times/min (RRF).

Table 1. Feature descriptions of heart rate variability, heart rate, and respiration.

Category, abbreviation of feature	Definition
Heart rate variability [25]	
SDNN	Standard deviation of heartbeat interval
TP	Total spectral power of heartbeat interval
LF	Low frequency band power of 0.04-0.15 Hz
HF	High frequency band power of 0.15-0.40 Hz
ULF	Ultralow frequency band power of <0.0033 Hz
VLF	Very low frequency band power of 0.0033-0.04 Hz
LF/HF	Ratio of low frequency band power over high frequency band power
Heart rate	
HR	Heart rate mode during night sleep (beats/min)
Respiration	
RR	Respiratory rate mode during night sleep (times/min)
RRF (min)	Cumulative duration of respiratory rate faster than 21 times/min

Since the bed sensors continuously monitor during night, HRV, HR, and respiration can be continuously monitored over weeks without affecting the lifestyle of the recruited patients. Considering that the features of HRV, HR, and respiration obtained in a single night sleep monitoring could be influenced by the daytime emotions and health status of the patients, we propose to use the statistical average of the continuously monitored features (HRV, HR, and respiration) over weeks instead of that recorded in a single day for further analysis. Numerically, the statistical average of the features HRV, HR, and respiration over t consecutive days are calculated by the following method. We first calculated the sum of the values of HRV, HR, and respiration for t consecutive days; then subtracted the maximum and minimum values in the values of HRV, HR, and respiration for t consecutive days; and then obtained the statistical means by dividing by t–2. In this study, we set t=1, 7, 14, and 30 days as the cutoff observation time for analysis.

Sample Size Estimation

The purpose of this study was to evaluate the diagnostic value of continuous HRV and respiration monitoring in patients with COPD and control groups by using pulmonary function examination as the diagnostic gold standard. We expected continuous HRV and respiration monitoring to diagnose COPD with sensitivity of 90%, tolerance error of 10% of sensitivity, specificity of 90%, tolerance error of 10% of specificity, and the test criteria α=.01. The sample size for patients with COPD and control groups was set at 1:2, and the dropout rate was set as 10%. Finally, we used PASS 11.0 software (NCSS Statistical Software) to calculate the sample size, and we found that there should be at least 37 patients with COPD and 74 controls, for a total of 111 participants.

Statistical Analysis

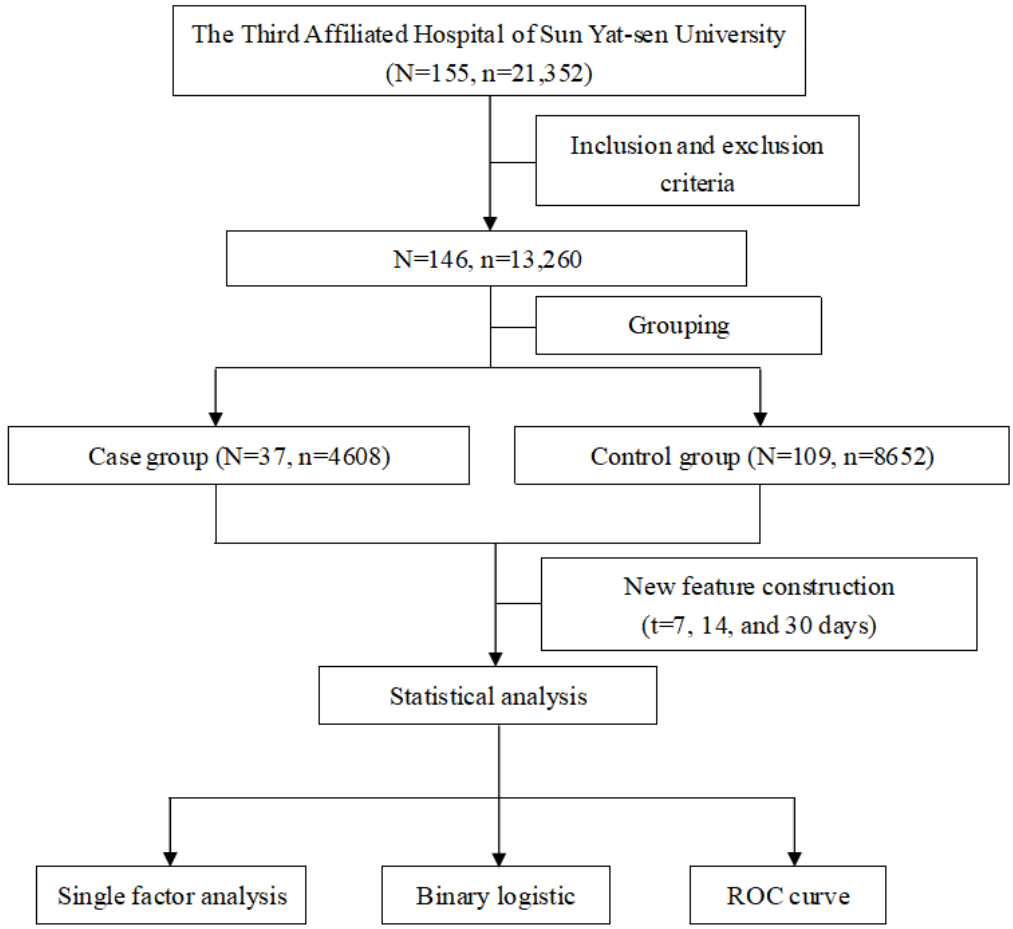
Statistical analysis was conducted using SPSS software (version 20.0; IBM Corp). The main statistical methods included descriptive analysis, χ^2 test, nonparametric test, binary logistic regression, and receiver operating characteristic (ROC) curves. For statistical description, counting data were described by frequency and rate, and metrological data were described by median and quartile. For statistical tests, the χ^2 test was employed to compare the 2 groups in terms of the counting data, with the χ^2 value serving as the test statistic. The metrological data were analyzed using a nonparametric test, specifically the Mann-Whitney U test. The z score was employed as the test statistic. In the multifactor analysis, the classical binary logistic regression was applied for classification, and the ROC curve was used to evaluate the diagnosis of each variable. The determination of the cutoff values relied on the maximum Youden index as the tangent point. The test criterion was set as α=.05, and P<.05 was considered statistically significant. The classification cutoff point was set as 0.5.

Data Cleaning and Analysis Flowchart

The experimental process of this study is shown in Figure 2. We recruited 155 participants corresponding to a total number

of 21,352 night-sleep monitored samples (person-time) of patients with COPD and healthy participants in the Third Affiliated Hospital of Sun Yat-sen University, in accordance with the abovementioned inclusion criteria for the population. After data cleaning based on the abovementioned exclusion criteria, 146 patients corresponding to a final number of 13,260 samples (person-time) were included, corresponding to case and control groups of 4608 and 8652 samples, respectively. Subsequently, the statistically averaged features of long-term HRV, HR, and respiration over the observation time of 7, 14, and 30 days were constructed for further analysis.

Figure 2. Flowchart of data cleaning and analysis. N: number of participants; n: number of person-times; ROC: receiver operating characteristic; t: cutoff observation time.



Results

Baseline Characteristics

A total of 146 individuals who met the inclusion and exclusion criteria were enrolled in this study, comprising 37 cases and 109 controls. There were 12 mild to moderate cases and 25 severe cases. The overall compliance rate among the participants was 0.73 (IQR 0.50-0.86), with no statistically significant difference between the 2 groups (z score=–1.5; P=.15). Moreover, the average number of continuous night-sleep monitoring days per person was found to be 56.5 (IQR 32.0-113.0) days, showing no significant difference between the 2 groups again (z score=0.5; P=.65). Therefore, compliance and monitoring duration between the 2 groups were comparable. Significant differences were observed between the 2 groups in terms of age, gender, and BMI (P<.001). Specifically, the case group exhibited a higher proportion of older individuals, males, and those with low BMI in comparison to the control group (as depicted in Table 2).

Table 2. Overall baseline characteristics of the total population.

Variables	Total (N=146)	Control group (n=109)	Case group (n=37)	*z* score/χ^2 (*df*)	*P* value
Compliance rate, median (IQR)	0.73 (0.50-0.86)	0.76 (0.53-0.86)	0.65 (0.43-0.83)	–1.5	.15
Monitoring per person, median (IQR)	56.5 (32.0-113.0)	56.0 (34.0-107.0)	78.0 (25.0-164.0)	0.5	.65
Age (years), median (IQR)	59.0 (52.0-68.0)	55.0 (51.0-63.0)	73.0 (66.0-77.0)	7.0	<.001
Sex, n (%)				21.8 (1)	<.001
Male	70 (47.9)	40 (36.7)	30 (81.1)		
Female	76 (52.1)	69 (63.3)	7 (18.9)		
BMI (kg/m^2), n (%)				18.9 (2)	<.001
<18.5	8 (5.5)	2 (1.8)	6 (16.2)		
18.5-24	95 (65)	67 (61.5)	28 (75.7)		
≥24	43 (29.5)	40 (36.7)	3 (8.1)		

Statistical Comparisons of the Features Between Case and Control Groups

This study consists of 13,260 samples (person-time), corresponding to 8652 in the control group and 4608 in the case group. During the monitoring period of 7, 14, and 30 days, there was a total of 12,396, 11,422, and 9404 person-times recorded, respectively. In 4 cutoffs of the observation time of 1, 7, 14, and 30 days, the distributions of HRV, HR, and respiration features, that is, SDNN, TP, LF, HF, VLF, ULF, HR, RR, and RRF, averaged over different timescales were of significant statistical difference, as shown in Tables S1-S4 of Multimedia Appendix 1. Specifically, the values of all the averaged features in the different timescales in the case group were significantly higher than those of the control group (P<.001). Interestingly, no significant difference was observed in the distribution of the features LF/HF between the 2 groups (P>.05). Consequently, when performing a binary logistic regression analysis, all features of LF/HF averaged over different timescales (t= 1, 7, 14, and 30 days), that is, LF/HF_1, LF/HF_7, LF/HF_{14}, and LF/HF_{30} were excluded, while the rest 36 features were included.

Classification Performance of Multidimensional Vital Sign Features

HRV, HR, and respiratory features at different timescales are represented by features and timescales, where timescales are in subscripts. The overall classification accuracy of HRV, HR, and respiration at 1, 7, 14, and 30 days achieved a high level of 0.958, with an impressive Youden index of 0.904 and an excellent area under the ROC curve (AUC) of 0.989 (as presented in Table 3 and Figure 3A).

We combined the vital signs features based on 4 different timescales (t= 1, 7, 14, and 30 days) and subsequently classified the participants into case and control groups. Figure 4 demonstrates that the classification performance of HRV, HR, and respiration improved as the monitoring duration increased, with HRV, HR, and respiration at 30 days exhibiting superior performance (accuracy 0.956, Youden index 0.899, AUC 0.989). Specifically, the ranking of the performance was presented as follows: HRV, HR, and respiration in 30 days > HRV, HR, and respiration in 14 days > HRV, HR, and respiration in 7 days > HRV, HR, and respiration in 1 day.

We combined them again based on the vital signs category and subsequently classified those participants into the case and control groups. Figure 4 demonstrates an improving trend in the classification efficacy of respiration, HRV, and HR, surpassing that of single-day monitoring. On the one hand, when evaluating the classification effect using accuracy and Youden index, we observed an improvement in the classification effect of HRV with the increasing duration of monitoring, where HRV_{30} demonstrated the highest performance (accuracy 0.866, Youden index 0.717). Similarly, as the monitoring duration increased, respiration also exhibited enhanced performance, with respiration at 30 days achieving the best results (accuracy 0.913, Youden index 0.792). However, there was minimal variation in the classification performance of HR across different timescales, maintaining an accuracy rate of approximately 0.67 and a Youden index of around 0.15. On the other hand, AUC was utilized to be evaluated. The findings demonstrated that all AUC values exceeded 0.5 and exhibited statistical significance. Notably, the classification performance of HRV improved with longer monitoring duration, specifically presented as HRV_{30} > HRV_{14} > HRV_7 > HRV_1. Similarly, the performance of respiration gradually tended to increase with extended monitoring duration, indicated by respiration over 30 days/respiration over 14 days > respiration over 14 days/respiration over 7 days > respiration over 1 day. However, there were no statistically significant differences between respiration over 14 days and respiration over 7 days or between respiration over 14 days and respiration over 30 days. Furthermore, the performance of HR also progressively enhanced with prolonged monitoring time ($HR_{30}/HR_{14}/HR_7$ > HR_1). Nevertheless, there were no statistically significant differences among HR_7, HR_{14}, and HR_{30}, as shown in Table 3.

Table 3 and Figures 3B-3E demonstrate that the classification performance, as evaluated by the accuracy, Youden index, and AUC, consistently showed respiration > HRV > HR over a single timescale. This consistent pattern was observed over all 4 timescales. In addition, the sensitivity of HR remained consistently low over the 4 timescales, ranging from

approximately 0.2 to 0.3, indicating a high potential for missed diagnoses rate (1–sensitivity) of up to 0.7-0.8. Specifically, the sensitivity of HR_1 was 0.177, with a corresponding missed diagnosis rate of 0.823. These findings highlight that only using HR features for the diagnosis of patients with COPD may lead to a significant increase in missed diagnosis rates exceeding 80%.

Table 3. The model evaluation index for the diagnosis of chronic obstructive pulmonary disease by using a combination of multidimensional features.

Features (time=1, 7, 14, 30 days)	Accuracy	Sensitivity	Specificity	Youden index	Area under the curve (95% CI)	*P* value
(HRV[a], HR[b], respiration) in 1, 7, 14, and 30 days	0.958	0.924	0.980	0.904	0.989 (0.987-0.991)	<.001
(HRV, HR, respiration) in 1 day	0.915	0.813	0.970	0.783	0.959 (0.955-0.963)	<.001
(HRV, HR, respiration) in 7 days	0.941	0.876	0.976	0.852	0.976 (0.973-0.979)	<.001
(HRV, HR, respiration) in 14 days	0.946	0.892	0.978	0.870	0.982 (0.980-0.985)	<.001
(HRV, HR, respiration) in 30 days	0.956	0.919	0.980	0.899	0.989 (0.987-0.990)	<.001
HRV in 1 day	0.802	0.619	0.899	0.518	0.863 (0.856-0.869)	<.001
HRV in 7 days	0.841	0.732	0.900	0.632	0.907 (0.901-0.912)	<.001
HRV in 14 days	0.851	0.773	0.896	0.669	0.919 (0.914-0.924)	<.001
HRV in 30 days	0.866	0.823	0.894	0.717	0.932 (0.927-0.937)	<.001
HR in 1 day	0.670	0.177	0.932	0.109	0.646 (0.636-0.656)	<.001
HR in 7 days	0.674	0.227	0.920	0.147	0.669 (0.659-0.679)	<.001
HR in 14 days	0.672	0.266	0.905	0.171	0.677 (0.666-0.687)	<.001
HR in 30 days	0.648	0.325	0.858	0.183	0.688 (0.677-0.699)	<.001
Respiration in 1 day	0.889	0.721	0.978	0.699	0.914 (0.908-0.920)	<.001
Respiration in 7 days	0.902	0.767	0.975	0.742	0.930 (0.924-0.935)	<.001
Respiration in 14 days	0.907	0.789	0.975	0.764	0.937 (0.932-0.943)	<.001
Respiration in 30 days	0.913	0.819	0.973	0.792	0.944 (0.939-0.950)	<.001

[a]HRV: heart rate variability.

[b]HR: heart rate.

Figure 3. Receiver operating characteristic curve of multidimensional features based on 4 timescales (1, 7, 14, and 30 days; shown in subscripts in the graph). HR: heart rate; HRV: heart rate variability.

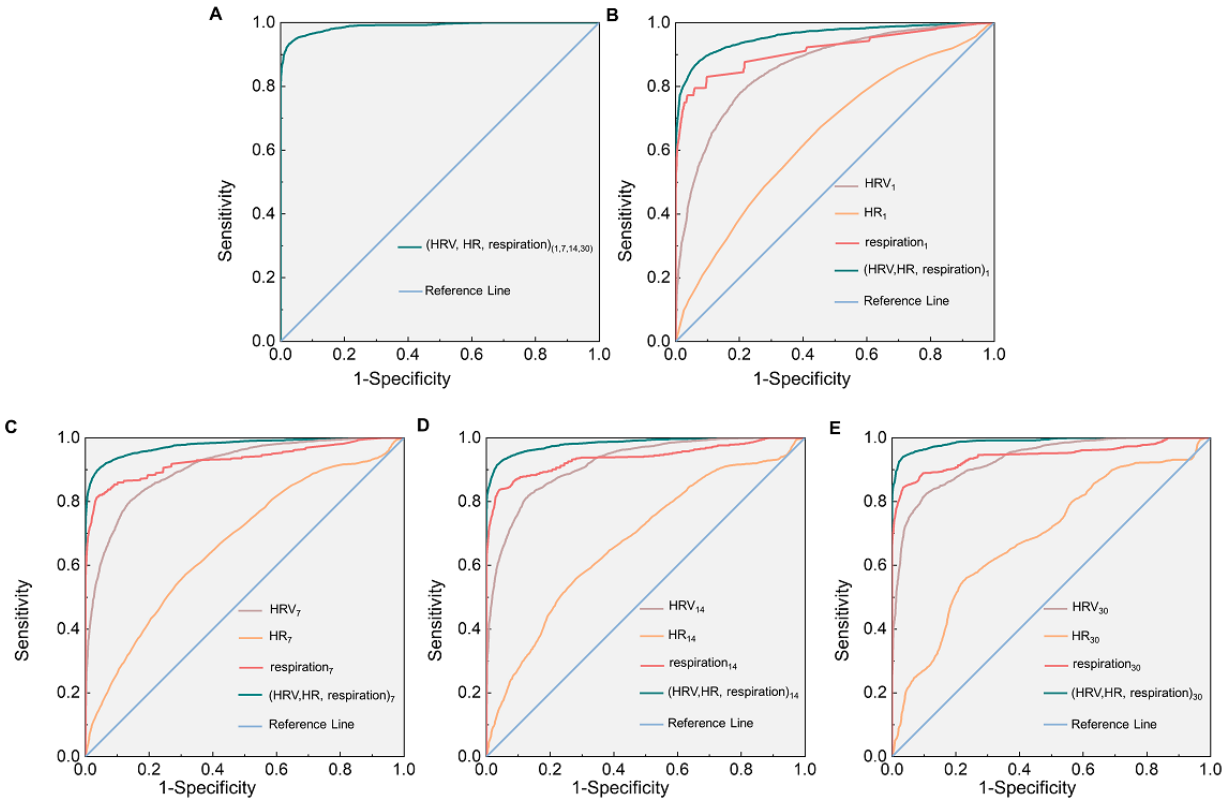
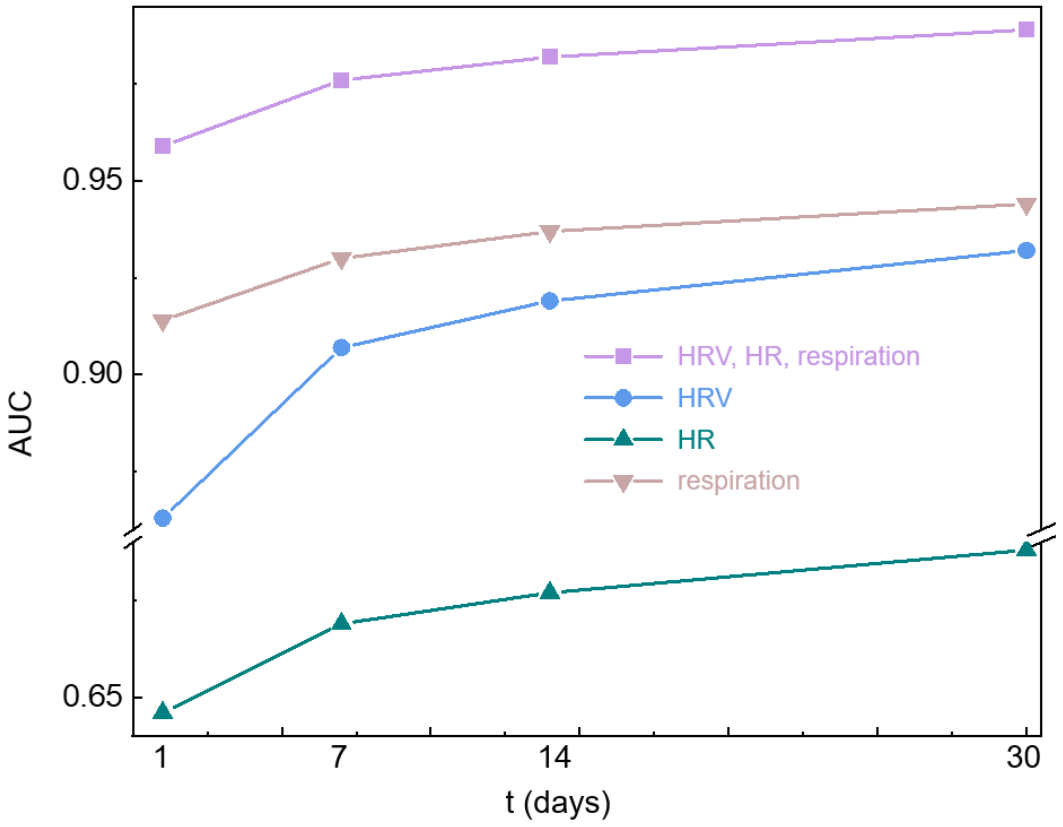


Figure 4. Relationship between area under the curve and monitoring duration of multidimensional statistical features. AUC: area under the curve; HR: heart rate; HRV: heart rate variability; t: time.



Classification Performance of Single Features for 7 Days Long-Term Monitoring

The results shown in Figure 4 indicate that the classification performance using the 7-day monitored features (HRV, HR, respiration; AUC=0.976) is already comparable to that of the 30-day monitored features (HRV, HR, respiration; AUC=0.989). Therefore, considering that the performance of monitoring longer than 7 days tends to saturate, while increasing the monitoring time leads to an increase in the dropout rate, we recommend continuous monitoring for 7 days during daily home life, which can already provide valuable assistance in the auxiliary diagnosis of COPD.

We classified the case and control groups by a single measured vital sign feature during the 7-day long-term monitoring period. As shown in Table 4, the results revealed that the HF_7 exhibited the most effective classification performance (accuracy 0.830, Youden index 0.602), followed by TP_7 (accuracy 0.792, Youden index 0.474) among the 6 features (HRV) evaluated by the accuracy and Youden index. Among the features set (HRV, HR, respiration)$_7$, RRF_7 demonstrated the best performance (accuracy 0.898, Youden index 0.733), followed by HF_7, while HR_7 showed the poorest performance (accuracy 0.674, Youden index 0.147). Furthermore, both HR_7 and ULF_7 exhibited very low sensitivity with the missed diagnosis rate exceeding 70%, indicating that using HR_7 or ULF_7 for classification and prediction could lead to higher chances of missed diagnosis.

Additionally, AUC was used to evaluate the effect of the features on the classification performance. As depicted in Figure 5, the AUCs of all 7-day time-scale monitored features were greater than 0.5 and demonstrated statistical significance (P<.001). The highest AUC value was observed for RRF_7 at 0.955. The order of the classification effects was as follows: $RRF_7 > HF_7 > RR_7 > LF_7/TP_7 > SDNN_7 > VLF_7 > ULF_7/HR_7$. Notably, there were no statistically significant differences between LF_7 and TP_7, ULF_7, and HR_7.

Table 4. Model evaluation index of the single statistical average of features based on 7 days monitoring.

Features (time=7 days)	Accuracy	Sensitivity	Specificity	Youden index	Area under the curve (95% CI)	*P* value
SDNN[a]	0.745	0.465	0.899	0.364	0.768 (0.759-0.777)	<.001
TP[b]	0.792	0.549	0.925	0.474	0.832 (0.825-0.840)	<.001
LF[c]	0.788	0.555	0.917	0.472	0.829 (0.821-0.837)	<.001
HF[d]	0.830	0.701	0.901	0.602	0.890 (0.884-0.896)	<.001
VLF[e]	0.746	0.414	0.929	0.343	0.724 (0.714-0.734)	<.001
ULF[f]	0.708	0.286	0.940	0.226	0.688 (0.677-0.698)	<.001
HR[g]	0.674	0.227	0.920	0.147	0.669 (0.659-0.679)	<.001
RR[h]	0.830	0.651	0.927	0.578	0.869 (0.862-0.876)	<.001
RRF[i]	0.898	0.759	0.974	0.733	0.955 (0.950-0.959)	<.001

[a]SDNN: standard deviation of heartbeat interval.

[b]TP: total spectral power of heartbeat interval.

[c]LF: low frequency band power of 0.04-0.15 Hz.

[d]HF: high frequency band power of 0.15-0.40 Hz.

[e]VLF: very low frequency band power of 0.0033-0.04 Hz.

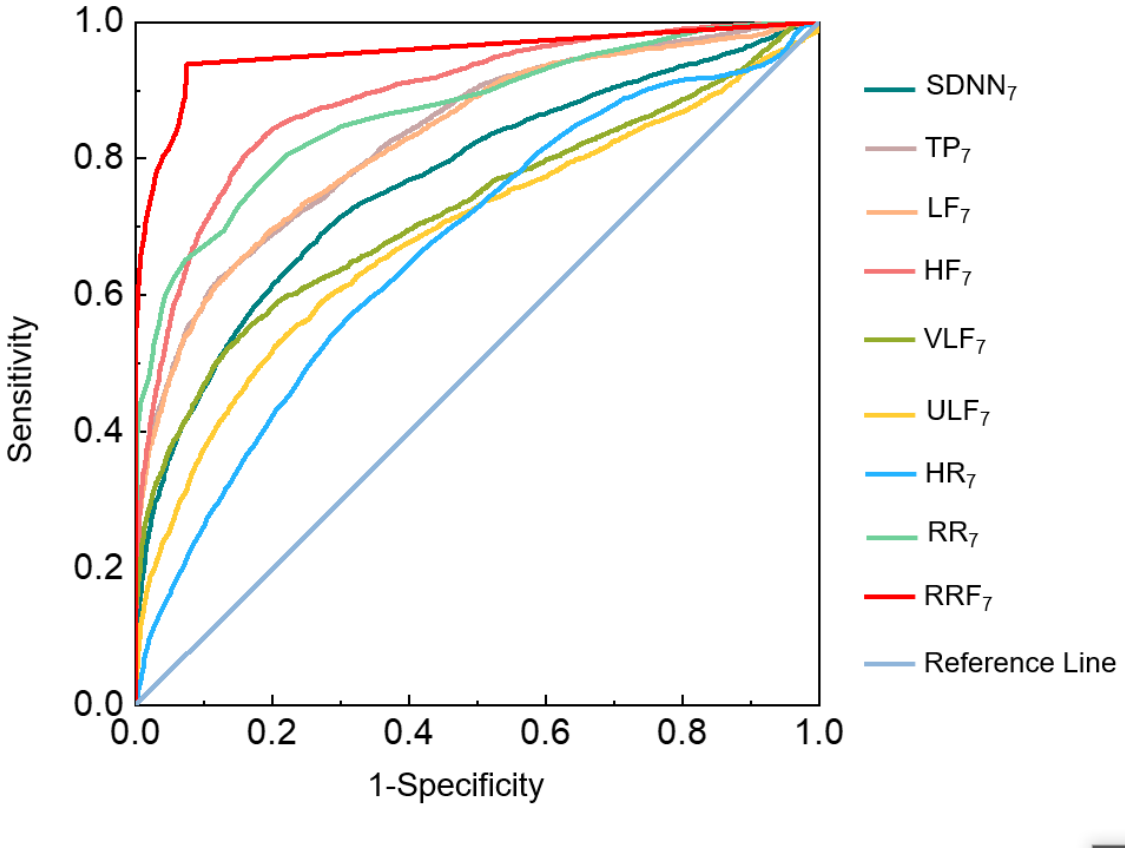
[f]ULF: ultralow frequency band power of <0.0033 Hz.

[g]HR: heart rate.

[h]RR: respiratory rate.

[i]RRF: respiratory rate faster than 21 times/min.

Figure 5. Receiver operating characteristic curve of single statistical average of features based on 7 days monitoring. HF: high frequency band power of 0.15-0.40 Hz; HR: heart rate; LF: low frequency band power of 0.04-0.15 Hz; RR: respiratory rate; RRF: respiratory rate faster than 21 times/min; SDNN: standard deviation of heartbeat interval; TP: total power of heartbeat interval; ULF: ultralow frequency band power of <0.0033 Hz; VLF: very low frequency band power of 0.0033-0.04 Hz.



Cutoff Values

Considering that the classification performance of HF_7, RR_7, and RRF_7 was superior to that of other features, which reflected the potential values of such features for the diagnosis of COPD, we further analyzed the cutoff values based on the maximum Youden index for HF_7, RR_7, and RRF_7. This serves as a reference for auxiliary diagnosis in patients with COPD. The determined cutoff value for HF_7 was 1316.30 nU, which corresponded to sensitivity and specificity values of 0.836 and 0.810, respectively. The cutoff value for RR_7 was 16.3 times/min, accompanied by sensitivity and specificity values of 0.783 and 0.803, respectively. As for RRF_7, the designated cutoff value was set at 0.1 minute with corresponding sensitivity and specificity values of 0.938 and 0.925, respectively. Samples exceeding these respective cutoff values would be assigned to the case group, while the others would be assigned to the control group (as illustrated in Table 5).

Table 5. Cutoff values of statistical average of respiration rate faster than 21 times/min, high frequency band power of 0.15-0.40 Hz in heart rate variability, and respiration rate.

Features (time=7 days)	Cutoff point	Sensitivity	Specificity	Youden index
High frequency band power of 0.15-0.40 Hz	1316.30	0.836	0.810	0.646
Respiratory rate	16.30	0.783	0.803	0.585
Respiratory rate faster than 21 times/min	0.10	0.938	0.925	0.864

Discussion

Principal Findings

The primary purpose of this study was to explore the feasibility of bed sensor–aided continuous night-sleep monitoring of vital signs (HRV, HR, and respiration) to assist in the remote diagnosis of COPD. Our findings show that continuous night-sleep monitoring has significant potential in the diagnosis of COPD, and the statistical means of RRF, HF, and RR are the crucial features. By extracting statistical features from the continuously recorded HRV, HR, and respiration over 7, 14, and 30 days, the classification performance of the vital signs for COPD diagnosis was evaluated and compared in different timescales, and the cutoff values of the most important features were obtained as a reference point for in-home diagnosis of COPD. Our prospective cohort experimental analysis verified

that the classification performance of the diagnosis of COPD increases directionally proportional to the monitoring duration of vital signs recording. This outcome is expected due to the influence of various factors such as environment, daytime emotions, and activities on the recorded vital signs during a single night sleep [26,27]. Comparatively, by increasing the statistical timescale with regard to feature extraction, the derived statistical features can reflect the status of individual health more accurately [28,29]. Under the tradeoff between the performance of diagnosis and the feasibility of continuous monitoring, it is recommended that a 7-day daily home-based continuous monitoring is sufficient to assist in the diagnosis of COPD.

According to previous research, the resting HR and respiration of patients with COPD are higher than those of healthy individuals. However, there is still a lack of literature support regarding the comparison of the correlation between these indicators and patients with COPD [30]. In this study, we discovered that long-term monitored features such as RR, HF, and RRF were more significant for diagnosing COPD than HR alone. Using either RR or HF as features, the classification performance accuracy already exceeded 80%. Furthermore, when utilizing RRF for classification, the accuracy reached 90%. In comparison, if HR was used for classification purposes only, the accuracy dropped to 67%, with a missed diagnosis rate exceeding 70%.

The resting RR for healthy individuals ranges from 12 to 20 beats per minute [31]. However, our findings showed that RR exceeded 16 beats per minute, which can be viewed as an important feature for the diagnosis of COPD. This finding is expected since patients with COPD are more prone to experiencing rapid breathing [32]. The cutoff value for RRF in this study was determined as 0.1 minute, indicating that any presence of rapid breathing during night sleep significantly increases the probability of a COPD diagnosis. The corresponding sensitivity and specificity values reach up to 94% and 93%, respectively.

Patients with COPD exhibit abnormal autonomic nervous function and reduced HRV [33,34]. Studies have demonstrated a decrease in sympathetic regulation among stable patients with COPD, suggesting a potential shift toward parasympathetic regulation [35]. By contrast, our findings show that patients with COPD exhibit an elevated HRV compared to healthy controls. The possible reasons are 2-fold. On the one hand, the HRV indicators of patients with COPD were increased due to cardiac compensation since all patients with severe cardiovascular diseases were excluded in this study. This phenomenon was also validated by a pioneer study [36]. On the other hand, as pointed out by [37], parasympathetic regulation is more active in patients with acute exacerbation of COPD compared to those with stable COPD. In this study, although the events of acute exacerbation in the case group were excluded before analysis, potential labelling errors of the acute exacerbation cases existed, since the recording of periodic follow-up might have been incomplete due to out-of-hospital visits or recall bias, and small amount of monitoring data during acute exacerbations were included, which could account for the observed increase in HRV among patients with COPD.

LF reflects the regulation of the sympathetic nerve, HF reflects the activity of the parasympathetic nerve, and LF/HF reflects the relative activity of the sympathetic nerve and the parasympathetic nerve, that is, the balance between the sympathetic nerve and the parasympathetic nerve [36]. The results of our study showed that there was no statistically significant differences in the LF/HF characteristics between the case group and the control group, which indicated that in patients with COPD, the LF and HF both increased, thus resulting in the relatively constant LF/HF indices. This is also consistent with previous studies, which demonstrated no significant difference between the LF/HF of patients with COPD and the control groups [37].

Strengths and Limitations

This study is the first prospective observational study based on continuous vital signs monitoring during night sleep to assist in-home diagnosis of COPD. With the help of noncontact sensors placed under the mattress, we performed continuous night sleep monitoring for individuals with COPD and healthy individuals over weeks and employed the statistical means of the monitored HR, respiration, and HRV at different timescales as features instead of the short-term measurement in a single day. Comparatively, the proposed long-term monitoring could reduce the uncertainty and inaccuracy caused by various interference factors present in short-term monitoring sessions and thereby improve the performance of the auxiliary diagnosis of COPD.

This study has some limitations. The majority of the individuals in the case group were patients with COPD recruited from the hospital. Due to the characteristics of COPD, the proportion of old and lean men in the case group was higher, while individuals in the control group were normal primarily. Statistical comparisons revealed uneven distributions regarding age, sex, and BMI at baseline between the 2 groups, potentially introducing a selection bias. To address this issue, we accounted for these factors during data analysis by adjusting them prior to analysis. Consequently, we observed an improvement in the classification performance while observing minor impacts on the comparison of the classification effects for different features among our patients.

Conclusions

In this paper, we report the utilization of bed sensors for continuous recording of vital signs, including HR, respiration, and HRV during nighttime, thereby aiming to provide auxiliary diagnosis for COPD. Compared to single-time and single-day measurement, the long-term statistical features derived from continuous vital signs recordings demonstrate superior performance in the diagnosis of COPD. Moreover, the longer the timescale used for constructing the statistical features, the better is the classification performance for COPD diagnosis. Feature importance analysis showed that the statistical features of respiration and HRV contributed significantly more than the HR features in the COPD diagnosis. Specifically, RR, RRF, and HF exhibited the highest contribution to COPD diagnosis and hold important reference for home-assisted diagnosis of COPD. Additionally, this study determines the cutoff values for important features, which can provide a reference range for

clinical COPD diagnosis. In principle, this design may also be applicable to other cardiovascular diseases such as heart failure, but the important features for disease diagnosis may be different.

Acknowledgments

This research was supported by Guangzhou Key Research and Development Plan project (202206010127, 202206010099). We thank the Third Affiliated Hospital of Sun Yat-sen University and Guangzhou Zhongke Xinzhi Co Ltd for providing the data and the monitoring equipment.

Authors' Contributions

HZ (zhanghan@scnu.edu.cn), YZ (zhouyuqi@mail.sysu.edu.cn), and ZL (lizw@bjmu.edu.cn) are the co-corresponding authors. HZ and XC contributed to the study design, data analysis, interpretation of the findings, and preparation of the manuscript. YZ contributed to the study design and identified the clinical cases. ZL contributed to data analysis and interpretation of the findings. SL contributed to identifying the clinical cases.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary data.
[DOCX File , 47 KB - mhealth_v12i1e56226_app1.docx]

References

1. Zhong N, Wang C, Yao W, Chen P, Kang J, Huang S, et al. Prevalence of chronic obstructive pulmonary disease in China. Am J Respir Crit Care Med 2007 Oct 15;176(8):753-760. [doi: 10.1164/rccm.200612-1749oc]
2. Fang L, Gao P, Bao H, Tang X, Wang B, Feng Y, et al. Chronic obstructive pulmonary disease in China: a nationwide prevalence study. Lancet Respir Med 2018 Jun;6(6):421-430 [FREE Full text] [doi: 10.1016/S2213-2600(18)30103-6] [Medline: 29650407]
3. Wang C, Xu J, Yang L, Xu Y, Zhang X, Bai C, China Pulmonary Health Study Group. Prevalence and risk factors of chronic obstructive pulmonary disease in China (the China Pulmonary Health [CPH] study): a national cross-sectional study. Lancet 2018 Apr 28;391(10131):1706-1717. [doi: 10.1016/S0140-6736(18)30841-9] [Medline: 29650248]
4. GBD 2019 DiseasesInjuries Collaborators. Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet 2020 Oct 17;396(10258):1204-1222 [FREE Full text] [doi: 10.1016/S0140-6736(20)30925-9] [Medline: 33069326]
5. Safiri S, Carson-Chahhoud K, Noori M, Nejadghaderi SA, Sullman MJM, Ahmadian Heris J, et al. Burden of chronic obstructive pulmonary disease and its attributable risk factors in 204 countries and territories, 1990-2019: results from the Global Burden of Disease Study 2019. BMJ 2022 Jul 27;378:e069679 [FREE Full text] [doi: 10.1136/bmj-2021-069679] [Medline: 35896191]
6. Van Gestel AJR, Kohler M, Steier J, Teschler S, Russi EW, Teschler H. Cardiac autonomic dysfunction and health-related quality of life in patients with chronic obstructive pulmonary disease. Respirology 2011 Aug;16(6):939-946. [doi: 10.1111/j.1440-1843.2011.01992.x] [Medline: 21564403]
7. Chronic Obstructive Pulmonary Disease Group of Chinese Thoracic Society, Chronic Obstructive Pulmonary Disease Committee of Chinese Association of Chest Physician. [Guidelines for the diagnosis and management of chronic obstructive pulmonary disease (revised version 2021)]. Zhonghua Jie He He Hu Xi Za Zhi 2021 Mar 12;44(3):170-205. [doi: 10.3760/cma.j.cn112147-20210109-00031] [Medline: 33721932]
8. Bernardo AFB, Vanderlei LCM, Garner DM. HRV analysis: a clinical and diagnostic tool in chronic obstructive pulmonary disease. Int Sch Res Notices 2014;2014:673232 [FREE Full text] [doi: 10.1155/2014/673232] [Medline: 27433490]
9. Al-Halhouli A, Al-Ghussain L, Khallouf O, Rabadi A, Alawadi J, Liu H, et al. Clinical evaluation of respiratory rate measurements on COPD (male) patients using wearable inkjet-printed sensor evaluation of respiratory rate measurements on COPD (male) patients using wearable inkjet-printed sensor. Sensors (Basel) 2021 Jan 11;21(2):468 [FREE Full text] [doi: 10.3390/s21020468] [Medline: 33440773]
10. Warnier MJ, Rutten FH, Numans ME, Kors JA, Tan HL, de Boer A, et al. Electrocardiographic characteristics of patients with chronic obstructive pulmonary disease. COPD 2013 Feb;10(1):62-71. [doi: 10.3109/15412555.2012.727918] [Medline: 23413894]
11. Chee Y, Han J, Youn J, Park K. Air mattress sensor system with balancing tube for unconstrained measurement of respiration and heart beat movements. Physiol Meas 2005 Aug;26(4):413-422. [doi: 10.1088/0967-3334/26/4/007] [Medline: 15886436]

12. Liu J, Miao F, Yin L, Pang Z, Li Y. A noncontact ballistocardiography-based IoMT system for cardiopulmonary health monitoring of discharged COVID-19 patients. IEEE Internet Things J 2021 Nov 1;8(21):15807-15817. [doi: 10.1109/jiot.2021.3063549]
13. Li KHC, White FA, Tipoe T, Liu T, Wong MC, Jesuthasan A, et al. The current state of mobile phone apps for monitoring heart rate, heart rate variability, and atrial fibrillation: narrative review. JMIR Mhealth Uhealth 2019 Feb 15;7(2):e11606 [FREE Full text] [doi: 10.2196/11606] [Medline: 30767904]
14. Wen X, Huang Y, Wu X, Zhang B. A feasible feature extraction method for atrial fibrillation detection from BCG. IEEE J Biomed Health Inform 2020 Apr;24(4):1093-1103. [doi: 10.1109/jbhi.2019.2927165]
15. Chen W, Zhu X, Nemoto T, Kanemitsu Y, Kitamura K, Yamakoshi K. Unconstrained detection of respiration rhythm and pulse rate with one under-pillow sensor during sleep. Med Biol Eng Comput 2005 Mar;43(2):306-312. [doi: 10.1007/BF02345970] [Medline: 15865143]
16. Naranjo-Hernández D, Talaminos-Barroso A, Reina-Tosina J, Roa LM, Barbarov-Rostan G, Cejudo-Ramos P, et al. Smart vest for respiratory rate monitoring of COPD patients based on non-contact capacitive sensing. Sensors (Basel) 2018 Jul 03;18(7):2144 [FREE Full text] [doi: 10.3390/s18072144] [Medline: 29970861]
17. Bellos CC, Papadopoulos A, Rosso R, Fotiadis DI. Identification of COPD patients' health status using an intelligent system in the CHRONIOUS wearable platform. IEEE J Biomed Health Inform 2014 May;18(3):731-738. [doi: 10.1109/JBHI.2013.2293172] [Medline: 24808219]
18. Newandee D, Reisman S, Bartels M, De MR. COPD severity classification using principal component and cluster analysis on HRV parameters. 2003 Presented at: 2003 IEEE 29th Annual Proceedings of Bioengineering Conference; August 4; Newark, NJ, USA p. 134-135. [doi: 10.1109/nebc.2003.1216028]
19. Rahman MJ, Nemati E, Rahman MM, Nathan V, Vatanparvar K, Kuang J. Automated assessment of pulmonary patients using heart rate variability from everyday wearables. Smart Health 2020 Mar;15:100081. [doi: 10.1016/j.smhl.2019.100081]
20. Tiwari A, Liaqat S, Liaqat D, Gabel M, de Lara E, Falk TH. Remote COPD severity and exacerbation detection using heart rate and activity data measured from a wearable device. Annu Int Conf IEEE Eng Med Biol Soc 2021 Nov;2021:7450-7454. [doi: 10.1109/EMBC46164.2021.9629949] [Medline: 34892818]
21. Global strategy for the diagnosis, management and prevention of chronic obstructive pulmonary disease (2022 report). Global Initiative for Chronic Obstructive Lung Disease. URL: https://goldcopd.org/wp-content/uploads/2021/12/GOLD-REPORT-2022-v1.1-22Nov2021_WMV.pdf [accessed 2021-11-22]
22. Mai Y, Chen Z, Yu B, Li Y, Pang Z, Han Z. Non-contact heartbeat detection based on ballistocardiogram using UNet and bidirectional long short-term memory. IEEE J Biomed Health Inform 2022 Aug;26(8):3720-3730. [doi: 10.1109/jbhi.2022.3162396]
23. Zhang H, Zhu W, Ye S, Li S, Yu B, Pang Z, et al. Monitoring of non-invasive vital signs for detection of sleep apnea. J Mech Med Biol 2021 Apr 07;21(05):2140007. [doi: 10.1142/s0219519421400078]
24. Liang J, Huang J, Mu L, Yu B, Chen P, Pang Z, et al. An effective algorithm for beat-to-beat heart rate monitoring from ballistocardiograms. J Med Imaging Hlth Inform 2020 Mar 01;10(3):633-640. [doi: 10.1166/jmihi.2020.2910]
25. Santos-de-Araújo AD, Bassi-Dibai D, Camargo PF, Marinho RS, Goulart CDL, Dourado IM, et al. Inter- and intrarater reliability of short-term measurement of heart rate variability on rest in chronic obstructive pulmonary disease (COPD). Heart Lung 2023;62:64-71. [doi: 10.1016/j.hrtlng.2023.06.004] [Medline: 37327614]
26. Čukić M, Savić D, Sidorova J. When heart beats differently in depression: review of nonlinear heart rate variability measures. JMIR Ment Health 2023 Jan 17;10:e40342 [FREE Full text] [doi: 10.2196/40342] [Medline: 36649063]
27. Hawthorne G, Greening N, Esliger D, Briggs-Price S, Richardson M, Chaplin E, et al. Usability of wearable multiparameter technology to continuously monitor free-living vital signs in people living with chronic obstructive pulmonary disease: prospective observational study. JMIR Hum Factors 2022 Feb 16;9(1):e30091 [FREE Full text] [doi: 10.2196/30091] [Medline: 35171101]
28. Wu R, Liaqat D, de Lara E, Son T, Rudzicz F, Alshaer H, et al. Feasibility of using a smartwatch to intensively monitor patients with chronic obstructive pulmonary disease: prospective cohort study. JMIR Mhealth Uhealth 2018 Jun 14;6(6):e10046 [FREE Full text] [doi: 10.2196/10046] [Medline: 29903700]
29. Buekers J, Theunis J, De Boever P, Vaes AW, Koopman M, Janssen EV, et al. Wearable finger pulse oximetry for continuous oxygen saturation measurements during daily home routines of patients with chronic obstructive pulmonary disease (COPD) over one week: observational study. JMIR Mhealth Uhealth 2019 Jun 06;7(6):e12866 [FREE Full text] [doi: 10.2196/12866] [Medline: 31199331]
30. Shah SA, Velardo C, Farmer A, Tarassenko L. Exacerbations in chronic obstructive pulmonary disease: identification and prediction using a digital health system. J Med Internet Res 2017 Mar 07;19(3):e69 [FREE Full text] [doi: 10.2196/jmir.7207] [Medline: 28270380]
31. Tavares LP, Peh HY, Tan WSD, Pahima H, Maffia P, Tiligada E, et al. Granulocyte-targeted therapies for airway diseases. Pharmacol Res 2020 Jul;157:104881 [FREE Full text] [doi: 10.1016/j.phrs.2020.104881] [Medline: 32380052]
32. Ballal T, Heneghan C, Zaffaroni A, Boyle P, de Chazal P, Shouldice R, et al. A pilot study of the nocturnal respiration rates in COPD patients in the home environment using a non-contact biomotion sensor. Physiol Meas 2014 Dec;35(12):2513-2527. [doi: 10.1088/0967-3334/35/12/2513] [Medline: 25402668]

33. Roque AL, Valenti VE, Massetti T, da Silva TD, Monteiro CBDM, Oliveira FR, de Almeida Junior, et al. Chronic obstructive pulmonary disease and heart rate variability: a literature update. Int Arch Med 2014;7:43 [FREE Full text] [doi: 10.1186/1755-7682-7-43] [Medline: 25945125]
34. Park S, Saiphoklang N, Jung D, Gomez D, Phillips JE, Dolezal BA, et al. Use of a wearable biosensor to study heart rate variability in chronic obstructive pulmonary disease and its relationship to disease severity. Sensors (Basel) 2022 Mar 15;22(6):2264 [FREE Full text] [doi: 10.3390/s22062264] [Medline: 35336436]
35. Serrão NF, Porta A, Minatel V, Castro AAM, Catai AM, Sampaio LMM, et al. Complexity analysis of heart rate variability in chronic obstructive pulmonary disease: relationship with severity and symptoms. Clin Auton Res 2020 Apr;30(2):157-164. [doi: 10.1007/s10286-019-00659-z] [Medline: 31938978]
36. Kabbach EZ, Mazzuco A, Borghi-Silva A, Cabiddu R, Agnoleto AG, Barbosa JF, et al. Increased parasympathetic cardiac modulation in patients with acute exacerbation of COPD: how should we interpret it? Int J Chron Obstruct Pulmon Dis 2017;12:2221-2230 [FREE Full text] [doi: 10.2147/COPD.S134498] [Medline: 28814850]
37. Zamarrón C, Lado MJ, Teijeiro T, Morete E, Vila XA, Lamas PF. Heart rate variability in patients with severe chronic obstructive pulmonary disease in a home care program. THC 2014 Feb 01;22(1):91-98. [doi: 10.3233/thc-140777]

Abbreviations

AUC: area under the receiver operating characteristic curve
COPD: chronic obstructive pulmonary disease
GOLD: Global Initiative for Chronic Obstructive Lung Disease
HF: high frequency band power of 0.15-0.40 Hz
HR: heart rate
HRV: heart rate variability
LF: low frequency band power of 0.04-0.15 Hz
ROC: receiver operating characteristic
RR: respiratory rate
RRF: respiratory rate faster than 21 times/min
SDNN: standard deviation of heartbeat interval
TP: total spectral power of heartbeat interval
ULF: ultralow frequency band power of <0.0033 Hz
VLF: very low frequency band power of 0.0033-0.04 Hz

Edited by L Buis; submitted 12.01.24; peer-reviewed by T Kamio, M Liu; comments to author 23.05.24; revised version received 07.06.24; accepted 18.06.24; published 18.07.24.

Please cite as:
Chen X, Zhang H, Li Z, Liu S, Zhou Y
Continuous Monitoring of Heart Rate Variability and Respiration for the Remote Diagnosis of Chronic Obstructive Pulmonary Disease: Prospective Observational Study
JMIR Mhealth Uhealth 2024;12:e56226
URL: https://mhealth.jmir.org/2024/1/e56226
doi:10.2196/56226
PMID:

Monitoring Adolescent and Young Adult Patients With Cancer via a Smart T-Shirt: Prospective, Single-Cohort, Mixed Methods Feasibility Study (OncoSmartShirt Study)

Emma Balch Steen-Olsen[1], MD; Helle Pappot[1,2], MD; Maiken Hjerming[1], BSN; Signe Hanghoej[1], MA, MPH, PhD; Cecilie Holländer-Mieritz[1,3], MD, PhD

[1]Department of Oncology, Centre for Cancer and Organ Diseases, Copenhagen University Hospital - Rigshospitalet, Copenhagen, Denmark

[2]Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark

[3]Department of Oncology, Zealand University Hospital, Naestved, Denmark

Corresponding Author:
Cecilie Holländer-Mieritz, MD, PhD
Department of Oncology, Centre for Cancer and Organ Diseases, Copenhagen University Hospital - Rigshospitalet, , Copenhagen, , Denmark

Abstract

Background: Wearables that measure vital parameters can be potential tools for monitoring patients at home during cancer treatment. One type of wearable is a smart T-shirt with embedded sensors. Initially, smart T-shirts were designed to aid athletes in their performance analyses. Recently however, researchers have been investigating the use of smart T-shirts as supportive tools in health care. In general, the knowledge on the use of wearables for symptom monitoring during cancer treatment is limited, and consensus and awareness about compliance or adherence are lacking.

Objectives: The aim of this study was to evaluate adherence to and experiences with using a smart T-shirt for the home monitoring of biometric sensor data among adolescent and young adult patients undergoing cancer treatment during a 2-week period.

Methods: This study was a prospective, single-cohort, mixed methods feasibility study. The inclusion criteria were patients aged 18 to 39 years and those who were receiving treatment at Copenhagen University Hospital - Rigshospitalet, Denmark. Consenting patients were asked to wear the Chronolife smart T-shirt for a period of 2 weeks. The smart T-shirt had multiple sensors and electrodes, which engendered the following six measurements: electrocardiogram (ECG) measurements, thoracic respiration, abdominal respiration, thoracic impedance, physical activity (steps), and skin temperature. The primary end point was adherence, which was defined as a wear time of >8 hours per day. The patient experience was investigated via individual, semistructured telephone interviews and a paper questionnaire.

Results: A total of 10 patients were included. The number of days with wear times of >8 hours during the study period (14 d) varied from 0 to 6 (mean 2 d). Further, 3 patients had a mean wear time of >8 hours during each of their days with data registration. The number of days with any data registration ranged from 0 to 10 (mean 6.4 d). The thematic analysis of interviews pointed to the following three main themes: (1) the smart T-shirt is cool but does not fit patients with cancer, (2) the technology limits the use of the smart T-shirt, and (3) the monitoring of data increases the feeling of safety. Results from the questionnaire showed that the patients generally had confidence in the device.

Conclusions: Although the primary end point was not reached, the patients' experiences with using the smart T-shirt resulted in the knowledge that patients acknowledged the need for new technologies that improve supportive cancer care. The patients were positive when asked to wear the smart T-shirt. However, technical and practical challenges in using the device resulted in low adherence. Although wearables might have potential for home monitoring, the present technology is immature for clinical use.

Trial Registration: ClinicalTrials.gov NCT05235594; https://clinicaltrials.gov/study/NCT05235594

International Registered Report Identifier (IRRID): RR2-10.2196/37626

(JMIR Mhealth Uhealth 2024;12:e50620) doi:10.2196/50620

KEYWORDS

smart T-shirt; AYA; oncology; home monitoring; patients' perspective; perspective; perspectives; experiences; experience; youth; adolescent; adolescents; smart; monitoring; biometric; sensor; sensors; young adult; young adults; feasibility; cancer; cancers; electrode; electrodes; adherence; mobile phone

Introduction

Patients with cancer can be exposed to several treatments (eg, surgery, radiation, chemotherapy, and hormone therapy), individually or in combination, depending on their disease and stage. Cancer treatment is known to cause acute side effects [1-3]. The degree of symptoms and side effects depends on the type of cancer, the treatment modality, and the pre-existing comorbidity [1,4,5]. During oncological treatment, the patients may need acute hospitalization due to side effects, while in other cases, side effects are related to poor treatment compliance and reduced quality of life (QoL) [6].

Several studies have emphasized that patients and health care professionals can assess and perceive symptoms and side effects differently [7,8]. This is exemplified by the fact that health care professionals tend to underestimate patients' symptoms [9]. The development of side effects and symptoms often results in a deterioration of the patient's health condition, affecting the patient's QoL [10,11]. This applies especially to adolescents and young adults (AYAs) with cancer [12].

There is an increased focus on home monitoring to help patients manage their symptoms and side effects. Patient-generated health data can provide health care professionals with valuable information. One type of patient-generated health data is biometric sensor data, which are typically collected by wearables [13-15]. A wearable device is a noninvasive wireless sensor that monitors and collects health parameters [13,16]. A newer type of wearable is a smart T-shirt with biometric sensors embedded in the fabric. Wearables allow health professionals to monitor an increased number of health parameters on various biometric data points.

The data collected from wearables are predicted to be exact and comparable to data collected from conventional medical measuring devices [17-19]. These new technologies allow for the more extensive passive monitoring of patients in their home environment and may minimize the burden resulting from hospital visits [6,16,20-23]. In addition, wearables ensure exact information without recall and reporting bias, which hopefully results in better cancer treatment [20,24-26]. However, studies that investigate the use of wearables in an oncological setting are limited [24,27-29]. Furthermore, it has been stated that there is a lack of consensus and awareness about compliance with or adherence to wearables. These are essential parts of using and comparing collected biometric sensor data [30].

Many existing and new technologies are not developed or evaluated based on users' perspectives and sometimes do not adequately meet the needs of their target groups [31-33]. The AYA patient group frequently uses new technologies, such as wearables [34]. AYAs thus have unique and beneficial knowledge, which is why patients' involvement in the study design and in feasibility assessment can be extremely useful [35].

The aim of this study was to evaluate feasibility based on adherence to and experiences with using a smart T-shirt for the remote monitoring of biometric sensor data among AYA patients undergoing cancer treatment during a 2-week test period.

Methods

Ethical Considerations

This study was an investigator-driven partnership between the Department of Oncology, Rigshospitalet, and Chronolife and was registered at ClinicalTrials.gov (trial number: NCT05235594). This study conformed to the General Data Protection Regulation guidelines and was registered at the Capital Region of Denmark (registration number: P-2021-357). The trial was approved by the local division for IT and Medico Technology in the Capital Region of Denmark and was a collaboration between the Department of Oncology, Rigshospitalet; the Department of Innovation, Rigshospitalet; and the Telemedical Knowledge Center, Capital Region of Denmark. Approval from the National Committee on Health Research Ethics was not required for this trial in the Danish context when this study was conducted. Informed consent was obtained from all patients involved in this study. The patients received verbal and written information. Written informed consent was required, and patients were informed that it was possible to withdraw from this study at any time during the study period. No financial compensation was provided.

Study Design

The OncoSmartShirt study was a prospective, single-cohort, mixed methods study that investigated the feasibility of using the Chronolife smart T-shirt (Keesense) during cancer treatment. This smart T-shirt was designed with multiple fully embedded sensors and electrodes, which engendered 6 different measurement flows continuously [25]. Before this study was conducted, the project was presented to a group of AYAs with cancer at a social meeting with the "Kræftværket" cancer network group. The participants provided the researchers with their inputs and perspectives on the study design to make it relevant and feasible. The acceptance and comfort of wearing the Chronolife smart T-shirt throughout the day (8 h/d) for 2 weeks (14 d) were investigated among all enrolled patients.

The inclusion criteria were young patients with cancer aged 18 to 39 years (defined as *AYAs*) and those who were receiving antineoplastic treatment at the Department of Oncology and Department of Haematology of the Centre for Cancer and Organ Diseases, Copenhagen University Hospital - Rigshospitalet, Copenhagen, Denmark. Other inclusion criteria were having the ability to read and speak Danish and having no serious cognitive deficits. There were no requirements regarding specific cancer diagnoses, and both patients in curative care and patients in palliative care could be included. Inclusion in this study did not interfere with the planned oncological treatment.

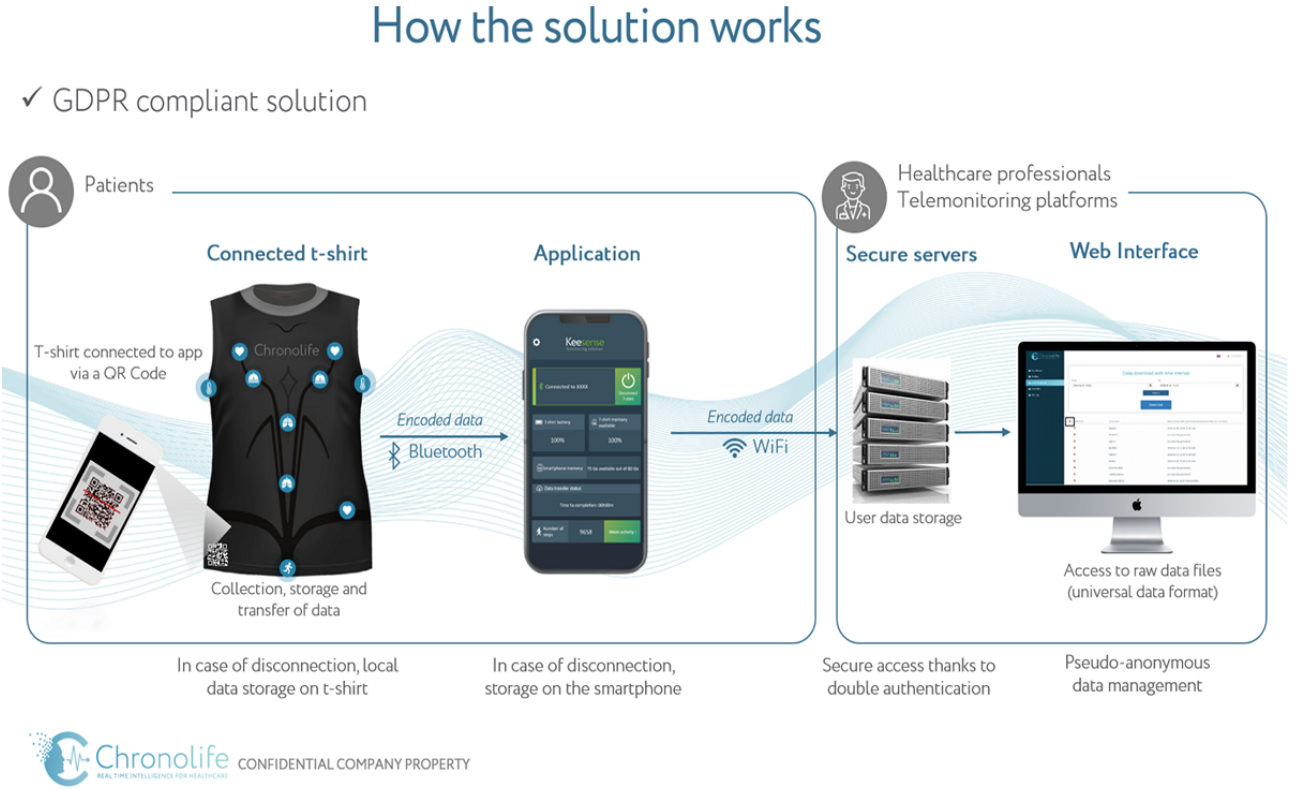
Further details on the OncoSmartShirt study can be reviewed in the previously published protocol article [36]. This paper reports results from 10 patients with cancer aged under 39 years (defined as *AYAs*). The decision to have a sample size of 10 AYA patients was influenced by the feasibility study design, which does not require a formal power calculation but aligns with the sample sizes used for similar studies in the literature [36]. In the previously published protocol, the plan was to also include 10 patients with cancer older than 65 years (defined as *elderly*). However, due to the results from the AYA cohort, the

research group omitted the inclusion of the second cohort of older patients.

Device

The device in this study consisted of the following four units: a washable smart T-shirt from Chronolife; a companion smartphone app; a secure, accredited data hosting server; and a web interface (Figure 1) [37]. The Chronolife smart T-shirt was designed for everyday use. It had electrical sensors embedded, allowing for the detection of the following six physiological parameters: electrocardiogram (ECG) measurements (beats/min), thoracic respiration (respirations/min), abdominal respiration (respirations/min), thoracic impedance (kΩ), physical activity (steps), and skin temperature (°C) [37]. A rechargeable battery powered the sensors. Additionally, a memory card that stores data and a Bluetooth interface that transmits data were fully integrated into the smart T-shirt and sealed in water-resistant coatings. The smart T-shirt was commercialized and Conformité Européenne–marked for the consumer market.

Figure 1. Framework for the OncoSmartShirt study (previously published figure from our protocol article [36]). GDPR: General Data Protection Regulation.



The smart T-shirt connected to the smartphone app via a QR code located on the smart T-shirt. Bluetooth Low Energy transmitted collected health data to the connected smartphone app designed for storage. The app transmitted data to a data hosting server that stored and provided data for a web interface, which was used for analysis and algorithm training [37].

The devices—the smart T-shirt and connected smartphone—were supplied by the hospital and returned after study termination. For data safety reasons, the patients could not use their own smartphones. Each smart T-shirt was produced based on individual patients' size measurements. No biometric data on health parameters collected by the smart T-shirt were sent to health care professionals during this study. The patients could contact a study assistant if they experienced technical difficulties, but this was not a requirement.

Patient Feedback

Immediately after the intervention period, qualitative individual telephone interviews were performed with the participants, allowing them to elaborate on their experiences with using the smart T-shirt. A semistructured interview guide was used. The interview guide was developed by SH based on research in the field [38,39] and consisted of questions about the pros and cons of using the smart T-shirt, material and appearance, identity and social stigma, behavioral changes, and ethics [36]. The interviews were performed by SH, who did not have prior knowledge about the participants and vice visa.

Patients were also asked to complete a quantitative paper questionnaire (handed out by a study assistant) concerning their experiences with wearing the smart T-shirt. The questionnaire consisted of 21 items, of which 14 were statements; patients could use a 5-point scale to indicate whether they agreed or disagreed with the statements. The remaining items consisted of free-text questions and items for gathering data on patients' estimates of wear time. The patients were asked to complete the questionnaire once. The questionnaire was prepared by Chronolife.

Variables

The primary end point was to assess the feasibility of using the Chronolife smart T-shirt based on adherence, which was defined as the number of included patients who used the smart T-shirt

(wear time) at least 8 hours per day during the 2-week period. The wear times were obtained from the processed data provided by Chronolife. Wear time was not self-reported. Secondary end points were patient feedback, which was obtained as described in the *Patient Feedback* section, and technical feasibility in a Danish health care system, including data quality. Explorative end points were changes in the health parameters monitored.

Data Analysis

Chronolife conducted the analysis of the collected health data. Qualitative data from the telephone interviews were transcribed and analyzed thematically, according to Braun and Clarke's [40] approach. Text coding involved reading and rereading the transcriptions to identify and categorize concepts across data. Concepts relevant to the research question were highlighted using colored marks in the transcriptions and then sorted into themes. One researcher did the coding (SH), which was discussed thoroughly with a senior researcher (HP) until a consensus was reached.

Results

Baseline Demographics and Characteristics

The recruitment of patients for the project was organized by using a Facebook post (Meta Platforms Inc; May 2021) about the project, which was posted in a closed group for young patients with cancer. The patients thus contacted the research groups if they wanted to participate in the project. A total of 10 patients (female: n=5; male: n=5) aged 22 to 30 (median 27, IQR 24.5-29.5) years were included during the inclusion period (March to June 2022). The types of cancer were leukemia (n=3), lymphoma (n=2), breast cancer (n=2), central nervous system cancer (n=1), testis cancer (n=1), and malignant melanoma (n=1).

Feasibility and Data Quality

As shown in Table 1, the number of days with wear times of >8 hours during the study period varied from 0 to 6 (mean 2 d). Only 3 patients had a mean wear time of >8 hours during each of their days with data registration. The number of days with any data registration varied from 0 to 10 (mean 6.4 d). No one managed to wear the smart T-shirt 8 hours per day for 14 days straight. Further, 4 patients had no data registrations at all. For 3 of these patients, the connection between the smart T-shirt and the smartphone app was unsuccessful because of technical issues or malfunctions, and for the last patient, the connection was successful, but the patient did not use the smart T-shirt due to disease-related issues.

Table . Wear time data for the smart T-shirt.

	Number of sessions[a]	Number of days with data registration	Number of days with a wear time of >8 h	Maximum wear time per session (h)	Wear time per day (h), mean[b]	Total wear time (h)	Data quality (%), mean[b]
Patient 1	20	10	6	19.1	10.4	110.4	63
Patient 2	—[c]	—	—	—	—	—	—
Patient 3	—	—	—	—	—	—	—
Patient 4	0	0	0	0	0	0	—
Patient 5	—	—	—	—	—	—	—
Patient 6	2	2	0	0.08	0.065	0.13	92
Patient 7	15	9	4	14.4	8	76	90
Patient 8	12	10	2	20	4.2	34	87
Patient 9	16	9	3	14.4	10.4	93.8	65
Patient 10	5	5	1	8.6	4.1	20.4	63

[a]The number of sessions is the total number of times the smart T-shirt was worn.

[b]SDs were not available from Chronolife.

[c]Not available.

The heart rate was calculated based on an ECG segment that was considered reliable by the manufacturer's data cleaning algorithm. The data quality was defined as the following ratio: the length of the session with heart rate values available divided by the total length of the session. The mean data quality for patients from whom data were collected varied from 63% to 92% (mean 77%). In this study, the data quality value was based on the quality of the heart rate values. The data quality can vary with multiple factors. The most important factor was the fit of the smart T-shirt; if the smart T-shirt was too large, the electrodes for the ECG would not have optimal contact with the skin, which would cause noise and artifacts in the collected data.

As variable compliance and data quality were noted, the analysis of the collected health data has been omitted due to the risk of the misinterpretation of the results.

Feedback From Patients

Interviews

Overview of Thematic Analysis

The thematic analysis was based on telephone interviews. The interviews lasted between 8 and 21 (mean 12) minutes. The thematic analysis pointed to the following three main themes: (1) the smart T-shirt is cool but does not fit patients with cancer, (2) the technology limits the use of the smart T-shirt, and (3) the monitoring of data increases the feeling of safety.

Theme 1: The Smart T-Shirt Is Cool but Does Not Fit Patients With Cancer

The smart T-shirt was described as "soft," and some participants did not even notice wearing it. All participants agreed that the smart T-shirt had a nice design, which they did not associate with anything patient-like. The term "cool" reflects the look of the smart T-shirt rather than the sensation experienced while wearing it. There were different opinions about how the body's temperature was affected by the smart T-shirt. A few participants got extremely hot and could not bear to wear it for a long time, and this worsened when these participants were physically active. However, others thought it felt cool in terms of body temperature. The smart T-shirt was closed with a zipper at the side. Most participants had a hard time zipping it by themselves. Some described the smart T-shirt as a bit too long, and several were bothered by the transverse bands on the smart T-shirt that contained the measuring equipment, which was not elastic like the rest of the smart T-shirt. For the sake of measurements, the smart T-shirt was designed to fit very tightly, which was a problem for several participants who had undergone surgery or were experiencing medical side effects. Moreover, the women could not wear a bra under the smart T-shirt. The consequence of all of these challenges was that it was too difficult for the participants to wear the smart T-shirt as prescribed.

Theme 2: The Technology Limits the Use of the Smart T-Shirt

The participants generally had challenges with getting the technology in the smart T-shirt and smartphone to work. Several participants described problems with charging the smart T-shirt. The smartphone's connection to Wi-Fi was associated with some problems. Additionally, 1 participant noted that the app stopped working one night when his smartphone made automatic updates. In general, the participants requested the smart T-shirt to connect to their own smartphones, so that they did not have to carry 2 smartphones simultaneously. The technical challenges were experienced as barriers to wearing the smart T-shirt. One of the participants explained that as a patient with cancer, they had very little energy to overcome everyday things (eg, technology that does not work).

Theme 3: The Monitoring of Data Increases the Feeling of Safety

All participants described the smart T-shirt as a useful and important invention that met their need for safety when being released from the hospital and not being monitored by health professionals anymore. Most participants believed that the smart T-shirt could be used optimally if the health professionals could track their health parameters from the hospital and then contact them if something looked abnormal (eg, in cases of changes in heart rate or breathing). One of the participants expressed that the monitoring was like "bringing the hospital home" because it provided him with the same feeling of safety as when they were hospitalized. Another explained that he felt less ill when he was at home and that the smart T-shirt could play a key role in getting home under safe conditions. However, there was also a participant who explained that constant monitoring from the hospital would require the patients to be introduced to their tracking data, so that they would know what was normal and what was not. Some participants explained that being released from the hospital could be very concerning, especially because of their increased focus on their body and whether it behaved differently. In general, the participants strongly desired to follow bodily signs during cancer treatment. One participant believed that it would have been more motivating to use the smart T-shirt if he could follow and view the collected data on the smartphone.

Quantitative Questionnaire

Of the 10 patients, 8 responded to the quantitative questionnaire, providing 165 out of 210 (item response rate: 78.6%) possible answers. The patients' answers to the 14 statements in the questionnaire are illustrated in Table 2. The answers were very different among the patients, but in general, the patients had confidence in the product and believed that their physician could use the collected health data. In addition, several patients were concerned about whether the smart T-shirt worked correctly and whether it would limit their daily activities.

Table . The patients' responses to the questionnaire concerning their experiences with the smart T-shirt (Keesense). The questionnaire was prepared by the smart T-shirt manufacturer (Chronolife).

Questionnaire statements	Patients' responses, n					Total responses, N
	1 (strongly disagree)	2 (disagree)	3 (neutral)	4 (agree)	5 (strongly agree)	
"Before starting the project, I had no fears regarding the medical device - the smart t-shirt and the phone"	0	0	0	1	7	8
"I think using this device will help my doctors monitor my condition more closely"	0	0	0	4	4	8
"I am afraid that using the Keesense medical device may affect my daily activities"	1	4	1	1	1	8
"I think using the Keesense medical device can help me be more active"	0	4	3	0	1	8
"I am concerned about a possible malfunction of the Keesense medical device"	3	2	1	2	0	8
"After being shown and taught how to use the Keesense medical device, I was confident that I could then use it"	0	0	1	4	3	8
"I think the smart t-shirt was easy to use"	0	1	2	4	1	8
"I felt comfortable with the use of the Keesense medical device"	0	3	1	3	1	8
"I easily forget that I am wearing the smart t-shirt"	1	2	2	1	2	8
"I can accomplish my daily activities with the smart t-shirt"	0	0	3	1	4	8
"I can easily do physical activity with the smart t-shirt"	0	1	4	1	2	8
"I find the Keesense medical device easy to use"	0	2	2	2	1	7[a]

Questionnaire statements	Patients' responses, n					Total responses, N
	1 (strongly disagree)	2 (disagree)	3 (neutral)	4 (agree)	5 (strongly agree)	
"I sweat abnormally while wearing the Keesense smart t-shirt"	0	2	4	0	1	7[a]
"I have skin itching and/or irritation"	4	2	0	1	0	7[a]

[a]One patient did not provide an answer for the statement.

Discussion

Principal Findings

The OncoSmartShirt study was a feasibility study that tested a smart T-shirt for the home monitoring of AYA patients with cancer from a public health care hospital in Denmark. To investigate adherence, we had predefined a preferred wear time of 8 hours daily for 2 weeks [36]. Unfortunately, none of the included patients achieved this. This finding is similar to the results from a comparable study conducted by Höllander-Mieritz et al [41] that investigated adherence to a smartwatch during radiotherapy among patients with head and neck cancer. Specific literature reviews show that it is possible to achieve high adherence to wearable technology in an oncology setting [30,42]. In general however, there is a tendency for compliance to decrease as the length of the study period increases. In addition, the exact desired wear time per day is not specified in several studies. This can contribute to the fact that it can be difficult to compare adherence across different studies.

We included qualitative and quantitative data in this study, and even though the wear time target was not met, this feasibility study resulted in knowledge about patients' experiences with the smart T-shirt and why the patients did not use the smart T-shirt in the predetermined time. In the qualitative interviews, we identified a discrepancy between the need for the smart T-shirt and the design of the smart T-shirt. The T-shirt met the patient's needs in terms of monitoring their health, but at the same time, it was not designed for patients with cancer experiencing treatment-induced side effects, such as gastrointestinal problems, increased body heat, and scars.

Another explanation for the lack of compliance is the technical challenges that some of the participants encountered. Prior to the completion of this study, we assumed that the group of AYAs would have greater technical ability and thus experience fewer technical problems when compared to older patients with cancer [43]. Based on the questionnaire, it appeared that most participants were confident with the technology and could use the smart T-shirt and the smartphone. However, in reality, 3 participants never connected their smart T-shirts to the smartphone app correctly, and several others experienced technical issues. In addition, it emerged from the interviews that patients with cancer do not have the energy to deal with technical problems associated with, for example, a smart T-shirt.

Responses to the questionnaire varied widely among participants and spanned the entire scale. Nevertheless, it is important to emphasize that the smart T-shirt manufacturer (Chronolife) prepared the questionnaire, which probably increased the risk of a ceiling effect among the answers [44].

To our knowledge, no studies have investigated the use of a smart T-shirt in an oncology setting, but there have been a few studies in the field of cardiology that primarily investigated if the monitoring of ECGs via a smart T-shirt can replace Holter monitoring [17,18,45]. In general, compliance is better in cardiology studies, but unlike the participants in the OncoSmartShirt study, the participants in those aforementioned cardiology studies were healthy and were not undergoing treatment. Thus, healthy participants in cardiology studies probably do not have the same challenges and annoyances as those among AYA patients with cancer [17,18]. Further, because of the low adherence and the patient experiences identified in patient interviews and the quantitative questionnaire, the research group excluded the preplanned group of older patients from this study. We believed that compared to the AYA patients included in this study, older participants would have experienced the same amount of issues (if not more) with the smart T-shirt and the setup with an extra phone. Therefore, we did not find it ethical to proceed with their inclusion in this study. However, we believe that age would not be an issue with a less demanding technical setup.

This study highlights the importance of investigating practical and technical feasibility. Practical feasibility refers to the specific wearable chosen, including the design, comfort, number of connected devices, and need for charging and maintenance. The smart T-shirt was not comfortable for our patient population. The patients had to have an extra phone to secure the setup and ensure safe data transfer. Both the smart T-shirt and the phone required charging, and the smart T-shirt also needed to be washed, which added to the patients' tasks. Technical feasibility refers to the setup for securing the data transfer, data quality, and reliable data. In this study, the data were safely transferred, but the data quality fluctuated.

For future studies investigating wearables, we suggest that the device be simple, comfortable, and minimally disturbing for the patient. The technical setup must also be simple; a possibility could be using the bring-your-own-device study design if the data can be securely transferred and if the data quality is sufficient. However, caution is advised for conducting bring-your-own-device studies in research due to the potential

imbalance [46]. We also recommend having a technical backup team that patients can contact and providing resources for electronic health education to patients and health care professionals. Finally, studies must assist in determining the potential uses of the data collected.

Although wearables might have the potential to be used in selected patient groups who need monitoring for a period of time, it is essential that the wearables can be worn and accepted by the patients and that the technical setup is as convenient as possible.

Conclusion

The OncoSmartShirt study was a feasibility study that investigated the use of a Chronolife smart T-shirt for the home monitoring of vital parameters among AYA patients with cancer during treatment. This study showed that AYA patients with cancer could not wear a smart T-shirt 8 hours per day for 2 weeks. However, this study revealed new and important perspectives and knowledge, which, among other things, pointed to why it can be challenging to achieve high compliance in this type of study. Furthermore, this study, as well as the patients, emphasized that wearables have potential. However, this area requires more research to develop the proper setup with minimal effort on the part of patients. Hopefully, in the long term, wearables can help improve the QoL for patients with cancer.

Conflicts of Interest

None declared.

References

1. Reilly CM, Bruner DW, Mitchell SA, et al. A literature synthesis of symptom prevalence and severity in persons receiving active cancer treatment. Support Care Cancer 2013 Jun;21(6):1525-1550. [doi: 10.1007/s00520-012-1688-0] [Medline: 23314601]
2. Kim JEE, Dodd MJ, Aouizerat BE, Jahan T, Miaskowski C. A review of the prevalence and impact of multiple symptoms in oncology patients. J Pain Symptom Manage 2009 Apr;37(4):715-736. [doi: 10.1016/j.jpainsymman.2008.04.018] [Medline: 19019626]
3. Fisch MJ, Lee JW, Weiss M, et al. Prospective, observational study of pain and analgesic prescribing in medical oncology outpatients with breast, colorectal, lung, or prostate cancer. J Clin Oncol 2012 Jun 1;30(16):1980-1988. [doi: 10.1200/JCO.2011.39.2381] [Medline: 22508819]
4. Moreira-Pinto J, Leão I, Palmela C, et al. Febrile neutropenia in patients with solid tumors undergoing intravenous chemotherapy. Oncol Res Treat 2020;43(11):605-612. [doi: 10.1159/000506109] [Medline: 32818937]
5. Niraula S, Seruga B, Ocana A, et al. The price we pay for progress: a meta-analysis of harms of newly approved anticancer drugs. J Clin Oncol 2012 Aug 20;30(24):3012-3019. [doi: 10.1200/JCO.2011.40.3824] [Medline: 22802313]
6. Kofoed S, Breen S, Gough K, Aranda S. Benefits of remote real-time side-effect monitoring systems for patients receiving cancer treatment. Oncol Rev 2012 Jun 5;6(1):e7. [doi: 10.4081/oncol.2012.e7] [Medline: 25992209]
7. Sonn GA, Sadetsky N, Presti JC, Litwin MS. Differing perceptions of quality of life in patients with prostate cancer and their doctors. J Urol 2013 Jan;189(1 Suppl):S59-S65. [doi: 10.1016/j.juro.2012.11.032] [Medline: 23234635]
8. Litwin MS, Lubeck DP, Henning JM, Carroll PR. Differences in urologist and patient assessments of health related quality of life in men with prostate cancer: results of the CaPSURE database. J Urol 1998 Jun;159(6):1988-1992. [doi: 10.1016/S0022-5347(01)63222-1] [Medline: 9598504]
9. Xiao C, Polomano R, Bruner DW. Comparison between patient-reported and clinician-observed symptoms in oncology. Cancer Nurs 2013;36(6):E1-E16. [doi: 10.1097/NCC.0b013e318269040f] [Medline: 23047799]
10. Burrows M, Dibble SL, Miaskowski C. Differences in outcomes among patients experiencing different types of cancer-related pain. Oncol Nurs Forum 1998 May;25(4):735-741. [Medline: 9599356]
11. Ferreira KASL, Kimura M, Teixeira MJ, et al. Impact of cancer-related symptom synergisms on health-related quality of life and performance status. J Pain Symptom Manage 2008 Jun;35(6):604-616. [doi: 10.1016/j.jpainsymman.2007.07.010] [Medline: 18362059]
12. Quinn GP, Gonçalves V, Sehovic I, Bowman ML, Reed DR. Quality of life in adolescent and young adult cancer patients: a systematic review of the literature. Patient Relat Outcome Meas 2015 Feb 17;6:19-51. [doi: 10.2147/PROM.S51658] [Medline: 25733941]
13. Dias D, Cunha JPS. Wearable health devices—vital sign monitoring, systems and technologies. Sensors (Basel) 2018 Jul 25;18(8):2414. [doi: 10.3390/s18082414] [Medline: 30044415]
14. Izmailova ES, Wood WA. Biometric monitoring technologies in cancer: the past, present, and future. JCO Clin Cancer Inform 2021 Jun;5:728-733. [doi: 10.1200/CCI.21.00019] [Medline: 34236887]
15. Fonseka LN, Woo BK. Consumer wearables and the integration of new objective measures in oncology: patient and provider perspectives. JMIR Mhealth Uhealth 2021 Jul 15;9(7):e28664. [doi: 10.2196/28664] [Medline: 34264191]
16. Ajami S, Teimouri F. Features and application of wearable biosensors in medical care. J Res Med Sci 2015 Dec;20(12):1208-1215. [doi: 10.4103/1735-1995.172991] [Medline: 26958058]

17. Steinberg C, Philippon F, Sanchez M, et al. A novel wearable device for continuous ambulatory ECG recording: proof of concept and assessment of signal quality. Biosensors (Basel) 2019 Jan 21;9(1):17. [doi: 10.3390/bios9010017] [Medline: 30669678]
18. Fouassier D, Roy X, Blanchard A, Hulot JS. Assessment of signal quality measured with a smart 12-lead ECG acquisition T-shirt. Ann Noninvasive Electrocardiol 2020 Jan;25(1):e12682. [doi: 10.1111/anec.12682] [Medline: 31339208]
19. Tsukada YT, Tokita M, Murata H, et al. Validation of wearable textile electrodes for ECG monitoring. Heart Vessels 2019 Jul;34(7):1203-1211. [doi: 10.1007/s00380-019-01347-8] [Medline: 30680493]
20. Beg MS, Gupta A, Stewart T, Rethorst CD. Promise of wearable physical activity monitors in oncology practice. J Oncol Pract 2017 Feb;13(2):82-89. [doi: 10.1200/JOP.2016.016857] [Medline: 28387544]
21. Jim HSL, Hoogland AI, Brownstein NC, et al. Innovations in research and clinical care using patient-generated health data. CA Cancer J Clin 2020 May;70(3):182-199. [doi: 10.3322/caac.21608] [Medline: 32311776]
22. Chung AE, Jensen RE, Basch EM. Leveraging emerging technologies and the “Internet of Things” to improve the quality of cancer care. J Oncol Pract 2016 Oct;12(10):863-866. [doi: 10.1200/JOP.2016.015784] [Medline: 27624946]
23. Piwek L, Ellis DA, Andrews S, Joinson A. The rise of consumer health wearables: promises and barriers. PLoS Med 2016 Feb 2;13(2):e1001953. [doi: 10.1371/journal.pmed.1001953] [Medline: 26836780]
24. Cox SM, Lane A, Volchenboum SL. Use of wearable, mobile, and sensor technology in cancer clinical trials. JCO Clin Cancer Inform 2018 Dec;2:1-11. [doi: 10.1200/CCI.17.00147] [Medline: 30652590]
25. Basch E, Deal AM, Kris MG, et al. Symptom monitoring with patient-reported outcomes during routine cancer treatment: a randomized controlled trial. J Clin Oncol 2016 Feb 20;34(6):557-565. [doi: 10.1200/JCO.2015.63.0830] [Medline: 26644527]
26. Komarzynski S, Wreglesworth NI, Griffiths D, et al. Embracing change: learnings from implementing multidimensional digital remote monitoring in oncology patients at a district general hospital during the COVID-19 pandemic. JCO Clin Cancer Inform 2021 Feb;5:216-220. [doi: 10.1200/CCI.20.00136] [Medline: 33606562]
27. Lu TC, Fu CM, Ma MHM, Fang CC, Turner AM. Healthcare applications of smart watches: a systematic review. Appl Clin Inform 2016 Sep 14;7(3):850-869. [doi: 10.4338/ACI-2016-03-R-0042] [Medline: 27623763]
28. Reeder B, David A. Health at hand: a systematic review of smart watch uses for health and wellness. J Biomed Inform 2016 Oct;63:269-276. [doi: 10.1016/j.jbi.2016.09.001] [Medline: 27612974]
29. Low CA, Dey AK, Ferreira D, et al. Estimation of symptom severity during chemotherapy from passively sensed data: exploratory study. J Med Internet Res 2017 Dec 19;19(12):e420. [doi: 10.2196/jmir.9046] [Medline: 29258977]
30. Beauchamp UL, Pappot H, Holländer-Mieritz C. The use of wearables in clinical trials during cancer treatment: systematic review. JMIR Mhealth Uhealth 2020 Nov 11;8(11):e22006. [doi: 10.2196/22006] [Medline: 33174852]
31. Wesley KM, Fizur PJ. A review of mobile applications to help adolescent and young adult cancer patients. Adolesc Health Med Ther 2015 Aug 18;6:141-148. [doi: 10.2147/AHMT.S69209] [Medline: 26316835]
32. Wolpin S, Stewart M. A deliberate and rigorous approach to development of patient-centered technologies. Semin Oncol Nurs 2011 Aug;27(3):183-191. [doi: 10.1016/j.soncn.2011.04.003] [Medline: 21783009]
33. Majeed-Ariss R, Baildam E, Campbell M, et al. Apps and adolescents: a systematic review of adolescents' use of mobile phone and tablet apps that support personal management of their chronic or long-term physical conditions. J Med Internet Res 2015 Dec 23;17(12):e287. [doi: 10.2196/jmir.5043] [Medline: 26701961]
34. Lenhart A. Teens, social media & technology overview 2015. Pew Research Center. 2015 Apr 9. URL: https://www.pewresearch.org/internet/2015/04/09/teens-social-media-technology-2015/ [accessed 2024-04-04]
35. Elsbernd A, Hjerming M, Visler C, et al. Cocreated smartphone app to improve the quality of life of adolescents and young adults with cancer (Kræftværket): protocol for a quantitative and qualitative evaluation. JMIR Res Protoc 2018 May 10;7(5):e10098. [doi: 10.2196/10098] [Medline: 29748162]
36. Steen-Olsen EB, Pappot H, Green A, Langberg H, Holländer-Mieritz C. Feasibility of monitoring patients who have cancer with a smart T-shirt: protocol for the OncoSmartShirt study. JMIR Res Protoc 2022 Oct 3;11(10):e37626. [doi: 10.2196/37626] [Medline: 36190744]
37. Keesense. Chronolife. URL: https://www.chronolife.net/product/keesense/ [accessed 2024-04-04]
38. Bergmann JHM, Chandaria V, McGregor A. Wearable and implantable sensors: the patient’s perspective. Sensors (Basel) 2012 Dec 5;12(12):16695-16709. [doi: 10.3390/s121216695] [Medline: 23443394]
39. Papi E, Belsi A, McGregor AH. A knee monitoring device and the preferences of patients living with osteoarthritis: a qualitative study. BMJ Open 2015 Sep 7;5(9):e007980. [doi: 10.1136/bmjopen-2015-007980] [Medline: 26346873]
40. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
41. Holländer-Mieritz C, Steen-Olsen EB, Kristensen CA, Johansen C, Vogelius IR, Pappot H. Feasibility of using wearables for home monitoring during radiotherapy for head and neck cancer—results from the OncoWatch 1.0 study. Cancers (Basel) 2023 Jan 9;15(2):422. [doi: 10.3390/cancers15020422] [Medline: 36672370]
42. Huang Y, Upadhyay U, Dhar E, Kuo LJ, Syed-Abdul S. A scoping review to assess adherence to and clinical outcomes of wearable devices in the cancer population. Cancers (Basel) 2022 Sep 13;14(18):4437. [doi: 10.3390/cancers14184437] [Medline: 36139602]

43. Anderson M, Perrin A. Technology use among seniors. Pew Research Center. 2017 May 17. URL: https://www.pewresearch.org/internet/2017/05/17/technology-use-among-seniors/ [accessed 2024-04-04]
44. Voutilainen A, Pitkäaho T, Kvist T, Vehviläinen-Julkunen K. How to ask about patient satisfaction? the visual analogue scale is less vulnerable to confounding factors and ceiling effect than a symmetric Likert scale. J Adv Nurs 2016 Apr;72(4):946-957. [doi: 10.1111/jan.12875] [Medline: 26689434]
45. Fukuma N, Hasumi E, Fujiu K, et al. Feasibility of a T-shirt-type wearable electrocardiography monitor for detection of covert atrial fibrillation in young healthy adults. Sci Rep 2019 Aug 13;9(1):11768. [doi: 10.1038/s41598-019-48267-1] [Medline: 31409855]
46. Cho PJ, Yi J, Ho E, et al. Demographic imbalances resulting from the bring-your-own-device study design. JMIR Mhealth Uhealth 2022 Apr 8;10(4):e29510. [doi: 10.2196/29510] [Medline: 34913871]

Abbreviations

AYA: adolescent and young adult
ECG: electrocardiogram
QoL: quality of life

Edited by L Buis; submitted 06.07.23; peer-reviewed by A Khudair, F Keusch, KL Du; revised version received 25.01.24; accepted 01.02.24; published 01.05.24.

Please cite as:
Steen-Olsen EB, Pappot H, Hjerming M, Hanghoej S, Holländer-Mieritz C
Monitoring Adolescent and Young Adult Patients With Cancer via a Smart T-Shirt: Prospective, Single-Cohort, Mixed Methods Feasibility Study (OncoSmartShirt Study)
JMIR Mhealth Uhealth 2024;12:e50620
URL: https://mhealth.jmir.org/2024/1/e50620
doi:10.2196/50620

Impact of Remote Blood Pressure Monitoring Device Connectivity on Engagement Among Pregnant Individuals Enrolled in the Delfina Care Platform: Observational Study

Mia Charifson[1,2], BA, MA; Timothy Wen[1,3], MPH, MD; Bonnie Zell[1], MD; Priyanka Vaidya[1], MBA; Cynthia I Rios[1,4], MD; C Funsho Fagbohun[1,4], MD, PhD; Isabel Fulcher[1], PhD

[1]Delfina Care Inc, San Francisco, CA, United States

[2]Department of Population Health, New York University, New York, NY, United States

[3]Department of Obstetrics, Gynecology, and Reproductive Sciences, University of California San Francisco, San Francisco, CA, United States

[4]Center for Women's Health, Richmond, TX, United States

Corresponding Author:
Mia Charifson, BA, MA
Delfina Care Inc, , San Francisco, CA, , United States

Abstract

User engagement with remote blood pressure monitoring during pregnancy is critical to optimize the associated benefits of blood pressure control and early detection of hypertensive disorders of pregnancy. In our study population of pregnant individuals, we found that *connected* blood pressure cuffs, which automatically sync measures to a monitoring platform or health record, increase engagement (2.13 [95% CI 1.36 - 3.35] times more measures per day) with remote blood pressure monitoring compared to *unconnected* cuffs that require manual entry of measures.

(JMIR Mhealth Uhealth 2024;12:e55617) doi:10.2196/55617

KEYWORDS

blood pressure; hypertension; remote patient monitoring; pregnancy; digital health; remote monitoring; user engagement; users; connected; unconnected; comparison; patient engagement; prospective pregnancy cohort; device; devices; female; females; women; logistic regression; Poisson

Introduction

Hypertensive disorders of pregnancy affect 5% - 10% of pregnancies and increase the risk of adverse pregnancy and postpartum outcomes [1,2]. Typically, management of these disorders involves blood pressure (BP) monitoring and initiation of antihypertensive therapy. Remote BP monitoring (RBPM) enables at-home BP monitoring to inform clinical decision-making in a timely manner [2-4]. Previous studies have shown that RBPM facilitates earlier detection of elevated BP and reduces prenatal hospitalizations and clinic visits [2,3]. Despite studies establishing feasibility and patient acceptability of RBPM [4-9], best practices for integrating RBPM within prenatal care have not yet been established. One open question is whether connected BP cuffs, which automatically sync measures to a health record, improve data quality and frequency over standard BP cuffs, which require users to manually enter measures [3-8]. The goal of this study was to compare user engagement with RBPM between *connected* and *unconnected* BP devices among users of a pregnancy care platform.

Methods

Study Population

The study population consisted of pregnant individuals enrolled in Delfina Care [10], a comprehensive pregnancy care platform, at a community practice in Texas, USA, between January and July 2023. Initially, these users were provided with unconnected BP devices at their provider's discretion, with a recommendation to record 2 BP measures a day per internal expert clinical consensus. Connected devices were introduced in April 2023 as part of a quality improvement initiative. Differences in user experiences between the connected and unconnected device groups are further described in Multimedia Appendix 1.

The outcome of interest was user engagement, defined as the number of daily BP measures taken, and as a binary indicator of completing the daily recommended BP measures (ie, 2 distinct BP measures per day). Analogous engagement outcomes at the weekly level were also reported. To avoid inflated differences due to repeated measures from connected device users, we considered multiple entries within 1 hour as a single measurement (Multimedia Appendix 2). The exposure of interest was receiving a connected versus unconnected device. Clinical

and demographic characteristics were collated from user-reported questionnaire data and electronic health records.

Statistical Analysis

Poisson and logistic regressions were fit for the daily number of BP measures and ≥2 daily BP measures, respectively. Both models controlled for relevant confounders (ie, age, parity, weeks since enrollment, primary clinic, and preferred language) and included a random effect for users across enrollment days.

Ethical Considerations

The study team received an institutional review board exemption waiver of HIPAA (Health Insurance Portability and Accountability Act) authorization on August 22, 2022, from WIRB-Copernicus Group (WCG) Institutional Review Board (protocol number 202208 - 001).

Results

During the study period, 164 users with BP cuffs were enrolled in the Delfina Care platform. Restricting to those with covariate data, the analytic sample consisted of 163 users (97 unconnected device users and 66 connected device users). Users with connected devices had more mean BP measure entries per day (0.51 vs 0.32) and a higher proportion completed the recommended ≥2 daily BP measures (12% vs 7%) compared to users with unconnected devices (Table 1). At the weekly level, the mean entries per week (3.37 vs 2.07) and the proportion of users who completed the recommended ≥1 weekly BP measure (63% vs 47%) were higher for connected users than for unconnected device users (Table 1).

Adjusting for confounders, users with connected devices had 2.13 (95% CI 1.36 - 3.35) times more measures per day and 5.62 (95% CI 2.28 - 13.83) times the odds of meeting the recommendation of ≥2 daily BP measures than unconnected device users (Table 2).

Table . Study population characteristics.

Characteristics	Overall (n=163)	Unconnected device users (n=97)	Connected device users (n=66)	*P* value
Age in years, mean (SD)	28.73 (6.02)	28.71 (5.61)	28.76 (6.61)	.96
Language, n (%)				.32
English	139 (85.3)	80 (82.5)	59 (89.4)	
Spanish	24 (14.7)	17 (17.5)	7 (10.6)	
Parity, mean (SD)	1.37 (1.44)	1.36 (1.32)	1.38 (1.60)	.99
Primary clinic, n (%)				.07
A	35 (21.5)	20 (20.6)	15 (22.7)	
B	71 (43.6)	49 (50.5)	22 (33.3)	
C	57 (35.0)	28 (28.9)	29 (43.9)	
User engagement[a], mean (SD)				
Weekly entries	2.59 (3.02)	2.07 (2.78)	3.37 (3.20)	.007
Daily entries	0.40 (0.46)	0.32 (0.42)	0.51 (0.49)	.008
Proportion ≥1 weekly entries	0.53 (0.35)	0.47 (0.35)	0.63 (0.33)	.003
Proportion ≥2 daily entries	0.09 (0.17)	0.07 (0.15)	0.12 (0.19)	.048

[a]User engagement metrics were first averaged within a user and then averaged across users for each device type and overall.

Table . Effect estimates and 95% CIs by outcome model.

	Number of BP measures per day		Completed ≥2 BP measures per day	
Predictors	Incidence rate ratio	95% CI	Odds ratio	95% CI
Connected versus unconnected device	2.13	1.36 - 3.35	5.62	2.28 - 13.83
Age (years)	1.06	1.02 - 1.11	1.11	1.03 - 1.20
Weeks since enrollment	0.97	0.97 - 0.98	1.00	0.98 - 1.01
English vs Spanish language	1.25	0.66 - 2.35	0.72	0.22 - 2.38
Parity (number of live births)	0.88	0.74 - 1.05	0.79	0.57 - 1.10
Clinic B vs A	1.44	0.79 - 2.61	4.18	1.25 - 14.03
Clinic C vs A	1.54	0.83 - 2.86	4.06	1.16 - 14.20

Discussion

We observed user engagement with RBPM was significantly higher among those with connected devices than those with unconnected devices. Previous studies among pregnant individuals have shown that the recommended frequency of BP measures ranges from several times daily to once weekly [4,7,8,11]. In our cohort, the proportion of users meeting the twice-daily recommendation was low, but the majority of users completed the once-weekly entry at least, and connected device users still had higher utilization than unconnected device users (63% vs 47%; *P*=.003). Compared to in-clinic BP measures, all users had a higher average number of weekly readings than the in-clinic average (2.59 vs 0.50 BP readings/week). Our study corroborated the feasibility of at-home RBPM during pregnancy and highlights the potential advantages of device connectivity on user engagement.

Limitations

Our findings are limited by the lack of true randomization to device types, which we addressed by controlling for potential confounders. We also addressed potential time-related confounding via a sensitivity analysis with no change in findings (Multimedia Appendix 2).

Conclusion

This study highlights how connected BP devices can improve patient engagement to RBPM during the prenatal period. Other aspects of RBPM, such as recommended frequency and patient education, should be further investigated to ensure users are able to successfully engage with monitoring technologies.

Acknowledgments

This study was supported by grant funding from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) under grant number R43HD114360.

Conflicts of Interest

MC reports serving as a data science consultant for Delfina, Inc. TW reports serving as the associate chief medical officer for Delfina, Inc. BZ reports serving as the chief medical officer for Delfina, Inc. PV reports serving as the chief product officer for Delfina, Inc. IF reports serving as the chief scientific officer for Delfina, Inc

Multimedia Appendix 1
Additional information on Delfina Care.
[DOCX File, 344 KB - mhealth_v12i1e55617_app1.docx]

Multimedia Appendix 2
Additional analyses.
[DOCX File, 15 KB - mhealth_v12i1e55617_app2.docx]

References

1. Hutcheon JA, Lisonkova S, Joseph KS. Epidemiology of pre-eclampsia and the other hypertensive disorders of pregnancy. Best Pract Res Clin Obstet Gynaecol 2011 Aug;25(4):391-403. [doi: 10.1016/j.bpobgyn.2011.01.006] [Medline: 21333604]
2. Nathan HL, Duhig K, Hezelgrave NL, Chappell LC, Shennan AH. Blood pressure measurement in pregnancy. Obstet Gynaecol 2015 Apr;17(2):91-98. [doi: 10.1111/tog.12173]

3. Yeh PT, Rhee DK, Kennedy CE, et al. Self-monitoring of blood pressure among women with hypertensive disorders of pregnancy: a systematic review. BMC Pregnancy Childbirth 2022 May 31;22(1):454. [doi: 10.1186/s12884-022-04751-7] [Medline: 35641913]
4. Aquino M, Munce S, Griffith J, Pakosh M, Munnery M, Seto E. Exploring the use of telemonitoring for patients at high risk for hypertensive disorders of pregnancy in the antepartum and postpartum periods: scoping review. JMIR Mhealth Uhealth 2020 Apr 17;8(4):e15095. [doi: 10.2196/15095] [Medline: 32301744]
5. van den Heuvel JFM, Kariman SS, van Solinge WW, Franx A, Lely AT, Bekker MN. SAFE@HOME - feasibility study of a telemonitoring platform combining blood pressure and preeclampsia symptoms in pregnancy care. Eur J Obstet Gynecol Reprod Biol 2019 Sep;240:226-231. [doi: 10.1016/j.ejogrb.2019.07.012] [Medline: 31330428]
6. Marko KI, Krapf JM, Meltzer AC, et al. Testing the feasibility of remote patient monitoring in prenatal care using a mobile app and connected devices: a prospective observational trial. JMIR Res Protoc 2016 Nov 18;5(4):e200. [doi: 10.2196/resprot.6167] [Medline: 27864167]
7. Tucker KL, Taylor KS, Crawford C, et al. Blood pressure self-monitoring in pregnancy: examining feasibility in a prospective cohort study. BMC Pregnancy Childbirth 2017 Dec 28;17(1):442. [doi: 10.1186/s12884-017-1605-0] [Medline: 29284456]
8. Runkle JD, Sugg MM, McCrory S, Coulson CC. Examining the feasibility of smart blood pressure home monitoring: advancing remote prenatal care in rural Appalachia. Telemed Rep 2021;2(1):125-134. [doi: 10.1089/tmr.2020.0021] [Medline: 35720741]
9. Khalil A, Perry H, Lanssens D, Gyselaers W. Telemonitoring for hypertensive disease in pregnancy. Expert Rev Med Devices 2019 Aug;16(8):653-661. [doi: 10.1080/17434440.2019.1640116] [Medline: 31282755]
10. Delfina - intelligent pregnancy care for all. URL: https://www.delfina.com/ [accessed 2024-06-28]
11. Chappell LC, Tucker KL, Galal U, et al. Effect of self-monitoring of blood pressure on blood pressure control in pregnant individuals with chronic or gestational hypertension: the BUMP 2 randomized clinical trial. JAMA 2022 May 3;327(17):1666-1678. [doi: 10.1001/jama.2022.4726] [Medline: 35503345]

Abbreviations

BP: blood pressure
HIPAA: Health Insurance Portability and Accountability Act
NICHD: National Institute of Child Health and Human Development
RBPM: remote blood pressure monitoring
WCG: WIRB-Copernicus Group

Edited by L Buis; submitted 19.12.23; peer-reviewed by B Senst, CP Lau; revised version received 03.05.24; accepted 12.06.24; published 12.07.24.

Please cite as:
Charifson M, Wen T, Zell B, Vaidya P, Rios CI, Fagbohun CF, Fulcher I
Impact of Remote Blood Pressure Monitoring Device Connectivity on Engagement Among Pregnant Individuals Enrolled in the Delfina Care Platform: Observational Study
JMIR Mhealth Uhealth 2024;12:e55617
URL: https://mhealth.jmir.org/2024/1/e55617
doi:10.2196/55617

Commentary

The Evaluation of Generative AI Should Include Repetition to Assess Stability

Lingxuan Zhu[1*], MD; Weiming Mou[2*], MD; Chenglin Hong[1], MD; Tao Yang[3], MD; Yancheng Lai[1], MD; Chang Qi[4], MEng; Anqi Lin[1], MD; Jian Zhang[1], MD; Peng Luo[1], MD

[1]Department of Oncology, Zhujiang Hospital, Southern Medical University, Guangzhou, China

[2]Department of Urology, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

[3]Department of Medical Oncology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

[4]Institute of Logic and Computation, TU Wien, Austria

[*]these authors contributed equally

Corresponding Author:
Peng Luo, MD
Department of Oncology
Zhujiang Hospital
Southern Medical University
253 Industrial Avenue
Guangzhou
China
Phone: 86 020 61643888
Email: luopeng@smu.edu.cn

Related Article:

Comment on: https://mhealth.jmir.org/2024/1/e51526/

Abstract

The increasing interest in the potential applications of generative artificial intelligence (AI) models like ChatGPT in health care has prompted numerous studies to explore its performance in various medical contexts. However, evaluating ChatGPT poses unique challenges due to the inherent randomness in its responses. Unlike traditional AI models, ChatGPT generates different responses for the same input, making it imperative to assess its stability through repetition. This commentary highlights the importance of including repetition in the evaluation of ChatGPT to ensure the reliability of conclusions drawn from its performance. Similar to biological experiments, which often require multiple repetitions for validity, we argue that assessing generative AI models like ChatGPT demands a similar approach. Failure to acknowledge the impact of repetition can lead to biased conclusions and undermine the credibility of research findings. We urge researchers to incorporate appropriate repetition in their studies from the outset and transparently report their methods to enhance the robustness and reproducibility of findings in this rapidly evolving field.

(JMIR Mhealth Uhealth 2024;12:e57978) doi:10.2196/57978

KEYWORDS

large language model; generative AI; ChatGPT; artificial intelligence; health care

Since OpenAI released ChatGPT-3.5, there has been a growing interest within the medical community regarding the prospective applications of this general pretrained model in health care [1-7]. Using ChatGPT as a search keyword in the PubMed database, the results show that 2075 papers discussing ChatGPT were published in 2023. As the leading journal in the field of digital medicine, JMIR Publications Inc published a total of 115 papers related to ChatGPT in the year 2023. It should be noted that this is a quick and simple search that may not comprehensively capture all relevant articles, but it provides a general reflection of the growing interest and research on ChatGPT in the medical field. For example, Gilson et al [8] explored the performance of ChatGPT on the United States Medical Licensing Examination (USMLE) step 1 and step 2 exams, discovering that ChatGPT's performance exceeded the passing score for third-year medical students in step 1. More studies are exploring

ChatGPT's performance on other medical exams, such as the Japanese and German Medical Licensing Examinations [9,10], the Otolaryngology-Head and Neck Surgery Certification Examinations [11], and the UK Standardized Admission Tests [12]. Beyond examinations, many articles have discussed the potential applications of ChatGPT in medicine from various perspectives. Shao et al [13] examined the suitability of using ChatGPT for perioperative patient education in thoracic surgery within English and Chinese contexts. Cheng et al [14] investigated whether ChatGPT could be used to generate summaries for medical research, and Hsu et al [15] evaluated whether ChatGPT could correctly answer basic medication consultation questions. However, we would like to point out that as a relatively new technology, there are some differences in evaluating the potential application of generative artificial intelligence (AI) like ChatGPT in health care that require additional attention from researchers.

The most significant difference affecting the evaluation of ChatGPT compared to traditional AI models known to people is the randomness inherent in the responses generated by ChatGPT. Common perception holds that for a given input, an AI model should produce the same output consistently each time. However, for natural language models like ChatGPT, this is not the case. ChatGPT generates a response by predicting the next most likely word, followed by each subsequent word. The process of generating responses involves a certain degree of randomness. If you access ChatGPT using the application programming interface, you can also control the degree of randomness in the generated responses with the temperature parameter. Even with the same input, the responses provided by ChatGPT will not be the same, and sometimes may even be completely contradictory. Therefore, when evaluating ChatGPT's performance, it is necessary to generate multiple responses to the same input and assess these responses collectively to explore ChatGPT's performance accurately; otherwise, there is a high likelihood of drawing biased conclusions. For example, as one of the earliest studies published, Sarraju et al [4] asked the same question three times and assessed whether the three responses given by ChatGPT to the same question were consistent. As OpenAI made the ChatGPT application programming interface accessible, it became feasible to ask the same question many more times. In a recent study investigating whether ChatGPT's peer-review conclusions are influenced by the reputation of the author's institution, von Wedel et al [16] conducted 250 repeated experiments for each question to mitigate the effects of ChatGPT's randomness. However, not all researchers have recognized this aspect. For instance, in a study where ChatGPT was asked to answer the American Heart Association Basic Life Support and Advanced Cardiovascular Life Support exams, they found that ChatGPT could not pass either examination [17]. However, that study only asked the question once without repeating, which means that the randomness of ChatGPT could have had an impact on the experiment, affecting the reliability of the conclusions. In another improved study, researchers acknowledged the impact of ChatGPT's randomness, asking each question three times. Compared to earlier results, ChatGPT's performance in this study significantly improved, and it could pass the Basic Life Support exam [18], further underscoring the importance of repetitions. Therefore, it is inappropriate to evaluate ChatGPT's performance based on a single response if one aims to draw rigorous, scientifically meaningful conclusions. Just as biological experiments typically require three repetitions for validity, without repetition, it becomes challenging to determine whether the observed phenomenon is an inherent characteristic of the model or merely a random occurrence. Additionally, for models intended for clinical practice applications, whether for patient education, diagnosis, or support in clinical documentation writing, we hope that ChatGPT can always provide correct and harmless responses. Repetition also allows us to evaluate the model's stability and further assess its application value. However, we noticed that many recent manuscripts we reviewed were not aware of this, thus affecting the reliability of the conclusions.

Therefore, in research on the application of generative AI like ChatGPT in health care, appropriate repetition should be included to comprehensively evaluate the model's performance by assessing the stability of the model in the task set by the author. This should be considered from the beginning of the research. Since models like ChatGPT will continue to be upgraded, if the authors only realize the need for repetition when revising the manuscript, there will be a considerable time gap between the authors' supplementary analysis and the original analysis. The model has likely been upgraded during this period, introducing new uncertainties into the research. Alternatively, the authors need to completely redo the analysis from scratch during the manuscript revision process, wasting time and effort. Therefore, we hope that future researchers will recognize the necessity of repeated experiments from the start and report in the manuscript how the repetition was carried out in the study [19].

Conflicts of Interest

None declared.

References

1. Grünebaum A, Chervenak J, Pollet SL, Katz A, Chervenak FA. The exciting potential for ChatGPT in obstetrics and gynecology. Am J Obstet Gynecol 2023 Jun;228(6):696-705. [doi: 10.1016/j.ajog.2023.03.009] [Medline: 36924907]
2. Howard A, Hope W, Gerada A. ChatGPT and antimicrobial advice: the end of the consulting infection doctor? Lancet Infect Dis 2023 Apr;23(4):405-406. [doi: 10.1016/S1473-3099(23)00113-5] [Medline: 36822213]
3. Zhu L, Mou W, Luo P. Potential of large language models as tools against medical disinformation. JAMA Intern Med 2024 Apr 01;184(4):450. [doi: 10.1001/jamainternmed.2024.0020] [Medline: 38407861]

4. Sarraju A, Bruemmer D, Van Iterson E, Cho L, Rodriguez F, Laffin L. Appropriateness of cardiovascular disease prevention recommendations obtained from a popular online chat-based artificial intelligence model. JAMA 2023 Mar 14;329(10):842-844 [FREE Full text] [doi: 10.1001/jama.2023.1044] [Medline: 36735264]
5. Zhu L, Mou W, Chen R. Can the ChatGPT and other large language models with internet-connected database solve the questions and concerns of patient with prostate cancer and help democratize medical knowledge? J Transl Med 2023 Apr 19;21(1):269 [FREE Full text] [doi: 10.1186/s12967-023-04123-5] [Medline: 37076876]
6. Ali SR, Dobbs TD, Hutchings HA, Whitaker IS. Using ChatGPT to write patient clinic letters. Lancet Digit Health 2023 Apr;5(4):e179-e181 [FREE Full text] [doi: 10.1016/S2589-7500(23)00048-1] [Medline: 36894409]
7. Patel SB, Lam K. ChatGPT: the future of discharge summaries? Lancet Digit Health 2023 Mar;5(3):e107-e108 [FREE Full text] [doi: 10.1016/S2589-7500(23)00021-3] [Medline: 36754724]
8. Gilson A, Safranek CW, Huang T, Socrates V, Chi L, Taylor RA, et al. How does ChatGPT perform on the United States Medical Licensing Examination (USMLE)? The implications of large language models for medical education and knowledge assessment. JMIR Med Educ 2023 Feb 08;9:e45312 [FREE Full text] [doi: 10.2196/45312] [Medline: 36753318]
9. Meyer A, Riese J, Streichert T. Comparison of the performance of GPT-3.5 and GPT-4 with that of medical students on the written German Medical Licensing Examination: observational study. JMIR Med Educ 2024 Feb 08;10:e50965 [FREE Full text] [doi: 10.2196/50965] [Medline: 38329802]
10. Yanagita Y, Yokokawa D, Uchida S, Tawara J, Ikusaka M. Accuracy of ChatGPT on medical questions in the National Medical Licensing Examination in Japan: evaluation study. JMIR Form Res 2023 Oct 13;7:e48023 [FREE Full text] [doi: 10.2196/48023] [Medline: 37831496]
11. Long C, Lowe K, Zhang J, Santos AD, Alanazi A, O'Brien D, et al. A novel evaluation model for assessing ChatGPT on Otolaryngology-Head and Neck Surgery Certification Examinations: performance study. JMIR Med Educ 2024 Jan 16;10:e49970 [FREE Full text] [doi: 10.2196/49970] [Medline: 38227351]
12. Giannos P, Delardas O. Performance of ChatGPT on UK standardized admission tests: insights from the BMAT, TMUA, LNAT, and TSA examinations. JMIR Med Educ 2023 Apr 26;9:e47737 [FREE Full text] [doi: 10.2196/47737] [Medline: 37099373]
13. Shao C, Li H, Liu X, Li C, Yang L, Zhang Y, et al. Appropriateness and comprehensiveness of using ChatGPT for perioperative patient education in thoracic surgery in different language contexts: survey study. Interact J Med Res 2023 Aug 14;12:e46900 [FREE Full text] [doi: 10.2196/46900] [Medline: 37578819]
14. Cheng S, Tsai S, Bai Y, Ko C, Hsu C, Yang F, et al. Comparisons of quality, correctness, and similarity between ChatGPT-generated and human-written abstracts for basic research: cross-sectional study. J Med Internet Res 2023 Dec 25;25:e51229 [FREE Full text] [doi: 10.2196/51229] [Medline: 38145486]
15. Hsu H, Hsu K, Hou S, Wu C, Hsieh Y, Cheng Y. Examining real-world medication consultations and drug-herb interactions: ChatGPT performance evaluation. JMIR Med Educ 2023 Aug 21;9:e48433 [FREE Full text] [doi: 10.2196/48433] [Medline: 37561097]
16. von Wedel D, Schmitt RA, Thiele M, Leuner R, Shay D, Redaelli S, et al. Affiliation bias in peer review of abstracts by a large language model. JAMA 2024 Jan 16;331(3):252-253. [doi: 10.1001/jama.2023.24641] [Medline: 38150261]
17. Fijačko N, Gosak L, Štiglic G, Picard CT, John Douma M. Can ChatGPT pass the life support exams without entering the American Heart Association course? Resuscitation 2023 Apr;185:109732. [doi: 10.1016/j.resuscitation.2023.109732] [Medline: 36775020]
18. Zhu L, Mou W, Yang T, Chen R. ChatGPT can pass the AHA exams: open-ended questions outperform multiple-choice format. Resuscitation 2023 Jul;188:109783. [doi: 10.1016/j.resuscitation.2023.109783] [Medline: 37349064]
19. Chen J, Zhu L, Mou W, Liu Z, Cheng Q, Lin A, et al. STAGER checklist: Standardized Testing and Assessment Guidelines for Evaluating Generative AI Reliability. arXiv. Preprint posted online on Decemeber 8, 2023 2024. [doi: 10.48550/arXiv.2312.10074]

Abbreviations

AI: artificial intelligence
USMLE: United States Medical Licensing Examination

Edited by L Buis; submitted 01.03.24; this is a non–peer-reviewed article;accepted 30.04.24; published 06.05.24.

Please cite as:
Zhu L, Mou W, Hong C, Yang T, Lai Y, Qi C, Lin A, Zhang J, Luo P
The Evaluation of Generative AI Should Include Repetition to Assess Stability
JMIR Mhealth Uhealth 2024;12:e57978
URL: https://mhealth.jmir.org/2024/1/e57978
doi:10.2196/57978
PMID:38688841

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org